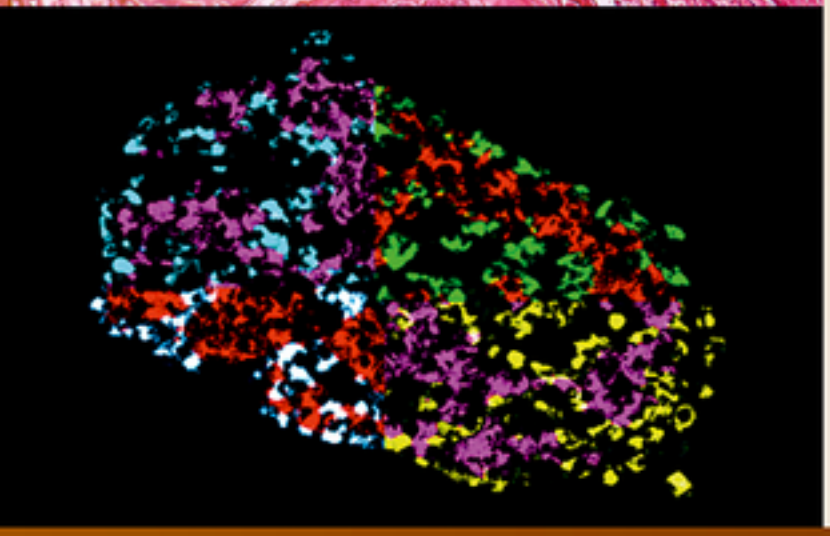
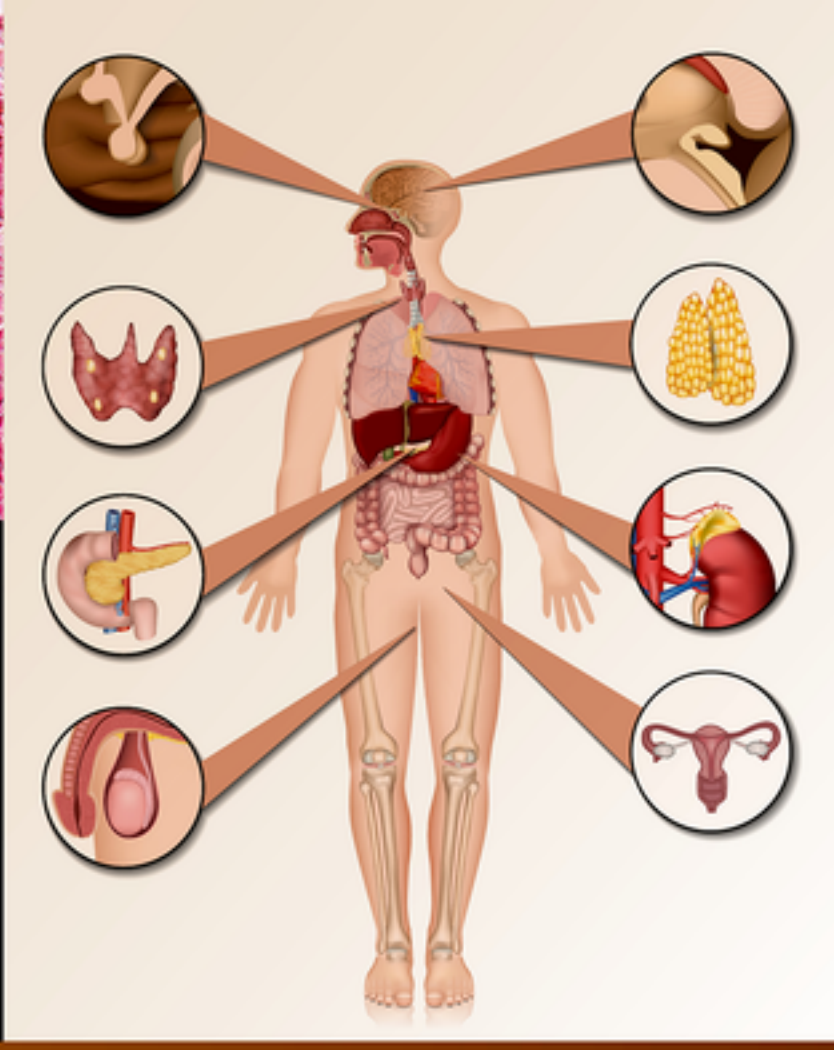




Oxford Textbook of Endocrinology & Diabetes

THIRD EDITION

EDITED BY
John Wass
Wiebke Arlt
Robert Semple

DIABETES SECTION EDITED BY
James Shaw
Desmond Johnston

Oxford Textbook of

Endocrinology and Diabetes

Oxford Textbook of

Endocrinology and Diabetes

THIRD EDITION

Volume 1: Sections 1–6

EDITED BY

John A.H. Wass

Professor of Endocrinology, Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Wiebke Arlt

William Withering Chair of Medicine, Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

Robert K. Semple

Professor of Translational Molecular Medicine, University of Edinburgh Centre for Cardiovascular Science, Queen's Medical Research Institute, Edinburgh, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2002

Second Edition published in 2011

Third Edition published in 2022

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2021951488

Volume 1 ISBN 978–0–19–880232–7
Volume 2 ISBN 978–0–19–887018–0
Pack ISBN 978–0–19–887019–7

DOI: 10.1093/med/9780198802327.001.0001

Printed in Great Britain by
Bell & Bain Ltd., Glasgow

Preface

It is now over a hundred and seventy years since Arnold Berthold demonstrated that endocrine glands convey their effects via the bloodstream and a hundred and ten years since the word 'hormone' was coined by Ernest Starling. Amazingly, it is just under a hundred years since one of the nascent specialty's most spectacular Nobel Prize-winning triumphs, the purification and use of insulin to transform Type I diabetes from a rapidly fatal wasting disease of childhood into a manageable chronic condition. Similarly, seventy years ago the Nobel Prize recognised the life-saving impact of glucocorticoids both for patients with adrenal insufficiency and as immunosuppressive therapy for chronic inflammatory disease. Since then endocrinology has continued to ride the crest of a wave of technological advances, most prominent recently in areas such as molecular genetics, imaging, immunotherapy and rational drug design.

The first edition of the *Oxford Textbook of Endocrinology & Diabetes* was published in 2002. The landscape of endocrine disease has evolved enormously since then, driven largely by burgeoning obesity and by population ageing. The spectrum of endocrinology has been further broadened by the increasing use of immunomodulatory therapies with endocrine complications, evolving patterns of recreational drug use impacting endocrine systems, and widespread exposure to endocrine disrupting chemicals. The therapeutic armamentarium of the endocrinologist has also expanded at pace, not only through development of novel small molecule and biological therapies, but also through step changes in e-technology and their application to chronic disease management and "precision" medicine.

We now comprehensively update the second edition of the *Textbook* published in 2012. The expanded first part of the book is devoted to overview chapters focusing on principles underpinning the science and practice of endocrinology. Elsewhere, other sections have also been enlarged to capture exciting developments in subspeciality practice, with separate sections now devoted to endocrine disease in pregnancy and transgender endocrinology.

The diabetes section has been extensively reorganised to reflect rapid advances in understanding the molecular pathogenesis of diabetes, step changes in the sophistication of technologies used for metabolic monitoring and insulin delivery, and innovations in immunotherapy, behaviour-focused and cell-based therapies. New chapters are devoted to current urgent responses to diabetes as a public health and economic challenge.

Fascination in the science underlying endocrinology continues to endure, and clinical endocrinologists have an ever more sophisticated ability to transform the length and quality of life of those with hormone-related diseases. This book aims both to illuminate the emerging scientific concepts that underlie endocrinology, and to provide an accessible and authoritative account of cutting edge endocrine practice.

We are very grateful indeed to our national and international colleagues who have kindly, expertly and cerebrally contributed to all sections. We are proud of this *magnum opus* and thank them sincerely for their significant efforts.

We also thank Claire Brankin, James Oates and Helen Liepman from Oxford University Press who have expertly and efficiently guided us through this whole process.

This book should be available to every endocrinologist, trainee and researcher and we hope that it provides as much enjoyment and intellectual satisfaction in the reading as it did in putting it together.

John A.H. Wass
Wiebke Arlt
Robert K. Semple

Contents

Volume 1

SECTION 4
Parathyroid, Calcium and Bone Metabolism Disorders

Section editor: John Bilezikian

SECTION 5
Adrenal Diseases

Section editor: Wiebke Arlt

Volume 2

SECTION 7
Disorders of Growth, and Development and Transition

Section editor: Peter Clayton

SECTION 8
Female Reproductive Endocrine Disorders

Section editor: Bulent Okan Yildiz

SECTION 9
Endocrine Disorders of Pregnancy

Section editors: Kristien Boelaert and Cathy Williamson

SECTION 10
Male Reproductive Endocrine Disorders

Section editors: Frederick Wu and Mathis Grossmann

Symbols and Abbreviations

AAS	androgenic anabolic steroids
AC	adenylate cyclase (*also* arachnoid cysts)
ACE	American College of Endocrinology
ACE	angiotensin-converting enzyme
ACTH	adrenocorticotrophic hormone
AD	autosomal dominant
ADH	autosomal dominant hypocalcaemia
ADIS	agonist-driven insertional signalling
ADT	androgen deprivation therapy
AF	atrial fibrillation
AGHDA	assessment of GH deficiency in adults
AgRP	agouti-related peptide
AHA	Anterior hypothalamic area
AHO	Albright's hereditary osteodystrophy
AI	alcohol-induced (*also* aromatase inhibitors)
AIDS	acquired immunodeficiency syndrome
AIDS	advanced HIV disease
AIP	aryl hydrocarbon receptor-interacting protein
AIRE	autoimmune regulator
AKAP	A-kinase-anchoring proteins
ALS	acid-labile subunit
ALSPAC	Avon Longitudinal Study of Parents and Children
ANCA	antineutrophil cytoplasmic antibodies
ANF	atrial natriuretic factor
AP	anterior pituitary
APECED	autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy
APT	aggressive pituitary tumours
AR	androgen receptor
ARC	arcuate nucleus
ARE	androgen response elements
ART	assisted reproduction technologies
ASD	autism spectrum disorders
ATA	American Thyroid Association
ATD	antithyroid drug
ATMA	antithyroid microsomal antibody
ATOR	Australian thyroid-associated orbitopathy research
AVP	arginine-vasopressin (*also* antidiuretic hormone)
BMAD	bone mineral apparent density
BMD	bone mineral density
BMI	body mass index
BP	blood pressure
BST	bed nucleus of the stria terminalis
CAH	congenital adrenal hyperplasia
CAIS	complete androgen insensitivity syndrome
CAPTEM	capecitabine and temozolomide
CART	cocaine- and amphetamine-related transcript
CAS	clinical activity score
CAS	Court of Arbitration for Sport
CaSR	calcium-sensing receptor
CBG	corticosteroid-binding globulin
CBG	cortisol-binding globulin
CC	clivus chordoma (*also* clomiphene citrate)
CCCR	calcium-to-creatinine clearance ratio
CCH	central congenital hypothyroidism
CDI	central diabetes insipidus
CDK	cyclin-dependent kinase
CDP	constitutional delay of puberty
CE	calcium excretion
CEA	carcinoembryonic antigen
CG	chorionic gonadotrophin
CGTT	clinical genetics think tank
CH	cavernous haemangiomas (*also* congenital hypopituitarism)
CHH	congenital hypogonadotropic hypogonadism
CIRCI	critical illness-related corticosteroid insufficiency
CKD	chronic kidney disease
CLIP	corticotropin-like intermediate peptide
CLOCK	circadian locomotor output cycles kaput
CME	clathrin-mediated endocytosis
CNS	central nervous system
CNV	copy number variant
COPD	chronic obstructive pulmonary disease
CPAP	continuous positive airway pressure
CPHD	combined pituitary hormone deficiency
CPI	checkpoint inhibitors
CRE	CAMP-response elements
CREB	CAMP-Response Element Binding
CREB	cyclic AMP-response element binding
CRH	corticotropin-releasing hormone
CRP	C-reactive protein
CRPC	castration-resistant prostate cancer
CSF	cerebrospinal fluid
CSHI	continuous subcutaneous hydrocortisone infusion
CSM	cavernous sinus meningiomas
CT	computed tomography
DA	dopamine agonists
DAA	direct-acting antivirals
DALY	disability-adjusted life year
DBD	DNA-binding domain
DCV	dense-cored vesicles
DDT	dichloro-diphenyl-trichloroethane

DEXA	dual-energy X-ray absorptiometry
DHAES	dehydroepiandrosterone sulphate
DI	diabetes insipidus
DMD	Duchenne muscular dystrophy
DMH	dorsomedial hypothalamus
DON	dysthyroid optic neuropathy
DS	diaphragma sellae
DSA	digital subtraction angiography
DSD	disorder of sex development
dSPZ	dorsal subparaventricular zone
DXA	dual-energy X-ray absorptiometry
ED	endocrine disruptor (*also* erectile dysfunction)
EDC	endocrine disrupting compounds
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EMA	European Medicines Agency
EMAS	European Male Ageing Study
ENIGI	European Network for the Investigation of Gender Incongruence
EPP	ectopic posterior pituitary
EQA	external quality assurance
ER	endoplasmic reticulum
ERR	excess relative risk
ESE	European Society of Endocrinology
ESPGHAN	European Society of Paediatric Gastroenterology, Hepatology and Nutrition
ESPR	European Society of Paediatric Research
ESR	erythrocyte sedimentation rate
ESRD	end stage renal disease
ETA	European Thyroid Association
EUGOGO	European Group on Graves' Orbitopathy
FAP	familial amyloidotic polyneuropathy
FDA	US Food and Drug Administration
FDH	familial dysalbuminaemic hyperthyroxinaemia
FEO	food-entrainable oscillator
FHH	familial hypocalciuric hypercalcaemia
FHPP	familial hypokalaemic periodic paralysis
FIH	familial isolated hyperparathyroidism
FISH	fluorescent *in situ* hybridization
FN	follicular neoplasm
FNA	fine-needle aspiration
FNAB	fine-needle aspiration biopsy
FSH	follicle-stimulating hormone
GABA	gamma aminobutyric acid
GAD	glutamic acid decarboxylase
GAH	gender-affirming hormones
GBD	gracile bone dysplasia
GD	gender dysphoria (*also* Graves' disease)
GDNF	glial cell-derived neurotrophic factor
GDR	German Democratic Republic
GFR	glomerular filtration rate
GH	growth hormone
GHBP	growth hormone-binding protein
GHD	growth hormone deficiency
GHR	growth hormone receptor
GHRH	growth hormone-releasing hormone
GHS	growth hormone secretagogues
GK	Gamma Knife
GNAS	guanine nucleotide-binding protein G(S)
GnRH	gonadotropin-releasing hormone
GO	Graves' orbitopathy
GPCR	G-protein-coupled receptor
GREAT	Graves' Recurrent Events After Therapy
GRP	gastrin-releasing peptide
GTR	gross total resection
GTV	gross tumour volume
GWAS	genome-wide association studies
H&E	haematoxylin & eosin
HCG	human chorionic gonadotropin
HDL	high-density lipoprotein
HDR	hypoparathyroidism, deafness, and renal
HGPIN	high-grade prostatic intraepithelial neoplasia
HH	hypogonadotropic hypogonadism (*also* hypothalamic hamartoma)
HIF	hypoxia-inducing factor
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HOMA	homeostasis model assessment
HOS	hypo-osmotic swelling
HP	hypothalamo–pituitary
HPA	hypothalamic–pituitary–adrenal
HPF	high power field
HPT	hypothalamic-pituitary-testicular
HPT	hypothalamic–pituitary–thyroid
HPV	human papilloma virus
HRE	hormone response element
HRT	hormone replacement therapy
HSP	heat shock protein
HT	Hashimoto's thyroiditis
HyOb	hypothalamic obesity
IAD	isolated ACTH deficiency
IAPP	islet amyloid polypeptide
ICSI	intracytoplasmic sperm injection
ICU	intensive care unit
IDD	iodine deficiency disorders
IGF	insulin-like growth factor
IGF-1	insulin-like growth factor 1
IGF1R	insulin-like growth factor 1 receptor
IGHC	integrated growth hormone concentration
IGT	impaired glucose tolerance
IHH	idiopathic hypothalamic hypogonadism
IIH	idiopathic infantile hypercalcaemia
IJV	internal jugular vein
ILP	interstitial laser photocoagulation
IMRT	intensity-modulated radiotherapy
INSR	insulin receptor
IOC	International Olympic Committee
IOM	Institute of Medicine
IPEX	immunodysregulation polyendocrinopathy enteropathy X-linked
IPSC	induced pluripotent stem cells
IQ	intelligence quotient
IR	insulin resistance
IRAE	immune-related adverse effects
IRD	inner ring deiodination

ISE	ion-specific electrode
ISRS	International Stereotactic Radiosurgery Society
ITT	insulin tolerance test
IUI	intrauterine insemination
IVF	*in-vitro* fertilization
JNK	Jun N-terminal kinase
KS	Kallmann syndrome (*also* Klinefelter syndrome)
LAR	long-acting release
LATS	long-acting thyroid stimulator
LBD	ligand-binding domain
LC	liquid chromatography
LC	locus coeruleus
LCH	Langerhans cell histiocytosis
LDL	low-density lipoprotein
LDT	laterodorsal tegmental nucleus
LEPR	leptin receptor
LGBT	lesbian, gay, bisexual, and transgender
LH	luteinizing hormone
LHA	lateral hypothalamic area
LHRH	luteinizing hormone-releasing hormone
LIF	leukaemia inhibitory factor
LINAC	linear accelerator
LN	lymph node
LOH	local osteolytic hypercalcaemia
LOH	loss of heterozygosity
LS	Lugol's solution
LS	Lynch syndrome
LVMI	left ventricular mass index
MACE	major adverse cardiovascular events
MAI	mycobacterium avium intracellulare
MAP	mitogen-activated protein
MAPK	mitogen-activated protein kinase
MC2R	melanocortin-2 receptor
MCH	melanin-concentrating hormone
MCT	monocarboxylate transporter
MDD	major depressive disorder
MDT	multidisciplinary team
MEN	multiple endocrine neoplasia
MES	mineralocorticoid excess syndrome
MFB	medial forebrain bundle
MFS	metastasis-free survival
MHC	major histocompatibility complex
MHRA	Medicines and Healthcare products Regulatory Agency
MI	myocardial infarction
MIF	migration inhibition factor
MIP	minimally invasive parathyroidectomy
MMI	methyl-mercapto-imidazole
MMSE	Mini-Mental State Examination
MNG	multinodular goitre
MPB	male pattern baldness
MPN	medial preoptic nucleus
MPO	medial preoptic area
MR	magnetic resonance
MRI	magnetic resonance imaging
MSH	melanocyte-stimulating hormone
MTC	medullary thyroid cancer
MTC	medullary thyroid carcinoma
MTP	mitochondrial trifunctional protein
NAFLD	non-alcoholic fatty liver disease
NCD	non-communicable diseases
NCRP	National Council of Radiation Protection
NEFA	non-esterified fatty acids
NFPA	non-functioning pituitary adenomas
NGS	next generation sequencing
NHL	non-Hodgkin's lymphoma
NICE	National Institute for Clinical Excellence
NLS	nuclear localization sequence
NNRTI	non-nucleoside reverse transcriptase inhibitor
NOD	non-obese diabetic
NOGG	National Osteoporosis Guideline Group
NRTI	nucleoside reverse transcriptase inhibitors
NTCP	Na-taurocholate cotransporting polypeptide
NTD	N-terminal domain
NTD	N-terminal transcriptional regulation domain
NTE	neuropathy target esterase
NTG	non-toxic goitre
NTI	non-thyroidal illness
NYHA	New York Heart Association
OAR	organ at risk
OATP	organic anion transporting polypeptide
ODS	osmotic demyelination syndrome
OF	orbital fibroblasts
OGTT	oral glucose tolerance test
OHG	optico-hypothalamic gliomas
OI	osteogenesis imperfecta
OIS	oncogene-included senescence
ONH	optic nerve hypoplasia
OPG	optic pathway gliomas
OR	odds ratio
ORD	outer ring deiodination
OS	overall survival
OSA	obstructive sleep apnoea
OXT	oxytocin
PA	pituitary adenomas
PASS	pheochromocytoma of the adrenal gland scaled score
PC	pituitary carcinoma (*also* prohormone convertase)
PCB	polychlorinated biphenyls
PCOS	polycystic ovary syndrome
PCR	polymerase chain reaction
PCS	petroclival chondrosarcomas
PDX	patient-derived xenograft
PEARS	Parathyroid Epidemiology and Audit Research Study
PEG	polyethylene glycol
PET	positron emission tomography
PeVN	periventricular nucleus
PFS	progression-free survival
PH	pleckstrin homology
PI	protease inhibitor
PIA	proliferative inflammatory atrophy
PIC	pars intermedia cysts
PIN	prostatic intraepithelial neoplasia
PLAP	placental alkaline phosphatase
PPT	pedunculopontine tegmental nucleus
PREGO	presentation of Graves' orbitopathy

PROTAC	PROteolysis TArgeting Chimera
PRRT	peptide receptor radionuclide therapy
PRV	planning risk volume
PSA	prostate-specific antigen
PSIS	pituitary stalk interruption syndrome
PSMA	prostate-specific membrane antigen
PTC	papillary thyroid carcinoma
PTH	parathyroid hormone
PTHrP	parathyroid hormone-related protein
PTM	post-translational modification
PTPR	papillary tumours of the pineal region
PTSD	post-traumatic stress disorder
PTTG	pituitary tumour-transforming gene
PTV	planning target volume
PVH	periventricular hypothalamus
PVN	paraventricular nucleus
QC	quality control
QoL	quality of life
RAAS	renin angiotensin aldosterone system
RANKL	RANK ligand
RCC	Rathke's cleft cysts
RCOG	Royal College of Obstetrics and Gynaecology
RCT	randomized clinical trial
REM	rapid eye movement
REMS	risk evaluation and mitigation strategy
RL	RAS-like
RNI	reference nutrient intake
RR	relative risk
RRM	RNA recognition motifs
RS	radiosurgery
RTH	resistance to thyroid hormone
RTK	receptor tyrosine kinases
RXR	retinoid X receptor
SAH	subarachnoid haemorrhage
SCCM	Society of Critical Care Medicine
SCN	suprachiasmatic nucleus
SD	standard deviation
SDR	spontaneous dwarf rats
SDS	standard deviation score
SE	spin echo
SERM	selective oestrogen receptor modulator
SGK	serum glucocorticoid-regulated kinase
SH	subclinical hyperthyroidism
SHBG	sex-hormone-binding globulin
SIADH	syndrome of inappropriate antidiuretic hormone
SIBO	small intestinal bacterial overgrowth
SINE	selective inhibitors of nuclear export
SMR	standardized mortality ratio
SNP	single-nucleotide polymorphism
SOCS	suppressor of cytokine signalling
SOD	septo-optic dysplasia
SON	supraoptic nucleus
SPECT	single photon emission computed tomography
SRIF	somatotropin release inhibiting factor
SRL	somatostatin receptor ligands
SRS	sex reassignment surgery
SSA	senile systematic amyloidosis
SSRI	selective serotonin reuptake inhibitors
SST	short synacthen test
StAR	steroidogenic acute regulatory
SWS	slow-wave sleep
TAO	thyroid-associated orbitopathy
TBI	traumatic brain injury
TBS	trabecular bone score
TC	thyroid cancer
TCR	T-cell receptor
TDF	Tenofovir disoproxil fumarate
TED	thyroid eye disease
TGCC	testicular germ cell cancer
TGF	transforming growth factor
TH	thyroid hormone
THDP	thyroid hormone distributor protein
THR	thyroid hormone receptor
TKI	tyrosine kinase inhibitor
TMC	total motile count
TMD	transmembrane domains
TMN	tuberomammillary nucleus
TPP	thyrotoxic periodic paralysis
TR	thyroid hormone receptor
TR	thyroid receptor
TRH	thyrotropin-releasing hormone
TRT	testosterone replacement therapy
TSH	thyroid-stimulating hormone
TSHR	thyroid-stimulating hormone receptor
TSM	tuberculum sellae meningioma
TSS	transcriptional start site
TSS	transsphenoidal surgery
TVDT	tumour volume doubling time
UFC	urine-free cortisol
USI	universal salt iodization
VDDR	vitamin D-dependent rickets
VDR	vitamin D receptor
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
VEP	visual evoked potential
VFA	vertebral fracture assessment
VIP	vasoactive intestinal peptide
VLPO	ventrolateral preoptic area
VMH	ventromedial hypothalamus
VMN	ventromedial nucleus
vSPZ	ventral subparaventricular zone
VTA	ventral tegmental area
WADA	World Anti-Doping Agency
WBS	whole-body scan
WGS	whole genome sequencing
WHO	World Health Organization
WHR	waist/hip ratio
WT	Wilm's tumour
XLH	X-linked hypophosphataemia

Section Editors

Wiebke Arlt Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

John P. Bilezikian International Education and Research, Metabolic Bone Diseases Unit; Department of Medicine, Division of Endocrinology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Kristien Boelaert University Hospital Birmingham NHS Foundation Trust, Birmingham, UK

Peter Clayton Faculty of Biology, Medicine & Health, University of Manchester; Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Machester, UK

Mathis Grossmann Department of Endocrinology Austin Health, The University of Melbourne, Heidelberg, Victoria, Australia

Ken K.Y. Ho The Garvan Institute of Medical Research and St. Vincent's Hospital, Darlinghurst, Sydney, Australia

Desmond Johnston Department of Diabetes Endocrinology and Metabolic Medicine, Imperial College London, UK

John Newell- Price Department of Oncology and Metabolism, University of Sheffield; Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

David Ray Division of Diabetes, Endocrinology & Gastroenterology, Manchester Institute for Collaborative Research on Ageing, Lydia Becker Institute of Immunology and Inflammation, University of Manchester, UK

Robert K. Semple Professor of Translational Molecular Medicine, University of Edinburgh Centre for Cardiovascular Science, Queen's Medical Institute, Edinburgh, UK

James Shaw Newcastle Diabetes Centre, Newcastle University, Freeman Hospital, UK

Guy T'Sjoen Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

John A.H. Wass Professor of Endocrionolgy, Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Wilmar M. Wiersinga Department of Endocrinology and Metabolism, University of Amsterdam, Amsterdam, The Netherlands

Catherine Williamson Guy's and St Thomas' NHS Foundation Trust; King's College London, UK

Frederick Wu Division of Diabetes, Endocrinology & Gastroenterology, Faculty of Biology, Medicine and Health, University of Manchester, UK

Bulent Okan Yildiz Division of Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, Hacettepe University, Turke

Contributors

Rochan Agha-Jaffar Consultant, Imperial College Healthcare NHS Trust, London, UK

Shahzada Ahmed Department of Ear, Nose and Throat, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

S. Faisal Ahmed University of Glasgow, Royal Hospital for Sick Children, Glasgow, UK

Salma R. Ali Developmental Endocrinology Research Group, School of Medicine, Dentistry and Nursing, University of Glasgow; and Office for Rare Conditions, Royal Hospital for Children and Queen Elizabeth University Hospital, Glasgow, UK

Zain Al-Safi Department Obstetrics & Gynecology, University of California, Los Angeles, USA

Stephanie A. Amiel Department of Diabetes, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, UK

Nobuyuki Amino Amino Thyroid Research Laboratory, Osaka, Japan

Panagiotis Anagnostis Unit of Reproductive Endocrinology, First Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Vivian Anastasiou School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Bradley D. Anawalt Professor and Vice Chair of Medicine, University of Washington, Seattle, WA, USA

Stig Andersen Department of Geriatrics and Clinical Medicine, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

Bo Angelin Metabolism Unit and Integrated CardioMetabolic Centre, Department of Endocrinology, Metabolism & Diabetes, Department of Medicine, Karolinska Institute at Karolinska University Hospital Huddinge, Stockholm, Sweden

Leen Antonio Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven; Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium

Naoko Arata Division of Internal Medicine, National Center for Child Health and Development, Tokyo, Japan

Jon Arcelus Nottingham Centre for Transgender Health, Nottingham University School of Medicine, Nottingham, UK

Laleh Ardeshirpour Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Wiebke Arlt William Withering Chair of Medicine, Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

Eleni Armeni ENETS Centre of Excellence, Royal Free Hospital, London, UK

Baris Ata Department of Obstetrics and Gynaecology, School of Medicine, Koc University, Ankara, Turkiye

Richard J. Auchus Division of Metabolism, Endocrinology, and Diabetes, Department of Internal Medicine, Department of Pharmacology, and the Program for Disorders of Sexual Development, University of Michigan, Ann Arbor, MI, USA

Anna Aulinas Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona. Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Ricardo Azziz Center for Androgen Related Disorders, Cedars-Sinai Medical Center, and Department of Obstetrics and Gynecology and Department of Medicine, The David Geffen School of Medicine at UCLA, Los Angeles, USA

Muriel Babey Diabetes, Endocrinology and Metabolism Fellow, Division of Endocrinology and Metabolism, Department of Medicine, University of California, San Francisco, CA, USA

Corin Badiu Department of Endocrinology, C. I. Parhon National Institute, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

Michael N. Badminton Department of Medical Biochemistry and Immunology, University Hospital of Wales; School of Medicine, Cardiff University, Wales, UK

Clifford J. Bailey Aston Pharmacy School, School of Life and Health Sciences, Aston University, Birmingham, UK

Adam H. Balen Leeds Teaching Hospitals Leeds Fertility, Seacroft Hospital, Leeds, UK

Stephen G. Ball The Medical School, Newcastle University, and Newcastle Hospitals NHS Trust, Newcastle, UK

Irina Bancos Mayo Clinic, Rochester, MN, USA

Indraneel Banerjee Royal Manchester Children's Hospital, University of Manchester, UK

María-José Barahona Department of Endocrinology/Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona; Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Edouard Bardou-Jacquet Liver Unit, Pontchaillou University Hospital, Rennes, France

Luigi Bartalena Department of Medicine and Surgery, University of Insubria, Endocrine Unit, ASST dei Sette Laghi, Varese, Italy

Shehzad Basaria Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Waiel Bashari Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Ayat Bashir Northern Institute Forfor Cancer Research, Newcastle-upon-Tyne, UK

Fulvio Basolo Department of Surgery, Medical, Molecular, and Critical Area Pathology, University of Pisa, Pisa, Italy

J.H. Duncan Bassett Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Rachel L. Batterham Centre for Obesity Research, Department of Medicine, University College London; Bariatric Centre for Weight Management and Metabolic Surgery, University College London Hospital, NHS Foundation Trust; National Institute for Health Research, Biomedical Research Centre, University College London Hospital, London, UK

Adrian Bauman Sydney School of Public Health, Sydney University, Australia

Paul Bech Department of Investigative Medicine, Hammersmith Hospital, Imperial College, London, UK

Finn N. Bennedbæk Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen; Department of Endocrinology, Herlev and Gentofte Hospital, Herlev, Denmark

Emelie Benyi Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Sarah L. Berga Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT, USA

Jérôme Bertherat Endocrinology Department, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris; Endocrinology, Diabetes and Metabolism Department, INSERM U1016, Institut Cochin and Université Paris Descartes, Paris, France

Rudi Beschorner Diagnostic and Interventional Neuroradiology, Department of Radiology, University of Tuebingen, Tuebingen, Germany

Rachel Besser Consultant in Paediatric Diabetes and Endocrinology at Oxford University Hospitals NHS Foundation Trust; and Honorary Senior Clinical Lecturer at the University of Oxford, Oxford, UK

Ragini Bhake Department of Diabetes and Endocrinology, University Hospital of Leicester NHS Trust, Leicester Royal Infirmary, Leicester, UK

N.R. Biermasz Department of Medicine, Division of Endocrinology, and Center for Endocrine Tumors, Leiden, The Netherlands

John P. Bilezikian International Education and Research, Metabolic Bone Diseases Unit; Department of Medicine, Division of Endocrinology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Anita Boelen Department of Endocrinology and Metabolism, Academic Medical Center, University of Amsterdam, AZ Amsterdam, Netherlands

Steen Joop Bonnema Professor and Consultant, University of Southern Denmark, Department of Endocrinology, Odense University Hospital, Odense, Denmark

Fidéline Bonnet-Serrano Hormonal Biology Department, Cochin Hospital, APHP; Paris Descartes University, Paris, France

Jean-François Bonneville Department of Endocrinology, CHU Sart-Tilmann, Liège, Belgium

Charlotte Boughton University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Sheerazed Boulkroun INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Andrew J.M. Boulton Division of Endocrinology, Diabetes & Gastroenterology, Faculty of Medicine, Biology and Health, University of Manchester; Consultant Physician, Manchester Royal Infirmary, Manchester, UK; Visiting Professor, Diabetes Research Institute, University of Miami, Miami, FL, USA

Walter Pierre Bouman Nottingham Centre for Transgender Health, Nottingham University School of Medicine, Nottingham, UK

Frank Lee Bowling Podiatric Consultant Surgery, Division of Surgery, Manchester Royal Infirmary and Reader in Translational Medicine; Division of Endocrinology, Diabetes & Gastroenterology, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK

Gurkan Bozdag Department of Obstetrics and Gynaecology, School of Medicine, Hacettepe University, Istanbul, Turkiye

Georg Brabant FRCP Exp. & Clin Endocrinology, University of Lübeck, Germany

Glenn D. Braunstein Department of Medicine, Cedars-Sinai Medical Center and University of California, Los Angeles, CA, USA

Gemma V. Brierley University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science; National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Augustin Brooks Bournemouth Diabetes and Endocrine Centre, Royal Bournemouth Hospital, Bournemouth, UK

Morris J. Brown Barts and the London School of Medicine; St Bartholomew's Hospital, London, UK

Gary Butler Consultant in Paediatric and Adolescent Endocrinology, University College London Hospital, Honorary Clinical Professor in Child and Adolescent Health, UCL Great Ormond Street Institute of Child Health, London, UK

Marilia Calcia Diabetes Psychiatry and Psychology Service, King's College Hospital, London, UK

Chris Callaghan Guy's and St Thomas' NHS Foundation Trust, London, UK

Fiona Campbell Children's Diabetes Centre, Leeds Children's Hospital, Leeds Teaching Hospitals, Leeds, UK

Ernesto Canalis Departments of Orthopaedic Surgery and Medicine, and The UConn Musculoskeletal Institute, UConn Health, Farmington, CT, USA

Martyn E. Caplin Department of Gastroenterology, Royal Free Hospital, London, UK

Cesare Carani University of Modena and Reggio Emilia, Modena, Italy

Jean-Claude Carel Pediatric Endocrinology Diabetology Department, Reference Center for Growth and Development Endocrine Diseases, Université de Paris; Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1141, Assistance Publique-Hôpitaux de Paris, Robert-Debré University Hospital, Paris, France

Thomas O. Carpenter Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Paul V. Carroll Consultant Endocrinologist and Honorary Senior Lecturer, Department of Endocrinology, Guy's & St. Thomas' NHS Foundation Trust; Faculty of Life Sciences & Medicine, King's College London, UK

Ruth T. Casey Department of Medical Genetics, University of Cambridge, Cambridge, UK

Manuela Cerbone University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Filomena Cetani Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Ghassan Chamseddine School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Li F. Chan Centre for Endocrinology, William Harvey Research Institute, London, UK

R. Jeffrey Chang Division of Reproductive Endocrinology and Infertility, University of California, San Diego, CA, USA

Krishna Chatterjee Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, UK

Steven Chernausek Department of Pediatrics, Section of Diabetes and Endocrinology, University of Oklahoma Health Science Center, College of Medicine, Oklahoma City, OK, USA

C.L. Cheung Department of Therapeutics, The University of Hong Kong, Hong Kong, China

Jennifer Chin Department of Obstetrics, Gynecology, and Women's Health, University of Hawaii, Honolulu, HI, USA

Luca Chiovato Department of Endocrinology, University of Pisa, Pisa, Italy

Yeoungjee Cho Department of Nephrology, Princess Alexandra Hospital, Brisbane, Australia

Pratik Choudhary Department of Diabetes, King's College London, London, UK

Frank Claessens Department of Cellular and Molecular Medicine, Laboratory of Molecular Endocrinology, KU Leuven, Leuven, Belgium

Orlo H. Clark Endocrine Surgery Section, Department of Surgery, University of California, San Francisco, CA, USA

Robert Clarke Department of Epidemiology and Public Health, University of Oxford, Oxford, UK

Bart L. Clarke Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, Minnesota, USA

Philip Clarke Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford; School of Population Health, University of Melbourne, Melbourne, UK

Dominique Clement Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Kings College Hospital, London, UK

Stephen Colagiuri Boden Institute of Obesity, Nutrition, Exercise & Eating Disorders, University of Sydney, Sydney, Australia

Trevor Cole Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

Anthony P. Coll University Lecturer/Honorary Consultant Physician, University of Cambridge Metabolic Research Laboratories, MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Cambridge, UK

Cristina Colom Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona; Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Sarah J. Conner Centre for Human Reproductive Science, IMSR, College of Medical and Dental Sciences, University of Birmingham; ICS, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK Consultant Gastroenterologist, Kings Health Partners ENETS Centre of Excellence, London, UK

Gerard S. Conway Professor of Reproductive Endocrinology, Institute for Women's Health, University College London, UK

Giovanni Corona Endocrinology Unit, Medical Department, AziendaUsl Bologna Maggiore-Bellaria Hospital, Bologna, Italy

Fernanda A. Correa Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular (LIM/42) do Hospital das Clinicas, Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Jeremy Cox Department of Endocrinology and Diabetes, St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Richard M. Cubbon Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT Laboratories, University of Leeds, Leeds, UK

Janine A. Danks Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Australia, Australia

Mehul T. Dattani University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Chantal Daumerie Department of Endocrinology, Université catholique de Louvain, Cliniques Universitaires Saint-Luc, Brussels, Belgium

Terry F. Davies Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Melanie J. Davies Diabetes Research Centre, University of Leicester, Leicester General Hospital, Leicester, UK

Julian Davis Department of Endocrinology, Manchester Royal Infirmary; School of Medical Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK

Colin Dayan Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK

Chloë De Roo Department of Reproductive Medicine, Ghent University Hospital, Corneel Heymanslaan, Ghent, Belgium

Sunita M.C. De Sousa Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Justine Defreyne Department of Endocrinology, Center for Sexology and Gender, Ghent University Hospital, Corneel Heymanslaan, Ghent, Belgium

Yves Deugnier Liver Unit, Pontchaillou University Hospital, Rennes, France

Ketan Dhatariya Consultant in Diabetes, Endocrinology and General (Internal) Medicine Honorary Professor, Norwich Medical School, Elsie Bertram Diabetes Centre, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK

Whaljit Dhillo Department of Investigative Medicine, Hammersmith Hospital, Imperial College, London, UK

Evanthia Diamanti-Kandarakis Department of Endocrinology and Diabetes, HYGEIA Hospital, Athens, Greece

Ruxandra Dobrescu Parhon National Endocrinology Institute in Bucharest, Bucharest, Romania

Adrian Dobs Divisions of Endocrinology, Metabolism and Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Martin Drage Guy's and St Thomas' NHS Foundation Trust, Great Ormond Street Hospital, London, UK

William M. Drake Department of Endocrinology, St Bartholomew's Hospital; Barts and the London School of Medicine, London, UK

Daniel Dumesic Department Obstetrics & Gynecology, University of California, Los Angeles, USA

Richard Eastell Academic Unit of Bone Metabolism, Metabolic Bone Centre, Northern General Hospital, Sheffield, UK

David A. Ehrmann The University of Chicago, Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Chicago, IL, USA

Mustapha El Lakis Endocrine Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Rossella Elisei Department of Endocrinology and Metabolism, University of Pisa, Pisa, Italy

Jackie Elliott Sheffield Teaching Hospitals Foundation Trust, University of Sheffield, Sheffield, UK

Cara E. Ellis Laboratory of Molecular and Cellular Medicine, Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Charis Eng Genomic Medicine Institute, Lerner Research Institute; Taussig Cancer Institute, Cleveland Clinic; Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine; Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Cleveland, OH, USA

Ulrike Ernemann Department of Neuropathology, University of Tuebingen, Tuebingen, Germany

Mohammed Eslam Storr Liver Centre, The Westmead Institute for Medical Research, University of Sydney and Westmead Hospital, Westmead, Australia

I. Sadaf Farooqi University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Maritza T. Farrant University of Washington School of Medicine, Seattle, WA

Alan P. Farwell Division of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts, USA

Fabio R. Faucz Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Ulla Feldt-Rasmussen Department of Medical Endocrinology and Metabolism, Rigshospitalet, Copenhagen University, Copenhagen, Denmark

Fabio Fernandes-Rosa INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Luca Ferraro School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Naomi Fersht Department of Oncology, University College London Hospital, London, UK

Nicholas Finer National Centre for Cardiovascular Prevention and Outcomes, Institute of Cardiovascular Science, University College London, London, UK; Senior Principle Clinical Scientist, Global Medical Affairs Management, Novo Nordisk A/S

Sarah B. Fisher Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Anneliese Flatt Newcastle University, Newcastle-upon-Tyne, UK

Karin Frank-Raue Endokrinologische Gemeinschaftspraxis, Heidelberg, Germany

Steve Franks Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College; St. Mary's and Hammersmith Hospitals, London, UK

Bernard Freudenthal Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Dagmar Führer Professor of Medicine, Department of Endocrinology, Diabetes and Metabolism, University Hospital Essen, University of Duisburg-Essen, Germany

John W. Funder Prince Henry's Institute of Medical Research, Monash Medical Centre, Clayton, Australia

John S. Fuqua Professor of Clinical Pediatrics, Division of Pediatric Endocrinology and Diabetology, Indiana University School of Medicine, Riley Hospital for Children, Indianapolis, IN, USA

Thiago Gagliano-Jucá Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Sonya Galcheva Department of Paediatrics, Varna Medical University/University Hospital 'St. Marina', Varna, Bulgaria

Alessandra Gambineri Endocrinology Unit, Department of Medical and Surgical Sciences, St Orsola-Malpighi Hospital, Alma Mater University of Bologna, Bologna, Italy

Hoong-Wei Gan University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Chris Garrett Diabetes and Mental Health Group, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

Amalia Gastaldelli Institute of Clinical Physiology, CNR, Pisa, Italy

Minu George Department of Pediatrics, Section of Diabetes and Endocrinology, University of Oklahoma Health Science Center, College of Medicine, Oklahoma City, OK, USA

Jacob George Storr Liver Centre, The Westmead Institute for Medical Research, University of Sydney and Westmead Hospital, Westmead, Australia

Nitin Narayan Gholap Consultant in Diabetes and Endocrinology, Departments of Diabetes, Endocrinology, and Metabolism, University Hospitals Coventry and Warwickshire, Coventry, UK; Diabetes Research Centre, University of Leicester, Leicester, UK

James Gibney Tallaght University Hospital, Dublin, Ireland

Helena K. Gleeson Department of Endocrinology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; Diabetes Research Centre, University of Leicester, Leicester, UK

Luigi Gnudi School of Cardiovascular Medicine & Science, King's College London, London, UK

David Goltzman Metabolic Disorders and Complications (MeDiC), Room # EM1.3220, RI-McGill University Health Centre, Glen Site, 1001 Décarie Blvd, Montréal, QC H4A 3J1, Canada

Phillip Gorden National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

Dimitrios G. Goulis Unit of Reproductive Endocrinology, First Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Hans Graf Professor of Endocrinology at the Federal University of Paraná, Brazil

Antonio R.M. Granata Department of Medical Specialties, Azienda Ospedaliero-Universitaria of Modena, Modena, Italy

Claus H. Gravholt Department of Endocrinology, Aarhus University, Aarhus, Denmark

Alice Greenhalgh Institute of Cancer Sciences, University of Manchester, Manchester, UK

Louise C. Gregory Genetics and Genomic Medicine Programme, UCL Great Ormond Street Institute of Child Health, London, UK

Steven K. Grinspoon Professor of Medicine, Harvard Medical School; MGH Program in Nutritional Metabolism and Neuroendocrine Unit, Boston, MA, USA

Ashley B. Grossman Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Royal Free Hospital, London, UK

Mathis Grossmann Professor, Department of Medicine (Austin Health), The University of Melbourne, Department of Endocrinology, Austin Health, Heidelberg, Australia

Peter Guest X-ray Department, Radiology, University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital Medical Centre, Birmingham, UK

Jan Gunst Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven, Belgium

Mark Gurnell Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Stefanie Hahner Department of Medicine, Endocrinology and Diabetology, University Hospital Würzburg, Germany

Peter Hammond Harrogate and District NHS Foundation Trust, Lancaster Park Road, Harrogate, North Yorkshire, UK

David J. Handelsman ANZAC Research Institute, and Department of Andrology, Concord Hospital, Sydney, Australia

Fadil M. Hannan Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, Faculty of Health & Life Sciences, University of Liverpool, Liverpool; Academic Endocrine Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Andrew Hattersley Professor of Molecular Genetics, College of Medicine and Health, University of Exeter; and Honorary Consultant in Diabetes and Endocrinology, The Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Laszlo Hegedüs Professor and Consultant, University of Southern Denmark, Department of Endocrinology, Odense University Hospital, Odense, Denmark

Simon Heller School of Medicine and Biomedical Sciences, University of Sheffield, Sheffield, UK

Christel Hendrieckx School of Psychology, Deakin University, Geelong; The Australian Centre for Behavioural Research in Diabetes, Diabetes Victoria, Melbourne, Victoria, Australia

Geoffrey N. Hendy† Calcium Research Laboratory, Royal Victoria Hospital, Montreal, Canada

David Henley Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Perth, Australia

Claire E. Higham Consultant Endocrinologist, Christie Hospital NHS Foundation Trust; University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

Olaf Hiort Division of Paediatric Endocrionolgy and Diabetes, Department of Paediatric and Adolescent Medicine, University of Lübeck, Germany

Ken K.Y. Ho The Garvan Institute of Medical Research, St. Vincent's Hospital and the University of New South Wales, Sydney, Australia

Anita C.S. Hokken-Koelega Erasmus University Medical Center, Rotterdam, The Netherlands

Richard I.G. Holt Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton Faculty of Medicine, Southampton, UK

Jürgen Honegger Department of Neurosurgery, University of Tuebingen, Tuebingen, Germany

Roman Hovorka University of Cambridge Metabolic Research Laboratories, Wellcome Trust- MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Wenyu Huang Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Ilpo Huhtaniemi Institute of Reproductive and Developmental Biology, Imperial College London; Institute of Biomedicine, Research Centre for Integrative Physiology and Pharmacology, University of Turku

Khalid Hussain Department of Paediatric Medicine, Division of Endocrinology, Sidra Medicine, Doha, Qatar

Daniela Ibarra-Gasparini Endocrinology Unit, Department of Medical and Surgical Sciences, St Orsola-Malpighi Hospital, Alma Mater University of Bologna, Bologna, Italy

Radha Indusekhar Department of Obstetrics and Gynaecology, North Staffordshire Hospital, The University Hospital of North Midlands NHS Trust, Royal Stoke University Hospital, Staffordshire, UK

Khalida Ismail Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

Peter Jacob Department of Diabetes, King's College London, London, UK

Susie Jacob Leeds Centre for Reproductive Medicine, Seacroft Hospital, Leeds, UK

Barbara Jarzab M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Holger Jäschke Department of Endocrinology, Diabetes and Metabolism, University Hospital Essen, University of Duisburg-Essen, Germany

Friedrich C. Jassil Centre for Obesity Research, Department of Medicine, University College London; Bariatric Centre for Weight Management and Metabolic Surgery, University College London Hospital, NHS Foundation Trust, London, UK

Channa Jayasena Departments of Medicine and Andrology, Imperial College London, Hammersmith Hospital; Department of Obstetrics & Gynecology, St Mary's Hospital, London, UK

David W. Johnson Department of Nephrology, Princess Alexandra Hospital, Brisbane, Australia

Peter M. Jones Department of Diabetes, King's College London, London, UK

Alexander A.L. Jorge Unidade de Endocrinologia Genética, Laboratório de Endocrinologia Celular e Molecular (LIM/25), Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Jens O.L. Jørgensen Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark

Anne Jouinot Oncology Department, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris; Endocrinology, Diabetes and Metabolism Department, INSERM U1016, Institut Cochin and Université Paris Descartes, Paris, France

Eleni A. Kandaraki Department of Endocrinology and Diabetes, HYGEIA Hospital, Athens, Greece

Bliss Kaneshiro Department of Obstetrics, Gynecology and Women's Health, University of Hawaii, Honolulu, HI, USA

Niki Karavitaki Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Jean-Marc Kaufman Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

Mark T. Kearney Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT laboratories, University of Leeds, Leeds, UK

Electron Kebebew Department of Surgery and Stanford Cancer Institute, Stanford University, Stanford, CA, USA

Brian Keevil Clinical Biochemistry, Manchester University Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Bernard Khoo ENETS Centre of Excellence, Royal Free Hospital, London, UK

Kamlesh Khunti Diabetes Research Centre; Leicester Clinical Trials Unit, University of Leicester, Leicester Diabetes Centre; NIHR Collaborations for Leadership in Applied Health Research and Care, Leicester, UK

Timothy J. Kieffer Laboratory of Molecular and Cellular Medicine, Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Na Ri Kim Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven, Leuven, Belgium

Jackson C. Kirkman-Brown Centre for Human Reproductive Science, IMSR, College of Medical and Dental Sciences, University of Birmingham; Birmingham Women's Fertility Centre, Birmingham Women's & Children's NHS Foundation Trust, Birmingham, UK

Anne Klibanski Neuroendocrine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Daniel Klink ZNA Queen Paola Children's Hospital Antwerp, Antwerp, Belgium

Joanna Klubo-Gwiezdzinska Metabolic Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

Karen E. Knudsen Departments of Cancer Biology, Medical Oncology, Urology, Radiation Oncology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Josef Köhrle Institut für Experimentelle Endokrinologie und Endokrinologisches Forschungs-Centrum der Charité EnForCé, Charité Universitätsmedizin Berlin, CVK, Berlin, Germany

Márta Korbonits Centre for Endocrinology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London, UK

Tim I.M. Korevaar Department of Internal Medicine, Academic Center for Thyroid Diseases, Erasmus Medical Center, Rotterdam, The Netherlands

Olympia Koulouri Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Kalman Kovacs Department of Laboratory Medicine, Division of Pathology; The Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

Jolanta Krajewska M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Jelena Kravarusic Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Nils P. Krone Academic Unit of Child Health, Department of Oncology and Metabolism, University of Sheffield, Sheffield Children's Hospital, Sheffield, UK

Annie W.C. Kung Department of Medicine, The University of Hong Kong, Hong Kong, China

Robin H. Lachmann Charles Dent Metabolic Unit, National Hospital for Neurology and Neurosurgery, London, UK

Lies Langouche Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium

Michael Laßmann Klinik und Poliklinik für Nuklearmedizin, Universitätsklinikum Würzburg, Würzburg, Germany

Francesco Latrofa Department of Clinical and Experimental Medicine, Endocrinology Unit I, University of Pisa; University Hospital of Cisanello, Pisa, Italy

Deepthi Lavu Department of Obstetrics and Gynaecology, Health Education West Midlands, Birmingham, UK

Edward R. Laws Jr. Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

John H. Lazarus Thyroid Research Group, Cardiff University, Cardiff, Wales, UK

Juliane Léger Pediatric Endocrinology Diabetology Department, Reference Center for Growth and Development Endocrine Diseases, Université de Paris; Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1141, Assistance Publique-Hôpitaux de Paris, Robert-Debré University Hospital, Paris, France

Michael A. Levine Professor of Pediatrics and Medicine, Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia and Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Miles J. Levy Department of Diabetes and Endocrinology, University Hospital of Leicester NHS Trust, Leicester Royal Infirmary, Leicester, UK

Rossella Libè Department of Endocrinology, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France

Steven A. Lietman Department of Orthopaedic Surgery, Geisinger Medical Center, Danville, PA, USA

Stafford L. Lightman Henry Wellcome Laboratories for Integrative Neuroscience & Endocrinology, University of Bristol, Bristol, UK

Kok Haw Jonathan Lim Department of Medical Oncology, The Christie NHS Foundation Trust, Manchester, UK

Siew Lim Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Jonathan Z.M. Lim University of Liverpool, Liverpool, UK

Kirstie Lithgow Clinical Research Fellow, Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Lynn Loriaux Department of Medicine, Oregon Health Sciences University, Portland, Oregon, USA

Sandra Lowe Royal Hospital for Women; School of Women's and Children's Health, University of New South Wales, Sydney, Australia

Thomas Lung The George Institute for Global Health, University of New South Wales, Kensington, Australia

Markus Luster Department of Nuclear Medicine, University Hospital Marburg, Marburg Germany

Djuro Macut Clinic of Endocrinology, Diabetes and Metabolic Diseases, and Faculty of Medicine, University of Belgrade, Belgrade, Serbia

Mario Maggi Andrology and Sexual Medicine Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy

Eamonn R. Maher Department of Medical Genetics, University of Cambridge, Cambridge, UK

Ingrid Y.F. Mak Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Royal Free Hospital, London, UK

Noemi Malandrino National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

Shreya Malhotra Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Crystal Man Ying Lee Boden Institute of Obesity, Nutrition, Exercise & Eating Disorders, University of Sydney, Sydney; School of Psychology and Public Health, La Trobe University, Melbourne, Australia

Prakash Manoharan Department of Radiology, The Christie NHS Foundation Trust, Manchester, UK

Wasat Mansoor Department of Medical Oncology, The Christie NHS Foundation Trust, Manchester, UK

Claudio Marcocci Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

John C. Marshall Center for Research in Reproduction, University of Virginia, Charlottesville, VA, USA

Sally M. Marshall Translational and Clinical Research Institute (Diabetes), Faculty of Clinical Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

George Mastorakos Endocrine Unit, University of Athens, Medical School, 'Aretaieion' Hospital, Athens, Greece

Alvin M. Matsumoto Department of Medicine, Division of Gerontology and Geriatric Medicine, University of Washington, Seattle VA Puget Sound Health Care System, Seattle, Washington, USA

Gherardo Mazziotti Endocrine Unit, ASST Carlo Poma, Mantua, Italy

Christopher R. McCartney Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, VA, USA

Ann McCormack Department of Endocrinology, St Vincent's Hospital; Garvan Institute of Medical Research; University of New South Wales, Darlinghurst, NSW, Australia

Karim Meeran Department of Investigative Medicine, Imperial College London, London, UK

Shlomo Melmed Pituitary Center, Cedars-Sinai Medical Center, Los Angeles, California

Francesca Menconi Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Radu Mihai Department of Endocrine Surgery, Churchill Cancer Centre, Oxford University Hospitals Foundation Trust, Oxford, UK

Paul D. Miller Distinguished Clinical Professor of Medicine, University of Colorado Health Sciences Center, Medical Director, Colorado Center for Bone Research at Panorama Orthopedics and Spine Center, Golden, Colorado, USA

Karen K. Miller Neuroendocrine Unit, Massachussetts General Hospital; Faculty of Medicine at Harvard Medical School, Boston, MA, USA

Mark E. Molitch Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Carla Moran Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, UK

Lisa Moran National Health and Medical Research Council Centre for Research Excellence in PCOS, University of Adelaide, Adelaide; Monash Centre for Health Research and Implementation, Monash Public Health and Preventive Medicine, Monash University and Monash Health, Melbourne, Vic., Australia

John F. Morris Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford, UK

Laura Morrison Faculty of Medicine, Cancer Center, Yale School of Medicine, New Haven, CT, USA

Calum Moulton Department of Psychological Medicine, King's College London, UK

Aya Mousa Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Sezcan Mumusoglu Department of Obstetrics and Gynecology, School of Medicine, Hacettepe University, Ankara, Turkey

Alia Munir Sheffield Teaching Hospitals NHS Foundation Trust, University of Sheffield, Sheffield, UK

Helen R. Murphy Cambridge University Hospitals NHS Foundation Trust, Cambridge; Women's Health Academic Centre, Division of Women's and Children's Health, King's College London, London; Norwich Medical School, University of East Anglia, Norwich, UK

Robert D. Murray Consultant Endocrinologist, Leeds Teaching Hospital NHS Trust; Honorary Associate Professor, Leeds Institute for Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK

Sonia Nagi Faculty of Medicine, National Institute of Neurology, University of Tunis El Manar, Tunis-Tunisia

Radhika Narla Department of Medicine, Division of Metabolism, Endocrinology and Nutrition, University of Washington, Seattle VA Puget Sound Health Care System, Seattle, Washington, USA

Melissa Nataatmadja Department of Nephrology, Sunshine Coast University Hospital and Health Service; Faculty of Medicine, University of Queensland, Brisbane, Australia

John Newell-Price Department of Oncology and Metabolism, University of Sheffield; Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Markella Nezi Endocrine Unit, University of Athens, Medical School, 'Aretaieion' Hospital, Athens, Greece

Vania Nosé Professor of Pathology, Harvard Medical School, Director of Anatomic and Molecular Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA

Birte Nygaard Department of Clinical Medicine, Herlev-Gentofte Hospital, University of Copenhagen, Denmark

Shaughn O'Brien Department of Obstetrics and Gynaecology, Keele University, Keele; Nuffield Health North Staffordshire Hospital, Newcastle-under-Lyme, UK

Stephen O'Rahilly Department of Endocrinology, Beaumont Hospital Dublin and Royal College of Surgeons in Ireland (RCSI), Ireland

Michael W. O'Reilly Department of Endocrinology, Beaumont Hospital Dublin and Royal College of Surgeons in Ireland (RCSI), Ireland

Kutluk Oktay Innovation Institute for Fertility Preservation and In Vitro Fertilization, New York, NY; Laboratory of Molecular Reproduction and Fertility Preservation, Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Nick Oliver Department of Metabolism, Digestion and Reproduction, Imperial College Healthcare NHS Trust, London, UK

Richard A. Oram Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK

Jacques Orgiazzi Université Claude-Bernard Lyon-1, and Hospices Civils de Lyon, Service d'Endocrinologie-Diabétologie-Maladies Métaboliques, Groupement Hospitalier Sud, Pierre-Bénite, France

Henrik Oster Institute for Neurobiology, University of Lübeck, CBBM, Germany

Katharine R. Owen Associate Professor, University of Oxford, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Oxford, UK

Catherine J. Owen Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, UK

Furio Pacini Departments of Internal Medicine, Endocrinology and Metabolism, and Biochemistry, University of Siena, Siena, Italy

Stephanie T. Page University of Washington School of Medicine, Seattle, WA

Fausto Palazzo Hammersmith Hospital and Imperial College London, UK

Maria Papagianni Pediatric Endocrinology Unit, Third Department of Pediatrics, 'Hippokrateion' General Hospital of Thessaloniki, Aristotle University of Thessaloniki School of Medicine, Thessaloniki, Greece

Socrates E. Papapoulos Department of Endocrinology and Metabolic Diseases, Leiden University Medical Centre, Leiden, The Netherlands

Thomas G. Papathomas Clinician Scientist in Molecular Pathology, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

Paolo Parini Division of Clinical Chemistry, Department of Laboratory Medicine, Metabolism Unit, Department of Endocrinology, Metabolism & Diabetes, Department of Medicine, Karolinska Institute at Karolinska University Hospital and Inflammation and Infection Theme, Karolinska University Hospital Huddinge, Stockholm, Sweden

Stavroula A. Paschou Assistant Professor of Endocrinology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

Peysh A. Patel Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT laboratories, University of Leeds, Leeds, UK

Kashyap A. Patel Wellcome Turst Clinical Research Career Development Felllow, College of Medicine and Health, University of Exeter; and Honorary Consultant in Diabetes and Endocrinology, The Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Gilberto Paz-Filho Medical & Scientific Affairs, Janssen Australia and New Zealand

Michael Pazianas Senior Research Associate, Institute of Musculoskeletal Sciences, Oxford University, Oxford, UK

Elizabeth N. Pearce Division of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts, USA

Simon H.S. Pearce Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, UK

Inge Bülow Pedersen Department of Endocrinology and Medicine, Aalborg Hospital, Århus University Hospital, Aalborg, Denmark

Robin P. Peeters Department of Internal Medicine, Academic Center for Thyroid Diseases, Erasmus Medical Center, Rotterdam, The Netherlands

David L. Penn Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Nancy D. Perrier Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Shanta J. Persaud Department of Diabetes, King's College London, London, UK

John R. Petrie Professor of Diabetic Medicine, Institute of Cardiovascular and Medical Sciences, BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK

John Pickup Diabetes Research Group, King's College London School of Medicine, Guy's Hospital, London, UK

Duarte Pignatelli Faculty of Medicine of Porto and I3S - Institute of Investigation and Innovation in Health, University of Porto, and Centro Hospitalar Universitário S João, Porto, Portugal

Terhi Piltonen Department of Obstetrics and Gynecology, PEDEGO Research Unit, Medical Research Center, Oulu University Hospital, University of Oulu, Oulu, Finland

Biljana Popovic-Todorovic Centre for Reproductive Medicine, UZ Brussels, Belgium

Kris Poppe Endocrine Unit Centre Hospitalier Universitaire Saint Pierre; Université Libre de Bruxelles (ULB), Brussels, Belgium

Iulia Potorac Endocrinology, CHU Sart-Tilmann, Liège, Belgium

L.D.K.E. Premawardhana Thyroid Research Group, Cardiff University, Cardiff, Wales, UK

Jackie Price Public Health Sciences Section, Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, UK

Hermione Price Research and Development Department, Tom Rudd Unit, Moorgreen Hospital, Southampton, UK

Margit G. Proescholdt University of Basel, Psychiatric Hospital, Division of Substance Use Disorders, Switzerland

Dalal Qanaq School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Shwetha Ramachandrappa Centre for Endocrinology, William Harvey Research Institute, London, UK

Giulia Rastrelli Andrology and Sexual Medicine Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy

Friedhelm Raue Endokrinologische Gemeinschaftspraxis, Heidelberg, Germany

David W. Ray Division of Digestion, Endocrinology and Metabolism, The University of Manchester, Manchester; NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital; Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

Gerry Rayman The Ipswich Diabetes Centre and Research Unit, Ipswich Hospital NHS Trust, Suffolk, UK

Narendra Reddy Department of Endocrinology, University Hospital of Leicester NHS Trust, Leicester, UK

Caroline S. Repetti Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Samantha J. Richardson School of Science and School of Health and Biomedical Sciences, RMIT University, Melbourne, Australia

Guido Rindi Section of Anatomic Pathology, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Roma, Italia; Section of Anatomic Pathology, Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italia; Roma European NeuroEndocrine Tumor Society (ENETS) Center of Excellence, Roma, Italia

Rachel E. Roberts Department of Obstetrics and Gynaecology, Queen Charlotte's and Chelsea Hospital, Imperial College Healthcare NHS Trust, London, UK

Cemre Robinson Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA; Assistant Professor of Pediatrics, Division of Endocrinology and Diabetes, Icahn School of Medicine at Mount Sinai, New York, USA

Stephen Robinson Department of Endocrinology and Diabetes, St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Vincenzo Rochira Unit of Endocrinology, Department of Biomedical, Metabolic, and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy; Department of Medical Specialties, Azienda Ospedaliero-Universitaria of Modena, Modena, Italy

Alan D. Rogol Emeritus Professor of Pediatrics, Section of Pediatric Endocrinology and Diabetology, University of Virginia School of Medicine, Charlottesville, VA, USA

Stefano Romeo Institute of Medicine, Sahlgrenska Academy, Department of Molecular and Clinical Medicine, University of Gothenburg; Cardiology Department, Sahlgrenska University Hospital, Gothenburg, Sweden; Clinical Nutrition Unit, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy

Richard Ross Department of Oncology and Metabolism, University of Sheffield, UK

Fabio Rotondo Department of Laboratory Medicine, Division of Pathology; The Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

Francesco Rubino School of Life Course Sciences, Department of Diabetes, King's College London; Bariatric and Metabolic Surgery, London, UK

Dagmara Rusinek M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Susan Sam The University of Chicago, Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Chicago, IL, USA

Ferruccio Santini Department of Endocrinology and Metabolism, University of Pisa, Pisa, Italy

Massimo Santoro Dipartimento di Medicina Molecolare e Biotecnologie Mediche, University 'Federico II', Naples, Italy

Federica Saponaro Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Naveed Sattar Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, Scotland, UK

David B. Savage Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Lars Sävendahl Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Clark Sawin† Veterans Health Administration, Washington, DC, USA

Renata C. Scalco Unidade de Endocrinologia Genética, Laboratório de Endocrinologia Celular e Molecular (LIM/25), Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo; Disciplina de Endocrinologia da Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo, Brazil

Peter H. Scanlon Gloucestershire and Oxford Eye Units; Harris Manchester College, University of Oxford; University of Gloucestershire, UK

Matthew J. Schiewer Department of Cancer Biology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Danja Schulenburg-Brand Department of Medical Biochemistry and Immunology, University Hospital of Wales; School of Medicine, Cardiff University, Wales, UK

Peter Senior Division of Endocrinology and Metabolism, Department of Medicine, University of Alberta, Edmonton, Canada

Amna Sheri Breast Unit, Royal Marsden Hospital, London, UK

Dolores M. Shoback Professor of Medicine, Endocrine Research Unit, Department of Medicine, San Francisco Veterans Affairs Medical Center, University of California, San Francisco, CA, USA

Soulmaz Shorakae Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Angela Shore Institute of Biomedical and Clinical Science, Peninsula Medical School, University of Exeter, Hevitree, UK

Kenneth Siddle University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science; National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Craig Sineath Rollins School of Public Health, School of Medicine, Emory University, Atlanta, Georgia, USA

Timothy C. Skinner La Trobe University, Bendigo, Victoria, Australia; Institute for Psychology, University of Copenhagen; Steno Diabetes Centre Copenhagen, Denmark

Francesca Soldà Department of Oncology, University College London Hospital, London, UK

Peter Sonksen St Thomas' Hospital and King's College, London; University of Southampton Faculty of Medicine, Southampton, UK

Jane Speight School of Psychology, Deakin University, Geelong; The Australian Centre for Behavioural Research in Diabetes, Diabetes Victoria, Melbourne, Victoria, Australia

Raj Srirajaskanthan Department of Gastroenterology, Kings College Hospital, London

Marietta Stadler Diabetes Department, King's College London Faculty of Life Sciences and Medicine, London, UK

Irinel Stanciu Clinical Assistant Professor of Medicine, University of Colorado Health Sciences Center Panorama Orthopedics and Spine Center

Takara L. Stanley Associate Professor in Pediatrics, Harvard Medical School, Boston, MA, USA

Anna Stears Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Constantine A. Stratakis Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Agnieszka Święcicka Department of Endocrinology, Manchester Royal Infirmary; School of Medical Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK

Luis V. Syro Department of Neurosurgery, Hospital Pablo Tobon Uribe and Clinica Medellin, Medellin, Colombia

Guy T'Sjoen Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

Vin Tangpricha School of Medicine, Emory University, Atlanta, Georgia, USA

Juha Tapanainen Department of Obstetrics and Gynecology, University of Helsinki, Helsinki, Finland

Danijela Tatovic North Bristol NHS Trust, University of Bristol, Bristol; Cardiff University School of Medicine, Cardiff, UK

Enes Taylan Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Roy Taylor Professor of Medicine and Metabolism, Magnetic Resonance Centre, Institute of Cellular Medicine, Newcastle University, UK

Peter N. Taylor Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK

Shu Teng Chai Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Massimo Terzolo Department of Clinical and Biological Sciences, San Luigi Gonzaga Hospital, University of Turin, Orbassano, Italy

Solomon Tesfaye Consultant Diabetologist and Honorary Professor of Diabetic Medicine University of Sheffield, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Rajesh V. Thakker Academic Endocrine Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Henri Timmers Department of Internal Medicine, Division of Endocrinology, Radboud University Medical Centre, Nijmegen, The Netherland

Yaron Tomer Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Jeremy W. Tomlinson Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, UK

Massimo Tonacchera Department of Endocrinology, University of Pisa, Pisa, Italy

Jorma Toppari Institute of Biomedicine, Research Centre for Integrative Physiology and Pharmacology, University of Turku; Department of Pediatrics, Turku University Hospital, Turku, Finland

David J. Torpy Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Christos Toumpanakis Neuroendocrine Tumour Unit, Centre for Gastroenterology, Royal Free Hospital, London, UK

Herman Tournaye Centre for Reproductive Medicine, UZ Brussels, Belgium

Peter J. Trainer Department of Endocrinology, The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Nicholas A. Tritos Neuroendocrine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

A.S. Paul van Trotsenburg Department of Pediatric Endocrinology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Engin Türkgeldi Department of Obstetrics and Gynaecology, School of Medicine, Koc University, Istanbul, Turkiye

Clara Ugolini Department of Laboratory Medicine, Section of Pathology Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy

David Unuane Department of Endocrinology, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels, Belgium

Thomas Upton University of Bristol, Bristol, UK; University of Otago, Dunedin, New Zealand

Bijay Vaidya Department of Endocrinology, Royal Devon & Exeter Hospital and University of Exeter Medical School, Exeter, UK

Jonathan Valabhji St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Juan W. Valle Department of Medical Oncology, The Christie NHS Foundation Trust; Division of Molecular and Clinical Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

Greet Van den Berghe Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven, Belgium

Wouter R. van Furth Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands

Mark P.J. Vanderpump Consultant Physician and Endocrinologist, OneWelbeck Endocrine and Diabetes Partners, The Physicians' Clinic, London, UK

Dirk Vanderschueren Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven; Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium

Irina A. Vasilevskaya Department of Cancer Biology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Flora Veltri Endocrine Unit Centre Hospitalier Universitaire Saint Pierre; Université Libre de Bruxelles (ULB), Brussels, Belgium

Bruno Vergès Endocrinology-Diabetology Department, University-Hospital and INSERM, Medicine University; Service Endocrinologie, Diabétologie et Maladies Métaboliques CHU-Dijon, Dijon, France

Russell M. Viner Great Ormond Street Institute of Child Health, University London College Hospital, London, UK

W. Edward Visser The Rotterdam Thyroid Center and Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

Paolo Vitti Department of Clinical and Experimental Medicine, Endocrinology Unit I, University of Pisa, Italy

Robert Volpé† Department of Medicine, The Wellesley Hospital, Toronto, Canada

Mark Walker Institute of Cellular Medicine, Newcastle University, UK

Marc Walter University of Basel, Psychiatric Hospital, Division of Substance Use Disorders, Switzerland

Nicholas J. Wareham MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK

Leonard Wartofsky Department of Medicine, Washington Hospital Center, Washington, DC, USA

John A.H. Wass Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Laura M. Watts Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Susan M. Webb Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona. Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Anthony P. Weetman Faculty of Medicine, Dentistry, and Health, The University of Sheffield Medical School, Sheffield, UK

Ralf Werner Division of Paediatric Endocrionolgy and Diabetes, Department of Paediatric and Adolescent Medicine, and Institute for Molecular Medicine, University of Lübeck, Germany

Philip J. Weston Diabetes Centre, Royal Liverpool University Hospital, Liverpool, UK

Clare Whicher Research and Development Department, Tom Rudd Unit, Moorgreen Hospital, Southampton, UK

Billy White Great Ormond Street Institute of Child Health, University London College Hospital, London, UK

Michael G. White Institute of Cellular Medicine, Newcastle Medical School, Newcastle University, Newcastle Upon Tyne, UK

Michael P. Whyte Center for Metabolic Bone Disease and Metabolic Research, Shriners Hospital for Children, and Division of Bone and Mineral Diseases, Departments of Medicine, Pediatrics, and Genetics, Washington University School of Medicine, St Louis, Missouri, USA

Wilmar M. Wiersinga Department of Endocrinology and Metabolism, University of Amsterdam, Amsterdam, The Netherlands

Sarah Wild Usher Institute, University of Edinburgh, Edinburgh, UK

John P. Wilding Department of Obesity and Endocrinology, University of Liverpool, UK

Graham R. Williams Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Bryan Williams University of Leicester School of Medicine, and Leicester Blood Pressure Clinic, University Hospitals of Leicester NHS Trust, Leicester, UK

Peter Winocour ENHIDE, East and North Herts Institute of Diabetes and Endocrinology, East and North Hertfordshire NHS Trust, UK

Douglas Wiseman Endocrine Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Gary Wittert Professor and Head, Discipline of Medicine, Director, Freemasons Foundation Centre for Men's Health, University of Adelaide, Senior Consultant Endocrinologist, Royal Adelaide Hospital, Adelaide, Australia

F. Susan Wong Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

Rohana J. Wright Department of Diabetes and Endocrinology, St John's Hospital, Livingston, UK

Jing Wu Department of Endocrinology, Xiangya Hospital, Central South University, Changsha, China

Swaytha Yalamanchi Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, The Johns Hopkins University, Baltimore, Maryland, USA

Jennifer M. Yamamoto Department of Medicine, Division of Endocrinology, University of Calgary, Alberta, Canada

Shunichi Yamashita Faculty of Medicine, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

Ephia Yasmin Consultant Gynaecologist, sub-specialist in Reproductive Medicine, Institute for Women's Health, University College London, UK

Bu B. Yeap Professor, The Medical School, University of Western Australia, Endocrinologist, Department of Endocrinology and Diabetes, Fiona Stanley Hospital, Perth, Western Australia

Lamis Yehia Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Bulent Okan Yildiz Division of Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, Hacettepe University, Turkey

Dorina Ylli Faculty of Medicine, University of Tirana, Radiology and Clinical Semeiotics, Tirana, Albania

Xuefei Yu Institute of Cellular Medicine, Newcastle University, UK

Sagen Zac-Varghese ENHIDE, East and North Herts Institute of Diabetes and Endocrinology, East and North Hertfordshire NHS Trust, UK

Maria Christina Zennaro INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité; Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Service de Génétique, Paris, France

Michael B. Zimmermann The Human Nutrition Laboratory, ETH Zürich, Switzerland

Michael Zitzmann University Clinics of Muenster, Center for Clinical Andrology, Domagkstrasse, Münster, Germany

Nitash Zwaveling-Soonawala Department of Pediatric Endocrinology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Free personal online access for five years

Individual purchasers of this book are also entitled to free personal access to the online edition for five years via oxfordmedicine.com/otendocrinology3. Please refer to the access token for instructions on token redemption and access.

Accessing this content online allows you to print, save, cite, email, and share content; download high-resolution figures as PowerPoint® slides; save often-used books, chapters, or searches; annotate; and quickly jump to other chapters or related material on a mobile-optimised platform.

SECTION 1

Principles of Basic and Clinical Endocrinology

1.1

Endocrine Practice Fundamentals

Lynn Loriaux

Introduction

The 'essential' theme of the physician's life is the commitment to expend time in a focused effort to improve the life of patients (πάσχειν—paskhein—to suffer, sufferer). This is a tall order. Many forces come to bear in the effort, forces that can hinder or facilitate the process. Understanding these forces is essential for the complete physician: fear, trust, knowledge, and threat.

Fear

It is not an exaggeration to say that all people fear the doctor. It starts in childhood when going to see the doctor meant pain and separation. It is still that way. How can we manage fear?

In my transition between the Brigham Hospital and the National Institutes of Health (NIH), I had decided that medicine has too many shibboleths, like the white coat, which frighten people. I decided that I would dress like any average person at work. I began to wear a Pendleton shirt in place of the white coat. Nobody seemed to care. A few months later a head nurse corralled me and dragged me into the clean utility room.

'What is the matter with you?' she asked.
'I am not sure. What do you think?'
'Your patients are afraid of you!' she said.
'What? They are not!'
'You don't look like a doctor. You look like a cowboy. Where are you from anyway?'
'New Mexico.'
'That explains it. Get rid of that shirt, get some white ones and sport jacket.'
I did what she said.

A month later, walking down the hall, she took my arm and said, 'your patients are not afraid of you now. And neither are we!'

Michael DeBakey, the famous heart surgeon, used patient expectations to fight fear in a different way. He wore only green scrub suits, everywhere he went. His scrubs were monogrammed, MED. He always wore a scrub hat and had a mask around his neck. When patients met him, it was like an audience with St. Peter or Gandhi. Fear melted into adoration. Comfort filled the room. His patients had no fear and did better than the other doctor's patients. Fear is everpresent. The best doctors learn to manage it for the good of their patients. Calling their name in the waiting room is a good start.

Trust

Without trust there could be no Western medicine. It is the bedrock of the physician–patient interaction.

It begins in the waiting room, sometimes at the bedside. The physician's eye meets the patient's eye and thus begins an association based on mutual trust and an unspoken commitment to work together to find more hours of tolerable life than would have happened without the encounter. There is a trade of some of the physician's 'good hours' for some of the patient's 'bad' ones. Both are energized by a 2500-year record of spectacular success in Western medicine. There is nothing else in life quite like this covenant. It occurs no place else in human experience. A 'hand off' of some of the labours of suffering in return for some of the burdens of uncertainty. In its train comes trust and hope.

At the core of the physician patient relationship is the Oath of Hippocrates. The patient knows that the physician will always work in the patients' best interest and can be trusted with nakedness of the flesh and nakedness of the soul. Short of catastrophe, this covenant will endure.

The physician washes hands. In this gesture the physician demonstrates a commitment to the health of the patient and to their own health. It happens at every meeting that requires touch. The barrier to touch is breached. The pulse and respirations are counted and temperature is estimated. There is an assessment of anxiety, depression, indifference, pain, and sorrow. 'What brings you here?' It is the last open-ended question the physician will ask.

One night, many years ago, I was closing the admitting clinic doors when I saw a small middle-aged woman sitting in the corner wrapped in a care worn coat of wool.

'Why are you still here?' I asked.
'To see the doctor,' she said.
'How long have you been here?' I asked.
'A long time. Nobody called me in.'
'Did you register?'
'No. They could see me.'

The nurses were in no mood for another patient, but they did their duty and got her in a gown and in an examining room. I could find nothing on physical examination. She was tired, five children, abusive husband, not enough money. I told her I could find nothing to worry about. Maybe some iron pills.

'That is all right doctor,' she said. 'I will be fine. I have iron pills at home.'
She got up to dress and I said, 'When was your last pelvic examination?'
She stopped and looked at me for what seemed a long time.
'It has been years,' she said.
'We should do one now,' I said.
The nurses were seething.
She said, 'All right.'
I left the room and the nurses set up a pelvic tray and got her into the stirrups.
When I returned, large tears were running down her cheeks.
'What is wrong?' I asked. 'Are you afraid?'
'No. I am not afraid.'
'Should we go ahead?'
'Yes.' The tears rolled on.

Before I could start the examination, I found that her vulva would not admit the entrance of even a single finger. Her pelvis was filled with an ovarian cancer that would be obvious to any physician contemplating a pelvic examination. No one had ever looked. I got up and moved to the head of the table.

'How long has it been like this?'
'A while,' she said.
'No other doctors have examined you?'
'No. They always said everything was all right down there. I just needed iron and vitamins. It did seem to help.' She wept.

She died of ovarian cancer not too many weeks hence. She knew she was sick and she was reasonably sure of what the problem was, but she trusted the doctors. If they could find nothing wrong, nothing was wrong. It gave her some relief from fear until she could tolerate it no longer and then she would go to a different doctor. She reasoned that any competent doctor would find the problem. Not so. Fear of the truth.

A complete physical examination is the single most important thing the physician can do for the patient. William Osler famously said, 'It is the responsibility of the consultant to perform the rectal examination.'

Knowledge—Docere—To Teach

Hippocrates spends the first third of the oath clarifying who can be a physician and the ever-present obligation to teach. In this case, the physician teaches the patient. Patient understanding is essential for success.

The physician explains the examination that is about to occur. When it is clear that the patient understands, the patient sits on the examination table, the physician stands up and turns the computer off. It will not be turned on again until the patients leaves. The process begins with an examination of the hair. The physician explains to the patient that in this kind of specialty examination, a thorough physical examination is of utmost importance. For completeness sake, it is helpful if the patient will get into a hospital gown. They almost never say no. This is the best time to ask if the patient would like to have a nurse with us during the examination. A family member cannot play this role. The physician steps out of the room while the patient changes into the gown. This is the time to update the medication list and verify what the patient actually is taking. It is the third most important activity of the day.

The physician must be able to see all of the patient, physically and mentally. The lemma is that 'nothing can be between the patient and the stethoscope but air'. Patients want a thorough physical examination, but they fear it. Something serious could be found bringing an end to the façade. The patient is at the point of maximum vulnerability.

There is the focused examination. The sprained ankle. There is the complete examination—head to toe. And then, there is the effective examination. The physician looks and sees, listens and hears, and touches and feels. Every patient knows intuitively what an effective examination is. Every patient deserves an effective examination, even if incomplete. At the end of the examination, the patient is asked to change back into their street clothes.

While the patient changes, the physician slips into the hall and jots down the important findings. Re-entering the examination room, the patient is asked if they would like anybody to join us in the discussion of the findings. It is important that another person participates. There is always more than one person can remember, much less understand.

After discussing the findings, you explore your plans to verify the 'working diagnosis'. You discuss various options you might use in case the first plan doesn't work out. Laboratory tests are sure to play an important role in this discussion. The physician must understand the tests that are likely to be used. The physician must know the dangers of the test, the normal ranges, the sensitivity and specificity, the positive and negative predictive values, the effect of 'time of day', fasting or fed, what medications or foods can affect the accuracy and precision of the test, how much blood will be drawn? One of the most egregious examples of the damage that ignoring laboratory details can cause is the saga of the 1 mcg Cosyntropin® stimulation test. In the early 1990s a group of investigators became interested in finding the smallest dose of Cosyntropin® that could effect secretion of cortisol from the adrenal glands. The usual dose of Cosyntropin® in this test is 250 mcg given intravenously. These investigators examined doses of 0.6, 0.8, 1.0, and 250 mcg of Cosyntropin®. They examined these doses in a group of ten normal volunteers with blood drawn at 30 minutes. The resultant data were not normally distributed so that the mean and standard deviation could not be calculated. However, the response of cortisone to the 1 mcg dose was identical to the 250 mcg dose. The investigators proposed that because of this finding the 1 mcg test could be interpreted using the 250 mcg standard curve. The Cortrosyn test performed in a test

with normally distributed data, however, shows that almost 50% of subjects are 'abnormal'. The diagnosis of a mild form of adrenal insufficiency was made. Subjects that failed this test were told that they had adrenal insufficiency and most were treated with cortisol. When the studies were repeated in assays providing a normal data distribution, all of the tested volunteers were normal. The 1 mcg test has been abandoned. The point is, that when you understand the tests that you use, this kind of disaster will not occur. You must understand all of the hormonal measurements and the provocative tests that you use on a day-by-day basis.

'Where will I find the time to do all of these things?' you ask. You will create the time by asking no open-ended questions.

Threat

Job Satisfaction

Physician job satisfaction is always highest at the beginning of a new job. As time goes by, most of the reasons that induced the physician to accept the job will change. If the change is good, it is soon forgotten. When the change is bad, it is never forgotten and frequently comes to mind. Most problems are related to patient volume, pay, patient satisfaction, and assigned clerical duties.

Patient Volume

Physicians lost control over their work volume years ago. The volume is now regulated by business people. They know how to 'optimize' your time, how to 'trim' your product. However, they do not know how to put these changes into effect.

Currently, most endocrine practices look like this: consultants see all urgent cases in the day the consultation is received. Non-urgent consultations must be seen in 3 weeks. The full patient load is seven clinics a week. Hospital-attending duties are usually added on without supplemental pay. Most practices schedule new patients with two return visits. The first visit is the diagnostic visit. The second visit is the planning visit, and the third visit is the discharge visit. The clinic sessions are usually three new patients in the morning and six return patients in the afternoon. This means that there will be alternation of six clinics per week with eight clinics per week, giving an average load of seven clinics a week. Necessary paperwork such as the electronic medical record fills the remaining time. It is never enough. This results in doing chart work from home. The workload from home never ever goes down. It continually goes up. This problem is tethered to patient volume, and can be approached from that direction, but with great difficulty. Dogma prevails.

Problems with compensation are almost always caused by failure of the 'compensation plan' to keep pace with changes in the value of the 'dollar'. Money doubles every 10 years. Compensation plans never double every 10 years. At 5 years, the physician is 'seeing' more patients for less money than when the job was negotiated. The doctor will be forced to ask for a raise. The 'compensation boss' will first try to convince you that the 'bonus' feature of the compensation plan will retain your income in a competitive range. Changes in the bonus calculation, however, will ensure that equity can never occur. The vehicle is usually an emergency tax that is essential for the 'health of the institution'. You will always be short unless you negotiate a new job in a new place at great cost. If you do, the cycle begins anew.

Next, the patient satisfaction survey. Like all great hospitals and clinics, your great institution will expect 100% patient satisfaction, and is proud of it!! The problem is that the expectation and the highest possible grade are the same, guaranteeing that no physician will ever meet the expectation of 100%. If a physician does get 100% patient satisfaction more than a few times, that physician is almost certainly pandering to patients, giving them what they want, instead of what they need. Never go to that physician, nor refer any of your patients in that direction. These doctors are usually 'the apple in the eye of the CFO' [chief financial officer]. They meet all the targets, even when it is mathematically impossible. They can be dangerous doctors, even lethal.

There was a time when business people worked for the doctors. They were charged with making medicine more profitable. Physicians paid their wages. Sometime around 1968–1969, and the advent of Medicare and Medicaid, business people were hired to manage the money, and in a few years, the doctors worked for the business people. Things have never been the same. They want us to treat every patient the same and to pay all of the doctors the same salaries, surgeons at the top, and cognitive specialist at the bottom. Now, we work for them. The quality of medicine has deteriorated. At some point in the future, the structure will collapse, and we will have a chance to rebuild along the principles of patient-centred care. We will have to move fast. They are not afraid. They believe we will not leave the bedside and hence ignore the struggle. They are right. Hope for the best!

You may attribute most of this to the 'ravings of an old man'. True, but most of the recommendations were made when I was much younger. They have survived, especially the no 'open-ended questions', and the '100% satisfaction' doctor. If all fails to help with your particular problem, there is this:

'Never be the first, nor the last, to use a new medication or technique!' If you can remember this, and do it, you will probably be alright.

1.2

Hormones and Receptors

Fundamental Considerations

John W. Funder

Background

The original endocrine physiologists viewed hormones as responses to homoeostatic challenge, any signal a call to arms; the word is thus derived from the classical Greek ωρμαειν—'to arouse'. In the twenty-first century a hormone is a molecule—small or large, protein or lipid—secreted in a regulated fashion from one organ and acting on another. The definition is firmly based on the anatomy of the seventeenth century, the histology of the nineteenth, and the physiology of the twentieth. It has been shaped by convention and clinical specialization: gut hormones are the marches between endocrinology and gastroenterology, and the adrenal medulla the territory of the cardiovascular physician. It has been refined by concepts of paracrine—where the secretion of one cell type in a tissue acts on another cell type in the same tissue—and autocrine, where a particular cell type both secretes and responds to a particular signal. Inherent in the concepts of paracrine and autocrine are that the signal is not secreted into blood or lymph, to be distributed more or less throughout the body, but is made locally to act locally. A very good example of a signalling system with both paracrine and autocrine activities is the neuronal synapse.

Inherent in the concept of the signal is that of a receptor: a signal without a receptor is the sound of one hand clapping. Inherent in the concept of a receptor are two functions: that of being able to discriminate between different signals, and to propagate the signal by activating cell membrane or intracellular signal transduction pathways. Discrimination by a receptor between different circulating potential signals is, in the first instance, a function of the likelihood of a particular signal being able to interact with the receptor, for a period of time sufficient to alter the confirmation of the receptor and thus to trigger propagation. This interaction is commonly referred to as binding, and thus the circulating hormone as a ligand (that which is bound). If the structures of ligand and receptors are such that the initial interaction is followed by formation of strong intermolecular bonds between the two, lessening the possibility of dissociation and the receptor returning to an unliganded state, the receptor is said to have high affinity for the ligand (and vice versa). If the binding is followed by propagation of the 'appropriate' signal the ligand is classified as an agonist, or active hormone; if a molecule occupies the binding site on the receptor but does not so alter its structure as to propagate a signal, it is often called a hormone antagonist (and, more accurately, a receptor antagonist). In the past couple of decades, the concepts of 'agonist' and 'antagonist' have needed to be refined, as noted subsequently in this chapter.

Hormones and Receptors: Binding

In symbols, the reversible interaction between hormone and receptor can be simply written as follows;

$$[H] \cdot [R] \underset{}{\overset{K_1}{\rightleftharpoons}} [HR] \qquad \text{(Equation 1)}$$

where [H] is the concentration of hormone, [R] the concentration of empty or unliganded receptor, and [HR] the concentration of occupied receptor (i.e. hormone-receptor complexes). The forward (to the right by convention) or association reaction is equally a function of hormone and receptor concentrations; the association rate constant [K_1] is a reflection of the likeliness of apposition/goodness of fit of hormone and receptor, reflecting their structures plus extrinsic factors such as temperature, ionic strength of the milieu, and unstirred layers (close to the receptor, where hormone is less likely to diffuse). The actual rate of the forward reaction is thus

multifactorial, a function of the rate constant, the concentration of hormone, and the concentration of receptor, or

$$\text{foreword rate (or on-rate)} = K_{-1}[H][R]. \qquad \text{(Equation 2)}$$

The dissociation of hormone-receptor complexes [HR] is driven by one thing, and one thing only, the dissociation rate constant [K_{-1}], a measure of the inherent probability of the two entities falling apart, under particular conditions of temperature, ionic strength, and so on. The actual rate of dissociation is thus the product of K_{-1}, the dissociation rate constant, and the concentration of hormone-receptor complexes, or

$$\text{reverse rate}\left(\text{or off-rate}\right) = K_{-1}[HR]. \qquad \text{(Equation 3)}$$

At equilibrium, by definition, the rates of the forward and reverse reactions are equal, i.e. for every molecule of hormone that associates with a receptor molecule, a preformed hormone-receptor complex dissociates, or

$$K_1[H][R] = K_{-1}[HR]. \qquad \text{(Equation 4)}$$

By simple rearrangement, this can be rewritten as

$$\frac{K_{-1}}{K_1} = \frac{[H][R]}{[HR]}. \qquad \text{(Equation 5)}$$

The quotient of the two rates constants (K_{-1}/K_1) is termed the dissociation constant or K_d; its reverse (K_1/K_{-1}) is the less commonly used K_a or association constant of the reaction. The key outcome of all this relatively simple mathematics is to put a value on K_d, as a measure of affinity, or overall probability of the hormone-receptor complex being in existence, as follows:

$$K_d = \frac{K_{-1}}{K_1} = \frac{[H][R]}{[HR]}. \qquad \text{(Equation 6)}$$

If we were to choose a concentration of hormone which would half saturate the receptors, then [R] would equal [HR]. Under such circumstances the two terms can be cancelled in Equation 6, and

$$K_d = [H], \qquad \text{(Equation 7)}$$

where K_d equals [H], the hormone concentration at which half maximal receptor occupancy is achieved, and which has the dimensions of concentration, that is, molar.

From Equations 1 to 7 there are a number of things that flow. First, in a simple binding system the dissociation of hormone from receptor is not accelerated by addition of excess hormone. What this does, when, for instance, 1000-fold non-radioactive hormone is added to a system containing tracer hormone-receptor complexes, is to operationally prevent (i.e. dilute 1000-fold) reassociation of tracer to receptor. Under such conditions then, the disappearance of tracer–receptor complexes over time thus provides an accurate estimate of the dissociation rate. There are receptors that oligomerize: in such circumstances binding of ligand can increase or decrease the affinity of the other binding sites for hormone, termed positive and negative cooperativity, respectively. Dissociation of bound tracer, for instance, is accelerated in systems displaying negative cooperativity.

Secondly, dissociation constants can only be derived from equilibrium studies, that is, those in which the rates of forward and backward reactions are equal. The association rate constant and dissociation rate constant are often very different, and are constant for a given set of physical circumstances; the actual rates of association and dissociation are determined by not just these constants, but also by the concentration of reactants, as noted earlier. Where this concept of equilibrium comes into play is in situations where binding is covalent, or essentially irreversible; under such circumstances Scatchard analysis, for example, is inappropriate for determining K_d. A practical case in point is triamcinolone acetonide (TA), a powerful synthetic glucocorticoid in clinical use, which (in contrast with dexamethasone or the physiological glucocorticoids) requires approximately 24 h to come into equilibrium in glucocorticoid receptor binding systems *in vitro* at 4 C; exposure for shorter time points will consistently underestimate the affinity of TA for the glucocorticoid receptor. Third, different binding systems respond differently to changes in physical conditions. Cortisol, for example, binds transcortin with an order of magnitude higher affinity at 4°C than at 37°C, across a number of species, and shows clear differences in transcortin binding at physiologically relevant temperatures. In contrast, cortisol binding to glucocorticoid receptors is not particularly temperature dependent, but if anything is of a higher affinity at physiological than at lower temperatures.

Finally, there is the inherent bias of endocrinology, that of seeing high-affinity binding as good ('binds well to the receptor . . .'), and lower affinity binding as less good ('binds poorly . . .'). The underpinnings of this bias are twofold, one theoretical and the other practical. Practically, particularly in often unstable broken cell preparations, the absence of high-affinity binding equates to experimental failure, a powerful driver of emotive language. Even if no experiment ever failed, however, an endocrinologist's bias is to regard high-affinity binding as good, for the following reason. The higher the affinity the lower the concentration of signal required to half-maximally occupy, and, other things being equal, activate the 'cognate' receptor. There are two consequences of this, one of which appears to be biologically sound, the other less so. The latter is a notion of economy; that it is better for an organ to make less rather than more signal, in that it poses less of a demand on precursors and metabolism. This is experientially not the case; every molecule of thyroglobulin, with a molecular weight in excess of 600 000 yields 4–16 molecules of thyroxine, at first sight an example of conspicuous biological extravagance. The other concept underlying the bias has more biological purchase, in that the higher the concentration required to activate cognate receptors, the more likely is the hormone to cross-react with other receptors, acting as an agonist or antagonist, and thus reducing the specificity of the signalling system. It is, of course, entirely possible that there have evolved circumstances in which such 'cross-reactivity' may reflect physiology, and that our bias is Ockham's razor cutting too close to the bone: on the whole, however, such a degree of cautious reductionism appears justified.

Hormones and Neurotransmitters

In contrast with the previous discussion, if we take a broader biological view that low-affinity binding can be 'good'—for example, when it enables rolling of platelets or leucocytes on endothelium, giving them time to 'sniff the wind' in terms of damage or inflammation. It is also not only advantageous, but functionally required, within the nervous system where low-affinity binding of signal to receptor is a necessity for the time constants of neurotransmission.

When the electrical impulse underlying nerve conduction is translated into a chemical signal at a synapse or neuroeffector junction, minute quantaties of neurotransmitter are released. Because the space into which the neurotransmitter is released is even more minute, the concentration of neurotransmitter becomes very high, so that receptors are rapidly occupied and activated. To achieve this, the 'on-rate' of neurotransmitter-receptor binding must be very rapid; and the off-rate (in contrast with hormone-receptor interactions) must also be very rapid, to enable the receptor to return to ground zero. Signal is rapidly cleared by reuptake, diffusion, and metabolism, so that quantal release of signal is followed essentially stochastically by a single response.

To achieve this rapid-onset-rapid-offset binding and activation by neurotransmitters, receptors have to be low affinity, to allow the time constants that characterize neurotransmission. The nervous system does it by mass, 'brute-forcing' occupancy of low-affinity receptors, with a restricted spatial distribution of the mass of signal to allow the very high concentrations required, and very efficient mechanisms of rapidly reducing signal concentration. Reflecting this difference, hormones have time constants of minutes, hours, and days compared with the nervous system's milliseconds; the endocrine system sacrifices time to allow its signals to be distributed all over the body, to 'arouse' the diversity of cells that express receptors to which the particular signal can bind. Its signals are broadcast like radio, in contrast with the nervous system landline telephone network.

One striking anthropomorphic illustration of this difference may be worth a thousand words of theoretical justification [1]. First, picture a hummingbird in the *National Geographic*, its wings still blurred despite shutter speeds of 1/500 or 1/1000 of a second. If acetylcholine had the same high affinity for its receptors at the neuromuscular junction as progesterone has for progesterone receptors, then a hummingbird could beat its wings twice a minute—aerodynamically challenging and clearly no evolutionary advantage. Even less of an evolutionary advantage accrues if progesterone receptors had the same affinity for progesterone as cholinergic receptors for acetylcholine. Unless the efficiency of steroidogenesis were vastly improved, the placenta would need to be considerably larger: to maintain plasma progesterone at the levels required, other things equal, it would need to be the size of a 14 cubic foot refrigerator. Other evolutionary considerations would be the 9 months of somnolence that such levels of progesterone would almost certainly produce, difficult to reconcile with the additional 25 000 calories per day required to maintain the requisite levels of progesterone biosynthesis required.

Mineralocorticoid Receptors: A Case Study

We have mercifully evolved otherwise, and evolution has exploited a range of interactions between signals and receptors in terms of growth, development, homoeostasis, and cognition. Sometimes we can second-guess nature, perhaps to our own disadvantage in terms of realizing our own physiology.

One example, within the author's area of experience, is that of the mineralocorticoid receptor. Mineralocorticoid hormones were defined in 1961 by Jean Crabbé as promoting unidirectional transepithelial sodium transport [2], a definition that has stood the test of time. The principal mineralocorticoid hormone, aldosterone, is secreted from the zona glomerulosa of the adrenal cortex in response to elevated plasma potassium concentrations, increased levels of angiotensin II, or acutely, adrenocorticotropic hormone (ACTH). In response to sodium deficiency, volume depletion, or potassium loading, aldosterone incontestably acts via mineralocorticoid receptor in kidney and colon, salivary gland and sweat gland to retain sodium, and excrete potassium thus acting as a classic homoeostatic hormone. And yet . . .

When human mineralocorticoid receptors were first cloned [3], the highest levels of mRNA were found in the hippocampus, not a classical site of aldosterone action, and recapitulating earlier binding studies on rat tissue extracts [4]. Second, in both studies, mineralocorticoid receptors were shown to have equivalent affinity for the physiological glucocorticoids (cortisol, corticosterone) as for aldosterone, raising obvious questions of how aldosterone ever occupies epithelial mineralocorticoid receptors, given the orders of magnitude for higher circulating concentrations of glucocorticoids.

The answer to this question appears to be the coexpression, in epithelial tissues, of the enzyme 11β-hydroxysteroid dehydrogenase [5, 6], which converts cortisol and corticosterone to their inactive 11-keto metabolites cortisone/11-dehydrocorticosterone. Aldosterone is not similarly metabolized, because its signature aldehyde group at C18 cyclizes with the hydroxyl at C11, forming a stable hemiacetal which is not susceptible to enzyme attack by 11β-hydroxysteroid dehydrogenase 2.

The enzyme is expressed at high abundance in aldosterone target cells ($3–4 \times 10^6$ molecules/cell), and its operation—by metabolizing glucocorticoids and clearly by other mechanisms [7, 8]—appears sufficient to confer aldosterone selectivity on the epithelial mineralocorticoid receptor. When it is congenitally deficient, as in the autosomal recessive syndrome of apparent mineralocorticoid excess [9], cortisol activates epithelial mineralocorticoid receptors, leading to uncontrolled sodium retention and severe hypertension.

The enzyme 11β-hydroxysteroid dehydrogenase 2 is not found in non-epithelial tissues in which mineralocorticoid receptors are expressed at high (hippocampus) or modest (heart) abundance, and which aldosterone thus has little chance of occupying. An inescapable corollary of the last sentence is that such mineralocorticoid receptors are physiologically high-affinity glucocorticoid receptors.

A second, quite different mechanism modulating the activity of mineralocorticoid receptors has recently been demonstrated [10]. Classically, mineralocorticoid receptors were believed to be confined to principal cells in the tubule; in 2013 Shibata and his colleagues

showed that renal intercalated cells also express mineralocorticoid receptors, normally held inactive by phosphorylation of serine 843 in the ligand binding site. In volume depletion, however, angiotensin dephosphorylates the receptors, which can then be activated by aldosterone or by cortisol. Their response to glucocorticoid underpins the difference between the aldosterone synthase null mouse and the mineralocorticoid receptor mouse [11]. The latter does not survive sodium restriction [12], whereas the former does [13]. Although a complete tissue scan is yet to be done, to date this mechanism appears confined to the renal intercalated cell.

Hormones and Receptors: Evolutionary Considerations

In the syndrome of glucocorticoid remediable aldosteronism ([14]: now termed familial hyperaldosteronism type-1), aldosterone is secreted primarily in response to adrenocorticotrophic hormone (ACTH), with aldosterone synthase activity expressed throughout the adrenal cortex. The underlying genetic defect is a chimeric gene in which the 5′ end of the gene for 11β-hydroxylase is fused with the 3′ end of the gene coding for aldosterone synthase. This can happen because the two parent genes lie next to one another, on chromosome 8, and because they are 94% identical in terms of nucleotide sequence. What the condition reflects is the product of an unequal crossing over at meiosis in an ancestral gamete, reflecting the relatively small misalignment required (gene proximity) and the possibility of realignment (sequence homology). In evolutionary terms, however, what the condition illustrates is the probability that the two genes (for 11β-hydroxylase and aldosterone synthase) share a relatively recent ancestor, and that their degree of identity and juxtaposition represent a relatively recent gene duplication event. Compare this with the gene coding for the mineralocorticoid receptor (chromosome 4) and the glucocorticoid receptor (chromosome 5).

Mineralocorticoid receptors and glucocorticoid receptors have one area of high (about 90%) sequence identity, the DNA-binding domain, and another of considerable homology, the ligand binding domain, with 57% identity: the majority of the two molecules, including major activation domains, have minimal (less than 15%) identity. It would thus appear that the mineralocorticoid receptor and glucocorticoid receptor are rather more evolutionarily distant than are the enzymes 11β-hydroxylase and aldosterone synthase. Although classically mineralocorticoid and glucocorticoid receptors were thought to share a common immediate 'corticoid' receptor ancestor [15], more recently evidence has emerged for mineralocorticoid receptors being the first of the mineralocorticoid/glucocorticoid/androgen/progestin receptor subfamily to branch off [16].

In evolutionary terms aldosterone is thus a Johnny-come-lately, pressed into service as organisms became amphibious, to activate a pre-existing high-affinity glucocorticoid receptor (which we now term the mineralocorticoid receptor). Mineralocorticoid receptor selectivity in epithelial aldosterone target tissues is produced by coexpression of the enzyme 11β-hydroxysteroid dehydrogenase 2 at high abundance, and the integrity of a system for Na^+ retention out of seawater obtained by the expression of aldosterone synthase being yoked to surrogates of Na^+ deficiency (angiotensin II, K^+) rather than primarily to the brain hormone ACTH. To call aldosterone the cognate ligand for mineralocorticoid receptor—and the ascription 'mineralocorticoid receptor' itself—is thus understandable in terms of our historical knowledge of aldosterone, but it fails to recognize the previous, and current, physiological roles for mineralocorticoid receptors net of aldosterone. The rainbow trout, for instance, does not synthesize aldosterone. In an attempt to clone rainbow trout androgen receptors, an rtMR sequence was identified, related to rtGR but with much higher identity with mammalian mineralocorticoid receptors [17]. Its physiologic role(s), like the pathophysiologic roles of mammalian non-epithelial mineralocorticoid receptors, await exploration.

A final fundamental consideration might thus be as follows. There are currently 49 members of the extended steroid/thyroid/retinoid/orphan receptor superfamily of ligand activated transcription factors in the human genome, evidence for enormous evolutionary scope and flexibility. One might thus be pardoned for asking why a 'specific' mineralocorticoid receptor did not evolve, responsive to a ligand with levels inversely related to Na^+ status, rather than the complicated system of highly reactive C18 aldehyde groups and epithelial 11β-hydroxysteroid dehydrogenase 2. This is in fact an impertinent question, bluntly put this way: what is the appropriate question to ask, is where is the evolutionary gain in the system developing as it did?

Receptor Activation, Receptor Blockade

For aldosterone and mineralocorticoid receptors, the past decade has provided more questions than answers. Among the latter, for the hormone, is the acceptance that aldosterone can have both genomic and acute, non-genomic effects [18], and that most but probably not all such rapid effects are via the classic mineralocorticoid receptor [19]. In addition, there is now general consensus that the syndrome of primary aldosteronism represents ≥ 10% of all 'essential hypertension' [20], and that such patients show higher cardiovascular morbidity and mortality than age-, sex-, and blood pressure-matched patients with essential hypertension [21]. For the receptor, the RALES, EPHESUS, and 4E trials [22–24] have shown the beneficial effects of mineralocorticoid receptor blockade in heart failure and essential hypertension. The functions and roles of non-epithelial mineralocorticoid receptors, constitutively (90–99%) occupied by glucocorticoids, have hardly begun to be properly addressed. The mechanisms whereby the physiological glucocorticoids show bivalent activity when bound to mineralocorticoid receptors—normally antagonist, but agonist (in the sense of mimicking aldosterone) in the context of redox change (11β-hydroxysteroid dehydrogenase 2 blockade, reactive oxygen species generation [7, 8]) similarly remain to be established.

In fact, the terms agonist and antagonist need to be seen for what they are—effector definitions, like that proposed for mineralocorticoids almost half a century ago by Jean Crabbé. For most hormone-receptor systems, the last 20 years has seen the growing emergence of tissue selective agents, agonist in some organs, antagonist is others. While most microarray analyses have provided a formidable list of genes, expression of which is doubled or halved by a classical agonist, similar lists can be complied for classical antagonists. Some classical antagonists (e.g. spironolactone for epithelial

mineralocorticoid receptors), demonstrate inverse agonist activity in experimental myocardial infarction [25]. Aldosterone and cortisol aggravate the infarct area; spironolactone at low (EC50 3–5 nm) concentration reduces the infarct area, in the absence of any other steroid.

Spironolactone thus has its 'antagonist' effects in the context of cardiac damage not just by competing with agonist steroids for occupancy of mineralocorticoid receptors, but by inducing expression of protective genes and lowering that of proapoptotic genes. It does this at relatively low concentrations, evidence that not all mineralocorticoid receptors, or even a majority, need to be occupied for such an effect. Even before the advent of microarray, it was clear that the effects on enzyme induction, for example, in cultured cells could show distinct dose–response curves, evidence for maximal effects on some readouts at submaximal receptor occupancy. This has not been widely incorporated into consideration of the clinical roles of aldosterone and mineralocorticoid receptors. An example of the former is the demonstration that relatively mild elevations of aldosterone in primary aldosteronism, which would have minimal incremental effects on non-epithelial receptor occupancy, are accompanied by demonstrable cardiovascular damage [26], even in the absence of an elevated blood pressure [27], compared with age-, sex-, and blood pressure-matched controls. An example of the latter would appear to be the otherwise curiously low dose (x = 26 mg/day) of spironolactone which, when added to standard care, produced a remarkable 30% improvement in survival, and 35% lower hospital admission rate, in the RALES trial [22].

ENV0I

Given the achievements of the human genome project, we are faced with a mass of information of daunting proportions. This brief chapter has attempted to raise questions, and thus help shape the mindsets of those who face the exciting but very challenging task of reconciling the enormity of information with the demands of clinical endocrinology, from individual patients through populations. For a chance of success, we need a degree of comfort with the underlying mathematics, the biology, and as best we can guess from the historical record, the evolution.

REFERENCES

1. Funder J. Receptors, refrigerators and hummingbirds. *News in Physiological Sciences*, 1987; **2**: 231–2
2. Crabbé J. Stimulation of active sodium transport by the isolated toad bladder with aldosterone *in vitro*. *J Clin Investig*, 1961; **40**: 2103–10.
3. Arriza JL, Weinberger C, Cerelli G, *et al.* Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocortiod receptor. *Science*, 1987; **237**: 268–75.
4. Krozowski ZS, Funder JW. Renal mineralocorticoid receptors and hippocampal corticosterone-binding species have identical intrinsic steroid specificity. *Proc Natl Acad Sci USA*, 1983; **880**: 6036–60.
5. Funder JW, Pearce P, Smith R, Smith AL. Mineralocorticoid action: target-tissue specificity is enzyme, not receptor, mediated. *Science*, 1988; **242**: 583–5.
6. Edwards CR, Stewart PM, Burt D, *et al.* Localisation of 11 beta-hydroxsteriod dehydrogenase–tissue specific protector of the mineralocorticoid receptor. *Lancet*, 1988; **2**: 986–9.
7. Funder JW, Myles K. Exclusion of corticosterone from epithelial mineralocorticoid receptors is insufficient for selectivity of aldosterone action: *in vivo* binding studies. *Endocrinology*, 1996; **137**: 5264–8.
8. Funder JW. Is aldosterone bad for the heart? *Trends Endocrinol Metab*, 2004; **15**: 139–42.
9. Wilson RC, Krozowski ZS, Li K, *et al.* A mutation in the HSD11B2 gene in a family with apparent mineralocorticoid excess. *J Clin Endrocrinol Metab*, 1995; **80**: 2263–6.
10. Shibata S, Rinehart J, Zhang J, *et al.* Mineralocorticoid receptor phosphorylation regulates ligand binding and renal response to volume depletion and hyperkalemia. *Cell Metab*, 2013; **18**: 660–71.
11. Funder JW. Angiotensin retains sodium by dephosphorylating mineralocorticoid receptors in renal intercalated cells. *Cell Metab*, 2013, **18**: 609–10.
12. Berger S, Bleich M, Schmid W, *et al.* Mineralocorticoid receptor knockout mice: pathophysiology of Na+ metabolism. *Proc Natl Acad Sci U S A*, 1998; **95**: 9424–9.
13. Makhanova N, Sequeira-Lopez MLS, Gomez RA, Kim H-S, Smithies O. Disturbed homeostasis in sodium restricted mice heterozygous and homozygous for aldosterone synthase gene disruption. *Hypertension*, 2006; **48**: 1151–9.
14. Lifton RP, Dluhy RG, Powers M, *et al.* A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocortiod-remediable aldosteronism and human hypertension. *Nature*, 1992; **355**: 262–5.
15. Mangelsdorf DJ, Thummel C, Beato M, *et al.* The nuclear receptor superfamily: the second decade. *Cell*, 1995; **83**: 835–9.
16. Kassahn KS, Ragan MA, Funder JW. Mineralocorticoid receptors: evolutionary and pathophysiological considerations. *Endocrinology*, 2011; **152**:1883–90.
17. Colombe L, Fostier A, Bury N, Pakdel P, Gurgen YA. Mineralocorticoid receptor in the rainbow trout, *Oncorhynchus Mykiss,* cloning and characterization of its steroid binding domain. *Steroids*, 2000; **65**: 319–28.
18. Funder JW. The nongenomic actions of aldosterone. *Endocrine Reviews*, 2005; **26**: 313–21.
19. Ruhs S, Nolze A, Hubschmann R, Grossmann C. 30 years of the mineralocorticoid receptor: nongenomic effects via the mineralocorticoid receptor. *J Endocrinol*, 2017; **234**: T107–24.
20. Funder JW, Carey RM, Mantero F, *et al.* The management of primary aldosteronism: case detection diagnosis and treatment: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 1889–916.
21. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Renal outcomes in Medically and surgically treated primary aldosteronism. *Hypertension*, 2018; **72**: 658–66.
22. Pitt B, Zannad F, Remme WJ, *et al.* The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *N Engl J Med*, 1999; **341**: 709–17.
23. Pitt B, Remme W, Zannad F, Neaton J, Martinez F, Roniker B, *et al.* Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N Engl J Med*, 2003; **348**: 1309–21.
24. Pitt B, Reichek N, Willenbrock R, *et al.* Effects of eplerenone, enalapril, and eplerenone/enalapril in patients with essential hypertension and left ventricular hypertrophy: the 4E-left ventricular hypertrophy study. *Circulation*, 2003; **108**: 1831–8.

25. Mihailidou AS, Loan Le Ty, Mardini M, Funder JW. Glucocorticoids activate cardial mineralocorticoid receptors during experimental myocardial infarction. *Hypertension*, 2009; **54**: 1211–12.
26. Milliez P, Girard X, Plouin P-F, Blacher J, Safar ME, Mourad J-J. Evidence for an increased rate of cardiovascular events in patients with primary aldosteronism. *J Am Coll Cardiol*, 2005; **45**: 1243–8.
27. Stowasser M, Sharman J, Leano R, *et al*. Evidence for abnormal left ventricular structure and function in normotensive individuals with familial hyperaldosteronism type 1. *J Clin Endocrinol Metab*, 2005; **90**: 5070–6.

1.3

Molecular Aspects of Hormone Regulation

Kenneth Siddle and Gemma V. Brierley

Introduction

Hormones are chemically and structurally diverse. Most are either small proteins (e.g. insulin, glucagon), shorter peptides derived from precursor proteins (e.g. CRH, melanocortins), cholesterol-derived steroids (e.g. testosterone, cortisol), amino acid derivatives (e.g. catecholamines, thyroid hormones), or fatty acid derivatives (e.g. eicosanoids). All bind highly selective or specific receptors, either on the surface, in the cytoplasm, or in the nucleus of target cells to elicit biological responses. Hormone receptors enable cells to receive information and to process, decode, and respond to that information. Decades of research has established that information is transmitted from many hormone receptors via signalling 'cascades' of sequential enzymic reactions, giving rise to the widespread notion of 'signalling pathways'. It has progressively become clear, however, that specific biological effects often do not arise solely through linear information transfer, but rather involve complex interactions among multiple pathways, so that many responses are better understood in terms of signalling 'networks'. Distinct biological outcomes result from integration of protein–protein interactions, modulation of enzymatic activity, sometimes generating 'second messengers', multiple post-translational protein modifications including phosphorylation, and changes in RNA and protein biosynthesis. These processes are all subject to temporal, spatial, and threshold constraints. Similar mechanisms govern the actions of growth and differentiation factors and cytokines that act in a paracrine or autocrine manner. This chapter gives a brief overview of molecular mechanisms involved in cellular responses to hormones, using examples to illustrate the underlying principles and complexities (see **Box 1.3.1**).

Classes of Hormone Receptor

Nuclear Hormone Receptors

Steroid and thyroid hormones, which are lipophilic, cross cell membranes by passive diffusion, by facilitated diffusion via membrane-bound transport proteins, or by other mechanisms which continue to be elucidated. Once inside target cells, they bind intracellular receptors. These are either constitutively located within the nucleus (e.g. thyroid hormone receptors) or translocated to the nucleus following hormone binding in the cytosol (e.g. sex hormone receptors). This class of receptors serve as ligand-activated transcription factors, acting in concert with other proteins that include so-called corepressors and coactivators. Receptors bind via zinc-finger motifs to hormone response elements (HREs) within enhancer and promoter regions of genes. Ligand binding causes a conformational change that alters the repertoire of proteins interacting with the receptor to drive gene transcription. The pattern of gene expression that ensues is not stereotyped, with different hormones or ligands regulating distinct but overlapping sets of genes even when acting on the same receptor in the same cell. Cellular context is also important, with differential expression of the proteins that interact with receptors, and differential accessibility of HREs due to chromatin conformation modulating transcriptional responses to receptor activation. Although the direct transcriptional actions of nuclear hormone receptors bound to HREs are best studied, they are also capable of indirect actions that do not require the receptors themselves to be DNA bound, but are mediated through interactions with other transcription factors. Nuclear hormone receptors may also exert fast-acting non-genomic actions via initiation of second messenger cascades similar to those discussed next for cell surface receptors.

Subclassification of Nuclear Receptors

Nuclear hormone receptors are grouped into subtypes based on cellular location, internal structure of cognate HREs, and mechanism of their action on gene transcription. Type I hormone receptors, such as those for androgens and oestrogens, reside in the cytoplasm,

Box 1.3.1 'Call out' box

Glucose transporter 4 (GLUT4) storage vesicle (GSV): insulin-responsive storage and transport vesicle containing glucose transporter 4 protein.

GTPases: proteins that switch between GTP-bound active form to GDP-bound inactive form by hydrolysis of GTP via their intrinsic GTPase activity.

GTPase-activating proteins (GAPs): enzymes that inactivate GTPases by catalysing GTP hydrolysis.

Guanine nucleotide exchange factors (GEFs): enzymes that activate GTPases by catalysing GDP release followed by GTP binding.

Histones: proteins that package DNA into chromatin.

Histone acetyltransferase (HAT): enzyme that transfers acetyl groups to histones to relax the structure of chromatin.

Histone deacetylase (HDAC): enzyme that removes acetyl groups from histones to condense chromatin.

Hormone response element (HRE): a short sequence of DNA that binds specific nuclear hormone receptors.

Nuclear localization sequence (NLS): an amino acid sequence within a protein that identifies the protein for import into the nucleus.

Phosphatase: enzyme that catalyses the removal of phosphate groups from proteins.

Phosphotyrosine-binding (PTB) domain: protein domain that binds phosphorylated tyrosine residues in proteins.

Pleckstrin homology (PH) domain: protein domain that (usually) binds phosphoinositol lipids within membranes.

(Protein) kinase: enzyme that catalyses the transfer of phosphate groups to serine, threonine, or tyrosine residues on proteins.

Src homology 2 (SH2) domain: protein domain that binds phosphorylated tyrosine residues in proteins and differs from PTB domains.

where they are sequestered by binding to chaperone proteins. Ligand binding by a type I receptor causes conformational changes within the receptor leading to its dissociation from the chaperone protein and subsequent homodimerization. This produces further conformational changes exposing a nuclear localization sequence (NLS) that permits entry into the nucleus. Hormone-bound type I receptors in the nucleus bind palindromic inverted repeat HREs and associate with multiprotein complexes called transcriptional coactivators to increase target gene transcription [1]. Type III hormone receptors function similarly to type I receptors, but recognize HREs with a directly repeated sequence.

Type II hormone receptors reside within the nucleus bound to HREs with direct repeat sequences, usually as heterodimers with the retinoid X receptor (RXR). The spacing between the HRE direct repeats confers receptor specificity [2]. Unliganded type II hormone receptors, including thyroid hormone receptors (THR), are bound to HREs near the transcriptional start site (TSS) of target genes (e.g. growth hormone, GH) and act as transcriptional repressors in the basal state through association with multiprotein corepressor complexes (Figure 1.3.1). These corepressor complexes include histone deacetylases that remove acetyl side groups from histones, favouring chromatin condensation, limiting transcription factor access, and repressing transcription [2]. Binding of T_3 to THR promotes dissociation of the corepressor complex and recruitment of a coactivator complex. This opposes the action of corepressor complexes, favouring acetylation of histones and thus access to the TSS for core transcription factors and RNA polymerase (Figure 1.3.1).

Cell Surface Hormone Receptors

Protein, peptide, catecholamine, and some eicosanoid hormones, in contrast to steroid hormones, elicit cellular responses by binding to and activating receptors on the cell surface. Ligand binding to a receptor induces a conformational change in the receptor that transmits information across the plasma membrane lipid bilayer to the intracellular portion of the receptor, ultimately initiating relays of intracellular signals, or signalling 'cascades'. Termination of such

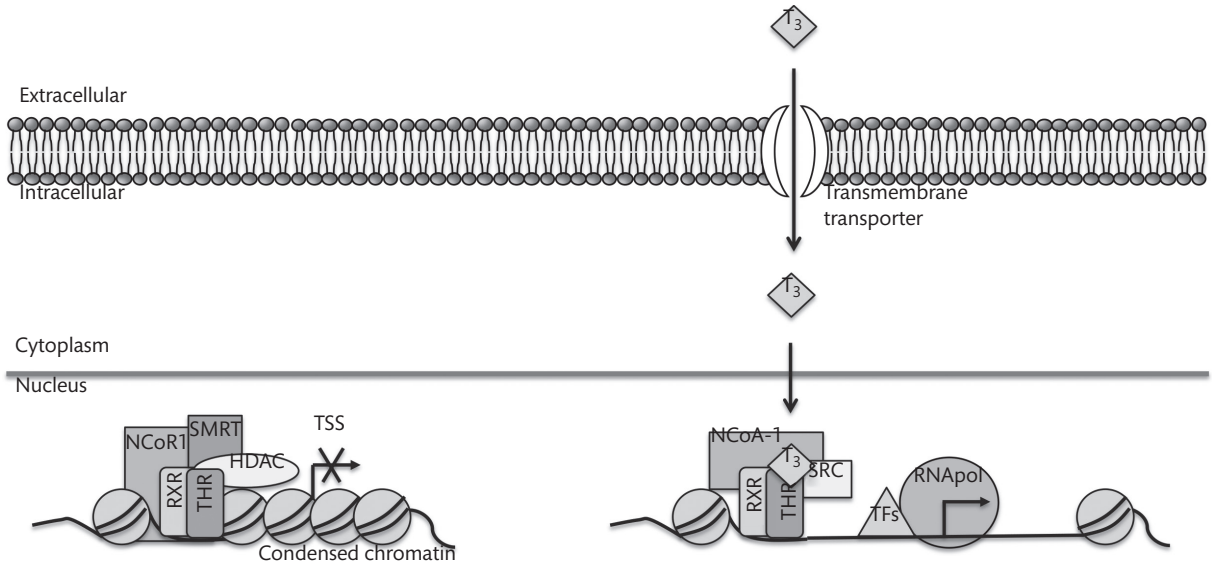


Figure 1.3.1 Triiodothyronine (T_3)-mediated regulation of growth hormone gene transcription. T_3 binds the thyroid hormone receptor THR in the nucleus where it is bound to the hormone-responsive element in a heterodimer with RXR. This changes the conformation of THR allowing coactivators to bind, remodelling surrounding chromatin to allow transcription factors (TFs) and RNA polymerase to access the transcriptional start site (TSS) and leading in turn to expression of target genes such as GH. NCoR1, nuclear receptor corepressor-1; SMRT, silencing mediator for retinoid and thyroid receptor (SMRT); HDAC, histone deacetylases; SRC, steroid-receptor coactivator family; NCoA-1, nuclear coactivator 1 (NCoA-1).

signalling is essential if the receptor is to continue to serve a useful information transfer function, and so signalling is almost invariably transient, being terminated by enzymatic reactions and/or internalization of the activated receptor into the endosomal/lysosomal system. Cell surface receptors have diverse mechanisms of action, but these usually involve either the generation of intracellular small molecule 'second messengers', or the activation of protein kinase activity acting on intracellular substrates. This section will give an overview of the principles and general paradigms underlying intracellular signalling networks before focusing on the receptor families.

Covalent Modifications in Signalling Cascades

One major mechanism whereby activation of a hormone receptor propagates signalling inside a cell is through sequential activation of enzymes through their covalent modification. The most common and best studied such modification is phosphorylation, or addition of a phosphate group by kinase enzymes to amino acids with suitably receptive side chains, namely serine, threonine, or tyrosine. This often induces a conformational change that either disrupts or creates protein–protein interactions, or may create a new protein binding site without major conformational change, which in turn leads to activation of the target enzyme or transcription factors. Often, a sequence of several enzymes in a signalling cascade, each activating the next one specifically, serves to amplify the initial signal within the cell, and also provides multiple points for regulation and modulation of the response.

Although phosphorylation is the best studied so-called post-translational modification (PTM) in hormone signalling pathways, many other forms of PTM of target proteins are also important. These include other chemical modifications such as acylation, methylation, and acetylation. For example, the covalent modification of histones by methylation and acetylation is involved in the epigenetic regulation of gene expression by both altering chromatin structure and allowing the recruitment of other histone modifiers. Other PTMs are protein based and involve the attachment of the 76 amino acid protein ubiquitin in a process known as ubiquitylation. This can be extended to form polyubiquitin chains and the ubiquitin within these chains can be further modified by small ubiquitin-like (Ubl) modifier (SUMO) family proteins, acetylation, and phosphorylation to create a diverse signalling cipher known as the 'ubiquitin code'. The biological outcomes of ubiquitin signalling are many and various and affect almost all aspects of cell biology [3].

Second Messengers in Signal Transduction

As well as transferring information directly through the action of enzymes on protein substrates to alter their function, hormone signalling pathways also commonly rely on activation of enzymes that generate diffusible small molecules, or 'second messengers' that can either be cytosolic or membrane-anchored. These can exert effects on distant cellular components, or, conversely, can bind target proteins and thus concentrate them locally, close to substrate or targets. Second messengers are rapidly generated or released in response to extracellular signals and serve to relay and amplify the message. They are now known to come in many forms, including not only cyclic nucleotides (e.g. cAMP, cGMP) or ions (e.g. calcium, magnesium) acting within the cytosol, but also membrane-associated lipids or lipid-derived molecules. In non-stimulated cells second messenger concentrations are extremely low, but they are produced rapidly in response to receptor activation, binding and altering target proteins within the cell to propagate signalling [4]. As for all signalling processes, mechanisms must exist to extinguish the signal. In the case of second messengers this is usually through enzyme-mediated metabolism of organic second messengers, or sequestration of ionic messengers in membrane-bound organelles.

Cyclic Nucleotide Second Messengers

In the 1960s, Earl Sutherland first elucidated the role of 3'–5'-cyclic adenosine monophosphate (cAMP) as the 'second messenger' of catecholamine action (adrenaline/epinephrine itself being viewed as the first messenger), a breakthrough of such significance that Sutherland was awarded the 1971 Nobel Prize in Physiology or Medicine 'for his discoveries concerning the mechanisms of the action of hormones'. Membrane-bound adenylyl cyclases, which synthesize cAMP in response to activation, are particularly important in transmitting signals from the G-protein-coupled class of cell membrane receptors described next.

The major target protein of cAMP is protein kinase A (PKA), which through its kinase activity can then phosphorylate downstream targets (Figure 1.3.2). Many different proteins are substrates for PKA, and ensuring that the correct target is phosphorylated at the right time in response to the physiologically appropriate signal is governed both by the monomers constituting the PKA heterotetramer and by A-kinase-anchoring proteins (AKAPs) that function as scaffolds to anchor PKA to subcellular structures [4]. The phosphodiesterase family of enzymes play a critical role in modulating the duration and amplitude of signalling by cAMP (and cGMP, which has also been shown to have a second messenger role in distinct signalling pathways) by catalysing the hydrolysis of these two second messengers. Tissue-specific expression and discrete subcellular location of various phosphodiesterase family members allows tight control of compartmentalized gradients of cAMP and cGMP hence regulating their signalling pathways and downstream biological effects. Phosphodiesterases have thus been exploited pharmacologically for therapeutic effect and the development of drugs that target individual subfamilies has improved tolerability and efficacy of treatments for Cushing's disease, erectile dysfunction, asthma, and heart disease [5].

Lipid Second Messengers

Several types of membrane phospholipid serve as important substrates for generation of second messengers. Glycerophospholipids anchored on the cytoplasmic face of the cell membrane can be hydrolysed by phospholipase D and phospholipase C (PLC) to generate the second messengers, phosphatidic acid and diacylglycerol (DAG), respectively [4]. When phosphatidylinositol 4,5-bisphosphate (PIP_2) is hydrolysed by PLC in response to receptor activation, DAG remains in the cell membrane while the polar head group inositol trisphosphate (IP_3) is released into the cytosol. There it binds IP_3 receptors on organelles resulting in the release of calcium (itself a second messenger) into the cytosol (Figure 1.3.2). IP_3 can be further phosphorylated to IP_4 which activates chloride channels, and IP_6 which may have a role in apoptosis [4]. DAG acts as a second messenger to target a family of enzymes known collectively as protein

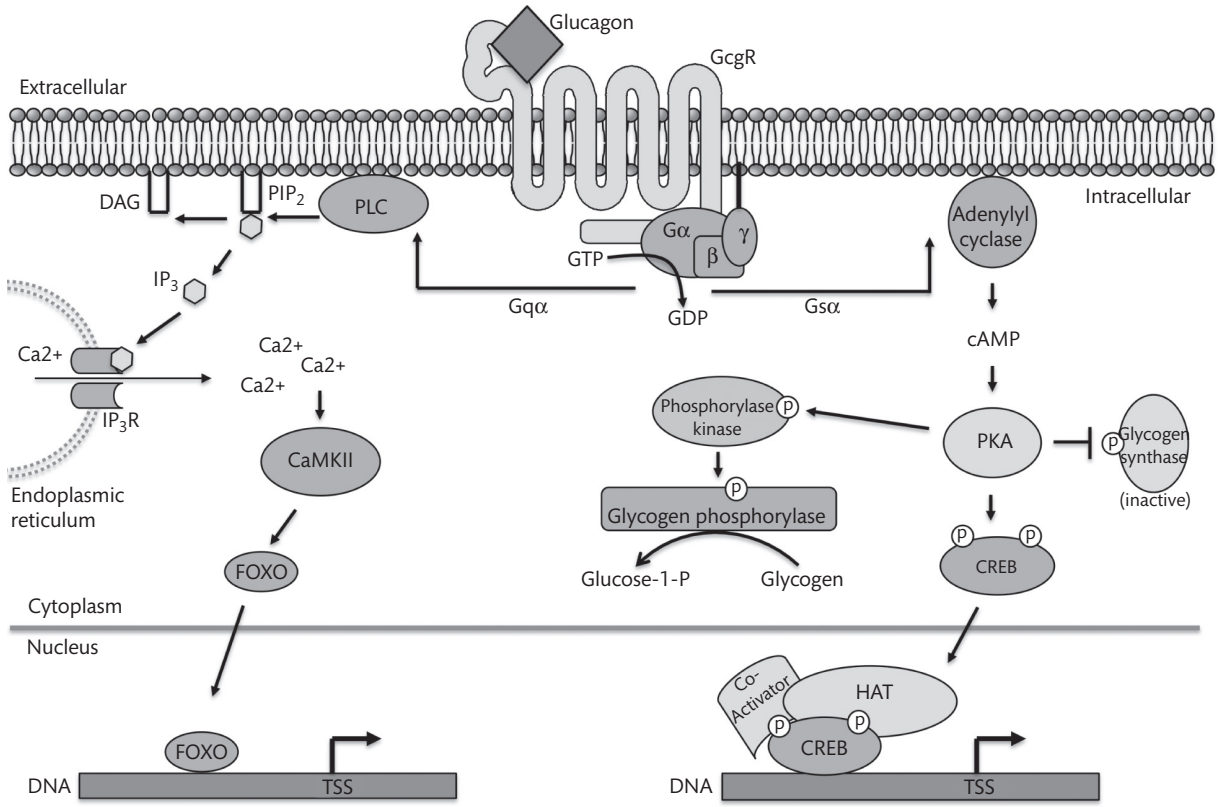


Figure 1.3.2 Glucagon signalling via its G-protein-coupled receptor. Glucagon signals via a GPCR (GCGR) that associates with either Gsα or Gqα heterotrimeric G-proteins. This activates pathways leading to transcription of genes involved in glycogenolysis and gluconeogenesis. Glucagon signalling also stimulates enzymatic changes that regulate glycogenolysis and the generation of glucose.

kinase C (PKC). Some PKC isoenzymes require both DAG and calcium for activation, giving rise to a level of specificity as PKC can transduce signals that trigger selective hydrolysis of PIP_2 due to the release of calcium caused by IP_3 mobilization. Other isoenzymes of PKC can be activated by DAG alone, while so-called atypical PKC isoenzymes do not respond to DAG or calcium. Signalling by PKC is terminated by the phosphorylation of DAG by DAG kinase which removes the second messenger [4].

Plasma-membrane-associated PIP_2 can also be phosphorylated by phosphoinositide 3-kinase (PI3K) to produce the second messenger phosphatidylinositol (3,4,5)-triphosphate (PIP_3), which is key in the PI3K-AKT signalling pathway. This is triggered by many receptor tyrosine kinases (RTKs) such as the insulin and IGF1 receptors (Figure 1.3.3). Activation of PI3K leads to conversion of PIP_2 to PIP_3. PIP_3 in turn provides a docking site for AKT and other effector proteins, serving to drive their localization to desired sites of action at the plasma membrane.

Calcium as a Second Messenger

In direct response to receptor activation, or as a consequence of second messenger action, calcium can rapidly enter the cytoplasm via activation of channels and transporters either at the cell surface or on organelles with a high internal calcium concentration. Calcium itself acts as a second (or third!) messenger and can rapidly propagate signals by regulating targets either by direct binding or by modulating calcium sensors and calcium-sensitive enzymes [4]. For example, as one component of glucagon signalling (Figure 1.3.2), in addition to activation of adenylate cyclase, PLC is activated and cleaves phosphatidylinositol 4,5-bisphosphate (PIP_2) generating IP_3 and DAG. IP_3 then acts on the IP_3 receptor calcium channel on the endoplasmic reticulum resulting in an increase in Ca^{2+} in the cytoplasm and activation of the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII). Downstream of CaMKII activation, the transcription factor FoxO is phosphorylated and translocated to the nucleus where it can stimulate expression of genes involved in hepatic glucose production [6]. Cells actively maintain gradients of calcium across lipid membranes and when calcium channels are opened, rapid diffusion of calcium down these gradients generates spatiotemporal waves and oscillations which, coupled to the subcellular organization of the calcium signalling machinery, allows for a vast array of signals to be encoded and propagated. Calcium actions are terminated by pumping calcium out of the cell or back into subcellular organelles [4].

Types of Plasma Membrane Hormone Receptor

G-Protein-Coupled Receptors

G-protein-coupled receptors (GPCRs) are a large family of proteins (numbering as many as 800 in mammals). They share core sequence and structural similarity, but the family has evolved to be activated by a myriad of different ligands including light-sensitive compounds, odorant chemicals, pheromones, and neurotransmitters as well as hormones. Hormones of clinical importance that bind

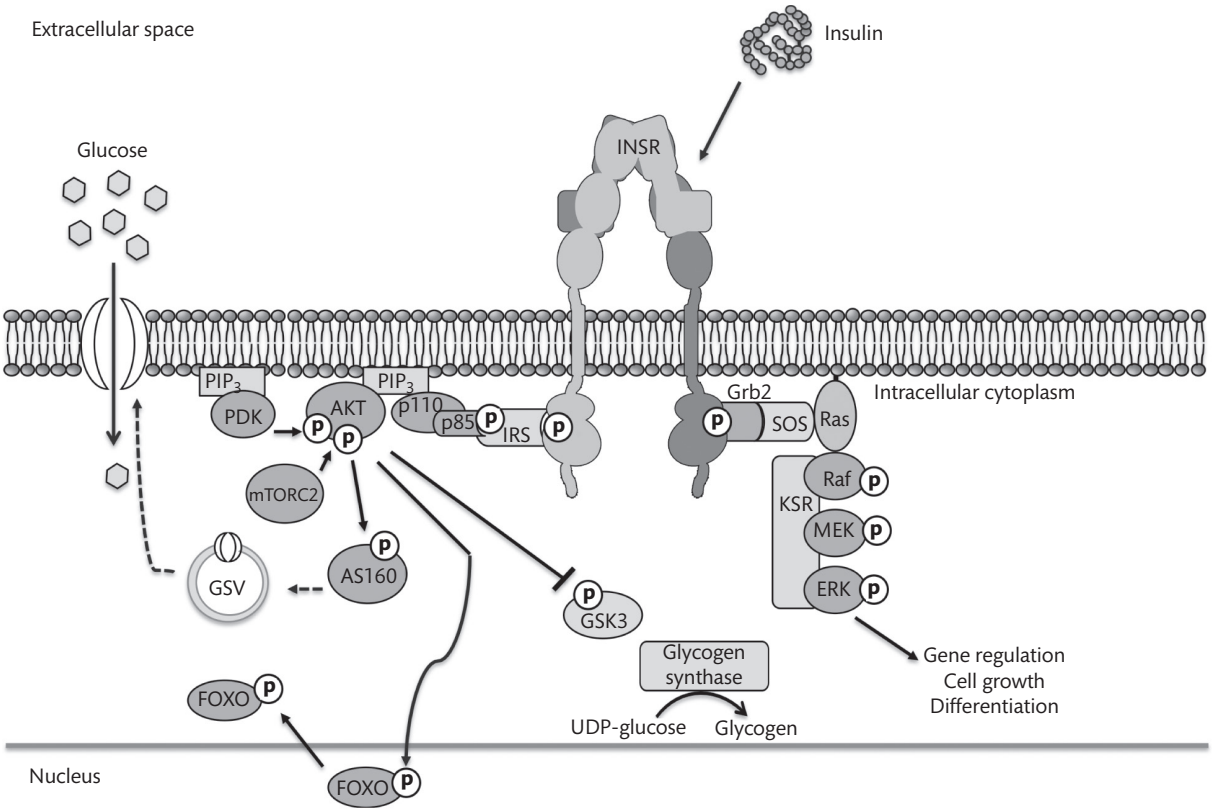


Figure 1.3.3 Signalling pathways downstream from insulin receptor activation. Insulin stimulates activation of both PI3K/AKT and Ras/ERK pathways resulting in a broad range of cellular outcomes including effects on glucose uptake and metabolism, protein synthesis, cell growth and differentiation, and prevention of apoptosis.

GPCRs include catecholamines, glucagon, ACTH, PTH, TSH, FSH, LH, and β-hCG, among many others. GPCRs are also the largest family of proteins targeted by approved drugs. All GPCRs include seven transmembrane regions between an extracellular N-terminal domain and a cytosolic C-terminal segment. Extracellular ligand binding causes a conformational change, reconfiguring the cytoplasmic loops to bind and activate a heterotrimeric G-protein complex that is responsible for signal transduction [7]. GPCRs are also capable of transactivating both RTKs and serine/threonine receptor kinases, providing one of many mechanisms of 'crosstalk' with other signalling pathways [7].

G proteins constitute one of the most important families of signalling proteins. Martin Rodbell and Alfred Gilman shared the 1994 Nobel Prize in Physiology or Medicine 'for their discovery of G-proteins and the role of these proteins in signal transduction in cells'. G-protein complexes comprise α, β, and γ subunits, for each of which multiple isoforms exist. Function is defined primarily by the Gα subunit. The combination of G-protein subunits determines which receptor the complex couples to, the second messenger generated, and the downstream effectors activated, allowing the cell to integrate information from a variety of extracellular stimuli. Stimulatory G protein (Gs_α) activates adenylyl cyclase, which generates the second messenger 3'–5'-cyclic adenosine monophosphate (cAMP), while inhibitory G protein (Gi_α) inhibits adenylyl cyclase. Gq_α activates phospholipase C (PLC), resulting in generation of inositol trisphosphate (IP_3) and diacylglycerol (DAG) as second messengers (Figure 1.3.2). Signalling from G_α is terminated by its intrinsic GTPase, which hydrolyses G_α-bound guanosine triphosphate (GTP) to guanosine diphosphate (GDP). This enables reassociation of G_α to $G\beta\boldsymbol{\gamma}$ [7]. GTP hydrolysis is enhanced by a variety of GTPase-activating proteins (GAPs), and nucleotide exchange is enhanced by Guanine nucleotide exchange factors (GEFs). Both these types of proteins serve as important sites of regulation of the action of many hormones. GPCR signalling is also regulated by the β-arrestin family of proteins. β-arrestins were initially identified for their role in switching off GPCR signalling by desensitizing the receptors and facilitating their internalization and recycling. More recently, β-arrestins have been shown to have wider roles in GPCR signalling, acting both as signal transducers by controlling phosphorylation of intracellular targets and as scaffold proteins for the recruitment of signalling cascades (such as ERK), thereby prolonging signalling by GPCRs internalized to the endosomes [8].

β-arrestins also play a key role in the 'signalling bias' shown by both natural ligands and pharmacological compounds that can act not only as full agonists and antagonists, but also as partial and inverse agonists by way of multidimensional allosteric or orthosteric interactions with the GPCR. Such 'signalling bias' is being exploited for therapeutic benefit as some ligands/compounds preferentially activate classical G-protein-dependent signal transductions, whereas others demonstrate preferential activation of β-arrestin mediated signalling. This allows selective control of pathway activation to minimize potential side effects [8].

Receptor Tyrosine Kinases

RTKs bind many important peptide hormones and growth factors. RTK ligand-binding domains reside in the extracellular portion of the receptor, a transmembrane domain spans the plasma membrane, and the intracellular portion of the receptor contains the tyrosine kinase domain. RTKs are commonly expressed at the cell surface as monomers that undergo dimerization upon binding of bivalent ligands which cross-links monomers to form active complexes [9]. Exceptions are the insulin receptor (INSR) and the closely related type I insulin-like growth factor receptor (IGF1R), which exist at the cell surface as preformed dimers, within which ligand binding causes conformational changes and activation currently best described by a 'harmonic oscillator' model [10].

While precise signalling mechanisms differ for the various members of the RTK family, ligand binding and the resulting conformational changes generally trigger the release of autoinhibition within the tyrosine kinase domain, resulting in activation of the RTK [9]. This leads to autophosphorylation of tyrosine residues in receptors themselves and on separate substrate proteins (e.g. INSR substrates, IRS-1 and IRS-2), which provide docking sites for adaptor proteins that recognize phosphotyrosines within specific sequence contexts. Docking and sometimes phosphorylation of these adaptor molecules propagates signalling cascades responsible for diverse cellular responses. For example, the effects that insulin exerts on glucose metabolism, cell growth, and differentiation are mediated mostly by signalling pathways involving phosphoinositide 3-kinase/AKT (PI3K/AKT) and mitogen-activated protein kinase/extracellular signal-related kinase (MAPK/ERK) [11], which are initiated by recruitment of regulatory subunits or proteins to IRSs (**Figure 1.3.3**). Generation of PIP_3 by PI3K leads to membrane binding of AKT and its colocalization with phosphoinositide-dependent kinase I (PDKI), also recruited to the cytoplasmic face of the cell membrane by PIP_3, and the mammalian target of rapamycin complex 2 (mTORC2). These phosphorylate and activate AKT, a kinase that in turn phosphorylates an array of substrates with important metabolic and growth-related functions [12]. Multiple enzymes contribute to the termination of RTK signalling, including protein tyrosine phosphatases (PTPs) that dephosphorylate INSR and IRS proteins, the phosphatase and tensin homologue (PTEN) that dephosphorylates PIP_3 and the protein phosphatases PHLPP and PP2A that dephosphorylate AKT.

Other Receptors and Signalling Pathways

In contrast to the RTKs and PS/TRKs (discussed next), cytokine-type cell surface receptors lack intrinsic kinase activity, instead associating with members of the intracellular Janus kinase (JAK) family to propagate signalling. The cytokine family includes a diverse array of proteins including growth hormone, leptin, interferons, chemokines, adipokines, lymphokines, and tumour necrosis factors with endocrine, paracrine, and autocrine actions. Ligand binding to cytokine receptors induces conformational changes favouring recruitment and activation of the associated JAK leading to phosphorylation of proteins called signal transducers and activators of transcription (STAT) [13]. Like the other cell surface receptors, a key mechanism for modulating cytokine signalling at the cell surface is internalization of the receptor upon ligand binding. Together with subsequent recycling back to the cell surface or degradation of the receptor, this modulates receptor availability to ligand. Cytokine receptor signalling is negatively regulated intracellularly by Src homology 2 domain-containing PTPs (SHPs) and the suppressor of cytokine signalling (SOCS) family of proteins [14].

The protein serine/threonine receptor kinases (PS/TRKs) respond to members of the superfamily of ligands that include the transforming growth factor beta family (TGFβ), bone morphogenic proteins (BMPs), activins, and growth and differentiation factors (GDFs) which have diverse roles in embryonic development and tissue homeostasis, the immune system, and cancer, as well as endocrine regulation of adipose tissue metabolic functions. Ligand binding causes the formation of a heterotetrameric complex of two different PS/TRKs (type I and type II receptors) resulting in an interaction where the type II receptor phosphorylates and activates the type I receptor allowing signal transduction via the phosphorylation of serine or threonine residues on target downstream signalling substrates [15]. The best-known pathway downstream from PS/TRKs involves the Smad family of transcription factors, although other signal transduction pathways such as ERK and c-Jun N-terminal kinase (JNK) can be activated, for example, in response to TGFβ [15].

Information Coding in Signalling Networks

For all the multitude of different cell surface receptors, the repertoire of signalling cascades initiated in response to receptor activation is relatively limited. Despite this they mediate a broad spectrum of cellular outcomes, and yet individual receptors produce distinct responses. Many mechanisms exist to permit such specificity of information transfer. Important factors include the cellular context of receptor expression, positive and negative crosstalk among signalling pathways, and feedback regulation within pathways. Spatial and temporal factors including subcellular localization and regulated intracellular trafficking of signalling components, cytoskeletal arrangements, and differential utilization of 'scaffold' and adaptor proteins are also critical. Such determinants of signalling specificity are discussed briefly next.

Hormone Receptor Signalling: 'Cascades' or 'Networks'?

Although hormone signalling pathways are commonly schematized as distinct, simple, and linear, this is a gross simplification. Feedback both within and between different pathways is a major feature of signalling. Negative feedback from later stages of a pathway to proximal steps is one prevalent mechanism which means that activation of a pathway inevitably leads, with a time delay, to its downregulation [16]. Positive feedback loops also exist within some signalling cascades, however, that can stabilize and amplify signals [17]. Many examples of both positive and negative feedback between distinct pathways also exist. The collective dynamics of such positive/negative feedback loops and oscillations in enzyme on/off states can affect the timescale and magnitude of biological responses. Sophisticated modern techniques, most involving mass spectroscopy, are beginning to define temporal complexity of signalling in response to different hormones; however, understanding how different types of

information are encoded remains limited. A complete account of the richness of information transmitted by hormone action will require further integration of such temporal considerations with the subcellular spatial localization of signalling molecules. Only once this considerable computational feat is achieved may it be possible to understand the diverse biological outcomes achieved with such a limited number of signalling pathways [17].

Selective Receptor Modulation and Biased Agonism

Hormone receptor signalling is sometimes viewed, for simplicity, as binary, with receptors acting essentially as 'switches' to trigger a stereotyped cellular response. However, as the preceding discussion has indicated, hormone receptor signalling is much more complex. Responses to hormones are usually graded, and at different levels of hormone stimulation responses may qualitatively vary, as different arms of the downstream pathway may require different degrees of stimulation to activate. Signalling responses are also context dependent, in part due to the accessory proteins that function with hormone receptors to modulate selectivity or to direct downstream responses. This is true of both nuclear hormone receptors and cell surface receptors. For example, the biased agonism of GPCRs depends on the context of the allosteric or orthosteric interactions of a particular ligand of a GPCR that lead to either preferential activation of G-protein-dependent signalling or preferential β-arrestin mediated signalling [8]. Biased signalling can also be observed with the receptors of the insulin-like growth factor (IGF) system and their various affinities for the IGFs and insulin, which is further complicated by the variety of hybrid receptors that can form between the IGF1R, INSR-A, and INSR-B isoforms [12]. Understanding the biophysical, structural, and pharmacological biology of receptors, their mechanism of actions and how these relate to distinct signalling outcomes has enabled the development of therapeutics that capitalize on selective pathway activation to minimize potential side effects and/or selectively target one target tissue over another. A prime example is the failed contraceptive tamoxifen, which is a selective oestrogen receptor modulator (SERM) and has mixed agonist and antagonist properties depending on tissue of action. Tamoxifen is used to treat breast cancer and has beneficial effects in postmenopausal women on bone mineral density and serum lipids, but also increases the risk of development of endometrial cancer due to its oestrogen receptor (ER) agonist activity in the uterus [18].

Spatial Regulation of Intracellular Signalling

Another mechanism which can serve to diversify biological outcomes, even using the same signalling process, is the localization of machinery involved to specific parts of the cell. This commonly involves so-called scaffold proteins which essentially serve to tether, orientate, and coordinate allosteric control of partner molecules, providing a flexible means of signalling regulation [19]. Scaffold proteins can either mediate linear signalling down a pathway or provide a branching point for diversification of the signal. An example of linear signalling can be observed with the kinase suppressor of Ras (KSR) scaffold protein that helps to tether and activate components of the Ras-Raf-MEK-ERK signalling pathway downstream of RTK activation, by facilitating the assembly of a multiprotein complex that brings the downstream substrates in close proximity to their upstream activators (Figure 1.3.3) [19, 20]. Scaffold proteins are themselves subject to regulation by PTMs such as phosphorylation, which can be either stimulatory or inhibitory to enable or block protein–protein or protein–lipid interactions that sequester the signalling molecules to distinct subcellular locations (such as the effect of AKAP scaffold proteins on PKA as discussed in the section on cAMP second messengers), or ubiquitination which targets substrates for proteosomal degradation [19]. Scaffold proteins are also pivotal in coordinating feedback loops involving phosphatases that are important for temporal dynamics and the termination of signalling pathways [17]. Not only do scaffold proteins function to integrate signalling networks arising from hormonal activation of cell surface receptors, they are also involved in the coordination of sequential passing of some type I and III steroid hormone receptors across their chaperone proteins prior to translocation to the nucleus [20].

Receptor Endocytosis and Signal Termination

Common to both GPCR and RTK signalling is that receptor activation at the cell surface is followed by endocytosis. The consequent decrease in cell surface expression helps prevent excessive downstream signal activation by attenuating the strength and duration of signalling or by providing a shift in the dose-response relationship between ligand and receptor such that a greater concentration of ligand is required to produce a cellular response [21]. Internalized receptors are trafficked through the endosomal/lysosomal pathway and this process is not only critical for the termination of signalling via receptor sequestration and degradation, but also allows recruitment of specific signalling components in different compartments. For example, full activation of the extracellular signal-regulated kinases (ERK1/2) by RTKs requires internalization of the receptor and complexing with scaffold proteins in the late endosome [21]. In contrast, full activation of AKT can occur at the plasma membrane prior to receptor internalization [21]. Similarly, GPCRs are capable of sustained signalling within endosomes that affects recruitment of downstream mediators [22].

Activated receptors generally undergo clathrin-mediated endocytosis (CME) [23] either by direct interaction with the adaptor complex AP2 (RTKs) or via interactions with β-arrestins (GPCRs). Ubiquitination can also target receptors for internalization and the level of receptor ubiquitination in endosomes affects recycling back to the plasma membrane or trafficking to the lysosomes for degradation [21]. Receptors that are internalized by clathrin-independent endocytosis may be degraded more rapidly than those that undergo CME, suggesting another level by which trafficking routes can influence cellular signalling. Various signalling kinases also have an effect on endosomal sorting and trafficking to lysosomes to reduce receptor recycling and terminate signalling [21, 22].

Hormone Resistance

Primary endocrine disorders may broadly be classified into those featuring loss or dysregulation of hormone production, and others featuring an abnormality of hormone action, most commonly hormone resistance. Hormone resistance as a clinical phenomenon has been recognized for many years, notably through the work of

Table 1.3.1 Genetic forms of hormone resistance

Examples of hormone resistance caused by hormone receptor mutations			
Gene	**Hormone**	**Receptor type**	**Syndrome**
INSR	Insulin	Receptor tyrosine kinase	Severe insulin resistance
IGF1R	IGF-1	Receptor tyrosine kinase	Short stature
GHR	Growth hormone	Cytokine	Laron dwarfism
LEPR	Leptin	Cytokine	Severe early onset obesity
GNRHR	GnRH	G-protein-coupled	Hypogonadotropic hypogonadism
TSHR	TSH	G-protein-coupled	Congenital hypothyroidism
MC2R	ACTH	G-protein-coupled	Glucocorticoid deficiency
THRB	Thyroid hormone	Nuclear receptor	Thyroid hormone resistance
AR	Androgens	Nuclear receptor	Androgen insensitivity
Examples of hormone resistance caused by mutations in signalling molecules downstream from hormone receptors*			
Gene	**Function of gene product**		**Syndrome**
PIK3R1	Regulatory subunit of phosphatidylinositide 3-kinase (PI3K)		SHORT syndrome (including short stature and insulin resistance)
GNAS	Stimulatory G protein coupled to many peptide hormone receptors		Pseudohypoparathyroidism
STAT5B	Transcription factor activated downstream from some cytokine receptors		Growth hormone resistance with immunodeficiency

* Examples of hormone resistance caused by loss-of-function mutations. Syndromes due to defects in signalling molecules are often more complex than those involving receptor defects due to sharing of signalling pathways. ACTH, adrenocorticotropic hormone; AR, androgen receptor; GHR, growth hormone receptor; GNAS guanine nucleotide-binding protein G(S); GnRH, gonadotropin releasing hormone; GNRHR, gonadotropin releasing hormone receptor; IGF-1, insulin-like growth factor 1; IGF1R, insulin-like growth factor 1 receptor; INSR, insulin receptor; LEPR, leptin receptor; MC2R, melanocortin-2 receptor; PIK3R1, phosphatidylinositol 3-kinase regulatory subunit alpha; SHORT, short stature, hyperextensibility, hernia, ocular depression, Rieger anomaly, and teething delay; STAT5B, signal transducer and activator of transcription 5B; THRB, thyroid hormone receptor beta; TSH, thyroid stimulating hormone; TSHR, thyroid stimulating hormone receptor.

Harold Himsworth on insulin in the 1930s, and the work of Fuller Albright on PTH in the 1940s. In the molecular genetic era, as the genes encoding components of signalling pathways have been identified, so the precise molecular bases of some forms of hormone resistance have been identified. Many of these lie in the receptor itself, while others lie in receptor signalling pathways downstream from the receptor (Table 1.3.1). For some truly pandemic conditions such as insulin resistance, however, the precise defect is imprecisely identified, and appears not, in general, to be due to single gene defects of insulin action. For other disorders, such as obesity, the very nature and existence of leptin resistance continues to be debated.

It follows from consideration of the complexity of receptor action that some forms of primary hormone resistance are not simply classifiable as loss of hormone action, but rather reflect selective loss of some aspects of hormone intracellular signalling. Depending on the feedback loops that regulate a pathway, this means that compensatory high levels of hormone may paradoxically drive unimpaired aspects of receptor signalling, and this may lead to clinical disease. Common insulin resistance may provide one such example, with significant evidence suggesting that some aspects of the 'insulin resistance syndrome' are actually driven by high levels of insulin signalling in some tissues, driven by hyperinsulinaemia. Disentangling perturbations of complex signalling networks is a major challenge for current efforts to understand many common endocrine disorders.

Conclusion

Methodological and conceptual advances over the last 40 years have provided major insight into the molecular mechanisms of hormone action. The anatomy of intracellular signalling networks has been revealed, and new insights gained into how cells decode, process, and respond to multiple external stimuli utilizing only a limited repertoire of signalling pathways. The concept of temporal and spatial compartmentalization influencing signalling network specificity is now established, however, much is still to be learned about the mechanisms that couple external stimuli to their intracellular 'roadmap' to a specific biological outcome. Integrating the computational methodologies of systems biology with biochemical experimentation [24] will be required to further enhance understanding of signalling specificity, and to enable more refined endocrine interventions, for example, in the form of selective receptor modulators, for human disease.

REFERENCES

1. Sever R, Glass CK. Signalling by nuclear receptors. *Cold Spring Harb Perspect Biol*, 2012; **5**: a016709.
2. Cheng S-Y, Leonard JL, Davis PJ. Molecular aspects of thyroid hormone actions. *Endocr Rev*, 2010; **31**: 139–70.
3. Swatek KW, Komander D. Ubiquitin modifications. *Cell Res*, 2016; **26**: 399–422.
4. Newton AC, Bootman MD, Scott JD. Second messengers. *Cold Spring Harbour Perspect Biol*, 2016; **8**: pii: a005926.
5. Lomas O, Zaccolo M. Phosphodiesterases maintain signaling fidelity via compartmentalisation of cyclic nucleotides. *Physiology*, 2014; **29**: 141–9.
6. Goldstein I, Hager GL. The three Ds of transcription activation by glucagon: direct, delayed, and dynamic. *Endocrinology*, 2018; **159**: 206–16.

7. Syrovatkina V, Alegre KO, Dey R, Huang XY. Regulation, signalling and physiological functions of G-proteins. *J Mol Biol*, 2016; **428**: 3850–68.
8. Reiter E, Ayoub MA, Pellissier LP, *et al.* Beta-arrestin signaling and bias in hormone-responsive GPCRs. *Mol Cell Endocrinol*, 2017; **449**: 28–41.
9. Lemmon MA, Schlessinger J. Cell signalling by receptor-tyrosine kinases. *Cell*, 2010; **141**: 1117–34.
10. Kiselyov VV, Versteyhe S, Gauguin L, De Meyts P. Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. *Mol Syst Biol*, 2009; **5**: 243.
11. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signaling pathways: insights into insulin action. *Nat Rev Mol Cell Biol*, 2006; **7**: 85–96.
12. Siddle K. Molecular basis of signalling specificity of insulin and IGF receptors: neglected corners and recent advances. *Front Endocrinol (Lausanne)*, 2012; **3**: 34.
13. Dodington DW, Desai HR, Woo M. JAK/STAT—emerging players in metabolism. *Trends Endocrinol Metab*, 2018; **29**: 55–65.
14. Liongue C, Sertori R, Ward AC. Evolution of cytokine receptor signalling. *J Immunol*, 2016; **197**: 11–18.
15. Wu MY, Hill CS. TGFβ superfamily signalling in embryonic development and homeostasis. *Dev Cell*, 2009; **16**: 329–43.
16. Hunter T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell*, 1995; **80**: 225–36.
17. Kholodenko BN. Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol*, 2006; **7**: 165–76.
18. Maximov PY, Lee TM, Jordan VC. The discovery and development of selective oestrogen receptor modulators (SERMs) for clinical practice. *Curr Clin Pharmacol*, 2013; **8**: 135–55.
19. Good MC, Zalatan JG, Lim QA. Scaffold proteins: hubs for controlling the flow of cellular information. *Science*, 2011; **332**: 680–6.
20. Echeverria PC, Picard D. Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochemica et Biophysica Acta*, 2010; **1803**: 641–9.
21. Sorkin A, von Zastrow M. Endocytosis and signaling: intertwining molecular networks. *Nat Rev Mol Cell Biol*, 2009; **10**: 609–22.
22. Pavlos NJ, Friedman PA. GPCR signalling and trafficking: the long and the short of it. *Trends Endocrinol Metab*, 2017; **28**: 213–26.
23. Robinson MS. Forty years of clathrin-coated vesicles. *Traffic*, 2015; **16**: 1210–38.
24. Janes KA, Lauffenburger DA. Models of signalling networks—what cell biologists can gain from them and give to them. *J Cell Sci*, 2013; **126**: 1913–21.

1.4

Endocrinology and Evolution

Lessons from Comparative Endocrinology

Janine A. Danks and Samantha J. Richardson

Introduction

Comparative endocrinology is the study of endocrine glands and hormones in different animal species. It is undergoing a renaissance driven by rapid technological advances facilitating, for example, DNA/RNA sequencing, profiling of DNA methylation and protein interactomes, and genome editing with CRISPR. Until relatively recently, characterization of hormones in 'lower' vertebrates relied upon biological assays and protein chemistry, whereas now genes are readily revealed in whole genome sequences, and specific antibodies and other reagents for sensitive assays can rapidly be developed, enabling revealing physiological experiments.

Endocrinology traditionally used animal species including many lower vertebrates. Comparative endocrinology became distinct only in the last 50 years as endocrinologists concentrated on rodents as their preferred model. In 1933, Riddle demonstrated that an avian pituitary factor promoting growth of the pigeon crop-sac was identical to a mammalian pituitary factor earlier found to initiate and maintain milk secretion. Riddle called this avian factor prolactin and the crop-sac response provided a sensitive assay for human prolactin detection in pituitary extracts. Pigeon prolactin was the first pituitary hormone to be crystallized and purified in 1937, leading to purification of mammalian prolactin. Prolactin has many roles in lower vertebrates, including a vital role as a hypercalcaemic factor in fish.

This chapter will focus on two paradigmatic examples of the value of comparative endocrinology. Calcium regulating factors including parathyroid hormone (PTH), parathyroid hormone-related protein (PTHrP) and stanniocalcin (STC) will first be considered, followed by THs (which have identical structures in all vertebrates from fish to humans) and transthyretin (a thyroid hormone distributor in blood and cerebrospinal fluid). Both groups of hormones are essential for human life during development and in adulthood, and pivotal insights have been gained from non-mammalian vertebrates.

Calcium Regulation

Regulation of calcium concentration in intracellular (c.0.1μmol/L) and extracellular compartments is fundamental in vertebrate physiology. Calcium regulation in body fluids demarcates invertebrates and vertebrates: marine invertebrates have plasma calcium concentrations around 10 mmol/L, close to the concentration in the environment. Primitive marine vertebrates (cyclostomes) have plasma calcium levels of 5 mmol/L in the same environment, and most cartilaginous and bony fish, regardless of environment, have

plasma calcium levels around 2 mmol/L, similar to mammals. Most aquatic vertebrates are less dependent on the skeleton as a calcium source, as the aquatic environment provides an inexhaustible calcium supply.

Fish use their scales instead of their skeleton as a rapidly accessible calcium store. In contrast, terrestrial vertebrates have skeletons serving both as a physical support and calcium reservoir. Dietary calcium supply is episodic, so calcium reserves must be labile and dynamic. All vertebrates possess several essential factors with complex interrelated actions controlling circulating ionic calcium.

Calcium Regulating Factors in 'Lower' Vertebrates

The major difference in calcium control between 'lower' and tetrapod vertebrates is the presence of the parathyroid gland and its product, PTH. A distinct parathyroid gland first appeared in amphibians, but no comparable gland has been observed in fish. PTH is critical for calcium homeostasis at all stages of tetrapod development, interacting with calcitonin and 1,25-dihydroxyvitamin D_3. PTH elevates serum calcium by causing calcium release from bone and inhibition of renal calcium excretion.

PTH

Despite the absence of a parathyroid gland, several reports describe an immunoreactive PTH-like protein in fish, detected by antisera raised against the mid-molecule portion of bovine PTH [1]. PTH-like immunoreactivity was demonstrated in brain and pituitary of goldfish and platyfish, the pituitary of eel and cod, and the plasma of trout and goldfish. Problems with these studies included insufficient controls and poor characterization of antibodies, however. Physiological studies by Parsons *et al.* [2] predicted a rapidly acting hypercalcaemic factor in cod pituitary with a PTH-like N-terminus, however existence of a fish PTH homologue was unproven until 2003, when a *PTH* gene was identified from the *Fugu rubripes* genome [3]. *Fugu* PTH has low overall homology with human PTH (32%) and *Fugu* PTH(1–34) has only 56% identity with human PTH(1–34), however *Fugu* PTH(1–34) was found to have *in vitro* biological activity resembling that of human PTH(1–34). A second *Fugu PTH* gene was later identified that is not present in the human genome, suggesting a gene duplication in bony fish. There are now three PTH genes in zebrafish [4], chicken and *Xenopus* [5], but the third PTH gene is absent in eutherian mammals, which generally have fewer chromosomes (e.g. 39 in chickens vs. 23 in humans).

Subsequent studies have identified two *PTH* genes in the elephant shark (*Callorhinchus milii*) [6]. The elephant shark is among the most ancient cartilaginous fish, the oldest living group of jawed vertebrates (approximately 450 million years old) (Figure 1.4.1). Its genome has been termed the 'reference vertebrate genome' as initial studies have shown it to resemble the human genome more than those of other fish (e.g. *Fugu* or zebrafish) [7]. The two elephant shark PTH proteins have different biological activity and the gene encoding elephant shark *PTH2* does not exist in other fish or humans [6].

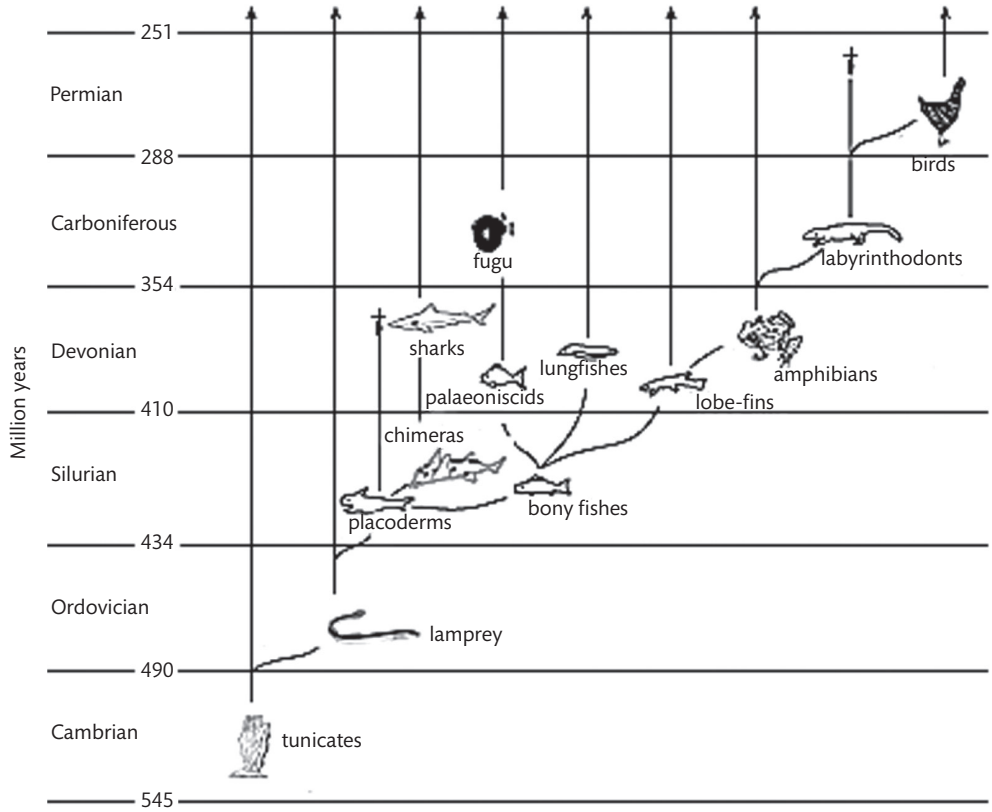


Figure 1.4.1 Classification of chordates, including tunicates and lower vertebrates, in relation to phylogenetic origins and palaeontological periods.

PTHrP

The earliest reports of PTH-like immunoreactivity in fish prompted studies in the early 1990s [8] demonstrating immunoreactive PTHrP in tissues including the pituitary as well as in the circulation of a marine bony fish. PTHrP is highly homologous to the N-terminus of PTH and may be the factor predicted by Parsons *et al.* [2] in cod pituitary. PTHrP was first isolated from human tumours [1, 9, 10] and identified as a factor causing humoral hypercalcaemia of malignancy (HHM), acting via the classical PTH receptor (PTH1R) in bone and kidney. PTH and PTHrP bind the receptor with different affinities. They share sequence homology only in the N-terminal, notably in the 1–34 region, with 8 of the first 13 amino acids being identical. The N-terminus of PTH and PTHrP are required for PTH1R binding. Subsequent studies indicated that PTHrP is synthesized in many normal adult and embryonic mammalian tissues, implying numerous potential functions, many autocrine or paracrine [1].

The *PTHrP* and *PTH* Genes

The *PTHrP* gene is more complex than the *PTH* gene (Figure 1.4.2). Rat and chicken *PTHrP* genes comprise five exons encoding a protein of 139 or 141 amino acids (Figure 1.4.2), while the human *PTHrP gene* has nine exons and encodes at least three isoforms of 139, 141 and 173 amino acids [1]. The *Fugu PTHrP* gene, in contrast, has only three exons, like the *PTH* gene. The *PTH* gene is conserved among these species, generating proteins of 84 or 88 amino acids. The increasing complexity and number of protein products of the *PTHrP* gene from birds to humans may reflect increasing numbers of roles for PTHrP through evolution. Indeed, *PTHrP* may be the original gene and *PTH* the copy. This is supported by the highly conserved gene structure of *PTH*, and the single encoded protein, suggesting a role(s) unchanged through evolution.

Duplication of the *PTHrP* and *PTH* Genes

The gene for human *PTHrP* is on chromosome 12 while the *PTH* gene is located on chromosome 11. The *PTH* and *PTHrP* genes are considered paralogous. That is, they have arisen through duplication of a single ancestral gene within a species [11]. This is opposed to orthologous genes, which are found in different species but have diverged from their common ancestral gene during speciation and separate evolution. This divergence has been documented for other gene families with members on both chromosomes 11 and 12, including the insulin-like growth factors, the aromatic amino acid hydroxylases, lactate dehydrogenases and the *Ras* gene family (Table 1.4.1). All these genes map to the same regions of these two chromosomes, suggesting a single duplication event. Chromosome 12 appears moreover to have a central position in the human genome [11], indicating that it might be a primordial chromosome. The genes on this chromosome may thus have a long evolutionary history.

The duplication event that generated two members of the PTH family may have been one of two tetraploidization events occurring between invertebrates and vertebrates [12]. Tetraploidization means

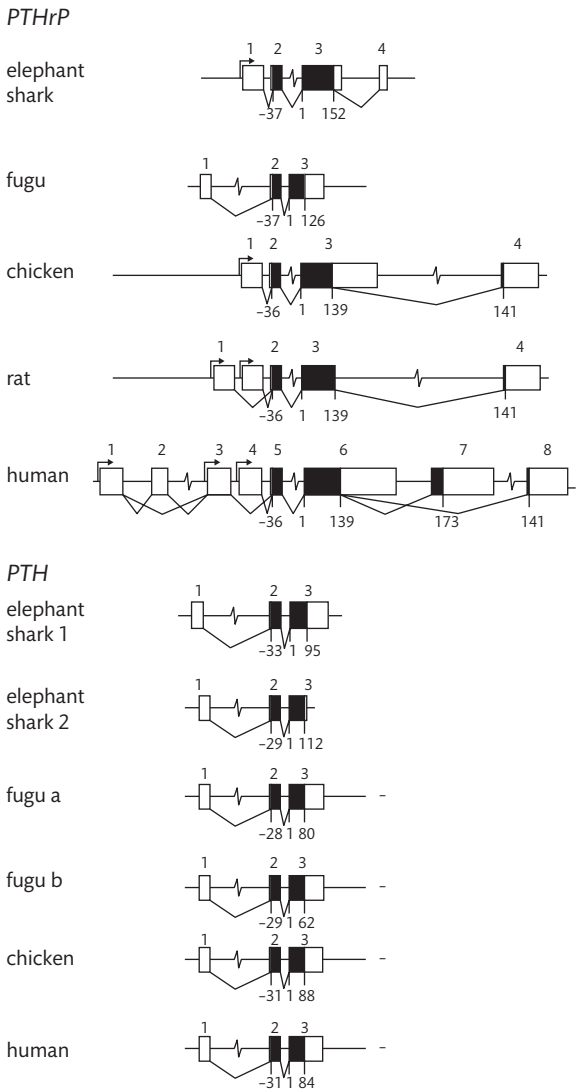


Figure 1.4.2 Comparison of the structure of elephant shark, *Fugu*, chicken, rat and human *PTHrP*, and *PTH* genes.

Adapted with permission from Liu Y, Ibrahim AS, Richardson SJ, Walker TI, Bell J, Ho PMW, Brenner S, Venkatesh V and Danks JA. (2008) The parathyroid gene family of the elephant shark (*Callorhinchus milli*). *J. Bone Miner. Res.* (2010) **25**: 2337–47. Copyright © 2010, John Wiley and Sons.

that every gene in the genome is duplicated, in contrast to regional duplication where only some loci are duplicated. Regional duplication is thought to have produced the *IGF-II* gene from *insulin* on chromosome 11, while the *IGF-I* gene on chromosome 12 is believed to have arisen from tetraploidization [11] (Table 1.4.1).

Implications of Gene Duplication

The two tetraploidization events early in vertebrate evolution expanded the genome, creating new genomic and phenotypic diversity and enhancing evolutionary survival. Such a genetic 'safety net' is exemplified by knockout mouse studies. These are based on the premise that nature is 'thrifty', with one gene responsible for a defect. Frequently, however, knockout animals do not have the expected phenotype, teaching molecular biologists what evolutionary biologists have long

Table 1.4.1 Paralogous genes in portions of human chromosomes 11 and 12

Chromosome 11	Chromosome 12
Lactate dehydrogenase A Lactate dehydrogenase C	Lactate dehydrogenase B
Hras	*Kras*
Parathyroid hormone	Parathyroid hormone-related protein
Glutathione S-transferase 3	Glutathione S-transferase 3-like Glutathione S-transferase 12
Tyrosine hydroxylase Tryptophan hydroxylase	Phenylalanine hydroxylase
Insulin Insulin-like growth factor 2	Insulin-like growth factor 1
Progesterone receptor	Vitamin D receptor Retinoic acid receptor G

known—nature is 'extravagant', and genes can sometimes fulfil multiple roles. This is logical given the imperative to ensure survival of the species. Consequently, many diseases are polygenic. Thus, several lines of knockout mice, each with a different gene deleted, may need to be interbred to yield defects originally attributed to only one gene.

Timing of the Duplication of *PTH* and *PTHrP* Genes

The *PTHrP* gene has been cloned from various mammals (human, rat, mouse, dog) and chicken and shows such high homology that its evolutionary development can only be discerned from the sequences in lower vertebrates, including fish and sharks. It appears likely that the divergence between the *PTHrP* and *PTH* genes occurred prior to evolution of cartilaginous fish. The presence of both genes in elephant shark indicates that clues to their evolutionary relationship and possible role(s) of their corresponding proteins in calcium metabolism and other functions may best be found in lampreys, a jawless fish.

Fish as Model Vertebrates

Fish constitute the major group of vertebrates, with numbers of species estimated from 25 000–35 000. They are highly successful, with physiologies evolved to survive in water of high or low ionic strength, and sometimes both. The two major groups are the jawed bony fish and the jawed cartilaginous fish, including sharks and rays (Figure 1.4.1). Surviving jawless fish, the lampreys, and hagfishes, also have a cartilaginous skeleton. It seems likely that among bony fish modification of calcium metabolism resulted in ossification of the internal skeleton. This was essential for the future evolution of terrestrial vertebrates, providing an internal calcium store.

Originally, all fish were marine and more dependent on hypocalcaemic factors to maintain tissue calcium levels lower than the surrounding water. As fish colonized fresh water, hypercalcaemic agents would have become essential for survival, and such a hypercalcaemic agent may have developed from a factor already fulfilling a different role in marine fish such as PTH or PTHrP. Over the last decade PTHrP has been identified in tissues of lower vertebrates by immunohistochemistry and *in situ* hybridization with antisera and probes to human PTHrP [13]. Tissues were examined from bony fish, including the lungfish (a 'living fossil' believed to link fish and amphibians), cartilaginous fish, and jawless fish. PTHrP protein and messenger RNA (mRNA) localized to kidney, epidermis, vertebral elements, and muscle, a distribution resembling that reported in higher vertebrates. Conservation of PTHrP localization indicates that PTHrP's function in these tissues is essential, corroborated by successful use of antisera and probes to human PTHrP, which indicates that the *PTHrP* gene and protein are highly conserved. PTHrP was also localized to tissues unique to lower vertebrates, such as gills, saccus vasculosus, and the rectal gland of sharks, all involved in mineral ion regulation. Further examination of PTHrP in these tissues could reveal new roles it plays. No PTH has been detected in tissues of lower vertebrates using a polyclonal antiserum to human PTH (1–34) due to low homology between human and fish PTH.

PTHrP is an onco-fetal hormone in mammals. Findings in bony fish argue that PTHrP may be a classical hormone in fish, however [8]. Comparative endocrinology allows examination of the normal physiological roles of PTHrP in simpler systems. The dissection of PTHrP's roles in fish has been assisted by the isolation and cloning of zebrafish *PTH1R* and *PTH2R* genes [14].

Stanniocalcin

Calcium regulation in most marine fish is dependent upon hypocalcaemic agents such as stanniocalcin (STC), a hormone produced by the corpuscles of Stannius, an organ unique to some bony and cartilaginous fish. Fish generally have abundant access to environmental calcium, and STC prevents hypercalcaemia by targeting gill and gastrointestinal tract calcium transport [13]. STC was thought to be unique to fish with corpuscles of Stannius, but in 1995 human STC was isolated and its gene cloned [15–17]. Fish STC mRNA expression is limited, however mammalian STC mRNA is found in many tissues, including ovary, prostate, and thyroid. Human STC protein is localized to the renal cortex adjacent to the glomerula and has been detected in the circulation. The mouse Stc protein has the same length as human STC, with very high sequence similarity.

The Function of Mammalian STC

STC is localized in the renal cortical brush-border membrane of rats, and decreases calcium absorption and increases phosphate absorption in rat duodenum. STC expression has also been found in developing mouse chondrocytes, acting as a probable autocrine/paracrine factor [16]. Mammalian STC may be a regulator of calcium and phosphate homeostasis, with autocrine/paracrine roles as well as an endocrine role in cell growth and differentiation. This evolutionary progression is also seen for PTHrP, a classical hormone in lower vertebrates with predominantly autocrine/paracrine roles in mammals [17].

STC-2

A second mouse and human STC (STC-2) gene and protein have been identified [17]. *STC-2* is expressed in human tissues ranging

from spleen, peripheral blood leucocytes, and small intestine, to ovary and testes. It is also expressed in rat kidney, skeletal muscle, liver, and brain. STC-2 protein has significant similarity with fish STCs and mammalian STC-1. The difference between the two mammalian STCs is 15 histidine residues in STC-2, four of them forming a cluster near the C-terminus of the protein [17]. Clusters of histidine residues often interact with metal ions such as Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}, although this has not yet been established for mammalian STC-2. A second fish STC was described in salmon corpuscles of Stannius in the same year [18]. Salmon STC-2 was less effective as an inhibitor of gill calcium transport in fish, supporting a non-calcium-related role for STC-2. Salmon STC-2 has fewer histidine residues than human, eel, or coho salmon STC-1. The increase in the number of histidine residues in STC-2 on transition from fish to mammals may reflect altered roles of STC through evolution.

Thyroid Hormones (THs)

The first clear demonstration of the function of THs was by Gudernatsch in 1912 [19], when he showed that diced horse thyroid gland prematurely turned tadpoles into 'mini-frogs'. In 1925 Allen [20] showed that removal of thyroid cells inhibited metamorphosis and, in 1926, Harington identified the main product of the thyroid gland as thyroxine (T4) [21]. The more metabolically active triiodothyronine (T3) was later identified by Hird and Trikojuus in 1948 [22] but, having been published in The Australian Journal of Science, this was not widely acknowledged. Hird visited Pitt-Rivers in London to discuss his experiments. The discovery of T3 is more commonly attributed to Gross and Pitt-Rivers in 1952 [23].

THs play a significant role in bone remodelling alongside the core systemic hormones previously discussed effecting calcium and bone homeostasis. THs are essential for normal human skeletal development, growth, and maintenance of adult bone mass [24]. THs are also more widely involved in regulation of development and metabolism, particularly of the brain. Both the timing and the quantity of TH delivered to the sites of action are crucial for normal development and metabolism. Insufficient TH during human gestation produces 'cretinism' and intellectual disability from birth, whereas reduced TH levels in adults can result in depression.

Key insights into the physiological and developmental functions of THs and their binding proteins in humans have come from studying evolution of TH distribution in vertebrates. This section will focus on insights into both TH metabolism and into forms of human amyloidosis caused by polymerization of the key TH binding protein transthyretin (TTR). We will show that for TH binding, mammalian TTRs are the exception, not the rule.

TH Distributor Proteins Ensure that THs get from the Thyroid Gland to Target Cells

THs are synthesized in the thyroid gland then secreted into the blood for distribution to all tissues by so-called thyroid hormone distributor proteins (THDPs), which are present in both cerebrospinal fluid (CSF) and blood [25]. More than 99% of TH in blood is THDP-bound. Older literature and some current textbooks state that THs are bound to THDPs due to low solubility in blood, however the solubility limit of THs at pH 7.4 is 2.3 μM, and the concentration of free TH in blood is 24 pM, or only 1/100 000 of the solubility limit [26]. THs are lipophilic, partitioning between lipid and aqueous phases at a ratio of 20 000:1 [27]. This means that THs require protein binding to counteract avid partitioning into cell membranes, to maintain a pool of circulating THs in the blood, and to ensure even tissue delivery [28]. In the bloodstream, THs can dissociate from the THDPs to enter cells, largely via membrane-bound TH transporter proteins such as MCT8, MCT10 (monocarboxylate transporters 8 and 10). THs are then subject to activation or inactivation by the family of deiodinases. In humans, the major form of TH circulating in the blood is T4 (the 'transport form'), whereas the form of the hormone with highest affinity for the nuclear receptors is T3 (the 'active form'). However, there is now a rapidly expanding literature of non-genomic actions of THs other than T3 (e.g. T4, rT3, T2). Deiodinases types 1, 2, and 3 can either active (e.g. T4 to T3) or inactivate (e.g. T4 to rT3; T3 to T2) THs within a cell. Thus, deiodinases confer a tissue-specific level of regulation of TH activity.

Within the cell, THs bind specific cytosolic proteins and are translocated into the nucleus where they bind TH nuclear receptors (TRs). T3 has the highest affinity for TRs and is the usual ligand. TRs are encoded by two genes: TRα and TRβ. Each of these genes produces at least two splice variants resulting in the four main products TRα1, TRα2, TRβ1, and TRβ2. However, TRα2 does not bind T3. Together with comodulator proteins, T3-TRs dimerize with retinoid X receptor (RXR) and bind to TH response elements (TREs) and either positively or negatively regulate expression of specific genes.

TTR is Responsible for Much of the TH Distribution in the Body and in the Brain

THs are distributed by the bloodstream. In humans, THDPs are albumin, thyroxine-binding globulin (TBG), and transthyretin (TTR). In the blood about 75% TH is bound to TBG, 15% to TTR and 10% to albumin. A small fraction is lipoprotein-bound. Albumin, TBG and TTR are synthesized by the liver. TBG is lowest in concentration but has highest affinity for T4, albumin is highest in concentration but has the lowest T4 affinity, whereas TTR has intermediate concentration and affinity for T4. T4 bound to albumin rapidly dissociates, T4 bound to TBG acts as a reservoir for T4 in the blood and TTR is responsible for the bulk of the delivery of T4 to tissues (for kinetic discussion of the free hormone hypothesis and the free hormone transport hypothesis, see [29]). This may be captured by the 'Goldilocks and the Three Bears' analogy: TBG binds THs so tightly it does not participate in biodelivery, albumin binds so weakly that it does not distribute THs effectively, but TTR has affinity for THs that is 'just right', allowing appropriate distribution and cellular delivery of THs [30].

Blood is separated from the cerebrospinal fluid (CSF) by the blood-CSF barrier. The choroid plexus, located in the lateral, third, and fourth ventricles of the brain, forms this barrier. The only THDP synthesized by the brain is TTR, which is synthesized in the choroid plexus and is involved in the transport of T4 from the blood into the CSF [27, 31, 32, 33]. Only a small amount of albumin and TBG are present in the CSF due to the leakiness of the blood-brain

barrier. Thus, TTR is responsible for the majority of TH distribution in the CSF.

Evolution of TTR Synthesis in the Body and Brain

TTR is synthesized by the livers of all classes of vertebrates at some stage during their life cycle (Figure 1.4.3) ([34]; for a review, see [29]). In some animals this is only during development (e.g. fish, amphibians, reptiles, monotremes and Australian polyprotodont marsupials) whereas in birds, diprotodont marsupial and eutherians ('placental mammals') TTR is synthesized in the liver throughout life. In general, animals that are homeothermic ('warm blooded') synthesize TTR in their livers throughout life. THs are known to be involved in the regulation of body temperature. Poikilothermic ('cold blooded') animals synthesize TTR in their livers only during specific stages of development that are regulated by THs. From comparative endocrinology, we see that the function of TTR synthesized by the liver is related to an increased demand for TH distribution: during TH-regulated development and homeothermy, which is governed largely by THs.

TTR synthesis by the choroid plexus probably began with the stem reptiles about 320 million years ago [35]. The stem reptiles are the common ancestors to the reptiles, birds and mammals (but not to amphibians and fish). Thus, TTR synthesis in the brain is a more recent event than TTR synthesis in the liver. The selection pressure for turning on the TTR gene in the choroid plexus could be the increase in brain volume, which occurred with the first traces of the cerebral cortex in the stem reptiles [36]. This increase in lipid volume could have been the selection pressure requiring a protein to better distribute TH around the brain via the CSF (see [26]). From comparative endocrinology, we see that the function of TTR synthesized by the brain is to counteract the partitioning of THs into the increased lipid pool and to ensure appropriate distribution of THs throughout the CSF.

Evolution of TTR Structure and Function

The amino acid sequences of TTRs from more than 20 vertebrate species (including fish, amphibians, reptiles, birds, marsupials, and eutherians) have been determined (Figure 1.4.4). The amino acids in the central channel of the TTR homo-tetramer that are involved in ligand binding were 100% conserved between species. A surprising finding, therefore, was that TTRs from lower vertebrates bind T3 with higher affinity than T4 [37]. Only TTRs from mammals bind T4 with higher affinity than T3. Given that the amino acids in the TH binding sites are identical in all TTRs sequenced to date, regions of the protein that changed during vertebrate evolution were sought and considered for influence of ligand binding. The region with the highest rate of evolution is the N-terminal region of the TTR subunit. This changed from longer and more hydrophobic in lower vertebrates to shorter and more hydrophilic in eutherian mammals. These structural characteristics correspond with ligand binding preferences: TTRs with longer and more hydrophobic N-termini bind T3 with higher affinity than T4, whereas TTRs with shorter and more hydrophilic N-terminal regions bind T4 with higher affinity than T3 [37]. A series of chimeric TTRs were generated, where N-terminal regions were swapped or deleted. These studies confirmed the hypothesis that the N-terminal region of the TTR subunit confers ligand preference and affinity [38, 39]. The mechanism for shortening the N-terminal region of the TTR subunit was found to be the stepwise shift in the position of the intron 1/exon 2 splice

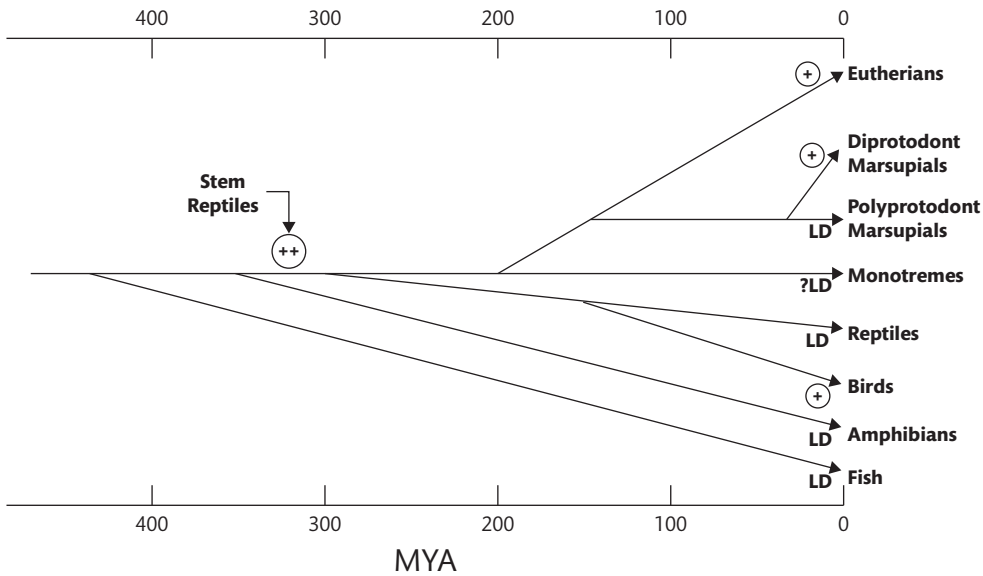


Figure 1.4.3 Evolutionary/developmental tree for TTR synthesis in the choroid plexus and liver of vertebrates. Evolutionary tree based on the fossil record indicating onset of TTR synthesis in vertebrates. ++, onset of TTR synthesis in the choroid plexus, in juveniles and in adult of extant species; LD, hepatic TTR synthesis during development only;?LD, possible onset of hepatic TTR synthesis during development only; +, hepatic TTR synthesis during development and in adult; MYA, millions of years ago.

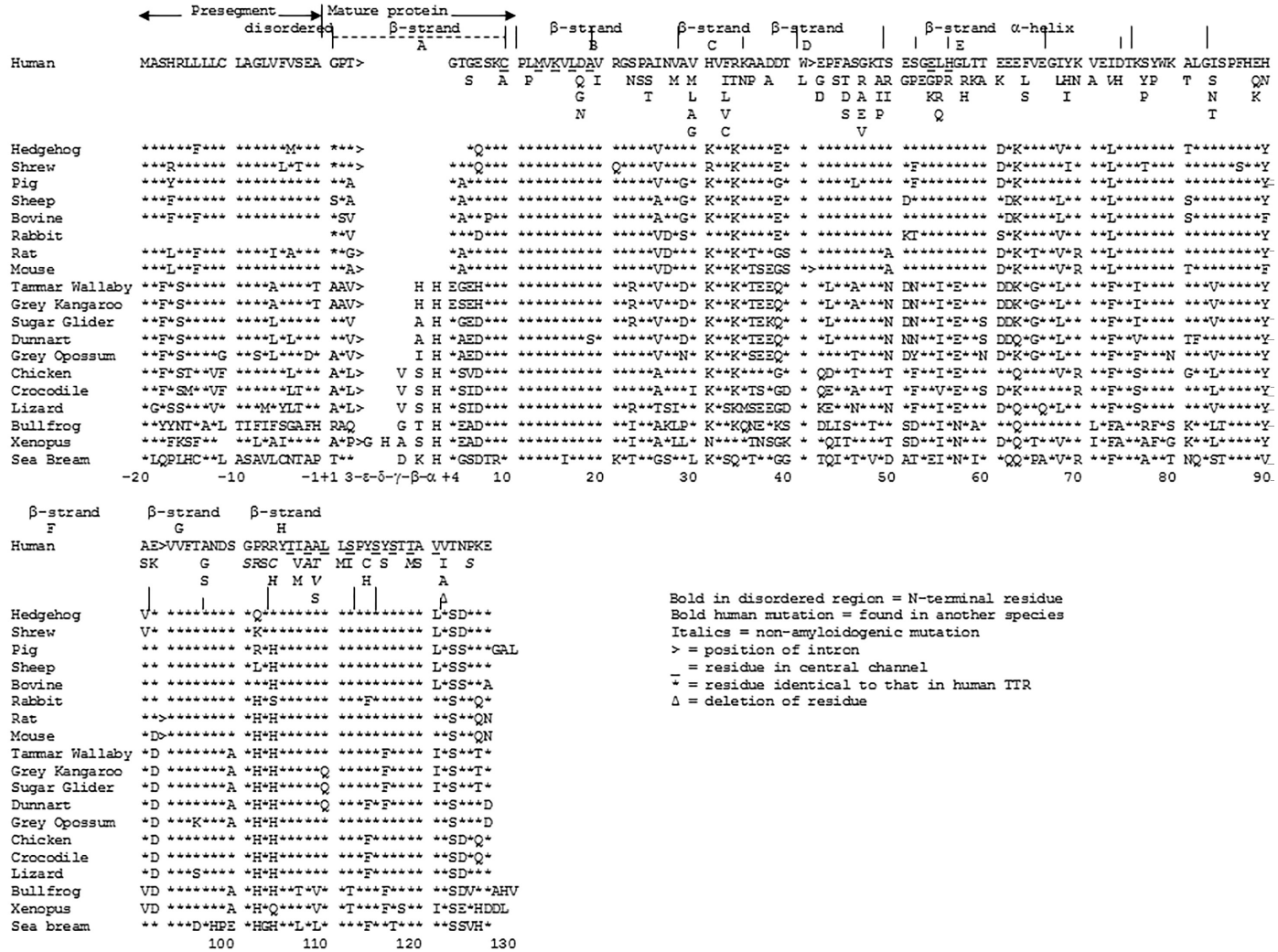


Figure 1.4.4 Human TTR mutations compared with evolutionary mutations in vertebrate TTRs. TTR amino acid sequences and those derived from cDNA sequences from 19 species aligned with human TTR. Secondary structural features for human TTR are indicated. Amino acids in other species identical to those in human TTR are indicated with asterisks. Point mutations detected in human TTRs are indicated below the human TTR sequence.

site in the 3' direction [40]. This situation is very unusual, as the TH binding site in TTR does not determine which form of TH is preferentially bound.

TTR Evolved from TTR-Like Protein, a 5-Hydroxyisourate Hydrolase

There is a very high degree of amino acid identity across all vertebrate classes, suggesting that the TTR gene evolved prior to the divergence of the vertebrates from the invertebrates [41]. Therefore, genomes of non-vertebrates were screened for open reading frames (ORFs) similar to TTR genes that could code for TTR-like proteins (TLPs). More than 100 ORFs coding for TLPs were identified, from all kingdoms [42]. It was proposed that the TTR gene evolved as a duplication of the TLP gene. Five motifs were identified as defining TLPs and three motifs were identified as defining the set of TTRs + TLPs. These motifs mapped to structurally conserved and functionally important regions of the proteins. Transcription of these ORFs was demonstrated in a bacterium, a plant, and an invertebrate animal, demonstrating their existence in nature. These TLPs had similar molecular weights and tetrameric structures to vertebrate TTRs. A subsequent study revealed the X-ray crystal structure of *Salmonella dublin* TLP to be extremely similar to that of vertebrate TTRs [43]. Thus, the 3D structure of TLP/TTR has not changed from bacteria to humans. The fact that the structure was so highly conserved throughout evolution implies that this protein serves a very important function. Even within vertebrates, the structure of TTR is extremely highly conserved, similarly to the conservation of the histones (involved in compacting DNA in eukaryotes), whose structures are among the highest conserved. This remarkable degree of conservation of structure was only revealed through comparative biochemistry (see [44]).

TLPs do not bind THs [45]). TLPs from *Salmonella*, mouse, and zebrafish are enzymes involved in the oxidation of uric acid to allantoin. TLP is a 5-hydroxyisourate hydrolase, which hydrolyses 5-hydroxyisourate (5-HIU) to 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) [43, 46, 47]. A series

of *Salmonella* TLPs with point mutations identified three conserved residues in the catalytic site that are essential for enzyme activity [43]. Comparison between the TH binding site in TTR and the equivalent region in *Salmonella* TLP revealed that the binding site was shallower and more positively charged in TLP compared to that of TTR, explaining why THs cannot be bound by TLPs [37]. TLP/TTR is a remarkable example of conservation of protein structure, but evolution of the protein's function: from a 5-HIUase to distributor of T3 to distributor of T4 (for greater detail, see [48]).

Evolution of TTR Structure Sheds Light on Human Amyloidoses: FAP and SSA

TTR can form amyloid deposits in humans, leading to two types of TTR amyloidosis: familial amyloidotic polyneuropathy (FAP) and senile systematic amyloidosis (SSA). The TTR tetramer is usually very stable. However, it can form amyloid fibrils naturally *in vivo*, and can be induced to form amyloid *in vitro* [49]. There are two types of TTR amyloid. SSA is an age-dependent disease and the TTR fibrils are formed from wild type protein. At least 65% of people over 70 years old have TTR SSA (see [50]). By contrast, FAP is a specific form of autosomal dominant hereditary polyneuropathy, which initially manifests as systemic deposition of amyloid in the peripheral nerves, but later effects many visceral organs. Many of the FAP mutations are named after their country of origin. Examples of this are being able to track the travels of the Vikings, as they left a trail of their FAP mutations in Scotland and in England; and the Portuguese explorers left a trail of their FAP mutations to Japan and back. Interestingly, there are about 200 words that are similar in Portuguese and Japanese languages.

There are at least 100 point mutations that have been documented in the 127 amino acids in the TTR subunit which result in FAP (see [51]). These mutations are evenly spread throughout the length of the TTR subunit, which is in sharp contrast to the evolutionary mutations, which are concentrated in the N-terminal regions of the subunit. TTR amyloidosis has not yet been described in a non-human species. Of the mutations in human TTR that result in amyloidosis, five are found in other species but do not result in amyloidosis: Val30Leu, Glu42Asp, Ile68Leu, Tyr69Ile and Ala81Thr (**Figure 1.4.4**) (see [29]). Of these, Leu30 is only found in sea bream TTR and Asp42 is only found in bullfrog TTR, both of which are evolutionarily quite distant from humans. However, the Ile68Leu substitution is found in seven mammalian species studied to date. Further investigation of these TTRs and comparison to the mutated human Leu68 could give valuable insight as to why Leu is tolerated in position 68 without amyloidosis formation in other mammalian species, but leads to amyloid formation in humans. This could result in understanding the molecular basis of TTR amyloid formation in humans.

All five point mutations in human TTR resulting in amyloidosis, that are the normal residue in that position in TTRs from other species, result in cardiac amyloid deposition in humans. This is highly intriguing and requires further investigation.

Implications of the Evolution of Calcium Regulation and Thyroid Hormone Distributors for Clinical Endocrinology

In summary, there has always been considerable synergy between comparative and mammalian endocrinology, with information from one field providing stimulus to the other. STC was originally identified in fish and subsequently found in mammals. Previous studies on fish stanniocalcin provided information about the potential roles of human STC-1 and the isolation of mammalian and fish STC-2 will stimulate the search for its endocrine and physiological roles in both fish and mammals. In contrast, PTHrP and PTH were originally identified in mammals and then the search for the orthologous genes in fish began. There has been considerable interest in calcium regulating factors from lower vertebrates because of the efficacy of salmon calcitonin in inhibiting bone resorption in mammals, including man. Certainly, a number of groups are interested in isolating fish PTH, or possibly PTHrP, as these fish factors or structural analogues could be potential treatments for osteoporosis in human subjects.

The identification of a second STC in both fish and mammals reflects the principle that Niall argued in 1982 [52] that nature duplicates a gene rather than creating a new gene. Through this mechanism, when a gene for a hormone is duplicated the copy is free to mutate and acquire new function(s) and receptors, leaving the first hormone carrying out its original role. This could also be the case with the duplication of PTH and PTHrP and calcitonin and calcitonin gene-related peptide. In his review, Ohno argues that for every gene present in invertebrates there could be up to four copies present in vertebrate genomes [12]. He bases his argument on the fact that there are 15 000 protein-coding genes in invertebrate genomes while there are 60 000 in vertebrate genomes as a result of two rounds of whole genome duplication. He argues that there could be up to four copies of each gene in jawed vertebrates compared to one copy of each gene in tunicates. The current evidence supports Ohno's theory. This one-to-four rule could apply to a number of other genes including the parathyroid hormone and calcitonin gene families. First it is important to establish if there are members of these gene families present in an invertebrate genome.

There are a number of other genes and hormones that have been identified in fish that have not yet been identified in mammals. One of these is somatolactin, a member of the growth hormone/prolactin family [53]. Its relationship with two of the major vertebrate hormones should stimulate the search for it in the human genome and uncover its function in both lower and higher vertebrates. This will be another project where interaction between the two fields of endocrinology will prove fruitful.

Now a number of whole genome sequencing projects, including the elephant shark, pufferfish (*Fugu rubripes)* and human are completed, and other genomes are currently being sequenced (e.g. Japanese lamprey) more information about the genes for calcium regulating factors as well as a number of other hormones are becoming accessible. When the complete sequences and genome structure are known, a shift in focus from gene to protein will be required leaving comparative endocrinology to answer the following intriguing questions: what are the roles of these factors in lower

vertebrates, compared to higher vertebrates, and how have evolutionary events modified these roles?

The evolution of the structure and function of TTR has given insights into the tissue-specific requirements of THs, both during development and during evolution. TLPs exist in all kingdoms, whereas TTRs only exist in vertebrates. The highly conserved structure of TLP was very slightly modified to allow binding of THs in the central channel. The N-terminal regions of each subunit were modified during evolution to change TTR from distributing T3 to distributing T4. This presumably resulted in an increased complexity of deiodinases at the tissue level. That certain amino acids in specific positions are tolerated in some species, but result in amyloidosis in human TTR should be exploited in order to understand the mechanism(s) of TTR amyloid formation in humans.

There are large amounts of data being generated by whole genome sequencing projects of a wide range of vertebrates and invertebrates currently being undertaken. There will be a number of exciting opportunities for collaboration between comparative and clinical endocrinology to learn more about the evolution of human endocrine conditions and possibly potential new treatments for diseases.

REFERENCES

1. Ingleton PM, Danks JA. Distribution and functions of parathyroid hormone-related protein in vertebrate cells. *Int Rev Cytol*, 1996; **166**: 231–80.
2. Parsons JA, Gray D, Rafferty B, Zanelli JM. Evidence for a hypercalcaemic factor in the fish pituitary immunologically related to mammalian parathyroid hormone. In: Copp DH, Talmage RV, eds. *Endocrinology of Calcium Metabolism*. Amsterdam: Excerpta Medica, 1979: 111–14.
3. Danks JA, Ho PM, Notini AJ, *et al.* Identification of a parathyroid hormone in the fish Fugu rubripes. *J Bone Miner Res*, 2003; **18**: 1326–31.
4. Suarez-Bregua P, Ankur Saxena E, *et al.* Pth4, an ancient parathyroid hormone lost in eutherian mammals, reveals a new brain-to-bone signaling pathway. *FASEB J*, 2017; **31**: 569–83.
5. Pinheiro PLC, Cardoso JCR, Power DM, Canario AVM. Functional characterization and evolution of PTH/PTHrP receptors: insights from the chicken *BMC Evolutionary Biology*, 2012; **12**: 110.
6. Liu Y, Ibrahim AS, Richardson SJ, *et al.* The parathyroid gene family in a cartilaginous fish, the elephant shark (*Callorhinchus milli*). *J Bone Miner Res*, 2010; **25**: 2613–23.
7. Yu WP, Rajasegaran V, Yew K, *et al.* Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: a comparative analysis of the protocadherin cluster. *Proc Natl Acad Sci USA*, 2008; **105**: 3819–24.
8. Danks JA, Devlin AJ, Ho PM, *et al.* Parathyroid hormone-related protein is a factor in normal fish pituitary. *Gen Comp Endocrinol*, 1993; **92**: 201–12.
9. Philbrick WM, Wysolmerski JJ, Galbraith S, *et al.* Defining the roles of parathyroid hormone-related protein in normal physiology. *Physiol Rev*, 1996; **76**: 127–73.
10. Wysolmerski JJ, Stewart AF. The physiology of parathyroid hormone-related hormone: an emerging role as a developmental factor. *Ann Rev Physiol*, 1998; **60**: 431–60.
11. Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosome regions in man and the house mouse. *Genomics*, 1993; **16**: 1–19.
12. Ohno S. The one-to-four rule and paralogues of sex-determining genes. *Cellular and Mol Life Sci*, 1999; **55**: 824–30.
13. Trivett MK, Officer RA, Clement JG, *et al.* Parathyroid hormone-related protein (PTHrP) in cartilaginous and bony fish tissues. *J Exp Zool*, 1999; **284**, 541–8.
14. Rubin DA, Hellman P, Zon LI, Lobb CJ, Bergwitz C, Jüppner H. A G-protein-coupled receptor from zebrafish is activated by human parathyroid hormone and not by human or teleost parathyroid hormone-related protein. *J Biol Chem*, 1999; **274**: 23035–42.
15. Yoshiko Y, Son A, Maeda S, *et al.* Evidence for stanniocalcin gene expression in mammalian bone. *Endocrinol*, 1999; **140**: 1869–74.
16. Chang AC, Janosi J, Hulsbeek M, *et al.* A novel human cDNA highly homologous to the fish hormone stanniocalcin. *Mol Cell Endocrinol*, 1995; **112**: 241–7.
17. Chang ACM, Reddel RR. Identification of a second stanniocalcin cDNA in mouse and human stanniocalcin 2. *Mol Cell Endocrinol*, 1998; **141**: 95–9.
18. Wagner GF, Jaworski EM, Haddad M. Stanniocalcin in the seawater salmon: structure, function, and regulation. *Am J Physiol*, 1998; **274**: R1177–85.
19. Gudernatsch JG. Feeding experiments on tadpoles I. The influence of specific organs given as food on growth and differentiation: a contribution to the knowledge of organs with internal secretion. *Arch Entwicklungsmech Org*, 1912; **35**: 457–81.
20. Allen BM. The effects of extirpation of the thyroid and pituitary glands upon the limb development of anurans. *J Exp Zool*, 1925; **42**: 13–30.
21. Harington CR. Chemistry of thyroxine: isolation of thyroxine from the thyroid gland. *Biochem J*, 1926; **20**: 293–9.
22. Hird FJR, Trikojus VM. Paper partition chromatography with thyroxine analogues. *Aust J Sci*, 1948; **10**: 185–7.
23. Gross J, Pitt-Rivers R. The identification of 3:5:3′-L-triiodothyronine in human plasma. *Lancet*, 1952; **1**: 439–41.
24. Bassett JH, Williams GR. The molecular actions of thyroid hormones in bone. *Trends Endocrinol Metab*, 2003; **14**: 356–64.
25. Alberts B, Johnson J, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*, 4th edn. New York: Garland Science, 2002: 840.
26. Schreiber G, Richardson SJ The evolution of gene expression, structure and function of transthyretin. *Comp Biochem Physiol*, 1997; **116B**: 137–60.
27. Dickson PW, Aldred AR, Menting JGT, Marley PD, Sawyer WH, Schreiber G. Thyroxine transport in choroid plexus. *J Biol Chem*, 1987; **262**: 13907–15.
28. Mendel CM, Weisiger RA, Jones AL, Cavalieri RR. Thyroid hormone-binding proteins in plasma facilitate uniform distribution of thyroxine within tissues: a perfused rat liver study. *Endocrinol*, 1987; **120**: 1742–9.
29. Richardson SJ. Cell and molecular biology of transthyretin and thyroid hormones. *Int Rev Cytol*, 2007; **258**, 137–93.
30. Alshehri B, D'Souza DG, Lee JY, Petratos S, Richardson SJ. The diversity of mechanisms influenced by transthyretin in neurobiology: development, disease and endocrine disruption. *J Neuroendocrinol*, 2015; **27**: 303–23.
31. Southwell BR, Duan W, Alcorn D, *et al.* Thyroxine transport to the brain: role of protein synthesis by the choroid plexus. *Endocrinol*, 1993; **133**: 2116–26.
32. Schreiber G, Aldred AR, Jaworowski A, Nilsson C, Achen MG, Segal MB. Thyroxine transport from blood to brain via

transthyretin synthesis in choroid plexus. *Am J Physiol*, 1990; **258**: R338–5.
33. Chanoine J-P, Alex S, Fang SL, *et al.* Role of transthyretin in the transport of thyroxine from the blood to the choroid plexus, the cerebrospinal fluid, and the brain. *Endocrinol*, 1992; **130**: 933–8.
34. Richardson SJ, Monk JA, Shepherdley CA, *et al.* Developmentally regulated thyroid hormone distributor proteins in marsupials, a reptile and fishes. *Am J Physiol*, 2005; **288**: R1264–72.
35. Achen MG, Duan W, Pettersson TM, *et al.* Transthyretin gene expression in choroid plexus first evolved in reptiles. *Am J Physiol*, 1993; **265**: R982–9.
36. Kent GC. *Comparative Anatomy of the Vertebrates*. St Louis: Time Mirror/Mosby College, 1987: 542.
37. Chang L, Munro SLA, Richardson SJ, Schreiber G. Evolution of thyroid hormone binding by transthyretins in birds and mammals. *Eur J Biochem*, 1999; **259**: 634–42.
38. Prapunpoj P, Richardson SJ, Schreiber G. Crocodile transthyretin: structure, function and evolution. *Am J Physiol*, 2002; **283**: R885–96.
39. Prapunpoj P, Leelawatwatana L, Schreiber G, Richardson SJ. Change in structure of the N-terminal region of transthyretin produces change in affinity of transthyretin to T4 and T3. *FEBS J*, 2006; **273**: 4013–23.
40. Aldred AR, Prapunpoj P, Schreiber G. Evolution of shorter and more hydrophilic transthyretin N-termini by stepwise conversion of exon 2 into intron 1 sequences (shifting the 3′ splice site of intron 1). *Eur J Biochem*, 1997; **246**, 401–9.
41. Prapunpoj P, Yamauchi K, Nishiyama N, Richardson SJ, Schreiber G. Evolution of structure, ontogeny of gene expression and function of *Xenopus laevis* transthyretin. *Am J Physiol*, 2000; **279**: R2026–41.
42. Hennebry SC, Wright HM, Likic V, Richardson SJ. Structural and functional evolution of transthyretin and transthyretin-like proteins. *Proteins: Struct, Funct Bioinformat*, 2006; **64**: 1024–45.
43. Hennebry SC, Buckle AM, Law RH, Richardson SJ, Whisstock JC. The crystal structure of the transthyretin-like protein from *Salmonella dublin* reveals the structural basis for 5-hidroxyisourate hydrolase activity. *J Mol Biol*, 2006; **359**: 1389–99.
44. McLean TR, Rank MM, Smooker PM, Richardson SJ. Evolution of thyroid hormone distributor proteins. *Mol Cell Endocrinol*, 2017; **459**: 43–52.
45. Eneqvist T, Lundberg E, Nilsson L, Abagyan R, Sauer-Eriksson AE. The transthyretin-related protein family. *Eur J Biochem*, 2003; **270**: 518–32.
46. Lee Y, Lee do H, Kho CW, *et al.* Transthyretin-related proteins function to facilitate the hydrolysis of 5-hydroxyisourate, the end product of the uricase reaction. *FEBS Letts*, 2005; **579**: 4769–74.
47. Ramazzina I, Folli C, Secchi A, Berni R, Percudani R. Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. *Nat Chem Biol*, 2006; **2**: 144–8.
48. Richardson SJ. Tweaking the structure to radically change the function: the evolution of transthyretin from 5-hydroxyisourate hydrolase to triiodothyronine distributor to thyroxine distributor. *Front Endocrinol (Lausanne)*, 2015; **5**: e245.
49. Colon W, Kelly JW. Partial denaturation of transthyretin is sufficient for amyloid fibril formation *in vitro*. *Biochemistry*, 1992; **31**: 8654–60.
50. Calkins E. Amyloidosis. In: Wintrobe WM, Thom GW, Adams RD, eds. *Harrison's Principles of Internal Medicine*, 7th edn. Toyko: McGraw-Hill Kogakusha, 1974: 644–7.
51. Benson MD. Genetics: clinical implications of TTR amyloidosis. In: Richardson SJ, Cody V, eds. *Recent Advances in Transthyretin Structure, Function and Biological Functions*. Basel: Springer, 2009: 173–89.
52. Niall HD. The evolution of peptide hormones. *Ann Rev Physiol*, 1982; **44**: 615–24.
53. Rand-Weaver M, Noso T, Muramoto K, Kawauchi H. Isolation and characterization of somatolactin, a new protein related to growth hormone and prolactin from Atlantic cod (*Gadus morhua*) pituitary glands. *Biochem*, 1991; **30**: 1509–15.

FURTHER READING

Ono S, Wolf U, Atkins NB. Evolution from fish to mammals by gene duplication. *Hereditas*, 1968; **59**: 169–87.
Richardson SJ, Cody V, eds. *Recent Advances in Transthyretin Structure, Function and Biological Functions*. Basel: Springer, 2009.

1.5

Hormones Across the Lifespan

James Gibney, Indraneel Banerjee, and Ken K.Y. Ho

Introduction

The endocrine system plays a central role in growth, development, metabolism, reproduction, and physical well-being throughout life. A complex developmental programme shapes distinct time-dependent changes in various components of the endocrine system during different phases of postnatal, childhood, and adult life. This permits context-dependent modulation of this multifaceted control system to subserve distinct functions according to the differing demands of different life stages. Regulation is critically dependent on feedback, which is configured distinctly in each of the reproductive, adrenal, thyroid, and growth axes from fetal to adult life.

There are three main points in the life course when pituitary function is remodelled. First, a remarkable programme of adaptation prepares the fetal endocrine system for extrauterine life no longer dependent on maternal factors, with a second shift marking the acquisition of sexual maturity at puberty. The much more gradual process of senescence in later life, which features particularly notable sexually dimorphic changes in the reproductive system (abrupt in women and gradual in men) constitutes a third phase of change.

The profiles of hormones throughout life are well characterized and provide a framework for assessing the integrity of the endocrine system in health and disease. This chapter will describe the profiles of the gonadal, adrenal, growth hormone, thyroid, and prolactin axes covering the paediatric and adult years of life. The paediatric phase spans the neonatal, childhood, and adolescent years. The endocrinology of fetal development is a major topic whose full treatment is beyond the scope of this chapter, however selected aspects are covered to provide an understanding of changes in the transition to postnatal life.

Gonadal Axis

Among the pituitary hormonal axes, the gonadal axis displays the greatest dynamic, developmental change, exhibiting distinct patterns in postnatal, childhood, and adult life that are moreover distinct in women during and at the end of reproductive life. The major measurable elements of this system are pituitary gonadotrophins and the gonadal products sex hormones and inhibin.

Development *In Utero* and Childhood

The hypothalamic–pituitary–gonadal axis (HPA) is fully intact by mid-fetal life. Its function is suppressed in the latter half of pregnancy by sex steroids produced from the fetal gonad under the influence of placental human chorionic gonadotropin (HCG) and rising maternal oestrogen levels. After placental restraint is removed at birth, gonadotrophin levels rise in early infancy, reaching a peak around 2–3 months, causing a parallel rise in sex steroid levels in early infancy (**Figures 1.5.1 and 1.5.2**). The term minipuberty is applied to this phase in early life when gonadotrophins and sex steroid levels are high, although sex hormone receptor insensitivity ensures lack of translation into secondary sexual characteristics [1]. Luteinizing hormone (LH), follicle-stimulating hormone (FSH), and sex steroid levels then decline by about six months in boys and 1–2 years in girls to prepubertal levels, entering the so-called juvenile pause, determined by neural mechanisms suppressing gonadotropin-releasing hormone (GnRH) secretion. The cause and significance of minipuberty, and the mechanisms determining the secondary inactivation of the axis, are not well understood.

The gonadal axis profile remains relatively quiescent throughout childhood, although very low amplitude pulsatile secretion of gonadotrophins does occur, with FSH levels tending to be higher than LH. Nocturnal sleep-entrained pulsatile secretion of hypothalamic GnRH activation marks the onset of puberty, driven by a hypothalamic pulse generator in which the neuropeptide kisspeptin plays a key role. During pubertal development, gonadotrophins rise, with LH levels higher than FSH, in concert with increases in the amplitude and frequency of episodic secretion extending into the

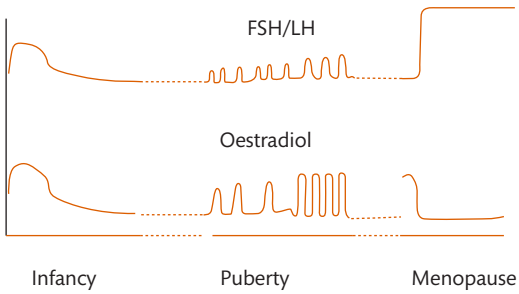


Figure 1.5.1 Schematic representation of gonadotropin (LH/FSH) and oestradiol levels in females throughout the lifespan. The figure shows concentrations in infancy through pubertal, reproductive, and menopausal phases. The LH/FSH levels depict the summated curve within a diurnal profile characterized by secretory pulses of varying amplitude and frequency. The waveform excursions during puberty represent changes during the menstrual cycle.

daytime. The progressive and sustained increase in gonadotrophin secretion leads to maturation of the gonads and production of sex steroids reaching a peak in late puberty, and occurring earlier in females than males (Figure 1.5.1).

Adulthood

Males

A gradual and progressive decrease in serum total testosterone occurs from age 20 to 80 years, with a recent study of 10 000 healthy men demonstrating a decline of 0.5% annually [2]. See Figure 1.5.2. The decline is more marked in those with chronic illness and in those who are overweight [3]. An associated increase in sex hormone-binding globulin occurs with ageing, resulting in a more marked reduction of free (bioavailable) testosterone.

The decline in testosterone results from reduced production rates and a fall in plasma clearance. Reduced testosterone production arises from changes at all levels of the HPA, including gonadotropin secretion and testicular steroidogenesis; There are also changes in negative and positive feedback interactions [4]. Testosterone displays a diurnal profile of high levels in the early morning, falling to a nadir in the late afternoon. The diurnal rhythm of testosterone production is less marked in older compared to younger men.

Mean LH and FSH levels increase with increasing age. However, the increase in LH is not proportionate to the decline in testosterone indicating alteration in the classical feedback loop. The sensitivity of pituitary LH secretion to androgen-mediated feedback inhibition is increased, and the ability of LH to stimulate testicular testosterone secretion is inhibited.

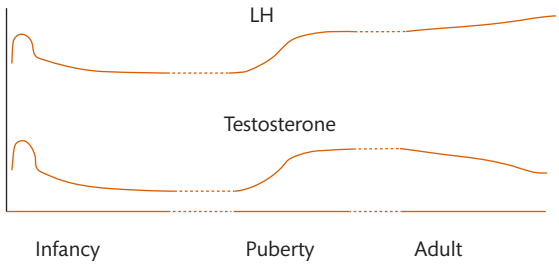


Figure 1.5.2 Schematic representation of gonadotropin LH and testosterone levels in males throughout the lifespan. The figure shows concentration in the neonatal and pubertal periods, and in adult life. The LH levels depict a summated curve within a diurnal profile characterized by secretory pulses of varying amplitude and frequency.

Changes in testicular morphology influence circulating hormone levels. Inhibin produced by Sertoli cells falls with age; in combination with increased FSH levels this indicates decline in Sertoli cell function. There is also a reduction in Leydig cell volume and number which probably contributes to reduced testosterone concentration.

As in younger subjects, interpretation of serum testosterone levels in older men is guided by timing and conditions of sampling, the validity of lower limit of the normal reference range, and whether symptoms of hypogonadism are present [5]. Current evidence that the lower level observed in later life merits supplementation to improve heath is poor.

Females

The phasic changes in gonadotrophins, oestrogen and progesterone concentrations that typify the female menstrual cycle are well covered elsewhere and will not be described here. See Figure 1.5.1.

In contrast to other hormonal axes in which there is a gradual change in hormone levels with ageing, a major change in female reproductive hormones occurs at the time of menopause. This stems from near exhaustion of the oocytes in the ovaries, which is the culmination of a gradual decrease in quantity and quality of oocytes in the ovarian cortex across reproductive life [6]. At the time of menarche, there are approximately 300 000 primordial follicles. During reproductive years, there is a gradual but accelerating decline, leaving fewer than 1000 follicles at the time of menopause [7]. The first detectable evidence of reproductive ageing is usually a lengthening of the menstrual cycle by 2 to 3 days. Subsequently insufficient availability of follicles further lengthens cycles, leading to missed periods or phases of cycle arrest. This is referred to as the menopausal transition and lasts until the final menstrual period. The average age of onset of the transition is 46 years, and the mean age of the final menstrual period is 51.

Increased FSH levels associated with decreased inhibin levels but normal LH levels accompany perimenopausal changes in menstrual cycle pattern, usually associated with slightly elevated oestradiol levels. These changes reflect decreasing ovarian reserve and are best detected on day 2 or 3 of the menstrual cycle. Serum oestradiol levels begin to decline less than a year before menopause. Ovarian production of inhibin begins to decrease after 30 years and this becomes more pronounced after age 40.

In the absence of oestrogen and inhibin from the menopause, LH and FSH levels rise sharply; LH pulse amplitude but not frequency is increased. FSH and LH levels peak a few years after menopause and slowly decline in subsequent years. Increased LH maintains significant production of androstenedione and testosterone from the ovary albeit at lower levels than in premenopausal women. Circulating oestradiol and oestrone are slightly less than average male levels postmenopausally and almost all of these steroids is derived from peripheral aromatization of androstenedione. These changes result in a marked increase in the ratio of androgen to oestrogen levels.

The function of the gonadal axes can be disrupted by many factors including stress, nutrition, body weight, and systemic illness in the absence of pathological lesions to the hypothalamus, pituitary, and gonads.

Adrenal Axis

The ontogeny of the pituitary–adrenal system features one progressive change, reflecting the diverse demands on the adrenal gland as part of electrolyte, metabolic, androgen, and cardiovascular homeostasis, and in systemic stress response.

ACTH

Development *In Utero* and Childhood

The fetal adrenal cortex is functionally active by 10 weeks [8]. It contributes significantly in the production of androgens, particularly dehydroepiandrosterone (DHEA) [9]. The fetal zone of the adrenal gland expresses high levels of sulfotransferase activity leading to a greater proportion of DHEA and DHEAS relative to cortisol and aldosterone. In turn DHEA and DHEAS are aromatized by the placental trophoblast to oestrogen, contributing significantly to the maternal pool.

The HPA axis is fully functional at birth in term infants. In babies born prematurely, adrenal function is often suboptimal, due to functional immaturity and the influence of antenatal maternal steroids [10]. Diurnal cortisol secretion within the fetal circulation is present by late gestation, controlled by maternal stimuli [11]. Following birth, cortisol secretion is unsynchronized for several months until the adrenal clock is entrained to the day–night cycle, mediated by ACTH (adrenocorticotropic hormone) secretion [12].

ACTH is secreted in a complex pattern featuring pulsatile and basal components; 12 to 30 pulses occur per day, with ACTH having a short half-life of 14–35 minutes. ACTH secretion in turn is driven by CRH rhythms; its diurnal pattern originates from the circadian pacemaker located in the suprachiasmatic nuclei of the hypothalamus. ACTH and cortisol secretion parallel one another, with a characteristic profile that peaks in the early morning, falling gradually through the day to a nadir around 12 midnight, before increasing again in the early morning.

Adulthood

Mean ACTH levels frequently sampled over 24 hours do not change with age. More subtle changes in the pattern of ACTH secretion do occur however; increasing age is associated with greater asynchrony of ACTH and cortisol secretion, and a greater proportion of brief sharp fluctuations in ACTH levels [13]. A recent intriguing finding has been the identification of a significant association between genes encoding the ACTH receptor and familial longevity [14].

Parallel age-related reductions in cortisol secretion rates and cortisol metabolic clearance result in serum cortisol levels that are unchanged or increased; clear changes in the diurnal pattern exist, however. Frequent sampling of cortisol levels over 24 hours has demonstrated an age-related phase advance of the circadian pattern with earlier nadir and peak levels (up to 3 hours), no difference in the interval between nadir and peak levels, and increased levels at the time of the nocturnal nadir, indicating reduced inhibition of nocturnal cortisol secretion. The largest study to address the effect of age on mean 24-hour cortisol levels report an increase of 20–50% between 20 and 80 years [15] although not all studies support this. Changes in the diurnal pattern of cortisol secretion reflect changes in the output of the circadian pacemaker, and potentially influence the sleep-wake cycle of older adults who tend to have earlier sleep onset, earlier morning awakening. and a more fragmented and shallow sleep.

The diurnal variation in axis function, with cortisol excursion varying across a 5–6-fold range means that caution must be exercised in assessing the integrity of the system using any isolated cortisol value, particularly if the sample time is unknown. The HPA axis plays a vital survival role by subserving the flight and fright response. The plasticity required for such immediate adaptation engenders enormous variability in ACTH-cortisol excursions in daily life.

DHEA

Childhood

Adrenal androgen production increases at around ages 5–8 years without a concomitant increase in cortisol secretion. This phenomenon is called adrenarche, and coincides with the emergence of mild androgenization including axillary and pubic hair appearance [16]. The trigger for adrenarche is poorly understood but is thought to arise from changes intrinsic to the biochemical machinery of the adrenal cortex that maximize DHEAS production. DHEAS levels later progressively rise through puberty, coinciding with appearance of androgen-driven secondary sexual characteristics. Activation of adrenal androgen secretion is usually clinically imperceptible, merging into centrally mediated puberty. DHEA and DHEAS levels peak later in adult life.

Adulthood

Serum DHEAS concentration in adult life is over 100-fold higher than the concentration of testosterone, and over 2000 times the concentration of oestrogens. DHEA and DHEAS form a pool of precursors that can be converted in various tissues to androgens and oestrogens. Most of the sex steroids in postmenopausal women are derived from DHEA and related adrenal precursors. There has been great interest in the biology of DHEAS in health because of observational studies linking higher levels to longevity and better cognition. In hypoadrenal patients, DHEA replacement improves sexual function, vitality, and quality of life [17]. The results from studies of DHEA supplementation in ageing populations with normal adrenal function have been conflicting [18, 19].

Among the adrenal hormones, the most marked age-related effects are seen for DHEA and DHEAS. Maximum levels occur at the beginning of the second decade of life before levels decrease to about 20–30% by the eighth decade, in association with selective involution of the zona reticularis of the adrenal cortex. A large Australian study which included approximately 1400 women demonstrated age-related declines in total and calculated free testosterone, DHEAS and androstenedione with no change in SHBG [20]. The decline was steepest in the early reproductive years, and flattened out in midlife, with a tendency to increase in later years. There was no effect of natural menopause on the rate of decline. Among postmenopausal women who had previously undergone oophorectomy, total and free testosterone levels but not androstenedione and DHEAS were lower than in the remainder of the cohort, indicating that the observed changes in the latter two steroids reflect reduction in adrenal production. The significance of DHEAS levels in the absence of adrenal disease is unknown. There is insufficient evidence warranting its measurement as an indicator of adrenal function.

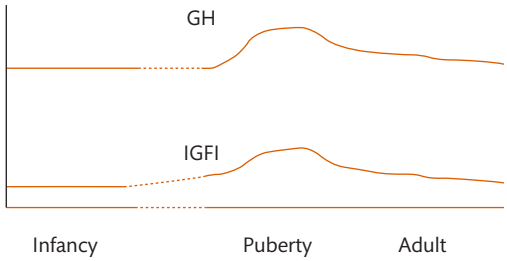


Figure 1.5.3 Schematic representation of growth hormone (GH) and IGF-I levels in males and females across the lifespan. The figure shows concentrations in the neonatal and pubertal periods, and in adult life. The GH levels depict a summated curve within a diurnal profile characterized by secretory pulses of varying amplitude and frequency.

GH Axis

Development *In Utero* and Childhood

The infantile pituitary secretes growth hormone at levels higher than in later childhood, however fetal growth is not dependent on adequacy of growth hormone secretion either from the pituitary or from the placenta. Postnatal growth is more dependent on nutrition than on growth hormone status, such that the manifestations of growth hormone deficiency are only evident after the first 2 years of life [21]. There is limited information on growth hormone (GH) secretion in infancy, but available evidence suggests that GH levels are relatively high in early infancy [22] before stabilizing during childhood until entry into puberty. There is increasing evidence that sensitivity of growth hormone-responsive tissues changes with age and plays a part in determining the magnitude of growth [23].

Spontaneous GH secretion is pulsatile with most occurring at night, triggered by onset of slow-wave sleep. Because secretion of GH is episodic, isolated basal plasma GH concentrations are a poor indicator of overall secretion. Studies employing frequent sampling over a 24-hour period have clearly demonstrated that GH secretion increases at onset of puberty, peaking by 2–3-fold at mid- to-late puberty (Figure 1.5.3). The increase in GH secretion occurs earlier in girls than in boys in parallel with entry and progression through puberty [24]. The changes in spontaneous GH secretion are attributable to changes in pulse amplitude and not frequency.

Adulthood

GH secretion starts to decline during the third decade of life and then progressively thereafter [25]. IGF-1 levels follow a similar trend to GH, increasing 2- to 3-fold at puberty in both sexes, progressively declining with advancing age to about 50% of that observed in the third decade (Figure 1.5.3). Both GH production rate and half-life decline with age, the former by 14% and the latter by 6% per decade [26]. GH secretion is reduced in obesity. Because obesity often increases with age, it is unclear the extent to which the fall in GH secretion is confounded by the age-related increase in adiposity.

The influence of age on stimulated GH secretion is less clear. Whether the GH response to insulin-induced hypoglycaemia changes with age is unresolved. The GH response to arginine, a GH secretagogue, and to GH-releasing hormone (GHRH) does not change significantly with age. In contrast, exercise-induced GH release is reduced with age and even in early middle age (mean age 42 years), the GH response to exhaustive exercise is greatly attenuated compared to younger (mean age 21 years) individuals [27].

Physical activity, sleep patterns, adiposity, and gonadal steroid status all change with age and have all been shown to regulate GH secretion. Age, body composition, and physical fitness are independent predictors of 24-hour integrated GH concentrations. GH secretion in response to arginine and clonidine in healthy adults is determined by body composition and physical fitness rather than by age while the GH response to exercise is influenced by age and physical fitness (VO_2max) but not by body fat [28]. These findings suggest that maintenance of physical fitness throughout life might attenuate the decline in GH secretion rates although notably training programmes, which improve physical fitness, do not appear to increase the GH response to exercise [27].

The age-related decline in GH secretion is mediated at the pituitary and hypothalamic or higher levels. Studies in animals have indicated that enhanced hypothalamic production of somatostatin, an inhibitor of GH secretion, is the principal mechanism for age-related hyposomatotropism, with reduced GHRH production involved to a lesser degree. Human studies using indirect approaches, however, have demonstrated an important role for reduced hypothalamic GHRH secretion in ageing. Firstly, the GH response to withdrawal of somatostatin infusion, which provides an estimate of GHRH release, is reduced in elderly compared to young women with a similar trend occurring in men. Secondly, GH pulse amplitude, a function of GHRH secretion, is reduced in elderly compared to young subjects. The suppression GH secretion by IGF-I infusion is reduced rather than enhanced in older people, implying that increased sensitivity to negative feedback by endogenous IGF-1 is not a mechanism through which this effect occurs.

In the absence of confounding factors, there is an age-dependent fall in GH and IGF-I levels that begins in the third decade well before the onset of physical frailty (Figure 1.5.3). The significance of the reduction in the activity of the GH axis is unknown; there has been enduring speculation that it may be pathological rather than adaptive biology of ageing. A beneficial effect of GH supplementation has not been established. Of interest are observations that patients with congenital isolated GH deficiency and GH receptor mutations appear to be protected from age-related diseases such as atherosclerosis, diabetes, and cancer [29, 30].

Prolactin

Prolactin is secreted by lactotrophs, the function and growth of which is stimulated by oestrogens. Prolactin secretion is under tonic inhibitory control of the hypothalamus.

Development *In Utero* and Childhood

Lactotrophs are present by 6 weeks' gestation and prolactin concentration increase gradually in the fetus throughout pregnancy. Prolactin levels in the newborn are approximately 10-fold higher than that seen in adult life in both boys and girls but gradually decrease within 3 months to a stable concentration in infancy and in childhood. The high prolactin levels are believed to be caused by the stimulatory effect of elevated maternal oestrogen levels. Prolactin levels increase slightly during puberty, to a greater extent in females

than males, and this gender difference is maintained virtually throughout adult life.

Adulthood

Prolactin is secreted in a diurnal pattern in both sexes, with episodic peaks and higher levels during the night than day. Prolactin secretion increases about 60–90 minutes after the onset of sleep coincident with non-rapid eye movement (REM) sleep and falls with the onset of REM sleep. Prolactin levels fluctuate during the menstrual cycle with higher levels at mid-cycle and lower concentrations during the luteal compared to follicular phase.

Prolactin levels decline gradually with age. In accordance with its pro-lactational function, levels rise progressively during pregnancy. Maternal prolactin levels start to rise by 6–8 weeks' gestation, increasing gradually throughout the course of pregnancy to 8–10 times higher than prepregnancy levels at term. Suckling is a potent stimulus to prolactin secretion, elevating blood levels by 3–5-fold. After termination of suckling, levels return to prenursing baseline by 1–2 hours. The suckling-evoked increase of prolactin secretion wanes with time post-partum particularly when the frequency of suckling diminishes with the introduction of formula feeds. Lactation can be maintained by regular suckling over years. On cessation, levels return to that seen before pregnancy by about 3–4 months. Prolactin has no known function in males.

Thyroid Axis

Development *In Utero* and Childhood

The fetus starts producing thyroid stimulating hormone (TSH) and thyroxine by the end of the first trimester. TSH levels remain low until 18 weeks, increasing steadily followed by rise in serum thyroxine levels through the mid and late trimesters.

TSH concentrations are higher in the fetus than in the mother throughout fetal life. At birth, TSH levels rise abruptly triggering a parallel increase in serum thyroxine levels achieving steady state values after 3–4 days of birth [31]. TSH levels fall rapidly shortly after the immediate postnatal rise, returning to baseline 48 hours after birth. The concentration of thyroxine in the neonatal period is considerably higher than in older children and adults, however, it falls gradually into the normal range of children and adults by the first month. Beyond the neonatal period, thyroid hormones levels remain stable and do not fluctuate with the onset and progress of puberty.

Newborn screening tests for the diagnosis of congenital hypothyroidism are reliant on TSH measurements in whole blood at day 5 of life, after the transient immediate postnatal surge. Because of the critical role of thyroid hormones for brain development and function, treatment of neonatal hypothyroidism aims to keep the TSH level in the normal range, including in premature babies with borderline TSH levels.

Adulthood

Interpretation of physiologic changes in thyroid hormones with ageing is complicated because rates of mild or subclinical thyroid dysfunction are high in elderly people, especially in women. Furthermore, acute and chronic non-thyroidal illness, poor nutritional status, and use of medications that influence thyroid function become more prevalent with age.

After the early neonatal period, the pituitary–thyroid axis changes minimally throughout life with TSH being the most robust indicator of thyroid status in the absence of pituitary disease. TSH is secreted constantly, exhibiting small episodic peaks and higher levels at night in both sexes while maintaining thyroid hormone levels within a tight range. With ageing, when patients with subclinical hypothyroidism or other confounding conditions are excluded from analysis, serum TSH remains within the normal range, although mean TSH levels are lower than at younger ages. Daily secretion of TSH is reduced with blunting of the nocturnal peak. It is not clear whether this reflects increased sensitivity to negative feedback or whether other mechanisms are involved.

Studies in subjects up to the 100–110 year age range report no change in total or free T4 levels, but an age-dependent reduction in Free T3 levels with levels continuing to decrease up to the eleventh decade, and age-dependent reduction in TSH [30]. Morphologic and functional changes are also seen in the ageing thyroid. There is a progressive decline in iodine uptake, while radioisotope studies have demonstrated reduced T4 distribution space, increased T4 half-life, and reduced T3 and T4 production rates. T4 degradation rates are also reduced (through reduced outer ring deiodination mediated by 5' deiodinase) with reduced generation of T3. The net effect of these physiologic changes likely explains the dissociation of T4 and T3 levels with ageing.

Conclusion

The profiles of hormones of the major pituitary axes across the lifespan differ from each other, with distinct responses seen at different stages of development and ageing, yet all are unified by a regulatory paradigm integrating peripheral feedback and converging neural cues. Fetal factors, maternal environment, and placental function shape the status and profile of each axis differently in postnatal life. When established, a central effector and peripheral feedback system provides for tight regulatory control. Developmental programming during life's transitions can bring dramatic changes of the pattern relationships between pituitary and target gland profiles within the regulated system. The hormone profiles are characterized by an orderly rhythmicity that is unique for each axis and which may not readily be discerned from single random hormone measurements. The hormone profiles of the pituitary gland provide key information on the integrity of the endocrine system across the human lifespan. An understanding of the properties and characteristics of each axis is critical for identifying disease and implementing appropriate treatment.

REFERENCES

1. Kuiri-Hanninen T, Sankilampi U, Dunkel L. Activation of the hypothalamic-pituitary-gonadal axis in infancy: minipuberty. *Horm Res Paediatr*, 2014; **82**: 73–80.
2. Handelsman DJ, Yeap B, Flicker L, Martin S, Wittert GA, Ly LP. Age-specific population centiles for androgen status in men. *Eur J Endocrinol*, 2015; **173**: 809–17.

3. Camacho EM, Huhtaniemi IT, O'Neill TW, *et al.* Age-associated changes in hypothalamic-pituitary-testicular function in middle-aged and older men are modified by weight change and lifestyle factors: longitudinal results from the European Male Ageing Study. *Eur J Endocrinol*, 2013; **168**: 445–55.
4. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocr Rev*, 2005; **26**: 833–76.
5. Bhasin S, Brito JP, Cunningham GR, *et al.* Testosterone therapy in men with hypogonadism: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
6. Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. *Endocr Rev*, 2009; **30**: 465–93.
7. Faddy MJ, Gosden RG. A model conforming the decline in follicle numbers to the age of menopause in women. *Hum Reprod*, 1996; **11**: 1484–6.
8. Goto M, Piper Hanley K, Marcos J, *et al.* In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development. *J Clin Invest*, 2006; **116**: 953–60.
9. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*, 2011; **32**: 81–151.
10. Ng PC, Lam CW, Lee CH, *et al.* Reference ranges and factors affecting the human corticotropin-releasing hormone test in preterm, very low birth weight infants. *J Clin Endocrinol Metab*, 2002; **87**: 4621–8.
11. Iwata O, Okamura H, Saitsu H, *et al.* Diurnal cortisol changes in newborn infants suggesting entrainment of peripheral circadian clock *in utero* and at birth. *J Clin Endocrinol Metab*, 2013; **98**: E25–32.
12. Rivkees SA. The development of circadian rhythms: from animals to humans. *Sleep Med Clin*, 2007; **2**: 331–41.
13. Veldhuis JD, Roelfsema F, Iranmanesh A, Carroll BJ, Keenan DM, Pincus SM. Basal, pulsatile, entropic (patterned), and spiky (staccato-like) properties of ACTH secretion: impact of age, gender, and body mass index. *J Clin Endocrinol Metab*, 2009; **94**: 4045–52.
14. Pilling LC, Kuo CL, Sicinski K, *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)*, 2017; **9**: 2504–20.
15. Van Cauter E, Leproult R, Kupfer DJ. Effects of gender and age on the levels and circadian rhythmicity of plasma cortisol. *J Clin Endocrinol Metab*, 1996; **81**: 2468–73.
16. Belgorosky A, Baquedano MS, Guercio G, Rivarola MA. Adrenarche: postnatal adrenal zonation and hormonal and metabolic regulation. *Horm Res*, 2008; **70**: 257–67.
17. Alkatib AA, Cosma M, Elamin MB, *et al.* A systematic review and meta-analysis of randomized placebo-controlled trials of DHEA treatment effects on quality of life in women with adrenal insufficiency. *J Clin Endocrinol Metab*, 2009; **94**: 3676–81.
18. Corona G, Rastrelli G, Giagulli VA, *et al.* Dehydroepiandrosterone supplementation in elderly men: a meta-analysis study of placebo-controlled trials. *J Clin Endocrinol Metab*, 2013; **98**: 3615–26.
19. Elraiyah T, Sonbol MB, Wang Z, *et al.* Clinical review: the benefits and harms of systemic dehydroepiandrosterone (DHEA) in postmenopausal women with normal adrenal function: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2014; **99**: 3536–42.
20. Davison SL, Bell R, Donath S, Montalto JG, Davis SR. Androgen levels in adult females: changes with age, menopause, and oophorectomy. *J Clin Endocrinol Metab*, 2005; **90**: 3847–53.
21. Karlberg J. A biologically-oriented mathematical model (ICP) for human growth. *Acta Paediatr Scand Suppl*, 1989; **350**: 70–94.
22. Hawkes CP, Grimberg A. Measuring growth hormone and insulin-like growth factor-I in infants: what is normal? *Pediatr Endocrinol Rev*, 2013; **11**: 126–46.
23. Stevens A, Hanson D, Whatmore A, Destenaves B, Chatelain P, Clayton P. Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks. *BMC Genomics*, 2013; **14**: 547.
24. Chowen JA, Frago LM, Argente J. The regulation of GH secretion by sex steroids. *Eur J Endocrinol*, 2004; **151** Suppl 3: U95–100.
25. Ho KY, Evans WS, Blizzard RM, *et al.* Effects of sex and age on the 24-hour profile of growth hormone secretion in man: importance of endogenous estradiol concentrations. *J Clin Endocrinol Metab*, 1987; **64**: 51–8.
26. Iranmanesh A, Lizarralde G, Veldhuis JD. Age and relative adiposity are specific negative determinants of the frequency and amplitude of growth hormone (GH) secretory bursts and the half-life of endogenous GH in healthy men. *J Clin Endocrinol Metab*, 1991; **73**: 1081–8.
27. Gibney J, Healy ML, Sonksen PH. The growth hormone/insulin-like growth factor-I axis in exercise and sport. *Endocr Rev*, 2007; **28**: 603–24.
28. Zaccaria M, Varnier M, Piazza P, Noventa D, Ermolao A. Blunted growth hormone response to maximal exercise in middle-aged versus young subjects and no effect of endurance training. *J Clin Endocrinol Metab*, 1999; **84**: 2303–7.
29. Menezes Oliveira JL, Marques-Santos C, Barreto-Filho JA, *et al.* Lack of evidence of premature atherosclerosis in untreated severe isolated growth hormone (GH) deficiency due to a GH-releasing hormone receptor mutation. *J Clin Endocrinol Metab*, 2006; **91**: 2093–9.
30. Guevara-Aguirre J, Balasubramanian P, Guevara-Aguirre M, *et al.* Growth hormone receptor deficiency is associated with a major reduction in pro-aging signaling, cancer, and diabetes in humans. *Sci Transl Med*, 2011; **3**: 70ra13.
31. Burrow GN, Fisher DA, Larsen PR. Maternal and fetal thyroid function. *N Engl J Med*, 1994; **331**: 1072–8.

1.6

Pituitary Assessment Strategy

William M. Drake, Brian Keevil, and Peter J. Trainer

Introduction

The optimum methods of testing anterior and posterior pituitary function and the interpretation of the results are subjects of continuing debate. The syndromes associated with and consequences of hypo- and hyperpituitarism; and the diagnosis and treatment of diabetes insipidus are all discussed elsewhere in this book. The intention of this chapter is to describe the physiological basis and evidence in favour of the various available tests of anterior pituitary function, while at the same time acknowledging their limitations and appreciating the importance of the clinical context of testing.

The aetiology, treatment, and complications of hypopituitarism are described elsewhere (see Chapter 2.3.5) but fundamental to achieving the desired outcomes is the early diagnosis of pituitary dysfunction as, untreated, the morbidity and mortality of hypopituitarism is high, mainly as a consequence of secondary adrenal insufficiency. There are two broad groups of patients for whom pituitary function tests are required. The first group comprises all patients with symptoms suggestive of hypopituitarism. This, in turn, includes patients with target organ failure (such as hypoadrenalism, hypothyroidism, and hypogonadism) in whom low levels of cortisol, thyroxine, and sex steroid respectively are not associated with an appropriate elevation of the relevant pituitary trophic hormone; patients with cranial diabetes insipidus; patients presenting with the mechanical consequences of pituitary tumours (possibly hypersecreting, e.g. acromegaly, such as headache and visual failure; and patients in whom a pituitary mass is found incidentally during the course of radiological investigation for an unrelated symptom. In recent years, increasing recognition that immune cancer therapies (such as 'checkpoint' inhibitors) and brain injury (traumatic or vascular) may cause hypopituitarism provide important alerts to the possibility that both chronic and acute ill health may have a pituitary aetiology. The second group of patients includes those with known hypothalamic–pituitary disease or a history of radiotherapy to the head or neck who are at risk of evolving endocrine deficiency (**Box 1.6.1**). The former should be under endocrine review and be having regular pituitary function testing but the concern is that in the latter group, there may be no endocrine follow-up and the risk of hypopituitarism may not be appreciated and testing not occur. Conditions in which progressive hypothalamo–pituitary destruction may occur, such as sarcoidosis or Langerhans cell histiocytosis, require monitoring as the development of symptomatic hypopituitarism may be subtle.

The order of development of pituitary hormone deficiency is predictable with, typically, growth hormone deficiency (GHD) preceding gonadotrophin deficiency with subsequent failure of adrenocorticotrophic hormone (ACTH) and thyroid-stimulating hormone (TSH) secretion. Traumatic brain injury and lymphocytic

Box 1.6.1 At-risk patients

In such patients (e.g. those who have received pituitary radiotherapy), pituitary function tests should be performed regularly in order to detect asymptomatic hypopituitarism, although there is a paucity of data on the optimum frequency with which this should be done. Our practice is to check thereafter basal pituitary function (7 to 9 a.m.) every 2 years, with a dynamic test of ACTH reserve if the basal serum cortisol is <400 nmol/L. If patients exhibit symptoms consistent with GHD, then the dynamic test of choice will be the ITT. GHD occurs early after radiotherapy and, once it has been proven, many physicians use sequential SSTs to document the subsequent evolution of ACTH deficiency. Again, data in this regard are scarce such that accurate; robust local reference ranges are essential.

hypophysitis (including checkpoint inhibitor-induced hypophysitis) are exceptions to the rule with the order of development of deficiencies being less predictable; isolated ACTH deficiency is well recognized with the latter. In patients receiving cytotoxic and other chemotherapies, the use of dexamethasone as an antiemetic is common; and many patients carry other diagnoses (such as airways disease and skin conditions) that mandate the use of inhaled or transdermal steroid preparations. A meticulous drug history is therefore crucial prior to ascribing a low serum cortisol to ACTH deficiency. In all patients, pituitary function testing may help identify those patients whose hypopituitarism is sufficiently severe to threaten their safety, irrespective of symptoms; or exclude hormonal deficiencies as a cause of symptoms such as fatigue. For example, an asymptomatic patient with a basal serum cortisol of 120 nmol/L rising to 175 nmol/L during insulin-induced hypoglycaemia is at risk of hypoadrenalism with intercurrent illness, and requires hydrocortisone replacement therapy plus education on the implications of glucocorticoid replacement therapy. In contrast, a patient with a pituitary mass and symptoms of lethargy who has normal thyroid hormone levels, a basal serum cortisol above 400 nmol/L and normal gonadal function does not require endocrine replacement therapy. A sound knowledge of the principles of pituitary function testing is mandatory for the accurate diagnosis and optimal treatment of pituitary failure.

The anterior pituitary gland secretes six known hormones: growth hormone (GH), ACTH, luteinizing hormone (LH), follicle-stimulating hormone (FSH), TSH, and prolactin, all of which are under regulatory feedback control. Regulation of secretion of each hormone is complex with, in most cases, at least two hypothalamic peptides directly acting on the appropriate pituitary cell type to influence secretion, which in turn may be pulsatile, with underlying circadian or ultradian rhythm. In the case of ACTH, for example, this phenomenon is striking with plasma levels of its target hormone, cortisol, varying in health between undetectable (<20 nmol/L) and 700 nmol/L over 24 hours. For other pituitary hormones, such as TSH, the circadian variation is modest and there is no diurnal change in serum thyroxine concentrations. This means that a single, random blood sample is unlikely to provide sufficient diagnostic information about the function of the pituitary–adrenal axis; whereas it is rare for investigations other than basal samples to be required in the assessment of the pituitary–thyroid axis. In general, the more dynamic the physiological system in health, the more likely will be the need for a dynamic/provocative test to investigate its possible malfunction in disease. In all cases, the fundamental question being posed is: is the functioning of this 'endocrine unit' adequate for this patient's health and, if not, does it require replacement/support? It is the authors' intention to describe the physiological basis of the various tests of anterior pituitary function, to discuss the evidence in favour of their interpretation and, ultimately, to produce a rational, reliable, and safe strategy for pituitary function testing.

General Principles of Pituitary Assessment

The diagnostic evaluation of pituitary function has several complementary limbs involving laboratory and radiological investigations. First, it is necessary to demonstrate target organ hormonal insufficiency, such as low levels of thyroid hormone or gonadal steroid. Paired testing of both hormones in the pituitary–target organ feedback loop, sometimes in combination with provocative testing, will prove that target organ failure is consequent upon lack of stimulation by the relevant pituitary trophic hormone. Additional tests may occasionally be performed in order to determine whether the pituitary itself is at fault, or whether pituitary failure is secondary to understimulation by the hypothalamus. However, this distinction is seldom useful clinically and is irrelevant to the need for hormone replacement therapy. Sophisticated radiological imaging is required to establish the aetiology, together with careful neuro-ophthalmological assessment. Lastly, in cases of pituitary failure where the cause is believed to be a systemic illness (such as sarcoidosis or tuberculous hypophysitis) more specific investigations may be needed. These are discussed in the relevant sections elsewhere (see Chapter 2.3.5).

It is important to recognize that it is erroneously simplistic to think of any pituitary hormone status as being either 'adequate' or 'deficient'; the reality is there are degrees of deficiency. The biochemical criteria propagated by consensus guidelines and similar are not necessarily predicated on rigorous science but rather the notion of 'best practice'. Furthermore, the decision about intervention and replacement therapy may depend on symptoms and other variables, an obvious example being gonadotrophin deficiency in a woman over 50 years of age, which may be indicative of hypopituitarism but does not necessarily mandate the use of oestrogen replacement therapy.

Types of Laboratory Test

The literature, and in particular clinical guidelines on the diagnosis of hypopituitarism rely on biochemical criteria to define deficiency. However, application of such international consensus criteria to local practice is complicated by bias in assay performance and it is inappropriate to blindly adopt such 'cut-offs' without an awareness of the performance of local assays.

Basal Pituitary Function Tests

Basal blood tests refer to samples taken with the patient resting, unstressed and with no physiological or pharmacological manipulation of the mechanisms that control the pituitary

Box 1.6.2 Basal pituitary function tests

New patients

Basal serum investigations, at 7–9 a.m.:
- cortisol
- free T_4, TSH
- Prolactin
- LH, FSH, testosterone/oestradiol, SHBG
- Insulin-like growth factor-I
- Urine/plasma osmolality

If basal serum cortisol is <400 nmol/L and/or GHD suspected, then proceed to ITT. If any abnormality in any of the aforementioned tests, then one should proceed to pituitary imaging.

TSH, thyroid-stimulating hormone; LH, luteinizing hormone; FSH, follicle stimulating hormone; SHBG, sex-hormone-binding globulin; GHD, growth hormone deficiency; ITT, insulin tolerance test.

cell–target cell interaction (Box 1.6.2). Hence basal samples are taken between 7 and 9 a.m., when serum cortisol and testosterone levels are highest. Given that the decision to proceed to dynamic testing is based on the results of basal samples, it is logical to maximize the chances of basal investigations yielding sufficient information to avoid the need for more complex tests. Paired measurement of both limbs of a pituitary hormone–target hormone loop is required for interpretation of the target hormone level. Low levels of target hormone in association with low or normal levels of the relevant pituitary trophic hormone indicate that target gland failure is consequent upon under stimulation by the pituitary.

Provocative Tests

Provocative (stimulation) tests are employed when hypofunction of a pituitary cell type is suspected and basal investigations have not yielded sufficient information. Such tests assess the ability of a given cell type to respond acutely to a stimulus, but do not necessarily provide information about the adequacy of day-to-day hormone production by that cell type under basal conditions. Two types of provocative tests are used: those that stimulate hormone release indirectly (such as the insulin tolerance and glucagon tests) and direct stimulation tests in which pharmacological doses of synthetically manufactured peptide are injected and the target cell hormone response measured. Examples of these include hypothalamic-releasing hormone tests and the short synacthen test (see next). The virtue of indirect provocation tests is that the integrity of an entire hypothalamo–pituitary–target cell loop is tested. Hypothalamic-releasing hormone tests with thyrotropin-releasing hormone (TRH), gonadotrophin-releasing hormone, and growth hormone-releasing hormone (GHRH) are discussed briefly in the relevant sections. When introduced to clinical practice, it was thought that they would facilitate the diagnosis of hypopituitarism and forewarn of insidious pituitary failure. However, experience has shown that they have no value in diagnosing hypopituitarism and they cannot be used to predict future pituitary failure. Historically they may occasionally be of value in differentiating hypothalamic from pituitary disease, but have been supplanted by high-resolution imaging.

Assessing ACTH Reserve (the Hypothalamic–Pituitary–Adrenal Axis)

Of all the aspects of pituitary function testing, this is the most important and controversial, mainly because assessment of the adequacy of the hypothalamic–pituitary–adrenal (HPA) axis and the provision of replacement therapy (if required) has the most far-reaching consequences of all the anterior pituitary hormones. The laboratory assessment of the HPA axis is performed in two distinct clinical settings, although the aim of establishing whether cortisol production is adequate is common to both. The first clinical scenario is that of a patient with symptoms suggestive of adrenal insufficiency (such as tiredness, listlessness, and malaise), where the question being posed is: 'are the symptoms due to cortisol deficiency?' The second clinical setting is one in which the patient is known to be at risk of developing secondary adrenal insufficiency, for example patients previously treated with supraphysiological doses of corticosteroids, known to have hypothalamic–pituitary disease who may have been treated with surgery and/or radiotherapy. In these 'at-risk' people, it is necessary to assess the adequacy of the patient's response to physiological stress, even in the absence of any symptoms of adrenal insufficiency. If the test used predicts that the patient will not be able to mount an adequate stress response, then the patient requires education about the implications of ACTH deficiency and adequate steroid cover must be provided in the event of emergency, although not necessarily day-to-day therapy.

A satisfactory method for assessing the adequacy of the day-to-day cortisol production rate has not yet been identified. Isotope dilution methods are accurate but complex and not widely available, while measurement of 24-hour urinary free cortisol requires complete collection and lacks the sensitivity to detect insufficiency. The inability to adequately assess basal, unstimulated cortisol production rate mean it is of paramount importance to consider patients' symptoms and well-being when contemplating initiation of lifelong glucocorticoid replacement therapy.

Physiological Background

Although there is considerable debate about many aspects of assessing ACTH reserve, the aim of establishing whether cortisol production is adequate for the patient's health is not disputed. Measurement of the target hormone (cortisol) is common to all tests of the hypothalamic–pituitary–adrenal axis. Cortisol has a multitude of actions, including regulation of protein and carbohydrate metabolism, maintenance of vascular tone, and modulation of the immune system. Its synthesis and secretion are controlled by ACTH, whose release from the pituitary is, in turn, regulated by hypothalamic corticotrophin-releasing hormone and vasopressin. Hypothalamic function is influenced by a complex array of factors including neural stimuli (particularly from the limbic system) and humoral inputs (such as inflammatory cytokines). A change in cortisol production rate is the 'final common pathway' for all of these complex modulatory factors: hence the use of cortisol levels for the assessment of ACTH reserve. It has the added practical advantage of ease of collection, as ACTH samples require cold centrifugation and flash-freezing whereas cortisol is measured in serum.

General Aspects of Cortisol Measurements

A common pitfall in the assessment of the HPA axis is failure to document exposure to exogenous glucocorticoids, whether administered orally, transdermally, by inhaler, by parenteral or intra-articular injection. Occasionally, patients may be unknowingly exposed to exogenous glucocorticoids, for example some so-called 'skin whiteners' have been shown to contain dexamethasone [1]. Coadministration of P450 CYP3A4 inhibitors, such as itraconazole, can potentiate the glucocorticoid action of exogenous steroid. Typically, such patients are asthmatic with fungal infections and are clinically Cushingoid with undetectable serum cortisol levels which reinforces the importance of interpreting the biochemical investigations in the clinical context [2].

Only 5–10% of cortisol is free and biologically active, the remainder being bound to cortisol-binding globulin and albumin. Variations in these proteins significantly impacts on measurement of total serum cortisol. Cortisol-binding globulin is synthesized in the liver and, like sex-hormone-binding globulin (SHBG), production is increased by oral oestrogen therapy. However, while SHBG is frequently measured along with testosterone in the assessment of the pituitary–gonadal axis, cortisol-binding globulin is not routinely measured along with cortisol. Total serum cortisol levels are therefore significantly raised in pregnancy and in patients taking oral oestrogens, although serum free cortisol is essentially unchanged. Hence, oral oestrogens should be discontinued prior to assessment of the HPA axis and most authorities accept that six weeks are required for their effect on cortisol-binding globulin levels to disappear completely [3]. Transdermal oestrogen therapy does not affect circulating cortisol-binding globulin (CBG) levels [4]. Accurate assessment of the HPA axis in pregnancy is extremely difficult. Other circumstances in which cortisol-binding globulin and albumin levels may complicate assessment of cortisol status axis include conditions of protein loss, such as nephrotic syndrome and protein losing enteropathy; and failure of protein synthesis, such as hepatic cirrhosis. GH also has a subtle effect on circulating CBG, with GH replacement therapy lowering circulating cortisol levels secondary to reduced circulating CBG [5], with the reverse seen with treatment of acromegaly.

Assessment of late-night salivary cortisol secretion is widely used in the diagnosis of Cushing's syndrome but there is increasing interest in salivary cortisol in the investigation of possible cortisol deficiency. Salivary cortisol concentrations are not affected by serum CBG levels (i.e. oestrogen status) and correlate closely with serum free cortisol [6]. Oxidation of cortisol to cortisone by the enzyme 11B-HSD2 in salivary glands results in cortisone being the predominate glucocorticoid in saliva which to varying degrees cross-reacts in cortisol immunoassays, and explains the significant 'between-assay' bias [7]. Mass spectrometry measurement of salivary steroids is unaffected by cross-reactivity and has the further virtue of being able to quantify cortisol and cortisone to lower levels than immunoassays. It is believed that as mass spectrometry becomes more widely available, salivary cortisol measurement will have an increasingly important role in the investigation of cortisol deficiency, although consensus is required around cut-off criteria [8].

Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) in Steroid Measurement

In the great majority of laboratories, cortisol is measured by immunoassay (IA), but LC-MS/MS is emerging as an alternative and superior technology as it is not blighted by the inherent limitation of IA, namely cross-reactivity of synthetic exogenous glucocorticoids such as prednisolone/prednisone, methylprednisolone and cortisol precursors, and metabolites such as 11 deoxycortisol. LC-MS/MS has the additional virtues of permitting measurement of multiple steroids at essentially the same cost as a single analyte and being able to quantify to lower limits than possible by IA.

Improvements in instrument performance have ensured that LC-MS/MS is increasingly becoming the method of choice for steroid hormone measurements, for example vitamin D, and drugs (e.g. tacrolimus and ciclosporin) due to smaller sample volumes, faster analysis times, and improved specificity compared to immunoassays. The main impediments to wider adoption of LC-MS/MS are the cost of instrumentation, greater technical complexity, speed, and turnaround of analysis. For these reasons and except for the larger specialized laboratories, the uptake of LC-MS/MS has been relatively poor in routine clinical laboratories, however once established and with high throughput MS is comparable in cost to the purchase and use of commercial IA kits.

The sample preparation process is not as simple as for IA as a clean-up/extraction is always needed before LC-MS/MS analysis to remove matrix and chemical interference. After extraction the sample needs to undergo separation on a liquid chromatography (LC) column to further remove potentially interfering substances before introduction into the mass spectrometer (**Figure 1.6.1**). The liquid eluate from the LC column is converted into either positively or negatively charged gaseous ions by the process of electrospray ionization in the mass spectrometer source. The tandem mass spectrometer instrument consists of two mass analysis quadrupoles separated by a collision cell. The strengths of this instrument lie in the ability to filter ions of a predetermined mass in the first quadrupole and then fragmentation of these ions in a compound specific way in the collision cell. Monitoring fragment ions, that have been selectively filtered in the second quadrupole, which can only have come from the parent ion, gives rise to high analytical specificity. The total process thus utilizes three different separation techniques, sample extraction, chromatography, and mass spectrometry which together confer a very high degree of specificity compared to automated immunoassay, which relies solely on the specificity of the antibody [9].

To further improve precision and accuracy, internal standards are added during the extraction process to control for ionization variability within the mass spectrometer source and the precision of the extraction procedure. The most appropriate internal standards are stable isotopes of the steroid being measured.

LC-MS/MS is particularly good at measuring urine free cortisol (UFC) with demonstrated enhanced specificity and sensitivity over IA methods [10] which results in significantly lower reference ranges. It can be debated whether the exclusion by MS of urinary cortisol metabolite measurement improves the diagnostic efficiency in Cushing's syndrome of UFC but it certainly improves the

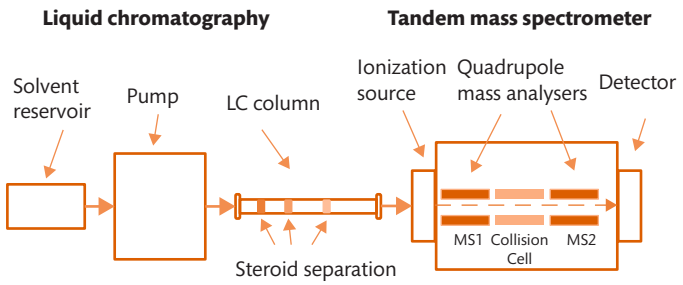


Figure 1.6.1 LC-MS/MS. Schematic of an LC-MS/MS system showing the principal components. Solvent is pumped through a liquid chromatography column to facilitate separation of the steroids of interest. The eluent from the column undergoes electrospray ionization in the mass spectrometer source, the ions generated are separated according to mass charge ratio by the first quadruple mass analyser and the specific molecular ion is selected. After fragmentation of the molecular ion in the collision cell, a specific fragment ion is selected in the second quadrupole mass analyser which is then used for quantitation.

value when used to monitor medical therapy with agents such as metyrapone [11].

Both cortisol and cortisone can be measured by LC-MS/MS in saliva with the added advantage that salivary cortisone may be a better diagnostic marker than cortisol because it agrees more closely with serum cortisol [12]. The LC-MS/MS serum methods for serum cortisol offer many advantages over IA methods when measuring samples from patients on metyrapone [11] or prednisolone therapy [13] and do not suffer some of the technical issues such as inadequate release of cortisol from CBG [14]. It also permits simultaneous measurement of cortisol and dexamethasone during a dexamethasone suppression test to ensure patient compliance.

While implementation of LC-MS/MS in the clinical laboratory has been slow, use in research laboratories has been expanding rapidly. Advances in mass spectrometer design have produced a boom in research not only in endocrinology but in other biological applications such as proteomics and metabolomics. In endocrinology, this work is being translated into the ability to measure panels of steroids adoption of which will improve diagnosis and monitoring disease [15, 16].

The ability of LC-MS/MS to perform multiplexed analysis of many steroids using similar sample extraction and instrument conditions, is often at no greater expense. Whereas, measuring steroids using IA one test at a time, adding in extra tests becomes more expensive.

Selective testing panels of steroids have been developed to assist in the management of various endocrine disorder such as congenital adrenal hyperplasia [17, 18], polycystic ovarian syndrome [19], and for the investigation of hypoadrenalism [20].

The enhanced specificity of LC-MS/MS should be contrasted with specificity problems with IA methods encountered with the move to perform 'direct' IA (without any extraction step) routinely on automated analysers. The elimination of the extraction step may streamline the assay procedure but at the expense of the potential for significant interference from other steroids.

Instrument manufacturers are bringing to market clinical mass spectrometry analysers with automation of the time-consuming sample extraction phase. These instruments currently have a limited test repertoire and are relatively expensive but offer an automated vision of the future that will undoubtedly make LC-MS/MS more attractive to a wider range of laboratories and should improve the availability of high quality tests to the clinician.

Defining an Adequate Peak Cortisol Response

The clinical spectrum of adrenal insufficiency ranges from cardiovascular collapse through to more subtle dysfunction that is apparent only during the physiological stress of sepsis, major surgery, or accidental trauma. The fundamental question is: What serum cortisol level is required under such circumstances? The defining study looked at the cortisol response to major abdominal surgery in normal and corticosteroid-treated patients. Peak intraoperative serum cortisol was closely correlated to the cortisol response to insulin-induced hypoglycaemia and that a peak serum cortisol of 580 nmol/L was associated with an adequate cortisol response during surgery [21]. Controlled iatrogenic hypoglycaemia has widely been advocated as the 'gold standard' by which to judge the ability to mount an adequate cortisol response to physiological stress.

The figure of 580 nmol/L was based on the fluorimetric method of Mattingly [22] for measurement of serum cortisol, a technique that, also detects cortisone, and with modern immunoassays a more appropriate 'cut-off' is probably in the range of 400–450 nmol/L depending on assay characteristics [23] and this has been confirmed using mass spectrometry methods [24].

Appreciation of assay performance and 'between-assay' bias is crucial to the appropriate application of cortisol measurement to the investigation and diagnosis of hypopituitarism. Bias in cortisol immunoassays arise because of the epitope-specificity of polyclonal antibodies resulting in cross-reactivity with other endogenous and exogenous steroids, matrix-differences between standards and samples, susceptibility to interference by binding proteins, and calibrator (reference-preparation) effects. In 1998, Clark *et al.* [25] documented a 26% bias between basal serum cortisol results using different assays, with highly significant differences also noted in serum cortisol levels 30 and 60 minutes after injection of synacthen. Twenty years later, little has changed with El-Farhan [24] reporting lower reference limits for six assays varying between 420 and 574 nmol/L, with differences between males and females. Evaluation of the more commonly used immunoassays against LC-MS/MS showed that much of this difference is caused by the varying ability of immunoassay kits to liberate cortisol from CBG, without which cortisol cannot bind to the antibody. This problem is exacerbated in pregnancy and those taking oral contraceptive pill in whom binding proteins are high [26]. Such findings make comparisons of cortisol responses between assays and laboratories extremely difficult and emphasize the need for centres to define/adapt appropriate reference ranges for their assay, as opposed to selecting rigid 'cut-off' values derived from studies using different methodologies.

The Testing of Cortisol Reserve

Having established a minimum serum cortisol level that is adequate for acute physiological stress, the question arises as to what is the

most appropriate stimulation test for the prediction that the patient will be able to achieve such a cortisol response when required. Several tests exist, each with its own merits and shortfalls, including measurement of basal serum cortisol, the insulin tolerance test, glucagon stimulation test, short synacthen test (standard and low dose), and the metyrapone test.

Measurements of Basal Serum Cortisol

In the interest of patient convenience and economics, it is preferable to maximize the value of basal unstimulated cortisol measurement with the goal of minimizing the number of dynamic tests undertaken. Endogenous HPA activity is maximal in the early morning and samples should be drawn between 8 and 9 a.m. or earlier if possible. In cases of suspected pituitary insufficiency, a basal morning serum cortisol of less than 100 nmol/L strongly indicates ACTH deficiency, dynamic testing is not necessary, and glucocorticoid replacement should commence immediately. In most such patients, the requirement for steroid replacement is likely to be permanent. However, in the case of ACTH deficiency prior to surgery for a pituitary tumour, recovery of ACTH reserve following surgical decompression may occur and so it is necessary to reassess the situation postoperatively.

In contrast, defining an unstimulated, basal cortisol that predicts an adequate response to stimulation is difficult. Several studies have confirmed that a close correlation exists between measurements of basal, unstressed serum cortisol to peak serum cortisol levels during insulin-induced hypoglycaemia. From these reports, it is clear that many patients can avoid a dynamic test on the basis of basal cortisol measurements, although the precise 'cut-off' that predicts an adequate response is challenging but it would be generally accepted that a baseline cortisol of 400 nmol/L or more precludes the need for dynamic testing [27–29].

Emergency Assessment of the Hypothalamic–Pituitary–Adrenal Axis

In the acutely sick patient with, for example, sepsis, trauma or checkpoint inhibitor therapy, and suspected hypoadrenalism the clinical situation dictates that a morning cortisol and/or a dynamic test of ACTH reserve are impractical and glucocorticoid support may need to be started immediately. In such a context, circadian variation of ACTH release will be absent and activity of the hypothalamic–pituitary–adrenal axis should be maximal. If adrenal insufficiency is suspected, then random serum cortisol and plasma ACTH measurements, prior to commencing glucocorticoid therapy, will suffice for an assessment of hypothalamic–pituitary–adrenal axis integrity. If, subsequently, the random cortisol level is shown to have been appropriate to the clinical situation (above 400 nmol/L) glucocorticoids may be withdrawn. If not, steroid support should continue, and dynamic testing must wait until the acute clinical situation has resolved. The plasma ACTH will indicate whether the adrenal insufficiency is primary or secondary. A serum cortisol below 200 nmol/L with a plasma ACTH above 200 pg/ml is diagnostic of primary adrenal failure.

The Insulin Tolerance Test

This test, first described in 1966 [30] seeks to simulate physiological 'stress' in a controlled, supervised environment by inducing hypoglycaemia with intravenous insulin. Hypoglycaemia is a powerful stress stimulus, which in the intact pituitary and hypothalamus induces ACTH and GH release and a rise in serum cortisol levels. It therefore assesses the integrity of the entire HPA axis and has traditionally been regarded as the 'gold standard' for this purpose. The cortisol response to hypoglycaemia has been shown to be reproducible in healthy volunteers and patients with pituitary disease [31, 32]. As discussed earlier, the assumption that the ability to respond to insulin-induced hypoglycaemia will translate into an appropriate cortisol rise in the event of acute illness or major surgery is supported by studies in which the peak cortisol levels of patients undergoing major surgery were comparable to those achieved during a preoperative insulin tolerance test [21].

Although the safety of the insulin tolerance test (ITT) has been questioned, particularly in children, the morbidity of this investigation in experienced hands within the setting of a designated metabolic investigation unit is reassuringly low, provided that the standard criteria are adhered to (ischaemic heart disease, epilepsy/unexplained blackouts, severe longstanding hypoadrenalism, glycogen storage disease are contraindications and patients should have normal thyroid function) [28]. The dose of fast-acting insulin used varies between centres. Most authorities recommend a dose of 0.1–0.15 IU/kg, with higher doses (typically 0.3 IU/kg) being required for patients with acromegaly or other conditions in which insulin resistance is a feature. Many physicians are uncomfortable with its use in older people and it requires careful supervision and monitoring of adrenergic and neuroglycopaenic symptoms.

The immediate counter-regulatory response to hypoglycaemia is characterized by catecholamine release, which, in turn, stimulates hepatic glycogenolysis and correction of hypoglycaemia. Glucocorticoids are not part of this phenomenon, although the laying down of hepatic glycogen stores does require pre-exposure to glucocorticoids. Thus, in patients with longstanding ACTH deficiency and consequent inadequate glycogen stores, recovery from hypoglycaemia may be delayed. It is therefore usual practice to administer oral glucose in the form of a sugary drink, together with a meal, at the conclusion of the test to guard against this eventuality.

A common reason to perform an ITT is to test the ability of the hypothalamic–pituitary–adrenal axis to respond to stress following the use of supraphysiological doses of corticosteroids. Such doses may lead to ACTH suppression, with secondary adrenal involution and loss of responsiveness. It is therefore essential, in this situation, to perform a short synacthen test in order to establish that the adrenals are capable of responding to ACTH, prior to an ITT. If the adrenals do not respond to ACTH, an ITT will yield no useful information.

Short Synacthen Test

This investigation was originally introduced in the 1960s [33] as a test for primary adrenal failure. It involves the injection of a pharmacological dose (250 μg) of synthetic ACTH, with measurement of the serum cortisol response, and it has been advocated as

an alternative to the ITT as a means of assessing ACTH reserve. The basis of its use, in the context of hypopituitarism, is that chronic underexposure of the adrenal glands to ACTH (either as a consequence of prolonged corticosteroid therapy or due to suspected or proven hypothalamic–pituitary disease) will result in a blunted cortisol response to exogenously administered ACTH. The test does not distinguish primary from secondary adrenal insufficiency, although clinical assessment (pigmentation) and measurement of basal plasma ACTH are usually sufficient in this regard. The major argument in favour of the short synacthen test (SST) is its simplicity, as it requires no specialist staff and takes only an hour to complete. The only reported side effect is allergy in patients with a history of atopy, although this is very rare. The SST does not assess GH reserve. It is universally accepted that the SST cannot be used for the assessment of ACTH reserve when acute hypopituitarism develops, such as following pituitary infarction (apoplexy) or the immediate postoperative assessment of the HPA axis. It takes at least 6 weeks for the adrenal zona fasciculata to involute following withdrawal of ACTH stimulation, during which time the adrenal cortex will remain responsive to supraphysiological doses of ACTH. In addition, it should be remembered, in the assessment of new patients with suspected hypothalamic–pituitary disease, that the duration of ACTH deficiency may be unknown and that, as following pituitary surgery or apoplexy, a falsely reassuring SST may result.

Two main aspects of the SST in assessing ACTH reserve have been debated: the peak serum cortisol versus increment and the level of serum cortisol that constitutes an adequate response. The increase in serum cortisol following synacthen is a poor index of adrenal responsiveness, as there is considerable overlap between normal volunteers and patients with secondary adrenal insufficiency [34]. Further, the cortisol increment is inversely correlated with the basal value and hence a smaller increment is seen in the early morning when plasma ACTH and serum cortisol levels are at their highest [35]. The peak serum cortisol response following synacthen shows no diurnal variation and is now the accepted index of adrenal responsiveness and, indirectly, endogenous ACTH exposure.

Excellent correlations exist between cortisol levels 30 min after injection of synacthen and the peak cortisol achieved during insulin-induced hypoglycaemia. This has led to the increasing use of the SST as a substitute for the ITT.

However, despite its widespread use as a method of assessing ACTH reserve, there is no study showing that a normal SST indicates that the HPA axis is capable of responding normally to major illness or stress. Critics of the use of the SST point to reports of patients with pituitary disease with symptoms and signs of adrenal failure, corrected by glucocorticoid replacement, having recently had a falsely reassuring 'normal' SST. This problem cannot be corrected by application of a more 'stringent' threshold of serum cortisol as in two such reported patients the peak serum cortisol value was more than 950 nmol/L 30 min after synacthen. However, reports exist of patients who have developed acute adrenal crisis following a reassuringly normal ITT.

The safety and simplicity of the SST have, in many units, made it the 'pragmatic default test' of ACTH and cortisol reserve with the ITT reserved for specific circumstances such as recent onset hypopituitarism or when GH and cortisol status require testing.

Low-Dose Short Synacthen Test

The use of a lower dose of ACTH (typically 1 μg) in the assessment of secondary adrenal failure has been extensively investigated over several decades. In health, the entire stored pool of pituitary ACTH is of the order of 600 μg, such that an injected bolus of 250 μg produces plasma concentrations that are unphysiological and beyond the top of the ACTH/cortisol dose–response curve. Proponents of the low-dose (LD) SST argue that chronically understimulated adrenal glands may mount a satisfactory cortisol response to the unphysiological concentration of ACTH provided by 250 μg of synacthen, but that only normal glands will respond to the small doses used in this test. Further, plasma ACTH levels following injection of 1 μg are comparable to those reached during an ITT in healthy volunteers [36]. The test is quick (a single sample only is required 30 min after injection of ACTH) and the test may be performed at any time of day. While theoretically attractive and the focus of many publications, a recent survey of 766 endocrinologists over 60 countries concluded that the 92% of centres were performing 250 μg SST, 43% low-dose, and 37% both, which by implication means very few centres relied exclusively on the LD SST. It also reported that 72% of paediatric endocrinologists used the LD SST. The attraction of a lower dose of ACTH in children is not surprising [37].

The study also reinforces one of the major concerns with the LD SST, namely the lack of a standardized, reproducible means of dilution of a 250 mcg in 1 ml vial of synthetic ACTH. The authors reported variation in dilution methods resulted in the dose of synacthen administered in a LD SST varying between 0.16 μg and 0.81 μg [37]. ACTH is prone to be adsorbed onto the plastic of syringes or saline bags and to date, there is no quality controlled 1 μg synthetic ACTH vials commercially available, so 250-fold dilution is required.

Although theoretically attractive, a compelling case has not been made that it is superior to the SST, and does not entail the additional complexity of dilution.

Glucagon Stimulation Test

The mechanism of glucagon mediated stimulation of cortisol and GH secretion is ill-understood. The subcutaneous injection of glucagon causes a transient rise in plasma glucose. During the subsequent fall in glucose levels, ACTH and GH are both released and this has led to its widespread use as a means of assessing the reserve of these two hormones. Glucagon is a less powerful or reliable stimulus to ACTH or GH release than hypoglycaemia and false negative results are common. Administration of glucagon routinely makes patients feel unwell with nausea and may cause abdominal pain and vomiting. The glucagon test has not been the subject of intense study and the interpretation of the serum cortisol response relies upon criteria established for insulin-induced hypoglycaemia. However, it remains a useful method of assessing the hypothalamic–pituitary–adrenal and GH axes, particularly when the ITT is contraindicated [38]. A 'normal' response to glucagon administration provides reassurance of adequate HPA axis function but caution must be shown before concluding a patient has inadequate cortisol reserve based on

this test and a second, confirmatory test is advisable in most patients.

Metyrapone Test

This test of adrenal reserve was first described in the 1950s [39] and its role in the assessment of ACTH reserve has therefore changed with the availability of plasma ACTH and serum cortisol assays. Metyrapone inhibits the conversion of 11-desoxycortisol to cortisol by the enzyme 11β-hydroxylase (CYP11b1). The principle of the test is that the metyrapone-induced fall in serum cortisol stimulates ACTH secretion, which in turn increases corticosteroidogenesis and in particular a rise in circulating 11-desoxycortisol. As 11-desoxycortisol has no glucocorticoid activity a rise in its level has no effect on ACTH secretion. In patients with secondary adrenal insufficiency, the metyrapone-induced fall in cortisol does not stimulate an increase in ACTH secretion and hence no rise in serum 11-desoxycortisol occurs.

A typical protocol entails oral administration of 30 mg/kg metyrapone in hospital at midnight. Simultaneous cortisol and 11-desoxycortisol levels are taken between 8 and 9 a.m. and then oral glucocorticoids are administered if the index of suspicion of ACTH deficiency is high. An 11-desoxycortisol level above 200 nmol/L (7 μg/dl) indicates normal adrenal function, irrespective of the simultaneous cortisol value. Levels less than 200 nmol/L, in the presence of a low serum cortisol level, strongly suggest secondary adrenal insufficiency [39]. A low serum cortisol level is required for the interpretation of the test as an indicator that a stimulus has been delivered to the pituitary. Anticonvulsant therapy such as phenytoin accelerates the metabolism of metyrapone and an alternative test of the hypothalamic–pituitary–adrenal axis should be used in such patients.

A major criticism of this investigation is that it is a test of the ACTH-cortisol feedback mechanism rather than of ACTH reserve. The test requires measurement of serum 11-deoxycortisol either by immunoassay or preferably by mass spectrometry, neither of which are routinely available in most centres. The test is rarely undertaken in the United Kingdom and appears to have fallen out of favour internationally, as it requires hospitalization, and specialist assays.

Conclusion

It is inevitable that the debate about the optimum method for the assessment of the hypothalamic–pituitary–adrenal axis will continue. Practical issues such as cost and staff availability will, to a large extent, affect local policy but the fundamental clinical issue of patient safety remains the same. Dynamic tests of the integrity of the hypothalamic–pituitary–adrenal axis support, rather than substitute for, clinical decisions and it is important to recognize that the use of sophisticated statistical methods for the comparison of serum cortisol levels in groups of people with or without endocrine disease can never substitute for clinical awareness in the individual patient. Even the ITT, thought for so long to be the 'gold standard' for assessing ACTH reserve cannot provide complete reassurance that an individual patient will not develop secondary adrenal insufficiency during physiological stress. Changes in methodology and variation in the assays used for cortisol measurements hinder comparisons between published experiences of hypothalamic–pituitary–adrenal testing and make it difficult to recommend a diagnostic cut-off that defines cortisol deficiency, but a cortisol level >450 nmol/L, in most circumstances, makes the diagnosis very unlikely. Endocrine physicians should always educate their patients about the possible implications of pituitary disease in terms of the stress response, particularly when it is anticipated that the functioning of the hypothalamic–pituitary–adrenal axis may change over a period of time, such as following pituitary irradiation. The ITT is the single most reliable test of the hypothalamic–pituitary–adrenal axis but should only be performed under close supervision in specialist centres. If there is any doubt about the adequacy of ACTH reserve, it is sensible to err on the side of caution with respect to the provision of emergency steroid cover; and to consider a trial of oral glucocorticoid replacement therapy in patients with symptoms suggestive of chronic adrenal insufficiency and an equivocal response to dynamic testing.

Pituitary–Thyroid Axis

Like other pituitary cell types, thyrotrophs interact with their target cells by feedback inhibition, although the marked diurnal and ultradian variations that characterize ACTH and GH release are not present. This means that investigations other than basal blood samples (free thyroxine (fT_4) and TSH) are rarely necessary for the assessment of thyroid status in the context of pituitary disease. The challenge in the diagnosis is that TSH is rarely undetectable in hypopituitarism and therefore the biochemical confirmation of deficiency relies on the context (i.e. other pituitary hormone deficiencies and falling or low circulating fT_4 levels). TSH deficiency can be the most difficult to diagnose pituitary hormone deficiency as it can insidious in onset (e.g. after radiotherapy), and free T_4 levels can fall by 50% (e.g. 20–10 pmol/L) and still be within the reference range. Secondary hypothyroidism is strongly suggested by low levels of circulating T_4 in the presence of a low or low normal TSH. TSH deficiency in the absence of gonadotrophin and GH deficiency is very unusual. Illness ('sick euthyroid' syndrome), thyroxine-binding globulin deficiency, supraphysiological doses of glucocorticoids, and drugs such as phenytoin may also produce a similar picture. The interpretation of the results is dependent on the overall clinical context. An important, practical, point is that serum free thyroxine levels vary very little through life, so the availability of previous measurements may be particularly informative if values lie close to the lower limit of the reference range but secondary hypothyroidism is clinically suspected.

A novel approach to the identification of TSH deficiency has been proposed by Jostel *et al.* [41]. The TSH Index (TSHI) is an algorithm-based approach that calculates a 'fT_4-adjusted TSH' that corrects for any physiological TSH suppression to provide a true estimation of thyrotrophic function. The TSHI, based on simple thyroid function tests (TFTs) and regardless of concomitant T_4 supplementation, correlates closely with other measures of the severity of pituitary dysfunction (i.e. GH and cortisol deficiency) in patients with a known risk of hypopituitarism. Potentially it permits the diagnosis of TSH deficiency when fT_4 and TSH values are within the quoted reference ranges. Further work is required but this approach combined with the rapid advances in machine-learning

offer the prospect of reducing delay in the initiation of thyroxine replacement therapy in patients with hypopituitarism.

TRH Testing

TRH testing is of no value in diagnosing secondary hypothyroidism or predicting imminent TSH deficiency.

Growth Hormone Axis

Recombinant GH became available in the 1985 and was approved for the treatment of GHD in adults in 1996. Tests of GH reserve in adults will be described in this chapter; the symptoms and signs and diagnosis of childhood and adult GHD and its treatment are discussed elsewhere (see Chapter 2.3.6). Due to the lack of normative data and bias in assay performance, it must be appreciated that the biochemical definition of GHD (peak GH <3 ng/ml) for the ITT and glucagon test [42] is, in part, an arbitrary cut-off but all the same important in some countries as it has been adopted by regulatory authorities to define patients' eligibility for replacement therapy. From a physiological and therapeutic stand-point, it may be more accurate to think of patients with a peak GH response of less than 3 ng/ml as having severe GHD and, based on the available data, being likely to benefit from GH replacement therapy. Patients with peak GH response of more than 3 ng/ml may still have degrees of GHD but evidence for the benefit of GH replacement therapy is less robust. In contrast to ACTH deficiency, GHD is not life-threatening and replacement therapy is not mandatory. It is only of value to test for adult GHD if GH therapy is a therapeutic option and the patient is willing to self-inject daily. In some countries, in addition to meeting the biochemical criteria for deficiency evidence of impaired quality of life is required to qualify for GH therapy, so it is clinically inappropriate to perform complex stimulation tests in patients with a normal quality of life.

Normal GH secretion is pulsatile, with four to six pulses per 24 h, mostly at night in association with stage III–IV rapid eye movement (REM) sleep, punctuated by long periods when GH levels in blood are undetectable (<0.3 ng/ml). As with the assessment of the hypothalamic–pituitary–adrenal axis, a single basal blood sample is unlikely to yield significant diagnostic information, unless the taking of the sample coincides with a GH surge. An attractive approach, therefore, might seem to be to measure 24 h spontaneous profiles, as has been employed in the diagnosis of GHD in childhood. However, this approach has proved disappointing, as there is considerable overlap in the integrated growth hormone concentration (IGHC) of normal subjects and those of hypopituitary patients and the test has little diagnostic value [43]. Other physiological methods of assessing GH secretion include sampling during sleep and exercise, both of which are associated with GH release. However, all three of these methods are prohibitively expensive and time-consuming for routine clinical use.

The symptoms, such as lethargy, of GH deficiency in adults are not specific and similarly the dynamic test of GH status described next do not have 100% specificity. It is important that testing for GHD is only undertaken in patients with a plausible cause of GHD, rather than in an indiscriminate manner in individuals with non-specific symptoms, such as fatigue or central adiposity, and no known reason for developing hypopituitarism as inevitably such an approach will result in a proportion of 'false positive' investigations and unnecessary treatment.

Most, if not all, actions of GH are mediated through the peptide hormone insulin-like growth factor-I (IGF-I). However, measurement of serum IGF-I is of limited value in the diagnosis of adult-onset GHD, as 30% of patients with unequivocal GHD may have a serum in the lower half of the age-related reference range [43]. However, in most patients with hypopituitarism GHD is the first hormone to become deficient and therefore the most recent Growth Hormone Research Society Guidelines [42] state that in the clinical context of multiple pituitary hormone deficits a very low serum IGF-I (<2 SD below the mean) is sufficient to diagnose severe GHD and that such patients do not require a dynamic test of GH reserve.

Most authorities accept that pharmacological stimulation of GH release is the most practical and reproducible method of assessing GH reserve. Hypoglycaemia is a powerful stimulus to GH secretion and, over the years, the ITT has been the most frequently employed test in this regard. It has the advantage that ACTH reserve can be assessed simultaneously and, in experience hands, is a safe investigation provided the exclusion criteria outlined earlier are adhered to. The criterion for profound GHD is met if the peak GH response to insulin-induced hypoglycaemia and glucagon tests is 3 ng/ml or less. As with cortisol measurement and for many of the same reasons, there is significant bias between different GH immunoassays which makes comparison of GH data between centres difficult and must be borne in mind when applying consensus guidelines to local practice. Similarly, the endocrine physician should be alert to the confounding effect of obesity on GH release which may lead to a false positive diagnosis of GHD [42].

Where the ITT is contraindicated, alternative provocative tests of GH reserve include the glucagon (see earlier), arginine, GH-Releasing Hormone (GHRH)+arginine.

Endocrine Society [45] and the GH Research Society guidelines [42] advocate the use of the combination of GHRH and arginine stimulation testing on the basis of evidence suggesting sensitivity and specificity comparable to that of the ITT. However, GHRH is not readily available and therefore arginine is often used on its own. Arginine is one of several amino acids (others include leucine and histidine) that stimulate GH release, with a peak occurring between 30 and 120 min after infusion. It is generally used when a second dynamic test of GH reserve is required and the ITT is contraindicated. It involves the intravenous infusion of 0.5 g/kg (maximum dose 30 g) in 100 ml normal saline over 30 minutes and sampling for 2 hours thereafter. It is frequently used as an alternative to the ITT in patients with suspected hypopituitarism following traumatic brain injury.

The GH-releasing peptides (GHRP) are a collection of synthetic peptides that act via an alternative mechanism than GHRH to be potent stimulators of GH secretion and have been studied for both the diagnosis and treatment of GHD. Studies failed to establish the therapeutic benefit of GHRPs in patients with GHD and so, while they may be of value in the diagnosis of GHD, they are no longer generally available. There is interest in the use of ghrelin mimetics as a means of testing of the GH axis but data are not yet available.

Clonidine testing is of no value in diagnosing GHD in adults.

Assessment of the Pituitary–Gonadal Axis

Assessment of the pituitary–gonadal axis differs from other aspects of pituitary function testing. First, regular menstruation in a woman implies normal gonadotroph function and measurement of serum gonadotrophins and oestradiol therefore add little to the clinical assessment. Associated ovulation is not necessarily implied by regular menstruation; measurement of luteal phase progesterone levels is required for the assessment of subfertility in a patient with pituitary disease and a regular cycle. Second, social and age-related factors may influence the need to correct any underlying gonadal deficiency in men and women. The avoidance of cardiovascular complications and loss of bone mineral density (BMD) consequent upon prolonged hypogonadism is obviously desirable, but must be set against the temporal relationship of normal physiology. For example, an 80-year-old patient with secondary hypogonadism is likely to feel differently about sex steroid replacement therapy than a patient of 30 years.

It is rare for tests other than basal measurements of gonadotrophin hormones and sex steroid levels to be required for assessment of the pituitary–gonadal axis. Both oestradiol and testosterone bind to sex-hormone-binding globulin (SHBG), such that simultaneous measurement of SHBG and gonadal steroid levels are required to assess 'free' (biologically active) levels of these hormones (see Chapter 2.3.5). Testosterone should be measured fasted between 8 and 9 a.m., as levels show considerable diurnal variation. Oestradiol is best measured in the follicular phase of the menstrual cycle (if female patients are menstruating). Ovulation is assessed by measurement of progesterone in the luteal phase (days 18–25) of the cycle.

Historically, a combination of clomiphene and gonadotrophin-releasing hormone (GnRH) tests were used in an attempt to distinguish hypothalamic from pituitary causes of secondary gonadal failure. However, such information has little clinical value in terms of therapy, and newer, more sophisticated imaging techniques are able to distinguish these two groups of causes in the majority of cases. Central hypogonadism can be isolated, occur in the context of a hypothalamo–pituitary tumour or its treatment, or be the earliest sign of incipient panhypopituitarism. Isolated gonadotrophin deficiency will either be congenital, as in Kallmann's syndrome and associated with delayed/absent pubertal development, or be acquired and secondary to systemic illness (e.g. AIDS), excessive exercise (long distance running) or psychological disturbance (anorexia nervosa). In all cases it is imperative to investigate pituitary function in detail.

Clomiphene Testing

This investigation has been used in the investigation of suspected gonadotrophin deficiency. Clomiphene citrate is a selective oestrogen receptor modulator (SERM) acting as a weak oestrogen receptor antagonist at the hypothalamus and pituitary, and as an oestrogen agonist at the liver. In healthy subjects its antagonistic action at the hypothalamus stimulates gonadotrophin-releasing hormone levels to rise, with consequent release of LH and FSH. The hepatic effect is to induce SHBG synthesis and a rise in measured total testosterone and oestradiol levels. Hence, such a rise may be misleading and does not necessarily indicate increased gonadotrophin release. Clomiphene should not be given to patients with liver disease because of its oestrogen agonist effects. The test is also contraindicated in depression, because of the risk of mood disturbance, and patients should be warned of the risk of alteration of peripheral vision. In normal women, gonadotrophin levels double by 10 days and menstruation usually accompanies a positive clomiphene test in women. However, the test is of very limited clinical value as the ability of a progesterone challenge to induce a menstrual bleed is highly predictive of the response to clomiphene. A normal response to clomiphene in a patient with amenorrhoea offers reassurance that the axis is intact and that the problem lies in the hypothalamus, but offers little indication of the aetiology. Patients with weight, exercise, and stress-induced amenorrhoea can have either a normal or absent response to clomiphene, presumably indicative of the severity of the suppression of gonadotrophin secretion.

Gonadotrophin-Releasing Hormone Testing

In an era of sophisticated imaging and an increasing array of genetic explanations of hypogonadotropic hypogonadism, GnRH testing is of very limited value in the investigation of gonadotrophin deficiency.

Prolactin

A clinical syndrome associated with prolactin deficiency is not recognized and the clinical consequences of hyperprolactinaemia are discussed elsewhere. The principle value in serum prolactin measurement is as a guide to the aetiology and severity of hypopituitarism. Prolactin physiology differs from that of other anterior pituitary hormones in that its secretion is principally under tonic inhibition by release of dopamine from the hypothalamus. Levels do not show significant diurnal variation and so tests other than basal measurements are very rarely required. Physiological stress and various medications that interfere with dopamine action, such as metoclopramide, prochlorperazine, and several antipsychotics, raise serum prolactin. TRH stimulates prolactin release but provides no additional clinically useful information compared to random serum prolactin measurements. Serum prolactin on three separate occasions minimizes the risk of falsely elevated, stress-induced hyperprolactinaemia.

Conclusion

Accurate assessment of anterior pituitary function requires a sound knowledge of its normal physiology together with careful integration of clinical and biochemical information. As discussed earlier, certain aspects of the optimum method of pituitary function testing, notably the assessment of ACTH reserve, are still disputed, with local circumstances and personal preference often dictating the final choice. Physicians are advised to acquaint

themselves with their local laboratory reference ranges and never to allow a single hormonal measurement in a single patient on a single day to substitute for clinical awareness, particularly where an evolving endocrinopathy is anticipated, such as following pituitary irradiation.

REFERENCES

1. Al-Saleh I, Elkhatib R, Al-Rouqi R, Al-Enazi S, Shinwari N. The dangers of skin-lightening creams. *Toxicological and Environmental Chemistry*, 2012; **94**: 195–219.
2. Woods DR, Arun CS, Corris PA, Perros P. Cushing's syndrome without excess cortisol. *BMJ (Clinical Research Ed.)*, 2006; **332**: 469–70.
3. Brien TG. Human corticosteroid binding globulin. *Clin Endocrinol*, 1981; 14: 193–212.
4. Qureshi AC, Bahri A, Breen LA, *et al*. The influence of the route of oestrogen administration on serum levels of cortisol-binding globulin and total cortisol. *Clin Endocrinol*, 2007; **66**: 632–5.
5. Rodriguez-Arnao J, Perry L, Besser GM, Ross RJM. Growth hormone treatment in hypopituitary GH deficient adults reduces circulating cortisol levels during hydrocortisone replacement therapy. *Clin Endocrinol*, 1996; **45**: 33–7.
6. Perogamvros I, Underhill C, Henley DE, *et al*. Novel corticosteroid-binding globulin variant that lacks steroid binding activity. *J Clin Endocrinol Metab*, 2010; **95**: E142–50.
7. Raff H, Homar PJ, Burns EA. Comparison of two methods for measuring salivary cortisol. *Clin Chem*, 2002; **48**: 207–8.
8. Mak IYF, Au Yeung BYT, Ng YW, *et al*. Salivary cortisol and cortisone after low-dose corticotropin stimulation in the diagnosis of adrenal insufficiency. *J Endocr Soc*, 2017; **1**: 96–108.
9. Adaway JE, Keevil BG, Owen LJ. Liquid chromatography tandem mass spectrometry in the clinical laboratory. *Ann Clin Biochem*, 2015; **52**: 18–38.
10. Raff H, Auchus RJ, Findling JW, Nieman LK. Urine free cortisol in the diagnosis of Cushing's syndrome: is it worth doing and, if so, how? *J Clin Endocrinol Metab*, 2015; **100**: 395–7
11. Monaghan PJ, Owen LJ, Trainer PJ, Brabant G, Keevil BG, Darby D. Comparison of serum cortisol measurement by immunoassay and liquid chromatography-tandem mass spectrometry in patients receiving the 11β-hydroxylase inhibitor metyrapone. *Ann Clin Biochem*, 2011; **48**(Pt 5): 441–6.
12. Blair J, Adaway J, Keevil B, Ross R. Salivary cortisol and cortisone in the clinical setting. *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 161–8
13. Owen LJ, Adaway JE, Davies S, *et al*. Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine service laboratory. *Ann Clin Biochem*, 2013; **50**(Pt 4): 345–52.
14. Hawley JM, Owen LJ, Lockhart SJ, *et al*. Serum cortisol: an up-to-date assessment of routine assay performance. *Clin Chem*, 2016; **62**: 1220–9.
15. Arlt W, Lang K, Sitch AJ, Dietz AS, *et al*. Steroid metabolome analysis reveals prevalent glucocorticoid excess in primary aldosteronism. *JCI Insight*, 2017; **2**: pii: 93136.
16. Eisenhofer G, Masjkur J, Peitzsch M, *et al*. Plasma steroid metabolome profiling for diagnosis and subtyping patients with Cushing syndrome. *Clin Chem*, 2018; **64**: 586–96.
17. Scwartz E, Liu A, Randall H, *et al*. Use of steroids profiling by UPLC-MS/MS as a second tier test in newborn screening of congenital adrenal hyperplasia to supplement newborn screening programmes. *Pediatr Res*, 2009; **66**: 230–5.
18. Hicks R, Yee JK, Mao CS, *et al*. Precursor to product ratios reflect biochemical phenotype in congenital adrenal hyperplasia. *Metabolomics*, 2014; **10**: 123–31.
19. O'Reilly, *et al*. 11-oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *JCEM* 2017; **102**: 840–8.
20. Peitzsch M, Dekkers T, Haase M, *et al*. Utility of LC-MS/MS based adrenal vein steroid profiling: should cortisol remain the gold standard for selectivity and comparisons to aldosterone? *Experimental and Clinical Endocrinology and Diabetes* 2014; **0**: 122–34.
20. Plumpton FS, Besser GM. The adrenocortical response to surgery and insulin-induced hypoglycaemia in corticosteroid-treated and normal subjects. *Br J Surg*, 1969; **56**: 216–19.
21. Mattingly D. A simple fluorimetric method for the estimation of free 11-hydroxycorticoids in human plasma. *Journal of Clinical Pathology*, 1962; **15**: 374–9.
22. Gaskell SJ, Collins CJ, Thorne GC, Groom GV. External quality assessment of assays for cortisol in plasma: use of target data obtained by gas chromatography/mass spectrometry. *Clin Chem*, 1983; **29**: 862–7.
23. El-Farhan N, Pickett A, Ducroq D, *et al*. Method-specific serum cortisol responses to the adrenocorticotrophin test: comparison of gas chromatography-mass spectrometry and five automated immunoassays. *Clin Endocrinol*, 2013; **78**: 673–80.
24. Clark PM, Neylon I, Raggatt PR, Sheppard MC, Stewart PM. Defining the normal cortisol response to the short Synacthen test: implications for the investigation of hypothalamic-pituitary disorders. *Clin Endocrinol*, 1998; **49**: 287–92.
25. Hawley JM, Owen LJ, Lockhart SJ, *et al*. Serum cortisol: an up-to-date assessment of routine assay performance. *Clin Chem*, 2016; **62**: 262–9.
26. Hurel SJ, Thompson CJ, Watson MJ, Harris MM, Baylis PH, Kendall-Taylor P. The short Synacthen and insulin stress tests in the assessment of the hypothalamic-pituitary-adrenal axis. *Clin Endocrinol*, 1996; **44**: 141–6.
27. Jones SL, Trainer PJ, Perry L, Wass JAH, Besser GM, Grossman A. An audit of the insulin tolerance test in adult subjects in an acute investigation unit over one year. *Clin Endocrinol*, 1994; **41**: 123–8.
28. Pavord SR, Girach A, Price DE, Absalom SR, Falconer-Smith J, Howlett TA. (1992). A retrospective audit of the combined pituitary function test, using the insulin stress test, TRH and GnRH in a district laboratory. *Clin Endocrinol*, 1992; **36**: 135–9.
29. Greenwood FC, Landon J, Stamp TC. The plasma sugar, free fatty acid, cortisol, and growth hormone response to insulin. I. In control subjects. *J Clin Invest*, 1966; **45**: 429–36.
30. Pfeifer M, Kanc K, Verhovec R, Kocijancic A. Reproducibility of the insulin tolerance test (ITT) for assessment of growth hormone and cortisol secretion in normal and hypopituitary adult men. *Clin Endocrinol*, 2001; **54**: 17–22.
31. Vestergaard P, Hoeck H, Jakobsen P, Laurberg P. Reproducibility of growth hormone and cortisol responses to the insulin tolerance test and the short ACTH test in normal adults. *Horm Metab Res*, 1997; **29**: 106–10.
32. Wood JB, Frankland AW, James VH, Landon J. A rapid test of adrenocortical function. *Lancet*, 1965; **1**: 243–5.
33. Speckart PF, Nicoloff JT, Bethune JE. Screening for adrenocortical insufficiency with cosyntropin (synthetic ACTH). *Arch Int Med*, 1971; **128**: 761–3.

34. May ME, Carey RM. Rapid adrenocorticotropic hormone test in practice. Retrospective review. *Am J Med*, 1985; **79**: 679–84.
35. Darmon P, Dadoun F, Frachebois C, *et al*. On the meaning of low-dose ACTH(1–24) tests to assess functionality of the hypothalamic-pituitary-adrenal axis. *Eur J Endocrinol*, 1999; **140**: 51–5.
36. Cross AS, Helen Kemp E, White A, *et al*. International survey on high- and low-dose synacthen test and assessment of accuracy in preparing low-dose synacthen. *Clin Endocrinol*, 2018; ***88***: 744–51.
37. Littley MD, Gibson S, White A, Shalet SM. Comparison of the ACTH and cortisol responses to provocative testing with glucagon and insulin hypoglycaemia in normal subjects. *Clin Endocrinol*, 1989; **31**: 527–33.
38. Liddle GW, Estep HL, Kendall JW, Williams WC, Townes AW. Clinical application of a new test of pituitary reserve. *J Clin Endocrinol Metab*, 1959; **19**: 875–94.
39. Spiger M, Jubiz W, Meikle AW, West CD, Tylor FH. Single-dose metyrapone test: review of a four-year experience. *Arch Int Med*, 1975; **135**: 698–700.
40. Jostel A, Ryder WDJ, Shalet SM. The use of thyroid function tests in the diagnosis of hypopituitarism: definition and evaluation of the TSH Index. *Clin Endocrinol*, 2009; **71**: 529–34.
41. Ho KKY. Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, J. *Eur J Endocrinol*, 2007; **157**: 695–700.
42. Shalet SM, Toogood A, Rahim A, Brennan BMD. The diagnosis of growth hormone deficiency in children and adults. *Endocrine Rev*, 1998; **19**: 203–23.
43. Hoffman DM, Nguyen TV, O'Sullivan AJ, Baxter RC, Ho KK. ITT and IGF-I in the diagnosis of growth hormone deficiency in adults. *Lancet*, 1994; **344**: 613–14.
44. Molitch AME, Clemmons DR, Malozowski S, Merriam GR, Vance ML. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 1587–609.

1.7

Endocrine Autoimmunity

Simon H.S. Pearce and Catherine J. Owen

Introduction

Autoimmune attack directed at endocrine target tissues is responsible for several conditions that present diverse management problems to the clinical endocrinologist, the most prevalent being the autoimmune thyroid diseases (AITDs) and type 1 diabetes mellitus (T1D). Addison's disease, premature ovarian insufficiency, hypophysitis, and hypoparathyroidism are less frequently encountered endocrine conditions that may also have an autoimmune basis. These conditions typically develop insidiously and their immunological hallmarks may be present for many years before any clinical presentation. Most of these conditions are inherited as complex genetic traits and knowledge about the immunogenetics underlying their susceptibility has advanced in recent years. In contrast, we remain largely in the dark about the molecular mechanisms through which both tissue-specific and environmental factors contribute to disease risk. At the moment most of these conditions are managed by replacing the missing hormone(s), however, immunomodulatory therapy aimed at modifying the natural history of these diseases may hold some hope for the future.

Diversity and Natural History of Endocrine Autoimmunity

With the exception of Graves' disease, autoimmune endocrine conditions present owing to clinical manifestations caused by reduced secretion of one or more hormones. Because of the multisystem influences of adrenal and thyroid hormones, the symptoms are frequently non-specific, reflecting disturbances to several bodily systems. In contrast, the features of ovarian insufficiency and insulin deficiency are relatively specific. These presentations with hormone deficiency or excess are different from those of other autoimmune diseases, such as rheumatoid arthritis or multiple sclerosis, which typically manifest with pain or reduced function at the site of autoimmune attack due directly to the associated inflammatory process. The inherent reserve capacity of hormone secretion and the regenerative capacity of many endocrine tissues means that these conditions don't present until the disease process has inflicted a critical amount of tissue damage leading to hormonal insufficiency.

Pancreatic islet cell antibodies are detectable in 70% of children who go on to develop T1D by the age of 5yrs, with about 11% of children with multiple islet cell antibodies developing overt hyperglycaemia per year [1]. Similarly, the rate of progression to overt hypothyroidism for people with positive thyroid peroxidase antibodies is around 5% per year [2]. In contrast, serum steroid 21-hydroxylase antibodies are rare in the population, but around 50% of adults found to have them will ultimately develop overt adrenal insufficiency [3]. In most conditions, these circulating autoantibodies are not believed to be a direct cause of tissue destruction/hormonal dysfunction, but are simply a marker for the autoimmune attack on the target tissue. The exception to this rule is Graves' disease, where antibodies that stimulate the thyroid-stimulating hormone (TSH) receptor directly lead to thyrocyte hypertrophy causing goitre and thyroid hormone hypersecretion [4]. For this reason, it seems likely that hyperthyroid Graves' disease becomes clinically apparent much earlier in the autoimmune process than other autoimmune diseases, because substantial tissue damage isn't necessary for its clinical presentation. This might explain its natural history to remit and relapse, whereas the majority of other autoimmune endocrinopathies seem to be monophasic, with a steady destruction of the target organ leading to hormonal insufficiency. Nevertheless, it has to be acknowledged that once patients are started on hormone replacement therapy, it is rare to re-evaluate whether there may be endocrine recovery at a later date. Thus, it remains possible that we are overtreating patients with several conditions, or treating for longer than necessary, if an unidentified proportion of patients ultimately recover function.

Epidemiology and Disease Clustering

AITDs are the commonest of the autoimmune disorders affecting around 4% of people over a lifetime with either Graves' disease or Hashimoto thyroiditis [2, 5] (Table 1.7.1). Both have a strong female proclivity (F to M 5:1) and a peak age of onset in the fifth decade of life. Autoimmune Addison's disease, while much rarer, has a similar age distribution of occurrence but the female proclivity is less strong, around 2:1. In contrast, type 1 diabetes presents in childhood, adolescence, and young adults, affects 0.5% of the population, and has no gender bias (Figure 1.7.1). Idiopathic autoimmune hypophysitis is most commonly found during the last trimester or in the year following pregnancy, however, drug-induced hypophysitis following immune checkpoint inhibitor therapy (ipilimumab) for cancer is rapidly changing the prevalence of this hitherto rare condition.

There is a strong tendency for each of these organ-specific autoimmune diseases to cluster with each other in an individual or in families. This is most marked for Addison's disease, where 50–65% of people will have a second autoimmune condition, broadly termed the autoimmune polyendocrinopathy syndrome type 2, or APS2 [6]. Autoimmune thyroid disease occurs in 70–90% of APS2 people and type 1 diabetes in 5–15%. On presentation with Addison's disease, type 1 diabetes already exists in around 10% and preventable deaths still occur in T1D patients with undiagnosed adrenal failure. Other conditions that may commonly cluster in APS2 patients in decreasing order of prevalence are pernicious anaemia, coeliac disease, vitiligo, and premature ovarian insufficiency.

Physiology of Immune Tolerance

In health, the two main arms of the adaptive immune system, T and B lymphocytes, work together via both cellular and humoral responses to protect the body against infections and tumours. Ensuring that the immune response is regulated to prevent autoimmunity is a complex process that starts early *in utero*. The immune system undergoes a process of 'education' in order to be able to exert an effective immune response while maintaining tolerance to self-antigens. This process is initially due to the development of central tolerance, however, it has recently become clear that this alone is insufficient to prevent autoimmunity and there are also important peripheral mechanisms involved.

Development of central tolerance: Immune tolerance begins *in utero* in the fetal thymus. Thymic medullary epithelial cells (mTECS) play a critical role in central tolerance—they have the ability to present antigens that are only otherwise expressed in a tissue-specific manner outside the thymus, such as glutamic acid decarboxylase (GAD) and insulin. These antigens are expressed together with costimulatory and major histocompatibility complex (MHC) molecules allowing engagement with T lymphocytes. This promiscuous gene expression in mTECs is partly controlled by the transcription factor AIRE (AutoImmune REgulator), which is responsible for the expression of the tissue-specific antigens enabling the thymus to represents the wide range of antigens that encompass the 'peripheral self' [6, 7].

The maturing T lymphocytes in the thymus (thymocytes) develop a diverse T-cell receptor (TCR) repertoire through numerous somatic recombination events. They then undergo a process of positive and negative selection by interacting with the mTECs that express high levels of MHC class I or II molecules associated with self-antigens. The nature of the T-cell interaction with MHC molecules determines its future, and if the CD4 + CD8+ thymocyte TCR is able to bind with sufficient affinity it receives survival signals (positive selection), ensuring that all mature T cells have some affinity for MHC molecules. Those thymocytes that interact with self-antigen–MHC class I develop into CD8+ cells, while those interacting with self-antigen–MHC class II complexes form CD4+ T cells. Those thymocytes whose TCRs do not bind with sufficient affinity do not receive any survival signals and become unresponsive (anergic). A negative selection process follows the positive selection, where potentially autoreactive thymocytes that bind the self-antigen–MHC complex with high affinity are either removed by apoptosis (known as clonal deletion), or differentiate into regulatory T cells [8]. Therefore, thymocytes that bind to the self-antigen–MHC complexes with low affinity will recognize antigen–MHC complexes but not become autoreactive, are allowed to develop as mature CD4+ or CD8+ T cells and leave the thymus. The CD4+ cells then develop into different T-cell subsets depending on the local cytokine environment. These include T-helper (Th) 1, Th2, and Th17 cells. In addition, some CD4+ thymocytes expressing TCRs that do bind with high affinity to self-antigen–MHC complexes express the forkhead box protein-P3 (FOXP3) and develop into T regulatory (Treg) cells [8, 9]. B cells undergo a similar process of positive and negative selection occurring in the bone marrow. Breakdown or dysregulation of this process at any of the aforementioned steps may lead to potentially autoreactive T cells escaping to the periphery, with the potential to initiate autoimmunity.

Peripheral tolerance: Although most self-reactive T cells are deleted in the thymus, some escape thymic negative selection and can be found in the periphery of most people. However, few of these go on to develop autoimmune disease, which supports the existence of peripheral tolerance which keeps these potentially 'auto reactive T cells' under control. The activation and proliferation of these cells in response to self-antigens is prevented by several different peripheral immune regulatory mechanisms, many of which are not fully understood. Some of these mechanisms are intrinsic to the T cell itself, such as activation-induced cell death, anergy [10], immunological ignorance, and phenotypic skewing. However, a key mechanism for peripheral immune tolerance is through the induction of regulatory T cells that inhibit effector T-cell proliferation.

Regulatory T cells are unique in that they develop both centrally (thymus derived) and peripherally. Central regulatory T cells form from CD4+ thymocytes that show high affinity interactions with self-peptide–MHC complexes in the thymus, and instead

Table 1.7.1 Prevalence of common endocrine autoimmune conditions

Condition	Prevalence per 100 000 population	% female
Graves hyperthyroidism	1151	88
Hashimoto's thyroiditis	791	95
Type 1 diabetes mellitus	192	48
Addison's disease	5	92

Data taken from Jacobson DL *et al.* 1997 (1)

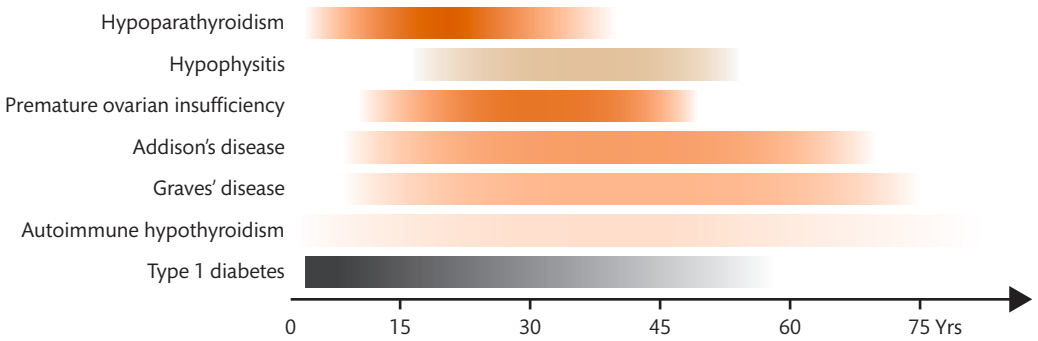


Figure 1.7.1 Age-related incidence of different autoimmune endocrinopathies.
Type 1 diabetes has the highest incidence in children and young people, and gets progressively less frequent with ageing. In contrast, autoimmune thyroid diseases and Addison's disease start to present around puberty, with a peak onset at around the age of 40 yrs. Hypoparathyroidism is almost always found in association with APS1, and commonly presents between the ages of 5 and 15 years in this context.

of undergoing apoptosis, if they express the transcription factor FOXP3+ and high levels of the IL-2 receptor alpha chain (CD25+), differentiate into the regulatory T-cell lineage [11, 12]. In addition, induced regulatory T cells, such as Tr1 cells, arise in the periphery after engagement in prolonged interactions with low affinity self-antigens (not present in the thymus) and in the gut following exposure to food antigens, bacterial flora and pathogens [13].

Regulatory T cells recognize a wide array of antigens and have a higher avidity for self-antigen/MHC complexes than other T cells, but show a limited capacity to proliferate when their TCR is engaged. Their role is to inhibit activation, proliferation, and cytokine synthesis by effector T and B cells. They exert this immunosuppressive effect through different mechanism such as the synthesis of anti-inflammatory cytokines such as interleukin (IL)-10, IL-35, and TGFβ or via cell contact. In summary, there are two functionally distinct populations of CD4+ T cells, one capable of mediating autoimmune disease and one dominantly inhibiting it, and the balance between these cell-types two maintains immune tolerance.

Breakdown of Tolerance: Autoimmunity

Endocrine autoimmunity is a consequence of the loss of immune tolerance towards tissue-specific self-antigens, and in most conditions this leads to tissue destruction and gland failure, which is mediated by cytotoxic T cells. In most conditions there is also a humoral immune response manifest as autoantibody production, although it is unclear whether this plays a direct pathogenic role or is just an epiphenomenon. The exception to this is Graves' disease, where circulating anti-TSH receptor antibodies are a prerequisite of hyperthyroidism. Nevertheless, histologically Graves' thyroid tissue shows a mixed T- and B-cell infiltrate, underlining that this isn't purely a B-cell mediated disorder. Furthermore, B-cell depletion therapy has been shown to ameliorate features of autoimmune disease in both the non-obese diabetic (NOD) mouse and in newly diagnosed type 1 diabetes patients, suggesting a significant role for B-cell at least in perpetuating the immune response [14].

Many organ-specific autoimmune diseases affect endocrine glands and it is hypothesized that this may be due to a critical role for thymic antigen presentation of low-abundance, specialist antigens such as those found in the biosynthesis, secretion, and regulation of the various hormones. This is best illustrated by the monogenic example of the autoimmune polyglandular syndrome type 1 (APS1), owing to loss of function of the AIRE gene (see next). In most other conditions, a complex interplay of multiple genetic variants interacting with environmental and endogenous factors seem to determine disease proclivity.

Genetic Factors

Monogenic Conditions

It is clear that the genetic basis of autoimmunity is complex in most autoimmune conditions, and we still do not have a clear understanding of how and why these conditions develop [9]. There are however a few monogenic forms of autoimmunity that have led to huge advances in our understanding of the development of self-tolerance and how autoimmunity develops, and further studies of these conditions are continuing to advance our knowledge. Two such monogenic endocrine polyendocrinopathies are autoimmune polyglandular syndrome type 1 (APS1) and the immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX)—these are discussed further next.

APS1: also known as the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome (APECED), is a rare and frequently debilitating disorder of childhood.

It is characterized by organ-specific autoimmunity affecting endocrine organs, skin, and the gastrointestinal (GI) tract. Most patients develop at least two of the three main components—chronic mucocutaneous candidiasis, autoimmune hypoparathyroidism, and autoimmune primary adrenal failure. In addition, over 40 other clinical manifestations have been associated with APS1, and many of these are autoimmune in nature [6, 15]. APS1 is inherited as an autosomal recessive condition and more than 120 different disease-causing mutations have now been identified by the defective autoimmune regulator gene, *AIRE*. It has been shown that *AIRE* encodes a nuclear transcription factor that plays a key role in the development of central immune tolerance. The highest levels of AIRE are found in mTECs where it induces the expression of tissue-specific antigens, which is an essential step in the development of central

immune tolerance. AIRE deficient mTECs show downregulation of the level of organ-specific gene transcripts compared with wild type mTECs, so in the absence of AIRE, the number of tissue-specific antigens expressed on mTECS is reduced, with consequent failure of negative selection of developing autoreactive thymocytes, which escape into the periphery [16]. AIRE deficient mTECs still express normal levels of some self-antigens such as GAD and C-reactive protein (CRP), so AIRE seems to only regulate a subset of transcripts. The full action of AIRE is still poorly understood and it may play further roles in the development of immune tolerance other than directing thymic antigen presentation. Evidence suggests AIRE may also have a role in thymocyte migration through modulating chemokine expression, helping to regulate mTEC apoptosis and differentiation [17]. In AIRE deficiency regulatory T cells express less FOXP3, so AIRE may affect peripheral immune tolerance as well by ensuring self-reactive thymocytes are directed into the regulatory T-cell lineage [18].

Most APS 1 patients are homozygous or compound heterozygous for loss of function mutations, so have no functional *AIRE* transcripts. Recently however, dominant-negative heterozygous AIRE mutations have been described [19], that are located in the AIRE PHD1 (first plant homeo-domain) zinc finger and are associated with a milder and later onset APS1 phenotype. Thus, it is now apparent that disease-causing mutations in *AIRE* are probably more common than previously appreciated and may cause a more variable autoimmune phenotype.

Immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX): IPEX is a rare and devastating X-linked condition of male infants, affecting immune regulation and resulting in multiple autoimmune disorders. Affected males develop severe diarrhoea, failure to thrive, type 1 diabetes, and thyroiditis, usually in the first year of life. Other features include eczema, other autoimmune conditions, and immunodeficiency. IPEX is due to mutations in the *FOXP3* gene, located at Xp11, encoding a transcription factor called 'scurfin' which belongs to the forkhead/winged-helix family. Several mutations have been reported in both the coding region and regulatory regions of FOXP3 [20, 21]. FOXP3 is specifically expressed in naturally arising CD4+CD25+ regulatory T cells, and appears to convert naïve T cells to this regulatory phenotype. Thus, FOXP3 is a critical regulator of Treg development and function [11]. Severe autoimmunity in FOXP3 deficiency is thought in part to be due to the inability of CD4+CD25+ Tregs to develop and therefore there is a lack of suppression of self-reactive effector T cells that have escaped to the periphery. Although female carriers of FOXP3 mutations appear to be healthy, a small number of cases of an IPEX-like syndrome have been reported in families with affected girls in whom no mutation was found. Deleterious mutations at several autosomal loci, including the IL-2 receptor subunit *CD25, CTLA4*, and *LRBA* have been shown to cause conditions with overlapping clinical features [22, 23]. This genetic heterogeneity may explain some of the clinical variation seen in this syndrome but, as yet, no obvious genotype–phenotype relationship has been identified, and other modifying genes, such as *HLA*, as well as environmental factors may influence the outcome. Importantly, the identification of the *AIRE* and *FOXP3* genes as defective in APS1 and IPEX have led to huge leaps in our understanding of the control of thymic T-cell selection and the development of regulatory T cells and how loss of function in these critical pathways leads to the development of autoimmunity.

Complex 'Polygenic' Conditions

Apart from the special cases of APS1 and IPEX just described, each of the individual autoimmune endocrinopathies and their 'combined' APS2 syndrome are inherited as complex conditions; meaning that multiple genetic susceptibility variants interact with environmental and/or endogenous factors to lead to the disease.

The MHC on the short arm of chromosome 6 is quantitatively the dominant susceptibility locus for all of the common autoimmune endocrinopathies. Strong association of type 1 diabetes, Graves' disease, Hashimoto thyroiditis, and Addison's disease with alleles of the human leukocyte antigen (HLA) genes within the class II MHC region, which includes *HLA-DRB, -DQA*, and *-DQB* has been consistently found [24]. Furthermore, some susceptibility alleles are common to more than one condition; for instance, *HLA-DRB1*0301* (*'DR3'*) confers genetic risk of type 1 diabetes, Graves' disease, Hashimoto's thyroiditis, and Addison's disease. Carrying one susceptibility HLA allele increases the risk of type 1 diabetes three- to sixfold [25], and for Graves' disease there is a twofold increase over the background population. Fine mapping to the amino acid level of the HLA-DRB molecules has demonstrated that an arginine at position β74 is the strongest determinant of susceptibility, although residues at other positions do have lesser influences. Variation in HLA molecules likely influences the spectrum of antigens that can be presented to T lymphocytes and the avidity of the HLA–TCR interaction, which may be acting during the thymic selection of T lymphocytes.

Several variations in other immune molecules have been associated with multiple autoimmune endocrine conditions. A series of genetic polymorphisms in the cell-surface receptor cytotoxic T-lymphocyte antigen-4 (*CTLA-4*) have been associated with type 1 diabetes, Graves' disease, Hashimoto's thyroiditis, Addison's disease, coeliac disease, and other related conditions [26]. CTLA-4 is a costimulatory receptor that binds T lymphocytes to antigen-presenting cells and which acts to downregulate T-lymphocyte signalling once antigen presentation has occurred to the TCR. It now seems probable that the disease-associated variations within the *CTLA-4* gene encode receptor isoforms with higher probabilities of being shed from the T-cell surface and that these soluble CTLA-4 molecules may saturate the antigen-presenting cell partner CD86 and prevent it from binding the structurally homologous but activating T-cell molecule, CD28 [27]. An additional molecular variant that is expressed in T lymphocytes and that has been widely associated with endocrine autoimmunity is known as protein tyrosine phosphatase non-receptor-22 (*PTPN-22*). A coding variant (*R620W*) in PTPN-22 has been associated with type 1 diabetes, Graves' disease, Hashimoto's thyroiditis, Addison's disease, and several other non-endocrine autoimmune conditions [28]. Interestingly, this PTPN-22 susceptibility allele has only been found to be associated with autoimmunity in populations of white European ancestry, whereas in East Asian populations, different, non-coding variations of PTPN-22 seem to confer susceptibility. Like CTLA-4 variants, PTPN-22 alleles may act as a brake to T-lymphocyte activation, by inactivating intracellular signalling molecules downstream of the TCR, such as Lck.

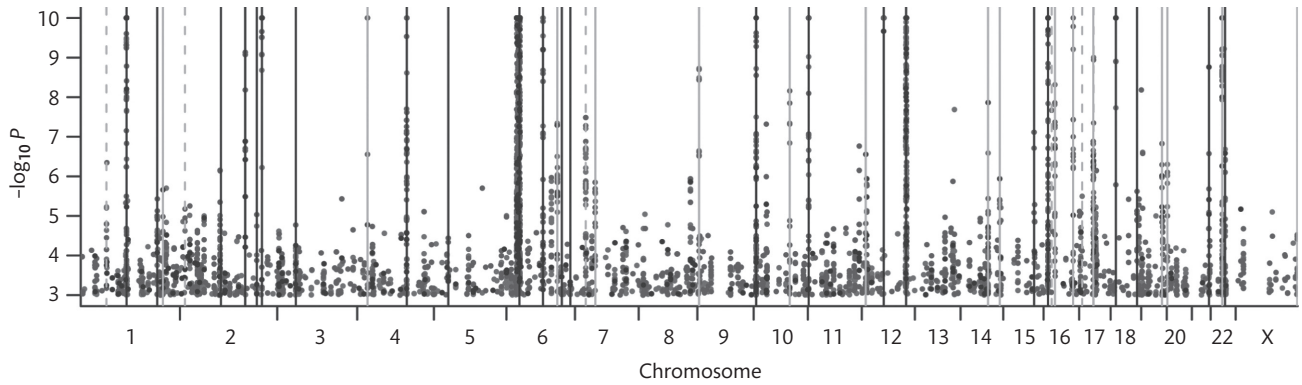


Figure 1.7.2 Manhattan plot of genetic associations for type 1 diabetes. Genomic single-nucleotide polymorphism (SNP) locations on human chromosomes 1–22, X are shown on the x-axis with the $-\log_{10}P$ value in favour of association on the y-axis. High densities of associated SNPs are found in many locations, for instance chromosome 1 (PTPN22), chromosome 6 (HLA), chromosome 12 (CD25).

Discovery-based genomic methods have now uncovered many additional genetic variants that each contribute a few per cent to the overall susceptibility to autoimmune diseases. Currently type 1 diabetes is the best studied of the conditions and genome-wide association scans have shown that allelic variants in more than 60 different genes underlie the proclivity to this disorder (Figure 1.7.2) [1, 29]. Knowledge of the diversity of the genes in which susceptibility variants have been found allows some insight into disease pathogenesis and highlights novel pathways that could be modified in attempts to ameliorate the underlying autoimmune disease process. For instance, variations in the *CD25* gene (a subunit of the interleukin 2-receptor) have been found in type 1 diabetes and AITDs, leading to current trials of low-dose IL2 therapy in recent-onset T1D [30]. Similarly, variants in B lymphocyte activating factor (BAFF) have been found in Graves' disease as well as multiple sclerosis and systemic lupus erythematosus, suggesting that specific anti-B lymphocyte therapies might be efficacious in these conditions. In addition, improved knowledge of the diverse genetic burden has begun to be used in multiallelic stratification scores to more accurately quantify disease risk in currently unaffected children who have affected family members with T1D [31]. The ultimate aim being that such stratification may be able to more precisely predict future development of T1D and that the highest risk individuals could be targeted for preventative therapies.

Antigen and Tissue-Specific Factors

As well as immune system variants, genetic studies have demonstrated that alleles of the insulin and TSH receptor genes are associated with T1D and Graves' disease, respectively [32, 33]. This establishes that variations in the target antigens, likely involving quantitative differences in expression pattern or regulation rather than qualitative alterations of the protein itself, are also important in determining the specific endocrine organ involved in autoimmunity. An important recent development is the discovery that T cells from the NOD mouse model of T1D and from T1D patients recognize hybrid insulin peptides, formed by fusion of proinsulin C-peptide and chromogranin or islet amyloid polypeptide (IAPP) fragments [34]. This finding of peptide rearrangement explains how such novel antigens can elude the thymic immune education process, and provides another layer of complexity to disease pathogenesis.

Environment and Endogenous Factors

Many different non-inherited factors have been shown to influence the risk of autoimmunity. Some are likely to be common to more than one autoimmune disorder such as stress or vitamin D insufficiency, and others may be highly specific for a certain condition such as a viral infection with a specific tissue tropism, or environmental iodine exposure which directly affects thyroid function.

Among the best characterized conditions are Graves' disease and type 1 diabetes. As mentioned earlier, dietary iodine supply is a significant factor in determining the prevalence of Graves' disease in a population. Iceland, where iodine intake is high, has more than double the rate of Graves' disease than found age-matched individuals in East Jutland, Denmark, where iodine intake was low [35]. Cigarette smoking has a clear association with thyroid-associated orbitopathy (TAO) and is also implicated in both prevalence of hyperthyroid Graves' disease and remission rate following antithyroid drug therapy. In contrast, smoking has been shown to protect against Hashimoto thyroiditis [36]. It is believed that thiocyanate compounds contained within cigarette smoke may perturb thyroid hormone production. Stress, in the form of adverse life events has been implicated in Graves' disease pathogenesis. Analysis of incidence during spells of national conflict have shown increased rates, although other, major changes that occur simultaneously could be confounding factors (e.g. dietary changes during wartime). In addition, significant life events (such as bereavement in the year before the onset of the disease) have been linked to the onset and remission rate of Graves' disease [37, 38]. In a prospective study, no correlation was found between stressful life events and thyroid peroxidase antibody (TPO Ab) status, suggesting that alterations in the cytotoxic or innate arms of the immune system might be involved [39]. Pregnancy is also a well-recognized risk factor for Graves' disease [40]. The mechanism for this presumably relates to a relative suppression of immune function during pregnancy and a

rebound of immune activity in the postpartum period. In contrast, oestrogen use protects against Graves' disease.

Several lines of evidence suggest that enteroviral infection, either maternal or in the infant, predisposes to type 1 diabetes. Enteroviral ribonucleic acid (RNA) has been identified in the blood of 10–50% of newly diagnosed T1D patients and serological evidence of enteroviral infection supports a high frequency of previous infection. Nevertheless, enteroviral infection is common in healthy controls and T1D patients have a distinct immunogenetic make up which might result in differences in viral clearance or antibody response that could be incidental to any pathogenic role for enteroviruses [41]. High birthweight, rapid early childhood growth, obesity, and other states of relative insulin resistance have been found to predispose to T1D. This gives rise to the accelerator hypothesis, which proposes that these increased insulin demands drive the autoimmune response against the pancreatic β cell [42]. Along similar lines to Graves' disease, adverse psychological stresses have been associated with T1D risk [43]. It seems possible that the metabolic consequences of this, including hypercortisolaemia and sleep disturbance could also lead to increased insulin demand ('β cell stress'), as well as direct effects on immune function. In addition, there are nutritional influences for T1D, including reductions in incidence in vitamin D supplemented children, and with diets containing more omega-3 polyunsaturated fats [43]. With the exception of vitamin D status, it remains largely unknown how these different environmental influences on T1D risk are mediated at a molecular level, and this remains an important subject for future research.β

Future Directions and Summary

Study of the rare monogenic polyendocrinopathy syndromes has unveiled fundamental insights into the mechanisms of thymic immune tolerance and peripheral T-cell regulation. In the commoner, complex conditions, several variants in both immune system and target antigen genes have been implicated. Several new treatment approaches are in development based on this understanding, including clinical trials of low-dose IL2, which may bolster Treg numbers and reduce cytotoxic T-cell responses in T1D [30]. In addition, knowledge of T-cell epitopes has led to promising studies of desensitization to target antigenic peptides [44]. Current therapies for autoimmune endocrinopathy still involve replacing the 'missing' hormone; it seems likely that specifically targeted immunomodulatory therapies may become a reality over the next few years.

REFERENCES

1. Pociot F, Lernmark Å. Genetic risk factors for type 1 diabetes. *Lancet*, 2016; **387**: 2331–9.
2. Vanderpump MP, Tunbridge WM, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol*, 1995; **43**: 55–68.
3. Coco G, Dal Pra C, Presotto F, *et al*. Estimated risk for developing autoimmune Addison's disease in patients with adrenal cortex autoantibodies. *J Clin Endocrinol Metab*, 2006; **91**: 1637–45.
4. Smith TJ, Hegedüs L. Graves' disease. *N Engl J Med*, 2016; **375**:1552–65.
5. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. *Clin Immunol Immunopathol*, 1997; **84**: 223–43.
6. Husebye ES, Anderson MS, Kämpe O. Autoimmune polyendocrine syndromes. *N Engl J Med*, 2018; **378**: 1132–41.
7. Metzger TC, Anderson MS. Control of central and peripheral tolerance by AIRE. *Immunol Rev*, 2011; **241**: 89–103.
8. Kalekar LA, Schmiel SE, Nandiwada SL, *et al*. CD4(+)T cell anergy prevents autoimmunity and generates regulatory T cell precursors. *Nat Immunol*, 2016; **17**: 304–14.
9. Rosenblum MD, Remedios KA, Abbas AK. Mechanisms of human autoimmunity. *J Clin Invest*, 2015; **125**: 2228–33.
10. Singh NJ, Schwartz RH. The strength of persistent antigenic stimulation modulates adaptive tolerance in peripheral CD4+ T cells. *J Exp Med*, 2003; **198**: 1107–17.
11. Lu L, Barbi J, Pan F. The regulation of immune tolerance by FOXP3. *Nat Rev Immunol*, 2017; **17**: 703–17.
12. Lee W, Lee GR. Transcriptional regulation and development of regulatory T cells. *Exp Mol Med*, 2018; **50**: e456.
13. Peron JP, de Oliveira AP, Rizzo LV. It takes guts for tolerance: the phenomenon of oral tolerance and the regulation of autoimmune response. *Autoimmun Rev* 2009; **9**: 1–4.
14. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, *et al.*; Type 1 Diabetes TrialNet Anti-CD20 Study Group. Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. *N Engl J Med*, 2009; **361**: 2143–52.
15. Bruserud Ø, Oftedal BE, Landegren N, *et al.* A Longitudinal Follow-up of Autoimmune Polyendocrine Syndrome Type 1. *J Clin Endocrinol Metab*, 2016; **101**: 2975–83.
16. Fierabracci A. Recent insights into the role and molecular mechanisms of the autoimmune regulator (AIRE) gene in autoimmunity. *Autoimmun Rev*, 2011; **10**: 137–43.
17. Passos GA, Mendes-da-Cruz DA, Oliveira EH. The role of AIRE, microRNAs and cell-cell interactions on thymic architecture and induction of tolerance. *Front Immunol*, 2015; **6**: 615.
18. Malchow S, Leventhal DS, Lee V, Nishi S, Socci ND, Savage PA. AIRE enforces immune tolerance by directing autoreactive T cells into the regulatory T cell lineage. *Immunity*, 2016; **44**: 1102–13.
19. Oftedal BE, Hellesen A, Erichsen MM, *et al*. Dominant mutations in the autoimmune regulator AIRE are associated with common organ-specific autoimmune diseases. *Immunity*, 2015; **42**: 1185–96.
20. Wildin RS, Freitas A. IPEX and FOXP3: clinical and research perspectives. *J Autoimmun*, 2005; **25**: Suppl: 56–62.
21. Owen CJ, Jennings CE, Imrie H, *et al*. Mutational analysis of the FOXP3 gene and evidence for genetic heterogeneity in the immunodysregulation, polyendocrinopathy, enteropathy syndrome. *J Clin Endocrinol Metab*, 2003; **88**: 6034–9.
22. Caudy AA, Reddy ST, Chatila T, Atkinson JP, Verbsky JW. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. *J Allergy Clin Immunol*, 2007; **119**: 482–7.
23. Lo B, Zhang K, Lu W, *et al*. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science*, 2015; **349**: 436–40.
24. Gough SC, Simmonds MJ. The HLA region and autoimmune disease: associations and mechanisms of action. *Curr Genomics*, 2007; **8**: 453–65.

25. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr Diab Rep*, 2011; **11**: 533–42.
26. Vaidya B, Pearce S. The emerging role of the CTLA-4 gene in autoimmune endocrinopathies. *Eur J Endocrinol*, 2004; **150**: 619–26.
27. Walker LS, Sansom DM. The emerging role of CTLA4 as a cell-extrinsic regulator of T cell responses. *Nat Rev Immunol*, 2011; **11**: 852–63.
28. Bottini N, Peterson EJ. Tyrosine phosphatase PTPN22: multifunctional regulator of immune signaling, development, and disease. *Annu Rev Immunol*, 2014; **32**: 83–119.
29. Roizen JD, Bradfield JP, Hakonarson H. Progress in understanding type 1 diabetes through its genetic overlap with other autoimmune diseases. *Curr Diab Rep*, 2015; **15**: 102.
30. Long SA, Buckner JH, Greenbaum CJ. IL-2 therapy in type 1 diabetes: 'trials' and tribulations. *Clin Immunol*, 2013; **149**: 324–31
31. Bonifacio E, Beyerlein A, Hippich M, *et al.*; TEDDY Study Group. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. *PLoS Med*, 2018; **15**: e1002548.
32. Barratt BJ, Payne F, Lowe CE, *et al.* Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes*, 2004; **53**: 1884–9.
33. Dechairo BM, Zabaneh D, Collins J, *et al.* Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur J Hum Genet*, 2005; **13**: 1223–30.
34. Delong T, Wiles TA, Baker RL, *et al.* Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science*, 2016; **351**: 711–4.
35. Laurberg P, Pedersen KM, Vestergaard H, Sigurdsson G. High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East- Jutland Denmark and Iceland. *J Intern Med*, 1991; **229**: 415–20.
36. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol*, 2013; **79**: 145–51.
37. Yoshiuchi K, Kumano H, Nomura S, *et al.* Stressful life events and smoking were associated with Graves' disease in women, but not in men. *Psychosom Med*, 1998; **60**: 182–5.
38. Fukao A, Takamatsu J, Murakami Y, *et al.* The relationship of psychological factors to the prognosis of hyperthyroidism in antithyroid drug-treated patients with Graves' disease. *Clin Endocrinol (Oxf)*, 2003; **58**: 550–5.
39. Strieder TGA, Prummel MF, Tijssen JGP, Brosschot JF, Wiersinga WM. Stress is not associated with thyroid peroxidase autoantibodies in euthyroid women. *Brain Behav Immun*, 2005; **19**: 203–6.
40. Strieder TGA, Prummel MF, Tijssen JGP, Endert E, Wiersinga WM. Risk factors for and prevalence of thyroid disorders in a cross-sectional study among healthy female relatives of patients with autoimmune thyroid disease. *Clin Endocrinol*, 2003; **59**: 396–401.
41. Stene LC, Rewers M. Immunology in the clinic review series; focus on type 1 diabetes and viruses: the enterovirus link to type 1 diabetes: critical review of human studies. *Clin Exp Immunol*, 2012; **168**: 12–23.
42. Wilkin TJ. The accelerator hypothesis: weight gain as the missing link between type I and type II diabetes. *Diabetologia*, 2001; **44**: 914–22.
43. Rewers M, Ludvigsson J. Environmental risk factors for type 1 diabetes. *Lancet*, 2016; **387**: 2340–8.
44. Alhadj Ali M, Liu YF, Arif S, *et al.* Metabolic and immune effects of immunotherapy with proinsulin peptide in human new-onset type 1 diabetes. *Sci Transl Med*, 2017; **9**: eaaf7779.

1.8

Common Features of Endocrine Tumours

Anne Jouinot, Fidéline Bonnet-Serrano, and Jérôme Bertherat

Introduction

There is a broad variety of endocrine tumours, defined by their localization, differentiation, and hormonal activity. They all share in common by definition to arise from endocrine cells. For this reason, they are mostly observed in endocrine glands. This is evident for tumours primarily located in an endocrine organ, like thyroid, or pituitary tumours. However, because endocrine cells can also be present in small quantities scattered in many non-endocrine organs they can also be observed in many non-endocrine organs. For instance, pulmonary neuroendocrine tumours or paraganglioma (extra-adrenal pheochromocytoma) are well-known examples of neuroendocrine tumours arising outside an endocrine gland. In some very rare instances primary endocrine tumours can be observed in various non-endocrine organs like breast or skin. The World Health Organization (WHO) 2017 classification of endocrine tumours has listed mainly the following entities that will be the focus of this chapter: tumours of the pituitary, the thyroid, the parathyroid, the adrenal cortex, the adrenal medulla, and extra-adrenal paraganglia, the neuroendocrine pancreas, and inherited tumour syndromes [1]. Their frequency is listed in Table 1.8.1.

The endocrine symptoms caused by hormone excessive secretion can be the revealing signs of an endocrine tumour. In this case, hormonal biology guided by clinical signs leads to the diagnosis of hormone excess, and subsequently clinical and imaging investigations lead to the identification of the causative tumour. However, there is a wide range of hormone secretion dysregulation from non-secretory to subclinical and overt-secreting tumours. For this reason, the tumour can also be found incidentally on medical imaging, or due to the tumour mass and finally more rarely distant metastasis. Incidentally discovered endocrine tumour is nowadays a very frequent and classic situation for adrenal tumours, thyroid tumours, and to a lesser extent pituitary and pancreatic neuroendocrine tumours. One of the major steps in the investigations of these 'incidentalomas' is the clinical and biological investigation in search of autonomous hormone secretion. The demonstration of hormone excess is both important for diagnosis as a kind of 'tumour marker', but also for patient management. For instance, a very high prolactin level (typically above 200 ng/ml) will classify a pituitary incidentaloma as a prolactinoma. Evidence of cortisol excess will be the main parameter to decide on surgery for an incidentally discovered benign adrenal adenoma: in case of overt-cortisol excess, surgery will be indicated; in case of modest autonomy of cortisol secretion surgery will be discussed, taking into account the potential consequences and the age of the patient; finally, in case of a non-secreting adenoma with a perfectly normal suppression of cortisol after dexamethasone, there is no indication for surgery [2].

Hormone Dysregulation in Endocrine Tumours

One main characteristic of endocrine tumours is usually the partial or complete loss of the normal control of hormone secretion. For this reason, endocrine tumours can be responsible for hormone excess in keeping in most cases with their cellular origin and differentiation. The main clinical symptoms due to hormone excess and their principal biological work-up are listed in Table 1.8.1.

Various mechanisms can be responsible for this hormone excess. This can be due to the loss of the negative feedback of the peripheral hormone, as for instance adrenocorticotropic hormone (ACTH) oversecretion escaping the negative feedback of glucocorticoids in ACTH-dependant Cushing syndrome. Another level of dysregulation is the relative dedifferentiation of the tumour cells, leading to mixed secretion as, for instance, prolactin, growth hormone (GH), and thyroid-stimulating hormone (TSH) cosecretion by plurihormonal pituitary adenomas. Processing of pro-hormone can also be incomplete in endocrine tumour cells, as for example pro-insulin maturation in insulinoma. This can be explained by the level of expression of the maturing enzyme that differs from the normal

Table 1.8.1 General characteristics of endocrine tumours

	Estimated prevalence in general population	%	Hormone secretion	Main clinical presentation	Main hormonal investigations for diagnosis	Main functional imaging used in routine
Pituitary gland	0.02–0.1%					
Functional pituitary adenoma		70%				
Somatotroph		13%	GH	Acromegaly	IGF1, GH suppression test	
Lactotroph		60%	PRL	Galactorrhoea	PRL	
Thyrotroph		<1%	TSH	Hyperthyroidism	T4L, TSH	
Corticotroph		6%	ACTH	Cushing's syndrome	Midnight cortisol, free urinary cortisol, cortisol suppression test, ACTH	
Plurihormonal and double adenomas		Unknown	PRL + GH (= most frequent)	Acromegaly and/or galactorrhoea	IGF1, GH suppression test, PRL	
Non-functional pituitary adenoma		30%				
Gonadotroph		15–20%	LH, FSH, SUA	Asymptomatic	LH, FSH, SUA	
Null cell adenoma		<5%	None	Asymptomatic	No	
Silent adenomas often aggressive		<5%	None, but possible tumour immunopositivity for ACTH and/or GH and/or or TSH	Headaches, acute visual loss	No	
Pituitary carcinoma		0.1–0.2%				
Thyroid	4–5% (palpable nodules) > 50% (ultrasonographic nodules)					
Benign tumour		90–95%				
Toxic benign adenoma		15% (palpable nodules)	T4L, T3L	Thyrotoxicosis	TSH, T4L	I123 scintigraphy
Non-functional benign adenoma		85% (palpable nodules)	None	Asymptomatic	No	
Malignant tumour	**1–10 /100 000 in men 2.5–20/100 000 in women**	5–10%				
Differentiated thyroid carcinoma		95%	None	Asymptomatic	Follow-up by TG	follow-up by I123 scintigraphy (or I131 in case of concomitant treatment)
Anaplastic thyroid carcinoma	1–9/1 000 000	<2%	None	Asymptomatic/local compression	Follow-up by TG	
Medullary thyroid carcinoma	1/14 300	<5%	TCT	Asymptomatic/flush and diarrhoea	TCT	Fluoro-DOPA-PET
Parathyroid gland	0.85/1000					
Parathyroid adenoma		99%	PTH	Primary hyperparathyroidism (hypercalcaemia symptoms and complications such as nephrolithiasis and osteoporosis)	Calcaemia, PTH, 24 hours calcinuria	MIBI scintigraphy
Parathyroid carcinoma		<1%		More severe primary hyperparathyroidism with possible malignant hypercalcaemia		

Table 1.8.1 *Continued*

	Estimated prevalence in general population	%	Hormone secretion	Main clinical presentation	Main hormonal investigations for diagnosis	Main functional imaging used in routine
Adrenal cortex	5%					
Adrenal adenoma		80%				
Non-secreting		75%	None	Asymptomatic		
Aldosterone secreting (Conn adenomas)		2.50%	Aldosterone	Hypertension and hypokalaemia	Renin and aldosterone (decubitus) +/- suppression test (salt load)	
Cortisol secreting		12%	Cortisol	Cushing syndrome	Midnight cortisol, free urinary cortisol, cortisol suppression test, ACTH	Noriodocholesteror scintigraphy
Adrenocortical carcinoma	**1/1 500 000**	2–25% depending on tumour size	Adrenal steroids (glucocorticoids, androgens, mineralocorticoids) and precursors	From asymptomatic to hyperandrogenia, Cushing syndrome, hypokalaemia hypertension		FDG-PET scan
Adrenal medulla and extra-adrenal paraganglia	1–9 /1 000 000					
Pheochromocytomas	1/500 000	70% (10% malignant)	Metanephrines and Normetanephrines	Menard triad: headaches, palpitations and tachycardia, sweating	Urinary and plasmatic metanephrines and normetanephrines	MIBG scintigraphy
Paragangliomas	1/1 000 000	30% (30% malignant)			Urinary and plasmatic metanephrines and normetanephrines	Fluoro-DOPA-PET
Neuroendocrine pancreas	2–3/100 000 Malignant: 0.8/100 000					
Non-functioning		60–90%		Asymptomatic	Chromogranin A	
Functioning		10–40%			Chromogranin A	Somatostatin receptor imaging
Insulinoma		39% (<10% malignant)	Insulin	Hypoglycaemic symptoms	Insulin and C-peptide during fasting trial or concomitant with spontaneous hypoglycaemia	(somatostatin receptor scintigraphy or PET/CT with 18Ga labelled somatostatin analogues)
Glucagonoma		2% (50–80% malignant)	Glucagon	Rash, glucose intolerance, weight loss	Glucagon	
Gastrinoma		25% (60–90% malignant)	Gastrin	Pain, diarrhoea, oesophageal symptoms	Gastrin	Fluoro-DOPA-PET
VIPoma		7% (40–70% malignant)	VIP (vasoactive intestinal peptide)	Diarrhoea, hypokalaemia, dehydration	VIP	
Somatostatinoma		<1% (>70% malignant)	Somatostatin	Diabetes mellitus, cholelithiasis, diarrhoea	Somatostatin	FDG-PET scan (malignant tumours)
Serotonin-producing tumours with and without carcinoid syndrome		25% (60–88% malignant)	Serotonin	Flush, diarrhoea, abdominal pain, cardiopathy	Serotonin	
ACTH-producing tumour with Cushing syndrome		Rare (>95% malignant)	ACTH	Cushing syndrome	Midnight cortisol, free urinary cortisol, cortisol suppression test, ACTH	

cells, especially in the case of malignant tumours. A classic example of that is the processing of pro-opio-melanocortin (POMC) by aggressive lung cancer causing high secretion of this non-matured ACTH precursor by comparison with pituitary corticotroph adenoma maturing POMC almost like a normal pituitary cell [3]. Activation of cell signalling pathways controlling hormone secretion has been explained by genetic defect of main component of this pathway in some endocrine tumours. For instance, the cAMP

pathway will be activated by somatic mutations of several of its components in different types of benign oversecreting endocrine tumours (i.e. the TSH receptor in toxic thyroid adenoma, the stimulatory G protein (GNAS) in somatotroph adenomas, and the catalytic alpha of the protein kinase A (PRKACA) in cortisol-secreting adrenal adenomas). Interestingly, some germline defects activating the same cAMP pathway, like mutations of the regulatory subunit of PKA (PRKAR1A) in Carney complex, cause various endocrine tumours including GH-secreting adenomas and primary pigmented nodular adrenal hyperplasia leading to cortisol excess (ACTH-independent Cushing's syndrome) [4].

In term of biological investigations, hormone excess will be demonstrated by basal assays and/or dynamic tests depending on the type secreting tumours (see Table 1.8.1). Basal assays can be sufficient for some tumour types and often in many other tumour types in case of a severe hormone excess. These different requirements reflect mostly the physiology of the investigated hormones: a hormone with a quite stable level of secretion during the nychthemeron (e.g. thyroid hormone) will not require dynamic testing. Thyroid tumours usually require only basal assay of TSH, fT4, fT3 and, in case of medullary thyroid cancer, calcitonin. By contrast, investigations of the highly variable corticotroph axis will in most cases require multiple dynamic testing for an accurate diagnosis. It can be also the demonstration of an inadequate level of an hormone by comparison with an increased level of its target that will prove hormone excess: for instance, a still in the normal reference range of the laboratory level of parathyroid hormone (PTH) with a high calcium level will be considered as not supressed and therefore inappropriately elevated. Dynamic tests to demonstrate hormone autonomous secretion required for various tumour types are mostly suppression tests. For instance, GH oversecretion in acromegaly will be demonstrated by a non-suppressed GH level after an oral glucose load, and cortisol excess in case of Cushing syndrome by a non-suppressed cortisol level during a dexamethasone suppression test. These tests are especially needed in case of a modest autonomy of hormone secretion and for this reason are also often important to exclude oversecretion. Multiple basal hormone assays can also help to demonstrate hormone dysregulation in some endocrine tumours; for instance, repeated assays of glucose and insulin during a fasting test in insulinoma or study of the circadian rhythm of cortisol in Cushing's syndrome.

Tumours with a neuroendocrine differentiation will also present abnormal secretion of granins. Granins are a major component of secretory granules. They participate in the maturation and exocytosis of secretory vesicles releasing active neuropeptides and hormones. Among them, chromogranin A (CgA) presents an increased secretion in patients with neuroendocrine tumours of various type and location [5]. CgA is therefore used as a tumour marker in these patients, both for diagnostic and follow-up, especially in case of malignant tumours.

Membrane Receptors in Endocrine Tumours: Implications for Pathophysiology, Imaging, and Treatment

Membrane receptors belonging to the family of seven transmembrane receptors coupled to G protein (RCPG) are present in endocrine cells. Specific ligands binding these receptors play a major role in endocrine gland development, maintenance, and hormone secretion, such as TSH receptor (TSH-R) in the thyroid, or growth hormone-releasing hormone (GHRH) receptor in the pituitary somatotroph tumours. Endocrine tumours in most cases still express these receptors, except in situations of dedifferentiated tumours that are mostly observed in some rare endocrine cancers. Dysregulation of these receptors can play a role in the pathophysiology of endocrine tumours, such as activating mutations of the luteinizing hormone (LH) receptors in Leydig cells tumours or illegitimate receptors expression in bilateral primary macronodular adrenal hyperplasia (PBMAH) [6]. These receptors can also be used for imaging in part of the extension work-up. For instance, many neuroendocrine tumours express the somatostatin receptor. Scintigraphy with radio-labelled somatostatin agonist has been developed for neuroendocrine tumour imaging [7]. Somatostatin receptor scintigraphy is probably the most advanced and used peptide receptor based medical imaging in endocrine tumours. But other receptors can also be targeted for medical imaging, like GLP-1 in insulinomas.

The presence of RCPG in endocrine tumours is also the basis of various therapeutic approaches. For instance, suppressive levothyroxine treatment in differentiated thyroid cancer aims at reducing the stimulation of the TSH-R by endogenous TSH. Somatostatin analogues are used in the treatment of various type of endocrine and neuroendocrine tumours mostly to inhibit hormone secretion, but also to control tumour growth.

Outcome and Prognosis of Endocrine Tumours

The majority of endocrine tumours are benign. These benign tumours often present a very slow progression when they are simply monitored, as it can be done in many thyroid benign tumours or some incidentally discovered non-secreting adrenal adenomas. In this case, symptoms will be mostly related to hormone excess or compression of adjacent organs, the latter being usually observed in large tumours.

Malignant tumours can be observed in all the endocrine glands, but with a different frequency and aggressiveness. For instance, pituitary carcinomas or parathyroid carcinomas are very rare with a reported prevalence of only 0.1% of all pituitary tumours and less than 1% of parathyroid tumours, respectively. By contrast, malignant pheochromocytoma or paraganglioma (PPGL) or thyroid cancer account for 10–20% of all PPGL and 5–10% of all thyroid tumours, respectively. Thyroid cancers are the most frequent endocrine cancer and are estimated to represent about 2% of the total cancer incidence. It is worth to note an increase in the incidence of thyroid cancer worldwide that might be at least in part explained by the increasing use of ultrasonography leading to more frequent diagnosis [8]. This is probably the most documented example of the increased incidence of incidentally discovered endocrine tumour.

Distinguishing malignant from benign endocrine tumours may be difficult even for the pathologist. This is because endocrine cancers are often well differentiated and share common pathologic characteristics with benign endocrine tumours. Pathological expertise is therefore a key for the diagnosis of many endocrine tumours, especially for rare tumours like adrenocortical cancer. For some tumour types like PPGL, malignancy can only be firmly affirmed by the presence of distant metastasis. The malignancy of a

pituitary tumour (i.e. pituitary carcinoma), is defined by cerebrospinal and/or systemic dissemination—the central nervous system, including brain, spinal cord, and meninges, representing the most common metastatic site. In parathyroid carcinomas, reliable indicators of malignancy, including invasion of adjacent tissues or metastatic spread, are rarely observed at the time of diagnosis. Other histological features such as capsular invasion, vascular invasion, thick fibrous bands, and mitotic activity orient towards malignancy but are not specific of cancers and can also be found in parathyroid adenomas. For some cancers pathological scores have been developed to improve the diagnosis of malignancy, including the Weiss score for the adrenal cortex and the PASS (pheochromocytoma of the adrenal gland scaled score) for the adrenal medulla. However, some approaches might be more common to several endocrine tumours. For instance, the Ki-67 index reflecting cell division and mitosis is used in various endocrine tumours for the diagnosis of malignancy and/or for the prognostication of malignant tumours.

Indeed, in endocrine cancers the prognosis varies greatly from one tumour type to another (Table 1.8.2). The prognosis of differentiated thyroid cancer is generally excellent with a 10-year survival rate of 90%, while the prognosis of pituitary or metastatic adrenocortical cancer is generally poor, with a median survival not exceeding 2 years. In most endocrine cancers, tumour stage at diagnosis will impact the prognosis. However, within a single tumour type and for similar tumour stage, prognosis can vary among patients. Furthermore, one specificity of endocrine cancers is that hormone related complications may also impact survival as well as tumour burden itself. Precise prognostication of endocrine cancer is often difficult. This could be explained by some specificity of endocrine tumours but also the rarity of endocrine cancers that is a limitation to design large clinical studies to investigated prognostic markers. The hope in the near future is that molecular markers will improve diagnosis and prognostication of endocrine cancers [9].

Genetics of Familial and Sporadic Tumours

Much progress in the genetics of endocrine tumours has been made since the identification of the proto-oncogene *RET* in multiple endocrine neoplasia type 2 [10]. Endocrine tumours can be observed in many genetic diseases responsible for familial neoplasias. In this case, a germline genetic defect of a tumour suppressor gene or, more rarely, an oncogene-activating mutation is present. Some of these genes predispose to various different types of endocrine tumours, like the proto-oncogene *RET* or the tumour suppressor menin (*MEN1*). Other tumour suppressor genes are apparently responsible for a single type of endocrine tumour like pituitary adenomas in patients with *AIP* mutations or adrenocortical tumours in patients with *ARMC5* mutations. Although inherited endocrine tumours can be diagnosed at various ages, the general rule is that the younger the patient is the more likely a genetic cause could be found. For instance, pituitary adenomas are more frequently of genetic origin before the age of 20-year-old and a genetic cause of primary

Table 1.8.2 Prognosis of endocrine cancers

Tumour type	Overall survival	Factors associated with poor prognosis					
		Age	Sex	Tumour extent	Hormonal secretion	Histological features	Genetic markers
Pituitary carcinoma	50% at 2 years (metastatic tumours)			Systemic metastases	Non-functional tumours		
Differentiated thyroid carcinoma	90% at 10 years	Older age		Large tumour size Local invasion Lymph node involvement Distant metastases	High postoperative thyroglobulin level	Non-papillary subtype Vascular invasion	*BRAF* V600E mutation
Anaplastic thyroid carcinoma	20% at 1 year	Older age	Male	Large tumour size Distant metastases			
Medullary thyroid carcinoma	75% at 5 years	Older age	Male	Local invasion Lymph node involvement Distant metastases	High postoperative thyroglobulin level Calcitonin doubling time < 6 months		*RET* M918T mutation
Parathyroid carcinoma	80% at 5 years	Older age	Male	(Large tumour size) (Lymph node involvement) Distant metastases	High calcium level		
Adrenocortical carcinoma	40% at 5 years	Older age		Local invasion Lymph node involvement Distant metastases	Cortisol-secreting tumours	Ki-67 index > 20% Mitotic count > 5/50 HPF	*BUB1B-PINK1* differential expression
Malignant pheochromocytoma-paraganglioma	50% at 5 years (metastatic tumours)			Large tumour size Visceral metastases	Catecholamines excess	Paraganglioma	*SDHB* mutation
Pancreatic neuroendocrine tumours	50% at 5 years			Large tumour size Lymph node involvement Distant metastases	Non-functional tumours	Glucagonoma/VIPoma Grade 3 (Ki-67>20%) Poorly differentiated tumours Vascular invasion	

HPF, high-power field.

hyperparathyroidism is very rare after 50-year-old. Endocrine cancers occur mostly in adults, even though some paediatric cases are described, mainly as part of tumour predisposition syndrome, such as hereditary PPGL (*SDH* germline mutation), multiple endocrine neoplasia type 1 (*MEN1*) or type 2 (*RET*), or Li Fraumeni syndrome (*TP53*).

Some endocrine tumours like pheochromocytoma and paraganglioma are of genetic origin in more than 30% of the cases. Depending on the causative gene, the outcome of the tumours might differ. It is therefore important in newly diagnosed patients with an endocrine tumour to consider the possibility of a genetic cause and to identify the patients to be referred to the geneticist, not only for familial screening but also for patient management [11].

The development of genomics, allowing us in a single experiment to study at the pangenomic level of genetic and epigenetic alterations, has been very helpful to progress in the molecular genetics of sporadic as well as familial endocrine tumours and the molecular classification of endocrine tumours. The integrated genomics of various endocrine cancers has been possible. It is due to these analyses that the burden of either mutation of driver genes or chromosomal alterations are recognised as important determinants of the classification of endocrine cancers and are associated with patient outcome [12].

Management of Benign Endocrine Tumours

Treatment is required to correct or prevent tumour growth and/or hormone excess. However, some benign non-secreting endocrine tumours considered after appropriate investigations to have no significant clinical consequences and not being at risk of rapid growth will be simply monitored. This is often the case for small incidentalomas of the thyroid, pituitary, adrenal cortex, or endocrine pancreas. In all the other situations, treatment will be required. In many cases, surgery will be indicated as the treatment of choice since complete tumour removal will also treat the cause of hormone excess. Surgery will usually definitely cure a benign endocrine tumour. However, recurrence might occur in cases of tumour remnant, or genetic or environmental factors favouring the development of a second benign tumour. Furthermore, in some endocrine glands (e.g. the pituitary) complete removal might be difficult to warrant because of local invasion to the adjacent structure and recurrence might be observed, sometime years after the initial surgery. For this reason, long-term follow-up is usually needed. Some benign secreting tumours might require a specific medical treatment to control hormone excess in patients that could not be operated upon because of comorbidity, severe consequences of hormone excess, or when complete tumour removal is not possible or expected. For some tumour types, surgery might not be always the first treatment of choice when a specific medical option might offer a definitive cure. This is, for instance, the case in thyroid toxic adenoma treated with radioiodine. Antisecretory treatment will depend on the type of endocrine tumour and its secretion, and will aim at inhibiting directly hormone hypersecretion from the causative tumour or indirectly on its targets. For instance, somatostatin analogues can be used to treat ACTH-secreting tumours, but anticortisolic treatment can also be used to inhibit, at the adrenal cortex level, the cortisol production driven by ACTH excess. Medical treatment of hormone excess can also rely on drugs blocking the peripheral action of hormones at the receptor level: for instance, spironolactone to inhibit the mineralocorticoid receptor in case of primary aldosteronism, or GH antagonists to block the GH receptor in case of GH-secreting pituitary adenomas. Symptomatic treatment of the clinical consequences of hormone excess is also important, like glucose-lowering drugs in case of diabetes secondary to acromegaly or Cushing's syndrome, β-blockers in case of hyperthyroidism, or antihypertensive drugs in case of pheochromocytoma.

Management of Malignant Endocrine Tumours

In endocrine cancers, treatment will aim at controlling tumour growth and metastasis. In secreting cancers, the treatment of hormone excess will often need to be associated with the antitumour treatment. Somatostatin analogues are used in various endocrine cancers to inhibit both hormone secretion and tumour growth [13]. Some more specific drugs like mitotane, used only in adrenocortical cancer, aim at both inhibiting steroid synthesis and tumour growth. Owing to the rarity of endocrine cancers, only few randomized controlled studies have investigated medical treatments for these diseases, and surgery is still the mainstay of treatment of most endocrine malignancies (**Table 1.8.3**). In localized tumours, surgery will be indicated in the majority of endocrine cancers, aiming at complete tumour removal. This often requires experienced surgeons in a multidisciplinary setting to offer the best chance of cure, at least for the rare forms of endocrine cancers. In metastatic diseases, surgery might still have a role in some type of endocrine cancers, when the number of metastatic organs is limited as well as the tumour load, and that the tumour progression is slow. The decision usually requires a multidisciplinary discussion in expert centres to offer the best standard of care. For the same reason, progressive endocrine cancers after initial treatment might benefit from loco-regional approaches, including repeated surgeries, radiofrequency ablation, or radiotherapy. This helps to reduce the tumour burden, prevent loco-regional complications, and symptoms related to tumour mass and can also improve hormonal control for secreting tumours. This is often the case for well-differentiated pancreatic endocrine cancers or malignant PPGL with a slow progression. In case of asymptomatic, low-tumour burden, non-progressive disease, a 'wait and see' strategy could even be discussed.

Some specific radionuclide treatments are given in endocrine cancers because of the capacity of the tumour cells to uptake the administered drug in a quite specific manner (for synthesis of its specific hormone or when the cell expresses significant amount of a membrane receptor). Radioiodine (I–131) therapy is recommended in differentiated thyroid cancers after surgery in patients with high-risk of recurrence as well as in patients with metastatic disease. metaiodobenzylguanidine (MIBG) therapy can be used in malignant PPGL and somatostatin analogues radionuclide therapy is in development in pancreatic neuroendocrine tumours.

Table 1.8.3 International recommendations for antitumour treatment of endocrine cancers

Tumour type	Tumour stage	Surgery	Wait and see	External radiotherapy	Radiopharmaceuticals	Hormonal treatment (for antitumour effect)	Cytotoxic chemotherapy	Targeted therapy
Pituitary carcinoma	Localized	For all tumours		Discussed for high-risk tumours				
	Advanced	For resectable tumours		For progressive tumours despite surgery		Somatostatin analogues or dopamine agonists for functional tumours	Temozolomide	
Differentiated thyroid carcinoma	Localized	For all tumours			Radioiodine for high-risk tumours	TSH suppression		
	Advanced	For all resectable tumours	For iodine-refractory slow-progressive tumours		Radioiodine	TSH suppression		Sorafenib or lenvatinib for iodine-refractory progressive tumours
Anaplastic thyroid carcinoma	Localized	For all tumours		After surgery			Taxanes, anthracyclines, or platinum regimens discussed after surgery	
	Advanced	For resectable tumours		After surgery or for unresectable tumours			Taxanes, anthracyclines, or platinum regimens	ALK or BRAF inhibitors
Medullary thyroid carcinoma	Localized	For all tumours		Discussed for extrathyroidal extension or after incomplete resection	Peptide receptor radionuclide therapy			Vandetanib or cabozantinib
	Advanced	For slow-progressive resectable tumours	For slow-progressive unresectable disease					
Parathyroid carcinoma	Localized	For all tumours						
	Advanced	For resectable tumours						
Adrenocortical carcinoma	Localized	For resectable tumours		After incomplete resection		Mitotane for high-risk tumours		
	Advanced					Mitotane	EDP	
Malignant pheochromocytoma-paraganglioma	Localized	For all tumours		discussed after incomplete resection				
	Advanced	For resectable tumours	For asymptomatic, slow-progressive disease		MIBG		CVD temozolomide	Sunitinib
Pancreatic neuroendocrine tumours	Localized	For all ≥2 cm tumours	For some low-grade, asymptomatic, <2 cm tumours					
	Advanced	For resectable tumours	For some low-grade, asymptomatic, slow-progressive tumours		Peptide receptor radionuclide therapy	Somatostatin analogues for grade 1–2 tumours	Streptozotocin regimen or TEM-CAP for well-differentiated tumours Platinum regimen for poorly differentiated tumours	Sunitinib or everolimus for grade 1–2 tumours

EDP, etoposide-doxorubicin-platinum; CVD, cyclophosphamide-vincristine-dacarbazine; TEM-CAP, temozolomide-capecitabine.

Most endocrine cancers are not good responders to classic systemic cytotoxic chemotherapy usually administrated in many non-endocrine cancers. For this reason, systemic chemotherapy is used in a limited number of aggressive endocrine cancers. For instance, chemotherapies are given in frontline treatment of advanced anaplastic thyroid cancer with taxanes, anthracyclines, and platinum-based regimens. Systemic chemotherapy with streptozotocin is used in progressive poorly differentiated endocrine pancreatic cancers. In metastatic adrenocortical cancer, the combination of mitotane and platinum chemotherapy represents the standard frontline treatment, even though this regimen shows limited efficacy [14]. Temozolomide monotherapy gives encouraging results in pituitary cancer and malignant PPGL [15, 16].

Because of the limited efficacy of systemic chemotherapy in endocrine cancers, efforts have been developed to see whether targeted therapy might offer a good alternative. Ideally, this should be based on molecular evidence of genetic alteration in the signalling pathway targeted by the drug. A good example of this approach is the differentiated thyroid cancer in which the mitogen-activated protein (MAP) kinase pathway is commonly activated, most often due to BRAF-activating mutations. The use of BRAF and mitogen-activated protein kinase kinase (MEK) inhibitors in thyroid cancers has proved to be efficient. Interestingly, in thyroid cancer refractory to radioiodine because they do not uptake radioiodine anymore, the MAP kinase pathway inhibition allows the re-differentiation of tumour cells and restores radioiodine uptake, and therefore, response to I–131 therapy [17]. Other targeted therapies include the multitargeted kinase inhibitors sorafenib and lenvatinib, acting on angiogenesis, which is also a hallmark of differentiated thyroid cancers. Activating mutations of *RET* observed in medullary thyroid cancer also offer therapeutic perspectives. The tyrosine kinase inhibitors vandetanib or cabozantinib, which target both *RET* and vascular endothelial growth factor receptor (VEGFR) kinases, can be used in first-line systemic therapy of medullary thyroid cancer. Finally, some anaplastic thyroid cancer may respond to antiangiogenic drugs and to ALK or BRAF inhibitors[18]. The mTOR inhibitor everolimus, as well as the antiangiogenic drug sunitinib, have proven their efficacy in pancreatic endocrine cancers [19, 20]. Antiangiogenic tyrosine kinase inhibitors seem also promising in malignant PPGL, and a phase III study with sunitinib is ongoing [21].

Conclusion

The broad variety of endocrine tumours clearly illustrates very different situations in terms of tumour growth and hormone excess. They share in common the need for a specialized imaging, biological, and sometimes pathological expertise for an accurate diagnosis. Treatment will depend on the tumour type and its secretory activities, going from simple monitoring to combination of various therapies aiming at controlling both secretion and tumour growth. For tumours requiring an active treatment, the best management will often have to be discussed in a multidisciplinary specialized team.

REFERENCES

1. Lloyd RV, Osamura RY, Klöppel G, Rosai J, eds. WHO Classification of Tumours of Endocrine Organs. *WHO Classification of Tumours*. Vol. 10, 4th edn. Geneva: WHO, 2017.
2. Fassnacht M, Arlt W, Bancos I, *et al.* Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2016; **175**: G1–34.
3. Raffin-Sanson ML, Massias JF, Dumont C, *et al.* High plasma proopiomelanocortin in aggressive adrenocorticotropin-secreting tumors. *J Clin Endocrinol Metab*, 1996; **81**: 4272–7.
4. Bonnet-Serrano F, Bertherat J. Genetics of tumors of the adrenal cortex. *Endocr Relat Cancer*, 2018; **25**: R131–52.
5. Hofland J, Zandee WT, de Herder WW. Role of biomarker tests for diagnosis of neuroendocrine tumours. *Nat Rev Endocrinol*, 2018; **14**: 656–69.
6. St-Jean M, Ghorayeb NE, Bourdeau I, *et al.* Aberrant G-protein coupled hormone receptor in adrenal diseases. *Best Pract Res Clin Endocrinol Metab*, 2018; **32**: 165–87.
7. Reubi JC, Maecke HR. Peptide-based probes for cancer imaging. *J Nucl Med* 2008; **49**: 1735–8.
8. Wiltshire JJ, Drake TM, Uttley L, *et al.* Systematic review of trends in the incidence rates of thyroid cancer. *Thyroid*, 2016; **26**: 1541–52.
9. Jouinot A, Bertherat J. Management of endocrine disease: adrenocortical carcinoma: differentiating the good from the poor prognosis tumors. *Eur J Endocrinol*, 2018; **178**: R215–R230.
10. Mulligan LM. 65 years of the double helix: exploiting insights on the RET receptor for personalized cancer medicine. *Endocr Relat Cancer*, 2018; **25**: T189–200.
11. Buffet A, Burnichon N, Amar L, *et al.* Pheochromocytoma: when to search a germline defect? *Presse Medicale*, 2018; **47**: e109–18.
12. Giordano TJ. 65 years of the double helix: classification of endocrine tumors in the age of integrated genomics. *Endocr Relat Cancer*, 2018; **25**: T171–87.
13. Cives M, Strosberg J. The expanding role of somatostatin analogs in gastroenteropancreatic and lung neuroendocrine tumors. *Drugs*, 2015; **75**: 847–58.
14. Fassnacht M, Dekkers O, Else T, *et al.* European Society of Endocrinology clinical practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2018; pii: EJE-18-0608.
15. Raverot G, Burman P, McCormack A, *et al.* European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumours and carcinomas. *Eur J Endocrinol*, 2018; **178**: G1–24.
16. Hadoux J, Favier J, Scoazec J-Y, *et al.* SDHB mutations are associated with response to temozolomide in patients with metastatic pheochromocytoma or paraganglioma. *Int J Cancer*, 2014; **135**: 2711–20.
17. Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2016; **26**: 1–133.

18. Haddad RI, Lydiatt WM, Ball DW, *et al.* Anaplastic thyroid carcinoma, version 2.2015. *J Natl Compr Cancer Netw*, 2015; **13**: 1140–50.
19. Falconi M, Eriksson B, Kaltsas G, *et al.* ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
20. Uri I, Grozinsky-Glasberg S. Current treatment strategies for patients with advanced gastroenteropancreatic neuroendocrine tumors (GEP-NETs). *Clin Diabetes Endocrinol*, 2018; **4**: 16.
21. Berruti A, Baudin E, Gelderblom H, *et al.* Adrenal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 2012; **23 Suppl 7**: vii131–8.

1.9

Genetic Aspects of Endocrine Disease

Trevor Cole

Historical Introduction

The recognition of a hereditary component in plant and animal phenotypes and human disease dates back many centuries but the start of genomic medicine may best be linked to the description of DNA's double helical structure reported by Watson and Crick in 1953 [1] (**Figure 1.9.1**). It is perhaps surprising that this actually predates other discoveries, now taken for granted, such as recognition of the correct number of human chromosomes (46 not 48), reported in 1956 by Tijo and Levan [2] and 3 years later the first papers reporting numerical chromosomal abnormalities in humans, accounting for three widely recognized clinical disorders: Down, Turner, and Klinefelter syndromes [3–5].

Karyotype analysis should be considered a genomic test in that it looks at the whole genetic architecture and the potential interrelationships therein, to identify their combined influence rather than the study of a focused single gene test which scrutinizes a specific genes sequence and function (adapted from the World Health Organization (WHO)'s definition) [6]. It took a series of significant technological advances over the next three decades to bring significant progress in sequencing power and enable the delivery of more widespread genetic and genomic testing. It is important for individuals wanting to incorporate these technologies into their clinical practice understand the differences between these two types of testing, which tests they should request and the differences in the information generated (**Table 1.9.1**).

The investigation of chromosomal disorders became more sensitive during the 1970s, firstly utilizing chromosome banding technologies and then in the 1980s and 1990s utilizing fluorescent labelled probes that could detect as subtle deletions, duplications, and translocations [7] (**Figure 1.9.1**). These remain important in detecting significant copy number variants (CNVs) in endocrinology, such as the 22q11 deletion resulting in hypoparathyroidism in DiGeorge syndrome (**Figure 1.9.2**) but will likely become redundant as microarray, and then whole genome analysis, becomes the universal platform.

In 1977 Fred Sanger described a technology, now known as Sanger sequencing, which utilizes radioactive or fluorescent tagging of chain-terminating dideoxynucleotides which are incorporated into DNA during its replication [8], and with further development of polymerase chain reaction (PCR) by Kary Mullis in 1983 [9] meant sequencing individual genes became more practical. The processes were, however, slow and expensive compared to currently available technology, as only a single sequence could be read on each analysis. Despite these limitations, in 1986 Royer-Pokora reported the first successful identification of a gene by positional cloning, a process by which the chromosomal location of a disease gene is continually refined by a series of experiments until a small enough number of candidates is reached within that region to enable individual sequencing of these genes. Three papers were published in 1986 and 1987 reporting the identification of the genes for chronic granulomatous disease, Duchenne muscular dystrophy and retinoblastoma [10–12].

Over the next two decades the rate and number of genes identified by these technologies gradually increased but there was a parallel technological development occurring in the non-clinical arena which would revolutionize human genetic discovery and service delivery. Massively parallel signature sequencing (MPSS) was being developed by a series of technological advances during the 1990s which subsequently resulted in the publication of Brenner *et al.* in 2000 [13]. Although it was several years before this became commercially available, the change in the diagnostic power are readily be illustrated by gene discovery before and after the introduction of this technology which we now know as next generation sequencing (NGS). Whereas Sanger sequencing analysed just one region of genetic coding in a single experiment, NGS could analyse many thousands of sequences simultaneously. There are currently several different NGS technologies available and readers wanting to understand more about the specific platforms are referred to the paper of Reuter *et al.* [14].

In 1989 Riordan *et al.* [15] reported the cloning and sequencing of the cystic fibrosis gene utilizing a Sanger-based approach. This was the culmination of research on thousands of families, by over 100 groups worldwide, and costing many millions of pounds. In

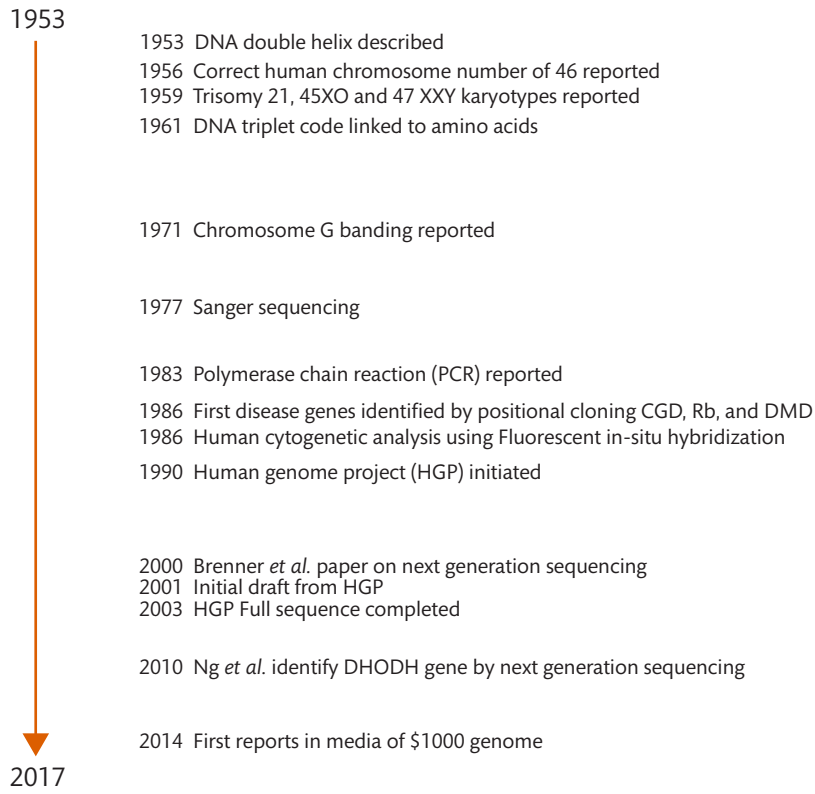


Figure 1.9.1 Timeline of genetic milestones.

2010, Ng *et al.* reported the identification of the Miller syndrome gene (DHODH), which was the work of a single collaborative group investigating four individuals in three families using NGS exome analysis, and was the first report of success using this approach [16]. The confirmed the suspected potential and many initiatives, including large scale national programmes were instigated.

In April 2011 the first patients were recruited into two national genomic projects in the United Kingdom [17] and Canada [18], which while not unique, they are illustrative of the rapid change occurring worldwide [19]. In the United Kingdom the Deciphering Developmental Disorders, was a national collaboration led by clinical geneticists and scientists from Sanger Institute in Cambridge [17] and in Canada the FORGE project (Finding of Rare Disease Genes in Canada) was led by genomic centres in Ontario, British Columbia and Quebec [18]. Within 18 months the FORGE project had discovered 41 novel genes not previously linked to human disease [18] and the DDD study has currently published over 100 papers [17]. Through projects such as these it rapidly became apparent the new challenge would not be sequencing our genome or identifying new genes but interpreting the 'avalanche' of data generated. How could this be delivered equitably, accurately and in a timely fashion across healthcare systems where it is estimated that 1 in 17 have a rare disease (a disease less frequently than 1 in 2000), of which the majority will have a genetic basis [20]?

Much of this work was underpinned by a multinational project planned to run from 1990 and 2005 to sequence the whole genome (The Human Genome Project) which published an initial draft in February 2001 and the full sequence in April 2003 [21]. This resulted in a re-evaluation of the number of human genes from the previously estimated number of over 50 000 to just over 22 000 but the function of many of these genes remains unclear. The cost of the project to sequence the initial genome of three billion base pairs was $3billion whereas now a whole human genome sequence can be generated for less than a $1000, although this does not necessarily include the larger cost of additional interpretation that is typically required (Figure 1.9.3).

The societal and political impact had also not gone unnoticed, for example, in 2009 the European Commission produced a directive to its 28 member states requiring them to develop and implement a Rare Disease Strategy for their citizens [22]. Without doubt this added impetus to numerous initiatives such as the 100 000 Genome Project which was launched in the United Kingdom in 2014 [20, 23] and a raft of clinical and laboratory initiatives across healthcare services worldwide. The challenge remains however how to deliver these aims across a whole population health system ensuring engagement of all health professionals not just those working within the specialism of clinical genetics [20, 24].

Delivery of Genetic Testing

There are significant challenges to the introduction of genetic investigation into the wider clinical setting. To date the majority of genetic testing has been organized through specialist clinical genetics services, only feasible as the number of individuals in whom testing was possible remained small. Genetic practitioners would evaluate the clinical history, assess the likelihood of an identifiable genetic cause being present based on the knowledge at that time and identify the most appropriate test. They would consent the appropriate member of the family for the initial test, typically an affected individual and not necessarily the initial consultant. The clinician would ensure individuals were aware of the medical and non-medical

Table 1.9.1 Examples of likely test selection dependent on presentation or disorder

Genetic mechanism	Disorder	Test type to order
Likely sex chromosome aneuploidy	Turner syndrome Klinefelter syndrome	Routine karyotype
Likely predictable copy number variant (CNV)	Disorder of sex development (DSD) and Wilm's tumour Hypoparathyroidism associated with DiGeorge syndrome	FISH for 11p13 deletion FISH for 22q11.2 deletion
Possible CNV—but not predictable	Developmental delay, short stature, and congenital anomalies	Microarray
Single gene disorder with known association with single gene [1]	MEN type 1 MEN2/MTC CAH due to 21 hydroxylase Deficiency Classical osteogenesis imperfecta (OI)[1] Achondroplasia short stature	Sanger sequence Menin Sanger sequence Ret oncogene Sanger sequence Cyp21A2 Sanger sequence Col1A1+ Col1A2 Sanger sequence FGFR3
Disorder that may be genetic but 'limited' heterogeneous aetiology—often less than 20 genes associated	Isolated hyperparathyroidism Isolated phaeochromocytoma DSD with undefined steroid pathway abnormality Atypical OI or bone fragility	NGS with interrogation of defined *HPT* gene panel NGS with interrogation of defined phaeochromocytoma gene panel NGS with interrogation of defined DSD gene panel NGS with interrogation of defined bone fragility gene panel
Disorder which may be genetic but extremely heterogeneous—often more than 100 genes	Syndromic short stature and developmental delay Syndromic obesity and/or overgrowth Undefined skeletal dysplaia[3]	[2]NGS with large panel of genes interrogated or alternatively Clinical exome or whole genome sequencing

[1] Some laboratories may do the sequencing on an NGS platform but only interpret specific relevant genes.
[2] Decision on whether NGS with extended panel analysis is by a clinical exome or whole genome may depend on laboratory informatics platforms.
[3] Decision on which test to select will be influenced by requesting clinicians diagnostic expertise, for example when investigating short stature due to a skeletal dysplasia the clinician might be expected to recognize and request specific test for achondroplasia rather than a skeletal dysplasia panel or clinical exome. For more complex clinical disorders, a local integrated genomic network (Figure 1.9.4) could help support and direct the most appropriate clinical test.

implications of the result to both affected family members, and those unaffected relatives who despite being asymptomatic may be at risk of developing the disorder in the future. This consent also needs to consider any relevance yet to be born individuals.

Such genetic counselling is both time-consuming and often outside the training and experience of non-genetics clinicians. The increased requirement for genetic or genomic testing, however, means that in future most diagnostic and precision management testing will have to be delivered by non-clinical geneticists. It is also pertinent that as more disorders and clinical presentations are recognized as having a monogenic aetiology or significant genetic component, the expert insight from non-clinical genetic specialists into the understanding of the natural history and management discussions, to support of clinical genetic colleagues will be equally vital if services are to be effective. The model of medical care delivery in the genomic medicine arena will therefore need to address the limited number of trained clinical and laboratory genetic specialists, necessary to deliver a service where up to 1 in 17 of the population may have a genetic disorder. Timely understanding of the genetic aetiology may have a direct impact on the optimal management, for example in some endocrine tumour or diabetic disorders. Reliance on clinical geneticists alone will delay the delivery of potential patient benefits from such expanding genomic knowledge.

Bowdin *et al.*, in their paper 'Recommendation for the integration of genomics in clinical practice' [25] provide some important and clear principles that would enable a model to be developed which is less heavily dependent on clinical genetic resources. They identify that the requesting clinician should be competent in the following broad areas: (i) obtaining an appropriate family history and genetic evaluation; (ii) ability to determine the most appropriate test; (iii) provide adequate information to obtain informed consent and iv the ability to interpret the genomic results and provide adequate post-test counselling, or identify a suitable source of such advice (Table 1.9.2).

In Tables 1.9.3 and 1.9.4 we provide some examples of the principles which may underpin the choice of test with some examples to illustrate these. The test and methodology of testing (Table 1.9.1) will be influenced not just by the evidence on detection rate and resources, but also the clinical validity and utility of the test, accessible funding, and potentially guidance within the healthcare system. Bowdin *et al.* [25] illustrate this by referring to the differences in the decision making between Canada and the United States, and even within insurance providers in the United States [25]. In other publicly funded services there may be a nationally agreed set of guidelines as documented in the United Kingdom on the UK Genetic Testing Network website [26], soon to be transferred to a national directory. Even with guidance on patient and test selection, nominating who should undertake testing, consent, and provide support mechanisms to help interpret and return results and develop patient pathways will remain a challenge. It is likely that integrated care networks or multidisciplinary team working with a 'local decision-making pathway' will improve the patient access and service delivery by identifying on-going professional education opportunities. It can also optimize data collection and provide clinical support to colleagues with hard to interpret or unexpected results (Figure 1.9.4).

Barriers and Solutions to Integrating Genomic Technologies into a Wider Healthcare Service

Evaluating the Cost

The cost is decreasingly seen as a barrier to genomic testing as the sequencing costs fall to levels at or below many other routinely ordered clinical tests. The milestone of whole genome sequencing (WGS) for under $1000 has been passed and some commercial ventures are advertising sequencing for half this figure or less which is similar to the

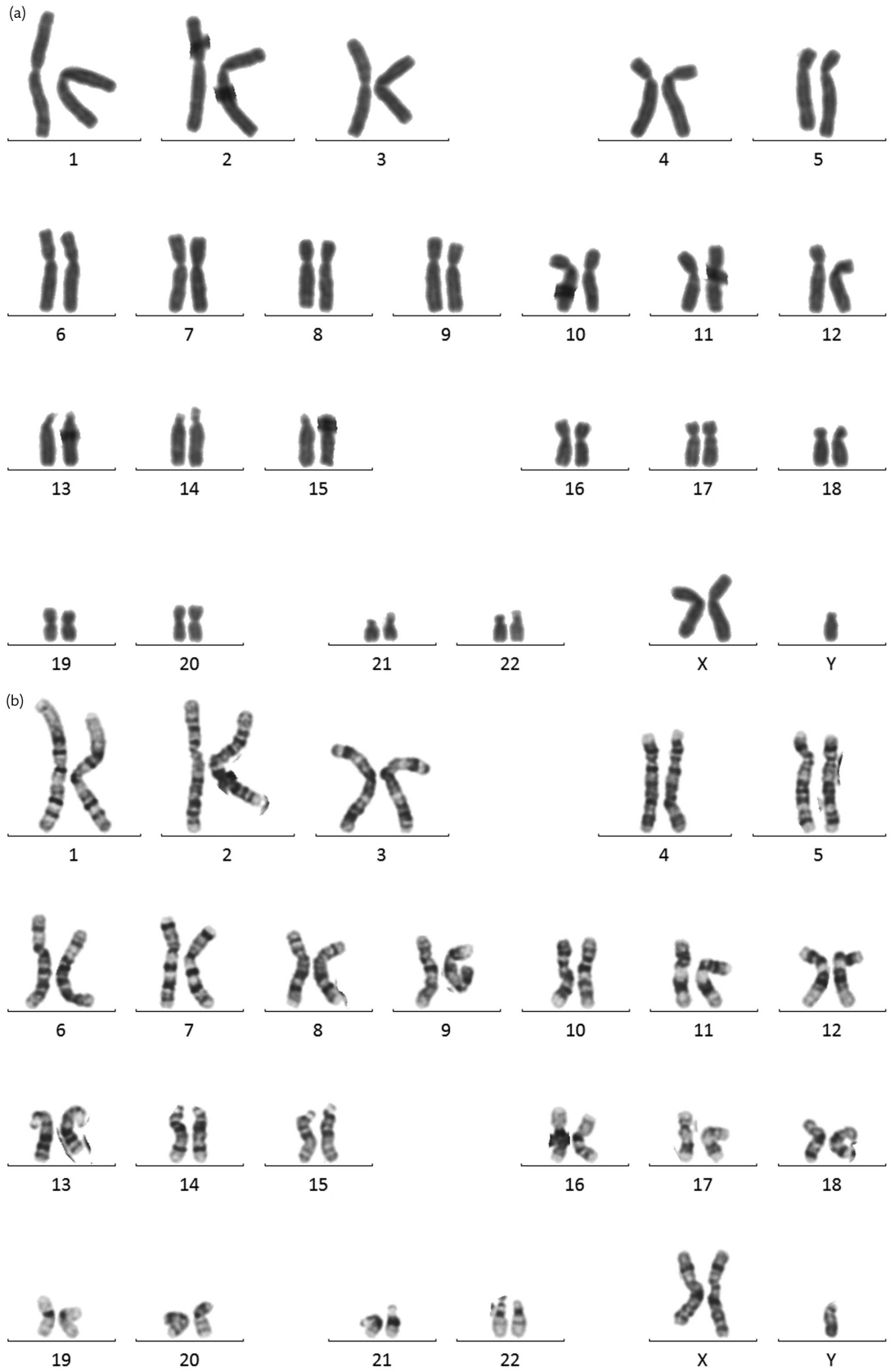


Figure 1.9.2 (a) Solid stain karyotype—47XXY Klinefelter syndrome; (b) G banded karyotyoe—47XXY; (c) Fluorescent in situ hybridization (FISH) —Di George syndrome del(22)(q11.2.q11.2). Inverted DAPI stain: the TUPLE1 probe is in red and for this hybridization, the 22q13.3 probe acting as a control in green. The FISH probe—Cytocell DiGeorge/VCFS TUPEL1 and 22q13.3 deletion syndrome.

Courtesy of Graham Fews.

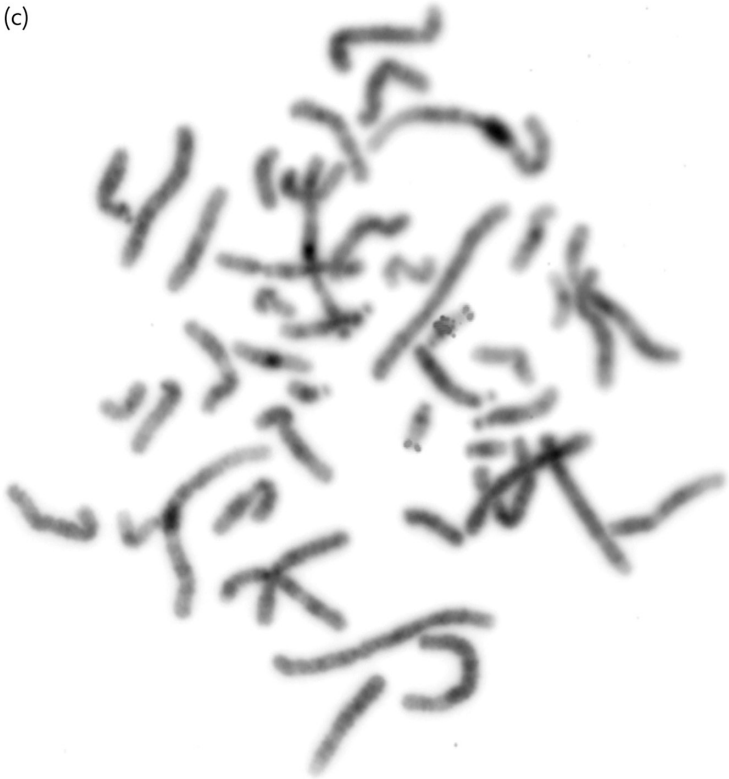


Figure 1.9.2 *Continued*

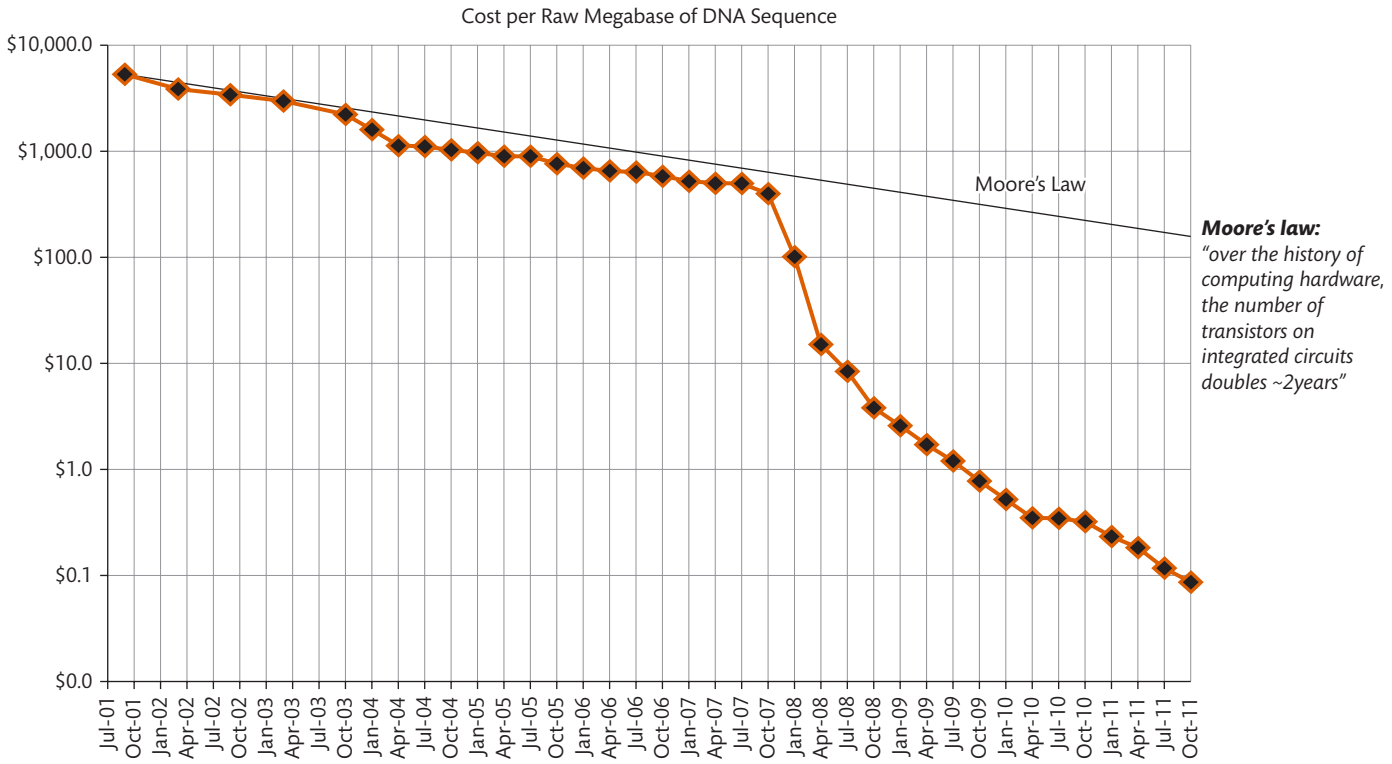


Figure 1.9.3 Comparison of falling cost of generating sequence data compared to Moore's law.

Reproduced with permission from Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Courtesy: National Human Genome Research Institute. Available at: https://www.genome.gov/sequencingcostsdata. Accessed November 15, 2018.

Table 1.9.2 Competencies recommended for clinicians ordering genomic tests as recommended by clinical genetics think tank (CGTT); Bowdin *et al.* (2016) and modified by Cole (2018)

Broad theme	Specific components
Clinical genetics evaluation	Take a family history Genetics focused history and exam Evaluate prior genetic tests
Determine which tests appropriate	Non-genomic—including metabolic Genomic—to include evaluation of specific genetic or genomic test
Pretest counselling and consent	Information on spectrum of disorder and natural history Possible test outcomes including primary and secondary coincidental findings Possible variants of unknown significance Make patient aware there may be non-medical implication such as insurance Obtain appropriately informed consent with a discussion that the information may be very relevant to other family members and ideally obtain permission to share this information in pretest consent
Post-test counselling and support	Personal ability to accurately interpret results and address patients' questions and concerns or alternatively have an identified route to provide such support with identified pathway in a network or with specialist genetic services

It is unrealistic to expect non-genetic specialist to be able to manage all these issues in all cases, but an identified mechanism to address them should be in place and the patient should be made aware that in some circumstances additional input may be required.

cost of MRI scans or endoscopies. These headline figures may not always clarify the extent of genes covered, the depth of coverage, or ability to detect CNVs, all factors which will significantly impact on the analytical validity and clinical utility of the test. Of at least equal importance when considering the cost to any health system, is the potential need for additional bioinformatic and clinical assessment and an increasing recognition that the clinical interpretation of many identified sequence variants still remains beyond our current knowledge. Therefore, a sound clinical focus on the testing that is required, at the current time remains a key component rather than relying solely on WGS as a 'scattergun' approach in all settings.

Despite these cautions the potential to personalize treatment can have a clear benefit as illustrated by a simple change of practice in the West Midlands region in the United Kingdom (population 5.5 million) (author's personal experience). In an 18-month period in 2006–2007, all apparently sporadic medullary thyroid cancer (MTC) cases were

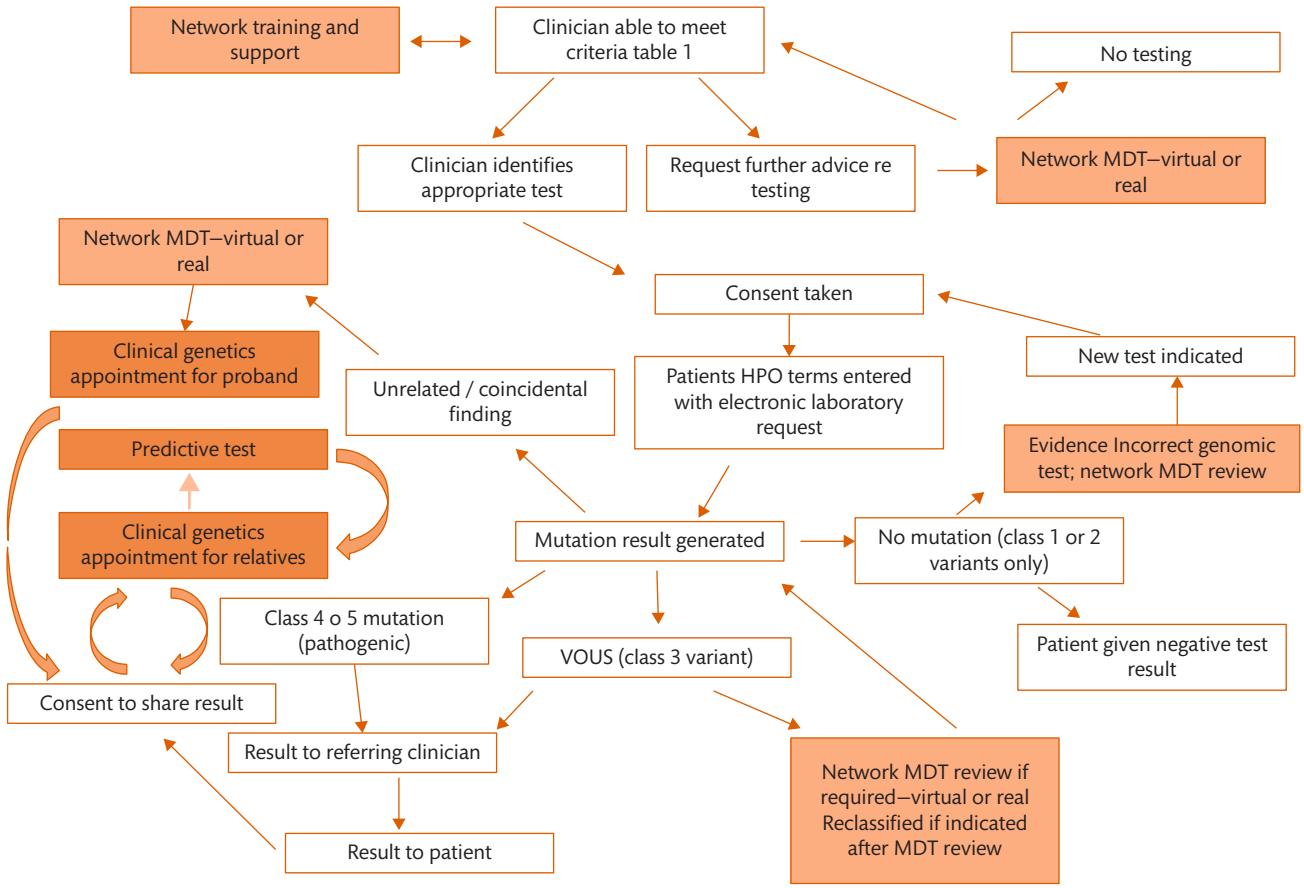


Figure 1.9.4 Example of integrated care pathway for genomic testing in mainstream healthcare. MDT active input in light orange filled boxes and from clinical genetics in dark orange filled boxes.

tested for the commonly recognized ret-oncogene mutations. This revealed pathogenic mutations in 3 of the 19 apparently sporadic MTC cases and cascade testing in the families, and uncovered 13 asymptomatic gene carriers of whom 11 underwent total thyroidectomies during this period. Pathological review identified 5 occult MTCs, 5 cases with C cell hyperplasia, and 1 normal result. Importantly, management was also modified to include parathyroid and adrenal surveillance. This illustrates the optimization of the test strategy, from confirming a previously unrecognized diagnosis (MEN2), delivering 'personalized management' and implementing effective and comprehensive cascade testing within the families (Table 1.9.3 and Box 1.9.1). While the initial cost to the health service from such systematic genetic testing is increased by subsequent consultant surveillance and surgery in asymptomatic patients, there are longer-term clinical and cost benefits, potentially many decades into the future, and typically in specialisms such as oncology and palliative care rather than clinical genetics and endocrinology where the management change was instigated and cost was incurred. All clinicians requesting genetic testing should strive to maintain the maximum clinical utility by utilizing all the potential information from informative results (Table 1.9.3 and Box 1.9.1).

Consent and Confidentiality

When requesting genetic testing, clinician instigating such testing should explain the potential outcomes and provides adequate information for the consultant, so that they, or their legal guardian in the case of children, can give informed consent. It is however important that we do not burden clinicians with a concept of 'genetic exceptionalism' such that they avoid genetic testing or refer all such cases to clinical genetics. Many blood, radiological, or pathological tests ordered by clinicians have unpredicted significances and impacts. The appropriate management of these possible outcomes, before and after results, are part of good holistic clinical care. Furthermore, a genetic diagnostic test may only be confirming a clinical diagnosis but can remove uncertainty and make possible choices for other family members. There are many instances of a diagnostic test only providing proof of the underlying condition, for example, patients presenting with an autosomal dominant (AD) family history and personal neurological diagnosis of Huntington's disease or an AD history of MTC indicative of MEN2. In these settings the genetic test does not reveal any 'unpredicted' information in the consultant, but can provide potential benefits to other family members. These tests therefore should be available to healthcare practitioners outside clinical genetics as long as they are able meet the requirements identified in Table 1.9.2, either personally or in collaboration with colleagues as part of a clinical pathway (Figure 1.9.3).

Asymptomatic family members requesting predictive testing will often have more complex issues to evaluate including the medical, psychological, and financial consequences the information may have for them and their families. In some instances, it may also have occupational or reproductive impacts and as such the counselling, consent, and testing may be better directed to clinical genetics services.

When obtaining consent for a genetic or genomic test the clinician should make the consultant aware that the result could well have an impact on other family members and where there is a clinical indication should, where possible, obtain the permission to share this information with relatives who may have the 'same genetic risk',

Table 1.9.3 Examples illustrating different impact of genetic results

Role of genetic testing	Example of disorder	Gene/Genes examined	Impact of identified mutation
Confirm diagnosis			
Example 1	Bone fractures	Col1A1 and Col1A2	Avoidance of incorrect diagnosis of non-accidental injury
Example 2	Disorder of sex development (DSD) Developmental delay (DD) Wilm's tumour (WT) and obesity	WT1 and contiguous gene deletion including BDNF	Confirm single unifying disorder
Precision management			
Example 1	Isolated Phaeochromocytoma	*SDH* and related genes, Ret-proto-oncogene, Von Hippel–Lindau (VHL), NF1	Surveillance for contralateral phaeochromocytoma and related tumour management, e.g. Thyroidectomy if Ret mutation +ve Renal surveillance if SDH or VHL mutation +ve Retinal and CNS management in VHL +ve
Example 2	Onset of diabetes in early adulthood	MODY panel of genes Including HNF1A	Trial of sulphonylurea rather than insulin
Example 3	DSD and DD	WT1 deletion	Surveillance for WT and renal function
Example 4	Isolated hypercalcaemia	CaSR	Avoidance of unnecessary investigation and inappropriate parathyroidectomy
Cascade testing and management			
Example 1	Isolated medullary thyroid cancer	Ret proto-oncogene	Cascade testing prophylactic thyroidectomy Parathyroid and adrenal surveillance
Example 2	For all scenarios above	First degree relative testing and cascade testing beyond if positive result in each tested relative	Reassure if de novo or individual tested is not a gene carrier. Implement appropriate management in proven gene carriers

Box 1.9.1 Three components to optimize clinical utility of genetic testing (somatic and germline findings)

Confirm precise diagnosis

Implement appropriate precision management (pharmacological or surgical intervention, surveillance, advice on the natural history or prognosis and genetic advice)

Implement cascade testing and management as appropriate to gene and genetic result in extended family. This includes scenarios of somatic genetic variant subsequently identified in the germline

and with anonymized databases to facilitate new interpretive knowledge. This is good practice and also provides a clear illustration of the potential impact of the possible outcomes to the consultant. These issues are often easier to discuss in the preresult setting rather than the post-test setting when the focus should be the consultant's own management. Patients may however refuse to share this information with family members. In our traditional medical model this right to confidentiality would be considered paramount but this absolute right has been questioned [26]. Non-disclosure could even result in one family member having information about a relative, of which the is relative themselves unaware. Examples might include obligatory female carriers of x-linked adrenoleukodystrophy, or a distant relative who recognizes the existence of an endocrine tumour syndrome in a second- or third-degree relative with non-penetrance in the intervening relatives, scenarios that could have profound medical implications.

Parker and Luccassen in 2018 [26] suggested that it may be inappropriate to see the information as belonging to the individual but rather belongs to the family, many of whom will share the same genetic information. The debate is illustrated by the range of published opinions, including the *British Medical Journal* (*BMJ*)'s 'head to head' in 2007 [27, 28] and nuanced differences between USA and Canadian patient surveys [29, 30]. The potential duty to inform family members, even after a patient has refused, is highlighted in the current UK high court case ABC versus St Georges Healthcare trust, in a family with Huntington's disease. This case and the broader issues are discussed by Mitchell *et al.* [31].

Thankfully such cases to date have been rare, but do illustrate that within the field of genomic medicine, some of our concepts around consent and confidentiality may be challenged. In many countries, professional guidance is currently being updated but in the meantime the clinician utilizing genomic investigations should be aware to the issues and the broader family considerations.

Our understanding of informed consent is also being challenged by what information can realistically be provided to the patient wanting to pursue WGS when offered as a diagnostic test. Coincidental findings, which are pathogenic mutations in genes unrelated to the clinical presentation, will become increasingly common on whole genome reports. The clinician cannot provide detailed information for consent on over 20 000 specific genes, especially as the function of many are not yet known and certainly many mutation specific phenotypes have yet to be identified. Future knowledge or family information may become available at a later date, such that variants that are difficult or impossible to assess at the current time may become readily interpretable. This should be included in the discussion with the patient.

Genomic investigation of broad phenotypic categories such as short stature or developmental delay, typical require the examination of hundreds or thousands of genes respectively. Within these huge panels there may be genes with unexpected consequences for the consultant, such as progressive degenerative disorders or tumour syndromes and even conditions which may be relevant to apparently unaffected parents. An example of the latter would be parents at risk of tumours due to carrying heterozygous BRCA2 mutations in a child with short stature and Fanconi's syndrome due to two pathogenic *BRCA2* mutations.

Microarrays also examine the whole genome and may identify CNVs in child with learning difficulties in an 'non-focused manner'. The coincidental finding of a 1 q deletion identified in a child with mild learning difficulties is illustrative. The deletion includes the gene *CDC73* which results in hyperparathyroid jaw tumour syndrome and is associated with a significant risk of parathyroid carcinoma (Figure 1.9.5). Without any previous history such a finding was completely unpredictable. This demonstrates the impossibility of pretesting consent covering all potential outcomes but the importance of making individuals aware of the possibility of unexpected or coincidental findings. There should be adequate post-test holistic service provision to address such scenarios. This is in lines with the principles laid out by Bowdin *et al.* [24].

Interpretation of Findings

Variants are Frequently Found but Commonly Not Significant

All humans have a unique pattern of genetic variation in both our coding and apparently non-coding DNA, for example in the paper on Ng *et al.* [16], each affected individual had over 1000 non-synonymous variants, a figure consistent with other large scale genomic sequencing projects [17]. Individuals not uncommonly have potentially pathogenic genetic changes in more than one gene which might account for the presenting phenotype [16, 17]. While it would be possible this represents a 'hybrid phenotype' or digenic inheritance (interaction of two non-allelic genes) a more likely explanation is that at least one variant is a coincidental finding. To aid accurate and consistent interpretation laboratory and clinicians need to draw on a number of tools as summarized in Table 1.9.4.

Epigenetics

Not all significant genetic variation will be detected by technologies such as microarrays, NGS, and methodologies to determine intragenic deletions or duplications. There are instances where the germline or somatic sequence remains unchanged but the gene expression is under epigenetic regulation. This typically results from modification by additions such as methyl side chains to the DNA or methyl or acetyl groups to the histones. These epigenetic changes maybe heritable or secondary to non-genetic influences and can alter many endocrine related genes [32] and are relative common mechanism in rare disorders of growth, hormone action and diabetes [33] (Table 1.9.5). Recognition of these clinical patterns may direct the clinician to alternative molecular genetic tests, such as looking at gene expression or methylation, to confirm the genetic aetiology.

Figure 1.9.5 Microarray result showing 1q deletion including CDC73. Courtesy of Graham Fews.

Coincidental Findings

The broader and less specific a genetic test the greater the potential for uncovering additional or unpredicted findings (Table 1.9.1). A single gene test for *MEN1* or *MEN2*, for example, if informative will only reveal information about those disorders. A panel test for isolated phaeochromocytoma may confirm a diagnosis of neurofibromatosis type 1, MEN2, or Von Hippel Lindau. As this later gene panel is relatively narrow, typically around 10–15 genes, these possibilities can be discussed in the consenting process. In comparison however when utilizing a gene panel for short stature or intellectual deficit, typically including between 100 and over 2000 genes, or a whole genome analysis counselling and consent cannot cover individual gene outcomes. The patient interaction should include an explanation of the possibility of either completely unexpected phenotypes, or in the case of whole genome analysis, the discovery of completely unrelated disorders. The latter are known as coincidental findings.

The clinician patient discussion should include not only the acknowledgement of such possibilities and their consequences, but also the potential strategy to deal with these. One possibility frequently proposed strategy to deal with such scenarios is an agreement not to look for additional findings or a patient request not to be informed of such coincidental findings, and informed only those findings related to the condition being investigated. Determining what might be relevant to the clinical presentation can

Table 1.9.4 Examples of tools to assist interpretation

Methodology	Example	Reference or link
Variant Interpretation and scoring	American College of Medical Genetics Consensus Guidance	[34]
Software for variant interpretation	Alamut SIFT Polyphen	https://www.interactive-biosoftware.com/ http://sift.jcvi.org/index.php http://genetics.bwh.harvard.edu/pph2/
Databases of normal and pathological variants	gnomAD ClinVar Decipher LOVD	http://gnomad.broadinstitute.org/ https://www.ncbi.nlm.nih.gov/clinvar/ https://decipher.sanger.ac.uk/ http://www.lovd.nl/3.0/home
Functional studies	*In vitro* expression RNA studies	

Table 1.9.5 Examples of imprinting disorders—may occur without sequence variant [30]

	Specific disorder	Underlying mechanism
Growth Disorders	Silver Russel l syndrome growth failure	Paternal imprinting centre 1 hypomethylation results in biallelic expression of H19 and silencing both copies of IGF2
	Beckwith Wiedemann syndrome (overgrowth)	Maternal imprinting centre 2 loss of methylation or gain of imprinting centre 1 methylation
Diabetic disorder	Transient neonatal diabetes mellitus	PLAG1 methylation defect
Calcium disorder	Pseudohypoparathyroidism 1b	GNAS abnormal maternal methylation

however be challenging. Furthermore, it leaves the managing clinician in the difficult situation of potentially having information on file, relevant to the patient's longer-term health, which has not been revealed. An added challenge to this solution is that coincidental findings may have significant medical and life-limiting implications to other family members. Failure to reveal significant genetic information to relatives has been the subject of past and present debates and legal cases worldwide [26, 27, 29–31]. This may potentially make such agreements less practical in the future.

An alternative approach is to ensure genetic testing is instigated only by clinicians able to explain these issues to the patient and working as part of a wider team with access to additional clinical and psychology support. Such additional expertise may only rarely be called upon but developing links to such a multidisciplinary team (MDT) will ensure rapid and appropriate access to clinical genetics, other medical-based specialties, and psychology services. The establishment of such MDTs or integrated care solutions will also support the discussion of hard to interpret genetic results and fulfil one of the suggested requirements from Bowdin *et al.* [25] (Table 1.9.2, Figure 1.9.3).

Conclusions

The increasing understanding of the genetic basis of much of human disease and falling cost of obtaining such information means genomics should be integrated in patient care. This will include both germline and somatic tissue testing which has the potential to make or refine a diagnosis, identify specific patient management or surveillance (personalized medicine) and alert members of the extended family to treatment or surveillance opportunities which can reduce their health own morbidity and risk of avoidable early death. The number of patients and wide spectrum of disorders where this is applicable means that delivery will be through a multitude of clinical specialities and settings. Many such services are not yet adequately developed nor are individual clinicians adequately trained or supported. The development of genomic MDT and integrated care pathways could address many of the delivery and training challenges and facilitate support and oversight by laboratory and clinical genetics with a process of gradually and iterative learning. It is vital that clinical services delivering both diagnostic and predictive genetic and genomic testing are aware of the complexity of issues highlighted earlier in this chapter and have in place solutions to address these.

Clinicians should not be swayed by unrealistic hype or blinkered to the remaining challenges and unknowns that lie in the field of genomics nor should they be deflected by fear or ignorance of the technologies available or uninformed statements that there are no proven benefits. Clinicians should be empowered to recognize the role of such testing, informed so they can select the most appropriate investigation or clinical root for such testing and apply the result to obtain the maximum clinical utility. This may be within their own clinical team or as part of a wider MDT dependent upon their knowledge and experience and the clinical scenario that is being managed. This is true in endocrinology and indeed all medical specialities.

REFERENCES

1. Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 1953; **171**: 737–8.
2. Tijo JH, Levan A. The chromosome number of man. *Hereditas*, 1956; **42**: 1–6.
3. Lejeune J, Gautier M, Turpin R. Les chromosomes somatique des enfants mongoliens. *Comptes Rend Acad Science Paris*, 1959; **248**: 1721.
4. Ford CE, Jones KW, Polani PE, de Almeida JC, Briggs JH. A sex chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet*, 1959; **1**: 711–13.
5. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex determining mechanism. *Nature*, 1959; **183**: 302–3.
6. World Health Organization. WHO definitions of genetics and genomics. Available at: http://www.who.int/genomics/geneticsVSgenomics/en/ (accessed: 24 October 2018).
7. Ferguson-Smith MA. History and evolution of cytogenetics. *Mol Cytogenet*, 2015; **8**: 19–26.
8. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*, 1977; **74**: 5463–7.
9. Mullis KB. The unusual origins of the polymerase chain reaction. *Sci Am*, 1990; **262**; 56–65.
10. Royer-Pokora B, Kunkel LM, Monaco AP, *et al.* Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature*, 1986; **322**: 32–8.
11. Koenig M, Hoffman EP, Bertelson CJ, *et al.* Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* 1987; **50**: 509–17.
12. Friend SH, Horowitz JM, Gerber MR, *et al.* Deletions of a DNA sequence in retinoblastomas and mesenchymal tumors: organization of the sequence and its encoded protein. *Proc Nat Acad Sci*, 1987; **84**: 9059–63.
13. Brenner S, Johnson M, Bridgham J, *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotech*, 2000; **18**: 630–4.
14. Reuter JA, Spacek DV, Snyder MP. High-throughput sequencing technologies. *Mol Cell*, 2015; **58**: 586–97.
15. Riordan JR, Rommens JM, Kerem B, *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 1989; **245**: 1066–73.

16. Ng SB, Buckingham KJ, Lee C, *et al*. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 2010; **42**: 30–5.
17. The Deciphering Developmental Disorders study. Available at: https://www.ddduk.org/ (accessed 24 October 2018).
18. Beaulieu CL, Majewski J, Schwartzentruber J, *et al.* FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet*, 2014: **94**: 809–17.
19. National Institutes of Health. US Department of Health and Human Services. NIH genome sequencing program targets the genomic bases of common, rare disease. Available at: https://www.nih.gov/news-events/news-releases/nih-genome-sequencing-program-targets-genomic-bases-common-rare-disease (accessed 14 January 2016).
20. Genomics England. The 100,000 Genomes Project. Available at: https://www.genomicsengland.co.uk/the-100000-genomes-project/ (accessed 24 October 2018).
20. Davies DS. Chief Medical Officer annual report 2016: Generation Genome. Department of Health and Social Care. Available at: https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome (last updated 20 July 2017).
21. National Human Genome Research Institute. International Consortium Completes Human Genome Project. Available at: https://www.genome.gov/11006929/2003-release-international-consortium-completes-hgp/(last reviewed 8 May 2012).
22. 2009/872/EC: Commission Decision of 30 November 2009 establishing a European Union Committee of Experts on Rare Diseases. *OJ L 315, 2.12.2009, p. 18–21*. Available at: https://eur-lex.europa.eu/eli/dec/2009/872/oj (accessed 14 October 2018).
23. Genomics in Medicine. PHG Foundation, June 2012. Available at: http://www.phgfoundation.org/documents/295_1341583712.pdf (accessed 24 October 2018).
24. Bowdin S, Gilbert A, Bedoukian E, Carew C, *et al*. Recommendations for the Integration of Genomics into Clinical Practice. *Genet Med*, 2016; **18**: 1075–84.
25. NHS UK Genetic Testing Network. Available at: https://ukgtn.nhs.uk/ (accessed 24 October 2018).
26. Parker M, Lucassen A. Using a genetic test result in the care of family members: how does the duty of confidentiality apply? *Eur J Hum Genet*, 2018; **26**: 955–9.
27. Lucassen A. Should families own genetic information? Yes. *BMJ*, 2007; **335**: 22.
28. Clarke A. Should families own genetic information? No. *BMJ*, 2007; **335**: 23.
29. Plantinga L, Natowicz MR, Kass NE, Hull SC, Gostin LO, Faden RR. Disclosure, confidentiality, and families: experiences and attitudes of those with genetic versus nongenetic medical conditions. *Am J Med Genet C Semin Med Genet*, 2003; **119C**: 51–9.
30. Kohut K, Manno M, Gallinger S, Esplen MJ. Should healthcare providers have a duty to warn family members of individuals with an HNPCC-causing mutation? A survey of patients from the Ontario Familial Colon Cancer Registry. *J Med Genet*, 2007; **44**: 404–7.
31. Mitchell C, Ploem C, Chico V, *et al*. Exploring the potential duty of care in clinical genomics under UK law. *Med Law Int*, 2017; **17**: 158–82.
32. Xiang Zhang and Shuk-Mei Ho Epigenetics meets endocrinology. *J Mol Endocrinol*. 2011; **46**: R11–R32
33. Eggermann T, Perez de Nanclares G, Maher ER, *et al.* Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin Epigenetics*, 2015; **7**: 123–41.
34. Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 2015; **17**: 405–24.

1.10

Environmental Influences on Endocrine Disease

George Mastorakos, Markella Nezi, Djuro Macut, and Maria Papagianni

Introduction

During the last 60 years, chemical substances such as plasticizers, pesticides, detergents, paints, those found in metal food cans, flame retardants, cosmetics, and chemical waste have proliferated worldwide. Many can interfere with human and animal endocrine systems. Many natural products can also interfere with endocrine regulation (i.e. phytoestrogens). These chemicals are collectively known as endocrine disruptors (EDs). The public health risks related to EDs have raised justified concern and EDs have become the focus of major research.

Definition

According to the International Programme on Chemical Safety an ED is a chemical substance or mixture of substances that alters endocrine function(s) and consequently adversely affects health in an intact organism or its progeny [1]. It is important to distinguish permanent from transient endocrine interference by EDs. It is plausible to suggest that characterization as an ED should require a substance to change endocrine function(s) permanently, or to cause inherited changes in endocrine physiology. EDs interfere with the synthesis, secretion, transport, binding, action, and/or elimination of hormones involved in homeostasis, reproduction, development and/or behaviour [1]. Many are biologically active at extremely low doses, and their effects on humans, wildlife, and the environment have been the subject of close scrutiny internationally. EDs may cause, directly or indirectly, effects including infertility, abnormal prenatal development, precocious or delayed puberty, thyroid dysfunction, obesity, type 2 diabetes mellitus, behavioural disorders, and cancer. They can also alter nervous and immune system functions. ED Research is focused on three areas: (i) the 'time-window' of exposure to EDs, which is critical for the adverse outcome; (ii) the dose responses of EDs and their intracellular action; (iii) the ED-induced genomic or epigenetic modifications which affect offspring of exposed individuals [2]. Some EDs cross the placenta and are concentrated in the fetal circulation, while others are transferred from mother to infant through breast milk.

Historical Background

The existence of EDs has been known since the 1930s, when the oestrogenic action of some chemicals (e.g. bisphenol A; BPA) was shown in animals, and in the 1950s, dichloro-diphenyl-trichloroethane (DDT), a chemical pesticide, was reported to feminize roosters. In 'Silent Spring', an influential environmental science book of 1962, Rachel Carson presented examples of impaired development in birds, mammalian gonadal dysfunction, and human tumorigenesis. During the 1970s, the use of diethylstilboestrol (DES, a synthetic oestrogen) for prevention of abortion was common. Later, it was found that children of DES-treated women developed disorders such as vaginal carcinomas and infertility. The use of DES is now prohibited [3].

Mechanisms of Action

EDs affect concentrations and/or actions of hormones or their receptors. The most common mechanism of ED action is agonism at a hormone receptor, particularly the oestrogen receptor (ER) (e.g. BPA, DES), androgen receptor (AR) (e.g. vinclozolin) and aryl hydrocarbon receptor (AhR) (e.g. dioxins). Other EDs are competitive or non-competitive hormone receptor antagonists. Competitive antagonism usually leads to total receptor deactivation, while non-competitive antagonism causes the receptor to react more slowly or less efficiently. Typical antagonists acting at hormone receptors are the herbicides linuron and vinclozolin and their metabolites. EDs can also inhibit enzymatic hormone interconversion (e.g. aromatization of testosterone to oestrogen), induce hormone-metabolizing enzymes (e.g. cytochrome P450) or antagonize hormone binding to transport proteins, increasing bioavailability. Finally, EDs can induce receptor down-regulation or degradation [4].

Most Common Categories of EDs and Exposure Routes

The most common categories of EDs (Table 1.10.1) and their exposure routes are described next.

Polychlorinated biphenyls (PCBs): PCBs are synthetic organic chemicals formerly used as coolants and lubricants in transformers, capacitors, and other electrical equipment. Their use was banned in 1977, when their negative health effects became clear. They can still be found in old microscope or hydraulic oil, old fluorescent lighting fixtures, or electrical devices that contain old PCB capacitors. They can be inhaled when released from electrical devices heated during operation, ingested in contaminated food, or be absorbed through the skin. Infants can be exposed to PCBs through breast milk.

Phthalate esters: Phthalate esters, commonly called plasticizers, are used in plastics, rendering them more flexible and harder to break. They are found in diverse products including wall coverings, tablecloths, floor tiles, furniture upholstery, shower curtains, garden hoses, swimming pool liners, rainwear, baby pants, squeeze toys and dolls, shoes, automobile upholstery, packaging film and sheets, medical tubing, and blood storage bags. They are also added to cosmetics. There is risk of exposure by inhalation of air containing phthalate vapours or dust contaminated with phthalate particles, but minimal risk of exposure from drinking water due to low solubility in water. Phthalates can enter the body during medical procedures such as blood transfusions, kidney dialysis, intravenous fluid administration, or artificial ventilation. Elevated exposure has been documented in neonatal intensive care units [5].

Phenols-Bisphenol A: Bisphenol A is a light plastic with unique toughness, optical clarity, and high heat and electrical resistance. It has been used widely since the 1960s in eyeglass lenses, medical equipment, water bottles, CDs and DVDs, mobile phones, computers, household appliances, reusable food and drink containers, safety shields, sports equipment, industrial floorings, industrial protective coatings, can coatings, and electrical equipment. Although BPA is considered biodegradable, it can leach out of the lining in cans, potentially contaminating foods and liquids therein (3). Medical equipment may also contain BPA, and elevated BPA exposure has also been documented in neonatal intensive care units [6]. In 2011, use of BPA in baby bottles was prohibited in Europe. In France, from 2013, BPA was banned in food contacting materials for children up to 3 years old and from 2015 in all packaging, containers, and utensils in direct contact with food. However, the European Plastics Regulation (EU) authorizes use of BPA as a

Table 1.10.1 Effects of EDs on humans

Substances	Effects
***In utero* exposure**	
PCBs	Neuromuscular disorders, lower IQ, hypothalamic–pituitary–testicular axis dysregulation
Dioxins	Low birth weight, skin discoloration, bronchitis, developmental retardation
Phenols	Irregular menstrual cycles
Phytoestrogens, xenoestrogens, other substances with oestrogenic bioactivity	Ambiguous genitalia, obesity later in life, sexual differentiation problems, hormone-dependent cancers
DES	Transplacental carcinogenesis (cervicovaginal cancer in female offspring)
DDT, DDE	Low T_4 levels in infants
PCP	Altered thyroid hormone levels and thus, neurodevelopmental deficits
Nitrofen (pesticide)	Lung hypoplasia
Phthalate esters	Morphological abnormalities of male reproductive tract
Disruption in pubertal timing	
Lead	Delayed pubertal onset
PCBs, phytoestrogens, pesticides, BPA	Precocious female reproductive tract development
DDE, DDT	Earlier menarche
Disruption in reproduction	
BPA	Oocyte meiotic disturbances (i.e. aneuploidy), PCOS
Phytoestrogen genistein	Altered cyclicity, prolonged and abnormal cycles
Dioxins (TCDD)	Endometriosis
DES	Suppression of lactation
DDE, PCBs	Reduction of duration of lactation
EDs and cancer development	
Oestrogen-mimicking compounds	Breast cancer, Testicular cancer
PCBs, arsenic	Prostate cancer
Pesticides (i.e. atrazine)	Ovarian cancer
EDs and thyroid function	
BPA, PCBs, phytoestrogens	Hypo- or hyperthyroidism
EDs and obesity	
PCBs, pesticides, phthalates, BPA, metals	Weight gain
EDs and various functions	
BPA, phthalates, dioxins	Alterations in blood glucose

monomer for production of plastic with a specific migration limit of 0.6 mg/kg of food [7].

Dioxins: Dioxins are highly persistent in the environment. The most toxic is the 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD), an unintentional by-product of many industrial processes involving chlorine, such as waste incineration, chemical and pesticide manufacturing, pulp, and paper bleaching. The major source of dioxins in the environment is waste incineration. Dioxin pollution is also encountered in paper mills where chlorine bleaching is used. People are exposed to dioxins by diet, especially meat and dairy products. Since dioxins are fat soluble, they bioaccumulate and climb the food chain. Dioxins can also cross the placenta and are present in breast milk [8].

Pesticides/Herbicides: Pesticides and herbicides such as atrazine, DDT, and trifluralin were once used widely. Atrazine is a white powder used to protect grasses and broadleaf plants from pests. It is water soluble and taken up by plants, and can be inhaled as dust or ingested through contaminated water, but is not absorbed through skin. DDT is extremely hydrophobic and strongly absorbed by soils, where its half-life ranges from 22 days to 30 years. Routes of loss and degradation include runoff, volatilization, photolysis, and aerobic or anaerobic biodegradation. In aquatic ecosystems it is quickly absorbed by organisms and soil or else evaporates, leaving little DDT in the water itself. Its breakdown products and metabolites, dichlorodiphenyldichloroethylene (DDE) and dichlorodiphenyldichloroethane (DDD), are also highly persistent and have similar properties. Trifluralin is a herbicide against grasses and some broad-leaved weeds. It is usually directly incorporated into soil, although some mixtures may be sprayed. Trifluralin may enter the aquatic environment predominantly *via* diffuse sources resulting from its recommended use (e.g. in agricultural run-offs bound mainly to soil particles). Industrial discharges, accidental spillages during transport, storage, and use are potential sources of trifluralin contamination.

Phytoestrogens: These are a diverse group of naturally occurring non-steroidal plant compounds with oestrogenic or antioestrogenic properties due to structural similarity with 17-β-oestradiol. They are important in plant defence systems, mainly against fungi. Foods with the highest phytoestrogen content are nuts and oilseeds, followed by soy products, cereals and breads, legumes, meat products, and other processed foods containing soy, vegetables, fruits, alcoholic, and non-alcoholic beverages.

Parabens: Parabens are esters of p-hydroxybenzoic acid used as preservatives in foods, drugs, and in cosmetics or deodorants; they may also be found in polyester fabrics [9]. The most commonly used in cosmetics are methylparaben, propylparaben, butylparaben, and ethylparaben. They can cross the skin and in animal studies parabens demonstrated weak oestrogenic activity. No direct link between parabens and cancer has been established despite numerous studies. According to the European Medicines Agency (EMA), methylparaben has not been associated with adverse effects on male or female reproductive organs in juvenile rats or in embryo-fetal development studies. The use of methylparaben in oral formulations up to 0.2% is not a concern for humans of any age. The EMA concluded that additional information for parabens is not necessary, however, further data are needed to evaluate potential risk of propylparaben in children below 2 years of age. The use of propylparaben-containing medicines in this age group must be justified on a case-by-case basis, by weighing the need for treatment against potential risk.

Polybrominated diphenyl ethers (PBDEs): These are used in manufacture of plastic products. Certain PBDE congeners are persistent, bioaccumulative, and toxic to both humans and the environment. The exposure of mice to low doses during intrauterine or neonatal brain growth induces irreversible changes in mental function of adults [10]. Polybrominated diphenyl ethers are not chemically bound to plastics, foam, fabrics, or other products in which they are used, making them more likely to leach out.

Effects of EDs on *In Utero* Development

Effects of *in utero* exposure to EDs are intensely studied. Exposure at critical 'time-windows' during development can cause various disorders, most of them irreversible. Several chemicals or chemical classes can cause neurodevelopmental abnormalities by interfering with neuroendocrine functions. These include PCBs, dioxins, metals, pesticides, phytoestrogens, synthetic steroids, and triazine herbicides. Any compound that mimics or antagonizes a neurotransmitter, hormone, or growth factor in the developing brain may adversely affect fetal neurodevelopment. The resulting deficit, which could include cognitive dysfunction, altered neurologic development, or sensory deficits, depends on the severity of the disturbance, for example of thyroid function, and the developmental period when exposure occurred (**Figure 1.10.1**) [11].

The classical example of endocrine disruption *in utero* producing adult disease is prenatal exposure to DES. Maternal treatment during pregnancy produces cervicovaginal cancer in daughters when in adulthood, and was the first demonstration of human transplacental carcinogenesis. Daughters of DES-exposed mothers also had functional and anatomical abnormalities of the uterus and fallopian tubes, with subfertility [12].

Exposure to DDE and its metabolites during fetal development is associated with lower cord serum T_4, emphasizing the need for further study of thyroid function, growth, and neural development in children exposed to high doses of DDT. Neuromuscular disorders have been reported in infants exposed to PCBs *in utero* and during lactation, and lower IQ in children of 4 to 11 years of age prenatally exposed. Another chemical, related to BPA and named pentachlorophenol (PCP), alters thyroid hormone concentrations in newborns and may cause neurodevelopmental defects.

In 1968 contamination of rice oil with dioxins in the Japanese city of Yusho led to low birth weight, skin discoloration, bronchitis, and developmental retardation in neonates, and hypoactivity and hypotony in infants. Intrauterine exposure to dioxins might derange thyroid function and thus newborn development [11].

Maternal exposure to environmental pollutants during pregnancy and high oestrogenic bioactivity in newborns' serum have implicated fetal exposure to EDs in ambiguous genitalia development [12]. Perinatal exposure to BPA causes irregular cycles in mice, although insufficient evidence exists to confirm this in humans [13]. Perinatal exposure to EDs with oestrogenic activity is moreover proposed to induce later life obesity [14].

Phyto- and xeno-oestrogens often inhibit steroidogenic enzymes including 3β- and 17β-hydroxyl-steroid-dehydrogenase, aromatase, sulphatases, and sulphotransferases. Both phyto- and

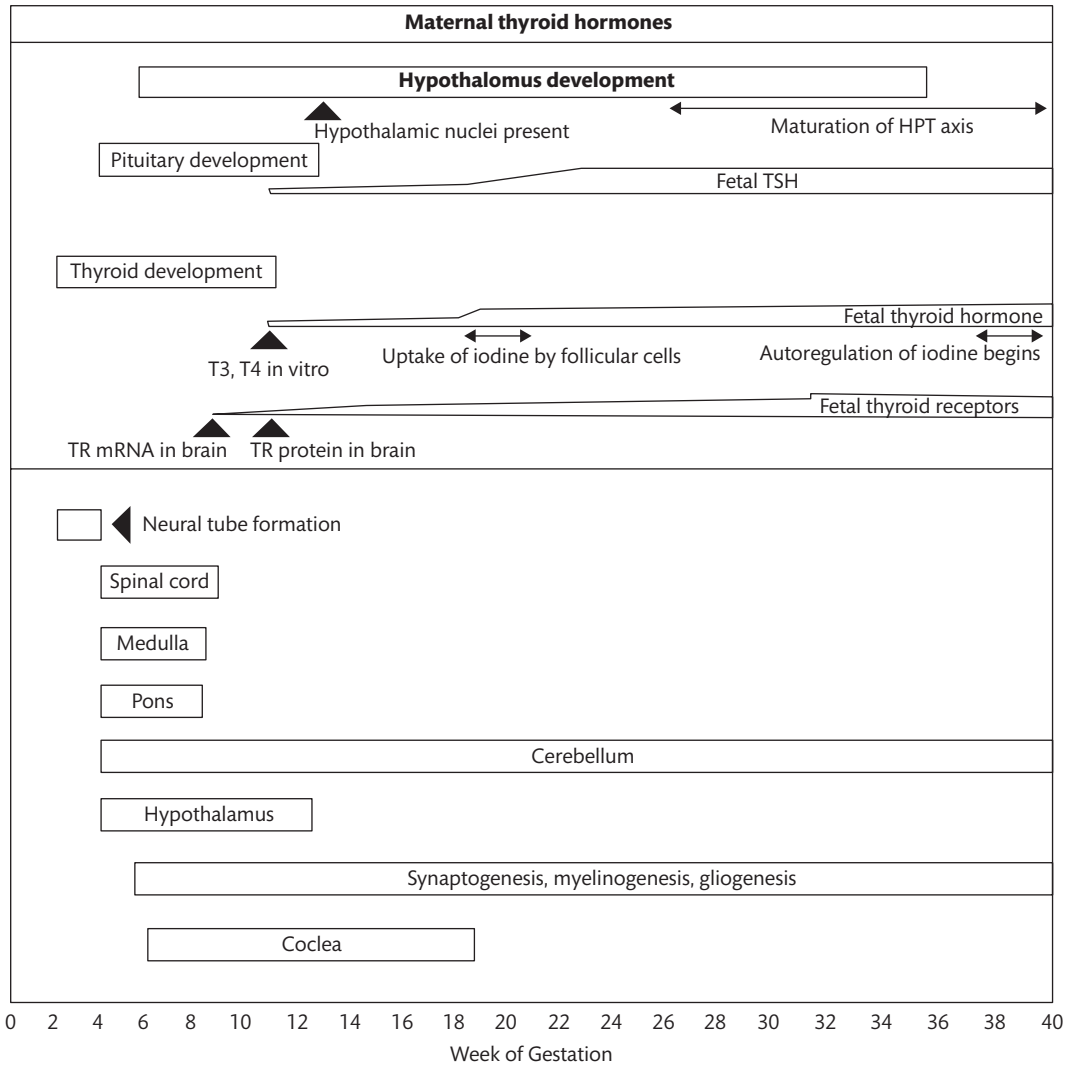


Figure 1.10.1 Fetal development of the hypothalamic–pituitary–thyroid axis during gestation with reference to development of thyroid hormone-dependent parts of the nervous system. Exposure to EDs at critical time-points may cause irreversible disorders.

xeno-oestrogens can also modulate intracellular signalling pathways, thereby inhibiting expression and activity of steroidogenic enzymes. Modulation of enzyme activity could impact sexual differentiation and development as well influencing risk of hormone-dependent cancers [15].

In utero exposure to phthalate esters is associated with morphological abnormalities of the male reproductive tract, including decreased anogenital distance, cryptorchidism, hypospadias, reduced Leydig cell numbers, and decreased testicular testosterone secretion. Testicular androgen signalling may be impaired through suppression of the hypothalamic–pituitary–testicular (HPT) axis. Exposure to DES impaired HPT signalling in rats, reducing plasma testosterone and increasing plasma follicle-stimulating hormone (FSH). The HPT axis is also disrupted in rats by PCB-169 exposure *in utero*, decreasing spermatogenesis, Leydig cell number, and plasma testosterone concentrations. Another herbicide, atrazine, with antiandrogenic and oestrogenic properties, exerts various adverse reproductive effects in rats including reduction of testicular testosterone secretion. Atrazine has low affinity for androgen and oestrogen receptors, reduces androgen synthesis, and enhances oestrogen production *via* increased intratesticular aromatization of androgens in both Leydig and Sertoli cells [16].

Numerous studies have assessed the relationship between EDs and birth weight, with inconsistent evidence of an association between reduced birth weight and exposure to insecticides, herbicides, or low-level PCBs but not DDT [17, 18]. Evidence of low birth weight related to phthalates, BPA and PBDE is similarly inconsistent [17]. Several studies have reported a relationship between exposure to PCBs, pesticides or phthalates and preterm delivery [19, 20], however further studies reported no association or even a positive association between phthalates and length of gestation [21].

Effects of EDs on Pubertal Timing

Falling age of onset of puberty and increased incidence of precocious puberty have been noted over recent decades in many countries, and cannot be wholly explained by factors such as increasing adiposity or changing living conditions. EDs are a suspected major contributing factor [22]. Recent studies have revealed an impact

of EDs on pubertal timing in both males and females, particularly for oestrogenic or antiandrogenic compounds [23]. Exposure may be gestational, lactational or juvenile, and EDs can disturb the hypothalamic–pituitary–gonadal axis through negative feedback to hypothalamus and pituitary as well as through direct gonadal effects. Lead exposure has been associated with delayed puberty, while exposure to phytoestrogens, PCBs, pesticides, or BPA was implicated in precocious female development. In male rodents, transient neonatal androgen exposure reduced testis weight and testosterone production [23].

A study of adolescent girls in Canada revealed that exposure to chemicals such as PCBs is associated with oestrogenic phenomena and may affect menarche. Lead was associated with a delayed median age at menarche, when controlling for other toxicants, age, and socioeconomic status. However different effects may occur by exposure to different doses of EDs [24]. Age at menarche is reduced in girls exposed to oestrogenic organochlorines, but discerning the exact contribution of these substances to precocious menarche is complicated by the numerous environmental variables influencing menarche. A study of women exposed to DDE through Great Lakes fish consumption found a reduction of one year in the age of menarche for each increase of 15 mg/L serum DDE, while in Chinese textile workers a 10 mg/L serum DDT increase was associated with 0.2 year reduction in menarchal age [13]. In the United States, in 1151 girls aged 6 to 8 years assessed for pubertal development, positive associations between low molecular weight urinary phthalate metabolites and breast and pubic hair development were demonstrated [25]. An inverse relationship between high molecular weight phthalates and pubic hair development was reported, but no association between BPA and pubertal development was found. A smaller study of Turkish boys demonstrated a greater concentration of high molecular weight phthalates in boys 11 to 15 years old with pubertal gynaecomastia compared to controls [26].

Effects of EDs on Female Reproduction

The impact of EDs on human fertility remains controversial. Ten years ago, because mice housed in damaged polycarbonate plastic cages were shown to exhibit a high frequency of oocytes with meiotic disturbances, oocyte-damaging effect of the oestrogenic plasticizer BPA was investigated. When BPA was added to water of mice in intact cages, similar oocyte meiotic disturbances were induced, some causing aneuploidy. Other EDs such as DES have also been implicated in meiotic disturbances.

Takeuchi *et al.* reported that plasma BPA concentrations were increased in both non-obese women with polycystic ovary syndrome (PCOS) and obese women with or without PCOS compared to non-obese women without PCOS [27]. The same authors reported an association between BPA exposure, complex endometrial hyperplasia, and endometrial cancer [28], although another study found lower BPA concentrations in patients with premalignant complex endometrial hyperplasia potential than in either control or simple endometrial hyperplasia groups. BPA may thus represent one of the many underlying causes of PCOS, perhaps acting as a modifier to worsen symptoms or contributing to the phenotype in genetically predisposed women [29]. A strong association between BPA and both plasma androgen concentrations and insulin resistance indices has been reported, implicating BPA potentially in these two major components of PCOS pathophysiology. It is also known that women with ovulatory dysfunction have higher serum BPA concentrations than normally ovulating women [27] and that both have lower BPA concentrations than men [30]. BPA binds sex hormone-binding globulin, competing with natural androgens and increasing bioactive androgen concentrations [31].

Menstrual cyclicity may be perturbed by EDs by other means also. Experimental studies of exposure of neonatal mice to physiologically relevant concentrations of the phytoestrogen genistein caused prolonged and abnormal cycles in adult animals. In humans, altered cyclicity and adult simultaneous exposure to persistent organic pollutants and pesticides have been linked. Studies examining the influence of organochlorine pesticide exposure on cyclicity and fecundity suggest that organochlorine exposure shortens menstrual cycles, while women exposed to hormonally active pesticides (non-organochlorine) have a 60–100% increased odds for long cycles, intermenstrual bleeding, and missed menstrual bleeds.

EDs acting upon the human uterus produce structural changes including in development of uterine leiomyomas, while on the other hand consumption of phytoestrogens, such as those in soy, in a Japanese study was associated to decreased hysterectomy rates, implying a potential uterine protective effect of modest phytoestrogen consumption.

Finally, there appears to be a correlation between ED exposure and lactation: Exogenous oestrogens such as DES effectively suppress lactation [12], and duration of lactation is reduced in women with elevated serum concentrations of PCBs and DDE in a dose-dependent manner [13].

Effects of EDs on Male Reproduction

Significant toxicologic data from laboratory and wild animal studies suggest that exposure to certain EDs is associated with male reproductive toxicity, manifesting as abnormalities of the reproductive tract (cryptorchidism, hypospadias), reduced semen quality, and impaired adult fertility [32]. Endocrine disruption of spermatogenesis may be accounted for by four mechanisms, including: (1) epigenetic changes to the genome; (2) apoptosis of germ cells; (3) dysregulation of androgenic signalling; and (4) disruption of Sertoli and other spermatogenesis supporting cells [16]. The effects of EDs in male reproductive anomalies are presented in **Table 1.10.2**.

The 'endocrine disruption hypothesis' suggests ED exposure during fetal, neonatal, and adult life perturbs development of male reproductive organs, altering semen quality and reproductive hormone production [33]. Many studies in rodents have supported this, demonstrating that prenatal and perinatal exposure to EDs such as BPA, phthalates, and alkylphenols causes genitourinary developmental anomalies [34], decreased epididymal [35], and increased prostate weight [35], lower sperm production [16] and, in some cases, cryptorchidism. Shorter anogenital distance and smaller testicular and penile size are also described [36]. Neonatal ED exposure can decrease sperm motility and change morphology [37], including induction of abnormalities in the acrosomal granule and nucleus in older spermatids and spermatozoa, with attendant decreased height of the seminiferous epithelium. ED exposure of pubertal rodents

Table 1.10.2 Effects of EDs on male fertility

Compound	Outcome
Dibromochloropropane	Azoospermia and oligospermia Decreased motility and morphology Elevated FSH and LH Deficit of male births
Ethylene dibromide	Decreased sperm counts
Chlordecone (kepone)	Oligospermia, decreased sperm motility
Perchloroethylene	Dose-related morphological changes
Carbaryl	Impaired semen quality
Ethylene glycol ethers	Decreased sperm counts Decreased fertility
TCDD	Reduced serum testosterone, increased LH Deficit of male births
p-Nitrophenol	Decreased sperm concentration Decreased percentage of motile sperm Increased serum LH

has been correlated with decreased epididymal spermatozoa [38], and testosterone concentrations [38], increased testicular apoptosis, and loss of seminiferous epithelium. Finally, adult ED exposed rats show decreased epididymal spermatozoa concentration, sperm count, epididymal weight and testosterone concentrations along with increased luteinizing hormone concentration [39].

In animal models, there is thus substantial evidence for adverse effects on male reproduction for several EDs, but direct evidence from human studies is required to prove that EDs cause adverse effects on the male reproductive system. Human studies are challenging, however, because of the complexity of the field, and a recent systematic review of published evidence was unable to clarify the impact of EDs on human male reproductive health [40].

Effects of EDs on Cancer Development

There is increasing concern about cancer development after ED exposure. Gestational and perinatal exposures may have long-term effects on the endocrine system that influence tumour development later in life. Not only does the synthetic oestrogen DES cause vaginal adenocarcinoma in female offspring and extensive prostatic squamous metaplasia in male offspring of exposed mothers, but elevated gestational concentrations of natural oestrogens have been associated with increased incidence of breast cancer in adulthood [41].

Many studies have investigated the effect of underarm cosmetic products (UCPs) on breast cancer. Active ingredients in most UCPs are aluminium-based compounds such as aluminium chloride and aluminium chlorohydrate. Frequent use of UCPs may lead to aluminium accumulation in breast tissue, and in an Austrian study, women who reported UCP use several times a day starting under 30 years old were reported to have increased risk of breast cancer [42]. Developmental mammary toxicants may increase mammary tumour risk if they alter circulating or intratissue hormone concentrations, receptor expression, hormone transport, or hormone metabolism to alter responses to endogenous hormones or growth factors. Many oestrogenic environmental chemicals have been measured in human breast tissue and could be associated with increased breast cancer risk, however, although animal models support this point of view, new studies are needed to clarify the role of EDs in the human breast. Investigation of the impact of multiple oestrogen-mimicking compounds, at varying doses is required. New research is also needed to clarify whether the type of ED is the key cancerogenic factor, or if the exposure time and route is rather most important in cancer development.

Studies in the last decade provided no evidence of association between PCB exposure and breast cancer. No association was found either between DDT, DDE, and breast cancer, but further studies are needed. The impact of phytoestrogens on breast cancer risk is controversial. Some studies demonstrate an association with breast cancer, while others suggest a protective effect of these compounds [43].

Initial suspicions that testicular germ cell cancer (TGCC) may be related to early life exposure to environmental oestrogens and/or antiandrogens, while logical, are currently supported by little evidence. The dramatic recent increase in TGCC incidence does indicate that environmental and lifestyle factors are likely important but no compelling evidence implicating environmental oestrogenic or otherwise hormonally active substances to this rise in Western nations has emerged. More studies are warranted, however, as the rarity of the cancer, the long lag time between the presumed fetal sensitive period and the appearance of the cancer, and the lack of a good animal model have greatly hindered understanding.

It has been suggested that the exposure to environmental chemicals such as phthalates may impair semen quality, and exposure may coincide with development of testicular cancer, cryptorchidism, and hypospadias [44]. A case–control study did not find any association between serum organochlorine concentrations and risk of TGCC, but did find that organochlorine concentrations in mothers' peripheral blood decades after their sons' birth were predictive of increased risk in their sons [45]. It is plausible that *in utero* exposure to endocrine disruptor compounds (EDCs) is the aetiological exposure window. The exact contribution of pesticide and ED exposure to development of testicular cancer has also not been fully elucidated, although genetic and environmental factors have been implicated in reduced sperm counts and cancer development. Once again, the 2–3-decade latency between fetal exposure and adult cancer appearance is a major challenge for study design, highlighting the value of large-scale epidemiological approaches [44].

Increasing evidence both from epidemiological and animal studies suggests that EDs may influence development or progression of prostate cancer. These effects appear largely linked to oestrogen signalling, either *via* interaction with ERs or influences on steroid metabolism and thus oestrogen concentrations. In human studies, exposure to PCBs or inorganic arsenic have been associated with elevated prostate cancer risk, although risk seems to be increased only by exposure during critical developmental 'time-windows' (*in utero*, neonatal, peripubertal). Thus, infants and children may be considered a population highly susceptible to increased risk of later prostate cancer due to ED exposure [46].

Ovarian cancer development may also be influenced by hormone action, and pesticides with relevant endocrine-disrupting activity remain widely used in some countries. Research to date most strongly suggests a link between atrazine and ovarian cancer risk, but other environmental and occupational exposures may also be associated and require further study [47].

Effects of EDs on Thyroid Function

The existence of many thyroid disrupting chemicals has been proven in animal and human studies. Disruption occurs at many different levels of thyroid hormone synthesis, binding, action, and metabolism. The most common EDs that affect thyroid function are BPA, pentachlorophenol, PCBs, and phytoestrogens. Most studies of EDs and thyroid physiology were performed in rodents, with only some in humans. A substantial amount of information, albeit not conclusive, has thus been gathered about HPT axis perturbation by EDs. They can target the Na/I symporter, deiodinases, transport proteins and nuclear thyroid receptor (TR) and have been implicated in hypo- or hyperthyroidism, thyroid nodule formation, and thyroid tumorigenesis. Disruption of thyroid hormones during fetal life by exposure to EDs *in utero* may also produce irreversible developmental disorders including structural and functional brain abnormalities with postnatal behavioural changes.

Isolated measurements made during pre- or postnatal period cannot accurately test the cumulative or synergistic properties of EDs during these periods. However, much of the T_4 circulating in the fetus is of maternal origin and chemicals altering maternal thyroid function seem to be associated with impaired fetal development. Thorough investigation needs to take into account maternal physiological changes during pregnancy required to meet the needs of the conceptus in terms of energy supply and waste elimination. Three different compartments—maternal, placental, and fetal—interact closely during pregnancy and they should all be approached as one entity [11]. Limited data to date do not permit firm conclusions about the effects of EDs on thyroid function, and more research will be of crucial importance [11].

Effects of EDs on Obesity Development

The role of environmental chemicals on obesity development is emerging as an area of research primarily focusing on identification of obesogens. Obesogens are defined as chemicals which influence energy balance and/or lipid homeostasis to promote adipogenesis and lipid accumulation, and experimental evidence supporting this action of numerous chemicals is growing [48]. Many mimic endogenous lipophilic hormones which exert effects via members of the nuclear receptor superfamily, and environmental oestrogenic chemicals such as BPA and nonylphenol have been shown to promote differentiation and/or proliferation of murine preadipocyte cell lines.

The hypothesis that environmental chemicals could be responsible for the obesity pandemic was first introduced by Paula Baillie-Hamilton in 2002, based in part on evidence from toxicologic studies in which low-dose chemical exposures were associated with weight gain in experimental animals [49]. More recent studies have shown that chemicals including pesticides, organophosphates, PCBs, polybrominated biphenyls, phthalates, BPA, heavy metals, and solvents might cause weight gain by interfering with weight-controlling hormones, by altering sensitivity to neurotransmitters or by altering activity of the sympathetic nervous system. However, more research is needed.

Biological effects of some EDs may be prolonged in obese individuals due to retention in adipose tissue. This can exacerbate disruptive effects of fat-soluble toxicants, such as PCBs and dioxin-type compounds.

Other Effects of EDs

Lately, it has been shown that EDs might affect other systems such as the immune system and the endocrine pancreas. For example, EDs may be involved in autoantibody production by B1 cells and could be an aetiologic factor in development of autoimmunity. Other studies suggest that concentrations of BPA, phthalates, dioxins, and persistent organic pollutants are correlated with altered blood glucose homeostasis in humans [50], and in other studies EDs have been connected with diabetes and cardiovascular diseases. Dioxins, pesticides, and BPA can also cause insulin resistance and altered β-cell function in animal models [51].

Conclusions

The impact of EDs has been a matter of concern for 60 years. However, research remains challenging due to the multiple environmental effects seen on humans, due to differences in genetic susceptibility, and due to varying routes and durations of exposure. The exact time-point of exposure is crucial. *In utero* exposure may cause irreversible outcomes. Experimental studies may furthermore not agree with human studies due to different responsiveness of laboratory animals used. More research is needed to clarify mechanisms of ED action. Ultimately a choice must be made between short-term benefits of technological advances and global quality of life rooted in respect for the environment.

REFERENCES

1. U.S. Environmental Protection Agency. Special report on environmental endocrine disruption: an effects assessment and analysis. EPA/630/R-96/012. Washington, DC: Office of Research and Development, 1997. Available at: http://www.epa.gov/raf/publications/pdfs/ENDOCRINE.PDF (accessed 14 December 2008).
2. Gore AC, Heindel JJ, Zoeller RT. Endocrine disruption for endocrinologists (and others). *Endocrinology*, 2006; **147**(6 Suppl): S1–3.
3. Solomon GM, Schettler T. Environment and health: 6. Endocrine disruption and potential human health implications. *CMAJ*, 1997; **163**: 1471–6.
4. Lintelmann J, Katayama A, Kurihara N, Shore L, Wenzel A. Endocrine disruptors in the environment. *Pure Appl Chem*, 2003; **75**: 50.
5. Green R, Hauser R, Calafat AM, *et al.* Use of di(2-ethylhexyl) phthalate-containing medical products and urinary levels of mono(2-ethylhexyl) phthalate in neonatal intensive care unit infants. *Environ Health Perspect*, 2005; **113**: 1222–5.
6. Calafat AM, Weuve J, Ye X, *et al.* Exposure to bisphenol A and other phenols in neonatal intensive care unit premature infants. *Environ Health Perspect*, 2009; **117**: 639–44.
7. Monneret C. What is an endocrine disruptor? *C R Biol*, 2017; **340**: 403–5.

8. Dioxin homepage. Available at: http://www.ejnet.org/dioxin/ (accessed 14 December 2008).
9. European Medicines Agency, Science Medicines Health. News: new recommendations on the use of phthalates and parabens as excipients in human medicines released for public consultation 08/05/2013. Available at: https://www.ema.europa.eu/en/news/new-recommendations-use-phthalates-parabens-excipients-human-medicines-released-public-consultation (accessed 23 January 2020).
10. Branchi I, Capone F, Alleva E, Copsta GL. *NeuroToxicology*, 2003; **24**: 449–62.
11. Mastorakos G, Karoutsou EI, Mizamtsidi M, Creatsas G. The menace of endocrine disruptors on thyroid hormone physiology and their impact on intrauterine development. *Endocrine*, 2007; **31**: 219–37.
12. McLachlan JA, Simpson E, Martin M. Endocrine disrupters and female reproductive health. *Best Pract Res Clin Endocrinol Metab*, 2006; **20**: 63–75.
13. Crain DA, Janssen SJ, Edwards TM, *et al*. Female reproductive disorders: the roles of endocrine-disrupting compounds and developmental timing. *Fertil Steril*, 2008; **90**: 911–40.
14. Newbold RR, Padilla-Banks E, Snyder RJ, Jefferson WN. Perinatal exposure to environmental estrogens and the development of obesity. *Mol Nutr Food Res*, 2007; **51**: 912–17.
15. Whitehead S, Rice S. Endocrine-disrupting chemicals as modulators of sex steroid synthesis. *Best Pract Res Clinic Endocrin Metab*, 2006; **20**: 45–61.
16. Phillips KP, Tanphaichitr N. Human exposure to endocrine disrupters and semen quality. *J Toxicol Environ Health B Crit Rev*, 2008; **11**: 188–220.
17. Wigle DT, Arbuckle TE, Turner MC, *et al*. Epidemiologic evidence of relationships between reproductive and child health outcomes and environmental chemical contaminants. *J Toxicol Environ Health B Crit Rev*, 2008; **11**: 373–517.
18. Govarts E, Nieuwenhuijsen M, Schoeters G, *et al.* Birth weight and prenatal exposure to polychlorinated biphenyls (PCBs) and dichlorodiphenyldichloroethylene (DDE): a meta-analysis within 12 European Birth Cohorts. *Environ Health Perspect*, 2012; **120**: 162–70.
19. Stillerman KP, Mattison DR, Giudice LC, Woodruff TJ. Environmental exposures and adverse pregnancy outcomes: a review of the science. *Reprod Sci*, 2008; **15**: 631–50.
20. Meeker JD, Hu H, Cantonwine DE, *et al*. Urinary phthalate metabolites in relation to preterm birth in Mexico city. *Environ Health Perspect*, 2009; **117**: 1587–92.
21. Adibi JJ, Hauser R, Williams PL, *et al.* Maternal urinary metabolites of Di-(2-Ethylhexyl) phthalate in relation to the timing of labor in a US multicenter pregnancy cohort study. *Am J Epidemiol*, 2009; **169**: 1015–24.
22. Wolff MS, Teitelbaum SL, Pinney SM, *et al.*; Breast Cancer and Environment Research Centers. Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. *Environ Health Perspect*, 2010; **118**: 1039–46.
23. Jacobson-Dickman E, Lee MM. The influence of endocrine disruptors on pubertal timing. *Curr Opin Endocrinol Diabetes Obes*, 2009; **16**: 25–30.
24. Denham M, Schell LM, Deane G, Gallo MV, Ravenscroft J, DeCaprio AP. Relationship of lead, mercury, mirex, dichlorodip henyldichloroethylene, hexachlorobenzene, and polychlorinated biphenyls to timing of menarche among Akwesasne Mohawk girls. *Pediatrics*, 2005; **115**: 127–34.
25. Wolff MS, Teitelbaum SL, Pinney SM, *et al.*; Breast Cancer and Environment Research Centers. Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. *Environ Health Perspect*, 2010; **118**: 1039–46.
26. Durmaz E, Ozmert EN, Erkekoglu P, *et al.* Plasma phthalate levels in pubertal gynecomastia. *Pediatrics*, 2010; **125**: e122–9.
27. Takeuchi T, Tsutsumi O, Ikezuki Y, Takai Y, Taketani Y. Positive relationship between androgen and the endocrine disruptor, bisphenol A, in normal women and women with ovarian dysfunction. *Endocr J*, 2004; **51**: 165–9.
28. Hiroi H, Tsutsumi O, Takeuchi T, *et al*. Differences in serum bisphenol A concentrations in premenopausal normal women and women with endometrial hyperplasia. *Endocr J*, 2004; **51**: 595–600.
29. Kandaraki E, Chatzigeorgiou A, Livadas S, *et al*. Endocrine disruptors and polycystic ovary syndrome (PCOS): elevated serum levels of bisphenol A in women with PCOS. *J Clin Endocrinol Metab*, 2011; **96**: E480–4.
30. Takeuchi T1, Tsutsumi O. Serum bisphenol A concentrations showed gender differences, possibly linked to androgen levels. *Biochem Biophys Res Commun*, 2002; **291**: 76–8.
31. Déchaud H, Ravard C, Claustrat F, de la Perrière AB, Pugeat M. Xenoestrogen interaction with human sex hormone-binding globulin (hSHBG). *Steroids*, 1999; **64**: 328–34.
32. Mantzouki C, Bliatka D, Iliadou PK, *et al*. Serum Bisphenol A concentrations in men with idiopathic infertility. *Food Chem Toxicol*, 2019; **7**: 562–5.
33. Toppari J, Larsen JC, Christiansen P, *et al.* Male reproductive health and environmental xenoestrogens. *Environ Health Perspect*, 1996; **104** Suppl 4: 741–803.
34. Salian S, Doshi T, Vanage G. Neonatal exposure of male rats to Bisphenol A impairs fertility and expression of sertoli cell junctional proteins in the testis. Toxicology, 2009; **265**: 56–67.
35. vom Saal FS, Cooke PS, Buchanan DL, *et al*. A physiologically based approach to the study of bisphenol A and other estrogenic chemicals on the size of reproductive organs, daily sperm production, and behavior. *Toxicol Ind Health*, 1998; **14**: 239–60.
36. Foster PM. Disruption of reproductive development in male rat offspring following *in utero* exposure to phthalate esters. *Int J Androl*, 2006; **29**: 140–7.
37. Aikawa H, Koyama S, Matsuda M, Nakahashi K, Akazome Y, Mori T. Relief effect of vitamin A on the decreased motility of sperm and the increased incidence of malformed sperm in mice exposed neonatally to bisphenol A. *Cell Tissue Res*, 2004; **315**: 119–24.
38. Herath CB, Jin W, Watanabe G, Arai K, Suzuki AK, Taya K. Adverse effects of environmental toxicants, octylphenol and bisphenol A, on male reproductive functions in pubertal rats. *Endocrine*, 2004; **25**: 163–72.
39. Tohei A, Suda S, Taya K, Hashimoto T, Kogo H. Bisphenol A inhibits testicular functions and increases luteinizing hormone secretion in adult male rats. *Exp Biol Med (Maywood)*, 2001; **226**: 216–21.
40. Bliatka D, Lymperi S, Mastorakos G, Goulis DG. Effect of endocrine disruptors on male reproduction in humans: why the evidence is still lacking? Andrology. 2017 May;5(3):404–7.
41. Soto A, Maffini M, Sonnenschein C. Neoplasia as development gone awry: the role of endocrine disruptors. Int J Androl 2008; **31**: 288–93.

42. Linhart C, Talasz H, Morandi EM, *et al*. Use of underarm cosmetic products in relation to risk of breast cancer: a case-control study. *EBioMedicine*, 2017; **21**: 79–85.
43. Salehi F, Turner MC, Phillips KP, Wigle DT, Krewski D, Aronson KJ. Review of the etiology of breast cancer with special attention to organochlorines as potential endocrine disruptors. *J Toxicol Environ Health B Crit Rev*, 2008; **11**: 276–300.
44. Nordkap L, Joensen UN, Blomberg Jensen M, Jørgensen N. Regional differences and temporal trends in male reproductive health disorders: semen quality may be a sensitive marker of environmental exposures. *Mol Cell Endocrinol*, 2012; **355**: 221–30.
45. Hardell L, Bavel B, Lindström G, Eriksson M, Carlberg M. *In utero* exposure to persistent organic pollutants in relation to testicular cancer risk. *Int J Androl*, 2006; **29**: 228–34.
46. Prins GS. Endocrine disruptors and prostate cancer risk. *Endocr Relat Cancer*, 2008; **15**: 649–56.
47. Salehi F, Dunfield L, Phillips K, Krewski D, Vanderhyden B. Risk factors for ovarian cancer: an overview with emphasis on hormonal factors. *J Toxic Environ Health B Crit Rev*, 2008; **11**: 301–21.
48. Grün F, Blumberg B. Minireview: the case for obesogens. *Mol Endocrinol*, 2009; **23**: 1127–34.
49. Baillie-Hamilton PF. Chemical toxins: a hypothesis to explain the global obesity epidemic. *J Altern Complement Med*, 2002; **8**: 185–92.
50. Ropero A, Alonso-Magdalena P, Garcia-Garcia E, Ripoll C, Fuentes E, Nadal A. Bisphenol A disruption of the endocrine pancreas and blood glucose homeostasis. *Int J Androl*, 2008; **31**: 194–200.
51. Alonso-Magdalena P, Quesada I, Nadal A. Endocrine disruptors in the etiology of type 2 diabetes mellitus. *Nat Rev Endocrinol*, 2011; **7**: 346–53.

1.11

Endocrinology, Sleep, and Circadian Rhythms

Georg Brabant and Henrik Oster

Introduction

Endogenous circadian rhythms enable organisms to prepare for predictable environmental changes brought about by the Earth's rotation around its axis by temporally modifying behavioural and physiological functions—including the endocrine system [1]. One of the most obvious examples of circadian behaviour is the sleep–wake cycle, closely linked to diurnal variations of locomotor activity, body temperature regulation, and water/food intake. Even subtle changes in these circadian cycles may lead to detrimental effects. Such causative relationship between these changes and adverse biological effects have been obtained not only from mutations characterized in genes responsible for the generation and the integration of circadian rhythms, but also from observational studies where circadian rhythmicity was experimentally changed. Life in modern societies tends to increasingly ignore natural time cues and these environmental insults are increasingly recognized as the underlying mechanism for many pathophysiological changes and a higher susceptibility to disease. Focusing on endocrine-related effects, this chapter will highlight our current understanding of the genetic mechanisms of circadian rhythms, their integration with the light–dark cycle, and their links to sleep-related changes.

Mechanisms Underlying Circadian Behaviour and Definitions

Definitions

Rhythmic circadian (from *circa diem* (lat.)—'about a day') behaviour is not restricted to humans, but can be detected in most animals, plants, and even in prokaryotes. Circadian rhythms synchronize biological processes to the day–night (or light–dark) cycle of the natural environment. In mammals, it was originally believed that only specialized neurons of the hypothalamic suprachiasmatic nucleus (SCN) are able to induce circadian behaviour, but recent detection of rhythmicity in most peripheral organs or cells have challenged this view. The detection of clock and clock output genes in these peripheral cells and of rhythmic behaviour when time cues from the SCN are missing suggests a general endogenous pattern. In the current model, the SCN integrates information about the external light–dark cycle and synchronizes a ubiquitous network of cellular clocks throughout the body with each other and with external time. At the cellular level, these clocks coordinate physiology through rhythmic activation of transcriptional programmes. Approximately 5–10% of genes in a given tissue show such cyclical circadian behaviour. It has been estimated that across the whole organism, 40–50% of all genes are rhythmic in at least one tissue [2].

The time interval between two peaks of an individual circadian rhythm defined as its *period* is synchronized to the 24-h cycle by external time cues, so-called *Zeitgebers*. In contrast to the SCN, which integrates time cues mainly from the light circle, peripheral cells are insensitive to light and are predominantly regulated by internal *Zeitgebers* such as hormones or autonomic signals, and the timing of

energy intake. When these time cues are missing, temporal coordination follows the endogenous sustained rhythm provided by the SCN and—under specific conditions—individual circadian rhythms may even dissociate following their own internal pace. These *free-running rhythms* have been detected in experiments with volunteers kept in long-term isolation from all *Zeitgebers*. Depending on the exact experimental conditions, free-running rhythms in humans are estimated at a period length of 24–25 h, with individual rhythms reported to show period lengths even beyond 30 h [3].

One of the most striking real-life conditions where synchronization of circadian rhythms is temporarily lost is *jetlag* due to a transmeridian flight travel. Under jetlag conditions, internal and external time do no longer match and the internal clock system transiently becomes uncoupled during the following adaptation phase. Typical jetlag symptoms involve fatigue, immune suppression, mood disorder, and gastrointestinal dysfunction. There are multiple other examples for a weakening of the coordination of circadian rhythms including (night) shift work or extreme lighting conditions during winter or summer in high-latitude regions—both of which favour the development of metabolic and mood disorders.

With the recent characterization of the molecular mechanisms' underlying circadian rhythmicity mutational changes have been described, affecting the circadian mechanism in all cells. This may result either in shorter or longer than expected underlying rhythms, alter coordination between internal and external rhythms or between different tissues, or external time cues may no longer be able to optimally synchronize circadian rhythmicity under specific conditions such as given earlier. The so-called *phase angle*, defined as interval between a reference event like the beginning of the night and the peak of a given rhythm, is altered. It can be shifted to an earlier time, i.e. become *phase advanced*, or may be prolonged and become *phase delayed*. Every individual appears to have a given phase angle of its sleep–wake rhythm relative to external time which determines its *chronotype*. Morning ('lark') or evening ('owl') types have been characterized in humans who differ by up to 4 h in their optimal cognitive function and other physiological peak and trough performances due to the setup of their molecular clock [4].

Molecular Mechanism

Seminal work in plants, flies, mice, and humans on the genetic basis of circadian behaviour led to the independent discovery of the first genes involved in circadian behaviour, from non-vertebrates to mammals. The first clock gene, *period*, was identified in a behavioural mutagenesis screen in the fruit fly *Drosophila melanogaster* by Seymor Benzer and his student Ronald Konopka in 1970 [5]. While the different components of the circadian clock system differ between species, the principle makeup appears very much conserved across most multicellular organisms, from plants to mammals. At its core, the circadian clock consists of interlocked transcriptional–translational feedback loops driving 24-h rhythms of transcriptional activity regulating cellular physiology. For their unravelling of this principle of molecular clockworks, Michael Rosbash, Jeffrey Hall, and Michael W. Young received the Nobel Prize for Physiology or Medicine 2017 (https://www.nobelprize.org/nobel_prizes/medicine/laureates/2017/).

The first mammalian clock gene was described in 1994 by Joseph S. Takahashi in mice [6]. Further molecular components of the mammalian circadian clock were unravelled in the following years building a current working model of how this clock functions at the molecular level, even though there are still many open questions remaining. In brief, two basic-loop-helix transcription factors, circadian locomotor output cycles kaput (CLOCK) and its binding partner brain and muscle ARNT-like 1 (BMAL1; also known as aryl hydrocarbon receptor nuclear translocator-like or ARNTL) drive the expression of three *Period* (*Per1–3*) and two *Cryptochrome* (*Cry1, Cry2*) genes during the day. As monomers PER and CRY proteins are highly unstable and, therefore, with a delay of several hours, PER and CRY proteins heterodimerize and translocate into the nucleus where they inhibit their own transcription following interaction with the CLOCK–BMAL1 complex. Subsequently, during the dark period, the PER-CRY repressor complex is degraded, and a new cycle of transcription is initiated. The period of one such cycle approximates 24 h. This primary feedback loop is stabilized by a second negative feedback through the nuclear hormone receptors REV-ERBα/β and RORα. REV-ERBα/β, direct targets of CLOCK-BMAL1, are strong inhibitors of *Bmal1* transcription while RORα (whose transcription is also activated by CLOCK-BMAL1) enhances *Bmal1* transcription. At least two further accessory loops have been described. This basic feedback regulation is modulated by a large number of additional factors that change the kinetics of the feedback by altering the stoichiometry of clock protein complexes. During the late afternoon and night, PER proteins are progressively phosphorylated through key kinases such as casein kinase 1δ and ε (CSNK1δ; CSNK1ε). These phosphorylation steps are crucial for the degradation of clock proteins via the proteasomal pathway. Mutants in any members of these regulators may alter the kinetic of the circadian process and result in either short or long periods. Figure 1.11.1 schematically illustrates this process. In addition, more recently the critical importance of circadian changes in histone H3 acetylation and chromatin remodelling for circadian transcription of CLOCK-BMAL1 target genes has been recognized. It supports an intimate link between the autoregulatory feedback loop and chromatin remodelling (for a recent review, see [7]).

Mechanisms of Circadian Synchronization

Role of Light and the SCN

While the current view prevails that biological clocks are present in most—if not all—cells of the body, the key importance of the hypothalamus for the integration of circadian rhythms is undisputed. Targeted ablation of hypothalamic nuclei including the SCN, the ventrolateral preoptic, and dorsomedial nucleus clearly indicates that a normal patterning of the sleep–wake cycle, locomotor behaviour, feeding, and the circadian secretion of hormones is no longer observed. Using the same molecular instrumentarium observed in many cells, the SCN appears to be the master regulator of circadian behaviour. This may be based on the extraordinarily tight coupling of rhythms between the cellular clocks in this tissue and the ability of these clocks to respond to light which has been shown for the expression of the clock genes, *Per1* and *Per2*. While SCN clock rhythms—including the rhythmic activation of *Per* genes—persists even in the absence of external light input, this photic regulation of *Per* expression is crucially involved in the alignment of internal and external rhythms termed *entrainment*. While during most of

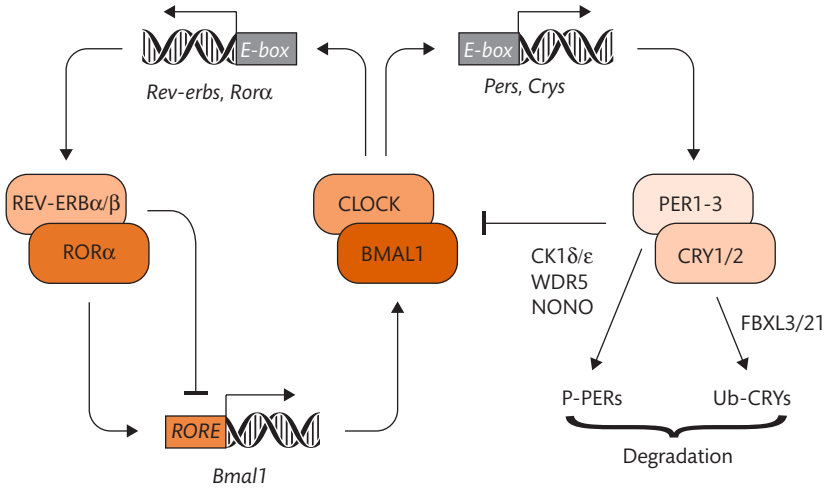


Figure 1.11.1 Current model of the mammalian circadian oscillator. The rhythm generating circuitry is based on a central molecular feedback loop with a positive (CLOCK, BMAL1) and a negative oscillator limb (PERs, CRYs) that is further stabilized via an accessory loop comprising the nuclear orphan receptors REV-ERBs and RORα. NONO, non-POU-domain-containing, octamer binding protein (an RNA-binding protein); RORE, retinoic acid-related orphan receptor response element; WDR5, WD repeat domain 5 (a histone methyltransferase-binding protein); FBXL3, F-box and leucine-rich repeat protein 3 (a ubiquitin-protein transferase) [9].

the *subjective* (i.e. internal) day the SCN clock is insensitive to light, during the first half of the subjective night light exposure elicits a strong induction of *Per1* and *Per2* transcription, while during the second part of the night only *Per2* remains light inducible. This out-of-phase activation of *Per2* mRNA and subsequent protein expression is integrated into the internal oscillator, thus resetting its phase and adjusting internal and external time. Similar responses in *Per* expression have been observed in peripheral tissues clocks, too [8]. However, here they are not part of photic entrainment but transmit internal time information provided by, for example, endocrine signals to local clocks. Both *Per* genes contain glucocorticoid (GC) response elements (*GREs*) in their promotors and, thus—as described further on in this chapter—the diurnal GC release rhythm is an important internal *Zeitgeber* of the circadian system.

The Role of Food in Circadian Entrainment

Studies in nocturnal rodents show that a shift in the availability of food has a strong effect on locomotor activity and may reset the circadian clocks of metabolic tissues such as liver, lung, adipose, and pancreas while having little effect on the SCN pacemaker which stays locked to the light–dark cycle [10]. In rats, such food-entrained rhythms persist under fasting conditions even in SCN-lesioned animals, supporting the notion that a food-entrainable oscillator (FEO) exists distinct from the well-characterized light-dependent entrainment of circadian rhythms. Ablation studies indicate that the dorsal medial hypothalamus may be involved in the food-dependent effects on circadian activity rhythms, but controversial data exist, and the current model favours a more distributed organization of the FEO. An involvement of a gut–brain communication, either via humoral or neural pathways, has been postulated as intestinal clock genes are largely independent of the light–dark cycle but respond rapidly to the timing of food intake similar to other peripheral tissues. In line with this, food-entrained circadian rhythms in physiology such as blood triglyceride levels persist even in the fasting state. The mechanisms of food-induced clock resetting are still poorly understood, but two endocrine factors, insulin and oxyntomodulin, have been implicated in aligning the liver clock and liver physiology with meal timing [11].

Sleep and Circadian Rhythms

The daily pattern of activity versus sleep is the most obvious circadian rhythm in humans. Sleep is controlled by two interacting processes: during the day, sleep propensity is low and this changes during the night (*process c*—for *circadian*). The second process increases sleep drive as a response to the duration of time spent awake. This sleep homeostat is called *process s* (for *sleep drive*). Together, process *c* and *s* control consolidated sleep profiles in humans, with most of the time spent awake during the day and asleep during the night. At the neuroanatomical level sleep–wake states are regulated by a flip-flop mechanism comprising neurons of the basal forebrain, the lateral hypothalamus, and the brainstem. Wakefulness is driven by the noradrenergic ascending arousal system of the brain stem which inhibits sleep-promoting galanergic and GABAergic neurons in the ventrolateral preoptic area (VLPO). The firing rate of these sleep promoting neurons, in turn, is high during sleep under the stimulation of adenosine inhibiting wake-promoting monoaminergic brain stem neurons. Orexins (or hypocretin) expressing neurons in the perifornical area and the lateral and posterior hypothalamus stabilize the wake state by inhibiting the VLPO and stimulating wake-inducing monoaminergic neurons [12]. Fitting to this pattern, it has been shown in animal models that orexin neurons fire during wake state. They are virtually completely inactive during rapid eye movement (REM) and non-REM sleep (NREMS). This complex interaction is highlighted by mutations of orexin receptors in animal models and humans. In both conditions, daytime sleepiness, narcolepsy, and obstructive sleep apnoea may be induced. Narcolepsy, with a prevalence of roughly 1 in 2000, is characterized by excessive daytime sleepiness but also

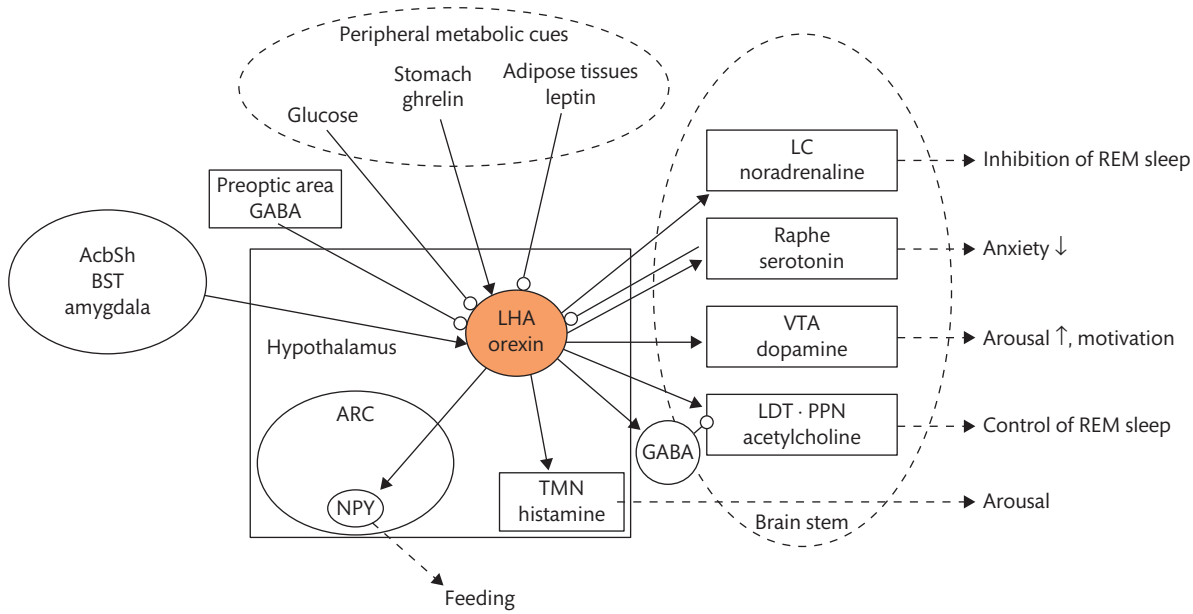


Figure 1.11.2 Orexin-centred view of sleep–wake regulation, energy homoeostasis, arousal, and locomotion [14]. ARC, arcuate nucleus; BST, bed nucleus of the stria terminalis; LC, locus coeruleus; LDT, laterodorsal tegmental nucleus; LHA, lateral hypothalamic area; PPT, pedunculopontine tegmental nucleus; TMN, tuberomammillary nucleus; VTA, ventral tegmental area.

sudden onset of weakness/atonia (*cataplexy*) fitting to orexin effects on muscle tone. Orexin promotes locomotor behaviour and energy homoeostasis. It further stimulates food intake via neuropeptide Y (NPY) but also exerts an inhibitory action on proopiomelanocortin (POMC) neurons to decrease energy expenditure.

The role of orexin highlights the intimate interplay between sleep–wakefulness, locomotion, and central as well as peripheral metabolic control into a sensitively regulated circadian system. It implies that any disturbance of the sleep–wake cycle, of energy homoeostasis, and of the interfering endocrine regulators may lead to substantial changes in other components of this highly integrated system (Figure 1.11.2).

Circadian Endocrine Rhythms and Their Relation to Sleep and Energy Homoeostasis

Multiple hormonal systems show pronounced circadian rhythmicity (see Figure 1.11.3 for examples and the link to other circadian rhythms) [13]. The physiological relevance of these rhythms is only in part elucidated so far. Multidirectional interactions between sleep, energy homoeostasis, and the endocrine system are currently best characterized, and the following section will thus focus on these interactions. As the capacity of endocrine signals to affect energy homoeostasis is reviewed in more detail in other parts of this book, the following will only review the relevance of energy shifts in relation to circadian behaviour and sleep [1].

Ghrelin and Leptin

Ghrelin and leptin are among the most important regulators of energy homoeostasis. Both hormones show a significant diurnal variation in lean subjects tied to circadian rhythms of food intake. Both in lean and obese subjects, a nightly increase of leptin has been shown with peak levels reached at around 02.00 h in the morning hours. Leptin levels decrease thereafter to reach a nadir in the hours between waking and noon. These changes have been viewed as important for the well-known regulatory effects of leptin on energy homoeostasis but also for leptin's action on other hypothalamic/pituitary hormones.

Leptin secretion is linked to another important hormone in appetite control, ghrelin. Ghrelin is released with a marked delay compared to leptin with peak ghrelin levels reached after waking in the morning. This diurnal pattern of ghrelin secretion is restricted to

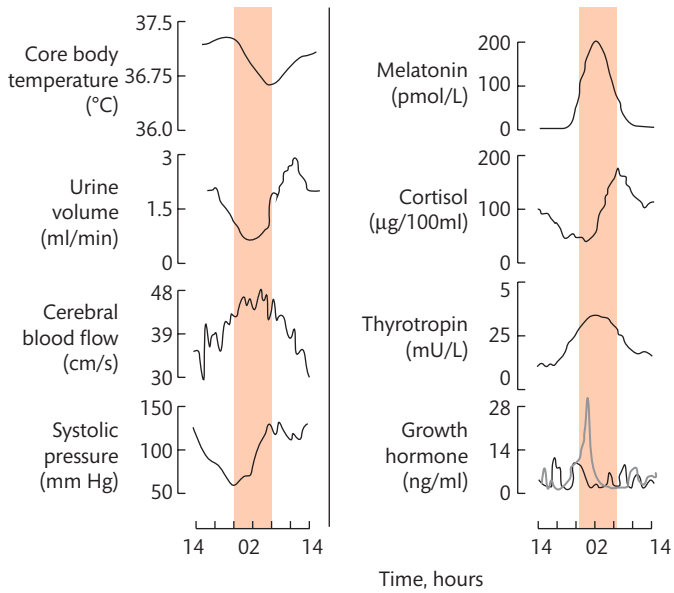


Figure 1.11.3 Examples of prominent endocrine and non-endocrine circadian rhythms [15].

lean persons only. It is interesting that the diurnal ghrelin pattern is lost in obese subjects.

Ghrelin is known to activate orexin neurons and thus plays an important role in the regulation of food searching behaviour and locomotive activity. Ghrelin also has direct effects on the machinery of circadian behaviour by inducing a phase advance and shifting *Per2* mRNA expression in dorsomedial hypothalamic neurons. On the contrary, leptin exerts an inhibitory influence on the firing of orexin neurons and counteracts feeding behaviour by activating POMC. Leptin as well acts directly on clock genes as shown in the example of mice deficient in *Per* or *Cry* in their osteoblasts. At least in this example clear phenotypic changes are found under leptin treatment, indicating that leptin acts on osteoblast proliferation via sympathetic nervous pathways and circadian gene activation.

Growth Hormone

With growth hormone-releasing hormone (GHRH) and somatostatin, ghrelin is an important regulator of growth hormone (GH). GH secretion shows a marked diurnal variation. Using high-frequency sampling techniques in several hundreds of volunteers and patients, it has been shown that GH is released in secretory pulses, and that these pulses form the basis of a circadian pattern. Modulation of frequency and amplitude of these secretory pulses shows circadian variations; a common pattern observed in many hormonal systems.

GH in humans is closely linked to sleep. Quantifying the amount of NREMS revealed that GH is robustly associated with the duration and deepness of NREMS but this relation depends on age and gender. It typically develops at about 3 months of age, reaches a peak in adolescence, and progressively decreases after 30 years of age. Despite the fact that there is a marked sex-related difference with a closer relation in males than in females, the decline in slow-wave sleep parallels the almost complete decline in nightly GH secretion above the age of 50 in both sexes.

These observational studies on a close link between GH and sleep were recently supported by investigations in mutant and transgenic animals. Growth hormone secretion in spontaneous dwarf rats (SDR) is almost completely lost due to a mutation of the *GH* gene. At variance to expectations NREMS is not reduced in these animals, but rather increased during the rest period. This suggests that growth hormone/insulin-like growth factor 1 (IGF-1) is only indirectly responsible for the reduction in spontaneous NREMS. The idea that a major part of this activity is mediated by GHRH-dependent pathways is supported by several rodent models with deletions of the GHRH receptor, such as the *lit/lit* mice or *dw/dw* rats. In both strains, GH and IGF-1 production is greatly decreased along with significantly reduced spontaneous NREMS. As no GHRH action is expected in these animals, chronic growth hormone replacement allows us to dissect GHRH action from growth hormone/IGF-1 responses. Despite successful correction of growth hormone deficiency, growth hormone replacement is unable to stimulate NREMS to normal indicating an important role of GHRH in the regulation of NREMS. This assumption fits to the detection of a circadian and sleep-related variation of GHRH in the hypothalamus. In contrast, REM sleep seems to be directly stimulated by GH secretion.

Adrenal Axis

Endogenous cortisol secretion rises sharply between 02.00 and 04.00 h at night with a peak serum concentration approximately 1 h after wakening. Under non-stressed conditions cortisol is secreted in a robust diurnal pattern overlying 90-min pulsatile patterns, both of which generally reflect the pattern of adrenocorticotropic activity (ACTH). Its high variation between nadir and peak secretions and its high reproducibility allow to use the circadian pattern as a window to evaluate changes in circadian rhythmicity in order to capture pathophysiology. Moreover, GCs have been shown to affect clock gene rhythms in target tissues, thus transmitting circadian time information across the body [16].

In Cushing's syndrome circadian cortisol variations are markedly dampened or even abolished. Whereas morning cortisol may still be within the normal range, typically midnight cortisol levels are increased. Similarly, circadian secretion is altered in normal ageing with an increased nightly secretion. This is a mild alteration whereas in depressive illnesses, as well as in chronic alcohol abuse, more pronounced shifts in circadian pattern reminiscent of Cushing's syndrome are observed. In cortisol deficiency stimulation of the entire corticotropin-releasing hormone (CRH)–ACTH–cortisol axis may be achieved by administration of the 11β-hydroxylase antagonist metyrapone, which blocks the conversion of 11-deoxycortisol to cortisol. The reduced negative feedback inhibition of cortisol on CRH and ACTH secretion can be used diagnostically in partial pituitary insufficiency. It is, however, highly dependent on circadian timing. Deoxycortisol has been shown to be maximally induced when the drug is applied at 20.00 h. This stimulation was significantly higher than after administration during the morning hours. Similarly, suppression of ACTH secretion by cortisol or synthetic analogues depends on the timing of their administration. Maximal inhibitory effects on ACTH secretion are observed just prior to the endogenous nightly rise in ACTH secretion. These effects have implications for the timing of corticosteroid treatment.

For physiological replacement therapy a slow-release preparation has recently been developed which is able to mimic the nightly cortisol increase. Application of the preparation when going to sleep will release peak cortisol levels at the physiological peak secretion time in the early morning hours. It is hoped, but still remains to be proven, that these promising preparations will improve the impaired quality of life of patients with Addison's disease. In pharmacotherapy with GCs, it is evident that the effectiveness depends on the timing. Nightly application improves the effect/dose, but side effects are higher as well. Bearing in mind interindividual variations in the sensitivity to GCs, there is no present consensus on which minimal dose is effective at which time of the day. In addition, this may vary due to disease specific factors.

Thyroid Hormone Axis

Thyrotropin (thyroid-stimulating hormone (TSH)) exhibits a marked circadian rhythm that governs a similarly phased—though far less developed—24-h rhythm of free triiodothyronine. Data on circadian thyroxine rhythms are less clear, probably due to the long half-life of several hours of the circulating hormone. There are early observations suggesting a light–dark cycle in total thyroxine but more recent data on free thyroxine do not confirm this. There are no data on the influence of light on TSH secretion but

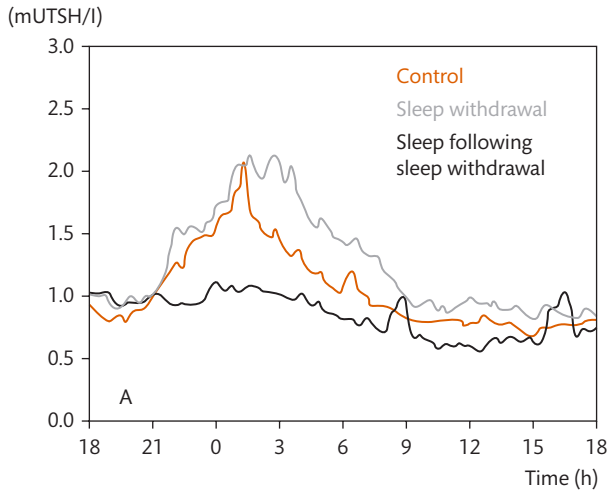


Figure 1.11.4 Effects of sleep modulation on TSH secretion in healthy volunteers. Comparing normal sleep to acute sleep withdrawal and to sleep in the night following sleep withdrawal [17].

Adapted with permission from Brabant G, Prank K, Ranft U et al. Physiological regulation of circadian and pulsatile thyrotropin secretion in normal man and woman. *J Clin Endocrinol Metab* 1990;**70**:403–9. Copyright © 1990, Oxford University Press [ref 17].

the important impact of sleep on TSH has been well investigated. Sleep withdrawal induces an acute increase and prolonged release of nocturnal TSH secretion (**Figure 1.11.4**) independent of total thyroid hormone levels. In contrast, TSH is almost completely suppressed if volunteers sleep significantly more and deeper in the night following a night of sleep withdrawal. This recovery of hormonal changes in acute total sleep deprivation is observed for other hormonal systems as well and raises the possibility that chronic sleep loss may result in long-term adverse endocrine effects via alterations in the circadian rhythmicity of hormone release. A direct link to energy homoeostasis is currently elusive. Short-term activation of the thyroid axis as in acute sleep withdrawal may, however, be linked to an activation of energy stores via the thyroid hormone system, an assumption fitting to data on fasting. Decreased energy availability following a 3-day fast almost completely suppresses circadian TSH release. Effects on the sympathetic nervous system are clearly important in this context but detailed studies are missing to date.

Melatonin

Melatonin, which is exclusively derived from the pineal gland, is secreted with a strong circadian rhythm peaking during the dark phase. Ganglionic photoreceptor cells in the retina integrate information on the light–dark cycle and signal via the retino-hypothalamic pathway to the SCN where duration, phase, and amplitude of melatonin hormone production are encoded. Nocturnal light exposure suppresses melatonin secretion, with blue light in the range 460–470 nm having the most pronounced effect (acting through retinal ganglionic melanopsin (OPN4) photoreceptors). Circulating melatonin levels show high interindividual variability which presumably is genetically determined. An age-related blunting of circadian melatonin rhythms has been suggested, but more recent data indicate that melatonin rhythms are preserved even in old age, albeit with large interindividual variations [18].

Melatonin exerts its physiological actions through G-protein-coupled specific cell membrane receptors, MT1 and MT2. The functions of the subtypes differ and are not restricted to sleep and circadian behaviour where MT1 receptor decreases neuronal firing rates and MT2 receptor regulates phase shifts. Like cortisol, melatonin is a major regulator of the circadian rhythm in peripheral tissues and of core body temperature in humans. This pattern is linked to sleep. In normal adults, the deepest level of sleep and the lowest core body temperature are reached simultaneously.

Without external light cues as in totally blind patients with a complete loss of light recognition entrainment of endogenous rhythms is lost and patients free-run with their own endogenous period (Non-24). Synchronized endogenous circadian rhythms are important for a normal quality of life. Totally blind persons lose this synchrony and exhibit cyclical sleep loss associated with daytime sleepiness and mood swings.

Exogenous melatonin affects sleep regulation largely through a phase-resetting mechanism. By its capability to readjust disturbed circadian rhythms to their correct phase position, melatonin strengthens endogenous rhythms, thus decreasing daytime sleepiness and restoring sleep quality. Circadian rhythm sleep disorders, either advanced or delayed sleep phase syndrome, have been successfully treated with melatonin [19]. A common denominator of these conditions is the loss of coordination between endogenous rhythms. Jet lag induced by a transmeridian flight across several time zones is a well-known but transient condition of such loss of entrainment as a mismatch occurs between the endogenous circadian rhythms and the new environmental light–dark cycle. Endogenous rhythms shift in the direction of the flight with a phase advance on eastbound flights but a phase delay when flying westwards. Symptoms typically include a disturbed night-time sleep, impaired daytime alertness and performance, irritability, distress, and appetite changes, along with other physical symptoms such as disorientation, fatigue, gastrointestinal disturbances, and lightheadedness. Pharmacological modification of melatonin secretion has been used to synchronize endogenous rhythms in jetlag and in shift workers. The convincing positive data in the totally blind on a coordinating role of melatonin have recently been paralleled in healthy volunteers treated with a melatonin agonist. The dose-dependent effect on sleep propensity supports an important coordinating function of melatonin on the sleep–wake cycle and on other circadian rhythms. To understand melatonin action in normal physiology, further mechanisms such as the activation of the sympathetic nerve system, which suppresses melatonin secretion from the pineal gland, may play a crucial role. Light, even dim light at night, leads to a suppression of melatonin and feedback on other rhythms. Recent data on melatonin secretion in postmenopausal women highlight this complex relation. Absolute 24-h melatonin secretion is enhanced in depressed postmenopausal women. In addition, the timing of melatonin secretion is altered, showing a delayed morning offset. This longer melatonin secretion window fits to results in seasonal affective disorders where—due to weak and noisy photic *Zeitgeber* conditions during winter—a dissociation of endogenous rhythms is observed. These patients also often show an increased melatonin secretion during the winter months. In both groups mood disturbance and sleep are improved by bright (blue-) light therapy [20].

Altered Sleep, Circadian Rhythmicity, and Metabolism

Increased melatonin secretion may further impact on the risk of insulin resistance and diabetes mellitus. Recent data from three independent groups reveal that a polymorphism of the melatonin receptor 1B, which leads to chronic overactivity of intracellular melatonin dependent signalling pathways, is an independent and strong risk factor for glucose intolerance and type-2 diabetes mellitus. The mechanism behind this is not entirely clarified but inhibition of insulin secretion from pancreatic β cells and a negative impact on incretins seem to cooperate.

Glucose tolerance, which critically depends on the ability of the pancreatic β cell to respond to a given glucose challenge, varies over the day in healthy individuals. It is much lower in the evening than in the morning. There is a further increase in plasma glucose when tested in the middle of the night, suggesting minimal glucose tolerance during sleep. Whereas reduced glucose tolerance in the evening hours is attributed to both a reduction in insulin sensitivity and a reduced insulin secretory response to glucose, the further deterioration of glucose tolerance during the night depends on sleep-related processes to maintain stable glucose levels during the extended overnight fast. During NREMS glucose utilization is lowest; it increases during REM sleep and is highest in the wake period. The underlying multifactorial causes of this circadian change in glucose tolerance are only partly unravelled. Insulin sensitivity decreases in the evening predominantly due to a decreased pancreatic insulin secretory response to glucose. Data from transgenic mice suggest that pancreatic insulin production itself is under control of the local clock gene machinery [21]. Melatonin-mediated effects may further contribute to this regulation. Glucose production and utilization fall in association with sleep during the first half of the night and increase again in the latter part. Insulin-dependent and -independent glucose disposal is reduced during sleep. In parallel, GH secretion is increased with the initiation of slow-wave sleep, cortisol is inhibited, sympathetic nerve activity is decreased, and vagal tone stimulated.

Moderate alteration of night sleep with a reduction to only 4 h/night over a period of 6 nights has been shown experimentally to profoundly affect energy metabolism. It acutely reduces insulin release predominantly via an increased sympathetic outflow and decreases peripheral insulin sensitivity on several levels. Importantly, counteractive hormone release is activated with augmented nightly GH, TSH, and cortisol secretion, and also with a higher level of cytokines and inflammatory markers. It is not surprising that testing for insulin resistance in such a state of sleeplessness revealed a metabolic state well comparable with metabolic syndrome and prediabetes. Similar data have subsequently been obtained in subjects where selective suppression of slow-wave sleep decreased the quality but not the duration of sleep.

Chronic sleep reduction as investigated in obstructive sleep apnoea is further associated with a dysregulation of the neuroendocrine control of appetite. A combined alteration of ghrelin, orexin, and leptin secretion is part of the pathomechanism leading to excessive food intake, decreased energy expenditure and, as recent data indicate, to hypertension (see Figure 1.11.5 for schematic integration of mechanisms). There is further evidence that metabolic changes in the polycystic ovary syndrome are a result of obstructive sleep apnoea. Experimental studies on partial sleep loss and their impact on energy conservation parallel epidemiological findings on the greatly increased risk of obesity and diabetes mellitus in societies. It is tempting to speculate, but remains to be proven, that sleep curtailment by modern lifestyle changes is a primary force behind the adverse metabolic effects via its impact on diurnal endocrine regulation. Mouse and human data suggest that even a short episode of mistimed and shortened sleep may profoundly affect circadian clocks in blood and peripheral tissues [22].

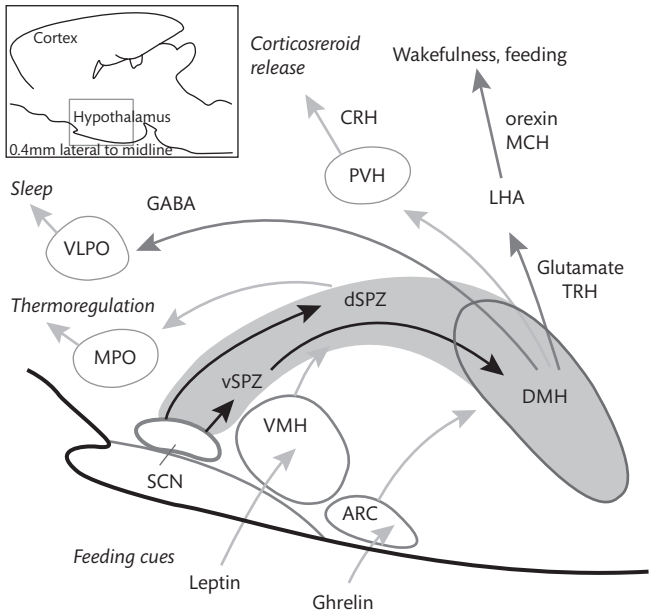


Figure 1.11.5 Integration of endocrine signals with sleep and food regulating hypothalamic circuits. ARC, arcuate nucleus; DMH, dorsomedial hypothalamus; dSPZ, dorsal subparaventricular zone; GABA, γ–aminobutyric acid; LHA, lateral hypothalamic area; MCH, melanin-concentrating hormone; MPO, medial preoptic area; PVH, periventricular hypothalamus; SCN, suprachiasmatic nucleus; TRH, thyrotropin-releasing hormone; VLPO, ventrolateral preoptic area; VMH, ventromedial hypothalamus; vSPZ, ventral subparaventricular zone [12].

Conclusion

In summary, these examples clearly demonstrate a powerful circadian regulation of the endocrine and metabolic system interlinked with sleep. Evidence is accumulating that the common curtailment of normal sleep in modern society has important consequences for metabolic and endocrine functions. Data on shift workers who most frequently experience gastrointestinal disturbances support the importance of food-entrained rhythms in addition to the light–dark cycle for the timing of many endocrine rhythms and sleep-associated cycles. Constant violation of these patterns may lead to detrimental effects. The example of treatment with melatonin and melatonin agonists suggests that a better understanding of the pathophysiology of the circadian system may help to develop new means to endocrinologically modulate these cycles for the benefit of patients.

REFERENCES

1. Tsang AH, Astiz M, Friedrichs M, *et al.* Endocrine regulation of circadian physiology. *J Endocrinol*, 2016; **230**: R1–11.
2. Zhang R, Lahens NF, Ballance HI, *et al.* A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci U S A*, 2014; **111**: 16219–24.
3. Aschoff J. Circadian rhythms in man. *Science*, 1965; **148**: 1427–32.
4. Phillips ML. Circadian rhythms: of owls, larks and alarm clocks. *Nature*, 2009; **458**: 142–4.
5. Konopka RJ, Benzer S. Clock mutants of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*, 1971; **68**: 2112–6.
6. Vitaterna MH, King DP, Chang AM, *et al.* Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science*, 1994; **264**: 719–25.
7. Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet*, 2017; **18**: 164–79.
8. Husse J, Eichele G, Oster H. Synchronization of the mammalian circadian timing system: light can control peripheral clocks independently of the SCN clock: alternate routes of entrainment optimize the alignment of the body's circadian clock network with external time. *Bioessays*, 2015; **37**: 1119–28.
9. Takahashi JS, Hong HK, Ko CH, *et al.* The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet*, 2008; **9**: 764–75.
10. Damiola F, Le Minh N, Preitner N, *et al.* Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev*, 2000; **14**: 2950–61.
11. Atger F, Mauvoisin D, Weger B, *et al.* Regulation of mammalian physiology by interconnected circadian and feeding rhythms. *Front Endocrinol (Lausanne)*, 2017; **8**: 42.
12. Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature*, 2005; **437**: 1257–63.
13. Angelousi A, Kassi E, Nasiri-Ansari N, *et al.* Clock genes alterations and endocrine disorders. *Eur J Clin Invest*, 2018; **48**: e12927.
14. Ohno K, Sakurai T. Orexin neuronal circuitry: role in the regulation of sleep and wakefulness. *Front Neuroendocrinol*, 2008; **29**: 70–87.
15. Maywood ES, O'Neill JS, Chesham JE, *et al.* Minireview: the circadian clockwork of the suprachiasmatic nuclei--analysis of a cellular oscillator that drives endocrine rhythms. *Endocrinology*, 2007; **148**: 5624–34.
16. Balsalobre A, Brown SA, Marcacci L, *et al.* Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science*, 2000; **289**: 2344–7.
17. Brabant G, Prank K, Ranft U, *et al.* Physiological regulation of circadian and pulsatile thyrotropin secretion in normal man and woman. *J Clin Endocrinol Metab*, 1990; **70**: 403–9.
18. Golombek DA, Pandi-Perumal SR, Brown GM, *et al.* Some implications of melatonin use in chronopharmacology of insomnia. *Eur J Pharmacol*, 2015; **762**: 42–8.
19. Rajaratnam SM, Polymeropoulos MH, Fisher DM, *et al.* Melatonin agonist tasimelteon (VEC-162) for transient insomnia after sleep-time shift: two randomised controlled multicentre trials. *Lancet*, 2009; **373**: 482–91.
20. Parry BL, Meliska CJ, Sorenson DL, *et al.* Increased melatonin and delayed offset in menopausal depression: role of years past menopause, follicle-stimulating hormone, sleep end time, and body mass index. *J Clin Endocrinol Metab*, 2008; **93**: 54–60.
21. Perelis M, Marcheva B, Ramsey KM, *et al.* Pancreatic β cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. *Science*, 2015; **350**: aac4250.
22. Archer SN, Oster H. How sleep and wakefulness influence circadian rhythmicity: effects of insufficient and mistimed sleep on the animal and human transcriptome. *J Sleep Res*, 2015; **24**: 476–93.

1.12

Principles of Hormone Replacement

Richard Ross

Introduction

The aim of hormone replacement is to replace the missing physiological effects of a deficient hormone. In principle this may best be achieved by replicating the physiological profile of that hormone in the bloodstream. This is easier said than done as there are many challenges to be met in order to achieve this, from defining the physiological profile of the hormone to designing and testing therapeutic formulations that can replicate that profile. This involves, accurately measuring the bioactive hormone, describing its actions, identifying biomarkers of those actions, accurately measuring those biomarkers, generating formulations of hormone or hormone mimetic drug that replace the physiological concentration and rhythm, providing a route of administration that yields bioavailability, and then demonstrating in clinical trials the short and long-term benefits of hormone replacement.

It is remarkable that in many cases providing a daily bolus of hormone is lifesaving even when it does not replicate the physiological profile, as was evident with the introduction of insulin and glucocorticoids in the last century. Thus, a daily subcutaneous injection of insulin can prevent ketoacidosis and death despite the fact that insulin in health is secreted from the pancreas into the portal circulation in a pulsatile manner under tight control of multiple nutrient sensing mechanisms, and coordinated with the secretion of glucagon. Hormone deficiencies are generally chronic diseases requiring lifelong therapy, often starting early in life. In the short term non-physiological daily administration may prevent death, but we now recognize that replacement needs to replicate physiology more faithfully to prevent poor long-term health outcomes. This is well demonstrated in type 1 diabetes mellitus, where poor glucose control over time results in large and small blood vessel disease with resultant ischaemic heart disease, neuropathy, retinopathy, and nephropathy. Thus, the emphasis in the management of diabetes mellitus in this century is on reconstituting physiology by providing basal insulin levels supplemented by mealtime boluses, or combining insulin with glucagon infusions, all with intensive glucose monitoring. Similarly, with glucocorticoid replacement in adrenal insufficiency we can prevent adrenal crises through administration of large doses of any glucocorticoid, but chronic excess glucocorticoid over time is associated with all the complications of Cushing's syndrome. This chapter explores the challenges associated with hormone replacement and through this exemplifies the principles of hormone replacement.

Physiology of Hormones

Understanding the physiological regulation and action of a hormone is essential to understanding how to provide its replacement. This is exemplified by gonadotropin-releasing hormone (GnRH), which, if given as a continuous infusion, results in downregulation of the pituitary gonadal axis. This is the opposite of what occurs in nature, where GnRH is secreted in a pulsatile manner, with the frequency and magnitude of those pulses regulating puberty and fertility (see Chapter 8.1.2). Characterizing hormone physiology is complex and requires not only an understanding of the synthesis and secretion of the hormone and its total concentration in the blood stream, but also an understanding of its binding proteins, which tissues it acts upon, what receptor(s) it acts through and its metabolism and excretion.

Some hormones are released in pulses, like GnRH, and the pulse and trough are important to the biological action, whereas parathyroid hormone is secreted at a fairly constant level. When increased in primary hyperparathyroidism parathyroid hormone causes osteoporosis, but, intriguingly, when given in pulses it has an anabolic action on bone (see Chapter 4.1). Some hormones are secreted in pulses superimposed on an underlying rhythm, exemplified by the hypothalamic–pituitary–adrenal (HPA) axis: cortisol production is controlled by the pulsatile release of adrenocorticotropic hormone (ACTH) from the pituitary, and shows an overall circadian rhythm underpinned by an ultradian rhythm. Both of these rhythms may be important in its biological actions [1].

Table 1.12.1 Current practice in hormone replacement

Hormone	Physiological profile	Replacement	Biomarkers	Future prospects
Growth hormone	3–4 hourly pulses; greatest at night	Daily s/c injection given in evening	Growth in childhood and IGF-I in adults	Long-acting formulations are in development but don't replicate physiology and therefore efficacy and safety important to demonstrate
Thyroid	T4 and T3 secretion from thyroid with circadian rhythm of T3 and peripheral conversion of T4 to T3	Thyroxine monotherapy	TSH	Development of new tissue specific biomarkers and T3 formulations that can provide the normal ratio of T4 to T3
Parathyroid	Constant secretion of parathyroid hormone (PTH) modulated by calcium level	Usually calcium and vitamin D supplementation. Subcutaneous PTH (1–84) recently licenced	Calcium level	Development of long-acting PTH supplementation as current treatment results in fluctuating calcium levels
Cortisol	Circadian cortisol rhythm with underlying ultradian rhythm	Hydrocortisone twice to thrice daily or modified release once daily	None	There is a need for tissue-specific biomarkers. There is a modified-release formulation in development that replaces the circadian overnight cortisol levels and work with s.c. pumps looking at the ultradian rhythm
Insulin	Baseline levels with postprandial surges	Short- and long-acting insulins	HbA1c	See Chapter 15.5.3
Oestrogen and progesterone	Gradually rise through puberty; varying levels during menstrual cycle	Oral and transdermal preparations; combined preparations	Menstrual cycle and dual energy X-ray absorptiometry (DEXA) scan for bone health	There is a need for biomarkers and to determine the optimal route and dose of replacement at different ages
Testosterone	Gradually rises in puberty; circadian rhythm	Parenteral or transdermal preparations	None	There is a need for biomarkers to determine optimal replacement doses and when to replace

Many hormones such as thyroid hormones and cortisol have binding proteins which influence both bioavailability and hormone measurement. Women taking oestrogens or when pregnant have a high plasma cortisol binding protein concentration and therefore have a high total cortisol level which needs to be recognized when assessing for adrenal insufficiency or measuring absorption of hydrocortisone.

Hormone receptors are variably expressed in different tissues and there may be different receptors for the same hormone in different tissues that can influence biological output of hormone exposure. As an example, there are two thyroid hormone receptors, THRα and THRβ. These are variably expressed with THRβ being more highly expressed in the hypothalamus and pituitary, mediating inhibition of thyroid-stimulating hormone (TSH), and so TSH alone may not be a suitable biomarker for thyroid action in other tissues [2]. Within cells there may be conversion of prohormones to either active or inactive forms, as seen in the deiodination of T4 to biologically active T3 or conversion of active cortisol to inactive cortisone by 11β-hydroxysteroid dehydrogenase. Finally, clearance of hormones varies between hormones in rate and site, and may feature renal clearance, hepatic metabolism, receptor-mediated tissue uptake, and/or proteolysis, all of which may influence the pharmacokinetics of a drug formulation.

Biomarkers of Hormone Replacement

A biomarker is defined as a characteristic that is objectively measured and evaluated as a surrogate indicator of a normal biological process, a pathogenic process, or a pharmacologic response to a therapeutic intervention [3]. Biomarkers are essential tools for monitoring hormone replacement and for developing new therapies for hormone replacement. Clinical endpoints, capturing how a patient feels, functions and survives, are ultimately the key objectives for hormone replacement, but these are usually not easily measurable in individual patients and are challenging to measure in clinical trials. In such trials we thus frequently depend on surrogate endpoints, or biomarkers that can accurately predict beneficial or harmful clinical outcomes.

Haemoglobin A1c (HbA1c) is an example of a validated surrogate endpoint for microvascular complication risk associated with diabetes mellitus, and has been used as the basis for approval of drugs intended to treat diabetes mellitus as well as being used in the management of individual patients. We have biomarkers for the actions of most hormones; however, few are validated as surrogate endpoints. Serum IGF-I is widely used as a biomarker in growth hormone (GH) replacement but correlates only weakly with clinical endpoints in GH treatment, although normalization of IGF-I is related to an improvement in mortality in patients with acromegaly [4]. TSH is routinely used as a biomarker of thyroid hormone replacement for the management of individual patients as well as in clinical trials. Nevertheless, in trials of T4 monotherapy at doses that normalized TSH in overt, primary hypothyroidism, not all systemic biological markers of thyroid hormone signalling were normalized, including serum cholesterol levels [2]. Identifying biomarkers that are validated endpoints is essential for management of hormone replacement both in the clinic and for developing optimized therapies and there is a need for new biomarkers in most fields of endocrinology.

Delivery of Hormonal Therapies

The route of administration for hormones depends on their intrinsic properties. Proteins such as insulin and GH have to be given

parenterally because of proteolysis in the gut. Most peptide hormones such as DDAVP (desmopressin) are also broken down in the gut, although when given at high dose sufficient hormone maybe absorbed to be biologically active. In contrast, amino acid-based hormones such thyroxine and steroid hormones such as hydrocortisone are well absorbed from the gut. The gonadal steroids testosterone and oestradiol, too, are well absorbed from the gut but undergo first pass metabolism in the liver, commonly necessitating parenteral or transdermal administration.

The route of administration will often dictate constraints in replicating physiology. Sophisticated subcutaneous pump technology performs well in diabetes, but this is generally impractical for most hormone replacement. GH is secreted in pulses approximately every 3 hours with the largest pulse occurring during deep sleep, however single daily subcutaneous GH injections given in the evening have proven effective in promoting growth in children. There are long-acting formulations of GH in development with the potential to offer weekly or less frequent dosing options. These GH formulations don't truly replicate endogenous GH profiles, but they do promote growth. Important to their development will be demonstration of a good safety profile [5]. It is also possible to modify release of drugs in the gut and in so doing to delay and/or sustain release of hormone, depending on what part of the gut the hormone is absorbed from. This approach has been used with hydrocortisone to mirror more faithfully the daily rhythm of cortisol concentration. There is now a dual-release formulation available that provides for once daily hydrocortisone replacement [6], and a delayed and sustained absorption formulation in development that replaces the overnight rise in cortisol and improved biochemical control of patients with congenital adrenal hyperplasia [7].

Summary

The last century identified most of the endocrine hormones, which can now be replaced when deficient; however, the challenge of the twenty-first century is to optimize replacement (see Table 1.12.1). The guiding principle in hormone replacement is replicating the natural levels and rhythms of hormones at different ages but this requires a good understanding of physiology. There is a need for better biomarkers of hormone actions and using these to develop new ways to deliver hormone replacement tailored to the individual. This chapter discusses current approaches to this problem.

REFERENCES

1. Porter J, Blair J, Ross RJ. Is physiological glucocorticoid replacement important in children? *Arch Dis Child*, 2017; **102**: 199–205.
2. McAninch EA, Rajan KB, Miller CH, Bianco AC. Systemic thyroid hormone status during levothyroxine therapy in hypothyroidism: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2018; **103**: 4533–42.
3. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Silver Spring, MD, Food and Drug Administration, 2016.
4. Johannsson G, Bidlingmaier M, Biller BMK, *et al.* Growth Hormone Research Society perspective on biomarkers of GH action in children and adults. *Endocr Connect*, 2018; **7**: R126–34.
5. Christiansen JS, Backeljauw PF, Bidlingmaier M, *et al.* Growth Hormone Research Society perspective on the development of long-acting growth hormone preparations. *Eur J Endocrinol*, 2016; **174**: C1–8.
6. Johannsson G, Nilsson AG, Bergthorsdottir R, *et al*. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab*, 2012; **97**: 473–81
7. Mallappa A, Sinaii N, Kumar P, *et al*. A phase 2 study of Chronocort(r), a modified-release formulation of hydrocortisone, in the treatment of adults with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2015: 100 (3), 1137–45.

1.13

Prevention in Endocrinology

Jonathan Valabhji and Rochan Agha-Jaffar

Introduction

At its inception in 1948, the World Health Organization defined health as 'a state of complete physical, mental, and social well-being and not merely the absence of disease and infirmity'. The widely accepted goals of healthcare systems are to promote and preserve health, to restore health when it is affected and to minimize suffering. The concept of disease prevention is embedded within these goals.

When considering disease prevention within the field of endocrinology, there has been a clear shift in focus reflecting changing epidemiology. Iodine deficiency disorders and congenital hypothyroidism were examples of conditions, which when left untreated had severe consequences and posed considerable public health challenges. However, the introduction of appropriate screening strategies and health policies has successfully increased early recognition and treatment, reducing the incidence of associated adverse outcomes. While the challenges do still exist, the epidemics of diabetes and obesity now observed have superseded both these and other preventable non-communicable diseases worldwide.

Disease prevention can take many forms ranging from public health campaigns addressing lifestyle factors to targeted interventions in at risk groups. Prevention can be classified as follows [1]:

1. **Primary prevention.** This focuses on improving health in populations' (e.g. salt iodization) programmes to prevent iodine deficiency disorders.
2. **Secondary prevention**. Here, disease development in at risk individuals is targeted (e.g. lifestyle modification in overweight individuals to prevent type 2 diabetes). The strength of this particular form of prevention is that the intervention can be matched to the needs of the individual allowing focused resource delivery.
3. **Tertiary prevention.** This aims to reduce the risk of complications associated with a disease in affected individuals (e.g. good glycaemic control to prevent microvascular complications in individuals affected by type 1 or type 2 diabetes).
4. **Quaternary prevention.** The most recent of the concepts relates to the actions taken to identify a group at risk of overmedicalization, thus protecting them from unnecessary invasive medical interventions and providing them with ethically acceptable care.

Screening

Effective secondary and tertiary prevention relies on practical screening strategies and the identification of clear modifiable factors, which when addressed, can mitigate the risk of either developing a disease or improving its associated complications.

The World Health Organization (WHO) defines screening as the 'presumptive identification of unrecognized disease or defects by means of tests, examinations, or other procedures that can be applied rapidly'. In an early series of papers on the subject, Rose and Barker stated that to ensure an effective screening programme, these three core principles should be addressed [2]:

1. Earlier treatment should improve the prognosis
2. The screening test should be valid and repeatable
3. There should be a high yield to the screening service

The WHO proposed these further recommendations for screening strategies in a clinical policy guideline [3]:

- The condition being screened for should be an important one
- Facilities for diagnosing and treating the condition should be available
- There should be an accepted treatment for those who ultimately develop the disease

- There should be a recognizable latent or early symptomatic stage
- The natural history of this condition including the progression from latent to declared disease should be adequately understood
- There should be a suitable test to screen for the condition and this test should be acceptable to the population
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure in medical care as a whole
- Case-finding should be on a continuing process and not a 'once and for all' project

When deciding on appropriate screening programmes, the sensitivity and specificity of the tests need to be considered together with the implications of false positive and false-negative results.

This chapter will explore the principle preventable diseases and their associated complications within the field of endocrinology as well as the strategies that have been employed to combat them. Two areas will be described in which public programmes of primary prevention have been implemented worldwide, the first relating to iodine deficiency disorders and the second to learning difficulties due to congenital hypothyroidism. The global epidemic of type 2 diabetes and its prevention will then be discussed, and potential prevention opportunities relating to gestational diabetes, type 1 diabetes, and osteoporosis will also be described.

Iodine Deficiency Disorders

Iodine deficiency is estimated to affect two thousand million people worldwide. Endemic goitre had been a well-recognized complication of iodine deficiency as early as the start of the twentieth century. However, it was not until much later that the links between iodine deficiency and mental as well as physical impairment were established, following which the term 'iodine deficiency disorders' was introduced (Table 1.13.1).

Dietary iodine is readily absorbed from the gastrointestinal tract, reaching the circulation in the form of iodide [4]. The latter is converted back to iodine in the thyroid gland and concentrated within the follicular cell. Iodine is an integral component of T4 and T3. A low dietary intake of iodine results in adaptive processes within the thyroid gland, which are mediated by the enhanced secretion of thyroid stimulating hormone (TSH) from the anterior pituitary gland. This in turn maintains T3 status as well as reducing the renal clearance of iodide. Despite these adaptive processes, iodine deficiency in the chronic stages results in total thyroid gland iodine depletion and hyperplasia of thyroid epithelial cells precipitating the development of goitre.

Table 1.13.1 Consequences of iodine deficiency in different age groups

Age group	Manifestations of iodine deficiency
Fetus	Spontaneous early pregnancy loss, congenital anomalies, stillbirth
Neonate	Cretinism (neurodevelopmental delay including mutism, squint, spastic diplegia), hypothyroidism, and short stature. Infant mortality
Childhood	Physical impairment, mental impairment
All age groups	Hypothyroidism, goitre, iodine-induced hyperthyroidism, increased susceptibility to nuclear radiation

In 1978, a study in the Chinese village of Jixian demonstrated that iodine supplementation in an iodine deficient population resulted in the following benefits over an eight year period: reduction in endemic goitre rates in schoolchildren (80% in 1978 versus 4.5% in 1986), reduction in incidence of cretinism (11% presupplementation versus 0% in 1986) and improvement in school performance (both school ranking and pupil failure rate) [5]. This, together with further similar studies provided evidence that mental impairment secondary to iodine deficiency could be prevented. The implications for translating this to a population level were therefore numerous particularly when considering the potential beneficial impact on socioeconomic development.

While iodine supplementation had been commonplace in countries such as the United States and Switzerland, it was not until the 1990s that a turning point in the global efforts to combat iodine deficiency was observed. This first involved concerted efforts to assess the prevalence of iodine deficiency and countries with a total goitre rate exceeding 10% were classed as iodine deficient. One hundred and ten countries fell into this remit exposing 1570 million people, equating to approximately 30% of the global population, to the risks of developing iodine deficiency disorders.

In 1990, the International Council for Control of Iodine Deficiency Disorders, UNICEF, and the World Health Organization collaborated on guidelines to address iodine deficiency, which was followed by the WHO recommending universal salt iodization (USI) programmes to combat the associated disorders. USI programmes were thought to represent the most cost-effective methodology by which to do so with the price per person per annum estimated at $0.002–0.006. The proportion of the global population consuming iodized salt increased from less than 20% in 1990 to approximately 70% in 2000. By 2004, WHO indicated that 54 countries were iodine deficient (relative to the 110 prior to the programme) indicating that the strategy was effective.

Longitudinal studies have now demonstrated that the implementation of USI programmes have improved endemic goitre rates, One example of this is the Chinese province of Hebei, where the rate has improved from 11.3% to 2.7% in school children aged 8–10 years over a 10-year period. Similar reductions in the incidence of cretinism have been observed.

However, not all regions required the implementation of USI programmes to improve iodine deficiency rate. Changes in farming practice have almost led to an accidental increase in iodine intake among the general population within certain parts of Europe and Australasia. Using the United Kingdom as an example, the endemic goitre rate was estimated at 30% in school children in the 1920s [6]. Changes to farming practice including provision of iodine-enriched feed to cattle to improve their reproductive performance led to an associated increase in the iodine content of milk. Subsequent to this, iodine deficiency rates improved across the country, which in turn was evidenced by both reductions in the endemic goitre rate (equating to approximately 11%) and increases in median urinary iodine concentrations, a surrogate marker of iodine intake. While cross-sectional surveys have suggested that iodine deficiency rates could once again be increasing in certain parts of the world, the true

prevalence, severity, and indeed long-term impact of this, has yet to be established.

Congenital Hypothyroidism

The screening programme for congenital hypothyroidism is well organized in many parts of the world and represents another important primary prevention programme within endocrinology. The condition affects approximately 1 in 3000 to 1 in 4000 neonates rendering it one of the most frequent congenital endocrine disorders. Untreated, congenital hypothyroidism results in severe intellectual disability, neurodevelopmental delay and growth impairment [7]. The majority of cases (80%) occur as a result of developmental defects within the thyroid gland itself otherwise known as thyroid agenesis or dysgenesis. In the remaining 20%, a normal or enlarged thyroid gland is found and the condition instead occurs due to disruptions in thyroid hormone biosynthesis.

The finding that normal mental development in a child with congenital hypothyroidism was contingent on both a timely diagnosis and initiation of treatment within the first three months of life prompted researchers to search for an effective screening strategy. However, screening only became viable when sensitive radioimmunoassays for TSH and T4 were introduced. The first two pilot screening studies were published in the 1970s. Subsequent to this, routine testing for congenital hypothyroidism has been introduced as part of neonatal screening programmes in many parts of the world and testing is generally undertaken between the first 48 hours and 4 days of life.

Type 2 Diabetes

The WHO has estimated that globally, 422 million adults were living with diabetes in 2014, compared to 108 million in 1980. Separate global estimates of diabetes prevalence for type 1 and type 2 do not exist. However, the majority of people with diabetes have type 2 diabetes, and policies aligned to attenuate the current trajectories of increasing incidence and prevalence of type 2 diabetes, as well as the associated complications of type 2 diabetes, are of global significance.

The aetiology of type 2 diabetes is multifactorial, with environmental and genetic risk factors, although to date only around 5% of the variability in predisposition to type 2 diabetes has been linked to specific genes. Nevertheless, family history and ethnicity are important non-modifiable risk factors, as is age. Modifiable risk factors principally involve the lifestyle risk factors relating to weight gain, poor quality nutrition, and reduced physical activity, although socioeconomic factors, with people from more deprived communities at significantly greater risk, may also be described as modifiable.

Type 2 diabetes is a disease to which all four potential prevention strategies can be applied:

1. **Primary prevention.** Type 2 diabetes is strongly associated with both overweight and obesity, the major modifiable risk factor, and reduced physical activity. Population-level interventions that address the obesogenic environment, and that encourage greater physical activity, are therefore attractive targets for interventions, interventions that have potential for wide reach across the whole population as well as carrying low unit cost. Examples include price increases applied to unhealthy food and drink choices, such as levies introduced on sugar-sweetened beverages in the United Kingdom (2018) and Mexico (2014); reformulation of products resulting in lower sugar or total calorie content; reducing promotions on unhealthy foods (e.g. buy one get one free), which have been shown to increase the amount of food and drink people buy in the United Kingdom by around 20%; limiting advertising of unhealthy foods—all forms of marketing consistently influence food preference, choice and purchasing in children and adults; promoting active transport; promoting healthy eating and physical activity in schools and in working environments; and requiring clear food labelling that indicates calorie, sugar, fat, and salt content. Although each intervention alone will carry only a small effect size, increasing the number of interventions in the national or regional portfolio of interventions will have a more significant effect. Examples of broader portfolios of interventions nationally or regionally are the UK Government's Childhood Obesity Plan (2016), the North Karelia Project in Finland, and the Amsterdam Healthy Weight Programme in the Netherlands.
2. **Secondary prevention.** It is now well established that onset of type 2 diabetes can be prevented or delayed in high-risk adults. At least five major randomized controlled trials, conducted in China [8], Finland [9], United States [10], Japan [11], and India [12] have documented 30–60% reductions in type 2 diabetes incidence in adults with impaired glucose tolerance through intensive lifestyle behavioural change programme interventions focusing on weight loss, better quality nutrition, and increased physical activity. The USA-based Diabetes Prevention Programme (DPP), the largest of the five studies, showed that an intervention targeting weight loss of around 5–7% of body weight and moderate physical activity of 150 minutes per week through a structured lifestyle change program reduced the cumulative incidence of type 2 diabetes by 58% at 2.8 years [10]. Real-world translations of these trial interventional approaches have established the effectiveness of similar group-based interventions [13], and interventions are now being established in some countries as part of routine care for those at high risk (the Healthier You: NHS DPP in England [14]; the US DPP [15]; and the Australian lifestyle intervention Programme 'Life' [16]).
3. **Tertiary prevention.** Preventing the complications of type 2 diabetes, including the microvascular complications of retinopathy, peripheral neuropathy and nephropathy, and the macrovascular complications of coronary heart disease, cerebrovascular disease, peripheral vascular disease, and renovascular disease, are of paramount importance in improving the quality of life and outcomes for people with type 2 diabetes as well as reducing the financial burden on health systems. In the United Kingdom, 80% of the direct costs associated with managing people with diabetes were associated with the management of complications [17]. Well-conducted randomized controlled trials have demonstrated the effectiveness of tighter glycaemic control in preventing the microvascular complications (UKPDS) [18], of blood pressure reduction in preventing the microvascular and macrovascular complications (UKPDS) [19], and of treatment

with statins in preventing the macrovascular complications (CARDS) [20].

4. **Quaternary prevention**. Both ageing and diabetes are recognized as important risk factors for the development of functional decline and disability. Frailty in those with type 2 diabetes, and its associated weight loss, can lead to overtreatment of glycaemic parameters and of blood pressure in those with type 2 diabetes if treatment regimes are continued without modification, resulting in hypoglycaemia and postural hypotension. Clinical recommendations that address these issues are now emerging [21].

Gestational Diabetes

Gestational diabetes is defined as 'diabetes diagnosed in the second or third trimester of pregnancy that is not clearly overt diabetes' [22]. Development of hyperglycaemia in pregnancy is associated with considerable risks to both the mother and developing baby. For the mother, these include a greater likelihood of undergoing a caesarean section, pre-eclampsia and the development of type 2 diabetes: for the baby, macrosomia, shoulder dystocia, and physiological and metabolic abnormalities such as neonatal hypoglycaemia, and obesity with insulin resistance in young adulthood. The risk factors for developing gestational diabetes mirror those for developing type 2 diabetes.

The incidence of gestational diabetes and its associated complications are increasing, reflecting the increasing prevalence of maternal overweight and obesity globally. Indeed, similar to type 2 diabetes, overweight and obesity is the major modifiable risk factor contributing to the development of hyperglycaemia in pregnancy. Preventing pathological hyperglycaemia during pregnancy has several potential benefits: reduction of associated immediate maternal and fetal adverse outcomes, potential improvements in the risk of long-term sequelae and reductions in the economic burden to healthcare systems worldwide.

Given the overlap in modifiable risk factors with type 2 diabetes, the primary prevention strategies outlined earlier, which address the obesogenic environment are also applicable to the prevention of gestational diabetes (e.g. introduction of the sugar-sweetened beverage tax in the United Kingdom). There are currently no clear proven or established secondary prevention strategies and interventions have been restricted to research settings [23]. Results from the randomized controlled trials investigating lifestyle intervention strategies in at risk cohorts have been conflicting. The reasons for this probably relate to the large degree of heterogeneity across the trials. Differences existed both in the demographics of the cohorts recruited as well as the criteria used to diagnose the condition.

Evidence relating to the tertiary prevention of complications associated with established gestational diabetes is more clearly defined. Achieving good glycaemic control mitigates the risk of many adverse maternofetal outcomes. In the first instance, dietary modification is recommended with pharmacological therapy in the form of either metformin or insulin advocated if glucose control is not achieved with the former.

The glycaemic thresholds used to diagnose gestational diabetes are the subject of debate. The Hyperglycaemia Adverse Pregnancy Outcomes study (HAPO) clearly demonstrated the linear relationship between glycaemia and fetal birth weight. On the basis of these results, the International Association of Diabetes in Pregnancy Study Group (IADPSG) and the WHO lowered the thresholds at which gestational diabetes should be diagnosed. However, there exists no evidence on treating glucose levels to the diagnostic criteria now formulated by these two organizations and concerns have been raised as to whether women who fall into lower categories of glycaemia could be overmedicalized [24].

Type 1 Diabetes

Type 1 Diabetes is a chronic autoimmune condition characterized by T-cell mediated progressive destruction of pancreatic β cells culminating in a complete deficiency of insulin. The aetiology is not completely understood. Environmental factors are thought to trigger the autoimmune process against a background of genetic susceptibility factors. The exact nature of the environmental triggers is still the subject of debate. Fetal exposure to rubella, enterovirus and maternal pre-eclampsia have been proposed as potential intrauterine factors that could contribute to later development of type 1 diabetes with early gluten exposure and infectious agents such as Coxsackie B virus thought to be possible environmental triggers in early childhood, although clear aetiological pathways have not been established.

The global incidence of type 1 diabetes has increased by 2% over the last 20 years. In the United Kingdom, this increase has been particularly pronounced and it now has one of the highest incidence rates for children under the age of 14, surpassed only by Finland, Sweden, Saudi Arabia, and Norway. Primary prevention strategies in the context of type 1 diabetes are not possible. A significant research effort is currently concentrating on potential secondary prevention strategies in at risk individuals, although at this point, identification of individuals within a latent or preclinical phase is challenging, and there are currently no translational opportunities of such research to routine clinical practice. [25]. The research includes immune modulation in the early preclinical phase, which may prevent complete β-cell destruction in individuals who are either antibody positive or who have a family history of type 1 diabetes.

Tertiary prevention is of paramount importance in type 1 diabetes. In the preinsulin era, type 1 diabetes was associated with a high mortality and was in fact almost universally fatal either acutely from complicating diabetic ketoacidosis or due to the chronic catabolic state that ensued. With the increase in life expectancy that was observed following the introduction of insulin therapy, the microvascular and macrovascular complications of the condition came to light. However, it was not until advances in insulin therapy were achieved that the efficacy of glycaemic control in preventing complications could really be assessed. The earliest trial to demonstrate this was the Diabetes Control and Complications Trial (DCCT), which was conducted in the 1980s and early 1990s. This demonstrated a clear reduction in the microvascular primary end-point of retinopathy (as well as the secondary endpoints of microalbuminuria and neuropathy) in those that had been randomized to intensive therapy compared to those in the conventional therapy arm: mean HbA1c ~7% in the intensive therapy group and ~9% in the conventional therapy group. The legacy effect of good glycaemic control has now also been established in long-term follow up studies of those originally recruited to the DCCT cohort. The first, published 20 years later, demonstrated that despite a loss in glycaemic separation, those randomized to the intensive therapy group had

a significant reduction in cardiovascular endpoints compared to those in the conventional therapy group [26]. The second described improved mortality rates with intensive therapy to the extent that they resembled those of the general adult population [27].

Similar, to type 2 diabetes, the goals of tertiary prevention in type 1 diabetes involve preventing either the onset of microvascular and macrovascular complications or their progression if established. Good glycaemic control is paramount for this with antihypertensive therapy and statin therapy also contributing significantly to reductions in microvascular and macrovascular complication risk. Screening programmes for the complications of type 1 and type 2 diabetes including retinal screening and regular foot checks, while microalbuminuria assessments are also vital in ensuring that complications are detected early and appropriate steps can therefore be taken promptly to prevent their progression.

Osteoporosis and Bone Health

Osteoporosis is defined by the WHO as a 'progressive systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture'. The clinical significance of osteoporosis relates to the fractures that arise as a result of its development. Data from the International Osteoporosis Foundation in 2017 indicated that more than 8.9 million osteoporotic fractures occurred in 1 year. In the United Kingdom alone, 536 000 fragility fractures occur each year leading to severe pain and disability on an individual basis and costing the NHS over £4.4 billion per annum [28].

During development, skeletal growth proceeds through coordinated bone deposition and resorption allowing bones to expand, otherwise known as periosteal apposition of cortical bone, and lengthen (endochondral ossification) [29]. This process begins *in utero* and continues until the third decade of life at which point peak bone density is achieved. Later development of osteoporosis has been linked to peak bone density mass and national guidelines around primary prevention; for example, those in the United Kingdom have recommended levels for physical activity, healthy lifestyle and adequate vitamin D and calcium intake in childhood and early adulthood to ensure maximum achievement in peak bone density.

In those with established osteoporosis, the following tertiary preventative measures are recommended to prevent fragility fractures (adapted from the UK Guidelines for Osteoporosis and the International Osteoporosis Guidelines):

1. Calcium (700–1200 mg daily) and vitamin D supplementation (800 units)
2. Regular weight-bearing exercise
3. Falls risk assessment if appropriate

Where appropriate, antiresorptive therapy (e.g. bisphosphonates) are recommended.

Endocrinology and Other Lifestyle Factors

In addition to considering the aforementioned lifestyle modifications (e.g. healthy diet implementation and increased physical activity), abstinence from smoking has benefits. In the first instance, smoking is important in the development of goitre. Cigarette smoke contains thiocyanate, an anion that competitively inhibits sodium iodide symporter thus preventing iodine transport and organification [30]. Within areas that are iodine deficient, thyroid goitre prevalence is closely related to smoking habits. Smoking also increases the risk of developing Graves' disease and there is more than a tenfold increase in associated Graves' orbitopathy in individuals who smoke. Cessation of smoking is therefore an important tertiary measure in preventing thyroid associated eye disease.

In addition, abstinence from smoking and smoking cessation in the context of both type 1 and type 2 diabetes has been shown to reduce the risk of macrovascular complications.

Conclusions

There is an increasing appreciation that focusing on prevention and causing health systems to be more pro-active rather than reactive, is likely to contribute to long-term sustainability of those health systems. The Global Burden of Disease Study 2013 suggested that around 40% of health services workload is potentially preventable, yet the proportion of health expenditure directed at prevention, although acknowledged to be hard to estimate reliably, was thought to be closer to 4% [31]. Prevention in endocrinology has already contributed to significant health gains globally, and is likely to continue to reap significant benefits for populations moving forward.

REFERENCES

1. Bonita R, Beaglehole R, Kjellstrom T, WHO. *Basic Epidemiology*, 2nd edn. Geneva: WHO, 2006.
2. Rose G, Barker D. Epidemiology for the uninitiated. Screening. *BMJ*, 1978; **2**: 1417–18.
3. Wilson J, Jungner G. *Principles and Practice of Screening for Disease*. Geneva: WHO, 1968.
4. Li M, Eastman CJ. The changing epidemiology of iodine deficiency. *Nat Rev Endocrinol*, 2012; **8**: 434–40.
5. Dunn JT, Van Der Haar F. *A Practical Guide to the Correction of Iodine Deficiency*. Ontario: International Council for Control of Iodine Deficiency Disorders, 1990.
6. Vanderpump MPJ, Lazarus JH, Smyth PP, *et al.* Iodine status of UK schoolgirls: a cross-sectional survey. *Lancet*, 2011; **377**: 2007–12.
7. Gruters A, Krude H. Detection and treatment of congenital hypothyroidism. *Nat Rev Endocrinol*, 2011; **8**: 104–13.
8. Pan X, Li G, Hu Y, *et al.* Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. *Diabetes Care*, 1997; **20**: 537–44.
9. Tuomilehto J, LindströM J, Eriksson Jg, *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *NEJM*, 2001; **344**: 1343–50.
10. The Diabetes Prevention Programme Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *NEJM*, 2002; **346**: 393–403.
11. Kosaka K, Noda M, Kuzuya T. Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. *Diabetes Res Clin Pract*, 2005; **67**: 152–62.
12. Ramachandran A, Snehalatha C, Mary S, *et al.* The Indian Diabetes Prevention Programme shows that lifestyle

modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia*, 2006; **49**: 289–97.
13. Public Health England. A systematic review and meta-analysis assessing the effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes mellitus in routine practice. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/456147/PHE_Evidence_Review_of_diabetes_prevention_programmes-_FINAL.pdf (accessed 23 January 2020).
14. Barron E, Clark R, Hewings R, Smith J, Valabhji J. Progress of the Healthier You: NHS Diabetes Prevention Programme: referrals, uptake and participant characteristics. *Diabet Med*, 2018; **35**: 513–18.
15. Ely EK, Gruss SM, Luman ET, *et al.* A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. *Diabetes Care*, 2017; **40**: 1331–41.
16. Dunbar JA, Jayawardena A, Johnson G, *et al.* Scaling up diabetes prevention in Victoria, Australia: policy development, implementation, and evaluation. *Diabetes Care*, 2014; **37**: 934–42.
17. Hex N, Bartlett C, Wright D, Taylor M, Varle D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Med*, 2012; **29**: 855–62.
18. UK Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylurea or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*, 1998; **352**: 837–53.
19. Idem. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. *BMJ*, 1998; **17**: 703–13.
20. Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HA, Livinstone SJ. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. *Lancet*, 2004; **364**: 685–96.
21. Strain WD, Hope SV, Green A, Kar P, Valabhji J, Sinclair AJ. Type 2 diabetes mellitus in older people: a brief statement of key principles of modern day management including the assessment of frailty. A national collaborative stakeholder initiative. *Diabet Med*, 2018; **35**: 838–45.
22. Alberti K, Zimmet P. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med*, 1998; **15**: 539–53.
23. Agha-Jaffar R, Oliver N, Johnston D, Robinson S. Gestational diabetes mellitus: does an effective prevention strategy exist? *Nat Rev Endocrinol*, 2016; **12**: 533–46.
24. The HAPO Study Cooperative Research Group. Hyperglycaemia and adverse pregnancy outcomes. *NEJM*, 2008; **358**: 1991–2002.
25. Roep BO. Primary prevention for type 1 diabetes mellitus? *Nat Rev Endocrinol*, 2015; **11**: 451–2.
26. Nathan DM & DCCT/EDIC Research Group. The Diabetes Control and Complications Trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*, 2014; **37**: 9–16.
27. The DCCT/EDIC Study Group. Mortality in type 1 diabetes in the DCCT/EDIC versus the general population. *Diabetes Care*, 2016; **39**: 1378–83.
28. Compston J, Cooper A, Cooper C, *et al.* UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos*, 2017; **12**: 43.
29. Weaver CM, Gordon CM, Janz KF, *et al.* The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporos Int*, 2016; **27**: 1281–386.
30. Krudsen N, Bulow I, Laurberg P, Ovesen L, Perrild H, Jorgensen. T. Association of tobacco smoking with goiter in a low-iodine-intake area. *Arch Int Med*, 2002; **162**: 439–43.
31. Newton JN, Briggs AD, Murray CJ, *et al.* Changes in Health in England with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 2015; **386**: 2257–74.

SECTION 2

Pituitary and Hypothalamic Diseases

2.1

Functional Anatomy of the Hypothalamus and Pituitary

John F. Morris

Introduction: The Hypothalamus and Its Many Functions

The hypothalamus forms only a very small part (0.3%) of the brain but one that is the crucial integrative centre for the control, via its autonomic and endocrine effector systems, of fluid and electrolyte balance, reproduction, thermoregulation, ingestive behaviour and energy balance, affective and emotional-response systems. To do this it has extensive, reciprocal, neural and vascular connections with all other parts of the central nervous system (CNS) and, via the autonomic and vascular systems with the rest of the body. Although this chapter must focus on the hypothalamo–pituitary relationships, it must also cover hypothalamo-autonomic relationships and the relationships between the hypothalamus, the alimentary tract, and appetite control. It also plays a major role in diurnal rhythm control.

Anatomy of the Hypothalamus and Pituitary Hypothalamus

See Figure 2.1.1. The hypothalamus is that part of the diencephalon forming the lateral and inferior walls of the third ventricle; extending from the hypothalamic sulcus (which separates it from the thalamus) to the base of the brain, and from the lamina terminalis rostrally to a plane posterior to the mammillary bodies and posterior commissure caudally. It is continuous anteriorly with the septal and orbitofrontal cortex and anterior perforated substance, anterolaterally with the optic tracts and substantia innominate, dorsolaterally with the anterior part of the subthalamus and the internal capsule, and posteriorly with the tegmentum of the midbrain. The floor of the third ventricle comprises the optic chiasm, the tuberal eminence which extends into the infundibulum/pituitary stalk, the tuber cinereum, the mammillary bodies, and the posterior perforated substance. The floor of the third ventricle is marked by a distinct infundibular recess, and by the median eminence, in the inner zone of which many peptidergic neuroendocrine fibres pass either on into the posterior pituitary or end on the fenestrated capillaries of the outer zone. Specialized ependymal cells (tanycytes) line the floor and adjacent lateral walls of the third ventricle, are sexually dimorphic, and have processes that pass through the outer zone of the median eminence to end on the fenestrated blood vessels of the outer zone.

Pituitary Gland

See Figure 2.1.1. The pituitary gland is located in the sella turcica on the dorsal aspect of the sphenoid. It is attached to the infundibulum by both the pituitary stalk (composed largely of magnocellular axons en route to the posterior pituitary) and by long portal vessels that supply the anterior pituitary. Beneath the pituitary and floor of the sella turcica is the roof of the sphenoid air sinus, a relationship exploited in the transethmoid, trans-sphenoid surgical approach to the pituitary. The bony anterior margin of the sella supports the optic chiasm; lateral to the midline on either side the internal carotid artery

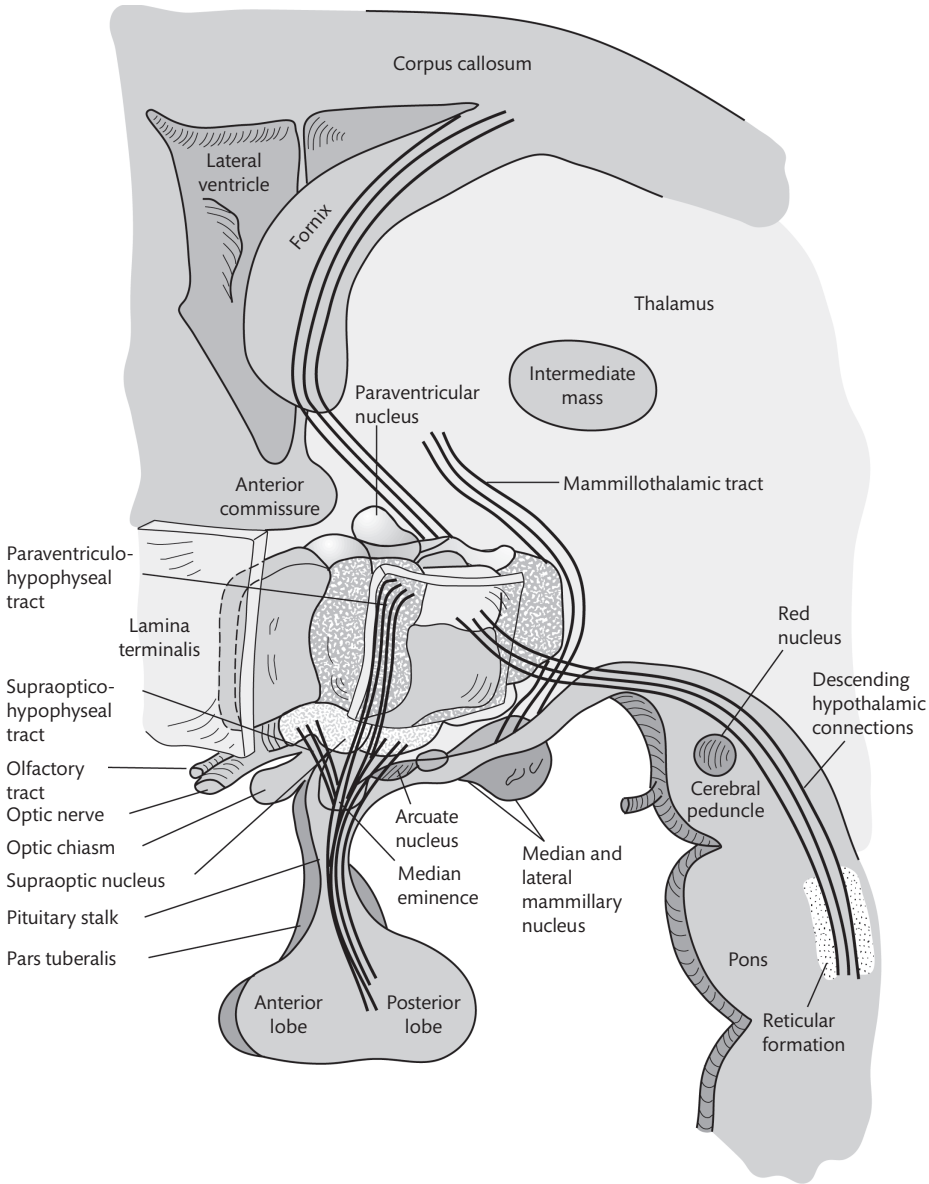


Figure 2.1.1 Hypothalamic nuclei and major tracts.

passes from the medial wall of the cavernous sinus into the subarachnoid space; lateral to the artery the anterior clinoid processes of the sphenoid give attachment to the meningeal layer of the dura forming the lateral wall of the cavernous sinus. The posterior margin of the sella is the midline dorsum sellae which ends laterally as the posterior clinoid processes to which dura forming the lateral wall of the cavernous sinus is attached. The meningeal layer of the dura continues over the pituitary as the sellar diaphragm which has a (usually) small central aperture for the pituitary stalk. The sella is lined by the periosteal layer of dura. The pituitary gland itself lacks leptomeninges which surround the stalk and below the level of the sellar diaphragm reflect on themselves forming the infradiaphragmatic pituitary cerebrospinal fluid (CSF) cistern. If this cistern is breached during transsphenoidal surgery in a person in which the diaphragmatic opening in the leptomeninges is large, persistent CSF rhinorrhoea may develop. The cavernous sinuses are situated on either side of the sella and body of the sphenoid. Each receives the superior and inferior ophthalmic veins (which communicate with the facial veins), the superficial middle cerebral vein and the sphenoparietal sinus which runs along the lesser wing of the sphenoid. The two cavernous sinuses communicate through veins anterior, posterior, and sometimes inferior to the pituitary gland. Cavernous sinus blood drains posteriorly through superior and inferior petrosal sinuses to reach the internal jugular vein and through the foramen lacerum to the pterygoid venous plexus. The oculomotor, trochlear, and maxillary and ophthalmic branches of the trigeminal pass through the lateral wall of the sinus, the internal carotid artery follows an S-shaped course in the medial wall of the sinus, with the abducent nerve on its lateral aspect. Within the cavernous sinus, the internal carotid artery gives a number of branches including fine branches which supply the sinus walls, a branch to the abducent nerve, and the posterior (but not anterior) pituitary via the inferior hypophyseal artery and pituitary capsule.

Development of the Hypothalamus and Pituitary

In the diencephalic part of the forebrain vesicle the hypothalamus develops in the floor; the thalamus develops in the roof, and eye

primordia in the lateral walls. Sonic hedgehog (Shh) secreted by the prechordal mesoderm situated below the ventral midline floor directs progenitors to hypothalamic and optic stalk fates but inhibits retinal differentiation (loss of Shh causes absent hypothalamus and cyclopia). A variety of transcription factors (including Nxk2.1, Nkx2.1, Rx, Six3) direct formation of the dorsal and ventral hypothalamus. The hypothalamus shows extensive spatial patterning with neurons becoming organized into 12–28 different nuclei. Wnt is critical in anterior-posterior patterning and BMP7 directs nuclear specification. With one exception (gonadotropin-releasing hormone (GnRH) neurons) there is little cell migration in and out of the hypothalamus after neurogenesis starts. Neurons in the lateral hypothalamic area are generated first, then neurons in the dorsomedial and ventromedial nucleus and finally those along the midline such as the arcuate nucleus. Gliogenesis occurs only after neurogenesis. GnRH neurons form from precursors which originate outside the CNS in the nasal placode and migrate in association with olfactory axon bundles. Once in the CNS, olfactory axons synapse in the olfactory bulb whereas GnRH neurons continue to migrate caudally on other axons to the hypothalamus. Mutation of a number of genes including KAL1 prevent this migration and cause Kallmann syndrome (anosmia with hypogonadotropic hypogonadism due to a failure of GnRH secretion).

The pituitary has a dual origin: the adenohypophysis (anterior lobe and intermediate zone) develops from ectoderm in the roof of the stomodeum which evaginates by week 3 of gestation to form Rathke's pouch. The dorsal portion of Rathke's pouch directly contacts the infundibulum which grows down from the floor of the developing hypothalamus to form the neurohypophysis (posterior lobe). When Rathke's pouch becomes separated from the stomodeal ectoderm by the developing sphenoid bone, remnants of the pouch form the pharyngeal pituitary. The development and differentiation of the various pituitary endocrine cell types involves the sequential and expression of several transcription factors, including PIT1 (POU1F1) and PROP 1. Corticotrophs are the earliest detectable fetal pituitary cell type, appearing by the eighth week of gestation. T-pit (Tbx19) has been identified as a critical transcription factor for both corticotroph cell differentiation and for *POMC* gene transcription (and subsequently melanotroph development in the intermediate zone). Prop-1 and Pit-1 direct the formation of somatotrophs and lactotrophs, and Pit-1 and GATA-2 that of most thyrotrophs. Finally, GATA-2 with steroidogenic factor 1 (SF1) directs the formation of gonadotrophs. Mutations in the genes encoding these transcription factors can produce pituitary hypoplasia or agenesis, and a combined deficiency of pituitary hormones (GH, PRL, TSH, FSH, LH). Mutations in the gene encoding T-pit cause congenital isolated adrenocorticotropic hormone (ACTH) deficiency. Postnatally, cell turnover in the anterior pituitary is very slow and regulated by both hypothalamic and feedback factors. The number of somatotrophs doubles during puberty and the number and size of lactotrophs changes markedly during pregnancy, lactation, and weaning. Progenitor stem cells in the anterior pituitary have been difficult to define but candidates have included the cells which line the pituitary cleft and certain types of folliculo-stellate (F-S) cells. Recently a tiny population of SOX2-expressing cells in the adult pituitary has been shown to be capable of differentiating into all the endocrine and F-S cell types. Because of the immunochemical resemblance of some F-S cells to dendritic cells a bone marrow origin has been proposed for these cells but this remains to be confirmed. Much less is known about the control of the formation of the neurohypophysis, but it has recently been shown that both Hes1 and Hes5 are required both for the evagination of the infundibulum to form the posterior pituitary, and also for the formation of pituicytes.

Blood Supply of the Hypothalamus and Pituitary

See Figure 2.1.2. The hypothalamus is supplied by small arteries which penetrate the base of the brain. These are derived from vessels forming the circle of Willis in the interpeduncular fossa. An anteromedial group from the anterior cerebral and anterior communicating arteries supply the anterior part of the hypothalamus including the optic chiasm, suprachiasmatic, and supraoptic regions and the anterior column of the fornix. A posteromedial group from the posterior communicating artery and proximal part of the posterior cerebral artery supply more posterior parts of the hypothalamus. Superior hypophyseal arteries usually arise from the internal carotid immediately after it pierces the sellar diaphragm. They form anterior and posterior branches and a plexus of vessels which surround the infundibulum. From these some branches pass toward the ependyma of the third ventricle in the median eminence where they are surrounded by a plexus of capillaries (gomitoli). In the median eminence arteries give rise to fenestrated capillaries that form an external (mantle) plexus and an internal (deep) plexus. From the external and internal plexus, long portal vessels descend along the stalk to the anterior pituitary where they branch repeatedly to form the vascular sinusoids that surround the anterior pituitary endocrine cells. Short portal vessels formed from the capillary plexus lower in the stalk also pass to the anterior pituitary. There are no capillaries among the cells of the pars intermedia. The anterior pituitary therefore has no direct arterial supply, but its fenestrated sinusoids are fed entirely by the long and short portal vessels. By contrast, the posterior pituitary and the lower part of the pituitary stalk receives a direct arterial supply from the inferior hypophyseal arteries, derived from the internal carotid artery within the cavernous sinus, and which pass directly into the posterior pituitary to form a plexus of fenestrated capillaries within that gland. A few small branches from the artery pass forward and give some supply to the periphery of the anterior pituitary. Collecting veins draining all the parts of the pituitary pass via the intercavernous and cavernous sinuses into the inferior petrosal sinus and from there into the internal jugular vein. Inferior petrosal sinus blood sampling can therefore be useful in the diagnosis of pituitary tumours.

Whereas the capillaries in most of the hypothalamus are continuous and, with the surrounding astrocytic end feet, form an effective blood–brain barrier, those in the median eminence are fenestrated, The fenestrations of the median eminence capillaries not only allow those capillaries to collect the hypophysiotropic hormones secreted by the terminals of the parvocellular neurosecretory neurons and to transfer them to the anterior pituitary, but they also provide a route whereby otherwise impermeant peptide hormones secreted by fat (e.g. leptin), the pancreas (e.g. insulin) and the gut (e.g. ghrelin) can reach neurons in the arcuate nucleus around the base of the third ventricle, where they are important signals in the homeostatic control of feeding behaviour. The median eminence is therefore one of the 'circumventricular organs' of the third ventricle (c.f. the organum vasculosum of the lamina terminalis,

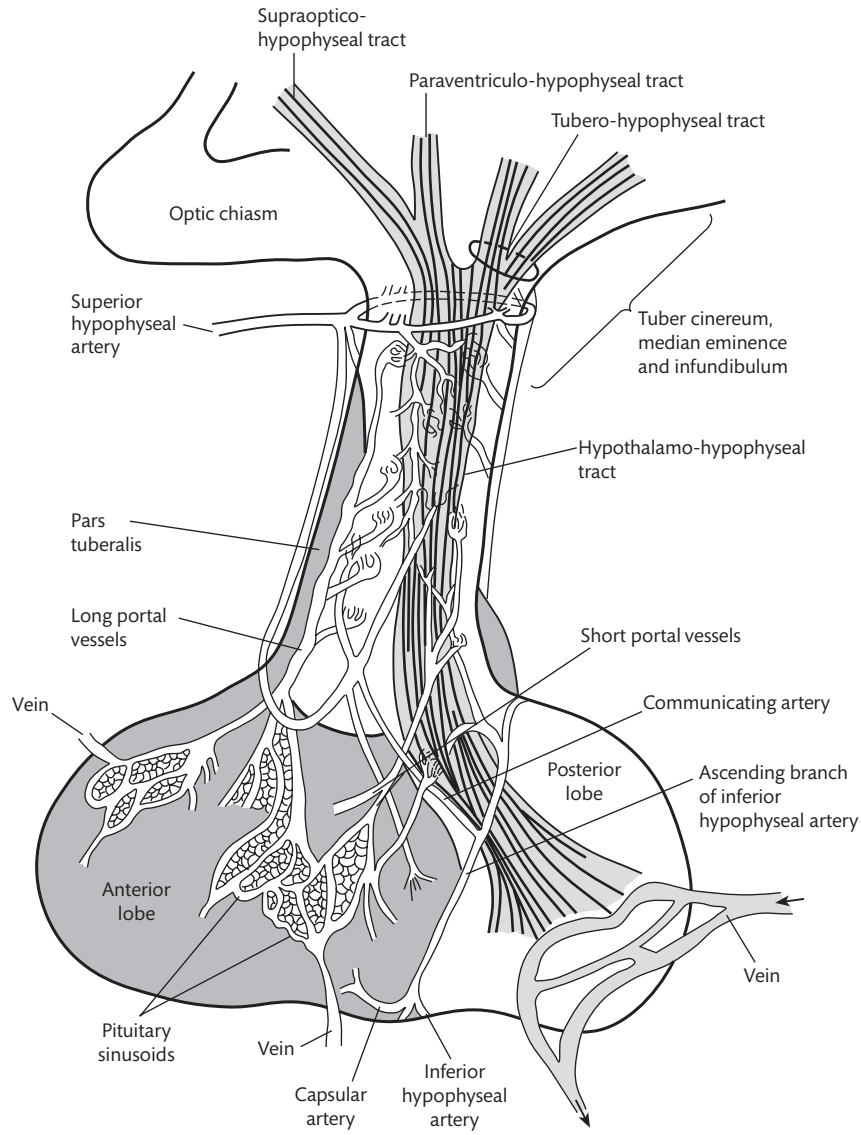


Figure 2.1.2 Vasculature of the pituitary gland.

subfornical organ, and area postrema) where the blood–brain barrier is permeable.

Cellular Structure of the Pituitary

Neurohypophysis

The neurohypophysis (posterior, or neural lobe) of the pituitary is a direct ventral extension of the infundibulum at the base of the hypothalamus. The majority of the gland is composed of the varicose axons of magnocellular neurons of the supraoptic and paraventricular nuclei which leave the nuclei and course together through the internal zone of the median eminence and pituitary stalk into the neural lobe where they branch repeatedly and form 'nerve endings' on the fenestrated capillaries. In the neural lobe the varicosities are often very large ('Herring bodies'; axonal 'swellings'). Vasopressin and oxytocin are synthesized as precursors in the magnocellular neuron cell bodies, and packaged in ~160 nm diameter dense-cored vesicles (DCV) which are transported along axonal microtubules to the neural lobe where they are released by exocytosis into the perivascular space of the capillaries. The DCV are found throughout the undilated axons, 'endings' and 'swellings' but the perivascular 'endings' are distinguished by the presence of a population of electron-lucent synaptic-like vesicles which contain glutamate the significance of which is unclear. Although the perivascular 'endings' may be the principal site of hormone release, exocytosis of DCV can occur from any part of the varicose axons. In some 'nerve swellings' there is massive lysosomal destruction of DCV, disposing of aged and effete DCV that have not been released. Intermingled among and surrounding the neurosecretory axons and also contacting the capillaries are pituicyte cells. These cells are characterized by the presence of lipid droplets and of glial acidic fibrillary protein (GFAP) an intermediate filament protein found in astrocytes. Like astrocytes, pituicytes can form gliomas ('pituicytomas'). Pituicytes express β-adrenergic and opioid receptors. When secretion of vasopressin or oxytocin is particularly active (dehydration; parturition) the extent which pituicytes enclose the neurosecretory axons is much reduced. They are also capable of

engulfing and destroying parts of the axons. Classical macrophages are located in the perivascular space of the neurohypophysis. These act to 'trim off' any neurosecretory axons that penetrate the perivascular basal lamina. The ability of neurosecretory axons to grow is also exemplified by what happens after traumatic pituitary stalk rupture. At first the distal segments of the axons die (releasing vasopressin and temporarily reducing urine output) but then, after a period of central diabetes insipidus, the increase in urine output decreases as the proximal parts of the neurosecretory axons grow and make functional contacts with capillaries in the median eminence and upper infundibulum.

Adenohypophysis

The adenohypophysis (anterior pituitary) forms about 80% of the pituitary; it comprises the large pars distalis, the very small pars intermedia, and the pars tuberalis the cells of which extend up around the infundibulum to the tuber cinereum.

The pars distalis is formed from five types of endocrine cells: corticotrophs, gonadotrophs, thyrotrophs, lactotrophs, and somatotrophs all of which synthesize and store hormone in dense-cored secretory vesicles; together with folliculo-stellate (F-S) cells—agranular, 'non-endocrine' cells, and the endothelial cells lining the vascular sinusoids. The endocrine cell types are not evenly distributed: somatotrophs and lactotrophs (acidophils) predominate laterally whereas the central zone is rich in basophil, PAS-positive corticotrophs, gonadotrophs, and thyrotrophs. Of the total, somatotrophs form about 50%, lactotrophs 20%, corticotrophs 15%, gonadotrophs 10% and thyrotrophs 5% of the endocrine cells. Somatotrophs, which secrete growth hormone (GH; 191 aa), are ovoid cells of medium size containing abundant spherical DCV of ~300 nm diameter. Somatotroph adenomas contain variable numbers of DCV, but those with few DCV tend to be more aggressive with a worse response to therapy. Lactotrophs secrete prolactin (PRL; 199aa) and exist in a number of different types, depending on the size and morphology of their DCV. Type I lactotrophs contain many large (300–900 nm) irregularly shaped DCV that are thought to represent greater storage; type II lactotrophs contain smaller (<200 nm) spherical DCV and are thought to be the more actively secreting form (though type I cells can be stimulated to exocytose their large irregular DCV). There are also intermediate forms, and some 'somatomammotrophs' produce both GH and prolactin which is not unexpected as both somatotrophs and lactotrophs are derived from a common precursor. Lactotrophs are particularly influenced by oestradiol, and in pregnancy the increased number and size of lactotrophs causes the entire pituitary to enlarge. Some lactotrophs are situated in close relationship with gonadotrophs. Increased prolactin has a marked inhibitory effect on gonadotrophs and prolactinomas are a common cause of infertility. As with somatotroph adenomas, lactotroph adenomas tend to have fewer, smaller DCV. Corticotrophs are medium sized irregular polygonal cells in which the DCV (150–250 nm) are largely situated at the cell periphery, just beneath the plasmalemma. Corticotrophs synthesize pro-opiomelanocortin (PMC) which is cleaved in corticotrophs to yield adrenocorticotrophic hormone (ACTH; 39aa), β-lipotropin, and endorphins. Most corticotroph adenomas are basophilic (many DCV) or chromophobe (few DCV) monoclonal tumours most of which secrete ACTH and cause Cushing's disease. When cortisol is excessive, corticotrophs develop Crooke's hyaline changes caused by an accumulation of microfilaments. Because the adenoma cells remain sensitive to and inhibited by cortisol, they remain small microadenomas. In contrast 'silent' adenomas or macroadenomas synthesizing high-mw POMC are aggressive and invasive.

Gonadotrophs are large basophilic cells with extensive cytoplasm; most produce both follicle-stimulating hormone (FSH; 204aa) and luteinizing hormone (LH; 204aa), though in some only one of the gonadotrophins can be detected. Their DCV tend to be scattered throughout the cytoplasm except at the time of the preovulatory gonadotrophin surge, when they are located more marginally. The DCV vary in appearance from small spherical (250 nm) vesicles with uniformly dense cores, to larger, more ovoid vesicles with a dense core and more lucent surround. The relative production of FSH and LH is determined by the frequency of GnRH pulses and the amount of inhibin. Functioning gonadotroph adenomas (mostly macroadenomas) are very rare. They can cause menstrual irregularity and ovarian hyperstimulation in premenopausal females, testicular enlargement in males, and precocious puberty in children. Most gonadotroph adenomas are 'silent' or non-functioning and the most common type of macroadenoma. They may be asymptomatic or cause neurological effects via increasing mass. Most commonly they produce FSH or the alpha subunit, but do so very inefficiently.

Thyrotrophs are the least common, the smallest and contain the smallest DCV (100–150 nm) of any pituitary endocrine cell. They produce thyrotropin (TSH; 201aa).

Folliculo-stellate cells are a heterogenous group of cells that comprise about 5% of anterior lobe cells and whose functions are only recently being revealed. They are characterized by a lack of dense-cored secretory vesicles, by the expression of S100 protein, and by their linkage together to form follicles and a network throughout the anterior pituitary coupled electrically by gap junctions and calcium waves. They produce many signalling molecules with paracrine effects on the endocrine cells including growth factors, cytokines, nitric oxide and annexin 1 (which is implicated in the early inhibitory effect of glucocorticoids on ACTH release). There is evidence that some folliculostellate (FS) cells may act as stem/progenitor cells for the anterior pituitary.

The pars tuberalis (pars infundibularis) is an upward extension of the anterior lobe along the pituitary stalk.

The pars intermedia is vestigial in human adults, but its scattered cells can produce POMC and these can intrude into the anterior margin of the neurohypophysis (basophilic invasion')

The pharyngeal pituitary is a small, but consistent embryologic remnant mass of tissue located either in the submucosa of the dorsal wall of the nasopharynx or in the sphenoid (i.e. along the line of Rathke's pouch contact with the forebrain vesicle). The cells are often chromophobe (few DCV) and inactive, but rarely form adenomas which cause Cushing's syndrome, acromegaly, hyperthyroidism, or prolactinoma.

Hypothalamic Neuroendocrine Neurons: Endocrine and Central Effects

For this volume it is particularly important to consider the neuroendocrine neurones. These form two major groups: magnocellular neurons of the supraoptic and paraventricular nuclei which project to the posterior pituitary and release oxytocin or vasopressin into the systemic circulation; and the more scattered parvocellular

neurosecretory neurons which produce the releasing- and release-inhibiting factors and project to the capillaries of the median eminence where they release their signalling molecules to control the anterior pituitary endocrine cells via the hypothalamo–pituitary portal veins. Although vasopressin and oxytocin in peripheral blood can be assayed for diagnostic purposes, the concentrations of the parvocellular releasing and release-inhibiting factors are normally too low for assay, and such measurements are useful only when investigating ectopic secretion by tumours.

Most magnocellular and parvocellular neurosecretory neurons secrete peptides. These are produced from larger precursors and packaged into DCV in the cell bodies. The DCV are transported along varicose axons of the neurosecretory neurons to their neurohaemal contacts (indeed, this defines the neurons as 'neuroendocrine'). The role of DCV in the dendrites of neurosecretory neurons was not understood until it was shown that they can be released from the dendrites and that dendritic release into the hypothalamus is controlled separately from neurohaemal axonal release and that the released peptides have important local autocrine and paracrine physiological effects. This has been demonstrated most clearly for magnocellular neurons secreting oxytocin or vasopressin, but it seems likely that it applies more generally to peptidergic neurons. Indeed, the two long processes of GnRH neurons both contain GnRH DCV and are contacted by synaptic boutons along their length, making them intermediate between axons and dendrites ('dendrons'). There is also now evidence that the oxytocinergic neurons of the posterior part of the paraventricular nucleus that project varicose axons down the spinal cord release oxytocin from non-synaptic varicosities to influence the lumbar gastrin-releasing peptide (GRP) neurons to facilitate male reproductive activity. It is clear, therefore, that peptidergic neuroendocrine neurons can have much more widespread effects in the CNS than those caused by the neurohaemal release of their products. This is consistent with the widespread distribution of receptors for the peptides, often far distant from relevant peptidergic axons. Neuroendocrine neurons are only one subset of peptidergic neurons—the extent to which these considerations apply generally to all peptidergic neurons remains to be determined.

Functional Hypothalamic Neuron Groups and Hypothalamo-Pituitary Systems

See Figure 2.1.1. The landmarks of the ventral surface of the hypothalamus are used to divide it into an anterior (preoptic and supraoptic), a middle (tuberal), and a posterior (mammillary) zone. On each side of the third ventricle is a narrow periventricular zone, a medial zone with relatively well-defined nuclei, and a lateral hypothalamic area where cell groupings are less clear. Some nuclei, such as the paraventricular nucleus, comprise a number of structurally and functionally different subnuclei. Such neuronal groupings usually reflect integrative centres from which the neurons exert efferent control.

Anterior Zone: Circadian Rhythm; Osmotic and Thermal Homeostasis; Stress Response

Preoptic region—osmotic and thermal sensing: The preoptic region comprises the periventricular grey matter around the rostral extremity of the third ventricle. One population of kisspeptin/neurokinin B neurons is located here in greater numbers in females than in males (see Figure 2.1.9). Two of the circumventricular organs where the blood–brain barrier is permeable are located the anterior region. Some organum vasculosum of the lamina terminalis (OVLT) neurons are intrinsically osmosensitive and essential for normal thirst responses (see Figure 2.1.4); others are thermo-, prostaglandin, and interleukin 1-sensitive cells involved in thermoregulation and the febrile response. The subfornical organ (SFO), located on the ventral aspect of the fornix just behind the interventricular foramen and the median preoptic nucleus are also involved in thirst and fluid regulation

Supraoptic Region

Suprachiasmatic nucleus (SCN)**—circadian control** [1]: See Figure 2.1.3. Immediately above the optic chiasm are the SCN, discrete bilateral groups of cells which form the body's principal circadian clock. Its neurons and astrocytes exhibit an autonomous timing mechanism based on transcriptional and post-translational feedback loops. It receives a direct input from light-sensitive ganglion cells of the retina which modify its intrinsic rhythmicity according to the light–dark cycle. The rhythmic output from the SCN projects dorsally into the periventricular zone and more widely, via both gamma aminobutyric acid (GABA) and neuropeptides including vasopressin and VIP, to centres in the medial hypothalamus and brainstem that control arousal, sleep, feeding, neuroendocrine behaviour, and autonomic status.

Supraoptic nucleus (SON) and **paraventricular nucleus** (PVN) [2]: Magnocellular neurons of the SON and PVN produce either vasopressin (ADH) for osmotic and volume regulation (Figure 2.1.4) or oxytocin for milk ejection and parturition (Figure 2.1.5), each with their respective neurophysin. The SON neurons lie dorsal and lateral to the optic chiasm and optic tract. Their varicose axons all project posteromedially to the inner zone of the median eminence and on into the neurohypophysis; their dendrites project ventrally to form a dendrite-rich zone on the ventral surface of the brain. PVN magnocellular neurons are located laterally in the PVN. Their axons pass laterally around the anterior columns of

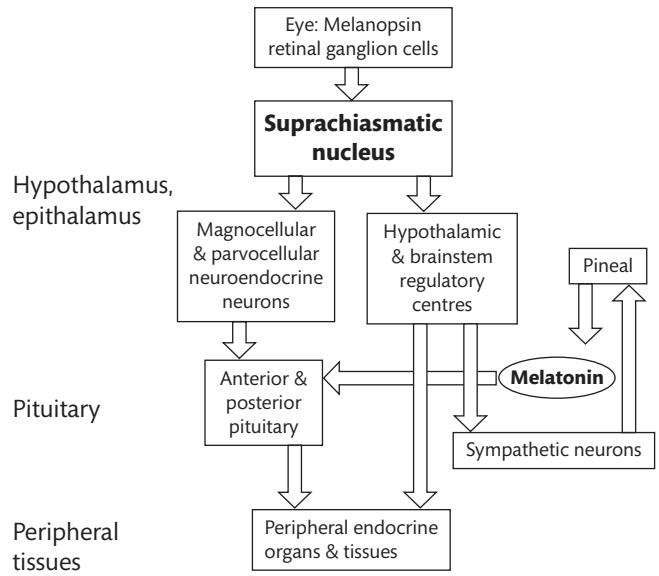


Figure 2.1.3 Control of circadian rhythms.

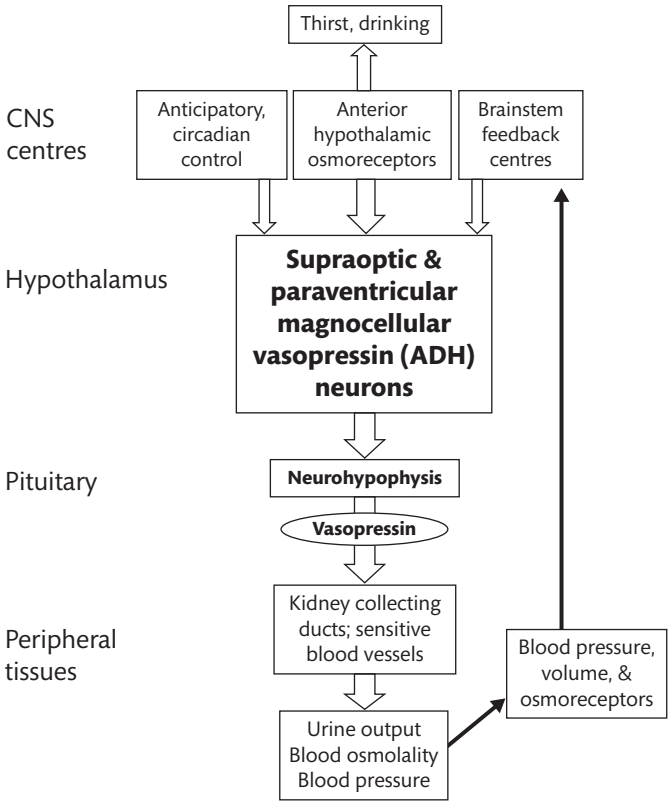


Figure 2.1.4 Osmotic and volume regulation.

fornix then ventrally to join the supraoptic axons to pass through the median eminence and into the neural lobe; their dendrites project medially and form a plexus beneath the ependyma of the third ventricle. Vasopressin neurons receive osmotic inputs from the anterior hypothalamus, volume and stress inputs from the brainstem; oxytocin neurons receive spinal inputs from the breast and uterus.

Whereas the SON is composed entirely of magnocellular neurosecretory neurons, the PVN contains many different types of neuron and forms a complex integrative centre. Parvocellular neurosecretory neurons of the PVN are located ventromedially, their axons passing ventrally to end on median eminence capillaries. They produce corticotrophin-releasing hormone (CRH; 41aa), vasopressin, or thyrotropin-releasing hormone (TRH; 3aa).

Parvocellular oxytocin neurons project to various brainstem nuclei and regions of the spinal cord. Recent evidence suggests that they involved in the modulation of feeding (Prader–Willi syndrome), cardiovascular and respiratory functions, inflammatory pain processing and, via axons ending in the sexually dimorphic nuclei of the lumbosacral cord, male sexual activity.

Hypothalamo–pituitary stress axis [3]: See Figure 2.1.6. CRH neurons are a key part of the hypothalamic–pituitary–adrenal (HPA) axis stress response system. Interestingly, limbic centres have minimal direct projections to the CRH neurons, but afferents from the hippocampus, prefrontal cortex, and amygdala mostly relay in the bed nucleus of the stria terminalis, hypothalamus, and brainstem to influence CRH neurons in the PVN. They are, however, directly responsive to glucocorticoids as are many of their inputs. Some CRH neurons also produce vasopressin when stressed and this acts synergistically with CRH to stimulate ACTH. Other parvocellular vasopressin neurons project widely in the hypothalamus and to the limbic system, the brainstem, and the spinal cord where they are involved in blood pressure and temperature regulation, including brown fat thermogenesis. Oxytocin neurons in the posterior part of the PVN project mainly to the brainstem and spinal cord where they have roles in gastric reflexes and penile erection. The PVN also act as a hypothalamic integrative centre for feeding.

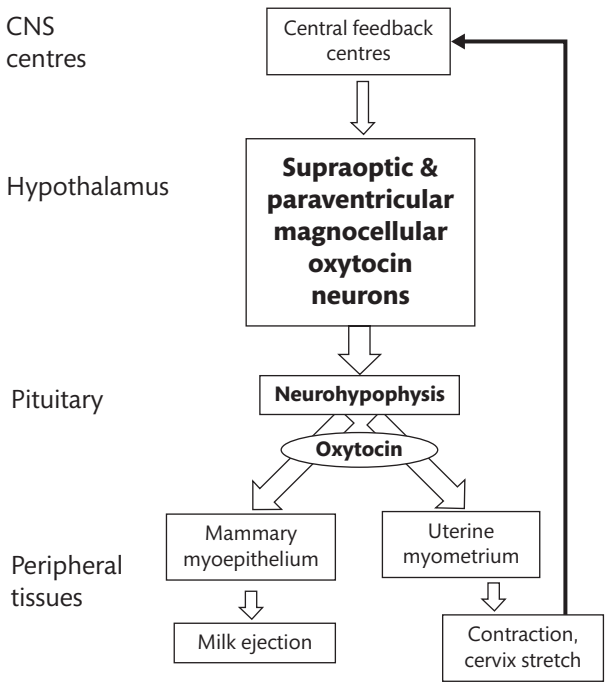


Figure 2.1.5 Oxytocin pathways and their effects.

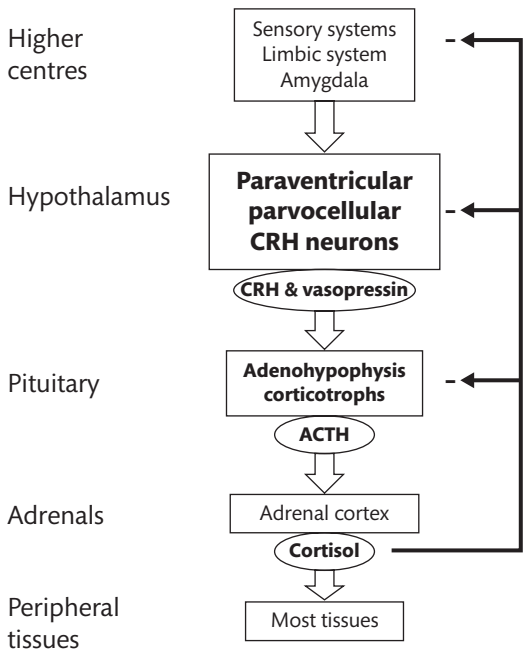


Figure 2.1.6 Endocrine and neural stress control system.

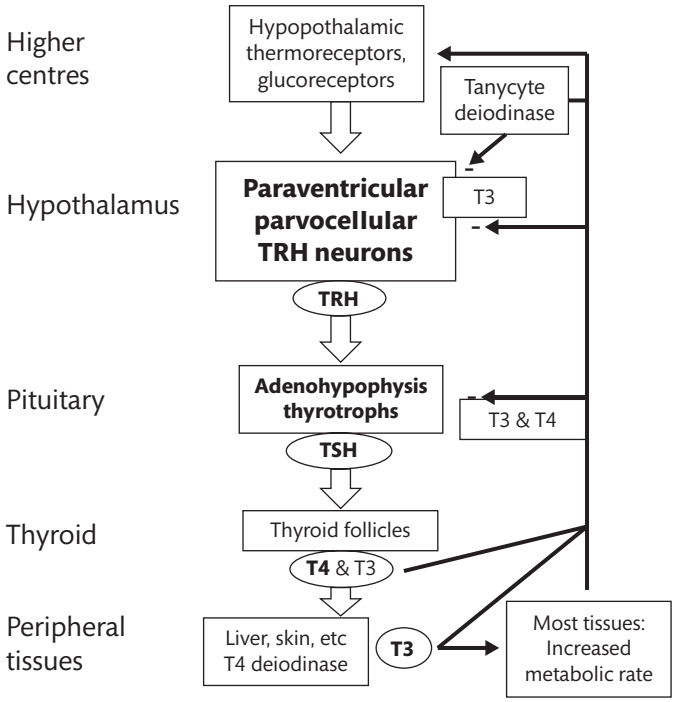


Figure 2.1.7 Hypothalamo–pituitary–thyroid system.

Hypothalamo–pituitary–thyroid axis [4]: See Figure 2.1.7. TRH-secreting neurons in the medial and periventricular parvocellular parts of the PVN act as a hypothalamic integrator of energy metabolism. They receive inputs from arcuate neurons involved in feeding control and have reciprocal connections with AHA neurons involved in thermoregulation. They express T3 receptors and hypothalamic T3 is produced by type 2 iodothyronine deiodinase in local tanycytes. Their axons project to the median eminence to control anterior pituitary TSH and thus thyroid activity and metabolism. TRH expression is reduced in fasting and by endotoxins to decrease metabolic rate.

Anterior and lateral hypothalamic areas: These areas contain less well-defined groups of neurons. The **anterior hypothalamic area** (AHA), which blends with the preoptic area and extends caudally into the middle (tuberal) zone, is the location of thermosensitive neurons that form the key body thermostat. Their thermal sensitivity is conferred by the expression of TRPM2 channels, but the neurons also receive and integrate inputs from peripheral temperature sensors and also from inflammation-derived cytokines. Fever is thought to be the result of prostaglandin E2 (PGE2) inhibition of warm-sensitive neurons hence the ability of aspirin to reduce fever. The outputs of these neurons influence the central autonomic control neurons. Warming the thermosensitive area induces heat loss via vasodilation, sweating, and behavioural adaptation; cooling triggers the opposite. It is therefore unsurprising that damage to the area (e.g. after surgery for large pituitary tumours) can cause loss of body temperature control. The neurons also activate the PVN CRH neurons in fever-induced stress. Neurons of the **lateral hypothalamic area** (LHA) are involved in a number of important functions: sleep, feeding, sex, and reward. Early lesion experiments associated the lateral hypothalamus with reduced feeding and increased sexual activity; it was therefore designated a 'feeding centre'. Studies have now shown that both wakefulness and feeding are associated with increased activity of LHA orexin neurons as well as brainstem arousal systems including the noradrenergic locus coeruleus. In contrast, LHA neurons expressing melanin-concentrating hormone (MCH) promote sleep and feeding by circuits that are still being elucidated, though both orexin and MCH neurons project to the PVN integrative centre and the LHA receives inputs from the central amygdala and from hypothalamic nutrient-sensing neurons.

Middle (tuberal zone): control of the anterior pituitary, growth, feeding and metabolism: The hypothalamus is widest in the tuberal region, and the anterior columns of fornix separate the medial group of nuclei from the LHA and medial forebrain bundle. Ventrally, the **arcuate (infundibular)** nucleus surrounds the base of the third ventricle; above it is a distinct **ventromedial** (VM) nucleus and above that a less distinct **dorsomedial** (DM) nucleus. Much recent attention has concentrated on the arcuate nucleus which has important functions in feeding (it receives peptide hormonal signals from the gastrointestinal tract and adipose tissue via the permeable capillaries in the median eminence), in reproduction (through its kisspeptin, GnRH, and dopamine neurons) and growth (though its growth hormone-releasing hormone (GHRH) and somatostatin neurons).

Feeding and metabolism control [5]: See Figure 2.1.8. Although early lesion experiments designated the ventromedial area as a 'satiety' centre, it was the discovery of leptin, and of arcuate neurons producing either neuropeptide Y (NPY) and agouti-related peptide (AgRP) which promote feeding or POMC (acting via alpha-MSH and the MC4 receptor) and CART (cocaine- and amphetamine-related transcript) which inhibit feeding, that dramatically increased understanding of homeostatic feeding control. Not only are many of the neurons of the area influenced by circulating levels of glucose, fatty acids, and some amino acids, but virtually all of the signalling molecules produced by the pancreas (insulin, amylin), the gastrointestinal tract (ghrelin, GLP-1, PYY, CCK) and adipose tissue (leptin, adiponectin, interleukins) influence arcuate neurons to signal nutritional status. Of these, ghrelin, secreted by the stomach when empty, stimulates feeding; the remainder inhibit feeding—the system has evolved to protect us against famine, not against obesity! The NPY/AgRP and POMC/CART neurons are reciprocally interconnected and connected with the LHA, the ventromedial

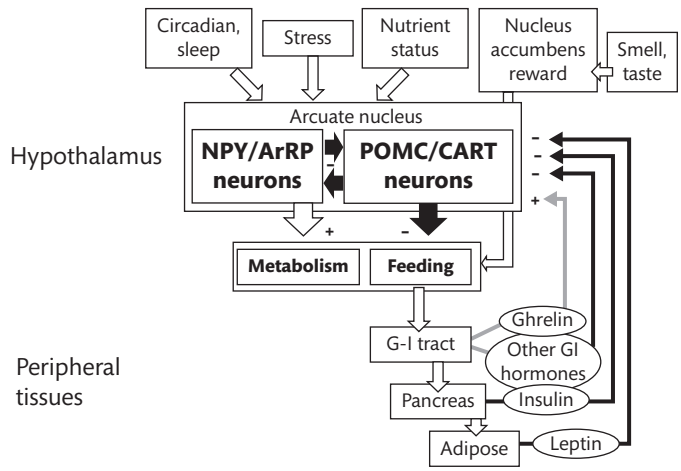


Figure 2.1.8 Feeding and metabolism control system.

hypothalamus (VMH) and dorsomedial hypothalamus (DMH), and the integrative centre in the PVN to provide a functional network that normally maintains nutritional homeostasis. However, it is also now clear that the forebrain nucleus accumbens reward centre, which responds to food taste, food novelty, energy content, and social pressures can override the hypothalamic homeostatic control of feeding—hence the current western world obesity epidemic. Mutations in leptin, the leptin receptor, POMC processing, and the MC4 receptor for alpha-MSH all produce major obesity syndromes; mutations in the other signalling systems have more minor effects because of compensation by other of the signals involved. Recently, a genome-wide scan identified the FTO (fat and obesity) gene as a major component of the polygenic control of obesity. Unsurprisingly, the FTO gene is expressed particularly in the arcuate nucleus and PVN. Hypothalamic mechanisms controlling feeding inevitably also influence the autonomic and TRH systems controlling metabolism, and the systems controlling growth and reproduction.

Gonadotrophin control [6]: See Figure 2.1.9. Reproductive function requires the pulsatile release of hypothalamic gonadotrophin-releasing hormone (GnRH; 10aa) to control FSH and LH secretion from the pituitary. During reproductive life GnRH is secreted in circhoral pulses; oestradiol speeds up the pulses in the preovulatory period whereas progesterone slows the pulses after ovulation. In primates, unlike other animals, a GnRH surge is not required to generate an LH surge; unvarying infusion of circhoral pulses of GnRH can restore fertility in Kallmann syndrome. Most GnRH neurons are located in the infundibular (arcuate) hypothalamus though scattered cells are located more anteriorly. Their essential projection is to the capillaries of the median eminence to stimulate pituitary gonadotrophs via the portal veins. Although sex hormones have feedback effects directly on pituitary gonadotrophs, the generation of a preovulatory LH surge in primate females requires a 'positive feedback' action of oestradiol in the hypothalamus. The action of progesterone to slow GnRH pulses after ovulation also occurs in the hypothalamus. However, the relevant oestradiol and progesterone receptors are located, not on GnRH neurons, but on kisspeptin (KP) neurons which are also mostly located in the arcuate nucleus. They stimulate the activity of GnRH neurons and play a vital role in the pulsatile secretion of GnRH. The failure of either GnRH or KP action causes hypothalamic hypogonadism and infertility, and constitutive activity of mutant KP receptors on GnRH neurons is a cause of precocious puberty. There appear to be at least two population of KP neurons, one of which is stimulated, and one inhibited by oestradiol; both also express dynorphin and neurokinin B. Kisspeptin neurons express leptin receptors and provide a link between low body weight, delayed puberty, and reproductive problems. Experimentally, stresses and prolactin can reduce KP expression and action suggesting a role for KP in mediating the suppression of reproductive function in hyperprolactinaemia and stress. Clinical studies in assisted reproduction suggest that administration of KP may confer less risk of ovarian hyperstimulation than hCG in assisted.

Prolactin control [7]: See Figure 2.1.10. Arcuate nucleus dopamine neurons (tubero-infundibular neurons) have long been known to project to median eminence capillaries and form the major negative control on prolactin (PRL) secretion from pituitary lactotrophs. About half also release GABA within the arcuate

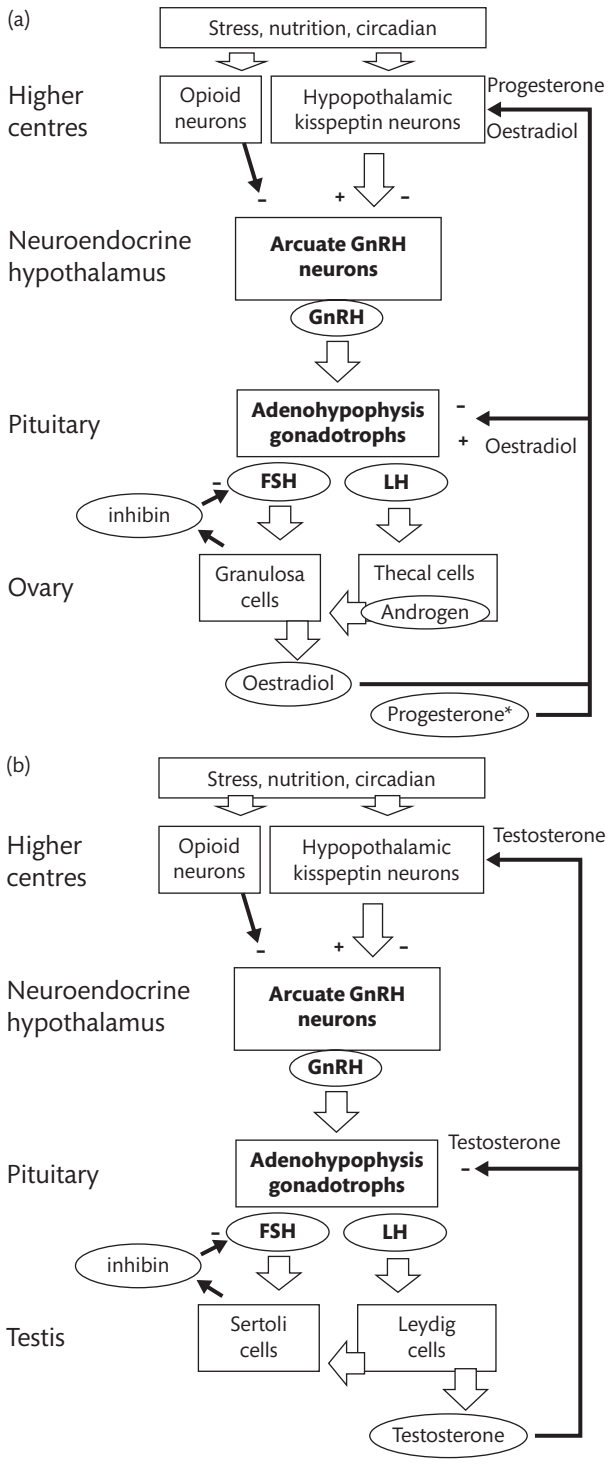


Figure 2.1.9 Reproductive control system.

nucleus and themselves express the D2 and PRL receptors creating an important a short-loop inhibitory feedback system. Humans with an inactivating mutation of the PRL receptor are hyperprolactinaemic. Exogenous administration of TRH and hypoglycaemia resulting from an insulin tolerance test both stimulates PRL secretion. Neither hypoglycaemia nor TRH are physiological regulators of systemic PRL but the TRH effect may explain the hyperprolactinaemia seen in hypothyroidism. Like GH, PRL is secreted in pulses, especially during stage 3 sleep. Circulating PRL is

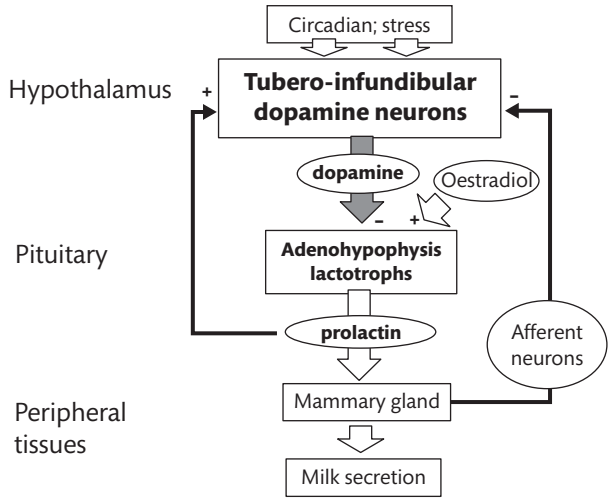


Figure 2.1.10 Growth control system.

generally maintained at low levels in men and non-pregnant women but, during pregnancy, increasing oestradiol stimulates significant lactotroph hyperplasia and prolactin secretion for mammary gland growth. During lactation, the suckling stimulus not only stimulates oxytocin secretion to cause milk ejection, but also inhibits dopamine release and thereby stimulates release of PRL to replenish mammary stores of milk. It is now clear that, like magnocellular neurons and substantia nigra dopamine neurons, arcuate dopamine neurons also release dopamine from their somata and dendrites to influence appetite, growth, and reproductive circuits in the arcuate nucleus. Hyperprolactinaemia resulting from lactotroph adenomas is well known to cause infertility in women and impotence in men through both central and systemic effects. Evidence showing widespread distribution of PRL receptors in the brain and transport of PRL into the brain via its vasculature suggest that many neural and neuroendocrine functions can be influenced by PRL.

Growth hormone control [8]: See Figure 2.1.11. The secretion of growth hormone (GH) is controlled by two hypothalamic hormones: growth hormone-releasing hormone (GHRH; 44aa) and somatostatin (14aa) and by the negative feedback of both GH and IGF-1 at the pituitary and hypothalamus. Both GHRH and somatostatin are also produced widely in splanchnic tissues (GHRH was first isolated from a pancreatic adenoma). GHRH stimulates both the synthesis and release of GH; somatostatin inhibits GH release but not GH synthesis; both are secreted in pulses. Shorter analogues of GHRH are used clinically. GH is normally released in pulses (8–12/day) with a large pulse during sleep, particularly in males; the generation of a GH pulse requires the coincidence of a GHRH pulse when somatostatin release is low. GHRH neurons are situated ventrolaterally in the arcuate nucleus; somatostatin neurones are more widespread, although the most are located in the periventricular zone of the AHA and infundibulum/arcuate nucleus. In addition to inhibiting GH release, hypothalamic somatostatin also inhibits TSH release. Like GnRH neurons, most GHRH neurons are fusiform with processes emerging from each pole; the processes of both the GHRH and somatostatin neurons form a dense network in the arcuate nucleus and contact the portal capillaries in the median eminence. Somatostatin fibres form multiple inhibitory contacts with GHRH perikarya, but GHRH fibres only rarely contact somatostatin perikarya. Feedback by GH to the hypothalamus is primarily negative feedback on somatostatin release. GH secretion is also regulated by peripheral signals such as T3, cortisol (inhibitory), androgens (stimulatory), nutrients such as glucose and fatty acids and metabolites. Hypoglycaemia and low free fatty acids (e.g. in starvation or as a result of an insulin tolerance test (ITT)) stimulate, whereas glucose infusion inhibits GHRH and thus GH release. In contrast, arginine and other amino acids stimulate GH release via hypothalamic sensors. Somatotroph adenomas cause excessive growth in children, acromegaly in adults. Although pulsatile secretion of GH stimulates growth (it increases at puberty and declines in old age) persistently raised GH causes insulin resistance. Synthetic growth hormone-releasing peptides (e.g. GHRP-6; hexarelin) which act on the ghrelin receptor can be used to stimulate GH secretion. Long-acting and selective somatostatin analogues (octreotide, lanreotide) are used to treat excessive growth and acromegaly.

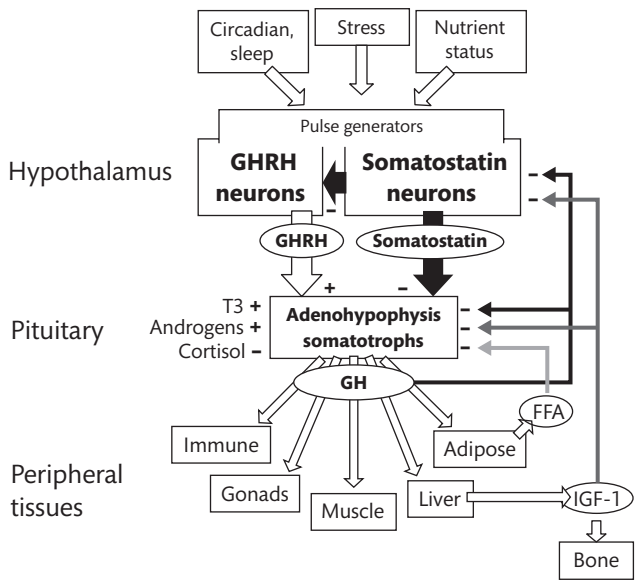


Figure 2.1.11 Prolactin control axis.

Ventromedial (VMH) Nucleus

The VMH forms a fairly well-defined group of neurons dorsal to the arcuate nuclei. Many fibres which pass between the arcuate and paraventricular nuclei pass through it, so it is not surprising that VMH lesions cause hyperphagia and hyposexuality as in Fröhlich's syndrome. Nuclear steroidogenic factor 1 plays a key role in development of the VMH and is exclusively expressed there. Like the arcuate nucleus, VMH neurons are involved in metabolic control, but VMH neurons do not produce specific peptides and its functions have been more difficult to define, though it plays a key role in controlling the sympathoadrenal response to hypoglycaemia. Experimentally, about half the VMH neurons are either stimulated or inhibited by glucose, many express the intracellular energy sensor AMP-kinase (AMPK) and many express insulin, leptin, ghrelin, and cannabinoid receptors. Other parts of the VMH express oestrogen receptor alpha and are involved in autonomic responses to aggressive or sexually attractive situations.

Dorsomedial Nucleus (DMH)

Like the VMH, the DMH is involved in the integration of many aspects of feeding, and in autonomic stress responses. It receives direct and indirect inputs from the SCN and sends projections to the sleep-promoting part of the preoptic nucleus and the orexin neurons of the LHA which promote wakefulness. It receives both neural and hormonal input from other areas important in the regulation of feeding and its activity increases at the expected time of a meal. Inputs from the amygdala signal acute stress and DMH efferents to the ventrolateral medulla and raphe nuclei mediate DMH cardiovascular effects via projections to spinal sympathetic preganglionic neurons.

Posterior Zone (Mamillary Region): Limbic and Autonomic Control

The posterior zone of the hypothalamus comprises the poorly defined posterior hypothalamic area and the well-defined mammillary bodies with various subnuclei. The posterior hypothalamus contains histaminergic neurons and orexin neurons activated in arousal. It receives inputs from diverse areas including the limbic forebrain and mid- and hind-brain areas involved in nociception. Stimulation of the posterior hypothalamus activates sympathetic responses.

Mammillary Bodies

The **mammillary bodies** are well-known as a target of hippocampal fornix fibres and an integral part of 'Papez circuit' with a role in memory. Lesions in this part of the diencephalon cause Korsakoff syndrome. The mammillary bodies comprise two main nuclei: medial and lateral. All mammillary neurons contribute axons to the mammillo-thalamic tract (MTT) that terminates in the anterior thalamus; lesions of the MTT cause anterograde amnesia and poor spatial memory. Mammillary neurons are connected with tegmental motor nuclei via the mammillo-tegmental tract.

Fibre Systems Connecting the Various Nuclei of the Hypothalamus

As would be expected of a part of the brain which integrates many different physiological functions, the various components of the hypothalamus have extensive, complex, and usually reciprocal connections with many regions of the forebrain, brainstem, and spinal cord. There are few large myelinated fibre tracts and only relatively recently have a variety of tracing techniques elucidated the more subtle connections.

Information about the light–dark cycle is passed to the hypothalamus via **retino-hypothalamic** fibres that originate from retinal ganglion cells and project to the suprachiasmatic nuclei. Olfactory inputs reach the hypothalamus from the pyriform cortex, amygdala, and orbitofrontal cortex; gustatory inputs from the brainstem and orbitofrontal cortex; both are involved in arousal and feeding behaviour. Visual and auditory inputs affect the hypothalamus after their significance has been assessed by the limbic system; nociceptive inputs influence hypothalamic function via brainstem connections and the **medial forebrain bundle** (MFB). The hypothalamus (especially the paraventricular nucleus) receives information from the viscera via ascending projections from the solitary nucleus of the medulla, which itself receives sensory input from the facial, glossopharyngeal, and vagus nerves. The MFB is actually a series of fibre bundles that run through the lateral part of the hypothalamus, interconnecting basal forebrain olfactory cortex, septum, and brainstem. The mesolimbic pathway of dopamine fibres from the ventral tegmental area to the nucleus accumbens is a component of the MFB and of the reward system. Together with mesocortical dopamine fibres it forms one of the 'fountains' that spread diffusely throughout the forebrain from the locus coeruleus (noradrenaline), raphe nuclei (serotonin), and basal nucleus of Meynert (acetylcholine). Brainstem reticular afferents from the central grey of the midbrain reach the hypothalamus via the **dorsal longitudinal fasciculus**.

The **fornix** is a large bundle that connects the hippocampal formation with the septal cortex, anterior thalamus, hypothalamus, and mamillary bodies. Most of its fibres originate from pyramidal hippocampal neurons. Commissural fibres link the two hippocampi. Anteriorly, precommissural fibres pass rostral to the anterior commissure to the septal nuclei, the preoptic and dorsal hypothalamic areas; postcommissural fibres pass caudally through the lateral part of the hypothalamus to the medial mammillary nucleus. The fornix also carries cholinergic axons from the basal nucleus to the hippocampus; degeneration of these is a feature of Alzheimer disease. The nuclei of the amygdala are connected to many parts of the hypothalamus by two distinct pathways: the **stria terminalis** (which parallels the course of the fornix) ends in the bed nucleus of the stria terminalis and medial hypothalamus; the **ventral amygdalofugal pathway** links the basolateral amygdala to the lateral hypothalamus and MFB. These convey olfactory and visceral information to the hypothalamus and are important in mediating autonomic and endocrine responses particularly to aversive stimuli ('fear reaction').

Apart from its neuroendocrine efferents, most neural efferent connections of the hypothalamus are reciprocal to its afferents. Fibre tracts link hypothalamic autonomic centres to brainstem and spinal cord preganglionic autonomic neurons; descending fibres controlling sympathetic activity pass through the lateral medulla and lateral medullary (Wallenberg) lesions cause a Horner syndrome. Neurons of the posterior part of the PVN project oxytocin fibres to the sex dimorphic nuclei of the lumbosacral cord.

REFERENCES

1. Gamble KL, Berry R, Frank, Young ME. Circadian clock control of endocrine factors. *Nat Rev Endocrinol*, 2014; **10**: 466–75.
2. Iovino M, Guastamacchia E, Giagulli VA, Lichelli B, Iovino E, Triggiani V. Molecular mechanisms involved in the control of neurohypophyseal hormones secretion. *Curr Pharmac Design*, 2014; **20**: 6702–13.
3. Herman JP, McKiveen JM, Ghosal S, *et al*. Regulation of the hypothalamic-pituitary-adrenocortical stress response. *Compr Physiol*, 2016: **6**: 603–21.
4. Fekete C, Lechan RM. Central regulation of hypothalamic-pituitary-thyroid axis under physiological and pathophysiological conditions. *Endocr Revs*, 2014; **35**: 159–94.
5. Morton GJ, Meek TH, Schwartz MW. Neurobiology of food intake in health and disease. *Nature Revs Neurosci*, 2014; **15**: 367–78.

6. Skorupskaite K, George JT, Anderson RA. The kisspeptin-GnRH pathway in human reproductive health and disease. *Human Reprod Update*, 2014; **20**: 485–500
7. Grattan DR. 60 years of neuroendocrinology. The hypothalamo-prolactin axis. *J Endocrinol*, 2015; **226**: T101–22.
8. Murray PG, Clayton PE. Endocrine control of growth. *Am J Med Genet Part C Semin Med Genet*, 2013; **163C**: 76–85.

FURTHER READING

Leng G. *The Heart of the Brain: The Hypothalamus and Its Hormones*. London: MIT Press, 2018.

Mai JK, Paxinos G. *The Human Nervous System*. San Diego, CA: Elsevier, 2012. [See chapters 16 (hypothalamus), 17 (hypophysis), 18 (circumventricular organs), 21 Sex differences in the forebrain), 22 (Amygdala).]

Molitch ME. Diagnosis and treatment of pituitary adenomas. A review. *J Am Med Assoc*, 2017; **317**: 516–24.

Mollard P, Hodson DJ, Lafont C, Rizzoti K, Drouin J. A tridimensional view of pituitary development and function. *Trends Endocrinol Metab*, 2012; **23**: 261–9.

Pearson CA, Placzek M. Development of the medial hypothalamus: forming a functional hypothalamic-neurohypophyseal interface. *Curr Topics Dev Biol*, 2013; **106**: 49–88.

2.2

The Neurohypophysis

Stephen G. Ball

Neuroanatomy, Molecular Biology, and Physiology of the Neurohypophysis

The neurohypophysis is a complex neurohumoral system with a key role in body fluid homeostasis and reproductive function. This chapter will concentrate on the physiology and pathophysiology of the two hormones made by the neurohypophysis, vasopressin (VP), and oxytocin (OT); outlining the roles of both hormones together with the molecular, cellular, and anatomical basis of their regulation and action.

The Neurohypophysis: Neuroanatomy

The neurohypophysis consists of three parts which together form a functional unit; the supraoptic and paraventricular nuclei of the hypothalamus, containing the cell bodies of the magnocellular, neurosecretory neurones that synthesize and secrete VP and OT; the supraoptico-hypophyseal tract, which includes the axons of these neurones; and the posterior pituitary, where the axons terminate on capillaries of the inferior hypophyseal artery (Figure 2.2.1).

The supraoptic nucleus (SON) is situated along the proximal part of the optic tract. It contains the cell bodies of discrete vasopressinergic and oxytocic magnocellular neurosecretory neurones which in turn, project to the posterior pituitary along the supraoptico-hypophyseal tract. In man, vasopressinergic neurones are found in the ventral SON, with oxytocic neurones situated dorsally. The paraventricular nucleus (PVN) also contains discrete vasopressinergic and oxytocic magnocellular neurones which also project to the posterior pituitary along the supraoptico-hypophyseal tract. In man, magnocellular neurones of the PVN synthesizing VP are found centrally in the nucleus, with oxytocic neurones in the periphery. The PVN contains additional, smaller parvicellular neurones projecting to the median eminence and additional extrahypothalamic areas including forebrain, brainstem, and spinal cord. A group of those projecting via the median eminence cosecrete VP and corticotrophin releasing hormone (CRH), and terminate in the hypophyseal-portal bed of the anterior pituitary. Together with other hypothalamic CRH neurones, these have a role in the regulation of adrenocorticotropin (ACTH) release.

VP magnocellular neurons in the SON and PVN coexpress the peptide apelin and the apelin receptor. Apelin is a 36 amino-acid peptide, though shorter forms (apelin-17 and apelin-13) have also been found that retain biological activity. Intracerebroventricular injection of apelin-17 inhibits the phasic firing of VP magnocellular neurons, reducing VP release. Hypertonic stress and water loading have reciprocal effects on plasma VP and apelin concentrations. Apelin receptors are also coexpressed in VP target cells in the renal collecting duct. VP and apelin are thus regulated in opposite directions and a 'yin and yang' relationship has been proposed for VP and apelin as of physiological regulators [1].

The posterior pituitary receives an arterial blood supply from the internal carotid artery via the inferior hypophyseal artery and the artery of the trabecular, which is a branch of the superior hypophyseal artery. The SON and PVN receive an arterial supply from the suprahypophyseal artery and from the anterior communicating, anterior cerebral, posterior communicating, and posterior cerebral arteries via the circle of Willis. Venous drainage of the neurohypophysis is through the dural, cavernous, and inferior petrosal sinuses.

Molecular and Cell Biology of Vasopressin and Oxytocin

Mammalian VP and OT are nonapeptides containing a closed loop, held together through a disulphide bridge between the cysteine residues at positions 1 and 6 (Figure 2.2.2). Most mammals have the amino-acid arginine at position 8. In the pig family, arginine is substituted by lysine. OT differs from VP by two amino acids. OT has isoleucine rather than phenylalanine at position 3; and leucine rather than arginine at position 8. The similarities between VP and OT probably reflects derivation from a common ancestral gene [2–4].

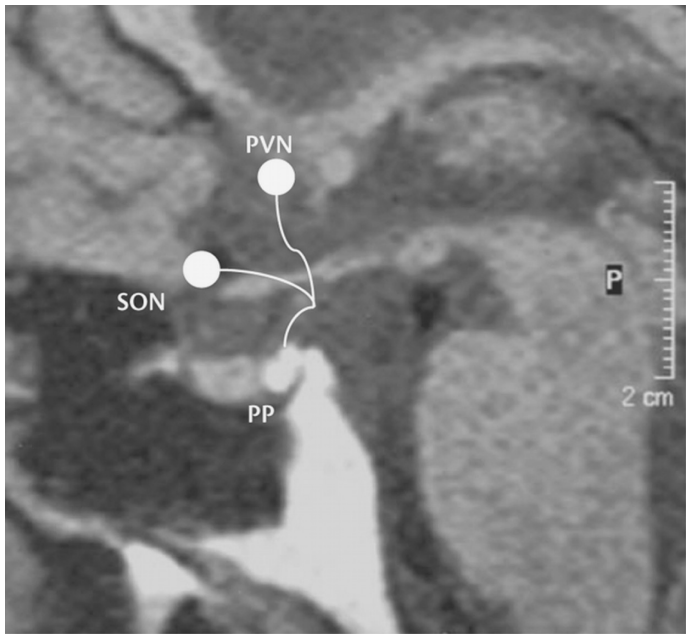


Figure 2.2.1 The neurohypophysis. MRI with overlay demonstrating relative positions of the paraventricular nucleus (PVN), supraoptic nucleus (SON) connecting to the posterior pituitary (PP) via the supraoptico-hypophyseal tract.

The Vasopressin and Oxytocin Genes

In man, the genes encoding VP and OT lie in head to head tandem array on chromosome 20p13, separated by some 12 Kb of DNA. Both are composed of three exons, encoding a polypeptide precursor with a common modular structure: an amino-terminal signal peptide; the VP or OT peptide; a hormone-specific mid-molecule peptide termed a neurophysin (NPI and NPII for OT and VP, respectively); and a carboxyl-terminal peptide known as copeptin (Figure 2.2.3). There is considerable homology between the NP sequences of both genes, positions 10–74 of the NP sequences being highly conserved at the amino acid level.

Regulatory control of *VP* gene expression is mediated through positive and negative elements in the proximal promoter. Several transcription factors bind to these elements. AP1, AP2, and CREB stimulate VP gene expression, while the glucocorticoid receptor (GR) represses expression. Cell-selective gene expression is mediated though regulatory elements in the intergenic region between the *VP* and *OT* genes, with the region –288 to –116 5' upstream of the *VP* gene promoter conferring magnocellular neuron specific expression of the *VP* gene. *VP* gene expression is also be regulated post-transcriptionally. Stability of VP mRNA is influenced by poly(A) tail length. Poly(A) tail length increases in response to water deprivation, increasing stability and enhancing translation. VP mRNA also contains a dendritic localization sequence (DLS). Interaction of the DLS with a multifunctional poly(A) binding protein (PABP) plays a further key role in ribonucleic acid (RNA) stabilization, initiation of translation, and translational silencing [2, 3].

Synthesis, Release, and Metabolism of Neurohypophyseal Hormones

Synthesis of VP and OT occur in the cell bodies of VP and OT-specific magnocellular neurosecretory neurones of the SON and PVN. Generation of both hormones entails substantial post-translational modification of a large primary precursor. Following translation, the carboxy terminal domains are glycosylated, and the precursors packaged in vesicles of the regulated secretory pathway. These migrate along their respective neuronal axons toward the nerve terminals in the posterior pituitary. Migration is microtubule dependent. During vesicle translocation, VP and OT precursors are cleaved by basic endopeptidases to produce the mature hormone and the respective NPs. These are stored as a complex in secretory granules within the nerve terminals of the posterior pituitary. An increase in the firing frequency of vasopressinergic and oxytocic neurones opens voltage-gated Ca_2^+ channels in the nerve terminals which results in fusion of the neurosecretory granules with the nerve terminal membrane and release of their contents into the circulation. The hormone and its NP are cosecreted into the systemic circulation in equimolar quantities. Apart from acting as carrier proteins for VP and OT during axonal migration, NPs appear to serve no specific biological function.

VP and OT circulate free, unbound to plasma proteins. However, VP does bind to specific receptors on platelets. VP concentrations in platelet-rich plasma are thus about fivefold higher than in platelet-depleted plasma. VP and OT have short are short half-life in the circulation: VP half-life being 5–15 minutes. VP and OT are degraded by several endothelial and circulating endo- and amino-peptidases. A specific placental cysteine aminopeptidase degrades VP and OT rapidly during pregnancy and the immediate postpartum period.

The Physiology of Vasopressin

Actions of Vasopressin

VP is the key regulator of renal water excretion and the hormone has a central role in fluid balance. However, the physiology of VP encompasses a wider context within the integrated neuro-humoral response to changes in cardiovascular status and stress [5–7].

Vasopressin Receptors

VP action is mediated by binding to specific receptors on the cell membrane of target cells. There are three VP receptor (V-R) subtypes, encoded by different genes (Table 2.2.1). All have seven

	Amino acid position 1 2 3 4 5 6 7 8 9	Distribution
Arginine vasopressin: Cys-Tyr-Phe-Glu(NH_2)-Asp(NH_2)-Cys-Pro-Arg-Gly(NH_2)		Most mammals
Lysine vasopressin	Phe Glu(NH_2) Lys	Pig family
Oxytocin	Ile Glu(NH_2) Leu	Mammals, birds

Figure 2.2.2 Amino acid sequences of vasopressin and oxytocin.

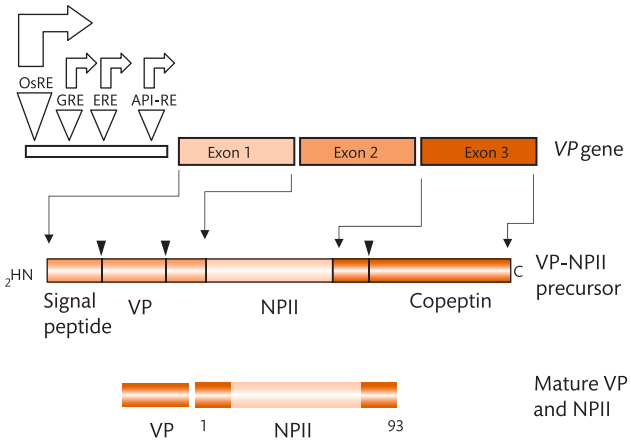


Figure 2.2.3 Functional organization of the VP gene. The VP gene consists of three exons encoding a large precursor which is cleaved to produce the mature peptide through post-translational modification. The VP 5′-promoter contains a number of response element sites that interact with transcription factors regulating VP gene expression. Os-RE, GRE, ERE, and API-RE represent the response elements for osmoregulation: the glucocorticoid receptor, the oestrogen receptor, and AP1, respectively.

transmembrane spanning domains, and all are G-protein coupled. They differ in tissue distribution, signal transduction mechanisms, and function. There is 70–80% human–rat subtype homology at the amino acid level. The human V2-R gene has been mapped to Xq28. The murine V2-R gene maps to a syntenic X-chromosome locus.

Renal Effects of VP

Although VP has multiple actions, its principle physiological effect is in the regulation of water reabsorption in the distal nephron. The hair-pin structure and electrolyte transport processes of the nephron allow the kidney to both concentrate and dilute urine in response to the prevailing circulating VP concentration. Active transport of solute out of the thick ascending loop of Henle generates an osmolar gradient in the renal interstitium which increases from renal cortex to inner medulla, a gradient through which distal parts of the nephron pass en route to the collecting system. This is the basis of the renal countercurrent osmolar exchange mechanism. The presence of selective water channel proteins (aquaporins) in the wall of the distal nephron allows reabsorption of water from the duct lumen along an osmotic gradient, and excretion of concentrated urine

A wide range of aquaporins (AQPs) has been identified. They act as passive pores for small substrates and are divided into two families: the water only channels; and the aquaglyceroporins that can conduct other small molecules such as glycerol and urea. As with other membrane channels, specific structural arrangements within the primary, secondary, and tertiary structure convey the three functional characteristics of permeation, selectivity, and gating. AQPs have a common structure consisting of two tandem repeats (each formed from three transmembrane domains), together with two highly conserved loops containing the signature motif asparagine-proline-alanine (NPA). AQP channels are formed by a tertiary complex of homo-tetramers within the cell membrane, providing four functionally independent pores with an additional central pore. Water can pass through all the four independent channels of water-permeable AQPs. The central pore may act as independent channel in some AQPs [8–10].

Seven AQPs are found in the kidney: AQP1–4 and AQP6–8. AQP1 is constitutively expressed in the proximal tubule and descending loop of Henle, where it facilitates isotonic fluid movement. Loss of function mutations of AQP1 in man lead to defective renal water conservation. AQP3 and AQP4 are both expressed in the basolateral membrane of collecting duct cells in the distal nephron, where they facilitate the movement of water from the collecting duct cells into the renal interstitium. AQP2 is expressed on the luminal surface of collecting duct cells and is the water channel responsible for VP-dependant water transport from the lumen of the nephron into the collecting duct cells. The binding of VP to V2-Rs on the interstitial surface of collecting duct cells produces increases in expression of AQP2. The increase is biphasic: one requiring new AQP2 protein synthesis and one which is independent of new protein synthesis. V2-R activation triggers an intracellular phosphorylation cascade leading activation of the nuclear transcription factor

Table 2.2.1 Vasopressin receptor subtypes

	Vasopressin receptor		
	V1a	V2	V3
Expression	• Vascular smooth muscle • Liver • Platelets • CNS	Basolateral membrane of distal nephron	Pituitary corticotroph
Amino acid structure	418 amino acids (human)	370 amino acids (human)	424 amino acids (human)
Second messenger system	Gq/11mediated phospholipase C activation: Ca^{2+}, inositol triphosphate & diacyl glycerol mobilization	Gα*s mediated adenylate cyclase activation: cAMP production and protein kinase A stimulation	As V1a
Physiological effects	• Smooth muscle contraction • Stimulation of glycogenolysis. • Enhanced platelet adhesion • Neurotransmitter and neuromodulatory function	Increased production and action of aquaporin-2	Enhanced ACTH release

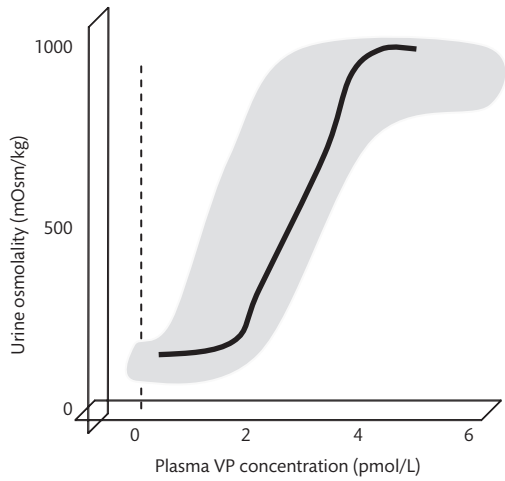


Figure 2.2.4 The relationship of plasma vasopressin concentration to urine concentrating ability. There is a sigmoid relationship between plasma vasopressin concentration and urine osmolality, with maximum urine concentration occurring at plasma vasopressin concentrations of 4–6 pmol/L. There is a range of response in the normal population depicted by the grey area, within which an individual response is demonstrated.

CREB and expression of c-Fos and stimulation of AQP2 gene expression through CRE and AP-1 response elements in the *AQP2* gene promoter. V2-R activation also stimulates an immediate increase in AQP2 activity by accelerating trafficking and assembly of presynthesized AQP2 protein into functional, homo-tetrameric water channels in luminal cell membranes [4–6].

As VP levels rise, there is a sigmoid relationship between plasma VP concentration and urine osmolality, with maximum urine concentration achieved at plasma VP concentrations of 3–4 pmol/L (Figure 2.2.4). Following persistent VP secretion, antidiuresis may diminish. Downregulation of both V2-R function and AQP2 expression may be responsible for this escape phenomenon. Maximum diuresis occurs at plasma VP concentrations of 0.5 pmol/L or less.

VP has additional effects at other sites in the nephron, which serve to augment renal concentrating ability. VP decreases medullary blood flow, stimulates active urea transport in the distal collecting duct and stimulates active sodium transport into the renal interstitium. VP also stimulates a bi-phasic up-regulation of bumetanide-sensitive sodium-potassium-chloride cotransporter (SLC12A1) expression in the thick ascending loop of Henle. VP both accelerates post-translational processing/trafficking of presynthesized SCLC12A1 and increases *SCLC12A1* gene expression [7].

Cardiovascular Effects of VP

VP is a potent pressor agent, its effects mediated via a specific membrane receptor (V1a-R). Systemic effects on arterial blood pressure are only apparent at high concentrations due to compensatory buffering haemodynamic mechanisms. Nevertheless, VP is important in maintaining blood pressure in mild volume depletion. The most striking vascular effects of VP are in the regulation of regional blood flow. The sensitivity of vascular smooth muscle to the pressor effects of VP vary according to the vascular bed; vasoconstriction of splanchnic, hepatic, and renal vessels occurring at VP concentrations close to the physiological range. Furthermore, there are differential pressor responses within a given vascular bed; selective effects on intrarenal vessels resulting in redistribution of renal blood flow from medulla to cortex. Such effects suggest that baroregulated VP release constitutes one of the key physiological mediators of the integrated haemodynamic response to volume depletion.

Effects of VP on the Pituitary

VP is an ACTH secretagogue, acting through pituitary corticotroph-specific V3-Rs. Though the effect is weak in isolation, VP and C-reactive protein (CRF) act synergistically. VP and CRF colocalize in neurohypophyseal parvicellular neurones projecting to the median eminence and the neurohypophyseal portal blood supply of the anterior pituitary. Levels of both VP and CRF in these neurones are inversely related to glucocorticoid levels, clearly suggesting a role in feedback regulation.

Behavioural Effects of VP

Vasopressinergic fibres and V-Rs are present in many areas of the brain, including the cerebral cortex and limbic system. These extensive neural networks are anatomically and functionally independent of the neurohypophysis. In rodents, these central vasopressinergic systems have key roles in mediating complex social behaviour such as mating patterns. There are similar emerging data in man. Association studies link *V1a-R* gene sequence variation with a range of behaviours. Dysregulation of central VP action may be a distal end point in conditions characterized by complex altered social and emotional behaviour [8–12].

Effects of VP on the Skeleton

Emerging data support a role for neurohypophyseal hormones in bone physiology [13]. VP inhibits osteoblast formation. This effect is opposed by OT, which has a reciprocal effect in stimulating osteoblast formation. Mice deficient in V1a-R (*Avpr1a*$^{-/-}$) have a high bone mass while both haploinsufficiency for OT and deletion of the OT receptor (*Oxr*$^{-/-}$) result in osteopaenia.

Miscellaneous Effects of VP

Several additional actions of VP are listed in Table 2.2.2.

Osmoregulation of in VP Release

In line with the hormone's key role in body fluid homeostasis, plasma osmolality is the most important determinant of VP secretion. Working in tandem, the combined osmoregulatory system for thirst and VP secretion maintain plasma osmolality within the narrow limits of 284–295 mOsml/kg. The osmoregulation of VP release and the physiological relationship between plasma osmolality

Table 2.2.2 Miscellaneous effects of VP

Action	Receptor involved
Coagulation/clotting cascade	
Factor VII release from hepatocytes	V-2r
von Willebrand factor release from vascular endothelium	V-2r
Liver metabolism	
Glycogen phosphorylase A activation	V-1ar

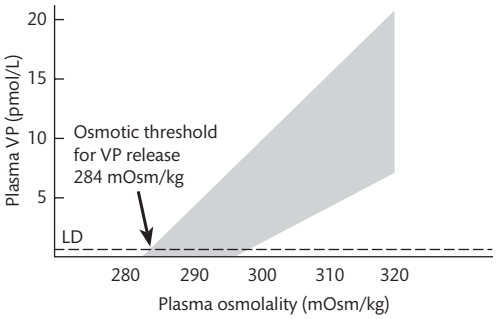


Figure 2.2.5 Relationship between plasma osmolality and plasma vasopressin (VP) concentration during progressive hypertonicity induced by infusion of 855 mmol/L saline in a group of healthy adults. LD represents the limit of detection of the assay, 0.3 pmol/L.

and plasma VP concentration is described by three characteristics; the linear relationship between plasma osmolality and plasma VP concentration; the osmotic threshold or 'set point' for VP release: and the sensitivity of the osmoregulatory mechanism.

Increases in plasma osmolality increase plasma VP concentrations in a linear manner (Figure 2.2.5). The abscissal intercept of this regression line, 284 mOsml/kg, indicates the mean 'osmotic threshold' for VP release: the mean plasma osmolality above which plasma VP starts to increase. Though there is no level of plasma osmolality below which VP release is completely suppressed, though such low levels of VP have little antidiuretic effect. The concept of a threshold of VP release thus remains a pragmatic means to characterize the physiology of osmoregulation; VP release being increased from a basal rate by activation of stimulatory osmoreceptor afferents, and decreased to minimal values by removal of this drive and the activation of synergistic inhibitory afferents. The slope of the regression line reflects the sensitivity of osmoregulated VP release. There are considerable interindividual variations in both threshold and sensitivity of VP release. Twin studies indicate a strong heritable component in this variation. However, over time, these parameters are remarkably reproducible within an individual [14].

There are several physiological situations where the tight relationship between plasma osmolality and VP concentration is lost. The act of drinking results in rapid suppression of VP release, independent of changes in osmolality. In addition, the rate of change of plasma osmolality can influence the VP response; rapid increases in plasma osmolality result in exaggerated VP release. The osmotic threshold for VP release is lowered in normal pregnancy, and a similar though smaller change occurs in the luteal phase of the menstrual cycle. Plasma VP concentrations increase with age, together with enhanced VP responses to osmotic stimulation. In contrast, thirst appreciation is blunted and fluid intake reduced. These changes, together with age-related decreases in renal handling of water loads and generation of maximal urine concentration, form the basis for the predisposition of older people to both hyper- and hyponatraemia.

Baroregulation of VP Release

As a principle determinant of fluid homeostasis, VP is a key player in maintaining haemodynamic integrity. In contrast to osmoregulated VP release, progressive reduction in blood pressure produces an exponential increase in plasma VP. Falls in arterial blood pressure of 5–10% are necessary to increase circulating VP concentrations in man. Importantly, baroregulated VP release can occur at low levels of plasma osmolality levels that would normally act to suppress VP production. This apparent 'hierarchy' of regulation is important when considering the integrated physiological response to volume depletion. It is a key pathophysiological mechanism underpinning the hyponatraemia resulting from effective circulating volume depletion.

Humoral Regulation of VP Release

Changes in circulating volume and blood pressure trigger an autonomic and endocrine cascade resulting in a coordinated physiological response. Baroregulated VP responses can be modified by other neurohumoral influences triggered as part of this coordinated response: atrial natriuretic peptide (ANP) inhibiting and norepinephrine augmenting baroregulated VP release. The renin–angiotensin system (activated in effective volume depletion) is also intricately involved in the regulation of VP production. Circulating angiotensin II stimulates VP secretion, acting centrally at forebrain centres influencing SON and PVN activation. In addition, angiotensin II stimulates VP release via a direct effect on VP magnocellular neurones, where type 2 angiotensin II receptors have been identified. In the rat, ANP inhibits both osmo- and baro-stimulated VP release centrally.

Other Regulatory Mechanisms of VP Release

Nausea and emesis are potent stimuli to VP release, independent of osmotic and haemodynamic status. Manipulation of abdominal contents is another powerful stimulus to VP release. Both contribute to the high plasma VP values and consequent impairment of water load excretion, observed after gastrointestinal surgery. VP release in response to these stimuli and others, such as neuroglycopaenia, justify its classification as a stress response hormone.

Thirst

Thirst, and the drinking response to thirst, are key components maintaining fluid homeostasis. While the two systems are discrete, it is no surprise that the physiology of VP and thirst are closely aligned.

There is a linear relationship between thirst, determined by visual analogue scale, and plasma osmolalities in the physiological range (Figure 2.2.6). The mean osmotic threshold for thirst perception is 281 mOsml/kg, similar to that for VP release. Thirst occurs when plasma osmolality rises above this threshold, the intensity varying in relation to the ambient plasma osmolality. The functional characteristics of osmoregulated thirst, just as VP release, remain consistent within an individual on repeated testing, despite wide variations between individuals [14].

As with osmoregulated VP release, there are also specific physiological situations in which the relationship between plasma osmolality and thirst breaks down. The act of drinking reduces osmo-stimulated thirst, just as it does VP release. There is a fall in the osmotic threshold for thirst in the luteal phase of the menstrual cycle. In contrast, thirst appreciation and fluid intake are blunted in older people. Thirst can be stimulated by extracellular volume depletion through volume sensitive cardiac autonomic afferents. In addition, hypovolaemia, and hypotension lead to the generation of circulating and intracerebral angiotensin II, a powerful dipsogen.

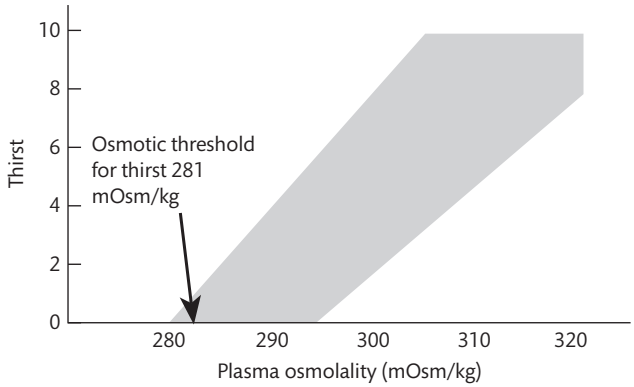


Figure 2.2.6 Relationship between thirst and plasma osmolality during progressive hypertonicity induced by infusion of 855 mmol/L saline in a group of healthy adults.

The Coordinated Regulation of VP and Thirst

The control of VP production and processes underpinning thirst perception and drinking are separate and discrete. Nevertheless, there are close functional links between the two systems: highlighted by the situations where they are activated and suppressed in parallel. It is therefore not surprising that the two systems are regulated by systems that are also closely related. The lamina terminalis within the forebrain is a key hub for integrating neural and humeral influences on VP production, thirst perception, and drinking behaviour.

The Neurobiology of Osmoregulation

The Lamina Terminalis: Anatomy, Neurophysiology, and Role in Central Osmoregulation

The lamina terminalis (LT) is a small region of the forebrain including areas just anterior and dorsal to the third ventricle. It is composed of three, interconnected nuclei: the subfornicular organ (SFO), the organum vasculosum (OVLT) and the median preoptic nucleus (MnPO). Due to local fenestrations in the blood-brain barrier, neurones within the SFO and OVLT are in direct contact with the circulation. They thus function as interoceptive sensory organs. The MnPO lies within the blood-brain barrier. While it does not have a sensory function, it is thought to have a key role in integrating the output of the LT to higher centres, including those involved in initiating drinking [15].

SFO and OVLT neurones have intrinsic osmosensitivity; neuronal firing rates increasing in line with extracellular fluid osmolality. These neurones constitute the principle central osmo-sensing mechanism regulating VP and thirst. Changes in the osmotic environment of these neurons alter their shape. This physical change alters the activity of the cell surface stretch-sensitive cationic channel TRPV1. TRPVI is the transduction mechanism linking osmolality to altered membrane potential and neurone firing frequency. Vasopressinergic neurons of the SON and PVN also express TRPV1: the basis for a direct effect of plasma osmolality on magnocellular neurone activity. Interestingly, osmoregulatory function is not lost in *Trpv*1$^{-/-}$ mice. Additional osmo-sensing pathways must therefore also be in operation [16].

Glutaminergic neurones projecting from the SFO and OVLT stimulate drinking. Glutaminergic pathways from the LT also project to the PVN and SON, providing a neuroanatomical basis for the parallel stimulation of thirst and VP release. In contrast, activation of GABAergic neurones within the LT reduces drinking, highlighting a balance of both positive and negative elements in the central neural circuitry controlling fluid balance (Figure 2.2.7).

Peripheral Osmoregulation

A related, but distinct osmo-sensory input feeds additional data on peripheral osmolar status. Sensory neurones in hepatic portal blood vessels are responsive to changes in the plasma osmolality. The transducing element in this peripheral osmosensitivity is the stretch-sensitive ion channel, TRPV4 [17].

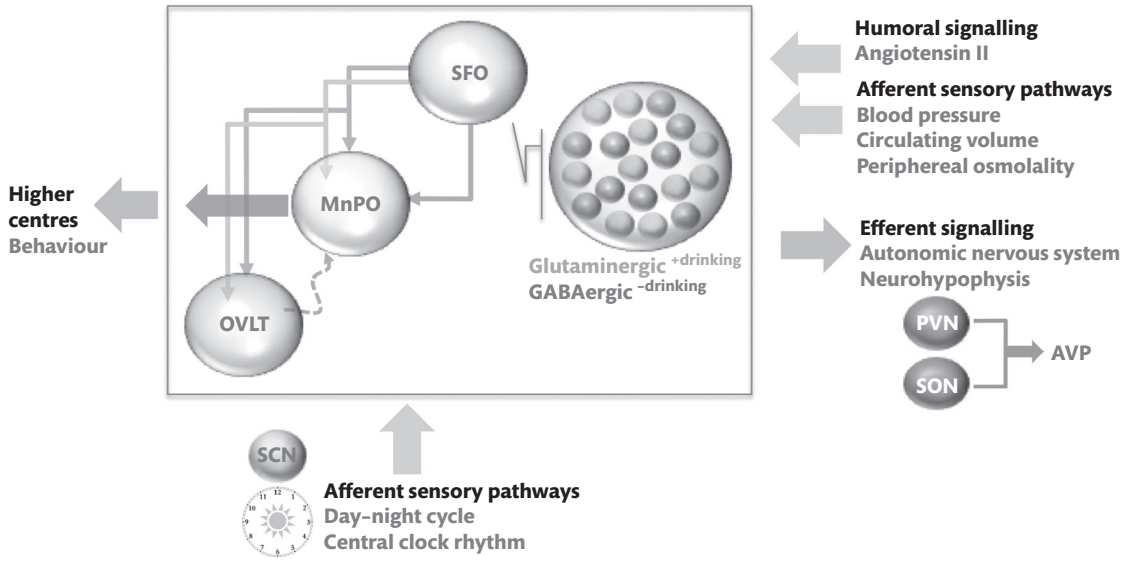


Figure 2.2.7 The role of the lamina terminalis in in the integration of thirst and AVP release. The lamina terminalis is key in the integration of sensory and humoral signals that regulate thirst and AVP release. MnPO, median preoptic nucleus; OVLT, organum vasculosum of the lamina terminalis; SCN, suprachiasmatic nucleus; SFO, subfornicular organ; SON, supraoptic nucleus; PVN, paraventricular nucleus.

Afferent Sensory Regulation of VP and Thirst

Baroregulatory influences on neurohypophyseal VP release derive from aortic arch, carotid sinus, cardiac atrial, and great vein stretch-sensitive afferents via cranial nerves IX and X. Ascending projections are via the nucleus tractus solitarius (NTS). From the NTS, further afferents project to the SON and PVN and the LT. These mechanisms act to integrate a number of afferent inputs that reflect volume status. Ascending baroregulatory pathways affect tonic inhibition of VP release, as their interruption increases plasma VP concentration.

Osmosensitivity of VP release is influenced by circadian rhythms. VP release increases during sleep, mediated by clock neurones projecting from the suprachiasmatic nucleus to the LT and neurohypophysis [18].

The act of drinking reduces thirst perception. This change occurs before any change in plasma osmolality. The effect is produced through three mechanisms: oropharyngeal sensory afferents; gastrointestinal stretch-sensitive afferents; and peripheral osmoreceptors in the hepatic portal vein. Thirst-promoting neurones in the SFO integrate sensory inputs from the oropharynx (drinking and food composition) with central osmolar status to influence thirst perception. This mechanism effectively allows anticipatory changes in water consumption that can precede changes in plasma osmolality.

Central Mechanisms Integrating Humeral Influences on VP and Thirst

Osmoregulated VP responses are modified by other components of the coordinated neurohumoral response to changes in circulating volume and blood pressure. Angiotensin II amplifies the proportional relationship between osmolality and neuronal firing in the SON. Angiotensin II triggers polymerization of intracellular actin filaments in osmosensitive neurones in the SON, altering cell shape in a manner that works in synergy with that produced by changes in extracellular osmolality. Angiotensin II thus enhances osmosensitivity. This mechanism underpins the changes in osmo-regulated VP release in hypo- and hypervolaemia: the osmotic threshold and sensitivity of VP release is lowered by hypovolaemia; while the converse is found in hypervolaemia and hypertension.

Vasopressinergic neurones within the neurohypophysis also contribute a direct 'short-loop' mechanism for the physiological coordination of VP production. Angiotensin II and relaxin excite both OT and VP magnocellular neurones. In contrast, ANP inhibits VP neuron activity. VP magnocellular neurones themselves have independent osmo-sensing properties. V-Rs are present on vasopressinergic neurons of both the PVN and SON, highlighting the potential for auto-control of VP release through direct osmoregulation and short-loop feedback.

Integrated Physiology of VP Release and Body Fluid Homeostasis

The physiological regulation of water balance is intimately linked with that of circulating volume; common systems are involved in both processes. As sodium is the major cationic osmolyte, the interrelationships of sodium and water excretion with circulating volume regulation are key to appreciating the position of VP and thirst-driven fluid intake in the physiology of fluid homeostasis.

At plasma osmolalities of 285–295 mOsml/kg, osmolar balance can be maintained by VP-dependent regulation of renal water loss. A rise in plasma osmolality within this range produces a progressive increase in plasma VP to a concentration of 3–4 pmol/L, and antidiuresis. Further increases in plasma osmolality stimulate further VP release, but this does not result in any further reduction of renal water excretion. Correction of plasma osmolality back to the range over which osmolar balance can be maintained by VP requires thirst-stimulated drinking. As the osmolar threshold for thirst is similar to that for VP release (284 mOsml/kg), the maintenance of water balance through a combination of osmoregulated VP release and thirst is clearly a seamless, coordinated process of subtle complexity.

If excessive fluid volumes are consumed, greater than those demanded by thirst, plasma VP levels are suppressed to below 0.3 pmol/L, resulting in maximum diuresis of up to 15–20 L/24 h. Ingestion of water in excess of this causes a reduction of plasma osmolality into the subnormal range, and hyponatraemia.

VP release is also regulated by other, non-osmotic stimuli. This complex regulation has a hierarchy, with significant physiological and pathophysiological sequelae. Hypovolaemia shifts the relationship of plasma osmolality with VP concentration to the left. During moderate hypovolaemia, osmoregulation is maintained around a lower osmolar set point. As the degree of hypovolaemia progresses, baroregulated VP release overrides the osmolar set point, and antidiuresis is maintained despite the potential for ensuing hyponatraemia. Coincident activation of the systemic and intracerebral renin–angiotensin systems stimulates drinking and augments VP release, in addition to independent pressor and antinatriuretic effects. The homeostatic response to hypovolaemia thus involves an integrated neurohumoral cascade, of which VP and thirst are key components.

The Physiology of Oxytocin (OT)

OT binds to specific G-protein coupled cell surface receptors (OT-Rs) on target cells to mediate a variety of effects largely concerned with reproductive function. Historically, attention has focused on the role of OT in the regulation of lactation; parturition; and reproductive behaviour. Data from animal models lacking oxytocin because of targeted disruption of the oxytocin gene have challenged this dogma, forcing a review of the physiological roles of the hormone.

Oxytocin and Lactation

In the rat, the stimulation of sensory afferents in the nipple by the act of suckling trigger a reflex cascade leading to synchronized firing bursts of oxytocic magnocellular neurones, with pulsatile OT release corresponding to this burst activity. The released OT acts on OT-Rs on smooth muscle cells lining the milk ducts of the breast, initiating milk ejection. OT is essential for completion of this milk ejection reflex in rodent. Mice lacking OT fail to transfer milk to their suckling young, and this deficit is corrected by injection of OT. In contrast, Women lacking posterior pituitary function can breast feed normally, illustrating that OT is not necessary for lactation in Humans.

Oxytocin and Parturition

OT is a powerful uterotonic agent. In many mammals there is both an increase in OT secretion during parturition, and an increase in uterine responsiveness to OT at term. These data suggest a key role for the hormone in the initiation and progression of labour.

It is believed that falling progesterone concentrations toward the end of pregnancy lead to up-regulation of uterine myometrial OT-Rs, enhanced contractility, and increased sensitivity to circulating OT. Stretching of the 'birth canal' during parturition leads to the stimulation of specific autonomic afferents, triggering increased burst firing of oxytocic magnocellular neurones and OT release. A positive feedback loop is formed, with OT both stimulating uterine contraction further and enhancing the production of local uterotonic mediators. It has been difficult to demonstrate increased circulating OT levels in women during labour, due in part to the difficulties of analysing pulsatile release coupled with the short circulating half-life of the hormone due to the action of placental cysteine aminopeptidase. In mice lacking OT, parturition is normal. Moreover, Women with absent posterior pituitary function can have a normal labour. However, the importance of OT in the birth process is highlighted by the effectiveness of OT antagonists in the management of preterm labour.

OT may have key additional offspring-facing roles in parturition. Maternal OT produces a switch in fetal central nervous system (CNS) neurotransmission with enhanced inhibitory GABAergic signalling. This increases fetal neuronal resistance to hypoxaemic damage that may occur during delivery, part of adaptive mother-fetal signalling during parturition in which OT is a major player [19].

Oxytocin and Behaviour

OT-R expression is widespread in the CNS of many species. There is clear evidence that OT has important influences on reproductive behaviour in rat; facilitating both lordosis and the development of maternal behaviour patterns. However, mice lacking OT exhibit normal sexual and maternal behaviour, indicating these effects may be species dependent. Central oxytocic transmission appears to reduce anxiety and hypothalamo-pituitary-adrenal stress responses in female rats. However, the same central oxytotic function may be required for normal adrenocorticotropin responses to stress. OT release from both dendrites and nerve terminals of hypothalamic magnocellular neurons can be regulated by other neuropeptides, highlighting the potential for magnocellular OT to integrate with central neurotransmission.

Central OT pathways have been implicated complex behaviour: social recognition, affiliation, and social bonding. This clustering has raised interest in the role of OT and OT pathways in the development of autism. A number of association studies have demonstrated linkage between autism and OT-R polymorphisms. Together, these data suggest a complex role for OT in the stress and other behavioural responses, with species and context-dependent differential effects [20, 21].

Integrated Physiology of Oxytocin

How are the proposed roles of OT in reproductive function reconciled with both human and mouse data that highlight normal function in the absence of the hormone? Clearly there are interspecies differences in OT-modulated processes that contribute important qualifications to the data. The mouse gravid uterus does not express OT-Rs, in contrast to human and rat. It is perhaps not surprising therefore, that parturition is normal in the OT null-mouse. Similarly, in contrast to rat, maternal behaviour evolves gradually in mouse, and is not acquired rapidly in the postpartum period. Mouse may therefore not be a good model for the uterine and behavioural effects of OT. Secondly, there is clearly variable redundancy in some of the physiological pathways in which OT is involved. This redundancy may vary between species. The extrapolation of OT's role in normal physiology from those responses found in its absence should thus be made with caution.

Clinical Endocrinology of the Neurohypophysis

There are no recognized clinical sequalae of OT deficiency in man. The pathophysiology of the neurohypophysis thus reflects the physiology of VP and the regulation of water excretion. Defects in VP production or action impact through disturbances in fluid and electrolyte balance. A second, related but less common group of conditions reflect primary defects in thirst. In some cases, the two may coincide, in line with the close anatomical and functional relationship of both processes.

Diabetes Insipidus

Classification

A pragmatic clinical definition of polyuria is the excretion of urine in excess of 3 l/24 h (over 40 ml/kg/24 h) in adults and over 100 ml/kg/24 h in infants). Diabetes Insipidus (DI) simply describes the excretion of large amounts of dilute urine. One of three mechanisms may be responsible.

1. Deficiency of VP: termed hypothalamic diabetes insipidus (HDI).
2. Renal resistance to VP: termed nephrogenic diabetes insipidus (NDI).
3. Inappropriate, excessive water drinking: termed dipsogenic diabetes insipidus (DDI) or primary polydipsia.

A classification of DI based on aetiology is outlined in Box 2.2.1.

Hypothalamic Diabetes Insipidus

HDI (also known as neurogenic, central, or cranial DI) is due to deficient osmoregulated VP secretion. In most cases it is a partial defect, with patients having inappropriately low plasma VP concentrations with respect to concomitant plasma osmolalities. Presentation with HDI implies destruction or loss of function of more than 80% of vasopressinergic magnocellular neurones. Though persistent polyuria can lead to dehydration, given free access to water, most patients can maintain water balance through an intact thirst mechanism. HDI is rare, with an estimated prevalence of 1:25 000 and equal gender distribution.

Aetiology

Most cases of HDI are acquired. Improvements in imaging and an appreciation of the varied presentation of inflammatory/autoimmune forms are responsible for fewer cases being designated idiopathic. Inherited/familial forms account for 5% of HDI.

Trauma, either as a result of head injury or surgery, can produce HDI through damage to the hypothalamus, pituitary stalk, or posterior pituitary. Pituitary stalk trauma may lead to a triphasic disturbance in water balance; an immediate polyuria characteristic of HDI followed within days by a more prolonged period of antidiuresis suggestive of VP excess. This second phase may last up to several weeks, and can be followed by reversion to HDI or recovery. Such

Box 2.2.1 Classification of polyuric syndromes

(A) Hypothalamic diabetes insipidus

- Primary
  - Genetic
    - DIDMOAD (Wolfram) syndrome
    - Autosomal dominant
    - Autosomal recessive
  - Developmental syndromes
    - Septo-optic dysplasia
  - Idiopathic
- Secondary/acquired
  - Trauma
    - Head injury
    - Post surgery (transcranial, transphenoidal)
- Tumour
  - Craniopharyngioma, germ cell tumour, metastases, pituitary macroadenoma
- Inflammatory
  - Granulomata
  - Sarcoidosis, histiocytosis
  - Infection; meningitis, encephalitis
  - Infundibuloneurohypophysitis
- Guillain–Barré syndrome
  - Autoimmune (anti-VP neurone antibodies)
- Vascular
  - Aneurysm
  - Infarction
  - Sheehan's syndrome
  - Sickle cell disease
  - Pregnancy (associated with vasopressinase)

(B) Nephrogenic diabetes insipidus

- Primary
- Genetic
  - X-linked recessive (V2-R defect)
  - Autosomal recessive (AQP2 defect)
  - Autosomal dominant (AQP2 defect)
- Idiopathic
- Secondary
  - Chronic renal disease
  - Polycystic kidneys
  - Obstructive uropathy
- Metabolic disease
  - Hypercalcaemia
  - Hypokalaemia
- Drug induced
  - Lithium
  - Demeclocycline
- Osmotic diuretics
  - Glucose
  - Mannitol
- Systemic disorders
  - Amyloidosis
  - Myelomatosis
- Pregnancy

(C) Dipsogenic diabetes insipidus

- Compulsive water drinking
- Associated with affective disorders
- Drug induced?
- Structural/organic hypothalamic disease
  - Sarcoid
  - Tumours involving hypothalamus
  - Head injury
- Tuberculous meningitis

a 'triple response' reflects initial magnocellular axonal damage; the subsequent unregulated release of large amounts of presynthesized VP; and ultimately, either recovery or development of permanent HDI, as determined by the degree of initial neuropraxia/axonal shearing and damage. Not all phases of the response may be apparent. Recent data suggest acute HDI can occur in up to 22% of non-selected patients presenting with traumatic brain injury (TBI), persisting in some 7% of the total TBI cohort on long-term follow-up [22, 23].

Circulating antibodies to VP secreting neurons can be found in 30% of patients classified previously as having HDI with no identifiable cause, implying an autoimmune aetiology. Presence of anti-VP neurone antibodies in patients with HDI is associated with pituitary stalk thickening on magnetic resonance imaging (MRI). However, anti-VP neurone antibodies can also be found at low prevalence in patients with HDI secondary to histiocytosis X and following pituitary surgery, suggesting the specificity of the test or the auto-antibody response is low. Evidence of wider organ-specific autoimmunity common in isolated HDI [24].

Primary pituitary adenoma rarely causes HDI. Hypothalamic tumours, such as craniopharygioma and germinoma, are more common causes of HDI. Presentation with HDI can precede the radiological appearance of germinoma. Hypothalamic tumours and developmental defects such as septo-optic dysplasia (SOD) account for up to 50% of HDI in children. Hypothalamic or pituitary metastases (e.g. from primary breast or lung cancer) can present with HDI, as can primary brain malignancy and primary CNS lymphoma.

HDI can present in pregnancy. Placental vasopressinase activity can decompensate pre-existing limited antidiuretic capacity in patients with partial HDI through increased VP degradation that cannot be matched by increased hormone release. This can revert to normal after delivery, though permanent HDI may ultimately develop if the natural history of the central defect is progressive.

Wolfram Syndrome (WS)

Genetic studies have identified two subtypes of WS. WS1 is caused by loss of function mutations in the *WFSI* gene on Ch.4p16. *WFSI* encodes an 890 amino-acid glycoprotein, wolframin. Wolframin expression is restricted to the endoplasmic reticulum (ER) where it regulates ER stress and calcium homeostasis. Loss of function of wolframin triggers apoptosis and cell death. Non-inactivating mutations of *WFSI* are associated with autosomal dominant sensorineural hearing loss. WS1 may thus represent one extreme of a spectrum disorder.

WS2 is characterized by optic atrophy and diabetes mellitus. There are additional features of peptic ulcer disease and bleeding tendency which are not seen in WS1. WS2 is caused by loss of function mutations in the *CISD2* gene. *CISD2* encodes a protein expressed on both ER and the outer mitochondrial membrane. Loss of function of CISD2 disrupts calcium flux between ER and mitochondria leading to autophagy and cell death in a manner similar to that seen in several other neurodegenerative diseases [25, 26].

Autosomal Dominant Familial HDI

Autosomal dominant familial HDI is caused by mutations in the *VP* gene on chromosome 20. While it typically presents in childhood, the age of presentation varies considerably, reflecting variation in

the progressive loss of VP secretion. A variety of different missense and nonsense mutations within exons 1, 2, and 3 of the *VP* gene have been identified in affected kindreds. Mutant VP precursors accumulate in the ER of magnocellular neurones, to which they are neurotoxic. This explains the both the progressive loss of VP release in the condition, the range of presentation and its dominant inheritance pattern: the product of a single mutated allele being sufficient to trigger neural degeneration. Growth failure may be an early clinical feature [27].

The inherited HDI of the Brattleboro (BB) rat is due to a frame shift in exon 2 of the *VP* gene, resulting in a VP precursor with an altered carboxy terminus which also accumulates in the ER of vasopressinergic neurones. Interestingly, the HDI of the BB rat is inherited in a recessive manner, in contrast to the equivalent condition in man.

Investigation of Polyuric States

The strategy of investigation of DI is to confirm the polyuric state, define its basis, and to explore possible primary aetiologies. After establishing significant polyuria of greater than 3 l/24 h in adults and excluding hyperglycaemia, hypokalaemia, hypercalcaemia, and significant renal insufficiency, attention should be focused on the VP axis. Direct tests that measure plasma VP in response to osmotic stimulation would be the ideal modality to differentiate HDI from other causes of polyuria. However, access to reliable VP assays have been limited. Because of this, an indirect test using a surrogate endpoint of VP release (the ability of the kidney to produce concentrated urine during osmotic stress) has historically been used to differentiate the cause of confirmed polyuria: the water deprivation test.

Indirect Tests of the Hypothalamo-Neurohypophyseal (AVP) Axis: The Water Derivation Test

The water deprivation test assesses the capacity to concentrate urine during the osmotic stress of controlled water deprivation. The period of water deprivation can be followed by evaluation of the antidiuretic response to exogenous VP: the aim being to confirm renal sensitivity to VP or establish renal resistance. A standard protocol is outlined in **Box 2.2.2**. HDI can be distinguished by urine osmolality less than 300 mOsml/kg, accompanied by plasma osmolality greater than 290 mOsml/kg after dehydration. Urine osmolality should rise above 750 mOsml/kg after desmopressin (DDAVP), indicating normal renal responsiveness. In contrast, failure to increase urine osmolality above 300 mOsml/kg after dehydration together with failure to respond to DDAVP is diagnostic of NDI. Patients with DDI should concentrate urine appropriately during dehydration, without significant rise in plasma osmolality.

In reality however, many patients have incomplete defects and manifest mild or moderate forms of DI. Moreover, prolonged polyuria of any type can impair urine concentrating ability through dissipation of the medullary interstitial concentration gradient, resulting in a partial functional NDI. The water deprivation test can be a poor discriminator in these circumstances. The water deprivation test has been shown to support the correct diagnosis of underlying the polyuric state in some 70% of patients overall; and in only 41% of patients with underlying primary polydipsia [28].

Box 2.2.2 Protocol for water deprivation/desmopressin test

- Preparation
  - Free access to fluid given overnight prior to test
  - Avoid caffeine and smoking
  - 07.50 h weigh patient
- Dehydration phase
  - 08.00 plasma and urine osmolality, and urine volume
  - Restrict fluids up to eight hours
  - Weigh patient at two hourly intervals
  - Plasma and urine osmolality, and volume measurements two hourly
  - Stop test if weight loss exceeds 5% of starting weight, or thirst is intolerable
  - Supervise patient closely to avoid non-disclosed drinking
- Desmopressin phase
  - Inject intramuscularly 1 mcg desmopressin
  - Allow patient to eat and drink up to 1.5–2.0 times the volume of urine passed during dehydration phase
  - Collect urine for osmolality and volume at 20.00 h
- Plasma and urine osmolality, and volume measurements at 09.00 h next day

Direct Tests Used in the Differential Diagnosis of DI: The Utility of Measuring VP and Copeptin

An accurate diagnosis of HDI can be made by direct measurement of plasma VP during the controlled osmotic stress of a hypertonic 5% sodium chloride infusion. Patients with HDI have either undetectable VP levels, or values falling to the right of the normogram relating plasma VP to plasma osmolality. In NDI, plasma VP is inappropriately high for the prevailing urine and plasma osmolality, indicating VP resistance. In DDI, the relationship of plasma VP to osmolality is normal. The test is not interpretable if the patient experiences nausea, a powerful non-osmotic stimulus of VP release, during the test. Importantly, access to reliable VP assays has been limited and there are significant preanalytic obstacles in the scaling of this test to wider use. The addition of VP measurements as an adjunct to the standard water deprivation test does not add value [28, 29].

Copeptin, the c-terminal fragment of the VP-NP precursor is released in equimolar amounts to VP. In comparison to VP, the peptide is much more stable in plasma, making it an attractive alternative analyte to VP to support the differential diagnosis of polyuric states. Copeptin concentrations greater than 4.9 pmol/L during osmotic stimulation with hypertonic fluid can reliably discriminate between HDI and primary polydipsia, with a diagnostic accuracy of 96%. As with VP, the addition of copeptin to the standard water deprivation test does not improve the diagnostic accuracy of the test [30].

Additional Tests to Support the Differential Diagnosis of Polyuric States

A pragmatic alternative to VP measurements during hypertonic stress if there is diagnostic uncertainty following water deprivation is a controlled therapeutic trial of DDAVP: 10–20 mcg of intranasal DDAVP per day for 2–4 weeks, with monitoring of plasma sodium every 2–3 days. Patients with DDI exhibit progressive dilutional hyponatraemia, whereas those with NDI remain unaffected. Patients with HDI experience improvement in polyuria and polydipsia, but remain normonatraemic.

In familial autosomal dominant HDI, sequencing of the *VP* gene can help to establish the diagnosis in at-risk individuals early in the natural history of the disease and at a time when the water deprivation test can be equivocal [31].

Neuroimaging in HDI

Imaging of the hypothalamus, pituitary, and surrounding structures is essential in patients with HDI. MRI is the modality of choice. HDI is associated with the loss of the normal hyperintense signal of the posterior pituitary on T1-weighted images. Signal intensity is correlated strongly with VP content of the gland. As some hypothalamic germ cell tumours can be slow growing, imaging should be repeated after 6–18 months if the initial scan shows no demonstrable lesion. A negative scan at this stage should be taken as reassuring in the absence of a change in clinical features. Importantly, the absence of a posterior pituitary bright spot on imaging is not specific and can occur in a range of physiological situations. The finding should therefore not be taken as diagnostic of HDI.

Treatment

Patients with a urine output of less than 4 l/24 h can be managed by advising adequate fluid intake. The treatment of choice for those with more severe symptoms is the synthetic, long-acting VP analogue DDAVP; given as an intranasal spray (5–100 mcg daily), parenterally (0.1–2.0 mcg daily), or orally (100–1000 mcg daily) in divided doses. There is wide individual variation in the dose required to control symptoms. DDAVP has twice the antidiuretic potency of VP, but has minimal vasopressor activity. It is well tolerated. Dilutional hyponatraemia is the most serious potential adverse effect. This can be avoided by omitting treatment on a regular basis (perhaps weekly), to allow a short period of breakthrough polyuria and thirst [32].

Acute onset of HDI following pituitary or neurosurgery is usually transient and may not require intervention. Hypernatraemia can develop if cognition is impaired or access to water limited. Polyuria should be managed with DDAVP (parenteral administration may be required). As the situation is dynamic, a single dose of DDAVP may be sufficient to manage the patient until normal posterior pituitary function returns. Additional doses can be administered if polyuria persists or recurs. Regular DDAVP may be used if symptoms persist beyond 48 hours [33].

Nephrogenic Diabetes Insipidus

NDI is due to renal resistance to the antidiuretic effects of VP. Most forms of NDI are acquired, secondary to a range of metabolic or drug effects. The final common endpoint in many cases is downregulation of AQP2 expression. NDI secondary to lithium is common. Lithium enters collecting duct cells through the amiloride-sensitive epithelial sodium channel (ENaC) expressed on the apical cell membrane. This leads to inhibition of glycogen synthase kinase-3β (GSK3β) signalling pathways and dysregulated AQP2 expression and trafficking along the whole collecting duct. There is also dysregulated expression of ENaC in the cortical collecting duct. NDI secondary to lithium toxicity can irreversible, persisting after drug withdrawal.

Primary familial forms of NDI are rare. There are two types. A 'pure' type is characterized by water loss alone. A second, more complex type is characterized by water loss and a wider tubulopathy. X-linked recessive familial NDI is caused by inherited loss of function mutations of the V2-R. Over 70 different mutations have been described that have a negative impact on receptor expression, ligand binding, or G-protein coupling. Most mutations lead to complete loss of function. A small number lead to partial function and a milder phenotype.

Some10% of familial NDI is inherited as an autosomal recessive trait in which V2-R function is normal. In contrast to the X-linked form, autosomal recessive NDI is caused by loss of function mutations of the *AQP2* gene. Most mutations occur in the region of the gene coding for the transmembrane domain of the AQP2 protein.

A further, rare form of NDI is inherited as an autosomal dominant trait. While this form is also due to a loss of function mutation in *AQP2*, mutations and mechanisms are distinct and separate from the autosomal recessive form. The *AQP2* mutation in these kindreds involves the portion of the gene encoding the carboxyl-terminal intracellular tail of AQP2. The mutant AQP2 protein is expressed but forms non-functional, mixed tetramers with the product of the wild type AQP2 allele acting in a dominant-negative manner.

Defects in thick ascending limb and distal tubular sodium chloride transport underlie a range of inherited tubulopathies producing combinations of NDI and renal solute loss, including inherited forms of Bartter syndrome.

The mainstay of management in acquired cases of NDI is removal of the underlying cause and maintenance of hydration. While additional measures can be used for persistent symptoms, these rarely reduce urine volumes by more than 50%. High-dose DDAVP (4 mcg im. bd) can produce a response in partial NDI. There are data to support the use of amiloride in NDI caused by lithium. Additional treatment options (used alone or in combination) include hydrochlorothiazide 25 mg/24 hours; non-steroidal anti-inflammatory drugs such as ibuprofen 200 mg/24 hours; and low salt diets. The selective phosphodiesterase inhibitor Sildenafil has been shown to improve aquaresis in X-linked NDI. The proposed mechanism of action involves bypassing defective V2-R signalling through enhancement of cyclic guanosine monophosphate-mediated apical membrane trafficking [34, 35].

Dipsogenic Diabetes Insipidus

DDI is a syndrome of excess fluid intake, and consequent polyuria. Though structural abnormalities may be the cause, more commonly it is a manifestation of primary hyperdipsia, psychiatric disease, or secondary to drug effects. DDI in the absence of other identifiable illness is compulsive water drinking. It is associated with abnormalities of thirst perception, including; a low osmotic threshold for thirst; an exaggerated thirst response to osmotic challenge; and an inability to suppress thirst at low osmolalities. The structural and/or functional basis for any of these abnormalities has not been identified. The association of DDI with affective disorders is well recognized. Up to 20% of patients with long-term, persisting schizophrenia have polydipsia. Although in some cases abnormal drinking is in response to adverse effects of medication producing a dry mouth, complex abnormalities in both osmoregulated VP release and osmoregulated thirst have been described. Whether these reflect long-term effects of drug therapy or a primary defect in the central integration of thirst, is unclear.

Confirmation of the diagnosis of DDI is through direct or indirect demonstration of normal osmoregulated VP release and antidiuretic action. As with many conditions, the treatment of DDI

should address the underlying disorder. Individuals with persistent DDI are at risk of hyponatraemia if treated with DDAVP, as fluid intake is maintained despite an obligate antidiuresis. In such cases a reduced fluid intake is the only rational treatment.

Syndrome of Inappropriate Antidiuresis

Hyponatraemia (serum sodium <135 mmol/L) is a feature in some 15–20% of non-selected emergency hospital admissions. It is associated with increased morbidity and mortality across a range of conditions and data support the association of correction of hyponatraemia with improved outcome. However, the relationship of serum sodium to clinical outcomes is not straightforward and may not be a causal one. Comorbidity and underlying disease severity may be the more significant contributors to outcomes in patients with hyponatraemia [36].

VP plays a key role in many clinical situations in that include hyponatraemia. However, when VP does play a role, VP production may not be inappropriate. Hyponatraemia may reflect an appropriate physiological response to volume depletion, with baroregulated VP release persisting despite plasma osmolalities below the normal osmolar threshold for VP release. In this situation, continued physiologically appropriate VP release can lead to hyponatraemia which can be persistent. Though clinical assessment can identify the extracellular volume status of some patients, it is unreliable and has poor sensitivity and specificity [37].

Pathophysiology of the Syndrome of Inappropriate Antidiuresis

An individual with hyperosmolar plasma but a normal circulating volume and in whom in whom the plasma VP concentration is high for the prevailing osmolality, has a syndrome of inappropriate antidiuresis (SIAD) due to VP excess. A variety of conditions are associated with SIAD. To date, four patterns of abnormal VP secretion have been identified (Table 2.2.3). Absolute plasma VP concentrations may not be strikingly high; the key finding is that that they are inappropriate for the prevailing plasma osmolality. When this obligate antidiuresis is not accompanied by decreased water intake, haemodilution is inevitable.

Aetiology

Many conditions have been reported to cause SIAD (Box 2.2.3). SIAD is a non-metastatic manifestation of small cell lung cancer and other malignancies. The mechanism(s) of inappropriate VP release in many cases of SIAD are not clear. Some tumours are an ectopic source of VP, and produce a type A syndrome. However, excessive posterior pituitary VP secretion also occurs in association with malignancy. In fact, the absence of an ectopic VP source suggests a lesion in the neurohypophysis or its regulatory afferent pathways. The similarities between SIAD type B and the changes in VP regulation in response to hypovolaemia and hypotension, suggest a single lesion in the baroregulatory afferent pathways. In contrast, the normal osmoregulated VP release found in the type D syndrome suggests an increase in renal sensitivity to VP, or the action of an as yet unidentified antidiuretic factor.

SIAD is a common mechanism of drug induced hyponatraemia. It can reflect direct stimulation of VP release from the hypothalamus; indirect action on the hypothalamus via effects on higher centres; or aberrant resetting of the hypothalamic osmostat.

Box 2.2.3 Causes of SIAD

- Neoplastic disease
  - Carcinoma (bronchus, duodenum, pancreas, prostate)
  - Thymoma
  - Mesothelioma
  - Lymphoma, leukaemia
  - Ewing's sarcoma
  - Carcinoid
  - Bronchial adenoma
- Neurological disorders
  - Head injury, neurosurgery
  - Brain abscess or tumour
  - Meningitis, encephalitis
  - Guillain–Barré syndrome
  - Cerebral haemorrhage
  - Cavernous sinus thrombosis
  - Hydrocephalus
  - Cerebellar and cerebral atrophy
  - Shy–Drager syndrome
  - Peripheral neuropathy
  - Seizures
  - Subdural haematoma
  - Alcohol withdrawal
- Chest disorders
  - Pneumonia
  - Tuberculosis
  - Empyema
  - Cystic fibrosis
  - Pneumothorax
  - Aspergillosis
- Drugs
  - Chlorpropamide
  - Opiates
  - Vincristine, vinblastine, cisplatin
  - Thiazides
  - Dopamine antagonists
  - Tricyclic antidepressants
  - MAOIs
  - SSRIs
  - 3,4-MDMA agents ('Ecstasy')
  - Anticonvulsants
- Miscellaneous
  - Idiopathic
  - Psychosis
  - Porphyria
  - Abdominal surgery

Table 2.2.3 Classification of SIAD

SIAD type	Characteristics
SIAD type A	Wide fluctuations in plasma VP concentration, independent of plasma osmolality. Accounts for 35% of SIAD
SIAD type B	Osmotic threshold for VP release subnormal. Patients osmoregulate around subnormal plasma osmolar set point. Accounts for 30% of SIAD
SIAD type C	Failure to suppress VP release at low plasma osmolality, normal response to osmotic stimulation
SIAD type D	Normal osmoregulated VP release, but unable to excrete a water load. Accounts for less than 10% of SIAD

Dopamine antagonists cause SIAD through stimulation of VP release. Antidepressant medication, including selective serotonin reuptake inhibitors (SSRIs), potentiate stimulatory central adrenergic input to VP-producing neurones. Opiates also stimulate inappropriate VP release through enhancing central adrenergic drive. SIAD is commonly associated with antiseizure medication. The frequency of hyponatraemia in patients treated with carbamazepine ranges from 4.8 to 40%, though the majority of such cases are asymptomatic. Carbamazepine increases both the sensitivity of central osmoceptors, and renal sensitivity to VP.

Clinical Features, Diagnosis, and Differential Diagnosis of SIAD

The major features in the diagnosis of SIAD are highlighted in Box 2.2.4. The most frequent problem in clinical practice is distinguishing SIAD from chronic, mild hypovolaemia. In both conditions, urine osmolality tends to be higher than plasma osmolality. Plasma VP will be detectable or elevated in both. Neither is therefore diagnostic of SIAD. The diagnosis hinges on confirming excretion of urine that is not maximally dilute in the context of a dilute plasma (i.e. urine concentration greater than 100 mOsml/Kg). Renal sodium excretion should be above 30 mmol/L to make a diagnosis of SIAD. Below this value, volume depletion should be considered the more likely cause of hyponatraemia. SIAD is often associated with urine sodium concentrations of 60 mmol/L or more, as persisting SIAD is a volume expanded state. Consistent with this, there is evidence of mild sodium loss as other regulators of volume homeostasis attempt to minimize volume expansion.

The role of VP production or action in producing hyponatraemia can be confirmed indirectly by assessing excretion of a standard water load over a fixed time: the water load test (Table 2.2.4). Normal subjects excrete 78–82% of the ingested water load in the 4 h observation period. This is reduced to 30–40% in the presence of constitutive VP production or action. The test is not essential to establish a diagnosis and is not recommended as part of routine clinical practice.

Exercise-Associated Hyponatraemia

Extreme endurance exercise is a profound physiological stress. Non-osmoregulated VP release is a central feature of extreme endurance exercise, reflecting this physiological stress and determined principally by the duration of the event and the effort involved. Combined with reduced renal blood flow (another feature of extreme endurance exercise) this VP release can lead to a marked antidiuretic state. In this situation, fluid intake in excess of water loss can lead to hyponatraemia. Those athletes developing hyponatraemia during endurance exercise demonstrate weight gain over the course of the event, clearly implicating water intake in excess of water and electrolyte loss as the cause. There is a positive correlation between the odds ratio for developing hyponatraemia during extreme endurance exercise and the length of time taken to complete the event. 'Occasional' runners are therefore particularly at risk of exercise-associated hyponatraemia. They should be advised to follow their thirst and avoid rigid, time-based fluid intake. Importantly, health professionals need to be aware of the problem of exercise-associated hyponatraemia and resist resuscitation with large volumes of hypotonic fluid in the absence of appropriate indications and without biochemical monitoring [38].

Box 2.2.4 Diagnosis of SIAD

- Hyponatraemia with appropriately low plasma osmolality
- Urine osmolality that is not maximally dilute in context of on-going hyponatraemia (i.e. urine osmolality >100 mOsm/kg)
- Urine sodium concentration >30 mmol/L
- Absence of hypotension, hypovolaemia, and oedema-forming states
- Normal renal and adrenal function

Table 2.2.4 Protocol for water load test

Preparation	• Free access to fluid overnight prior to test • Avoid caffeine and smoking • 07:30 h weigh patient • Cannulate patient • Rest patient 30 minutes
Water load phase	• 08:00 plasma and urine osmolality, plasma VP • Patient to drink 20 ml/kg water over 15 minutes • Measure hourly urine output for 4 hours • Measure urine osmolality, plasma osmolality, and plasma VP hourly for 4 hours
Recovery phase	• **Plasma sodium 2 hours after test completed** • Plasma sodium and osmolality 09:00 h next day

Nephrogenic Syndrome of Inappropriate Antidiuresis

While loss of function mutations of the V2-R are the cause of X-linked nephrogenic diabetes insipidus, rare individuals express the reciprocal problem: constitutively activating mutations in the V2-R that lead to VP-independent, but V2-R mediated, antidiuresis resulting in persistent hyponatraemia. This nephrogenic syndrome of inappropriate antidiuresis (NSIAD) can have a variable phenotype. Although initially described in male infants with persistent hyponatraemia, the condition is not limited to males and may manifest in adulthood. This is consistent with in the condition being X-linked but with variable expression in heterozygous females [39, 40].

Central Salt Wasting

This acquired, primary natriuretic state is described as a combination of hyponatraemia with hypovolaemia associated with neurological or (more often) neurosurgical pathologies. The definition, true prevalence, and underlying pathophysiology remain a subject of controversy. SIAD can occur in the same group of patients in whom central salt wasting (CSW) has been reported and both are associated with urine sodium concentrations greater than 40 mmol/L. Importantly therefore, the diagnosis of CSW hinges on the natural history: the development of hyponatraemia being preceded by natriuresis and diuresis with ensuing clinical and biochemical features of hypovolaemia. In contrast to SIAD, urea and creatinine are elevated and there may be postural hypotension.

The context and opposed cause-directed management approaches to CSW and SIAD can lead to significant tension in clinical practice. CSW is a particular concern for the neurosurgical patient in whom autoregulation of cerebral blood flow is disturbed and in whom

small reductions in circulating volume can reduce cerebral perfusion. The management of CSW is volume replacement with 0.9% saline; while the cause-directed approach to SIAD would often involve restriction of fluid. A cause-independent approach to the management of the neurosurgical patient with hyponatraemia needs to balance management of hyponatraemia with the need to avoid threatening cerebral perfusion and avoidable vasospasm. A prospective, single centre study including 100 cases of hyponatraemia developing after subarachnoid haemorrhage failed to identify a case of CSW. Rather, SIAD, glucocorticoid deficiency and inappropriate fluid administration were identified as the key mechanisms of hyponatraemia [41, 42].

Treatment of Hyponatraemia Secondary to SIAD

Independent of the underlying cause, the morbidity SIAD reflects the impact of hyponatraemia on CNS function from cerebral oedema and primary neuronal dysfunction (Box 2.2.5). The relationship between serum sodium and neurological function is not simple: patients with marked biochemical disturbance may have mild symptoms if hyponatraemia develops over a prolonged period. This reflects CNS adaptation to reduced osmolality through inorganic and organic ion efflux. These adaptive CNS mechanisms can complicate the management of hyponatraemia. Over rapid correction of serum sodium following the gradual development of hyponatraemia can lead to changes in brain volume as the osmolar gradient across the blood-brain barrier alters. This change in volume can trigger CNS osmotic demyelination syndrome (ODS): a rare but serious complication of hyponatraemia and its treatment. ODS develops within 1–4 days of rapid (>10–12 mmol/L per 24 hours) correction of serum sodium. Other factors play a role in susceptibility: hepatic failure, potassium depletion, and malnutrition. Neurological manifestations include quadriplegia, ophthalmoplegia, pseudo-bulbar palsy, and coma.

In SIAD, hyponatraemia with serum sodium concentrations greater than 130 mmol/L may not require specific treatment. More significant degrees of hyponatraemia, associated with symptoms, can require intervention

Management of SIAD Associated with Severe or Moderately Severe Symptoms

Hyponatraemia associated with severe or moderately severe symptoms requires urgent management. Treatment is cause-independent and should be given priority over establishing the aetiology of hyponatraemia. The principle of treatment is to reduce immediate risk through increasing serum sodium to a level that decreases morbidity but at a rate that does not result in harm through precipitating ODS. Critically, the immediate target should not be to normalize serum sodium.

Determining hypertonic fluid prescription by calculated serum sodium deficit is associated with a significant risk of over-rapid correction. A pragmatic approach is to use small boluses of hypertonic fluid to achieve a 5 mmol/L rise in serum sodium in the first hour of treatment. The increase in serum sodium should be limited to 10 mmol/L in the first 24 hours, and no more than 8 mmol/L per 24 hours thereafter. An increase above the recommended rate should prompt review and consideration of active management with hypotonic fluid, with or without concurrent use of DDAVP to limit or reverse the trajectory in serum sodium. Once the clinical situation is stable, investigations to establish the cause of hyponatraemia are recommended to guide cause-directed therapy [43].

Management of SIAD Associated with Mild to Moderate Symptoms

Cause-directed therapy is the foundation managing SIAD in this context. In parallel, fluid restriction of 0.5–1 L/day is a reasonable parallel intervention. The increase in serum sodium should be limited to 8–10 mmol/L per 24 hours. All fluids need to be included in the restriction. As SIAD is associated with a degree of natriuresis, sodium intake should be maintained. Fluid restriction may need to be maintained for several days before sodium levels normalize and it is important that a negative fluid balance is confirmed during this period. As cause-directed therapy progresses (e.g. treatment of underlying infection or removal of drug-causing SIAD), fluid restriction can appropriately relaxed.

Drug Treatment in SIAD

If hyponatraemia persists or recurs after initial intervention, it is important that the underlying diagnosis is reviewed the intervention reconsidered. Clinical balance is key. There may be situations where withdrawal of a causal agent for SIAD (e.g. drug) is not practical or in the overall benefits of the patient. In other situations, fluid restriction may be only partly effective, be poorly tolerated, or may prove non-sustainable. Clinicians may thus have to balance the merits of incremental intervention with those of tolerating mild, persisting hyponatraemia.

V2-R antagonists (Vaptans) are a rational approach to the management of hyponatraemia due to SIAD. They are pure aquaretics, increasing renal water excretion without impacting on renal electrolyte loss. Both selective (V2-R specific) and non-selective (V2- and V1a-R antagonism) increase serum sodium in patients with normal or increased plasma volume and can increase serum sodium within 4–6 hours. There use can be associated with over-rapid correction. In the absence of robust data on cost-utility and evidence on patient-orientated outcomes other than serum sodium, the role of V-R antagonists in the management of SIAD remains to be clarified [43, 44].

Box 2.2.5 Symptoms and signs of hyponatraemia secondary to SIAD

- Headache
- Nausea
- Vomiting
- Muscle cramps
- Lethargy
- Disorientation
- Seizure
- Coma
- Brainstem herniation
- Death

Adipsic and Hypodipsic Disorders

Adipsic and hypodipsic disorders are characterized by inadequate spontaneous fluid intake due to a primary defect in osmoregulated thirst. Patients can become hypovolaemic and dehydrated. Despite this, they deny thirst and do not drink. If the defect is mild, the

Table 2.2.5 Classification of adipsic/hypodipsic syndromes

Classification	Osmoregulated thirst	Osmoregulated VP release
Type A (essential hypernatraemia)	Osmotic threshold increased Normal sensitivity	Osmotic threshold increased, normal sensitivity Normal response to non-osmotic stimuli
Type B	Normal osmotic threshold Reduced sensitivity	Normal osmotic threshold, reduced sensitivity Normal response to non-osmotic stimuli
Type C	No thirst response to osmotic stimulation	Persistent low-level VP release, no response to osmotic stimulation or inhibition Normal response to non-osmotic stimulation
Type D	No thirst response to osmotic stimulation	Normal

Box 2.2.6 Causes of adipsic/hypodipsic syndromes

- Neoplastic (50%)
  - Primary
    - Craniopharyngioma
    - Germinoma
    - Meningioma
  - Secondary
    - Lung cancer
    - Breast cancer
- Vascular (15%)
  - Internal carotid ligation
  - Anterior communicating artery aneurysm
  - Intrahypothalamic haemorrhage
- Granulomatous (20%)
  - Histiocytosis
  - Sarcoidosis
- Miscellaneous (15%)
  - Hydrocephalus
  - Ventricular cyst
  - Trauma
  - Toluene poisoning

resultant hypernatraemia is often well tolerated. Severe disorders leading to marked electrolyte disturbances are tolerated poorly, and can lead to somnolence, seizures, coma, and renal failure. Because of the close anatomical relationship of the osmoregulatory centres for thirst and VP release, adipsic syndromes are often associated with defects in osmoregulated VP release and HDI. Understandably, this can exacerbate electrolyte and water balance problems and lead to significant and life-threatening decompensation.

Aetiology

Four distinct patterns of discorded thirst and associated VP release are recognized (Table 2.2.5; Figure 2.2.8). A range of conditions leading to or associated with adipsia/hypodipsia have been described (Box 2.2.6).

The type A syndrome can be mistaken for HDI, as patients are hyperosmolar with a dilute urine. Formal assessment of thirst confirms the diagnosis. Imaging is generally normal. While the precise nature of the problem remains unclear, normal VP responses to non-osmotic stimuli place the lesion responsible at the level of the osmoreceptor, rather than the VP magnocellular neurone. Patients effectively osmoregulate around a higher osmolar set point. They are protected from extreme hypernatraemia, as are those with the type B syndrome. In contrast, type C adipsia is associated with complete lack of osmoregulated thirst and VP release, consistent with effective loss of osmoreceptor function. Patients present with adipsic HDI. Specific precipitants include rupture and repair of anterior communicating artery (ACA) aneurysm. The OVLT receives its blood supply from perforating branches of the anterior cerebral artery and ACA. Surgery involving the anterior communicating artery can lead to damage at more distal sites, including the OVLT. Some patients with the type C syndrome have persistent, constitutive low-level VP release. The resultant obligatory antidiuresis thus also places such individuals at risk of dilutional hyponatraemia if large volumes of fluid are administered. Impaired osmoregulated thirst with normal osmoregulated VP release (type D adipsia) is very rare.

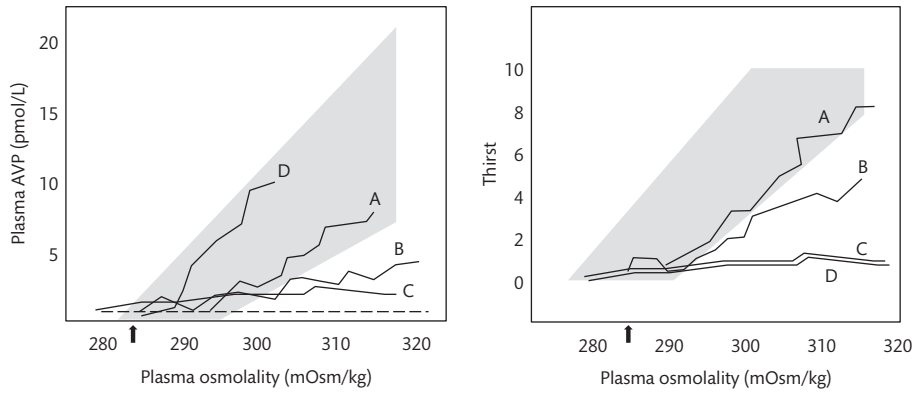


Figure 2.2.8 Patterns of plasma vasopressin and thirst responses to hypertonic stress in patients with adipsic syndromes. Normal range responses to osmolar stimulation are shown by the shaded areas. The four types of adipsic syndrome are demonstrated. Patients with the type A syndrome osmoregulate around a higher osmolar set point. Those with the type B syndrome mount vasopressin and thirst responses but with reduced sensitivity to increases in plasma osmolality. Patients with the type C syndrome have much reduced or absent vasopressin and thirst responses to osmolar stimulation while those with the type D syndrome demonstrate normal vasopressin responses to osmolar stimulation but much reduced thirst responses.

Treatment

Patients with type A and type B adipsia are protected from extreme hypernatraemia. Treatment is to recommend an obligate fluid intake of some 2 L/24 h, with appropriate adjustment for climate and season. Fluid balance may be difficult to maintain during intercurrent illness. In such circumstances, in-patient management may be required. The adipsic HDI of the type C syndrome can be difficult to manage. Structural and vascular causes of the type C syndrome may lead to associated cognitive deficits and problems with short term memory and task organization which can complicate any intervention. In addition, other problems associated with hypothalamic dysfunction may be present: such appetite and thermal balance dysregulation. The principle of management is to define an acceptable urine output (1–2 L/24 h) with regular DDAVP, and to vary the daily fluid intake depending on day to day fluctuation from a target weight at which the patient is euvolaemic and normonatraemic:

Daily fluid intake in litres = 1–2 L (i.e. the targeted urine output as dictated by the DDAVP dose set and taking into account insensible loss) + (target weight – daily weight in kg).

This formula, together with weekly checks of plasma sodium to avoid the creeping development of hyper- and hyponatraemia, can result in stable fluid balance [45, 46].

REFERENCES

1. Liorens-Cortes C, Moos F. Apelin and vasopressin: two work better than one. *J Neuroendocrinol*, 2012; **245**: 1085–6.
2. Ponzio TA, Fields RL, Rashid OM, *et al.* Cell-type specific expression of the vasopressin gene analyzed by AAV mediated gene delivery of promoter deletion constructs into the rat SON in vivo. *PLoS One*, 2012; **7**: e48860.
3. Mohr E, Kachele I, Mullin C, *et al.* Rat vasopressin mRNA: a model system to characterize cis-acting elements and trans-acting factors in dendritic mRNA sorting. *Prog Brain Res*, 2002; **139**: 211–24.
4. King LS, Yasui M. Aquaporins and disease: lessons from mice to humans. *Trends Endocrinol Metab*, 2002; **13**: 355–60.
5. Nielsen S, Kwon T-H, Frokiaer J, Agre P. Regulation and dysregulation of aquaporins in water balance disorders. *J Int Med*, 2007; **261**: 53–64.
6. Verkman AS. Aquaporins at a glance. *J Cell Sci*, 2011; **124**: 2107–12.
7. Knepper MA, Kwon T-H, Nielsen S. Molecular physiology of water balance. *N Engl J Med*, 2016; **372**: 1348–58.
8. Bielsky IF, Hu S-B, Szegda KL, *et al.* Profound impairment in social recognition and reduction in anxiety-like behaviour in vasopressin V1a receptor knockout mice. *Neuropsychopharmacology*, 2004; **29**: 483–93.
9. Walum H, Westberg L, Henningsson S, *et al.* Genetic variation in the vasopressin receptor 1a gene (*AVPR1A*) associates with pair-bonding behavior in humans. *Proc Natl Acad Sci USA*, 2008; **105**: 14153–6.
10. Albers HE. The regulaton of social recognition, social communication and aggression: vasopressin in the social behaviour neural network. *Horm Behav*, 2012; **61**: 283–92.
11. Dai L, Carter CS, Ying J, *et al.* Oxytocin and vasopressin are dysregulated in Williams syndrome, a genetic disorder affecting social behavior. *PLoS One*, 2012; **7**: e38513.23.
12. Stoop R. Neuromodulation by oxytocin and vasopressin. *Neuron*, 2012; **76**: 142–59.
13. Sun L, Tamma R, Yuen T, *et al.* Functions of vasopressin and oxytocin in bone mass regulation. *Proc Natl Acad Sci*, 2016; **113**: 164–9.
14. McKenna K, Thompson C. Osmoregulation in clinical disorders of thirst and thirst appreciation. *Clin Endocrinol*, 1998; **49**: 139–52.
15. Zimmerman CA, Leib DE, Knight ZA. Neural circuits underlying thirst and fluid homeostasis. *Nat Rev Neurosci*, 2017; **18**: 459–69.
16. Bourque CW. Central mechanisms of osmosensation and systemic osmoregulation. *Nat Rev Neurosci*, 2008; **9**: 519–31.
17. Lechner SG, Markworth S, Poole K, *et al.* The molecular and cellular identity of peripheral osmoreceptors. *Neuron*, 2011; **69**: 332–44.
18. Trudel E, Bourque CW. Circadian modulation of osmoregulated firing in rat supraoptic neurones. *J Neuroendocrinol*, 2012; **24**: 577–86.
19. Tyzio R, Cossart R, Khalilov I, *et al.* Maternal oxytocin triggers a transient Inhibitory switch in GABA signaling in the Fetal brain during delivery. *Science*, 2006; **15**: 1788–92.
20. Hammock EAD, Young LJ. Oxytocin, vasopressin and pair bonding: implications for autism. *Philos Trans R Soc Lond B Biol Sci*, 2006; **29**: 2187–98.
21. Windle RJ, Gamble LE, Kershaw YM, *et al.* Gonadal steroid modulation of stress-induced hypothalamo-pituitary-adrenal activity and anxiety behavior: role of central oxytocin. *Endocrinology*, 2006; **147**: 2423–1.
22. Lo JA, Verbali JG. Diabetes insipidus as a complication after pituitary surgery. *Nat Endocrinol Metab*, 2007; **3**: 489–94.
23. Aghar A, Thornton E, O'Kelly P, *et al.* Posterior pituitary dysfunction after traumatic brain injury. *J Clin Endocrinol Metab*, 2004; **89**: 5987–92.
24. Hannon MJ, Orr C, Moran C, Behan LA, Agha A, Ball SG, Thompson CJ. Anterior hypopituitarism is rare and autoimmune disease is common in adults with idiopathic central diabetes insipidus. *Clin Endocrinol*, 2012; **76**: 725–8.
25. Rigoli L, Lombardo F, Di Bella C. Wolfram syndrome and *WFS1* gene. *Clin Genet*, 2011; **79**: 103–17.
26. Chaussenot A, Bannwarth S, Rouzier C, *et al.* Neurologic features and genotype-phenotype correlation in Wolfram syndrome. *Ann Neurol*, 2011; **69**: 501–8.
27. Nijenhuis M, van den Akker ELT, Zalm R, *et al.* Familial neurohypophyseal diabetes insipidus in a large Dutch kindred: effect of the onset of diabetes on growth in children and cell biological defects of the mutant vasopressin prohormone. *J Clin Endocrinol Metab*, 2001; **86**: 3410–20.
28. Fenske W, Quinkler M, Lorenz D, *et al.* Copeptin in the differential diagnosis of the polydipsia-polyuria syndrome—revisiting the direct and indirect water deprivation tests. *J Clin Endocrinol Metab*, 2011; **96**: 1506–15.
29. Ball SG, Barber T, Baylis PH. Tests of posterior pituitary function. *J Endocrinol Invest*, 2003; **26** (suppl): 15–24.
30. Fenske W, Refardt J, Chifu I, *et al.* A Copeptin-based approach in the diagnosis of diabetes insipidus. *N Engl J Med*, 2018; **379**: 428–39.
31. Chitturi S, Harris M, Thomsett MJ, *et al.* Utility of AVP gene testing in familial neurohypophyseal diabetes insipidus. *Clin Endocrinol*, 2008; **69**: 926–30.
32. Behan LA, Sherlock M, Moyles P, *et al.* Abnormal plasma sodium concentrations in patients treated with desmopressin for cranial diabetes insipidus: results of a long-term retrospective study. *Eur J Endocrinol*, 2015; **172**: 243–50.

33. Garrahy A, Sherlock M, Thompson CJ. Management of endocrine disease: neuroendocrine surveillance and management of neurosurgical patients. *Eur J Endocrinol*, 2017; **176**: R217–r233.
34. Grunfeld J-P, Rossler BC. Lithium nephrotoxicity revisited. *Nat Rev Nephrol*, 2009; **5**: 2709–276.
35. Assadi F, Sharbaf FG. Sildenafil for the treatment of congenital nephrogenic Diabetes insipidus. *Am J Nephrol*, 2015; **42**: 65–9.
36. Chawla A, Sterns RH, Nigwekar SU, Cappuccio JD. Mortality and serum sodium: do patients die with or from hyponatraemia? *Clin J Am Soc Nephrol*, 2015; **6**: 960–5.
37. Fenske W, Maier SKG, Blechschmidt A, *et al*. Utility and limitations of the traditional diagnostic approach to hyponatremia: a diagnostic study. *Am J Med*, 2010; **123**: 62–657.
38. Rosner MH. Exercise associated hyponatremia. *Semin Nephrol*, 2009; **29**: 271–81.
39. Feldman BJ, Rosenthal SM, Vargas GA, *et al*. Nephrogenic syndrome of inappropriate antidiuresis. *N Engl J Med*, 2005; **352**: 1884–90.
40. Decaux G, Vandergheynst F, Bouko Y, *et al*. Nephrogenic syndrome of inappropriate antidiuresis in adults: high phenotypic variability in men and women from a large pedigree. *J Am Soc Nephrol*, 2007; **18**: 606–12.
41. Yee AH, Burns JD, Eelco FM, Wijdicks FM. Cerebral salt wasting: pathophysiology, diagnosis and treatment. *Neurosurg Clin N Am*, 2010; **221**: 339–52.
42. Hannon MJ, Behan LA, Brien MMC, *et al*. Hyponatraemia following mild/moderate subarachnoid haemorrhage is due to SIAD and glucocorticoid deficiency and not cerebral salt wasting. *J Clin Endocrinol Metab*, 2014; **99**: 291–8.
43. Spasovski G, Vanholder R, Allolio B, *et al*. Clinical practice guideline on diagnosis and treatment of hyponatraemia. *Eur J Endocrinol*, 2014; **170**: G1–G47.
44. Gross PA, Wagner A, Decaux G. Vaptans are not the mainstay of treatment in hyponatremia: perhaps not yet. *Kidney Int*, 2011; **80**: 594–600.
45. Ball SG, Vaidja B, Baylis PH. Hypothalamic adipsic syndrome: diagnosis and management. *Clin Endocrinol*, 1997; **47**: 405–9.
46. Cuesta M, Hannon MJ, Thompson CJ. Adipsic diabetes insipidus in adult patients. *Pituitary*, 2017; **20**: 372–80.

2.3

Aetiology, Pathogenesis, and Management of Disease of the Pituitary

2.3.1 Development of the Pituitary and Genetic Forms of Hypopituitarism

Louise C. Gregory and Mehul T. Dattani

Hypothalamo-Pituitary Development

Hypothalamo-pituitary (HP) development is dependent on the communication between the oral ectoderm and the overlying neural ectoderm. This occurs through a complex spatiotemporal genetic cascade of transcription factors and signalling molecules, intrinsic or extrinsic to the developing Rathke's pouch, the primordium of the anterior pituitary (AP) [1] (see **Figure 2.3.1.1**). The mature pituitary gland is situated at the base of the brain, within the sella turcica recess of the sphenoid bone, and centrally regulates growth, metabolism, reproduction, development, and homeostasis through the regulation of endocrine glands throughout the body [2]. It consists of three lobes; the anterior and intermediate lobes derived from the oral ectoderm, and the posterior lobe derived from the neural ectoderm [3, 4]. A series of tightly regulated steps resulting in cell proliferation and differentiation give rise to the five different specialized AP cell types that secrete the six AP hormones: somatotrophs [growth hormone (GH)], thyrotrophs [thyroid-stimulating hormone (TSH)], gonadotrophs [luteinizing hormone (LH) and follicle-stimulating hormone (FSH)], lactotrophs [prolactin (PRL)], and the corticotrophs [adrenocorticotropic hormone (ACTH)] [5]. Specific hypothalamic peptides travel via the hypophyseal portal system from the hypothalamus to the AP, where they bind to their specific receptors on each cell type to stimulate AP hormone production, that in turn target organs throughout the body to perform distinct roles in endocrine regulation.

Deficiencies in one or more of the six AP hormones is termed congenital hypopituitarism (CH), with GH deficiency being the most common and often seen in isolation [1]. CH severity varies widely, presenting in the early neonatal period or later in childhood, with severe cases resulting in fatality. Disordered embryogenesis can give rise to a range of midline craniofacial and brain structural abnormalities associated with CH, which may manifest as highly variable disorders, such as holoprosencephaly (HPE), septo-optic dysplasia (SOD), and Kallmann syndrome (KS), which is characterized by hypogonadotropic hypogonadism (HH) with anosmia [6].

Septo-Optic Dysplasia

SOD (n = 1/10,000 live births) is a heterogeneous variable disorder, characterized by optic nerve hypoplasia (ONH), midline neuroradiological abnormalities (such as agenesis of the corpus callosum and absence of the septum pellucidum), and pituitary hypoplasia with endocrine deficits [7, 8] with up to ~40% of SOD patients presenting with normal endocrinology. Other associated features are developmental delay, seizures, visual impairment, precocious puberty, obesity, anosmia, sensorineural hearing loss, and cardiac anomalies [9], illustrating the complexity of the disorder.

Intriguingly, SOD is associated with a younger maternal age when compared to mothers of children with isolated defects of the HP axis [10] for reasons unknown, although in some cases it may be due to excessive maternal drug and alcohol use [9, 11]. ONH occurs in 75–80% of SOD patients (unilateral [12%] or bilateral [88%]), with microphthalmia or anophthalmia occurring in rare severe cases [12]. Endocrine dysfunction may occur later [13]. The known association of midline abnormalities with hypopituitarism suggests a common developmental origin of the hypothalamus, pituitary, and midline structures within the brain [14]. Mutations in the transcriptional repressor *HESX1* were the first to be associated with the pathogenesis of rare cases of SOD, p.R160C being the first *HESX1* mutation described in association with the condition [15, 16]. Murine transgenic studies of *HESX1* revealed phenotypes that were highly reminiscent of SOD in the mice,

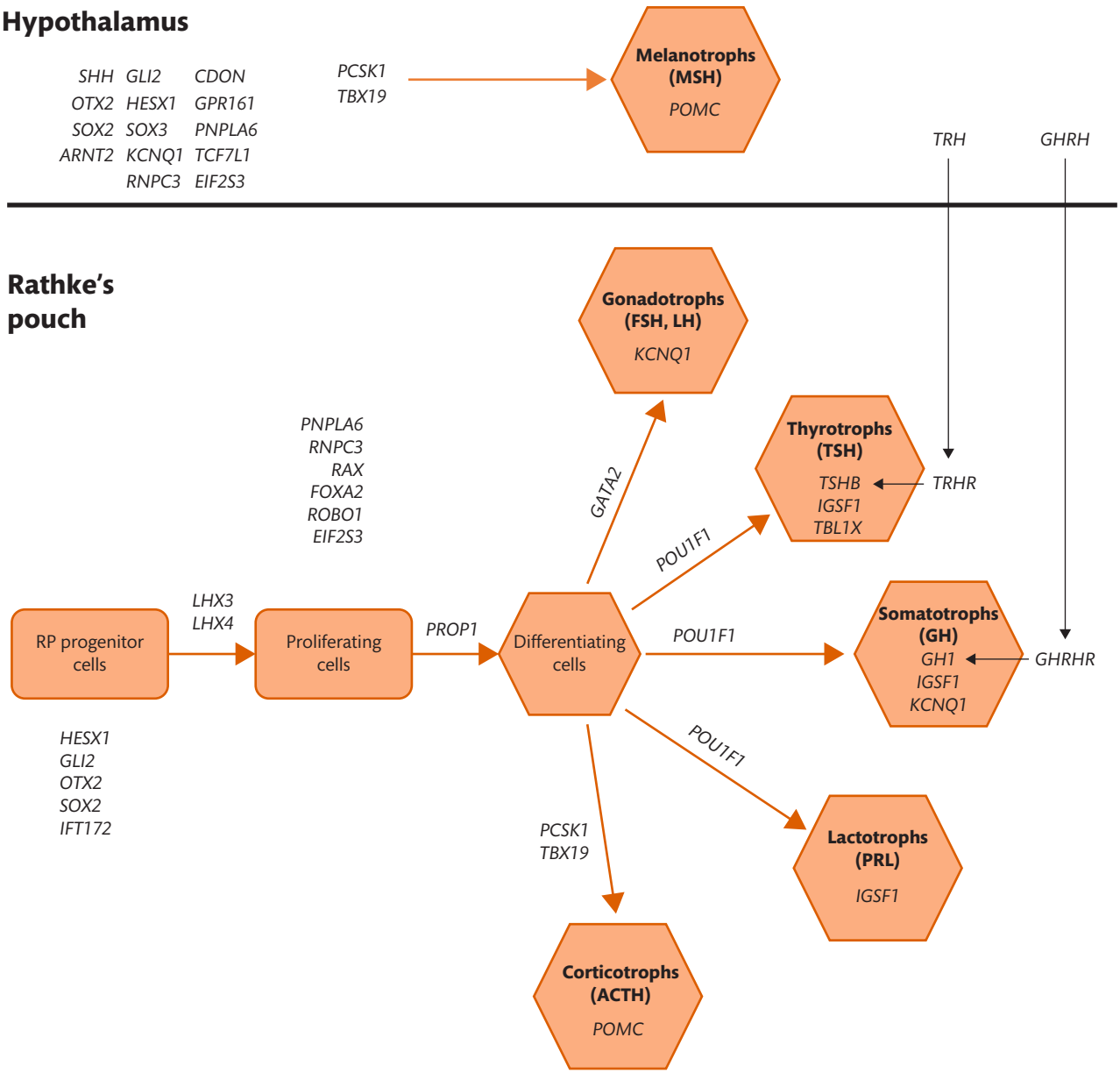


Figure 2.3.1.1 A genetic cascade of transcription factors and signalling molecules involved in hypothalamo-pituitary embryonic development.

supporting the requirement of the gene for midline and HP development, and a role for mutations in the gene in the aetiology of SOD. Subsequently, human mutations have been cloned into mouse models and studied in depth [17]. Since then, mutations in *SOX2* and *OTX2* were identified in SOD patients with severe bilateral eye defects including microphthalmia or anophthalmia, and *SOX3* mutations were identified in SOD patients with abnormalities of the hypothalamus, pituitary, infundibulum, and corpus callosum, respectively [18]. Mutations in genes implicated in KS have also been implicated in SOD. Three patients with an SOD phenotype, for example, were reported to have heterozygous mutations in *FGFR1* that altered receptor signalling, with one predicted to affect splicing. The same report described an SOD patient with an *FGF8* mutation affecting splicing and altered ligand binding [19]. Another *FGF8* mutation was identified in an SOD patient with microcephaly and neurological defects, and in a patient with semilobar HPE, diabetes insipidus, and TSH and ACTH insufficiency, respectively [6], making *FGF8* a candidate for SOD, and HPE as well as KS. Additionally, *KAL1* mutations and the *PROKR2* mutation, p.R268C, previously implicated in normosmic HH and KS, were reported in three unrelated pedigrees with SOD [20]. Heterozygous missense *TCF7L1* variants have been identified in two unrelated SOD patients [21], following studies using a conditional *Tcf7L1* murine deletion that resulted in forebrain defects and partially penetrant dwarfism [21], thus suggesting a role for *TCF7L1* in the aetiology of SOD. Table 2.3.1.1 summarizes the causative genes implicated in the aetiology of SOD and the subsequent HP-related disorders discussed in this chapter.

Table 2.3.1.1 Inheritance patterns of mutations in causative genes implicated in congenital hypopituitarism

Gene	Phenotype	Inheritance
GH1	IGHD type IA	Autosomal recessive
	IGHD type IB	Autosomal recessive
	IGHD type II	Autosomal dominant
GHRHR	IGHD type IB	Autosomal recessive Autosomal dominant (rare)
RNPC3	IGHD	Autosomal recessive
TSHB	TSHD	Autosomal recessive
TRHR	TSHD	Autosomal recessive
IGSF1	TSHD, hypoprolactinaemia, transient GHD; usually with macro-orchidism	X-linked
TBL1X	TSHD	X-linked
TBX19	IAD	Autosomal recessive
POMC	IAD; early-onset obesity and red hair pigmentation	Autosomal recessive
PCSK1	IAD, GHD, TSHD, DI	Autosomal recessive Autosomal dominant (obesity)
TCF7L1	SOD	Autosomal dominant
HESX1	IGHD	Autosomal dominant Autosomal recessive
	MPHD	
	SOD	
SOX2	HH, anophthalmia/microphthalmia	Autosomal dominant/de novo
	Hypothalamo-pituitary tumour	
SOX3	MPHD and absent infundibulum GHD	X-linked
OTX2	SOD	Autosomal dominant: haploinsufficiency or dominant negative
	MPHD	
	IGHD	
LHX3	MPHD, short neck with limited rotation	Autosomal recessive
LHX4	MPHD	Autosomal dominant Autosomal recessive
PROP1	MPHD	Autosomal recessive
POU1F1	MPHD	Autosomal dominant Autosomal recessive
PROKR2	HH/KS	Autosomal recessive Autosomal dominant
	SOD	
FGFR1	HH/KS	Autosomal dominant
	SOD	
FGF8	HH/KS	Autosomal dominant Autosomal recessive
	HPE	
	SOD	
KAL1	HH/KS	X-linked
	SOD	
	KS	
GLI2	HPE	Autosomal dominant: Haploinsufficiency
	IGHD/MPHD	
	HH	
CDON	PSIS	Autosomal dominant
GPR161	PSIS	Autosomal recessive

Table 2.3.1.1 *Continued*

Gene	Phenotype	Inheritance
ROBO1	PSIS	Autosomal dominant
ARNT2	MPHD	Autosomal recessive
PNPLA6	Oliver–McFarlane and Laurence–Moon syndrome	Autosomal recessive
KCNQ1	GHD, maternally inherited gingival fibromatosis	Autosomal dominant
IFT172	GHD, retinopathy, metaphyseal dysplasia, renal failure (ciliopathies)	Compound heterozygous

IGHD, isolated growth hormone deficiency; TSHD, thyroid-stimulating hormone deficiency; IAD, isolated adrenocorticosteroid hormone deficiency; DI, diabetes insipidus; SOD, septo-optic dysplasia; MPHD, multiple pituitary hormone deficiency; HH, hypogonadotropic hypogonadism; KS, Kallmann syndrome; HPE, holoprosencephaly; PSIS, pituitary stalk interruption syndrome.

Combined Pituitary Hormone Deficiency

Combined pituitary hormone deficiency (CPHD) is defined as the presence of at least two or more pituitary hormone deficiencies including GH, TSH, PRL, ACTH, LH, and FSH deficiencies. Mutations in the genes encoding transcription factors *PROP1, POU1F1, LHX3*, and *LHX4*, which are critical for normal pituitary development, often result in unique patterns of CPHD that reflect their differential expression during organogenesis [22]. Depending on the deficiencies in the patient, the phenotypic features may include those of hypothyroidism, delayed or absent puberty which may lead to infertility, intellectual disability, midline defects such as cleft lip or palate, and short stiff neck (specifically caused by *LHX3* mutations). The early acting transcription factors *LHX3* and *–4*, as opposed to the later acting *PROP1* and *POU1F1*, cause deficiencies of most, if not all, pituitary hormones, often referred to as panhypopituitarism or multiple pituitary hormone deficiency (MPHD). Mutated forms of all four genes affect somatotroph cell development and therefore all may give rise to severe short stature, reflecting why growth hormone deficiency (GHD) occurs with a higher prevalence. CPHD with differing endocrine deficits may also stem from mutations in *HESX1, SOX3*, and *OTX2* [18, 23, 24].

CPHD patients with mutations in genes that are implicated in KS have also been described, once again demonstrating genetic overlap with KS. For example, a CPHD patient with right microphthalmia, right renal aplasia and severe developmental delay had a hemizygous variant in *KAL1* that was predicted to be pathogenic [25]. The known *PROKR2* p.R85H mutation was identified in a CPHD patient with GH, TSH, ACTH, LH and FSH deficiencies, and a microphallus; suggesting neonatal gonadotropin-releasing hormone (GnRH) deficiency [19] and thus features of HH/KS. Furthermore, the *PROKR2* p.L173R variant initially identified in KS patients, that affects targeting to the cell surface receptor, has since been frequently identified in both CPHD and SOD patients [26]. A novel *PROKR2* variant, p.R248W, predicted to be deleterious in a CPHD patient, where a glutamine substitution at this highly conserved residue had previously been described, was identified in a patient with HH [27]. A novel loss of function mutation in *FGFR1*, p.R448W, was also recently identified in a patient with GH and TSH deficiency [28].

Furthermore, the transcription factor *GLI2*, a component of the sonic hedgehog (SHH) signalling pathway known to be implicated in HPE and other midline neurodevelopmental anomalies [29, 30], has now been implicated in CH in the absence of midline brain defects [31]. Interestingly, patients with *GLI2* mutations may have variable phenotypes ranging from isolated GHD (IGHD) to CPHD, in combination with variable polydactyly, cleft lip/palate, diabetes insipidus, dysmorphic features, and an ectopic posterior pituitary (EPP) on MRI [32–34]. This does not seem to be apparent for all components of the SHH pathway, for example mutations in the *SHH* gene itself seem to exclusively cause HPE, with no reported cases of CH to date [31].

Despite these reports implying a genetic overlap between CPHD and HH/KS, and HPE and CH cases respectively, digenic inheritance cannot be ruled out here, where another mutation may be responsible for the aetiology of at least one of the deficiencies in their CPHD. Incomplete or variable penetrance of a mutation often occurs in many of these families, especially when *GLI2* or *HESX1* mutations are present, where a heterozygous mutation with functional consequences in the child is present in the unaffected parent or a parent with a mild form of the disease, respectively [31, 35].

Isolated Growth Hormone Deficiency

Congenital IGHD has an incidence between 1/4000 and 1/10 000 live births and is the most prevalent isolated pituitary hormone deficiency. The majority of cases are sporadic with a small percentage (3–30%) of familial cases, and the aetiology of most patients is unknown [36, 37]. GH-releasing hormone (GHRH) released from the hypothalamus binds to its receptor (GHRHR) on somatotrophs resulting in the synthesis and release of GH, in the presence of the transcription factor *POU1F1* [38]. Children with IGHD manifest with moderate to severe short stature associated with a poor growth velocity and delayed skeletal maturation. Recombinant human GH (rhGH) is used to treat the condition and generally achieves a good response in patients [37].

There are four main genetic forms of IGHD. Type 1A autosomal recessive IGHD harbour complete deletions of *GH1* and these patients present with severe growth failure in the first 6 months of life with undetectable GH concentrations. These patients frequently develop anti-GH antibodies after receiving exogenous GH, which prevent the growth response expected on commencing rhGH therapy [39]. Heterogeneous homozygous *GH1* deletions (~6.7 Kb long), were the first, and remain the most common, *GH1* gene alteration in patients with IGHD type IA [40]. Since then, many other severe loss of function *GH1* mutations have subsequently been described.

Type IB GHD is associated with recessive *GH1* and *GHRHR* mutations [41], the latter also being known as Sindh dwarfism. GHD due to *GHRHR* mutations differs from the classical IGHD phenotype, as patients have limited facial dysmorphism and no microphallus, yet still manifest AP hypoplasia (APH) on MRI [42]. These patients

originate from consanguineous pedigrees in Brazil or the Indian subcontinent [43] and usually harbour loss of function *GHRHR* mutations. The first and most common *GHRHR* mutation is p.E72X, resulting in a truncated protein devoid of both the transmembrane and intracellular domains. The p.W273S, p.A176V [36, 44], and p.K329E substitutions have since been reported among many others, with the latter failing to show any cAMP response following administration of GHRH in *in vitro* studies [45]. However, there is some phenotypic variability; an unusually mild form of IGHD in two unrelated families with a novel partial loss of function homozygous *GHRHR* mutation, p.P79L, has recently been described. The patients were compound homozygous with a second variant in *GHRHR*, p.R4Q, that did not appear to compromise the cAMP pathway [46].

Type II GHD, the most common autosomal dominant form of the disease, is often caused by heterozygous *GH1* mutations that affect splicing, resulting in exon skipping [47]. The shorter 17.5kDa GH isoform, resulting from the skipping of exon 3, has been reported to exert a dominant negative effect on GH secretion, with expression levels directly related to severity of the disorder [48, 49]. Other heterozygous *GH1* missense mutations, such as p.E32A, p.R178H, and p.R183H have also been described in this form of GHD. These patients have variable height deficit (occasionally within the normal range) and severity, and may develop additional pituitary hormone deficiencies over time, including ACTH, TSH, and LH/FSH deficiencies [36]. To date, no mutations in *GHRH* have been described in association with IGHD. However, mutations in genes encoding early (*HESX1, SOX2, SOX3*, and *OTX2*) or late (*PROP1* and *POU1F1*) transcription factors implicated in murine and human pituitary development, have been reported in patients with IGHD [23, 37, 50].

Mutations in *RNPC3* have recently been associated with IGHD. The ribonucleic acid (RNA)-binding region (RNA recognition motifs [RRM]) containing 3 on chromosome 1 encodes a 65K protein component of the U12-type spliceosome. Its two RRM motifs suggest interaction with one of the small nuclear RNAs of the minor spliceosome [51]. Biallelic mutations in *RNPC3* have been described in three sisters with severe IGHD and pituitary hypoplasia, where anomalies were identified in the splicing of multiple U12-type introns in patient cells [52]. Despite the unknown direct mechanism underlying the GHD, a subset of 21 genes were found to be affected by this particular splicing. Interestingly, these genes encoded proteins with relevant functions in pituitary development, such as *SPCS2* and *SPCS3* that encode subunits of the signal peptidase complex, implicated in post-translational processing of preprohormones such as preproghrelin to proghrelin [52, 53].

An induced lethal point mutation in *rnpc3* in a zebrafish model resulted in the formation of aberrant U11- and U12-containing snRNAs that sufficiently impaired the efficiency of U12-type splicing, thus causing arrested development in the intestine, liver, and pancreas. This is the only *rnpc3* loss of function *in vivo* model reported to date. Analysis of the zebrafish transcriptome revealed that efficient mRNA processing is critical for the growth and proliferation of cells during vertebrate development [54].

GHD with Accompanying Features

Mutations in *KCNQ1* have recently been described in phenotypically variable GHD patients with mild craniofacial dysmorphic features and maternally inherited gingival fibromatosis [55]. This paternally imprinted gene, encoding the alpha subunit of the voltage-gated ion channel Kv7.1, is expressed in mouse and human somatotroph and gonadotroph cells in the postnatal pituitary, in hypothalamic GHRH neurons during murine development, and in the human hypothalamus [55]. Mutations in *KCNQ1* were previously associated with cardiac arrhythmia syndromes and other heart defects [56]. This suggests how critical ion channels act as regulators in the function of the human pituitary, thus supporting previous data implicating voltage-gated potassium channel currents in pituitary cells [57–59].

A patient with early growth retardation, APH, and an EPP on their MRI, harboured compound heterozygous mutations in the *IFT172* gene. The patient manifested retinopathy associated with metaphyseal dysplasia and hypertension with renal failure, indicative of a ciliopathy [60]. The *IFT172* gene encodes a subunit of the intraflagellar transport (IFT) subcomplex IFT-B, necessary for ciliary assembly and maintenance. Mutations in *IFT172* have previously been associated with skeletal ciliopathies with or without polydactyly, and with retinal, cerebellar, or hepatorenal malformations [61–63]. This signifies a possible role for ciliary function in pituitary development and the bridge between early-onset growth failure and ciliopathies. In further support of this theory, the *ALMS1* gene, which encodes a protein that localizes to the centrosomes and basal bodies of ciliated cells, was identified in a patient with Alström syndrome. This rare autosomal recessive disease is characterized by multiorgan dysfunction and associated with GHD [64].

Additional Isolated Hormone Deficiencies and Malformations

Congenital isolated hormone deficiencies, aside from IGHD, have been reported for all other pituitary cell lineages, such as isolated TSH deficiency (TSHD), isolated hypogonadotropic hypogonadism [IHH; LH and FSH deficiency] that may be part of KS (please see Chapter 2.4.1 'Hypothalamic Dysfunction'), isolated ACTH deficiency (IAD) and, very rarely, isolated PRL deficiency (PRLD) [65].

Defective stimulation of the thyroid gland by TSH causes inadequate thyroid hormone biosynthesis and subsequent central, or secondary, hypothyroidism in affected patients, thus TSHD. Detection of congenital hypothyroidism in neonates by T_4 and TSH screening is a highly sensitive method used to prevent intellectual disability, which may occur due to delayed diagnosis [66]. Rarely, mutations in genes that cause CPHD may present with isolated hormone deficiencies. For example, a homozygous *PROP1* frameshift was identified in a young patient with isolated central hypothyroidism [67].

Mutations in genes regulating TSH biosynthesis and secretion such as *TSHB, TRHR*, and more recently *IGSF1,* have been described in patients with isolated TSHD in rare cases [68, 69]. Patients with mutations in *TSHB*, encoding the TSHβ subunit, previously identified in hypothyroid patients, can have undetectable or highly variable TSH concentrations [70, 71]. The most frequent *TSHB* mutation is termed the homozygous 'hotspot', c.373delT (C105Vfs114X), that causes secondary hypothyroidism in several populations worldwide [71, 72].

Rare recessive biallelic inactivating *TRHR* mutations have been reported in patients with central congenital hypothyroidism (CCH), with absent TSH and prolactin responses to exogenous thyrotropin-releasing hormone (TRH) [73, 74]. The homozygous mutation, p.I131T, that decreases TRH affinity was identified in an overweight patient with CCH and normal stature [68]. Subsequently, the missense *TRHR* mutation (p.P81R) was identified in a patient with isolated CCH, and exemplified the importance of TRH receptor activation in normal TSH secretion [75].

Recently, two X-linked forms of CCH have been identified. Firstly, mutations in *IGSF1* have been described in CCH with macro-orchidism [69, 76]. *Igsf1* is expressed in murine pituitary thyrotroph, lactotroph, and somatotroph cell lineages [69], Leydig and germ cells in murine/human testes, and at low levels in Sertoli cells [77]. It has been suggested that IGSF1 stimulates transcription of TRHR thereby enhancing TSH production and function. In contrast, IGSF1 appears to inhibit the activin-Smad pathway, leading to reduced expression of *FSHB* secreted by gonadotrophs. A large hemizygous 207.873 Kb deletion on Chr. Xq26.2 incorporating *IGSF1* was identified in a patient with reduced TSH biopotency leading to hypothyroidism, macro-orchidism, and an increase in FSH secretion in neonatal minipuberty [77]. Many other *IGSF1* mutations have been reported in such patients, however it should be noted that macro-orchidism is not consistently present in every patient with *IGSF1* mutations [78], and intriguingly, heterozygous female carriers of *IGSF1* mutations may manifest a mild form of hypothyroidism [79]. The second X-linked form, identified in isolated CCH, is caused by mutations in the *TBL1X* gene, encoding a component of the thyroid hormone receptor-corepressor complex, previously associated with sensorineural hearing loss [80, 81].

In contrast to other isolated hormone abnormalities, an abnormal increase in prolactin as opposed to a decrease is more frequent in children with CH, particularly those with midline defects. Isolated PRLD, known as hypoprolactinaemia, clinically manifests only in women as puerperal alactogenesis, namely the failure of milk production during breastfeeding [82], although it remains extremely rare. Patients that have PRLD as part of CPHD are currently screened for mutations in candidate genes known to be involved in the lineage of lactotrophs, such as *POU1F1, PROP1, LHX3, LHX4, HESX1*, and *OTX2* [83]. However, the aetiology of isolated PRLD remains undetermined.

IAD is a rare, heterogeneous, and variable condition, and the diagnosis can be challenging. It can be fatal due to hypocortisolism or cholestasis if unrecognized, and is associated with neonatal hypoglycaemia, convulsions, hypercalcaemia, severe hyponatraemia, and an empty sella on MRI [84–87]. *TBX19*, formally known as *TPIT*, plays a critical role in the terminal differentiation of corticotrophs and melanotrophs; the pre-pro-opiomelanocortin (POMC) lineages. Severe loss of function *TBX19* mutations, most commonly missense in the DNA-binding Tbox domain, have been identified in ~65% of neonatal early-onset IAD cases [88, 89]. Truncated proteins formed by chromosomal deletions or aberrant splicing [90], and compound heterozygosity for the p.Arg222Lysfs*4 mutation biallelic with the known p.R286X mutation, have since been described in IAD patients. Recurrent infections may be a prominent clinical feature of adrenal insufficiency, and should alert clinicians so that diagnostic delay and subsequent life-threatening consequences can be avoided [91].

Recessive mutations in *POMC* have been reported in association with a distinct type of IAD. These patients present with early-onset obesity and red hair, often with adrenal insufficiency, hypocortisolism and hypoglycaemia [92]. Compound heterozygosity for nonsense mutations in *PCSK1*, encoding prohormone convertase (PC) 1, have previously been reported in patients with severe obesity with glucose dysregulation and gonadotrophin deficiency [93], and in isolated obesity [94], respectively. Patients with such mutations may also suffer from malabsorptive severe refractory neonatal diarrhoea, as described in an obese patient with hypoadrenalism, hypoglycaemia, and elevated circulating concentrations of prohormones [95]. Other PC1/3-deficient patients often manifest hypothyroidism and hypocortisolism diagnosed by their low blood T_4 and cortisol concentrations. They can also present with elevated TSH and ACTH, respectively [96], and in rare cases, GHD and diabetes insipidus, thus further expanding the disease manifestation in *PCSK1*-deficient patients [97]. Murine *Pcsk1* (PC1) knockouts present with growth retardation as opposed to obesity, and, together with studies using PC1-deficient human embryonic stem cells, reflect defective POMC and proinsulin processing as observed in human patient phenotypes when *PCSK1* is absent [98, 99].

Pituitary Stalk Interruption Syndrome

Pituitary stalk interruption syndrome (PSIS) is distinguished by a thin or discontinuous pituitary stalk, APH, and/or an EPP on MRI. Until recently, only genes known to cause CH, including *LHX4, OTX2, HESX1, SOX3*, and *PROKR2*, have been described to be mutated in rare cases of PSIS [23, 24, 100]. However, a novel mutation in *CDON*, another member of the SHH signalling pathway that causes HPE, has been reported in a PSIS patient. The patient presented with neonatal hypoglycaemia, cholestasis, and GH, TSH, and ACTH deficiencies, without HPE features [101]. Furthermore, two affected siblings from a consanguineous pedigree had a recessive variant in the *GPR161* gene, encoding the transmembrane orphan G protein-coupled receptor 161 suggesting a possible involvement in the aetiology of PSIS patients. Interestingly, GPR161 is predicted to interact with GLI2, GLI3, and other members of the SHH pathway [102], however functional studies on GPR161 are needed to determine its physiological role during development. These reports repeatedly demonstrate how the SHH pathway may elicit varying phenotypes when its components are mutated.

Another gene recently implicated in PSIS, is *ROBO1*; a receptor involved in Slit/Robo signalling critical for embryonic axon guidance and branching in the nervous system during development [103]. Ocular anomalies including hypermetropia with strabismus, and ptosis were present in the majority of PSIS patients harbouring these mutations [104].

Oliver–McFarlane and Laurence–Moon Syndromes

Mutations in the *PNPLA6* gene, encoding neuropathy target esterase (NTE), have recently been implicated in Oliver–McFarlane

and Laurence–Moon syndromes, two distinct neurodegenerative disorders. Aside from the characteristic phenotypic features in these patients (chorioretinopathy, spinocerebellar ataxia, spastic paraplegia, learning difficulties and trichomegaly), pituitary dysfunction in the form of variable GHD and HH, with a small AP on MRI, is usually present. Human expression studies alongside the generation of a *pnpla6*-morphant zebrafish model and analysis of patient-derived cells, signify that defective recessive *PNPLA6* alleles give rise to rare distinct pituitary-related phenotypes, when in combination with variable neurodegenerative disorders.

Variant Identification and Modern Techniques Used in Functional Analysis

The use of targeted multigene panels has become the initial preferential technique used in many laboratories to screen patients for mutations and novel variants in known and novel causative genes. If no variants are identified using this method then next generation sequencing (NGS) involving a multitude of revolutionized techniques may be used. NGS incorporates whole genome (WGS) or exome sequencing (WES), microarrays, and homozygosity mapping, with the latter technique usually employed when analysing consanguineous pedigrees. Different parameters and filters are used in the data analysis of WES to search for candidate genes and variants with different modes of inheritance. The filter settings are dependent on the phenotype and ethnic background of the patient. This has rapidly replaced the laborious and expensive Sanger sequencing approach to screening which to date has been the standard approach used in the majority of laboratories worldwide. Identification of any novel variants in highly conserved residues that are not present on control databases (e.g. gnomAD browser) may be screened in patients with similar phenotypes for further mutations in the gene of interest. Human embryonic expression, and *in vitro* functional studies investigating the significance of the novel variants may then be conducted.

WGS is the most efficient NGS technique in identifying variants when cost is not an issue. WGS interrogates all parts of the genome, potentially revealing single-nucleotide variants (SNVs), indels, structural variants (SVs), and copy number variants (CNVs) in both the ~1% part of the protein coding sequences, and the remaining ~99% of the non-coding sequence [105]. However, WGS is impractical to perform as a standard approach for every pedigree. Despite WES constituting such a low percentage of the entire genome, it still contains the majority of pathogenic mutations identified to date [106]. Therefore, WES remains the preferred choice as it costs significantly less than WGS due to it being confined to the cDNA only.

The generation of induced pluripotent stem cells (IPSC) from patient fibroblasts [107] allows the potential to differentiate cells into a multitude of relevant cell types, such as neuronal, pancreatic, adipocyte cells, and so on, dependent on the phenotype or pathway to be tested. This technique, although time-consuming, may then allow further investigation using relevant methods such as insulin secretion assays, apoptosis, and proliferation assays, using cells derived directly from the patient. This pioneering technology will limit the need for mouse models and will be more physiologically relevant than the current *in vitro* functional assays.

CRISPR-Cas9 gene editing is now routinely used to knock out a gene more efficiently and faster than previous methods using embryonic stem cells. The insertion of genetic variants such as point mutations may also be conducted using the CRISPR-Cas9 system in order to study the impact of specific variants identified in patients [108, 109]. Upon identification of any significant differences between patient and control cells following assays in patient-derived differentiated IPSC cell lines, rescue experiments may also be performed using the CRISPR-Cas9 system to correct certain variants that are under investigation. Subsequent assays may then be implemented on the rescued cells to see if the correct function or conditions have been restored.

Newer methods of interrogating patient samples and of manipulating the genome will revolutionize the practice of medicine in the future, and the aforementioned techniques mark the dawn of a new era in molecular medicine.

REFERENCES

1. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson IC, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev*, 2009; **30**: 790–829.
2. Davis SW, Castinetti F, Carvalho LR, *et al*. Molecular mechanisms of pituitary organogenesis: in search of novel regulatory genes. *Mol Cell Endocrinol*, 2010; **323**: 4–19.
3. Cohen LE. Genetic disorders of the pituitary. *Curr Opin Endocrinol Diabetes Obes*, 2012; **19**: 33–9.
4. Bancalari RE, Gregory LC, McCabe MJ, Dattani MT. Pituitary gland development: an update. *Endocr Dev*, 2012; **23**: 1–15.
5. Alatzoglou KS, Dattani MT. Genetic forms of hypopituitarism and their manifestation in the neonatal period. *Early Hum Dev*, 2009; **85**: 705–12.
6. McCabe MJ, Gaston-Massuet C, Tziaferi V, *et al*. Novel FGF8 mutations associated with recessive holoprosencephaly, craniofacial defects, and hypothalamo-pituitary dysfunction. *J Clin Endocrinol Metab*, 2011; **96**: E1709–18.
7. De Morsier G. [Studies on malformation of cranio-encephalic sutures. III. Agenesis of the septum lucidum with malformation of the optic tract]. *Schweizer Archiv fur Neurologie und Psychiatrie Archives Suisses de Neurologie et de Psychiatrie Archivio Svizzero di Neurologia e Psichiatria*, 1956; **77**: 267–92.
8. Brodsky MC, Glasier CM. Optic nerve hypoplasia. Clinical significance of associated central nervous system abnormalities on magnetic resonance imaging. *Arch Ophthalmol*, 1993; **111**: 66–74.
9. Webb EA, Dattani MT. Septo-optic dysplasia. *Eur J Human Genet*, 2010; **18**: 393–7.
10. McNay DE, Turton JP, Kelberman D, *et al*. HESX1 mutations are an uncommon cause of septooptic dysplasia and hypopituitarism. *J Clin Endocrinol Metab*, 2007; **92**: 691–7.
11. Lippe B, Kaplan SA, LaFranchi S. Septo-optic dysplasia and maternal age. *Lancet*, 1979; **2**: 92–3.
12. Kelberman D, Dattani MT. Septo-optic dysplasia—novel insights into the aetiology. *Hormone Res*, 2008; **69**: 257–65.
13. Kelberman D, Dattani MT. Genetics of septo-optic dysplasia. *Pituitary*, 2007; **10**: 393–407.
14. Mehta A, Hindmarsh PC, Mehta H, *et al*. Congenital hypopituitarism: clinical, molecular and neuroradiological correlates. *Clin Endocrinol*, 2009; **71**: 376–82.

15. McCabe MJ, Alatzoglou KS, Dattani MT. Septo-optic dysplasia and other midline defects: the role of transcription factors: HESX1 and beyond. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 115–24.
16. Dattani MT, Martinez-Barbera JP, Thomas PQ, *et al*. Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat Genet*, 1998; **19**: 125–33.
17. Sajedi E, Gaston-Massuet C, Signore M, *et al*. Analysis of mouse models carrying the I26T and R160C substitutions in the transcriptional repressor HESX1 as models for septo-optic dysplasia and hypopituitarism. *Dis Models Mech*, 2008; **1**: 241–54.
18. Woods KS, Cundall M, Turton J, *et al*. Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. *Am J Hum Genet*, 2005; **76**: 833–49.
19. Raivio T, Avbelj M, McCabe MJ, *et al*. Genetic overlap in Kallmann syndrome, combined pituitary hormone deficiency, and septo-optic dysplasia. *J Clin Endocrinol Metab*, 2012; **97**: E694–9.
20. McCabe MJ, Hu Y, Gregory LC, *et al*. Novel application of luciferase assay for the *in vitro* functional assessment of KAL1 variants in three females with septo-optic dysplasia (SOD). *Mol Cell Endocrinol*, 2015; **417**: 63–72.
21. Gaston-Massuet C, McCabe MJ, Scagliotti V, *et al*. Transcription factor 7-like 1 is involved in hypothalamo-pituitary axis development in mice and humans. *Proc Natl Acad Sci U S A*, 2016; **113**: E548–57.
22. Pfaffle R, Klammt J. Pituitary transcription factors in the aetiology of combined pituitary hormone deficiency. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 43–60.
23. Thomas PQ, Dattani MT, Brickman JM, *et al*. Heterozygous HESX1 mutations associated with isolated congenital pituitary hypoplasia and septo-optic dysplasia. *Hum Mol Genet*, 2001; **10**: 39–45.
24. Diaczok D, Romero C, Zunich J, Marshall I, Radovick S. A novel dominant negative mutation of OTX2 associated with combined pituitary hormone deficiency. *J Clin Endocrinol Metab*, 2008; **93**: 4351–9.
25. Takagi M, Narumi S, Hamada R, Hasegawa Y, Hasegawa T. A novel KAL1 mutation is associated with combined pituitary hormone deficiency. *Human Genome Variation*, 2014; **1**: 14011.
26. McCabe MJ, Gaston-Massuet C, Gregory LC, *et al*. Variations in PROKR2, but not PROK2, are associated with hypopituitarism and septo-optic dysplasia. *J Clin Endocrinol Metab*, 2013; **98**: E547–57.
27. Asakura Y, Muroya K, Hanakawa J, *et al*. Combined pituitary hormone deficiency with unique pituitary dysplasia and morning glory syndrome related to a heterozygous PROKR2 mutation. *Clin Pediatr Endocrinol*, 2015; **24**: 27–32.
28. Correa FA, Trarbach EB, Tusset C, *et al*. FGFR1 and PROKR2 rare variants found in patients with combined pituitary hormone deficiencies. *Endocr Connect*, 2015; **4**: 100–7.
29. Roessler E, Du YZ, Mullor JL, *et al*. Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proc Natl Acad Sci U S A*, 2003; **100**: 13424–9.
30. Roessler E, Ermilov AN, Grange DK, *et al*. A previously unidentified amino-terminal domain regulates transcriptional activity of wild-type and disease-associated human GLI2. *Hum Mol Genet*, 2005; **14**: 2181–8.
31. Gregory LC, Gaston-Massuet C, Andoniadou CL, *et al*. The role of the sonic hedgehog signalling pathway in patients with midline defects and congenital hypopituitarism. *Clin Endocrinol*, 2015; **82**: 728–38.
32. Franca MM, Jorge AA, Carvalho LR, *et al*. Novel heterozygous nonsense GLI2 mutations in patients with hypopituitarism and ectopic posterior pituitary lobe without holoprosencephaly. *J Clin Endocrinol Metab*, 2010; **95**: E384–91.
33. Franca MM, Jorge AA, Carvalho LR, *et al*. Relatively high frequency of non-synonymous GLI2 variants in patients with congenital hypopituitarism without holoprosencephaly. *Clin Endocrinol*, 2013; **78**: 551–7.
34. Bear KA, Solomon BD, Antonini S, *et al*. Pathogenic mutations in GLI2 cause a specific phenotype that is distinct from holoprosencephaly. *J Med Genet*, 2014; **51**: 413–18.
35. Simm F, Griesbeck A, Choukair D, *et al*. Identification of SLC20A1 and SLC15A4 among other genes as potential risk factors for combined pituitary hormone deficiency. *Genet Med*, 2018; **20**: 728–36.
36. Alatzoglou KS, Kular D, Dattani MT. Autosomal dominant growth hormone deficiency (type II). *Pediatr Endocrinol Rev*, 2015; **12**: 347–55.
37. Alatzoglou KS, Webb EA, Le Tissier P, Dattani MT. Isolated growth hormone deficiency (GHD) in childhood and adolescence: recent advances. *Endocrine Rev*, 2014; **35**: 376–432.
38. Iguchi G, Okimura Y, Takahashi T, *et al*. Cloning and characterization of the 5'-flanking region of the human growth hormone-releasing hormone receptor gene. *J Biol Chem*, 1999; **274**: 12108–14.
39. Cogan JD, Phillips JA, 3rd. GH1 gene deletions and IGHD type 1A. *Pediatr Endocrinol Rev*, 2006; **3** Suppl 3: 480–8.
40. Wagner JK, Eble A, Hindmarsh PC, Mullis PE. Prevalence of human GH-1 gene alterations in patients with isolated growth hormone deficiency. *Pediatr Res*, 1998; **43**: 105–10.
41. Baumann G, Maheshwari H. The Dwarfs of Sindh: severe growth hormone (GH) deficiency caused by a mutation in the GH-releasing hormone receptor gene. *Acta Paediatr Suppl*, 1997; **423**: 33–8.
42. Shohreh R, Sherafat-Kazemzadeh R, Jee YH, Blitz A, Salvatori R. A novel frame shift mutation in the GHRH receptor gene in familial isolated GH deficiency: early occurrence of anterior pituitary hypoplasia. *J Clin Endocrinol Metab*, 2011; **96**: 2982–6.
43. Baumann G. Mutations in the growth hormone releasing hormone receptor: a new form of dwarfism in humans. *Growth Horm IGF Res*, 1999; **9**(Suppl B): 24–9; discussion 9–30.
44. Carakushansky M, Whatmore AJ, Clayton PE, *et al*. A new missense mutation in the growth hormone-releasing hormone receptor gene in familial isolated GH deficiency. *Eur J Endocrinol*, 2003; **148**: 25–30.
45. Salvatori R, Fan X, Mullis PE, Haile A, Levine MA. Decreased expression of the GHRH receptor gene due to a mutation in a Pit-1 binding site. *Mol Endocrinol*, 2002; **16**: 450–8.
46. Gregory LC, Alatzoglou KS, McCabe MJ, *et al*. Partial loss of function of the GHRH receptor leads to mild growth hormone deficiency. *J Clin Endocr Metab*, 2016; **101**: 3608–15.
47. Lee MS, Wajnrajch MP, Kim SS, *et al*. Autosomal dominant growth hormone (GH) deficiency type II: the Del32-71-GH deletion mutant suppresses secretion of wild-type GH. *Endocrinology*, 2000; **141**: 883–90.
48. McGuinness L, Magoulas C, Sesay AK, *et al*. Autosomal dominant growth hormone deficiency disrupts secretory vesicles *in vitro* and *in vivo* in transgenic mice. *Endocrinology*, 2003; **144**: 720–31.

49. Ryther RC, McGuinness LM, Phillips JA, 3rd, *et al*. Disruption of exon definition produces a dominant-negative growth hormone isoform that causes somatotroph death and IGHD II. *Hum Genet*, 2003; **113**: 140–8.
50. Ashkenazi-Hoffnung L, Lebenthal Y, Wyatt AW, *et al*. A novel loss-of-function mutation in OTX2 in a patient with anophthalmia and isolated growth hormone deficiency. *Hum Genet*, 2010; **127**: 721–9.
51. Zhao E, Li J, Xie Y, *et al*. Cloning and identification of a novel human RNPC3 gene that encodes a protein with two RRM domains and is expressed in the cell nucleus. *Biochem Genet*, 2003; **41**: 315–23.
52. Argente J, Flores R, Gutierrez-Arumi A, *et al*. Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. *EMBO Mol Med*, 2014; **6**: 299–306.
53. Yin X, Li Y, Xu G, An W, Zhang W. Ghrelin fluctuation, what determines its production? *Acta Biochim Biophys Sin (Shanghai)*, 2009; **41**: 188–97.
54. Markmiller S, Cloonan N, Lardelli RM, *et al*. Minor class splicing shapes the zebrafish transcriptome during development. *Proc Natl Acad Sci U S A*, 2014; **111**: 3062–7.
55. Tommiska J, Kansakoski J, Skibsbye L, *et al*. Two missense mutations in KCNQ1 cause pituitary hormone deficiency and maternally inherited gingival fibromatosis. *Nature Communications*, 2017; **8**: 1289.
56. Wang Q, Curran ME, Splawski I, *et al*. Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nat Genet*, 1996; **12**: 17–23.
57. Stojilkovic SS, Tabak J, Bertram R. Ion channels and signaling in the pituitary gland. *Endocrine Rev*, 2010; **31**: 845–915.
58. Stojilkovic SS, Bjelobaba I, Zemkova H. Ion channels of pituitary gonadotrophs and their roles in signaling and secretion. *Front Endocrinol*, 2017; **8**: 126.
59. Xu R, Roh SG, Loneragan K, Pullar M, Chen C. Human GHRH reduces voltage-gated K+ currents via a non-cAMP-dependent but PKC-mediated pathway in human GH adenoma cells. *J Physiol*, 1999; **520** Pt 3: 697–707.
60. Lucas-Herald AK, Kinning E, Iida A, *et al*. A case of functional growth hormone deficiency and early growth retardation in a child with IFT172 mutations. *J Clin Endocrinol Metab*, 2015; **100**: 1221–4.
61. Beales PL, Bland E, Tobin JL, *et al*. IFT80, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. *Nat Genet*, 2007; **39**: 727–9.
62. Bredrup C, Saunier S, Oud MM, *et al*. Ciliopathies with skeletal anomalies and renal insufficiency due to mutations in the IFT-A gene *WDR19*. *Am J Hum Genet*, 2011; **89**: 634–43.
63. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. *Pediatr Nephrol*, 2011; **26**: 1039–56.
64. Romano S, Maffei P, Bettini V, *et al*. Alström syndrome is associated with short stature and reduced GH reserve. *Clin Endocrinol*, 2013; **79**: 529–36.
65. Douchi T, Nakae M, Yamamoto S, Iwamoto I, Oki T, Nagata Y. A woman with isolated prolactin deficiency. *Acta Obstetricia et Gynecologica Scandinavica*, 2001; **80**: 368–70.
66. Buyukgebiz A. Newborn screening for congenital hypothyroidism. *J Clin Res Pediatr Endocrinol*, 2013; **5** Suppl 1: 8–12.
67. Wassner AJ, Cohen LE, Hechter E, Dauber A. Isolated central hypothyroidism in young siblings as a manifestation of PROP1 deficiency: clinical impact of whole exome sequencing. *Hormone Res Paediatr*, 2013; **79**: 379–86.
68. Garcia M, Fernandez A, Moreno JC. Central hypothyroidism in children. *Endocr Dev*, 2014; **26**: 79–107.
69. Sun Y, Bak B, Schoenmakers N, *et al*. Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat Genet*, 2012; **44**: 1375–81.
70. Hayashizaki Y, Hiraoka Y, Endo Y, Miyai K, Matsubara K. Thyroid-stimulating hormone (TSH) deficiency caused by a single base substitution in the CAGYC region of the beta-subunit. *EMBO J*, 1989; **8**: 2291–6.
71. Medeiros-Neto G, Herodotou DT, Rajan S, *et al*. A circulating, biologically inactive thyrotropin caused by a mutation in the beta subunit gene. *J Clin Invest*, 1996; **97**: 1250–6.
72. McDermott MT, Haugen BR, Black JN, Wood WM, Gordon DF, Ridgway EC. Congenital isolated central hypothyroidism caused by a 'hot spot' mutation in the thyrotropin-beta gene. *Thyroid*, 2002; **12**: 1141–6.
73. Collu R, Tang J, Castagne J, *et al*. A novel mechanism for isolated central hypothyroidism: inactivating mutations in the thyrotropin-releasing hormone receptor gene. *J Clin Endocrinol Metab*, 1997; **82**: 1561–5.
74. Bonomi M, Busnelli M, Beck-Peccoz P, *et al*. A family with complete resistance to thyrotropin-releasing hormone. *N Engl J Med*, 2009; **360**: 731–4.
75. Koulouri O, Nicholas AK, Schoenmakers E, *et al*. A novel thyrotropin-releasing hormone receptor missense mutation (P81R) in central congenital hypothyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 847–51.
76. Tenenbaum-Rakover Y, Turgeon MO, London S, *et al*. Familial central hypothyroidism caused by a novel IGSF1 gene mutation. *Thyroid*, 2016; **26**: 1693–700.
77. Garcia M, Barrio R, Garcia-Lavandeira M, *et al*. The syndrome of central hypothyroidism and macroorchidism: IGSF1 controls TRHR and FSHB expression by differential modulation of pituitary TGFbeta and Activin pathways. *Sci Rep*, 2017; **7**: 42937.
78. Hughes JN, Aubert M, Heatlie J, *et al*. Identification of an IGSF1-specific deletion in a five-generation pedigree with X-linked central hypothyroidism without macroorchidism. *Clin Endocrinol*, 2016; **85**: 609–15.
79. Joustra SD, Schoenmakers N, Persani L, *et al*. The IGSF1 deficiency syndrome: characteristics of male and female patients. *J Clin Endocrinol Metab*, 2013; **98**: 4942–52.
80. Bassi MT, Ramesar RS, Caciotti B, *et al*. X-linked late-onset sensorineural deafness caused by a deletion involving OA1 and a novel gene containing WD-40 repeats. *Am J Hum Genet*, 1999; **64**: 1604–16.
81. Heinen CA, Losekoot M, Sun Y, *et al*. Mutations in TBL1X are associated with central hypothyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 4564–73.
82. Kauppila A. Isolated prolactin deficiency. *Curr Ther Endocrinol Metab*, 1997; **6**: 31–3.
83. Iwama S, Welt CK, Romero CJ, Radovick S, Caturegli P. Isolated prolactin deficiency associated with serum autoantibodies against prolactin-secreting cells. *J Clin Endocrinol Metab*, 2013; **98**: 3920–5.
84. Ahn SW, Kim TY, Lee S, *et al*. Adrenal insufficiency presenting as hypercalcemia and acute kidney injury. *Int Med Case Rep J*, 2016; **9**: 223–6.
85. Alsaleem M, Saadeh L, Misra A, Madani S. Neonatal isolated ACTH deficiency (IAD): a potentially life-threatening but treatable cause of neonatal cholestasis. *BMJ Case Rep*, 2016; **2016**. DOI: 10.1136/bcr-2016-215032.

86. Bigos ST, Carnes TD. Isolated ACTH deficiency presenting as severe hypercalcemia. *Am J Med Sci*, 1982; **284**: 24–30.
87. Doroftei NA, de Rudder C, de Visscher N, Hanon F. Isolated ACTH deficiency in a patient with empty sella as revealed by severe hyponatremia. *Acta Clinica Belgica*, 2016; **71**: 451–4.
88. Metherell LA, Savage MO, Dattani M, *et al*. TPIT mutations are associated with early-onset, but not late-onset isolated ACTH deficiency. *Eur J Endocrinol*, 2004; **151**: 463–5.
89. Lamolet B, Pulichino AM, Lamonerie T, *et al*. A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell*, 2001; **104**: 849–59.
90. Couture C, Saveanu A, Barlier A, *et al*. Phenotypic homogeneity and genotypic variability in a large series of congenital isolated ACTH-deficiency patients with TPIT gene mutations. *J Clin Endocrinol Metab*, 2012; **97**: E486–95.
91. Akcan N, Serakinci N, Turkgenc B, Bundak R, Bahceciler N, Temel SG. A novel TBX19 gene mutation in a case of congenital isolated adrenocorticotropic hormone deficiency presenting with recurrent respiratory tract infections. *Front Endocrinol*, 2017; **8**: 64.
92. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Gruters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet*, 1998; **19**: 155–7.
93. Jackson RS, Creemers JW, Ohagi S, *et al*. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*, 1997; **16**: 303–6.
94. Philippe J, Stijnen P, Meyre D, *et al*. A nonsense loss-of-function mutation in PCSK1 contributes to dominantly inherited human obesity. *Int J Obesity (Lond)*, 2015; **39**: 295–302.
95. Jackson RS, Creemers JW, Farooqi IS, *et al*. Small-intestinal dysfunction accompanies the complex endocrinopathy of human proprotein convertase 1 deficiency. *J Clin Invest*, 2003; **112**: 1550–60.
96. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JW. PCSK1 mutations and human endocrinopathies: from obesity to gastrointestinal disorders. *Endocrine Rev*, 2016; **37**: 347–71.
97. Martin MG, Lindberg I, Solorzano-Vargas RS, *et al*. Congenital proprotein convertase 1/3 deficiency causes malabsorptive diarrhea and other endocrinopathies in a pediatric cohort. *Gastroenterology*, 2013; **145**: 138–48.
98. Zhu X, Zhou A, Dey A, *et al*. Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc Natl Acad Sci U S A*, 2002; **99**: 10293–8.
99. Wang L, Sui L, Panigrahi SK, *et al*. PC1/3 deficiency impacts pro-opiomelanocortin processing in human embryonic stem cell-derived hypothalamic neurons. *Stem Cell Rep*, 2017; **8**: 264–77.
100. Reynaud R, Jayakody SA, Monnier C, *et al*. PROKR2 variants in multiple hypopituitarism with pituitary stalk interruption. *J Clin Endocrinol Metab*, 2012; **97**: E1068–73.
101. Bashamboo A, Bignon-Topalovic J, Rouba H, McElreavey K, Brauner R. A nonsense mutation in the hedgehog receptor CDON associated with pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*, 2016; **101**: 12–15.
102. Karaca E, Buyukkaya R, Pehlivan D, *et al*. Whole-exome sequencing identifies homozygous GPR161 mutation in a family with pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*, 2015; **100**: E140–7.
103. Blockus H, Chedotal A. The multifaceted roles of Slits and Robos in cortical circuits: from proliferation to axon guidance and neurological diseases. *Curr Opin Neurobiol*, 2014; **27**: 82–8.
104. Bashamboo A, Bignon-Topalovic J, Moussi N, McElreavey K, Brauner R. Mutations in the human ROBO1 gene in pituitary stalk interruption syndrome. *J Clin Endocrinol Metab*, 2017; **102**: 2401–6.
105. Ng PC, Kirkness EF. Whole genome sequencing. *Methods Mol Biol*, 2010; **628**: 215–26.
106. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet*, 2014; **59**: 5–15.
107. Miller JD, Schlaeger TM. Generation of induced pluripotent stem cell lines from human fibroblasts via retroviral gene transfer. *Methods Mol Biol*, 2011; **767**: 55–65.
108. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, 2009; **17**: 904–12.
109. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 2014; **346**: 1258096.

2.3.2 Molecular Pathogenesis of Pituitary Tumours

Shlomo Melmed

Introduction

Although clinically apparent tumours are rarely encountered, pituitary adenomas are apparent in up to 25% of unselected autopsies. The pituitary gland is composed of differentiated cell types: somatotrophs, lactotrophs, corticotrophs, thyrotrophs, and gonadotrophs. Tumours may arise from any of these differentiated cell types and their hormone secretory products depend upon the cell of origin. Functional classification of pituitary tumours is based on identification of differentiation markers by immunostaining or mRNA detection, as well as measurement of circulating tumour and target organ hormone levels. Oversecretion of adrenocorticotropic hormone (ACTH) results in cortisol excess with Cushing disease. Growth hormone (GH) overproduction leads to acromegaly with typical acral overgrowth and metabolic abnormalities. Prolactin (PRL) hypersecretion results in hypogonadism and galactorrhoea. Rarely, thyroid-stimulating hormone (TSH) hypersecretion leads

to goitre and thyrotoxicosis, and gonadotropin excess results in gonadal dysfunction [1]. Mixed tumours cosecreting GH with PRL, TSH, or ACTH may also arise from single cells. Clinically nonfunctional tumours do not efficiently secrete their gene products, and most commonly are derived from gonadotroph cells. Pituitary tumours are defined radiographically as microadenomas (<1 cm in diameter) or macroadenomas (≥1 cm in diameter).

Pituitary tumours cause morbidity by both abnormal hormone secretion as well as compression of regional structures. As a significant proportion of patients do not achieve optimal therapeutic control of mass effects and/or hormone hypersecretion despite advances in therapeutic approaches, understanding pathogenesis and pituitary tumour growth patterns in individual patients enables identification of personalized treatment targets, ultimately decreasing tumour-related morbidity and mortality [2].

Determinants of initiation and progression of pituitary adenomas are not fully understood. This chapter describes a spectrum of mechanisms implicated in pituitary tumorigenesis, including the role of pituitary plasticity, imbalances in cell cycle regulation, transcription factors, signalling pathways, and angiogenesis (Figure 2.3.2.1). Molecular events related to tumorigenesis in human pituitary adenoma subtypes are summarized in Table 2.3.2.1. The causal role for selected genetic imbalances leading to development of pituitary tumours has been recapitulated in several transgenic mouse models (Table 2.3.2.2).

Pituitary Plasticity

Although commitment of pituitary cell function is under cell-specific transcriptional control resulting in differentiated mature cell types (Figure 2.3.2.2), the pituitary gland responds to central and peripheral signals that regulate plastic pituitary cell hormone production and proliferation. Under physiological conditions, hypothalamic and peripheral hormones act in concert to regulate pituitary trophic activity. Age (puberty), and pregnancy/lactation results in increased pituitary volume, and prolonged target gland failure (e.g. hypothyroidism) and oestrogen excess are recognized causes of pituitary hyperplasia. However, there is only modest direct evidence that pituitary hyperplasia is a necessary prerequisite for pituitary tumour development. For example, hyperplastic proliferation of PRL-secreting cells during pregnancy and lactation does not increase the frequency of prolactinomas, while untreated primary hypothyroidism and exogenous oestrogen administration are very infrequently associated with adenoma development. Pituitary hyperplasia caused by ectopic tumour production of growth hormone releasing hormone (GHRH) [3] is very rarely associated with discrete adenoma formation. In general, adenohypophyseal tissue surrounding pituitary tumours is normal, supporting the notion that multiple independent cellular events such as generalized hyperplasia do not necessarily precede adenoma formation. Excess pituitary hormone secretion is usually associated with invariably benign monoclonal adenomas arising from a specific cell type supporting intrinsic pituitary defect in the process of tumour development (Box 2.3.2.1) [4].

Hypothalamic hormones, local growth factors, and circulating sex steroid hormones are likely implicated in enabling a permissive environment which potentiates cell mutation and subsequent tumour growth. Trophic pituitary stimuli may influence the intrapituitary milieu to either enhance or attenuate expansion of a monoclonal tumour cell population.

Cell Cycle Regulation

Retinoblastoma Susceptibility Gene (*RB1*)

The protein encoded by this gene (pRB) is a negative regulator of the cell cycle and behaves as a tumour suppressor. In its active, hypophosphorylated form pRB binds the E2F transcription factors, restraining cell cycle progression from the G1 to S phase.

Mice with homozygous loss of *Rb1* are non-viable, however those with heterozygous *Rb1* inactivation develop pituitary tumours with high penetrance, and less frequently thyroid medullary carcinoma, and pheochromocytoma. Interestingly, mice with deregulated intermediate lobe E2F activity develop tissue hyperplasia that does not progress to tumour formation, likely because sustained E2F activity ultimately triggers premature senescence in a pRB, p16 and p19-dependent manner [5].

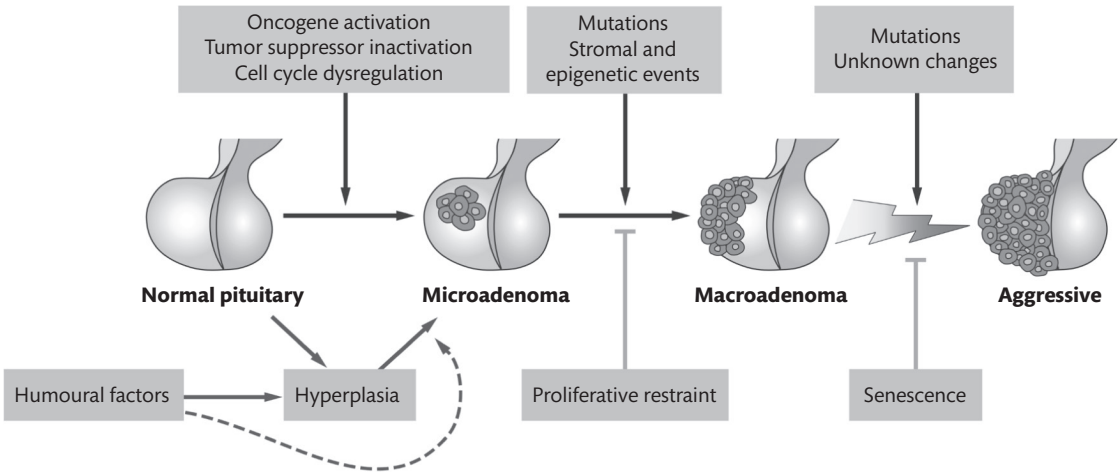


Figure 2.3.2.1 Cascade of pituitary tumorigenesis. Pituitary hyperplasia is usually reversible, as exemplified by the situation that occurs during pregnancy.

Table 2.3.2.1 Selected molecular events related to tumorigenesis in human pituitary adenoma subtypes

	Tumour type	Mechanism of activation/inactivation
Activating		
Gsp	GH adenomas	Activating mutation
CREB	GH adenomas	Increased Ser-phosphorylated CREB promoted by gsp overexpression
Cyclin B2 (*CCNB2*)	All tumour types examined	Overexpression
Cyclin D1 (*CCND1*)	Non-functioning	Overexpression
EGF/EGFR	Non-functioning	Overexpression
PTTG	All tumour types examined	Overexpression
Gal-3	Prolactinomas ACTH adenomas	Overexpression
HMGA2	Non-functioning ACTH adenomas Prolactinomas	Overexpression
FGF-4	Prolactinomas	Overexpression
USP8	Cushing	Mutation
Inactivating		
RB1	Negative pRB in ~25% GH adenomas	Promoter methylation
13q14	Aggressive tumours	13q14 loss of heterozygosity
AIP	15% of FIPA 2% sporadic GH adenomas	Inactivating mutation
MEN1	Prolactinomas in familial MEN1	Inactivating mutation
P16INK4a (*CDKN2A*)	All tumour types examined	Promoter methylation
P27Kip1 (*CDKN1B*)	All tumour types examined	Reduced expression
MEG3a	Non-functioning GH adenomas	Promoter methylation
GADD45-γ	Non-functioning GH adenomas Prolactinomas	Promoter methylation

FIPA, familial isolated pituitary adenomas.

Table 2.3.2.2 Transgenic mouse models for pituitary tumours

	Hyperplasia/Adenoma [b]
Gene overexpression [a]	
CMV.**HMGA1**	GH, PRL
CMV.**HMGA 2**	GH, PRL
Ubiquitin C.**hCG**	PRL
αGSU.**bLH**	Pit-1 lineage
GH.**galanin**	GH, PRL
PRL.**galanin**	PRL [c]
PRL.**TGFα**	PRL
αGSU.**PTTG1**	LH, GH, TSH
αGSU.**Prop1**	Non-functioning
PRL.**pdt-FGFR4**	PRL
Gene inactivation	
p27/Kip1 $^{-/-}$	ACTH, αMSH
p18/INK4c $^{-/-}$	ACTH, αMSH
Rb$^{+/-}$	ACTH, αMSH αGSU, GH, βTSH
D2R-deficient	PRL
Men1$^{+/-}$	PRL
PRL$^{-/-}$	Non-functioning

[a] Genes are listed in bold, and are preceded by the promoter that determines transcriptional control;
[b] Hormone immunoreactivity/secreting profile;
[c] Pituitary hyperplasia, with no tumour formation.

CMV, cytomegalovirus; PRL, prolactin; HMGA, high mobility group A; pdt-FGFR4, pituitary tumour-derived fibroblast growth factor receptor-4; TGFα, transforming growth factor-alpha; Men1, multiple endocrine neoplasia type 1; PTTG, pituitary transforming gene.

Reproduced with permission from Donangelo I, Melmed S. *Molecular Pathogenesis of Pituitary Tumors*. 2nd ed. Massachusetts: Blackwell Publishing; 2008.

Individuals who inherit a defective copy of *RB1* gene are at high risk for developing retinoblastoma at an early age, however, interestingly these patients do not exhibit a predisposition to development of pituitary adenomas. Aggressive human pituitary tumours and rarely encountered metastasis exhibit loss of heterozygosity of region 13q14 (*RB1* locus), however pRB usually remains expressed suggesting that a tumour suppressor gene other than *RB1* present in the same chromosomal region may be related to pituitary tumour progression. Studies based on immunodetection in tumour sections found normal expression of pRB in most non-functioning pituitary adenomas, however approximately 25% of GH-secreting adenomas exhibit loss of pRB expression, and this finding did not correlate with tumour behaviour. In some cases, decreased *RB1* expression correlated with promoter hypermethylation [6]. It is likely that *RB1* inactivation may be involved in human pituitary tumour development in a small subset of adenomas.

INK4 Family

Cell cycle is regulated by two families of cyclin-dependent kinase (CDK) inhibitors, the INK4 family and the Cip/Kip family. The INK4 family (p16INK4a, p15INK4b, and p18INK4c) inhibit G1/S progression by binding CDK4 and CDK6.

The protein p16INK4a, encoded by the *CDKN2A* gene on chromosome 9p21, maintains pRB unphosphorylated (active) by blocking CDK4. p16INK4a is not expressed due to promoter hypermethylation in most non-functioning pituitary tumours and in a smaller subset of other pituitary tumour subtypes. Loss of p16INK4a and pRB in tumours tend to be mutually exclusive, likely because functional pRB is required for cell cycle inhibition by p16INK4a and loss of both regulators of the cell cycle would not provide an additive growth advantage. Promoter hypermethylation of the *CDKN2B* gene that encodes p15INK4b was also noted in a subset of pituitary tumours. p18INK4c-deficient mice develop gigantism and widespread organomegalia, and proopiomelanocortin (POMC) intermediate lobe pituitary hyperplasia and tumours. P18INK4c is significantly underexpressed in human ACTH-secreting adenomas.

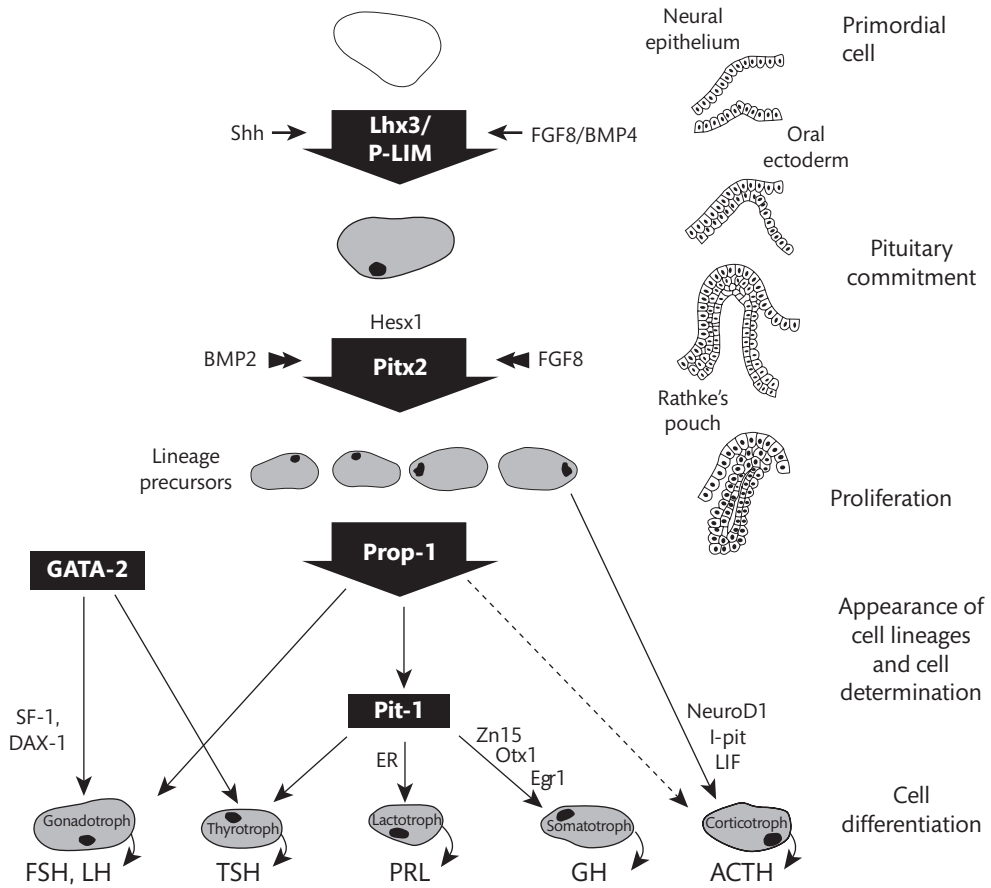


Figure 2.3.2.2 Model for development of human anterior pituitary cell lineage determination by a temporally controlled cascade of transcription factors. Trophic cells are depicted with transcription factors known to determine cell-specific human or murine gene expression.

Cip/Kip Family

Members of the Cip/Kip family (p21Cip1, p27Kip1, and p57Kip2) restrain cell cycle progression by associating with CDK1 and CDK2 complexes.

Mice with disrupted *CDKN1B* gene that encodes p27Kip1 develop overall increased body weight, multiorgan hyperplasia, female infertility, POMC intermediate lobe pituitary tumours, and hyperplasia of haematopoietic organs [7]. Intermediate lobe tumours derived from p27Kip1 null mice differ from those derived from $Rb^{+/-}$ mice in that they are more vascular, exhibit lower proliferation rates, and are oligoclonal or polyclonal (vs. monoclonal in $Rb^{+/-}$ mice).

Downregulation of p27Kip1 protein expression is common in corticotroph tumours and pituitary carcinomas, and p27Kip1 levels are lower in recurrent pituitary adenomas compared to non-recurrent adenomas. p27Kip1 mRNA expression is generally not decreased in pituitary tumours, suggesting that decreased p27Kip1 expression is likely due to post-translational factors.

P21-deficient mice develop increased risk for multiple neoplasia in late adult life, however, pituitary tumours are uncommon, and p21Cip deletion or mutation is not commonly encountered in human tumours. The role of p21Cip in oncogene-induced pituitary senescence is discussed below.

Box 2.3.2.1 Evidence for an intrinsic pituitary defect in the pathogenesis of pituitary tumours

- Pituitary adenomas are monoclonal
- Absence of pituitary hyperplasia in tissue surrounding pituitary adenomas
- Surgical resection of well-circumscribed pituitary adenomas controls >75% of patients
- Adenoma formation is rarely associated with generalized pituitary hyperplasia
- Unrestrained pituitary hormonal hypersecretion occurs independent of physiological hypothalamic feedback regulation
- Normalization of hormonal pulsatility pattern often occurs after adenoma resection

p53

p53 inhibits cell cycle progression or induces apoptosis, however, this tumour suppressor appears not to play a major role in the pathogenesis of pituitary adenomas.

Cyclins

Cyclins D and E regulate the G1 to S phase of cell cycle progression. Cyclin D1, D2, and D3 are the first wave of cyclins to be upregulated when quiescent cells (G0) enter the proliferative cell

cycle. Cyclin-CDK complexes induce phosphorylation (inactivation) of pRB, releasing E2F to prompt cell cycle transition to S phase. *CCND1* encoding cyclin D1 was studies in pituitary tumours. Allelic imbalance at the *CCND1* locus was found to be more frequent among invasive tumors, while cyclin D is more expressed in aggressive and non-functioning pituitary adenomas, than in GH-secreting adenomas [8]. However, *CCND1* allelic imbalance and cyclin D1 expression may not coexist in the same tumour, suggesting that it may not be a primary event in pituitary tumorigenesis. Similarly, cyclin A, B, and E are also more abundant in larger, highly proliferative pituitary adenomas.

Classic oncogene mutations have not been encountered in pituitary tumours [9, 10] but genomic profiling shows altered copy number variation with evidence for genomic instability [11]. Copy number variations are more prominently encountered in hormone-secreting tumours, and those also harbouring *GNAS* mutations [12].

PTTG

Overexpression of the pituitary tumour-transforming gene (*PTTG*), isolated from rat GH-secreting pituitary tumour cells, induces cellular transformation *in vitro* and tumour formation in nude mice. PTTG is a mammalian securin, a key regulator of metaphase to anaphase transition during mitosis, and overexpression causes aneuploidy by inhibiting sister chromatid separation [13]. PTTG also plays a role in pathways responsible for DNA break repair [14–16]. Because PTTG abundance correlates with pituitary gland trophic status, regulation of this gene may subserve a mechanism for affecting tumour formation (Figure 2.3.2.3). Global *Ptg* inactivation results in hypotrophic effects (i.e. pituitary, pancreatic β-cell, splenic, and testicular hypoplasia). *Ptg* inactivation in $Rb^{+/-}$ protects mice from pituitary tumour development, and combined $Rb^{+/-}$ and targeted pituitary *PTTG* overexpression further enhances pituitary hyperplasia and tumour prevalence [17].

PTTG is abundantly expressed in most pituitary tumours and PTTG tumour content is determined by RSUME-mediated sumoylation [18]. *PTTG1* mRNA was overexpressed in 54 pituitary tumours assessed by RT-PCR (23/30 of non-functioning, 13/13 GH-secreting, 9/10 prolactinoma, and 1/1 ACTH-secreting), *PTTG* expression correlates well with clinical tumour invasiveness [19].

Transgenic mice with human *PTTG* targeted to the pituitary under the α-subunit of glycoprotein hormone (αGSU) promoter exhibited gonadotroph, thyrotroph, and somatotroph focal hyperplasia and small adenomas, with elevated serum LH, testosterone, GH, and/or IGF-I levels, and prostate and seminal vesicle hypertrophy. Pituitaries derived from compound double transgenic αGSU.*PTTG1;* $Rb^{+/-}$ are enlarged as early as 2 months of age, and the incidence of anterior lobe tumours increased 3.5-fold (Figure 2.3.2.4c), suggesting that targeted *PTTG* overexpression in anterior lobe αGSU cells facilitates tumour formation [17].

Transcription Factors

The process of adenohypophyseal differentiation is a highly specific and temporally regulated series of events. Expression of transcription factors, such as PROP1, Pit-1 (POU1F1), and DAX-1 in pituitary adenomas reflects the origin of tumour cells, and possibly their level of differentiation. However, whether or not dysregulation of these transcription factors plays a causal role in the development of human pituitary tumours remains unclear.

Signalling Pathways

Guanine Nucleotide-Activating α-subunit *(GNAS)*

The McCune–Albright syndrome comprises defects in bony skeleton and skin, precocious puberty, thyrotoxicosis, acromegaly, gigantism, or Cushing syndrome. The molecular defect is a mutation in the *GNAS* gene that encodes Gsα protein, termed oncogene *gsp* which induces constitutive adenylate cyclase activation.

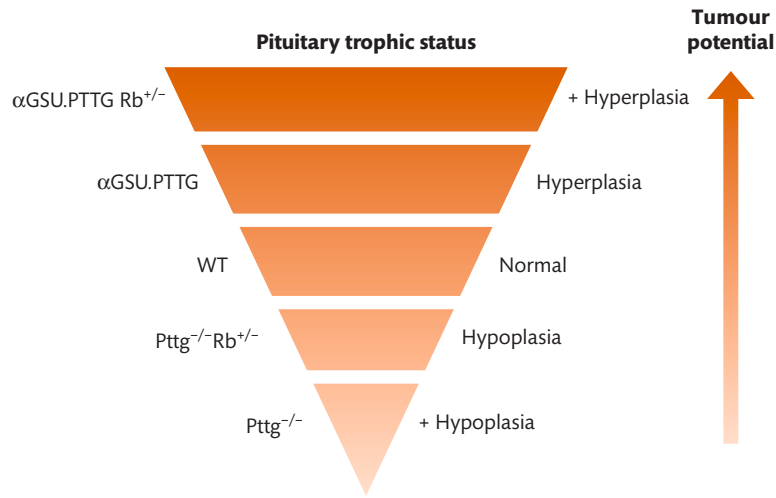


Figure 2.3.2.3 Pituitary PTTG content correlates with gland plasticity and with tumour formation potential. On the left side of the inverted triangle are listed mouse models with descending pituitary PTTG content, with or without the combination with tumorigenic $Rb^{+/-}$. Horizontal bars composing the inverted triangle represent the observed effects of the different genotypes on pituitary trophic status, which correlates with pituitary tumorigenic potential (arrow).

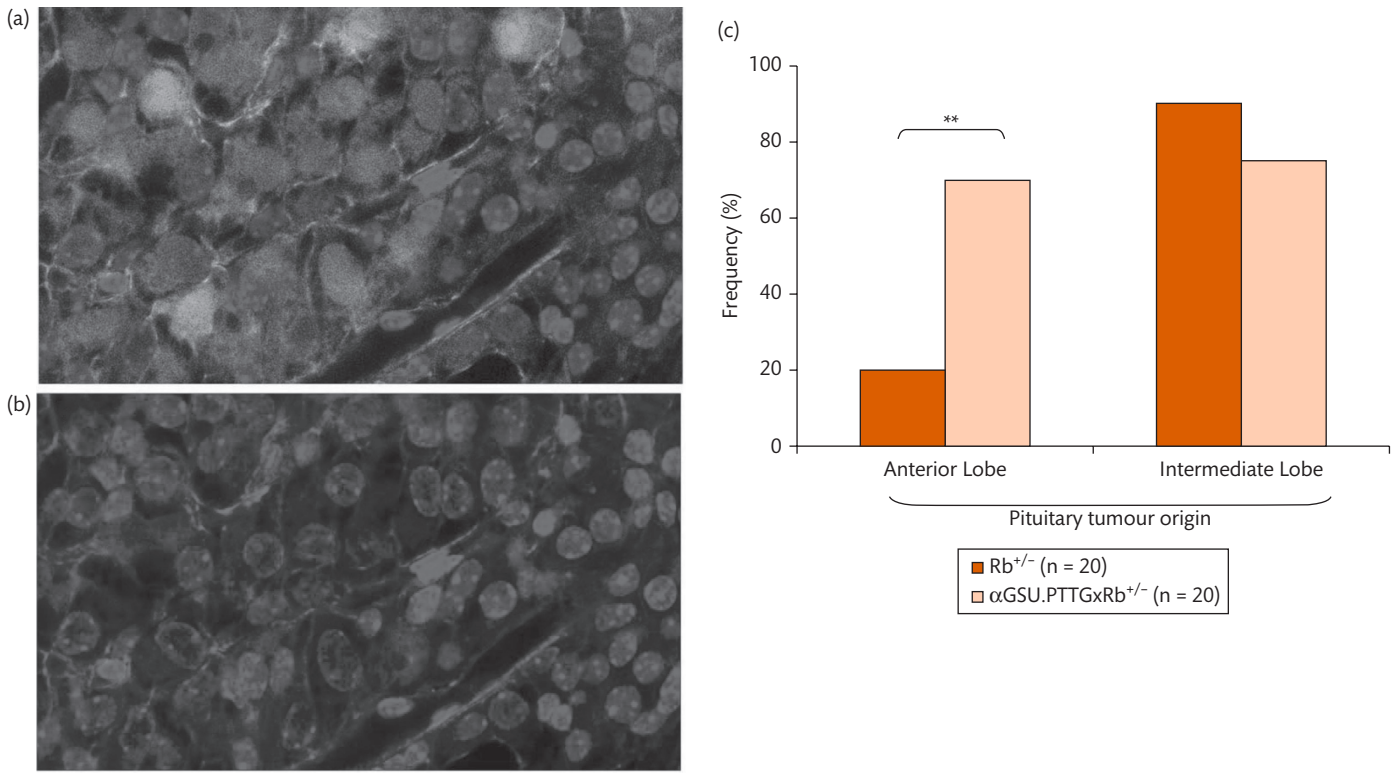


Figure 2.3.2.4 Targeted *PTTG* overexpression to anterior lobe pituitary cells results in cell hyperplasia and increased tumour formation. Panels (a) and (b) are duplicates of the same image, overview of pituitary cells expressing *αGSU.PTTG1.IRESeGFP* transgene. (a) is the untouched image, and in (b) the *green layer* (eGFP) has been hidden for better visualization of nuclear morphology. Contrast between eGFP positive (overexpressing *PTTG*) and eGFP negative (normal *PTTG* content) can be appreciated, notably presence of macronuclei and reorganization of chromatin suggestive of hyperplastic cells. Panel (c) depicts that bitransgenic *αGSU.PTTG;Rb*$^{+/-}$ mice exhibit higher prevalence of anterior lobe and similar prevalence of intermediate lobe pituitary tumours when compared with *Rb*$^{+/-}$ mice. Pathological analysis of pituitary tumours reveals that frequency of tumours arising from anterior lobe is higher in *αGSU.PTTG;Rb*$^{+/-}$ (white bars) than in *Rb*$^{+/-}$ (black bars) pituitary tumours (**, $P = 0.0036$), but frequency of tumours arising from the intermediate lobe (where there was no *PTTG* overexpression) is similar. n, Total number of pituitary tumours analysed.

In GH-secreting cells, *gsp* activates GHRH postreceptor pathways (i.e. cell proliferation and hormone secretion without necessarily requiring ligand binding to the GHRH receptor).

Although *gsp* mutations are reported in 30–40% of GH-secreting adenomas, major clinical differences between *gsp*+ and *gsp*– pituitary tumours have not been identified. *GNAS* is imprinted in normal pituitary tissue and only the maternal allele is expressed. In GH-secreting adenomas, *gsp* activating mutations mostly occur in the maternal allele. The *gsp* oncogene has been detected in <10% of non-functioning or ACTH-secreting tumours.

Activated cAMP-Response Element Binding Proteins (CREB)

The direct mechanism by which cAMP stimulates somatotroph GH-transcription may be mediated by the cAMP-responsive nuclear transcription factor (CREB), which binds as a dimer to cAMP-response elements. Transgenic mice overexpressing a phosphorylation-deficient and transcriptionally inactive mutant of CREB in the anterior pituitary exhibit dwarfism and somatotroph hypoplasia, indicating that phosphorylated CREB plays a role as a biochemical intermediate in the somatotroph proliferative response. Significantly higher amounts of Ser^{133}-phosphorylated, and hence, activated CREB was detected in a series of GH-secreting pituitary tumours compared to a group of non-functioning tumours, suggesting that constitutively activated CREB, possibly promoted by Gsα overexpression, may be another factor facilitating somatotroph transformation.

Dopamine Receptors

The dopamine 2 receptor (D2R) mediates inhibitory effects of dopamine on pituitary PRL synthesis and secretion. D2R-deficient mice exhibit hyperprolactinemia and lactotroph hyperplasia with late progression to pituitary tumours, suggesting that loss of dopamine inhibition induces murine neoplastic transformation. This finding likely cannot be extrapolated to pituitary tumour development in humans as D2R mutations have not been identified. Decreased D2R expression has been linked to dopamine agonist resistance in prolactinomas.

Somatostatin Receptors

Somatostatin receptors (SST) 1 through 5 are expressed in all pituitary tumour types [20]. Prolactinomas exhibit high SST1 expression compared to other tumour types. SST3 is frequently detected in non-functioning adenomas, while SST4 expression is relatively infrequent in pituitary tumours. In GH-secreting adenomas, SST2 and SST5 expression does not correlate with

tumour behaviour; however lower SST2 content is observed in adenomas less responsive to somatostatin analogue therapy. A germline mutation in the coding sequence of SST5 was reported in a single patient with acromegaly resistant to somatostatin analogue therapy, and no mutations have been identified in other SST subtypes. SST2 and SST5 variants do not correlate with responsiveness to somatostatin analogue therapy. The role of these receptors in the pathogenesis of pituitary adenomas remains unproven.

Somatotroph Signalling

GH production is induced by hypothalamic GHRH as well as by intracellular factors. Mice bearing a GHRH transgene develop mammosomatotroph hyperplasia that may convert to adenomas in older mice. Extrahypothalamic tumours secreting ectopic GHRH induce somatotroph hyperplasia and acromegaly that rarely progress to somatotroph adenomas [21]. The GHRH receptor (GHRHR) may be overexpressed in GH-secreting adenomas compared to normal pituitary tissue [22], but the significance of increased expression of GHRHR in the pathogenesis of GH-secreting adenomas is not clear. A related protein, GPR101 is overexpressed in a subset of short stature patients (X-linked acrogigantism) with Xq26.3 microduplication [23]. Pituitary STAT3 abundance is induced by GH and participates in a closed loop autocrine stimulation of GH leading to excess tumour hormone production. This pathway offers another target for subcellular therapy of these tumours [24].

Angiogenesis and Growth Factors

Vascularization is decreased in pituitary tumours compared to normal tissue, in marked contrast to the pattern observed in other tumour types where cancer development is linked to increased angiogenesis. In contrast to the normal pituitary which is predominantly supplied by the hypothalamic-pituitary portal vein, pituitary adenomas receive a direct systemic blood supply, and the relatively low vascular density in these tumours may occur in association with an in-growth of systemic capillaries which may dilute intrapituitary concentrations of hypothalamic factors, especially dopamine.

Microvascular density is decreased in pituitary adenoma tissue, and macroprolactinomas, invasive prolactinomas, or pituitary carcinomas have higher microvascular density compared, respectively, to microprolactinomas, non-invasive prolactinomas, or pituitary adenoma. Conversely, angiogenesis is not enhanced with tumour growth in GH-secreting adenomas, consistent with the notion that microprolactinomas may represent a pathological and clinical entity distinct from macroprolactinomas, whereas different-sized GH-secreting adenomas are components of a similar disease spectrum [25].

Vascular Endothelial Growth Factor (VEGF)

VEGF increases proliferation and migration of endothelial cells, and functions as an antiapoptotic factor promoting vessel endothelial cells survival. VEGF is detected in both the normal pituitary gland and in pituitary adenomas, and levels correlate with tumour behaviour, as VEGF is more abundant in carcinomas and macroprolactinomas. Pituitary glands derived from DR2-knockout female mice have increased VEGF expression compared with wild-type mice.

Fibroblast Growth Factors (FGFs)

FGFs participate in cell development, growth, and angiogenesis, and basic FGF (bFGF or FGF-2) is abundantly expressed in the pituitary. During the hyperplastic phase of prolactinoma development, pituitary expression of both PTTG and bFGF is increased in a time- and dose-dependent manner in oestrogen-treated rats, and bFGF synthesis is induced in NIH-3T3 cells overexpressing PTTG. In the rodent, pituitary bFGF has been localized primarily to folliculostellate cells and regulates GH, prolactin, and TSH secretion. In murine folliculostellate cells, FGF-2 induced positive autofeedback with protein kinase C mediated FGF-2 auto-induction, and stimulates cell proliferation and increased Pttg expression.

FGF-4 is encoded by heparin-binding secretory transforming (*hst*) gene and is expressed in about 30% of PRL-secreting pituitary adenomas. FGF-4 expression in prolactinomas correlates with tumour invasiveness, and GH4 cells transfected with *hst* form more aggressive tumours *in vivo* [26]. *hst* gene rearrangement have not been detected in human prolactinomas, and mechanisms by which *hst*/FGF-4 complex initiates or promotes lactotroph proliferation and prolactin secretion are unclear.

ErbB Receptors

Epidermal growth factor (EGF) has potent mitogenic activity in pituitary cells, and both EGF and epidermal growth factor receptor (EGFR) may be overexpressed in pituitary tumours, particularly in non-functioning adenomas. Examination of ErbB2 and ErbB3 expression in prolactinomas revealed positive results for both ErbB2 (7 of 8 tumours) and ErbB3 (4 of 8 tumours) especially in aggressive, recurrent tumours [27].

EGF activates EGFR and ErbB signalling in lactosomatotroph rat GH3 cells, with resulting increased PRL and GH expression. EGFR antagonist gefitinib decreases GH3 cell proliferation and PRL secretion and attenuates growth and hormone production of GH3 derived tumours in nude mice through blockage of EGFR/ERK signalling pathway [28]. Cushing disease is associated with enhanced EGFR signalling [29] and dominant *USP8* mutations lead to increased EGFR de-ubiquitination with increased ACTH production in Cushing disease [30]. These results suggest that targeted ErbB receptor inhibition could be a potential therapeutic alternative.

Familial Syndromes

Multiple Endocrine Neoplasia Type 1

Pituitary adenomas occur in a familial setting in about 5% of all cases, and over half of the cases are due to multiple endocrine neoplasia type 1 (MEN1). MEN1 is an autosomal dominant genetic disease characterized by parathyroid adenoma, pancreatic endocrine tumours, and pituitary adenomas (Table 2.3.2.3). Pituitary adenomas occur in approximately 25% of patients and these tumours may be larger and more aggressive than sporadic counterparts. Most secrete PRL, with or without secretion of excess GH, followed by those secreting GH alone, non-functional tumours, and those secreting excessive ACTH. MEN1 is caused by inactivating mutations of the tumour suppressor gene, *MEN1* [31]. *MEN1* encodes menin, a nuclear protein expressed in all organs and tissues of the body that interacts with both nuclear and cytoplasmic partners to regulate gene transcription, DNA repair, and cytoskeletal organization.

Table 2.3.2.4 Pituitary tumour genetic syndromes

	Gene	Chromosomal locus	Main clinical characteristics
Carney complex	*PRKAR1A*	17q24.2	Skin pigmentation; cardiac and cutaneous myxomas; thyroid, testis and adrenal tumours; GH-cell hyperplasia or adenoma
Familial isolated pituitary adenomas	*AIP*	11q13.2	Young familial invasive GH adenomas, often therapy resistant
Gigantism	*GPR101*	Xq26.3	X-linked gigantism due to somatotroph hyperplasia or adenoma
McCune–Albright	*GNAS*	20q13.32	Polyostotic fibrous dysplasia, café-au-lait spots, and precocious puberty with GH and/or PRL excess
MEN1	*MEN1*	11q13.1	Pancreatic, pituitary, and parathyroid gland tumours
MEN4	*CDKN1B*	12q13.1	MEN1 like, usually with GH-secreting adenomas

Hundreds of inactivating *MEN1* mutations have been identified, and it is unclear why *MEN1* mutations cause selected endocrine tumours while menin is ubiquitously expressed. Homozygous murine *Men1* deletions result in embryonic lethality while heterozygous *Men1* deletion results in pituitary tumour formation.

Rarely p27Kip1/CDKN1B functions as a tumour suppressor in patients with clinical MEN1 but without *MEN1* mutations [32]. Rare mutations in other CDK inhibitor genes, including p15INK4b/CDKN2B, p18INK4c/CDKN2C, and p21Cip1/CDKN1A, were identified in families with MEN1 phenotypes with no identifiable germline *MEN1* mutations.

Carney Complex

Carney complex is rare autosomal dominant genetic disease with myxomas of the heart, skin hyperpigmentation, and endocrine overactivity. GH-secreting adenomas are the most common pituitary tumours encountered in these patients. Up to 75% of patients have elevated levels of GH, IGF-1, or PRL, and 10% of patients exhibit clinical acromegaly. Somatotroph hyperplasia is followed by GH-secreting tumour formation associated with inactivating mutations of *PRKAR1A*, the regulatory subunit isoform 1A of protein kinase A (PKA). Inactivating *PRKAR1A* mutations result in constitutive activation of PKA catalytic subunit. Mice with a specific homozygous deletion of *PRKAR1A* develop somatotroph hyperplasia, elevated GH levels, and GH-secreting adenomas. In some patients, and in animal models, the wild-type *PRKAR1A* allele is retained in tumour tissue and it appears that decreased expression (i.e. haplo-insufficiency), rather than absence of the PKA regulatory subunit is sufficient to cause tumorigenesis.

Aryl Hydrocarbon Receptor-Interacting Protein (AIP)

Familial isolated pituitary adenoma is a syndrome defined as two or more members in a family harbouring anterior pituitary tumours without evidence of MEN1 or Carney complex. Whole-genome single-nucleotide polymorphism genotyping was performed on three Finnish families with very-low-penetrance susceptibility to pituitary adenomas. Linkage analysis provided evidence for linkage in chromosome 11q12–11q13, a region previously implicated in isolated familial somatotropinomas (IFS) [33]. No mutations or altered expression in *MEN1* were detected in this cohort. Mapping of the linked chromosomal region identified *AIP*, or aryl hydrocarbon receptor-interacting protein gene, with loss of heterozygosity detected in pituitary adenomas with *AIP* germline mutations, suggesting that AIP may behave as a tumour suppressor gene. Of 73 families with the syndrome of familial isolated pituitary adenomas, 11 (15.1%) harbour at least 10 different germline mutations in the *AIP* gene [34]. Most of these patients exhibit a young age of onset for GH oversecretion with macro-adenomas [35]. However, ~2% of patients with sporadic pituitary tumours have *AIP* mutations in germline DNA, suggesting that the majority of sporadic pituitary tumours are not associated with *AIP* mutations [36, 37].

Other Molecular Events

Growth Arrest and DNA Damage-Inducible Gene 45γ (*GADD45-γ*)

GADD45-γ identified as a candidate tumour suppressor gene for pituitary adenomas by cDNA-representational difference analysis (cDNA-RDA), is expressed in the normal pituitary gland, but absent in most non-functioning, GH-, and PRL-secreting adenomas, as well as in immortalized pituitary cell lines. *GADD45-γ* a p53-responsive gene induced by DNA damage mediates growth suppression and apoptosis. Introducing *GADD45-γ* into a rat pituitary tumour cell line decreases cell proliferation and anchorage-independent colony formation. Silencing of *GADD45-γ* in pituitary tumours likely occurs by epigenetic changes (i.e. methylation of CpG islands in the *GADD45-γ* promoter [23]).

Maternally Expressed 3 Gene (*MEG3*)

An isoform of *MEG3* contains an extra exon (*MEG3a*), and has been identified by cDNA-RDA, and is expressed in normal human pituitary, brain, and other tissues, but is diminished or absent in pituitary tumours and human cancer cell line [38]. MEG3a is undetectable in both non-functioning and GH-secreting pituitary adenomas, while introduction of *MEG3a* in cancer cells inhibits proliferation and decreases colony formation. Hence, loss of *MEG3a* in pituitary adenomas likely confers a tumour growth advantage. *MEG3* silencing in non-functioning pituitary adenomas is likely due to promoter hypermethylation.

High Mobility Group A (*HMGA*)

The HMGA family includes the related HMGA1 and HMGA2 abundantly expressed during embryogenesis, but not in normal adult tissues, including the pituitary gland. Transgenic *HMGA1* and *HMGA2* overexpression in mice causes GH-secreting adenomas and prolactinomas. Trisomy of chromosome 12, which harbours *HMGA2*, represents the most frequent cytogenetic alteration in human PRL-secreting pituitary adenomas, and *HMGA2*

overexpression was detected in a number of prolactinomas harbouring rearrangement of regions 12q14–15. *HMGA2* expression was present in 38 of 98 (39%) pituitary adenomas, and noted to be more frequent in FSH/LH cell adenomas (15/22, 68%), prolactinomas (5/15, 31%), and ACTH-secreting adenomas (12/18, 18%), however it was rarely detected in GH or mixed GH/PRL-secreting adenomas [39]. High *HMGA2* levels correlate with tumour size, invasiveness, and cell proliferation marker. *HMGA2* tumorigenic effects may be mediated by stimulation of cyclin B2 expression by HMGA2 binding to the *CCNB2* promoter [40] and by activation of the E2F pathway. There is also evidence that *HMGA2* is suppressed by microRNA *Let-7*, a putative tumour suppressor, and *HMGA2* and *Let-7* expression correlate inversely in human pituitary adenoma samples [39].

Senescence

Cellular senescence is characterized by cell growth arrest induced by diverse mechanisms including age linked telomere shortening, DNA damage, oxidative stress, chemotherapy, and oncogene activation, and has emerged as a potential new component of cancer-protective response to oncogenic events. In response to oncogene activation, protective cellular mechanisms subject the cell to apoptosis and senescence. Apoptosis of cells with oncogene removes them from the tissue population thereby completely preventing tumorigenesis. Oncogene-included senescence (OIS) is a largely irreversible process in which proliferative arrest is mediated thorough upregulation of cell cycle inhibitors including p16INK4A, p15INK4B, p21Cip1, p53, and pRB. OIS also involves participation of cytokine and chemokine pathways which may be protective for malignant transformation [41].

Cellular senescent markers are noted to be elevated in benign tumours but not in malignant carcinomas. Indeed, p21Cip1 and senescence activity were found to be elevated in human GH-secreting adenomas and in rat GH3 pituitary cells [28]. In 38 GH-secreting adenomas, 29 exhibited strong and 9 weak p21 staining. In contrast, p21 was not detected in GH-producing pituitary carcinomas, non-secreting pituitary oncocytomas, or null cell adenomas, or in aggressive breast carcinoma. Senescence-associated β-galactosidase activity (SA-β-gal), a marker of senescence, is strongly positive in GH-secreting adenomas (**Figure 2.3.2.5**). P21 and PTTG levels strongly correlated in these pituitary tumours.

Activation of the p21/p53 senescence pathway is noted to occur with both pituitary PTTG overexpression and deficiency. *Pttg*-null mice exhibit pituitary activation of senescence-like features, including

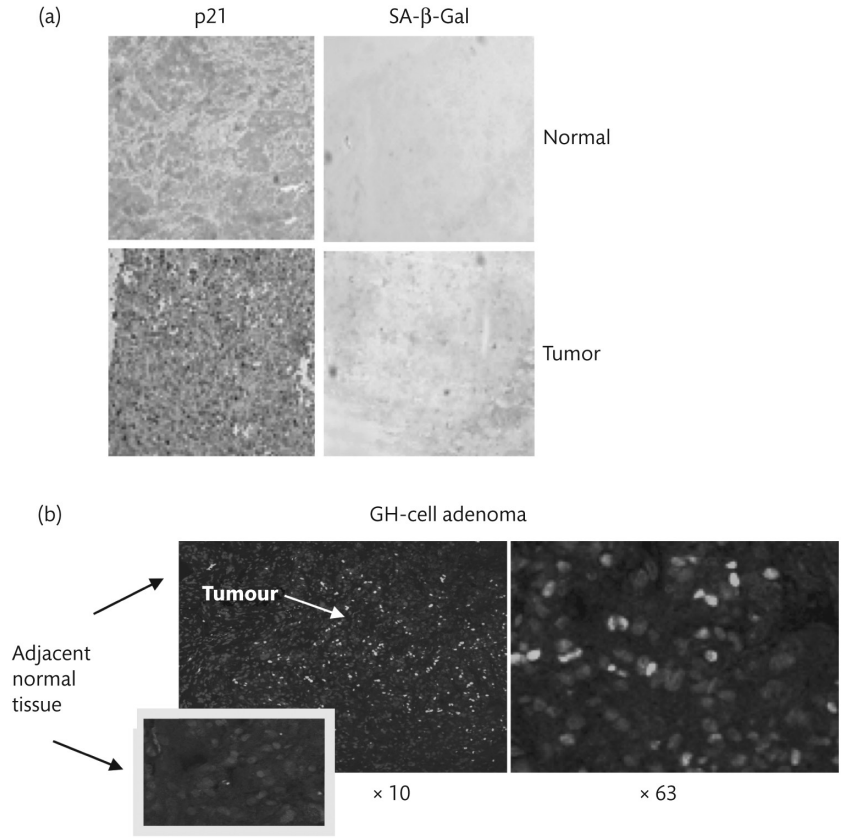


Figure 2.3.2.5 Senescence markers in human GH-producing pituitary adenomas. (a) Immunohistochemistry of the same GH-secreting human adenoma sections stained for p21 (brown) and SA-β-gal activity (blue). (b) Confocal image of double fluorescence immunohistochemistry of p21 (green) and β-galactosidase (red) proteins coexpression in human pituitary adenoma but not in normal adjacent tissue (left). High resolution (×63) image of the same slide (right). *For a colour version of this figure, please see colour plate section.*

increased SA-β-gal activity, upregulation of p21, pRB, and p19, and apoptosis blockage, which result in the pituitary hypoplasia phenotype. Telomere shortening was not noted in *pttg*-null pituitaries, indicating that activation of senescence pathway is not due to early cell ageing. Upregulation of p21 is associated with relative protection from pituitary formation in $Rb^{+/-}$ mice with *pttg* deletion ($Rb^{+/-}Pttg^{-/-}$ mice). In both *PTTG* deletion and overexpression, activation of p21/p53 senescence pathway occurs on a background of aneuploidy and DNA damage. Furthermore, autocrine IL-6 action was shown to be a determinant of pituitary senescence, with clonal tumorigenicity occurring in cells escaping senescence [42]. Taken together, these findings suggest that p21 may exert a tissue-specific tumour-suppressing function, unmasked under conditions where other pituitary genetic alterations or stresses are present. Activation of the pituitary senescence pathway may constrain pituitary tumour growth, and provides an explanation for the invariably benign nature of these tumours.

REFERENCES

1. Melmed S. Mechanisms for pituitary tumorigenesis: the plastic pituitary. *J Clin Invest*, 2003; **112**: 1603–18.
2. Melmed S. Pituitary medicine from discovery to patient-focused outcomes. *J Clin Endocrinol Metab*, 2016; **101**: 769–77.
3. Thorner MO, Perryman RL, Cronin MJ, *et al*. Somatotroph hyperplasia. Successful treatment of acromegaly by removal of a pancreatic islet tumor secreting a growth hormone-releasing factor. *J Clin Invest*, 1982; **70**: 965–77.
4. Alexander JM, Biller BM, Bikkal H, Zervas NT, Arnold A, Klibanski A. Clinically nonfunctioning pituitary tumors are monoclonal in origin. *J Clin Invest*, 1990; **86**: 336–40.
5. Lazzerini Denchi E, Attwooll C, Pasini D, Helin K. Deregulated E2F activity induces hyperplasia and senescence-like features in the mouse pituitary gland. *Mol Cell Biol*, 2005; **25**: 2660–72.
6. Simpson DJ, Hibberts NA, McNicol AM, Clayton RN, Farrell WE. Loss of pRb expression in pituitary adenomas is associated with methylation of the RB1 CpG island. *Cancer Res*, 2000; **60**: 1211–16.
7. Kiyokawa H, Kineman RD, Manova-Todorova KO, *et al*. Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). *Cell*, 1996; **85**: 721–32.
8. Hibberts NA, Simpson DJ, Bicknell JE, *et al*. Analysis of cyclin D1 (CCND1) allelic imbalance and overexpression in sporadic human pituitary tumors. *Clin Cancer Res*, 1999; **5**: 2133–9.
9. Valimaki N, Demir H, Pitkanen E, *et al*. Whole-genome sequencing of growth hormone (GH)—secreting pituitary adenomas. *J Clin Endocrinol Metab*, 2015; **100**: 3918–27.
10. Lecoq AL, Kamenicky P, Guiochon-Mantel A, Chanson P. Genetic mutations in sporadic pituitary adenomas—what to screen for? *Nat Rev Endocrinol*, 2015; **11**: 43–54.
11. Hage M, Viengchareun S, Brunet E, *et al*. Genomic alterations and complex subclonal architecture in sporadic GH-secreting pituitary adenomas. *J Clin Endocrinol Metab*, 2018; **103**: 1929–39.
12. Bi WL, Greenwald NF, Ramkissoon SH, *et al*. Clinical identification of oncogenic drivers and copy-number alterations in pituitary tumors. *Endocrinology*, 2017; **158**: 2284–91.
13. Pei L, Melmed S. Isolation and characterization of a pituitary tumor-transforming gene (PTTG). *Mol Endocrinol*, 1997; **11**: 433–41.
14. Romero F, Multon MC, Ramos-Morales F, *et al*. Human securin, hPTTG, is associated with Ku heterodimer, the regulatory subunit of the DNA-dependent protein kinase. *Nucleic Acids Res*, 2001; **29**: 1300–7.
15. Kim DS, Franklyn JA, Smith VE, *et al*. Securin induces genetic instability in colorectal cancer by inhibiting double-stranded DNA repair activity. *Carcinogenesis*, 2007; **28**: 749–59.
16. Vlotides G, Eigler T, Melmed S. Pituitary tumor-transforming gene: physiology and implications for tumorigenesis. *Endocr Rev*, 2007; **28**: 165–86.
17. Donangelo I, Gutman S, Horvath E, *et al*. Pituitary tumor transforming gene overexpression facilitates pituitary tumor development. *Endocrinology*, 2006; **147**: 4781–91.
18. Fuertes M, Sapochnik M, Tedesco L, *et al*. Protein stabilization by RSUME accounts for PTTG pituitary tumor abundance and oncogenicity. *Endocr Relat Cancer*, 2018; **25**: 665–76.
19. Xiao JQ, Liu XH, Hou B, *et al*. Correlations of pituitary tumor transforming gene expression with human pituitary adenomas: a meta-analysis. *PLoS One*, 2014; **9**: e90396.
20. Gunther T, Tulipano G, Dournaud P, *et al*. International Union of Basic and Clinical Pharmacology. CV. Somatostatin receptors: structure, function, ligands, and new nomenclature. *Pharmacol Rev*, 2018; **70**: 763–835.
21. Nasr C, Mason A, Mayberg M, Staugaitis SM, Asa SL. Acromegaly and somatotroph hyperplasia with adenomatous transformation due to pituitary metastasis of a growth hormone-releasing hormone-secreting pulmonary endocrine carcinoma. *J Clin Endocrinol Metab*, 2006; **91**: 4776–80.
22. Lopes MB, Gaylinn BD, Thorner MO, Stoler MH. Growth hormone-releasing hormone receptor mRNA in acromegalic pituitary tumors. *Am J Pathol*, 1997; **150**: 1885–91.
23. Trivellin G, Daly AF, Faucz FR, *et al*. Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. *N Engl J Med*, 2014; **371**: 2363–74.
24. Zhou C, Jiao Y, Wang R, Ren SG, Wawrowsky K, Melmed S. STAT3 upregulation in pituitary somatotroph adenomas induces growth hormone hypersecretion. *J Clin Invest*, 2015; **125**: 1692–702.
25. Turner HE, Harris AL, Melmed S, Wass JA. Angiogenesis in endocrine tumors. *Endocr Rev*, 2003; **24**: 600–32.
26. Shimon I, Huttner A, Said J, Spirina OM, Melmed S. Heparin-binding secretory transforming gene (hst) facilitates rat lactotrope cell tumorigenesis and induces prolactin gene transcription. *J Clin Invest*, 1996; **97**: 187–95.
27. Vlotides G, Cooper O, Chen YH, Ren SG, Greenman Y, Melmed S. Heregulin regulates prolactinoma gene expression. *Cancer Res*, 2009; **69**: 4209–16.
28. Vlotides G, Siegel E, Donangelo I, Gutman S, Ren S, Melmed S. Rat Prolactinoma cell growth regulation by epidermal growth factor receptor ligands. *Cancer Res*, 2008; **68**: 6377–86.
29. Fukuoka H, Cooper O, Ben-Shlomo A, *et al*. EGFR as a therapeutic target for human, canine, and mouse ACTH-secreting pituitary adenomas. *J Clin Invest*, 2011; **121**: 4712–21.
30. Reincke M, Sbiera S, Hayakawa A, *et al*. Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet*, 2015; **47**: 31–8.
31. Agarwal SK, Kennedy PA, Scacheri PC, *et al*. Menin molecular interactions: insights into normal functions and tumorigenesis. *Horm Metab Res*, 2005; **37**: 369–74.
32. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, *et al*. Germline mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *Proc Natl Acad Sci U S A*, 2006; **103**: 15558–63.
33. Vierimaa O, Georgitsi M, Lehtonen R, *et al*. Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science*, 2006; **312**: 1228–30.

34. Daly AF, Vanbellinghen JF, Khoo SK, *et al.* Aryl hydrocarbon receptor-interacting protein gene mutations in familial isolated pituitary adenomas: analysis in 73 families. *J Clin Endocrinol Metab*, 2007; **92**: 1891–6.
35. Schofl C, Honegger J, Droste M, *et al.* Frequency of AIP gene mutations in young patients with acromegaly: a registry-based study. *J Clin Endocrinol Metab*, 2014; **99**: E2789–93.
36. Rostomyan L, Beckers A. Screening for genetic causes of growth hormone hypersecretion. *Growth Horm IGF Res*, 2016; 30–1: 52–7.
37. Caimari F, Korbonits M. Novel genetic causes of pituitary adenomas. *Clin Cancer Res*, 2016; **22**: 5030–42.
38. Zhang X, Zhou Y, Mehta KR, *et al.* A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *J Clin Endocrinol Metab*, 2003; **88**: 5119–26.
39. Qian ZR, Asa SL, Siomi H, *et al.* Overexpression of HMGA2 relates to reduction of the let-7 and its relationship to clinicopathological features in pituitary adenomas. *Mod Pathol*, 2009; **22**: 431–41.
40. De Martino I, Visone R, Wierinckx A, *et al.* HMGA proteins up-regulate CCNB2 gene in mouse and human pituitary adenomas. *Cancer Res*, 2009; **69**: 1844–50.
41. Kuilman T, Michaloglou C, Vredeveld LC, *et al.* Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell*, 2008; **133**: 1019–31.
42. Sapochnik M, Haedo MR, Fuertes M, *et al.* Autocrine IL-6 mediates pituitary tumor senescence. *Oncotarget*, 2017; **8**: 4690–702.

2.3.3 Histopathology of Pituitary Tumours

Luis V. Syro, Fabio Rotondo, and Kalman Kovacs

Introduction

The human pituitary gland consists of two major components: the adenohypophysis and the neurohypophysis. The adenohypophysis comprises the pars anterior, pars intermedia, and pars tuberalis [1]. The pars anterior, also known as pars glandularis or pars distalis, is the largest portion of the gland. In contrast to most mammalian species, the human gland has no anatomically distinct pars intermedia. The exclusively pro-opiomelanocortin (POMC)-producing cells of the pars intermedia are sandwiched between the anterior and posterior lobes in the majority of mammals, whereas in humans they are incorporated within the pars anterior. The pars tuberalis is a minor upward extension of the adenohypophysis attached to the exterior of the lower pituitary stalk. Histologically, the adenohypophysis consists of a central median (or mucoid) wedge flanked by two lateral wings. The hormone-producing cell types are distributed in an uneven, but specific manner. The cells are arranged within evenly sized acini surrounded by a delicate but well-defined reticulin fibre network giving the pituitary its distinct architecture. In the centre of the acini is the pituitary follicle composed of the agranular nonendocrine folliculo-stellate cells. The neurohypophysis is composed of nerve fibres from the hypothalamic nuclei that give rise to the infundibulum, the pituitary stalk, and the posterior lobe which stores and releases the hypothalamic hormones oxytocin and vasopressin.

Cell Types

The mature endocrine cell types of the anterior pituitary gland emerge from a common primordium. Cytodifferentiation occurs in response to opposing signalling gradients that emanate from distinct organizing centres [2]. To progress beyond the initial patterning, each cell type requires additional specific transcription factors in determining its cytodifferentiation and hormone production [3] (see Table 2.3.3.1). Pituitary-specific transcription factor 1 (Pit-1) determines cells that can produce growth hormone (GH), prolactin (PRL), and thyroid-stimulating hormone (TSH), while T-box transcription factor TBX19 (Tpit) determines corticotrophs. Oestrogen receptor alpha (ER-α) cooperates with Pit-1 to enhance PRL secretion, therefore, lactotrophs coexpress both Pit-1 and ER-α. Guanine-Adenine-Thymine-Adenine binding protein 2 (GATA-2) appears to be an essential contributor to thyrotroph development and is coexpressed with Pit-1 in thyrotrophs while expression of steroidogenic factor 1 (SF1), ER-α and GATA-2 categorize gonadotrophs [4]. Apart from their role in pituitary gland development, pituitary transcription factors can serve as diagnostic markers of cell lineage in pituitary tumours [5, 6] (see Table 2.3.3.2).

GH cells or **somatotrophs** make up approximately 50% of the gland, occupying mainly the lateral wings. They show strong acidophilia by histology and GH hormone immunoreactivity by

Table 2.3.3.1 Cell lineage differentiation and transcription factors

Cell lineage	Transcription factor	Adenohypophyseal cell
Acidophilic lineage	Pit-1	Somatotrophs
	Pit-1, ER-α	Lactotrophs
	Pit-1, GATA-2	Thyrotrophs
Corticotroph lineage	Tpit	Corticotrophs
Gonadotroph lineage	SF-1, ER-α, GATA-2	Gonadotrophs

Pit-1, pituitary-specific transcription factor 1; Tpit, T-box transcription factor TBX19; GATA-2, GATA binding protein 2; ER-α, oestrogen receptor alpha; SF-1, steroidogenic factor 1.

Table 2.3.3.2 Classification of pituitary adenomas

Tumour	Transcription factors	Hormones
Pit-1 Family of tumours		
Somatotroph adenomas		
Densely granulated	Pit-1	GH, α-subunit
Sparsely granulated		GH (weak)
Mammosomatotroph adenomas	Pit-1, ER-α	GH, PRL, α-subunit
Mixed GH-PRL adenomas	Pit-1, ER-α	GH, PRL, α-subunit
Plurihormonal GH-producing adenomas	Pit-1, ER-α	GH, PRL, α-subunit, ß-TSH
Lactotroph adenomas		
Densely granulated	Pit-1, ER-α	PRL
Sparsely granulated		PRL, α-subunit
Acidophil stem cell adenomas		PRL, GH
Thyrotroph adenomas	Pit-1, GATA-2	ß-TSH, α-subunit
Monomorphous plurihormonal adenomas		
Silent subtype 3	Pit-1	GH, PRL, β-TSH
Tpit Family of tumours		
Corticotroph adenomas		
Densely granulated	Tpit	ACTH
Sparsely granulated		ACTH
Crooke cell adenoma		ACTH
SF-1 Family of tumours		
Hormone-positive gonadotroph adenomas	SF-1, ER-α, GATA-2	ß-FSH, ß-LH, α-subunit
Hormone-negative gonadotroph adenomas	SF-1, ER-α, GATA-2	None
Polymorphous plurihormonal adenomas		
Plurihormonal adenomas	Multiple	Multiple
Transcription factor and hormone-negative adenoma		
Null cell adenomas	None	None

GH, growth hormone; ACTH, adrenocorticotrophic hormone; PRL, prolactin; FSH, follicle-stimulating hormone; LH, luteinizing hormone; TSH, thyroid-stimulating hormone; Pit-1, pituitary-specific transcription factor 1; Tpit, T-box transcription factor TBX19; GATA-2, GATA binding protein 2; ER-α, Oestrogen receptor alpha; SF-1, steroidogenic factor 1.

immunohistochemistry. Contingents of somatotrophs also express PRL or the α-subunit of the glycoprotein hormones. PRL cells or **lactotrophs** account for 10–30% of the pituitary gland cell population. The chromophobic or slightly acidophilic and PRL-immunoreactive lactotrophs are evenly scattered with focal accumulation in the posterolateral rim of lateral wings. The majority of adrenocorticotropic hormone (ACTH) cells or **corticotrophs** reside within the median wedge of the pituitary gland. They are basophilic, Periodic acid–Schiff (PAS) positive, and immunoreactive for ACTH as well as for other pro-opiomelanocortin (POMC)-peptides (β-endorphin, β-lipotropic hormone (LPH), corticotropin-like intermediate peptide (CLIP)). Approximately 12% of adenohypophyseal cells are immunopositive for POMC-peptides, which include a small but undetermined percentage of pars intermedia derived POMC cells as well. TSH cells or **thyrotrophs**, occupy mainly the anterior one-third of the median wedge and represent approximately 5% of adenohypophyseal cells. The slightly basophilic, angular thyrotrophs are strongly immunoreactive for β-TSH and α-subunit. **Gonadotrophs** producing follicle-stimulating hormone (FSH), and luteinizing hormone (LH) account for an estimated 15–20% of cells evenly distributed throughout the pars distalis. They are immunoreactive for β-FSH, β-LH, and their α-subunit. The hormonal function of the adenohypophyseal cell types displays considerable flexibility depending on the functional demand placed on them. Reversible transdifferentiation has been shown to occur between members of the Pit-1 group: GH cells to PRL cells during pregnancy, and GH cells to TSH cells in hypothyroidism [7]. Pituitary cell types differ not only in their function, structure, and hormone content but also in their morphological responses to functional stimulation or suppression. The **folliculo-stellate** cells, which occupy up to 10% of the anterior pituitary, have no hormonal function but both *in vitro* biochemical and morphological studies have documented that these cells are not only unexpectedly versatile but indispensable [8].

Pituitary Adenomas

Pituitary adenomas are common, accounting for approximately 15% of intracranial tumours [1]. They are monoclonal, produced by genetic and epigenetic alterations—leading to activation of proto-oncogenes or inactivation of tumour suppressor genes—that impart growth and survival advantage to a single cell. While the majority of the tumours are sporadic, some are associated with familial syndromes exhibiting different genetic background and variable phenotype; germline, somatic, and mosaic mutations have been identified [9, 10] (see Table 2.3.3.3).

Pituitary tumours may be referred to as microadenomas (less than 10 mm in diameter), or macroadenomas (greater than 10 mm in diameter). The growth pattern of these tumours may be expansive resulting in a slowly growing mass exerting increasing pressure on the surrounding healthy gland and the bony sella. In contrast, invasive adenomas spread into the surrounding normal gland, dura or other parasellar structures (sphenoid sinus, cavernous sinus) regardless of their size. Adenomas extending into the suprasellar space may compress the optic chiasm causing visual disturbances, a frequent clinical manifestation of macroadenomas.

When interpreting pituitary pathology, the clinical and radiological information must be considered as a whole [11]. At the time of surgery, gentle tissue handling is recommended with prompt and appropriate fixation of the specimen in formalin for pathological study. It is optional that a small fragment of tissue be fixed in glutaraldehyde, osmicated, and embedded for electron microscopy for cases that are difficult to diagnosis. It may also be worthwhile to obtain DNA from the patient's blood before surgery and freeze a portion of the tumour for subsequent molecular and DNA analysis [12].

A logical order of steps must be followed to study the morphology of pituitary adenomas [11, 13]. It is essential to first establish whether the tissue obtained is from a normal pituitary gland, a pituitary

Table 2.3.3.3 Clinicopathologic features of germline, somatic, and mosaic mutations in pituitary adenomas

Syndrome	Gene	Morphologic characteristics
Germline mutations		
FIPA	*AIP*	*AIP*-positive 20% Somatotroph and lactotroph, rarely other types.
	GPR101	Somatotroph adenoma or pituitary enlargement suggestive of hyperplasia
MEN1	*MEN1*	Lactotroph adenomas (60%) Non-functional adenomas (15%) Somatotroph and corticotroph less common
MEN4	*MEN4*	Somatotroph adenoma most common
Carney complex	*PRKAR1A* *PRKACA* *PRKACB*	Somatotroph adenoma or pituitary hyperplasia
Succinate dehydrogenase complex genes syndrome 3P association	*SDHA/B/C/D* *SDHAF2*	Lactotroph adenomas most common Somatotroph and non-functional adenomas Vacuolar changes in most cells
Neurofibromatosis type 1	*NF1*	Not clear yet
DICER 1 syndrome	*DICER 1*	Pituitary blastoma Neonatal or early childhood age Cushing's disease
Somatic mutations		
Somatotroph adenomas	*GNAS*	No differences between subgroups
Corticotroph adenomas	*USP8*	Small tumours with high ACTH production
Oncocytic pituitary adenomas	Complex I genes	Less aggressive tumour phenotype
All types of pituitary adenomas	*PIK3CA*	20% to 40% of various types of pituitary adenomas
Mosaic mutations		
McCune-Albright	*GNAS* (Mosaic postzygotic mutation)	Somatotroph adenomas Somatotroph and lactotroph hyperplasia
X-LAG	*GPR101* (Germline or somatic mosaicism)	Somatotroph adenoma or pituitary enlargement suggestive of hyperplasia

FIPA, familial isolated pituitary adenoma; MEN1, multiple endocrine neoplasia type 1; MEN4, multiple endocrine neoplasia type 4; 3P, paraganglioma, phaeochromocytoma and pituitary adenoma association; ACTH, adrenocorticotropic hormone; X-LAG, X-linked acrogigantism syndrome.

tumour, or a different sellar lesion. Next, immunohistochemical classification based on hormone production by the cells is accomplished and, finally, prognostic information and treatment options can be determined by analysing various biomarkers and by molecular/genetic/epigenetic investigation (See Table 2.3.3.4).

Table 2.3.3.4 Panel approach for the evaluation and diagnosis of pituitary adenomas

Stain	Stain pattern	Indication
Haematoxylin-eosin (HE)	Acidophilic, basophilic, or chromophobic	Ancillary stain which has stood the test of time For initial classification of the tumour
Periodic acid–Schiff (PAS)	Cytoplasmic	To recognize corticotrophs and Crooke cells
Reticulin, Silver stain, Collagen IV	Acinar reticular pattern	To differentiate normal pituitary gland, hyperplasia, or tumour
Chromogranin, Synaptophysin	Secretory granules	Neuroendocrine markers
TTF-1	Nuclear	Positive in posterior pituitary gland and posterior pituitary tumours
GH, PRL, ACTH, TSH, FSH, LH, α-subunit	Cytoplasmic	To identify hormone secretion
LMWK Cam 5.2	Cytoplasmic Dot patterns Ring-like pattern	Positive in somatotrophs and corticotrophs To categorize pure somatotroph adenomas as densely or sparsely granulated To identify Crooke cells
Pit-1, Tpit, SF-1, ER-α, GATA-2	Nuclear	To identify cell lineage differentiation
E-cadherin	Membrane	Absent E-cadherin expression may be correlated to tumour invasiveness
Ki67, p53	Nuclear	Proliferation markers
SSTR2, SSTR5	Membrane	Predictive biomarkers for long-acting somatostatin analogues?
AIP	Cytoplasmic	Predictive biomarker for long-acting for somatostatin analogues?
MGMT	Nuclear	To evaluate DNA repair pathway, DR Predictive biomarker for temozolomide treatment in aggressive pituitary adenomas and carcinomas
MSH2, MSH6	Nuclear	To evaluate integrity of DNA repair pathway, MMR
MPG	Nuclear	To evaluate integrity of DNA repair pathway, BER
VEGF	Cytoplasmic	To assess angiogenic growth and progression of tumour

TTF-1, thyroid transcription factor 1; GH, growth hormone; PRL, prolactin; TSH, thyroid-stimulating hormone; FSH, follicle-stimulating hormone; LH, luteinizing hormone; ACTH, adrenocorticotropic hormone; LMWK, low molecular weight keratin; SSTR2, somatostatin receptor 2; SSTR5, somatostatin receptor 5; AIP, aryl-hydrocarbon receptor-interacting protein; MGMT, O6-methylguanine DNA methyltransferase; MSH2, MutS protein homolog 2; MSH6, mutS homolog 6; DNA, deoxyribonucleic acid; DR, direct repair; MMR, mismatch repair; MPG, N-methylpurine DNA glycosylase; BER, base excision repair; VEGF, vascular endothelial growth factor.

Modified with permission from Syro, L.V., Rotondo, F., Serna, C.A. et al. Pathology of GH-producing pituitary adenomas and GH cell hyperplasia of the pituitary. *Pituitary* (2017) 20: 84. Copyright © 2016, Springer Science Business Media New York.

By histology, adenomas are acidophilic, basophilic, or chromophobic. The tinctorial properties of tumours are largely unrelated to their hormonal function. After initial evaluation using the haematoxylin & eosin (H&E) stain, which allows distinction from other pathologies of the sellar region, the next step is to demonstrate the absence of reticulin fibres. This differentiates normal adenohypophysis from hyperplasia or adenoma. The normal acinar reticular pattern is enlarged but intact in cases of hyperplasia and, regardless of their size, a total dissolution of the acinar architecture is evident in every adenoma. PAS stain is used to recognize corticotrophs because of the carbohydrate moiety present in POMC which is the precursor of ACTH. Chromogranin and synaptophysin, broad-spectrum endocrine markers, are used to confirm that the lesion is of endocrine origin. Cam5.2, a low molecular weight keratin, is used to identify and confirm somatotroph and corticotroph adenomas. They are immunopositive, and the keratin pattern will help to clarify their subtype: diffuse perinuclear, paranuclear globular (dot pattern), mixed or ring-like. These characteristic patterns are specific to certain adenoma types.

The current World Health Organization (WHO) classification (2017) of pituitary adenomas recommends the use of pituitary adenohypophyseal cell lineages for their classification rather than according to their hormone production [5, 6, 14] (see Table 2.3.3.2). Accordingly, pituitary adenoma classification uses a wide range of monoclonal antibodies against pituitary hormones, and pituitary transcription factors, to differentiate the various tumour cell types. Electron microscopy, although useful, is currently only utilized in cases that are difficult to diagnose. The available immunohistochemical stains used for studying pituitary adenomas and their utility are presented in Table 2.3.3.4.

Pit-1 Family of Tumours

Somatotroph Adenomas

Somatotroph adenomas arise from Pit-1 pituitary cell lineage and express GH in excess. Approximately 15% of surgically removed pituitary adenomas represent pure GH cell adenomas [15]. Most of these tumours are associated with clinical signs of acromegaly, but silent forms have also been reported [16]. According to the density of GH-containing secretory granules in the cytoplasm of the cells, somatotroph adenomas can be densely or sparsely granulated. The clear distinction between them is crucial for proper prognosis and treatment [17].

Densely granulated somatotroph adenomas occur with the same frequency in both sexes peaking at the same time (in the sixth decade). The tumours display slow, expansive growth resulting in the typical 'ballooning of the sella' and may remain intrasellar for several years. Histologically, the densely granulated form is strongly acidophilic displaying diffuse or, less frequently, trabecular pattern. Extensive immunoreactivity for GH is usually accompanied by similarly strong positivity for α-subunit. Scattered immunopositivity for PRL and β-TSH is usually not associated with oversecretion of these hormones. Immunostaining for Cam5.2 demonstrates a diffuse perinuclear pattern.

Sparsely granulated somatotroph adenomas are diagnosed earlier, peaking in the fourth decade. The tumours tend to be macroadenomas at the time of diagnosis and are often invasive. Histology identifies chromophobic adenomas invariably displaying a diffuse pattern; GH immunoreactivity is often sparse. Nuclear pleomorphism may be evident, and adenoma cells harbour a homogeneous, spherical, juxtanuclear, practically unstained structure. This fibrous body is strongly immunopositive for Cam5.2 in a characteristic *dot* pattern and is the best histological marker for this tumour type since it is present in more than 70% of the cells. As opposed to the densely granulated type, multiple immunoreactivities for pituitary hormones are rarely noted in sparsely granulated somatotroph adenomas.

Mixed somatotroph and lactotroph adenomas, comprised of two distinct cell types, account for approximately 5% of surgically removed adenomas. The tumours are associated with acromegaly and varying degrees of hyperprolactinaemia. They tend to be aggressive and are difficult to treat. Mixed adenomas usually consist of densely granulated GH cells and sparsely granulated PRL cells. Accordingly, they are composed of acidophilic and chromophobic cells, displaying immunoreactivity for GH and PRL; α-subunit positivity is also frequent in this tumour type. Other combinations may occur, but they are rare.

The infrequent (2%) **mammosomatotroph cell adenomas** are monomorphous, consisting of one cell type displaying markers of both GH and PRL differentiation. Clinically they are associated with acromegaly and mild hyperprolactinaemia. These acidophilic tumours are immunoreactive for GH and α-subunit and, to a lesser extent, for PRL. At the ultrastructural level, the densely granulated tumour cells possess unusually large (up to 1000 nm and over) secretory granules and display lateral granule extrusions which is a PRL cell marker. The slow-growing mammosomatotroph adenomas show biological behaviour similar to densely granulated somatotroph adenomas.

Morphologic effects of treatment with long-acting somatostatin analogues are neither severe nor consistent. Remarkable shrinkage of cells and marked fibrosis are infrequent. Most common findings include the increase in size and number of secretory granules and increased lysosomal activity while significant fibrosis is less frequent. The close correlation observed between clinical response and tumour morphology in cases of prolactinomas treated with dopamine agonists does not exist in examples of somatostatin analogues therapy.

Both types of GH cell adenoma may engage in focal production of endocrine amyloid. Approximately 2–3% of morphologically typical sparsely granulated GH cell adenomas are clinically silent, the reason for which is unknown. Few cases of the sparsely granulated tumours contain variable amounts of nervous tissue (neuron-like cells and neuropil) as a likely result of neuronal differentiation within the adenoma. This variant is known as neuronal metaplasia.

Lactotroph Adenomas

Lactotroph adenomas arise from the Pit-1 pituitary cell lineage and express mainly PRL. They are classified into three histological subtypes: sparsely granulated lactotroph adenomas, densely granulated lactotroph adenomas, and acidophil stem cell adenomas.

Sparsely granulated lactotroph adenomas are the most common adenoma type accounting for 27–30% of pituitary tumours. They are associated with hyperprolactinaemia, primary or secondary amenorrhoea, galactorrhoea, and infertility in women. The less specific symptoms in men—decreased libido and impotence—usually mean a delay in diagnosis and development of macroadenomas

whereas the majority of tumours in women are small and classified as microadenoma. In both sexes, lactotroph adenomas are most frequent in the third and fourth decade of life. However, the incidence of this tumour type is significantly higher in women in their childbearing years. Lactotroph adenomas are associated with wide-ranging biological behaviours from indolent to highly aggressive, and in men usually present invasive radiological features. The PRL blood levels are roughly proportional to the tumour mass. Histologically, most lactotroph adenomas are chromophobic with a diffuse pattern. Immunostaining demonstrates strong PRL immunopositivity tracing the sacculi of the prominent Golgi apparatus (Golgi pattern). Ultrastructurally, the three salient features are masses of rough endoplasmic reticulum, prominent Golgi apparatus, and lateral extrusion of the small, sparse secretory granules—misplaced exocytosis. The latter is a specific marker of PRL differentiation.

Densely granulated lactotroph adenomas display strong and diffuse PRL immunoexpression with coexpression of Pit-1 and ER-α. They are rare and have been considered aggressive tumours.

Acidophil stem cell adenomas are rare (2%), monomorphous type tumours with morphological signs of PRL and GH differentiation. The tumours are associated chiefly with hyperprolactinaemia, but the serum PRL levels may be disproportionately low for its size. Physical stigmata of acromegaly and significant elevation of GH levels are infrequent. These tumours grow aggressively in young subjects with a tendency to invade intrasellar areas. The histological analysis demonstrates chromophobic adenomas with moderate to strong immunoreactivity for PRL. Immunopositivity for GH is weak or negative, but immunostaining for cytokeratin reveals the dot-like positivity of fibrous bodies. The striking ultrastructure is characterized by oncocytic change with the formation and increased number of giant mitochondria, sparse and small secretory granules with lateral extrusion (PRL marker), and fibrous bodies (GH marker).

Although calcification is extremely rare in other adenoma types, an estimated 10–15% of lactotroph adenomas display varying degrees of calcification. The alteration may be extreme ('pituitary stone'). Deposition of endocrine amyloid occurs infrequently.

Adequate therapeutic response with dopamine agonists is associated with striking morphological changes in adenoma cells: the nucleus becomes heterochromatic, the cytoplasm shows marked shrinkage due to loss and involution of the hormone-producing apparatus of the cell (rough endoplasmic reticulum, Golgi complex), and PRL immunoreactivity is reduced or lost. Theoretically, the morphological changes are reversible; however, portions of the neoplasm may permanently lose their responsiveness and retain their suppressed features even when treatment is discontinued. Long-term administration of dopamine agonists may also cause varying degrees of fibrosis and calcification—which presents as psammoma bodies.

Thyrotroph Adenomas

Thyrotroph adenomas express mainly TSH and also arise from Pit-1 pituitary cell lineage. They represent 1% of surgically removed adenomas and are associated either with hyperthyroidism or inappropriately elevated levels of TSH. Inexplicably, some tumours bearing immunohistochemical characteristics of thyrotroph adenomas occur in euthyroid subjects. At the time of diagnosis, these tumours are often macroadenomas with a tendency to invade. The morphology of the small group of thyrotroph adenomas exhibits surprising diversity. Histologically, the adenomas are chromophobic and negative, or mildly positive with PAS. They may be highly differentiated comprising elongated polar cells forming pseudorosettes around vessels. Alternatively, the pattern may be diffuse in some cases with considerable nuclear pleomorphism. However, another variant is markedly fibrotic. Minute calcifications may be evident as well. Immunoreactivity for TSH is variable; it is often patchy or scattered, rarely extensive. Scattered cells may exhibit immunoreactivity for GH, PRL, and α-subunit. Both Pit-1 and GATA2 are expressed. Thyrotroph adenomas possess somatostatin receptors and, in most cases, they are strongly positive for somatostatin receptor 2 (SSTR2) immunostaining.

Monomorphous Plurihormonal Adenomas

Plurihormonal adenomas produce more than one hormone. They can be monomorphous or plurimorphous, consisting of two or more different cell lineages. They include **plurihormonal Pit-1 positive adenomas** previously called silent subtype 3 adenomas. They have a frequency of approximately less than 1% and are clinically significant owing to their aggressive behaviour. This tumour type is equally frequent in both sexes, but it has a strikingly different age-related distribution. In men, they may occur at any age from the second to the seventh decade while in women it occurs between the second and fourth decade, peaking in the late twenties, but rarely occurring after 40 years of age. Plurihormonal Pit-1 positive adenomas can mimic lactotroph adenomas in women and are associated with low-grade hyperprolactinaemia. Some cases may present with acromegaly. Histologically, they are often acidophilic and may show mild PAS positivity. The large adenoma cells form a diffuse, or lobular pattern. Immunohistochemistry may demonstrate scattered, minor positivity for various adenohypophyseal hormones, but the majority of tumour cells are immunonegative for all known pituitary hormones, showing extensive nuclear expression of Pit-1.

Molecular Pathology of the Pit-1 Family of Tumours

Approximately 40–60% of sporadic somatotroph adenomas present somatic mutations in the gene encoding the α subunit of the stimulatory G-protein (*GNAS)*. These mutations are commonly found in densely granulated somatotroph adenomas. The somatic activating mutations, known as *gsp mutations*, prevent hydrolysis of guanosine-5'-triphosphate (GTP), leading to a constitutive activation of the cAMP pathway, which in somatotrophs increases GH secretion and cell proliferation [9, 18, 19]. A hallmark of somatotroph adenomas is the genetic and epigenetic alterations in the *GNAS* gene with the ensuing dysregulation of the cAMP pathway. Somatotroph adenomas with *gsp mutations* tend to have higher serum GH and IGF-I, but there are no differences between sparsely and densely granulated subtypes [20]. Epigenetic alterations that are common to both somatotroph and lactotroph adenomas include *CASP8, CDKN2A, DAPK1, MGMT, RB1*, and *TP73* [12].

AIP-mutation positive patients present with somatotroph and lactotroph adenomas in 85% of the cases [21]. In MEN 1, lactotroph adenomas are the most frequent, followed by clinically non-functioning adenomas, somatotroph and corticotroph adenomas [10]. In thyrotroph adenomas, inactivating mutations of *MEN1* gene have been reported.

Tpit Family of Tumours

Corticotroph Adenomas

Corticotroph adenomas arise from Tpit pituitary cell lineage and express ACTH and other POMC-derived peptides. They are classified into three subtypes: densely granulated corticotroph adenomas, sparsely granulated corticotroph adenomas, and Crooke cell adenomas. Corticotroph adenomas may also be clinically silent.

Densely granulated corticotroph adenomas account for 10–12% of surgically removed adenomas and are responsible for Cushing's disease. These tumours exhibit a marked female preponderance. The age-related occurrence of corticotroph adenomas is similar in both sexes peaking in the fourth decade. Most corticotroph lesions are small microadenomas causing florid Cushing's disease [22]. The tumours, often measuring only a few millimetres in diameter, may be too small to be conclusively detected by imaging or to be identified at surgery. Therefore, serial sectioning of the biopsied tissue fragments is often needed, which may not always result in the demonstration of the tumour. Histologically, corticotroph adenomas are basophilic and PAS-positive with a sinusoidal or diffuse pattern. Immunoreactivity can be demonstrated not only for ACTH, but for β-endorphin, β-LPH, and CLIP as well. Perinuclear bundles of cytokeratin filaments, characteristic for the human corticotrophs are demonstrated with Cam5.2 as cytoplasmic, perinuclear pattern.

In a minority of cases, pituitary-dependent Cushing's disease is brought about by larger tumours. These neoplasms are often associated with a milder form of hypercortisolism, but the tumours grow aggressively, often invade, and are frequently macroadenomas at the time of diagnosis. Histologically they exhibit variable, often weak PAS positivity and immunoreactivity for ACTH. A few cases of aggressive macroadenomas also display immunoreactivity for LH and α-subunit. Morphological features of corticotroph adenomas in cases of Nelson's syndrome are similar to those of densely granulated corticotroph adenomas in Cushing's disease with few or no cytokeratin filaments.

Sparsely granulated corticotroph adenomas are composed of faintly basophilic or chromophobic cells. They are PAS-positive with patchy immunopositivity for ACTH.

Crooke hyalinization (i.e. excessive accumulation of cytokeratin filaments) is the ubiquitous response of the normal human ACTH cells to the long-lasting elevation of circulating glucocorticoid levels. Accumulation of these cytokeratin filaments causes a glassy hyaline appearance to the cytoplasm and displacement of PAS and ACTH-positive granules to the cell periphery. These 'Crooke cells' can be identified using Cam5.2 immunostaining, which displays a strong ring-like pattern around the nucleus. In cases of Cushing's disease with persistent hypercortisolism after surgery, no evidence of a corticotroph adenoma at pathology, and only pituitary gland obtained, analysis for the presence of Crooke cells is pertinent. Presence of Crooke cells confirms the previous hypercortisolism and will dictate the next step of treatment [22]. Accordingly, Crooke hyalinization would be noted in (1) non-tumorous corticotrophs adjacent to corticotroph adenomas; (2) in ectopic ACTH/corticotrophin-releasing hormone (CRH) syndrome; (3) in patients with glucocorticoid secreting adrenocortical tumours; (4) and subjects taking glucocorticoids (factitious Cushing syndrome). In cases of pseudo-Cushing, Crooke cells will not be present in the pituitary tissue.

Crooke cell adenomas are a rare type of pituitary tumours. They produce ACTH causing Cushing's disease or may be endocrinologically silent. Crooke hyalinization is not expected to develop in corticotroph tumours yet, a minority of such adenomas contain a variable percentage of adenoma cells displaying this alteration, identical to the Crooke cells seen in the adenohypophysis of patients with glucocorticoid excess. If this change affects more than 50% of the cells, tumours are diagnosed as Crooke cell adenomas [23]. The reason for which Crooke cell adenomas produce ACTH while simultaneously displaying Crooke hyaline changes in response to increased glucocorticoid excess is not well understood. These tumours are usually invasive, may exhibit aggressive clinical behaviour, and often recur with a low success of cure after reoperation and radiotherapy. Due to their rarity, they present considerable difficulties in assessing prognosis, treatment, and clinical management [23].

Silent corticotroph adenomas are unassociated with functional activity and present as clinically non-functional tumours. There are two subtypes of silent corticotroph adenomas. **Silent corticotroph adenomas subtype 1**—densely granulated—show less female preponderance and different age-related occurrence than corticotroph adenomas associated with Cushing's disease. The tumours display a high propensity for haemorrhage and may present with pituitary apoplexy. Morphologically the adenomas have the same basophilia, PAS positivity, ACTH and β-endorphin immunoreactivity and ultrastructural features as corticotroph tumours associated with Cushing's disease. **Silent corticotroph adenomas subtype 2**—sparsely granulated—have a frequency of 1.5–2.0% and show marked male preponderance. The tumours appear as non-functioning masses and are usually diagnosed at the macroadenoma stage. Histology reveals chromophobic tumours comprised of small cells, which exhibit only modest PAS positivity and scattered immunoreactivity for ACTH and β-endorphins. No cytokeratin filaments are present with Cam5.2. They can present with aggressive clinical behaviour [5]. These two adenoma types are possibly derived from cells of the pars intermedia.

Molecular Pathology of the Tpit Family of Tumours

In approximately 40–60% of corticotroph adenomas, somatic mutations in Ubiquitin Specific Peptidase 8 gene (*USP8*)—resulting in increased EGFR expression and increased mRNA levels of POMC—have been demonstrated [24, 25]. Corticotroph adenomas with mutated *USP8* are commonly found in females, present as small tumours with high ACTH production, and are associated with better prognosis [24]. *USP8* mutations have not been confirmed in silent corticotroph tumours. Epigenetic mutations in *TP53* are linked to corticotroph adenomas.

MicroRNA (miRNA) studies revealed a large number of miRNAs that are upregulated and downregulated. Among those, it was shown that miR-26a plays a crucial role in cell cycle control of corticotroph adenomas [12].

Pituitary blastoma, in *DICER1* syndrome, is a rare developmental early childhood invasive neoplasm arising within the fetal adenohypophysis causing infantile-onset Cushing disease. It consists of a combination of Rathke-type epithelial rosettes, small

primitive appearing cells, and secretory cells expressing ACTH. Genetic analysis has revealed a germline *DICER1* mutation which is often accompanied by a somatic mutation affecting the RNase IIIb domain of *DICER1* in the cases studied [26].

SF-1 Family of Tumours

Gonadotroph Adenomas

Gonadotroph adenomas arise from SF-1 pituitary cell lineage and produce β-FSH, β-LH, and α-subunit. They can also express ER-α and GATA2. Their incidence in the surgical material is about 10%; they occur with similar frequency in both sexes. Most gonadotroph adenomas appear as slow-growing, expansive macroadenomas causing local symptoms. A discrepancy between clinical parameters and morphological signs of gonadotroph differentiation is common; tumours displaying FSH/LH immunoreactivity and signs of a high degree of functional differentiation, may be unassociated with elevated serum FSH or LH levels. Some clinically non-functioning pituitary adenomas are gonadotroph adenomas due to their lack of any endocrine activity. The morphology of gonadotroph adenomas is variable. Histology may reveal either polar cells forming pseudorosettes around vessels or a diffuse pattern. Oncocytic change, an excessive increase of number and volume density of mitochondria, is frequent. Immunoreactivity for FSH or LH is variable, often patchy; α-subunit, which is a useful clinical indicator, is not a reliable morphological marker.

Molecular Pathology of the SF-1 Family of Tumours

Molecular studies involving tumours from the SF-1 lineage found that they have a lower number of germline and somatic mutations compared to other pituitary adenomas. Silencing of *MEG3* located on chromosome 14q32 which has an antiproliferative role in adenomas was also found to have an important role in the development of gonadotroph adenomas. To date, in MEN-1, no gonadotroph adenomas have been reported. Both non-functioning adenomas and gonadotroph adenomas have shown to have loss of p16 protein (tumour suppressor encoded by *CDKN2A* gene) and hypermethylation of *CDKN2A*.

Polymorphous Plurihormonal Adenomas

Plurihormonal Adenomas

Polymorphous plurihormonal adenomas are rare tumours, often with unique ultrastructure. The most common combinations are GH-TSH-PRL or PRL-TSH, but other combinations of different cellular lineages can also occur.

Transcription Factor and Hormone-Negative Adenomas

Null Cell Adenomas

Null cell adenomas are composed of adenohypophyseal cells that do not display evidence of cell lineage differentiation by pituitary transcription factors or pituitary hormones. These hormonally inactive adenomas account for approximately 25% of surgically removed tumours. They are twice as common in males, and peak in the sixth decade in both sexes. Most of the tumours are slowly growing, expansive macroadenomas causing local symptoms and varying degrees of hypopituitarism. Low-grade hyperprolactinaemia may occur ('stalk section effect'). Histologically they are chromophobic with a predominantly diffuse staining pattern. Pseudorosette formation, characteristic of glycoprotein hormone-producing tumours, may also occur. Null cell adenomas are immunonegative for adenohypophyseal hormones and transcription factors. Oncocytic change can occur, and in this case, they are referred to as pituitary oncocytomas.

Oncocytomas characteristically have a diffuse staining pattern by histology. The adenoma cells are larger than null cells and may display acidophilia due to non-specific binding of acidic stains by mitochondria. The pattern of immunoreactivity is similar to null cell or gonadotroph adenomas. By electron microscopy, the sole ultrastructural marker is the extensive accumulation of mitochondria, whereas the other organelles are poorly developed.

Molecular Pathology of Null Cell Adenomas and Oncocytomas

Promoter methylation has been reported in many genes (*LGAL3*, *PLAG1*, and long non-coding RNA *MEG3* and *WIF1*) that function as tumour suppressors with antiproliferative activity. Recent studies reported that the *ENC1* gene, which has an essential role in tumour invasion, may have an important role in null cell adenomas and oncocytomas. Patients with MEN1 have an increased risk of developing null cell adenomas. Approximately 60% of oncocytic pituitary adenomas show somatic mutations in mitochondrial DNA genes, coding for various respiratory complex I components. These mutations contribute not only to the development of oncocytic change, but also to a less aggressive tumour phenotype [27].

Pituitary Carcinomas

Pituitary carcinomas can be diagnosed only when a pituitary neoplasm gives rise to distant, craniospinal, or less frequently, extracranial metastasis. Such tumours are extremely rare and have been associated with poor prognosis. They usually present as invasive macroadenomas and grossly they appear similar to other pituitary adenomas. To date, no criteria have been found to quickly and easily distinguish between an invasive adenoma and one that has the potential to become a carcinoma. Most pituitary carcinomas produce either PRL or ACTH. Other types, including those unassociated with signs of hormone overproduction, are exceptionally rare. Pituitary carcinomas are not accompanied by specific histological features; enhanced mitotic activity, nuclear and cellular pleomorphism do not necessarily indicate malignancy, and vice versa. Neoplasms with bland features may give rise to metastasis. Application of the cell proliferation marker, Ki-67 is more useful; carcinomas display nuclear labelling consistently higher than adenomas [28]. Immunoreactivities of pituitary carcinomas follow the pattern of their non-malignant phenotype, although the degree of immunopositivity may be variably reduced. The p53 expression has been used as a marker for tumour behaviour and high p53 expression is associated with malignancy in both the primary lesion and metastases. Studies have not been able to conclusively confirm the reliability of p53 pertaining to pituitary carcinomas since studies have shown that p53 expression may be negligible or absent in both lesions [29].

Molecular Pathology of Pituitary Carcinomas

Although several studies have reported molecular and genetic defects in both pituitary carcinoma metastases and primary lesions, to date, there is no definitive data on the molecular or genetic profiles of pituitary carcinomas.

Aggressive Pituitary Adenomas

Clinically aggressive pituitary adenomas comprise a broad, and not yet clear, group of tumour types which are diagnosed later, do not respond well to therapy, and are generally associated with a poor prognostic outcome. Aggressive tumours can be either microadenomas or macroadenomas, have the potential to invade into surrounding structures, and are characterized by a higher rate of recurrence compared to their benign adenoma counterparts [30]. A specific histopathological description of the multiple pituitary adenomas may assist in recognizing characteristics suggestive of aggressiveness. According to WHO classification of pituitary adenomas (2017), sparsely granulated somatotroph adenomas, lactotroph adenomas in men, silent corticotroph adenomas, Crooke cell adenomas, and plurihormonal Pit-1 positive adenomas are considered as high risk (of recurrence) adenomas [6]. Apart from the histologic and morphologic methods utilized to classify the tumour types, no single feature can provide a reliable assessment to predict aggressive behaviour. To date, the most reliable and accepted immunohistochemical analysis to evaluate aggressive behaviour in pituitary adenomas is the assessment of the Ki-67 nuclear labelling. If the Ki-67 nuclear labelling index is more than 10%, the tumour may be classified as aggressive, although there is no agreement on this [3]. Another histologic marker used to correlate tumour behaviour is the p53 expression. Studies have shown that tumour behaviour is associated with p53 expression since it is absent in non-invasive pituitary adenomas but large, invasive pituitary adenomas demonstrate strong, diffuse nuclear immunopositivity for p53, although, as with Ki-67, studies have not conclusively confirmed the reliability of p53 as a marker for aggressive behaviour [30]. A clinicopathological classification, which considers proliferation markers, invasion, and size, has been proposed [31], and, recognizing the variable behaviour of pituitary tumours in patients, a recent suggestion to replace the term pituitary adenoma to pituitary neuroendocrine tumour (PitNET) has also been made [32].

Immunohistochemical study of MGMT expression may be used as a prognostic marker for aggressive pituitary adenomas. Studies have shown that progression and recurrence of pituitary adenomas are often associated with loss of MGMT expression [4, 33, 34] and that low MGMT expression may be both a predictive marker for temozolomide response, and a prognostic marker of tumour recurrence and poor prognosis in pituitary adenomas.

REFERENCES

1. Melmed S. Pathogenesis of pituitary tumors. *Nat Rev Endocrinol*, 2011; **7**: 257–66.
2. Scully KM, Rosenfeld MG. Pituitary development: regulatory codes in mammalian organogenesis. *Science*, 2002; **295**: 2231–5.
3. Zhu X, Gleiberman AS, Rosenfeld MG. Molecular physiology of pituitary development: signaling and transcriptional networks. *Physiol Rev*, 2007; **87**: 933–63.
4. Syro LV, Rotondo F, Ramirez A, *et al*. Progress in the diagnosis and classification of pituitary adenomas. *Front Endocrinol (Lausanne)*, 2015; **6**: 97.
5. Osamura RY, Grossman A, Korbonits M, *et al*. Pituitary adenoma. In: Lloyd RV, Osamura RY, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. Lyon: IARC Press, 2017: 14–18.
6. Osamura RY, Lopes, MBS, Grossman A, *et al*. Tumours of the pituitary gland. Introduction. In: Lloyd RV, Osamura, RY, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. Lyon: IARC Press, 2017: 13.
7. Vidal S, Horvath E, Kovacs K, Cohen SM, Lloyd RV, Scheithauer BW. Transdifferentiation of somatotrophs to thyrotrophs in the pituitary of patients with protracted primary hypothyroidism. *Virchows Arch*, 2000; **436**: 43–51.
8. Allaerts W, Vankelecom H. History and perspectives of pituitary folliculo-stellate cell research. *Eur J Endocrinol*, 2005; **153**: 1–12.
9. Caimari F, Korbonits M. Novel genetic causes of pituitary adenomas. *Clin Cancer Res*, 2016; **22**: 5030–42.
10. Syro LV, Rotondo F, Kovacs K, Korbonits M. Clinicopathologic features of familial pituitary adenomas. *Diagnostic Histopathology*, 2016; **22**: 85–91.
11. Syro LV, Builes CE, Di Ieva A, Sav A, Rotondo F, Kovacs K. Improving differential diagnosis of pituitary adenomas. *Exp Rev Endocrinol Metab*, 2014; **9**: 377–86.
12. Rotondo F, Kovacs K, Lloyd RV. The role of molecular pathology in the classification and clinicopathologic evaluation of pituitary neoplasms. *Diagnostic Histopathology*, 2018; **24**: 95–103.
13. Kleinschmidt-DeMasters BK, Lopes MB, Prayson RA. An algorithmic approach to sellar region masses. *Arch Pathol Lab Med*, 2015; **139**: 356–72.
14. Lopes MBS. The 2017 World Health Organization classification of tumors of the pituitary gland: a summary. *Acta Neuropathol*, 2017; **134**: 521–35.
15. Syro LV, Rotondo F, Serna CA, Ortiz LD, Kovacs K. Pathology of GH-producing pituitary adenomas and GH cell hyperplasia of the pituitary. *Pituitary*, 2017: **20**: 84–92.
16. Cooper O, Melmed S. Subclinical hyperfunctioning pituitary adenomas: the silent tumors. *Best Pract Res Clin Endocrinol Metab*, 2012; **26**: 447–60.
17. Cuevas-Ramos D, Carmichael JD, Cooper O, *et al*. A structural and functional acromegaly classification. *J Clin Endocrinol Metab*, 2015; **100**: 122–31.
18. Matsumoto R, Izawa M, Fukuoka H, *et al*. Genetic and clinical characteristics of Japanese patients with sporadic somatotropinoma. *Endocr J*, 2016; **63**: 953–63.
19. Ronchi CL, Peverelli E, Herterich S, *et al*. Landscape of somatic mutations in sporadic GH-secreting pituitary adenomas. *Eur J Endocrinol*, 2016; **174**: 363–72.
20. Larkin S, Reddy R, Karavitaki N, Cudlip S, Wass J, Ansorge O. Granulation pattern, but not GSP or GHR mutation, is associated with clinical characteristics in somatostatin-naive patients with somatotroph adenomas. *Eur J Endocrinol*, 2013; **168**: 491–9.
21. Daly AF, Tichomirowa MA, Petrossians P, *et al*. Clinical characteristics and therapeutic responses in patients with germ-line AIP mutations and pituitary adenomas: an international collaborative study. *J Clin Endocrinol Metab*, 2010; **95**: E373–83.
22. Syro LV, Rotondo F, Cusimano MD, *et al*. Current status on histological classification in Cushing's disease. *Pituitary*, 2015; **18**: 217–24.

23. Di Ieva A, Davidson JM, Syro LV, *et al*. Crooke's cell tumors of the pituitary. *Neurosurgery*, 2015; **76**: 616–22.
24. Ma ZY, Song ZJ, Chen JH, *et al*. Recurrent gain-of-function USP8 mutations in Cushing's disease. *Cell Res*, 2015; **25**: 306–17.
25. Reincke M, Sbiera S, Hayakawa A, *et al*. Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet*, 2015; **47**: 31–8.
26. de Kock L, Sabbaghian N, Plourde F, *et al*. Pituitary blastoma: a pathognomonic feature of germ-line DICER1 mutations. *Acta Neuropathol*, 2014; **128**: 111–22.
27. Kurelac I, MacKay A, Lambros MB, *et al*. Somatic complex I disruptive mitochondrial DNA mutations are modifiers of tumorigenesis that correlate with low genomic instability in pituitary adenomas. *Hum Mol Genet*, 2013; **22**: 226–38.
28. McCormack AI, Dekkers O, Petersenn S, *et al*. Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*, 2018; **178**: 265–76.
29. Kumar K, Macaulay RJ, Kelly M, Pirlot T. Absent p53 immunohistochemical staining in a pituitary carcinoma. *Can J Neurol Sci*, 2001; **28**: 174–8.
30. Di Ieva A, Rotondo F, Syro LV, Cusimano MD, Kovacs K. Aggressive pituitary adenomas--diagnosis and emerging treatments. *Nat Rev Endocrinol*, 2014; **10**: 423–35.
31. Trouillas J, Roy P, Sturm N, *et al*. A new prognostic clinicopathological classification of pituitary adenomas: a multicentric case-control study of 410 patients with 8 years post-operative follow-up. *Acta Neuropathologica*, 2013; **126**: 123–35.
32. Asa SL, Casar-Borota O, Chanson P, *et al*. From pituitary adenoma to pituitary neuroendocrine tumor (PitNET): an International Pituitary Pathology Club proposal. *Endocr Relat Cancer*, 2017; **24**: C5–8.
33. Dai C, Sun B, Liu X, *et al*. O-6-Methylguanine-DNA methyltransferase expression is associated with pituitary adenoma tumor recurrence: a systematic meta-analysis. *Oncotarget*, 2017; **8**:19674–83.
34. Micko ASG, Wohrer A, Hoftberger R, *et al*. MGMT and MSH6 immunoexpression for functioning pituitary macroadenomas. *Pituitary*, 2017; **20**: 643–53.

2.3.4 Imaging of the Pituitary

Jean-François Bonneville, Sonia Nagi, and Iulia Potorac

Adenomas

Non-Functioning Pituitary Macroadenomas

Non-functioning pituitary macroadenomas are by far the most common macroadenomas. They are also named non-secreting pituitary adenomas and, more recently, pituitary neuroendocrine tumours (PitNET). They are divided into eight subtypes, based on the immunohistochemical expression of anterior pituitary hormones and pituitary transcription factors [1].

They are revealed by symptoms of mass effect, such as headache and visual field defects, less frequently by anterior pituitary insufficiency. On the contrary, the presence of diabetes insipidus allows in nearly all cases to reject the diagnosis of pituitary adenoma. Non-functioning pituitary macroadenomas are also discovered by chance in more than 30% of the cases.

The role of MRI is essential:

- to make the proper diagnosis and eliminate other sellar masses
- to evaluate the optic chiasm and optic nerve compression
- to localize the normal pituitary gland
- to search for signs of aggressivity, such as cavernous sinus invasion
- to look for remnants or recurrence after surgery.

Pituitary macroadenomas are centred by an enlarged sella turcica and frequently present with lobular contours. Tumoural extension is usually directed upwards to the suprasellar cistern, with compression and /or thinning of the optic chiasm and amputation of the inferior recesses of the third ventricle. Hyperintensity of the whole optic chiasm or a rail-like pattern on coronal T2-weighted images is a criterion of severity of nerve fibre lesion/damage and must prompt rapid surgical decompression. It has been demonstrated that an angle less than 114 degrees measured between the optic nerve and the floor of the anterior fossa is a sign of poor prognosis despite surgery (Figure 2.3.4.1).

Giant non-functioning adenomas can extend to the frontal or temporal lobes. Blockade of the foramens of Monro can be responsible for hydrocephalus. Extension can also be directed inferiorly within the sphenoid sinus or laterally in the cavernous sinus, most of the times unilaterally.

Non-functioning pituitary macroadenomas usually present with a moderate hypointense signal in T1 and a hyperintense signal on T2-W images. In large pituitary adenomas, T1 as well as T2 signal is usually heterogeneous, with areas of T1 hyperintensity and/or T2 hyper- or hypointensity revealing haemorrhagic or necrotic phenomena. A fluid-fluid level indicates an old haemorrhage. In contrast, meningiomas most of the time display a homogeneous signal on T1 as well as on T2-W images. Post-gadolinium enhancement of the solid part of the tumour is variable but usually moderate. A characteristic microcystic pattern is observed on T2-W images in 50% of silent corticotropic adenomas. The normal residual anterior pituitary gland is visualized on T2 coronal images and after gadolinium infusion. It is compressed and pushed laterally on one side, against the cavernous sinus and superiorly, but almost never inferiorly. The opposite is true for meningiomas, craniopharyngiomas, and Rathke cleft cysts, where the normal gland is pushed down and can be seen inferiorly, below the tumour (Figure 2.3.4.2). The posterior pituitary bright spot is flattened and remains visible on axial T1 sequences; nevertheless, as soon as the tumour height reaches 20 mm, pituitary

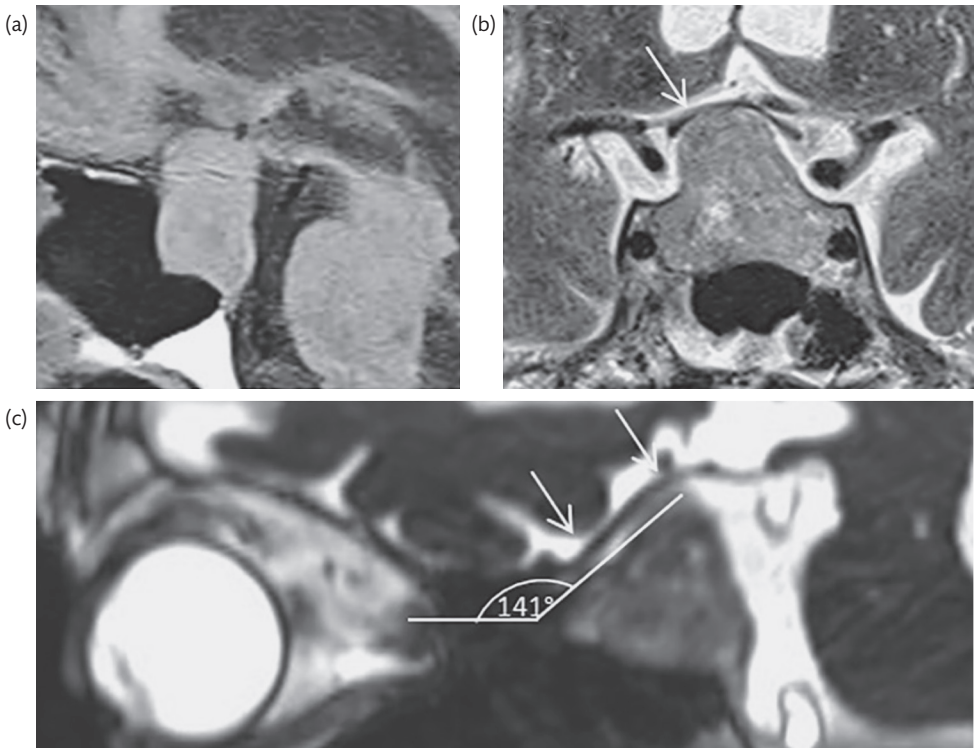


Figure 2.3.4.1 Non-functioning pituitary macroadenoma on sagittal T1 (a) and coronal T2 (b) W images. A sagittal oblique contrast enhanced FIESTA (c) demonstrates the angle between optic nerve and anterior cranial fossa.

stalk compression leads to an ectopic vasopressin storage at the infundibular level.

Sequential coronal MRI exams obtained rigorously with the same projection, for instance in our practice perpendicularly to the subcallosal line, are mandatory to detect recurrence as early as possible (Figure 2.3.4.3).

Cavernous Sinus Invasion

Invasion of the cavernous sinus is usually obvious: enlargement of the laterosellar space, bulging of its lateral wall, blurring of the intracavernous veins with encasement but without narrowing of the intracavernous internal carotid artery make the diagnosis clear. Subtle invasion occurs initially at the posterior part of the cavernous sinus and is better identified on axial views. Integrity of the medial dural wall, more frequently visible on coronal or axial T2-W images at 3.0 T, excludes cavernous sinus invasion (Figure 2.3.4.4). In non-functioning pituitary adenomas, the cavernous sinus invasion, which represents a sign of aggressiveness, is usually unmodified by the surgeon/usually persists after surgery and is a potential source of tumoural regrowth years after surgery [2].

Apoplexy

Apoplexy is a severe complication of pituitary adenomas resulting from an ischaemia or a brutal haemorrhagic event. Apoplexy is symptomatic, unlike subacute haemorrhage from which it has to be differentiated. Most pituitary apoplexies complicate pituitary middle-sized macroadenomas but rarely giant adenomas, as if haemorrhagic infarction preferentially occurs when the responsible

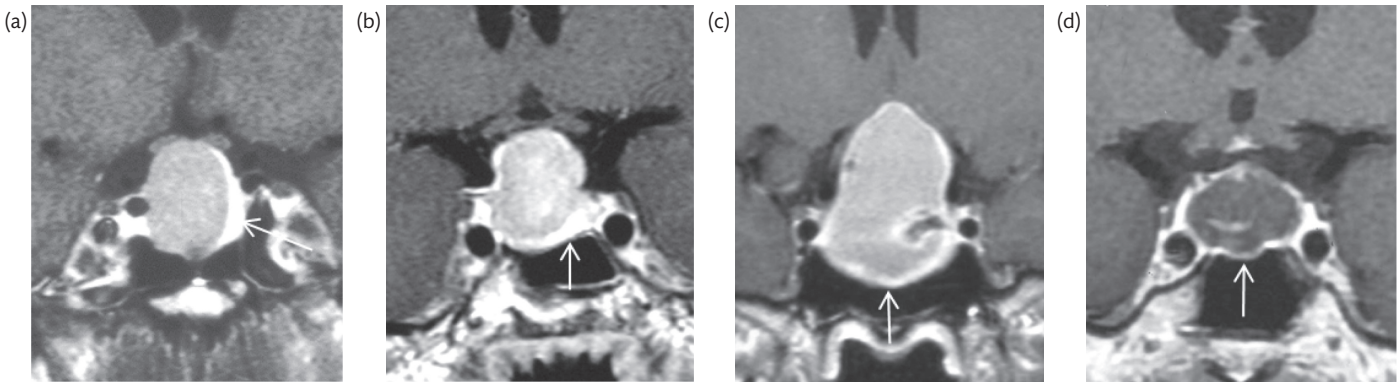


Figure 2.3.4.2 Coronal enhanced T1-W images demonstrating the normal anterior pituitary gland (arrows) in pituitary adenoma (a), sellar meningioma (b), craniopharyngioma (c), and Rathke cleft cyst (d).

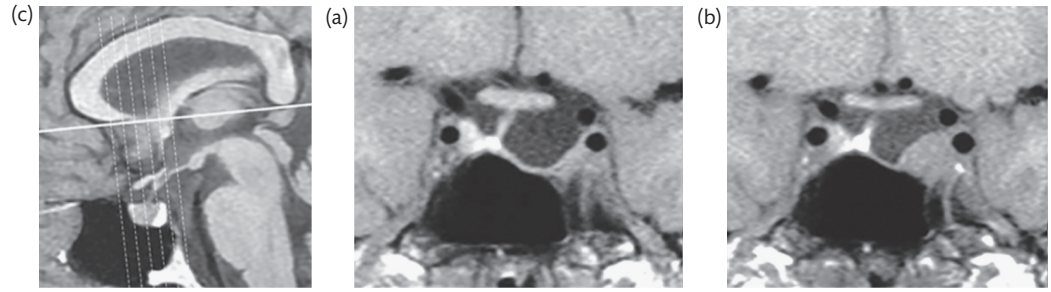


Figure 2.3.4.3 Sequential coronal T1-W images (a, b) are easily compared when obtained perpendicularly to the subcallosal line drawn on (c), sagittal T1-W image.

pituitary adenoma is still contained by the sellar diaphragm. Not infrequently, apoplexy reveals the pituitary adenoma. Apoplexy of an adenoma remnant can also occur years after surgery [3]. Diagnosis may be difficult in the early stage, the classical predominant hyperintensity on T1 being frequently absent; the T2 sequence is more sensitive by demonstrating heterogeneous low signal intensity. Diffusion-weighted images can show an increased signal intensity if compared with the normal brain and a low apparent diffusion coefficient. Thickening of the sphenoid sinus mucosa is present from the early stage (Figure 2.3.4.5). After gadolinium injection, peripheral enhancement is noticed with no or minimal enhancement of the central part of the sellar content.

Sequential MRI will demonstrate the gradual increase in the T1 hyperintense signal, from the periphery towards the centre of the mass. If surgery is not indicated, as is usually the case, MRI will show shrinkage of the sellar content in several weeks or months.

The Postoperative Sella

In the immediate postoperative period, some neurosurgeons request for an MRI to eliminate a possible complication. But blood products and packing material can obscure the sellar content and possible residual tumour in the first weeks after surgery. Postoperative MRI is then usually not obtained before 3 months after surgery. At this time, blood and most of the packing material, if any, are resorbed at the exception of fat grafts which can persist many years after surgery. Coronal high-resolution 2 mm thick T2-W sequence is the most informative, whereas contrast enhanced MRI can frequently be spared. Remodelling of the normal pituitary gland occurs 6 months after surgery at the latest; it appears as a homogeneous triangular or oval small mass in contact with the sellar floor and the medial wall of the cavernous sinus, most of the times isointense to the brain cortex on T2-weighted sequences. The T2 signal of the tumoural remnant, if any, is quite always more or less hyperintense if compared with that of the residual normal pituitary tissue and of course the same as preoperatively. In subsequent years, MRIs rigorously performed with the same parameters and the same inclination will be able to demonstrate stability or an early slight increase in volume of the remnant and no change of the normal pituitary tissue (Figure 2.3.4.6) [4, 5].

Prolactin-Secreting Pituitary Adenomas

In women with amenorrhea and galactorrhoea and a prolactin level above 35 ng/ml, the chances to find a prolactinoma on MRI are very high. These adenomas display a T1 hypointense signal and a more or less hyperintense signal on T2 (Figure 2.3.4.7). When the clinical presentation is clear and the diagnosis confirmed by T1/T2 MRI sequences, the protocol can be simplified and gadolinium injection avoided. Most are microprolactinomas and there is a good correlation between prolactin level and adenoma size. For instance, a prolactin level of 60–90 ng/ml usually corresponds to an adenoma of

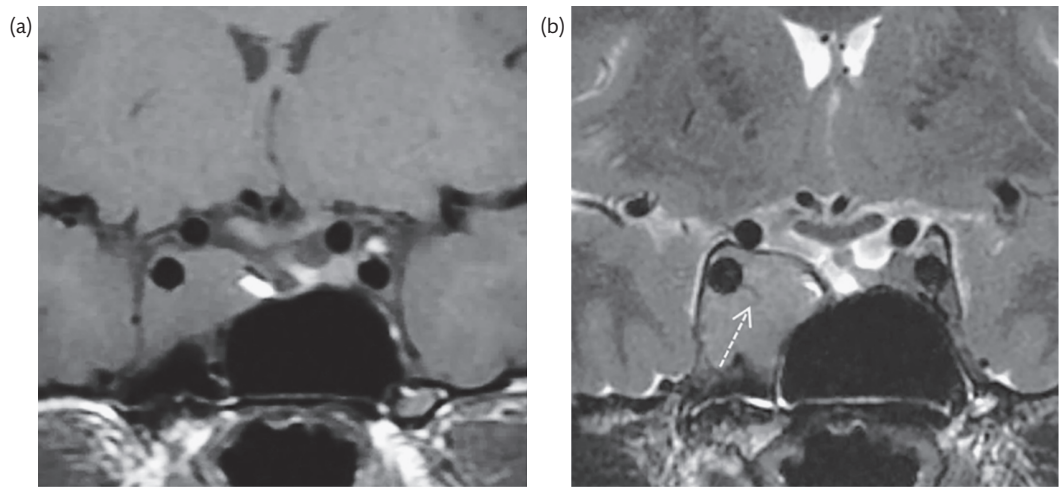


Figure 2.3.4.4 Non-functioning pituitary adenoma years after surgery with intrasellar and intracavernous tumoural remnant. T1- (a) and T2-W images (b). Flap of the medial wall of cavernous sinus (arrow).

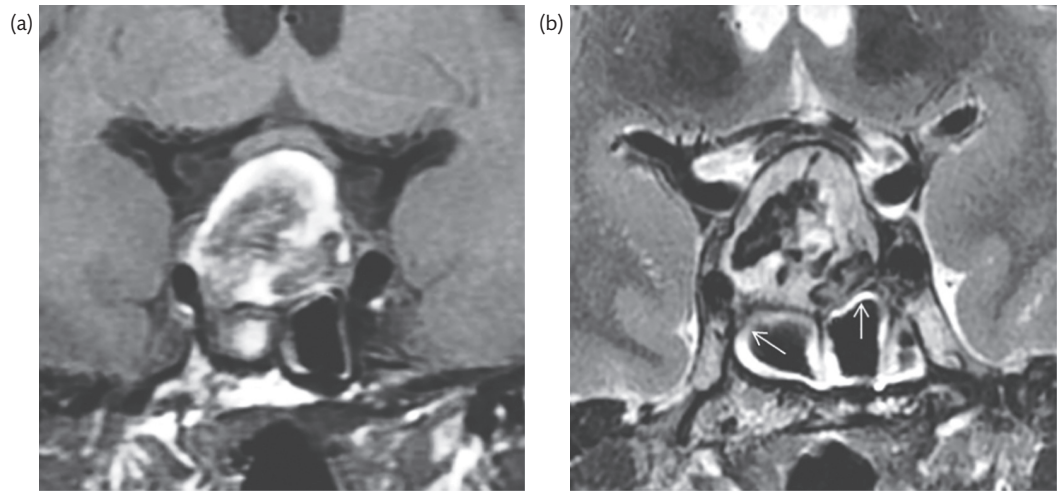


Figure 2.3.4.5 Pituitary apoplexy on coronal T1 (a) and T2 (b) W images. Thickened sphenoid sinus mucosa (arrows).

7–10 mm in diameter. Nevertheless, some adenomas do not obey this rule. A frank T2 hyperintensity reflecting degenerative changes or a T1 hyperintensity indicating a partial haemorrhagic transformation is less secreting and correlate with a lower prolactin level. Such a haemorrhage is usually asymptomatic and frequently seen at diagnosis as well as after cabergoline treatment. Prolactin levels above 150–200 ng/ml correspond to pituitary macroadenomas with suprasellar extension and/or cavernous sinus invasion.

After cabergoline treatment, shrinkage of the adenoma occurs quickly, as soon as a few days or weeks and the hyperintense T2 signal increases. If pregnancy is obtained and cabergoline withdrawn, the size of the adenoma roughly doubles when compared with its size before treatment. This has to be kept in mind for the management of prolactinomas during pregnancy. As a general rule, pregnancy has no adverse effect for microprolactinomas. Nevertheless, a close supervision has to be discussed in haemorrhagic adenomas, in prolactinomas with unusually T2 hypointense signal or T2 microcystic pattern or in women with a small sella [6].

In men, most prolactinomas are macroadenomas and even giant adenomas. In contrast with the usual T2 hyperintensity of prolactinomas, some harbour T2-hypointense areas corresponding to calcifications or amyloid deposits. Prolactin levels are very high, and can reach thousands of ng/ml. Cavernous sinus invasion is very frequent. Extension to the entire sphenoid sinus can simulate an ear/nose/throat (ENT) cancer (Figure 2.3.4.8). Cabergoline is usually efficient with a rapid tumour shrinkage but can lead to a cerebrospinal fluid (CSF) fistula in case of erosion of the sphenoid sinus walls. Initiation of treatment with low-dose dopamine agonists does not always allow to avoid this complication [7].

Growth Hormone-Secreting Pituitary Adenomas

Growth hormone (GH)-secreting pituitary adenomas or somatotropinomas represent the third most frequent type of pituitary adenomas following prolactinomas and non-functioning tumours. Their prevalence ranges from 8 to 16% of all types of adenomas with an incidence of 3–4 cases/million inhabitants/year.

Most of the times, somatotropinomas are revealed during the imaging studies that follow the diagnosis of acromegaly. Rarely, they can be an incidental finding on an imaging examination performed for a different reason in a patient that was not known to suffer from acromegaly.

Studies performed on large series of patients with acromegaly have shown that somatotropinomas are macroadenomas in over two-thirds of cases at diagnosis. The largest diameter of most GH-secreting adenomas is between 1 and 2 cm. Giant tumours are rare. There seems to be an inverse relationship between adenoma size

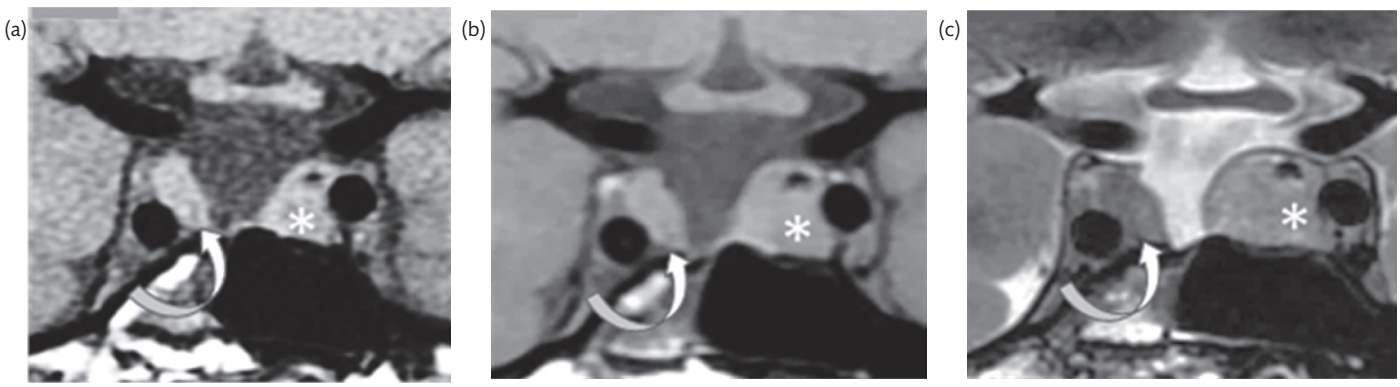


Figure 2.3.4.6 Pituitary adenoma remnant. Sequential T1-W images (a, b) show an enlargement of the remnant (asterisk). In c, coronal T2-W image, the normal pituitary gland (curved arrow) is unchanged and less intense than that of the tumoural remnant.

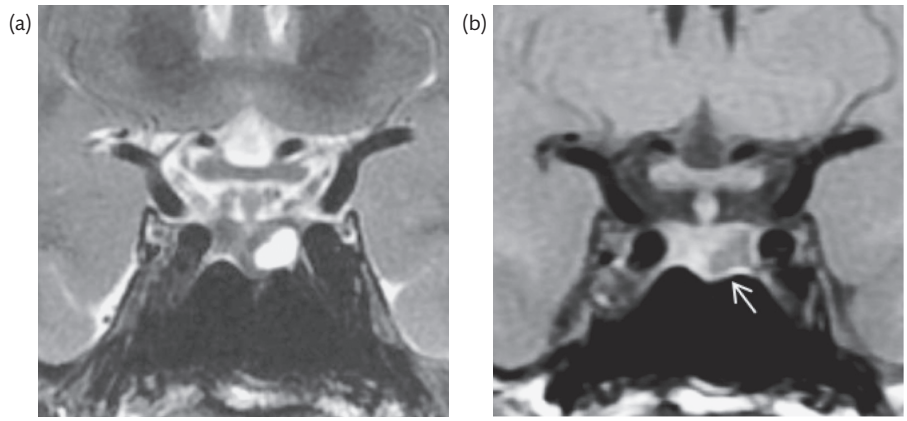


Figure 2.3.4.7 Microprolactinoma hyperintense on T2 (a) and hypointense on T1-W images (b).

and patient age at diagnosis, with younger patients often developing larger tumours. Among these younger patients, tumours develop more frequently in a genetic context and, when appearing before the closure of the growth plates, will lead to gigantism. The biochemical anomalies (GH and IGF1 increases) seem to also be related to tumour size with bigger tumours being responsible for higher levels of GH and IGF1. This is valid for adenomas measuring up to 20 mm of largest diameter. Beyond this threshold, correlations between secretory levels and tumour size are no longer found.

A particularity of the natural evolution of GH-secreting adenomas is their tendency to extend towards the sphenoid sinus. The reasons behind this propensity towards infrasellar development are not yet clear and may involve local effects of the somatotropinoma or the consequences of GH secretion on connective tissue. Suprasellar extension with optic chiasm compression is rarer than infrasellar extension, which explains why visual field defects are not very common at the diagnosis of acromegaly.

Unlike other types of pituitary adenomas, GH-secreting ones exhibit a T2-hypointense signal in around 50% of cases, whereas prolactinomas and non-functioning adenomas are more rarely T2-hypointense. Although different definitions of T2-weighted signal intensity have been employed in the existing literature, we believe

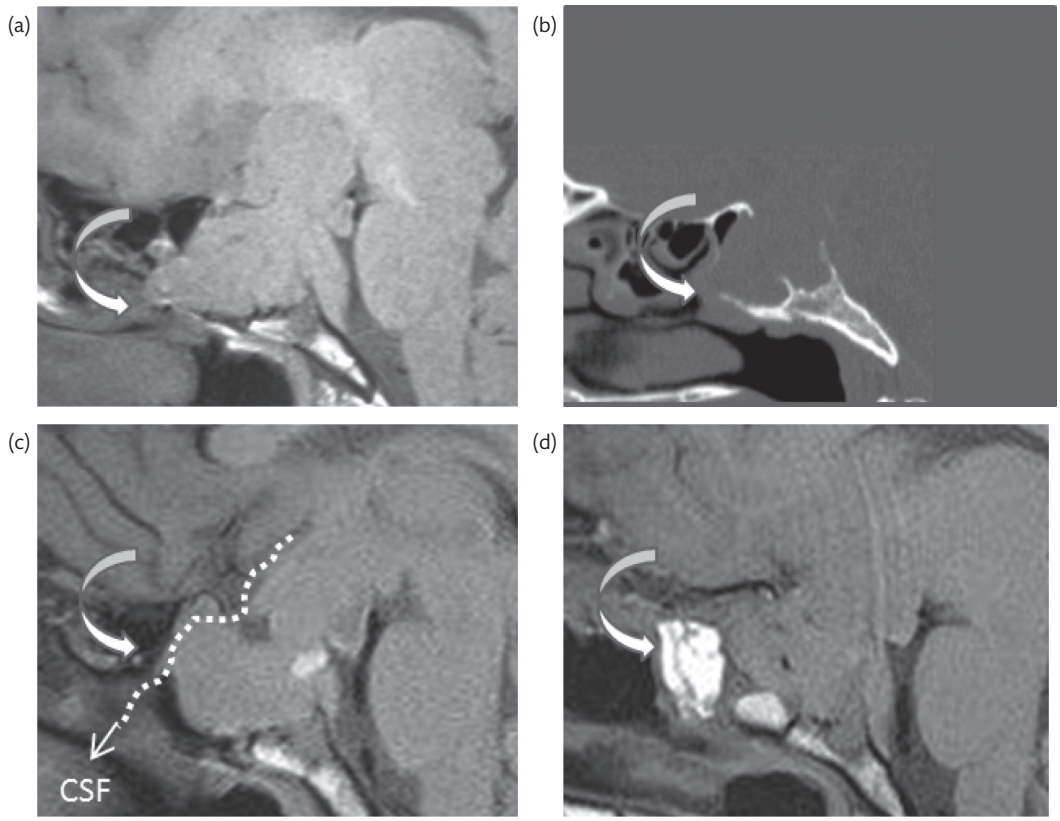


Figure 2.3.4.8 Macroprolactinoma invading sphenoid sinus in a male. Prolactin level is 10 000 ng/ml. (a) Sagittal T1-W image. Huge tumoural process extending upwards, downwards, and posteriorly. The inferior surface of the sphenoid bone is eroded as demonstrated on (b), sagittal CT (curved arrow). In (c), after low dose cabergoline, limited tumoural shrinkage leading to a CSF fistula. (d) Fat graft interposition to cure the fistula.

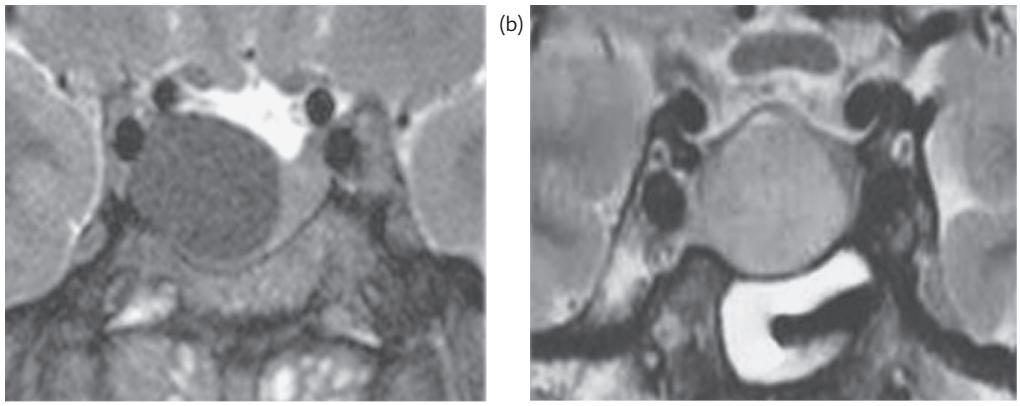


Figure 2.3.4.9 Growth hormone-secreting adenomas on T2-W images: T2 hypointense (a) and T2 hyperintense (b).

that the most accurate one is comparison with normal pituitary tissue and, only when the latter is not visible, with the grey matter of the temporal lobe. T2-hypointense somatotropinomas present different characteristics compared to T2-hyper and isointense ones. T2-hypointense adenomas seem to be smaller, less invasive tumours, with higher secretory levels (Figure 2.3.4.9). These T2-hypointense GH-secreting adenomas respond better to somatostatin analogue treatment. This is true both for primary treatment in terms of decreases of GH and IGF1 and of tumour reduction and for adjunctive therapy following surgical failure with hormonal control being achieved more frequently in patients with T2-hypointense tumours.

Current research is focusing on the relationships between T2-W signal intensity in somatotropinomas and histologic parameters. These might explain on one hand the different T2-W signal intensity of somatotropinomas and, on the other, the differences in adenoma behaviour, whether spontaneous or in response to somatostatin analogue treatment. Although not confirmed on large studies, it is believed that the densely granulated pattern of somatotropinomas correspond to T2-hypointensity, whereas T2-hyperintense adenomas are mostly sparsely granulated. Iron, fibrous tissue, or amyloid content do not seem to influence T2-W signal intensity. One study found that T2-hypointensity is correlated to expression of the somatostatin receptor type 5 (SSTR5), whereas no difference seems to exist with regards to SSTR2. Further detailed studies on larger series are required in order to clarify the matter [8].

Corticotropic Pituitary Adenomas

Corticotropic pituitary adenomas are responsible for adrenocorticotropic hormone (ACTH)-dependent hypercorticism. Their detection is of major importance for the treatment of Cushing's disease and represents a difficult challenge for radiologists. Less of half of corticotropic pituitary adenomas are macroadenomas. Most are microadenomas and a large number are picoadenomas (i.e. less than 3–4 mm in diameter). Moreover, a significant number escape MRI detection even at 3.0 T. In these cases, inferior petrosal sinus blood sampling for plasma ACTH measurements has an important role. 11C methionine PET-computed tomography (CT) combined with 3D MRI is also very promising but for now is not able to detect the tiniest pituitary adenomas. Macrocorticotropic adenomas do not differ from other adenomas. They are hypointense on T1 and more or less hyperintense on T2-weighted sequences. Micro and picoadenomas can be located centrally or laterally in the sella, more readily posteriorly and in contact with the posterior lobe. They are suspected through the presence of a small or even millimetric T2 hyperintense area and confirmed by a localized enhancement defect on 3D MRI. Cavernous sinus invasion can be present even in microadenomas. Optimized MRI sequences are here mandatory, but increase the risk of false positive findings. Dynamic MRI has been proposed to distinguish ACTH-producing adenomas from non-functioning pituitary adenomas, the time-intensity curve of enhancement featuring a rapidly enhancing and slow washout pattern in the first ones. However, dynamic imaging can lead to false positive images, particularly if a laterally positioned posterior lobe is not recognized (Figure 2.3.4.10). Delayed washout of corticotropic adenomas could also be demonstrated on a late postcontrast FLAIR sequence. Administration of corticotropin-releasing hormone (CRH) during 3T MRI has also been proposed to improve the detection of corticotropic adenomas. These data have to be confirmed in larger series of patients [9].

Silent corticotroph pituitary adenomas are clinically and biologically silent, but harbour ACTH-secreting cells at pathology. Because they are endocrinologically silent, silent corticotroph adenomas are either diagnosed when large or incidentally and behave like non-functioning macroadenomas [10]. A cystic or microcystic pattern on T2-W images is present in more than 50% of silent corticotroph adenomas (Figure 2.3.4.11).

Rathke Cleft Cysts

Rathke cleft cysts (RCC) are intra and/or suprasellar lesions believed to derive from remnants of the Rathke pouch. If 'microRCCs' are considered, they represent the most frequent lesions of the sellar region revealed on MRI and have to be evoked first when an incidentaloma is revealed. Characteristic imaging patterns make their diagnosis easy, provided that MRI sequences are carefully selected.

RCC's can be classified by location, content, and related symptoms.

RCC's Originate Intra–Or Suprasellarly

Intrasellar RCCs are by far the most frequent, diagnosed by their location, just in front and in close contact with the posterior lobe. Most are hyperintense on T1 and can be mistaken for the

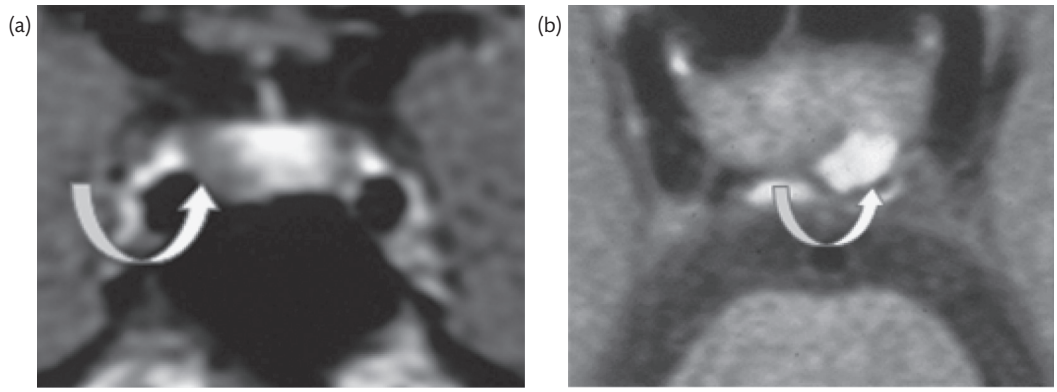


Figure 2.3.4.10 False positive diagnosis of a pituitary microadenoma on early phase of dynamic imaging: the pseudo right defect in enhancement (curved arrow in a, coronal T1-W image) is related to a lateral normal position of the T1 hyperintense posterior lobe visible on (b) axial T1-W image.

posterior lobe or even for a fatty dorsum sellae: axial T1 sequence with fat saturation solves the problem. They are strictly midline and symmetrical with a curved regular anterior border, except when coexisting with a pituitary adenoma (Figure 2.3.4.12). Most intrasellar RCCs are asymptomatic, but can expand into the chiasmatic cistern.

Suprasellar RCCs are said to originate from the pars tuberalis. They are seen on the upper surface of the pituitary gland or embedded in the pituitary as an egg in an eggcup, on the midline (Figure 2.3.4.13a) very rarely off midline. They can also involve the pituitary stalk and simulate a tumour (Figure 2.3.4.13b). These pituitary stalk RCCs never prevent the normal vasopressin storage within the posterior lobe. Suprasellar RCCs can be hypo- or hyperintense on T1 but usually less hyperintense than when intrasellarly located and more or less hyperintense on T2-weighted sequences. Cyst signal is related to its proteic content and can vary with mucus secretion from its wall and dehydration. Cyst volume can also vary with time; it can increase, decrease, or even disappear spontaneously. In about 70% of the cases, a T2 hypointense hyperproteinic nodule is able to make the diagnosis of RCC indisputable; this nodule can be tiny or involve the whole RCC (Figure 2.3.4.14). After gadolinium infusion, the thin monocellular cyst wall does not enhance in asymptomatic cysts [11].

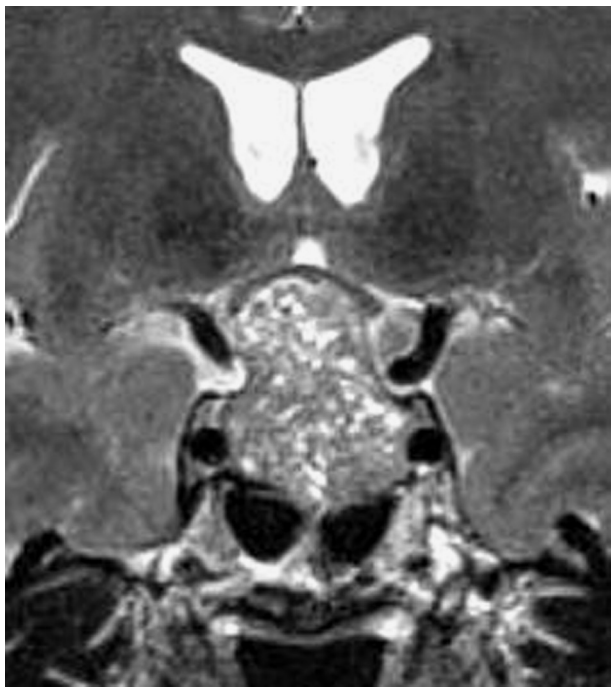


Figure 2.3.4.11 Silent corticotropic adenoma. Microcystic pattern on coronal T2-W MRI.

We Used to Differentiate Symptomatic from Complicated RCCs

Symptomatic RCCs are responsible for mass effects due to their size, mainly visual field defects if the optic chiasm is threatened, headache, rarely anterior pituitary deficits [12]. Mild hyperprolactinemia due to stalk compression can be encountered. **Complicated RCCs** can present with haemorrhage, infection or rupture and are mainly responsible for anterior and/or posterior pituitary deficits, sometimes after one or several short episodes of severe headache, probably related to a reactional hypophysitis. These pituitary deficits usually occur in a short period of time. In these cases of complicated RCCs, the cyst wall is thickened and appears enhanced with gadolinium infusion. The distinction between wall enhancement and normal pituitary tissue encircling the cyst has to be made; dynamic imaging could help to this purpose (Figure 2.3.4.15).

The characteristic pattern of RCCs presented here render the differential diagnosis with pituitary adenomas usually simple. Moreover, mass effect, such as enlargement of the sella is less frequent in RCCs than in pituitary adenomas. Finally, the position of the normal pituitary gland is important to consider: unlike what is observed in pituitary adenomas, the normal gland in RCCs can be typically seen between the cyst and the sellar floor (Figure 2.3.4.2d).

Perisellar Tumours

Craniopharyngioma

Craniopharyngioma is an epithelial tumour, classified World Health Organization grade 1. They represent approximately 5% of all intracranial tumours. Craniopharyngiomas are derived from remnants of the Rathke's pouch and can occur anywhere along the course of the craniopharyngeal duct, from the nasopharynx to the third ventricle. They are characterized by a bimodal age distribution, with a main peak in children (5–14 years) and a second peak in adults (fifth–seventh

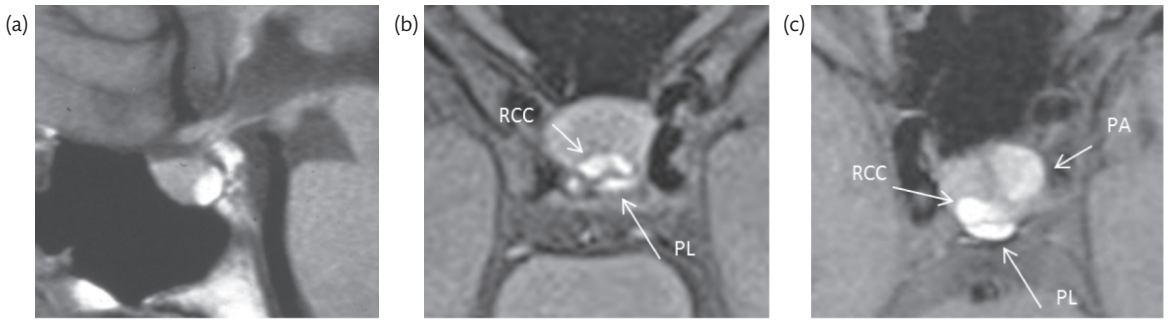


Figure 2.3.4.12 Micro Rathke cleft cyst on sagittal T1 (a) and axial T1 Fatsat (b) W images. Coexisting haemorrhagic pituitary adenoma on axial T1-W image on (c).

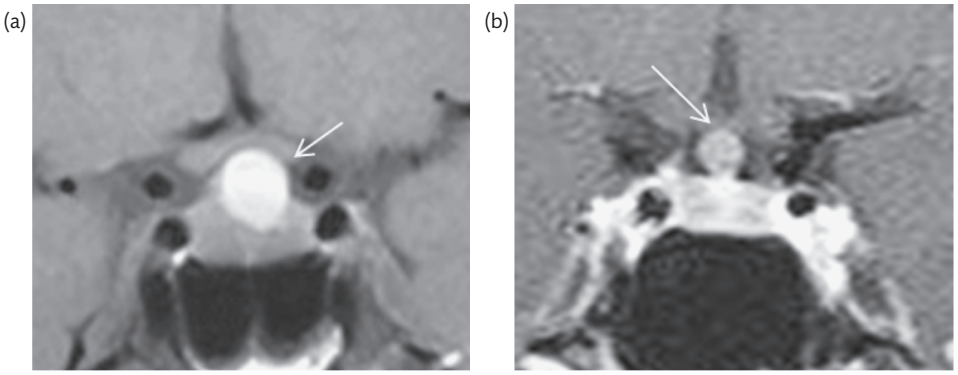


Figure 2.3.4.13 Rathke cleft cysts as an egg in an eggcup on coronal T1-W image (a) and of the pituitary stalk in (b) coronal T1-W-enhanced image.

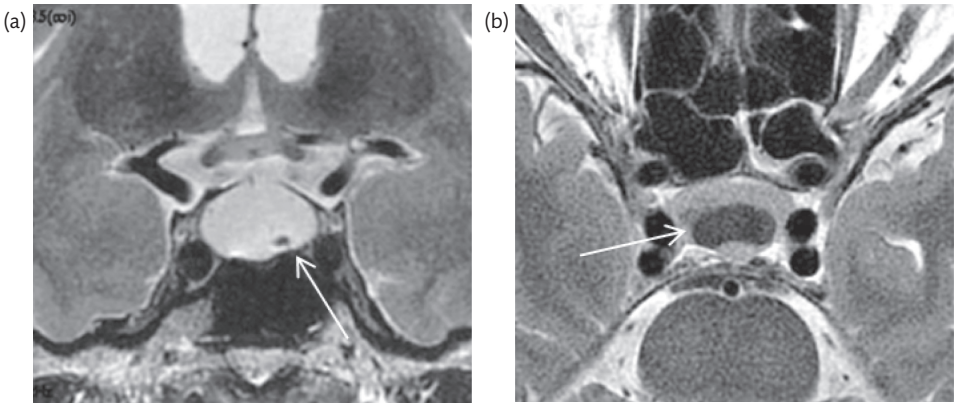


Figure 2.3.4.14 Rathke cleft cysts. Hyperproteinic pathognomonic T2 hypointense nodules, small in (a) coronal T2-W image, large and involving the whole cyst in (b) axial T2-W image.

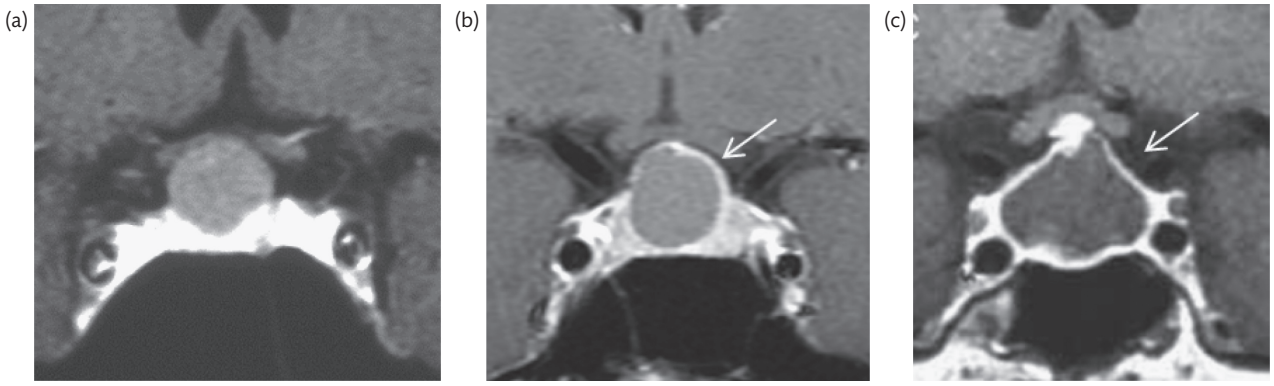


Figure 2.3.4.15 Rathke cleft cyst wall. (a–c) Coronal T1-W enhanced images. Non-visible and non-enhanced cyst wall in non-complicated RCC (a). Thin and incomplete enhanced wall is also a normal pattern (b). Thick and complete wall enhancement in complicated RCC (c).

decade). There are two histologic subtypes: the adamantinomatous type, the most common form, typically seen in children and adolescents, and the papillary type, seen almost exclusively in adults, usually in the third ventricle. They most commonly involve the suprasellar and sellar area, where they are either anterior to, or posterior to the optic chiasm. In fewer cases, their location is purely suprasellar or entirely intrasellar. This explains the variability of the clinical symptoms. Headache, visual field defects, decreased visual acuity, and hormone disturbances are the common clinical symptoms in the suprasellar location. The most common endocrine dysfunctions encountered in children are growth retardation, delayed or precocious puberty. Global hypopituitarism, hyperprolactinemia, and diabetes insipidus are the other endocrine disorders that can be encountered. Rare locations include the third ventricle and the infrasellar region (sphenoid bone, ethmoid bone, and nasopharynx) [13].

Adamantinomatous craniopharyngiomas typically present with three components: solid, cystic, and calcified portions which occupy the suprasellar cistern (Figure 2.3.4.16). The cystic component is single or multiple, usually hyperintense on T1, T2, and FLAIR weighted images reflecting high protein concentration, cholesterol, or methaemoglobin contents. Less commonly, a CSF-like signal pattern can be seen in huge craniopharyngiomas with a thick wall enhancement. The solid component has variable signal intensities and shows contrast enhancement. Calcifications are common, they are better seen on computerized tomography. The sella is often enlarged and may be eroded.

Papillary craniopharyngiomas usually appear as predominantly solid or mixed solid and cystic spherical tumours in the suprasellar region. They are characterized by a more upward growth towards the third ventricle. The solid parts classically show isointense signal intensity on T1- and T2-weighted images, hypointense signal on diffusion-weighted images reflecting the low-grade character and a homogeneous or reticular enhancement. They less frequently contain calcifications. Surgical removal is easier in this subtype than in the adamantinomatous type.

Although histologically benign, craniopharyngiomas are aggressive neoplasms and recurrences are common because complete surgical resection is difficult. Treatment usually consists of surgical resection with or without adjuvant radiation therapy.

Optochiasmatic and Hypothalamic Glioma

Chiasmatic and hypothalamic gliomas are childhood tumours, 75% occurring in the first decade of life, with a similar gender distribution. In children, low-grade astrocytoma is the most frequent type of lesion, and 10–70% of these patients have neurofibromatosis type 1 (NF1). Visual loss with optic atrophy and growth delay are the most common signs. Intracranial hypertension syndrome with hydrocephalus is seen in large tumours. Optic glioma associated with NF1 is frequently asymptomatic. Optic and hypothalamic gliomas appear as rounded or lobulated solid suprasellar masses hyperintense on T2-W images, hypointense on T1-W images with a variable, often heterogeneous enhancement. A posterior extension along the optic tracts is usual (Figure 2.3.4.17). Suprasellar

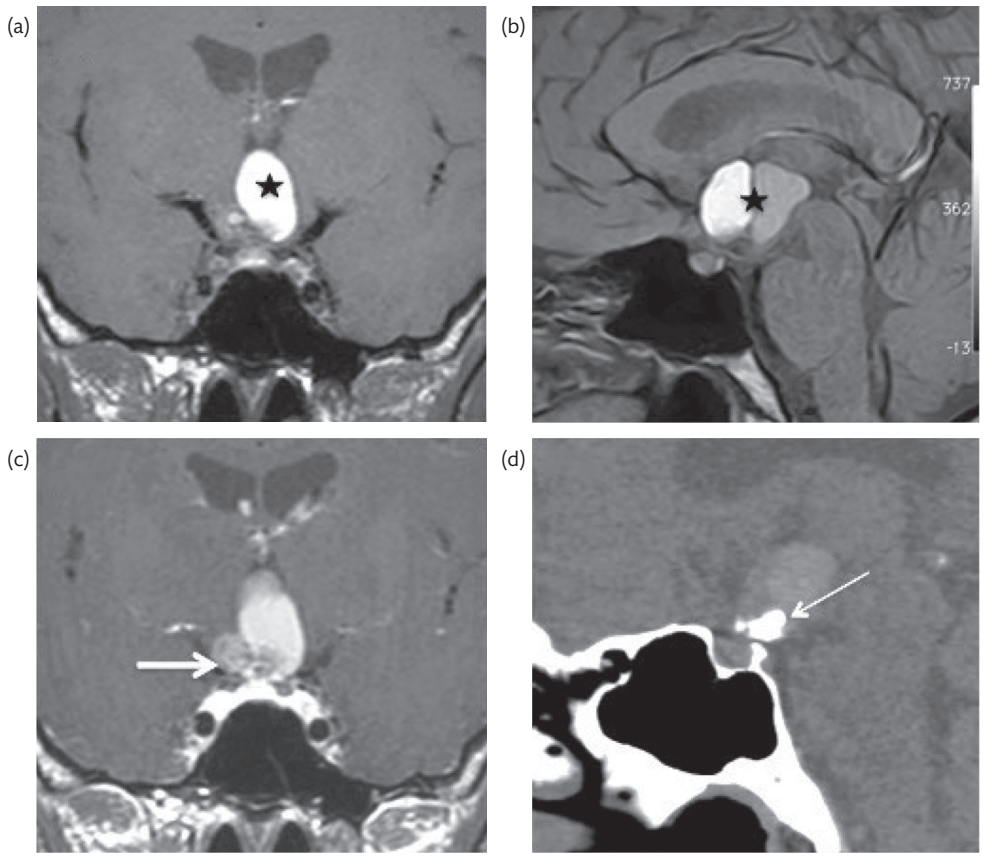


Figure 2.3.4.16 Suprasellar adamantinomatous craniopharyngioma on coronal (a) and sagittal (b) T1-W images, coronal T1-W image after contrast (c) and coronal CT-scan (d). Suprasellar multiloculated mixed mass with varying mainly bright signal intensities (asterisk) and solid calcified component heterogeneously enhanced (arrow). Hyperdense calcifications on CT-scan (thin arrow).

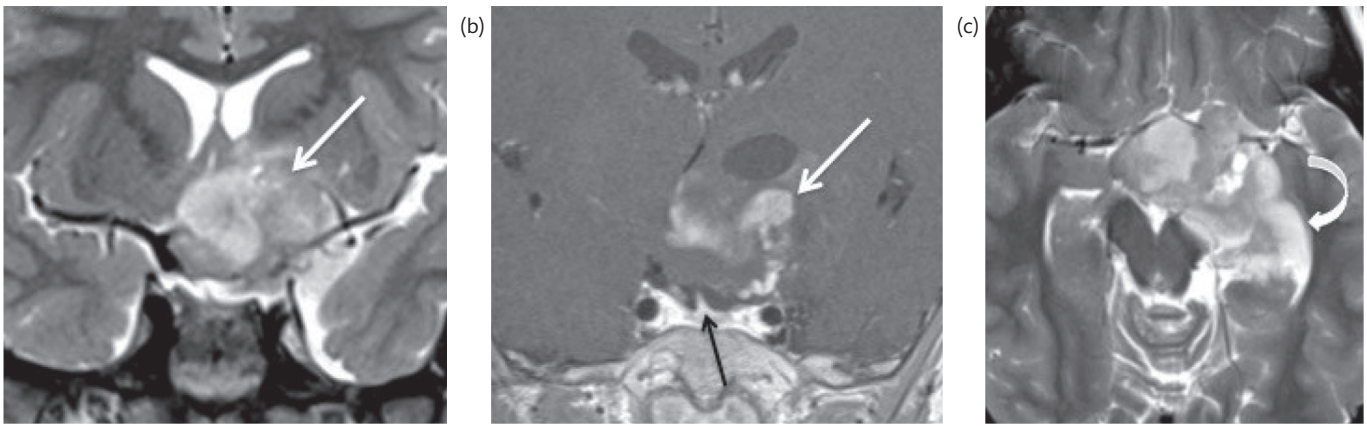


Figure 2.3.4.17 Suprasellar optochiasmatic glioma. Coronal T2 (a) and T1 after gadolinium (b) W images. Suprasellar multiloculated mass hyperintense and heterogeneous on T2 with strong and heterogeneous enhancement (arrow). Axial T2-W image (c). Posterior tumoural extension along the left optic tract (curved arrow).

extension compressing the third ventricle may be responsible for hydrocephalus [14].

Suprasellar Germinoma

Germinoma is the most common germ cell tumour of the suprasellar region. It can be a primitive or a metastatic lesion from a pineal germinoma. It is most commonly seen in adolescents and young adults with no gender preference. Clinical symptoms include diabetes insipidus, precocious puberty, or growth failure and visual loss. Tumour markers such as alpha-fetoprotein and human chorionic gonadotropin can be found in the serum and/or CSF. The most precocious radiological sign is the infundibular thickening with the absence of the posterior lobe bright spot on T1-W images. As it grows, the lesion appears as a well-delineated round or lobulated lesion markedly enhanced. Intraventricular extension and infiltration of the basal ganglia and the corpus callosum can occur. Leptomeningeal spread with drop metastases along the spinal cord are reported, so before treatment, a complete examination of the brain and spinal canal is required (Figure 2.3.4.18). Early diagnosis may reduce the risk of dissemination and the morbidity of treatment [15].

Meningioma

Meningiomas are the most common intracranial tumours in adults, with 20% occurring at the skull base and arising from the arachnoidal cells of the meninges. The common sites of origin are the planum sphenoidale, the tuberculum sellae, the clinoid processes, the sellar diaphragm, and the cavernous sinus.

Pre- and suprasellar meningiomas can be revealed by anosmia, frontal syndrome, and visual disturbances when the optic chiasm or optic nerves are compressed. Meningioma appears iso to hyperintense on T2-W images and is hypo- to isointense on T1-W image with a marked homogeneous enhancement after gadolinium. Hyperostosis is commonly associated. Upward convexity of the jugum sphenoidale in the area near the insertion of the meningioma, termed blistering, can be seen (Figure 2.3.4.19). Posterior extension can completely cover the sellar diaphragm, making the identification of the site of tumour insertion difficult. Diagnosis of meningiomas of the sellar diaphragm can be tricky. They should be differentiated from the underlying normal pituitary gland (Figure 2.3.4.2b) [16].

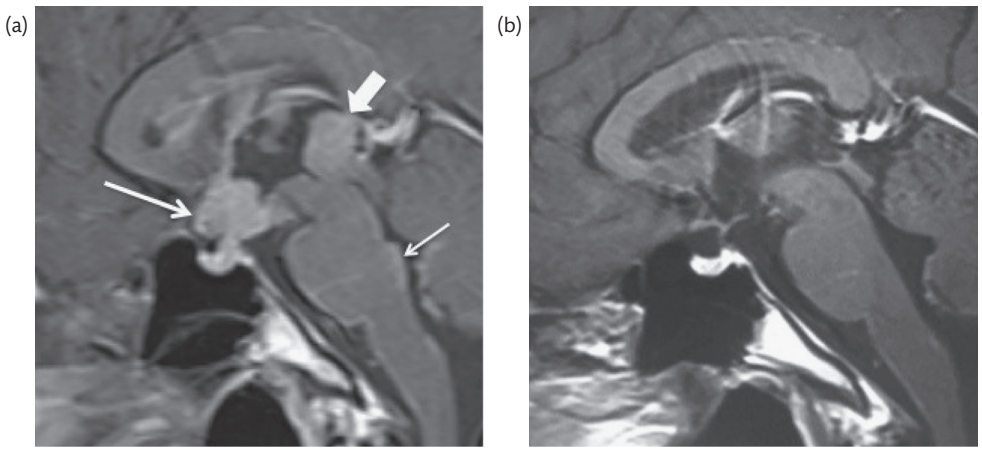


Figure 2.3.4.18 Suprasellar and pineal germinoma. Sagittal T1-W images after gadolinium before (a) and after treatment (b). Suprasellar tumour strongly enhanced involving the optic chiasm (arrow). Contrast enhanced pineal mass (thick arrow). Metastatic leptomeningeal linear and micronodular enhancement (short arrow). After treatment, complete radiological remission.

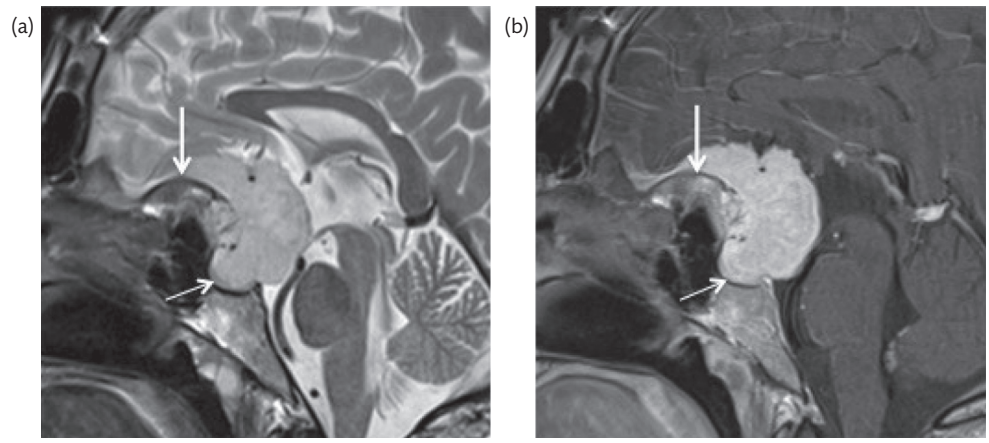


Figure 2.3.4.19 Presellar meningioma with posterior extension in the suprasellar cistern. Sagittal T2 (a) and T1 after contrast (b) W images. The tumour is hyperintense to the grey matter on T2 with strongly and homogeneously enhancement (asterix). Blistering of the jugum sphenoidale (arrow). The pituitary gland is pushed inferiorly (thin arrow).

Cavernous sinus meningiomas are revealed by oculomotor deficit, ptosis, diplopia, anisocoria, or complete ophthalmoplegia. Compression of the internal carotid artery can be responsible for ischaemic events. Some of them arise from the lateral dural wall, while others are developed exclusively inside the cavernous sinus. These frequently extend into the sella turcica, pushing the pituitary gland towards the opposite side (Figure 2.3.4.20).

Metastasis

Pituitary metastases are rare. They are found particularly in breast and lung cancer patients. The pituitary gland can be reached via hematogenous spread, via the skull base, or by meningeal spread. The posterior lobe is the most frequently involved, probably because of its direct arterial blood supply by the inferior hypophyseal arteries. Diabetes insipidus is the most frequent symptom. Oculomotor nerves palsies are seen if the cavernous sinus is involved. On MRI, loss of hyperintensity in the posterior pituitary is constant. We can also see a non-specific isolated thickening of the infundibulum, a large intra- and suprasellar dumbbell-shaped tumour with or without invasion of the cavernous sinus. Erosion or destruction of the dorsum sellae is better appreciated on CT. Rapid growth of the lesion on sequential MRIs suggests a malignant disease [17].

Lymphoma

Primary central nervous system lymphoma (PCNSL) is rarely limited to the sellar region. Sellar and suprasellar locations represent a diagnostic challenge. The most common presentation is hypopituitarism, followed by headache, diplopia, diabetes insipidus, hyperprolactinemia, and fever. Prognosis is generally poor and the disease is rapidly progressing.

The common MRI finding is a T2 hypointense mass lesion resulting from the dense cellularity and high nucleus-to-cytoplasm ratio of lymphoma. Contrast enhancement is typically intense and homogeneous in immunocompetent patients, but inhomogeneous or ring-like in immunocompromised patients. Diffusion, perfusion, and MR spectroscopy may help differentiate central nervous system (CNS) lymphomas from other brain lesions. Peritumoral oedema is generally less pronounced than in metastases. Confirmation of diagnosis is most frequently obtained after surgery [18, 19].

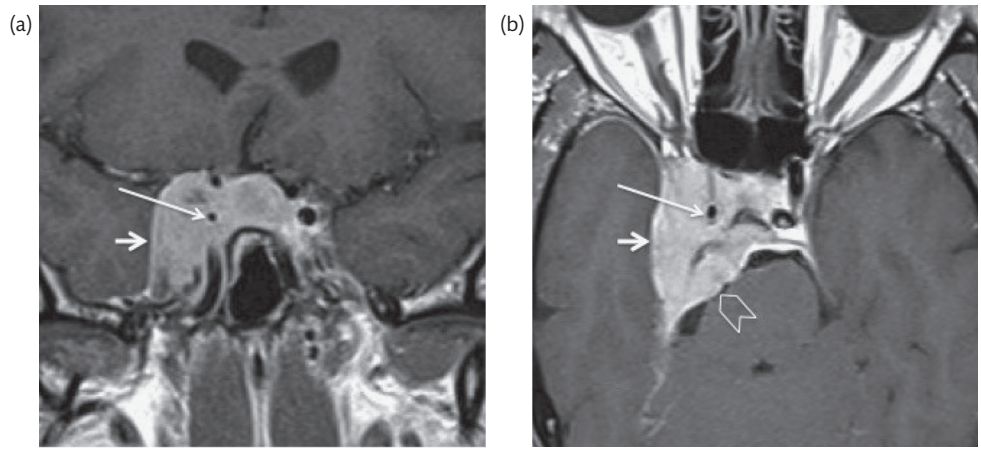


Figure 2.3.4.20 Meningioma of right cavernous sinus with intrasellar extension. Coronal (a) and axial (b) T1-W images. Bulging of the lateral wall of the right cavernous sinus (short arrow) and convexity of the upper pole of the sellar content; meningioma and normal pituitary gland are not differentiated. Severe reduction of the right intracavernous internal carotid artery (arrow). Posterior extension in the prepontine cistern (arrowhead).

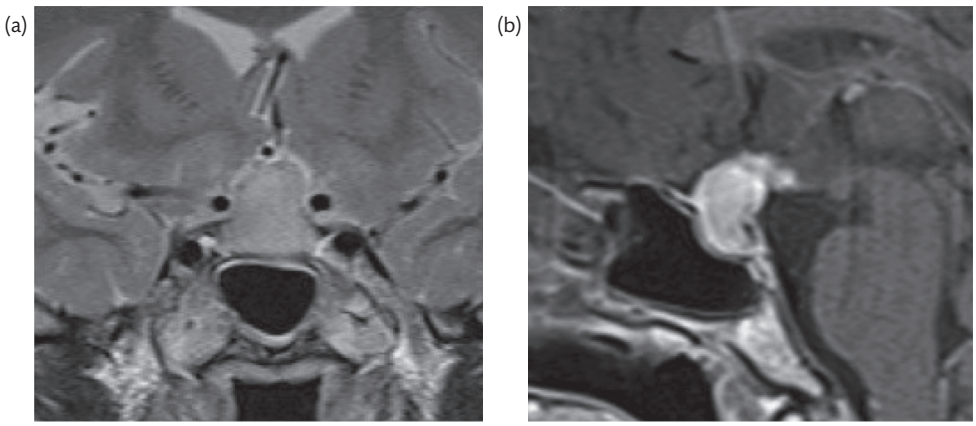


Figure 2.3.4.21 Pituicytoma. Coronal T2-W image (a) and coronal T1-W image after gadolinium (b). Hyperintense intrasellar mass with suprasellar expansion and strong homogeneous enhancement.

Primary Neurohypophyseal Glial Tumours

The pituitary stalk and the posterior pituitary gland, which are composed of specialized glial cells called pituicytes, can host neoplastic glial processes. We distinguish astrocytomas, tanycytomas, and granular cell tumours. Astrocytomas (pituicytomas and pilocytic astrocytomas) and tanycytomas are specific to the neurohypophysis, whereas granular cell tumours (GCT) can also be encountered elsewhere in the central nervous system.

Pituicytomas, formerly called infundibulomas, are considered grade I glial neoplasms by the 2007 World Health Organization (WHO) classification. They are seen in adults with a mean age of 50 years old without gender predilection. The main clinical signs are visual impairment, headache, hypopituitarism, and fatigue. Paradoxically, diabetes insipidus is an uncommon symptom in pituicytomas. On MRI scans, the normal bright spot on T1 of the posthypophysis is typically absent. The pituicytoma appears as a solid, well-circumscribed mass with a size ranging between 5 and 30 mm. It can be suprasellar, intra- and suprasellar, or rarely purely intrasellar, occasionally with cavernous invasion. This tumour is hyperintense on T2-W images, iso- or hypointense to grey matter on T1-W images, with typically an intense and homogeneous enhancement. Sellar enlargement is uncommon. The final diagnosis is made by immunohistochemistry and electron microscopy (Figure 2.3.4.21). A total surgical resection is curative with no recurrence or need for adjuvant therapy. **Pituitary pilocytic astrocytomas** are indistinguishable from pituicytomas on imaging. The differential diagnosis is made on anatomopathology.

Tanycytomas are hypothalamic–suprasellar tumours, clinically aggressive, with high recurrence rates. They are characterized by their large size, their encasement of the circle of Willis with possible infarction. They appear hyperintense on T2-W images, hypointense on T1-W images with cystic and solid components. The latter is intensely enhanced. Prognosis in children is worse than in adults.

GCTs, previously called granular cell myoblastomas or choristomas, are rare and benign tumours, considered WHO grade I glial neoplasms. They are usually diagnosed in the fifth decade and are twice as common in females than in males. In most cases, GCTs are asymptomatic. When symptomatic, the signs consist of visual disturbances, hypopituitarism, hyperprolactinemia, and headache. The appearance of GCT on MRI is non-specific, with an enhancing suprasellar or supra and intrasellar mass, which is isointense to grey matter on both T2- and T1-W sequences. Homogeneous or heterogeneous but intense enhancement is possible and reflects its high vascularity. Absence of the normal pituitary bright spot may be a clue for the neurohypophyseal origin, but this finding is not specific. Treatment for symptomatic GCT is surgical; postoperative radiation is controversial.

Germ cell tumours are discussed earlier [20].

Spindle Cell Oncocytomas

Spindle cell oncocytomas arise from the adenohypophysis. Visual disturbances, headache, and panhypopituitarism are the most common symptoms. Diabetes insipidus has never been reported. On MRI, they appear as intra- and suprasellar masses. It is difficult to distinguish them from pituitary adenomas or lymphocytic hypophysitis based on imaging features alone [20].

Perisellar Pseudotumours

Hamartomas of the Tuber Cinereum

Hamartomas of the tuber cinereum are not true tumoural lesions, but rather congenital malformations formed by the accumulation of normal neurons and glia in an unusual location, the tuber cinereum, located between the optic chiasm and the mammillary bodies. They are usually revealed by a very early precocious puberty, gelastic seizures; less frequently, the only clinical manifestation is a developmental delay in relation with isolated GH deficit or panhypopituitarism. There is no correlation between the clinical symptoms and the size of the hamartoma. On MR studies, the hypothalamic hamartoma appears as a round suprasellar mass, sessile or pedunculated. The size is variable, from a few millimetres to 5 cm in diameter. The signal is similar to that of grey matter on all sequences without enhancement after gadolinium injection (Figure 2.3.4.22). However, T2 hyperintense foci are occasionally seen in large hamartomas [21].

Aneurysms

Cerebral aneurysms are focal outpouchings of the wall of the cerebral arteries. They mainly develop close to the circle of Willis around the

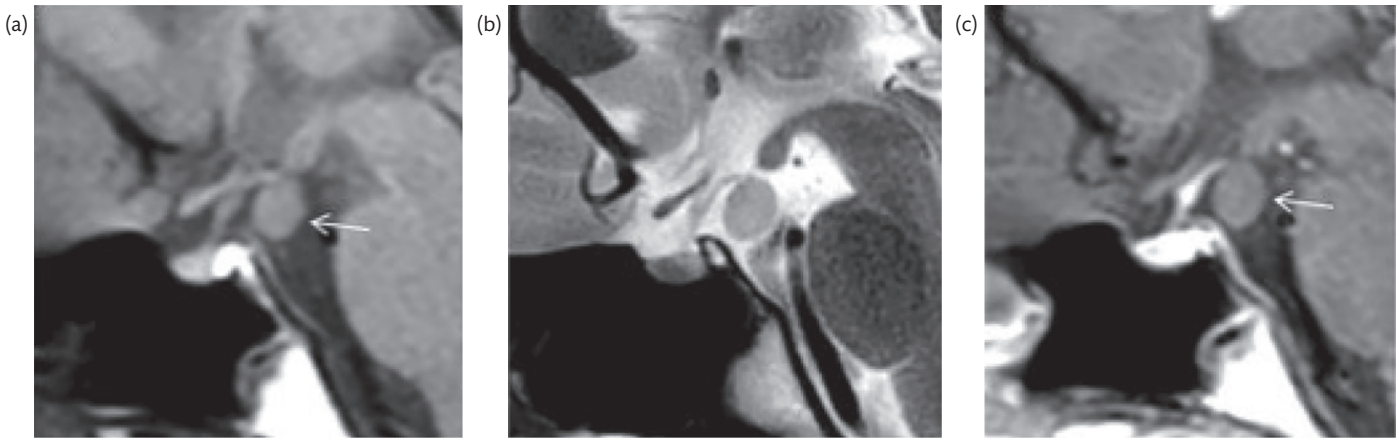


Figure 2.3.4.22 Precocious puberty in a 3-year-old male. Sagittal T1 (a), T2 (b) and T1 after gadolinium (c) W images. Large hamartoma of the tuber cinereum (arrow), between the optic chiasm and the mammillary bodies. No enhancement is noted.

sellar region and can be intra, supra, or laterosellar. They can exert mass effect on the pituitary gland, the stalk, or the cranial nerves or may be discovered incidentally. On MRI, they appear as well-delineated, round, hypointense lesions on all spin-echo sequences (Figure 2.3.4.23). They may enhance on gradient-echo sequences. Aneurysms may also be partially thrombosed, appearing heterogeneous on both T1- and T2-W images. An aneurysm may coexist with a pituitary adenoma incidentally or in a context of acromegaly in which the risk to develop an intracranial aneurysm is increased. When an aneurysm is suspected, it should be confirmed by magnetic resonance angiography before any biopsy or surgery [11].

Secondary Hypophysitis

Secondary hypophysitis can be due to systemic diseases (e.g. sarcoidosis, Langerhans' cell histiocytosis, Wegener granulomatosis) and systemic inflammatory processes and systemic infections (e.g. tuberculosis, syphilis). It can also be associated with a local process such as Rathke's cleft cyst rupture or with an autoimmune process in ipilimumab-induced hypophysitis.

Hypophysitis and Main Systemic Diseases

Systemic diseases affect multiple organs and sometimes associate sellar region involvement. They are numerous and encompass various autoimmune diseases, connective tissue diseases, granulomatosis, and vasculitis. They can be located within the anterior pituitary lobe and/or in the stalk or in the infundibulum or in the parasellar meninges, cavernous sinus, skull base, or neighbouring brain. The diagnosis is easy when the systemic disease is already known. Otherwise, the radiologist has to evoke the diagnosis on particular patterns suggestive of each disease. Here, granulomatous hypophysitis such as sarcoidosis, histiocytosis, or Wegener granulomatosis are detailed.

Sarcoidosis is a multisystemic granulomatosis, characterized by the presence of epithelioid granulomas, without caseous necrosis. Intracranial sarcoidosis lesions develop on the meninges, the brain, and the cranial nerves, with a predilection for the optic chiasm and hypothalamus-pituitary regions. Diabetes insipidus is the most frequent clinical symptom, followed by hyperprolactinemia and hypogonadism. On MRI, lesions of neurosarcoidosis are often multiple and diffused, affecting the

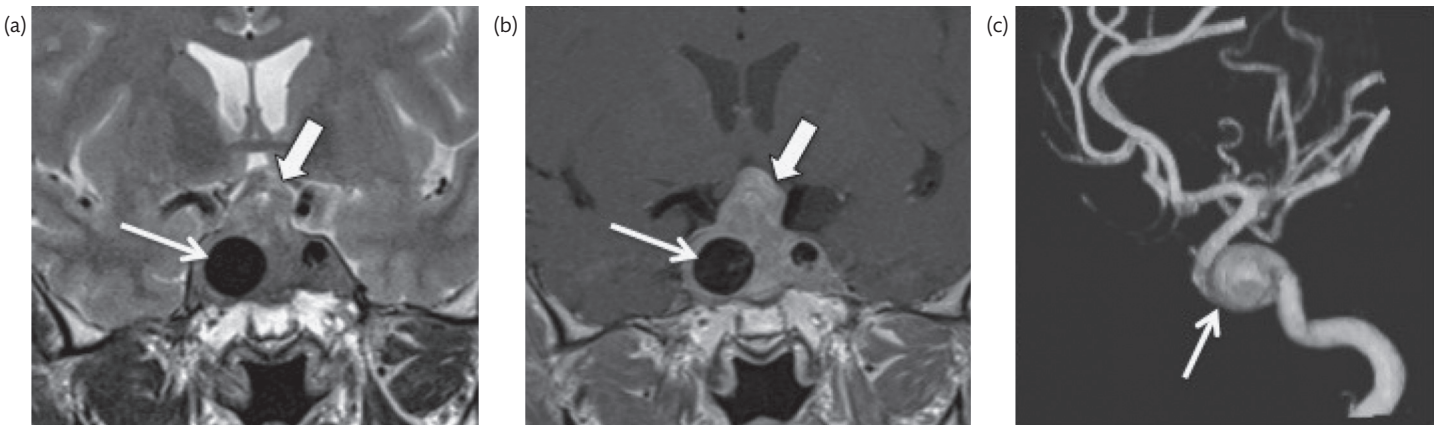


Figure 2.3.4.23 Right intracavernous internal carotid artery aneurysm with intra and suprasellar macroadenoma. Coronal T2 (a) and T1 spin-echo after gadolinium (b) W images. 3D volume rendering reformatted MRA(c). The aneurysm appears as a well-demarcated lesion within the right cavernous sinus and demonstrates marked hypointensity on T2 and T1 spin-echo after gadolinium W images (arrow). MRA confirms the flow in this aneurysm (arrow). Associated intra and suprasellar macroadenoma (thick arrow).

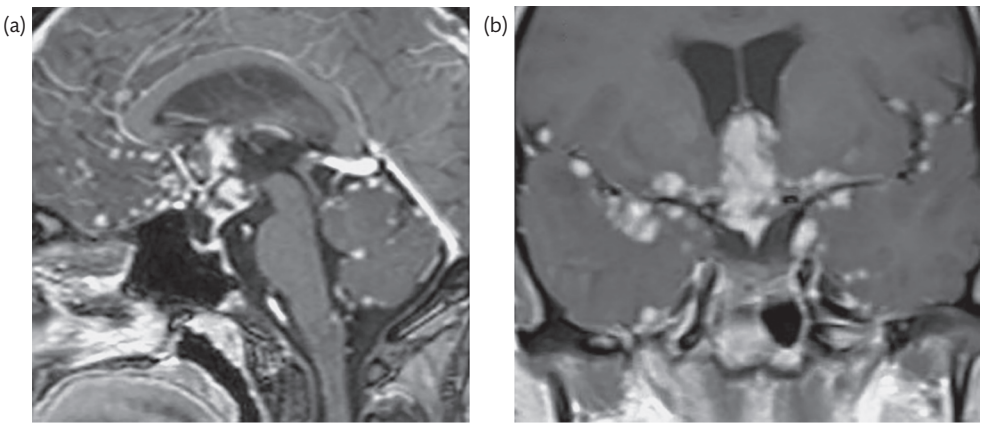


Figure 2.3.4.24 Neurosarcoidosis. Sagittal (a) and coronal (b) T1-W images after gadolinium. Typical leptomeningeal micronodular enhancing lesions along the hypothalamus-pituitary axis and the surface of the base of the brain.

meninges, the brain, and the cranial nerves at variable degrees. In the sellar region they can involve the hypothalamus, the infundibulum, the stalk, the adenohypophysis, the cavernous sinus, and the optic chiasm. The micronodular or linear gadolinium uptake along the surface of the brain, with a predilection for the suprasellar area and the base of the brain is highly suggestive (Figure 2.3.4.24). Intracerebral and spinal lesions are also frequently observed. All these lesions are presumed to regress partially or fully under treatment with corticosteroids or other immunosuppressive therapy [22].

Langerhans' cell histiocytosis. Different types of histiocytosis are associated with tissue accumulation of histiocytes: Langerhans cell histiocytosis (LCH) and non-Langerhans histiocytosis. LCH mainly develops in children aged 1–5 years. It is a multiorgan disease that mostly affects the cranial region, liver, spleen, lymph nodes, and bone. It also touches the endocrine system, principally the hypothalamic-pituitary axis, with a predilection for the pituitary stalk, causing anterior or both anterior and posterior hypopituitarism: diabetes insipidus, hyperprolactinemia, hypopituitarism, and hypogonadotropic hypogonadism. The MRI can reveal the absence of the posterior pituitary bright spot with or without an associated suprasellar abnormality: stalk enlargement (Figure 2.3.4.25a), infundibular thickening (Figure 2.3.4.25b), small nodule, or large hypothalamic mass. As in sarcoidosis, such lesions exhibit T2 hypointensity, T1 isointensity, and a marked enhancement after gadolinium. A peripheral dark rim surrounding the adenohypophysis may be observed on T2-W images. This sign, described in lymphocytic hypophysitis, may be consistent with fibrosis. Specific additional lesions may help to evoke the diagnosis: supratentorial intra-axial masses and cerebellar hemisphere hyperintensity in all histiocytosis, infiltrative lesion in the pons and lytic lesions of the vault in LCH and intraorbital enhancing mass lesions and osteosclerosis of the walls of the sinus in Erdheim–Chester disease. Radiation therapy, chemotherapy, or a combination of both are usually proposed to patients with histiocytosis [23].

Wegener granulomatosis is a granulomatosis with polyangiitis (GPA) associating necrotizing small vessel vasculitis and giant cell necrotizing extravascular granulomas. It is known to involve the respiratory tract, the kidneys, and the ENT area. Pituitary involvement is usually a late manifestation causing diabetes insipidus and gonadotropin deficiency. MRI will reveal pituitary enlargement mimicking an adenoma, stalk thickening or loss of the posterior lobe pituitary bright spot [24].

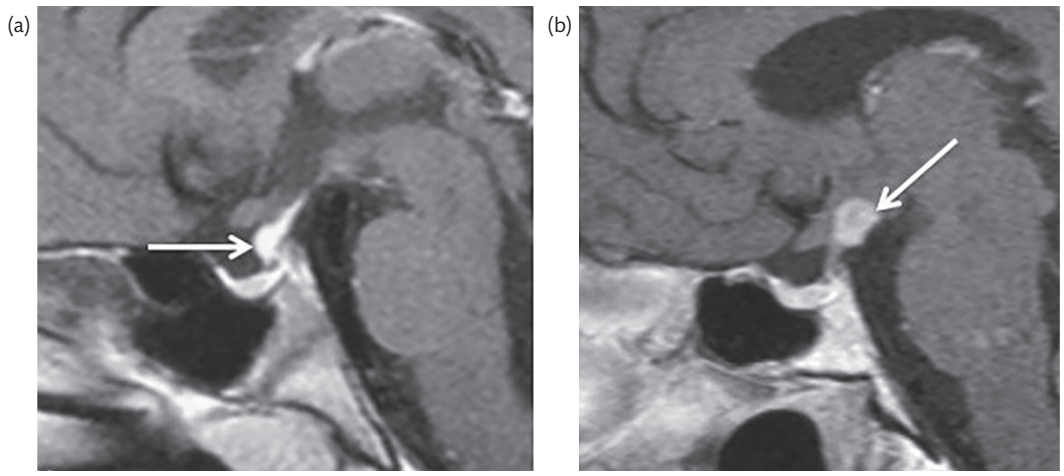


Figure 2.3.4.25 Histiocytosis in two different patients with diabetes insipidus. Sagittal T1-W images after gadolinium. Stalk enlargement (a), infundibular thickening mass lesion (b) which enhance homogeneously after gadolinium.

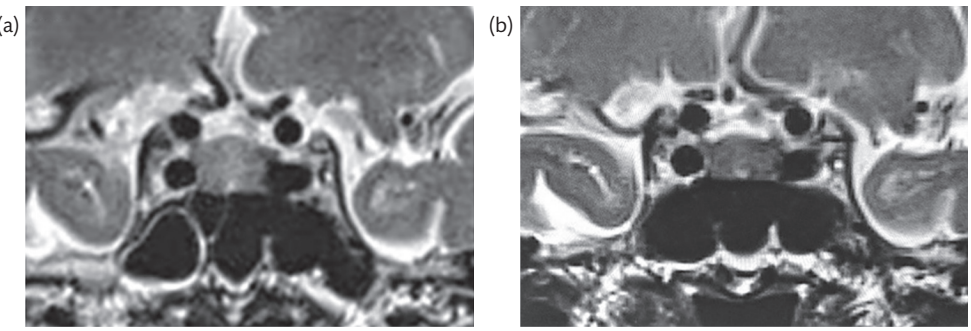


Figure 2.3.4.26 Immunotherapy induced hypophysitis in the acute phase (a) and one month after replacement therapy (b) on T2-W images. Enlargement and T2 hyperintense signal of the pituitary gland resolve quickly.

Ipilimumab-Induced Hypophysitis

Ipilimumab-induced secondary hypophysitis [25] is a newly described entity induced by administration of ipilimumab which is a human monoclonal antibody blocking cytotoxic T-lymphocyte antigen 4 (CTLA-4). This treatment is efficient in certain types of cancer, but with some side effects including enterocolitis, hepatitis, uveitis, dermatitis, arthritis, and different endocrinopathies such as hypophysitis that can lead to panhypopituitarism. MRI findings are not specific, characterized by a diffuse enlargement of the pituitary gland with homogeneous or less commonly heterogeneous enhancement. Follow-up imaging shows resolution after drug cessation, steroid therapy, and hormone replacement (Figure 2.3.4.26). Endocrinologists, neurooncologists, and neuroradiologists should be aware of the potential side effects of ipilimumab to induce immune-mediated hypophysitis [25].

Primary Hypophysitis

Primary hypophysitis comprises five histologic forms: lymphocytic, granulomatous, xanthomatous, necrotizing, and IgG4 plasmacytic. The two main types are lymphocytic and granulomatous hypophysitis. IgG4-related hypophysitis is a newer variant [26].

The role of MRI in primary hypophysitis is to eliminate the main alternative diagnosis of a pituitary mass, which is adenoma. Diffuse and symmetric enlargement of the pituitary gland is seen in **adenohypophysitis,** with frequent superior extension. This tumoural syndrome contrasts with a normal-sized pituitary fossa. Moreover, the pseudocapsule formed by the compressed normal pituitary gland is not seen after gadolinium administration. Signal intensity on T2- and T1-weighted images before and after contrast administration is usually homogeneous and intense but can be moderate and heterogeneous (Figure 2.3.4.27). A non-specific perisellar dural enhancement can be seen. Another more specific but later sign has been described in lymphocytic hypophysitis, appearing 2 to 20 months after the initial MR exam. It consists of a dark rim around of the gland and in the cavernous sinuses on T2-weighted images, due to fibrosis (Figure 2.3.4.28). Anterior pituitary atrophy with an empty sella is the common final outcome of adenohypophysitis. Thickened pituitary stalk and loss of the posterior pituitary bright spot on the precontrast images are seen in **infundibuloneurohypophysitis**. At a later stage, pituitary stalk thickening usually resolves and atrophy takes place, with slow decrease in size of the anterior lobe. However, recurrence of the stalk thickening can be seen. Loss of the posterior lobe bright spot is usually permanent. In cases of **panhypophysitis**, signs of both entities will be seen. Exploration of the brain and spinal cord is recommended, as it could show other signs that will help guide diagnostic orientation. Primary hypophysitis management

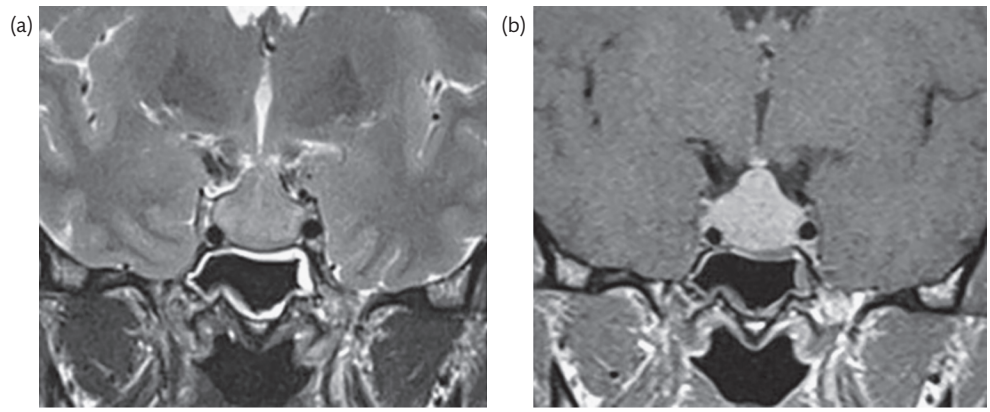


Figure 2.3.4.27 Primary adenohypophysitis. Coronal T2 (a) and T1 after gadolinium (b) W images. Diffuse and symmetric enlargement of pituitary gland in a normal-sized pituitary fossa with suprasellar extension and optic chiasm compression. The mass is slightly hyperintense on T2 with a homogeneous and moderate enhancement. No normal pituitary gland is seen.

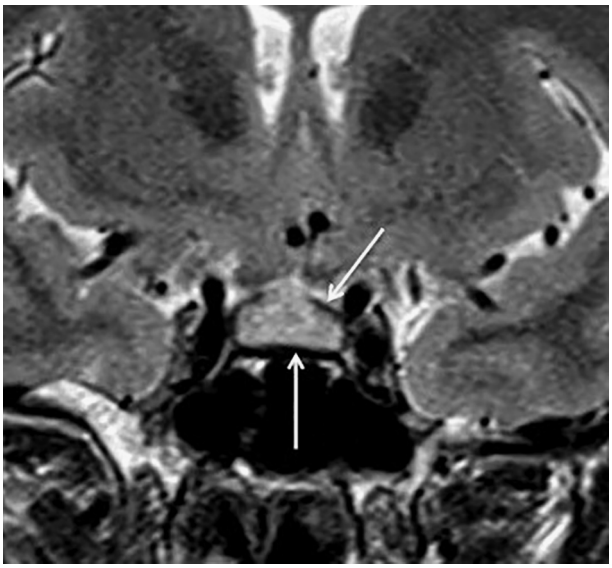


Figure 2.3.4.28 Lymphocytic adenohypophysitis 6 months after acute episode. Coronal T2-W image. Hypointense encircling the pituitary gland.

should be done under clinical, biological, and neuroradiological supervision. The main differential diagnoses are pituitary adenoma and non-adenomatous pituitary masses such as germinomas, choristomas, and pituitary metastases and pituitary hyperplasia secondary to hypothyroidism.

Distinguishing primary hypophysitis from secondary hypophysitis is a diagnosis of exclusion. Investigations will have to rule out the numerous aetiologies cited earlier.

Pituitary Hyperplasia, Pregnancy, and Primary Hypothyroidism

The normal pituitary gland enlarges gradually during pregnancy because of oestrogen-stimulated hyperplasia and hypertrophy of prolactin cells. Its height increases linearly by 0.08–0.1 mm/week (i.e. by 3–4 mm at term) and can reach 10 mm during the last trimester and 12 mm in the immediate postpartum period, and approach the optic chiasm. At the same time, T1 signal intensity of the pituitary gland increases, particularly during the last trimester; this is clear when comparing the T1 signal of the anterior pituitary with that of the temporal white matter.

Pituitary enlargement in primary hypothyroidism in adults and children is rare, and can result from both acquired and congenital hypothyroidism. On imaging studies, pituitary hyperplasia is characterized by a spherical enlargement of the gland with suprasellar extension. The T2 signal can be slightly hyperintense; enhancement is homogeneous after gadolinium injection. Shrinkage of the mass occurs very quickly, as soon as a few weeks after thyroxine hormone replacement therapy [27].

Sheehan Syndrome

Sheehan syndrome corresponds to pituitary necrosis following severe postpartum haemorrhage and hypovolemia [28]. It may manifest itself very early after delivery, with headache and hypopituitarism responsible for lack of milk production immediately or some weeks later. More frequently, Sheehan syndrome is diagnosed years later. In the following days after the ischaemic infarct, MRI typically shows an enlarged, non-haemorrhagic pituitary gland, hypointense on T1-W images, heterogeneous with hyperintense foci on T2-W images and a peripheral enhanced rim. Afterwards, the pituitary gland becomes atrophic, making the passage of CSF within the sella possible, which leads to the so-called empty sella.

REFERENCES

1. Molitch ME. Nonfunctioning pituitary tumors. *Handb Clin Neurol*, 2014; **124**: 167–84.
2. Cao L, Chen H, Hong J, *et al*. Magnetic resonance imaging appearance of the medial wall of the cavernous sinus. *J Neuroradiol*, 2013; **40**: 245–51.
3. Briet C, Salenave S, Chanson P. Pituitary apoplexy. *Endocrinol Metab Clin North Am*, 2015; **44**: 199–209.
4. Bonneville JF, Bonneville F, Schillo F, *et al*. Follow-up MRI after transsphenoidal surgery. *J Neuroradiol*, 2003; **30**: 268–79.
5. Stofko DL, Nickles T, Sun H, Dehdashti AR. The value of immediate postoperative MR imaging following endoscopic endonasal pituitary surgery. *Acta Neurochir (Wien)*, 2014; **156**: 133–40.
6. Bonneville JF, Bonneville F, Cattin F. Magnetic resonance imaging of pituitary adenomas. *Eur Radiol*, 2005; **15**: 543–8.
7. Iglesias P, Bernal C, Villabona C, *et al*. Prolactinomas in men: a multicentre and retrospective analysis of treatment outcome. *Clin Endocrinol (Oxf)*, 2012; **77**: 281–7.
8. Potorac I, Petrossians P, Daly AF, *et al*. T2-weighted MRI signal predicts hormone and tumor responses to somatostatin analogs in acromegaly. *Endocr Relat Cancer*, 2016; **23**: 871–81.
9. Bonneville JF, Cattin F, Bonneville F, *et al*. Pituitary gland imaging in Cushing disease. *Neurochirurgie*, 2002; **48**: 173–8.
10. Cazabat L, Dupuy M, Boulin A, *et al*. Silent, but not unseen: multimicrocystic aspect on T2-weighted MRI in silent corticotroph adenomas. *Clin Endocrinol*, 2014; **81**: 566–72.
11. Bonneville F, Cattin F, Marsot-Dupuch K, *et al*. T1 signal hyperintensity in the sellar region: spectrum of findings. *Radiographics*, 2006; **26**: 93–113.
12. Bonneville JF. Rathke cleft cyst: symptomatic and complicated. In: Bonneville J-F, Bonneville F, Cattin F, Nagi S, eds. *MRI of the Pituitary Gland.* Cham: Springer, 2016; 135–44.
13. Choi SH, Kwon BJ, Na DG, *et al*. Pituitary adenoma, craniopharyngioma, and Rathke cleft cyst involving both intrasellar and suprasellar regions: differentiation using MRI. *Clin Radiol*, 2007; **62**: 453–62.
14. Raelson C, Chiang G. Chiasmatic-hypothalamic masses in adults: a case series and review of the literature. *J Neuroimaging*, 2015; **25**: 361–4.
15. Kanagaki M, Miki Y, Takahashi JA, *et al*. MRI and CT findings of neurohypophyseal germinoma. *Eur J Radiol*, 2004; **49**: 204–11.
16. Saloner D, Uzelac A, Hetts S, Martin A, Dillon W. Modern meningioma imaging techniques. *J Neurooncol*, 2010; **99**: 333–40.
17. He W, Chen F, Dalm B, *et al*. Metastatic involvement of the pituitary gland: a systematic review with pooled individual patient data analysis. *Pituitary*, 2015; **18**: 159–68.

18. Kaufmann T, Lopes MB, Laws ER, *et al.* Primary sellar lymphoma: radiologic and pathologic findings in two patients. *Am J Neuroradiol*, 2002; **23**: 364–67.
19. Nagi S, Mahmoud M. Sellar and suprasellar lymphoma. In: Bonneville J-F, Bonneville F, Cattin F, Nagi S, eds. *MRI of the Pituitary Gland*. Cham: Springer, 2016; 235–39.
20. Covington MF, Chin SS, Osborn AG. Pituicytomas, spindle cell oncocytoma and granular cell tumor. Clarification and meta-analysis of the world literature since 1893. *Am J Neuroradiol*, 2011; **32**: 2072.
21. Martin DD, Seeger U, Ranke MB, Grodd W. MR imaging and spectroscopy of a tuber cinereum hamartoma in a patient with growth hormone deficiency and hypogonadotropic hypogonadism. *Am J Neuroradiol*, 2003; **24**: 1177–80.
22. Langrand C, Bihan H, Raverot G, *et al.* Hypothalamo-pituitary sarcoidosis: a multicenter study of 24 patients. *Q J Med*, 2012; **105**: 981–95.
23. Kurtulmus N, Mert M, Tanakol R, Yarman S. The pituitary gland in patients with Langerhans cell histiocytosis: a clinical and radiological evaluation. *Endocrine*, 2014; **48**: 949–56.
24. De Parisot A, Pue'chal X, Langrand C, *et al.* Pituitary involvement in granulomatosis with polyangiitis. Report of 9 patients and review of the literature. *Medicine*, 2015; **94**: e748.
25. Carpentier KJ, Murtagh D, Lilienfeld H, *et al.* Ipilimumab-induced hypophysistis: MR imaging findings. *Am J Neuroradiol*, 2009; **30**: 1751–3.
26. Khare S, Jaktab VS, Budyal SR, *et al.* Primary (autoimmune) hypophysitis: a single center experience. *Pituitary*, 2015; **18**: 16–22.
27. Kanza RE, Gagnon S, Villeneuve H, *et al.* Spontaneous ovarian hyperstimulation syndrome and pituitary hyperplasia mimicking macroadenoma associated with primary hypothyroidism. *World J Radiol*, 2013; **5**: 20–4.
28. Kaplun KJ, Fratila C, Farenczi A, *et al.* Sequential pituitary MR imaging in Sheehan syndrome: report of two cases. *Am J Neuroradiol*, 2009; **30**: 1751–3.

2.3.5 Hypopituitarism

Replacement of Adrenal, Thyroid, and Gonadal Axes

Miles J. Levy, Ragini Bhake, and Narendra Reddy

Adrenal Replacement in ACTH Deficiency

Choice and Timing of Glucocorticoid Replacement

Hydrocortisone, the generic pharmaceutical name for cortisol, is the standard form of glucocorticoid replacement for ACTH deficiency, and directly replaces the missing active hormone. Cortisone acetate was previously widely used, but is metabolized to cortisol to achieve its glucocorticoid activity, so that its onset of action is slower than hydrocortisone (a slight disadvantage) and biological half-life slightly longer (potentially a slight advantage).

The normal pattern of diurnal cortisol secretion is difficult to mimic precisely with oral therapy and there is no universal agreement regarding the appropriate dose, timing, and monitoring of hydrocortisone replacement, although the need for close attention has been highlighted. Cortisol diurnal rhythm in normal individuals is regulated by the suprachiasmatic nucleus in hypothalamus (the body clock). Cortisol rises between 03.00 hrs and 04.00 hrs, peaks within an hour of waking, falls over the day with a quiet period from 19.00 hrs, and nadirs at 00.00 hrs [1]. In addition to this diurnal variation, variable peaks of cortisol secretion (ultradian rhythm) may occur due to other factors such as in response to stress, meal, and exercise. There is currently no ideal glucocorticoid regime available to mimic this rhythm. The best approximation to this pattern is with hydrocortisone in three divided doses per day: 10 mg of hydrocortisone on awakening, 5 mg at lunch and 5 mg in the early evening [2]. This is widely believed to achieve 'physiological' plasma cortisol levels compared to a traditional twice daily regime, which usually results in very low cortisol levels in late afternoon before the evening dose.

Patients with ACTH deficiency may still experience symptoms of hypoadrenalism on hydrocortisone leading to development of two compounds namely, Duocort® or Plenadren® [3] and Chronocort® [4], purported to mimic normal circadian rhythm more closely. Plenadren® is a dual-release once daily hydrocortisone formulation (available in 20 mg and 5 mg tablets), currently available for prescription in the United Kingdom and rest of Europe. It has an immediate release coating with an extended release core obviating the need for midday dosing when taken first thing in the morning. It has been shown to reduce cardiovascular risk factors (weight, HbA1c, and blood pressure) and fatigue [5]. Its prime advantage, due to 20% lower bioavailability than conventional hydrocortisone, may include reduction in cardiovascular risk factors, although dose adjustment may be required. The main limitation however is same as that for hydrocortisone as it too can only be taken after waking. Chronocort®, on the other hand, with its pH-sensitive enteric coated modified-release multiparticulate hydrocortisone capsule formulation, is reported to have an edge in this regard, it provides an overnight cortisol rise and peak around awakening. It is taken twice daily, 20 mg at 23.00 hrs and 10 mg 07.00 hrs. The potential clinical advantage of this compound may be alleviation of early morning fatigue experienced by some of these patients.

Another potentially useful route of hydrocortisone administration, currently under evaluation, is continuous subcutaneous hydrocortisone infusion (CSHI) via a programmable pump [6]. This has been used with variable success, and needs more evidence and safety data prior to routine clinical use.

Some groups have advocated the use of low-dose prednisolone (3–4 mg daily) instead of hydrocortisone in patients with hypoadrenalism. Prednisolone has a longer half-life, given as a once daily preparation, and may have a comparable cardiovascular risk profile to hydrocortisone and provide better patient satisfaction [7]. The lack of widely available prednisolone assays for monitoring purposes along with reported adverse cardiovascular and bone health outcomes [8], has led to some concerns about the use of prednisolone, although this may simply be due to higher doses of prednisolone being used in these studies. Currently, it may be reasonable to use prednisolone 3–5 mg/day in patients who have difficulty in adhering to multiple-dose regimen, non-availability of hydrocortisone, or in patients who have a poor quality of life despite optimization of their hydrocortisone regime [9]. Dexamethasone should be avoided as a replacement glucocorticoid due to the high incidence of adverse glucocorticoid effects.

Assessment of Hydrocortisone Replacement

Conventionally, glucocorticoid replacement with hydrocortisone is monitored using plasma cortisol measurements at multiple times throughout the day—a hydrocortisone day curve (HCDC). Studies using frequent sampling for plasma cortisol have identified wide interindividual variations in plasma cortisol levels obtained after the same dose of hydrocortisone and highlighted the need for individual adjustment of hydrocortisone dose, but such frequent sampling is rarely possible or necessary in routine practice.

One such conventional practice is to monitor hydrocortisone replacement with a simple HCDC involving collection of a 24 h urine for free cortisol (UFC) on the day prior to the test, and three plasma cortisols during a day-case attendance [2]. However, this strategy is now not recommended as the practice was based on underpowered short term-studies; given lack of objective evidence, guidelines currently advocate basing hydrocortisone dose adjustments on clinical criteria alone [9, 10].

The criteria for deciding optimum hydrocortisone regimes are inevitably a compromise between theory, practicality, and patient convenience. A thorough history of patients' daily habits should be accounted for dose adjustments, including sleep timings, shift-working nature, dips in energy, daytime somnolence, and general well-being. Clinical markers of over-replacement are weight gain, insomnia, and peripheral oedema. Markers of insufficient replacement include nausea, lethargy, postural dizziness, weight loss, and lack of appetite.

Even thrice daily regimes fail to accurately mimic the normal physiological circadian rhythm of cortisol and attention has been focused on the differences in circulating cortisol levels overnight in patients on hydrocortisone replacement (where levels are low or undetectable throughout the night) compared to normal individuals (where levels rise substantially during the last hours of sleep). These differences would be hard to avoid using standard preparations of hydrocortisone, since few patients would be prepared to wake to take a tablet during the night, but certainly represent an unphysiological feature of current replacement regimes, may contribute to an overnight deficiency of a variety of metabolic fuels and is potentially an area where the newer delayed preparations described earlier, particularly Chronocort®, may have a role. For now, in practical terms, patients should be advised to take hydrocortisone as soon as possible after wakening to avoid prolonged activity with low circulating cortisol levels, and those who habitually wake in the early hours some time before rising might benefit from taking their hydrocortisone dose at that time. Patients with ACTH deficiency should not require mineralocorticoid replacement, since the renin-angiotensin-aldosterone axis is not disrupted by pituitary disease.

Replacement During Intercurrent Illness

During intercurrent illness, glucocorticoid replacement is mandatory, and doses need to be increased for all but the most minor illness in order to mimic the normal increase in ACTH and cortisol secretion, which occurs during stress and illness. Appropriate patient education on this aspect of replacement is a vital part of management of hypoadrenalism. Patients must be advised to double their normal oral dose of hydrocortisone during common pyrexial illnesses, and understand the need for parenteral glucocorticoid replacement if illness, operation, vomiting, or diarrhoea prevents the effective administration or absorption of oral glucocorticoid. Patients should seek medical advice if symptoms worsen in spite of increased oral hydrocortisone, should keep an 'emergency' ampoule of hydrocortisone at home, and if possible they or their family should be taught to give the injection if medical help is unavailable. Some patients also find a symptomatic need for increased glucocorticoid replacement during psychological stress, but this is much more difficult to define or regulate. During severe intercurrent illness or major surgery, immediate parenteral injection of hydrocortisone 100 mg (50 mg/m^2 for children) should be administered, followed by parenteral hydrocortisone 200 mg/day in divided doses every 6 hourly [9]. This regime is noted to provide consistent, high levels of circulating cortisol comparable with those found in normal individuals during such stress in cases of primary adrenal insufficiency. Some studies indicate lower hydrocortisone dose requirements (25–75 mg/24 hrs) in the first 24 hours after non-pituitary surgery, in secondary adrenal insufficiency. However, the former regime is recommended based on common practice, as it is the safer approach. Intravenous boluses of hydrocortisone produce wide swings in cortisol levels and are therefore less desirable, but stable, high plasma cortisol levels can be achieved by an intravenous infusion of hydrocortisone 5 mg/hr (preceded by a 25 mg intravenous bolus), although this is only appropriate in circumstances where an intravenous infusion can be reliably maintained and monitored.

Patient support groups have developed clear formatted guidance on replacement for intercurrent illness and during surgery and other procedures, which are readily available on the internet and useful for providing advice to patients and surgical colleagues [11].

In summary, following should be made mandatory for care providers in all healthcare settings dealing with steroid dependent patients:

1. Verbal and written patient education information should be provided about drug compliance, risk of hypoadrenal crisis, steroid sick-day rules, and emergency hydrocortisone injection technique at each clinical encounter.
2. Patients should be encouraged to carry an emergency card, wear a medical alert bracelet, or pendant, and carry a rescue-pack of hydrocortisone at all times.

3. Patients should be issued an emergency intramuscular hydrocortisone kit designed for self-injection; enquired if there is a need for reissue in case of expired ampoules; patient confidence and satisfaction with their injection technique should be ascertained at each consultation.

Adverse Effects of Hydrocortisone Replacement

It is challenging to balance adverse effects of chronic supraphysiological glucocorticoid exposure, which are serious and well known, against those of deficiency. In theory perfect physiological replacement should lead to no adverse effects, since circulating cortisol levels would be no different from normal subjects, but close attention to replacement doses is essential in order to achieve this.

Gross Cushingoid side effects and symptoms of severe hypoadrenalism are usually clinically obvious and most endocrinologists can avoid such extremes of inappropriate replacement. Minor degrees of over- or under-replacement may be clinically undetectable, resulting in morbidity (poor quality of life, cardiovascular risks, osteoporosis) or even mortality. Several studies support this view: glucose tolerance and insulin secretion alter with hydrocortisone replacement, and blood pressure rises with replacement therapy. The European Adrenal Insufficiency Registry data suggest that the type of glucocorticoid, dose, and regimen are very variable. One in eight patients are noted to be on hydrocortisone ≥30 mg/day, a cut-off level that increases the risk of endothelial dysfunction, higher arterial stiffness, poor quality of life, and dose-related high cardiovascular and metabolic consequences [12].

Therefore, if minor over-replacement caused slight worsening cardiovascular risk factors such as glucose intolerance, central obesity, or blood pressure, then this might be undetectable in an individual yet have a significant influence on overall cardiovascular morbidity. Studies of cortisol production rates using stable nucleotides show daily secretion of 5–7 mg/m^2, equivalent of 9–11 mg/day of hydrocortisone [13]. Unlike Addison's disease, where severe cortisol deficiency is noted, hypopituitary individuals often have some residual cortisol secretion. Hydrocortisone 10 mg am/5 mg pm regimen showed improved quality of life, beneficial cardiovascular profile secondary to weight loss (average 7 kg weight loss when reduced from higher doses), reduced arterial stiffness, and a more physiological nocturnal blood pressure dip. In summary, it appears the lower the hydrocortisone dose the better in terms of metabolic and cardiovascular risk. It may be appropriate to use doses of hydrocortisone as low as 10 mg/day in patients with partial adrenal insufficiency while monitoring for hypothalamic–pituitary–adrenal (HPA) axis recovery.

It is well recognized that bone health is adversely affected with overexposure of glucocorticoids. Increased incidence of vertebral fractures was noted in hypopituitarism individuals on hydrocortisone 30 mg/day dose, despite correction of gonadotrophin deficiency. Reduction of hydrocortisone to 15 mg/day showed increased bone formation and better bone remodelling ability [14].

In summary, there is a need to avoid even subclinical glucocorticoid over-replacement and aim for the lowest total dose of hydrocortisone replacement compatible with good health. Conversely avoidance of very low cortisol levels before the next dose seems advisable to minimize the risk of hypoadrenalism if intercurrent illness or stress occurs at that time.

Thyroid Replacement in TSH Deficiency

Choice of Replacement Therapy

The synthetic version of endogenous human thyroid hormone, levothyroxine is the routine replacement used for treatment of thyroid-stimulating hormone (TSH) deficiency. Its long plasma half-life (6–10 days) ensures stable levels of thyroid hormones on once daily administration, and conversion to T_3 *in vivo* results in appropriate blood levels of both T_4 and T_3. Liothyronine (T_3; triiodothyronine) can be used, but has no advantage in most circumstances, and may raise safety concerns due to elevated free T_3 levels indicating supraphysiological T_3 exposure. Combined T_4 and T_3 replacement and the use of 'natural' thyroid extracts has been advocated in print and on the internet by a variety groups, but clinical trial evidence is limited and where available suggests no benefit [15].

Commencing Levothyroxine Replacement

The starting dose and regime of levothyroxine replacement depends on the clinical circumstances. Prior to commencing levothyroxine, it is essential to know the status of the ACTH-adrenal axis since starting levothyroxine without glucocorticoid replacement in a patient with severe ACTH deficiency may precipitate a hypoadrenal crisis. If ACTH deficiency is present, hydrocortisone must be started before levothyroxine. Thereafter, many patients with TSH deficiency have serum free T_4 levels only slightly below the reference range, and in patients with such mild deficiency and no evidence of cardiovascular disease replacement can be simply commenced with a near-full replacement dose of levothyroxine 100 μg once daily. This dosage is comparable to the study that increased the dose from 1 mcg/kg/day to 1.6 mcg/kg/day, which led to weight loss, alleviation of hypothyroid symptoms, reduced body mass index (BMI), low density lipoprotein (LDL) cholesterol and serum creatine kinase levels. In patients with more profound reductions of serum free T_4, a lower starting dose of 50 μg daily, increased after a few weeks to 100 μg may be better tolerated. In elderly patients, or any patient with known cardiovascular disease—particularly ischaemic heart disease, more caution is required: in most cases a starting dose of 25 mcg will be well tolerated, increased slowly in 25 mcg increments over several weeks until the target dose is achieved.

Monitoring Levothyroxine Replacement

Defining the optimal replacement dose of levothyroxine in TSH deficiency is problematic, and little scientific evidence is available to guide the clinician. Unlike primary hypothyroidism, where serum TSH is a sensitive marker of under- or over-replacement, there is no biochemical marker to indicate precise physiological levels of replacement for an individual patient—indeed serum TSH may be low, normal, or even slightly elevated in untreated TSH deficiency. Therefore, instead of relying on 'reflex TSH strategy', adjustment is based on the clinical response and on measurement of circulating thyroid hormone levels, which are

limited by the very wide reference ranges in the normal population [16]. Serum free T_4 appears the most appropriate marker with which to adjust the levothyroxine dose and the conventional recommendation is to maintain free T_4 in the middle or upper part of the reference range for normal individuals, although this is certainly not 'evidence-based' and begs the question of the criteria used to define TSH deficiency since it implies that patients with pituitary disease (and indeed 50% of the normal population!) with a free T_4 in the lower half of the normal range might benefit from levothyroxine replacement therapy. Some workers have advocated using a levothyroxine dose based on body weight (1.6 μg/kg) but in doing so increased free T_4 levels close to the upper limit of the reference range.

On comparison of free T4 levels of secondary hypothyroid patients to that of patients with primary thyroid disease in our centre, it was found that TSH deficiency was substantially underdiagnosed and under-treated [17]. The twentieth to eightieth centile range for controls on levothyroxine was a free T_4 level of 14–19 pmol/L (normal range 9–26 pmol/L) in our laboratory (in general, middle third of the range) and we propose to use this as target range for replacement levels in TSH deficiency in future.

Ultimately, such controversies regarding appropriate levothyroxine replacement dosage and monitoring can only be answered by a controlled trial. In the meantime, a serum free T_4 anywhere in the middle centiles of the normal range when the patient is asymptomatic is accepted, with an option to move the free T_4 into the upper part of the reference range if the patient experiences symptoms suggestive of hypothyroidism.

Adverse Effects of Levothyroxine Replacement

Although thyrotoxicosis of the primary gland is a well-documented risk factor for osteoporosis, bone density remains normal even in patients on deliberate supraphysiological replacement with levothyroxine. However, thyrotoxicosis is also a risk factor for cardiovascular disease and, although there is no direct evidence of such adverse effects of levothyroxine replacement, the association of suppressed TSH levels and risk of atrial fibrillation in older patients in population-based studies indicates the need for caution to avoid unnecessary over-replacement.

Gonadal Steroid and Gonadotrophin Replacement in Gonadotrophin Deficiency

In the context of hypogonadism, consideration has to be given to both sex-hormone replacement and fertility potential, while accounting for individual's choice of route of administration, timing, social, and cultural beliefs. The beneficial evidence noted for sex-hormone replacement therapy in primary hypogonadism is extended to secondary hypogonadism, as there is a lack of direct evidence in the latter group. Unlike adrenal and thyroid replacement, replacement of gonadotrophin deficiency offers an extensive choice of gonadal steroid therapy for both sexes, and the choice of gonadotrophin or gonadotrophin-releasing hormone (GnRH) therapy if and when fertility is desired. In most cases, the choice between different treatment modalities is largely a matter of patient and physician preference rather than being 'evidence-based'.

Female: Choice of Replacement Regime

All female patients of premenopausal age with gonadotrophin deficiency require oestrogen replacement to maintain bone mass, to avoid vasomotor symptoms of oestrogen deficiency (hot flushes and night sweats), and to improve vaginal atrophy, urinary frequency and dysuria [18]. Cyclical progestogen is also essential in patients with an intact uterus to avoid endometrial hyperplasia and neoplasia; oestrogen alone suffices in women who have had a previous hysterectomy.

Oral

The choice of oral replacement is extensive, including all forms of oral oestrogen marketed for postmenopausal oestrogen deficiency and all combined oestrogen-progesterone contraceptive pills, with little or no objective evidence to choose between regimes. In younger women the low dose (20–35 mcg ethinyloestradiol) oestrogen-progesterone pill is often preferred since there are extensive data on its safety in long-term use in women of this age, particularly in context of hypogonadism. It may also be preferred due to its lower cost, and acceptability as in some cases taking the 'pill' may feel more 'normal' psychologically than taking hormone replacement therapy (HRT). 17-β-oestradiol containing preparations can potentially be monitored by blood oestrogen measurements, but this is unreliable with most preparations, therefore not recommended. Most preparations and regimes provide adequate levels of oestrogen to avoid effects of deficiency, and in an individual patient a regular menstrual withdrawal bleed is considered an adequate bioassay of oestrogen effect. When future pregnancy is planned it may be appropriate to monitor uterine size and endometrial thickness by ultrasound to ensure that the uterus is adequately oestrogenized.

Transdermal

A variety of transdermal oestradiol delivery systems are available—some of which also deliver transdermal progestogen. Gel preparations are available but have not been widely used.

Such regimen have the advantage of achieving near normal physiological levels of natural 17-β-oestradiol with administration of lower doses as they bypass the first pass metabolism and directly enter the blood stream. Disadvantages include skin reactions, unsatisfactory skin adherence in some patients, unacceptability due to visibility, and of increased cost of some preparations. This form of oestrogen replacement has advantages as first line in patients with complex pituitary disease since it avoids the effects of oral oestrogen on other hormone binding proteins and has less interaction with growth hormone replacement. Transdermal oestrogen may also be particularly appropriate in patients where direct hepatic effects need to be avoided (e.g. in rare patients who develop hepatic adenomas on oral oestrogen, or in young patients with known thrombophilia or previous thromboembolic disease). However, studies comparing the effects of oral combined oestrogen-progestin contraceptive pill and transdermal HRT in the context of central hypogonadism are lacking.

Risk-Benefit Analysis

In recent years large studies of gonadal hormone replacement in postmenopausal women have suggested that, although HRT

certainly reduces the risk of osteoporosis and probably bowel cancer, the increased risk of thromboembolism and cardiovascular disease at postmenopausal age means that the overall effect on health may be trending towards negative [18]. These conclusions do not however apply to younger women with oestrogen deficiency, including those with gonadotrophin deficiency, since the risk of osteoporosis is greater and the risk of cardiovascular events lower due to younger age. Indeed, it is now noted as previously hoped that oestrogen replacement reduces the risk of premature vascular disease and mortality [19]. Resolution of local and systemic symptoms of oestrogen deficiency provides additional drivers for routine oestrogen replacement in all gonadotrophin deficient women of premenopausal age. Balanced against these positive effects is a slight increase in risk of venous thromboembolism compared to the deficient state (which nevertheless remains a very low absolute risk), and mixed evidence on a possible slight increase in risk of breast cancer. In addition, some women will suffer from unwelcome cyclical changes similar to those that may be experienced in the normal cycle.

In the younger women the balance of risks seems overwhelmingly in favour of routine oestrogen replacement. In women of postmenopausal age, the choice of whether or not to take oestrogen replacement is ultimately no different to that in women without pituitary disease.

Male

Hypogonadism treatment in men has two goals: elevating testosterone levels to physiological levels, and induction of spermatogenesis. Androgen replacement therapy is usually aimed at improving libido, sexual function, energy levels, sense of well-being, muscle mass, reduce fat mass, reversal of anaemia and osteoporosis secondary to testosterone deficiency [20]. It is also used for inducing and maintaining secondary sexual characteristics. Caution needs to be exercised if planning for fertility within 6 to 12 months as exogenous testosterone suppresses spermatogenesis, which though reversible may lead to prolonged recovery of the hypothalamic-pituitary-testes axis for men aiming for non-assisted fertility success.

Choice of Replacement Regime

Intramuscular (IM) Depot Testosterone

Depot Testosterone ester preparations were the traditional form of androgen replacement, typically given as testosterone enanthate, 250 mg IM every 3 weeks (range 2–4 weeks) or mixed testosterone esters (Sustanon) IM every 3 weeks. These regimes are the cheapest preparations but do not mimic normal physiology, resulting in high or high normal levels of plasma testosterone in the days immediately following injection falling to low normal or subnormal levels before the next dose. Most patients will notice some changes in mood, libido, muscle strength or general 'drive' consistent with these hormonal fluctuations, which some find troublesome. The need for IM injections also usually necessitates regular visits to the surgery, which may be inconvenient, although occasionally patients may self-administer. Newer forms of testosterone replacement that have become available in recent years have become the preferred methods of replacement in many countries.

A depot preparation of testosterone undecanoate 1000 mg administered deep IM every 3 months (Nebido®) is a popular alternative, which gives more stable and physiological testosterone levels, achieved more rapidly if the second injection is given after 6 weeks. Timing of doses can be adjusted by measuring trough levels of testosterone prior to injection.

Oral

Testosterone undecanoate, a 17α-hydroxyl ester, is also active orally since its highly polar side chain and oily vehicle allow direct absorption into the lymphatic system bypassing hepatic metabolism. Its half-life is relatively short, and multiple daily doses are necessary—typically 40 mg, 2–3 times daily—but oral therapy is preferred by some patients. This preparation is not currently approved for use in the United States given weak efficacy, drug interactions, and gynaecomastia, nevertheless is available in some countries.

Other oral formulations of testosterone are either ineffective or have an unacceptable incidence of hepatic side effects and are no longer used.

Implant

Testosterone pellets (Testopel) can be implanted subcutaneously by trocar and provide stable testosterone levels, typically using testosterone 600 mg every 6 months. The need for repeated surgical procedures means that long-acting depots now largely supersede this form of replacement.

Transdermal

Transdermal gel preparations (typically 50 mg in 5 ml of 1% gel) have become increasingly popular in recent years [21]. They are available in a variety of formulations including sachet, tube, and pump dispenser—all of which have advantages and disadvantages in terms of convenience and ability to adjust doses. In each case the patient applies the gel to shoulders, arms, or abdomen after bathing or showering, typically first thing in the morning. Contact of gel, or gel treated areas, with females or children needs to be avoided. This form of treatment is convenient for many patients but can be problematic for individuals who need to wash or shower frequently.

Transdermal patch (Androderm) preparations are discontinued in United Kingdom; still available in some countries, but their use has been largely overtaken by gel preparations. The patch may cause troublesome skin reactions in 5–10% of patients.

All forms of transdermal therapy achieve physiological peak levels of testosterone and can even mimic the normal testosterone circadian rhythm and hence preferred by some endocrinologists.

Monitoring Androgen Replacement

Testosterone levels can be measured in blood on all forms of replacement and is generally aimed at mid-normal range as that of healthy male population. With conventional intramuscular depot, the majority advocate measuring levels one week after injection (which should be in the upper part of the reference range for normal individuals) and just prior to injection (which will usually be low normal); frankly subnormal levels prior to injection may indicate the need for more frequent injection, while very high peak levels may necessitate a reduction in dose. On the 3-monthly depot,

a preinjection nadir level is most informative, and should be at the lower end of mid-normal reference range. Testosterone levels 2–8 hours post-transdermal preparations' application should be in the mid-reference range. Random testosterone measurements are often low or low normal on oral testosterone undecanoate, but the short half-life makes these measurements less reliable, and dihydrotestosterone levels may be preferentially raised.

Sexual Function

Severe androgen deficiency causes a reduction in libido and erectile potency, which are normally restored by appropriate replacement therapy. In patients who have been hypogonadal for many years counselling of the patient and their partner will be necessary before starting therapy. However, a proportion of men presenting with sexual dysfunction will also be found to have testosterone levels in the borderline low range, without elevation of gonadotrophins, but with no other evidence of pituitary disease; improvement in libido and particularly potency is far less certain with replacement in these patients. Testosterone trials have shown that androgen replacement in 'age-related' hypogonadism may improve sexual function, bone density and anaemia but does not improve vitality or cognition [22].

Adverse Effects of Androgen Replacement Therapy

Randomized and open-label trials indicate that replacement dose testosterone in young men has a low frequency of side effects, but common drug-related side effects are acne, oily skin, and breast tenderness. Excess androgen replacement may cause polycythaemia, and this can also occur in elderly men, in chronic obstructive airways disease and in sleep apnoea with high normal replacement doses, so that haemoglobin levels need to be monitored in such patients. The possibility of long-term adverse effects on the prostate remains controversial. Hypogonadal men have lower prostate volume and prostate-specific antigen levels than the normal population, and induction of hypogonadism in patients with established prostate cancer induces temporary disease remission and reduction in prostate volume—thus raising the possibility that testosterone replacement might increase the incidence of prostate cancer and prostatic hypertrophy. While it seems likely that replacement will raise the incidence of these conditions to those of the normal population there is currently no evidence to suggest an absolute increased risk of either condition with appropriate replacement. In spite of this, and the current lack of evidence that screening for prostate cancer is beneficial in the normal population, some form of prostate monitoring is usually recommended in older men on testosterone replacement. One such strategy is to monitor serum prostate-specific antigen (PSA), with further necessity of urological evaluation only if PSA levels rise above normal, but some recommend repeated rectal examination and/or prostate imaging. Some clinicians advocate testosterone treatment in fully resected gland-confined cancer, rendered 'cured' with or without adjuvant therapy with normal PSA level for 2 years post-treatment. Lack of long-term outcome data to suggest that this strategy is safe is a limitation to make a general recommendation. Hence this warrants close liaison with Urologist and discussion with patient regarding benefits and risks prior to treatment commencement. The link between testosterone and cardiovascular events is controversial. Previous evidence to suggest that low testosterone itself may be associated with the metabolic syndrome and increased cardiovascular risk and that testosterone replacement may therefore be beneficial to cardiovascular health remains to be proven conclusively as to whether the low testosterone is an association or cause to effect relationship exists. Although number of metanalyses have examined the causal relationship between exogenous testosterone and MACE (major adverse cardiovascular events) and deaths in RCTs [23], given variety of limitations including heterogeneity, lack of CV events adjudication, different dosing and formulations, and so on, there are insufficient data to establish this causal link. Nevertheless, the Food and Drugs Administration (FDA) in United States following a citizen petition has mandated pharmaceutical companies to label their products with a warning about potential increased risks of CV events associated with androgen replacement whereas Medicines and Healthcare products Regulatory Agency (MHRA) and European Medicines Agency (EMA) have maintained that the evidence is inconclusive to mandate such warning.

Puberty Induction

Where gonadotrophin deficiency develops before puberty, special care is required to induce pubertal developmental at an appropriate speed in both sexes. Simply commencing full adult replacement doses will result in inappropriately rapid pubertal development with insufficient time for usual psychological adaptation, in less satisfactory secondary sexual development and in possible attenuation of the pubertal growth spurt and final height.

Gonadal steroid replacement is commenced in low doses in both sexes. In females, low doses of oral oestrogen (e.g. ethinyloestradiol 10 mcg daily, or even lower doses where available) or transdermal oestradiol (e.g. 25 mcg or less, twice weekly) is commenced at very low dose, and gradually increased to full adult replacement dosage over 2.5 years with eventual addition of cyclical progestogen [24]. Males may commence with testosterone enanthate 50–100 mg IM every 3–4 weeks for many months before increasing towards adult replacement dosage; similar low-dose regimes can be devised using transdermal testosterone or oral testosterone undecanoate but have not been widely accepted. Other factors sometimes require attention: girls with prepubertal panhypopituitarism may fail to develop pubic hair, and use of topical testosterone creams has been described if normal hair growth is desired; boys may need reassurance that pubertal gynaecomastia is common, and that normal adult facial and body hair develop slowly over many years once testosterone levels have reached the normal adult range.

Growth

Normal individuals show a pubertal growth spurt associated with gonadal steroids, particularly in males, yet gonadal steroids are also responsible for epiphyseal fusion and the cessation of linear growth. Appropriate adjustment of replacement doses of gonadal steroid and growth hormone are therefore essential in patients with hypopituitarism during induction of puberty. Overall, studies indicate that although early puberty (spontaneous or induced) reduces final height, deliberately delaying puberty induction to allow increased time for growth probably does not increase final height significantly, and may be associated with obvious adverse psychological consequences.

Fertility Induction

Gonadal steroid replacement does not induce fertility, but ovulation and spermatogenesis can be stimulated by therapy with gonadotrophin injections (initially human menopausal, now increasingly recombinant) or by pulsatile GnRH administered subcutaneously by infusion pump.

Female

The use of gonadotrophin therapy and pulsatile GnRH are both highly successful, resulting in fertility rates approximating normal levels with repeated cycles of treatment in expert hands [25]. GnRH therapy is mostly used where the gonadotrophin deficiency is considered primarily 'hypothalamic', but may also be successful in patients considered to have primary pituitary disease. As always, such therapy should only be undertaken with close biochemical and ultrasound monitoring of the ovarian response, and in centres with extensive experience of ovarian stimulation techniques, and precise description of treatment regimes is beyond the scope of this book.

Male

Gonadotrophin and GnRH therapy can both induce spermatogenesis, but induction of adequate spermatogenesis takes a minimum of 3 months and may require 1–2 years. luteinizing hormone (LH) therapy is usually used first in the form of human chorionic gonadotropin (hCG) typically 1000–2000 IU, 2–3 times/week which should result in adequate testosterone levels and may sometimes be sufficient to allow spermatogenesis, but follicle-stimulating hormone (FSH) activity is usually required for adequate fertility. A wide variety of regimes have been recommended with successful fertility in a majority of patients, which may occur at surprisingly low total sperm counts. Coexistent primary testicular defects (e.g. related to cryptorchidism) may cause failure of therapy. The need to wear an infusion pump or attend for regular intramuscular injections over many months may be a disadvantage, but self-administration of low doses of gonadotrophins subcutaneously may also be successful in both induction of fertility and increase in testicular size [26]. When pregnancy is achieved spermatogenesis may occasionally be maintained by testosterone replacement alone although usually continued or repeated gonadotrophin therapy is required. Sperm may also be frozen after successful treatment for use in future attempts at fertility.

Hypopituitary Crisis

Acute Hypopituitarism Crisis or Pituitary Apoplexy

An acute hypopituitary crisis or pituitary apoplexy, derived from the Greek word *apoplēxia* meaning 'striking away', is a potentially life-threatening condition due to acute ischaemic infarction or haemorrhage of the pituitary gland characterized by sudden onset headache, vomiting, visual impairment, ophthalmoplegia, and altered consciousness. Apoplexy is often the first presentation of a pre-existing pituitary tumour resulting in acute hypopituitarism causing haemodynamic instability, mainly due to ACTH deficiency. Prompt corticosteroid replacement along with intravenous fluids is vital to treat haemodynamic instability, followed by hormonal, radiological and visual status evaluation [27] as demonstrated in Figure 2.3.5.1.

Surgery versus conservative management for treatment of apoplexy remains controversial. A pituitary apoplexy scoring system using neuro-ophthalmic criteria to assess apoplexy severity to guide the management may be useful [28]. The majority of patients do well with conservative management, but need close monitoring of neurological and visual status, and surgical decompression is to be considered if neuro-ophthalmic signs persist or worsen.

Chronic Hypopituitarism Crisis or Pituitary Coma

Chronic hypopituitarism or pituitary coma is a rare, but life-threatening presentation of severe, long-standing, untreated hypopituitarism, usually precipitated by stress including infection, trauma, surgery, or infarction. Treatment is firstly full replacement with parenteral hydrocortisone, as described previously in this chapter, followed by correction of other factors which may precipitate or worsen coma including hypothermia, salt and water depletion due to hypoadrenalism and/or diabetes insipidus (DI), hyponatraemia due to excessive desmopressin or hypothalamic dysfunction and gradual replacement of hypothyroidism. Radiation-induced hypopituitarism may cause a hypopituitarism crisis if patients are suboptimally managed, as the time of onset of new pituitary deficiencies varies. In the majority of cases, this takes few years to develop, therefore requiring patient education of the symptoms to look out for, together with life-long monitoring and prompt treatment of deficiencies to avoid hypoadrenal crisis and/or death.

Interactions with Other Therapies

Steroid Replacement

Adrenal insufficiency can mask the presence of partial DI as glucocorticoid deficiency impairs free renal water clearance. Therefore, patients should be warned to look out for symptoms of polyuria and polydipsia when they are started on steroids. Typical examples are dexamethasone commencement following cranial radiotherapy for a variety of brain tumours, and hydrocortisone commencement for ACTH deficiency from pituitary radiotherapy or hypophysitis.

Drugs that induce hepatic CYP450 isoenzyme activity (e.g. many anticonvulsants, rifampicin, and so on) may modestly reduce prednisolone concentrations, requiring cautious dose increments in such circumstances. This effect is usually clinically insignificant in case of hydrocortisone, except in case of oxcarbazepine, which accelerates cortisol elimination via cytochrome CYP450 3A4 induction. Nevertheless, it is good practice to be vigilant for symptoms of under-replacement when such drugs are used.

Growth hormone (GH) suppresses the conversion of cortisone to cortisol, and concomitant GH replacement therapy has been shown to reduce cortisol levels in patients on a stable dose of hydrocortisone. This effect appears to be mediated by a reduction in levels of corticosteroid-binding globulin (CBG) and so it is important to look out for symptoms of hypoadrenalism after GH initiation, particularly in cases of partial ACTH deficiency.

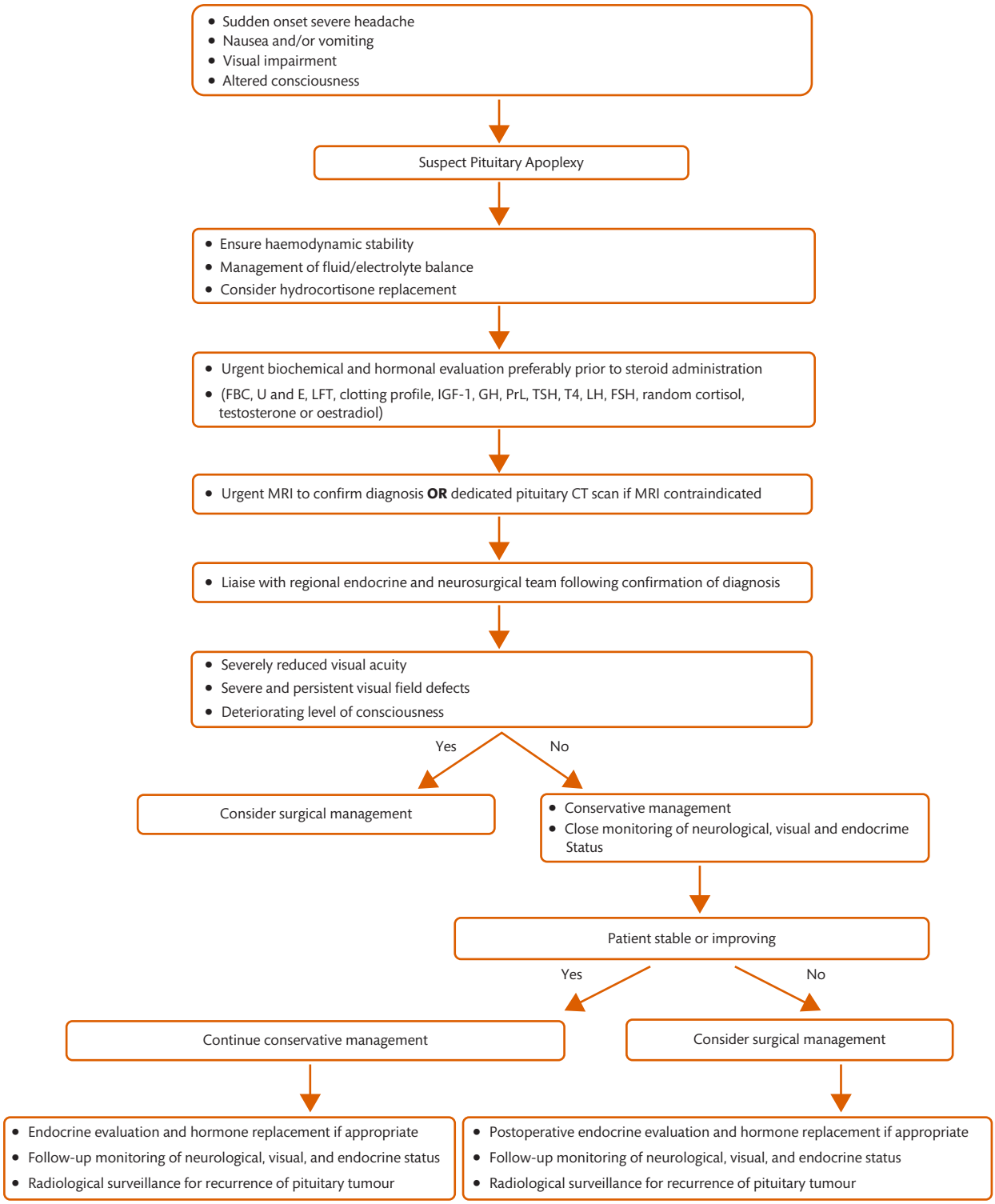


Figure 2.3.5.1 Algorithm for the management of pituitary apoplexy.

Thyroid hormone accelerates endogenous cortisol clearance and could unmask adrenal insufficiency, hence HPA axis should be assessed prior to commencement of levothyroxine to avert a hypoadrenal crisis [16].

Thyroxine Replacement

The simultaneous use of proton pump inhibitors or antacids, calcium carbonate, and ferrous sulphate may reduce absorption of levothyroxine and result in altered dose requirements. It is therefore advised to take levothyroxine on an empty stomach without any interference from food, beverages, or any other drugs for at least 30 minutes.

Administering GH in GH deficient individuals can reduce free T_4 levels thereby unmasking central hypothyroidism. Therefore, the hypothalamic–pituitary–thyroid axis should be tested prior to GH initiation. Central hypothyroidism should be fully treated prior to diagnosing GH deficiency.

Sex Steroid Replacement

Oral oestrogen increases CBG and in turn elevates total, but not free cortisol level but vigilance towards HPA axis re-assessment or hydrocortisone dose evaluation would be appropriate. This effect does not seem to apply to the transdermal route of oestrogen replacement. Similarly, oral oestrogen raises plasma total T_4 level by increasing thyroid-binding globulin. This does not influence active levels of thyroid hormone or change the necessary dose of replacement, but does make biochemical monitoring of thyroid reserve and/or replacement dose more difficult. It is therefore usually best to fully assess both thyroid and adrenal axes and/or optimize replacement levels before starting or during oral oestrogen therapy, or to use transdermal oestradiol as an alternative.

Oral oestrogen also causes a slight reduction in IGF-1 and rise in GH levels, which may require dose adjustment to concomitant GH replacement therapy in order to maintain IGF1 in the target range.

Special Circumstances

Management of Hypopituitarism in Pregnancy

Natural pregnancy in the context of hypopituitarism is rare because secondary hypogonadism resulting in infertility is quite common. Assisted Conception is often necessary, and with appropriate monitoring, women with hypopituitarism usually have an uneventful pregnancy.

Hydrocortisone is degraded by placental11-β-hydroxy steroid dehydrogenase 2, and hence does not cross the placenta, making it the preferred glucocorticoid of choice in pregnancy. Dexamethasone should not be used, as the placenta does not inactivate it. HPA axis is activated in pregnancy with a resulting twofold increase in total cortisol and eightfold increase in aldosterone [29]. Given that total cortisol levels are falsely 'normal' due to raised CBG, there is a danger of masking of frank adrenal insufficiency in the first trimester when clinical features may be mistaken for hyperemesis or that of normal pregnancy. A cortisol of <100 nmol/L with classical symptoms should raise suspicion of adrenal insufficiency. The ideal hydrocortisone replacement dose in pregnancy is not well defined and dose adjustment is largely based on clinical symptoms, but often the dose is increased as pregnancy progresses and may require an increase by 20–40% towards end of third trimester [29]. Parenteral hydrocortisone of 50 mg or 100 mg during second stage of labour, 100 mg at induction of anaesthesia, and subsequent 8-hourly 100 mg dosing for caesarean section is recommended [9].

Secondary hypothyroidism may not require levothyroxine dose escalation as done routinely for primary hypothyroidism in pregnancy. This is due to human chorionic gonadotrophin activating the inert thyroid gland, due to common α-subunit similar to TSH, to produce thyroid hormones, particularly in first trimester. It is advisable to dose adjust based on thyroid hormone levels which should be taken at 4 to 6 weekly intervals. Some clinicians advocate measurement of total T_4 as many free T_4 assays are found to be less precise in pregnancy, and local assay protocols should be evaluated to choose the preferred monitoring test. TSH is unreliable in secondary hypothyroidism, and is not helpful in this context.

Pregnancy may unmask mild DI as vasopressinase enzyme produced by the placenta can inactivate endogenous vasopressin (antidiuretic hormone, ADH). Hence there may be a need to increase DDAVP (i.e. desmopressin) particularly towards the end of pregnancy, and DDAVP is generally safe in pregnancy and for the baby even during lactation periods. The majority of guidelines advocate discontinuation of GH treatment due to lack of safety or efficacy data, but an observational study showed no harm with normal pregnancy outcomes on continuing GH in about 100 pregnancies [30].

Cancer Immunotherapy-Induced Hypopituitarism

A novel class of cancer immunotherapy with monoclonal antibodies targeting both T cell and tumour cells has been found effective in wide range of cancers. They enhance antitumour immunity by blocking negative regulators (checkpoints) on the T cell, and are therefore called immune checkpoint inhibitors (CPIs). The main checkpoints targeted are cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) (ipilimumab and tremelimumab), programmed cell death-1 receptor (PD-1) (nivolumab and pembrolizumab) and its ligand PD-L1 (avelumab), resulting in improved survival rates in variety of cancers such as melanoma, non-small cell lung cancer, renal cancer, bladder cancer, non-Hodgkin's lymphoma among others [31]. Introduced since 2011, this class of immunotherapy has resulted in increasingly recognized autoimmune toxicities of various systems, including that of endocrine system. These side effects are called immune related adverse events (IRAEs) and the main endocrinopathies described are thyroiditis (15%), hypophysitis (9%), adrenalitis (1%) and type 1 diabetes mellitus (<1%) [32]. Typically, the onset of endocrine IRAEs following initiation of CPIs is between 6 to 12 weeks but the range can be from 3 weeks to as late as 3 years. Endocrine IRAEs have a slight male preponderance; often do not resolve on CPI cessation, particularly hypoadrenalism. The clinical features of CPI-induced hypopituitarism are similar to hypophysitis, although interestingly DI does not seem to be a feature. Identification and

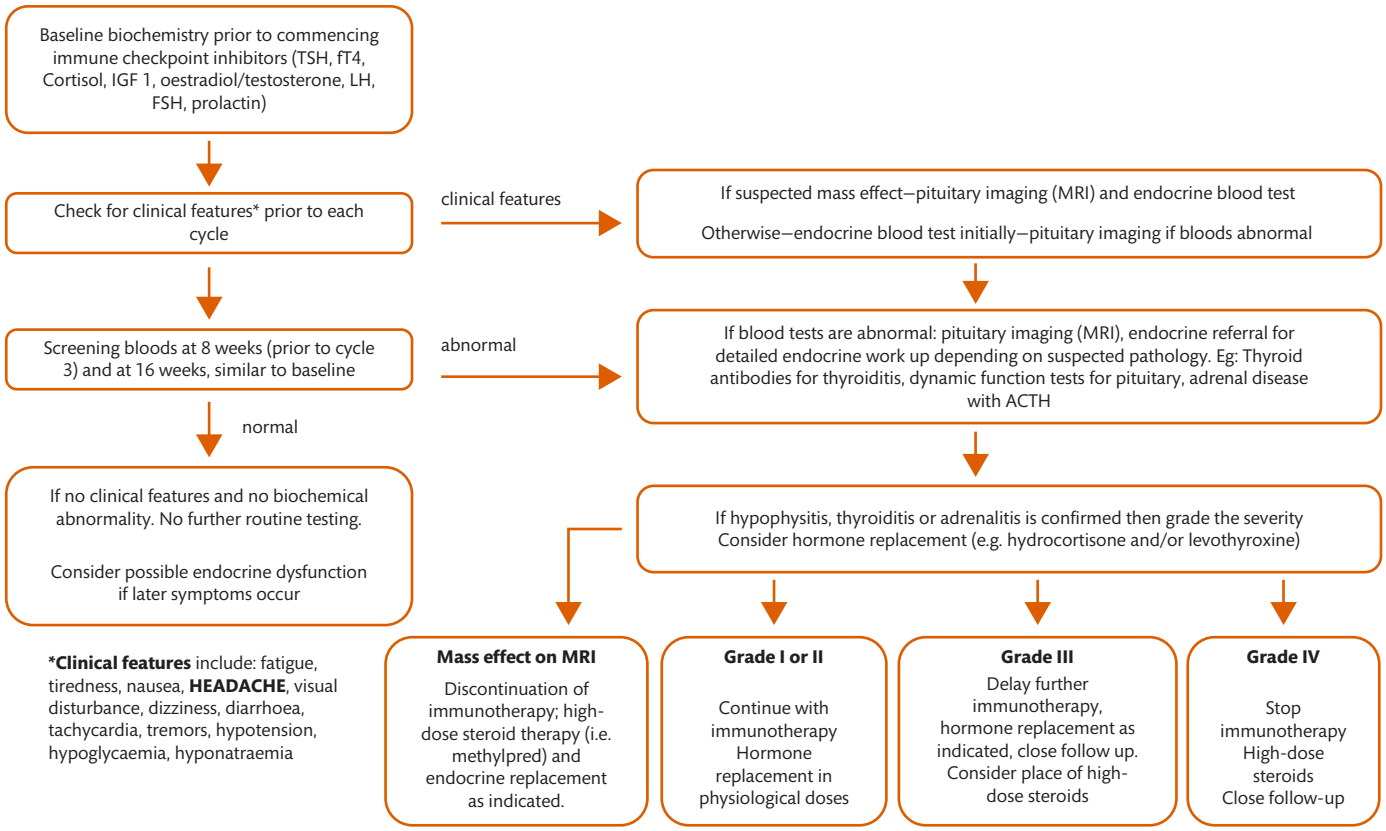


Figure 2.3.5.2 Algorithm of management of checkpoint inhibitor-induced endocrinopathies.
Reproduced with permission from Joshi MN, Whitelaw BC, Palomar MTP, Wu Y, Carroll PV. Immune checkpoint inhibitor-related hypophysitis and endocrine dysfunction: clinical review. *Clinical Endocrinology* Vol. 85, Issue 3 (2016). Copyright 2016 © John Wiley & Sons Ltd.

treatment of individual endocrinopathies including hypopituitarism secondary to CPIs is the same as any other aetiology that causes isolated or combined endocrine pathologies [32], and is outlined in Figure 2.3.5.2.

Some clinicians believe CPIs should be discontinued if hypophysitis occurs, but current clinical practice is to continue with CPIs with the aim of malignancy regression. If there is evidence that vision has been affected by an enlarged inflamed pituitary gland, a pituitary MRI scan should be performed, and steroid immunosuppression started. Intravenous methylprednisolone is the usual steroid of choice and at the same time any hormonal deficits should be replaced. The mechanism by which CPIs induce endocrinopathy warrants further research as these new molecular targets and pathways may shed light on these old endocrine diseases, resulting in a better understanding of the pathophysiology of hypopituitarism.

Traumatic Brain Injury

Hypopituitarism following traumatic brain injury (TBI) has received increasing attention in the past few years. Evidence suggests that this is a significant cause of reversible morbidity in patients who may remain symptomatic for many years after the injury [33]. One or all of the pituitary axes can be deficient following TBI, and a high index of clinical suspicion is required in order to trigger appropriate testing and endocrine replacement. A summary of identification, treatment and monitoring of hypopituitary patients resulting from any aetiology including that of TBI is depicted in Table 2.3.5.1. It is good practice for endocrinologists to liaise with the specialist looking after patients with TBI in order to design simple endocrine screening protocols in this group of patients.

Patient Education and Participation

Pituitary disease is best managed in specialized centres possessing dedicated pituitary multidisciplinary team equipped with robust patient pathways and protocols. It has become almost necessary to mandate this, quite appropriately, given complexity of hypopituitarism, growing pituitary tumours, risk of pituitary apoplexy, desire for fertility, multiple comorbidities, polypharmacy, poor quality of life, and so on. Optimal management is only possible when the patient is fully aware of the nature and consequences of the disorder and can participate actively in the adjustment of replacement therapy. Patient education is therefore an essential part of management, and patients should also be encouraged to obtain further information and support from groups such as the *Pituitary Foundation* (UK) (https://www.pituitary.org.uk) or *Pituitary Network Association* (USA) (https://www.pituitary.org).

Table 2.3.5.1 Management of hypopituitarism

Pituitary axis deficiency	Clinical features	Diagnosis	Treatment	Monitoring
GH	Fatigue Reduced exercise capacity Decreased lean body mass/increased fat mass Muscle weakness Osteoporosis Hypertension Hyperlipidaemia Insulin resistance Impaired cardiac function Premature atherosclerosis Impaired sleep quality Depression	• Low IGF-1 • >11 score on AGHDA (Assessment of GH deficiency in adults) symptom questionnaire • Inadequate GH response upon dynamic pituitary testing: (a) Insulin tolerance test or Glucagon stimulation test (expected > 3–5 mcg/L depending upon BMI); (b) GHRH + Arginine test (expected > 4 mcg/L; obesity may blunt GH response)	Recombinant human GH	• IGF-1 in upper third of normal range of normal healthy population • AGHDA symptom questionnaire
LH/FSH Male	Fatigue Reduced exercise capacity Low libido Erectile dysfunction Hot flushes Infertility Pallor Loss of body hair Muscle weakness Osteoporosis Impaired sleep quality Depression	• Low testosterone • Low or inappropriately normal LH • Low or inappropriately normal FSH • Low sperm counts, morphology, and function	• Androgen replacement therapy • HCG, gonadotrophin	Fasting 08.00–09.00 serum Testosterone, PSA, Haemoglobin & haematocrit, Lipids, sperm analysis in context of infertility treatment
LH/FSH Female	Fatigue Reduced exercise capacity Low libido Hot flushes Infertility Vaginal dryness Pallor Loss of body hair Muscle weakness Osteoporosis Oligo/amenorrhoea Impaired sleep quality Depression	• Low oestradiol • Low or inappropriately normal LH • Low or inappropriately normal FSH	Oestrogen replacement until age of 50 years Cyclical progesterone combined with oestrogen if uterus intact Stimulation of ovaries with gonadotrophin, GnRH, assisted conception techniques for infertility treatment	Serum oestradiol level generally not useful
TSH	Fatigue Weight gain Cold intolerance Dry skin Pallor Loss of body hair Muscle weakness Reduced exercise capacity Dyspnoea constipation Oligo/amenorrhoea Infertility Hypertension Bradycardia Hyperlipidaemia Insulin resistance Impaired cardiac function Premature atherosclerosis Cognitive decline Osteoporosis impaired sleep quality Depression	Free T_4	Levothyroxine	Free T_4 in mid-range of normal healthy population

Table 2.3.5.1 *Continued*

Pituitary axis deficiency	Clinical features	Diagnosis	Treatment	Monitoring
ACTH	Fatigue Weight loss Reduced exercise capacity Cognitive decline Dry skin Pallor Loss of body hair Orthostatic hypotension Hypoglycaemia Impaired cardiac function Dyspnoea Anorexia Nausea/vomiting Diarrhoea/loose stools Osteoporosis Oligo/amenorrhoea	• Inadequate cortisol response (<500 nmol/L) on insulin tolerance test • 250 mcg corticotropin stimulation test 30 or 60 min cortisol response of <500 or 550 nmol/L depending upon the assay	Hydrocortisone Prednisolone	• Clinical features of excess: weight gain, insomnia, and peripheral oedema • Clinical features of insufficient replacement: nausea, lethargy, postural dizziness, weight loss, and lack of appetite
ADH	Polydipsia Polyuria/nocturia Clinical evidence of dehydration	• High serum osmolality • Low urine osmolality • Hypernatraemia • Hyperuricaemia	Desmopressin (DDAVP) nasal spray/oral/sublingual	Serum osmolality Urine osmolality Serum urea & electrolytes

Data derived from Fleseriu M, et al. Hormonal Replacement in Hypopituitarism in Adults: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2016 Nov; 101(11):3888–3921. Copyright © 2016, Oxford University Press.

REFERENCES

1. Krieger DT, Allen W, Rizzo F, Krieger HP. Characterization of the normal temporal pattern of plasma corticosteroid levels. *J Clin Endocrinol Metab*, 1971; **32**: 266–84.
2. Howlett TA. An assessment of optimal hydrocortisone replacement therapy. *Clin Endocrinol*, 1997; **46**: 263–8.
3. Johannsson G, Bergthorsdottir R, Nilsson AG, Lennernas H, Hedner T, Skrtic S. Improving glucocorticoid replacement therapy using a novel modified-release hydrocortisone tablet: a pharmacokinetic study. *Eur J Endocrinol*, 2009; **161**: 119–30.
4. Whitaker M, Debono M, Huatan H, Merke D, Arlt W, Ross RJ. An oral multiparticulate, modified-release, hydrocortisone replacement therapy that provides physiological cortisol exposure. *Clin Endocrinol*, 2014; **80**: 554–61.
5. Johannsson G, Nilsson AG, Bergthorsdottir R, *et al*. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab*, 2012; **97**: 473–81.
6. Oksnes M, Bjornsdottir S, Isaksson M, *et al*. Continuous subcutaneous hydrocortisone infusion versus oral hydrocortisone replacement for treatment of addison's disease: a randomized clinical trial. *J Clin Endocrinol Metab*, 2014; **99**: 1665–74.
7. Smith DJF, Prabhudev H, Choudhury S, Meeran K. Prednisolone has the same cardiovascular risk profile as hydrocortisone in glucocorticoid replacement. *Endocr Connect*, 2017; **6**: 766–72.
8. Frey KR, Kienitz T, Schulz J, Ventz M, Zopf K, Quinkler M. Prednisolone is associated with a worse bone mineral density in primary adrenal insufficiency. *Endocr Connect* 2018; **7**: 811–18.
9. Fleseriu M, Hashim IA, Karavitaki N, *et al*. Hormonal replacement in hypopituitarism in adults: an Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 3888–921.
10. Bornstein SR, Allolio B, Arlt W, *et al*. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 364–89.
11. Addisons Disease Self Help Group. Surgical guidelines, 2017. Available at: https://www.addisons.org.uk (accessed 29 January 2020).
12. Murray RD, Ekman B, Uddin S, *et al*. Management of glucocorticoid replacement in adrenal insufficiency shows notable heterogeneity—data from the EU-AIR. *Clin Endocrinol*, 2017; **86**: 340–6.
13. Kraan GP, Dullaart RP, Pratt JJ, Wolthers BG, Drayer NM, De Bruin R. The daily cortisol production reinvestigated in healthy men. The serum and urinary cortisol production rates are not significantly different. *J Clin Endocrinol Metab*, 1998; **83**: 1247–52.
14. Behan LA, Kelleher G, Hannon MJ, *et al*. Low-dose hydrocortisone replacement therapy is associated with improved bone remodelling balance in hypopituitary male patients. *Eur J Endocrinol*, 2014; **170**: 141–50.
15. Grozinsky-Glasberg S, Fraser A, Nahshoni E, Weizman A, Leibovici L. Thyroxine-triiodothyronine combination therapy versus thyroxine monotherapy for clinical hypothyroidism: meta-analysis of randomized controlled trials. *J Clin Endocrinol Metab*, 2006; **91**: 2592–9.
16. Persani L. Clinical review: central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*, 2012; **97**: 3068–78.
17. Koulouri O, Auldin MA, Agarwal R, *et al*. Diagnosis and treatment of hypothyroidism in TSH deficiency compared to primary thyroid disease: pituitary patients are at risk of under-replacement with levothyroxine. *Clin Endocrinol*, 2011; **74**: 744–9.
18. Rossouw JE, Anderson GL, Prentice RL, *et al*. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, 2002; **288**: 321–33.
19. Rossouw JE. Hormones, genetic factors, and gender differences in cardiovascular disease. *Cardiovasc Res*, 2002; **53**: 550–7.
20. Snyder PJ, Peachey H, Berlin JA, *et al*. Effects of testosterone replacement in hypogonadal men. *J Clin Endocrinol Metab*, 2000; **85**: 2670–7.

21. Bhasin S, Basaria S. Diagnosis and treatment of hypogonadism in men. *Best Pract Res*, 2011; **25**: 251–70.
22. Snyder PJ, Ellenberg SS, Farrar JT. Testosterone treatment in older men. *N Engl J Med*, 2016; **375**: 90.
23. Alexander GC, Iyer G, Lucas E, Lin D, Singh S. Cardiovascular risks of exogenous testosterone use among men: a systematic review and meta-analysis. *Am J Med*, 2017; **130**: 293–305.
24. Matthews D, Bath L, Hogler W, Mason A, Smyth A, Skae M. Hormone supplementation for pubertal induction in girls. *Arch Dis Child*, 2017; **102**: 975–80.
25. Martin KA, Hall JE, Adams JM, Crowley WF, Jr. Comparison of exogenous gonadotropins and pulsatile gonadotropin-releasing hormone for induction of ovulation in hypogonadotropic amenorrhea. *J Clin Endocrinol Metab*, 1993; **77**: 125–9.
26. Jones TH, Darne JF. Self-administered subcutaneous human menopausal gonadotrophin for the stimulation of testicular growth and the initiation of spermatogenesis in hypogonadotrophic hypogonadism. *Clin Endocrinol*, 1993; **38**: 203–8.
27. Rajasekaran S, Vanderpump M, Baldeweg S, *et al*. UK guidelines for the management of pituitary apoplexy. *Clin Endocrinol*, 2011; **74**: 9–20.
28. Reddy NL, Rajasekaran S, Han TS, *et al*. An objective scoring tool in the management of patients with pituitary apoplexy. *Clin Endocrinol*, 2011; **75**: 723.
29. Suri D, Moran J, Hibbard JU, Kasza K, Weiss RE. Assessment of adrenal reserve in pregnancy: defining the normal response to the adrenocorticotropin stimulation test. *J Clin Endocrinol Metab*, 2006; **91**: 3866–72.
30. Vila G, Akerblad AC, Mattsson AF, *et al*. Pregnancy outcomes in women with growth hormone deficiency. *Fertil Steril*, 2015; **104**: 1210–17; e1211.
31. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*, 2016; **16**: 275–87.
32. Joshi MN, Whitelaw BC, Palomar MT, Wu Y, Carroll PV. Immune checkpoint inhibitor-related hypophysitis and endocrine dysfunction: clinical review. *Clin Endocrinol*, 2016; **85**: 331–9.
33. Klose M, Feldt-Rasmussen U. Chronic endocrine consequences of traumatic brain injury—what is the evidence? *Nat Rev Endocrinol*, 2018; **14**: 57–62.

2.3.6 Adult Growth Hormone Deficiency

Jens O.L. Jørgensen

Introduction

A growth-promoting activity of anterior pituitary extracts was discovered a century ago, and growth hormone (GH) was isolated in 1944 and tested in human subjects ten years after [1]. The protein anabolic and lipid catabolic effects of GH were documented in both children and adults, and it was hypothesized that GH replacement therapy in adults with hypopituitarism could be beneficial [2, 3]. Growth hormone for clinical use originally derived from human cadaveric pituitaries and the limited supply was restricted to treat hypopituitary children with severe growth retardation. The introduction of biosynthetic human GH enabled other indications such as GH replacement in adult patients with GH deficiency (GHDA).

The first placebo-controlled trial included young adult patients with childhood-onset GHDA in whom prior GH treatment discontinued several years before [4]. A significant increase in muscle mass and reduction in fat mass were recorded together with improved aerobic exercise capacity. The study also revealed a marked increase in the circulating and urinary levels of bone remodelling biomarkers [5, 6], increased extrathyroidal conversion of T4 to T3 [7], and increased sweating [8]. The second placebo-controlled trial comprised adult patients with adult-onset GHD due to a pituitary tumour and its treatment [9]. Lean body mass increased significantly after 6 months GH replacement together with a significant reduction in fat mass. This was associated with an increase in energy expenditure and exercise capacity. These pivotal findings were confirmed by numerous groups and led to the approval of GHDA as an indication for GH replacement.

Pathophysiology

It is important to emphasize that GHDA as a disease entity only exists within a clinical context of overt pituitary pathology. The reasons for this are several. First, GH secretion in the general population declines gradually as a function of age and adiposity, and the symptoms and signs of GHDA are unspecific. Second, even though diagnostic biochemical criteria exist, they are not sufficiently specific to exclude the diagnosis in elderly individuals with simple obesity. Third, a clinical benefit of GH replacement has only been documented in patients with a clinical history of pituitary disease [10].

One appropriate clinical context is a history of childhood-onset GHDA, which accounts for approximately 30% of GHDA. The cardinal symptom of GH deficiency (GHD) in childhood is longitudinal growth retardation and the underlying cause in this group is frequently idiopathic.

The second clinical context is adult-onset GHDA caused by a pituitary mass lesion or its treatment, of which pituitary adenomas constitute 70% (**Box 2.3.6.1**).

Box 2.3.6.1 Aetiology of GHDA

Cause

Adult-onset disease (70%):
Pituitary adenoma (40%)
Craniopharyngioma (8%)
Non-pituitary intracranial tumours (7%)
Other causes[a] (15%)

Childhood-onset disease (30%):
Idiopathic (15%)
Craniopharyngioma (7%)
Non-pituitary intracranial tumours (5%)
Congenital (3%)

[a] Including pituitary apoplexy, empty sella, Sheehan's syndrome, granulomatous disease, hypophysitis, irradiation, brain trauma.

GH is typically the first pituitary hormone to become deficient in response tumour pressure, pituitary surgery, cranial irradiation, or any other trauma [11]. Traumatic brain injury (TBI), which is a very frequent occurrence, may cause hypopituitarism, but the prevalence of permanent GHD after TBI is only 1% [12]. The annual incidence and the prevalence of adult-onset GHDA are approximately 15/million and 300/million, respectively [13].

Clinical Features

The syndrome of GHDA is used as a term to describe the clinical features and the effects of GH replacement [14]. The syndrome overlaps with the metabolic syndrome as regards visceral obesity, hyperlipidaemia, and atherosclerosis (Box 2.3.6.2). Moreover, premature cardiovascular morbidity and mortality are prevalent in GH-untreated hypopituitary adults [15, 16]. Untreated GHDA is accompanied by reduced total body water and extracellular fluid volume [17], which is reversed by GH replacement [18, 19]. Impaired thermoregulation in response to the ambient outside temperature [20] and during strenuous exercise [21, 22] are also present and partly attributed to reduced sweating capacity [8].

Insulin resistance, which is a hallmark of the metabolic syndrome, is not part of the GHDA syndrome, rather the opposite. GH antagonizes the effects of insulin on glucose metabolism in both the liver and skeletal muscle, which is causally linked to the lipolytic effects of GH [23]. Indeed, increased insulin sensitivity and reactive hypoglycaemia is characteristic of children with GHD, whereas the opposite is true for active acromegaly. The insulin-antagonistic effect of GH is rapidly reversible and in normal physiology, it operates in the fasting state, where insulin activity is low [24, 25].

Box 2.3.6.2 The clinical features of GHDA

Fatigue and impaired quality of life

Abnormal body composition:
- Increased fat mass, particularly central and visceral fat deposition
- Decreased lean body mass
- Decreased total body water

Reduced bone mineral density

Abnormal lipid profile and increased cardiovascular risk profile:
- Increased levels of total and low-density lipoprotein (LDL) cholesterol and reduced high-density lipoprotein (HDL) cholesterol
- Increased carotid intima media thickness
- Increased C-reactive protein levels

Impaired cardiac function:
- Reduced left ventricular wall mass
- Left ventricular systolic function

Decreased sweating

Reduced muscle strength and aerobic exercise capacity

Diagnosis of Adult GHD

In the absence of specific clinical indicators to discriminate GHDA from the normal population, the diagnosis is based on the results of biochemical testing within an appropriate clinical context [10]. For childhood-onset GHD, retesting in adulthood is necessary in patients with isolated idiopathic GHD, since spontaneous normalization of GH status is recorded in almost 70% [26]. Retesting is not required for those with an identified genetic cause and those with more than three pituitary hormone deficits.

Since GH is secreted in a pulsatile fashion, a single random GH measurement is only informative if a high value is incidentally encountered. Therefore, a stimulation test is necessary to perform. The insulin tolerance test (ITT) is the 'gold standard' but is contraindicated in patients with ischaemic heart disease or seizures [10]. Alternative tests include the glucagon stimulation test, GH releasing hormone (GHRH) with arginine, and GHRH with growth hormone-releasing peptide 6. The diagnostic threshold for the ITT and the GST is less than 3 μg/L. The response to GHRH + arginine is reduced in obesity, wherefore threshold levels stratified by body mass index (BMI) are recommended (Box 2.3.6.3).

The anabolic effects of GH are mainly mediated by insulin-like growth factor I (IGF-I), which is present in the circulation and not subject to diurnal fluctuations. In the absence of severe concomitant illness, a low serum IGF-I level is a strong predictor of GHD, but a normal serum IGF-I level does not rule out the diagnosis. Thus,

Box 2.3.6.3 Diagnosis of GHDA

Who to test
- **Only patients with a history of hypothalamic-pituitary disease:**
  - Adult-onset evidence of structural lesions and trauma
  - Adolescent and adult patients with childhood-onset idiopathic GHD

How to test
- **A single stimulation test is sufficient:**
  - Insulin tolerance test (ITT) is gold standard (cut-off: <3 μg/L)
  - Glucagon stimulation test if ITT is contraindicated (cut-off: 3 μg/L)
  - GHRH and arginine stimulation test if ITT is contraindicated (cut-off: <11 μg/L in normal weight patients; reduced cut-off in obese patients)
- **Patients with three or more pituitary deficits and a subnormal IGF-I level do not need a test**

Outcomes	No. of trials	Treatment		Weighted mean (SD) change (GH-placebo)
		GH	Placebo	
Lean B mass	19	473	474	2.82 kg (2.68)
Fat mass	13	352	345	−3.05 kg (3.29)
BMI	8	134	134	−0.12 kg/m^2 (1.40)
TG	11	202	203	0.07 mmol/L (0.36)
HDL Chol.	13	267	261	0.06 mmol/L (0.09)
LDL Chol.	13	255	248	−0.53 mmol/L (0.29)
Total Chol.	15	310	306	−0.34 mmol/L (0.31)
DBP	10	200	201	−1.80 mm Hg (3.77)
SBP	9	190	191	2.06 mm Hg (5.34)
Insulin	11	192	194	8.66 pmol/L (6.98)
Glucose	13	254	257	0.22 mmol/L (0.14)

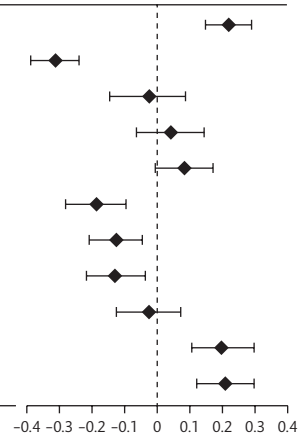


Figure 2.3.6.1 Results of meta-analysis of GH effects on cardiovascular risk factors *from Maison et al. (28).* Abbreviations: Lean B mass, Lean body mass; TG, triglycerides; Chol, cholesterol; DBP, diastolic blood pressure; SBP, systolic blood pressure; ns, non-significant.

in patients at high à priori risk of GHD, a subnormal serum IGF-I value may substitute for a GH stimulation test [27].

Effects of GH Replacement

Several meta-analyses of placebo-controlled trials of GH replacement in GHDA have been published on outcomes such as cardiovascular risk factors [28], muscle strength and exercise capacity [29, 30], bone mineral density [31, 32], body composition [33], and cardiac function [34] (Figure 2.3.6.1).

The studies confirm and substantiate beneficial effects of GH replacement on body composition, bone mineral density, cardiac function, and exercise capacity, as well as side effects attributable to fluid retention and insulin resistance [28, 33].

GHDA patients are highly sensitive to GH in terms of serum IGF-I generation and side effects [35], and male patients are more responsive to GH as compared to females [36, 37].

It remains an open question whether GH replacement therapy improves patient-reported outcomes such as quality of life (QoL) or cognitive function in the adult patient, since neither original studies nor meta-analyses provide unambiguous answers [33, 38–40]. Most QoL studies have utilized generic or disease-specific questionnaires, which mainly record and depend on the respondents' remembrance and it is possible that improvements in remembered QoL wane with time. It is interesting that the most compelling beneficial effects of adult GH replacement on QoL was recorded in a placebo-controlled crossover study, in which the spouse of the patient was asked to score the patient [41]. The National Institute for Health and Care Excellence (NICE) in the United Kingdom requires impaired pretreatment QoL in order to initiate adult GH replacement [42].

Transition from Childhood to Adulthood

Normal puberty marks the transition from childhood to adulthood and is dominated by the pubertal growth spurt and the development of secondary sexual characteristics and reproductive capacity. Muscle and bone mass increase markedly during the transition and depend in part on a marked physiological increase in GH secretion and action resulting in grossly elevated IGF-I levels [43] (Figure 2.3.6.2).

Serum IGF-I levels remain elevated 2–5 years after puberty suggesting additional physiological actions of GH on muscle and bone mass accrual.

Previously, GH replacement in childhood patients terminated when a target height was reached, but the availability of biosynthetic GH enabled continuation during the entire transition phase.

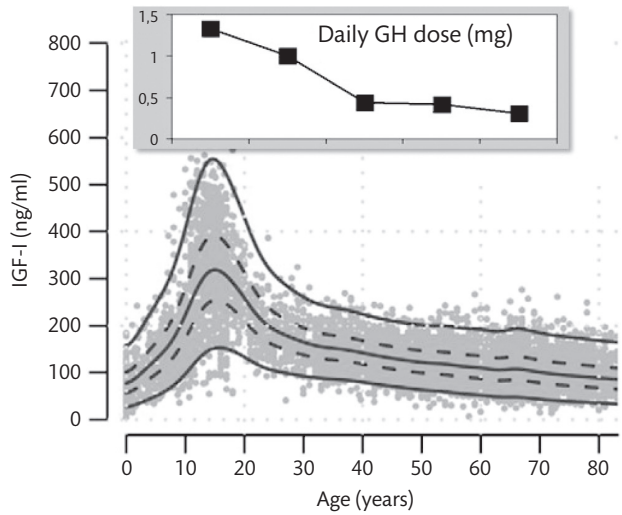


Figure 2.3.6.2 Serum IGF-I levels (2.5%, 50%, and 97.5% percentiles) as a function of age. The insert shows the corresponding daily GH dosage targeted to achieve a serum IGF-I level within the upper normal range.

A placebo-controlled study comparing continuation vs. discontinuation of GH in transition patients revealed that GH discontinuation result in decreased IGF-I levels and increased body fat, which reversed by resumption of GH therapy [44, 45]. Therefore, childhood-onset GH replacement should continue during the transition period and into adulthood, but it is important to retest GH status in the absence of either a genetic cause or additional pituitary deficiencies.

The Senescence

The decline in endogenous GH and IGF-I production with age [46] runs in parallel with senescent changes in body composition and physical performance (Figure 2.3.6.2). Interestingly, in mid-life adults, abdominal adiposity is the strongest and negative determinant of endogenous GH secretion [47]. These associations have led to speculations about a causal link between reduced GH production and the physical frailties of ageing, which has been coined 'somatopause'. This concept is dubious and a meta-analysis of GH treatment studies in elderly subjects without overt pituitary disease record only limited positive effects and a high prevalence of GH-related side effects [48].

Despite the age-associated decline GH secretion, GHDA patients aged 60–80 years have reduced GH and IGF-I levels compared to age-matched controls [49] and seem to respond to respond to GH replacement in the same manner as younger patients [50–52].

GH Dosing and Side Effects

The goal of GH replacement in GHDA is to correct the abnormalities associated with adult GHD with a minimum of side effects [27]. The daily starting dose of GH in young men and women are 0.2 and 0.3 mg, respectively, and in older patients 0.1 mg (Figure 2.3.6.2).

Serum IGF-I is a useful biomarker of GH replacement and the aim is to achieve a level within the upper half of the age-related normal range. GH is administered as subcutaneous self-injections in the evening. Quantification of body composition and bone mineral density by dual X-ray absorptiometry (DXA) should be considered on an individual basis, and body composition can also be assessed using anthropometric measures such waist to hip ratio or skinfold thickness, and bioelectrical impedance. Regular measurements of cardiovascular risk factors such as plasma lipids are also recommended, since they may be abnormal at baseline and modified favourably by GH replacement. Growth hormone replacement is considered life-long, but a trial of withdrawal can be considered if a patient perceives no benefit. In the United Kingdom, NICE advises that GH replacement should be discontinued in patients who fail to demonstrate improvement in QoL the first 9 months of therapy [42]. This recommendation, however, is pragmatic rather than based on evidence from placebo-controlled trails.

Oral oestrogen treatment suppresses hepatic IGF-I production and the GH dose requirement may change accordingly. GH replacement may modify the dose of levothyroxine or unmask incipient central hypothyroidism by increasing the peripheral conversion of T4 to T3 [7]. Finally, GH reduces the enzymatic conversion of cortisone to cortisol, wherefore GH replacement occasionally necessitates a dose increase of glucocorticoid replacement in a patient with concomitant adrenocortical insufficiency.

Reversible fluid retention is a frequent but reversible and dose-dependent side effect of GH replacement in GHDA. The mechanism involves sodium retention. Of note, this increase in hydration accounts for a part of the GH-induced changes in body composition [53]. The direct insulin-antagonistic effects of GH do not compromise glucose tolerance during physiological conditions, since endogenous GH levels are suppressed by food intake. Daily subcutaneous GH injections in the evening, however, is unable to fully imitate the endogenous GH pattern [54], wherefore GH replacement therapy may result in moderate elevations in fasting levels pf glucose and insulin [25] despite favourable changes in body composition [28].

Absolute contraindications for adult GH replacement therapy include active malignancy and proliferative retinopathy.

Although early pregnancy is not a contraindication, GH replacement is discontinued after the first trimester, where placental GH production sets in [27].

GH Replacement and Mortality

Increased mortality in hypopituitary patients due to cardiovascular disease is well established and has been attributed to unsubstituted GHDA [16, 55]. However, numerous underlying mechanisms may be equally—or more—likely, e.g. the underlying disease, treatment complications, and suboptimal substitution of additional pituitary deficiencies. It is also noteworthy, that mortality and cancer incidence are increased in acromegaly [56, 57], and epidemiological human studies suggest a U-shaped association between serum IGF-I levels and all-cause mortality in the general population [58]. Controlled studies of GH replacement therapy with mortality as an endpoint do not exist, but observational studies in GHDA suggest that mortality is reduced in GH replaced patients as compared to GH-untreated patients [55, 59–61].

Conclusion

Metabolic effects of GH were documented in hypopituitary adult patients more than half a century ago, and GH replacement in GHDA has been routinely used for 15 years. The change in body composition with reduced fat mass and increased lean body mass is the most robust effect, together with improved aerobic exercise capacity and cardiac function. Bone turnover is acutely stimulated and there is evidence to suggest increased bone mineral density (BMD) after prolonged GH replacement therapy. Whether this reduces the risk of osteoporotic fractures is uncertain, but observational studies suggest a reduced fracture risk [62]. The impact on QoL is less certain, since the results from placebo-controlled trials are ambiguous, and the results from open and observational studies are likely biased. Cancer risk is not increased with GH treatment, and mortality [63], if anything, is reduced [59, 60]. Side effects in terms of fluid retention and impaired insulin sensitivity are recognized, but are dose-dependent, rapidly reversible, and probably of limited concern. Nevertheless, caution is mandated to avoid overtreatment, not least of the elder patient.

REFERENCES

1. Beck JC, McGarry EE, Dyrenfurth I, Venning EH. Metabolic effects of human and monkey growth hormone in man. *Science*, 1957; **125**: 884–5.
2. Raben MS. Growth hormone. *N Engl J Med*, 1962; **266**: 31–5.
3. Raben MS. Growth hormone. *N Engl J Med*, 1962; **266**: 82–6.
4. Jørgensen JO, Pedersen SA, Thuesen L, *et al*. Beneficial effects of growth hormone treatment in GH-deficient adults. *Lancet*, 1989; **1**: 1221–5.
5. Johansen JS, Pedersen SA, Jørgensen JO, *et al*. Effects of growth hormone (GH) on plasma bone Gla protein in GH-deficient adults. *J Clin Endocrinol Metab*, 1990; **70**: 916–19.
6. Schlemmer A, Johansen JS, Pedersen SA, Jørgensen JO, Hassager C, Christiansen C. The effect of growth hormone (GH) therapy on urinary pyridinoline cross-links in GH-deficient adults. *Clin Endocrinol (Oxf)*, 1991; **35**: 471–6.
7. Jørgensen JO, Pedersen SA, Laurberg P, Weeke J, Skakkebaek NE, Christiansen JS. Effects of growth hormone therapy on thyroid function of growth hormone-deficient adults with and without concomitant thyroxine-substituted central hypothyroidism. *J Clin Endocrinol Metab*, 1989; **69**: 1127–32.
8. Pedersen SA, Welling K, Michaelsen KF, Jørgensen JO, Christiansen JS, Skakkebaek NE. Reduced sweating in adults with growth hormone deficiency. *Lancet*, 1989; **2**: 681–2.
9. Salomon F, Cuneo RC, Hesp R, Sonksen PH. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. *N Engl J Med*, 1989; **321**: 1797–803.
10. Consensus guidelines for the diagnosis and treatment of adults with growth hormone deficiency: summary statement of the Growth Hormone Research Society Workshop on Adult Growth Hormone Deficiency. *J Clin Endocrinol Metab*, 1998; **83**: 379–81.
11. Toogood AA, Beardwell CG, Shalet SM. The severity of growth hormone deficiency in adults with pituitary disease is related to the degree of hypopituitarism. *Clin Endocrinol (Oxf)*, 1994; **41**: 511–16.
12. Klose M, Stochholm K, Janukonyte J, *et al*. Prevalence of posttraumatic growth hormone deficiency is highly dependent on the diagnostic set-up: results from The Danish National Study on Posttraumatic Hypopituitarism. *J Clin Endocrinol Metab*, 2014; **99**: 101–10.
13. Stochholm K, Gravholt CH, Laursen T, *et al*. Incidence of GH deficiency—a nationwide study. *Eur J Endocrinol*, 2006; **155**: 61–71.
14. Cuneo RC, Salomon F, McGauley GA, Sonksen PH. The growth hormone deficiency syndrome in adults. *Clin Endocrinol (Oxf)*, 1992; **37**: 387–97.
15. Markussis V, Beshyah SA, Fisher C, Sharp P, Nicolaides AN, Johnston DG. Detection of premature atherosclerosis by high-resolution ultrasonography in symptom-free hypopituitary adults. *Lancet*, 1992; **340**: 1188–92.
16. Rosen T, Bengtsson BA. Premature mortality due to cardiovascular disease in hypopituitarism. *Lancet*, 1990; **336**: 285–8.
17. Rosen T, Bosaeus I, Tolli J, Lindstedt G, Bengtsson BA. Increased body fat mass and decreased extracellular fluid volume in adults with growth hormone deficiency. *Clin Endocrinol (Oxf)*, 1993; **38**: 63–71.
18. Moller J, Frandsen E, Fisker S, Jørgensen JO, Christiansen JS. Decreased plasma and extracellular volume in growth hormone deficient adults and the acute and prolonged effects of GH administration: a controlled experimental study. *Clin Endocrinol (Oxf)*, 1996; **44**: 533–9.
19. Moller J, Jørgensen JO, Frandsen E, Laursen T, Christiansen JS. Body fluids, circadian blood pressure and plasma renin during growth hormone administration: a placebo-controlled study with two growth hormone doses in healthy adults. *Scand J Clin Lab Invest*, 1995; **55**: 663–9.
20. Juul A, Skakkebaek NE. Growth hormone deficiency and hyperthermia. *Lancet*, 1991; **338**: 887.
21. Juul A, Behrenscheer A, Tims T, Nielsen B, Halkjaer-Kristensen J, Skakkebaek NE. Impaired thermoregulation in adults with growth hormone deficiency during heat exposure and exercise. *Clin Endocrinol (Oxf)*, 1993; **38**: 237–44.
22. Juul A, Hjortskov N, Jepsen LT, *et al*. Growth hormone deficiency and hyperthermia during exercise: a controlled study of sixteen GH-deficient patients. *J Clin Endocrinol Metab*, 1995; **80**: 3335–40.
23. Nielsen S, Moller N, Christiansen JS, Jørgensen JO. Pharmacological antilipolysis restores insulin sensitivity during growth hormone exposure. *Diabetes*, 2001; **50**: 2301–8.
24. Rabinowitz D, Zierler KL. A metabolic regulating device based on the actions of human growth hormone and of insulin, singly and together, on the human forearm. *Nature*, 1963; **199**: 913–15.
25. Moller N, Jørgensen JO. Effects of growth hormone on glucose, lipid, and protein metabolism in human subjects. *Endocr Rev*, 2009; **30**: 152–77.
26. Nicolson A, Toogood AA, Rahim A, Shalet SM. The prevalence of severe growth hormone deficiency in adults who received growth hormone replacement in childhood [see comment]. *Clin Endocrinol (Oxf)*, 1996; **44**: 311–16.
27. Ho KK. Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, Japan Endocrine Society, and Endocrine Society of Australia. *Eur J Endocrinol*, 2007; **157**: 695–700.
28. Maison P, Griffin S, Nicoue-Beglah M, Haddad N, Balkau B, Chanson P. Impact of growth hormone (GH) treatment on cardiovascular risk factors in GH-deficient adults: a metaanalysis of blinded, randomized, placebo-controlled trials. *J Clin Endocrinol Metab*, 2004; **89**: 2192–9.
29. Rubeck KZ, Bertelsen S, Vestergaard P, Jørgensen JO. Impact of GH substitution on exercise capacity and muscle strength in GH-deficient adults: a meta-analysis of blinded, placebo-controlled trials. *Clin Endocrinol (Oxf)*, 2009; **71**: 860–6.
30. Widdowson WM, Gibney J. The effect of growth hormone replacement on exercise capacity in patients with GH deficiency: a metaanalysis. *J Clin Endocrinol Metab*, 2008; **93**: 4413–17.
31. Barake M, Klibanski A, Tritos NA. Effects of recombinant human growth hormone therapy on bone mineral density in adults with growth hormone deficiency: a meta-analysis. *J Clin Endocrinol Metab*, 2014; **99**: 852–60.
32. Xue P, Wang Y, Yang J, Li Y. Effects of growth hormone replacement therapy on bone mineral density in growth hormone deficient adults: a meta-analysis. *Int J Endocrinol*, 2013; 2013: 216107.
33. Hazem A, Elamin MB, Bancos I, *et al*. Body composition and quality of life in adults treated with GH therapy: a systematic review and meta-analysis. *Eur J Endocrinol*, 2012; **166**: 13–20.
34. Maison P, Chanson P. Cardiac effects of growth hormone in adults with growth hormone deficiency: a meta-analysis. *Circulation*, 2003; **108**: 2648–52.
35. Cuneo RC, Judd S, Wallace JD, *et al*. The Australian multicenter trial of growth hormone (GH) treatment in GH-deficient adults. *J Clin Endocrinol Metab*, 1998; **83**: 107–16.

36. Johannsson G, Bjarnason R, Bramnert M, *et al*. The individual responsiveness to growth hormone (GH) treatment in GH-deficient adults is dependent on the level of GH-binding protein, body mass index, age, and gender. *J Clin Endocrinol Metab*, 1996; **81**: 1575–81.
37. Burman P, Johansson AG, Siegbahn A, Vessby B, Karlsson FA. Growth hormone (GH)-deficient men are more responsive to GH replacement therapy than women. *J Clin Endocrinol Metab*, 1997; **82**: 550–5.
38. Deijen JB, Arwert LI, Witlox J, Drent ML. Differential effect sizes of growth hormone replacement on quality of life, well-being and health status in growth hormone deficient patients: a meta-analysis. *Health Qual Life Outcomes*, 2005; **3**: 63.
39. Arwert LI, Deijen JB, Witlox J, Drent ML. The influence of growth hormone (GH) substitution on patient-reported outcomes and cognitive functions in GH-deficient patients: a meta-analysis. *Growth Horm IGF Res*, 2005; **15**: 47–54.
40. Falleti MG, Maruff P, Burman P, Harris A. The effects of growth hormone (GH) deficiency and GH replacement on cognitive performance in adults: a meta-analysis of the current literature. *Psychoneuroendocrinology*, 2006; **31**: 681–91.
41. Burman P, Broman JE, Hetta J, *et al*. Quality of life in adults with growth hormone (GH) deficiency: response to treatment with recombinant human GH in a placebo-controlled 21-month trial. *J Clin Endocrinol Metab*, 1995; **80**: 3585–90.
42. NICE Clinical Guidelines. Available at: https://www.nice. org. uk/guidance/ ta64/documents/ growth-hormone-deficiency-adults-human-growth-hormone-appendix-b-proposal-paper-presented-to-the-institutes-guidance-executive2
43. Bidlingmaier M, Friedrich N, Emeny RT, *et al*. Reference intervals for insulin-like growth factor-1 (igf-i) from birth to senescence: results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations. *J Clin Endocrinol Metab*, 2014; **99**: 1712–21.
44. Vahl N, Juul A, Jørgensen JO, Orskov H, Skakkebaek NE, Christiansen JS. Continuation of growth hormone (GH) replacement in GH-deficient patients during transition from childhood to adulthood: a two-year placebo-controlled study. *J Clin Endocrinol Metab*, 2000; **85**: 1874–81.
45. Norrelund H, Vahl N, Juul A, *et al*. Continuation of growth hormone (GH) therapy in GH-deficient patients during transition from childhood to adulthood: impact on insulin sensitivity and substrate metabolism. *J Clin Endocrinol Metab*, 2000; **85**: 1912–17.
46. Veldhuis JD, Roelfsema F, Keenan DM, Pincus S. Gender, age, body mass index, and IGF-I individually and jointly determine distinct GH dynamics: analyses in one hundred healthy adults. *J Clin Endocrinol Metab*, 2011; **96**: 115–21.
47. Vahl N, Jørgensen JO, Skjaerbaek C, Veldhuis JD, Orskov H, Christiansen JS. Abdominal adiposity rather than age and sex predicts mass and regularity of GH secretion in healthy adults. *Am J Physiol*, 1997; **272**(6 Pt 1):E1108–16.
48. Liu H, Bravata DM, Olkin I, *et al*. Systematic review: the safety and efficacy of growth hormone in the healthy elderly. *Ann Int Med*, 2007; **146**: 104–15.
49. Toogood AA, O'Neill PA, Shalet SM. Beyond the somatopause: growth hormone deficiency in adults over the age of 60 years. *J Clin Endocrinol Metab*, 1996; **81**: 460–5.
50. Toogood AA, Shalet SM. Growth hormone replacement therapy in the elderly with hypothalamic-pituitary disease: a dose-finding study. *J Clin Endocrinol Metab*, 1999; **84**: 131–6.
51. Fernholm R, Bramnert M, Hagg E, *et al*. Growth hormone replacement therapy improves body composition and increases bone metabolism in elderly patients with pituitary disease. *J Clin Endocrinol Metab*, 2000; 85: 4104–12.
52. Franco C, Johannsson G, Bengtsson B-Ak, Svensson J. Baseline characteristics and effects of growth hormone therapy over two years in younger and elderly adults with adult onset GH deficiency. *J Clin Endocrinol Metab*, 2006; **91**: 4408–14.
53. de Boer H, Blok GJ, Van der Veen EA. Clinical aspects of growth hormone deficiency in adults. *Endocr Rev*, 1995; **16**: 63–86.
54. Jørgensen JO. Human growth hormone replacement therapy: pharmacological and clinical aspects. *Endocr Rev*, 1991; **12**: 189–207.
55. Pappachan JM, Raskauskiene D, Kutty VR, Clayton RN. Excess mortality associated with hypopituitarism in adults: a meta-analysis of observational studies. *J Clin Endocrinol Metab*, 2015; **100**: 1405–11.
56. Dal J, Leisner MZ, Hermansen K, *et al*. Cancer incidence in patients with acromegaly: a cohort study and meta-analysis of the literature. *J Clin Endocrinol Metab*, 2018; **103**: 2182–8.
57. Holdaway IM, Bolland MJ, Gamble GD. A meta-analysis of the effect of lowering serum levels of GH and IGF-I on mortality in acromegaly. *Eur J Endocrinol*, 2008; **159**: 89–95.
58. Burgers AM, Biermasz NR, Schoones JW, *et al*. Meta-analysis and dose-response metaregression: circulating insulin-like growth factor I (IGF-I) and mortality. *J Clin Endocrinol Metab*, 2011; **96**: 2912–20.
59. Stochholm K, Berglund A, Juul S, Gravholt CH, Christiansen JS. Socioeconomic factors do not but GH treatment does affect mortality in adult-onset growth hormone deficiency. *J Clin Endocrinol Metab*, 2014; **99**: 4141–8.
60. Olsson DS, Trimpou P, Hallen T, *et al*. Life expectancy in patients with pituitary adenoma receiving growth hormone replacement. *Eur J Endocrinol*, 2017; **176**: 67–75.
61. Berglund A, Gravholt CH, Olsen MS, Christiansen JS, Stochholm K. Growth hormone replacement does not increase mortality in patients with childhood-onset growth hormone deficiency. *Clin Endocrinol (Oxf)*, 2015; **83**: 677–83.
62. Mo D, Fleseriu M, Qi R, *et al*. Fracture risk in adult patients treated with growth hormone replacement therapy for growth hormone deficiency: a prospective observational cohort study. *Lancet Diabetes Endocrinol*, 2015; **3**: 331–8.
63. Stochholm K, Johannsson G. Reviewing the safety of GH replacement therapy in adults. Growth hormone & IGF Res, 2015; **25**: 149–57.

2.3.7 Surgery of Pituitary Tumours

David L. Penn, Caroline S. Repetti, and Edward R. Laws Jr

Introduction

Among the numerous different pathologies, pituitary tumours represent the second most common primary brain tumour found in human beings, with an incidence of 10–15% [1, 2]. More interestingly, there exists an extremely wide variety of tumour types deriving from this small organ located in the anterior portion of the skull base, near the centre of the brain (Table 2.3.7.1). These neoplasms were recently re-classified in the 2017 World Health Organization Classification of Tumours of the Endocrine Organs [3]. By far the most prevalent type are the pituitary adenomas, a neuroendocrine tumour, which occurs with an incidence ranging from 14.4–22.5% [1, 4]. Pituitary adenomas fall into two general categories, functional tumours that present from hypersecretion of different hormone products and non-functioning tumours that present with symptoms caused by mass effect exerted on surrounding structures. Similar to non-functioning pituitary adenomas, there are a category of non-adenomatous tumours and cysts from and within the pituitary gland that present with symptoms of mass effect. Among the more common of these include craniopharyngiomas and Rathke's cleft cysts [2, 5–8]. Lastly, there are a number of relatively less common pituitary lesions that can present with various forms of hypopituitarism, such as adrenal insufficiency or diabetes insipidus, from destruction of the normal gland, or again from mass effect as the tumour grows and compresses the normal surrounding anatomy.

Table 2.3.7.1 Tumours of the sellar region

Tumours of the sellar region	
Neuroendocrine tumours	• Pituitary adenoma – Null cell adenoma – Gonadotroph adenoma – Somatotroph adenoma – Corticotroph adenoma – Lactotroph adenoma – Thyrotroph adenoma • Pituitary carcinoma
Non-neuroendocrine tumours	• Pituicytoma • Granular cell tumour • Spindle cell oncocytoma • Gangliocytoma • Astrocytoma (posterior pituitary)
Tumours of non-pituitary origin	• Craniopharyngioma – Adamantinomatous – Papillary • Meningioma • Chordoma • Langerhans cell histiocytosis • Metastatic tumour • Germinoma • Schwannoma
Cystic lesions	• Rathke's cleft cyst • Arachnoid cyst • Epidermal/dermoid cyst
Inflammatory lesions	• Lymphocytic hypophysitis • Granulomatous hypophysitis • Xanthomatous hypophysitis • Sarcoidosis
Vascular lesions	• Cavernous angioma • Aneurysm • Pituitary tumour apoplexy

Although a few pituitary tumours are managed primarily with medical or radiation therapy, for a majority of them, surgical resection still represents the first-line treatment [9–11]. The primary role of surgical management is for maximal safe resection and decompression of mass effect on the surrounding neurovascular structures, most importantly, the optic nerves and chiasm that course superiorly to the tumours that arise within the sella and extend into the suprasellar space. A secondary role for surgical resection is the removal of a hyperfunctioning pituitary adenoma. With the exception of prolactinomas, surgical resection remains the primary treatment modality for functioning tumours, such as suspected growth hormone (GH) and adrenocorticotrophic hormone (ACTH) adenomas [9, 11, 12]. These subtypes of tumours do not always present with symptoms of mass effect secondary to excessive growth, but will present secondary to the downstream effects of hypersecretion of their respective hormones: acromegaly and Cushing's disease, respectively. In these cases, to offer patients the best chance of remission from these syndromes caused by hormonal excess, it is imperative to attempt to remove the entirety of the tumour, balancing the risks of gross total removal with possible complications of an aggressive operation. Lastly, surgery of pituitary tumours plays a role in diagnosis of some of the less common tumour entities that have equivocal clinical and radiographic presentations. In this setting, biopsy and histopathological analysis can help guide clinicians to proceed with the most appropriate medical management or radiotherapies to achieve optimal patient outcomes.

Multiple different surgical approaches exist for operating upon pituitary tumours, often chosen based on individual surgeon preference, that include transcranial and endonasal approaches. The most common procedure for surgery of the pituitary gland is by far the endonasal transsphenoidal approach, using either the operative microscope or endoscope for visualization. Although some of the subtleties of these two techniques differ, both provide entry to the sella and the pituitary gland through the sphenoid sinus, minimizing retraction of normal brain and providing direct access to the pathology.

This chapter will discuss the indications, goals, and outcomes of surgical intervention for tumours of the pituitary gland, with particular focus on the most common tumour types. The role of multiple different surgical approaches in the treatment of pituitary tumours will be reviewed to provide both medical and surgical providers with an overview of how these different techniques play a role in delivering safe and efficacious treatment to patients affected with these lesions.

Clinical Presentation and Surgical Indications

Pituitary tumours can manifest in multiple ways, ranging from symptoms caused by mass effect on the surrounding, normal neurovascular structures and pituitary gland, as well as the plethora of symptoms and comorbidities from hormonal hypersecretion. For the majority of pituitary tumours, symptomatic lesions warrant surgical resection as first-line treatment, with the exception of prolactinomas, which are often successfully managed medically with dopamine agonists [9–12]. Given that the vast majority of pituitary tumours are benign, asymptomatic lesions, especially the pituitary adenomas and Rathke's cleft cysts, many can be managed

conservatively with interval neuroimaging studies [13]. Some more aggressive lesions, such as craniopharyngiomas, should undergo attempted gross total resection, even if they are asymptomatic; however, because of the relatively low incidence of these lesions, clinical guidelines have not been delineated and are a subject of continuing debate [14–17]. Lastly, any pituitary lesion where the diagnosis is unclear and may be amenable to medical management or chemo/radiotherapy may require surgical biopsy for definitive diagnosis.

Mass Effect

As lesions gradually grow in size, they can exert effects on the normal surrounding structures in the sellar and suprasellar space. Initially, as the tumour grows it can begin to stretch the surrounding dural layers, in particular, the diaphragma sellae, which can result in intractable headaches [18]. The pathophysiological mechanism for headaches is based on the fact that the dura has nociceptive fibres within its layers transmitting pain through the sensory distribution of the trigeminal nerve. Although pituitary patients can present with almost any type of headache, headaches caused by lesions arising from the sella are often referable to the frontal region or behind the eyes, and sometimes to the occipital region. Work up of any type of headache, including migrainous or tension-type headaches, can uncover a sellar lesion, but often these headache types are not believed to be related to these lesions. Headache can be a chief complaint of 16.0–62.1% of patients found to have non-functional adenomas [19–21].

Pituitary tumours can often cause neurological deficit as they exert pressure on the surrounding cranial nerves. In particular, the optic nerves that travel superiorly to the sella are at risk. Compression of the optic nerves classically can result in a bitemporal hemianopsia; however, based on the growth patterns of pituitary tumours and their relation to the optic nerves and chiasm many different types of visual field and optic nerve deficits can be detected. Bitemporal hemianopsia is the result of mid-chiasmal compression and classically progresses from the superior temporal fields into the inferior temporal fields as the chiasm becomes thinned and more compressed. Compression of the anterior chiasm can cause junctional scotoma and pressure on the posterior chiasm or optic tracts can cause an asymmetric homonymous hemianopsia. Of patients with non-functioning adenomas, up to 60% present with visual deficits [19–21]. In addition to visual field defects, patients can experience decreased visual acuity and loss of red-green distinction when the optic nerves are compressed. Pituitary tumours invading the cavernous sinus can compress cranial nerves III through VI which run within this venous structure resulting in complete or partial ophthalmoplegia or loss of facial sensation, or facial pain. Very large tumours can exert mass effect on the brain stem posteriorly or obstruct cerebrospinal fluid (CSF) flow causing obstructive hydrocephalus.

Compression of the normal pituitary gland by enlarging tumours can result in signs of hypopituitarism and is found in 30–40% of patients with non-functioning adenomas [19–21]. In general, gonadotroph dysfunction is the first manifestation of hypopituitarism causing decreased sexual function or libido in men or amenorrhoea in women. As more damage is done to the pituitary gland, thyrotrophic, somatotropic, and corticotropic function can be lost. It is very uncommon for pituitary adenomas to cause symptomatic compression of the posterior pituitary gland, resulting in diabetes insipidus (DI). Patients presenting with DI are more likely to have lesions originating from the remnant Rathke's pouch, such as craniopharyngiomas or Rathke's cleft cysts, or invasive pituitary tumours that cause destruction of the gland. Of patients found to have craniopharyngiomas, 17–27% can present with DI, which is less prevalent with Rathke's cleft cysts [16, 22, 23]. In contrast, one presentation of mass effect on the normal pituitary gland is hyperprolactinemia which is caused by interruption of inhibitory input from the hypothalamus through distortion or compression of the infundibulum as it travels from the hypothalamus to the anterior pituitary gland. Hyperprolactinemia often results in galactorrhoea and amenorrhoea in women.

Hypersecretion

There exist a number of subtypes of pituitary adenomas that produce hormone products causing elevated systemic levels of the hormone, most commonly somatotroph, corticotroph, and lactotroph adenomas, less commonly thyrotrophs and gonadotrophs. Somatotroph adenomas cause excess growth hormone levels and acromegaly. Symptoms of acromegaly include coarsening of facial features, increased hand or foot size, joint pains, skin tags, snoring, diabetes mellitus, hypertension, and cardiomegaly. Patients with corticotroph adenomas, or Cushing's disease, experience elevated levels of ACTH and cortisol causing central obesity, moon facies, thinning of the skin and easy bruising, joint pains, violaceous striae, increased susceptibility to infections, amenorrhoea in women, hypertension, osteopaenia, and diabetes mellitus, which can greatly affect quality of life and lead to medical morbidity and premature mortality [24, 25]. Lastly, lactotroph adenomas present with hyperprolactinemia, which, as previously mentioned, can induce galactorrhoea and breast tenderness in both men and women, amenorrhoea in women, and signs of hypogonadism in men. Although hyperprolactinemia can be a result of both a hypersecreting tumour and a tumour-causing stalk effect, serum prolactin levels in prolactinomas are generally considered to be much higher, with the existing literature claiming levels greater than 150 to 250 as pathognomonic [26–28] of tumour rather than stalk effect. In contrast to non-functioning tumours, hypersecretory tumours commonly present smaller in size and with fewer symptoms of mass effect. The symptoms caused by the clinical syndromes they produce are often detected and diagnosed before they grow to such a size.

Incidentalomas

With continually increasing utilization of imaging modalities for a multitude of symptoms, as well as improved resolution of these techniques, the incidence of incidentally discovered pituitary tumours is increasing [13]. Tumours that are small in size, do not show signs of imminent compression of the optic chiasm, and are determined to not be attributable to the patient's presenting symptoms can be managed conservatively, following them for clinical changes and interval neuroimaging studies to ensure that the lesion is not enlarging.

Pituitary Tumour Apoplexy

Pituitary tumours, in particular pituitary adenomas, as they grow can outstrip or compress their blood supply, causing the tumour itself to become necrotic and swell in size or to haemorrhage from defective tumour blood vessels. The occurrence of this phenomenon can result in a syndrome of acute onset, with severe headache, visual

loss, nausea/vomiting, and altered mental status, known as pituitary apoplexy [29, 30].

Although uncommon, this presentation represents a clinical emergency, requiring hormone replacement, especially of the adrenal axis, and when presenting with neurological deficit, such as visual loss or ophthalmoplegia, should be managed with emergent surgical decompression and resection of the tumour [31, 32].

Clinical Evaluation and Preoperative Planning

All patients with pituitary lesions should undergo a full endocrine laboratory panel to check for signs of hypopituitarism or for the presence of a potential functional pituitary adenoma. Serum prolactin levels should be performed with dilution to ensure that falsely low levels are not the result of the 'Hook Effect', a phenomenon caused by flooding of the detection assay with extraordinarily high prolactin levels. Patients with signs and symptoms of hypopituitarism should be managed appropriately with hormone replacement. A particular situation is when hypothyroidism or adrenal insufficiency is noted. Rapid thyroid repletion without regard to the cortisol status may increase metabolic stress and precipitate adrenal insufficiency which must be optimized preoperatively.

When feasible, non-urgent patients with clinical vision loss should undergo a full neuro-ophthalmologic evaluation to establish a preoperative baseline. Evaluation should include pupillary responsiveness, range of extraocular muscle movement, Humphrey automated perimetric assessment, and optical coherence tomography to assess thinning of the retinal nerve fibre layer.

As part of both the diagnostic and preoperative evaluation, all patients with suspected pituitary tumours should undergo high-resolution imaging. Although computed tomography (CT) can be helpful in the initial detection of pituitary tumours, and can be helpful in appreciating calcifications within the tumour to help determine whether the lesion may represent a craniopharyngioma, pituitary focused magnetic resonance imaging (MRI) is the ideal study (Figure 2.3.7.1). MRI of the brain with and without gadolinium and with high-resolution sequences through the sella provides the best anatomic delineation of the lesion and the surrounding normal structures, particularly the carotid arteries adjacent to the sella. This allows for improved surgical planning in order to determine the best surgical approach to safely remove the lesion in its entirety, as well as allowing surgeons to recognize obstacles such as a narrow intercarotid artery distance or the relationship between the tumour and the optic apparatus that could make surgery more challenging. Additionally, high-resolution sequences can be used with many computerized neuronavigation systems which allow surgeons to localize instruments and maintain orientation to vital structures in real time throughout the operation.

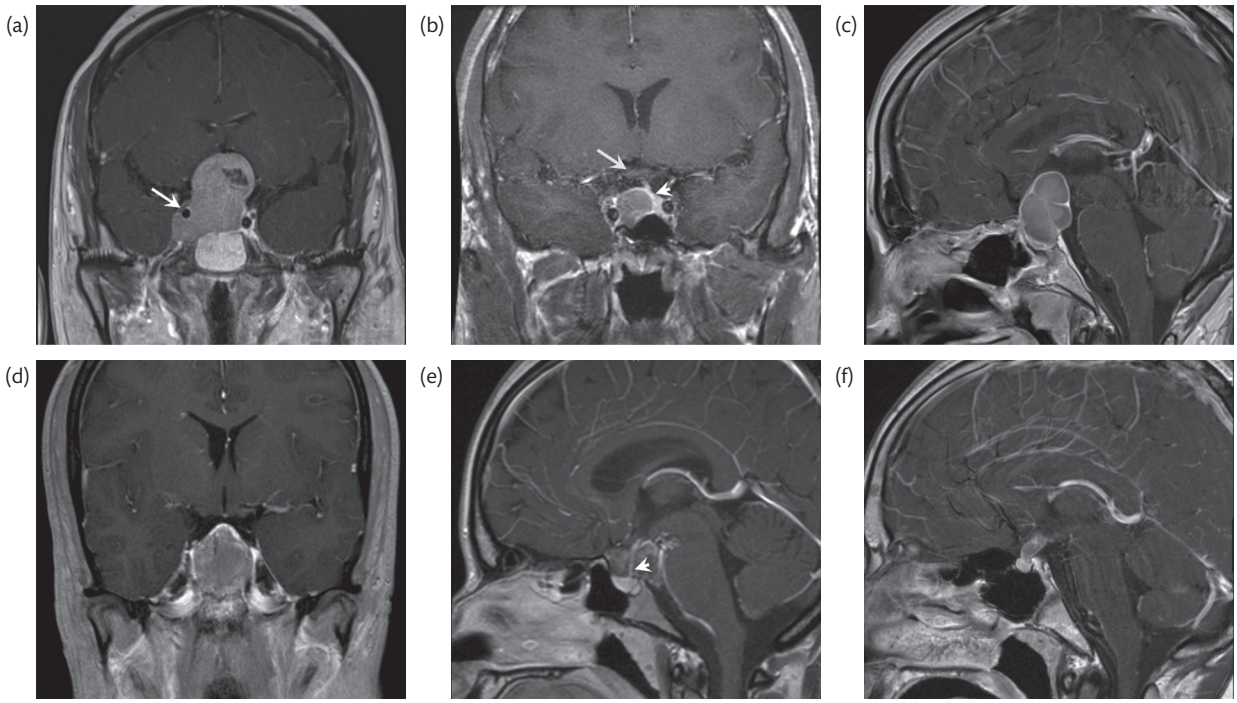


Figure 2.3.7.1 Contrast enhanced MRI Brain demonstrating a number of different pituitary tumours. (a) Shows a large homogenously enhancing gonadotroph adenoma with a hypointense cystic component representing prior haemorrhage. The lesion encases the carotid artery on the right (white arrow) and displaces the optic chiasm superiorly (which is not visible radiographically secondary to severe thinning). (b) Illustrates a somatotroph adenoma growing from the pituitary gland, displacing the normal gland to the left (white arrowhead). There is no compression of the optic chiasm (White arrow). (c) Reveals a large septated hypointense Rathke's cleft cyst causing compression of the optic apparatus and bitemporal hemianopsia. (d) Demonstrates a heterogeneously enhancing craniopharyngioma. The normal pituitary gland is not visible because it has been engulfed by the tumour. (e) Exhibits a heterogeneously enhancing, dermoid cyst in the suprasellar space, surrounding the infundibulum (white arrowhead). (f) Displays a hypointense lesion, expanding the pituitary stalk, with small punctate points of enhancement, consistent with suprasellar germinoma.

Surgical Approaches

There are two major groups of surgical approaches for lesions of the pituitary gland: transcranial and endonasal. Although the endonasal transsphenoidal approach is by far the most popular, there still exist indications and a role for transcranial surgery of the pituitary gland.

Transcranial Surgery

Transcranial surgery for the pituitary gland involves a scalp incision and creation of a craniotomy to access the pituitary gland in the sella. These approaches take advantage of fissures between the lobes of the brain and the subarachnoid spaces to avoid transgression through normal brain parenchyma. A number of common approaches used for transcranial surgery of the pituitary gland include the pterional craniotomy and its transsylvian approach, the cranio-orbitozygomatic approach accessing the pituitary gland via the subfrontal or transsylvian route, and bifrontal craniotomy via the subfrontal route [33, 34], and the so-called supraorbital subfrontal approach through an eyebrow incision. Transcranial surgery for tumours of the pituitary gland is most useful in lesions with extensive suprasellar extension. Examples include giant pituitary adenomas, craniopharyngiomas with invasion posteriosuperiorly within the third ventricle and/or hypothalamus, biopsy for tumours or lesions of the pituitary stalk, and recurrent Rathke's cleft cysts requiring resection of the cyst wall. The advantages of craniotomy for pituitary tumours include increased exposure and control of the major arteries of the circle of Willis and improved visualization for dissecting tumour planes away from major neurovascular structures that are more poorly visualized from an endonasal approach. The unique disadvantages of these approaches include increased risk to normal brain from retraction, and the need for a visible incision on the scalp; however, the overall risk of major neurovascular injury or mortality is approximately the same as with the endonasal approach.

Endonasal Transsphenoidal

This surgical approach involves exposing and entering the sphenoid sinus, and subsequently the sella through the nasal cavities (Figure 2.3.7.2). Enlargement of the opening in the sphenoid sinus allows access laterally to tumours that have invaded the cavernous sinuses, superiorly for tumours that extend to the tuberculum sellae, and inferiorly to the clivus for tumours invading the clival bone or with extension to the pons. This approach allows for removal of very large lesions given that they tend to arise from within the sella. As they grow, they tend to extend superiorly, displacing normal anatomy. In particular, pituitary adenomas, as well as Rathke's cleft cysts, have a very soft and sometimes liquid consistency that allows relatively easy removal from the smaller cranial opening created by transsphenoidal surgery. The tumour can then be separated from the normal surrounding gland using a number of blunt curettes, minimizing damage to the surrounding normal gland. Further advances in the transsphenoidal technique have allowed for effective resection of the more solid and adherent tumours; however, preoperative imaging should be carefully studied for appropriate selection of these cases to ensure that the tumour can be removed without significant risk to the surrounding neurovascular structures, such as the cranial nerves and carotid arteries. For instance, retrochiasmatic craniopharyngiomas with suprasellar extension, tumours arising superiorly to the chiasm from the infundibulum, and giant pituitary adenomas with growth into the temporal lobe, should be considered for a transcranial approach to avoid crossing these vital structures with instruments, placing the patient at risk for a major complication.

The endonasal transsphenoidal operation can be performed with the operating microscope or endoscope (2D or 3D) for visualization. Both techniques place the surgeon at the face of the sella, operating through the nasal cavities; however, there exist subtleties in the steps of the procedure and the visualization offered between the two. Performing the procedure with the operating microscope involves placement of a nasal speculum along the perichondral surface of the septal mucosa. The speculum is placed within the sphenoid sinus at the face of the sella, and opened to enter the sinus and provide appropriate exposure. With the speculum in place, the microscope offers a lighted but necessarily narrowed corridor to the sellar opening. Technically, there are a few disadvantages with this approach, such as a smaller field of vision, limited illumination, and collision of instruments with the microscope in place. When using the endoscopic approach, the endoscope is placed through the nasal cavity and the posterior bony septum is removed posterior to anterior. The sphenoid sinus is opened and then subsequently the sellar floor is removed. Disadvantages with the 2D endoscopic approach include decreased three-dimensional perspective when operating while looking at a two-dimensional monitor; however, this disadvantage is mitigated by the use of three-dimensional endoscopes [35]. Advantages to the endoscope include improved illumination with the light source at the tip of the camera and panoramic visualization with the image coming directly from the operative field. Furthermore, advances in technique using the endoscope have allowed for the development of extended endoscopic approaches with more access to the anterior and middle skull base, allowing for the removal of larger and more extensive and invasive tumours than previously.

Both approaches are safe and effective for the resection of appropriately selected pituitary tumours. The use of the operative microscope or the endoscope has become largely one of surgeon preference, recognizing that there is a lack of Level I evidence to prove the superiority of one technique over another. Studies comparing the two techniques in patients with pituitary adenomas have shown no statistically significant differences in outcomes, including extent of resection, endocrinologic function, visual improvements, and postoperative CSF leak rates [35–38]. Similar results were found comparing techniques in patients with secretory tumours, such as ACTH adenomas, as well as similar remission and recurrence rates [39].

Postoperative Care

Patients undergoing surgery for pituitary tumours have a number of unique considerations relating to the immediate postoperative period. At our institution, patients undergoing transsphenoidal surgery are transferred to the postoperative anaesthesia unit and then our neurosurgical floor, while patients undergoing transcranial surgery are monitored in the intensive care unit. Neurological status,

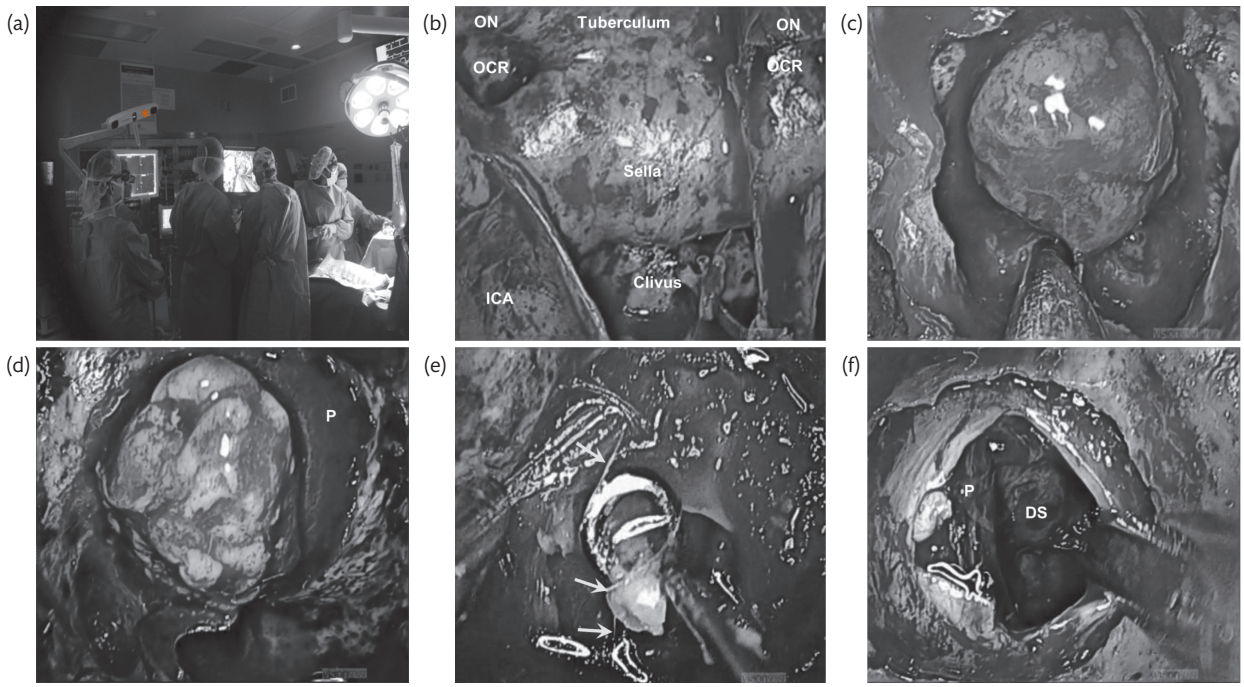


Figure 2.3.7.2 Intraoperative photographs. (a) Demonstrates the operative team at our institution and the set-up of the operating room. With an assistant holding the endoscope through a binostral approach, the primary surgeon can use bimanual technique to perform the operation. The monitor for the endoscope is positioned directly in front of the surgeons with the neuro-navigation system (orange asterix) to the left and the instrument table and nursing team to the right. (b) Exhibits the face of the sella and relevant anatomy. ON, optic nerve; OCR, opticocarotid recess; ICA, internal carotid artery. (c) Shows the removal of a gonadotroph macroadenoma. Intracranial pressure aids in tumour removal pushing the tumour out of the sella and down from the suprasellar space. (d) Illustrates resection of a right-sided, somatotroph adenoma. Note the classic, white appearance of the tissue and the normal pituitary gland (P) on the right. (e) Shows resection of a suprasellar dermoid cyst using blunt curettage. The classic fatty, white appearance of the contents is shown, as well as a number of hairs growing within the tumour (white arrows). (f) Shows the resection cavity once the tumour has been removed. Prolapse of the diaphragma sellae (DS) is a good sign that the nerves superiorly have been decompressed. Normal pituitary gland (P) can be observed on the right.

in particular visual fields and acuity, is monitored closely for signs of changes.

One of the primary concerns following pituitary surgery is the development of either DI or the syndrome of inappropriate antidiuretic hormone secretion (SIADH). During the initial postoperative period, strict fluid intake and output measurements are recorded, and patients undergo regular serum sodium evaluations and urinalysis for determination of specific gravity. Mild DI is managed conservatively with oral fluid replacement, as this is often temporary and self-limited. Patients with more severe DI or DI that is significantly interrupting sleep are treated with oral or IV desmopressin that is titrated to the patient's symptoms. Signs of SIADH, as indicated by progressive decreases in or less than normal serum sodium values are managed with strict fluid restriction. In severe cases, hypertonic 3% saline may also be administered. Patients are at highest risk for SIADH between postoperative days 2 and 10 [40]. Because our patients are normally hospitalized for two to three days, we prophylactically discharge patients on a 1 L per day fluid restriction, which has brought our hospital readmission rate for hyponatraemia to zero [41].

Patients with preoperative laboratory evidence of adrenal insufficiency receive steroid replacement at the time of surgery with hydrocortisone. Hydrocortisone is then tapered to physiological levels over the next few days. The total duration of hydrocortisone replacement is determined based on the clinical and biochemical severity of adrenal insufficiency and need for replacement. Conversely, patients not requiring steroid replacement are monitored daily with fasting morning cortisol levels to ensure that hypopituitarism does not develop as a complication of the operation.

After discharge from the hospital, patients return to the office, one week after their procedure, for a serum sodium check, as well as potentially other short-term laboratory studies that may affect clinical management, such as fasting morning cortisol. At 6 weeks postoperatively, patients return for follow-up with a full endocrine panel and at three months for a postoperative MRI of the brain, to assess the extent of tumour resection and to establish a baseline for annual routine imaging going forward.

Outcomes

Overall, the surgical management of pituitary adenomas is successful in relieving symptoms of mass effect, such as headache and visual loss, as well as achieving control of symptoms from functional tumours, with acceptable morbidity and mortality. In patients with non-functioning adenomas, relief of symptoms caused by mass effect are reported as high as 90% [19, 20, 42–46]. Restoration of normal pituitary gland function has been reported as high as 48% [31, 47, 48]. Furthermore, one study demonstrated superiority of the transsphenoidal approach in improving and preserving pituitary function [49]. Although surgical intervention provides excellent results for symptoms of mass effect, functional adenomas can be

more complex, since relief of symptoms and achievement of lasting remission may require removal of every single hyperfunctioning cell, which can be extremely difficult for aggressive and invasive tumours. Despite the challenge, surgery is still effective for Cushing's disease, with initial remission rates reported as high as 95% in selected cases, and acromegaly, with remission rates as high as 70% [50–55]. Recurrence rates can be affected by multiple factors, including extent of resection, mitotic index, invasiveness, tumour type, and the number of prior procedures; however, for pituitary adenomas, the rate of recurrence is generally accepted to be approximately 10% at 10 years, with ACTH adenomas being among the more challenging, with long-term recurrence rates approximating as high as 30% [20, 56–60].

Outcomes for other common tumour types, such as craniopharyngiomas, are more complex. Similar to pituitary adenomas, surgery is effective for relief of symptoms caused by mass effect. Rates of visual improvement have been reported to be as high as 69.2% [61, 62]. Because of the tumours' intimate association with the pituitary gland, infundibulum, and hypothalamus, the rates of pituitary gland preservation are lower, as evidenced by the rates of DI being reported as high as 93% [14, 23, 61, 63]. In some cases, in order to achieve definitive resection of the tumour, the infundibulum must be sharply divided which can result in panhypopituitarism. Dissection of the tumour from the hypothalamus can also result in hormonal dysfunction and subsequent obesity, especially in paediatric patients [61, 62, 64]. In addition, the extent of resection greatly influences recurrence rates. Following subtotal resection, ultimate progression rates can be as high as 90% [61, 65].

Complications

Although complications exist in transcranial and endonasal approaches for pituitary surgery, the notable complications in transcranial surgery are neurovascular injury to surrounding structures, which is common to many neurosurgical procedures. Severe complications associated with endonasal transsphenoidal surgery include injury to the major vasculature, most commonly the internal carotid arteries that lie laterally to the pituitary tumour. When performing extended approaches for larger pituitary tumours, other vessels can be damaged including the basilar artery or any of the vessels comprising the circle of Willis; however, this is much less common. In addition, damage can occur to the cranial nerves that course adjacent to the sella, primarily the optic nerves that lie superiorly. Visual loss occurs secondary to a number of mechanisms, including direct injury from dissection of the tumour, haemorrhage, overpacking of closure material, or devascularization of the nerves from the minute vessels originating from the superior hypophyseal artery. In addition to visual loss, ophthalmoplegia or facial numbness can occur when dissecting tumours from the cavernous sinus. Other complications that can occur include infection/meningitis, epistaxis, and chronic sinusitis. One large series of 506 patients undergoing transsphenoidal surgery for non-functioning adenomas revealed an overall complication rate of 9.1%, with the most common complication being CSF leak and severe complications such as visual loss or vascular injury being less than 1–2% [66].

In addition to neurovascular injury, the normal pituitary gland can be damaged when attempting to remove pituitary tumours. Temporary or permanent dysfunction of the anterior and/or posterior pituitary gland may occur resulting in hypopituitarism, such as isolated adrenal insufficiency or DI, or even panhypopituitarism. Rates of hypopituitarism after endonasal transsphenoidal surgery vary, and can be heavily influenced by the pathology present. For instance, invasive or inflammatory pathologies, such as metastatic tumours, lymphocytic hypophysitis, or germinomas, cause inherent damage to the gland, and therefore, removal of these lesions can result in greater levels of hypopituitarism, further influenced by the extent of removal.

One complication more uniquely associated with all types of endonasal surgery is persistent postoperative CSF leak. As tumours from the pituitary gland enlarge, they can stretch the diaphragma sellae, causing this normally thickened, protective layer to become thinned and fragile, or invasive tumour pathologies can erode the membrane or produce holes in it. When the tumour is removed, CSF leak can occur either from damage to the diaphragma sellae from surgical manipulation or removal of the tumour that was previously plugging a defect that can now leave a communication. CSF leak needs to be managed at the time of surgery to decrease the risk of complications such as meningitis, pneumocephalus, or intracranial haemorrhage. There are a number of techniques for sellar closure in the event of a leak to prevent persistent fistulization, such as autologous abdominal fat or fascia lata, biologic glues, synthetic and xenograft dural substitutes, free nasal mucosal flaps, and vascularized nasoseptal flaps [67–69]. Although a number of factors can affect the rate of persistent postoperative CSF leak, the incidence of complications can be as high as 17% for pituitary adenomas and 33% for large lesions that create larger dural defects and high-flow spinal fluid leakage [70–72].

Conclusions

Although there are a multitude of different tumours that can arise from the pituitary gland and the surrounding structures related to the sella, surgery is generally considered the first-line treatment for a vast majority of these pathologies. Growing lesions can result in headaches, visual loss, and hypopituitarism, and hyperfunctioning adenomas can cause symptoms and comorbidities from excess circulating hormones. Although there are a number of ways to surgically approach these lesions, the transsphenoidal approach is the most popular, and use of the endoscope is rapidly becoming the standard method for treating these tumours. Symptomatic relief via this approach is largely successful, with acceptable morbidity and low rates of serious complications from neurovascular injury. The most prevalent complication is persistent CSF leak; however, advancing techniques have greatly diminished its occurrence. Significant and continued surgical experience improves outcomes and decreases complication.

REFERENCES

1. Gittleman H, Ostrom QT, Farah PD, *et al*. Descriptive epidemiology of pituitary tumors in the United States, 2004–2009. *J Neurosurg*, 2014; **121**: 527–35.

2. Ostrom QT, Gittleman H, Farah P, *et al*. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neuro Oncol*, 2013; **15** Suppl 2: ii1–56.
3. Lloyd RV, Osamura RY, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, Vol. 10, 4th edn. Lyon: IARC Publications, 2017.
4. Ezzat S, Asa SL, Couldwell WT, *et al*. The prevalence of pituitary adenomas: a systematic review. *Cancer*, 2004; **101**: 613–19.
5. Dho YS, Jung KW, Ha J, *et al*. An updated nationwide epidemiology of primary brain tumors in Republic of Korea, 2013. *Brain Tumor Res Treat*, 2017; **5**: 16–23.
6. Kim CW, Hwang K, Joo JD, Kim YH, Han JH, Kim CY. Spontaneous involution of Rathke's cleft cysts without visual symptoms. *Brain Tumor Res Treat*, 2016; **4**: 58–62.
7. Simmons JD, Simmons LA. Spontaneous regression of a pituitary cyst. *Neuroradiology*, 1999; **41**: 27–29.
8. Muller HL. Craniopharyngioma. *Endocr Rev*, 2014; **35**: 513–43.
9. Katznelson L, Laws ER, Jr., Melmed S, *et al*. Acromegaly: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 3933–51.
10. Lucas JW, Bodach ME, Tumialan LM, *et al*. Congress of Neurological Surgeons systematic review and evidence-based guideline on primary management of patients with nonfunctioning pituitary adenomas. *Neurosurgery*, 2016; **79**: E533–5.
11. Nieman LK, Biller BM, Findling JW, *et al*. Treatment of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2015; **100**: 2807–31.
12. Melmed S, Casanueva FF, Hoffman AR, *et al*. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
13. Scangas GA, Laws ER, Jr. Pituitary incidentalomas. *Pituitary*, 2014; **17**: 486–91.
14. Caldarelli M, Massimi L, Tamburrini G, Cappa M, Di Rocco C. Long-term results of the surgical treatment of craniopharyngioma: the experience at the Policlinico Gemelli, Catholic University, Rome. *Childs Nerv Syst*, 2005; **21**: 747–57.
15. Karavitaki N, Brufani C, Warner JT, *et al*. Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endocrinol (Oxf)*, 2005; **62**: 397–409.
16. Mortini P, Gagliardi F, Boari N, Losa M. Surgical strategies and modern therapeutic options in the treatment of craniopharyngiomas. *Crit Rev Oncol Hematol*, 2013; **88**: 514–29.
17. Rajan B, Ashley S, Gorman C, *et al*. Craniopharyngioma—a long-term results following limited surgery and radiotherapy. *Radiother Oncol*, 1993; **26**: 1–10.
18. Rizzoli P, Iuliano S, Weizenbaum E, Laws E. Headache in patients with pituitary lesions: a longitudinal cohort study. *Neurosurgery*, 2016; **78**: 316–23.
19. Chen L, White WL, Spetzler RF, Xu B. A prospective study of nonfunctioning pituitary adenomas: presentation, management, and clinical outcome. *J Neurooncol*, 2011; **102**: 129–38.
20. Dallapiazza RF, Grober Y, Starke RM, Laws ER, Jr., Jane JA, Jr. Long-term results of endonasal endoscopic transsphenoidal resection of nonfunctioning pituitary macroadenomas. *Neurosurgery*, 2015; **76**: 42–52; discussion 52–43.
21. Iglesias P, Arcano K, Trivino V, *et al.* Non-functioning pituitary adenoma underwent surgery: a multicenter retrospective study over the last four decades (1977–2015). *Eur J Intern Med*, 2017; **41**: 62–7.
22. Muller HL. Childhood craniopharyngioma. Recent advances in diagnosis, treatment and follow-up. *Horm Res*, 2008; **69**: 193–202.
23. Hoffman HJ, De Silva M, Humphreys RP, Drake JM, Smith ML, Blaser SI. Aggressive surgical management of craniopharyngiomas in children. *J Neurosurg*, 1992; **76**: 47–52.
24. Colao A, Cozzolino A, Pivonello R. Quality of life in patients with Cushing's disease: a modern approach. *Clin Endocrinol (Oxf)*, 2012; **76**: 776–7.
25. Guignat L, Bertherat J. The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline: commentary from a European perspective. *Eur J Endocrinol*, 2010; **163**: 9–13.
26. Casanueva FF, Molitch ME, Schlechte JA, *et al*. Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. *Clin Endocrinol (Oxf)*, 2006; **65**: 265–73.
27. Smith MV, Laws ER, Jr. Magnetic resonance imaging measurements of pituitary stalk compression and deviation in patients with nonprolactin-secreting intrasellar and parasellar tumors: lack of correlation with serum prolactin levels. *Neurosurgery*, 1994; **34**: 834–9; discussion 839.
28. Zaidi HA, Cote DJ, Castlen JP, *et al*. Time course of resolution of hyperprolactinemia after transsphenoidal surgery among patients presenting with pituitary stalk compression. *World Neurosurg*, 2017; **97**: 2–7.
29. Bi WL, Dunn IF, Laws ER, Jr. Pituitary apoplexy. *Endocrine*, 2015; **48**: 69–75.
30. Bills DC, Meyer FB, Laws ER, Jr., *et al*. A retrospective analysis of pituitary apoplexy. *Neurosurgery*, 1993; **33**: 602–8; discussion 608–9.
31. Ebersold MJ, Quast LM, Laws ER, Jr., Scheithauer B, Randall RV. Long-term results in transsphenoidal removal of nonfunctioning pituitary adenomas. *J Neurosurg*, 1986; **64**: 713–19.
32. Semple PL, Jane JA, Jr., Laws ER, Jr. Clinical relevance of precipitating factors in pituitary apoplexy. *Neurosurgery*, 2007; 61: 956–61; discussion 961–52.
33. Du C, Feng CY, Yuan XR, *et al*. Microsurgical management of craniopharyngiomas via a unilateral subfrontal approach: a retrospective study of 177 continuous cases. *World Neurosurg*, 2016; **90**: 454–68.
34. Zhang Y, Xu B, Jiang J, *et al*. [Treatment for giant pituitary adenomas through transcranial approach in a series of 112 consecutive patients]. *Zhonghua Wai Ke Za Zhi*, 2015; **53**: 197–201.
35. Laws ER, Jr., Barkhoudarian G. The transition from microscopic to endoscopic transsphenoidal surgery: the experience at Brigham and Women's Hospital. *World Neurosurg*, 2014; **82**(6 Suppl): S152–54.
36. Eseonu CI, ReFaey K, Rincon-Torroella J, *et al*. Endoscopic versus microscopic transsphenoidal approach for pituitary adenomas: comparison of outcomes during the transition of methods of a single surgeon. *World Neurosurg*, 2017; **97**: 317–25.
37. Karppinen A, Kivipelto L, Vehkavaara S, *et al*. Transition from microscopic to endoscopic transsphenoidal surgery for nonfunctional pituitary adenomas. *World Neurosurg*, 2015; **84**: 48–57.
38. Singh H, Essayed WI, Cohen-Gadol A, Zada G, Schwartz TH. Resection of pituitary tumors: endoscopic versus microscopic. *J Neurooncol*, 2016; **130**: 309–17.
39. Qiao N. Outcome of endoscopic vs microsurgical transsphenoidal resection for Cushing's disease. *Endocr Connect*, 2018; **7**: R26–R37.

40. Olson BR, Gumowski J, Rubino D, Oldfield EH. Pathophysiology of hyponatremia after transsphenoidal pituitary surgery. *J Neurosurg*, 1997; **87**: 499–507.
41. Burke WT, Cote DJ, Iuliano SI, Zaidi HA, Laws ER. A practical method for prevention of readmission for symptomatic hyponatremia following transsphenoidal surgery. *Pituitary*, 2018; **21**: 25–31.
42. Comtois R, Beauregard H, Somma M, Serri O, Aris-Jilwan N, Hardy J. The clinical and endocrine outcome to trans-sphenoidal microsurgery of nonsecreting pituitary adenomas. *Cancer*, 1991; **68**: 860–6.
43. Dekkers OM, de Keizer RJ, Roelfsema F, *et al*. Progressive improvement of impaired visual acuity during the first year after transsphenoidal surgery for non-functioning pituitary macroadenoma. *Pituitary*, 2007; **10**: 61–5.
44. Laws ER, Jr., Trautmann JC, Hollenhorst RW, Jr. Transsphenoidal decompression of the optic nerve and chiasm. Visual results in 62 patients. *J Neurosurg*, 1977; **46**: 717–22.
45. Trautmann JC, Laws ER, Jr. Visual status after transsphenoidal surgery at the Mayo Clinic, 1971–1982. *Am J Ophthalmol*, 1983; **96**: 200–8.
46. Fleseriu M, Yedinak C, Campbell C, Delashaw JB. Significant headache improvement after transsphenoidal surgery in patients with small sellar lesions. *J Neurosurg*, 2009; **110**: 354–8.
47. Berkmann S, Fandino J, Muller B, Kothbauer KF, Henzen C, Landolt H. Pituitary surgery: experience from a large network in Central Switzerland. *Swiss Med Wkly*, 2012; **142**: w13680.
48. Berkmann S, Fandino J, Muller B, Remonda L, Landolt H. Intraoperative MRI and endocrinological outcome of transsphenoidal surgery for non-functioning pituitary adenoma. *Acta Neurochir (Wien)*, 2012; **154**: 639–47.
49. Nomikos P, Ladar C, Fahlbusch R, Buchfelder M. Impact of primary surgery on pituitary function in patients with non-functioning pituitary adenomas—a study on 721 patients. *Acta Neurochir (Wien)*, 2004; **146**: 27–35.
50. Cebula H, Baussart B, Villa C, *et al*. Efficacy of endoscopic endonasal transsphenoidal surgery for Cushing's disease in 230 patients with positive and negative MRI. *Acta Neurochir (Wien)*, 2017; **159**: 1227–36.
51. De Tommasi C, Vance ML, Okonkwo DO, Diallo A, Laws ER, Jr. Surgical management of adrenocorticotropic hormone-secreting macroadenomas: outcome and challenges in patients with Cushing's disease or Nelson's syndrome. *J Neurosurg*, 2005; **103**: 825–30.
52. Keskin FE, Ozkaya HM, Bolayirli M, *et al*. Outcomes of primary transsphenoidal surgery in Cushing disease: experience of a tertiary center. *World Neurosurg*, 2017; **106**: 374–81.
53. Anik I, Cabuk B, Gokbel A, *et al*. Endoscopic transsphenoidal approach for acromegaly with remission rates in 401 patients: 2010 consensus criteria. *World Neurosurg*, 2017; **108**: 278–90.
54. Chen CJ, Ironside N, Pomeraniec IJ, *et al*. Microsurgical versus endoscopic transsphenoidal resection for acromegaly: a systematic review of outcomes and complications. *Acta Neurochir (Wien)*, 2017; **159**: 2193–207.
55. Jane JA, Jr., Starke RM, Elzoghby MA, *et al*. Endoscopic transsphenoidal surgery for acromegaly: remission using modern criteria, complications, and predictors of outcome. *J Clin Endocrinol Metab*, 2011; **96**: 2732–40.
56. Fernandez Mateos C, Garcia-Uria M, Morante TL, Garcia-Uria J. Acromegaly: surgical results in 548 patients. *Pituitary*, 2017; **20**: 522–8.
57. Alexandraki KI, Kaltsas GA, Isidori AM, *et al*. Long-term remission and recurrence rates in Cushing's disease: predictive factors in a single-centre study. *Eur J Endocrinol*, 2013; **168**: 639–48.
58. Dimopoulou C, Schopohl J, Rachinger W, *et al*. Long-term remission and recurrence rates after first and second transsphenoidal surgery for Cushing's disease: care reality in the Munich Metropolitan Region. *Eur J Endocrinol*, 2014; **170**: 283–92.
59. Elshazly K, Kshettry VR, Farrell CJ, Nyquist G, Rosen M, Evans JJ. Clinical outcomes after endoscopic endonasal resection of giant pituitary adenomas. *World Neurosurg*. 2018; **114**: e447–56.
60. Shirvani M, Motiei-Langroudi R, Sadeghian H. Outcome of microscopic transsphenoidal surgery in Cushing disease: a case series of 96 patients. *World Neurosurg*, 2016; **87**: 170–5.
61. Yamada S, Fukuhara N, Yamaguchi-Okada M, *et al*. Therapeutic outcomes of transsphenoidal surgery in pediatric patients with craniopharyngiomas: a single-center study. *J Neurosurg Pediatr*, 2018: **21**: 549–62.
62. Patel VS, Thamboo A, Quon J, *et al*. Outcomes after endoscopic endonasal resection of craniopharyngiomas in the pediatric population. *World Neurosurg*, 2017; **108**: 6–14.
63. Elliott RE, Wisoff JH. Surgical management of giant pediatric craniopharyngiomas. *J Neurosurg Pediatr*, 2010; **6**: 403–16.
64. Andereggen L, Hess B, Andres R, *et al*. A ten-year follow-up study of treatment outcome of craniopharyngiomas. *Swiss Med Wkly*, 2018; **148**: w14521.
65. Becker G, Kortmann RD, Skalej M, Bamberg M. The role of radiotherapy in the treatment of craniopharyngioma—indications, results, side effects. *Front Radiat Ther Oncol*, 1999; **33**: 100–13.
66. Halvorsen H, Ramm-Pettersen J, Josefsen R, *et al*. Surgical complications after transsphenoidal microscopic and endoscopic surgery for pituitary adenoma: a consecutive series of 506 procedures. *Acta Neurochir (Wien)*, 2014; **156**: 441–9.
67. Ziu M, Jimenez DF. The history of autologous fat graft use for prevention of cerebrospinal fluid rhinorrhea after transsphenoidal approaches. *World Neurosurg*, 2013; **80**: 554–62.
68. Hadad G, Bassagasteguy L, Carrau RL, *et al*. A novel reconstructive technique after endoscopic expanded endonasal approaches: vascular pedicle nasoseptal flap. *Laryngoscope*, 2006; **116**: 1882–6.
69. Fishpool SJ, Amato-Watkins A, Hayhurst C. Free middle turbinate mucosal graft reconstruction after primary endoscopic endonasal pituitary surgery. *Eur Arch Otorhinolaryngol*, 2017; **274**: 837–44.
70. Strickland BA, Lucas J, Harris B, *et al*. Identification and repair of intraoperative cerebrospinal fluid leaks in endonasal transsphenoidal pituitary surgery: surgical experience in a series of 1002 patients. *J Neurosurg*, 2018: **129**: 425–9.
71. Harvey RJ, Parmar P, Sacks R, Zanation AM. Endoscopic skull base reconstruction of large dural defects: a systematic review of published evidence. *Laryngoscope*, 2012; **122**: 452–9.
72. Kassam AB, Prevedello DM, Carrau RL, *et al*. Endoscopic endonasal skull base surgery: analysis of complications in the authors' initial 800 patients. *J Neurosurg*, 2011; **114**: 1544–68.

2.3.8 Pituitary Radiotherapy

Naomi Fersht and Francesca Soldà

Introduction

Pituitary adenomas are the most common pituitary disorder in the adult population [1].

Their progressive growth within and outside the sellar region is responsible for the onset of characteristic mass-related symptoms (cranial nerve dysfunction and optic pathway compression), while a hypersecretory endocrine state is the common modality of disease presentation in functioning lesions.

The optimal management of pituitary adenomas involves a multidisciplinary collaboration between neurosurgeons, endocrine specialists and clinical (radiation) oncologists, and often requires a combination of treatments.

The main goal of all treatments is to provide long-term tumour control and/or to normalize the hormonal hypersecretion. While surgery has been historically considered the first approach in all symptomatic pituitary tumours, radiotherapy maintains an important and long-established role in the management of secreting and non-functioning adenomas, with a considerable number of patients receiving external beam irradiation in course of their disease.

In patients with non-functioning tumours, radiotherapy is withheld until evidence of further tumour progression after surgery, unless significant concerns are raised by new symptoms (visual impairment), or by specific histological features (atypical features, silent corticotroph adenoma). Often a second surgery is considered before radiotherapy. In progressive, non-functioning adenomas radiotherapy achieves tumour control in over 90% of patients at 10 years, and in 85–92% at 20 years [2–7].

For functioning tumours, radiotherapy is indicated when rapidly progressive or recurrent lesions are detected or in those patients failing to achieve hormonal normalization after other treatments (previous surgeries and medical therapy). Hormone levels decline slowly, normalizing in the great majority of patients over months or years after radiotherapy, depending on pretreatment hormone values [8, 9].

Modern radiotherapy techniques allow the delivery of an effective dose of radiation to the treatment target while minimizing the dose delivered to surrounding normal tissues thus potentially decreasing the long-term risk of normal tissue damage.

Radiotherapy for Pituitary Adenomas: Treatment Plan and Fractionation

Regardless of the technology used to deliver it, pituitary radiotherapy aims to:

- Accurately identify the treatment target within the sella and the adjacent structures;
- Calculate the dose to deliver to the treatment target by using a computerized treatment planning system;
- Accurately deliver the predefined dose of radiation in one or multiple fractions (radiosurgery or fractionated radiotherapy);
- Minimize the dose received by the surrounding uninvolved brain hence reducing the risk of normal tissue injury.

In general, the treatment target or GTV (gross tumour volume) is identified on high resolution imaging (thin cuts MRI sequences without contrast) accurately fused with computed tomography (CT) imaging acquired for planning purposes. During the planning scan the patient head is immobilized in a closely fitting relocatable frame (generally used in fractionated radiotherapy) or with the use of more invasively fixed frames (mainly employed in single fractionated treatments).

The GTV is manually contoured in three planes to generate a 3D volume encompassing the macroscopically visible lesion. An isotropic margin is added to account for areas of uncertainty in volume delineation, the trans-sphenoidal surgical route, and any set-up variation relative to the single radiotherapy centre. The generated volume is defined as the planning target volume (PTV). A careful delineation of all the organs surrounding the PTV and at potential risk of irradiation (OARs, organs at risk) is mandatory. The entire optic pathway, the brainstem, and the hypothalamic structures are routinely outlined. All OARs are expanded to create a planning risk volume (PRV) which margin, as for the PTV, should reflect the accuracy of daily local set-up (**Figure 2.3.8.1**).

The treatment dose is calculated with a sophisticated computerized planning system that defines the best arrangement of the radiotherapy beams to deliver the dose homogenously to the PTV while minimizing the dose received by non-target structures.

In conventional radiotherapy, this is achieved by dividing the total dose over multiple sessions (fractionated radiotherapy). Fractionated pituitary irradiation is given to total dose of 45 to 50.4 Gy at 1.8 Gy per fraction, once a day, five days per week. Fractionation is used for lesions larger than the ones treated with single fraction radiotherapy—often referred as radiosurgery—or for tumours abutting or compressing the optic apparatus.

Radiotherapy Techniques

3D Conformal Fractionated Radiotherapy

Conventional radiotherapy for pituitary adenoma is delivered with photons using a linear accelerator (LINAC), the most common machine used to deliver external beam radiotherapy.

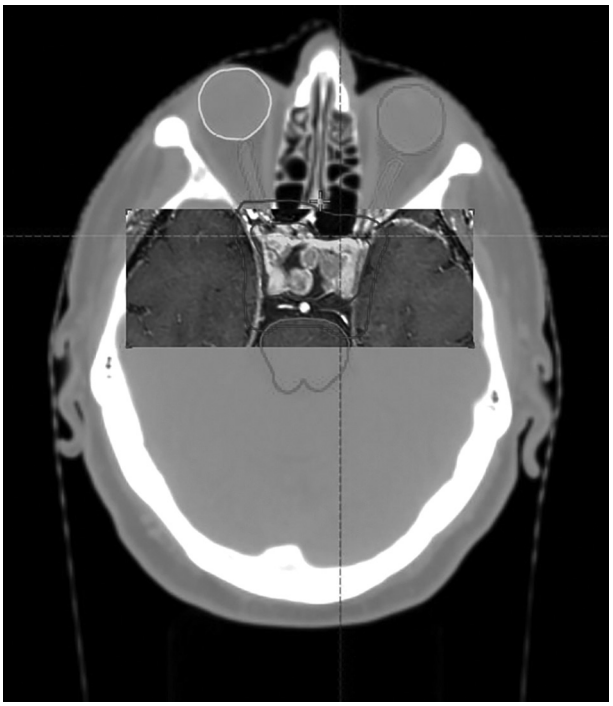


Figure 2.3.8.1 Target delineation of a pituitary adenoma. Gross tumour volume/planning target volume (GTV/PTV) and organs at risk are outlined. GTV (brown)/ PTV (red); left optic nerve (pink); right optic nerve (light orange); brainstem (orange). *For a colour version of this figure, please see colour plate section.*

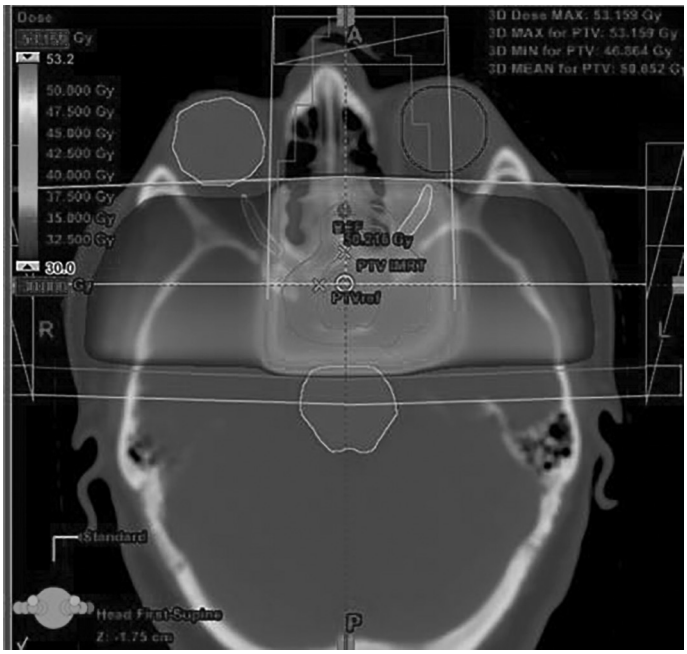


Figure 2.3.8.2 Example of beam arrangement and dose distribution in a conventional three fields CRT plan (an anterior oblique and two lateral opposed beams).

Three-dimensional (3D) conformal fractionated radiotherapy (CRT) is used to treat functioning and non-functioning adenomas of small or large volumes.

To achieve reproducible head position and aid treatment precision, a custom-made closely fitting thermoplastic mask is applied and moulded directly to the patient's face in the treatment planning process. The repositioning accuracy of the mask during treatment is around 3–5 mm [10], and can be improved to 2–3 mm, by using tighter fitting masks [11].

Imaging for treatment planning is performed with the patient lying in supine position with the head in the immobilization device. Planning CT and pre- and postoperative MRI scans are coregistered on the planning system for 3D target delineation (GTV) with an isotropic margin of 5–10 mm added to account for uncertainties in planning or treatment delivery.

In 3D-CRT, three fixed fields of radiation are commonly used to irradiate the PTV: an anterior oblique field directed to the pituitary area through the forehead and two lateral opposed beams coming from the temporal regions (Figure 2.3.8.2). If a different beam arrangement is adopted, the planning system defines the optimal number and orientation of each field to achieve a conformal dose distribution within PTV, limiting the dose to the adjacent normal OARs. Each beam of radiation is shaped to conform to the profile of the PTV with a multileaf collimator (MLC): an array of computer-controlled parallel leaves moving in and out to create an adjustable aperture through which the beam is delivered. The use of a dynamic MLC (with the leaves moving during the gantry rotation) also allows to alter the intensity of the radiation during its path (intensity-modulated radiotherapy, IMRT) resulting in a more homogeneous dose distribution within targets of any shape. A technique of arcing IMRT (described as volumetric modulated arc therapy (VMAT) or RapidArc) offers a fast way of delivering complex IMRT treatments and is increasingly in use as an alternative to conventional fixed field 3D-CRT. Arc techniques are used to minimize the dose bilaterally to the temporal lobes with the aim of reducing the impact of treatment on patients' cognitive function.

Stereotactic Radiotherapy

The stereotactic techniques represent an evolution of 3D-CRT where the combination of more robust patient immobilization and on-treatment imaging (image guidance) guarantee accurate treatment delivery.

Originally, the term 'stereotactic' referred to the use of a system of external 3D coordinates, generally a neurosurgically fixed head frame, adopted to determine the location of a lesion within the brain. While in some contexts the stereotactic localization has been superseded by the integration of modern imaging within the radiotherapy planning system, the term 'stereotactic' is still widely in use.

With more precise target localization and accurate dose delivery compared with 3D-CRT, stereotactic radiotherapy (SRT) results in a reduction of the volume of normal brain tissue irradiated to high doses of radiation.

SRT can be delivered in daily fractions over few days (3–5 fractions, **hypofractionated radiotherapy—hSRT**), or over few weeks (up to 25–28 fractions in pituitary adenoma, conventionally **fractionated stereotactic radiotherapy—fSRT**), or in a single fraction (**radiosurgery—RS**).

In deciding between these methods, several considerations must be taken into account. Large adenomas (3 cm or more in maximum diameter) or adenomas close to radiation-sensitive structures (such as the optic chiasm) are more safely treated with fractionated radiation (3D-CRT, fSRT, or less commonly hSRT). At the standard total doses of 45–50.4 Gy delivered in fractions of less than 2 Gy, the

risk of radio-induced damage to the optic pathway is around 1%. As the risk of optic neuropathy is dose-dependent, a single large dose of radiation delivered with RS results in an unacceptable toxicity if the adenoma abuts or displaces the surrounding critical structures (optic apparatus) [12, 13].

Fractionated Stereotactic Radiotherapy (fSRT)

fSRT is delivered on LINACs with patients immobilized in relocatable frames with a reported relocation accuracy of 1–2 mm [14, 15], or in precisely fitting thermoplastic mask systems with an accuracy of 2–3 mm [11].

A smaller PTV margin is used in fSRT (generally 3–5 mm) as a result of the more precise patient immobilization devices adopted and the concomitant use of frequent imaging during the course of radiation (image guidance). On-treatment imaging (orthogonal X-rays or cone beam CT scans) enables instant, real-time correction for patient positioning deviations with respect to the planning scans, thus improving the precision of the daily treatment.

fSRT is delivered using multiple static beams of radiation (usually 4–6), each beam conforming to the shape of the PTV using an MLC (mini MLCs with a 5 mm width or micro MLCs with 3 mm width). As for 3D-CRT, the use of MLC allows to modulate the beam intensity during its delivery (IMRT). VMAT or RapidArc is also possible on dedicated LINACs (Figure 2.3.8.3).

Adapted LINACs mounted on a robotic arm capable of movements in six degrees of freedom (Cyberknife) can be used to treat pituitary adenomas either in a single fraction or by hypofractionating the total dose of radiation in three to five fractions.

Single Fraction Stereotactic Radiotherapy (Radiosurgery–RS)

Single fraction radiotherapy, usually referred as radiosurgery (RS), is a high precision localized technique of irradiation used as a non-invasive alternative to surgical excision (hence the misleading term of 'radiosurgery') in patient with both malignant and benign brain conditions.

RS for pituitary adenomas was first delivered on the Gamma Knife unit (GK), a multiheaded cobalt device in which multiple cobalt-60 sources (192 in the current version) are arranged in a hemispherical distribution. The generated multiple beams of radiation converge with high precision at a single point (isocentre) to achieve a circumscribed dose distribution within the PTV. For more complex shaped adenomas, a multiple isocentre arrangement is often chosen resulting in a combination of several spherical dose volumes encompassing the treatment target.

In the planning phase of the procedure, an MRI-compatible frame is neurosurgically attached to the skull of the patient in four points, under local anaesthesia. A system of fiducial markers coupled to the frame during the planning imaging allows a frame-based coordinate system, where any point within the brain is located with submillimetric precision. The number and the position of the isocentres are calculated by the planning system to shape the prescription dose to the lesion profile (marginal dose). Due to the invasive nature of the frame, the treatment planning and delivery procedures necessitate been carrying out and completing within a single day.

RS for pituitary adenomas is given to doses of 12–35 Gy to the tumour margin with maximum doses to the optic pathway and the cranial nerves in the cavernous sinus limited to 8–10 Gy and 16–18 Gy, respectively (Figure 2.3.8.4).

LINACs initially modified with external mechanical supports to deliver RS have been replaced by LINACs with a high level of precision not requiring adaptation for RS. Non-invasive relocatable frames coupled to image-guided techniques can be used in alternative to fixed frames. In this so called 'frameless RS', the high single dose of radiation is delivered with single or multiple arcs of radiation (VMAT) or multiple fixed fields. With Cyberknife (CK) LINACs, the freedom in movements of the robotic arm allows the delivery of a wide range of single non-coplanar beams intersecting one other to create a uniform dose distribution in the target.

Proton Beam Radiotherapy (PBT)

Proton devices are increasingly used to treat brain tumours and the renewed interest in proton beam radiotherapy (PBT) is associated to the physic properties of these heavy particles. Unlike photons, protons are able to deliver most of their energy when they hit the target. The very low entry dose and minimal exit dose beyond the target boundaries allow to deliver the prescribed treatment also to

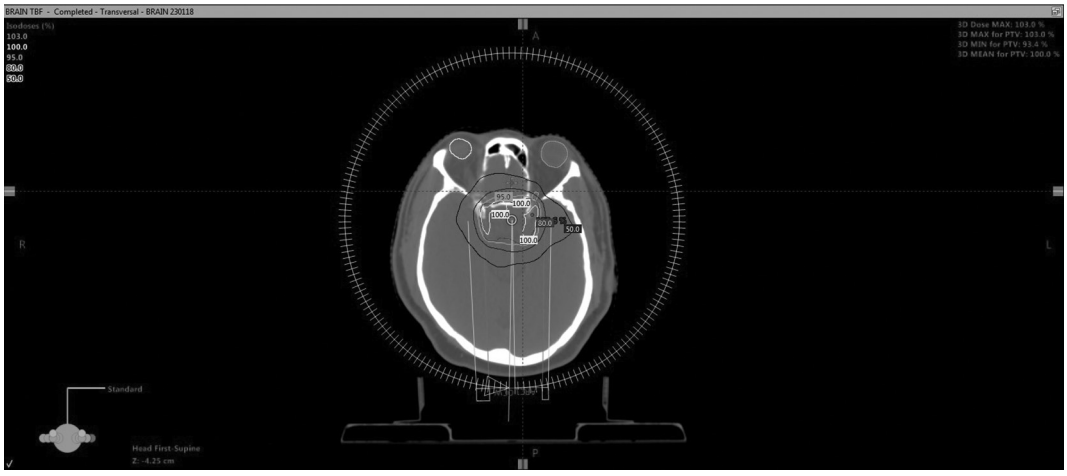


Figure 2.3.8.3 Example of isodose distribution in a VMAT plan. (in red the PTV) *For a colour version of this figure, please see colour plate section.*

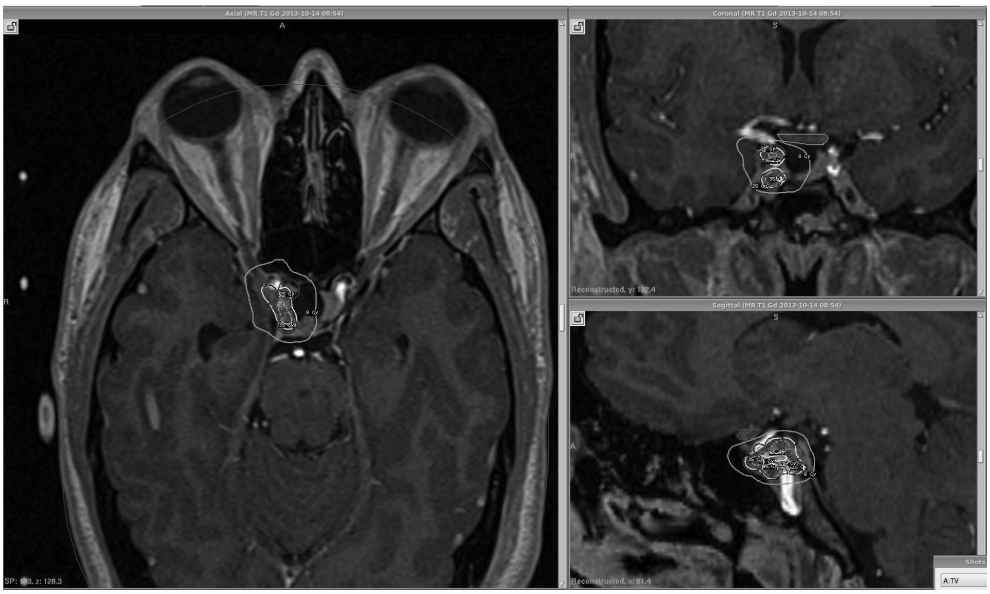


Figure 2.3.8.4 Example of Gamma Knife plan for a residual secreting pituitary adenoma within the right cavernous sinus (marginal dose 25 Gy—yellow). White line shows the 8 Gy isodose distribution with respect to the optic pathway. *For a colour version of this figure, please see colour plate section.*

large, irregular volumes with optimal sparing of the surrounding healthy tissues [16].

Protons are produced by large particle accelerators (cyclotrons or synchrotrons) where the atom is stripped of its electron to produce heavy positive charged particles. The steps involved in PBT are similar to photon-based treatments with relocatable frames commonly used for patient immobilization.

Current indications for the use of protons within the UK Specialised Commissioning Team include the treatment of pituitary adenomas up to the age of 24 years old based on theoretical reduction in the possible late side effects of brain radiation, such as second malignancy, neurocognitive deficits and cerebrovascular disease (Figure 2.3.8.5) [17].

Toxicity of Radiotherapy

Hypopituitarism is the most common consequence of radiation treatments for pituitary adenomas, after both fractionated and single dose radiation therapy. It is more frequent in patients with previous insults to the pituitary gland (post-surgery, from the tumour itself) and its clinical manifestations can gradually develop over several years from treatment (the increased risk of hormonal deficiency is as high as 80% by 10–15 years).

Permanent cranial nerve injuries secondary to radiotherapy are far less common than hormonal failure in pituitary adenomas treated with fractionated irradiation, where the risk of neurotoxicity increases markedly only at doses >60 Gy at ≥1.8 Gy/fraction. In single fraction RS, the risk of radiation-induced optic neuropathy is estimated at 1.7%, 1.8%, 0%, and 6.9% for a Dmax of <8, 8–10, 10–12, and >12 Gy, to the optic apparatus, respectively, so the frequency of complications should be low if an appropriate patient selection for RS is adopted [12].

Patients with pituitary adenoma treated with radiotherapy also seem to have a significantly increased risk of cerebrovascular complications (CVAs) compared to the general population [18], although these data are not confirmed in all studies [19, 20].

Radiation is recognized to be a contributing factor in the development of second brain tumours and an increased risk of developing a radiation-induced malignancy in pituitary adenomas treated with fractionated radiotherapy has been reported [21, 22]. The relative risk of radiation-induced tumours after RS has not been precisely determined and the relatively short follow-up of most published series does not allow for meaningful comparison to fractionated radiotherapy.

Clinical Outcome of Pituitary Radiotherapy

Despite several differences in treatments-related parameters between LINAC, GK, and CK radiotherapy in pituitary adenomas, the published data on the clinical efficacy of the various techniques is based on the results of retrospective case studies reporting on a single technique and non-randomized comparative studies.

The major limitation of these types of studies is the difficulty in understanding the real efficacy of an intervention due to the lack of control over factors that influence the outcome being measured, making the presumed superiority of one technique against other still a matter of debate.

The clinical efficacy of radiotherapy for pituitary adenomas should be measured in terms of overall survival, actuarial tumour control (progression-free survival, PFS), and quality of life. The most frequently reported endpoints of radiation treatment for non-functioning pituitary adenomas are local tumour control and long-term morbidity.

Hormone normalization following treatment should be described in relationship to pretreatment hormone levels and defined as the time to reach 50% of pretreatment values.

Furthermore, although the total dose delivered with fractionated meanings of irradiation is largely consistent within different

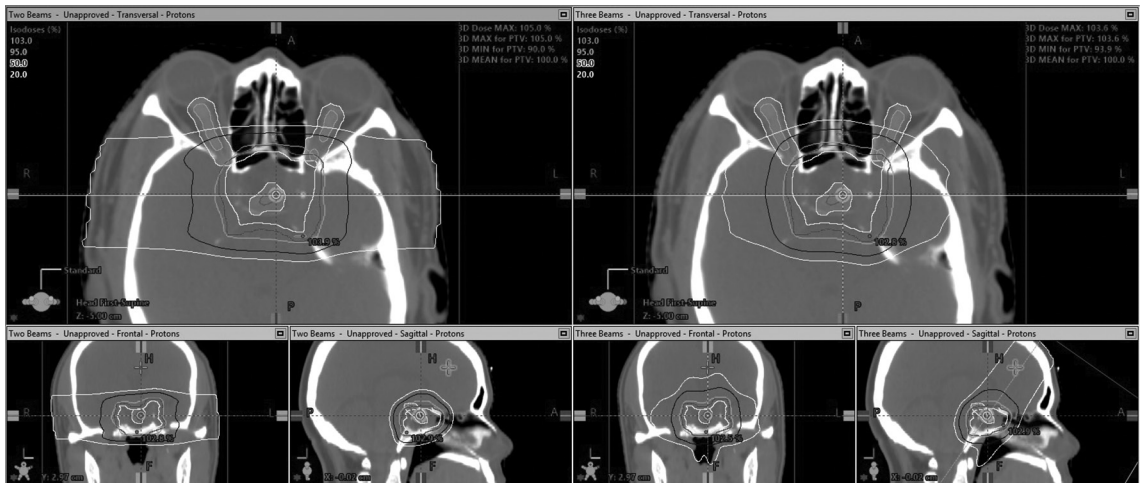


Figure 2.3.8.5 Example of two-beam (lateral opposed) and three-beam (superior-oblique and two laterally opposed) arrangement in a recurrent pituitary adenoma treated with fractionated PBT (50.4 Gy/28#).

publications (45–50.4 Gy), the range of dose prescriptions between secretory and non-functioning adenomas treated with RS tends to be different (from 12 to 30 Gy). The rationale behind this practice is based on the observation that a more rapid hormone normalization was reported in single studies using higher doses to treat secreting tumours. In absence of a strong radiobiological model and of prospective randomized studies in support, the relationship between dose delivered and endocrine remission warrants further investigation.

Outcome of 3D-CRT

The long-term outcomes from case series published in the literature for 3D-CRT in pituitary adenomas are shown in Table 2.3.8.1. The actuarial PFS varies from 87% to 98% at 10 years [3, 23–25] with the majority of the studies providing consistent long-term follow-up data and the largest study also reporting figures on PFS at 20 years [2]. In secreting tumours, normalization of the pretreatment values is achieved at variable times from irradiation. Stabilization of growth hormone (GH) hypersecretion mainly depends on pre-RT levels, with GH/IGF-1 normalization in up to 70% of patients 10–15 years after RT. A reduction to 50% of pretreatment levels is detected after 2 years from treatment [26, 27]. Biochemical remission with normalization of cortisol secretion has been reported in 73%, 78%, and 84% of patients at 3, 5, and 10 years, respectively. Urinary free cortisol (UFC) is reduced to 50% of the pretreatment levels after an interval of 6–12 months, and plasma cortisol after around 12 months [28]. Large recurrent prolactinomas are reported to be less sensitive to radiotherapy but data are scarce because of its limited use. The reported 10-year tumour and hormone control rates are 90% and 50%, respectively [29–31].

Data about toxicity of 3D-CRT show low rates of complications after treatment. The most common long-term complication is hypopituitarism, occurring in 30–80% of patients at 10 years from radiotherapy. The incidence of symptomatic radio-induced optic neuropathy following 3D-CRT is 1–3%, but long-standing compression of the visual pathway in large tumours might contribute to these figures.

Cerebrovascular mortality has been found increased in patients with pituitary adenomas treated with radiotherapy compared with the general population [18, 32]. The metabolic and cardiovascular effects of hormone imbalance and the consequences of surgery might play a significant role in the long-term occurrence of these conditions.

There is a slight but significant 2–2.7% increased risk of secondary brain tumours after conventional radiotherapy for pituitary adenoma reported in various studies [21, 33, 34].

Outcome of Fractionated Stereotactic Radiotherapy (fSRT)

The outcome of 1166 functioning and non-functioning pituitary adenomas treated with fSRT is reported in Table 2.3.8.2.

The length of follow-up after fSRT is shorter than reported for cohorts of patients treated with 3D-CRT.

At a corrected median follow-up of 58 months (range 9–152 months), tumour control was achieved in 96% of patients with the largest study reporting a PFS of 99% at 5 years [35]. Few studies reported the same data at 10 years with similar results as per 3D-CRT.

Only limited data are available on the efficacy of fSRT in secreting tumours. Normalization of GH levels was reported in 21 (84%) and 6 (33%) acromegalic patients after a median follow-up of 26 months and 39 months, respectively [24, 36]. Data on biochemical remission in ACTH-secreting lesions are limited but it is unlikely to be different from those obtained with CRT using the same dose fractionation.

The insufficient number of patients with prolactinoma who fail previous medical or surgical treatment does not allow definitive conclusion on the efficacy of fSRT in achieving hormone normalization.

The incidence of complications after fSRT is in line with the same figures reported for CRT although a 20% mean rate of hypopituitarism might still reflect the short duration of follow-up for this group of patients.

Outcome of Single Fraction Stereotactic Radiotherapy (Radiosurgery–RS)

The results of GK SRS in non-functioning pituitary adenomas have been summarized in Table 2.3.8.3.

Table 2.3.8.1 Summary of results of published series on 3D-CRT for pituitary adenomas

Authors	Type of adenoma	Number of patients	Follow-up (median years)	Actuarial progression-free survival (PFS) (%)	Late toxicity (%) | Visual hypopituitarism	
Grigsby at al.,1989	NFA, SA	121	11.7	89.9 at 10 years	1.7	NA
McCollough *et al.*, 1991	NFA, SA	105	7.8	95 at 10 years	NA	NA
Brada *et al.*, 1993	NFA, SA	411	10.8	94 at 10 years 88 at 20 years	1.5	30 at 10 years
Tsang *et al.*, 1994	NFA, SA	160	8.7	87 at 10 years	0	23**
Zierhut *et al.*, 1995	NFA, SA	138	6.5	95 at 5 years	1.5	27**
Estrada *et al.*, 1997	SA (ACTH)	30	3.5	73 at 2 years*	0	48**
Rush *et al.*, 1997	NFA, SA	70	8	NA	NA	42**
Breen *et al.*, 1998	NFA	120	9	87.5 at 10 years	1	NA
Gittoes *et al.*, 1998	NFA	126	7.5	93 at 10 and 15 years	NA	NA
Barrande *et al.*, 2000	SA (GH)	128	11	53 at 10 years*	0	50 at 10 years
Biermasz *et al.*, 2000	SA (GH)	36	10	60 at 10 years*	0	54 at 10 years
Sasaki *et al.*, 2000	NFA, SA	91	8.2	93 at 10 years	1	NA
Epaminonda *et al.*, 2001	SA (GH)	67	10	65 at 15 years*	0	NA
Minniti *et al.*, 2005	SA (GH)	45	12	52 at 10 years*	0	45 at 10 years
Langsenlehner *et al.*, 2007	NFA, SA	87	15	93 at 15 years	0	88 at 10 years
Jenkins *et al.*, 2006 [63]	SA (GH)	656	7	NA	0	18% LH/FSH, 15% ACTH, 27% TSH at 10 years
Minniti *et al.*, 2007	SA (ACTH)	40	9	78 and 84 at 5 and 10 years*	0	62 at 10 years
Rim *et al.*, 2011	NFA, SA	60	5.6	96 at 10 years (NFA), 66 at 10 years (SA)	0	76 at 10 years
Kim *et al.*, 2016	NFA, SA	73	8	98 at 10 years	0	NA
Patt *et al.*, 2016	SA (GH)	36	4.9 (mean)	89 at 5 years	0	33

NFA, non-functioning adenoma; SA, secreting adenoma; NA, not assessed, ACTH-Cushing, GH-acromegaly, * hormone concentration normalization, ** no time specified.
Updated with permission from Soldà F, Fersht N. Radiotherapy for Pituitary Tumours. [Updated 2017 Dec 5]. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK278955/ Copyright © 2017, MDText.com, Inc.

The actuarial 5-year control rate (PFS) following GK is 95%, with limited data at 10 years. The reported treatment doses vary from 12 to 20 Gy (mean dose) [37, 38]. Although outcome results are in line with the corresponding figures for 3D-CRT and fSRT, smaller tumours are generally considered for treatment with GK.

Hormonal response following GK RS for GH-secreting tumours is shown in Table 2.3.8.4.

Data published up to 2016 on 34 available studies showed that 41% of patients achieved normalization of serum GH, after a median follow-up of 46 months. A large study on 136 patients reported higher actuarial remission rates (normal age and gender-matched serum IGF-1 level and, in some patients, a GH level less than 1 ng/ml after oral glucose tolerance test (OGTT)) at 2, 4, 6, and 8 years after RS of 31.7, 64.5, 73.4, and 82.6%, respectively. The strongest prognostic factor for endocrine remission was a lower pre-RS IGF-1 [39].

The reported rate of hypopituitarism at 5 years varies from 8% to 33% in the studies with longer follow-up data.

The outcome of GK RS in patients with Cushing's disease is reported in Table 2.3.8.5.

Biochemical remission (defined as normalization of plasma cortisol and 24-hour UFC level) was achieved in 52% of patients at a corrected median follow-up of 50 months after RS.

In a recent international multi-institutional study on 278 patients (mean follow-up 5.3 years), the reported rate of new hypopituitarism was 25% [40]. The same study also reported a recurrence rate up to 20% after initial disease remission, suggesting the importance of a continuous follow-up even in patients with early hormonal normalization.

Small prolactinomas in patients resistant or intolerant to dopamine agonist can be considered for RS.

In 24 studies reporting on 692 patients, 33% of them had normalization of serum prolactin concentrations at a corrected median follow-up of 43 months (median range 6–60 months), following RS (Table 2.3.8.6). There are insufficient data to assess the rate of decline of prolactin following GK RS in comparison to data following 3D-CRT or fSRT.

The most commonly reported complication following GK RS is hypopituitarism, with a crude incidence ranging from 0% to 66%, thus largely comparable to fSRT and 3D-CRT. The incidence of cranial nerve deficits after RS is generally low (in the range of 0–6%), although higher rates of complications are occasionally reported in selected studies suggesting a potential inappropriate selection of patients for treatment (large tumours/tumours close to the optic apparatus) [41, 42].

Table 2.3.8.2 Summary of results on published studies on fSRT for pituitary adenomas

Authors	Number of patients	Follow-up median (months)	Tumour growth control rate (%)	Late toxicity (%) | Visual hypopituitarism	
Coke *et al.*, 1997	19*	9	100	0	0
Mitsumori *et al.*, 1998	30*	33	86 at 3 years	0	20
Milker-Zabel *et al.*, 2001	68*	38	93 at 5 years	7	5
Paek *et al.*, 2005	68	30	98 at 5 years	3	6
Colin *et al.*, 2005	110*	48	99 at 5 years	2	29 at 4 years
Minniti *et al.*, 2006	92*	32	98 at 5 years	1	22
Selch *et al.*, 2006	39*	60	100	0	15
Kong *et al.*, 2007	64*	37	97 at 4 years	0	11
Snead *et al.*, 2008	100*	6.7 years	98 and 73 at 10 years for NFA and SA	1	35
Roug *et al.*, 2010	34*	34	91 (50% hormonal normalization)	–	–
Schalin-Jantti *et al.*, 2010	30	5.3 years	100	0	23
Weber *et al.*, 2011	27*	72.4	96	4	8
Wilson *et al.*, 2012	67	5.12 years	88	2	6
Kim *et al.*, 2013	76*	6.8 years	97.1 at 7 years	0	48 (one or more hormone)
Kopp *et al.*, 2013	37	57	91.9	5	43
Liao *et al.*, 2014	34	36.8 (mean)	100	0	NA
Minniti *et al.*, 2015	68	75	97 and 91 at 5 and 10 years	0	26
Puataweepong *et al.*, 2015	94*	72	95	3	9.6
Diallo *et al.*, 2015	34*	152 (mean)	97	0	39
Barber *et al.*, 2016	75*	47.5 (mean)	100	1.5	28

* Case series includes secreting adenomas.

The risk of RS-associated secondary brain tumours has not been precisely determined and longer follow-up data are required.

Outcome of LINAC-Based Single Fraction Radiotherapy (Radiosurgery–RS)

A limited number of studies reported the results of LINAC-based RS for both functioning and non-functioning pituitary adenomas (Table 2.3.8.7), with results in terms of tumour control and hormone normalization broadly in the range of those achieved with GK RS.

Similar considerations are valid for series on robotic RS (CK) (Table 2.3.8.8). The main limitation of both groups of data is the duration of the follow-up provided that does not allow drawing definitive conclusions on the superiority of these techniques over fractionated approaches.

Outcome of Proton Therapy

The use of PBT, both fractionated and single fraction, has been described to treat functioning and non-functioning lesions [53–57]. A recent study on 47 patients treated with fractionated PBT reported a high rate of post-treatment complications including a case of temporal lobe necrosis, 3 patients with new significant visual deficits, and 11 patients with new hypopituitarism. Early data on tumour control rate are not dissimilar from those reported with 3DCRT and stereotactic treatments (87% at 6 months) [56].

The largest study on stereotactic PBT reported on the biochemical response of a population of secreting adenomas (74 ACHT, 50 GH, 9 PRL, 8 Nelson, 3 TSH). At a median follow-up of 52 months, 42% of patients did not achieve endocrine control with acromegalic patients having the longer time to biochemical response (49% at 5 years). The risk of developing hypopituitarism was 62% at 5 years. Four patients (3%) experienced post-treatment temporal lobe seizures, with associated temporal lobe changes on imaging (1 month to 9 years from proton treatment) [57].

Pituitary Re-irradiation

Re-irradiation of recurrent/progressive adenomas unsuitable for surgical resection should take into careful consideration the position of the adenoma with respect to the optic pathway or other

Table 2.3.8.3 Summary of results of published series on RS for non-functioning pituitary adenomas

Authors	Number of patients	Follow-up median (months)	Tumour control growth rate (%)	Late toxicity (%) | Visual hypopituitarism	
Martinez *et al.*, 1998	14	26–45	100	0	0
Pan *et al.*, 1998	17	29	95	0	0
Ikeda *et al.*, 1998	13	45	100	0	0
Mokry *et al.*, 1999	31	20	98	NA	NA
Sheehan *et al.*, 2002	42	31*	97	2.3	0
Wowra *et al.*, 2002	45	55	93 at 3 years	0	14
Petrovich *et al.*, 2003	56	36	94 at 3 years	4	NA
Pollock *et al.*, 2003	33	43	97 at 5 years	0	28 and 41 at 2 and 5 years
Losa *et al.*, 2004	56	41*	88 at 5 years	0	24
Iwai *et al.*, 2005	34	60	93 at 5 years	0	6
Mingione *et al.*, 2006	100	45*	92	0	25
Liscak *et al.*, 2007	140	60	100	0	2
Pollock *et al.*, 2008	62	63	95 at 3 and 7 years	0	32 at 5 years
Kobayashi *et al.*, 2009	60	>3 years	97	4.3	8.2 worsening
Gopalan *et al.*, 2011	48	80.5	83	9.4	39
Park *et al.*, 2011	125	62	94 at 5 years and 76 at 10 years	1	24 at 2 years
Wilson *et al.*, 2012	51	4.17 years	100	0	0
Runge *et al.*, 2012	61	83	98	0	9.8
Starke *et al.*, 2012	140	4.2 years	97 at 5 and 87 at 10 years	12.8	30.3
El-Shehaby *et al.*, 2012	38	44*	97	0	0
Sheehan *et al.*, 2013+	512	36	95 at 5 years	6.6	21
Lee *et al.*, 2014	41	48	94 at 5 and 85 at 10 years	2.4	24.4
Xu *et al.*, 2014	34	56	73 at 3 years	24	29
Hasegawa *et al.*, 2015	16	98	100	0	6

* Mean follow-up; + multicentre study, 34 patients had prior fSRT. NA, not available.

Updated with permission from Soldà F, Fersht N. Radiotherapy for Pituitary Tumours. [Updated 2017 Dec 5]. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK278955/ Copyright © 2017, MDText.com, Inc.

cranial nerves, the previous dose of treatment received and the time from primary irradiation.

The effects of a previous irradiation might take several years to develop (especially in secreting tumours) and a further dose of radiation might be associated with an unacceptable increase in treatment-related complications. Few retrospective studies on small groups of selected patients treated with highly conformal fractionated radiotherapy or RS have been reported [58–60]. A recent study evaluating the outcome and toxicity of re-irradiation in 15 secreting and non-secreting adenomas reported an early occurrence of radiation-induced neuropathy (13% at 14 months) and a high rate of temporal lobe necrosis (13%) [61]. Conversely, no visual sequelae or further loss of anterior pituitary function at a median follow-up of 33 months were described in another study reporting on 21 patients retreated with LINAC-based RS [62].

Based on the paucity of available data, a judicious selection of patients for pituitary re-irradiation must be considered. The use of 'stereotactic' techniques and fractionation could theoretically reduce the incidence of late sequelae.

Conclusions

Both fractionated and single fraction treatments achieve excellent rates of tumour control with figures as high as 90% at 5 and 10 years also reducing hormone hypersecretion in approximately 50% of patients at 10 years. Normalization of excessive secretion is generally variable and might require a long interval after treatment.

Careful selection of patients for RS must be considered in view of the potential risks of radiation-induced toxicity following a high single dose of radiation on the surrounding sensitive structures, with small tumours away from the optic chiasm usually preferred for RS.

Despite the large number of reports on the outcome of the different available techniques of irradiation, there are no prospective comparative studies supporting the clinical efficacy relative to other treatment and mature follow-up data are needed to fully evaluate the rate of long-term side effects of stereotactic meanings of treatments in comparison with conventional fractionated radiotherapy.

Table 2.3.8.4 Summary of results of published series on RS for GH-secreting pituitary adenomas

Authors	Number of patients	Follow-up median (months)	Hormone normalization* (%)	Late toxicity (%) | Visual hypopituitarism	
Thoren *et al.*, 1991	21	64	10	0	15
Martinez *et al.*, 1998	7	26–45	NA	0	0
Pan *et al.*, 1998	15	29	NA	0	0
Morange-Ramos *et al.*, 1998	15	20	20	6	16
Lim *et al.*, 1998	20	26	30	5	5
Kim *et al.*, 1999	11	27	35	NA	NA
Landolt *et al.*, 1998	16	17	50	0	16
Mokry *et al.*, 1999	16	46	31	0	NA
Hayashi *et al.*, 1999	22	>6	41	0	0
Inoue *et al.*, 1999	12	>24	58	0	0
Zhang *et al.*, 2000	68	>12	40	NA	NA
Izawa *et al.*, 2000	29	>6	41	0	0
Pollock *et al.*, 2002	26	36	47	4	16
Attanasio *et al.*, 2003	30	46	23	0	6
Choi *et al.*, 2003	12	43	30	0	0
Jane *et al.*, 2003	64	>18	36	0	28
Petrovich *et al.*, 2003	6	36	100	0	NA
Castinetti *et al.*, 2005	82	49.5*	17	0	18
Gutt *et al.*, 2005	44	22	48	NA	NA
Kobayashi *et al.*, 2005	67	63	17	0	NA
Jezkova *et al.*, 2006	96	54	50	0	26
Pollock *et al.*, 2007	46	63	11 and 60 at 2 and 5 years	0	33 at 5 years
Jagannathan *et al.*, 2009	95	57 *	53	5~	34 (new)
Kobayashi, 2009	49	63	17 (normal or nearly normal)	11	15
Wan *et al.*, 2009	103	60 (minimum)	37	0	1.7**
Castinetti *et al.*, 2009	27	60 (minimum)	42 at 50 months	1.3**	23**
Iwai *et al.*, 2010	26	84	38	0	8
Hayashi *et al.*, 2010	25	36*	40	0	0
Erdur *et al.*, 2011	22	60	55	0	29
Sheehan *et al.*, 2011	130	30	53 at 30 months	0	34
Franzin *et al.*, 2012	103	71	56.9 at 5 years	0	7.8 (new)
Liu *et al.*, 2012	40	72	57.5	0	40 (new)
Zeiler *et al.*, 2013	21	33	30	3.9	13.2
Lee *et al.*, 2014	136	61.5	64.5 and 82.6 at 4 and 8 years	3	33.1

* mean follow-up; ~ had previous RT, ** whole series. NA: not assessed.

Updated with permission from Soldà F, Fersht N. Radiotherapy for Pituitary Tumours. [Updated 2017 Dec 5]. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK278955/

Table 2.3.8.5 Summary of results of published series on RS for ACTH-secreting pituitary adenomas

Authors	Number of patients	Follow-up median (months)	Tumour growth control rate (%)	Hormone normalization*%	Late toxicity (%) | Visual hypopituitarism	
Degerblad *et al.*, 1986	29	3–9 years	76	48	NA	55
Ganz *et al.*, 1993	4	18	NA	NA	0	NA
Seo *et al.*, 1995	2	24	100	NA	0	NA
Martinez *et al.*, 1998	3	26–45	100	100	0	0
Pan *et al.*, 1998	4	29	95	NA	0	0
Morange-Ramos *et al.*, 1998	6	20	100	66	0	16
Lim *et al.*, 1998	4	26	NA	25	2	2
Mokry *et al.*, 1999	5	26	93	20	0	2
Kim *et al.*, 1999	8	26	100	60	NA	NA
Hayashi *et al.*, 1999	10	>6	100	10	0	5
Inoue *et al.*, 1999	3	>24	100	100	0	0
Izawa *et al.*, 2000	12	>6	100	17	NA	0
Sheehan *et al.*, 2000	43	44	100	63	2	16
Hoybye *et al.*, 2001	18	17 years	100	83	0	66
Kobayashi *et al.*, 2002	20	60	100	35	NA	NA
Choi *et al.*, 2003	9	43	100	55	0	0
Jane *et al.*, 2003	45	>18	100	63	1	31
Petrovich *et al.*, 2003	4	36	NA	50	0	NA
Devin *et al.*, 2004	35	35	91	49	0	40
Castinetti *et al.*, 2007	40	54	100	42	0	NA
Jagannathan *et al.*, 2009	90	45	96	54	6	22
Pollock *et al.*, 2009	8	73	100	87 at 4 years	0	26 at 4 years
Kobayashi, 2009	25	64 (mean)	100	35	NA	NA
Wan *et al.*, 2009	68	60 (minimum)	90	28	0	1.7
Castinetti *et al.*, 2009	18	60 (minimum)	–	50 at 28 months	1.3**	23**
Hayashi *et al.*, 2010	13	36 (mean)	97	38	0	0
Sicignano *et al.*, 2012	15	60	97.7	64	NA	12.3
Zeiler *et al.*, 2013	8	35	100	50	3.9	32
Sheehan *et al.*, 2013	96	48	98	70	4	36
Marek *et al.*, 2015	26	78	90.9 at 5 and 10 years	80.7	0	23

* time not specified; ** whole series. NA, not assessed.

Table 2.3.8.6 Summary of results of published series on RS for prolactin secreting pituitary adenomas

Authors	Number of patients	Follow-up median (months)	Hormone normalization*%	Late toxicity (%) | Visual hypopituitarism	
Ganz *et al.*, 1993	3	18	0	0	NA
Martinez *et al.*, 1998	5	26–45	0	0	0
Pan *et al.*, 1998	27	29	30	0	0
Morange-Ramos *et al.*, 1998	4	20	0	0	16
Lim *et al.*, 1998	19	26	50	NA	NA
Mokry *et al.*, 1999	21	31	57	0	19
Kim *et al.*, 1999	18	27	16	NA	NA
Hayashi *et al.*, 1999	13	>6	15	NA	5
Inoue *et al.*, 1999	2	>24	50	0	0
Landolt *et al.*, 2000	20	29	25	0	NA
Pan *et al.*, 2000	128	33	41	0	NA
Izawa *et al.*, 2000	15	>6	16	0	NA
Pollock *et al.*, 2002	7	26	29	0	16**
Choi *et al.*, 2003	21	43	23	0	0
Jane *et al.*, 2003	19	>18	11	0	21
Petrovich *et al.*, 2003	12	36	83	0	NA
Pouratian *et al.*, 2006	23	55	26	7	28
Jezkova *et al.*, 2009	35	96	80	NA	NA
Kobayashi, 2009	27	37 (mean)	17	0	0
Wan *et al.*, 2009	176	60 (minimum)	23	0	1.7
Castinetti *et al.*, 2009	15	60 (minimum)	46 at 24 months	1.3**	23**
Tanaka *et al.*, 2010	22	60	18	4	42 at 4 years
Liu *et al.*, 2013	22	36	27	NA	4.5
Cohen-Inbar *et al.*, 2015	38	42.3	50	NA	30.3

** whole series.

Updated with permission from Soldà F, Fersht N. Radiotherapy for Pituitary Tumours. [Updated 2017 Dec 5]. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK278955/ Copyright © 2017, MDText.com, Inc.

Table 2.3.8.7 Summary of results of published series on LINAC-based RS for functioning and non-functioning pituitary adenomas

Authors	Tumour type	Number of patients	Follow-up mean (months)	Tumour control (%)	Hormone normalization (%)	Late toxicity (%) | Visual hypopituitarism	
Voges *et al.*, 2006 [43]	37 NFA, 64 GH, 17 ACTH, 9 Nelson, 13 PRL, 2 TSH	142	81.9	96.5	51 at 5 years	1.4	12.3
Sallabanda *et al.*, 2011 [44]	4 NFA, 11 GH, 9 ACTH, 2 PRL, 4 mixed	30**	102.6	93	65.4	NA	13.3
Runge *et al.*, 2012 [45]	NF	61	83	98	–	0	10
Wilson *et al.*, 2012 [46]	NF	51	50	100	–	0	0
Wein *et al.*, 2012 [47]	ACTH	17	23	94.1	58.8	0	11.8
Yan *et al.*, 2013 [48]	GH	22	98	95	68.2	0	22.7
Wilson *et al.*, 2013 [49]	GH	86	66	96	18.6	1.2	19.8
Bostrom *et al.*, 2015 [50]	GH	21	96	97.1	23	5	46.4
Wilson *et al.*, 2014 [51]	ACTH	36	66	97	25	0	13.9
Puataweepong *et al.*, 2015 [52]	4 NFA, 17 GH, 2 ACTH, 1 PRL	21	62	93 at 6 years	30 at 5 years*	4.7	10

* includes pts treated with fSRT; ** includes 6 re-treatments. NFA, non-functioning adenoma; NA, not assessed.

Table 2.3.8.8 Summary of results of published series on Cyberknife RS for functioning and non-functioning pituitary adenomas

Authors	Tumour type	Number of patients	Follow-up mean (months)	Tumour control (%)	Hormone normalization (%)	Late toxicity (%) | Visual hypopituitarism	
Kajiwara *et al.*, 2005	14 NFA, 3 PRL, 2 GH, 2 ACTH	21	35.3	95.2	50	4.7	9.5
Adler *et al.*, 2006	12 NFA, 4 GH, 2 ACTH, 1 PRL	19	46	94	NA	5.2	NA
Roberts *et al.*, 2007	GH	9	25.4	NA	44.4	0	33
Killory *et al.*, 2009	14 NFA, 4GH, 1 PRL, 1 TSH	20	26.6	100		0	NA
Cho *et al.*, 2009	17 NFA, 3 PRL, 6 GH	26	30	92.3	44	7.6	0
Iwata *et al.*, 2011	NFA	100	33 median	98	–	1	4%
Puaweepong *et al.*, 2015	27 NFA, 7 GH, 5 PRL, 1 ACTH	40	38.5 median	97.5	54	0	0
Iwata *et al.*, 2016	GH	52	60 median	100	20.4	0	2.2

NFA, non-functioning adenoma; NA, not assessed.

Updated with permission from Soldà F, Fersht N. Radiotherapy for Pituitary Tumours. [Updated 2017 Dec 5]. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK278955/ Copyright © 2017, MDText.com, Inc.

REFERENCES

1. Famini P, Maya MM, Melmed S. Pituitary magnetic resonance imaging for sellar and parasellar masses: ten-year experience in 2598 patients. *J Clin Endocrinol Metab*, 2011; **96**: 1633–41.
2. Brada M, Rajan B, Traish D, *et al*. The long-term efficacy of conservative surgery and radiotherapy in the control of pituitary adenomas. *Clin Endocrinol (Oxf)*, 1993; **38**: 571–8.
3. Tsang RW, Brierley JD, Panzarella T, Gospodarowicz MK, Sutcliffe SB, Simpson WJ. Radiation therapy for pituitary adenoma: treatment outcome and prognostic factors. *Int J Radiat Oncol Biol Phys*, 1994; **30**: 557–65.
4. Zierhut D, Flentje M, Adolph J, Erdmann J, Raue F, Wannenmacher M. External radiotherapy of pituitary adenomas. *Int J Radiat Oncol Biol Phys*, 1995; **33**: 307–14.
5. Gittoes NJ, Bates AS, Tse W, *et al*. Radiotherapy for non-function pituitary tumours. *Clin Endocrinol*, 1998; **48**: 331–7.
6. Rush S, Cooper PR. Symptom resolution, tumor control, and side effects following postoperative radiotherapy for pituitary macroadenomas. *Int J Radiat Oncol Biol Phys*, 1997; **37**: 1031–4.
7. Breen P, Flickinger JC, Kondziolka D, Martinez AJ. Radiotherapy for nonfunctional pituitary adenoma: analysis of long-term tumor control. *J Neurosurg*, 1998; **89**: 933–8.
8. Minniti G, Brada M. Radiotherapy and radiosurgery for Cushing's disease. *Arquivos Brasileiros de Endocrinologia e Metabologia*, 2007; **51**: 1373–80.
9. Minniti G, Clarke E, Scaringi C, Enrici RM. Stereotactic radiotherapy and radiosurgery for non-functioning and secreting pituitary adenomas. *Rep Pract Oncol Radiother*, 2016; **21**: 370–8.
10. Khoo VS, Oldham M, Adams EJ, Bedford JL, Webb S, Brada M. Comparison of intensity-modulated tomotherapy with stereotactically guided conformal radiotherapy for brain tumors. *Int J Radiat Oncol Biol Phys*, 1999; **45**: 415–25.
11. Karger CP, Jakel O, Debus J, Kuhn S, Hartmann GH. Three-dimensional accuracy and interfractional reproducibility of patient fixation and positioning using a stereotactic head mask system. *Int J Radiat Oncol Biol Phys*, 2001; **49**: 1493–504.
12. Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. Radiation dose-volume effects of optic nerves and chiasm. *Int J Radiat Oncol Biol Phys*, 2010; **76**(3 Suppl): S28–35.
13. Hiniker SM, Modlin LA, Choi CY, *et al*. Dose-response modeling of the visual pathway tolerance to single-fraction and hypofractionated stereotactic radiosurgery. *Semin Radiat Oncol*, 2016; **26**: 97–104.
14. Gill SS, Thomas DG, Warrington AP, Brada M. Relocatable frame for stereotactic external beam radiotherapy. *Int J Radiat Oncol Biol Phys*, 1991; **20**: 599–603.
15. Graham JD, Warrington AP, Gill SS, Brada M. A non-invasive, relocatable stereotactic frame for fractionated radiotherapy and multiple imaging. *Radiother Oncol*, 1991; **21**: 60–2.
16. Paganetti H, Schmitz T. The influence of the beam modulation technique on dose and RBE in proton radiation therapy. *Phys Med Biol*, 1996; **41**: 1649–63.
17. Specialised Commissioning Team NE. B01/P/c Proton Beam Therapy for Cancer in Teenagers and Young Adults 2015. Available at: https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/10/b01-pc-prtn-bm-thrpy-teens-yng-oct15.pdf [accessed 30 January 2020].
18. Brada M, Burchell L, Ashley S, Traish D. The incidence of cerebrovascular accidents in patients with pituitary adenoma. *Int J Radiat Oncol Biol Phys*, 1999; **45**: 693–8.
19. Brown PD, Blanchard M, Jethwa K, *et al*. The incidence of cerebrovascular accidents and second brain tumors in patients with pituitary adenoma: a population-based study. *Neuro-oncol Pract*, 2014; **1**: 22–8.
20. van Westrhenen A, Muskens IS, Verhoeff JJC, Smith TRS, Broekman MLD. Ischemic stroke after radiation therapy for pituitary adenomas: a systematic review. J Neuro-oncol, 2017; **135**: 1–11.

21. Brada M, Ford D, Ashley S, *et al*. Risk of second brain tumour after conservative surgery and radiotherapy for pituitary adenoma. *BMJ*, 1992; **304**: 1343–6.
22. Erridge SC, Conkey DS, Stockton D, *et al*. Radiotherapy for pituitary adenomas: long-term efficacy and toxicity. *Radiother Oncol*, 2009; **93**: 597–601.
23. Grigsby PW, Simpson JR, Emami BN, Fineberg BB, Schwartz HG. Prognostic factors and results of surgery and postoperative irradiation in the management of pituitary adenomas. *Int J Radiat Oncol Biol Phys*, 1989; **16**: 1411–17.
24. Minniti G, Traish D, Ashley S, Gonsalves A, Brada M. Fractionated stereotactic conformal radiotherapy for secreting and nonsecreting pituitary adenomas. *Clin Endocrinol*, 2006; **64**: 542–8.
25. Kim MY, Kim JH, Oh YK, Kim E. Long-term outcomes of surgery and radiotherapy for secreting and non-secreting pituitary adenoma. *Radiat Oncol J*, 2016; **34**: 121–7.
26. Barrande G, Pittino-Lungo M, Coste J, *et al*. Hormonal and metabolic effects of radiotherapy in acromegaly: long-term results in 128 patients followed in a single center. *J Clin Endocrinol Metab*, 2000; **85**: 3779–85.
27. Biermasz NR, Dulken HV, Roelfsema F. Postoperative radiotherapy in acromegaly is effective in reducing GH concentration to safe levels. *Clin Endocrinol*, 2000; **53**: 321–7.
28. Minniti G, Osti M, Jaffrain-Rea ML, Esposito V, Cantore G, Maurizi Enrici R. Long-term follow-up results of postoperative radiation therapy for Cushing's disease. *J Neuro-oncol*, 2007; **84**: 79–84.
29. Tsagarakis S, Grossman A, Plowman PN, *et al*. Megavoltage pituitary irradiation in the management of prolactinomas: long-term follow-up. *Clin Endocrinol*, 1991; **34**: 399–406.
30. Johnston DG, Hall K, Kendall-Taylor P, *et al*. The long-term effects of megavoltage radiotherapy as sole or combined therapy for large prolactinomas: studies with high definition computerized tomography. *Clin Endocrinol*, 1986; **24**: 675–85.
31. Mehta AE, Reyes FI, Faiman C. Primary radiotherapy of prolactinomas. Eight- to 15-year follow-up. *Am J Med*, 1987; **83**: 49–58.
32. Brada M, Ashley S, Ford D, Traish D, Burchell L, Rajan B. Cerebrovascular mortality in patients with pituitary adenoma. *Clin Endocrinol*, 2002; **57**: 713–17.
33. Tsang RW, Laperriere NJ, Simpson WJ, Brierley J, Panzarella T, Smyth HS. Glioma arising after radiation therapy for pituitary adenoma. A report of four patients and estimation of risk. *Cancer*, 1993; **72**: 2227–33.
34. Erfurth EM, Bulow B, Svahn-Tapper G, *et al*. Risk factors for cerebrovascular deaths in patients operated and irradiated for pituitary tumors. *J Clin Endocrinol Metab*, 2002; **87**: 4892–9.
35. Colin P, Jovenin N, Delemer B, *et al*. Treatment of pituitary adenomas by fractionated stereotactic radiotherapy: a prospective study of 110 patients. *Int J Radiat Oncol Biol Physics*, 2005; **62**: 333–41.
36. Milker-Zabel S, Debus J, Thilmann C, Schlegel W, Wannenmacher M. Fractionated stereotactically guided radiotherapy and radiosurgery in the treatment of functional and nonfunctional adenomas of the pituitary gland. *Int J Radiat Oncol Biol Phys*, 2001; **50**: 1279–86.
37. Liscak R, Vladyka V, Marek J, Simonova G, Vymazal J. Gamma knife radiosurgery for endocrine-inactive pituitary adenomas. *Acta Neurochir (Wien)*, 2007; **149**: 999–1006.
38. Lee CC, Kano H, Yang HC, *et al*. Initial gamma knife radiosurgery for nonfunctioning pituitary adenomas. *J Neurosurg*, 2014; **120**: 647–54.
39. Lee CC, Vance ML, Xu Z, *et al*. Stereotactic radiosurgery for acromegaly. *J Clin Endocrinol Metab*, 2014; **99**: 1273–81.
40. Mehta GU, Ding D, Patibandla MR, *et al*. Stereotactic radiosurgery for Cushing disease: results of an international, multicenter study. *J Clin Endocrinol Metab*, 2017; **102**: 4284–91.
41. Kobayashi T. Long-term results of stereotactic gamma knife radiosurgery for pituitary adenomas. Specific strategies for different types of adenoma. *Prog Neurol Surg*, 2009; **22**: 77–95.
42. Pouratian N, Sheehan J, Jagannathan J, Laws ER, Jr., Steiner L, Vance ML. Gamma knife radiosurgery for medically and surgically refractory prolactinomas. *Neurosurgery*, 2006; **59**: 255–66; discussion 266.
43. Voges J, Kocher M, Runge M, *et al*. Linear accelerator radiosurgery for pituitary macroadenomas: a 7-year follow-up study. *Cancer*, 2006; **107**: 1355–64.
44. Sallabanda K, Usychkin S, Puebla F, *et al*. Stereotactic radiosurgery in pituitary adenomas: long-term single institution experience and role of the hypothalamic-pituitary axis. *J Radiosurg SBRT*, 2011; **1**: 213–20.
45. Runge MJ, Maarouf M, Hunsche S, *et al*. LINAC-radiosurgery for nonsecreting pituitary adenomas. Long-term results. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft [et al]*, 2012; **188**: 319–25.
46. Wilson PJ, De-Loyde KJ, Williams JR, Smee RI. A single centre's experience of stereotactic radiosurgery and radiotherapy for non-functioning pituitary adenomas with the linear accelerator (LINAC). *J Clin Neurosci*, 2012; **19**: 370–4.
47. Wein L, Dally M, Bach LA. Stereotactic radiosurgery for treatment of Cushing disease: an Australian experience. *Int Med J*, 2012; **42**: 1153–6.
48. Yan JL, Chang CN, Chuang CC, *et al*. Long-term follow-up of patients with surgical intractable acromegaly after linear accelerator radiosurgery. *J Formos Med Assoc*, 2013; **112**: 416–20.
49. Wilson PJ, De-Loyde KJ, Williams JR, Smee RI. Acromegaly: a single centre's experience of stereotactic radiosurgery and radiotherapy for growth hormone secreting pituitary tumours with the linear accelerator. *J Clin Neurosci*, 2013; **20**: 1506–13.
50. Bostrom JP, Kinfe T, Meyer A, *et al*. Treatment of acromegaly patients with risk-adapted single or fractionated stereotactic high-precision radiotherapy: high local control and low toxicity in a pooled series. *Strahlentherapie und Onkologie: Organ der Deutschen Rontgengesellschaft [et al]*, 2015; **191**: 477–85.
51. Wilson PJ, Williams JR, Smee RI. Cushing's disease: a single centre's experience using the linear accelerator (LINAC) for stereotactic radiosurgery and fractionated stereotactic radiotherapy. *J Clin Neurosci*, 2014; **21**: 100–6.
52. Puataweepong P, Dhanachai M, Hansasuta A, *et al*. Outcomes for pituitary adenoma patients treated with linac-based stereotactic radiosurgery and radiotherapy: a long term experience in Thailand. *Asian Pac J Cancer Prev*, 2015; **16**: 5279–84.
53. Ludecke DK, Lutz BS, Niedworok G. The choice of treatment after incomplete adenomectomy in acromegaly: proton versus high voltage radiation. *Acta Neurochirurg*, 1989; **96**: 32–8.
54. Petit JH, Biller BM, Coen JJ, *et al*. Proton stereotactic radiosurgery in management of persistent acromegaly. *Endocr Pract*, 2007; **13**: 726–34.

55. Petit JH, Biller BM, Yock TI, *et al*. Proton stereotactic radiotherapy for persistent adrenocorticotropin-producing adenomas. *J Clin Endocrinol Metab*, 2008; **93**: 393–9.
56. Ronson BB, Schulte RW, Han KP, Loredo LN, Slater JM, Slater JD. Fractionated proton beam irradiation of pituitary adenomas. *Int J Radiat Oncol Biol Phys*, 2006; **64**: 425–34.
57. Wattson DA, Tanguturi SK, Spiegel DY, *et al*. Outcomes of proton therapy for patients with functional pituitary adenomas. *Int J Radiat Oncol Biol Phys*, 2014; **90**: 532–9.
58. Schoenthaler R, Albright NW, Wara WM, Phillips TL, Wilson CB, Larson DA. Re-irradiation of pituitary adenoma. *Int J Radiat Oncol Biol Phys*, 1992; **24**: 307–14.
59. Edwards AA, Swords FM, Plowman PN. Focal radiation therapy for patients with persistent/recurrent pituitary adenoma, despite previous radiotherapy. *Pituitary*, 2009; **12**: 30–4.
60. Swords FM, Monson JP, Besser GM, *et al*. Gamma knife radiosurgery: a safe and effective salvage treatment for pituitary tumours not controlled despite conventional radiotherapy. *Eur J Endocrinol*, 2009; **161**: 819–28.
61. Verma J, McCutcheon IE, Waguespack SG, Mahajan A. Feasibility and outcome of re-irradiation in the treatment of multiply recurrent pituitary adenomas. *Pituitary*, 2014; **17**: 539–45.
62. Swords FM, Allan CA, Plowman PN, *et al*. Stereotactic radiosurgery XVI: a treatment for previously irradiated pituitary adenomas. *J Clin Endocrinol Metab*, 2003; **88**: 5334–40.
63. Jenkins PJ, Bates P, Carson MN, Stewart PM, Wass JA. Conventional pituitary irradiation is effective in lowering serum growth hormone and insulin-like growth factor-I in patients with acromegaly. *J Clin Endocrinol Metab*, 2006; **91**: 1239–45.

2.3.9 Prolactinomas and Hyperprolactinaemia (Including Macroprolactinaemia)

Nicholas A. Tritos and Anne Klibanski

Introduction

Prolactin is a single chain (23 kDa) polypeptide hormone, which is secreted by anterior pituitary lactotroph cells [1]. Prolactin secretion is predominantly under tonic inhibition by hypothalamic dopamine working through the D2 receptor [2]. In addition, a variety of factors, including thyrotropin-releasing hormone (TRH), vasoactive intestinal peptide (VIP), serotonin, and other neuropeptides and neurotransmitters influence prolactin secretion to lesser degrees. Prolactin secretion is also affected by oestrogen. Prolactin exerts its actions by binding to dimerized forms of its cognate receptor, a member of the cytokine receptor superfamily [3]. Prolactin has a central role in reproduction and lactation.

Extrapituitary sources of prolactin have been described, including the decidua and other sites [4]. Prolactin synthesis in the decidua may promote maternal tolerance to fetus *in utero*. A host of metabolic functions and paracrine effects have been proposed for extrapituitary prolactin on the basis of animal studies [5]. However, the function of prolactin synthesized outside the pituitary remains obscure in humans.

Hyperprolactinemia can be a consequence of increased prolactin secretion and/or decreased prolactin clearance. There are numerous causes of hyperprolactinemia as discussed in a separate section [6]. Prolactin-secreting pituitary adenomas (prolactinomas) are the most frequent functioning pituitary tumours.

Hyperprolactinemia may cause central hypogonadism, infertility, and bone loss (in hypogonadal patients). Galactorrhoea may occur (especially in women) and gynaecomastia may develop in men. However, prolactin excess can also be asymptomatic. Large pituitary tumours associated with hyperprolactinemia may additionally cause mass effect, including headache, vision loss, and anterior hypopituitarism.

The diagnosis of hyperprolactinemia is made by measuring serum prolactin in specimens optimally obtained in the fasting state while minimizing physiologic stress [7]. The hook effect, heterophilic antibodies, exogenous biotin, and macroprolactinaemia may confound the interpretation of the results of prolactin immunoassays, as will be detailed in a separate section. Pituitary imaging is generally advised in patients with hyperprolactinemia in the absence of a clear-cut, alternative explanation (such as pregnancy).

Medical management of hyperprolactinemia with dopamine agonists, including bromocriptine and cabergoline, is generally first-line therapy for this condition [6, 7]. However, asymptomatic patients with hyperprolactinemia who do not have macroadenomas can often be carefully monitored without therapy, as will be subsequently discussed.

The management of women with hyperprolactinemia and prolactinomas during preconception and pregnancy raises special concerns regarding the effects of prolactin excess during preconception, the impact of pregnancy on prolactinomas and the safety of medical therapy [8, 9].

Pituitary surgery and radiation therapy generally represent second- and third-line therapies, respectively, in patients with hyperprolactinemia and prolactinomas, as will be detailed in separate sections [7].

Causes and Epidemiology of Hyperprolactinemia

Hyperprolactinemia is fairly common in the general population, wherein its prevalence is estimated to be approximately 0.4% [6, 10]. There is a large number of causes of hyperprolactinemia, which can be grouped into physiologic states, medication-related, systemic illnesses, and pituitary lesions (please see **Box 2.3.9.1**). In addition, hyperprolactinemia can be idiopathic.

There are diverse physiologic stimuli of prolactin secretion, including food consumption, sleep, exercise, nipple stimulation and coitus [6]. Serum prolactin levels rise transiently after meals or during slow wave sleep. Exercise leads to an acute, albeit transient, increase in prolactin levels, which is not sustained long-term in individuals who exercise frequently. Nevertheless, it is prudent to obtain blood specimens intended for prolactin assay after an overnight fast while minimizing exercise, stress, or other factors that might trigger prolactin secretion. Pregnancy and lactation represent major physiologic causes of hyperprolactinemia [9]. Stimulated by oestrogen of placental origin, pituitary lactotrophs undergo substantial hyperplasia during pregnancy. As a consequence, serum prolactin levels rise progressively during gestation and may reach or exceed 200–300 ng/ml (8700–13 000 pmol/L). Prolactin secretion has a central role in inducing differentiation of the breast epithelium, thus preparing the mammary gland for lactation after delivery. Prolactin levels remain elevated for several months after delivery in women who are nursing. There are also secretory prolactin spikes that occur in response to each suckling episode.

Box 2.3.9.1 Causes of hyperprolactinemia

Physiologic
- Pregnancy
- Lactation
- Food consumption
- Sleep
- Exercise
- Coitus
- Nipple stimulation
- Other physiologic stress (seizure, hypoglycaemia, etc.)

Medications
- Phenothiazines
- Haloperidol
- Risperidone, paliperidone
- Tricyclic antidepressants
- Fluoxetine
- Monoamine oxidase inhibitors
- Metoclopramide, domperidone
- Alpha methyldopa
- Verapamil
- Opioids
- Cocaine

Systemic conditions
- End-stage renal disease
- Cirrhosis
- Primary hypothyroidism
- Primary adrenal insufficiency
- Polycystic ovary syndrome
- Ectopic secretion (rarely)
- Genetic (prolactin receptor mutation)
- Spinal cord injury
- Chest wall lesions (burns, zoster, etc.)

Pituitary lesions
- Prolactin-secreting adenomas
- Cosecreting tumours (acromegaly; rarely Cushing's disease or thyrotropinoma)
- Non-functioning pituitary adenomas
- Cystic sellar lesions
- Meningiomas
- Sarcoidosis
- Hypophysitis
- Idiopathic

Many medications can cause hyperprolactinemia, including antipsychotic agents (both typical and atypical), tricyclic antidepressants, fluoxetine, monoamine oxidase inhibitors, metoclopramide, domperidone, opioids, cocaine, alpha methyldopa, and verapamil [11]. Among antipsychotic drugs, older agents (such as chlorpromazine and haloperidol) and some of the atypical antipsychotic drugs (such as risperidone and paliperidone) may raise serum prolactin levels substantially, sometimes exceeding 100 ng/ml (4300 pmol/L) as a result of dopamine receptor inhibition. In contrast, other atypical agents, including aripiprazole and ziprasidone, are less likely to cause hyperprolactinemia, in part because of their weaker inhibition of dopamine receptors and their partial agonist effects. Although some patients can be switched to aripiprazole as an antipsychotic medication, add-on aripiprazole therapy may suppress hyperprolactinemia due to other neuroleptic agents as a consequence of its partial dopamine agonist effect [12–14]. Aripiprazole has also been used to treat patient with known prolactinomas who require antipsychotic therapy in whom dopamine agonists such as cabergoline or bromocriptine may be contraindicated.

Several systemic illnesses and conditions, including primary hypothyroidism, (rarely) primary adrenal insufficiency, chronic renal failure, cirrhosis, polycystic ovary syndrome, spinal cord injuries, chest wall lesions (such as zoster or burns) have been associated with hyperprolactinemia [6]. In addition, ectopic prolactin secretion has been reported in a very small number of patients with a variety of tumours (ovarian teratoma, gonadoblastoma, bronchogenic carcinoma, and others). Germline, heterozygous inactivating mutation of the prolactin receptor gene were reported in 3 sisters with hyperprolactinemia who had no evidence of a sellar mass; 2 of these patients had oligomenorrhea and 1 had infertility [15].

A large number of sellar masses can cause hyperprolactinemia, either as a consequence of prolactin secretion by tumorous lactotrophs (prolactinomas) or as a result of disinhibition of prolactin secretion by normal lactotrophs, induced by interference with dopamine outflow from the hypothalamus to the pituitary ('stalk effect') [6]. Examples of the latter include clinically non-functioning pituitary adenomas, cystic sellar lesions (Rathke's cleft cysts, craniopharyngiomas, and others), meningiomas, hypophysitis, sarcoidosis, and Langerhans cell histiocytosis. Stalk transection and radiation therapy to the sella are also associated with hyperprolactinemia, likely as a consequence of hypothalamic dysfunction.

Prolactinomas are the most common type of secretory (functioning) pituitary adenomas. Earlier studies suggested that the

incidence and prevalence of prolactinomas is about 10 cases per million per year and 100 cases per million in the general population, respectively [16]. In contrast, several recent reports have found that the incidence and prevalence of clinically detected prolactin-secreting tumours is about five times higher than previously reported (please see Table 2.3.9.1) [17–21]. Such differences may be partly explained based on improved detection rates in the recent past. Autopsy studies have found that clinically unsuspected pituitary adenomas are present in approximately 11% of subjects autopsied, almost all of them (>99%) microadenomas [22]. Of note, about 40% of adenomas detected at autopsy stained for prolactin on immunohistochemistry; whether these tumours were truly asymptomatic during life is unclear [23].

Most prolactinomas are sporadic. However, the multiple endocrine neoplasia 1 (MEN 1) syndrome, caused by germline inactivating mutations of *MENIN* (a tumour suppressor gene) accounts for a minority of cases [24, 25]. The coexistence of a pituitary adenoma and primary hyperparathyroidism (a constellation that defines an index case for MEN 1) was reported in 14% of patients with prolactinomas who were screened for hypercalcaemia and did not have a positive family history for the syndrome. In addition, a small minority of patients with prolactinomas may have a familial predisposition as a consequence of germline inactivating mutations in the *aryl hydrocarbon receptor interacting protein (AIP)* gene [26, 27]. This is an autosomal dominant trait with incomplete penetrance. Patients with *AIP* mutations have an increased risk of developing somatotropinomas, prolactinomas, or cosecreting tumours. Moreover, these patients generally present at a younger age and are more likely to harbour locally aggressive macroadenomas than patients with sporadic disease.

Microprolactinomas (<10 mm in greatest diameter) predominate at the time of clinical diagnosis. Macroadenomas (≥10 mm in greatest diameter) represent a larger proportion of prolactinomas in men than in women, as well as in children/adolescents than in adults.

Microprolactinomas have a relatively low growth potential. In earlier studies, tumour growth was found in 9 out of 139 women (6.5%) followed up to 8 years [28–31]. Macroprolactinomas have a higher growth potential than microadenomas and can be locally aggressive. Rarely, these tumours may evolve into carcinomas, defined by the development of metastases within the skull or at distant sites (liver, lymph nodes, bone) [32]. A minority of prolactin-secreting pituitary adenomas may cosecrete growth hormone or occasionally thyrotropin (TSH), corticotropin (ACTH), or follicle-stimulating hormone (FSH).

Idiopathic hyperprolactinemia is characterized by the absence of a demonstrable cause despite thorough diagnostic evaluation. These patients likely represent a diverse group. It is indeed possible that some patients have very small prolactinomas that are below the resolution of magnetic resonance imaging (MRI). Other patients may have dysregulated prolactin secretion. Idiopathic hyperprolactinemia has been reported to resolve in about one-third of patients without treatment. Pituitary microadenomas may occasionally develop during follow-up of patients with idiopathic hyperprolactinemia (reported in 23 out of 199 patients who were followed for 2–6 years) [28, 33].

Clinical Features of Hyperprolactinaemia and Prolactinomas

In patients with prolactinomas, the magnitude of prolactin elevation is generally proportionate to tumour size [6, 7]. This is an important concept because modest degrees of prolactin excess in patients with large sellar masses demonstrated on MRI scan may not represent prolactinomas, or in some cases a diagnosis other than a pituitary adenoma. Patients with hyperprolactinemia can be either asymptomatic or demonstrate clinical manifestations of prolactin excess (please see Table 2.3.9.2). Hyperprolactinemia often leads to central hypogonadism in patients of both genders, primarily caused by suppression of hypothalamic gonadotropin-releasing hormone (GnRH) pulsations [34].

In women, hyperprolactinemia is often associated with galactorrhoea, irregular menstrual cycles (oligomenorrhea or secondary amenorrhea) and infertility [6, 7]. Bone loss may occur as a consequence of long-standing hypogonadism. In contrast, hyperprolactinaemic women with regular menses are unlikely to develop osteoporosis because of an elevated prolactin level unless prior amenorrhea had occurred [35, 36]. Women may develop hyperandrogenism (hirsutism, acne) as a consequence of stimulation of adrenal androgen secretion by prolactin excess. In men, hyperprolactinemia may lead to low libido, erectile dysfunction, and infertility [6, 7]. Gynaecomastia may develop in some men; however, galactorrhoea is quite uncommon in men. Anaemia and bone loss may occur in men with long-standing hypogonadism. Children and adolescents of both genders may present with delayed puberty, including primary amenorrhea in girls [37].

Patients with large sellar masses (including macroprolactinomas) may develop anterior hypopituitarism and require thorough evaluation of anterior pituitary function. However, diabetes insipidus

Table 2.3.9.1 Epidemiology of prolactinomas based on the findings of several recent studies*

Author (year)	Country	Population size	Incidence (per million)	Prevalence (per million)
Daly *et al.*, 2006	Belgium	72 000	NR	620
Fernandez *et al.*, 2010	United Kingdom	81 500	NR	444
Raappana *et al.*, 2010	Finland	730 000	22	NR
Tjornstrand *et al.*, 2014	Sweden	1 600 000	16	NR
Agustsson *et al.*, 2015	Iceland	330 000	36 (women); 14 (men)	544

*Data were extracted from references 17–21; NR, not reported.

Table 2.3.9.2 Clinical manifestations of hyperprolactinemia and prolactinomas

System or area	Women	Men	Children/adolescents
Pituitary gonadal axis	Secondary amenorrhea or oligomenorrhea, infertility	Erectile dysfunction, low libido, infertility	Delayed puberty, primary amenorrhea (in girls)
Breast	Galactorrhoea	Gynaecomastia, galactorrhoea (very uncommon)	Galactorrhoea (more common in girls); gynaecomastia (in boys)
Adrenals	Androgenic manifestations (hirsutism, acne)		
Skeleton	Osteoporosis	Osteoporosis	Delayed growth
Sella	Headache, visual field defects, ophthalmoplegia (rare), anterior hypopituitarism	Headache, visual field defects, ophthalmoplegia (rare), anterior hypopituitarism	Headache, visual field defects, anterior hypopituitarism

is distinctly rare in patients with pituitary adenomas who did not undergo pituitary surgery [38].

Additional evidence of mass effect may occur in patients with macroadenomas [6]. Headache may occur as a result of pressure exerted on the dura. Patients with pituitary apoplexy typically present with severe headache of acute onset as a result of haemorrhage within a pituitary adenoma. Patients whose tumours extend into the suprasellar cistern may experience visual field deficits as a consequence of compression of the optic apparatus. Central scotomas may occur as a result of prechiasmatic optic nerve compression, whereas impingement on the optic chiasm may result in bitemporal vision loss; compression of one of the optic tracts may lead to homonymous hemianopsia.

Pituitary tumours may also extend laterally into the cavernous sinuses where they may impinge upon cranial nerves III, IV, V1, V2, VI [6]. However, the vast majority of patients with cavernous sinus involvement have no clinical evidence of cranial nerve dysfunction, with the exception of patients with pituitary apoplexy or occasional patients with aggressive pituitary adenomas, who may experience ophthalmoplegia, facial pain or numbness. Very infrequently, large pituitary tumours compressing the temporal lobes may cause complex partial seizures. Giant prolactinomas that extend into the third ventricle may lead to obstructive hydrocephalus [39, 40]. Tumours that extend inferiorly may erode the sphenoid and extend into the sphenoid sinus [39, 40]. These patients may rarely experience rhinorrhoea as a result of cerebrospinal fluid (CSF) leak, precipitated by dopamine agonist therapy that causes shrinkage of a macroprolactinoma and unmasks rents in the dura and bone, and require surgical intervention to correct the leak.

A minority of patients have manifestations of associated conditions, including patients with the MEN 1 syndrome, who often have primary hyperparathyroidism (90%) or entero-pancreatic neuroendocrine tumours (30–40%). These conditions are discussed in separate chapters. To identify the possibility of MEN 1 syndrome in the absence of a pertinent family history, it is prudent to measure serum calcium levels in all patients presumed to have prolactinomas.

Diagnostic Evaluation

The diagnosis of hyperprolactinemia is made by documenting elevated serum prolactin levels using a prolactin immunoassay. Pituitary imaging is generally advisable in these patients in the absence of an obvious explanation for hyperprolactinemia (such as pregnancy), as will be detailed next.

Laboratory Testing

Prolactin secretion is pulsatile; about 14 peaks in prolactin secretion occur every 24 hours [41]. The pulse amplitude increases during slow wave sleep. In patients with suspected hyperprolactinemia, blood specimens are obtained by venipuncture while avoiding physiologic stimuli of prolactin secretion (such as recent food consumption or exercise). Borderline prolactin elevations should be confirmed by repeat testing.

Prolactin is measured in serum specimens using two-site immunoassays, which feature a capture antibody that collects the antigen (prolactin) onto a solid surface (such as microbeads or coated tubes), and a reporter antibody that binds to the antigen (prolactin) at a different site and subsequently yields a detection signal by means of a luminescent or radiolabelled tag [42, 43].

Several caveats need to be considered in the interpretation of the results of prolactin immunoassays, including the hook effect, artefacts related to the presence of heterophilic antibodies, the presence of macroprolactin or exogenous biotin (please see Box 2.3.9.2) [44]. Immunoassay artefacts should be considered when serum prolactin levels do not decrease in a linear manner upon serial serum dilutions or in situations wherein there appears to be a discrepancy between the clinical picture and serum prolactin levels.

The hook effect is an immunoassay artefact that may lead to substantial underreporting of true prolactin values in specimens containing very high concentrations of the analyte (prolactin, in this case) [45]. This can rarely occur in patients with larger macroadenomas, wherein it is a consequence of the presence of vast stoichiometric excess of prolactin over the concentration of the two antibodies (capture and reporter) used in the immunoassay. In these rare cases, the concentration of the two antibody

Box 2.3.9.2 Caveats and artefacts that may confound the interpretation of prolactin levels

Hook effect
Heterophilic antibodies
Macroprolactin
Exogenous biotin

molecules is insufficient to allow formation of the heterotrimeric complex (consisting of capture antibody—prolactin—reporter antibody) that is required for prolactin detection. This artefact appears to be uncommon but is clinically significant, since patients with presumed prolactinomas are typically offered medical therapy as a first line intervention, whereas patients with minimally elevated prolactin levels (as a result of stalk effect) are referred for surgery. The hook effect can be circumvented by diluting the serum specimen (1:10 or 1:100), which helps restore an optimal stoichiometric relationship between prolactin and the two antibodies *in vitro*. Thus, it is advisable to request that the laboratory run serum prolactin specimens in dilution, particularly in patients with large tumours.

Heterophilic antibodies, such as human antimouse antibodies, can be present in human serum and falsely raise the results of prolactin immunoassays [46]. In these cases, the heterophilic antibody forms a bridge between capture and reporter antibodies in the absence of prolactin, generating a false detection signal. To avoid this artefact, immunoassays often include 'blocking' animal serum or heterophilic antibody blocking tubes.

Macroprolactin denotes the presence of prolactin aggregates in the serum, consisting of several prolactin monomers, either alone or together with immunoglobulin (IgG) [47–49]. In healthy individuals, macroprolactin accounts for less than 10% of prolactin in the serum, the rest being monomeric prolactin. Macroprolactin species can be detected and resolved by gel filtration. Since this is a labour-intensive process, it is used in research but very rarely in routine clinical laboratories. In daily practice, polyethylene glycol (PEG) is added to precipitate macroprolactin in serum specimens after centrifugation. Prolactin is then reassayed in the supernatant; the difference between prolactin concentration in the original serum specimen minus that in the supernatant can then be taken to quantify the presence of macroprolactin in the original specimen. Of note, hyperprolactinemia can be entirely attributed to the presence of macroprolactin only if the prolactin level is normal in the supernatant after PEG-induced macroprolactin precipitation.

Macroprolactin appears to have decreased biologic activity in some bioassays [47]. Women with macroprolactinaemia who were treated with dopamine agonist therapy did not show clear-cut improvement in menstrual patterns [48, 50]. As a corollary, the presence of macroprolactinaemia should be suspected in patients with hyperprolactinemia who have no evident symptoms that are typically associated with prolactin excess. Macroprolactinaemia has been reported in patients with either pituitary tumours or idiopathic hyperprolactinemia [51]. Therefore, pituitary imaging should be considered in patients with macroprolactinaemia. Overall, the clinical utility of macroprolactin assays is very limited in symptomatic patients with hyperprolactinemia.

Biotin ingestion in large doses can introduce artefacts, if the two-site immunoassay employs the biotin streptavidin reaction, wherein the capture antibody is biotinylated and collects prolactin onto streptavidin-coated microbeads [52, 53]. The presence of exogenous biotin in excess can interfere with this assay step, thus leading to substantial underreporting of prolactin levels. To avoid this artefact, patients can be cautioned to avoid consuming biotin for several days before submitting blood specimens for testing.

Once hyperprolactinemia is established, additional testing is required in order to clarify the underlying aetiology and assess other aspects of the patient's pituitary function [6]. It is advisable to consider the possibility of pregnancy in hyperprolactinaemic women of reproductive age and obtain a pregnancy test. In addition, laboratory evaluation of kidney and thyroid function is advised to exclude renal dysfunction and primary hypothyroidism, respectively, as underlying causes of hyperprolactinemia. Laboratory testing for hypogonadism is generally advised in hyperprolactinaemic patients (with the obvious exception of pregnancy or women who maintain menses despite elevated prolactin levels). The possibility of growth hormone cosecretion should be considered in hyperprolactinaemic patients with a sellar mass and serum IGF-1 levels be measured. In addition, all patients with macroadenomas require evaluation for hypopituitarism.

Imaging

Pituitary imaging is recommended in hyperprolactinaemic patients in order to detect an underlying sellar mass unless another compelling aetiology (such as pregnancy or severe primary hypothyroidism) is identified [54]. High resolution pituitary imaging should be obtained using a dedicated imaging protocol, preferably using MRI. Computed tomography (CT) is usually being advised as an alternative in patients who have contraindications to MRI, although the resolution of CT is more limited.

Patients taking medications that may raise serum prolactin levels cannot be always presumed to have medication-induced hyperprolactinemia. In most of these patients, baseline prolactin levels (obtained before the initiation of the medication in question) are not available. However, if the potentially offending medication can be safely discontinued, then prolactin levels should be rechecked 4 days later [6]. If hyperprolactinemia resolves on follow-up testing, then the medication can be reasonably presumed to be the culprit. However, it is often inadvisable to discontinue medications suspected to induce hyperprolactinemia (such as antipsychotics) and coordination with the prescribing physician is essential. Pituitary imaging is advised in these patients to exclude an underlying sellar lesion.

A thorough review of imaging data can be very helpful in narrowing down the differential diagnosis in hyperprolactinaemic patients, besides yielding important information on lesion size and its relationship to surrounding structures (please see **Figure 2.3.9.1** and **Figure 2.3.9.2**) [6, 7].

In patients with prolactinomas, there is a good correlation between tumour size and the magnitude of prolactin elevation. Patients with macroprolactinomas generally have serum prolactin levels that exceed 200 ng/ml (8700 pmol/L). Prolactin levels may not be proportionate to tumour size in patients with cystic prolactinomas, or poorly functioning prolactinomas wherein only a minority of cells secrete prolactin. The former can be diagnosed based on MRI scan characteristics. In contrast, patients with non-functioning lesions that cause stalk effect generally raise prolactin levels rather modestly (usually below 150 ng/ml [6500 pmol/L]). As a corollary, patients with macroadenomas and mildly elevated prolactin levels are likely to have a non-functioning lesion provided that the hook effect is excluded.

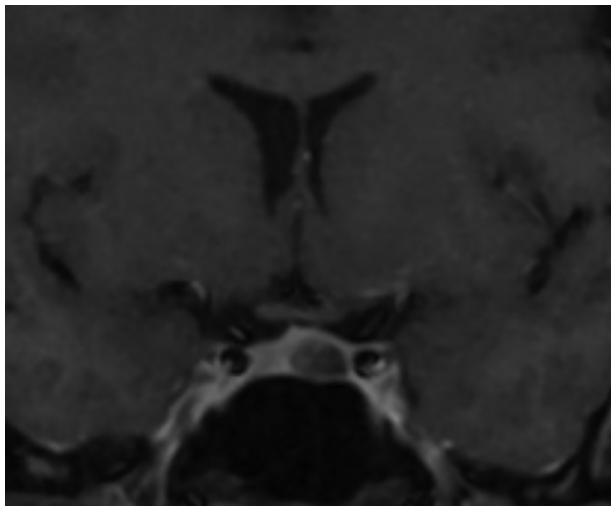


Figure 2.3.9.1 Coronal T1–weighted MRI image of the sella of a 34-year-old woman with a microadenoma, measuring 5 mm in greatest diameter (image obtained after gadolinium administration). She presented with secondary amenorrhoea. Her serum prolactin level was 82 ng/ml (3560 pmol/L) [normal, up to 20 ng/ml (870 pmol/L)]. Additional tests showed evidence of central hypogonadism. She was treated with bromocriptine, leading to restoration of menses and prolactin normalization. Tumour size decreased substantially on follow-up imaging.

Follow-up pituitary imaging is helpful in assessing tumour response to medical therapy and establishing the underlying aetiology [6, 7]. Prolactinomas generally decrease in size in response to dopamine agonist therapy, whereas non-functioning sellar masses are unlikely to shrink and may even grow after the institution of dopamine agonist therapy.

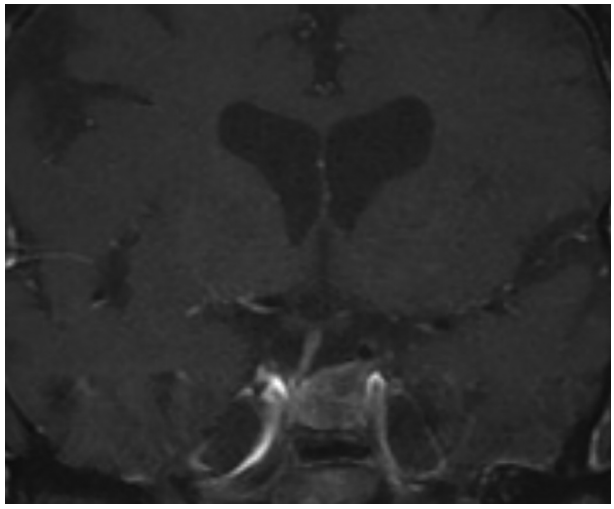


Figure 2.3.9.2 Coronal T1–weighted, gadolinium–enhanced MRI image of the sella of a 52-year-old man with a macroprolactinoma, measuring 2 cm in greatest diameter. He presented with new onset headache, low libido, and erectile dysfunction. His serum prolactin level was 302 ng/ml (13 130 pmol/L) [normal, up to 15 ng/ml (650 pmol/L)]. Evidence of central hypogonadism was present. He was treated with cabergoline, leading to symptomatic improvement, prolactin normalization, resolution of hypogonadism, and decrease in tumour size on follow-up imaging.

Management

The goals of therapy include control of hyperprolactinemia and amelioration of related symptoms, including resolution of hypogonadism. In addition, tumour control, relief of mass effect and restoration of pituitary function are important goals in patients with macroprolactinomas [6, 7].

Dopamine agonist therapy is generally first line therapy in patients who require treatment [6, 7]. Pituitary surgery and radiation therapy are second- and third-line interventions in selected patients, as will be discussed next.

Asymptomatic patients with microadenomas or no evident sellar mass, who are eugonadal and are not actively pursuing pregnancy can be followed expectantly without therapy [6, 7]. These patients require periodic follow-up, monitoring of prolactin levels, and periodic pituitary imaging to detect any evidence of progression (such as the development of symptoms or tumour growth) that would require the institution of medical therapy.

Medical Therapy

Medical therapy with dopamine agonists, including bromocriptine, cabergoline, or quinagolide (the latter being available in some countries outside the United States), represents the cornerstone of management in hyperprolactinaemic patients. As shown in Box 2.3.9.3, indications for dopamine agonist therapy in patients with hyperprolactinemia/prolactinoma include the presence of central hypogonadism, current interest in fertility, bothersome galactorrhoea, gynaecomastia, androgenic manifestations in women (hirsutism, acne), and tumour control in patients with presumed macroprolactinomas (with or without mass effect) [6, 7]. Of note, hyperprolactinaemic patients with stable microadenomas whose only symptoms are related to central hypogonadism and are not pursuing pregnancy can also be treated with sex steroid replacement and forego dopamine agonist therapy [6, 7].

Bromocriptine and cabergoline are both ergot alkaloid derivatives, whereas quinagolide is a non-ergot dopamine agonist [6, 7]. Pergolide was previously used off-label in hyperprolactinaemia, but was withdrawn from the United States and other countries because of the increased risk of cardiac valvulopathy associated with its use in high doses among patients with Parkinson's disease.

Bromocriptine is the first dopamine agonist introduced to treat hyperprolactinemia and prolactinomas [6, 7]. Bromocriptine therapy leads to prolactin normalization in about 75% of patients; most treated women resume menstruating regularly [1].

Box 2.3.9.3 Possible indications for dopamine agonist therapy in patients with hyperprolactinemia/prolactinomas

Central hypogonadism
Infertility
Galactorrhoea (if bothersome)
Gynaecomastia
Hirsutism, acne
Large tumour size (macroprolactinoma)
Mass effect (headache, visual field defects, etc.)

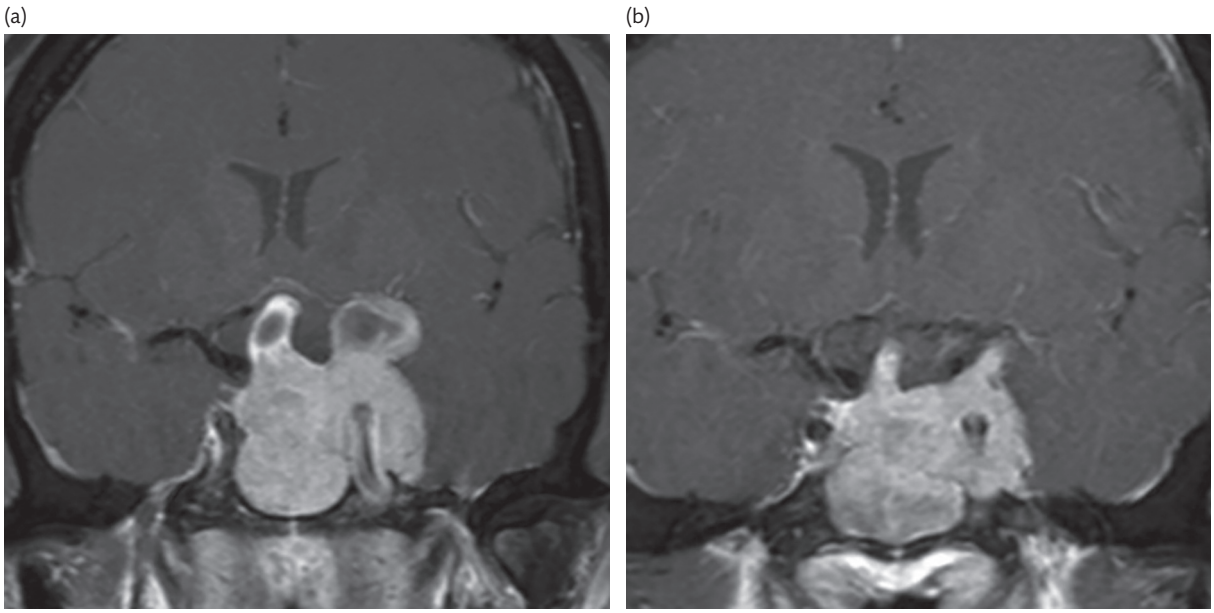


Figure 2.3.9.3 Coronal T1—weighted, gadolinium—enhanced MRI images of the sella of a 38-year-old man with a macroprolactinoma, measuring 3.5 cm in greatest diameter at presentation. Images were obtained before (a) and 3 weeks after (b) the institution of cabergoline therapy. He had presented with frequent headache of new onset, erectile dysfunction and bitemporal visual field deficits. At presentation, his serum prolactin level was 9487 ng/ml (412 475 pmol/L) [normal, up to 15 ng/ml (650 pmol/L)]. Evidence of central hypogonadism was present. He was treated with cabergoline, leading to resolution of visual field defects, improvement in prolactin levels, and decrease in tumour size on follow-up imaging. Testosterone replacement was ultimately advised since hypogonadism persisted despite prolactin normalization.

Cabergoline therapy is a selective D2 receptor agonist and is more effective than bromocriptine in inducing prolactin normalization and restoring eugonadism. In a randomized clinical trial of 459 women (most of whom harboured microadenomas), cabergoline administration normalized prolactin levels in 83% and restored eugonadism in 72% of cases, whereas bromocriptine normalized prolactin levels in in 59% and restored gonadal function in 52% of cases [55]. Cabergoline was also shown to have a more favourable side-effect profile than bromocriptine (as will be detailed subsequently).

Dopamine agonists also exert antiproliferative effects on prolactin-secreting tumours and cause involution of cellular organelles, accounting for their effects on tumour size (1). Bromocriptine therapy leads to a decrease in size of 76% of treated prolactinomas. Cabergoline has been reported to decrease tumour size in 96% of cases among naïve patients. In addition, cabergoline administration induces a decrease in tumour size in the majority of patients whose tumours are resistant to bromocriptine [56]. The extent of tumour shrinkage exceeds 50% in most patients treated with cabergoline over a variable time course (typically ranging from a few weeks to many months). An example of tumour response to cabergoline therapy in a patient with a macroprolactinoma is shown in Figure 2.3.9.3. Dopamine agonist therapy can ultimately lead to disappearance of visible tumour in some cases over a period of several years. Of note, improvement in mass effect can be very rapid in patients with macroprolactinomas treated with dopamine agonists, who often experience an improvement in visual field deficits within 1–3 days from treatment initiation [1]. In addition to gonadal dysfunction, other anterior pituitary hormone deficits may also resolve in treated patients as mass effect is relieved.

Bromocriptine has a relatively short half-life. Treatment is usually started at a dose of 0.625–1.25 mg daily and is advanced every 3–4 days to 2.5 mg daily. It is usually advisable to take the medication at bedtime with a small snack to minimize intolerance. The drug is titrated every 4 weeks towards normal prolactin levels, provided that it remains well-tolerated. Bromocriptine is generally given in divided doses (2–3 times daily) if a total dose over 2.5 mg daily is required.

Cabergoline has a longer half-life, owing to its slow release from binding sites and enterohepatic cycling [1]. Cabergoline is usually commenced at 0.25–0.5 mg weekly and is usually titrated every 4–6 weeks towards achieving normal prolactin levels, provided that the medication remains well-tolerated. It is sometimes advisable to advance cabergoline dose more quickly, including patients with giant macroprolactinomas. The majority of patients receiving cabergoline therapy require no more than 3.5 mg per week. However, resistance to all dopamine agonists is present in approximately 10% of patients, who may then need additional treatments (surgery, radiation therapy, or in the case of aggressive prolactinomas, temozolomide) [1, 57].

Dopamine agonists are generally well-tolerated. However, careful patient monitoring is advised to detect possible adverse effects (please see Box 2.3.9.4). Common adverse effects include nausea, dizziness, orthostasis [6, 7]. These adverse effects appear to be less frequent or severe with gradual dose titration (especially in the case of bromocriptine) and often improve over a period of several weeks. Some women may tolerate dopamine agonists better after intravaginal, rather than oral, administration although this is typically limited to short-term use only in patients seeking fertility. Less frequent adverse effects include nasal congestion, headache, constipation, nightmares, vivid dreams, digital vasospasm (akin to Raynaud's disease). As stated earlier, cabergoline is generally better tolerated than bromocriptine and can often be used to successfully treat patients who are unable to take bromocriptine because of intolerance [56].

Box 2.3.9.4 Adverse effects associated with dopamine agonist therapy

Common
Nausea
Dizziness, orthostasis

Less common
Headache
Nasal congestion
Constipation
Digital vasospasm
Vivid dreams, nightmares

Infrequent
Anxiety, depression, psychosis
Impulsivity–related manifestations

Rare
Fibrotic reactions
Cardiac valvulopathy*

* Reported in patients with Parkinson's disease on high doses of cabergoline or pergolide.

Tumour fibrosis has been reported in patients exposed to bromocriptine [58]. Very rarely, mediastinal and retroperitoneal fibrosis have been reported. In addition, patients with Parkinson's disease on very high cabergoline or pergolide doses (but not bromocriptine) were found to have a higher risk of cardiac valvulopathy [59, 60]. Such fibrotic reactions are thought to be a consequence of serotonin (5HT2B) receptor activation. Available data suggest that the risk of cardiac valvulopathy is related to the cumulative cabergoline dose. In contrast to the worrisome findings in patients with Parkinson's disease, patients with hyperprolactinemia who are generally treated with lower cabergoline doses (up to 2.0 mg weekly) have not been shown to be at risk for this complication [61]. Nevertheless, it is prudent to inform patients of this theoretical risk, particularly in patients who are taking high doses of dopamine agonists. Echocardiographic monitoring can be considered in such patients.

Rare adverse effects include the development of impulsivity (such as gambling, compulsive shopping, hypersexuality, and others) or psychosis, which can be serious but resolve after the offending drug is discontinued [62, 63]. Dopamine agonists should be avoided in patients with psychosis. However, some antipsychotics, such as aripiprazole, have both dopamine antagonist as well as partial dopamine agonist properties, and could be used in this setting in collaboration with the patient's psychiatrist. Patients with depression or anxiety can often be treated with dopamine agonists as long as the patient's psychiatrist has been consulted.

Patients who have shown good responses to dopamine agonist therapy can often be gradually downtitrated, as tumours may become more sensitive to the drug's effects as they get smaller. In some cases, the dopamine agonist can be withdrawn entirely after 2–3 years of therapy [64–66]. Good candidates for medication withdrawal are those who have maintained normal prolactin levels and have little or no visible remaining tumour on pituitary imaging. In one study of patients with normoprolactinaemia on cabergoline therapy, the risk of recurrence of hyperprolactinemia was higher in patients with visible tumour on MRI than those who had no residual tumour; in addition, patients with macroadenomas were at higher risk of recurrence of hyperprolactinemia than those with microadenomas [65]. If dopamine agonist therapy is withdrawn, patient education and careful monitoring are needed to detect possible recurrence.

Pituitary Surgery

Pituitary surgery is generally performed via the transsphenoidal approach [67]. Among patients with prolactinomas, pituitary surgery is generally second-line therapy (please see Box 2.3.9.5). Indications for pituitary surgery include lack of response to dopamine agonist therapy, including patients with persistent mass effect or tumour growth on medical therapy [67]. Cystic prolactinomas that are not responsive to dopamine agonists may be amenable to surgical resection [68]. Patients with resistance to medical therapy or intolerance to dopamine agonists and patients in whom these medications are contraindicated (including patients with psychosis) are also considered for surgery. In addition, patients with pituitary apoplexy (occurring either de novo or in response to dopamine agonist therapy) and those with CSF leak occurring in response to medical therapy are surgical candidates [67]. Patient preference should also be considered, as some patients with microprolactinomas may wish to avoid long-term medical therapy [69]. The role of surgery in the preconception setting is discussed in a separate section.

Surgical outcomes depend on tumour size and location [67]. In addition, surgical expertise is critical in order to achieve optimal outcomes. Published data typically represent the outcomes of experienced pituitary surgeons with a high procedural volume [70]. Surgery is very effective in relieving mass effect, particularly visual field deficits, by decompressing the optic apparatus. In aggregated case series, normoprolactinaemia was achieved in 75% of 2137 patients with microprolactinomas but only 34% out of 2226 patients with macroprolactinomas who were evaluated early postoperatively [1]. On long-term postoperative follow-up, normoprolactinaemia was reported in 60% of patients with microprolactinomas and 25% of patients with macroprolactinomas as a consequence of recurrence of hyperprolactinemia, particularly in patients with macroadenomas.

Pituitary function may improve in about 35% of patients with anterior hypopituitarism, whereas new anterior pituitary function

Box 2.3.9.5 Indications for pituitary surgery in patients with prolactinomas

Persistent mass effect (including visual field defects) despite dopamine agonist therapy
Tumour growth despite dopamine agonist therapy
Resistance to dopamine agonist therapy
Intolerance to dopamine agonist therapy
Contraindications to dopamine agonist therapy (including psychosis)
Pituitary apoplexy
Cerebrospinal fluid leak (rhinorrhoea)
Women with macroprolactinomas seeking fertility
Patients' preference (e.g. those with microprolactinomas who wish to avoid medical therapy)

Table 2.3.9.3 Safety of dopamine agonist therapy in the preconception setting*

	General population (%)	Patients who received bromocriptine [n (%)]	Patients who received cabergoline [n (%)]
Pregnancies		6239	789
Miscarriages	10–15%	620 (9.9%)	60 (7.6%)
Terminations	20%	75 (1.2%)	59 (7.5%)
Preterm deliveries**	12.7%	519 (12.5%)	67 (11.6%)
Congenital malformations***	3.0%	93 (1.8%)	21 (3.2%)

* Data were extracted from reference 8.
** Data were available on 4139 women who received bromocriptine and 547 women who received cabergoline before pregnancy.
*** Data were available on 5123 women who received bromocriptine and 664 women who received cabergoline before pregnancy.

deficits may occur in up to 20% of cases postoperatively [67]. Diabetes insipidus (DI) is often transient; permanent DI is uncommon in patients operated upon by experienced neurosurgeons. Other complications include cerebrospinal fluid rhinorrhoea, epistaxis, tumour haemorrhage, stroke, and meningitis [67]. Lower perioperative morbidity and mortality rates are achieved by expert neurosurgeons [71].

Radiation Therapy

Radiation therapy is reserved for patients with macroprolactinomas who cannot achieve adequate tumour control with dopamine agonist therapy and surgery [72]. Radiation therapy can be administered either as conventional radiation therapy or via a stereotactic approach (including Gamma Knife™, Cyber Knife ™, or proton beam radiation therapy). Stereotactic radiation therapy can be given in a single session ('radiosurgery') in patients whose tumours are distant from the optic apparatus. Radiation therapy is very effective for tumour control. However, control of hyperprolactinemia is slowly achieved over several years. All patients who have received radiation therapy are at lifelong risk of anterior hypopituitarism and require long-term follow-up [73]. Uncommonly, vision loss, ophthalmoplegia, stroke, temporal lobe necrosis, or secondary tumour formation may occur; whether stereotactic radiation therapy is safer than conventional radiation therapy has not been established.

Chemotherapy

Temozolomide is an alkylating agent that is absorbed orally and has been used off-label as therapy for locally aggressive pituitary adenomas or carcinomas (including prolactin-secreting tumours) [74–76]. It is generally used in patients whose tumours do not adequately respond to dopamine agonists, surgery, and radiation therapy. Using either tumour size or prolactin levels as treatment endpoints, temozolomide appears to benefit up to 73% of patients with locally aggressive adenomas. In some studies, tumour expression of O6-methylguanine-DNA methyltransferase (MGMT), a DNA repair enzyme, has been reported to show a negative association with response rates to temozolomide therapy [74–76]. Escape from the beneficial effects of temozolomide often occurs over time.

Hyperprolactinaemia and Pregnancy

Hyperprolactinemia and prolactinomas raise several concerns in women who are planning to conceive or are pregnant. Important considerations include subfertility in hyperprolactinemia, the safety of dopamine agonist therapy in the preconception setting and the effect of pregnancy on prolactinomas.

Hyperprolactinemia often impairs fertility in patients of both genders as a consequence of central hypogonadism [8, 9]. In addition to anovulation, the length of the luteal phase of the menstrual cycle is shortened, which can decrease the likelihood of implantation after oocyte fertilization. Dopamine agonist therapy which results in a normal prolactin level improves fertility in patients of both genders [77].

The safety of any medical therapy is of prime importance in the preconception setting as well as during pregnancy. Compiled, retrospectively collected data on many thousands of pregnancies have shown that bromocriptine therapy administered during the preconception period does not raise the risk of adverse maternal or fetal outcomes during a subsequent pregnancy (please see Table 2.3.9.3) [8, 9, 78]. Data on the safety of cabergoline are also reassuring. However, the published experience is substantially larger for

Table 2.3.9.4 Risk of prolactinoma growth during pregnancy*

Tumour size before gestation	Number of patients	Pituitary surgery and/or radiation therapy before gestation	Patients with symptomatic tumour growth during pregnancy [n (%)]
Macroprolactinoma	214	No	49 (22.9%)
Macroprolactinoma	148	Yes	7 (4.8%)
Microprolactinoma	658	No	18 (2.7%)

*Data were extracted from reference 8.

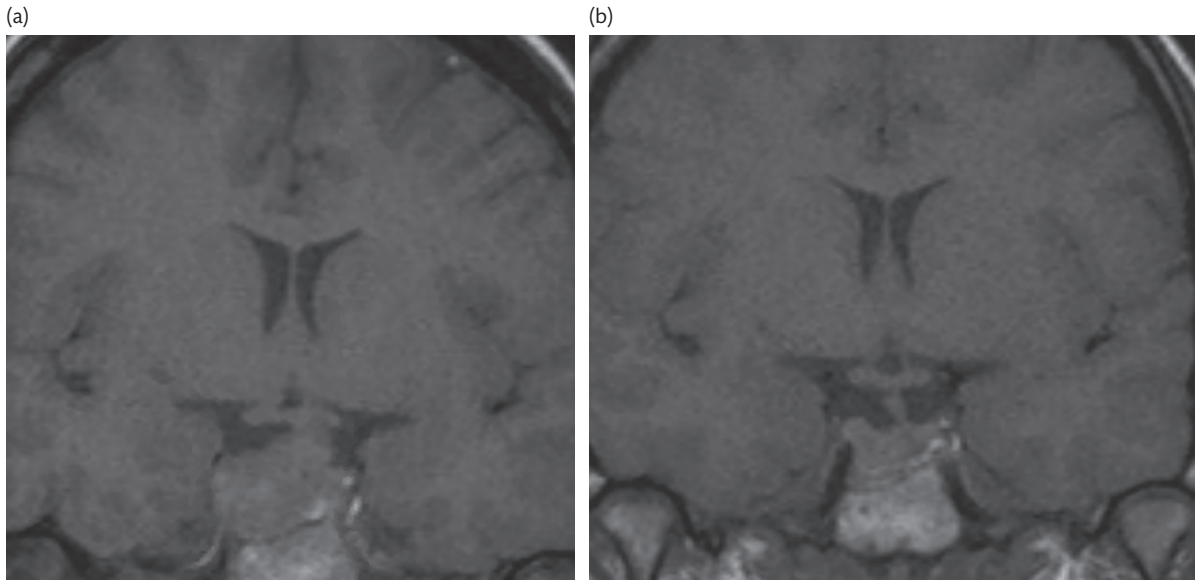


Figure 2.3.9.4 Coronal T1–weighted MRI images of the sella of a 23-year-old woman with a macroprolactinoma, measuring 2 cm in greatest diameter, which caused temporal visual field defects during the third trimester of pregnancy (a). Institution of bromocriptine therapy during pregnancy led to resolution of visual field defects as a result of tumour shrinkage (b).

bromocriptine than cabergoline [8, 9, 79]. As a corollary, bromocriptine is the agent of choice in the preconception setting or during gestation.

During pregnancy, oestrogen is copiously secreted by the placenta and may potentially promote prolactinoma growth. In aggregate, retrospective data suggest that the risk of microprolactinoma growth during pregnancy is low (please see Table 2.3.9.4) [8, 9, 80]. In contrast, the risk of tumour growth during gestation is higher in women with macroprolactinomas [8, 9]. A case of a patient with a macroprolactinoma that enlarged during pregnancy is shown in Figure 2.3.9.4. The risk of tumour growth during gestation is mitigated in women whose macroprolactinomas were resected before conception.

Hyperprolactinaemic women who are planning to conceive are offered bromocriptine therapy; cabergoline can be used in patients who are intolerant to bromocriptine. Pituitary surgery is considered in women with macroprolactinomas whose tumours do not substantially shrink in response to medical therapy in order to mitigate their risk of tumour growth during a subsequent pregnancy.

In patients with microprolactinomas, it is advisable to discontinue dopamine agonist therapy when pregnancy is established. During pregnancy, these patients can be followed clinically. Monitoring serum prolactin levels is of no diagnostic value during gestation. In the unlikely event that mass effect develops (including visual field deficits), limited views of the sella are obtained via non-contrast MRI and dopamine agonist therapy (preferably bromocriptine) is reinstituted. Limited data from case series suggest that bromocriptine and cabergoline are often effective in relieving mass effect during pregnancy and are not associated with increased risk [8, 9].

Medical management is individualized in patients with macroprolactinomas, who are at higher risk for tumour growth during gestation. In most cases, dopamine agonist therapy is discontinued when pregnancy is established. In some cases, including tumours close to the optic apparatus, dopamine agonists are continued during gestation to minimize the risk of tumour growth during that period. All pregnant patients with macroadenomas require careful follow-up, including frequent visual field-testing during gestation. If evidence of mass effect develops, then dopamine agonist (preferably bromocriptine) therapy is reinstituted and can be effective in relieving mass effect [8, 9, 81]. Alternatively, pituitary surgery can be performed to restore vision in patients who are unresponsive to dopamine agonist therapy. However, any surgical procedure performed during gestation is associated with a risk of miscarriage [8, 9].

Asymptomatic women with prolactinomas should be encouraged to nurse. Dopamine agonists are avoided in asymptomatic women who are nursing, since they suppress milk production and, in cases of partial suppression of lactogenesis, may expose the infant to dopamine agonists which are present in breast milk. There is no evidence that nursing is associated with an increased risk of tumour growth postpartum [8, 9].

REFERENCES

1. Gillam MP, Molitch ME, Lombardi G, Colao A. Advances in the treatment of prolactinomas. *Endocr Rev*, 2006; **27**: 485–534.
2. Ben-Jonathan N, Hnasko R. Dopamine as a prolactin (PRL) inhibitor. *Endocr Rev*, 2001; **22**: 724–63.
3. Horseman ND, Gregerson KA. Prolactin actions. *J Mol Endocrinol*, 2014; **52**: R95–106.
4. Bao L, Tessier C, Prigent-Tessier A, *et al*. Decidual prolactin silences the expression of genes detrimental to pregnancy. *Endocrinology*, 2007; **148**: 2326–34.
5. Christensen HR, Murawsky MK, Horseman ND, Willson TA, Gregerson KA. Completely humanizing prolactin rescues infertility in prolactin knockout mice and leads to human prolactin expression in extrapituitary mouse tissues. *Endocrinology*, 2013; **154**: 4777–89.
6. Klibanski A. Clinical practice. Prolactinomas. *N Engl J Med*, 2010; **362**: 1219–26.

7. Melmed S, Casanueva FF, Hoffman AR, *et al*. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
8. Molitch ME. Prolactinoma in pregnancy. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 885–96.
9. Molitch ME. Endocrinology in pregnancy: management of the pregnant patient with a prolactinoma. *Eur J Endocrinol*, 2015; **172**: R205–13.
10. Biller BM, Luciano A, Crosignani PG, *et al*. Guidelines for the diagnosis and treatment of hyperprolactinemia. *J Reprod Med*, 1999; **44**(12 Suppl): 1075–84.
11. Molitch ME. Drugs and prolactin. *Pituitary*, 2008; **11**: 209–18.
12. Chen JX, Su YA, Bian QT, *et al*. Adjunctive aripiprazole in the treatment of risperidone-induced hyperprolactinemia: a randomized, double-blind, placebo-controlled, dose-response study. *Psychoneuroendocrinology*, 2015; **58**: 130–40.
13. Lee BH, Kim YK, Park SH. Using aripiprazole to resolve antipsychotic-induced symptomatic hyperprolactinemia: a pilot study. *Prog Neuropsychopharmacol Biol Psychiatry*, 2006; **30**: 714–17.
14. Yasui-Furukori N, Furukori H, Sugawara N, Fujii A, Kaneko S. Dose-dependent effects of adjunctive treatment with aripiprazole on hyperprolactinemia induced by risperidone in female patients with schizophrenia. *J Clin Psychopharmacol*, 2010; **30**: 596–9.
15. Newey PJ, Gorvin CM, Cleland SJ, *et al*. Mutant prolactin receptor and familial hyperprolactinemia. *N Engl J Med*, 2013; **369**: 2012–20.
16. Ciccarelli A, Daly AF, Beckers A. The epidemiology of prolactinomas. *Pituitary*, 2005; **8**: 3–6.
17. Agustsson TT, Baldvinsdottir T, Jonasson JG, *et al*. The epidemiology of pituitary adenomas in Iceland, 1955–2012: a nationwide population-based study. *Eur J Endocrinol*, 2015; **173**: 655–64.
18. Daly AF, Rixhon M, Adam C, Dempegioti A, Tichomirowa MA, Beckers A. High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium. *J Clin Endocrinol Metab*, 2006; **91**: 4769–75.
19. Fernandez A, Karavitaki N, Wass JA. Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). *Clin Endocrinol (Oxf)*, 2010; **72**: 377–82.
20. Raappana A, Koivukangas J, Ebeling T, Pirila T. Incidence of pituitary adenomas in Northern Finland in 1992–2007. *J Clin Endocrinol Metab*, 2010; **95**: 4268–75.
21. Tjornstrand A, Gunnarsson K, Evert M, *et al*. The incidence rate of pituitary adenomas in western Sweden for the period 2001–2011. *Eur J Endocrinol*, 2014; **171**: 519–26.
22. Molitch ME. Nonfunctioning pituitary tumors and pituitary incidentalomas. *Endocrinol Metab Clin North Am*, 2008; **37**: 151–71, xi.
23. Buurman H, Saeger W. Subclinical adenomas in postmortem pituitaries: classification and correlations to clinical data. *Eur J Endocrinol*, 2006; **154**: 753–8.
24. Burgess JR, Shepherd JJ, Parameswaran V, Hoffman L, Greenaway TM. Prolactinomas in a large kindred with multiple endocrine neoplasia type 1: clinical features and inheritance pattern. *J Clin Endocrinol Metab*, 1996; **81**: 1841–5.
25. de Laat JM, Dekkers OM, Pieterman CR, *et al*. Long-term natural course of pituitary tumors in patients with MEN1: results from the DutchMEN1 Study Group (DMSG). *J Clin Endocrinol Metab*, 2015; **100**: 3288–96.
26. Daly AF, Tichomirowa MA, Petrossians P, *et al*. Clinical characteristics and therapeutic responses in patients with germ-line AIP mutations and pituitary adenomas: an international collaborative study. *J Clin Endocrinol Metab*, 2010; **95**: E373–83.
27. Vierimaa O, Georgitsi M, Lehtonen R, *et al*. Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science*, 2006; **312**: 1228–30.
28. Schlechte J, Dolan K, Sherman B, Chapler F, Luciano A. The natural history of untreated hyperprolactinemia: a prospective analysis. *J Clin Endocrinol Metab*, 1989; **68**: 412–18.
29. March CM, Kletzky OA, Davajan V, *et al*. Longitudinal evaluation of patients with untreated prolactin-secreting pituitary adenomas. *Am J Obstet Gynecol*, 1981; **139**: 835–44.
30. Sisam DA, Sheehan JP, Sheeler LR. The natural history of untreated microprolactinomas. *Fertil Steril*, 1987; **48**: 67–71.
31. Weiss MH, Teal J, Gott P, *et al*. Natural history of microprolactinomas: six-year follow-up. *Neurosurgery*, 1983; **12**: 180–3.
32. Kaltsas GA, Nomikos P, Kontogeorgos G, Buchfelder M, Grossman AB. Clinical review: diagnosis and management of pituitary carcinomas. *J Clin Endocrinol Metab*, 2005; **90**: 3089–99.
33. Pontiroli AE, Falsetti L. Development of pituitary adenoma in women with hyperprolactinaemia: clinical, endocrine, and radiological characteristics. *Br Med J (Clin Res Ed)*, 1984; **288**: 515–18.
34. Grattan DR, Jasoni CL, Liu X, Anderson GM, Herbison AE. Prolactin regulation of gonadotropin-releasing hormone neurons to suppress luteinizing hormone secretion in mice. *Endocrinology*, 2007; **148**: 4344–51.
35. Klibanski A, Biller BM, Rosenthal DI, Schoenfeld DA, Saxe V. Effects of prolactin and estrogen deficiency in amenorrheic bone loss. *J Clin Endocrinol Metab*, 1988; **67**: 124–30.
36. Klibanski A, Neer RM, Beitins IZ, Ridgway EC, Zervas NT, McArthur JW. Decreased bone density in hyperprolactinemic women. *N Engl J Med*, 1980; **303**: 1511–14.
37. Steele CA, MacFarlane IA, Blair J, *et al*. Pituitary adenomas in childhood, adolescence and young adulthood: presentation, management, endocrine and metabolic outcomes. *Eur J Endocrinol*, 2010; **163**: 515–22.
38. Di Iorgi N, Napoli F, Allegri AE, *et al*. Diabetes insipidus--diagnosis and management. *Horm Res Paediatr*, 2012; **77**: 69–84.
39. Corsello SM, Ubertini G, Altomare M, *et al*. Giant prolactinomas in men: efficacy of cabergoline treatment. *Clin Endocrinol (Oxf)*, 2003; **58**: 662–70.
40. Shimon I, Sosa E, Mendoza V, *et al*. Giant prolactinomas larger than 60 mm in size: a cohort of massive and aggressive prolactin-secreting pituitary adenomas. *Pituitary*, 2016; **19**: 429–36.
41. Veldhuis JD, Johnson ML. Operating characteristics of the hypothalamo-pituitary-gonadal axis in men: circadian, ultradian, and pulsatile release of prolactin and its temporal coupling with luteinizing hormone. *J Clin Endocrinol Metab*, 1988; **67**: 116–23.
42. St-Jean E, Blain F, Comtois R. High prolactin levels may be missed by immunoradiometric assay in patients with macroprolactinomas. *Clin Endocrinol (Oxf)*, 1996; **44**: 305–9.
43. Sturgeon CM, Viljoen A. Analytical error and interference in immunoassay: minimizing risk. *Ann Clin Biochem*, 2011; **48**(Pt 5): 418–32.
44. Smith TP, Kavanagh L, Healy ML, McKenna TJ. Technology insight: measuring prolactin in clinical samples. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 279–89.

45. Barkan AL, Chandler WF. Giant pituitary prolactinoma with falsely low serum prolactin: the pitfall of the 'high-dose hook effect': case report. *Neurosurgery*, 1998; **42**: 913–15; discussion 5–6.
46. Gulbahar O, Konca Degertekin C, Akturk M, *et al*. A case with immunoassay interferences in the measurement of multiple hormones. *J Clin Endocrinol Metab*, 2015; **100**: 2147–53.
47. Samson SL, Hamrahian AH, Ezzat S, AACE Neuroendocrine and Pituitary Scientific Committee, American College of Endocrinology (ACE). American Association of Clinical Endocrinologists, American College of Endocrinology disease state clinical review: clinical relevance of macroprolactin in the absence or presence of true hyperprolactinemia. *Endocr Pract*, 2015; **21**: 1427–35.
48. Suliman AM, Smith TP, Gibney J, McKenna TJ. Frequent misdiagnosis and mismanagement of hyperprolactinemic patients before the introduction of macroprolactin screening: application of a new strict laboratory definition of macroprolactinemia. *Clin Chem*, 2003; **49**: 1504–9.
49. Corenblum B. Asymptomatic hyperprolactinemia resulting from macroprolactinemia. *Fertil Steril*, 1990; **53**: 165–7.
50. Strachan MW, Teoh WL, Don-Wauchope AC, Seth J, Stoddart M, Beckett GJ. Clinical and radiological features of patients with macroprolactinaemia. *Clin Endocrinol (Oxf)*, 2003; **59**: 339–46.
51. Tritos NA, Guay AT, Malarkey WB. Asymptomatic 'big' hyperprolactinemia in two men with pituitary adenomas. *Eur J Endocrinol*, 1998; **138**: 82–5.
52. Wijeratne NG, Doery JC, Lu ZX. Positive and negative interference in immunoassays following biotin ingestion: a pharmacokinetic study. *Pathology*, 2012; **44**: 674–5.
53. Li D, Radulescu A, Shrestha RT, *et al*. Association of biotin ingestion with performance of hormone and nonhormone assays in healthy adults. *JAMA*, 2017; **318**: 1150–60.
54. Bayrak A, Saadat P, Mor E, Chong L, Paulson RJ, Sokol RZ. Pituitary imaging is indicated for the evaluation of hyperprolactinemia. *Fertil Steril*, 2005; **84**: 181–5.
55. Webster J, Piscitelli G, Polli A, Ferrari CI, Ismail I, Scanlon MF. A comparison of cabergoline and bromocriptine in the treatment of hyperprolactinemic amenorrhea. Cabergoline Comparative Study Group. *N Engl J Med*, 1994; **331**: 904–9.
56. dos Santos Nunes V, El Dib R, Boguszewski CL, Nogueira CR. Cabergoline versus bromocriptine in the treatment of hyperprolactinemia: a systematic review of randomized controlled trials and meta-analysis. *Pituitary*, 2011; **14**: 259–65.
57. Vasilev V, Daly AF, Vroonen L, Zacharieva S, Beckers A. Resistant prolactinomas. *J Endocrinol Invest*, 2011; **34**: 312–16.
58. Landolt AM, Osterwalder V. Perivascular fibrosis in prolactinomas: is it increased by bromocriptine? *J Clin Endocrinol Metab*, 1984; **58**: 1179–83.
59. Schade R, Andersohn F, Suissa S, Haverkamp W, Garbe E. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med*, 2007; **356**: 29–38.
60. Zanettini R, Antonini A, Gatto G, Gentile R, Tesei S, Pezzoli G. Valvular heart disease and the use of dopamine agonists for Parkinson's disease. *N Engl J Med*, 2007; **356**: 39–46.
61. Valassi E, Klibanski A, Biller BM. Clinical review#: potential cardiac valve effects of dopamine agonists in hyperprolactinemia. *J Clin Endocrinol Metab*, 2010; **95**: 1025–33.
62. Bancos I, Nannenga MR, Bostwick JM, Silber MH, Erickson D, Nippoldt TB. Impulse control disorders in patients with dopamine agonist-treated prolactinomas and nonfunctioning pituitary adenomas: a case-control study. *Clin Endocrinol (Oxf)*, 2014; **80**: 863–8.
63. Barake M, Evins AE, Stoeckel L, *et al*. Investigation of impulsivity in patients on dopamine agonist therapy for hyperprolactinemia: a pilot study. *Pituitary*, 2014; **17**: 150–6.
64. Passos VQ, Souza JJ, Musolino NR, Bronstein MD. Long-term follow-up of prolactinomas: normoprolactinemia after bromocriptine withdrawal. *J Clin Endocrinol Metab*, 2002; **87**: 3578–82.
65. Colao A, Di Sarno A, Cappabianca P, Di Somma C, Pivonello R, Lombardi G. Withdrawal of long-term cabergoline therapy for tumoral and nontumoral hyperprolactinemia. *N Engl J Med*, 2003; **349**: 2023–33.
66. Colao A, Di Sarno A, Guerra E, *et al*. Predictors of remission of hyperprolactinaemia after long-term withdrawal of cabergoline therapy. *Clin Endocrinol (Oxf)*, 2007; **67**: 426–33.
67. Swearingen B. Update on pituitary surgery. *J Clin Endocrinol Metab*, 2012; **97**: 1073–81.
68. Faje A, Chunharojrith P, Nency J, Biller BM, Swearingen B, Klibanski A. Dopamine agonists can reduce cystic prolactinomas. *J Clin Endocrinol Metab*, 2016; **101**: 3709–15.
69. Tampourlou M, Trifanescu R, Paluzzi A, Ahmed SK, Karavitaki N. Therapy of endocrine disease: surgery in microprolactinomas: effectiveness and risks based on contemporary literature. *Eur J Endocrinol*, 2016; **175**: R89–96.
70. Casanueva FF, Barkan AL, Buchfelder M, *et al*. Criteria for the definition of Pituitary Tumor Centers of Excellence (PTCOE): a Pituitary Society statement. *Pituitary*, 2017; **20**: 489–98.
71. Barker FG, 2nd, Klibanski A, Swearingen B. Transsphenoidal surgery for pituitary tumors in the United States, 1996–2000: mortality, morbidity, and the effects of hospital and surgeon volume. *J Clin Endocrinol Metab*, 2003; **88**: 4709–19.
72. Loeffler JS, Shih HA. Radiation therapy in the management of pituitary adenomas. *J Clin Endocrinol Metab*, 2011; **96**: 1992–2003.
73. Agha A, Sherlock M, Brennan S, *et al*. Hypothalamic-pituitary dysfunction after irradiation of nonpituitary brain tumors in adults. *J Clin Endocrinol Metab*, 2005; **90**: 6355–60.
74. Ghazi AA, Rotondo F, Kovacs K, *et al*. Treatment of invasive silent somatotroph pituitary adenoma with temozolomide. Report of a case and review of the literature. *Endocr Pathol*, 2015; **26**: 135–9.
75. Liu JK, Patel J, Eloy JA. The role of temozolomide in the treatment of aggressive pituitary tumors. *J Clin Neurosci*, 2015; **22**: 923–9.
76. Whitelaw BC, Dworakowska D, Thomas NW, *et al*. Temozolomide in the management of dopamine agonist-resistant prolactinomas. *Clin Endocrinol (Oxf)*, 2012; **76**: 877–86.
77. Eftekhari N, Mohammadalizadeh S. Pregnancy rate following bromocriptine treatment in infertile women with galactorrhea. *Gynecol Endocrinol*, 2009; **25**: 122–4.
78. Krupp P, Monka C. Bromocriptine in pregnancy: safety aspects. *Klin Wochenschr*, 1987; **65**: 823–7.
79. Glezer A, Bronstein MD. Prolactinomas, cabergoline, and pregnancy. *Endocrine*, 2014; **47**: 64–9.
80. Divers WA, Jr., Yen SS. Prolactin-producing microadenomas in pregnancy. *Obstet Gynecol*, 1983; **62**: 425–9.
81. Liu C, Tyrrell JB. Successful treatment of a large macroprolactinoma with cabergoline during pregnancy. *Pituitary*, 2001; **4**: 179–85.

2.3.10 Acromegaly

John A.H. Wass, Peter J. Trainer, and Márta Korbonits

Definition

Acromegaly almost always is the result of an anterior pituitary tumour and is a consequence of growth hormone (GH) and insulin-like growth factor-I (IGF-1) excess. It is characterized by enlargement of the hands and feet (Greek: *akron*, extremities; *megas*, great). Gigantism is the juvenile (prior to epiphyseal fusion) counterpart of acromegaly and results in excessive linear growth.

History

Giants are the material of legend and fairy tale. The earliest plausible description of a giant is Goliath (alleged to have been 290 cm tall), whose death, while fighting David, has been speculated, to have been a result of the disadvantage caused by a temporal field defect as a result of his pituitary tumour. The skeleton of Irish giant Charles Byrne (234 cm) is exhibited in the Royal College of Surgeons of England and has twice provided insight into our understanding of the pathophysiology of acromegalic gigantism. Harvey Cushing in 1910 had the skull opened to confirm an enlarged pituitary fossa, and more recently, a tooth was extracted to allow isolation of DNA which resulted in a genetic explanation for Byrne's disease as well as for the folklore of the Irish Giants [1]. The tallest man recorded was Robert Wadlow (272 cm), an American who died in 1940 at the age of 22 years, most likely suffered from X-linked acrogigantism (see next). Comprehensive historical and illustrated descriptions of acromegaly and gigantism are available [2].

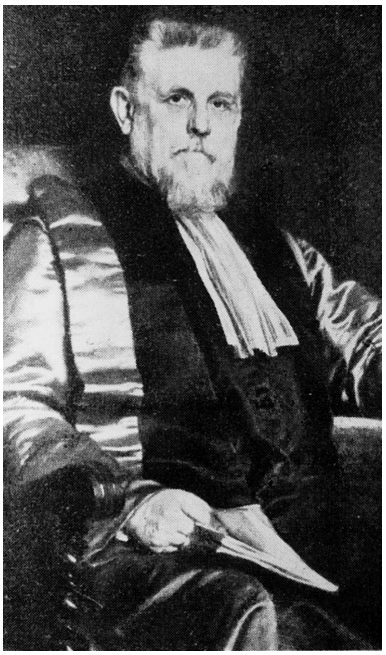


Figure 2.3.10.1 Pierre Marie, the describer of acromegaly.

Although there had been excellent previous descriptions, for example, the first description by a Dutch physician Johannes Wier in 1567 of a giant female patient, it was Pierre Marie (1886) who gave the name to the condition (Figure 2.3.10.1). He did not, at the time, appreciate that the pituitary was the cause of the problem. The recognition of the link between an enlarged pituitary and acromegaly is attributed to Minkowski (1887). Familial acromegaly was depicted by the illustrator John Kay of the Irish giant Charles Byrne with his twin giant cousins, the Knipe twins at the end of the 18th century [1]. Its clinical description was a century later [3]. The first gene identified causing acromegaly was *GNAS*, both as a change in the pituitary only [4] or as an embryonic mutation as part of McCune–Albright syndrome [5]. The first MEN1 syndrome patient had gigantism [6] with the gene responsible identified in 1997 [7]. Most young-onset acromegaly families have *AIP* (aryl hydrocarbon receptor interacting protein gene) mutations [8], while the tallest giants have *GPR101* duplication as a germline [9] or mosaic [10] condition. The first attempt at surgical treatment was by Caton and Paul in Liverpool (1893) who attempted to relieve the headache by surgical removal of part of the skull vault. Transsphenoidal operation on a patient with acromegaly was first reported von Eiselsberg in Vienna 1908. Harvey Cushing was convinced that acromegaly was a form of hyperpituitarism and operated on numerous cases via the transsphenoidal route. Béclère (1909) was the first to use pituitary radiation therapy. The modern era of management was marked by the development of radioimmunoassays for GH in the 1960s which provided the ability to measure disease activity. Landmarks in the medical management of acromegaly are represented by the introduction of dopamine agonists in 1972 [11], somatostatin analogues in 1985 [12], and the GH receptor antagonist pegvisomant in 2000 [13].

Epidemiology

The prevalence ranges from 28 to 137 per million with an incidence between 2 and 11 cases per million [14]. Acromegaly occurs in all races with an approximately equal sex incidence. The average age at diagnosis is 44 years, but patients with acromegaly can present at any age. The mean time between development of symptoms to diagnosis is 5 years according to the latest data. Young patients typically have larger, more aggressive tumours and grossly elevated serum GH levels (see next) while a more subtle phenotype with smaller and less aggressive tumours is typically encountered in older patients.

Aetiology

Over 99% of patients with acromegaly have a GH-secreting adenoma, which, in up to a third of cases, may cosecrete prolactin. Very rarely GH and thyroid-stimulating hormone (TSH) are cosecreted, causing acromegaly and thyrotoxicosis with detectable TSH (**Box 2.3.10.1**).

Less than 1% of patients with acromegaly have a GH-releasing hormone (GHRH)-secreting tumour, as a result of a neuroendocrine tumour of either the pancreas (often associated with MEN1 syndrome) or the lung, with extremely rare cases from a phaeochromocytoma. Ectopic GHRH results in a global enlargement of the pituitary as a consequence of somatotroph hyperplasia and presents a diagnostic challenge. Hypothalamic GHRH-producing tumours have been also described, such as a gangliocytoma. Pituitary carcinomas represent 0.1% of pituitary tumours, with GH-secreting carcinomas being the third most common subtype, after prolactinomas and corticotrophinomas [15]. Most often the metastases are found in the cerebrospinal axis, but they have been described outside the central nervous system.

Acromegaly can occur as part of a genetic condition either as part of a syndromic disease (MEN1, MEN4, Carney complex, pituitary adenoma associated with paragangliomas or McCune–Albright syndrome) or as an isolated condition (familial isolated pituitary adenoma, FIPA), as discussed next.

Pathology

Somatotroph cells in the normal pituitary are usually located in the lateral region of the pituitary, and at least in rodents this pool of somatotrophs expands significantly in puberty, possibly explaining the characteristic cavernous sinus invasion of these adenomas. Somatotroph adenomas can either be densely granulated with strong GH staining and diffuse distribution of cytokeratin or sparsely granulated with characteristic dot-like cytokeratin staining [16]. The sparsely granulated somatotroph adenomas occur more often in young patients, have a greater tendency to tumour invasiveness and respond poorly to somatostatin analogue therapy. Expression of somatostatin receptors can be assessed by immunohistochemistry, with subtype 2 and 5 being the most prominent, and their correlation with response to first- and second-generation somatostatin analogues has been studied [17]. Somatotroph hyperplasia can be seen in 25% of patients with X-linked acrogigantism, in 70% of patients with Carney complex, in patients with McCune–Albright syndrome, if the pituitary area is affected, with or without adenoma development and in patients with GHRH-secreting tumours. Silent somatotroph tumours (clinically non-functioning pituitary adenomas with GH staining on histology) have been observed in 7–8% of somatotroph adenomas [18].

Box 2.3.10.1 Lesions associated with excessive secretion of GH and IGF-1

- Pituitary
  - Growth hormone-secreting adenoma
  - Growth hormone and prolactin mixed adenoma
  - Growth hormone and TSH mixed adenoma
  - Plurihormonal adenoma
  - Growth hormone-secreting carcinoma
  - Somatotroph hyperplasia
- GHRH-producing neuroendocrine tumour (pancreas and lung)
- GHRH-producing hypothalamic lesions such as gangliocytoma
- GH-secreting ectopic tumours (only three cases described)

Molecular Endocrinology of GH-Secreting Pituitary Adenomas

The molecular pathogenesis of sporadic GH-secreting pituitary tumours is best considered by discussing changes which activate factors leading to increased tumour formation (e.g. oncogenes) or alterations which inactivate cell proliferation controlling genes (e.g. tumour suppressor genes). The GHRH—GHRH receptor—stimulatory G-protein α-subunit—cAMP—protein kinase A—CREB—Pit-1 pathway is key for somatotroph cell function. Activating genetic alterations include the stimulatory guanine nucleotide-binding protein (G-protein) α-subunit gene (*GNAS*), the orphan G-protein associated 7 transmembrane receptor GPR101, cyclin D (*CCDN1*), fibroblast growth factor receptor 4 (*FGFR4*), and pituitary tumour transforming gene (*PTTG*).

The stimulatory G protein ($G_s\alpha$) is involved in the activation of adenylate cyclase which mediates the regulatory actions of GHRH to stimulate GH synthesis and secretion. Missense mutations of *GNAS* at codons 201 and 227 (termed 'gsp' mutations) result in inhibition of the intrinsic GTPase activity of the α-subunit of the G protein, while its adenyl cyclase activating capacity is intact resulting in high intracellular levels of cyclic adenosine monophosphate (AMP). The downstream signalling pathway includes increased protein kinase A and cyclic AMP-response element binding protein (CREB) activity, increased binding, and activation of the *POU1F1* promoter resulting in activate GH synthesis and release. This results in autonomous GH secretion (**Figure 2.3.10.2**). Somatic *GNAS* mutations have been demonstrated in 40% of human GH-secreting pituitary adenomas and are the most commonly described genetic defect. If the gsp mutation occurs in embryonic stage and is found in a mosaic form in various organs contributing to activation of various G_s-coupled receptors, the patient develops McCune–Albright syndrome (see next). Patients with *GPR101* mutation develop infant onset somatotroph or somatomammotroph hyperplasia or tumours. The role of GPR101 is unclear in the somatotroph adenoma tumorigenesis. This cAMP-coupled receptor is normally expressed in the hypothalamus and could be upstream of GHRH. Embryonic overstimulation of GHRH and therefore GH/PRL-secreting cells may lead to tumorigenesis, a phenomenon

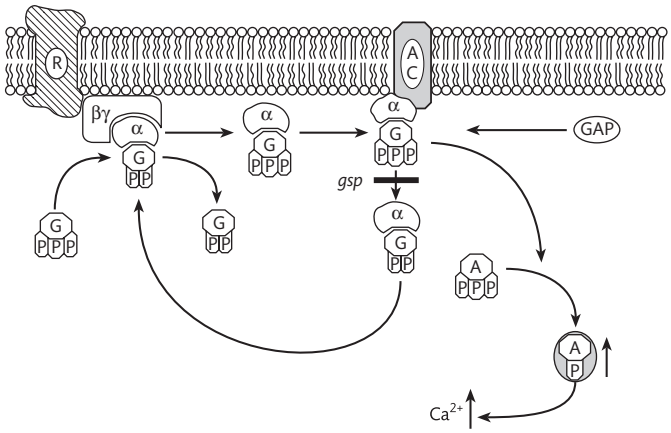


Figure 2.3.10.2 The G-protein abnormality seen in the pituitary of 40% of Caucasian patients with acromegaly.

observed in animal models. Increased *PTTG1* mRNA expression has been demonstrated in somatotroph adenomas and correlates with tumour size. FGFR4 and cyclin D overexpression have been described in pituitary tumours; however, this is not specific for somatotroph adenomas.

Tumour suppressor genes that may be involved in pituitary tumour pathogenesis include the retinoblastoma gene, cyclin-dependant kinase inhibitors, such as p27 (*CDKN1B*) and p16 (*CDKN2A*) as well as growth arrest and DNA damage-inducible protein (*GADD45γ*) and maternal imprinting gene 3 (*MEG3*). Some of these proteins are lost in pituitary tumours due to epigenetic mechanisms such as hypermethylation. p27 expression is reduced in all types of pituitary adenomas including somatotrophs. GADD45γ is a pro-apoptotic factor which is lost in GH-secreting adenomas. *MEG3* is an imprinted gene encoding a non-coding RNA that suppresses tumour cell growth; it is lost in non-functioning pituitary adenomas but not in somatotroph tumours. Germline *AIP* mutations have been described in patients with sporadic or familial isolated pituitary adenomas and *in vitro* studies confirm that loss of function of this protein is in the pathogenesis of these adenomas. AIP may play a role in the somatostatin induced inhibitory pathway regulating the tumor suppressor gene *ZAC1* or the inhibitory G protein $G_i\alpha 2$ [19].

Genetic Alterations Associated with Acromegaly

McCune–Albright Syndrome

This syndrome is characterized by polyostotic fibrous dysplasia, hyperpigmented cutaneous patches, and endocrinological abnormalities including precocious puberty, thyrotoxicosis, acromegaly or gigantism, and Cushing's syndrome. The extent of the mosaic activating mutations determine the severity of the disease in the various organs where $G_s\alpha$ plays an important role. The pituitary and the surrounding tissues are affected in 20–30% of the cases. A third of patients present in childhood and may develop gigantism, especially if there is hypogonadism due to excess prolactin, although the combination of early puberty and GH excess may result in a normal final height. The pituitary shows hyperplasia in just under half of the cases with micro or macroadenomas in the rest [20]. These patients present a challenge for treatment as the surrounding fibrous dysplasia in the bones renders difficult surgical access. Combination treatment with dopamine, somatostatin agonists, and GH-receptor antagonist can provide biochemical control.

Multiple Endocrine Neoplasia Type 1 and Type 4

MEN1 and MEN4 are an autosomal dominant disorders which are described elsewhere (see Section 4, 'Parathyroid, Calcium and Bone Metabolism Disorders') (see pg. 632). Following prolactinomas and clinically non-functioning adenomas, acromegaly is third commonest pituitary tumour type to occur in MEN1 syndrome, although plurihormonal adenomas are also characteristic of MEN1. In patients with MEN1-related GH excess, especially in childhood, GHRH-secreting pancreas neuroendocrine tumours should be considered. MEN4 has been described in a few patients with an MEN1-like phenotype and germline mutations in cell cycle inhibitor genes, such as p27 (*CDKN1B*), p15 (*CDKN2B*), p18 (*CDKN2C*), and p21 (*CDKN1A*). The most common tumour type is somatotroph adenomas, including a case with childhood-onset disease.

Carney Complex

This is an autosomal dominant condition caused by a loss-of-function mutation in the protein kinase A regulatory subunit gene (*PRKAR1A*) in over 70% of the cases, with others mapping to 2p16. A single patient with duplication of the activating subunit of protein kinase A (*PRKACB*) has also been described. It is characterized by spotty cutaneous pigmentation, cardiac and other myxomas, and endocrine overactivity, particularly Cushing's syndrome due to nodular adrenal cortical hyperplasia. Abnormal GH dynamics can be detected in two third of cases and with 10% of the patients developing true adenomas [21].

Familial Isolated Pituitary Adenomas

Patients with FIPA usually have no other endocrine abnormality. The genetic background and the clinical characteristics are heterogeneous.

X-linked Acrogigantism (XLAG)

Patients with germline or mosaic duplication in the *GPR101* gene, which is located on the X chromosome, develop infant onset acromegaly diagnosed <5 years of age with symptoms starting < 2 years [9]. Patients have pituitary hyperplasia or adenoma, secreting usually both GH and prolactin. The majority of the patients are females with a *de novo* germline mutations, while males are either have mosaic mutation or inherited the mutation from their mother. Unless treated, these patients develop extreme tall stature. As the whole gland is involved, treatment is challenging with either total hypophysectomy, combination medical treatment or radiotherapy.

Aryl Hydrocarbon Receptor Interacting Protein

Loss-of-function heterozygous germline mutations in the *AIP* gene cause young-onset acromegaly with incomplete (20–30%) penetrance [22]. Most patient have acromegaly manifesting in the second decade, with prolactinomas occurring in 10% of the cases and other pituitary tumour types rarely. Patients with *AIP*

mutations usually have early-onset, with 50% reporting onset of symptoms in the second decade [23]. The invariably sparsely granulated adenomas may show invasive growth and respond poorly to somatostatin analogue therapy. Genetic cascade screening can identify carrier family members leading to earlier diagnosis and better chance of a cure.

FIPA Families Currently Without a Known Mutation

The majority (89%) of the FIPA families do not harbour known disease-causing mutations. Patients in these families have an age of onset similar to sporadic pituitary adenoma patients and the associated tumour types are wider than in AIP or XLAG cases. About half of the families are homogenous (i.e. all affected family members have the same pituitary adenoma type) with the other half presenting with heterogeneous type of adenomas, acromegaly and prolactinoma being the most common tumour types [23]. While penetrance is 100% in XLAG and 20–30% in AIP kindreds, it is somewhat lower in AIP-negative FIPA families.

Symptoms of Acromegaly

Gross, long-standing acromegaly is easily recognized, the challenge is recognizing the more subtle phenotype and shortening the interval between the onset of disease and diagnosis. The most noticeable feature is usually a change in facial appearance. Vague symptoms such as fatigue may predominate, while increased sweating and sebaceous activity with acne can be problems. There is enlargement of the supraorbital ridges, prognathism, and macroglossia (see Figures 2.3.10.3 and 4); interdental separation occurs. This, together with the obvious changes in the hands and feet, often makes the diagnosis easy (see Box 2.3.10.2). Headache is a typical symptom, more commonly than other types of pituitary

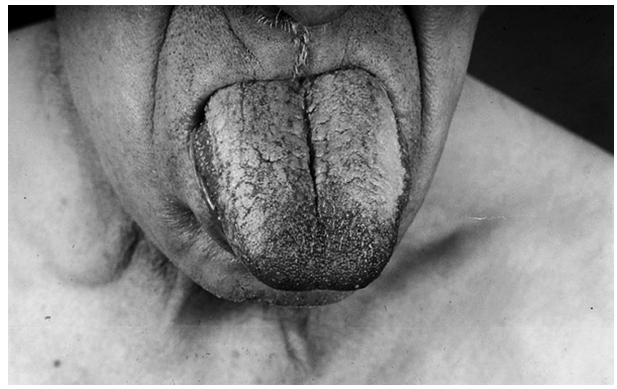


Figure 2.3.10.4 Macroglossia in a patient with acromegaly. The patient had to have surgical tongue reduction.

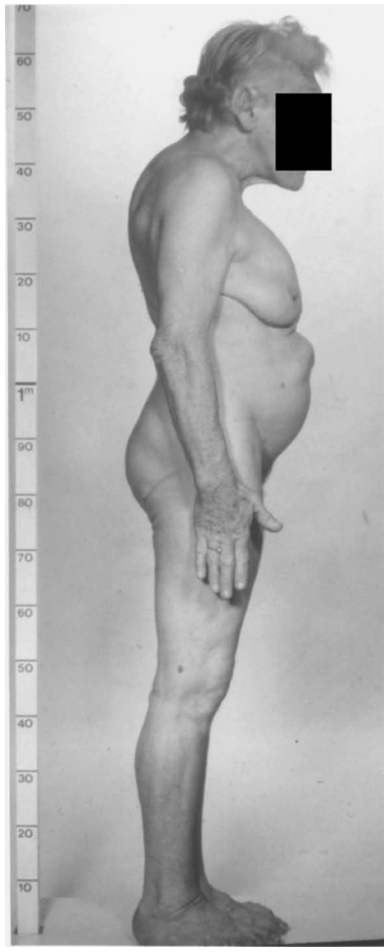


Figure 2.3.10.3 Kyphosis in a patient with acromegaly.

Box 2.3.10.2 Clinical features of acromegaly

Acral enlargement
- Increased hand-, shoe-, and ring size
- Prominent nasolabial fold
- Frontal bossing
- Prominent supraorbital ridge

Skin
- Increased sweating
- Oiliness and increased sebaceous activity
- Hirsutism
- Hyperpigmentation
- Thickened skin (detected on the dorsum of the hand)
- Skin tags
- Furrowing of the skin on the forehead or back of the head (cutis verticis gyrate)

Cardiovascular
- Hypertension
- Congestive heart failure
- Ventricular hypertrophy
- Cardiomyopathy [25]

Respiratory
- Sleep apnoea, snoring
- Enlarged sinuses
- Deep voice

Musculoskeletal
- Arthropathy, knee, hip, lumbar spine [26]
- Kyphosis (Figure 2.3.10.3)
- Prognathism
- Dental malocclusion

Box 2.3.10.2 *Continued*

- Muscle weakness
- Increased height

Alimentary

- Macroglossia (Figure 2.3.10.4)
- Visceromegaly (liver, spleen, kidney, bowels, salivary glands)
- Colonic polyps, constipation

Neurological

- Headache
- Sudden severe headache often with ptosis (pituitary apoplexy)
- Carpal tunnel syndrome (Figure 2.3.10.5)
- Paraesthesia

Reproductive

- Amenorrhoea
- Galactorrhoea
- Impotence
- Prostatic hypertrophy

Metabolic alterations

- Increased insulin resistance, diabetes mellitus
- Hypercalciuria, hypercalcaemia [27]

Endocrine system

- Cosecretion of prolactin or TSH
- Galactorrhoea
- Hypopituitarism
- Multinodular goitre

Psychological effects

- Anxiety due to distorted body image
- Tiredness, fatigue, apathy, depression

Local tumour effects

- Headache
- Visual field defects (bitemporal hemianopsia)
- Cranial nerve palsy

adenoma. Patients may present to rheumatologists, orthopaedic surgeons (joint pain and abnormalities), dentists (separation of the teeth, jaw pain), neurologists (carpal tunnel syndrome), gynaecologists (menstrual disturbance), ophthalmologists (visual problems), cardiologists (high blood pressure, heart disease) or to physicians treating the patient's sleep apnoea, hypertension, or diabetes mellitus [24]. Often the symptoms are present and progress insidiously over several years. It can be useful to review serial old photographs to show the presence and progress of subtle facial appearances.

Symptoms of an enlarged pituitary fossa are the same as with non-functioning tumours and are discussed elsewhere and include visual field defects, headache, and pituitary apoplexy (more often in younger patients).

The Phenotypic Changes of Acromegaly

Skin on the back of the hand is thickened and excessive sweating occurs in 80% with patients looking older than their years.

In the cardiovascular system, hypertension is present in 50% due to a direct effect of GH on sodium absorption, and there is also increased left ventricular muscle mass. Ischaemic heart disease is also present, possibly exacerbated by insulin resistance and is a major cause of morbidity and mortality. Myocardial hypertrophy with fibrosis leading to ventricular dilatation and biventricular failure are features of an acromegalic cardiomyopathy.

Respiratory symptoms are also common and account for part of the increased mortality of the condition. Sleep apnoea may result from significant airway obstruction caused by prognathism, macroglossia, and hypertrophied nasal structures and narcolepsy may be a presenting symptom in patients with acromegaly. Difficulty in tracheal intubation can be a problem in patients with acromegaly undergoing anaesthesia.

In the alimentary tract macroglossia, constipation and visceromegaly are common. A high prevalence of colonic polyps in acromegaly is reported and these may progress to colonic carcinoma. GH and/or IGF-1 may possess direct mitogenic effects on colonic epithelial cells.

Metabolic Consequences of Acromegaly

Increased insulin resistance occurs because of direct anti-insulin effects of GH and can result in diabetes mellitus and carbohydrate tolerance, which is reversible with successful treatment of the GH hypersecretion. Hypercalciuria occurs in 80% of patients because of GH being facultative in the synthesis of 1,25-dihydroxyvitamin D increases calcium absorption from the gut and reabsorption in the kidney. Hyperphosphataemia may occur due to the direct effect of GH/IGF-1 on renal phosphate reabsorption. If hypercalcaemia is detected, hyperparathyroidism and MEN1 syndrome (3%) needs to be investigated. Multinodular goitre occurs with increased frequency in acromegaly. IGF-1 is a major determinant of thyroid cell growth. Thyroid dysfunction (hyperthyroidism) occurs in acromegaly and is most commonly due to a multinodular goitre, but TSH secretion from a mixed pituitary tumour should be considered if the TSH is inappropriately normal/elevated in association with thyrotoxicosis. See Figure 2.3.10.5.

Cardiovascular and Respiratory Risk

This increased risk relates to hypertension and diabetes. There is no characteristic lipid disturbance in acromegaly. Before 1966, 50% of patients with acromegaly died before the age of 50, cardiovascular

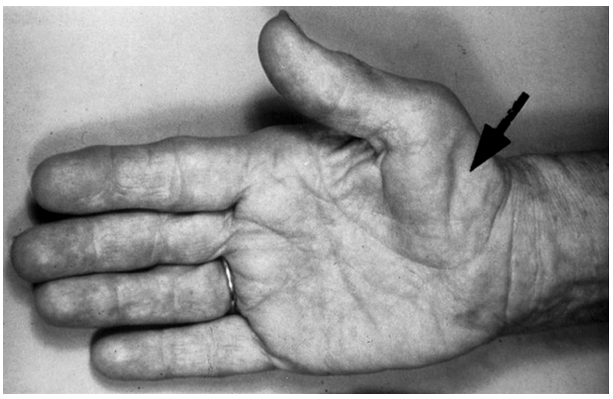


Figure 2.3.10.5 Carpal tunnel syndrome in acromegaly. Thenar wasting is clearly seen (arrow).

disease being the commonest cause of death. Cardiovascular disorders accounted for about 25% of deaths, followed by respiratory (20%) and cerebrovascular disease (15%). More recent data suggest a twofold risk of cardiovascular disease and no increased respiratory mortality [17].

Malignancy in Acromegaly

Whether acromegaly increases the risk of malignant disease is a controversial issue. The large German registry found no increased risk [28], while a Danish study found slightly increased risk [29]. A recent meta-analyses highlighted the contradictory findings of the various studies and acknowledged that some may have had positive bias but equally recognized that as patients with acromegaly live longer, the effect of GH/IGF-1 excess on cancer risk may become more apparent [29–31].

Mortality

Untreated acromegaly is associated with decreased life expectancy. This was first documented in the 1950s and was attributed to increased cardiovascular and respiratory mortality. Overall mortality of untreated disease is approximately double of the normal rate, while mortality in patients who are well controlled is not significantly different from that of the general population [32]. More recently the possibility of increased mortality due to malignant disease has been raised [31].

Biochemical Confirmation of Acromegaly

The greatest challenge in the diagnosis of acromegaly is for the disease to be considered; biochemical confirmation is rarely a challenge and relies upon a combination of an elevated, age-controlled IGF-1 level and failure of GH to suppress to <0.4 μg/L during an oral glucose tolerance test (OGTT) [33]. In patients with acromegaly, there may even be a paradoxical rise in GH in response to OGTT, and this have been attributed to increased expression of the receptor for gastrointestinal inhibitory hormone (GIPR) [34]. False positives do occur (**Boxes 2.3.10.3 and 4**), but few conditions apart from adolescence and pregnancy are likely to cause diagnostic confusion. In tall adolescents, possibly as a consequence of large GH pulses, GH levels may not adequately suppress during an OGTT, and during puberty IGF-I levels may overlap with those encountered in acromegaly. Careful examination for phenotypic changes and a pituitary scan should clarify the situation.

Box 2.3.10.3 Conditions associated with a failure of suppression after a glucose load

- Adolescence
- Diabetes mellitus
- Liver failure
- Renal failure
- Malnutrition
- Laron dwarfism
- Anorexia nervosa

Box 2.3.10.4 Investigation of acromegaly

Establish diagnosis
- 75 g OGTT
- IGF-1

Metabolic consequences of high GH
- OGTT (for glucose)
- HbA1c
- 24-h urine calcium

Pituitary function
- Prolactin
- LH/FSH, testosterone/oestradiol
- fT4, TSH
- Cortisol
- Insulin-tolerance test (short synacthen test of cortisol reserve)

Pituitary anatomy
- MRI
- Visual fields

Other (coexistent) diagnoses
- Serum calcium (multiple endocrine neoplasia)
- Plasma metanephrines (phaeochromocytoma)
- Sleep apnoea
- Hypertension
- Spine X-ray

fT4, free thyroxine; LH/FSH, luteinizing hormone/follicle-stimulating hormone; OGTT, oral glucose tolerance test; TSH, thyroid-stimulating hormone.

Initial Investigations

Imaging

MRI of the pituitary is essential and will demonstrate a macroadenoma (>1 cm) in approximately 60–70% of patients. Visual fields should be objectively documented if there is any suggestion of visual pathway compression. Plasma GHRH should be measured if an ectopic source of acromegaly is suspected, or if occasionally, pituitary histology reveals hyperplasia. A thyroid ultrasound should be done if there are thyroid nodules.

Other Associations

Essential hypertension is common, often associated with an increase in intravascular volume and low renin and increased aldosterone secretion [27]. Phaeochromocytoma is not normally associated with acromegaly, except in exceptional cases (genetic condition causing paraganglioma/phaeochromocytoma and pituitary adenoma, or ectopic GHRH secretion from a phaeochromocytoma [35]); however, it is important to exclude a phaeochromocytoma in a hypertensive patient with acromegaly, particularly prior to surgery.

Treatment

Ideal Treatment

The ideal treatment will render GH secretion normal, completely ablate the pituitary tumour mass, while preserving normal pituitary

function and reverse the acral and other systemic complications, while ensuring there is no biochemical or tumour recurrence [36, 37]. The reality is that some phenotypic changes, such as osteoarthritis of the weight-bearing joints or hypopituitarism may be irreversible, and all treatments have limitations and the potential for complications. The greatest improvement in outcomes would come from earlier diagnosis. Deaths due to mass effects of the pituitary tumour are rare in acromegaly, with mortality and morbidity being largely a consequence of the metabolic complications of long-term elevation of circulating GH and IGF-1 levels and hypopituitarism. Several studies have suggested that lowering IGF-1 levels into age-related reference ranges is associated with restoration of normal life expectancy [31, 32, 37].

Modes of Treatment

Transsphenoidal surgery, undertaken by a pituitary specialist surgeon, is the treatment of choice. In patients not cured by surgery, medical treatment can be effective at achieving biochemical disease control. Modern radiotherapy is extremely effective at controlling tumour growth and will with time control GH secretion (Box 2.3.10.5) A multidisciplinary team, consisting of, at a minimum, a pituitary endocrinologist, specialist pituitary surgeon, radiotherapy oncologist, neuroradiologist, and histopathologist should be involved in all major therapeutic decisions, while clinical geneticist, nurse practitioner, pituitary pathologist also contribute to patient care.

Defining the Biochemical Goals of Treatment

An expert consensus conference defined biochemical diagnosis and successful treatment as a nadir GH during an OGTT of <0.4 μg/L and an IGF-1 within a vigorously established, age-related reference range [37]. The adoption of a GH nadir value of <0.4 μg/L reflects an era of ultrasensitive GH assay which for the first time permits routine quantification of GH as low as 0.05 μg/L, which is lower than the nadir values encountered after a glucose load in healthy subjects [38]. Historical cut-off criteria, such as 1 μg/L, were not underpinned by an evidence base [39], were known to be associated with significant numbers of patients with discrepant GH and IGF-1 results and in particular resulted in reports of patients with unequivocal acromegaly with elevated IGF-1 values but nadir GH values <1 μg/L [40, 41].

The pulsatile nature of GH secretion, even in acromegaly [42], means that it is unwise to rely on a single random measure of GH. The nadir GH during an OGTT and the mean of five GH samples taken over the course of a single day have both been widely used. Both means of assessment are closely correlated to serum IGF-1 levels, but recent consensus criteria have predominately relied upon the OGTT. IGF-1 is the preferred measure of disease activity as it has dual advantages over GH: (i) convenience of a single, randomly-timed sample and (ii) it is a measure of GH action rather than tumour secretion that better correlates to biological end-points such as glucose tolerance [43]. That said, it needs to be recognized that saturation of IGF-1 generation occurs at GH levels of approximately 20 μg/L [44]. This implies that for example partial surgical resection of an adenoma may lower circulating GH levels from 100 to 50 μg/L, but that change may have little impact on circulating IGF-1. The relationship between mean GH levels and IGF-1 is also modulated by oestrogen therapy, prior radiotherapy, age, and liver disease, therefore when possible both GH and IGF-1 status should be assessed.

A major challenge in applying international consensus criteria to local practice is bias in assay performance which is an issue with both GH & IGF-1. Endocrinologists need to be aware of the characteristics of the assays used in their laboratories.

Box 2.3.10.5 Modes of treatment of acromegaly

Surgery
- Transsphenoidal
- Transfrontal

Medical therapy
- Dopamine agonists
- Somatostatin analogues
  - First generation (octreotide or lanreotide)
  - Second generation (pasireotide)
- Growth hormone receptor antagonist

Stereotactic radiotherapy
- Multifractional
- Radiosurgery (single fraction), e.g. Gamma Knife®
- Proton beam therapy

Transsphenoidal Surgery

Factors Affecting Outcome

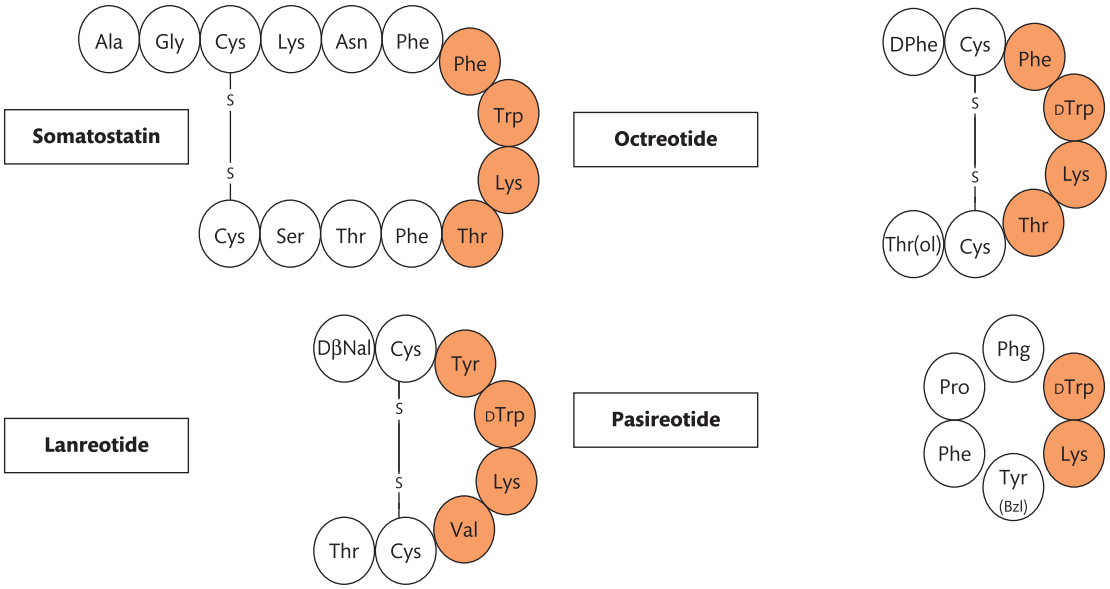
Pretreatment GH levels have been shown to affect outcome such that high levels are associated with a less successful surgical outcome [45]. The presence of micro- or macroadenoma and extent of cavernous sinus invasion has a key role affecting success of surgery [46, 47]. As opposed to tumour size and GH levels, the determinant of outcome that an endocrinologist can influence is the choice of surgeon. The outcome of surgery is very dependent on the skill and experience of the surgeon with significant differences being reported both in terms of disease control and complications is crucial to outcome [48]. The outcome is less good when large numbers of surgeons doing a small number of operations, and several centres have reported considerably improved outcomes following transition to having one or two surgeons to do all pituitary surgery. Complications are also less common with experienced surgeons [47].

Postoperative Assessment

The impact of surgery on GH status can be assessed 3 days postoperatively, but it can take three months for IGF-1 to reaches its nadir. The first routine postoperative scan should be undertaken at 6 months. The available figures in the best surgical hands show that between 70% and 90% of microadenomas and between 45% and 66% of macroadenomas can reach GH and IGF-1 levels in the safe range with surgery [47].

Complications

New hypopituitarism develops in between 12% and 18% of patients undergoing transsphenoidal surgery for acromegaly and implies



	SSTR1	SSTR2	SSTR3	SSTR4	SSTR5
Somatostatin	0.9	0.2	0.6	1.5	0.3
Octreotide	280.0	0.4	7.1	>1000	6.3
Lanreotide	2129	0.7	98	1826	5.2
Pasireotide	9.3	1.0	1.5	>1000	0.2

Figure 2.3.10.6 Left panel: receptor-binding profile of somatostatin analogues IC50 (nM) [52, 53]. Right panel: the structure of somatostatin-14 and the three somatostatin analogues currently used for the therapy of acromegaly.

the need for lifelong pituitary hormone replacement therapy. Occasionally pituitary function may recover [49]. Other complications include transient or permanent diabetes insipidus, cerebrospinal fluid leaks, haemorrhage, and meningitis. Recurrence of acromegaly occasionally occurs (5.5% at 3 years).

Transcranial Surgery

Developments in transsphenoidal techniques, particular the endoscope, mean that transcranial surgery is rarely indicated. A transcranial approach may be required when there is a very large suprasellar extension or a tumour extending out laterally which is unreachable transsphenoidally. In these cases, the reduction of GH to desired levels is virtually never obtained.

In patients with larger extrasellar tumours, surgical debulking of the tumour should take place which will lower GH and IGF-1 levels and improve the efficacy of medical treatment [50].

Medical Therapy

For the great majority of patients first line therapy is surgery, and medical therapy is reserved for patients with persisting disease activity after surgery.

Somatostatin Analogues

For more than 25 years first generation somatostatin analogues octreotide and lanreotide have been the mainstay of medical therapy for acromegaly. Both are synthetic octapeptide analogues of somatostatin but with longer duration of action than the native peptide with activity only on the type 2 somatostatin receptor (SSTR2) (Figure 2.3.10.6). Comparison of octreotide long-acting release (LAR) and lanreotide Autogel show similar numbers of patients obtaining disease control [51]. Recognition that GH secretion is regulated by both SSTR2 and 5 resulted in the development of second-generation pasireotide, that has high affinity for SSTRs 1, 2, 3 and, particularly, 5, for patients unresponsive to lanreotide and octreotide.

The significant proportion of patients respond to somatostatin analogue therapy to a greater or lesser extent, although controversy persists on the proportion that normalize IGF-1 and/or GH levels. The lack of clarity in part reflects the evermore stringent definition of disease control. As discussed earlier, the goal in terms of GH, in response to a glucose load, has been lowered from 1 to 0.4 μg/L, while better validated, age-related IGF-1 reference ranges have resulted in significantly lowering of the target range for IGF-1. In 1998, Newman *et al.*

reported control of IGF-1 at 3 years in 73% of patients with octreotide as primary therapy [54]; however, the target range for IGF-1 of 0.4 – 2.2×10^3 U/L in women and 0.4 – 1.9×10^3 U/L in men is unrecognizable by modern standards and not age-adjusted. Using strict criteria (GH <2.5 μg/L and normalized age-related IGF-1) in unselected patients, full control is only achieved in 25% [55]. A meta-analysis found that approximately 40% of patients achieve IGF-1 normalization with octreotide or lanreotide [56]. A prospective, randomized, double-blind, multicentre, 12-month study in a total of 358 patients reported control of IGF-1 in approximately 23% of patients treated with octreotide and 38% of patients on pasireotide [57]. Patients on somatostatin analogue therapy can be followed by IGF-1 and probably mean GH levels rather than the response to OGTT [58]. There are several oral somatostatin analogues under development. They are unlikely to result in improved disease control but will offer patients an alternative to the existing parenteral preparations.

Acromegaly related headache can improve in response to somatostatin analogue therapy. Short-acting octreotide injections can relieve headache in some patients resistant to other treatment modalities [59].

Effect on Carbohydrate Tolerance

The impact of octreotide and lanreotide can be variable and unpredictable in individual patients [60]. The balance between the benefit of lower GH levels against the impact of suppression of insulin secretions means octreotide and lanreotide therapy can result in either improved or deterioration of glucose tolerance. In contrast, pasireotide results in hyperglycaemia or diabetes mellitus in 70% of patients [61].

Side Effects

Diarrhoea and abdominal pain occur in 30% of patients to a mild or moderate degree initially but in the vast majority these usually settle. The most important chronic side effect is gallstones, which complicates long-term therapy with octreotide and the somatostatin analogues. The rate varies widely between 14% and 60% and probably depends on the length of treatment. They develop because octreotide decreases gallbladder contractility by suppressing cholecystokinin release. Bile also becomes abnormal, possibly in relation to prolonged intestinal transit and altered bacterial flora. The abrupt withdrawal of octreotide may be associated with the development of acute pancreatitis or gallstone colic [62]. Otherwise, gallstones developing on somatostatin analogues very rarely cause symptoms [63]. Antibody formation occurs but rarely and is very infrequently significant in terms of altering GH levels. Patients with acromegaly related headache using short-acting octreotide have been found to develop dependency to octreotide [59].

Place of Treatment

Most frequently somatostatin analogues are used postoperatively following 'non-curative' surgery. There has been interest in the use of somatostatin analogues either as an alternative to surgery (primary medical therapy) or for a limited time preoperatively with the desire to reduce morbidity and possibly, by shrinking the tumour, to improve the surgical cure rate. The proponents of primary medical therapy have pointed to the poor results of surgery undertaken by inexperienced surgeons and associated morbidity and subsequent need for long-term postoperative somatostatin analogue therapy and argued that there is little to be gained if surgery is not curative. However, with the recognition, as discussed earlier, that only a modest proportion of patients achieve biochemical control with somatostatin analogue therapy and open-ended, lifelong somatostatin analogue therapy is extremely expensive, a more appropriate solution is to ensure access to surgeons with documented excellent outcomes. Furthermore, there is evidence that the fall in GH levels from debulking, non-curative increases the prospect of subsequent somatostatin analogue therapy achieving disease control [50, 64, 65]. Primary therapy should be confined to patients not fit for surgery.

Presurgical somatostatin can result in significant reduction in tumour size of at least 20% in 75% of the patients. Although a meta-analysis suggest preoperative somatostatin analogue therapy for macroadenomas increases the chance of biochemical control of acromegaly 3 months after surgery [66], with long-term follow-up there is no significant benefit [67, 68]. The apparent explanation is that at three months the effect of the prior somatostatin analogue therapy has not fully washed-out. Presurgical somatostatin analogue therapy can be indicated in patients to reduce peri-operative morbidity and reduce soft tissue changes that can make intubation difficult. To date no study, however, has shown clear results in these regards.

Dopamine Agonists

Pharmacology

Bromocriptine, cabergoline, and quinagolide are selective agonist at the D_2 dopamine receptors. Their administration results in the paradoxical fall of GH levels in patients with acromegaly [69, 70], while in normal subjects they stimulate GH levels. Bromocriptine is the only dopamine agonist licensed for the treatment of acromegaly but cabergoline is the most potent and best tolerated and probably the most widely used in these circumstances.

Effects on Growth Hormone

It is not possible to predict the response to bromocriptine or cabergoline. Overall between 10% and 20% of patients have GH levels that are safe on treatment with bromocriptine (usually 20–40 mg daily) or cabergoline (1–3 mg weekly). The best responses are obtained with patients who before treatment have only marginal elevation of GH & IGF1 [70]. Carbohydrate tolerance improves because of the lowering of GH levels and prolactin levels are suppressed to below normal.

Side Effects

Acute postural hypotension, nausea, and vomiting. Usually these settle with time. Very rarely, particularly on high doses, psychosis and digital vasospasm may also develop. Impulse control disorders are a well-recognized but rare complication and can manifest as pathological gambling, shopping, eating, and hypersexuality. Patients should be warned of the risk. Cardiac fibrosis has been described with the high doses of cabergoline used for Parkinson's disease and it is recommended that all patients treated with high-dose

ergot-derived dopamine agonists (e.g. cabergoline) have an echocardiogram. Although patients with acromegaly and prolactinoma are routinely given much lower doses, recent recommendations suggest pretreatment baseline echocardiography, with a repeat at 5 years if cabergoline dose ≤ 2 mg/week, or yearly if dose is > 2 mg/week [71].

Place of Treatment

Dopamine agonists are less expensive than somatostatin analogues and orally administered. When medical treatment is indicated it should theoretically be the case that dopamine agonists are tried first. In practice, because the response rate is low this does not happen. However, it should be noted that occasionally patients who are not responsive to a somatostatin analogue may respond to a dopamine agonist.

Growth Hormone Receptor Antagonist (Pegvisomant)

Pegvisomant is a pegylated recombinant growth hormone analogue that inhibits GH action by binding the GH receptor without inducing signal transduction and IGF-1 induction [72]. As it does not lower circulating GH, serum IGF-1 is the principal biochemical means of monitoring effectiveness of treatment. With appropriate dose titration, serum IGF-1 level can be controlled in over 90% of patients [13]. The maximum licensed dose is 30 mg per day, although doses up 60 mg per day have been used. Although marketed as a daily injection, the long half-life of pegvisomant means that it is effective when administered once or twice weekly, although the volume of injection can be a challenge at higher doses [73].

Due to its mode of action the ability of pegvisomant to lower IGF-1 is independent of any tumour characteristics and is effective in patients resistant to somatostatin analogues and dopamine agonists. Pegvisomant may also be used in combination with somatostatin analogues and dopamine agonists [74, 75].

The drug is generally well tolerated, although liver function disturbance, specifically elevation of transaminase shortly after initiation of treatment, has been reported. Lipohypertrophy may occur at the site of injection.

When first introduced there was concern that pegvisomant therapy would be associated with pituitary tumour expansion but close follow-up of patients on long-term therapy has provided reassurance, although annual MRI is recommended.

Place of Treatment

Pegvisomant is indicated for patients with persistent disease despite expert surgery and optimal somatostatin analogue treatments, with the choice being whether to add it to ongoing somatostatin analogue treatment or substitute pegvisomant in its place. The decision depends on individual patient circumstances (e.g. good tumour shrinkage with a somatostatin analogue would be a reason for combination treatment while deteriorating glucose tolerance argues for monotherapy) [37].

Radiotherapy

Indications

There are two indications for radiotherapy in patients with acromegaly: either to control tumour growth, and/or to lower GH secretion with the goal of achieving biochemical disease control.

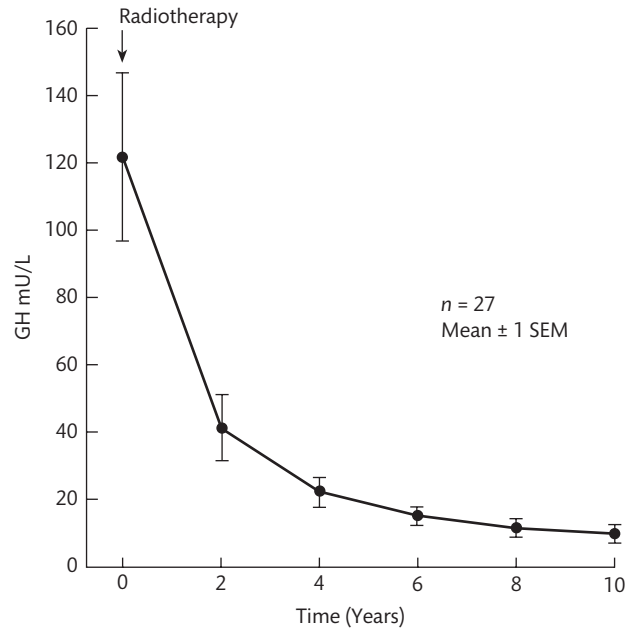


Figure 2.3.10.7 The exponential fall in GH levels after radiotherapy in patients with acromegaly studied over 10 years.

Modern radiotherapy is an effective, safe, and appropriate third-line treatment in patients with persisting hormonal excess or an expanding tumour after surgery and optimized medical therapy. The frequently cited papers describing the complications of pituitary radiotherapy largely pertain to conventional radiotherapy delivered many years ago using what is now obsolete technology. The latency of complications is long and therefore it is difficult to make definitive statements, but the combination of MRI and conformal radiotherapy allows size, shape and exact location of the tumour to be defined and beam shaping to the precise tumour contours which greatly reduces normal tissue doses. All modern radiotherapy is stereotactic and can be delivered in multiple small doses (fractionated stereotactic radiotherapy) or in a single session (stereotactic radiosurgery).

Radiotherapy is extremely effective at controlling tumour growth in 91–100% of patients [76]. ln general terms, radiotherapy can be expected to reduce GH levels by 50% in two years and by further 25% by 5 years [77]. See Figure 2.3.10.7.

Side Effects

As discussed elsewhere (see Chapter 2.3.8), external beam radiotherapy is well tolerated. Hypopituitarism is common and after exclusion of patients with preirradiation hormone deficiency; post-radiotherapy, gonadal, adrenal, and thyroid deficiency occur in 50, 35, and 35%, respectively, at 10 years. Pituitary function deterioration develops gradually so that it is necessary for regular (usually annual) assessment. Visual loss and radiation-induced late malignancy are discussed elsewhere and there is no specific increased risk associated with acromegaly. The effect of external

pituitary radiotherapy on memory and mental function requires further study.

Acromegaly and Pregnancy

Fertility is impaired in patients with acromegaly, but pregnancy is generally safe in terms of hormonal and tumoural stability, although there is an increased risk of gestational diabetes and hypertension. The hormonal changes of acromegaly, and their assessment, during pregnancy are complex [78, 79]. The elevated oestrogen levels of pregnancy antagonize GH action, while placental GH secretion, GH-V, has similar biological characteristics as 22 kD GH but is variably detected by routine GH assays. Typically IGF-1 remains satisfactory for the duration of pregnancy. The lack of normative data for GH and IGF-1 during pregnancy makes biochemical assessment challenging and in most cases is not necessary. Visual fields should be monitored in case of tumour enlargement.

Somatostatin analogue exposure may increase the chance of low birth weight, so it is recommended that drug treatment is ceased during pregnancy and for 2 months prior to a planned pregnancy. An exception to this could be the development of severe headache during pregnancy. There are limited data of on the safety of pegvisomant, available data did not reveal any apparent drug-related complications, but pegvisomant and it should be stopped in any patient trying to conceive or on confirmation of pregnancy [80].

Treatment Paradigms

Optimal outcomes require the input of a multidisciplinary team coordinated by a pituitary endocrinologist and including a specialist pituitary surgeon, neuroradiation oncologist, neuro-radiologist, and access to adequately validated GH and IGF-1 assays.

Patients with microadenomas should be 'cured' by transsphenoidal surgery. The outcome in macroadenomas is dependent on the size and invasiveness of the tumour but, in general terms, should be 'curative' in around 55%. Surgery is usually performed for macroadenoma even if surgical cure is unlikely, because debulking surgery improves the outcome of treatment with somatostatin analogues [50]. As approximately 60–70% patients have a macroadenoma, a significant number of patients require further therapy with somatostatin analogue or dopamine agonist therapy. Somatostatin analogue therapy offers the better prospect of achieving biochemical disease control, but occasionally, in non-somatostatin analogue responsive patients, dopamine agonists may be effective. In patients with persisting active disease the choice lies between pasireotide and pegvisomant, with the latter being used either as an alternative to octreotide or lanreotide or in combination. Comorbidities should be longitudinally monitored (e.g. blood pressure, carbohydrate tolerance, arthritis, and sleep apnoea).

It is important to consider radiotherapy in patients in whom there is evidence of tumour expansion or if medical therapy is ineffective. Radiotherapy will take several years to adequately control GH and IGF-1 levels but does offer the prospect of patients eventually being able to discontinue high-cost medical therapy.

Differential Diagnosis of Acromegaly Phenotype

Acromegaly is a characteristic disease and once the possibility of its diagnosis has been suggested, it is relatively easy to make a firm clinical and biochemical diagnosis. There are a few conditions, however, when clinical or biochemical features resembling acromegaly are present but there is another underlying diagnosis, these conditions can be summarized as pseudoacromegaly [81]. Acromegaloidism (insulin-mediated pseudoacromegaly) refers to the development of acromegaly-like features (e.g. jaw, hand, and feet enlargement) together with acanthosis nigricans caused by very severe insulin resistance. GH and IGF-1 values are normal. Pachydermoperiostosis (OMIM 1671002), is a familial condition associated with acromegaloid facial features and enlarged hands and feet with characteristic clubbing of the fingers and toes. Cantu syndrome also shares some facial features with acromegaly. The differential diagnosis of tall stature (constitutional or due to another disease) and gigantism can also sometimes cause problems, especially as the biochemical diagnosis of true abnormal GH excess is challenging in teenagers. Biochemical alterations of elevated GH with normal or low IGF-1 can be seen in some conditions, but clinically these do not represent differential diagnostic problems (e.g. kidney disease).

A unique, recently identified condition due to *IGFS1* mutation in some patients, despite the lack of pituitary adenoma, can present with elevated GH and IGF-1, some acromegalic features and somewhat taller stature (therefore not fulfilling the pseudoacromegaly criteria) [82]. There is no explanation currently for the abnormal GH axis in this condition.

REFERENCES

1. Chahal HS, Stals K, Unterlander M, *et al.* AIP mutation in pituitary adenomas in the 18th century and today. *N Engl J Med*, 2011; **364**: 43–50.
2. de Herder WW. Acromegaly and gigantism in the medical literature. Case descriptions in the era before and the early years after the initial publication of Pierre Marie (1886). *Pituitary*, 2009; **12**: 236–44.
3. Fraentzel O. *Ueber Akromegalie. Deutsche Medizinische Wocheschrift Deutsche Medizinische Wochenschrift*, 1888.
4. Landis CA, Masters SB, Spada A, Pace AM, Bourne HR, Vallar L. GTPase inhibiting mutations activate the a chain of Gs and stimulate adenylyl cyclase in human pituitary tumors. *Nature*, 1989; **340**: 692–6.
5. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*, 1991; **325**: 1688–95.
6. Erdheim J. Zur normalen und pathologischen Histologie der Glandula thyreoidea, parathyroidea und Hypophysis. *Beitr Pathol Anat*, 1903, **33**: 158–65.
7. Chandrasekharappa SC, Guru SC, Manickam P, *et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science*, 1997; **276**: 404–7.
8. Vierimaa O, Georgitsi M, Lehtonen R, *et al.* Pituitary adenoma predisposition caused by germline mutations in the AIP gene. *Science*, 2006; **312**: 1228–30.

9. Trivellin G, Daly AF, Faucz FR, *et al*. Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. *N Engl J Med*, 2014; **371**: 2363–74.
10. Rodd C, Millette M, Iacovazzo D, *et al*. Somatic GPR101 duplication causing X-linked acrogigantism (XLAG)-diagnosis and management. *J Clin Endocrinol Metab*, 2016; **101**: 1927–30.
11. Liuzzi A, Chiodini PG, Botalla L, Cremascoli G, Silvestrini F. Inhibitory effect of L-Dopa on GH release in acromegalic patients. *J Clin Endocrinol Metab*, 1972; **35**: 941–3.
12. Lamberts SW, Uitterlinden P, Verschoor L, van Dongen KJ, del Pozo E. Long-term treatment of acromegaly with the somatostatin analogue SMS 201–995. *N Engl J Med*, 1985; **313**: 1576–80.
13. Trainer PJ, Drake WM, Katznelson L, *et al*. Treatment of acromegaly with the growth hormone-receptor antagonist pegvisomant. *N Engl J Med*, 2000; **342**: 1171–7.
14. Lavrentaki A, Paluzzi A, Wass JA, Karavitaki N. Epidemiology of acromegaly: review of population studies. *Pituitary*, 2017; **20**: 4–9.
15. Heaney AP. Clinical review: pituitary carcinoma: difficult diagnosis and treatment. *J Clin Endocrinol Metab*, 2011; **96**: 3649–60.
16. Larkin S, Ansorge O. Pathology and pathogenesis of pituitary adenomas and other sellar lesions. In: Feingold KR, Anawalt B, Boyce A, *et al.*, eds. Endotext. South Dartmouth (MA)2017.
17. Iacovazzo D, Carlsen E, Lugli F, *et al.*, eds. Expression of somatostatin receptor 2 & 5 and AIP: useful tools to predict responsiveness to conventional somatostatin analogs and pasireotide treatment in acromegaly. *Endocr Rev*, 2015; 2015.
18. Chinezu L, Vasiljevic A, Trouillas J, Lapoirie M, Jouanneau E, Raverot G. Silent somatotroph tumour revisited from a study of 80 patients with and without acromegaly and a review of the literature. *Eur J Endocrinol*, 2017; **176**: 195–201.
19. Tuominen I, Heliovaara E, Raitila A, *et al*. AIP inactivation leads to pituitary tumorigenesis through defective Galphai-cAMP signaling. *Oncogene*, 2015; **34**: 1174–84.
20. Salenave S, Boyce AM, Collins MT, Chanson P. Acromegaly and McCune-Albright syndrome. *J Clin Endocrinol Metab*, 2014; **99**: 1955–69.
21. Stratakis CA, Salpea P, Raygada M. Carney complex. In: Pagon RA, Adam MP, Ardinger HH, *et al.*, eds. Seattle, WA: GeneReviews(R), 2018.
22. Daly AF, Tichomirowa MA, Petrossians P, *et al*. Clinical characteristics and therapeutic responses in patients with germ-line AIP mutations and pituitary adenomas: an international collaborative study. *J Clin Endocrinol Metab*, 2010; **95**: E373–83.
23. Hernández-Ramírez LC, Gabrovska P, *et al*. Landscape of familial isolated and young-onset pituitary adenomas: prospective diagnosis in AIP mutation carriers. *J Clin Endocrinol Metab*, 2015; **100**: E1242–54.
24. Kreitschmann-Andermahr I, Siegel S, Kleist B, *et al*. Diagnosis and management of acromegaly: the patient's perspective. *Pituitary*, 2016; **19**: 268–76.
25. Sharma AN, Tan M, Amsterdam EA, Singh GD. Acromegalic cardiomyopathy: Epidemiology, diagnosis, and management. *Clin Cardiol*, 2018; **41**: 419–25.
26. Mazziotti G, Maffezzoni F, Frara S, Giustina A. Acromegalic osteopathy. *Pituitary*, 2017; **20**: 63–9.
27. Kamenicky P, Mazziotti G, Lombes M, Giustina A, Chanson P. Growth hormone, insulin-like growth factor-1, and the kidney: pathophysiological and clinical implications. *Endocr Rev*, 2014; **35**: 234–81.
28. Petroff D, Tonjes A, Grussendorf M, *et al*. The incidence of cancer among acromegaly patients: results from the German acromegaly registry. *J Clin Endocrinol Metab*, 2015; **100**: 3894–902.
29. Dal J, Leisner MZ, Hermansen K, *et al*. Cancer incidence in patients with acromegaly: a cohort study and meta-analysis of the literature. *J Clin Endocrinol Metab*, 2018; **103**: 2182–8.
30. Boguszewski CL, Ayuk J. Management of endocrine disease: acromegaly and cancer: an old debate revisited. *Eur J Endocrinol*, 2016; **175**: R147–56.
31. Bolfi F, Neves AF, Boguszewski CL, Nunes-Nogueira VS. Mortality in acromegaly decreased in the last decade: a systematic review and meta-analysis. *Eur J Endocrinol*, 2018; **179**: 59–71.
32. Holdaway IM, Bolland MJ, Gamble GD. A meta-analysis of the effect of lowering serum levels of GH and IGF-I on mortality in acromegaly. *Eur J Endocrinol*, 2008; **159**: 89–95.
33. Freda PU, Reyes CM, Nuruzzaman AT, Sundeen RE, Bruce JN. Basal and glucose-suppressed GH levels less than 1 microg/L in newly diagnosed acromegaly. *Pituitary*, 2003; **6**: 175–80.
34. Hage M, Chaligne R, Viengchareun S, *et al*. Hypermethylator phenotype and ectopic gip receptor in GNAS mutation-negative somatotropinomas. *J Clin Endocrinol Metab*, 2019; **104**: 1777–87.
35. Ghazi AA, Amirbaigloo A, Dezfooli AA, *et al*. Ectopic acromegaly due to growth hormone releasing hormone. *Endocrine*, 2013; **43**: 293–302.
36. Katznelson L, Laws ER, Jr., Melmed S, *et al*. Acromegaly: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 3933–51.
37. Melmed S, Bronstein MD, Chanson P, *et al*. A consensus statement on acromegaly therapeutic outcomes. *Nat Rev Endocrinol*, 2018; **14**: 552–61.
38. Schilbach K, Gar C, Lechner A, *et al*. Determinants of the growth hormone nadir during oral glucose tolerance test in adults. *Eur J Endocrinol*, 2019; **181**: 55–67.
39. Trainer PJ. Editorial: acromegaly—consensus, what consensus? *J Clin Endocrinol Metab*, 2002; **87**: 3534–6.
40. Butz LB, Sullivan SE, Chandler WF, Barkan AL. 'Micromegaly': an update on the prevalence of acromegaly with apparently normal GH secretion in the modern era. *Pituitary*, 2016; **19**: 547–51.
41. Dimaraki EV, Jaffe CA, DeMott-Friberg R, Chandler WF, Barkan AL. Acromegaly with apparently normal GH secretion: implications for diagnosis and follow-up. *J Clin Endocrinol Metab*, 2002; **87**: 3537–42.
42. Ribeiro-Oliveira A, Jr., Abrantes MM, Barkan AL. Complex rhythmicity and age dependence of growth hormone secretion are preserved in patients with acromegaly: further evidence for a present hypothalamic control of pituitary somatotropinomas. *J Clin Endocrinol Metab*, 2013; **98**: 2959–66.
43. Clemmons DR, Van Wyk JJ, Ridgway EC, Kliman B, Kjellberg RN, Underwood LE. Evaluation of acromegaly by radioimmunoassay of somatomedin-C. *N Engl J Med*, 1979; **301**: 1138–42.
44. Barkan AL, Beitins IZ, Kelch RP. Plasma insulin-like growth factor-I/somatomedin-C in acromegaly: correlation with the degree of growth hormone hypersecretion. *J Clin Endocrinol Metab*, 1988; **67**: 69–73.
45. Anthony JR, Alwahab UA, Kanakiya NK, *et al*. Significant elevation of growth hormone level impacts surgical outcomes in acromegaly. *Endocr Pract*, 2015; **21**: 1001–9.
46. Micko AS, Wohrer A, Wolfsberger S, Knosp E. Invasion of the cavernous sinus space in pituitary adenomas: endoscopic verification and its correlation with an MRI-based classification. *J Neurosurg*, 2015; **122**: 803–11.

47. Buchfelder M, Schlaffer SM. The surgical treatment of acromegaly. *Pituitary*, 2017; **20**: 76–83.
48. Wang YY, Higham C, Kearney T, Davis JR, Trainer P, Gnanalingham KK. Acromegaly surgery in Manchester revisited—the impact of reducing surgeon numbers and the 2010 consensus guidelines for disease remission. *Clin Endocrinol (Oxf)*, 2012; **76**: 399–406.
49. Ahmed S, Elsheikh M, Stratton IM, Page RC, Adams CB, Wass JA. Outcome of transsphenoidal surgery for acromegaly and its relationship to surgical experience. *Clin Endocrinol (Oxf)*, 1999; **50**: 561–7.
50. Karavitaki N, Turner HE, Adams CB, *et al*. Surgical debulking of pituitary macroadenomas causing acromegaly improves control by lanreotide. *Clin Endocrinol (Oxf)*, 2008; **68**: 970–5.
51. Murray RD, Melmed S. A critical analysis of clinically available somatostatin analog formulations for therapy of acromegaly. *J Clin Endocrinol Metab*, 2008; **93**: 2957–68.
52. Bruns C, Lewis I, Briner U, Meno-Tetang G, Weckbecker G. SOM230: a novel somatostatin peptidomimetic with broad somatotropin release inhibiting factor (SRIF) receptor binding and a unique antisecretory profile. *Eur J Endocrinol*, 2002; **146**: 707–16.
53. Gunther T, Tulipano G, Dournaud P, *et al*. International union of basic and clinical pharmacology. CV. Somatostatin receptors: structure, function, ligands, and new nomenclature. *Pharmacol Rev*, 2018; **70**: 763–835.
54. Newman CB, Melmed S, George A, *et al*. Octreotide as primary therapy for acromegaly. *J Clin Endocrinol Metab*, 1998; **83**: 3034–40.
55. Mercado M, Borges F, Bouterfa H, *et al*. A prospective, multicentre study to investigate the efficacy, safety and tolerability of octreotide LAR (long-acting repeatable octreotide) in the primary therapy of patients with acromegaly. *Clin Endocrinol (Oxf)*, 2007; **66**: 859–68.
56. Colao A, Auriemma RS, Pivonello R. The effects of somatostatin analogue therapy on pituitary tumor volume in patients with acromegaly. *Pituitary*, 2016; **19**: 210–21.
57. Colao A, Bronstein MD, Freda P, *et al*. Pasireotide versus octreotide in acromegaly: a head-to-head superiority study. *J Clin Endocrinol Metab*, 2014; **99**: 791–9.
58. Carmichael JD, Bonert VS, Mirocha JM, Melmed S. The utility of oral glucose tolerance testing for diagnosis and assessment of treatment outcomes in 166 patients with acromegaly. *J Clin Endocrinol Metab*, 2009; **94**: 523–7.
59. Levy MJ, Matharu MS, Meeran K, Powell M, Goadsby PJ. The clinical characteristics of headache in patients with pituitary tumours. *Brain*, 2005; **128**(Pt 8): 1921–30.
60. Koop BL, Harris AG, Ezzat S. Effect of octreotide on glucose tolerance in acromegaly. *Eur J Endocrinol*, 1994; **130**: 581–6.
61. Gadelha MR, Bronstein MD, Brue T, *et al*. Pasireotide versus continued treatment with octreotide or lanreotide in patients with inadequately controlled acromegaly (PAOLA): a randomised, phase 3 trial. *Lancet Diabetes Endocrinol*, 2014; **2**: 875–84.
62. Paisley AN, Roberts ME, Trainer PJ. Withdrawal of somatostatin analogue therapy in patients with acromegaly is associated with an increased risk of acute biliary problems. *Clin Endocrinol (Oxf)*, 2007; **66**: 723–6.
63. Veysey MJ, Thomas LA, Mallet AI, *et al*. Prolonged large bowel transit increases serum deoxycholic acid: a risk factor for octreotide induced gallstones. *Gut*, 1999; **44**: 675–81.
64. Petrossians P, Borges-Martins L, Espinoza C, *et al*. Gross total resection or debulking of pituitary adenomas improves hormonal control of acromegaly by somatostatin analogs. *Eur J Endocrinol*, 2005; **152**: 61–6.
65. Fahlbusch R, Kleinberg D, Biller B, *et al*. Surgical debulking of pituitary adenomas improves responsiveness to octreotide lar in the treatment of acromegaly. *Pituitary*, 2017; **20**: 668–75.
66. Nunes VS, Correa JM, Puga ME, Silva EM, Boguszewski CL. Preoperative somatostatin analogues versus direct transsphenoidal surgery for newly-diagnosed acromegaly patients: a systematic review and meta-analysis using the GRADE system. *Pituitary*, 2015; **18**: 500–8.
67. Shen M, Shou X, Wang Y, *et al*. Effect of presurgical long-acting octreotide treatment in acromegaly patients with invasive pituitary macroadenomas: a prospective randomized study. *Endocr J*, 2010; **57**: 1035–44.
68. Fougner SL, Bollerslev J, Svartberg J, Oksnes M, Cooper J, Carlsen SM. Preoperative octreotide treatment of acromegaly: long-term results of a randomised controlled trial. *Eur J Endocrinol*, 2014; **171**: 229–35.
69. Sandret L, Maison P, Chanson P. Place of cabergoline in acromegaly: a meta-analysis. *J Clin Endocrinol Metab*, 2011; **96**: 1327–35.
70. Kuhn E, Chanson P. Cabergoline in acromegaly. *Pituitary*, 2017; **20**: 121–8.
71. Steeds R, Stiles C, Sharma V, Chambers J, Lloyd G, Drake W. Echocardiography and monitoring patients receiving dopamine agonist therapy for hyperprolactinaemia: a joint position statement of the British Society of Echocardiography, the British Heart Valve Society and the Society for Endocrinology. *Clin Endocrinol (Oxf)*, 2019; **90**: 662–9.
72. Tritos NA, Biller BM. Pegvisomant: a growth hormone receptor antagonist used in the treatment of acromegaly. *Pituitary*, 2017; **20**: 129–35.
73. Higham CE, Thomas JD, Bidlingmaier M, Drake WM, Trainer PJ. Successful use of weekly pegvisomant administration in patients with acromegaly. *Eur J Endocrinol*, 2009; **161**: 21–5.
74. Trainer PJ, Ezzat S, D'Souza GA, Layton G, Strasburger CJ. A randomized, controlled, multicentre trial comparing pegvisomant alone with combination therapy of pegvisomant and long-acting octreotide in patients with acromegaly. *Clin Endocrinol (Oxf)*, 2009; **71**: 549–57.
75. van der Lely AJ, Hutson RK, Trainer PJ, *et al*. Long-term treatment of acromegaly with pegvisomant, a growth hormone receptor antagonist. *Lancet*, 2001; **358**: 1754–9.
76. Minniti G, Scaringi C, Amelio D, Maurizi Enrici R. Stereotactic irradiation of GH-secreting pituitary adenomas. *Int J Endocrinol*, 2012; 2012: 482861.
77. Jenkins PJ, Bates P, Carson MN, Stewart PM, Wass JA. Conventional pituitary irradiation is effective in lowering serum growth hormone and insulin-like growth factor-I in patients with acromegaly. *J Clin Endocrinol Metab*, 2006; **91**: 123945.
78. Caron P, Broussaud S, Bertherat J, *et al*. Acromegaly and pregnancy: a retrospective multicenter study of 59 pregnancies in 46 women. *J Clin Endocrinol Metab*, 2010; **95**: 4680–7.

79. Abucham J, Bronstein MD, Dias ML. Management of endocrine disease: acromegaly and pregnancy: a contemporary review. *Eur J Endocrinol*, 2017; **177**: R1–R12.
80. van der Lely AJ, Gomez R, Heissler JF, *et al*. Pregnancy in acromegaly patients treated with pegvisomant. *Endocrine*, 2015; **49**: 769–73.
81. Marques P, Korbonits M. Pseudoacromegaly. *Front Neuroendocrinol*, 2019; **52**: 113–43.
82. Joustra SD, Roelfsema F, van Trotsenburg ASP, Schneider HJ, Kosilek RP, Kroon HM, *et al*. IGSF1 deficiency results in human and murine somatotrope neurosecretory hyperfunction. *J Clin Endocrinol Metab*. 2020; **105**(3).

2.3.11 Clinically Non-Functioning Pituitary Tumours and Gonadotropinomas

Nienke Biermasz and Wouter R. van Furth

Introduction

Non-functioning pituitary tumours (NFPAs) are pituitary adenomas without clinically overt hormone secretion. The vast majority of NFPAs (80%) arises from gonadotroph cells and are monoclonal expansions. They show positive staining for glycoprotein hormones, usually luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Only 20% of adenomas have no evidence of any hormone staining and this subgroup may be called 'null-cell tumour'.

In this chapter the term 'clinically non-functioning pituitary adenomas' is used to describe pituitary tumours, which in most instances produce no, or low quantities of hormones causing no clinically recognizable symptomatology. In the very few instances, in which such tumours produce intact gonadotropins that activate testicular or ovarian activity, the term 'gonadotropinomas' is used.

Box 2.3.11.1 Presenting features of clinically non-functioning pituitary adenomas

Compression of the optic chiasm (bitemporal hemianopsia, decreased visual acuity, nerve palsy—rare)
Hypopituitarism (e.g. fatigue, hypogonadism, Addisonian crisis)
Galactorrhoea
Chronic headache
Liquorrhoea (rare)
Acute headache with neurological symptoms (rare, in case of apoplexia)
Incidental finding

Clinical Presentation

The clinical presentation is highly variable, from completely asymptomatic to severe symptoms (see Box 2.3.11.1). Due to the lack of hypersecretion of hormones and due to a slow growth pattern, non-functioning pituitary adenomas classically present due to mass effects and compression of surrounding tissues. Patients typically report that in hindsight, symptoms developed gradually in the years prior to diagnosis and a delay in diagnosis is usually at least 1–2 years.

Hypopituitarism, of the anterior lobe, due to compression may give rise to a variety of symptoms (e.g. fatigue, amenorrhea, infertility, hypogonadism). The classical order in which deficiencies occur is thought to be GH, LH, FSH, thyroid-stimulating hormone (TSH), and ultimately adrenocorticotropic hormone (ACTH). However, in individual cases, all axes need to be tested.

Galactorrhoea by hyperprolactinemia can result from mass effects of the NFPA with compression of the pituitary stalk and the loss of inhibition of the dopaminergic tone. Prolactin values due to stalk compression are typically only slightly elevated (2–5 times ULN), in contrast to clearly elevated levels in case of (macro)prolactinoma.

Suprasellar extension may lead to compression of the optic chiasm and optic nerves resulting in the classical bitemporal hemianopsia and ultimately decreased visual acuity or blindness.

Headache can be a symptom of pressure, typically midline frontally, however, there is an extensive differential diagnosis. Acute onset, severe headache can be a presenting symptom of apoplexia, an ischaemic or haemorrhagic event, in a pre-existing pituitary adenoma.

Lateral, parasellar extension in the cavernous sinus, may result in dysfunction of the third, fourth and sixth cranial nerve resulting in ophthalmoplegia (diplopia and ptosis), which is rare outside apoplexia or aggressively growing tumours. Trigeminal neuralgia is exceedingly rare.

Inferior extension may result from invasive growth and destruction of the sellar floor and growth into the sphenoid sinus. An exceptional complication is liquorrhoea.

Pituitary incidentalomas are pituitary lesions without clinical features of pituitary disease, discovered on imaging done for unrelated reasons. With the advancing radiological techniques and scan indications, they are increasingly found. Incidentalomas can be microadenoma, macroadenoma, or invasive macroadenoma. They also need diagnostic work-up to exclude hormone overproduction and to assess indications for treatment.

Non-functioning adenoma may be 'giant' (>4 cm) and sporadically tumours of exceptional size may give rise to neurological symptomatology (e.g. temporal lobe epilepsy), obstruction hydrocephalus and, for example, hypothalamic dysfunction.

Epidemiology

Non-functioning pituitary adenoma account for 14–54% of pituitary adenoma depending on the population studied. The prevalence is 7–41.3/100 000 and standardized incidence rate 0.65–2.34/100 000. The peak occurrence is in middle age and elderly (40–80 years) [1]. It is the second most prevalent pituitary tumour after prolactinoma. However, when macroadenomas are considered there is a preponderance of NFPA over prolactinoma.

Pituitary incidentalomas are quite prevalent, in unselected populations they are reported to be 10–38% for microadenoma and 0.16–0.3% for macroadenoma on MRI scanning and 11% in autopsy series. So, the vast majority of incidentalomas is small, not hormone producing, and will not require treatment, however, needs follow-up and therefore burden for patients and healthcare system.

A small subset of NFPAs are a manifestation of a hereditary syndrome (MEN1, familial isolated pituitary adenoma, SDH). The majority of pituitary adenoma of hereditary syndromes are functioning lesions. Systematic screening with pituitary scans of genetic carriers will reveal more non-functioning microadenomas. The presence of a hereditary syndrome should be considered in younger patients and those with a positive family history. Non-functioning pituitary tumours in children are extremely rare.

The growth pattern of NFPA is highly variable and considered non-linear. Generally, it can be stated that the majority of macroadenomas will show gradual growth in years, resulting in treatment indication to prevent symptoms by compression of optic chiasm over time. In contrast, only a small minority of microadenomas will show clinically relevant growth.

Pathogenesis and Aetiology

In the late 1970s and early 1980s, much work has been done on defining the pathological properties of pituitary tumours. The work of Asa and Kovacs is specially known for the accurate description of the microscopic findings of non-functioning pituitary adenomas [2]. These tumours are morphologically classified into two groups, those which have hormone immunoreactivity and ultrastructural features of known adenohypophyseal cell types but are clinically silent, and those composed of cells that do not resemble non-tumorous adenohypophyseal cell types (null-cell adenomas). It is now known that non-functioning adenomas represent a heterogeneous group with staining by immunocytochemistry in the majority of cases. The pathogenesis is complex and include genetic and epigenetic events, hormonal stimulation, growth factor overproduction, pituitary stem cells derangements, microRNA deregulation promoting cell growth and proliferation [3].

Gene mutations involved are MEN1 gene, the tumour suppressor gene CDKN1B (MEN4) and the oncogene AIP.

Diagnostic Evaluation

Most sellar masses will be pituitary adenomas. However, many other neoplastic, inflammatory, infectious, and vascular lesions, may affect the sellar region and mimic pituitary tumours. These lesions must be considered in a differential diagnosis of non-functioning pituitary adenomas. The diagnosis of such lesions involves a multidisciplinary approach, and detailed endocrinological, ophthalmological, neuroimaging, neurological, and finally histological studies may be required.

Radiological Investigations

MRI dedicated to the pituitary region with T1- and T2-weighted images and gadolinium contrast series is the gold standard for evaluation and differential diagnosis of sellar and suprasellar masses. A pituitary adenoma usually appears hypo- or isointense compared to pituitary tissue on T1. Acute haemorrhage will appear hyperintensive. Cystic lesions can be identified.

The size and extension of the adenoma are evaluated with the Hardy-Wilson classification (microadenoma (I), macroadenoma (II), invasive adenoma into sellar floor (III/IV)) and suprasellar (A, B, C) and parasellar (D, E) extension. The Knosp classification [4] evaluates the cavernous sinus invasion in relation to the carotid artery, which is related to the chance of success of surgical resection, particularly relevant for hormone producing tumours. The relation of the tumour with and the severity of displacement of the optic chiasm are described. An important feature of pituitary macroadenoma is an enlarged sella. Radiological differential diagnosis of other sellar masses are for example craniopharyngioma, meningioma, cysts, hypophysitis, germ cell tumour, aneurysm.

Neuro-Ophthalmological Investigations

In particular, in NFPA in the vicinity of the optic chiasm or optic nerves, careful ophthalmological investigation is warranted [5]. Visual symptoms include visual field defects, loss of central vision, and motility problems. Evaluation of visual fields by automated static perimetry, visual acuity, and assessment by an ophthalmologist is the cornerstone of evaluation. Usually abnormalities in these assessments precede clinical symptoms. Other assessments include functional (visual evoked potential (VEP)) and anatomic (disc appearance and optical coherence tomography (OCT)). The findings of these modalities will be concordant in many cases, however, there may be disparities due to the innate differences in tests. Large-scale data on prediction of visual recovery after chiasm decompression by NFPA resection of the newer tests (e.g. OCT), are not yet available. Postoperatively, tests are repeated to document recovery up to stabilization with 3 to 6 monthly intervals and thereafter less frequently.

Endocrine Diagnosis

The endocrine work-up of NFPA is directed to the exclusion of hormone oversecretion, including a high dose hook effect to exclude prolactinoma, on the one hand, and detection of hypopituitarism on the other hand.

Initial assessment includes: IGF-I, cortisol, prolactin, FSH/LH, oestradiol or testosterone, Ft4 and TSH. In emergency clinical practice, basal evaluation of adrenal and thyroid function may suffice as preoperative evaluation, and evaluation of growth hormone deficiency and dynamic corticotrope function is postponed to the postoperative situation [1].

Histopathological Classification

The World Health Organization (WHO) classification of pituitary adenomas was revised in 2017 and a new name of PitNETs (pituitary neuroendocrine tumours) was proposed [6].

Non-functioning pituitary tumours represent a heterogeneous group, and they are being considered as a variant of functioning tumours with the exception of the null-cell adenomas without any hormone staining.

Based on the expression of anterior pituitary hormones and pituitary-specific transcription factors, gonadotroph tumours dominate within the group of clinically non-functioning tumours, followed by corticotroph type; however, also other less common types of the non-functioning tumours can be identified on the basis of staining.

Assessment of mitotic account, proliferative Ki-67 index, and tumour invasiveness is important to identify potentially aggressive tumours. Application of pituitary-specific transcription factors (Pit-1, T-Pit, SF1) plays an important role in recognition of less differentiated tumour types, some of which may demonstrate aggressive behaviour.

Predictive markers may be included in the classification (somatostatin receptors, dopamine receptors, low MGMT expression), particularly when pharmacological therapy is under consideration. There are three categories: typical adenoma (most frequent), high risk pituitary adenoma, pituitary carcinoma (exceptional) There are eight subtypes based on the immunohistochemical expression and pituitary-specific transcription factors.

First Management

The aims of treatment will differ depending on the clinical situation, individual patient's characteristics, preferences, and symptomatology on the one hand, and tumour size and extension, compression of critical structures, and growth rate on the other hand.

Aims for treatment may be symptom reduction: visual field deficits, headache; prevention of symptoms in case of tumour growth approaching the visual system, or future pregnancy wish; and/or diagnostic, to obtain histopathological diagnosis.

Therapeutic modalities for non-functioning pituitary adenoma include observation (wait and scan), surgery, radiotherapy, and medical therapy (see Figure 2.3.11.1).

Observation

It is good to bear in mind the natural course of growth of NFPA. Microadenomas (<10 mm) rarely grow (10%) and convert to a macroadenoma. Macroadenomas, in contrast, tend to grow and the tumour volume of macroadenomas increases gradually in approximately 25–50% of patients [7, 8]. Radiological surveillance could be performed annually for 3 years and thereafter with longer intervals (2–5 years). Those larger adenomas (>2 cm) and those abutting the chiasm should be followed more closely and also with visual field tests. In case of growth or visual field defects surgery should be considered.

Surgery

Indications for surgical resection are: mass effect resulting in compression and functional deterioration of the optic nerves, chiasm, or other neurological structures, (growing) tumours with anatomic relation threatening the optic chiasm, pituitary apoplexia resulting in the aforementioned deficits or extreme headache [9–12]. Relative indications include pituitary insufficiency and headache potentially caused by the mass effect. In recent years, several guidelines on surgical management have been published. Urgent indications are severe and progressive visual symptoms and apoplexia with intractable headache and compressive symptoms. Major goals of surgery are preserving pituitary function.

Pituitary surgery is primarily performed by the transnasal transsphenoidal approach (Figure 2.3.11.2) with endoscopic or

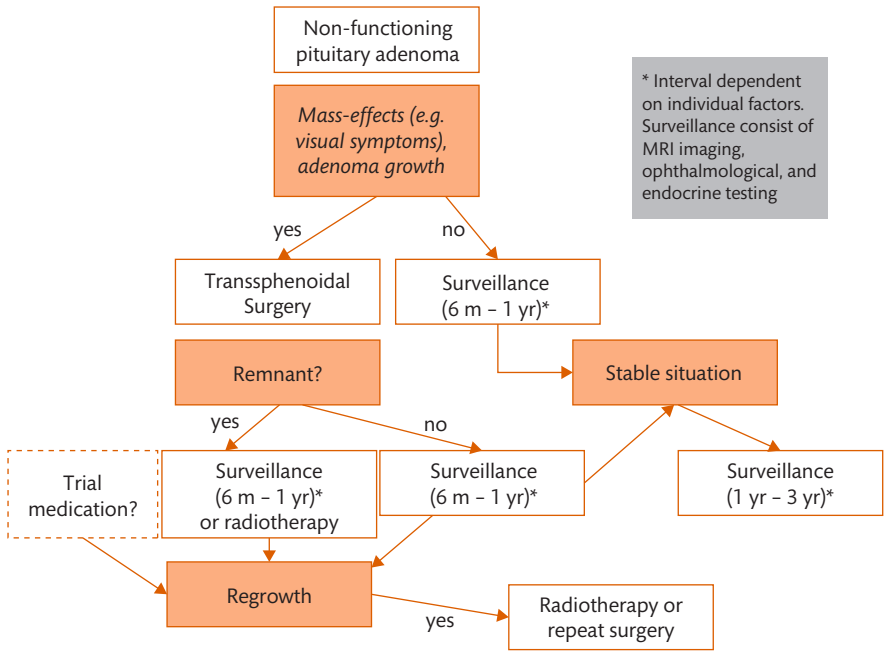


Figure 2.3.11.1 Treatment and surveillance algorithm for non-functioning pituitary adenoma.

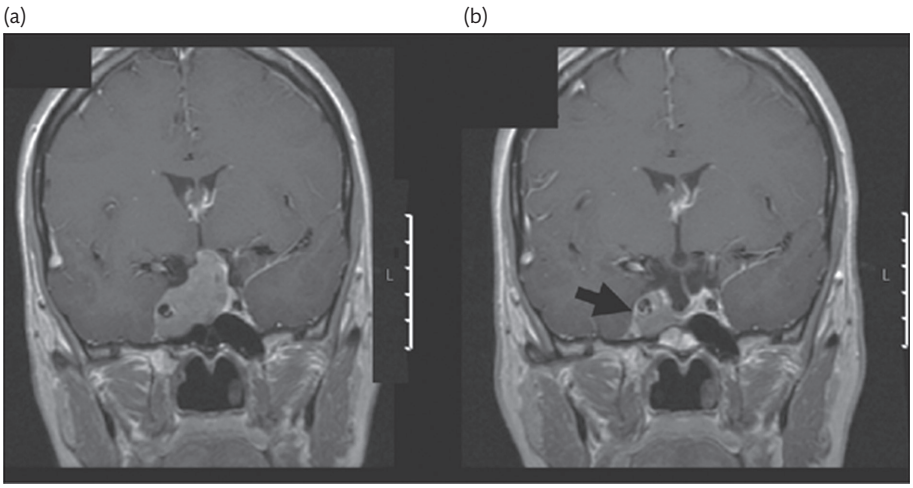


Figure 2.3.11.2 A 48-year-old woman with suspicion of a clinically non-functioning pituitary macroadenoma with compression of the optic chiasm leading to impaired visual acuity and bitemporal visual field defects. Postoperatively, visual function recovered while anterior pituitary function remained intact. (a) Baseline MR image. (b) Following transsphenoidal subtotal resection. Note the tumour remnant encasing the right carotid artery (arrow). However, significant tumour reduction has been achieved with complete decompression of the optic chiasm.

microscopic visualization. Advantages of endoscopic surgery are a wider and closer view of the surgical area, also of the supra- and parasellar regions, with less nasal trauma, and usual no need for postoperative nasal packing. During the intervention, the sphenoid sinus is entered and subsequently the sellar bone is opened. Also tumours with large suprasellar and parasellar compartments can successfully be resected via the transsphenoidal route. Tumours usually have a soft consistency and blunt curettes are used for removal of those parts. Leakage of cerebral spinal fluid during surgery needs to be anticipated. This is depending on the tumour size and invasion; however, it may occur in any surgery. There are multiple techniques available to treat leaks, including abdominal fat, fascia lata implants, and nasal septal flaps.

A transcranial approach via the pterional or subfrontal route is needed in exceptional cases with a dominant extrasellar tumour compartment and a small sella turcica, a large eccentric tumour extension into the middle, anterior, or posterior cranial fossa, or a coexisting aneurysm of the carotid artery. Combination of transcranial and trans-sphenoidal approach may be indicated for giant adenomas (>4 cm diameter).

The trans-sphenoidal surgery is usually very well tolerated, with some nasal morbidity in the early postoperative phase. The perioperative period is best done by a multidisciplinary team, involving both endocrinologist and neurosurgeon. Main complication to monitor is disturbance of fluid balance by evaluation of urine output and gravity, plasma sodium concentrations, and patient thirst experience because of transient diabetes insipidus or postoperative hyponatraemia that may occur up to 10 days. Adrenal function and other pituitary function require careful monitoring. Depending on tumour size and preoperative situation the chance for new onset hypopituitarism is considered to be about 1–5%. Length of hospital stay is about 2–5 days. Severe complications like intratumoral bleeding or meningitis are very rare and depend mainly on tumour size. Surgical mortality is less than 1% in experienced hands.

After trans-sphenoidal resection, improvement of visual function is achieved in 60–90% of patients. The rapidity of visual recovery depends in part on the duration of optic nerve compression. Recovery of visual function can already be observed in the first days postoperatively and can continue up to a year after surgery. In patients with pre-existing (partial) hypopituitarism, pituitary function is not likely to restore after resection of the adenoma, although in some patients improvement can be demonstrated, and figures vary between 2% and 50%. Headache may improve in 80–100% of cases operated. Still, performing an operation for headache alone is controversial since headache may be unrelated to pituitary adenoma. Gross total resection of tumour can be achieved in 50–70% of patients, depending on tumour size and extension. Regrowth op tumour remnant is described in up to 20% patients, for which repeat surgery can be considered or alternatively radiotherapy.

Radiotherapy

Complete removal of a macroadenoma cannot be achieved in all patients and up to 40–60% of patients will have a small or a clear remnant post-surgery. Since the goal of surgery is often to decompress vital structures, while preserving the pituitary gland, surgical strategy is to avoid risks associated with resection of the most distal parasellar tumour parts. In addition, some patients (up to 10–50%) particular those with a remnant in situ will face regrowth of tumour in time and will require additional therapy.

External beam radiotherapy has been used for pituitary adenoma for 50 years [13]. There is no doubt that excellent tumour volume control can be established with irradiation in NFPA. However, toxicity of radiotherapy is not negligible and therefore the indications to irradiate these benign, usually slow growing remnants, need to be carefully weighed against the negative effects.

So, the optimal treatment strategy of patients with a residual tumour after trans-sphenoidal surgery (example shown in Figure 2.3.11.2b) is still a matter of debate. Based on observational studies on the natural course of macroadenoma remnants it is clear that tumour remnant growth or recurrence may occur in long-term follow-up, usually at a slow growth rate and that there are only some predictive factors such as parasellar invasion before surgery and (the degree of) suprasellar extension of the postoperative remnant adenoma, but that we are unable to predict reliably for clinical practice. Unfortunately,

no morphological tumour features or molecular markers of cell proliferation are available that predict tumour growth.

Postoperative external radiotherapy is applied in non-functioning pituitary adenoma to achieve long-term tumour control by induction of tumour shrinkage or stabilization. In patients who receive conventional radiotherapy, progression free survival at 10 years is more than 90%, significantly higher compared to patients who were observed alone. Overall, no randomized trials have been performed that compare radiotherapy with a wait and see policy. It is important to note that radiotherapy cannot control tumour growth in each patient. Studies vary with respect to follow-up duration, number of invasive tumours, and other aspects, but regrowth of residual adenomas in patients treated with radiotherapy has been observed, so radiological surveillance remains to be done also after radiotherapy. Adjuvant radiotherapy has beneficial effects on tumour regrowth but this should be balanced against the complications of radiotherapy (i.e. radiation damage to the optic nerves with visual impairment, development of (partial) hypopituitarism in up to 50–75% of patients, increased risk of cerebrovascular events, and an increased risk on secondary brain tumours).

Therefore, the treatment strategy in patients with a residual adenoma after trans-sphenoidal surgery should be individualized and factors such as age, comorbidity, remnant tumour size, tumour distance to the optic chiasm, and status of pituitary function should be involved in the decision on adjuvant radiotherapy. Patients not treated with radiotherapy should carefully be observed with MRI and ophthalmological evaluation. If regrowth occurs, radiotherapy is still effective, but repeat surgery can also be considered. Some centres generally favour a strategy to irradiate remnants, while others wait for progression before radiotherapy is considered. The former has the advantage that radiological surveillance will need to be less stringent and critical but that side effects needs to be taken into account. The latter requires critical evaluation of future scans in a multidisciplinary team with expert neuroradiological expertise setting to decide on the scan intervals and the best moment to intervene in case of significant remnant growth. There is no clear advantage of any of the policies with respect to tumour growth control. Reasonable candidates for the observational approach are patients that present with 'not too big' tumours at presentation, small residual lesions, and no signs of aggressive growth. In patients with larger remnants, who had bigger tumours at presentations, or have suspected faster growth rate, it is probably wise to consider earlier radiotherapy.

There are various radiation techniques and innovations in the field, such as better imaging, better software for planning of dose delivery, probably have an improved balance between efficacy and safety because the surrounding tissues get less radiation. 3D conformal radiotherapy has evolved to stereotactic radiotherapy. Fractionated stereotactic radiotherapy is delivered in many small fractions for 20–25 days. Stereotactic radiosurgery with a Cobalt-60 gamma knife or a cyberknife or LINAC (linear accelerator), delivers the full dose in a single session. However, long-term results of new techniques need to be awaited to draw conclusions on long-term safety. Proton beam radiation of pituitary adenomas have been done in some centres, showing similar outcomes compared to photon radiotherapy.

Radiosurgery is suitable for small adenomas with sufficient distance to the optic nerves and optic chiasm, which are radiosensitive tissues. Fractionated stereotactic radiotherapy can be applied in larger tumours and tumours with a closer proximity to the optic chiasm. With both forms of stereotactic radiotherapy tumour control can be achieved in more than 90%. No data are available yet that compare recurrence rates and long-term safety and complications of conventional radiotherapy and stereotactic radiotherapy. There are no comparative studies between stereotactic fractionated radiotherapy and radiosurgery.

Medical Treatment

In other pituitary adenoma, particular in somatotropinoma and prolactinoma, there is a clear role for medical treatment for both tumour volume control, if needed, and control of hormone oversecretion. The role of medical treatment in NFPA is not established [14]. Potentially available drugs are dopamine agonists, somatostatin analogues, and temozolomide. When used in NFPA, it is off-label use.

The treatment goal of medical treatment could be tumour growth stabilization, or shrinkage, in the presence of few side effects and with acceptable costs, since long-term treatment is anticipated. Research in this field is difficult because of the variable and slow growth pattern of tumours and remnants requiring long-term studies to draw conclusions. Nevertheless, there is a clinical need for medical strategies to control NFPA, because of the limitations of surgery (resection of difficult parts of tumour in cavernous sinus and risk of complications) and radiotherapy (late toxicity particular induction of hypopituitarism and limited dose application once the tumour is in the vicinity of the optic nerves). Several different drugs (combinations) have been investigated. In the presence of a new classification for pituitary adenomas, in which the majority of NFPA will harbour some hormone staining, and receptors for medical treatment (dopamine receptors and somatostatin analogue receptor subtypes can be assessed, rationale molecular based medical treatment could be anticipated); however, at present no large-scale supporting data are available.

Based on the available literature, mainly uncontrolled studies, dopamine agonists have the best available evidence that they have a positive effect on tumour stabilization and are generally well tolerated. Some advocate a role for dopamine agonists (DAs) in patients with a postoperative remnant, others consider Das in selected cases with observed growth, there is no consensus.

Dopamine Agonists

In physiology, dopamine exerts tonic inhibition on prolactin-producing cells via the D2-receptor. Many pituitary adenomas, also NFPA express D2 receptors. Both tumour volume and hormone oversecretion of prolactinomas is very responsive to DA. It is reasonable to think that also other tumours expressing D2-receptors could respond as well to DA. There are several Das; bromocriptine and quinagolide, and the more potent specific D2 receptor agonist cabergoline.

The largest study to date included 79 patients [15], treated with a target dose of 2 mg/week and two-thirds of patients were treated. Tumour control was achieved in 87% (DA) vs. 47% (control), and DA decreased the need for additional treatment to 13% (DA) vs. 42% (control). Less favourable results were obtained in younger and male patients and those with a large preoperative tumour. D2R expression was not associated with responsiveness and so, there are no predictors of responsiveness to DA.

If DA is indicated, a 2 mg/wk cabergoline dose is most often used, aiming at stabilization of tumour size and so avoiding additional treatments (e.g. repeat surgery and radiotherapy). Logically, careful radiological surveillance is still needed, at an interval of 6 months for 2 years or so and thereafter yearly. A higher dose is not reasonable based on the limited data available. In patients who experience side effects, orthostatic hypotension, or psychiatric side effects, alternative treatment options are preferable especially because of the lack of firm evidence.

Somatostatin Analogues

Somatostatin analogues have antiproliferative and antisecretory effects through binding to one or more of the five subtypes of the somatostatin receptor. Lanreotide and octreotide are preparations that preferentially bind to the SSTR2. Main indication is acromegaly, but also TSH adenomas can be treated with somatostatin analogues. Not all tumours express SSTR2 and a substantial number of patients will be partially or fully resistant to these drugs. Pasireotide is a next generation somatostatin analogue that binds somatostatin receptor (SSTR) more universally, and especially the binding to the SSTR5. It has been shown to be effective in acromegaly and in Cushing, albeit with the drawback of a high incidence of diabetes mellitus.

Somatostatin receptors are expressed in non-functioning pituitary adenoma, with a predominance of the somatostatin receptor subtype 3 (sst_3), followed by sst_2 and sst_5. Although *in vitro* there was a significant antiproliferative and antisecretory effect of somatostatin in NFPA, clinical treatment has not shown to be efficient. Small and short-term studies showed a 12% tumour size reduction, 5% growth, and an 83% stable tumour mass, which is inconclusive in view of a study duration of 6 months on average. Preselection of patients for example based on positive octreotide receptor scintigraphy could be a strategy to select patients with better response rate; however, larger scale, long-term, and well-designed studies are needed to justify this form of (expensive) treatment for this indication. There are no clinical studies on pasireotide treatment and the safety profile is not attractive in a long-term treatment setting as for NFPA.

Gonadotropin-Releasing Hormone Agonists and Antagonists

The release of gonadotrophins by normal anterior pituitary cells is regulated by pulsatile secretion of gonadotropin-releasing hormone by the hypothalamus. The chronic administration of gonadotropin-releasing hormone to normal individuals produces an initial rise in gonadotropin levels, followed by gonadotroph desensitization, leading to efficient suppression of gonadotrophin release. Several case reports describe the occurrence of pituitary apoplexy after the administration of gonadotropin-releasing hormone as a test agent or as an agonist for the treatment of prostate cancer. Long-term treatment with gonadotropin-releasing hormone analogues had no effect on tumour size or visual fields in patients with non-functioning pituitary adenoma. This treatment modality is currently not anymore under clinical investigation.

Temozolomide

Temozolomide has been proposed as a treatment option for pituitary carcinomas and aggressive pituitary adenomas. The clinical experience is rather limited, but this is the modality of choice in exceptional pituitary adenomas with aggressive behaviour. Overall response rate is 22–50%. Non-functioning pituitary tumours tend to respond less than prolactinomas and corticotroph tumours [16], with partial complete response in 22% (6/27 patients), stable disease in 48% (13/27), and progression in 30% (8/27) according to a recent review of literature. In a recent survey of the European Society of Endocrinology, 47 patients with clinical non-functioning tumours were reported: 2% had complete regression, 15% had partial regression, 49% had stable disease, and 34% progression [17].

Management of Remnants/Regrowth

Radical removal of NFPA cannot be accomplished in all patients, for example because of growth into the cavernous sinus. There is a delicate balance between trying to be as radical as possible versus sparing normal pituitary gland and complication avoidance in general. In clinical practice, the postoperative situation is usually evaluated 3 to 6 months post-surgery. Postoperative remnants are identified, although small remnants may be difficult to distinguish from postoperative changes. This initial postoperative MRI scan is used as a reference scan. All patients need radiological surveillance, since there is a chance for regrowth in the years after surgery, especially in non-irradiated patients. In some centres, postoperative radiotherapy is applied on remnants, which is effective in preventing regrowth. However, it has side effects (e.g. common hypopituitarism) and rarely cerebrovascular disease, secondary brain tumour development, neurocognitive dysfunction, and so on.

In an off-label setting for medical treatment, primarily Das are considered as preventive medication for recurrence; however, these strategies are expert-based and not evidence-based.

The postoperative gross total resection rate is about 50%, so 50% of patients will have some remnant. Regrowth is observed in 10–50% of patients in long-term follow-up. Predictive factors of regrowth are presence of a remnant, young age, initial tumour size, especially invasive growth pattern. Tumour growth-free survival is twice as low in patients without versus with a remnant. Residual tumour doubling time has been investigated to be 3.4 years, confirming the need for long-term monitoring. It is important that new scans are not only compared with the previous scan but always also with the initial postoperative, reference scan. Tumour volume doubles times is a measure of growth velocity, the time it takes for a tumour to double in volume [18]. Pituitary tumours may show either an exponential growth, or linear or logistic growth. Exponential growth is usually more rapidly.

Pathological subtype is also a potential predictive factor to consider. Silent corticotrophtumours are considered to be more aggressive. With respect to the pathological classification the Ki-67 index is most used, with a cut-off of 3% revealing unusually high proliferation index and 10% being suspicious for malignant behaviour. Studies on gene expression have not been able to identify a set of prognostic markers for NFPA. The oestrogen receptor alpha is of potential interest because the suggestion that males with low expression had higher reintervention rates. In summary, for clinical practice these developments have not yet proven their benefit. So, patients operated for NFPA, especially those with remnants require long-term radiological surveillance [19].

Clinical Outcome of Patients

In general patients with a pituitary condition have significant morbidity, increased mortality, and decreased quality of life [20, 21]. Although this may be more pronounced in patients with functioning adenoma, also NFPA will have long-term consequences of their disease. From literature it is thought that intervention is able to improve quality of life scores, but that impairments persist despite optimal treatment including hormone replacement strategies. We assume hypopituitarism as important factor determining clinical outcome, in particular adrenal insufficiency. However, also visual impairments, headache, and psychosocial factors may be of relevance. To ensure optimal outcome a multidisciplinary team is required to address patient's issues not only at the medical level but also addressing psychosocioeconomical issues.

Gonadotropinomas

Hyperstimulation by excessive FSH secretion of gonadotropinomas has been described in only a few patients. An example of this is the observation of high serum FSH concentrations, but with normal luteinizing hormone and testosterone and large testes in four men with pituitary macroadenomas. After pituitary surgery there were decreases in serum gonadotropin and testosterone levels, which were accompanied by decreases in testicular volume. In females, a similar example was the description of a woman whose gonadotroph adenoma caused supranormal serum concentrations of FSH, which resulted in the development of multiple ovarian cysts, persistent elevation of her serum oestradiol concentration, and endometrial hyperplasia.

Future Perspective and Research Agenda

The variation in clinical course and presentation of non-functioning pituitary adenoma is tremendous (e.g. the new WHO classification in which only a small subset of clinically non-functioning pituitary adenomas are null-cell adenomas, and most have variable immune staining). The association with subtle clinical hormone secretion, clinical course, and presentation needs further research. A change in policy to be more restrictive in postoperative radiotherapy poses challenges to the multidisciplinary team, since careful surveillance of non-irradiated remnants is key in the high-quality care of this patient group.

New drug options are interesting but require long-term follow-up to evaluate added value; and although ideally randomized trials are needed, the variable and slow growth pattern will need large and long-term studies. In this respect, future registries for example on aggressive tumours may help to answer their questions.

REFERENCES

1. Ntali G, Wass JA. Epidemiology, clinical presentation and diagnosis of non-functioning pituitary adenomas. *Pituitary*, 2018; **21**: 111–18.
2. Asa SL, Kovacs K. Clinically non-functioning human pituitary adenomas. *Can J Neurol Sci*, 1992; **19**: 228–35.
3. Zatelli MC. Pathogenesis of non-functioning pituitary adenomas. *Pituitary*, 2018; **21**: 130–7.
4. Micko AS, Wohrer A, Wolfsberger S, Knosp E. Invasion of the cavernous sinus space in pituitary adenomas: endoscopic verification and its correlation with an MRI-based classification. *J Neurosurg*, 2015; **122**: 803–11.
5. Newman SA, Turbin RE, Bodach ME, *et al*. Congress of neurological surgeons systematic review and evidence-based guideline on pretreatment ophthalmology evaluation in patients with suspected nonfunctioning pituitary adenomas. *Neurosurgery*, 2016; **79**: E530–2.
6. Manojlovic-Gacic E, Engstrom BE, Casar-Borota O. Histopathological classification of non-functioning pituitary neuroendocrine tumors. *Pituitary*, 2018; **21**: 119–29.
7. Dekkers OM, Hammer S, de Keizer RJ, *et al*. The natural course of non-functioning pituitary macroadenomas. *Eur J Endocrinol*, 2007; **156**: 217–24.
8. Karavitaki N, Collison K, Halliday J, *et al*. What is the natural history of nonoperated nonfunctioning pituitary adenomas? *Clin Endocrinol (Oxf)*, 2007; **67**: 938–43.
9. Penn DL, Burke WT, Laws ER. Management of non-functioning pituitary adenomas: surgery. *Pituitary*, 2018; **21**: 145–53.
10. Kuo JS, Barkhoudarian G, Farrell CJ, *et al*. Congress of neurological surgeons systematic review and evidence-based guideline on surgical techniques and technologies for the management of patients with nonfunctioning pituitary adenomas. *Neurosurgery*, 2016; **79**: E536–8.
11. Lucas JW, Bodach ME, Tumialan LM, *et al*. Congress of neurological surgeons systematic review and evidence-based guideline on primary management of patients with nonfunctioning pituitary adenomas. *Neurosurgery*, 2016; **79**: E533–5.
12. Freda PU, Beckers AM, Katznelson L, *et al*. Pituitary incidentaloma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 894–904.
13. Minniti G, Flickinger J, Tolu B, Paolini S. Management of nonfunctioning pituitary tumors: radiotherapy. *Pituitary*, 2018; **21**: 154–61.
14. Even-Zohar N, Greenman Y. Management of NFAs: medical treatment. *Pituitary*, 2018; **21**: 168–75.
15. Greenman Y, Cooper O, Yaish I, *et al*. Treatment of clinically nonfunctioning pituitary adenomas with dopamine agonists. *Eur J Endocrinol*, 2016; **175**: 63–72.
16. Halevy C, Whitelaw BC. How effective is temozolomide for treating pituitary tumours and when should it be used? *Pituitary*, 2017; **20**: 261–6.
17. McCormack A, Dekkers OM, Petersenn S, *et al*. Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*, 2018; **178**: 265–76.
18. Oystese KA, Evang JA, Bollerslev J. Non-functioning pituitary adenomas: growth and aggressiveness. *Endocrine*, 2016; **53**: 28–34.
19. Reddy R, Cudlip S, Byrne JV, Karavitaki N, Wass JA. Can we ever stop imaging in surgically treated and radiotherapy-naive patients with non-functioning pituitary adenoma? *Eur J Endocrinol*, 2011; **165**: 739–44.
20. Andela CD, Lobatto DJ, Pereira AM, van Furth WR, Biermasz NR. How non-functioning pituitary adenomas can affect health-related quality of life: a conceptual model and literature review. *Pituitary*, 2018; **21**: 208–16.

21. Ntali G, Capatina C, Fazal-Sanderson V, *et al*. Mortality in patients with non-functioning pituitary adenoma is increased: systematic analysis of 546 cases with long follow-up. *Eur J Endocrinol*, 2016; **174**: 137–45.

2.3.12 Thyrotropinomas

Mark Gurnell, Olympia Koulouri, and Waiel Bashari

Introduction

Thyrotropin (TSH)-secreting pituitary adenomas (also known as thyrotropinomas or TSHomas) have traditionally been considered rare tumours, sometimes referred to as the 'one in a million pituitary adenoma', and accounting for approximately 1% of all pituitary tumours in historical surgical and pathological series [1]. However, just as the overall prevalence of clinically relevant pituitary adenomas has been shown to be higher than previously suspected (between 1:1000 and 1:1500 of the general population), findings from a national Swedish registry covering the interval 1990 to 2010 indicate (i) an incidence rate of 0.15 and (ii) a prevalence rate of 2.8 clinically overt thyrotropinomas per million population [2]; interestingly, the detection rate increased fivefold during the 20 year study interval (1991–2011). Consistent with this, following the first case report in 1960, fewer than 300 cases had been described by the mid-1990s; however, in the last two decades TSH-secreting pituitary adenomas have been recognized and reported with increasing frequency. Several factors are likely to account for this, including: (i) availability in routine clinical laboratories of ultrasensitive and highly specific TSH assays; (ii) increasing use of combined measurement of free thyroid hormone and TSH levels to screen for suspected thyroid dysfunction; (iii) improvements in (together with more widespread use of) pituitary magnetic resonance (MR) imaging and (iv) increased awareness of the condition. In parallel, a change in the proportion of microadenomas versus macroadenomas has emerged. Early case series reported macroadenomas in more than 80% of patients with TSH-secreting tumours, with up two-thirds manifesting extrasellar extension. However, recent studies indicate that micro-TSHomas are being increasingly diagnosed [2, 3] and, with the advent of more sensitive cross-sectional and alternative (e.g. positron emission tomography (PET)) techniques for detecting small functioning tumours [4], it seems likely that the balance will be even further redressed. The majority of patients are diagnosed in middle age (with both genders equally affected), but case series involving both paediatric and elderly (>80 years) subjects have been described [5].

Increased awareness among clinicians, coupled with improved diagnostics, has the potential to reduce the delay between the onset of hyperthyroidism and definitive treatment, thus mitigating the risk of complications such as atrial fibrillation, osteoporosis, and compression of the optic chiasm. At the same time, it increases the likelihood that a surgically-resectable adenoma will be identified. Equally as important however, is the need to reliably distinguish central hyperthyroidism due to a thyrotropinoma from the much more common causes of primary hyperthyroidism (e.g. Graves' disease, toxic nodular goitre, toxic adenoma) in order to avoid inappropriate thyroid ablation, and from the numerous other conditions that can masquerade as hyperthyroxinaemia with non-suppressed TSH [6].

Central hyperthyroidism caused by a TSH-secreting pituitary adenoma is sometimes referred to as neoplastic inappropriate secretion of TSH. This differentiates it from non-neoplastic inappropriate TSH secretion, as is seen in the syndrome of resistance to thyroid hormone (RTH) caused by mutations in *THRB* (encoding the β isoform of the thyroid hormone receptor). However, the term central hyperthyroidism is generally preferred as it more clearly identifies the site of the primary abnormality (autonomous pituitary TSH secretion in the case of a thyrotropinoma).

The utility of long-acting somatostatin analogues (SSA), both in the diagnostic algorithm and as an adjunct to surgery, is gaining prominence with several groups reporting examples of complete normalization of thyroid function and significant tumour shrinkage (even resolution) following SS therapy. Indeed, most patients can be managed through a combination of transsphenoidal surgery and medical therapy, with radiotherapy increasingly reserved for the small proportion of tumours that display aggressive behaviour and/or resistance to SSA treatment.

Aetiopathogenesis and Pathology

Normal thyrotrophs represent less than 5% of all anterior pituitary cells [7], which may explain why thyrotropinomas are so rare. In normal physiology, the set-point of the hypothalamic-pituitary-thyroid axis is maintained within a tight range, and reflects a balance between positive (e.g. hypothalamic thyrotropin-releasing hormone [TRH]) and negative (e.g. somatostatin [SA]; triiodothyronine [T3]) regulation (**Figure 2.3.12.1**). The critical role of thyroid hormone in constraining thyrotroph expansion/growth is emphasized by the development of thyrotroph hyperplasia in patients who are chronically under-replaced or non-compliant with thyroxine replacement therapy.

The aetiopathogenesis of thyrotropinomas remains poorly understood. While the majority arise within the sella, a small number of cases of TSH-secreting adenomas located within the nasopharynx have been reported [8]. Although they have been observed in the context of multiple endocrine neoplasia type I [9] and familial isolated pituitary adenoma (due to mutation in the aryl hydrocarbon receptor-interacting protein (AIP)) [10], such cases are rare and the majority appear to arise sporadically. Data from analyses of

all pituitary tumour subtypes (including a small number of TSH-secreting adenomas) point to clonal expansion from a single transformed cell [11]. Evidence for aberrant transcriptional regulation of thyrotrophs (with consequent TSH hypersecretion and unrestrained growth) has been sought but, to date, most studies have proved largely uninformative (including those examining proto-oncogenes (e.g. Gsp, ras) and tumour-suppressor genes (e.g. Rb or p53)). The pituitary-specific transcription factor Pit-1 is overexpressed in some TSH-secreting pituitary adenomas, and might play a role in cell proliferation as well as promoting hormone cosecretion in this subgroup of adenomas.

The hallmark of TSH-secreting pituitary adenomas is constitutive and autonomous secretion of TSH with failure of negative regulation by thyroid hormone. A somatic mutation of *THRB* (encoding TRβ), as well as aberrant expression of a TRβ4 isoform, have been proposed as potential mechanisms for the impaired regulation of TSH by triiodothyronine (T3) [12, 13]. Thyrotropin-releasing hormone (TRH) receptors are present on most tumour cells. However, the attenuated/absent response of TSH and α-subunit to exogenous TRH administration in the majority of patients with TSH-secreting pituitary adenomas suggests that the TRH receptor signalling pathway is impaired/non-functional in these tumours. Most thyrotropinomas express somatostatin receptor (SSTR) subtypes 1, 2, 3, and 5, with SSTR2 the predominant isoform. Native somatostatin negatively regulates TSH secretion and has been implicated in cell cycle arrest. Depot somatostatin analogue therapy rapidly normalizes thyroid function in most patients with thyrotropinomas and can promote tumour shrinkage [14], suggesting that the primary molecular defect is unlikely to involve this pathway, although somatostatin analogue (SSA) resistant tumours are recognized.

Using an alternative approach, a recent study of whole-exome sequencing of a small number of thyrotropinomas similarly failed to identify a single common driver, suggesting that discrete primary events may underlie tumorigenesis in different patients [15].

Normal thyrotrophs share a common lineage with somatotrophs and lactotrophs, and are dependent on expression of Pit-1. It is not surprising then that cosecretion of growth hormone (GH) and/or prolactin (PRL) is observed in a significant proportion of cases (up to 30%) (Figure 2.3.12.2). Intriguingly, not all patients with biochemical evidence of abnormal GH secretion manifest overt acromegaly. In addition, a subgroup of tumours demonstrates positive immunohistochemical staining for GH and/or PRL, but do not exhibit clinical or biochemical acromegaly or hyperprolactinaemia. Similarly, so-called silent thyrotropinomas (clinically non-functioning pituitary adenomas, but with abundant positive TSH staining on immunohistochemistry) are well recognized [16]. TSH-secreting macroadenomas are often accompanied by unbalanced hypersecretion of α-subunit, and double-immunostaining studies have suggested the existence of two different types of cells: one secreting α-subunit alone and another cosecreting α-subunit and β-TSH [17].

TSH-secreting pituitary adenomas are almost always benign, with only a handful of malignant tumours (with metastases to brain, lung, liver, and/or bone) reported to date [18, 19]. A subgroup of thyrotropinomas are more fibrotic than other pituitary tumours, which has been linked to local secretion of basic fibroblast growth factor by the adenoma cells [20]. In this context, both microscopic and macroscopic calcification may be seen and can confound the interpretation of MR imaging, and make surgical resection more challenging (Figure 2.3.12.2).

Finally, not all TSH-secreting pituitary adenomas are associated with overt elevation of serum TSH levels—indeed it has been estimated that between 30% and 40% of cases have a TSH level which repeatedly falls within the reference range, but it is inappropriately normal for the ambient raised thyroid hormone levels. This may be explained, at least in part, by an increased bioactivity of neoplasm derived TSH [1].

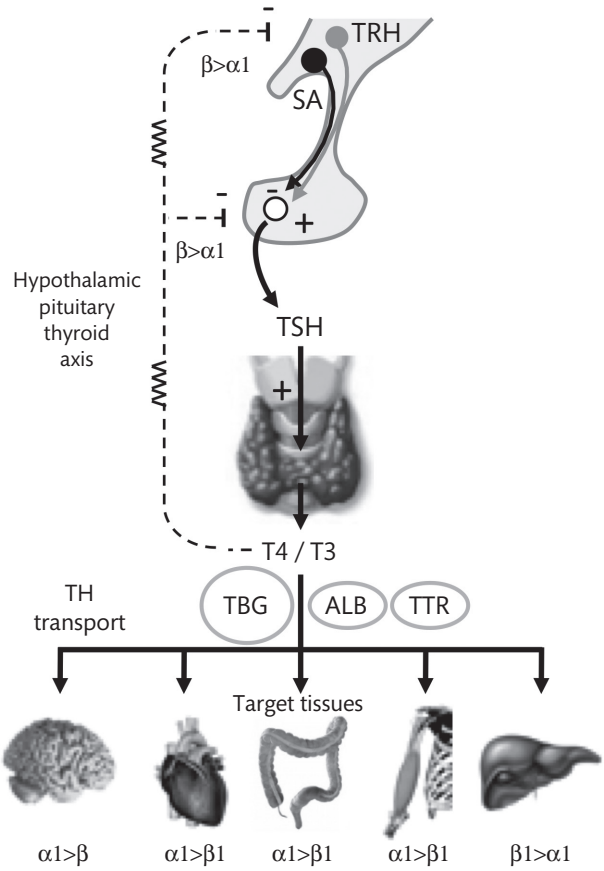


Figure 2.3.12.1 Schematic representation of the hypothalamic-pituitary-thyroid axis and major target organs. Key: α, thyroid hormone receptor α; β, thyroid hormone receptor β; ALB, albumin; SA, somatostatin; T3, triiodothyronine; T4, thyroxine; TBG, thyroxine-binding globulin; TR, thyroid receptor; TRH, thyrotropin-releasing hormone; TSH, thyrotropin; TTR, transthyretin.

Clinical Presentation

Clinical manifestations of central hyperthyroidism secondary to a TSH-secreting pituitary adenoma resemble those seen in primary hyperthyroidism (including palpitations/arrhythmias, tremor, heat intolerance, weight loss, sleep disturbance, etc.) but without disease-specific (e.g. Graves' ophthalmopathy or dermopathy) manifestations. Symptoms and signs may be progressive (although some patients report little change over intervals as long as a decade). However, in a surprisingly large number of patients the clinical features are milder, with some individuals appearing relatively asymptomatic (Koulouri, Bashari, and Gurnell, unpublished observation).

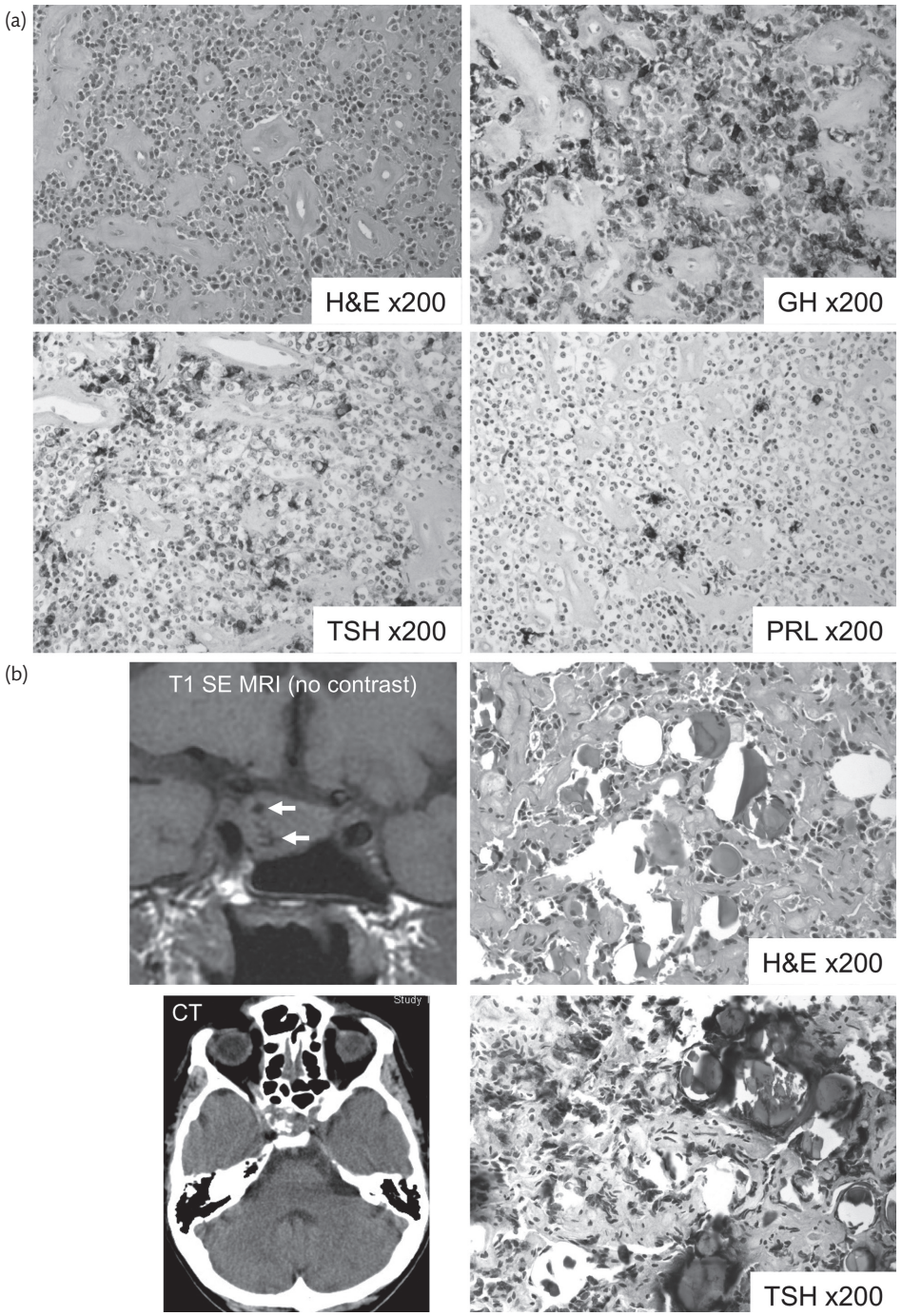


Figure 2.3.12.2 Spectrum of histological appearances in thyrotropinomas. (a) pituitary adenoma demonstrating positive immunohistochemical staining for growth hormone (GH), TSH and (to a lesser extent) prolactin (PRL). (b) thyrotropinoma with significant fibrosis and calcification—note the hypodense areas within the right side of the sella (arrows) on MRI (which may be mistaken for vascular structures, but which are confirmed as foci of calcification on CT). *For a colour version of this figure, please see colour plate section.*

Histology images provided courtesy of Dr Kieren Allinson, Department of Neuropathology, Addenbrooke's Hospital, Cambridge.

Importantly however, following successful treatment many of these patients report an improvement in their well-being, coinciding with resolution of previously unrecognized clinical symptoms. Estimates of the mean latency between onset of hyperthyroidism and establishment of the diagnosis are typically around 4–5 years.

Goitre and/or thyroid nodules may be present and are likely due to sustained TSH stimulation of thyroid follicular cells over many years. The co-occurrence of differentiated thyroid carcinoma has been described in a small number of cases [21]. Autoimmune thyroid dysfunction may also coexist and give rise to diagnostic uncertainty [22, 23].

In older case series, compressive symptoms (such as visual field defects (80%) and headaches (20%)), prevailed over those due to hyperthyroidism. However, this is no longer the case, as many patients present at an earlier stage and the proportion of microadenomas is increasing. Unilateral exophthalmos due to

Box 2.3.12.1 Causes of hyperthyroxinaemia with non-suppressed TSH

1 Raised thyroid hormone-binding capacity (leading to increased total T4 and total T3, but typically normal free hormone levels, e.g. due to oral oestrogen therapy)
2 Assay interference
  a) Falsely elevated free thyroid hormones
    i Heterophilic antibodies
    ii Anti-iodothyronine antibodies
    iii Familial dysalbuminaemic hyperthyroxinaemia (FDH)*
  b) Falsely non-suppressed TSH
    i Heterophilic or human antianimal antibodies
    ii Macro-TSH
3 Thyroxine replacement therapy (including poor compliance)
4 Medications (e.g. amiodarone, heparin)
5 Non-thyroidal illness (NTI) (including acute psychiatric disorders; neonatal period)
6 TSH-secreting pituitary adenoma
7 Resistance to thyroid hormone (RTH)
8 Disorders of thyroid hormone transport or metabolism

Key: *FDH is also associated with a true increase in total thyroid hormone levels as the mutant albumin exhibits increased binding capacity; at the same time artefactual elevation in free T4 (and in some instances free T3) may be observed on some laboratory platforms.

orbital invasion by the pituitary tumour is very rare, but must be distinguished from asymmetric exophthalmos of Graves' disease. In mixed secreting tumours, clinical findings are dependent on the nature of the hormone cosecreted (e.g. acromegaly features, amenorrhoea, and/or galactorrhoea). In macroadenomas, concomitant hypopituitarism (e.g. hypogonadism, hypoadrenalism) may be present. TSH-secreting pituitary adenomas have been reported in pregnant women, in patients with multiple endocrine neoplasia type I, and in atypical McCune–Albright syndrome [9, 24, 25].

Differential Diagnosis

Since the advent of ultrasensitive TSH assays, the classical biochemical signature of TSH-secreting pituitary adenomas is readily recognized: elevated or detectable (i.e. non-suppressed) TSH in the face of biochemical hyperthyroidism (with increased free thyroxine (fT_4) and triiodothyronine (fT_3)). However, this laboratory picture is not exclusive to thyrotropinomas and it is important to consider other possible causes before proceeding to further investigation (Box 2.3.12.1).

Elevated total thyroid hormone (TT4, TT3) levels, with inappropriately normal or elevated TSH, may be observed when thyroid hormone-binding capacity is increased, e.g. during pregnancy when thyroxine-binding globulin (TBG) levels are raised, or in individuals with familial dysalbuminaemic hyperthyroxinaemia (FDH), where point mutations in the *ALB* gene increase albumin's affinity for thyroid hormone [26]. FDH may also perturb commonly used fT_4 (± fT_3) assays to yield falsely elevated levels. Close liaison with the laboratory is therefore required to exclude these and other conditions (e.g. heterophilic antibodies or anti-iodothyronine antibodies) which are capable of producing false positive TSH or fT_4 (± fT_3) assay readouts. In addition, careful consideration must be given to other conditions and medications (e.g. heparin, amiodarone) that may yield a blood profile suggestive of central hyperthyroidism [27].

Investigation of Hyperthyroxinaemia with Non-Suppressed TSH

Step 1—Clinical Assessment

Even when a patient appears to have been extensively investigated, it is always important to revisit the clinical history and examination findings to: (i) identify potential confounding factors (intercurrent illness; amiodarone, heparin); (ii) determine the thyroid status of the patient; and (iii) check for other clues that might point to a TSH-secreting adenoma (e.g. coexistent acromegaly or hyperprolactinaemia; local compressive features), or indeed suggest the possibility of assay artefact (e.g. dysthyroid eye signs in a patient with primary thyroid dysfunction and an unreliable TSH readout).

Step 2—Exclusion of Laboratory Assay Interference

Close liaison with the Clinical Biochemistry laboratory is critical to avoid further inappropriate investigation in cases where either hyperthyroxinaemia or a non-suppressed TSH is due to assay interference. Confirmation that free thyroid hormone levels (free T4 and T3) are genuinely raised may require referral to a reference laboratory with the ability to perform a two-step assay or equilibrium dialysis to detect conditions such as FDH or anti-iodothyronine antibodies. TSH dilution studies, polyethylene glycol (PEG) precipitation or even gel filtration chromatography may be required to exclude interference in the TSH assay [6].

Step 3—Distinguishing a TSH-Secreting Pituitary Adenoma from *THRB* RTH

Once genuine hyperthyroxinaemia with non-suppressed TSH has been confirmed, the key challenge is to distinguish a TSH-secreting pituitary adenoma from rare genetic disorders of thyroid hormone transport, metabolism, and action (in particular RTHβ due to mutations in *THRB*). The degree of clinical thyrotoxicosis and the magnitude of elevation of thyroid hormone and TSH levels show considerable overlap when comparing cohorts of patients with thyrotropinomas and *THRB* RTH. Accordingly, other investigations are required to reliably distinguish between the two conditions.

Family History/Genetic Screening

Although a small number of individuals with thyrotropinomas have been found to harbour mutations in genes associated with familial pituitary disease (MEN type 1 and familial isolated pituitary adenoma), most TSH-secreting adenomas are sporadic. In contrast, *THRB* RTH is an autosomal dominant disorder and affected family members can be readily identified by performing thyroid function testing, thus guiding subsequent genetic screening. However, between 10% and 15% of patients with RTHβ do not have a demonstrable mutation in *THRB* (see Chapter 3.4.5, 'Syndromes of Resistance to Thyroid Hormone').

Alpha Subunit (ASU) and ASU:TSH Molar Ratio

Classically, an increased plasma ASU level and increased ASU:TSH molar ratio were considered important diagnostic markers for thyrotropinoma. The ASU:TSH molar ratio has been recommended to help avoid misinterpretation due to raised gonadotrophin levels (e.g. in postmenopausal women), reflecting the shared common alpha subunit with luteinizing hormone (LH) and follicle-stimulating hormone (FSH). However, in this context, the population reference range for ASU:TSH molar ratio is so wide [28] as to render it of limited value for the detection of all but the most obvious macroadenoma. As a general rule, a raised ASU (and/or ASU:TSH molar ratio) may help to support suspicions of a thyrotropinoma, but a normal result certainly does not exclude the possibility.

Sex Hormone-Binding Globulin (SHBG)

Hepatic production of SHBG is directly regulated by thyroid hormone operating via the TRβ1 isoform of the receptor. In RTHβ the liver is therefore resistant and SHBG levels remain within the age- and gender-related reference range. In contrast, in the presence of a TSH-secreting tumour, raised thyroid hormone levels cause hepatic thyrotoxicosis with raised SHBG. Unfortunately, however, normal SHBG levels may sometimes be seen, especially when there is cosecretion of GH by a thyrotropinoma. Similarly, other factors (e.g. concomitant oral oestrogen therapy) may cause an elevated SHBG in RTH.

Dynamic Testing

The use of dynamic testing can be particularly helpful in discriminating between *THRB* RTH and a TSH-secreting pituitary adenoma:

(a) TRH test (200 mcg i.v.):
- Thyrotropinoma: attenuated or absent response of TSH following exogenous TRH stimulation
- *THRB RTH:* preserved or exaggerated response

(b) Depot somatostatin analogue (Lanreotide Autogel® 90–120 mg or Sandostatin LAR® 20–30 mg) every 4 weeks for three injections):
- Thyrotropinoma: rapid and sustained normalization of free T3, free T4, and TSH levels in more than 90% of patients
- *THRB RTH:* a transient reduction in free thyroid hormone levels may be observed, but returns to baseline by the conclusion of the test

(c) Liothyronine (L-T3) suppression test (various protocols exist, including progressive dose escalation or fixed dosing over a 10-day period):
- Thyrotropinoma: absent or only slight reduction of TSH
- *THRB RTH:* TSH suppression is marked (although not always complete)

However, in practice the L-T3 suppression test is often reserved for those cases in whom results remain equivocal despite extensive investigation, or for detecting low level residual or recurrent disease following pituitary surgery. It is important to note that patients with known or suspected cardiac disease (arrhythmias, ischaemia or heart failure) and those with a history of psychiatric disease should not be exposed to high doses of thyroid hormone as is required for the L-T3 suppression test.

Table 2.3.12.1 summarizes the typical findings in TSH-secreting pituitary adenoma and *THRB* RTH across the various tests.

Step 4—Pituitary Imaging

Once the diagnosis of a TSH-secreting pituitary adenoma has been confirmed, then attention turns to tumour localization to help guide management. It is important however, to bear in mind that the presence of a pituitary microadenoma in a patient with inappropriate secretion of TSH, although strongly suggestive, is

Table 2.3.12.1 Distinguishing between TSH-secreting pituitary adenoma and TRβ resistance to thyroid hormone (*THRB* RTH)

	TSH-secreting adenomas	Resistance to thyroid hormone (*THRB*)
Family history	Usually absent	Possible
Genetic analysis	Usually negative	Positive in 85–90% of cases
Increased free thyroid hormones	Yes	Yes
Normal or high TSH	Yes	Yes
Circadian TSH secretion	Absent	Preserved
SHBG	Increased (but may be normal if cosecretion of GH)	Normal
α-subunit level	Increased/normal	Normal
α-subunit/TSH molar ratio	Increased/normal	Normal
Response to TRH stimulation test	Absent/attenuated	Preserved/exaggerated
Response to L-T3 suppression test	Absent/markedly attenuated	Present, although may not be complete
Response to depot SSA trial	Rapid and sustained suppression of TSH in majority of cases resulting in restoration of euthyroidism	No effect
GH and or prolactin cosecretion	+	–
MRI	Adenoma (although some microadenomas may not be visualized on standard clinical MR sequences)	Normal (but incidentalomas in 10–15%)

Key: GH, growth hormone; L-T3, Lio-thyronine; MRI, magnetic resonance imaging; SHBG, sex hormone-binding globulin; SSA, somatostatin analogue; TRH, thyrotropin-releasing hormone; TSH, thyroid-stimulating hormone.

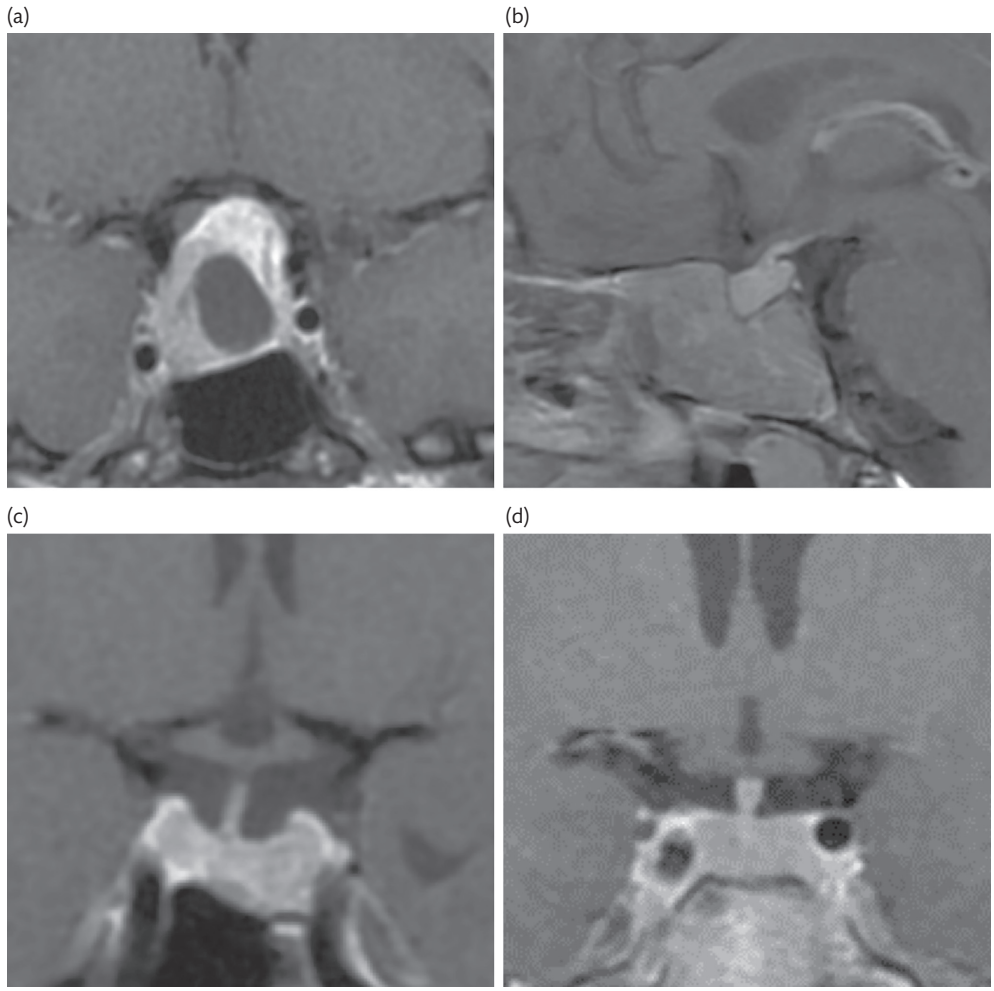


Figure 2.3.12.3 Spectrum of TSH-secreting pituitary adenomas. (a) Macroadenoma with predominant suprasellar extension compressing the optic chiasm. (b) Macroadenoma filling the sphenoid sinus—note how the adenoma appears discrete from the normal pituitary gland in the sella, raising the possibility of an extrasellar ectopic origin. (c) Left-sided microadenoma. (d) No visualized adenoma.

not diagnostic of a TSH-secreting pituitary tumour given that pituitary incidentalomas are found on MRI in ~10–15% of the general population. In contrast, incidental macroadenomas are much less commonly encountered. In patients with TSH-secreting pituitary adenomas there is no correlation between serum TSH levels and tumour size. In older case series, patients commonly presented with macroadenomas with suprasellar and sphenoidal extension, whereas in more recent series microadenomas feature much more frequently. Indeed, it is now clear that thyrotropinomas exhibit a spectrum ranging from large invasive (even giant) tumours through to microadenomas that cannot be readily visualized using conventional MR sequences (analogous to some Cushing's microadenomas) (Figure 2.3.12.3). For the latter, various approaches have been tried to aid localization of the tumour (including inferior petrosal sinus sampling with TRH stimulation and octreotide scintigraphy). However, the most promising approach appears to be functional imaging with ^{11}C-methionine PET-CT coregistered with volumetric MRI, which may aid precise localization of the tumour (Figure 2.3.12.4), especially when performed pre-and post-somatostatin analogue therapy [4].

Management

In patients with TSH-secreting pituitary adenomas, the goal of therapy is to restore euthyroidism and, in patients with large tumours, to eliminate any symptoms of mass effect. Early diagnosis and correct treatment of these rare tumours helps prevent complications (such as visual failure and hypopituitarism), and may also improve surgical cure rates.

The success of treatment depends on the criteria used. It has been suggested that an early test of cure might be an undetectable TSH concentration seven days after surgery (when the normal thyrotrophs remain suppressed as a consequence of prolonged exposure to raised circulating thyroid hormone levels) [29]. However, many patients are now treated with somatostatin analogues prior to surgery, often with rapid normalization of hyperthyroidism, which may therefore allow recovery of normal thyrotroph function prior to theatre. Normalization of dynamic tests (e.g. TRH test, T3 suppression test) might therefore represent a better (later) test of cure. In any case, long-term follow-up is necessary to detect relapse and recurrence.

Transsphenoidal surgery remains the treatment of choice in many patients with TSH-secreting pituitary adenomas. Historically, pituitary surgery alone has been reported to result in normalization

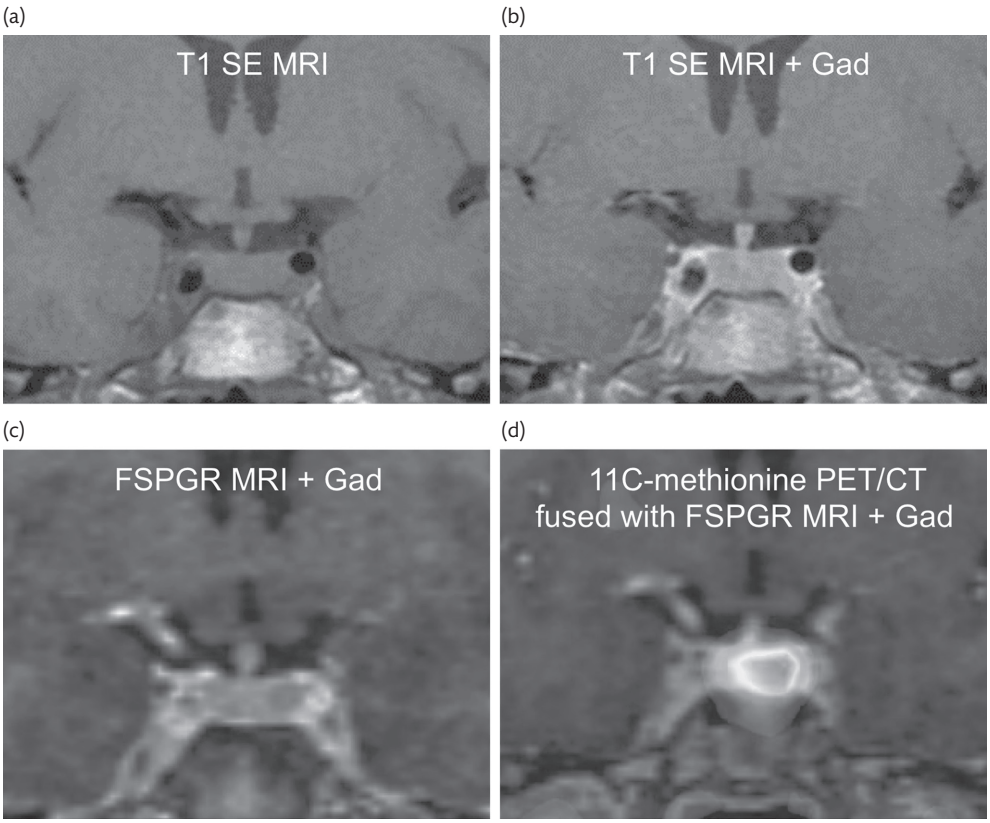


Figure 2.3.12.4 Confirmation of the site of a microadenoma using volumetric MRI and functional imaging with 11C-methionine PET. (a, b) Standard clinical spin echo (SE) MR sequences pre- and post-gadolinium (Gad) fail to identify a thyrotropinoma. (c) Volumetric (fast spoiled gradient recalled acquisition (FSPGR)) MR sequences (1 mm slices) identify a possible left-sided microadenoma adjacent to the cavernous sinus. (d) ^{11}C-Methionine PET/CT coregistered with FSPGR MRI demonstrates avid tracer uptake at the site of the suspected lesion. Transsphenoidal surgery and histological analysis confirmed a left-sided TSH-secreting microadenoma. *For a colour version of this figure, please see colour plate section.*

of thyroid hormone secretion and resolution of the pituitary mass in approximately 50% of patients, achieve normalization of thyroid parameters despite incomplete tissue removal in approximately 25% of patients, and is unsuccessful in nearly 30% of cases. However, an increase in the surgical cure rate has been reported in more recent series, probably reflecting improved surgical techniques and earlier diagnosis [29].

Medical treatment is an alternative option to pituitary surgery in patients with TSH-secreting pituitary tumours. Dopamine agonists are an effective treatment in true mixed TSH/prolactin-secreting pituitary adenomas, whereas success is limited when bromocriptine is used in patients with pure TSH-secreting pituitary adenomas [30]. In contrast, depot somatostatin analogues (octreotide LAR or Lanreotide Autogel®) suppress TSH secretion in more than 90% of TSH-secreting pituitary adenomas, and normalize thyroid hormone levels in approximately 75–90% of patients. However, significant shrinkage of the adenoma is only observed in approximately 50% of cases, and total resolution of a macroadenoma with maintenance of long-term cure in response to somatostatin analogue therapy has been reported in only a single case [31]. As a general rule, the effects of somatostatin analogues are reversible, with the need for long-term administration, with possible tachyphylaxis requiring increasing doses of the drug to maintain good control of the disease in up to 10% of cases, and with true resistance in a small number of patients [32]. The second generation somatostatin receptor ligand pasireotide may offer an alternative for treatment resistant cases, although experience to date is limited to a single patient in whom it was not clear whether there was true resistance to first line SSA therapy [33]. Somatostatin analogues remain expensive and may be associated with side effects such as cholelithiasis and carbohydrate intolerance. Accordingly, although some patients and their clinicians might elect to pursue long-term primary SSA therapy, for many patients these remain an important bridge to surgery (allowing anaesthesia to proceed safely) and as adjunctive therapy when surgery is unable to achieve full remission, or while awaiting the beneficial effects of radiotherapy.

It is appropriate to briefly mention conventional antithyroid drug (ATD) therapy, which may be used for a short duration in preparation for surgery. However, concerns remain that longer term use may predispose to tumour expansion (Nelson's-like phenomenon) [1], and for that reason permanent thyroid ablation is also generally best-avoided, although may occasionally be indicated in those with recalcitrant arrhythmias or cardiac failure.

Conventional radiotherapy and stereotactic radiosurgery have also been employed in the management of TSH-secreting pituitary tumours, especially when surgery and/or medical therapy has proved unsuccessful [34]. However, the number of reported cases, especially using modern radiotherapy techniques, remains relatively small. As with other pituitary tumour subtypes, and allowing for the long time period necessary for the full effect to be realized, and potential late side effects (e.g. hypopituitarism, second tumour, possible increased cerebrovascular disease), radiotherapy is generally reserved for the small number of cases in whom combination

surgery and medical therapy is unable to achieve satisfactory biochemical and/or tumour control.

Conclusions

TSH-secreting pituitary adenomas are a rare cause of hyperthyroidism. The main prognostic factors of these adenomas are size and invasiveness of the tumours, duration of symptoms, and intensity of hyperthyroidism. Early recognition of central hyperthyroidism, and localization of the causative lesion should now be readily achieved in most cases, reflecting the availability of ultrasensitive TSH assays and improvements in pituitary imaging. Reflecting this, the spectrum of disease and approach to the treatment of these rare pituitary tumours has changed significantly during the past decade. Patients typically present with mild or moderate symptoms and signs of hyperthyroidism, hormonal evaluation shows increased free thyroid hormone concentrations with detectable serum TSH levels, and MRI increasingly identifies patients with micro- rather than macroadenomas. Trans-sphenoidal surgery remains the treatment of choice, although depot somatostatin analogue therapy is now established as a key component in the management pathway: either pre- and, in selected cases, post-surgery, or sometimes as long-term primary medical therapy.

REFERENCES

1. Beck-Peccoz P, Persani L, Mannavola D, Campi I. Pituitary tumours: TSH-secreting adenomas. *Best Pract Res Endocrinol Metab*, 2009; **23**: 597–606.
2. Önnestam L, Berinder K, Burman P, *et al*. National incidence and prevalence of TSH-secreting pituitary adenomas in Sweden. *J Clin Endocrinol Metab*, 2013; **98**: 626–35.
3. Socin HV, Chanson P, Delemer B, *et al*. The changing spectrum of TSH-secreting pituitary adenomas: diagnosis and management in 43 patients. *Eur J Endocrinol*, 2003; **148**: 433–42.
4. Koulouri O, Hoole AC, English P, *et al*. Localisation of an occult thyrotropinoma with 11C-methionine PET-CT before and after somatostatin analogue therapy. *Lancet Diabetes Endocrinol*, 2016; **4**: 1050.
5. Tjörnstrand A, Nyström HF. Diagnostic approach to TSH-producing pituitary adenoma. *Eur J Endocrinol*, 2017; **177**: R183–97.
6. Koulouri O, Moran C, Halsall D, Chatterjee K, Gurnell M. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 745–62.
7. Kovacs K. Tumours of the pituitary gland. In: *Atlas of Tumour Pathology* (Second Series, Fascicle 21). Washington DC: Armed Forces Institute of Pathology, 1986.
8. Jin Z, Zhang Z, Li F, *et al*. Hyperthyroidism caused by an ectopic thyrotropin-secreting tumor of the nasopharynx: a case report and review of the literature. *Thyroid*, 2013; **23**: 1172–7.
9. Burgess JR, Shepherd JJ, Greenaway TM. Thyrotropinomas in multiple endocrine neoplasia type 1 (MEN-1). *Aust N Z J Med*, 1994; **24**: 740–1.
10. Hernández-Ramírez LC, Gabrovska P, *et al*. Landscape of familial isolated and young-onset pituitary adenomas: prospective diagnosis in AIP mutation carriers. *J Clin Endocrinol Metab*, 2015; **100**: E1242–54.
11. Ma W, Ikeda H, Watabe N, Kanno M, Yoshimoto T. A plurihormonal TSH-producing pituitary tumor of monoclonal origin in a patient with hypothyroidism. *Horm Res*, 2003; **59**: 257–61.
12. Tagami T, Usui T, Shimatsu A, *et al*. Aberrant expression of thyroid hormone receptor β isoform may cause inappropriate secretion of TSH in a TSH-secreting pituitary adenoma. *J Clin Endocrinol Metab*, 2011; **96**: E948–52.
13. Ando S, Sarlis NJ, Oldfield EH, Yen PM. Somatic mutation of TRβ can cause a defect in negative regulation of TSH in a TSH-secreting pituitary tumor. *J Clin Endocrinol Metab*, 2001; **86**: 5572–6.
14. Tosaka M, Horiguchi K, Umezawa R, *et al*. Somatostatin receptor subtypes mRNA in TSH-secreting pituitary adenomas: a case showing a dramatic reduction in tumor size during short octreotide treatment. *Endocr J*, 2007; **54**: 371–8.
15. Sapkota S, Horiguchi K, Tosaka M, Yamada S, Yamada M. Whole-exome sequencing study of thyrotropin-secreting pituitary adenomas. *J Clin Endocrinol Metab*, 2017; **102**: 566–75.
16. Kirkman MA, Jaunmuktane Z, Brandner S, Khan AA, Powell M, Baldeweg SE. Active and silent thyroid-stimulating hormone-expressing pituitary adenomas: presenting symptoms, treatment, outcomes, and recurrence. *World Neurosurg*, 2014; **82**: 1224–31.
17. Terzolo M, Orlandi F, Bassetti M, *et al*. Hyperthyroidism due to a pituitary adenoma composed of two different cell types, one secreting α-subunit alone and another cosecreting α-subunit and thyrotropin. *J Clin Endocrinol Metab*, 1991; **72**: 415–21.
18. Colandrea JM, Taubenberger JK, Doppman JL, *et al*. Thyrotropin-secreting pituitary carcinoma. *J Clin Endocrinol Metab*, 2014; **76**: 529–33.
19. Brown RL, Muzzafar T, Wollman R, Weiss RE. A pituitary carcinoma secreting TSH and prolactin: a non-secreting adenoma gone awry. *Eur J Endocrinol*, 2006; **154**: 639–43.
20. Smyth HS, Asa SL, Ezzat S, Horvath E, Singer W, Kovacs K. Basic fibroblast growth factor expression by two prolactin and thyrotropin-producing pituitary adenomas. *Endocr Pathol*, 2007; **6**: 125–34.
21. Ünlütürk U, Sriphrapradang C, Erdoğan MF, *et al*. Management of differentiated thyroid cancer in the presence of resistance to thyroid hormone and TSH-secreting adenomas: a report of four cases and review of the literature. *J Clin Endocrinol Metab*, 2013; **98**: 2210–17.
22. Arai N, Inaba M, Ichijyo T, Kagami H, Mine Y. Thyrotropin-producing pituitary adenoma simultaneously existing with Graves' disease: a case report. *J Med Case Rep*, 2017; **11**: 9.
23. Losa M, Mortini P, Minelli R, Giovanelli M. Coexistence of TSH-secreting pituitary adenoma and autoimmune hypothyroidism. *J Endocrinol Invest*, 2006; **29**: 555–9.
24. Caron P, Gerbeau C, Pradayrol L, Simonetta C, Bayard F. Successful pregnancy in an infertile woman with a thyrotropin-secreting macroadenoma treated with somatostatin analog (octreotide). *J Clin Endocrinol Metab*, 1996; **81**: 1164–8.
25. Gessl A, Freissmuth M, Czech T, Matula C, Buchfelder M, Vierhapper H. Growth hormone-prolactin-thyrotropin-secreting pituitary adenoma in atypical McCune–Albright syndrome with functionally normal Gsα protein. *J Clin Endocrinol Metab*, 1994; **79**: 1128–34.
26. Schoenmakers N, Moran C, Campi I, *et al*. A novel albumin gene mutation (R222I) in familial dysalbuminemic hyperthyroxinemia. *J Clin Endocrinol Metab*, 2014; **99**: E1381-6.

27. Laji K. Abnormal serum free thyroid hormone levels due to heparin administration. *QJM*, 2002; **94**: 471–3.
28. Beck-Peccoz P, Persani L, Faglia G. Glycoprotein hormone α-subunit in pituitary adenomas. *Trends Endocrinol Metab*, 1992; **3**: 41–5.
29. Yamada S, Yamaguchi-Okada M, Takeuchi Y, *et al.* Clinicopathological characteristics and therapeutic outcomes in thyrotropin-secreting pituitary adenomas: a single-center study of 90 cases. *J Neurosurg*, 2014; **121**: 1462–73.
30. Teramoto A, Sanno N, Tahara S, Osamura YR. Pathological study of thyrotropin-secreting pituitary adenoma: plurihormonality and medical treatment. *Acta Neuropathol*, 2004; **108**: 147–53.
31. Fliers E, van Furth WR, Bisschop PH. Cure of a thyrotrophin (TSH)-secreting pituitary adenoma by medical therapy. *Clin Endocrinol (Oxf)*, 2012; **77**: 788–90.
32. Beck-Peccoz P, Brucker-Davis F, Persani L, Smallridge RC, Weintraub BD. Thyrotropin-secreting pituitary tumors. *Endocr Rev*, 1996; **17**: 610–38.
33. van Eersel MEA, Meeuwisse–Pasterkamp SH, Muller Kobold AC, *et al.* Treatment of a thyrotropin-secreting pituitary adenoma (TSH-oma) with pasireotide LAR. *Clin Endocrinol*, 2017; **87**: 877–9.
34. Malchiodi E, Profka E, Ferrante E, *et al.* Thyrotropin-secreting pituitary adenomas: outcome of pituitary surgery and irradiation. *J Clin Endocrinol Metab*, 2014; **99**: 2069–76.

2.3.13 Pituitary Carcinoma

Ann McCormack

Definition and Epidemiology

Pituitary carcinomas (PC) are tumours originating from cells of the adenohypophysis with evident craniospinal or systemic metastatic disease [1]. They are rare, accounting for 0.12–0.2% of all pituitary tumours with an incidence of 4 per 1 000 000 person years [2–4]. However, the incidence is likely underestimated as frequently metastatic deposits are only detected at autopsy [5–7]. Most commonly PC presents in the fourth to sixth decades of life, with a mean age at diagnosis of 44 [7–9]. Paediatric cases are very rare but have been described [10, 11]. There is a slight male predominance [2, 8]. More than 80% of PC are clinically functioning, the majority corticotroph and lactotroph tumours, less commonly somatotroph and thyrotroph PC [8]. Notably, 20% of PC evolve from initially clinically silent to functional tumours. Silent hormone expressing PC are more frequently encountered among the clinically non-functioning PC with true gonadotroph and null cell PC comprising a small subset [8]. PC has been described in the setting of familial cancer syndromes, including patients with germline mutations in *MEN1, SDHB*, and DNA mismatch repair genes (Lynch syndrome) [12–15]. However, there is no evidence that PC are more common in patients with familial predisposition to pituitary tumours.

Clinical Presentation and Diagnosis

Pituitary tumours destined to become PC almost always begin as macroadenomas. However, size at presentation does not correlate with potential for malignant behaviour as giant lactotroph tumours can be highly sensitive to dopamine agonist treatment [16]. Commonly there is a history of relatively indolent behaviour with emergence of aggressive disease years later preceding the identification of metastases. There is considerable variability in the latency to development of metastatic disease from months to more than 20 years from diagnosis with a mean of 7 years [9]. Clinically aggressive tumour behaviour may be evident at presentation with fulminant progression [17].

Mass effect symptoms, such as visual impairment, headache, and cranial nerve palsies, tend to dominate the clinical presentation of PC. Unusual neurological signs, such as ataxia or motor weakness, should raise concern about the possibility of PC. Diabetes insipidus (DI) is rarely encountered in PC, and in general the presence of DI should alert to an alternative diagnosis, such as metastasis to the sellar arising from a non-pituitary primary [18].

Systemic metastases occur most commonly (47%), are more frequent among lactotroph compared with corticotroph tumours and most often involve the liver, lung, and bone, with rare sites of metastatic disease including the orbit, heart, pancreas, kidney, ovary, myometrium, and pelvic lymph nodes [9]. Systemic dissemination occurs through haematogenous spread via the cavernous sinuses and/or jugular veins [3]. Metastases to cervical lymph nodes are also seen and thought to result from invasion of lymphatics of the skull base and soft tissues [3]. Craniospinal metastases also occur frequently (40%), spread occurring through cerebrospinal fluid dissemination. Both craniospinal and systemic metastases are seen in 15% of patients [9].

The primary diagnostic modality is structural imaging with magnetic resonance imaging (MRI) or computed tomography (CT) with attention to site-specific symptoms such as neck or back pain that may indicate a site of metastatic disease. Functional imaging modalities may be useful in detecting otherwise occult metastases. Ga68 Dotatate and F18 FDG-PET/CT may identify different sites of metastatic disease when used concurrently, with potential therapeutic and prognostic implications [19, 20]. According to 2017 WHO criteria, the following need to be met in order to diagnose a PC: (1) the primary lesion must be a histologically proven adenohypophyseal tumour; (2) an alternative primary lesion must be excluded; (3) discontinuous spread must be present in the craniospinal axis; and (4) the pathological features of metastases should be similar to those of the primary pituitary tumour [21].

Thus, while imaging may suggest PC, a surgical biopsy may be required to confirm the diagnosis. However, fine needle aspiration biopsy of metastases within cervical lymph nodes, liver and lung have been successfully performed [22, 23].

Only a minority of aggressive pituitary tumours (APT) ultimately manifest metastases. However, recognition of an APT is vital to identify tumours with malignant potential. Furthermore, the morbidity and mortality of APT are increased even in patients without metastases [24]. An APT demonstrates invasion and displays an unusually rapid growth rate, or continued growth despite multimodal treatment (surgery, radiotherapy, standard medical therapy) [24]. Serial MRI imaging is required to determine tumour growth rates, with accurate reporting of tumour dimensions and ideally tumour volume measurements. In a large study of 153 pituitary tumours (functioning and non-functioning), the mean tumour volume doubling time (TVDT) was 1147 +/– 870 (SD) days (range 60–3478). Tumours with a more rapid growth rate (below 1 SD TVDT) had a TVDT under 1 year [25]. Younger patients with non-functioning pituitary adenomas (NFPA) and females with functioning tumours have been found to have faster growth rates [25, 26]. The rapid growth of a corticotroph pituitary tumour following bilateral adrenalectomy (Nelson's syndrome) is also cause for concern and has been found in 67% of corticotroph PC [18]. Metastatic disease should be suspected when there are discordant biochemical and radiological findings, for example rising prolactin (PRL), adrenocorticotropin hormone (ACTH) or growth hormone (GH) levels in the setting of low volume sella-based disease [24].

Functioning PC typically exhibit very high hormone levels, however there is significant overlap with levels found in pituitary adenomas (PA). One hallmark that may indicate malignant transformation of a PA is the emergence of resistance to medical treatment. This is particularly seen with prolactinomas, where resistance to dopamine agonist (DA) therapy may be a feature present early in the disease course but can also be acquired during treatment [27]. Transformation of a PA from silent to functioning may also herald malignant progression, well described for silent corticotroph adenomas but also reported in silent somatotroph tumours and in one case of an initially hormone-immunonegative tumour that subsequently transformed into a thyroid stimulating hormone (TSH) and PRL-secreting PC with brain metastases [28–31]. Conversely, tumour de-differentiation is thought to account for cases where reduced hormone secretion has been observed [18, 32].

Pathological Evaluation

There are no histopathological, immunohistochemical, or ultrastructural features to reliably differentiate PC from PA. Microscopically, PC usually appear as well-differentiated neuroendocrine tumours akin to benign PA. While histological features of hypercellularity, nuclear pleomorphism, necrosis, haemorrhage, or invasion may be present in PC, they are also common in PA. Neuronal metaplasia has been reported in PC but is rare. Metastatic deposits often have more atypical cytological features compared to the primary pituitary tumour [21]. Mitoses are more frequently seen in PC compared with PA, although there are cases of PC without significant mitoses [7]. In one study, mitoses were present in 3.9% of non-invasive PA, 21.4% of invasive PA and 66% of PC, while in the large European Society of Endocrinology (ESE) survey cohort of APT a raised mitotic count (>2 mitoses/10 high powered field (HPF)) was found in 90% of PC compared with 63% of APT [8, 33].

The neuroendocrine origin of PC can be confirmed on immunohistochemical analysis with expression of markers such as chromogranin A and synaptophysin. Pituitary hormone expression does not differentiate PC from PA but is important to determine tumour subtype. Tumour subclassification, utilizing immunohistochemical analysis of transcription factors and low-molecular-weight cytokeratin, can assist in identifying a tumour with an increased potential for more aggressive behaviour (Table 2.3.13.1) [34]. Clinically non-functioning but immunohistochemically classified Pit-1 positive plurihormonal adenomas (previously termed silent subtype 3 PA) are particularly recognized for their inherently aggressive behaviour. Cytokeratin staining identifies the fibrous bodies of a sparsely granulated somatotroph adenoma and also hyaline changes of Crooke's cell adenomas [34]. Ultrastructural evaluation is no longer necessary in the routine classification of pituitary tumours [21].

Measurement of Ki67, using the MIB-1 antibody, provides an assessment of the number of cells in the S-phase of the cell cycle, and is an important marker of the rate of cellular proliferation [7]. Metastatic deposits frequently exhibit higher Ki67 indices than the primary pituitary tumour [21]. Higher Ki67 indices have been correlated with invasive adenomas, risk of recurrence and malignant behaviour [8, 33, 35, 36]. The mean Ki67 in one series was 1.37% for non-invasive PA compared with 4.66% for invasive adenomas and 11.91% for PC [37]. Similarly, in the ESE survey cohort 46% of APT

Table 2.3.13.1 Pituitary tumour subtypes with potential for aggressive behaviour

Tumour subtype	Hormone	Cytokeratin	Transcription factor
Sparsely granulated somatotroph adenoma	GH +/- PRL	Dot-like (fibrous bodies)	Pit-1
Lactotroph macroadenoma in men	PRL +/– GH (acidophilic stem cell adenoma)	+/– fibrous bodies (acidophilic stem cell adenoma)	Pit-1 ER-α (sparsely or densely granulated)
Crooke cell adenoma	ACTH	Ring-like pattern	Tpit
Silent corticotroph adenoma	ACTH	Diffuse pattern	Tpit
Plurihormonal Pit-1-positive adenoma	GH, PRL, TSHβ +/– α-subunit	N/A	Pit-1

Modified with permission from Osamura RY, Lopes MBS, Grossman A, Kontogeorgos G, Trouillas J. *Pituitary Adenomas*. In: Lloyd RV KG, Rosai J, editor. WHO classification of tumours of endocrine organs, 4th edn. Lyon: IARC Press; 2017.

and PC were found to have a Ki67 ≥ 10% [38]. A threshold Ki67 ≥ 3% has commonly been used as a marker associated with an increased risk of tumour recurrence while a Ki67 ≥ 10% is considered by some authors to indicate a malignant pituitary tumour [21, 24]. However, there is substantial variability in Ki67 reported among PC from 0% to 34% [7, 8]. This may result from tumour heterogeneity as well as differences in interpretation of Ki67 labelling [9].

Used in isolation, the Ki67 measurement has limited prognostic potential [39, 40]. Increased expression of p53 has been reported in 100% of a cohort of PC compared with 7.1% of PA, while in other studies p53 expression has progressively increased during the progression from adenoma to carcinoma [7, 9]. It has also been found to be absent in PC and considerable debate exists over its reliability [7, 9, 24, 41]. The combination of Ki67 >3% and extensive p53 immunoreactivity has been associated with an increased recurrence rate of PA in some studies, but not in others [42, 43]. In the 2004 WHO classification of pituitary tumours, an 'atypical pituitary adenoma' was defined by the presence of Ki67 >3%, p53 overexpression, and increased mitotic rate, however this category was abandoned in the 2017 WHO classification given the lack of clinical validation as a prognostic tool [1, 21]. Nevertheless, measurement of Ki67, along with p53 immunostaining and evaluation of mitotic count, remain recommended in the assessment of an APT [24]. The utility of these proliferative markers may be most useful in determining risk of recurrence for invasive PA. This formed the basis for a clinicopathological grading classification system in which invasive and proliferative PA demonstrated a 12-fold increased risk of tumour progression compared with non-invasive, non-proliferative PA [44].

Pathogenesis and Genetics

The exact pathogenesis of PC is uncertain, but progression from PA to PC is thought to arise from an accumulation of molecular alterations or development of a subclone of tumour cells with critical genetic and epigenetic change(s) that form metastatic deposits [9, 21]. The multistep genetic progression from benign to malignant pituitary tumours has been illustrated in rat studies utilizing SMtTW lineages derived from spontaneous PRL tumours [45]. The most frequent somatic genetic event affecting PA are chromosomal alterations affecting whole chromosome arms [46]. In PC, there are an increased number of chromosomal aberrations, particularly gains, with an average of 8.3 imbalances per tumour [47]. Chromosomal gains in 1q, 3p, 5, 8, 14q, and 19p have been reported in PC, and loss of chromosome 11 may be important in malignant transformation [48]. Pituitary tumour transforming gene (*Pttg*), is a securin protein critical in the spindle checkpoint machinery and error-free mitosis. *Pttg* overexpression results in genetic instability by causing DNA breaks and is a key oncogene driving pituitary tumour development and progression [49]. Increased *Pttg* expression correlates with higher levels of Ki67 and is associated with more aggressive pituitary tumour behaviour [50]. In a rat model of malignant prolactinoma, markedly elevated nuclear *Pttg* levels was postulated to be important in the development of an aggressive tumour phenotype [48]. *Pttg* has been identified as a key gene involved in tumour metastasis within solid malignancies [51].

Unlike many solid cancers, PA are not characterized by recurrent focal genetic driver mutations. Activation of classic oncogene genes, such as *Ras*, and mutations in tumour suppressor genes, such as *p53*, are not found in PA but have been described in PC [52–54]. In fact, *p53* and *p21* upregulation are key features of senescence, an essential mechanism controlling cellular proliferation in benign PA [49]. Reduced expression of important cell cycle inhibitors, such as *p21* and *p27(Kip1)*, appear to be critical in pituitary tumour progression, with undetectable levels seen in PC [55, 56]. Similarly, reduced expression of other proteins, such as retinoblastoma (*Rb*) gene and *nm23*, also involved in regulation of cell cycle progression have been described in PC [7]. Consequent upregulation of cell cycle proteins, such as cyclin D1, and increased cellular proliferation are central events in the malignant transformation of pituitary tumours [48]. Increased cyclin D1 levels have also been proposed to play a role in stimulating cell migration, angiogenesis, transcription factor activation, chromosomal instability, and miRNA expression [57]. Concurrent with increased cellular proliferation, rates of apoptosis are higher among PC compared with PA which have been correlated with decreased expression of the antiapoptotic protein Bcl-2 [58]. However, increased telomerase activity, responsible for cellular immortality, has been described in association with PC development [59].

Proteins involved in DNA repair have also been implicated in pituitary tumour progression. DNA topoisomerase IIα (Topo IIα) plays an important role in DNA replication and mitosis. It is linked with cell cycle activity and rates of cellular proliferation [60]. High expression of Topo IIα has been found in PC, although its correlation with other measures of cellular proliferation in PA, such as Ki67, has been reported in some but not all studies [60, 61]. Low expression of a DNA repair protein, 06-methylguanine-DNA methyltransferase (MGMT) has been reported more frequently in aggressive PA and PC [62, 63]. In a microarray study, low MGMT-expressing pituitary tumours were found to have upregulation of gene sets involved in DNA repair and transcription [64]. A reduction in expression of mismatch repair genes, a family of proteins that correct DNA replication errors, have been implicated in pituitary tumour proliferation and reduced apoptosis [65]. In particular, loss of MSH6 immunoexpression has been described in cases of PC [66, 67].

Activation of epidermal growth factor receptor (EGFR) signalling is an important feature in APT, with both epidermal growth factor (EGF) and EGFR highly expressed in PC [68]. Furthermore, *HER-2/neu* (a member of the EGF receptor family) is commonly amplified in PC [7]. Increased EGFR signalling results in upregulated expression of matrix metalloproteinase 9 (MMP-9), which is critically involved in the degradation of the extracellular matrix and tumour cell migration and is upregulated in PCs and recurrent invasive PAs [69]. MMP-9 also plays a role in stimulating angiogenesis, another attribute of tumours with malignant potential. Vascular endothelial growth factor (VEGF), a potent angiogenic factor involved in endothelial cell proliferation, vascular permeability, and cell mobility, is stimulated by MMP-9 [57, 70]. VEGF expression, and other contributors to tumour angiogenesis including cyclooxygenase-2 (Cox-2) and hypoxia-inducible factor 1-alpha (HIF-1α) have also found to be higher among PC compared with PA [71–73]. Tumour angiogenesis is commonly assessed by studying microvascular density, and while found to be higher in PC compared to PA, substantial overlap exists, so in isolation it may not be a useful prognostic marker [7, 74, 75]. Furthermore, measures of cell proliferation,

such as Ki67, are not positively correlated with markers of angiogenesis, suggesting these two processes are not aligned in malignant transformation [76].

MicroRNA are also thought to play a role in malignant transformation of pituitary tumours. In one study, increased miR-20a, miR-106b, and miR-17-5p were found in metastatic deposits of a PC and six atypical PA and predicted to reduce PTEN and TIMP2 expression thereby promoting metastasis [77]. Utilizing an integrative genomic approach, down-regulation of miR-183 was found in a small cohort of aggressive PA and demonstrated to target KIAA0101 resulting in cell cycle activation [78]. The overexpression of miR-122 and miR-493 have also been found to be upregulated in ACTH-PC [79]. In fact, one of the predicted targets of miR-493 is *LGALS3* mRNA, which regulates the expression of galectin-3 (Gal-3), a β-galactoside-binding protein widely implicated in cancer biology with roles in tumour proliferation, adhesion, angiogenesis, apoptosis and metastasis [80, 81]. Gal-3 is thought to have an important role in pituitary tumour progression and has been found to be increased in ACTH and PRL-PC compared with PA [82, 83].

Management

The management of PC is challenging requiring a multimodal approach overseen by an expert multidisciplinary team. Surgery, radiotherapy, and systemic therapies all have a role in the management of PC. Treatment decisions should consider factors such as disease burden (single metastatic deposits or widely disseminated disease), tumour location, prior therapies, and patient comorbidities.

Surgical Treatment

In patients with PC, surgical resection is rarely curative but substantial tumour debulking can be achieved with significant relief of compressive symptoms [9]. In the case of isolated metastatic disease, complete surgical excision may result in long-term disease-free progression particularly when combined with adjuvant radiotherapy [84, 85]. This has been reported not only for craniospinal disease but also for bone and liver metastatic deposits [85–87]. In addition, repeated surgical resections of recurrent metastases may prolong survival [88]. While advances in endoscopic skull base techniques may assist in wider control of sellar-based disease, frequently transcranial approaches are still required to manage other intracranial disease.

Radiation Therapy

Radiation therapy is most often used in a palliative approach as there is no evidence that radiotherapy improves survival in PC [9]. However, radiotherapy may arrest or slow tumour growth in PC and is often used in an adjuvant setting following surgical excision or debulking of metastatic sites of disease. Chemotherapy with adjuvant radiotherapy is increasingly considered in PC [8]. Not infrequently patients with PC have already had at least one radiotherapy treatment over the course of their disease and further doses may be limited due to the risk of brain necrosis [9]. There has been more experience with use of fractionated external beam radiotherapy (EBRT) in PC but stereotactic radiosurgery (SRS) has also been reported with effect, most commonly with Gamma Knife (GK) [84, 87–90]. The choice of radiation technique and modality (linear accelerator, GK) is often based on safety considerations, such as proximity to optic chiasm, volume of disease and local centre availability [24, 91].

Medical Therapy

Standard medical therapies should be used in conjunction with other treatment in PC at maximally tolerated doses [24]. Lactotroph PC typically display complete resistance to DA, although there are case reports of reduction in prolactin levels and less commonly tumour size [18, 92–94]. High dose DA should be trialled (3.5–11 mg per week) although side effects including nausea and postural hypotension may limit the dose [9, 24]. Cabergoline is more effective compared with bromocriptine, and there is limited experience with quinagolide although it does not appear to be superior in PC [18, 24]. DA resistance may result from decreased dopamine D2 receptor expression, although other mechanisms may also exist [24]. Use of DA has been associated with a reduction in GH and IGF-1 levels and symptomatic improvement in rare cases of somatotroph PC but with no effect on overall disease progression, however use in corticotroph and thyrotroph PC is ineffective [5, 32, 95, 96]. Treatment with tamoxifen has unsuccessful in lactotroph PC [18].

A lack of response to first-generation somatostatin analogue (SSA) therapy (octreotide, lanreotide) is typical in PC, although temporary benefit in a thyrotroph PC case has been described [18, 32]. Resistance to octreotide may correlate in some cases with reduced somatostatin receptor (sst) 2 expression [97]. Sparsely granulated somatotroph PA, typically more aggressive than their densely granulated counterparts, and corticotroph PA more frequently express sst5 and may exhibit better response to second-generation SSA (pasireotide) with broader sst action, particularly against sst5 [24, 98]. In patients with acromegaly not responding to octreotide, a switch to pasireotide, can result in biochemical improvement [99]. However, there is no data about its use in malignant somatroph tumours. In limited cases of corticotroph PC pasireotide has not been effective [24].

In addition to attempts to control tumour growth in PC, medical therapy may be required to reduce hormonal hypersecretion. This is particularly important in patients with corticotroph carcinomas in whom unrestrained hypercortisolism is frequently a direct cause of death [9]. Adrenal-directed inhibitors of glucocorticoid synthesis, such as ketoconazole and metyrapone, are frequently used in this setting.

Chemotherapy

When disease burden is unable to be controlled with surgery and/or radiotherapy, chemotherapy should be considered. Almost always this requires systemic therapy although hepatic chemoembolization afforded temporary local control of liver metastases in a corticotroph PC [100]. Aggressive treatment of patients with PC with chemotherapy has been associated with improved survival [18]. Historically, many different chemotherapy regimens have been used with variable responses, most often limited to partial and transitory disease control [9]. Commonly employed agents have included lomustine (CCNU) in combination with 5-FU or dual cisplatin (or carboplatin) and etoposide [24].

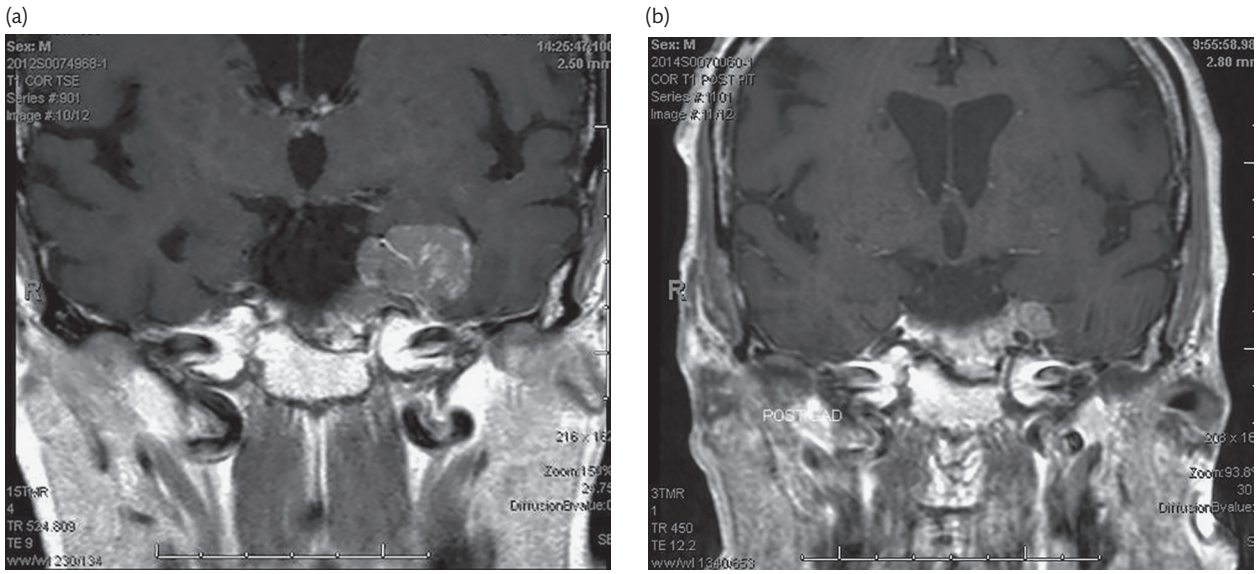


Figure 2.3.13.1 Sustained response to temozolomide.

However, the last decade has seen a major advance in the chemotherapeutic option for PC with the demonstration of response to temozolomide (TMZ) in a significant proportion of cases (Figure 2.3.13.1) [8, 101–103]. TMZ is an oral second-generation imidotetrazine alkylating agent which forms toxic methyl adducts with DNA bases and resulting ineffective DNA repair leads to DNA-strand breaks, cell cycle arrest and apoptosis [104]. In the largest published cohort of PC, comprising 40 patients, 43% of patients demonstrated a complete or partial response to temozolomide, with an additional 18% exhibiting stable disease [8]. Functioning tumours are more likely to respond than non-functioning tumours (Table 2.3.13.2) [8]. TMZ is now recommended as first-line chemotherapy for patients with PC [24]. It is delivered as outpatient oral chemotherapy, most commonly as monotherapy however some centres have proposed using TMZ following capecitabine pretreatment (CAPTEM) although evidence is lacking regarding superior efficacy with this or other combination regimes [24, 105]. The most common TMZ dosing regimen employed is 150–200 mg/m² for 5 consecutive days every 28 days. TMZ in combination with adjuvant radiotherapy may be more effective when maximal radiotherapy doses have not been reached [8, 24]. Response to TMZ therapy, most often concurrent biochemical and radiological improvement, is seen within 3–6 months. The appropriate duration of therapy has not been determined, but it is reasonable to continue for 6 months with longer duration considered in patients receiving clinical benefit. Generally well-tolerated, common side effects include fatigue, nausea/vomiting, and myelosuppression [24]. Sustained responses to TMZ therapy are frequent following treatment discontinuation, although subsequent tumour regrowth occurs in the majority of patients with a median time to progression of 12 months [8]. Low tumour MGMT expression may predict a positive response to TMZ, however cases with low MGMT expression and lack of response have been described [24]. Loss of expression of MSH6 has been reported in such a case [67]. Conversely, high MGMT expression is frequently associated with a poor response to TMZ, but in the absence of other effective chemotherapeutic options a trial of TMZ in such cases is still currently advocated [24].

Second-line therapy for patients progressing during or following cessation of TMZ, remains a challenge. Second courses of TMZ in patients who initially respond is unsuccessful in the majority of cases [24]. Drugs that inhibit the action of EGF/EGFR (e.g.

Table 2.3.13.2 Response to temozolomide based on subtype

Functional subtype	Pathological subtype	Complete response	Partial response	Stable disease	Progressive disease
Functioning (n = 33)	Corticotroph (n = 18)	2	7	1	8
	Lactotroph (n = 14)	1	6	4	3
	Gonadotroph (n = 1)	0	0	1	0
Non-functioning (n = 7)	Corticotroph (n = 1)	0	0	0	1
	Lactotroph (n = 1)	0	0	0	1
	Somatotroph (n = 2)	1	0	0	1
	Immunonegative (n = 3)	0	0	1	2

Adapted with permission from results of ESE survey in McCormack A, Dekkers OM, Petersenn S, Popovic V, Trouillas J, Raverot G, *et al.* Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*. 2018;178(3):265–76.

lapatinib, erlotinib) and VEGF/VEGFR (sunitinib, bevacizumab) pathways have encouraging *in vitro* results and reports of response in a few patients with APT [8, 106–108]. Promising preclinical studies in PA using inhibitors of the PI3K/Akt/mTOR pathway, such as everolimus, have translated into only occasional modest benefit when trialled in patients with APT [8, 109–112]. Finally, immunotherapy may prove useful in management of PC with a recent report of significant tumour shrinkage in a patient with an ACTH-secreting PC [113].

Peptide Receptor Radionuclide Therapy (PRRT)

There is good theoretical rationale in the use of PRRT to treat PC given sst expression and uptake of 68Ga-DOTATATE within pituitary tumours [20]. 111Indium-DPTA-octreotide, 177Lutetium-DOTATATE, 90Yttrium-DOTATOC, and 177Lutetium-DOTATOC have all been reported in a limited collective experience with PRRT for APT, with tumour regression seen in only a couple of patients thus far [24].

Prognosis

Prior to the TMZ era, the mean survival of patients with PC was less than 4 years, although there were cases of long-term survivors [7, 88, 90, 114, 115]. Patients with corticotroph PC and those with systemic compared with central nervous system (CNS) metastases have poorer survival [7]. In patients treated with TMZ, 3–4-year overall survival is reported between 50% and 60% [103, 116]. There is a clear survival benefit in patients who respond to TMZ therapy with a median survival of 44 months among responders compared with 16 months in non-responders [116]. The highest mortality is seen among patients progressing during TMZ therapy [8].

Conclusion

PC are rare and the diagnosis requires the documented presence of metastatic disease remote from the pituitary. Typically, PC arise many years following the diagnosis of PA. During the course of malignant transformation aggressive behaviour is exhibited which is important for clinicians to recognize because of the associated substantial morbidity and mortality. Unusually rapid tumour growth, hormonal hypersecretion in a previously silent PA, resistance to standard medical therapy and markers of increased proliferation such as elevated Ki67 should raise concern for a tumour with malignant potential. Management of PC requires multimodal therapy, including surgery, radiotherapy, and chemotherapy. Prognosis is poor although improved in patients responding to TMZ. Identification of other effective therapeutic options in PC remains an important area for future research and requires collaboration among physicians caring for these rare cancers. The use of tyrosine kinase inhibitors, antiangiogenic agents, immunotherapy, and PRRT may hold promise in the management of PC. An improved understanding of tumour genetics and biomarkers of response may prove essential in patient selection to effective therapy modalities.

REFERENCES

1. Lloyd RV KK, Young WF. *Pituitary Tumors*. DeLellis RH, ed. Lyon: IARC Press; 2004.
2. van der Zwan JM, Mallone S, van Dijk B, *et al*. Carcinoma of endocrine organs: results of the RARECARE project. *Eur J Cancer*, 2012; **48**: 1923–31.
3. Scheithauer BW, Kurtkaya-Yapicier O, Kovacs KT, Young WF, Jr., Lloyd RV. Pituitary carcinoma: a clinicopathological review. *Neurosurgery*, 2005; **56**: 1066–74; discussion 74.
4. Saeger W, Ludecke DK, Buchfelder M, Fahlbusch R, Quabbe HJ, Petersenn S. Pathohistological classification of pituitary tumors: 10 years of experience with the German Pituitary Tumor Registry. *Eur J Endocrinol*, 2007; **156**: 203–16.
5. Mountcastle RB, Roof BS, Mayfield RK, *et al*. Pituitary adenocarcinoma in an acromegalic patient: response to bromocriptine and pituitary testing: a review of the literature on 36 cases of pituitary carcinoma. *Am J Med Sci*, 1989; **298**: 109–18.
6. Negishi K, Suzuki T, Masuda Y, Masugi Y, Teramoto A, Ohama E. [Two autopsy cases of primary pituitary carcinoma]. *No To Shinkei*, 1988; **40**:491–6.
7. Kaltsas GA, Nomikos P, Kontogeorgos G, Buchfelder M, Grossman AB. Clinical review: diagnosis and management of pituitary carcinomas. *J Clin Endocrinol Metab*, 2005; **90**: 3089–99.
8. McCormack A, Dekkers OM, Petersenn S, *et al*. Treatment of aggressive pituitary tumours and carcinomas: results of a European Society of Endocrinology (ESE) survey 2016. *Eur J Endocrinol*, 2018; **178**: 265–76.
9. Heaney AP. Clinical review: pituitary carcinoma: difficult diagnosis and treatment. *J Clin Endocrinol Metab*, 2011; **96**: 3649–60.
10. AbdelBaki MS, Waguespack SG, Salceda V, *et al*. Significant response of pituitary carcinoma to carboplatin, leucovorin and fluorouracil chemotherapy: a pediatric case report and review of the literature. *J Neurooncol*, 2017; **135**: 213–15.
11. Huang AP, Yang SH, Yang CC, Kuo MF, Wu MZ, Tu YK. Malignant prolactinoma with craniospinal metastasis in a 12-year-old boy. *J Neurooncol*, 2008; **90**: 41–6.
12. Philippon M, Morange I, Barrie M, *et al*. Long-term control of a MEN1 prolactin secreting pituitary carcinoma after temozolomide treatment. *Ann Endocrinol (Paris)*, 2012; **73**: 225–9.
13. Tufton N, Roncaroli F, Hadjidemetriou I, *et al*. Pituitary carcinoma in a patient with an SDHB mutation. *Endocr Pathol*, 2017; **28**: 320–5.
14. Bengtsson D, Joost P, Aravidis C, *et al*. Corticotroph pituitary carcinoma in a patient with Lynch syndrome (LS) and pituitary tumors in a nationwide LS cohort. *J Clin Endocrinol Metab*, 2017; **102**: 3928–32.
15. Scheithauer BW, Kovacs K, Nose V, *et al*. Multiple endocrine neoplasia type 1-associated thyrotropin-producing pituitary carcinoma: report of a probable de novo example. *Hum Pathol*, 2009; **40**: 270–8.
16. Maiter D, Delgrange E. Therapy of endocrine disease: the challenges in managing giant prolactinomas. *Eur J Endocrinol*, 2014; **170**: R213–27.
17. Dudziak K, Honegger J, Bornemann A, Horger M, Mussig K. Pituitary carcinoma with malignant growth from first presentation and fulminant clinical course--case report and review of the literature. *J Clin Endocrinol Metab*, 2011; **96**: 2665–9.
18. Kaltsas GA, Grossman AB. Malignant pituitary tumours. *Pituitary*, 1998; **1**: 69–81.

19. Garmes HM, Carvalheira JBC, Reis F, *et al*. Pituitary carcinoma: a case report and discussion of potential value of combined use of Ga-68 DOTATATE and F-18 FDG PET/CT scan to better choose therapy. *Surg Neurol Int*, 2017; **8**: 162.
20. Xiao J, Zhu Z, Zhong D, Ma W, Wang R. Improvement in diagnosis of metastatic pituitary carcinoma by 68Ga DOTATATE PET/CT. *Clin Nucl Med*, 2015; **40**: e129–31.
21. Osamura RY, Lopes MBS, Grossman A, Kontogeorgos G, Trouillas J. Pituitary adenomas. In: Lloyd RV KG, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. Lyon: IARC Press, 2017.
22. Wang H, Liang J, Yong WH, Sullivan P. Metastatic pituitary carcinoma to cervical lymph node: diagnosis by fine needle aspiration and review of the literature. *Acta Cytol*, 2017; **61**: 242–6.
23. Yakoushina TV, Lavi E, Hoda RS. Pituitary carcinoma diagnosed on fine needle aspiration: report of a case and review of pathogenesis. *Cytojournal*, 2010; **7**: 14.
24. Raverot G, Burman P, McCormack A, *et al*. European Society of Endocrinology Clinical Practice Guidelines for the management of aggressive pituitary tumours and carcinomas. *Eur J Endocrinol*, 2018; **178**: G1–G24.
25. Monsalves E, Larjani S, Loyola Godoy B, *et al*. Growth patterns of pituitary adenomas and histopathological correlates. *J Clin Endocrinol Metab*, 2014; **99**: 1330–8.
26. Tanaka Y, Hongo K, Tada T, Sakai K, Kakizawa Y, Kobayashi S. Growth pattern and rate in residual nonfunctioning pituitary adenomas: correlations among tumor volume doubling time, patient age, and MIB-1 index. *J Neurosurg*, 2003; **98**: 359–65.
27. Brue T, Pellegrini I, Priou A, Morange I, Jaquet P. Prolactinomas and resistance to dopamine agonists. *Horm Res*, 1992; **38**: 84–9.
28. Karavitaki N, Ansorge O, Wass JA. Silent corticotroph adenomas. *Arq Bras Endocrinol Metabol*, 2007; **51**: 1314–18.
29. Brown RL, Wollman R, Weiss RE. Transformation of a pituitary macroadenoma into to a corticotropin-secreting carcinoma over 16 years. *Endocr Pract*, 2007; **13**: 463–71.
30. Brown RL, Muzzafar T, Wollman R, Weiss RE. A pituitary carcinoma secreting TSH and prolactin: a non-secreting adenoma gone awry. *Eur J Endocrinol*, 2006; **154**: 639–43.
31. Batisse M, Raverot G, Maqdasy S, *et al*. Aggressive silent GH pituitary tumor resistant to multiple treatments, including temozolomide. *Cancer Invest*, 2013; **31**: 190–6.
32. Mixson AJ, Friedman TC, Katz DA, *et al*. Thyrotropin-secreting pituitary carcinoma. *J Clin Endocrinol Metab*, 1993; **76**: 529–33.
33. Thapar K, Yamada Y, Scheithauer B, Kovacs K, Yamada S, Stefaneanu L. Assessment of mitotic activity in pituitary adenomas and carcinomas. *Endocr Pathol*, 1996; **7**: 215–21.
34. Lopes MBS. The 2017 World Health Organization classification of tumors of the pituitary gland: a summary. *Acta Neuropathol*, 2017; **134**: 521–35.
35. Gejman R, Swearingen B, Hedley-Whyte ET. Role of Ki-67 proliferation index and p53 expression in predicting progression of pituitary adenomas. *Hum Pathol*, 2008; **39**: 758–66.
36. Matsuyama J. Ki-67 expression for predicting progression of postoperative residual pituitary adenomas: correlations with clinical variables. *Neurol Med Chir (Tokyo)*, 2012; **52**: 563–9.
37. Thapar K, Kovacs K, Scheithauer BW, *et al*. Proliferative activity and invasiveness among pituitary adenomas and carcinomas: an analysis using the MIB-1 antibody. *Neurosurgery*, 1996; **38**: 99–106; discussion 7.
38. Trouillas J, Burman P, McCormack AI, *et al*. Aggressive pituitary tumours and carcinomas: two sides of the same coin? *Eur J Endocrinol*, 2018; **178**: C7–9.
39. Turner HE, Wass JA. Are markers of proliferation valuable in the histological assessment of pituitary tumours? *Pituitary*, 1999; **1**: 147–51.
40. Losa M, Franzin A, Mangili F, *et al*. Proliferation index of nonfunctioning pituitary adenomas: correlations with clinical characteristics and long-term follow-up results. *Neurosurgery*, 2000; **47**: 1313–8; discussion 8–9.
41. Kumar K, Macaulay RJ, Kelly M, Pirlot T. Absent p53 immunohistochemical staining in a pituitary carcinoma. *Can J Neurol Sci*, 2001; **28**: 174–8.
42. Kim JS, Lee YS, Jung MJ, Hong YK. The predictive value of pathologic features in pituitary adenoma and correlation with pituitary adenoma recurrence. *J Pathol Transl Med*, 2016; **50**: 419–25.
43. Hentschel SJ, McCutcheon E, Moore W, Durity FA. P53 and MIB-1 immunohistochemistry as predictors of the clinical behavior of nonfunctioning pituitary adenomas. *Can J Neurol Sci*, 2003; **30**: 215–19.
44. Trouillas J, Roy P, Sturm N, *et al*. A new prognostic clinicopathological classification of pituitary adenomas: a multicentric case-control study of 410 patients with 8 years postoperative follow-up. *Acta Neuropathol*, 2013; **126**: 123–35.
45. Wierinckx A, Auger C, Devauchelle P, *et al*. A diagnostic marker set for invasion, proliferation, and aggressiveness of prolactin pituitary tumors. *Endocr Relat Cancer*, 2007; **14**: 887–900.
46. Bi WL, Horowitz P, Greenwald NF, *et al*. Landscape of genomic alterations in pituitary adenomas. *Clin Cancer Res*, 2017; **23**: 1841–51.
47. Rickert CH, Scheithauer BW, Paulus W. Chromosomal aberrations in pituitary carcinoma metastases. *Acta Neuropathol*, 2001; **102**: 117–20.
48. Wierinckx A, Roche M, Raverot G, *et al*. Integrated genomic profiling identifies loss of chromosome 11p impacting transcriptomic activity in aggressive pituitary PRL tumors. *Brain Pathol*, 2011; **21**: 533–43.
49. Chesnokova V, Melmed S. Pituitary senescence: the evolving role of Pttg. *Mol Cell Endocrinol*, 2010; **326**: 55–9.
50. Filippella M, Galland F, Kujas M, *et al*. Pituitary tumour transforming gene (PTTG) expression correlates with the proliferative activity and recurrence status of pituitary adenomas: a clinical and immunohistochemical study. *Clin Endocrinol (Oxf)*, 2006; **65**: 536–43.
51. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet*, 2003; **33**: 49–54.
52. Karga HJ, Alexander JM, Hedley-Whyte ET, Klibanski A, Jameson JL. Ras mutations in human pituitary tumors. *J Clin Endocrinol Metab*, 1992; **74**: 914–19.
53. Pei L, Melmed S, Scheithauer B, Kovacs K, Prager D. H-ras mutations in human pituitary carcinoma metastases. *J Clin Endocrinol Metab*, 1994; **78**: 842–6.
54. Pinto EM, Siqueira SA, Cukier P, Fragoso MC, Lin CJ, de Mendonca BB. Possible role of a radiation-induced p53 mutation in a Nelson's syndrome patient with a fatal outcome. *Pituitary*, 2011; **14**: 400–4.
55. Arzt E, Chesnokova V, Stalla GK, Melmed S. Pituitary adenoma growth: a model for cellular senescence and cytokine action. *Cell Cycle*, 2009; **8**: 677–8.
56. Lidhar K, Korbonits M, Jordan S, *et al*. Low expression of the cell cycle inhibitor p27Kip1 in normal corticotroph cells, corticotroph tumors, and malignant pituitary tumors. *J Clin Endocrinol Metab*, 1999; **84**: 3823–30.

57. Yang Z, Zhang T, Gao H. Genetic aspects of pituitary carcinoma: a systematic review. *Medicine (Baltimore)*, 2016; **95**: e5268.
58. Kulig E, Jin L, Qian X, *et al*. Apoptosis in nontumorous and neoplastic human pituitaries: expression of the Bcl-2 family of proteins. *Am J Pathol*, 1999; **154**: 767–74.
59. Harada K, Arita K, Kurisu K, Tahara H. Telomerase activity and the expression of telomerase components in pituitary adenoma with malignant transformation. *Surg Neurol*, 2000; **53**: 267–74.
60. Vidal S, Kovacs K, Horvath E, *et al*. Topoisomerase II alpha expression in pituitary adenomas and carcinomas: relationship to tumor behavior. *Mod Pathol*, 2002; **15**: 1205–12.
61. Saeger W, Schreiber S, Ludecke DK. Cyclins D1 and D3 and topoisomerase II alpha in inactive pituitary adenomas. *Endocr Pathol*, 2001; **12**: 39–47.
62. Salehi F, Scheithauer BW, Kros JM, *et al*. MGMT promoter methylation and immunoexpression in aggressive pituitary adenomas and carcinomas. *J Neurooncol*, 2011; **104**: 647–57.
63. Lau Q, Scheithauer B, Kovacs K, Horvath E, Syro LV, Lloyd R. MGMT immunoexpression in aggressive pituitary adenoma and carcinoma. *Pituitary*, 2010; **13**: 367–79.
64. McCormack A, Kaplan W, Gill AJ, *et al*. MGMT expression and pituitary tumours: relationship to tumour biology. *Pituitary*, 2013; **16**: 208–19.
65. Uraki S, Ariyasu H, Doi A, *et al*. Reduced expression of mismatch repair genes MSH6/MSH2 directly promotes pituitary tumor growth via the ATR-Chk1 pathway. *J Clin Endocrinol Metab*, 2018; **103**: 1171–9.
66. Hirohata T, Asano K, Ogawa Y, *et al*. DNA mismatch repair protein (MSH6) correlated with the responses of atypical pituitary adenomas and pituitary carcinomas to temozolomide: the national cooperative study by the Japan Society for Hypothalamic and Pituitary Tumors. *J Clin Endocrinol Metab*, 2013; **98**: 1130–6.
67. Murakami M, Mizutani A, Asano S, *et al*. A mechanism of acquiring temozolomide resistance during transformation of atypical prolactinoma into prolactin-producing pituitary carcinoma: case report. *Neurosurgery*, 2011; **68**: E1761–7; discussion E7.
68. Onguru O, Scheithauer BW, Kovacs K, *et al*. Analysis of epidermal growth factor receptor and activated epidermal growth factor receptor expression in pituitary adenomas and carcinomas. *Mod Pathol*, 2004; **17**: 772–80.
69. Turner HE, Nagy Z, Esiri MM, Harris AL, Wass JA. Role of matrix metalloproteinase 9 in pituitary tumor behavior. *J Clin Endocrinol Metab*, 2000; **85**: 2931–5.
70. Zachary I. VEGF signalling: integration and multi-tasking in endothelial cell biology. *Biochem Soc Trans*, 2003; **31**(Pt 6): 1171–7.
71. Alexandraki KI, Munayem Khan M, Chahal HS, *et al*. Oncogene-induced senescence in pituitary adenomas and carcinomas. *Hormones (Athens)*, 2012; **11**: 297–307.
72. Onguru O, Scheithauer BW, Kovacs K, *et al*. Analysis of Cox-2 and thromboxane synthase expression in pituitary adenomas and carcinomas. *Endocr Pathol*, 2004; **15**: 17–27.
73. Vidal S, Horvath E, Kovacs K, Kuroki T, Lloyd RV, Scheithauer BW. Expression of hypoxia-inducible factor-1alpha (HIF-1alpha) in pituitary tumours. *Histol Histopathol*, 2003; **18**: 679–86.
74. Turner HE, Harris AL, Melmed S, Wass JA. Angiogenesis in endocrine tumors. *Endocr Rev*, 2003; **24**: 600–32.
75. Vidal S, Kovacs K, Horvath E, Scheithauer BW, Kuroki T, Lloyd RV. Microvessel density in pituitary adenomas and carcinomas. *Virchows Arch*, 2001; **438**: 595–602.
76. Turner HE, Nagy Z, Gatter KC, Esiri MM, Wass JA, Harris AL. Proliferation, bcl-2 expression and angiogenesis in pituitary adenomas: relationship to tumour behaviour. *Br J Cancer*, 2000; **82**: 1441–5.
77. Wei Z, Zhou C, Liu M, *et al*. MicroRNA involvement in a metastatic non-functioning pituitary carcinoma. *Pituitary*, 2015; **18**: 710–21.
78. Roche M, Wierinckx A, Croze S, *et al*. Deregulation of miR-183 and KIAA0101 in aggressive and malignant pituitary tumors. *Front Med (Lausanne)*, 2015; **2**: 54.
79. Stilling G, Sun Z, Zhang S, *et al*. MicroRNA expression in ACTH-producing pituitary tumors: up-regulation of microRNA-122 and -493 in pituitary carcinomas. *Endocrine*, 2010; **38**: 67–75.
80. Califice S, Castronovo V, Van Den Brule F. Galectin-3 and cancer (Review). *Int J Oncol*, 2004; **25**: 983–92.
81. Zhang HY, Jin L, Stilling GA, *et al*. RUNX1 and RUNX2 upregulate Galectin-3 expression in human pituitary tumors. *Endocrine*, 2009; **35**: 101–11.
82. Riss D, Jin L, Qian X, Bayliss J, *et al*. Differential expression of galectin-3 in pituitary tumors. *Cancer Res*, 2003; **63**: 2251–5.
83. Righi A, Jin L, Zhang S, *et al*. Identification and consequences of galectin-3 expression in pituitary tumors. *Mol Cell Endocrinol*, 2010; **326**: 8–14.
84. Park KS, Hwang JH, Hwang SK, Kim S, Park SH. Pituitary carcinoma with fourth ventricle metastasis: treatment by excision and Gamma-knife radiosurgery. *Pituitary*, 2014; **17**: 514–18.
85. Ayuk J, Natarajan G, Geh JI, Mitchell RD, Gittoes NJ. Pituitary carcinoma with a single metastasis causing cervical spinal cord compression. Case report. *J Neurosurg Spine*, 2005; **2**: 349–53.
86. Joehlin-Price AS, Hardesty DA, Arnold CA, Kirschner LS, Prevedello DM, Lehman NL. Case report: ACTH-secreting pituitary carcinoma metastatic to the liver in a patient with a history of atypical pituitary adenoma and Cushing's disease. *Diagn Pathol*, 2017; **12**: 34.
87. van der Klaauw AA, Kienitz T, Strasburger CJ, Smit JW, Romijn JA. Malignant pituitary corticotroph adenomas: report of two cases and a comprehensive review of the literature. *Pituitary*, 2009; **12**: 57–69.
88. Pernicone PJ, Scheithauer BW, Sebo TJ, *et al*. Pituitary carcinoma: a clinicopathologic study of 15 cases. *Cancer*, 1997; **79**: 804–12.
89. Gaffey TA, Scheithauer BW, Lloyd RV, *et al*. Corticotroph carcinoma of the pituitary: a clinicopathological study. Report of four cases. *J Neurosurg*, 2002; **96**: 352–60.
90. Koyama J, Ikeda K, Shose Y, Kimura M, Obora Y, Kohmura E. Long-term survival with non-functioning pituitary carcinoma—case report. *Neurol Med Chir (Tokyo)*, 2007; **47**: 475–8.
91. Lenders N, McCormack A. Malignant transformation in non-functioning pituitary adenomas (pituitary carcinoma). *Pituitary*, 2018; **21**: 217–29.
92. Hurel SJ, Harris PE, McNicol AM, Foster S, Kelly WF, Baylis PH. Metastatic prolactinoma: effect of octreotide, cabergoline, carboplatin and etoposide; immunocytochemical analysis of proto-oncogene expression. *J Clin Endocrinol Metab*, 1997; **82**: 2962–5.
93. Muhr C, Bergstrom M, Lundberg PO, *et al*. Malignant prolactinoma with multiple intracranial metastases studied with positron emission tomography. *Neurosurgery*, 1988; **22**: 374–9.
94. Gollard R, Kosty M, Cheney C, Copeland B, Bordin G. Prolactin-secreting pituitary carcinoma with implants in the

cheek pouch and metastases to the ovaries. A case report and literature review. *Cancer*, 1995; **76**: 1814–20.
95. Yamashita S, Izumi M, Nagataki S. Acromegaly and pituitary carcinoma. *Ann Intern Med*, 1992; **117**: 1057–8.
96. Kaiser FE, Orth DN, Mukai K, Oppenheimer JH. A pituitary parasellar tumor with extracranial metastases and high, partially suppressible levels of adrenocorticotropin and related peptides. *J Clin Endocrinol Metab*, 1983; **57**: 649–53.
97. Casar-Borota O, Heck A, Schulz S, *et al*. Expression of SSTR2a, but not of SSTRs 1, 3, or 5 in somatotroph adenomas assessed by monoclonal antibodies was reduced by octreotide and correlated with the acute and long-term effects of octreotide. *J Clin Endocrinol Metab*, 2013; **98**: E1730–9.
98. Iacovazzo D, Carlsen E, Lugli F, *et al*. Factors predicting pasireotide responsiveness in somatotroph pituitary adenomas resistant to first-generation somatostatin analogues: an immunohistochemical study. *Eur J Endocrinol*, 2016; **174**: 241–50.
99. Bronstein MD, Fleseriu M, Neggers S, *et al*. Switching patients with acromegaly from octreotide to pasireotide improves biochemical control: crossover extension to a randomized, double-blind, phase III study. *BMC Endocr Disord*, 2016; **16**: 16.
100. Lormeau B, Miossec P, Sibony M, Valensi P, Attali JR. Adrenocorticotropin-producing pituitary carcinoma with liver metastasis. *J Endocrinol Invest*, 1997; **20**: 230–6.
101. Raverot G, Sturm N, de Fraipont F, *et al*. Temozolomide treatment in aggressive pituitary tumors and pituitary carcinomas: a French multicenter experience. *J Clin Endocrinol Metab*, 2010; **95**: 4592–9.
102. Bush ZM, Longtine JA, Cunningham T, *et al*. Temozolomide treatment for aggressive pituitary tumors: correlation of clinical outcome with O6-methylguanine methyltransferase (MGMT) promoter methylation and expression. *J Clin Endocrinol Metab*, 2010; **95**: E280–90.
103. Losa M, Bogazzi F, Cannavo S, *et al*. Temozolomide therapy in patients with aggressive pituitary adenomas or carcinomas. *J Neurooncol*, 2016; **126**: 519–25.
104. McCormack AI, Wass JA, Grossman AB. Aggressive pituitary tumours: the role of temozolomide and the assessment of MGMT status. *Eur J Clin Invest*, 2011; **41**: 1133–48.
105. Zacharia BE, Gulati AP, Bruce JN, *et al*. High response rates and prolonged survival in patients with corticotroph pituitary tumors and refractory Cushing disease from capecitabine and temozolomide (CAPTEM): a case series. *Neurosurgery*, 2014; **74**: E447–55; discussion E55.
106. Fukuoka H, Cooper O, Ben-Shlomo A, *et al*. EGFR as a therapeutic target for human, canine, and mouse ACTH-secreting pituitary adenomas. *J Clin Invest*, 2011; **121**: 4712–21.
107. Cooper O, Mamelak A, Bannykh S, *et al*. Prolactinoma ErbB receptor expression and targeted therapy for aggressive tumors. *Endocrine*, 2014; **46**: 318–27.
108. Ortiz LD, Syro LV, Scheithauer BW, *et al*. Anti-VEGF therapy in pituitary carcinoma. *Pituitary*, 2012; **15**: 445–9.
109. Lee M, Wiedemann T, Gross C, *et al*. Targeting PI3K/mTOR signaling displays potent antitumor efficacy against nonfunctioning pituitary adenomas. *Clin Cancer Res*, 2015; **21**: 3204–15.
110. Chanal M, Chevallier P, Raverot V, *et al*. Differential effects of PI3K and dual PI3K/mTOR inhibition in rat prolactin-secreting pituitary tumors. *Mol Cancer Ther*, 2016; **15**: 1261–70.
111. Donovan LE, Arnal AV, Wang SH, Odia Y. Widely metastatic atypical pituitary adenoma with mTOR pathway STK11(F298L) mutation treated with everolimus therapy. *CNS Oncol*, 2016; **5**: 203–9.
112. Zhang D, Way JS, Zhang X, *et al*. Effect of everolimus in treatment of aggressive prolactin-secreting pituitary adenomas. J Clin Endocrinol Metab, 2019.
113. Lin AL, Jonsson P, Tabar V, Yang TJ, Cuaron J, Beal K, *et al*. Marked response of a hypermutated adrenocorticotropic hormone-secreting pituitary carcinoma to ipilimumab and nivolumab. *J Clin Endocrinol Metab*, 2019; **104**: 1929–36.
114. Landman RE, Horwith M, Peterson RE, Khandji AG, Wardlaw SL. Long-term survival with ACTH-secreting carcinoma of the pituitary: a case report and review of the literature. *J Clin Endocrinol Metab*, 2002; **87**: 3084–9.
115. Tanaka T, Kato N, Aoki K, *et al*. Long-term follow-up of growth hormone-producing pituitary carcinoma with multiple spinal metastases following multiple surgeries: case report. *Neurol Med Chir (Tokyo)*, 2013; **53**: 707–11.
116. Lasolle H, Cortet C, Castinetti F, *et al*. Temozolomide treatment can improve overall survival in aggressive pituitary tumors and pituitary carcinomas. *Eur J Endocrinol*, 2017; **176**: 769–77.

2.3.14 Pituitary Incidentalomas

Niki Karavitaki, Shu Teng Chai, and Shahzada Ahmed

Definition

A pituitary incidentaloma is defined strictly as a 'totally asymptomatic non-functional tumour, clinically and biochemically silent, which was discovered incidentally in a patient who is asymptomatic' or, less strictly, a pituitary mass discovered in the course of evaluation for an unrelated problem [1–3]. Based on the second definition, the term incidentaloma may not be appropriate to many of these lesions, as an incidentally detected macroadenoma may still be clinically relevant.

Box 2.3.14.1 Differential diagnosis of pituitary incidentalomas

Pituitary adenomas (functioning, non-functioning)
Germ cell tumours (germinoma, dermoid, teratoma)
Gliomas
Craniopharyngioma
Rathke's cleft cyst
Meningioma
Chordoma
Primary lymphoma
Pituitary carcinoma
Metastasis
Pituitary hyperplasia
Arachnoid cyst
Inflammatory lesions (e.g. sarcoidosis, hypophysitis, histiocytosis X)
Infectious lesions (e.g. abscess, tuberculosis)
Haemorrhage
Aneurysm
Cavernous angioma
Cavernous sinus thrombosis

Frequency of Detection

Autopsy Studies

The mean frequency of adenoma detection in the largest collection of autopsy case series, which examined 18 902 pituitaries, was 10.7% (ranging from 1.5% to 31%). All but seven of these lesions had a diameter smaller than 10 mm, indicating that almost all of the pituitary adenomas found in autopsy series are microadenomas. There was also no age or gender preponderance [4]. A detailed analysis of immunohistochemical staining of 334 pituitary adenomas in 316 pituitaries out of 3048 autopsy cases revealed 39.5% immunopositivity for prolactin, followed by 13.8% for ACTH, 6.6% for gonadotropins, 2.1% for growth hormone (GH), and 0.6% for thyroid-stimulating hormone (TSH) and α-subunit, respectively. Negative immunostaining was reported in 22.5%, while plurihormonal adenomas accounted for 2.7% [5].

Imaging Studies

The detection of incidental pituitary lesions on imaging depends on the modality used, the administration of contrast agents and

Table 2.3.14.1 Natural history of pituitary incidentalomas

Series	Number of subjects	Size of lesion	Follow-up duration	Outcome
Reincke *et al.*, 1990 [16]	14	≥1 cm, n = 7 <1 cm, n = 7	Median 22 months	Increase (all with intact visual acuity): • 14% of those with mass <1 cm • 29% of those with mass >1 cm Regression: • 14% of those with mass <1 cm • 0% of those with mass >1 cm
Donovan & Corenblum, 1995 [17]	31	>1 cm, n = 16 <1 cm, n = 15	<1 cm mean 6.7 years >1 cm mean 6.1 years	Increase: • 25% of those with mass >1 cm (one developed visual deterioration and was treated surgically) • 0% of those with mass <1 cm
Nishizawa *et al.*, 1998 [18]	28	>1 cm, n = 28	Mean 5.6 years	Increase: • 7%
Feldkamp *et al.*, 1999 [19]	50	>1 cm, n = 19 <1 cm, n = 31	Mean 2.7 years	Increase (none developed visual field defect): • 3% of those with mass <1 cm • 26% of those with mass >1 cm Decrease: • 3% of those with mass <1 cm • 5% of those with mass >1 cm
Sanno *et al.*, 2003 [20]	242	–	Mean 26.9 months	Increase: • 12.4% (among them, 67% had initial size >1 cm and none had shown visual disturbance) Decrease: • 12% No change: • 74.4%
Day *et al.*, 2005 [21]	46	>1 cm, n = 29 <1 cm, n = 17	Mean 3.2 years	Among those not treated surgically or medically at the outset, significant tumour growth was seen in: • 9% of those with mass <1 cm • 14% of those with mass >1 cm
Arita *et al.*, 2006 [22]	42	<1 cm, n = 5 ≥1 cm, n = 37	61.9 months	Lesion's height surpassing 110% of its initial measured height (48% had visual deterioration, diplopia, or hypopituitarism): • 40% of those with mass <1 cm • 51% of those with mass >1 cm Apoplexy developed in 9.5% of total group

Table 2.3.14.2 Natural history of non-operated presumed non-functioning pituitary adenomas

Series	Number of subjects	Size of lesion	Follow-up duration	Outcomes
Dekkers *et al.*, 2007 (23)	28	All macroadenomas	85 months	Increase: 50% (out of which 64% had worsening visual field defect) Decrease: 29% Stable: 21%
Karavitaki *et al.*, 2007 [24]	40	Macroadenoma n = 24 Microadenoma n = 16	Mean 42 months	Increase: • 50% of macroadenomas (67% of them had visual field defects) • 12.5% of microadenomas 48-month probability of enlargement: • 44% for macroadenomas • 19% for microadenomas Stable: • 33.3% of macroadenomas • 81.3% of microadenomas Decrease: • 16.7% of macroadenomas • 6.3% of microadenomas
Lenders *et al.*, 2016 [25]	50	Macroadenoma n = 23 Microadenoma n = 27	Mean 36 months	≥20% increase in volume: • 39% of macroadenomas • 7% of microadenomas

the slice thickness used [6]. Pituitary incidentalomas have been reported with increasing frequency paralleling the advances in imaging techniques and the wider use of brain scans. The prevalence of pituitary incidentalomas found by CT ranges from 3.7% to 20% and of those found by MRI is 10% [7]. Similar to autopsy series, radiological studies on subjects undergoing imaging for reasons not related to a pituitary problem have found frequency of macroadenomas between 0.16% and 0.20% [2].

Differential Diagnosis

The most common aetiology of pituitary incidentalomas is pituitary adenomas, the majority (>75%) of which are non-functioning [8]. Apart from adenomas, the differential diagnosis of pituitary incidentaloma is extensive (Box 2.3.14.1).

Diagnostic Evaluation

A cost-effective approach is required to exclude potentially harmful conditions, as well as to decrease patient anxiety.

Imaging techniques are helpful in the diagnostic evaluation of an incidentally found pituitary mass. MRI has been proven to be more sensitive than CT for the detection of pituitary adenomas [9]. CT is superior in detecting calcifications and is, therefore, helpful for the diagnosis of craniopharyngiomas and meningiomas [10]. There are no studies providing correlation between MRI and pathological features of pituitary masses. Nevertheless, specific MRI features of a number of sellar masses (including meningiomas, metastatic disease, craniopharyngiomas, Rathke's cleft cysts, arachnoid cysts, hypophysitis, abscess) may be useful in the differentiation from an adenoma [11–13].

Given that the most common lesion in the sellar area is a pituitary adenoma, assessment for hormonal hypersecretion is recommended. This includes clinical evaluation for relevant manifestations combined with biochemical screening for hormonal excess:

a) Prolactin (2–3 measurements) for prolactinomas.
b) Insulin-like growth factor-I, which in cases of suspected acromegaly could be combined with an oral glucose tolerance test.
c) 24-hour urine free cortisol and overnight dexamethasone suppression test for Cushing's disease.
d) TSH, free T4, and free T3 for TSH-secreting tumour.
e) FSH, LH, alpha subunit, oestradiol, or testosterone for functioning gonadotroph adenomas.

Smaller incidentalomas (5 mm or less in diameter) do not generally compromise the pituitary function and routine screening may not be necessary; however, evaluation for hypopituitarism is recommended for patients with larger lesions [3]. Assessment for the presence of diabetes insipidus is suggested particularly when there is suspicion of craniopharyngiomas [14] and pituitary metastases [15].

All patients who have a pituitary lesion abutting or compressing the optic nerves or chiasm on MRI, should undergo a formal visual field (VF) examination [3].

Natural History

The studies on the natural history of pituitary incidentalomas are limited. The shortcomings of currently available literature include patient heterogeneity, relatively small patient series, and lack of prolonged follow-up. Their results including masses of various pathologies are summarized in Table 2.3.14.1. Overall, the published data suggest that lesions smaller than 1 cm follow a benign course. In contrast, masses greater than 1 cm are associated with higher risk of enlargement often leading to pressure effects and requiring neurosurgical intervention. These data are in accord with the reported outcomes of non-operated presumed non-functioning pituitary adenomas (Table 2.3.14.2 and Figure 2.3.14.1). Indeed, a systematic review and meta-analysis of 11 cohort studies of significant heterogeneity showed that non-functioning pituitary adenomas and incidentalomas which were bigger than 1 cm and solid had a greater event rate of tumour

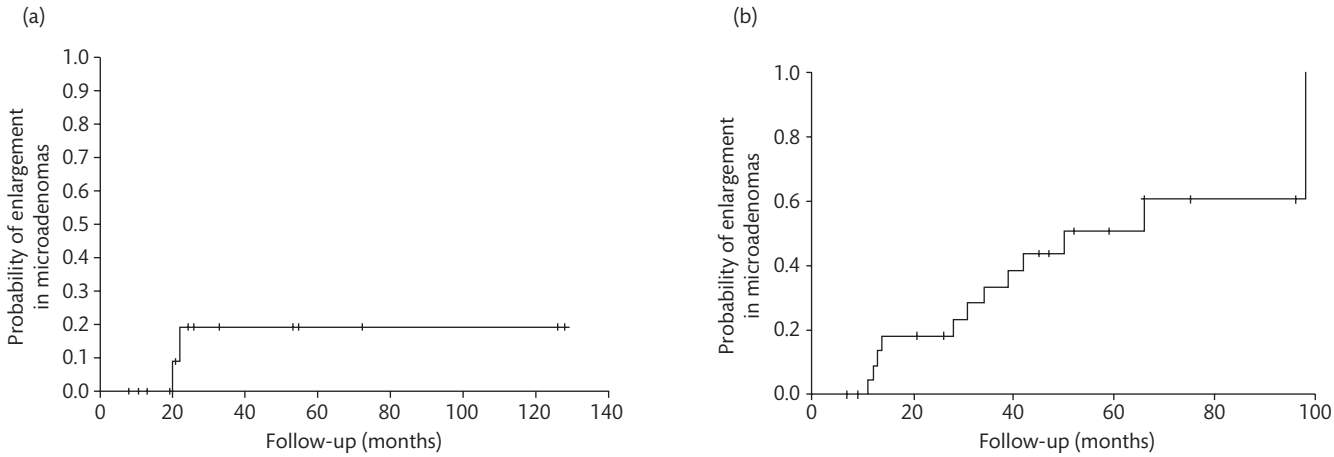


Figure 2.3.14.1 Natural history of (a) microadenoma versus (b) macroadenoma.
From Karavitaki N, Collison K, Halliday J, et al. What is the natural history of nonoperated nonfunctioning pituitary adenomas? *Endocrinol (Oxf)*. 2007; 67:938–943. Reproduced with permission from John Wiley and Sons.

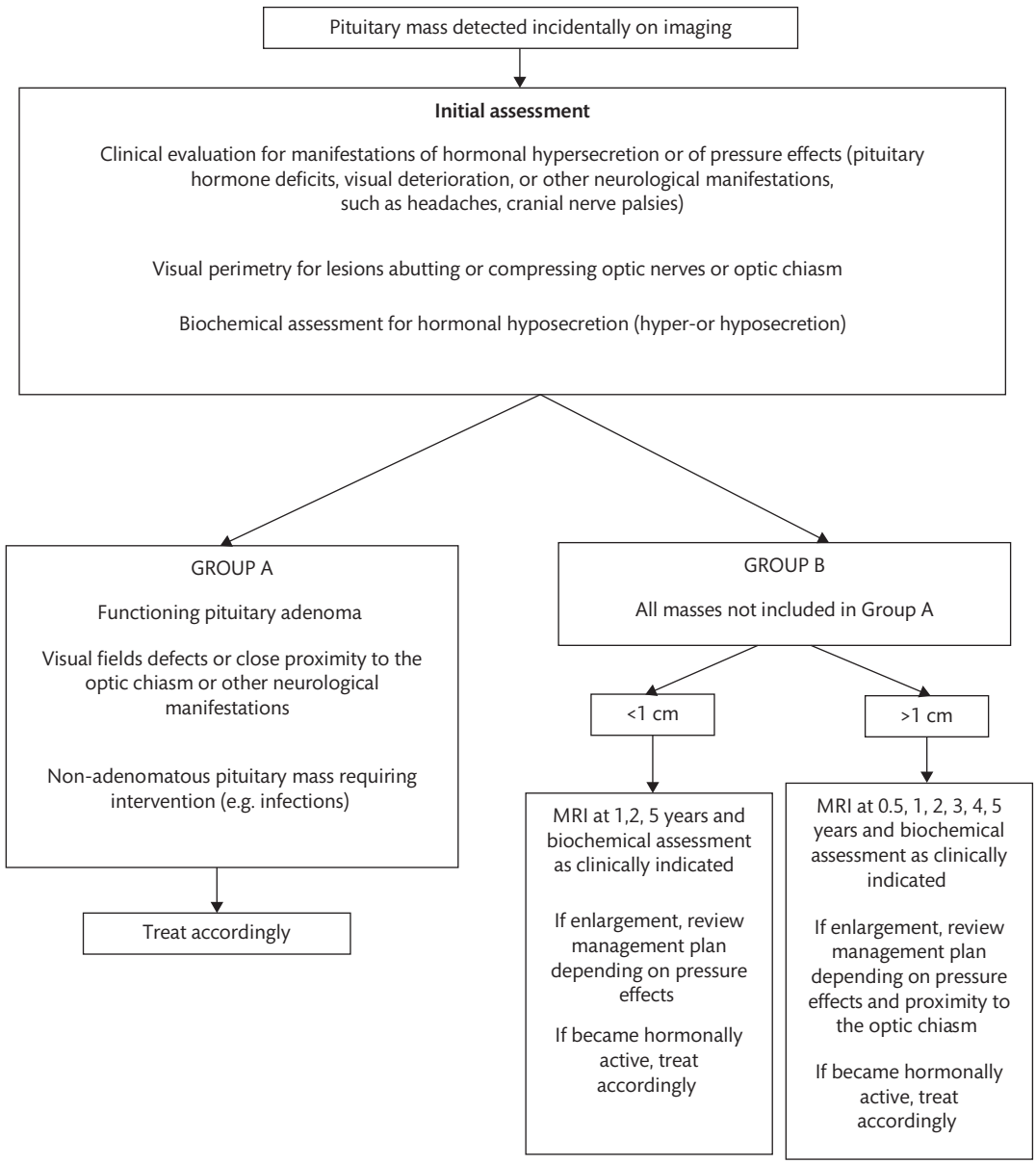


Figure 2.3.14.2 Algorithm for the evaluation and management of pituitary incidentalomas.

growth compared to smaller and cystic lesions, albeit the quality of evidence was deemed low [26].

It is noteworthy that majority of the reported studies have a follow-up period of less than 5 years, a duration which is deemed short considering the natural history of pituitary tumours. Hence, more studies with longer observation periods are needed to clarify the behaviour and outcomes of pituitary incidentalomas.

Despite the rare occurrence of pituitary apoplexy, its risk should always be taken into consideration, particularly in patients exposed to predisposing factors (e.g. anticoagulation).

The potential of incidentally found adenomas to become hormonally active at a later stage has not been fully elucidated and reliable data on the chance of developing relevant endocrinopathy is lacking.

Management

The Endocrine Society Clinical Practice Guideline recommends surgery by an experienced pituitary surgeon for patients with a pituitary incidentaloma who present with VF deficit or other visual abnormalities due to the lesion, lesion abutting, or compressing optic nerves or chiasm on MRI, pituitary apoplexy or hypersecreting tumours other than prolactinomas [3]. In cases not meeting the aforementioned criteria, conservative management with regular follow-up is recommended [3]. As the long-term natural history of incidentally detected pituitary masses is still unclear, the current consensus is largely based on expert opinion [3, 12].

Our proposed algorithm for the initial evaluation and management of these lesions is shown in **Figure 2.3.14.2**. Since the optimum duration of imaging surveillance is unknown, decisions on monitoring of a mass not showing enlargement 5 years after its detection should be individualized. Finally, cost remains an important factor in determining the optimum follow-up strategy of pituitary incidentaloma. Randall *et al.* reported the total cost of endocrine tests, imaging, and physician fees incurred during a year of follow-up of pituitary incidentaloma to be USD 6215.28 for males and USD 6061.78 for females. This was expected to account for a conservative estimate of nearly USD 7 million on the United States healthcare system in year 2005 [27]. It is important to note that this estimate does not include expenditure involved during further follow-ups. Hence, the cost-effectiveness of the suggested or other approaches remains to be elucidated.

REFERENCES

1. Mirilas P, Skandalakis JE. Benign anatomical mistakes: incidentaloma. *Am Surg*, 2002; **68**: 725–40.
2. Krikorian A, Aron D. Evaluation and management of pituitary incidentalomas—revisiting an acquaintance. *Nat Clin Pract Endocrinol Metab*, 2006; **2**: 138–45.
3. Freda PU, Beckers AM, Katznelson L, *et al.* Pituitary incidentaloma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 894–904.
4. Molitch ME. Pituitary tumours: pituitary incidentalomas. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 667–75.
5. Buurman H, Saeger W. Subclinical adenomas in postmortem pituitaries: classification and correlations to clinical data. *Eur J Endocrinol*, 2006; **154**: 753–8.
6. Turner HE, Moore NR, Byrne JV, Wass JAH. Pituitary, adrenal and thyroid incidentalomas. *Endocr Relat Cancer*, 1998; **5**: 131–50.
7. Aron DC, Howlett TA. Pituitary incidentalomas. *Endocrinol Metab Clin North Am*, 2000; **29**: 205–21.
8. Famini P, Maya MM, Melmed S. Pituitary magnetic resonance imaging for sellar and parasellar masses: ten-year experience in 2598 patients. *J Clin Endocrinol Metab*, 2011; **96**: 1633–41.
9. Johnson MR, Hoare RD, Cox T, *et al.* The evaluation of patients with suspected pituitary microadenoma: computed tomography compared to magnetic resonance imaging. *Endocrinol (Oxf)*, 1992; **36**: 335–8.
10. Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. *Endocr Rev*, 2006; **27**: 371–97.
11. Connor SE, Penney CC. MRI in the differential diagnosis of a sellar mass. *Clin Radiol*, 2003; **58**: 20–31.
12. Galland F, Vantyghem MC, Cazabat L, *et al.* Management of nonfunctioning pituitary incidentaloma. *Annales d'Endocrinologie*, 2015; **76**: 191–200.
13. Vasilev V, Rostomyan L, Daly AF, *et al.* Pituitary 'incidentaloma ': neuroradiological assessment and differential diagnosis. *Eur J Endocrinol*, 2016; **175**: R171–84.
14. Müller HL. Craniopharyngioma. *Endocr Rev*, 2014; **35**: 513–43.
15. Komninos J, Vlassopoulou V, Protopapa D, *et al.* Tumors metastatic to the pituitary gland: case report and literature review. *J Clin Endocrinol Metab*, 2004; **89**: 574–80.
16. Reincke M, Allolio B, Saeger W, *et al.* The 'incidentaloma' of the pituitary gland. Is neurosurgery required? *JAMA*, 1990; **263**: 2772–6.
17. Donovan LE, Corenblum B. The natural history of the pituitary incidentaloma. *Arch Intern Med*, 1995; **155**: 181–3.
18. Nishizawa S, Ohta S, Yokoyama T, Uemura K. Therapeutic strategy for incidentally found pituitary tumours ('pituitary incidentalomas'). *Neurosurgery*, 1998; **43**: 1344–50.
19. Feldkamp J, Santen R, Harms E, *et al.* Incidentally discovered pituitary lesions: high frequency of macroadenomas and hormone-secreting adenomas-results of a prospective study. *Endocrinol (Oxf)*, 1999; **51**: 109–13.
20. Sanno N, Oyama K, Tahara S, *et al.* A survey of pituitary incidentaloma in Japan. *Eur J Endocrinol*, 2003; **149**: 123–7.
21. Day PF, Guitelman M, Artise R, *et al.* Retrospective multicentric study of pituitary incidentalomas. *Pituitary*, 2005; **7**: 145–8.
22. Arita K, Tominaga A, Sugiyama K, *et al.* Natural course of incidentally found non-functioning pituitary adenoma, with special reference to pituitary apoplexy during follow-up examination. *J Neurosurg*, 2006; **104**: 884–91.
23. Dekkers OM, Hammer S, de Keizer RJW, *et al.* The natural course of nonfunctioning pituitary macroadenomas. *Eur J Endocrinol*, 2007; **156**: 217–24.
24. Karavitaki N, Collison K, Halliday J, *et al.* What is the natural history of nonoperated nonfunctioning pituitary adenomas? *Endocrinol (Oxf)*, 2007; **67**:938–43.
25. Lenders N, Ikeuchi S, Russel AW, *et al.* Longitudinal evaluation of the natural history of conservatively managed nonfunctioning pituitary adenomas. *Clin Endocrinol (Oxf)*, 2016; **84**: 222–8.
26. Ferna´ndez-Balsells MM, Murad MH, Barwise A, *et al.* Natural history of nonfunctioning pituitary adenomas and incidentalomas: a systematic review and metaanalysis. *J Clin Endocrinol Metab*, 2001; **96**: 905–12.
27. Randall BR, Kraus KL, Simard MF, Couldwell WT. Cost of evaluation of patients with pituitary incidentaloma. *Pituitary*, 2010; **13**: 383–4.

2.4

Aetiology, Pathogenesis, and Management of Diseases of the Hypothalamus

2.4.1 Hypothalamic Dysfunction (Hypothalamic Syndromes)

Hoong-Wei Gan, Manuela Cerbone, and Mehul T. Dattani

Background

The hypothalamus consists of a dense congregation of nuclei lying superior to the pituitary gland and sella turcica, inferior to the thalamus, and surrounding the third ventricle, extending from the optic chiasm anteriorly to the mammillary bodies posteriorly. Despite its small dimensions (approximately the size of an almond in an adult), it has numerous axonal projections to the cerebral cortex, brainstem, reticular formation, limbic system and autonomic nervous system, and is responsible for regulating a wide variety of functions, including appetite, metabolism, thirst, circadian rhythms, arousal, memory, and behaviour. The ill definition of some of these nuclei, and the limited resolution of current neuroradiological techniques, means that the circuitry of these various hypothalamic pathways is incompletely understood. Hypothalamic gene expression studies result in downstream effects on multiple neuronal subtypes, and, in some instances, lethality, making it additionally difficult to decipher the role of these various genes in the early development of this region [1, 2]. The hypothalamus is the 'master regulator' of the neuroendocrine system, and damage to or maldevelopment of this area can have widespread consequences on a variety of homeostatic mechanisms.

Development of the Hypothalamus

Like pituitary development, much of our current understanding of hypothalamic development in humans is based on mouse and rat models, although they are less well-characterized than that of the pituitary (see Chapter 2.3.1) [1, 2]. Additionally, despite several rodent models with hypothalamic gene mutations already described in the literature, only two of these genes have thus far been demonstrated to be associated with human hypothalamic disease [3, 4].

The primordium of the hypothalamus and infundibulum originates from the ventral diencephalon, and is located posterior to the anterior pituitary placode which appears as a localized ectodermal thickening in the anterior neural ridge of the developing embryo at mouse embryonic day E7.5 [5, 6]. Sonic hedgehog (*Shh*) signalling has been shown to be crucial for the early development of the hypothalamus [7], and is subsequently downregulated with the concomitant upregulation of bone morphogenetic proteins (BMPs) [8], *Fgf8*, and *Fgf10* [1]. The hypothalamic primordium itself becomes evident in the rostral part of the neural plate at E9.5, and neurogenesis commences at E10, concomitant with the expression of the hypothalamic patterning genes *Sim1, Arnt2, Otp, Nkx2.1*, and *Nr5a1* which peak at E12–14 [9]. Neurogenesis is complete by E16, but development is only completed in the postnatal period [9].

Sim1 is expressed in the paraventricular (PVN), supraoptic (SON) and anterior periventricular nuclei (PeVN) from E10.5 and continues postnatally. Homozygous mutants demonstrate significant anterior hypothalamic hypoplasia and early postnatal death [10]. SIM1 heterodimerizes with ARNT2, and homozygous *Arnt2* knockout mice therefore demonstrate a very similar phenotype [11, 12]. Both *Sim1* and *Arnt2* mutants additionally exhibit absent *Brn2* expression, which encodes a transcription factor required for differentiation of corticotrophin-releasing hormone (CRH), arginine vasopressin (AVP), and oxytocin (OXT)-producing neurons [1, 2, 11].

The role of other early hypothalamic developmental genes is beginning to be elucidated through other mouse models. *Otp* null

mice demonstrate absence of somatostatin, AVP, OXT, CRH, and thyrotropin-releasing hormone (TRH), and a failure of *Brn2* expression [13, 14]. *Nkx2.1* null mice demonstrate agenesis of Rathke's pouch and the ventromedial (VMN) and arcuate (ARC) nuclei, as well as lung and thyroid defects [15]. *Nr5a1* null mice demonstrate failure of migration of VMN precursors and disorganization of the dorsomedial VMN, alongside the well-described defects in pituitary gonadotroph development, and agenesis of adrenal and gonadal tissue [2, 16, 17]. Double knockout $Hmx2^{-/-}/Hmx3^{-/-}$ mice demonstrate severe growth hormone-releasing hormone (GHRH) deficiency in the ARC and postnatal dwarfism [18]. *Sox3*, a member of the Sry-related high mobility group (HMG) box group of transcription factors, has been well-described in association with X-linked congenital hypopituitarism, but is more highly expressed in the developing hypothalamus than the anterior pituitary [2]. Changes in *SOX3* dosage in humans have been described to be associated with not just anterior pituitary hypoplasia, but also ectopic posterior pituitary and an absent infundibulum, indicating its role in hypothalamic development [19].

The close association between the developing Rathke's pouch (the prospective anterior pituitary) and the ventral diencephalon (the prospective infundibulum and posterior pituitary) is crucial to the development of both the hypothalamus and pituitary gland [1, 20]. As the Rathke's pouch invaginates from the oral ectoderm, the infundibulum evaginates from the ventral diencephalon at E10.5 to come into contact with it; this apposition is maintained thereafter throughout development [1, 2]. Terminal axonal projections from the prospective supraoptic and paraventricular hypothalamic nuclei travel with the infundibulum to the developing posterior pituitary, where they ultimately secrete AVP and OXT directly into the peripheral circulation [20]. Gonadotrophin-releasing hormone (GnRH) neurons migrate from the olfactory placode to the developing hypothalamus, with GnRH being expressed in the hypothalamus from E10.5 [1, 2, 21, 22]. This is followed by dopamine at E10.5 [23], GHRH at E11 [21, 24], TRH at E13 [25, 26], CRH at E13.5 [24, 27], and somatostatin at E17 [28].

Structure of the Mature Human Hypothalamus

The mature adult hypothalamus is a highly neuron-dense area measuring about 0.7 cm^3 on either side of the third ventricle [29]. It is subdivided into the preoptic, anterior, tuberal and mammillary regions rostrocaudally, and into the periventricular, medial, and lateral regions mediolaterally [2, 29]. The various nuclei can be structurally classified by their location within the mediolateral divisions of the hypothalamus (Table 2.4.1.1).

The hypothalamus is intimately connected to the pituitary gland via two routes, both of which are carried by the pituitary stalk (infundibulum). The first of these, the hypothalamo-hypophyseal portal vascular system, arising from the superior hypophyseal arteries and draining into the vein of Galen, is responsible for transporting hypothalamic stimulatory and inhibitory releasing hormones secreted by parvocellular neurons to the anterior pituitary where they regulate secretion of pituitary hormones into the peripheral circulation. Other parvocellular neurons project to other areas of the brain to regulate appetite, metabolism, thirst, circadian rhythms, arousal, memory, and behaviour. The second pathway consists of direct axonal projections from magnocellular neurons arising from the hypothalamic PVN and SON to the posterior pituitary gland, where the neurohypophyseal hormones arginine-vasopressin (AVP) and oxytocin (OXT) are secreted directly into the peripheral circulation.

Table 2.4.1.1 Hypothalamic nuclei subdivided by their mediolateral, rostrocaudal then supero-inferior positions within the hypothalamus and their roles in controlling various bodily functions

Division	Nucleus	Role
Periventricular	Suprachiasmatic nucleus (SCN)	Control of circadian rhythms
	Paraventricular nucleus (PVN)	Release of TRH, CRH, oxytocin, and AVP
	Periventricular nucleus (PeVN)	Release of somatostatin
	Arcuate nucleus (ARC)	Release of GHRH and dopamine, control of appetite
Medial	Medial preoptic nucleus (MPN)	Release of GnRH, thermoregulation
	Anterior hypothalamic nucleus (AHN)	Thermoregulation
	Supraoptic nucleus (SON)	Release of AVP and oxytocin
	Dorsomedial nucleus (DMN)	Control of the autonomic system
	Ventromedial nucleus (VMN)	Control of appetite and satiety
	Mamillary nucleus	Memory
Lateral	Preoptic area (POA)	Control of thirst
	Lateral hypothalamic area (LHA)	Control of feeding and arousal via release of orexin
	Tuberomammillary nucleus (TMN)	Control of arousal

Clinical Presentations and Management of Hypothalamic Dysfunction

See Box 2.4.1.1.

Central Diabetes Insipidus (CDI)

CDI is the archetypal form of hypothalamo-pituitary endocrine dysfunction arising distinctly from the hypothalamus, as a result

Box 2.4.1.1 Clinical features of the hypothalamic syndrome

Endocrine features
- Hypopituitarism (especially central diabetes insipidus, central precocious puberty, hypogonadotropic hypogonadism)
- Hypothalamic obesity
- Diencephalic syndrome/hypothalamic anorexia

'Non-endocrine' features
- Temperature dysregulation
- Sleep dysregulation
- Behavioural difficulties
- Memory dysfunction
- Autonomic dysfunction

of congenital or acquired AVP neuronal deficiency, dysfunction, or damage. The resulting lack of ability to concentrate one's urine results in polyuria and polydipsia, which can often lead to patients drinking several times their total circulating volume. In most cases, where thirst is intact and patients have unrestricted access to fluid, plasma sodium concentration and osmolality are normal. True CDI must be differentiated from other causes of polyuria and polydipsia, such as diabetes mellitus, hypercalcaemia, renal disease, and primary polydipsia. The latter diagnosis may be difficult to differentiate from partial CDI (where some residual but insufficient AVP secretion is retained), and water deprivation tests may not sufficiently dehydrate the baseline hypoosmolar state present in these patients to confirm normal urine concentrating ability. Concurrent measurement of plasma copeptin, the C-terminal cleavage product of AVP secretion, may help with the diagnosis, with concentrations of ≥4.9 pmol/L after water deprivation having a 94% sensitivity and 96% specificity in determining the diagnosis [30]. In patients with multiple pituitary hormone deficiencies, coexistent central adrenal insufficiency may mask CDI, as cortisol has a permissive effect on free water clearance by suppressing AVP secretion.

Management is by AVP supplementation in the form of desmopressin (1-deamino-8-D-arginine vasopressin, DDAVP), which can be administered parenterally, enterally, or intranasally. Patients must be allowed free access to water, and careful fluid management is required whenever this is restricted (e.g. preoperatively). Given the distinct hypothalamic locations of AVP neurons (PVN and SON) and the thirst centre (POA), the concurrent presence of hypothalamic adipsia is rare, and where present, requires that a mandatory fluid intake requirement be set. Care is particularly required when CDI and adrenocorticotrophic hormone (ACTH) deficiency coexist; the management of an adrenal crisis with coexisting CDI needs very careful handling of glucocorticoid and DDAVP replacement with frequent measurement of electrolytes and other biochemical parameters.

Hypogonadotropic Hypogonadism and Central Precocious Puberty

Unlike other hypothalamic neurons, GnRH-secreting neurons arise from outside the hypothalamus in the olfactory placode and migrate along the olfactory nerve, crossing the cribriform plate to the olfactory bulbs to settle in their final position in the medial preoptic nucleus (MPN) at around the sixth week of gestation [31]. Congenital disruption of this process results in hypogonadotropic hypogonadism, although the timing of presentation can vary from neonatal (micropenis, cryptorchidism) to adulthood (subfertility). Some causes of hypothalamo-pituitary endocrine deficits such as septo-optic dysplasia and suprasellar tumours can also cause (in addition to or in isolation) central precocious puberty (testicular volume >4 ml in boys at <9 years of age, breast Tanner stage B2 in girls at <8 years of age), for reasons that are poorly understood [32, 33]. It is worth noting that a history of central precocious puberty in this situation does not preclude the future possibility of hypogonadotropic hypogonadism [33]. Treatment for the former is by GnRH analogue injections, while the latter is managed with gonadal steroid replacement. The use of recombinant β-hCG and/or gonadotrophins, particularly in children, is still untested in high quality trials and cannot therefore be currently recommended.

Hypothalamic Obesity

Hypothalamic obesity (HyOb) is a syndrome of intractable weight gain observed frequently after any hypothalamic damage [34] (septo-optic dysplasia, 50%; craniopharyngioma, 77%, low-grade glioma, 50%) [33, 35]. Despite having been described over 100 years ago, management remains complex and as yet there is no effective treatment [36]. Current understanding of the hypothalamic-gut-adipose tissue circuitry governing the regulation of appetite and metabolism remains incomplete, although with the advent of molecular genetic techniques and the discovery of leptin in the *ob*/*ob* (obese) mouse in 1994 [37] the identification of other peptides involved in this complex network rapidly followed.

Metabolic homeostasis is governed by the balance between anorexigens (appetite-suppressing peptides) and orexigens (appetite-inducing peptides) which signal to the ARC and VMN via afferent pathways [34, 38]. In the fed state, the peripheral anorexigenic signals leptin (from adipose tissue) and insulin (from pancreatic β cells) signal to these nuclei, alongside the central anorexigen brain-derived neurotrophic factor (BDNF) to increase the processing of pro-opiomelanocortin (POMC) to α-melanocyte stimulating hormone (αMSH) and the synthesis of cocaine- and amphetamine-regulated transcript (CART). αMSH binds to melanocortin-3 (MC3R) and melanocortin-4 (MC4R) receptors in the ARC, VMN, PVN and lateral hypothalamic area (LHA) and is modulated by CART. Concurrently, glucagon-like peptide-1 (GLP-1) and peptide YY (PYY) are secreted by the gastrointestinal tract and bind to receptors in the PVN and on orexigenic neurons to increase the output of the efferent anorexigenic pathway while suppressing orexigenic output, respectively. Efferent neurons from the PVN and LHA then signal to the locus coeruleus and nucleus tractus solitarius via signalling peptides OXT and nesfatin-1 to increase sympathetic nervous system output and limit meal size (via the potentiation of the effect of cholecystokinin) [39–41]. This then leads to increased energy expenditure, including increased lipolysis and glucose metabolism, thereby closing the feedback loop.

Conversely, in the fasted state, the key peripheral orexigen ghrelin is secreted by the stomach and bind to receptors in the ARC and VMN [34, 38]. This leads to increased secretion of the central orexigens neuropeptide Y (NPY) and agouti-related peptide (AgRP), which in turn inhibit the POMC/CART pathway. This reduces MC3R and MC4R occupancy in the PVN and LHA, which project efferent neurons to the dorsal motor nucleus of the vagus nerve to increase anabolic processes, including lipogenesis, peristalsis, nutrient absorption, and insulin secretion, once again counterbalancing the effects of starvation.

The complexity and redundancy of the various circuits in this network means that the development of targeted treatments for HyOb has proven difficult. Patients with HyOb have classically been described to be hyperphagic, but food intake diary studies in these patients have suggested that the number of calories consumed is no different to that of BMI-matched controls [42]. In children with HyOb, parents' perceptions of the degree of hyperphagia are also not significantly different from BMI-matched controls (unpublished data). In Prader–Willi syndrome, the archetypal genetic HyOb disorder, weight gain usually precedes hyperphagia by months to years [43]. Taken together, these data suggest that HyOb is not primarily driven by an increase in appetite and caloric intake, and instead is

a disorder of energy expenditure. Indeed, the basal metabolic rate and degree of physical activity is reduced in HyOb [42, 44] but the mechanisms behind this are incompletely understood, with one study showing reduced urinary catecholamine secretion [45], while another showed no significant differences [46].

Thus far, treatments which have been used for HyOb have included pancreatic vagotomy [47], dextroamphetamine [48], supraphysiological doses of triiodothyronine [49], octreotide [50], sibutramine [51], caffeine and ephedrine [52], diazoxide and metformin [53], and liraglutide [54]. None of these studies have demonstrated long-term sustainability of weight loss or even weight maintenance in a large cohort of individuals with HyOb. In some cases (sibutramine and triiodothyronine), treatment can be associated with unwanted side effects. Bariatric surgery has been shown to be capable of inducing significant amount of weight loss, potentially by concurrently resetting the hypothalamic-gut axis, but such intervention should not be taken lightly in patients with multiple comorbidities particularly from hypopituitarism [55].

Syndromes of Emaciation

The Diencephalic Syndrome

Hypothalamic dysfunction can conversely present with syndromes of emaciation, the most well-described of which is the diencephalic syndrome [56]. This rare syndrome is typically seen in infants younger than 2 years of age as a presenting feature of a hypothalamic tumour, having been described most commonly in hypothalamic low-grade gliomas (<10%) [33, 57], but also high-grade gliomas [58], craniopharyngiomas [59], germinomas [60], teratomas [61], ependymomas [62], epidermoid cysts [63], and even in tumours away from the suprasellar region such as brainstem gliomas [64]. The original description of diencephalic syndrome defined four 'major' criteria: profound emaciation, preserved (or accelerated) linear growth, hyperactivity, and euphoria, and three 'minor' criteria: pallor without anaemia, hypoglycaemia, and hypertension [56]. Other features include nystagmus, papilloedema, optic atrophy, vomiting, ataxia, sweatiness, and tremulousness. Endocrine biochemical findings include high random plasma growth hormone (GH) concentrations with low-normal IGF-1 concentrations, the former not being suppressed by an oral glucose tolerance test, increased ghrelin, and reduced insulin and leptin [65–67]. The resting energy expenditure is increased [68]. Without treatment, the outcome is universally fatal, but with modern chemotherapy and molecular therapy techniques, weight is often regained, with a rapid disappearance of presenting clinical features [69]. Patients may eventually land up having HyOb [68].

Hypothalamic Anorexia

It is worth noting that anorexia is frequently reported in patients with hypothalamic space-occupying lesions, and due to the vagueness of the symptom, this may result in a delay in diagnosis by several years [70]. Central nervous system tumours outside the suprasellar region can also present with anorexia, alluding to our incomplete understanding of the neuroendocrine circuitry which governs appetite and the role of inflammatory cytokines (e.g. tumour necrosis factor-α (TNFα), interleukin-1 (IL-1), interleukin-6 (IL-6) and interferon-γ (IFNγ)) on these circuits [71–73]. Differentiating hypothalamic anorexia from anorexia nervosa may be difficult, particularly in the adolescent female patient presenting additionally with primary or secondary amenorrhoea which may be the result of organic hypogonadotropic hypogonadism [74]. Anorexia due to hypothalamic lesions usually resolves with treatment, while true anorexia nervosa should fulfil International Classification of Diseases criteria for diagnosis [75]. A full auxological, pubertal, and endocrine biochemical and radiological assessment should be performed to exclude neuroendocrine disease, particularly in male patients.

'Non-endocrine' Clinical Presentations of Hypothalamic Dysfunction

'Non-endocrine' clinical manifestations of hypothalamic dysfunction include temperature and sleep dysregulation, behavioural, memory and learning difficulties, and autonomic dysfunction. The exact prevalence and severity of these disorders as a consequence of hypothalamic dysfunction in the paediatric population is unknown, due to the fact that these clinical conditions are difficult to diagnose and grade in childhood. Accurate clinical diagnostic tools (e.g. neuroimaging techniques and biomarkers) are lacking and young children are mostly unable to accurately self-report these symptoms. The management of the 'non-endocrine' clinical manifestations of hypothalamic dysfunction is difficult and requires multidisciplinary approach. At present, aetiological treatments are lacking for these disorders and the management strategy mainly consists of supportive care and symptomatic management in specialized centres. A detailed description of these disorders and their management is outside the scope of this chapter. The following sections contain a summary of the role of the hypothalamus in temperature, sleep, behaviour, memory, and autonomic regulation and explains the pathophysiological basis of these disorders.

Temperature Dysregulation

Thermoregulation is an example of integrative functioning of the human hypothalamus requiring a combination of autonomic and behavioural responses as well as endocrine functions. The MPN of the hypothalamus is the key structure for integrating thermal inputs from the skin and from thermo-sensitive neurons within the brain, such as neurons in the preoptic regions. The spinal cord contains warm and cold-sensing neurons projecting to the para-brachial nucleus of the pons, the relay station for transferring visceral sensory information from the spinal and cranial nerves to the hypothalamus and from the latter to the autonomic reflex and motor cell groups. Experimental studies in rabbits have shown reversible temperature dysregulation after acute hypothalamic compression and decompression [76]. Clinical observations on patients after operations on suprasellar tumours showed similar results, with the degree of temperature dysregulation correlating with the degree of hypothalamic compression [76]. In patients with cerebral trauma and intracranial haemorrhage, temperature studies showed significant differences between surviving and non-surviving patients, and diencephalic syndrome seemed relevant to clinical deterioration [76]. Hypothalamic dysfunction can be associated both with hyper- and hypothermia, with the first finding being more commonly documented in children [77].

Sleep Dysregulation

Sleep is one of the most highly conserved physiological processes in mammals, but its actual function is as yet incompletely understood. Circadian rhythm of sleep-wakefulness is controlled

by the master clock located in the SCN. Several hypothalamic circuits for sleep and wakefulness have been discovered over the last 30 years. In particular, neurons in the posterior part of the hypothalamus have been shown to promote wakefulness, with different cellular types being extensively connected to the cerebral cortex by direct excitatory projections. Additionally, arousal-promoting neurons in the brainstem all pass through the lateral hypothalamus and project directly to the cerebral cortex. Conversely, sleep-promoting neurons have been identified in the ventrolateral POA and MPN.

Sleep homeostasis is regulated by a number of neurochemicals, with the hypothalamic neuropeptides hypocretin (HCRT, or orexin) A and B playing a crucial role [78]. HCRT neurons are found exclusively in the posterior lateral hypothalamus. Their signalling is strongly associated with wakefulness and it has been suggested that HCRT neuronal degeneration may cause narcolepsy [78, 79]. In addition to HCRT, melanin concentrating hormone (MCH)-producing neurons in the LHA are also implicated in the regulation of sleep/wakefulness. In contrast to HCRT, MCH neurons are active during sleep but become silent during wakefulness [79].

Sleep disorders have been associated with damage in the hypothalamo-pituitary area in children with brain tumours, regardless of the treatment received, with the primary determinant being the area of the brain that was damaged [80]. Sleep disruption is associated with substantial adverse short- and long-term health consequences, and has been shown to increase the risk of obesity and metabolic dysfunction [81, 82]. Hence, in children and adolescents with hypothalamic dysfunction, sleep disruption is likely to greatly impact on their health-related quality of life and to contribute to their increased mortality.

Behavioural Difficulties

The hypothalamic control of behaviour is complex. Specific hypothalamic centres and circuits regulate sexual behaviour, aggressiveness, and feeding behaviour [83]. Sexually dimorphic neurons in the ventromedial hypothalamus are known to govern mating in both sexes and aggression in males [84]. LHA HCRT neurons play a role in reward-seeking and addiction [85]. OXT is considered a 'social neuropeptide', as it is involved in bonding, attachment, and other complex social behaviours that are essential for effective social interactions. The concentration of plasma OXT is abnormal in autistic spectrum disorders (ASD) and this abnormality is linked to dysregulation of OXT signalling pathways [86]. Studies have also suggested that polymorphisms of the OXT receptor contribute to oxytocinergic dysfunction in ASD [87]. Functional and anatomical differences in the hypothalamus of ASD patients have been postulated to reflect abnormal oxytocinergic modulatory mechanisms [88, 89].

Memory Dysfunction

The hypothalamus is part of the limbic system and, together with the hippocampus, the amygdala, and several other brain areas nearby, is responsible for the formation of memories. Cellular activation of hypothalamic *Hcrt* neurons facilitates short-term spatial memory in mice [90]. Discrete populations of neurons within the hypothalamus (as well as the amygdala and the lateral septum) are specifically activated by auditory fear conditioning and may be directly involved in fear learning and memory [91]. Social memory deficits are associated with reduced OXT levels in the hypothalamus of mice [92].

Memory and learning difficulties are described in patients with hypothalamo-pituitary lesions from various aetiologies. Memory deficits and difficulties in the retrieval of learned information have been documented in children after surgical treatment of craniopharyngioma [93]. Similarly, individuals with a history of treated craniopharyngioma self-reported a high prevalence of problems in everyday memory and cognition [94]. Finally, patients with hypothalamic hamartoma also display deficits in a broad range of cognitive functions, including visual and verbal learning and memory [95].

Autonomic Dysfunction

The hypothalamus controls the autonomic nervous system via a set of neurons that, from the PVN, LHA, and ARC, directly innervate both the parasympathetic and sympathetic preganglionic neurons and various cell groups in the brainstem. The PVN has emerged as one of the most important autonomic control centres in the brain and it has been proposed as a potential target for integrative treatment of autonomic dysfunction [96]. Disrupted hypothalamic functional connectivity has been documented in patients with Parkinson's disease and autonomic dysfunction [97]. Autonomic dysregulation is one the clinical features of a syndrome associated with hypothalamic dysfunction, so called ROHHAD (rapid-onset obesity, hypothalamic dysfunction, hypoventilation, autonomic dysregulation). In a case series of 15 patients with ROHHAD syndrome, symptoms of autonomic dysregulation were consistently identified in all patients, with ophthalmologic manifestations being the most frequent (86%), followed by gastrointestinal dysmotility (66%) [98].

Congenital Causes of Hypothalamic Dysfunction

Monogenic 'Non-Syndromic' Hypothalamic Obesity Syndromes

The advent of the molecular genetic era has led to the discovery of several monogenic obesity syndromes involving mutations in genes encoding the various peptides participating in the hypothalamic-gut-adipose tissue circuitry regulating appetite and metabolism, as well as their receptors. The first of these to be discovered, congenital leptin deficiency, confirmed the equivalent phenotype of the *ob/ob* (obese) mouse in humans [99] with subcutaneous recombinant leptin replacement therapy being curative [100]. Since then, other monogenic causes of obesity in humans have been described (Table 2.4.1.2), all of which are characterized by marked hyperphagia. Although these forms of obesity have often been described as 'non-syndromic', additional phenotypic features are often observed.

Syndromic Forms of Hypothalamic Obesity

Syndromic obesity is defined as the presence of obesity along with characteristic additional pleiotropic clinical features such as developmental delay, dysmorphism, and other congenital abnormalities.

Table 2.4.1.2 Monogenic non-syndromic causes of hypothalamic obesity. αMSH, α-melanocyte stimulating hormone; ACTH, adrenocorticotrophic hormone; GH, growth hormone

Gene	Inheritance	Primary mechanism	Other clinical features
SIM1 [101, 102]	Dominant/heterozygous *de novo*	Disrupted hypothalamic development	Hypogonadotropic hypogonadism, facial dysmorphisms, behavioural difficulties, 'Prader–Willi syndrome'-like phenotype
Leptin (*LEP*) [99]	Recessive	Leptin (anorexigen) deficiency	Hypogonadotropic hypogonadism
Leptin receptor (*LEPR*) [103]	Recessive	Leptin resistance	GH deficiency, central hypothyroidism, hypogonadotropic hypogonadism
Pro-opiomelanocortin (*POMC*) [104, 105]	Recessive	POMC (αMSH (anorexigen) precursor) deficiency	ACTH deficiency, red hair, pale skin (GH deficiency, central hypothyroidism, hypogonadotropic hypogonadism)
Prohormone convertase 1 (*PCSK1*) [106, 107]	Recessive	Failure of cleavage of POMC to αMSH and ACTH (and other prohormones)	GH/ACTH deficiency, hypogonadotropic hypogonadism, impaired glucose tolerance/diabetes mellitus, diabetes insipidus, postprandial hypoglycaemia, malabsorptive diarrhoea
Melanocortin 4 receptor (*MC4R*) [108, 109]	Dominant/recessive	αMSH resistance	
Melanocortin 3 receptor (*MC3R*) [110]	Dominant (heterozygous *de novo*)	αMSH resistance	
Cocaine-amphetamine-regulated transcript (*CARTPT*) [111]	Dominant	CART (anorexigen) deficiency	
Brain-derived neurotrophic factor (*BDNF*) [112]	Dominant (heterozygous *de novo*)	BDNF (anorexigen) deficiency	Only described as part of 11p13–14 contiguous gene deletion WAGRO syndrome (Wilms' tumour, aniridia, genitourinary abnormalities, cognitive disability, obesity)
Tyrosine receptor kinase B (*NTRK2*) [113]	Dominant (heterozygous *de novo*)	BDNF resistance	Developmental delay
SH2B adaptor protein 1 (*SH2B1*) [114]	Dominant (heterozygous *de novo*)	Disruption of insulin/leptin (anorexigen) signalling	As part of 16p11.2 contiguous gene deletion syndrome, associated with developmental delay

A recent systematic review by Kaur *et al.* (2017) [115] identified 79 unique obesity syndromes in the published literature, of which only 19 had been fully genetically elucidated. In the vast majority of these, the biological mechanisms linking the specific mutation with obesity remain unknown. In some, contiguous gene deletions result in disruption of the hypothalamic-gut appetite-regulating circuit due to disruption of anorexigenic signalling (Table 2.4.1.2) [112, 114]. Two syndromes where hypothalamic dysfunction is known to play a major role in obesity are Prader–Willi syndrome and rapid-onset obesity, hypothalamic dysfunction, hypoventilation, autonomic dysregulation, and neural crest tumour (ROHHADNET) syndrome.

Prader–Willi Syndrome

Prader–Willi Syndrome is a contiguous gene deletion syndrome involving the paternal copies of several imprinted genes within the 15q11–13 region (*SNRPN, NDN, MAGEL2, MKRN3, SNORD116*), with a birth incidence of 1 in 30 000 [116, 117]. The deletion can either arise directly (75%), via maternal uniparental disomy (25%), or through an imprinting defect (<1%). Phenotypically, infants are born with profound hypotonia and diminished swallowing and sucking reflexes, resulting in the need for nutritional support in the first year of life. Dysmorphic features gradually become more apparent—a narrow bifrontal diameter, almond-shaped, upslanted eyes, strabismus, hypopigmentation, and full cheeks. Hypothalamic dysfunction manifests as GH deficiency, hypogonadotropic hypogonadism, central hypoadrenalism, early onset weight gain (from a median age of 2 years), and marked hyperphagia (from a median age of 4.5 years), the latter leading to parents often needing to lock fridges and cupboards at night to prevent nocturnal foraging [43, 118, 119]. Intellectual impairment and self-mutilation in the form of skin picking manifests later in adolescence and adulthood [120]. Management includes treatment with GH, not just to improve linear growth but also to improve body composition and hypotonia [121].

ROHHADNET

ROHHADNET syndrome was first described by Ize-Ludlow *et al.* (2007) [98]. The genetic cause of this extremely rare disorder remains unknown despite extensive study, and diagnosis remains clinical, with obesity and hypoventilation being the core features. Apart from early-onset obesity, all forms of hypopituitarism (particularly CDI or the syndrome of inappropriate antidiuretic hormone secretion (SIADH)) have been observed. Hypoventilation is both central and peripheral in aetiology and often requires respiratory support. Autonomic dysfunction manifests as ophthalmological symptoms, temperature dysregulation, gastrointestinal dysmotility, and hypoalgesia. A proportion of patients will have tumours of neural crest origin such as ganglioneuroblastoma and ganglioneuroma [122, 123]. Mortality is high, and usually from cardiorespiratory failure, with no cases reported beyond the age of 18 years.

Septo-Optic Dysplasia (SOD)

SOD is a rare disorder of development of the forebrain, eye and pituitary gland occurring in about 1 in every 10 000 births [124]. Its aetiology is multifactorial, and although single gene mutations in a variety of hypothalamo-pituitary transcription factors have been identified (e.g. *HESX1, FGF8, TCF7L1, PROKR2*) [125–129] in the majority of cases no single causative factor can be found. The

diagnosis is made in the presence of at least two components of the triad of optic nerve hypoplasia, midline forebrain defects (e.g. agenesis of the corpus callosum, absent septum pellucidum), and hypopituitarism [130]. Although the latter is not always present, abnormal pituitary imaging (e.g. anterior pituitary hypoplasia, ectopic posterior pituitary, absent infundibulum) is a risk factor [131] and endocrinopathies may evolve over time necessitating long-term follow-up [130]. GH deficiency is the commonest endocrinopathy observed, followed by thyroid-stimulating hormone (TSH) and ACTH deficiencies. Gonadotrophin secretion may remain intact, or patients may conversely present with central precocious puberty suggesting a degree of hypothalamic dysfunction.

Holoprosencephaly

Holoprosencephaly arises due to a developmental defect in the separation of the forebrain resulting in a spectrum of abnormalities from failure of division of the cerebral hemispheres to agenesis of the corpus callosum. It has an incidence of 1 in 10 000 to 20 000 in the general population [132, 133] and is associated with a wide variety of genetic mutations (e.g. *SHH, PTCH1, GLI2, FGF8*) and cytogenetic abnormalities [132, 134]. It can be associated with a variety of midline facial defects including cyclopia, anophthalmia, midfacial hypoplasia, hypotelorism, cleft lip and/or palate, and a single central incisor. Hypothalamic dysfunction most commonly manifests as CDI, but anterior hypothalamo-pituitary hormone deficiencies have also been reported [127, 132].

Kallmann Syndrome and Congenital Hypogonadotropic Hypogonadism

See Table 2.4.1.3. Given their developmental origins, defects in GnRH neuron formation or migration can therefore be expected to also affect the function of olfactory neurons, leading to the combination of congenital hypogonadotropic hypogonadism and anosmia, otherwise known as Kallmann syndrome. Numerous genes have been implicated in this condition, including *ANOS1* (previously *KAL1*, causing the X-linked form), *FGFR1, FGF8, PROK2,*

Table 2.4.1.3 Genes associated with Kallmann syndrome and hypogonadotropic hypogonadism

Gene	Inheritance	Clinical features
ANOS1 (previously *KAL1*)	X-linked recessive	Kallmann syndrome, unilateral renal agenesis, synkinesis
FGFR1	Autosomal dominant, variable penetrance	Kallmann syndrome, normosmic hypogonadotropic hypogonadism, cleft lip/palate, abnormalities of corpus callosum, limb/digital abnormalities, dental anomalies
FGF8	Autosomal dominant/recessive	Kallmann syndrome, normosmic hypogonadotropic hypogonadism, cleft lip/palate, camptodactyly, holoprosencephaly
FGF17	Haploinsufficiency, possibly digenic	Kallmann syndrome, normosmic hypogonadotropic hypogonadism
IL17RD	Haploinsufficiency, possibly digenic	Kallmann syndrome, hearing loss
DUSP6	Haploinsufficiency, possibly digenic	Kallmann syndrome
PROK2	Autosomal dominant/recessive	Kallmann syndrome, obesity, normosmic hypogonadotropic hypogonadism
PROKR2	Autosomal dominant/recessive	Kallmann syndrome, normosmic hypogonadotropic hypogonadism
CHD7	Autosomal dominant	Kallmann syndrome, normosmic hypogonadotropic hypogonadism, hearing loss, CHARGE association (coloboma, heart defects, atresia choanae, retardation of growth, genitourinary abnormalities, ear abnormalities)
HS6ST1	Autosomal dominant/recessive	Kallmann syndrome, normosmic hypogonadotropic hypogonadism
WDR11	Autosomal dominant	Kallmann syndrome
SEMA3A	Autosomal dominant	Kallmann syndrome
SEMA3E	Autosomal dominant	Kallmann syndrome
SOX10	Autosomal dominant	Kallmann syndrome, Waardenburg syndrome (deafness, skin/hair/iris hypopigmentation), hearing loss, learning difficulties
FEZF1	Autosomal recessive	Kallmann syndrome
SPRY4	Haploinsufficiency, possibly digenic	Kallmann syndrome
FLRT3	Haploinsufficiency, possibly digenic	Kallmann syndrome
NELF	Haploinsufficiency, possibly digenic	Kallmann syndrome
CCDC141	Biallelic	Kallmann syndrome
GNRH1	Autosomal recessive	Normosmic hypogonadotropic hypogonadism
GNRHR	Autosomal recessive	Normosmic hypogonadotropic hypogonadism
KISS1	Autosomal recessive	Normosmic hypogonadotropic hypogonadism
KISS1R (previously *GPR54*)	Autosomal recessive	Normosmic hypogonadotropic hypogonadism
LEP	Autosomal recessive	Normosmic hypogonadotropic hypogonadism, obesity
LEPR	Autosomal recessive	Normosmic hypogonadotropic hypogonadism, obesity
TAC3	Autosomal recessive	Normosmic hypogonadotropic hypogonadism
TAC3R	Autosomal recessive	Normosmic hypogonadotropic hypogonadism

PROKR2, CHD7, HS6ST1, WDR11, SEMA3A, SOX10, and *FEZF1* [135, 136]. A proportion of patients with these mutations will have isolated hypogonadotropic hypogonadism without anosmia, although this is more usually associated with a different subset of mutations defining gonadotroph development and function, including *GNRH1, GNRHR, KISS1, KISS1R, LEP*, and *LEPR* [136].

Patients may present in the neonatal period with micropenis and cryptorchidism in males, and delayed or arrested puberty and subfertility in adolescence and adulthood. Other features associated with specific mutations may also be present, including synkinesis and renal agenesis (*ANOS1*), coloboma, heart defects, choanal atresia, retardation of growth, genitourinary defects and ear abnormalities (the CHARGE association, *CHD7*) or marked obesity (*LEP, LEPR, PROK2*). A number of other syndromes associated with hypothalamic dysfunction may also be associated with hypogonadotropic hypogonadism. These include Prader–Willi syndrome, Warburg Micro syndrome, and 4H syndrome. Treatment involves sex steroid supplementation (not just for puberty but also bone mineralization) and induction of fertility (with pulsatile GnRH or hCG and FSH) [136].

Congenital Central Diabetes Insipidus (CDI)

Congenital CDI is rare, and is usually found in combination with other hypothalamo-pituitary hormone deficits such as SOD or holoprosencephaly. Isolated congenital CDI is usually due to mutations in the AVP preprohormone gene, *AVP-NPII*, and is inherited in an autosomal dominant manner [137]. Interestingly, although the genetic mutation is present from birth, CDI usually presents after the neonatal period, suggesting that it results in progressive degeneration of AVP neurons via a dominant negative effect [138]. In the Wolfram (DIDMOAD) syndrome, mutations in *WFS1* result not just in CDI but also diabetes mellitus, optic atrophy, sensorineural deafness, and progressive neurodegeneration [139].

Other Genetic Causes of Hypothalamic Dysfunction

ARNT2

ARNT2 mutations were first described in a highly consanguineous family with six children with postnatal microcephaly, frontotemporal lobe hypoplasia, hypopituitarism, central diabetes insipidus, seizures, global developmental delay, severe visual impairment, and congenital renal tract abnormalities including hydronephrosis, vesicoureteric reflux and neurogenic bladder [3]. Homozygosity mapping and whole exome sequencing revealed a homozygous frameshift mutation in *ARNT2* (c.1373_1374dupTC), a gene encoding a basic helix-loop-helix transcription factor which is expressed in the developing hypothalamus, pituitary, thalamus, retina, lung, stomach, and renal tubules in the human embryo. Patients with the mutation exhibited markedly reduced *ARNT2* mRNA and ARNT2 protein, confirming the mutation as functionally deleterious, and extending the spectrum of *ARNT2* mutation-related phenotypes from that of the *Arnt2* knockout mouse to include renal tract and visual anomalies.

Acquired Causes of Hypothalamic Dysfunction

Acquired causes of hypothalamic dysfunction are covered extensively in Chapters 2.4.2–2.4.4. The manifestations of hypothalamic dysfunction are similar to that of congenital causes, although the resolution of current neuroimaging techniques precludes our ability to predict the risk of developing clinically apparent disease based on the degree or location of structural damage to the region. Certain acquired causes of hypothalamic dysfunction tend to present with particular endocrinopathies, for reasons which are not completely understood. For instance, germinomas, Langerhans cell histiocytosis, granulomatous diseases, and lymphocytic hypophysitis all have a predilection for the infundibulum and therefore present with CDI. Similarly, hypothalamic hamartomas often present with central precocious puberty. Over time, however, multiple pituitary hormone deficiencies often ensue, with the concurrent risk of the hypothalamic syndrome in many of these patients. Given the lack of optimal treatments for many of the 'non-endocrine' aspects of the hypothalamic syndrome, there is increasing recognition of the need to avoid hypothalamic damage in the first instance in the treatment of these acquired causes as far as possible, such as by limiting the extent of neurosurgical resection of hypothalamic tumours.

Conclusion

Hypothalamic dysfunction is rare in childhood, but it is associated with significant morbidity and mortality. Given the variety of neuroendocrine functions displayed by the hypothalamus, it is not surprising that hypothalamic dysfunction can present with a wide range of clinical manifestations including endocrine disorders, temperature, and sleep dysregulation, behavioural and learning difficulties, memory deficits, and autonomic dysfunction. The diagnosis of hypothalamic dysfunction is challenged by the lack of accurate diagnostic tools for this condition. Targeted treatment for hypothalamic obesity and the 'non-endocrine' clinical manifestations of hypothalamic dysfunction has proven difficult and supportive care and symptomatic management are the only strategies currently available for these patients.

REFERENCES

1. McCabe MJ, Dattani MT. Genetic aspects of hypothalamic and pituitary gland development. *Handb Clin Neurol*, 2014; **124**: 3–15.
2. Szarek E, Cheah PS, Schwartz J, Thomas P. Molecular genetics of the developing neuroendocrine hypothalamus. *Mol Cell Endocrinol*, 2010; **323**: 115–23.
3. Webb EA, AlMutair A, Kelberman D, *et al.* ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. *Brain*, 2013; **136**(Pt 10): 3096–105.
4. Ramachandrappa S, Raimondo A, Cali AM, *et al.* Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest*, 2013; **123**: 3042–50.
5. Couly GF, Le Douarin NM. Mapping of the early neural primordium in quail-chick chimeras. I. Developmental relationships between placodes, facial ectoderm, and prosencephalon. *Dev Biol*, 1985; **110**: 422–39.
6. Couly GF, Le Douarin NM. Mapping of the early neural primordium in quail-chick chimeras. II. The prosencephalic neural plate and neural folds: implications for the genesis of cephalic human congenital abnormalities. *Dev Biol*, 1987; **120**: 198–214.

7. Szabo NE, Zhao T, Cankaya M, Theil T, Zhou X, Alvarez-Bolado G. Role of neuroepithelial Sonic hedgehog in hypothalamic patterning. *J Neurosci*, 2009; **29**: 6989–7002.
8. Manning L, Ohyama K, Saeger B, *et al*. Regional morphogenesis in the hypothalamus: a BMP-Tbx2 pathway coordinates fate and proliferation through Shh downregulation. *Dev Cell*, 2006; **11**: 873–85.
9. Shimogori T, Lee DA, Miranda-Angulo A, *et al*. A genomic atlas of mouse hypothalamic development. *Nat Neurosci*, 2010; **13**: 767–75.
10. Michaud JL, Rosenquist T, May NR, Fan CM. Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1. *Genes Dev*, 1998; **12**: 3264–75.
11. Hosoya T, Oda Y, Takahashi S, *et al*. Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice. *Genes Cells*, 2001; **6**: 361–74.
12. Keith B, Adelman DM, Simon MC. Targeted mutation of the murine arylhydrocarbon receptor nuclear translocator 2 (Arnt2) gene reveals partial redundancy with Arnt. *Proc Natl Acad Sci U S A*, 2001; **98**: 6692–7.
13. Acampora D, Postiglione MP, Avantaggiato V, *et al*. Progressive impairment of developing neuroendocrine cell lineages in the hypothalamus of mice lacking the Orthopedia gene. *Genes Dev*, 1999; **13**: 2787–800.
14. Wang W, Lufkin T. The murine Otp homeobox gene plays an essential role in the specification of neuronal cell lineages in the developing hypothalamus. *Dev Biol*, 2000; **227**: 432–49.
15. Kimura S, Hara Y, Pineau T, *et al*. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev*, 1996; **10**: 60–9.
16. Ingraham HA, Lala DS, Ikeda Y, *et al*. The nuclear receptor steroidogenic factor 1 acts at multiple levels of the reproductive axis. *Genes Dev*, 1994; **8**: 2302–12.
17. Shinoda K, Lei H, Yoshii H, *et al*. Developmental defects of the ventromedial hypothalamic nucleus and pituitary gonadotroph in the Ftz-F1 disrupted mice. *Dev Dyn*, 1995; **204**: 22–9.
18. Wang W, Grimmer JF, Van De Water TR, Lufkin T. Hmx2 and Hmx3 homeobox genes direct development of the murine inner ear and hypothalamus and can be functionally replaced by Drosophila Hmx. *Dev Cell*, 2004; **7**: 439–53.
19. Woods KS, Cundall M, Turton J, *et al*. Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. *Am J Hum Genet*, 2005; **76**: 833–49.
20. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson IC, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. Endo Rev, 2009; **30**: 790–829.
21. Markakis EA, Swanson LW. Spatiotemporal patterns of secretomotor neuron generation in the parvicellular neuroendocrine system. *Brain Res Brain Res Rev*, 1997; **24**: 255–91.
22. Wray S, Grant P, Gainer H. Evidence that cells expressing luteinizing hormone-releasing hormone mRNA in the mouse are derived from progenitor cells in the olfactory placode. *Proc Natl Acad Sci U S A*, 1989; **86**: 8132–6.
23. Son JH, Min N, Joh TH. Early ontogeny of catecholaminergic cell lineage in brain and peripheral neurons monitored by tyrosine hydroxylase-lacZ transgene. *Brain Res Mol Brain Res*, 1996; **36**: 300–8.
24. Daikoku S, Okamura Y, Kawano H, Tsuruo Y, Maegawa M, Shibasaki T. Immunohistochemical study on the development of CRF-containing neurons in the hypothalamus of the rat. *Cell Tissue Res*, 1984; **238**: 539–44.
25. Burgunder JM, Taylor T. Ontogeny of thyrotropin-releasing hormone gene expression in the rat diencephalon. *Neuroendocrinology*, 1989; **49**: 631–40.
26. Faivre-Bauman A, Grouselle D, Nemeskeri A, Tixier-Vidal A. Ontogenesis of thyroliberin in the mouse hypothalamus. *Brain Res*, 1978; **154**: 382–7.
27. Keegan CE, Herman JP, Karolyi IJ, O'Shea KS, Camper SA, Seasholtz AF. Differential expression of corticotropin-releasing hormone in developing mouse embryos and adult brain. *Endocrinology*, 1994; **134**: 2547–55.
28. Gross DS, Longer JD. Developmental correlation between hypothalamic somatostatin and hypophysial growth hormone. *Cell Tissue Res*, 1979; **202**: 251–61.
29. Lemaire JJ, Nezzar H, Sakka L, *et al*. Maps of the adult human hypothalamus. *Surg Neurol Int*, 2013; **4**(Suppl 3): S156–63.
30. Timper K, Fenske W, Kuhn F, *et al*. Diagnostic accuracy of copeptin in the differential diagnosis of the polyuria-polydipsia syndrome: a prospective multicenter study. *J Clin Endocrinol Metab*, 2015; **100**: 2268–74.
31. Casoni F, Malone SA, Belle M, *et al*. Development of the neurons controlling fertility in humans: new insights from 3D imaging and transparent fetal brains. *Development*, 2016; **143**: 3969–81.
32. Oatman OJ, McClellan DR, Olson ML, Garcia-Filion P. Endocrine and pubertal disturbances in optic nerve hypoplasia, from infancy to adolescence. *Int J Pediatr Endocrinol*, 2015; 2015: 8.
33. Gan HW, Phipps K, Aquilina K, Gaze MN, Hayward R, Spoudeas HA. Neuroendocrine morbidity after pediatric optic gliomas: a longitudinal analysis of 166 children over 30 years. *J Clin Endocrinol Metab*, 2015; **100**: 3787–99.
34. Lustig RH. Hypothalamic obesity after craniopharyngiomas: mechanisms, diagnosis and treatment. *Front Endocrinol*, 2011; **2**: 60.
35. Pinto G, Bussieres L, Recasens C, Souberbielle JC, Zerah M, Brauner R. Hormonal factors influencing weight and growth pattern in craniopharyngioma. *Hormone Res*, 2000; **53**: 163–9.
36. Babinski MJ. Tumeur du corps pituitaire san acromegalie et avec arret de developpement des organs genitaux. *Rev Neurol (Paris)*, 1900; **8**: 531–3.
37. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature*, 1994; **372**: 425–32.
38. Smith JK. Exercise, obesity and CNS control of metabolic homeostasis: a review. *Front Physiol*, 2018; **9**: 574.
39. Blevins JE, Ho JM. Role of oxytocin signaling in the regulation of body weight. *Rev Endocr Metab Disord*, 2013; **14**: 311–29.
40. Leng G, Onaka T, Caquineau C, Sabatier N, Tobin VA, Takayanagi Y. Oxytocin and appetite. *Prog Brain Res*, 2008; **170**: 137–51.
41. Maejima Y, Sedbazar U, Suyama S, *et al*. Nesfatin-1-regulated oxytocinergic signaling in the paraventricular nucleus causes anorexia through a leptin-independent melanocortin pathway. *Cell Metab*, 2009; **10**: 355–65.
42. Harz KJ, Muller HL, Waldeck E, Pudel V, Roth C. Obesity in patients with craniopharyngioma: assessment of food intake and movement counts indicating physical activity. *J Clin Endocrinol Metab*, 2003; **88**: 5227–31.
43. Miller JL, Lynn CH, Driscoll DC, *et al*. Nutritional phases in Prader–Willi syndrome. *Am J Med Genet A*, 2011; **155A**: 1040–9.
44. Shaikh MG, Grundy RG, Kirk JM. Reductions in basal metabolic rate and physical activity contribute to hypothalamic obesity. *J Clin Endocrinol Metab*, 2008; **93**: 2588–93.

45. Coutant R, Maurey H, Rouleau S, *et al*. Defect in epinephrine production in children with craniopharyngioma: functional or organic origin? *J Clin Endocrinol Metab*, 2003; **88**: 5969–75.
46. Cohen M, Syme C, McCrindle BW, Hamilton J. Autonomic nervous system balance in children and adolescents with craniopharyngioma and hypothalamic obesity. *Eur J Endocrinol*, 2013; **168**: 845–52.
47. Smith DK, Sarfeh J, Howard L. Truncal vagotomy in hypothalamic obesity. *Lancet*, 1983; **1**: 1330–1.
48. Mason PW, Krawiecki N, Meacham LR. The use of dextroamphetamine to treat obesity and hyperphagia in children treated for craniopharyngioma. *Arch Pediatr Adolescent Med*, 2002; **156**: 887–92.
49. Fernandes JK, Klein MJ, Ater JL, Kuttesch JF, Vassilopoulou-Sellin R. Triiodothyronine supplementation for hypothalamic obesity. *Metabolism*, 2002; **51**: 1381–3.
50. Lustig RH, Hinds PS, Ringwald-Smith K, *et al*. Octreotide therapy of pediatric hypothalamic obesity: a double-blind, placebo-controlled trial. *J Clin Endocrinol Metab*, 2003; **88**: 2586–92.
51. Danielsson P, Janson A, Norgren S, Marcus C. Impact sibutramine therapy in children with hypothalamic obesity or obesity with aggravating syndromes. *J Clin Endocrinol Metab*, 2007; **92**: 4101–6.
52. Greenway FL, Bray GA. Treatment of hypothalamic obesity with caffeine and ephedrine. *Endocrine Pract*, 2008; **14**: 697–703.
53. Hamilton JK, Conwell LS, Syme C, Ahmet A, Jeffery A, Daneman D. Hypothalamic obesity following craniopharyngioma surgery: results of a pilot trial of combined diazoxide and metformin therapy. *Int J Pediatr Endocrinol*, 2011; 2011: 417949.
54. Ando T, Haraguchi A, Matsunaga T, *et al*. Liraglutide as a potentially useful agent for regulating appetite in diabetic patients with hypothalamic hyperphagia and obesity. *Int Med*, 2014; **53**: 1791–5.
55. Wijnen M, Olsson DS, van den Heuvel-Eibrink MM, *et al*. Efficacy and safety of bariatric surgery for craniopharyngioma-related hypothalamic obesity: a matched case-control study with 2 years of follow-up. *Int J Obes (Lond)*, 2017; **41**: 210–6.
56. Russell A. A diencephalic syndrome of emaciation in infancy and childhood. *Arch Dis Childhood*, 1951; **26**: 270–5.
57. Gnekow AK, Falkenstein F, von Hornstein S, *et al*. Long-term follow-up of the multicenter, multidisciplinary treatment study HIT-LGG-1996 for low-grade glioma in children and adolescents of the German Speaking Society of Pediatric Oncology and Hematology. *Neuro-oncology*, 2012; **14**: 1265–84.
58. Stival A, Lucchesi M, Farina S, *et al*. An infant with hyperalertness, hyperkinesis, and failure to thrive: a rare diencephalic syndrome due to hypothalamic anaplastic astrocytoma. *BMC Cancer*, 2015; **15**: 616.
59. Sharma RR, Chandy MJ, Lad SD. Diencephalic syndrome of emaciation in an adult associated with a suprasellar craniopharyngioma—a case report. Br *J Neurosurg*, 1990; **4**: 77–80.
60. Mohan SM, Dharmalingam M, Prasanna Kumar KM, *et al*. Suprasellar germ cell tumor presenting as diencephalic syndrome and precocious puberty. *J Pediatr Endocrinol Metab*, 2003; **16**: 443–6.
61. Chipkevitch E, Fernandes AC. Hypothalamic tumor associated with atypical forms of anorexia nervosa and diencephalic syndrome. *Arq Neuropsiquiatr*, 1993; **51**: 270–4.
62. Addy DP, Hudson FP. Diencephalic syndrome of infantile emaciation. Analysis of literature and report of further 3 cases. *Arch Dis Child*, 1972; **47**: 338–43.
63. Eliash A, Roitman A, Karp M, *et al*. Diencephalic syndrome due to a suprasellar epidermoid cyst. Case report. *Child's Brain*, 1983; **10**: 414–18.
64. Maroon JC, Albright L. 'Failure to thrive' due to pontine glioma. *Arch Neurol*, 1977; **34**: 295–7.
65. Velasco P, Clemente M, Lorite R, *et al*. The role of leptin in diencephalic syndrome. *Pediatrics*, 2014; **133**: e263–6.
66. Fleischman A, Brue C, Poussaint TY, *et al*. Diencephalic syndrome: a cause of failure to thrive and a model of partial growth hormone resistance. *Pediatrics*, 2005; **115**: e742–8.
67. Hager A, Thorell JI. Studies on growth hormone secretion in a patient with the diencephalic syndrome of emaciation. *Acta Paediatr Scand*, 1973; **62**: 231–40.
68. Kilday JP, Bartels U, Huang A, *et al*. Favorable survival and metabolic outcome for children with diencephalic syndrome using a radiation-sparing approach. *J Neuro-oncol*, 2014; **116**: 195–204.
69. Lassaletta A, Guerreiro Stucklin A, Ramaswamy V, *et al*. Profound clinical and radiological response to BRAF inhibition in a 2-month-old diencephalic child with hypothalamic/chiasmatic glioma. *Pediatr Blood Cancer*, 2016; **63**: 2038–41.
70. Chipkevitch E. Brain tumors and anorexia nervosa syndrome. *Brain Dev*, 1994; **16**: 175–9, discussion 80–2.
71. De Vile CJ, Sufraz R, Lask BD, Stanhope R. Occult intracranial tumours masquerading as early onset anorexia nervosa. *BMJ*, 1995; **311**: 1359–60.
72. Houy E, Debono B, Dechelotte P, Thibaut F. Anorexia nervosa associated with right frontal brain lesion. *Int J Eating Disord*, 2007; **40**: 758–61.
73. Ramos EJ, Suzuki S, Marks D, Inui A, Asakawa A, Meguid MM. Cancer anorexia-cachexia syndrome: cytokines and neuropeptides. *Curr Opin Clin Nutr Metab Care*, 2004; **7**: 427–34.
74. Diamanti A, Ubertini GM, Basso MS, *et al*. Amenorrhea and weight loss: not only anorexia nervosa. *Eur J Obstet Gynecol Reprod Biol*, 2012; **161**: 111–12.
75. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*, 10th edn. Geneva: World Health Organization, 1992.
76. Clar HE. Disturbances of the hypothalamic thermoregulation. *Acta Neurochir*, 1985; **75**: 106–12.
77. Liu M, Du J, Sang Y, Wu Y, Yan J, Zhu C. Clinical analysis on 33 patients with hypothalamic syndrome in Chinese children. *J Pediatr Endocrinol Metab*, 2014; **27**: 291–7.
78. Schwartz JR, Roth T. Neurophysiology of sleep and wakefulness: basic science and clinical implications. *Curr Neuropharmacol*, 2008; **6**: 367–78.
79. Ono D, Yamanaka A. Hypothalamic regulation of the sleep/wake cycle. *Neurosci Res*, 2017; **118**: 74–81.
80. Rosen GM, Bendel AE, Neglia JP, Moertel CL, Mahowald M. Sleep in children with neoplasms of the central nervous system: case review of 14 children. *Pediatrics*, 2003; **112**(1 Pt 1): e46–54.
81. Medic G, Wille M, Hemels ME. Short- and long-term health consequences of sleep disruption. *Nat Sci Sleep*, 2017; **9**: 151–61.
82. McHill AW, Wright KP, Jr. Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. *Obes Rev*, 2017; **18** Suppl 1: 15–24.

83. Saper CB, Lowell BB. The hypothalamus. *Curr Biol*, 2014; **24**: R1111–16.
84. Yang CF, Chiang MC, Gray DC, *et al*. Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell*, 2013; **153**: 896–909.
85. Aston-Jones G, Smith RJ, Sartor GC, *et al*. Lateral hypothalamic orexin/hypocretin neurons: a role in reward-seeking and addiction. *Brain Res*, 2010; 1314: 74–90.
86. Jacobson JD, Ellerbeck KA, Kelly KA, *et al*. Evidence for alterations in stimulatory G proteins and oxytocin levels in children with autism. *Psychoneuroendocrinology*, 2014; **40**: 159–69.
87. Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. *Eur Child Adolesc Psychiatry*, 2010; **19**: 169–78.
88. Kurth F, Narr KL, Woods RP, *et al*. Diminished gray matter within the hypothalamus in autism disorder: a potential link to hormonal effects? *Biol Psychiatry*, 2011; **70**: 278–82.
89. Wolfe FH, Auzias G, Deruelle C, Chaminade T. Focal atrophy of the hypothalamus associated with third ventricle enlargement in autism spectrum disorder. *Neuroreport*, 2015; **26**: 1017–22.
90. Aitta-Aho T, Pappa E, Burdakov D, Apergis-Schoute J. Cellular activation of hypothalamic hypocretin/orexin neurons facilitates short-term spatial memory in mice. *Neurobiol Learn Mem*, 2016; **136**: 183–8.
91. Butler CW, Wilson YM, Gunnersen JM, Murphy M. Tracking the fear memory engram: discrete populations of neurons within amygdala, hypothalamus, and lateral septum are specifically activated by auditory fear conditioning. *Learn Mem*, 2015; **22**: 370–84.
92. Zhang JB, Chen L, Lv ZM, *et al*. Oxytocin is implicated in social memory deficits induced by early sensory deprivation in mice. *Mol Brain*, 2016; **9**: 98.
93. Carpentieri SC, Waber DP, Scott RM, *et al*. Memory deficits among children with craniopharyngiomas. *Neurosurgery*, 2001; **49**: 1053–7; discussion 7–8.
94. Waber DP, Pomeroy SL, Chiverton AM, *et al*. Everyday cognitive function after craniopharyngioma in childhood. *Pediatr Neurol*, 2006; **34**: 13–19.
95. Quiske A, Frings L, Wagner K, Unterrainer J, Schulze-Bonhage A. Cognitive functions in juvenile and adult patients with gelastic epilepsy due to hypothalamic hamartoma. *Epilepsia*, 2006; **47**: 153–8.
96. Ferguson AV, Latchford KJ, Samson WK. The paraventricular nucleus of the hypothalamus—a potential target for integrative treatment of autonomic dysfunction. *Exp Opin Ther Targets*, 2008; **12**: 717–27.
97. Dayan E, Sklerov M, Browner N. Disrupted hypothalamic functional connectivity in patients with PD and autonomic dysfunction. *Neurology*, 2018; **90**: e2051-e8.
98. Ize-Ludlow D, Gray JA, Sperling MA, *et al*. Rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation presenting in childhood. *Pediatrics*, 2007; **120**: e179–88.
99. Montague CT, Farooqi IS, Whitehead JP, *et al*. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, 1997; **387**: 903–8.
100. Farooqi IS, Matarese G, Lord GM, *et al*. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*, 2002; **110**: 1093–103.
101. Bonnefond A, Raimondo A, Stutzmann F, *et al*. Loss-of-function mutations in SIM1 contribute to obesity and Prader–Willi-like features. *J Clin Invest*, 2013; **123**: 3037–41.
102. Holder JL, Jr., Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Hum Mol Genet*, 2000; **9**: 101–8.
103. Clement K, Vaisse C, Lahlou N, *et al*. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*, 1998; **392**: 398–401.
104. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Gruters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet*, 1998; **19**: 155–7.
105. Clement K, Dubern B, Mencarelli M, *et al*. Unexpected endocrine features and normal pigmentation in a young adult patient carrying a novel homozygous mutation in the POMC gene. *J Clin Endocrinol Metab*, 2008; **93**: 4955–62.
106. O'Rahilly S, Gray H, Humphreys PJ, *et al*. Brief report: impaired processing of prohormones associated with abnormalities of glucose homeostasis and adrenal function. *N Engl J Med*, 1995; **333**: 1386–90.
107. Farooqi IS, Volders K, Stanhope R, *et al*. Hyperphagia and early-onset obesity due to a novel homozygous missense mutation in prohormone convertase 1/3. *J Clin Endocrinol Metab*, 2007; **92**: 3369–73.
108. Vaisse C, Clement K, Guy-Grand B, Froguel P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet*, 1998; **20**: 113–14.
109. Yeo GS, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S. A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet*, 1998; **20**: 111–12.
110. Lee YS, Poh LK, Loke KY. A novel melanocortin 3 receptor gene (MC3R) mutation associated with severe obesity. *J Clin Endocrinol Metab*, 2002; **87**: 1423–6.
111. del Giudice EM, Santoro N, Cirillo G, D'Urso L, Di Toro R, Perrone L. Mutational screening of the CART gene in obese children: identifying a mutation (Leu34Phe) associated with reduced resting energy expenditure and cosegregating with obesity phenotype in a large family. *Diabetes*, 2001; **50**: 2157–60.
112. Han JC, Liu QR, Jones M, *et al*. Brain-derived neurotrophic factor and obesity in the WAGR syndrome. *N Engl J Med*, 2008; **359**: 918–27.
113. Yeo GS, Connie Hung CC, Rochford J, *et al*. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci*, 2004; **7**: 1187–9.
114. Bochukova EG, Huang N, Keogh J, *et al*. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, 2010; **463**: 666–70.
115. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. *Obes Rev*, 2017; **18**: 603–34.
116. Horsthemke B, Wagstaff J. Mechanisms of imprinting of the Prader–Willi/Angelman region. *Am J Med Genet A*, 2008; **146A**: 2041–52.
117. Whittington JE, Holland AJ, Webb T, Butler J, Clarke D, Boer H. Population prevalence and estimated birth incidence and mortality rate for people with Prader–Willi syndrome in one UK Health Region. *J Med Genet*, 2001; **38**: 792–8.
118. Hoybye C, Hilding A, Jacobsson H, Thoren M. Metabolic profile and body composition in adults with Prader–Willi syndrome and severe obesity. *J Clin Endocrinol Metab*, 2002; **87**: 3590–7.

119. de Lind van Wijngaarden RF, Otten BJ, Festen DA, *et al*. High prevalence of central adrenal insufficiency in patients with Prader–Willi syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 1649–54.
120. Curfs LM, Wiegers AM, Sommers JR, Borghgraef M, Fryns JP. Strengths and weaknesses in the cognitive profile of youngsters with Prader–Willi syndrome. *Clin Genet*, 1991; **40**: 430–4.
121. Carrel AL, Myers SE, Whitman BY, Allen DB. Benefits of long-term GH therapy in Prader–Willi syndrome: a 4-year study. *J Clin Endocrinol Metab*, 2002; **87**: 1581–5.
122. Bougneres P, Pantalone L, Linglart A, Rothenbuhler A, Le Stunff C. Endocrine manifestations of the rapid-onset obesity with hypoventilation, hypothalamic, autonomic dysregulation, and neural tumor syndrome in childhood. *J Clin Endocrinol Metab*, 2008; **93**: 3971–80.
123. Patwari PP, Wolfe LF. Rapid-onset obesity with hypothalamic dysfunction, hypoventilation, and autonomic dysregulation: review and update. *Curr Opin Pediatr*, 2014; **26**: 487–92.
124. Patel L, McNally RJ, Harrison E, Lloyd IC, Clayton PE. Geographical distribution of optic nerve hypoplasia and septo-optic dysplasia in Northwest England. *J Pediatrics*, 2006; **148**: 85–8.
125. Dattani MT, Martinez-Barbera JP, Thomas PQ, *et al*. Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat Genet*, 1998; **19**: 125–33.
126. McCabe MJ, Gaston-Massuet C, Gregory LC, *et al*. Variations in PROKR2, but not PROK2, are associated with hypopituitarism and septo-optic dysplasia. *J Clin Endocrinol Metab*, 2013; **98**: E547–57.
127. McCabe MJ, Gaston-Massuet C, Tziaferi V, *et al*. Novel FGF8 mutations associated with recessive holoprosencephaly, craniofacial defects, and hypothalamo-pituitary dysfunction. *J Clin Endocrinol Metab*, 2011; **96**: E1709–18.
128. Raivio T, Avbelj M, McCabe MJ, *et al*. Genetic overlap in Kallmann syndrome, combined pituitary hormone deficiency, and septo-optic dysplasia. *J Clin Endocrinol Metab*, 2012; **97**: E694–9.
129. Gaston-Massuet C, McCabe MJ, Scagliotti V, *et al*. Transcription factor 7-like 1 is involved in hypothalamo-pituitary axis development in mice and humans. *Proc Natl Acad Sci U S A*, 2016; **113**: E548–57.
130. Webb EA, Dattani MT. Septo-optic dysplasia. *Eur J Hum Genet*, 2010; **18**: 393–7.
131. Mehta A, Hindmarsh PC, Mehta H, *et al*. Congenital hypopituitarism: clinical, molecular and neuroradiological correlates. *Clin Endocrinol*, 2009; **71**: 376–82.
132. Dubourg C, Bendavid C, Pasquier L, Henry C, Odent S, David V. Holoprosencephaly. *Orphanet J Rare Dis*, 2007; **2**: 8.
133. Orioli IM, Castilla EE. Epidemiology of holoprosencephaly: prevalence and risk factors. *Am J Med Genet C Semin Med Genet*, 2010; **154C**: 13–21.
134. Bendavid C, Dubourg C, Pasquier L, *et al*. MLPA screening reveals novel subtelomeric rearrangements in holoprosencephaly. *Hum Mutat*, 2007; **28**: 1189–97.
135. Laitinen EM, Vaaralahti K, Tommiska J, *et al*. Incidence, phenotypic features and molecular genetics of Kallmann syndrome in Finland. *Orphanet J Rare Dis*, 2011; **6**: 41.
136. Boehm U, Bouloux PM, Dattani MT, *et al*. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*, 2015; **11**: 547–64.
137. Christensen JH, Siggaard C, Corydon TJ, *et al*. Six novel mutations in the arginine vasopressin gene in 15 kindreds with autosomal dominant familial neurohypophyseal diabetes insipidus give further insight into the pathogenesis. *Eur J Hum Genet*, 2004; **12**: 44–51.
138. Ito M, Jameson JL, Ito M. Molecular basis of autosomal dominant neurohypophyseal diabetes insipidus. Cellular toxicity caused by the accumulation of mutant vasopressin precursors within the endoplasmic reticulum. *J Clin Invest*, 1997; **99**: 1897–905.
139. Strom TM, Hortnagel K, Hofmann S, *et al*. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum Mol Genet*, 1998; **7**: 2021–8.

2.4.2 Craniopharyngiomas

Niki Karavitaki

Epidemiology

Craniopharyngiomas are rare tumours with a reported incidence of 0.13 cases per 100 000 person-years. They account for 2–5% of all the primary intracranial neoplasms and 5.6–15% of the intracranial tumours in childhood populations, in which they are the commonest lesion involving the hypothalamo-pituitary region. They may be detected at any age, even in the pre- and neonatal periods and almost half of the total cases have been described in adults. They show a bimodal age distribution with peak incidence rates in children of 5–14 and adults of 50–74 years old [1].

Pathology and Pathogenesis

Craniopharyngiomas are epithelial tumours arising along the path of the craniopharyngeal duct (the canal connecting the stomodeal ectoderm with the evaginated Rathke's pouch). Based on the World Health Organization (WHO) classification, they are grade I tumours. Rare cases of malignant transformation (possibly triggered by previous irradiation) have been described. Two main pathological subtypes have been reported: the adamantinomatous and the papillary, but transitional or mixed forms have also been described [1, 2].

The adamantinomatous type is the most common subtype and may occur at all ages. Macroscopically they have cystic and/or solid components, necrotic debris, fibrous tissue, and calcification. The cysts may be multiloculated and contain liquid ranging from 'machinery oil' to

shimmering cholesterol-laden fluid consisting of desquamated squamous epithelial cells, rich in membrane lipids and cytoskeleton keratin. They tend to have sharp and irregular margins, often merging into a peripheral zone of dense reactive gliosis, with abundant Rosenthal fibre formation (consisting of irregular masses of granular deposits within astrocytic processes) in the surrounding brain tissue and the vascular structures. The epithelium of the adamantinomatous type is composed of three layers of cells: a distinct palisaded basal layer of small cells with darkly staining nuclei and little cytoplasm (somewhat resembling the basal cells of the epidermis of the skin), an intermediate layer of variable thickness composed of loose aggregates of stellate cells (termed stellate reticulum), whose processes traverse empty intercellular spaces and a top layer facing into the cyst lumen with abruptly enlarged, flattened and keratinized to flat plate-like squamous cells (Figure 2.4.2.1). The flat squames are desquamated singly or in distinctive stacked clusters and form nodules of 'wet' keratin, which are often heavily calcified and appear grossly as white flecks. The keratinous debris may elicit an inflammatory and foreign body giant cell reaction. The presence of the typical adamantinomatous epithelium or of the 'wet' keratin alone are diagnostic, whereas features only suggestive of the diagnosis in small or non-representative specimens include fibrohistiocytic reaction, necrotic debris, calcification, and cholesterol clefts [1].

The papillary variety has been almost exclusively described in adult populations (accounts for 14–50% of the adult cases and for up to 2% of the paediatric ones). Calcification is rare and the cyst content is usually viscous and yellow. It is generally well circumscribed and infiltration of adjacent brain tissue by neoplastic epithelium is less frequent than in the adamantinomatous type. It consists of mature squamous epithelium forming pseudopapillae and of an anastomosing fibrovascular stroma without the presence of peripheral palisading of cells or stellate reticulin (Figure 2.4.2.2). The differential diagnosis between a papillary craniopharyngioma and a Rathke's cleft cyst may be difficult, particularly in small biopsy specimens, as the epithelial lining of the Rathke's cysts may undergo squamous differentiation; however, the lack of a solid component and the presence of extensive ciliation and/or mucin production are suggestive of Rathke's [1, 2].

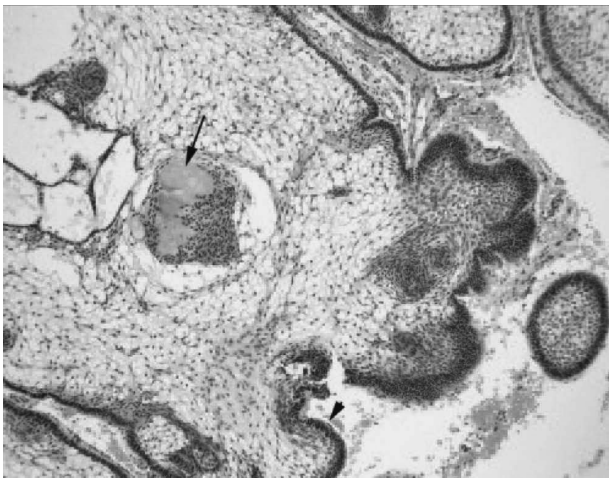


Figure 2.4.2.1 Adamantinomatous craniopharyngioma. The epithelium consists of a palisaded basal layer of cells (arrowhead), an intermediate stellate reticulum, and a layer of flattened, keratinized squamous cells. Nodules of 'wet' keratin (arrow) are also shown.

Reproduced with permission from Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. *Endocr Rev*. 2006 Jun;27(4):371–97. Copyright © 2006, Oxford University Press.

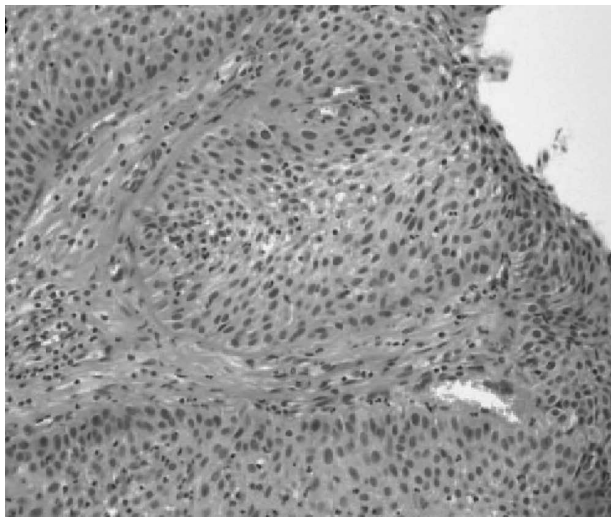


Figure 2.4.2.2 Papillary craniopharyngioma. The epithelium is mature squamous forming pseudopapillae downward into the underlying tissues.

Reproduced with permission from Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. *Endocr Rev*. 2006 Jun;27(4):371–97. Copyright © 2006, Oxford University Press.

Although the pathogenesis of craniopharyngiomas has not been fully elucidated, our understanding on this field has increased significantly in the recent years. B-catenin gene mutations have been identified in the adamantinomatous subtype affecting exon 3 which encodes the degradation targeting box of β-catenin; the mutant form is resistant to degradation leading to accumulation of nuclear β-catenin protein (a transcriptional activator of the Wnt signalling pathway). Furthermore, strong β-catenin expression has been shown in the adamantinomatous subtype indicating re-activation of the Wnt signalling pathway and subsequent de-regulation of several downstream pathways [3–5]. A number of studies using whole exome sequencing, next-generation panel sequencing, pyrosequencing and Sanger sequencing have shown the presence of activating mutations in *BRAF* (V600E) in the papillary subtype; the prevalence varies according to the sequencing method between 81% and 100% [4]. It has also been suggested that the two pathological subtypes have different epigenomic and transcriptomic signatures and that the cell clusters in the adamantinomatous subtype may have a functional role in the promotion of tumour invasion [4].

Location/Imaging

Most of the craniopharyngiomas are located in the sellar/parasellar region. The majority (94–95%) has a suprasellar component (purely suprasellar 20–41%/both supra- and intrasellar 53–75%), whereas the purely intrasellar ones represent the least common variety (5–6%. Other rare locations include the nasopharynx, the paranasal area, the sphenoid bone, the ethmoid sinus, the intrachiasmatic area, the temporal lobe, the pineal gland, the posterior cranial fossa, the cerebellopontine angle, the midportion of the midbrain or completely within the third ventricle [1, 6].

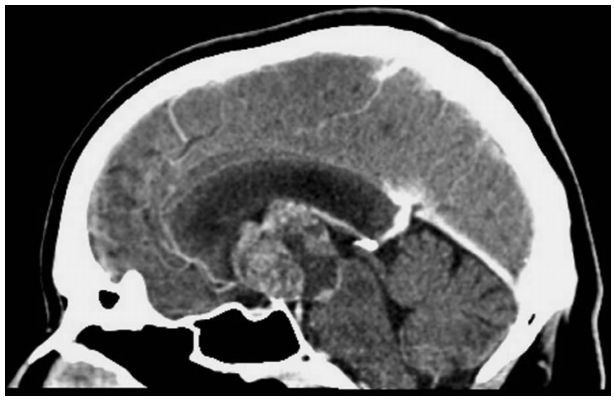


Figure 2.4.2.3 CT head: craniopharyngioma in the suprasellar area associated with mass effect on the third ventricle and hypothalamus. The lesion shows a multicystic appearance with calcifications and a marked inhomogeneous enhancement.

Imaging tools for the diagnosis of craniopharyngiomas include plain skull X-rays, CT, MRI, and occasionally, cerebral angiography. Plain skull X-rays, although seldom used nowadays, may show calcification and abnormal sella [1]. CT is helpful for the evaluation of the bony anatomy, the identification of calcifications, and the discrimination of the solid and the cystic components; they are usually of mixed attenuation, the cyst fluid has low density and the contrast medium enhances any solid portion, as well as the cyst capsule [1] (Figure 2.4.2.3). The MRI is particularly useful for the topographic and structural analysis of the tumour. The appearance of the craniopharyngioma depends on the proportion of the solid and cystic components, the content of the cyst(s) (cholesterol, keratin, haemorrhage) and the amount of calcification present. A solid lesion appears as iso- or hypointense relative to the brain on precontrast T_1-weighted images, shows enhancement following gadolinium administration and is usually of mixed hypo- or hyperintensity on T_2-weighted sequences. Large amounts of calcification may be visualized as areas of low signal on both T_1- and T_2-weighted images. A cystic element is usually hypointense on T_1- and hyperintense on T_2-weighted sequences. On T_1-weighted images a thin peripheral contrast-enhancing rim of the cyst is demonstrated. Protein, cholesterol, and methaemoglobin may cause high signal on T_1-weighted images, while very concentrated protein and various blood products may be associated with low T_2-weighted signal [1] (Figure 2.4.2.4).

The size of craniopharyngiomas has been reported as being larger than 4 cm in 14–20% of the cases, 2–4 cm in 58–76%, and smaller than 2 cm in 4–28%. Their consistency is purely or predominantly cystic in 46–64%, purely or predominantly solid in 18–39%, and mixed in 8–36%. Calcification has been demonstrated in 45–57% and is probably more common in children (78–100%). The calcification patterns vary from solid lumps to popcorn-like foci or less commonly, to an eggshell pattern lining the cyst wall. Hydrocephalus has been reported in 20–38% and is probably more frequent in childhood diagnosed disease (41–54%). There is no agreement on the radiological features discriminating the two histological subtypes. The differential diagnosis includes a number of sellar or parasellar lesions, including Rathke's cleft cyst, dermoid cyst, epidermoid cyst, pituitary adenoma, germinoma, hamartoma, suprasellar aneurysm, arachnoid cyst, suprasellar abscess, glioma, meningioma, sarcoidosis, tuberculosis, and Langerhans cell histiocytosis. Differentiation

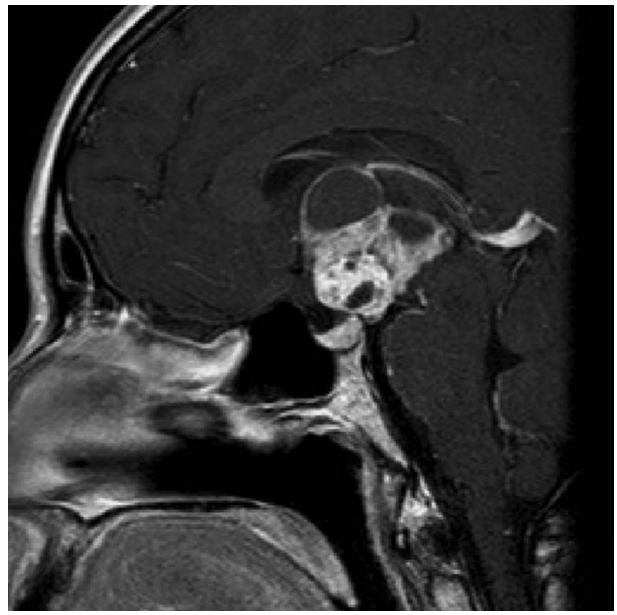


Figure 2.4.2.4 MRI of pituitary: large suprasellar craniopharyngioma with complex internal signal. There is cyst formation and enhancement after contrast.

from a Rathke's cleft cyst (typically small, round, purely cystic lesion lacking calcification), or from a pituitary adenoma (in the rare case of a homogenously enhancing solid craniopharyngioma) may be particularly difficult [1, 6–9].

Presenting Manifestations

Patients with craniopharyngioma may present with a variety of clinical manifestations attributed to pressure effects on vital structures of the brain (visual pathways, brain parenchyma, ventricular system, major blood vessels, and hypothalamo-pituitary system). Their severity depends on the location, the size, and the growth potential of the tumour. The duration of the symptoms until diagnosis ranges between 1 week to 372 months [1]. The presenting clinical manifestations (neurological, visual, hypothalamo-pituitary) and the pituitary function in a large series of cases are shown in Tables 2.4.2.1 and 2.4.2.2. Headaches, nausea/vomiting, visual disturbances, growth failure (in children) and hypogonadism (in adults) are the most frequently reported.

Treatment

Surgical Removal Combined or Not with External Beam Irradiation

Surgery combined or not with adjuvant external beam irradiation is currently one of the most widely used first therapeutic modalities for these tumours. Craniopharyngiomas remain challenging tumours, even in the era of modern neurosurgery. This is mainly attributed to their sharp, irregular margins and to their tendency to adhere to vital neurovascular structures making surgical manipulations potentially hazardous to vital brain areas. The attempted extent of excision has been a subject of significant debate and depends

Table 2.4.2.1 Presenting clinical features in children and adults with craniopharyngioma

	Children	Adults	Total
Headaches	78%	56%	64%
Menstrual disorders (adult women, n = 37)		57%	
Visual field defects	46%	60%	55%
Decreased visual acuity	39%	40%	39%
Nausea/vomiting	54%	26%	35%
Growth failure	32%		
Poor energy	22%	32%	29%
Impaired sexual function		28%	
Impaired secondary sexual characteristics (pts aged ≥13 years, n = 91)			24%
Lethargy	17%	26%	23%
Other cranial nerves palsies	27%	9%	15%
Polyuria/polydipsia	15%	15%	15%
Papilloedema	29%	6%	14%
Cognitive impairment (memory, concentration, orientation)	10%	17%	14%
Anorexia/weight loss	20%	8%	12%
Optic atrophy	5%	14%	10%
Hyperphagia/excessive weight gain	5%	13%	10%
Psychiatric symptoms/change in behaviour	10%	8%	8%
Somnolence	5%	10%	8%
Galactorrhoea		8%	
Decreased consciousness/coma	10%	4%	6%
Cold intolerance	0%	8%	5%
Unsteadiness/ataxia	7%	3%	4%
Hemiparesis	7%	1%	3%
Blindness	3%	3%	3%
Meningitis	0%	3%	2%

Adapted with permission from Karavitaki N, Brufani C, Warner JT, *et al.* (2005). Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endonol*, 62, 97–409.

Table 2.4.2.2 Pituitary function at presentation in children and adults with craniopharyngioma

	Children	Adults	Total
GH deficiency	100%	86%	95%
FSH/LH deficiency		74%	
ACTH deficiency	68%	58%	62%
TSH deficiency	25%	42%	36%
Hyperprolactinaemia		55%	
Diabetes insipidus	22%	17%	18%

Adapted with permission from Karavitaki N, Brufani C, Warner JT, *et al.* (2005). Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endonol*, 62, 397–409.

on the size (achieved in 0% of lesions >4 cm) and location (particularly difficult for retrochiasmatic or within the third ventricle) of the tumour, the presence of hydrocephalus, of more than 10% calcification and of brain invasion, as well as on the experience, the individual judgement during the operation and the general treatment policy (aggressive or not) adopted by each neurosurgeon [1, 8, 9]. Reasons for incomplete removal, as reported in 56 patients who underwent primary surgery, include firm adherence to hypothalamus (26.8%), obstructed view (21.4%), major calcifications (14.3%), adherence to perforating vessels (10.7%), adherence to major vessels (7.1%), severe bradycardia during dissection (5.4%), advanced age of the patient (3.6%), high blood loss because of co-existent aneurysm (1.8%), very thin capsule (1.8%) and impression of complete removal (7.1%) [8]. The perioperative morbidity ranges between 1.7% and 5.4% for primary operations [1, 6, 7, 10]. The irradiation of cystic craniopharyngiomas carries the risk of enlargement, which may later regress or necessitate further intervention [6, 11].

Recurrent tumours may arise even from small islets of craniopharyngioma cells in the gliotic brain adjacent to the tumour, which can remain even after gross total resection. The mean interval for their diagnosis after various primary treatment approaches ranges between 1 and 4.3 years and relapses as late as 36 years after initial therapy have been reported. Remote recurrences after apparent successful removal have been described with possible mechanisms including transplantation during the surgical procedures and dissemination by meningeal seeding or cerebrospinal fluid (CSF) spreading [1, 6].

Series with radiological confirmation of the extent of resection show that the recurrence rates following gross total removal range between 0% and 62% at 10 years follow-up. These are significantly lower than those reported after partial or subtotal resection (25–100% at 10 years follow-up). In cases of limited surgery, adjuvant radiotherapy improves significantly the local control rates (recurrence rates 10–63% at 10 years follow-up). Series with statistical comparisons of the recurrences achieved by gross total removal or combination of surgery and radiotherapy have not provided consistent results. Finally, radiotherapy alone, which however, can be offered to selected tumours, provides 10 years recurrence rates ranging between 0% and 23% [1, 6–16] (Table 2.4.2.3). Particularly focusing on series using proton therapy (aiming to reduce late toxicity) postoperatively, 5 years local control rates range between 85% and 100% [17].

In cases of predominantly cystic tumours, fluid aspiration provides relief of the obstructive manifestations and facilitates the removal of the solid tumour portion; the latter should not be delayed for more than a few weeks, as there is significant risk of cyst refilling (reported in up to 81% of the cases at a median period of 10 months) [6, 8]. The interpretation of the data on the effectiveness of each therapeutic modality has to be done with caution, since the published studies are retrospective, non-randomized, and often specialty biased. Although not widely accepted, it has been suggested that the tumour control correlates with the irradiation dose and doses below ≤5400 cGy are associated with poorer outcome. The growth rate of craniopharyngiomas varies considerably and reliable clinical, radiological, and pathological criteria predicting their behaviour are lacking. Thus, apart from significant impact of the treatment modality, attempts to identify other prognostic factors (age

Table 2.4.2.3 Recurrence rates at 10 years follow-up after treatment of craniopharyngioma by surgery and/or radiotherapy

Primary treatment	Range of 10-year recurrence rate
Gross total removal	0–62%
Partial/subtotal removal	25–100%
Partial/subtotal removal + Radiotherapy	10–63%
Radiotherapy	0–23%

group at diagnosis, sex, imaging features, pathological subtypes, immunoreactivity of the tumour proliferation marker MIB-1) have not provided consistent data [1].

The management of recurrent tumours remains difficult, as scarring/adhesions from previous surgeries or irradiation decrease the chance of successful excision. In such cases, total removal is achieved in a significantly lower rate when compared with primary surgery (0–25%) and is associated with increased perioperative morbidity and mortality (10.5–24%), suggesting that for many recurrent lesions palliative surgery is the most realistic target. The beneficial effect of radiotherapy (preceded or not by second surgery) in recurrent lesions has been clearly shown [1, 6, 15, 16].

Other Treatment Options

Intracavitary irradiation (brachytherapy) is a minimally invasive approach involving stereotactically guided instillation of β-emitting isotopes into cystic craniopharyngiomas and delivering higher radiation dose to the cyst lining than the one offered by external beam radiotherapy. It causes destruction of the secretory epithelial lining leading to elimination of the fluid production and cyst shrinkage. A number of beta- and gamma-emitting isotopes (mainly ^{32}phosphate, ^{90}yttrium, ^{186} rhenium, ^{198}gold) have been used; as none of them has the ideal physical and biological profile (i.e. pure β emitter with short half-life and with tissue penetrance limited to cover only the cyst wall), there is no consensus on which is the most suitable therapeutic agent. Based on studies with the largest series of patients and with relatively long follow-up periods, brachytherapy seems to offer a good prospect for the reduction/stabilization of cystic craniopharyngiomas. This combined with its reported low surgical morbidity and mortality render intracavitary irradiation an attractive option for predominantly cystic tumours, and particularly the monocystic ones. Its impact on the quality of survival and long-term morbidity (particularly vision, neuroendocrine and cognitive function) remain to be assessed [1, 18, 19].

The intracystic installation of the antineoplasmatic agent bleomycin has been proposed for the management of cystic tumours. However, in published reports the tumour control rates range between 0% and 100%. Direct leakage of the drug to surrounding tissues during the installation procedure, diffusion though the cyst wall or high drug dose have been associated with various toxic (hypothalamic damage, blindness, hearing loss, ischaemic attacks, peritumoral oedema), or even fatal effects. The value of this treatment option in the tumour control or even in the delaying of potentially harmful surgery and/or radiotherapy, as well as the optimal protocol and the clear-cut criteria predicting the long-term outcome remain to be established in large series with appropriate follow-up [1, 20, 21].

Stereotactic radiosurgery delivers a single fraction of high dose ionizing radiation on precisely mapped targets keeping the exposure of adjacent structures to a minimum. Tumour volume and close attachment to critical structures are limiting factors for its application, with 10 and 15 Gy being the maximum tolerated doses to the optic apparatus and the other cranial nerves, respectively. Published studies suggest that achieves tumour control in a substantial number of patients with small volume lesions (complete/partial resolution: 67–90%). Stereotactic radiosurgery may be particularly useful for well-defined residual disease following surgery or for the treatment of small solid recurrent tumours, particularly after failure of the conventional radiotherapy. In cases of large cystic portions multimodality approaches with instillation of radioisotopes or bleomycin may offer further benefits. Studies with long-term follow-up evaluating the optimal marginal dose, its role in the prevention of tumour growth and its effects on the neurocognitive and neuroendocrine functions are required [1, 22–24].

Systemic chemotherapy has been offered in a limited number of patients mainly with aggressive tumours with relative success [1]. Its application remains rather experimental and its place, particularly in the treatment of aggressive tumours, remains to be assessed.

The finding that most papillary craniopharyngiomas harbour a *BRAF* (V600E) mutation has opened avenues in using pharmacological agents specifically targeting and inhibiting mutant *BRAF* in resistant to treatment cases. Recent case reports have shown reduction in tumour volume suggesting the promising potential of this approach which, however, remains to be validated with clinical trials [4].

Long-Term Outcome After Surgery ± Conventional External Beam Irradiation

Morbidity

Patients with craniopharyngioma suffer from significant long-term morbidity (mainly endocrine, visual, hypothalamic, neurobehavioural, and cognitive) attributed to the damage of critical neuronal structures by the primary or recurrent tumour and/or to the adverse effects of the therapeutic interventions (Table 2.4.2.4). Notably, the severity of the radiation-induced late toxicity (endocrine, visual, hypothalamic, neurocognitive) is associated with the total and per fraction doses, the volume of the exposed normal tissue and the young age in childhood populations [1].

In series including subjects with various treatment modalities and follow-up periods, the frequency of pituitary hormone deficits ranges between 88% and 100% for growth hormone (GH), 80–95% for FSH/LH, 55–88% for ACTH, 39–95% for thyroid-stimulating hormone (TSH), and 25–86% for antidiuretic hormone (ADH). Apart from symptomatic diabetes insipidus (DI), which is probably more common in surgically treated patients, the long-term endocrine morbidity is not affected by the type of tumour therapy. Interestingly, restoration of pre-existing hormone deficits after surgical removal, is absent or uncommon. The phenomenon of growth without growth hormone

Table 2.4.2.4 Probability of various morbidities and compromised functional outcome at 10-years follow-up in patients with craniopharyngioma

Outcome	Rate at 10-year follow-up
Major visual filed defects (i.e. at least quadrantanopia)	48%
Hyperphagia-excessive weight gain	39%
Hemiparesis or monoparesis	11%
Epilepsy	12%
Complete dependency for basal daily activities	9%
Unable to work in previous occupation	23%
School status behind the expected level	28%
Depression or mood disorders necessitating treatment for various periods	15%
GH deficiency	88%
FSH/LH deficiency	90%
ACTH deficiency	86%
TSH deficiency	80%
Diabetes insipidus	65%

Adapted with permission from Karavitaki N, Brufani C, Warner JT, *et al.* (2005). Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endonol*, 62, 397–409.

has been reported in some children with craniopharyngioma, who show normal or even accelerated linear growth, despite their untreated GH deficiency. The pathophysiological mechanism has not been clarified; the obesity-associated hyperinsulinemia or the presence of hyperprolactinaemia have been proposed as factors stimulating growth by affecting serum concentrations of IGF-I or by binding directly to the IGF-I receptor. Finally, a number of studies support the view that GH replacement in children and adults does not increase the risk of tumour recurrence [1, 6, 25, 26].

Compromised vision has been reported in up to 62.5% of the patients treated by surgery combined or not with radiotherapy during observation period of 10 years. The visual outcome is adversely affected by the presence of visual symptoms at diagnosis and by daily irradiation doses above 2 Gy [1].

Hypothalamic damage may result in hyperphagia and uncontrollable obesity, disorders of thirst and water/electrolyte balance, behavioural, and cognitive impairment, loss of temperature control, and disorders in the sleep pattern.

Obesity is the most frequent manifestation affecting 26–61% of the patients treated by surgery combined or not with radiotherapy. It is a consequence of the disruption of the mechanisms controlling satiety, hunger, and energy balance and it often results in devastating metabolic and psychosocial complications. Possible contributing mechanisms include lack of sensitivity to endogenous leptin, vagally mediated hyperinsulinaemia, and autonomic imbalance, as well as reduced physical activity (rather than increased energy intake), which is exaggerated by the neurological defects, the visual failure, and the somnolence. Notably, in adults with craniopharyngioma, the basal metabolic rate adjusted to the total body weight is significantly lower compared with controls and children with surgically treated craniopharyngioma have decreased aerobic capacity during an exercise test related to hypothalamic involvement [27]. Hypothalamic obesity often results in devastating metabolic and psychosocial complications, necessitating provision of dietary and behavioural modifications, encouragement of regular physical activity and psychological counselling. Bariatric surgery is a management option for these patients. In a recent systematic review and meta-analysis of 21 cases of bariatric surgery for hypothalamic obesity in patients with craniopharyngioma (6 with adjustable gastric banding, 8 with sleeve gastrectomy, 6 with Roux-en-Y gastric bypass, and 1 with biliopancreatic diversion), the maximal mean weight loss was achieved in the gastric bypass group after 12 months [28]. Medical therapies including dextroamphetamine, combination of diazoxide and metformin (aiming to reduce the hyperinsulinaemia), octreotide (aiming to reduce hyperinsulinaemia and simultaneously enhance the insulin action) and glucagon-like peptide-1 analogues have been proposed as a medical approach [29]. Studies with large number of patients and longer follow-up are needed to establish the benefits and safety of these surgical and medical management options.

Diabetes insipidus with an absent or impaired sense of thirst confers a significant risk of serious electrolyte imbalance and is one of the most difficult complications to manage. In this group of patients, the maintenance of the osmotic balance has been shown to be precarious with recurrent episodes of hyper- or hyponatraemia contributing to morbidity and mortality. Careful fluid balance in and out and regular weighting are important. Factors associated with significant hypothalamic morbidity have been proposed to be young age at presentation in children, manifestations of hypothalamic disturbance at diagnosis, hypothalamic invasion, tumour height greater than 3.5 cm from the midline, attempts to remove adherent tumour from the region of hypothalamus (emphasizing the importance of conservative surgical resection when hypothalamic involvement is present), multiple operations for recurrence and hypothalamic radiation doses of more than 51 Gy [1, 6, 7, 9].

The compromised neuropsychological and cognitive function in patients with craniopharyngioma contributes significantly to poor academic and work performance, disrupted family and social relationships, and impaired quality of life. In a series of 121 patients treated by surgery with or without adjuvant radiotherapy and followed-up for a mean period of 10 years, 40% had poor outcome (the assessment was based on motor and visual deficits, dependence for activities of daily living, Karnofsky Performance Scale, school and work status, debilitating psychological or emotional problems) [7]. It has also been shown that the mean morbidity scores (based on endocrine deficiencies, vision, motor disorders and epilepsy, learning difficulties, behavioural problems, IQ, hypothalamic dysfunction) of children with additional surgery for recurrence were higher than the ones after their initial surgery and higher than those of children without recurrence [9]. There is no consensus on the therapeutic option with the least adverse impact on the neurobehavioural outcome necessitating prospective studies with formal neuropsychological testing and specific behavioural assessment prior and after any intervention [1]. These data are particularly important for the young children, in which the uncertainties of whether delaying irradiation is a reasonable policy, and whether the neurotoxicity of the recurrent disease and the subsequent surgery is higher than the one associated with irradiation offered to prevent relapse, need to be answered.

Mortality

The overall mortality rates of patients with craniopharyngioma have been reported to be 3–9 times higher than that of the general population with survival rates range between 83% and 92.7% at 10-years. Apart from the deaths directly attributed to the tumour (pressure effects to critical structures) and to the surgical interventions, the risk of cardio-/cerebrovascular and respiratory mortality is increased. It has also been suggested that in childhood populations the hypoadrenalism and the associated hypoglycaemia, as well as the metabolic consequences of ADH deficiency and absent thirst may contribute to the excessive mortality. The impact of tumour recurrence on the long-term mortality is widely accepted and the 10-year survival rates in such cases range between 29% and 70% (depending on the subsequent treatment modalities) [1, 6, 30].

Treatment Algorithm

The proposed treatment algorithm which is based on the significant available literature is shown in Figure 2.4.2.5 [1]. Surgical removal is suggested for all craniopharyngiomas causing compressive signs or symptoms (if a predominantly cystic lesion, the resection may be facilitated by previous aspiration of the cyst fluid). Gross total removal is a reasonable aim provided it is m performed by experienced neurosurgical hands and hazardous manipulations to vital brain structures (particularly the hypothalamus) are avoided. If residual tumour remains following surgery, adjuvant irradiation is recommended; this is because of the high risk of recurrence and its adverse impact on morbidity and mortality. Although this strategy may be debated for the young children, the radiation toxicity to the developing brain needs to be balanced with the consequences of relapse and subsequent possible multiple surgical procedures. In small tumours not causing pressure effects (visual, neurological, hypothalamic), radiotherapy (preceded by biopsy for confirmation of the diagnosis) is an attractive approach avoiding the risks of surgery. In predominantly cystic tumours, previous fluid aspiration may reduce the adverse sequelae of possible cyst enlargement during irradiation.

The treatment of recurrent disease depends on the previous interventions and the severity of the clinical manifestations. In recurrent lesions not previously irradiated, radiotherapy provides satisfactory local control rates. In view of the high morbidity and mortality of a second surgery, such an intervention is advocated only in cases of acute pressure effects. The treatment of further recurrence(s) should be individualized and could include gamma knife radiosurgery, cyst controlling procedures, surgical debulking (for significant solid life-threatening component) and systemic chemotherapy or for the papillary variant BRAF (V600E) inhibitors [1, 4].

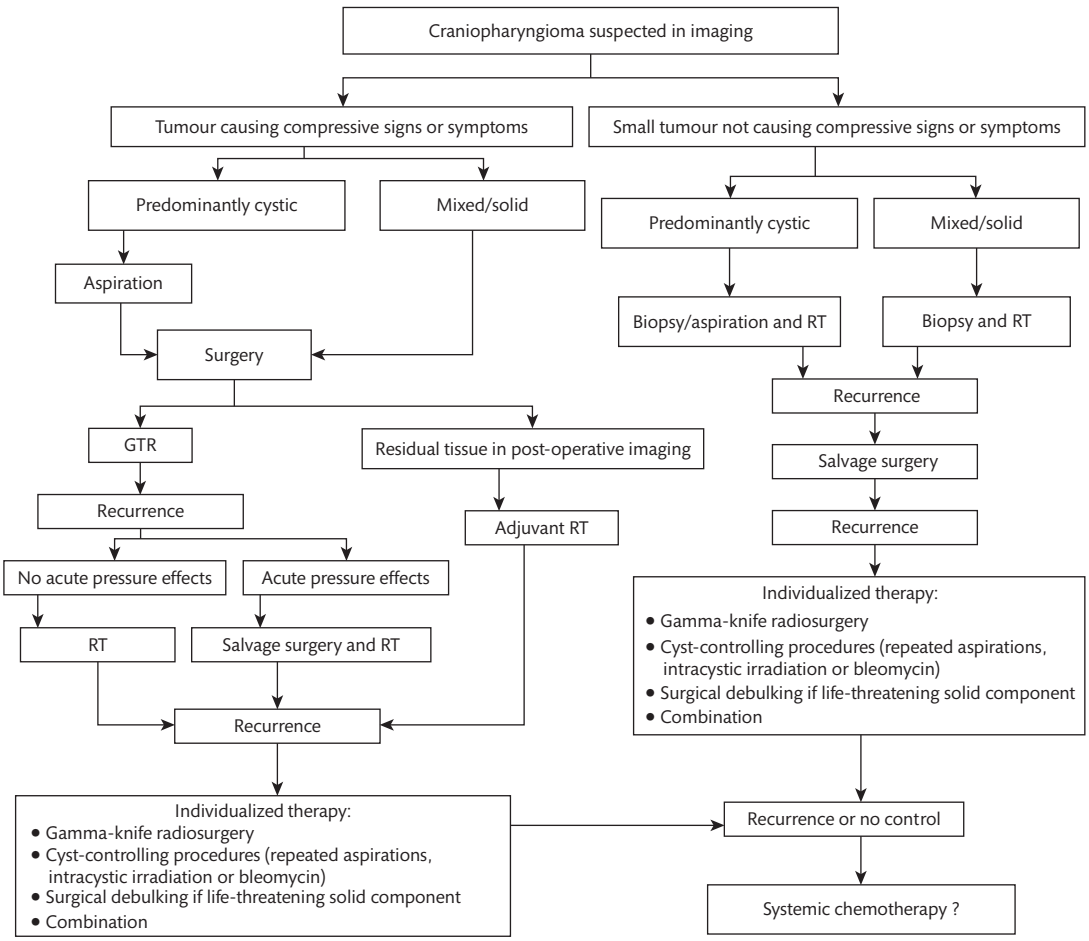


Figure 2.4.2.5 Treatment algorithm for craniopharyngiomas.

Modified with permission from Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. *Endocr Rev*. 2006 Jun;27(4):371–97. Copyright © 2006, Oxford University Press.

REFERENCES

1. Karavitaki N, Cudlip S, Adams CBT, Wass JAH. Craniopharyngiomas. *Endocr Rev*, 2006; **27**: 371–97.
2. Crotty TB, Scheithauer BW, Young WF, *et al*. Papillary craniopharyngioma: a clinico-pathological study of 48 cases. *J Neurosurg*, 1995; **83**: 206–14.
3. Buslei R, Nolde M, Hofman B, *et al*. Common mutations of beta-catenin in adamantinomatous but not in other tumours originating from the sellar region. *Acta Neuropathol (Berl)*, 2005; **109**: 589–97.
4. Larkin S, Karavitaki N. Recent advances in molecular pathology of craniopharyngioma. *F1000Res*, 2017; **24**: 1202.
5. Müller HL, Merchant TE, Puget S, Martinez-Barbera JP. New outlook on the diagnosis, treatment and follow-up of childhood-onset craniopharyngioma. *Nat Rev Endocrinol*, 2017; **13**: 299–312.
6. Karavitaki N, Brufani C, Warner JT, *et al*. Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endonol*, 2005; **62**: 97–409.
7. Duff JM, Meyer FB, Ilstrup DM, Laws ER, Jr., Scleck CD, Scheithauer BW. Long-term outcomes for surgically resected craniopharyngiomas. *Neurosurg*, 2000; **46**: 291–305.
8. Fahlbush R, Honegger J, Paulus W, Huk W, Buchfelder M. Surgical treatment of craniopharyngiomas: experience with 168 patients. *J Neurosurg*, 1999; **90**: 37–250.
9. De Vile CJ, Grant DB, Kendall BE, *et al*. Management of childhood craniopharyngioma: can the morbidity of radical surgery be predicted? *J Neurosurg*, 1996; **85**: 73–81.
10. Van Effenterre R, Boch AL. Craniopharyngioma in adults and children. *J Neurosurg*, 2002; **97**: 3–11.
11. Minniti G, Saran F, Traish D, *et al*. Fractionated stereotactic conformal radiotherapy following conservative surgery in the control of craniopharyngiomas. *Radiother Oncol*, 2007; **82**: 90–5.
12. Tomita T, Bowman RM. Craniopharyngiomas in children: surgical experience at Children's Memorial Hospital. *Childs Nerv System*, 2005; **21**: 729–46.
13. Rajan B, Ashley S, Gorman C, *et al*. Craniopharyngioma—long-term results following limited surgery and radiotherapy. *Rad Oncol*, 1993; **26**: 1–10.
14. Kim SK, Wang KC, Shin SH, Choe G, Chi JG, Cho BK. Radical excision of pediatric craniopharyngioma: recurrence pattern and prognostic factor. *Child's Nerv Syst*, 2001; **17**: 531–6.
15. Kalapurakal JA, Goldman S, Hsieh YC, Tomita T, Marymont MH. Clinical outcome in children with craniopharyngioma treated with primary surgery and radiotherapy deferred until relapse. *Med Pediatr Oncol*, 2003; **40**: 214–18.
16. Stripp DC, Maity A, Janss AJ, *et al*. Surgery with or without radiation therapy in the management of craniopharyngiomas in children and young adults. *Int J Rad Oncol Biol Phys*, 2004; **28**: 714–20.
17. O'steen L, Indelicato DJ. Advances in the management of craniopharyngioma. *F1000Res*, 2018; **11**: 1632.
18. Julow J, Backlund EO, Lanyi F, *et al*. Long-term results and late complications after intracavitary yttrium-90 colloid irradiation of recurrent cystic craniopharyngiomas. *Neurosurgery*, 2007; **61**: 288–95.
19. Hasegawa T, Kondzilka D, Hadjipanayis CG, Lunsford LD. Management of cystic craniopharyngiomas with phosphous-32 intracavitary irradiation. *Neurosurgery*, 2004; **54**: 813–22.
20. Takahashi H, Yamaguchi F, Teramoto A. Long-term outcome and reconsideration of intracystic chemotherapy with bleomycin for craniopharyngioma in children. *Childs Nerv Syst*, 2005; **21**: 701–4.
21. Hukin J, Steinbok P, Lafay-Cousin L, *et al*. Intracystic bleomycin therapy for craniopharyngioma in children: the Canadian experience. *Cancer*, 2007; **109**: 2124–31.
22. Chung WY, Pan DHC, Shiau CY, WY G, LW Wang. Gamma knife radiosurgery for craniopharyngiomas. *J Neurosurg*, 2000; **93**: 47–56.
23. Kobayashi T, Kida Y, Mori Y, Hasegawa T. Long-term results of gamma knife surgery for the treatment of craniopharyngioma in 98 consecutive cases. *J Neurosurg*, 2005; **103**: 482–8.
24. Gopalan R, Dassoulas K, Rainey J, Sherman JH, Sheehan JP. Evaluation of the role of Gamma Knife surgery in the treatment of craniopharyngiomas. *Neurosurg Focus*, 2008; **24**: E5.
25. Honegger J, Buchfelder M, Fahlbusch R. Surgical treatment of craniopharyngiomas: endocrinological results. *J Neurosurg*, 1999; **90**: 251–7.
26. Karavitaki N, Warner JT, Shine B, Ryan F, Turner HE, Wass JAH. GH replacement does not increase the risk of recurrence in patients with craniopharyngiomas. *Clin Endocrinol*, 2006; **64**: 556–60.
27. Karavitaki N. Management of craniopharyngiomas. *J Endocrinol Invest*, 2014; **37**: 219–28.
28. Bretault M, Boillot A, Muzzard L, *et al*. Clinical review: bariatric surgery following treatment for craniopharyngioma: a systematic review and individual-level data meta-analysis. *J Clin Endocrinol Metab*, 2013; **98**: 2239–46.
29. van Iersel L, Brokke KE, Adan RAH, Bulthuis LCM, van den Akker ELT, van Santen HM. Pathophysiology and individualized treatment for hypothalamic obesity following craniopharyngioma and other suprasellar tumors: a systematic review. *Endocr Rev*, 2019; **40**: 193–235.
30. Erfurth EM, Holmer H, Fjalldal SB. Mortality and morbidity in adult craniopharyngioma. *Pituitary*, 2013; **16**: 46–55.

2.4.3 Perisellar Tumours Including Cysts, Hamartomas, and Vascular Tumours

Jürgen Honegger, Ulrike Ernemann, and Rudi Beschorner

Introduction

Approximately 80% of symptomatic tumours in the pituitary region are pituitary adenomas and further 10% are craniopharyngiomas. Among the remaining 10%, a considerable number of rare tumour entities have to be considered (Table 2.4.3.1). Endocrinological, neuroradiological, and ophthalmological evaluation is the indispensable diagnostic triad to identify typical features in non-adenomatous perisellar tumours, and to provide diagnostic accuracy.

The typical clinical aspects of non-adenomatous perisellar tumours and the differential-diagnostic value of specific symptoms are presented in the following. The current therapeutic strategies and outcomes are described.

Table 2.4.3.1 Perisellar tumours

Pituitary and hypothalamic Tumours	Solitary fibrous tumour
	Spindle cell oncocytoma
Craniopharyngioma	**Extradural tumours**
Ganglioglioma	Bone 'tumour' (f.e. Paget's disease, fibrous dysplasia, osteoid osteoma)
Germ cell tumour (f.e. Germinoma)	
	Chordoma
Granular cell tumour	Chondrosarcoma
Haemangiopericytoma/solitary fibrous tumour	Chondrosarcoma
Hypothalamic hamartoma	Esthesioneuroblastoma
Lipoma, hypothalamic	Multiple myeloma
Lymphoma	Naso-pharyngeal carcinoma
Meningioma	Plasmacytoma
Metastasis	**Perisellar cysts**
Mixed ganglioglioma-adenoma	Arachnoid cyst
Neuroblastoma	Dermoid cyst
Optico-hypothalamic glioma	Epidermoid cyst
Paraganglioma	Rathke's cleft cyst
Pituicytoma	Sellar colloid cyst
Pituitary adenoma	**Vascular 'tumours'**
Pituitary carcinoma	Aneurysm
Pituitary blastoma	Cavernous sinus haemangioma
Schwannoma	Cavernous haemangioma of the optic chiasm
Sellar ependymoma	

Rathke's Cleft Cysts

It is assumed that Rathke's cleft cysts (RCC) are related to embryonal pituitary development and consist of remnants of Rathke's pouch. Microscopically small RCC are found during autopsies in 30% of normal pituitary glands. Patients with clinically symptomatic RCC usually present during adulthood.

Symptoms and endocrinological findings: Among the rare symptomatic cases, headache and visual loss are the most common presenting symptoms [1]. A meta-analysis showed hypopituitarism in 46% of the cases [2] (see Figure 2.4.3.1). The reported frequency of diabetes insipidus varies between 0% and 21% (Figure 2.4.3.2).

Neuroradiological imaging: The hyperintense appearance of the protein-rich cyst contents on T1-weighted MRI is characteristic but not obligatory. The two typical localizations are intrasellar with possible suprasellar extension and above the pituitary gland adjacent to the pituitary stalk (Figure 2.4.3.3).

Therapy: Indication for surgery is given for these benign cysts only when there is proof of deficits or a large space-occupying cyst is found. The vast majority of RCC can be removed by transsphenoidal surgery (TSS). The objective of surgery is drainage of the cyst and biopsy or partial excision of the cyst wall. Radical resection of the often adherent and thin capsule is usually not performed due to the increased morbidity (e.g. nasal cerebrospinal fluid (CSF) fistula, hypopituitarism, or diabetes insipidus). Histologically, the wall of the cyst is mainly lined by columnar ciliated and goblet cells as well as occasional pituitary hormone-producing cells. Additionally, squamous metaplasia may occur.

Outcome: Headache and visual impairment mostly abate postoperatively and hyperprolactinaemia frequently resolves [3, 4]. Preoperative hypopituitarism improves in some patients while diabetes insipidus usually persists. Recurrence rates of 0%-33% have been reported. RCC with squamous metaplasia are more likely to recur. Only a minority of recurrences become symptomatic and require re-operation [3].

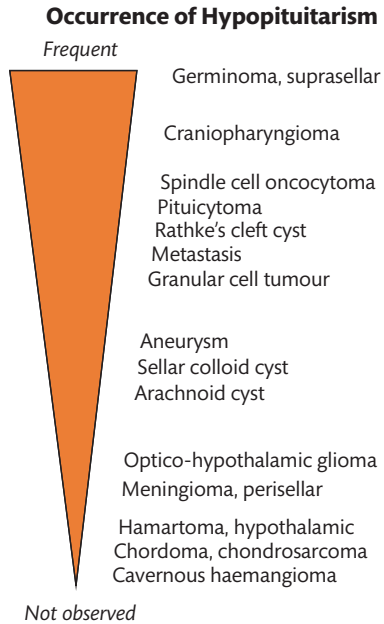


Figure 2.4.3.1 Frequency of hypopituitarism in perisellar tumours.

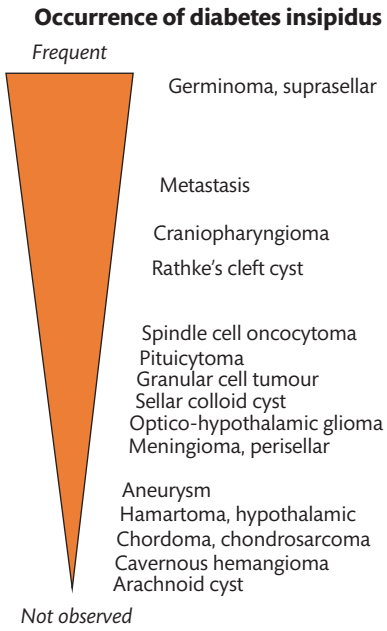


Figure 2.4.3.2 Frequency of diabetes insipidus in perisellar tumours.

Sellar Colloid Cysts

Sellar colloid cysts (SCC), arise from the intermediate lobe of the pituitary gland and must be regarded as 'pseudocysts'. They are relatively frequent sellar lesions and represent a clearly distinct clinical entity. It has been suggested that SCC are a result of cellular degeneration [5]. SCC are often misinterpreted as pituitary adenomas.

Symptoms and endocrinological findings: Usually, SCC are a chance finding and there are no endocrinological impairments.

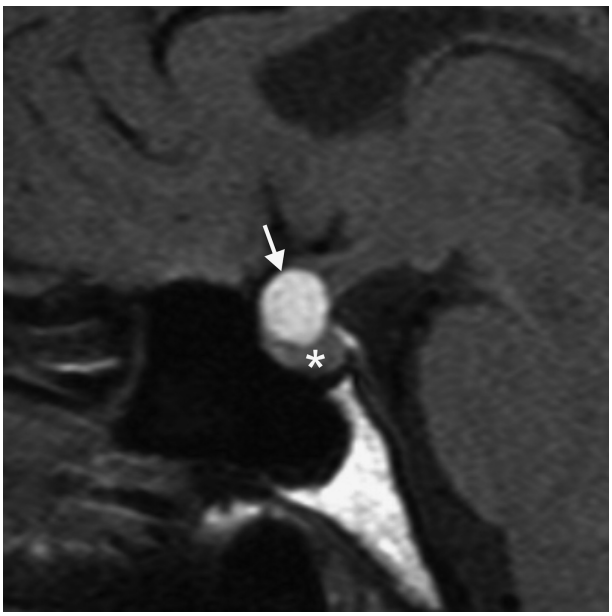


Figure 2.4.3.3 Suprasellar Rathke's cleft cyst. The sagittal T_1-weighted MRI shows the hyperintense signal of the cyst (arrow) above the pituitary gland (asterisk).

In neurosurgical series, the patients are mostly women with menstrual period disruption, galactorrhoea, and headache as the main clinical complaints. Endocrine deficits are usually mild. Formal endocrinological examination revealed hyperprolactinemia and hypogonadism in 72% of the symptomatic cases [5].

Neuroradiological imaging: The oval configuration (like an American football) and localization between the anterior lobe and the posterior lobe of the pituitary are characteristic and reliable in the differential diagnosis (see online Figure 2.4.3.4). No contrast enhancement is found. It should be looked for the characteristic dark signal of colloidal cyst content on T_2-weighted MRI.

Therapy and outcome: Surgical removal of colloidal cyst content is performed in symptomatic SCC by TSS. As they do not exhibit an epithelial lining, normal pituitary is usually found in the biopsy of the adjacent tissue [5]. Headaches often abate postoperatively. The endocrinological deficits mostly regress and normal prolactin levels are restored. No recurrence is to be anticipated.

Arachnoid Cysts

Arachnoid cysts (AC) constitute a non-proliferative anomaly of the arachnoidea. An expansive, 'tumour-like' cyst can arise due to a circumscribed collection of CSF within reduplication of the arachnoidal membrane. Intrasellar AC are relatively uncommon [6].

Symptoms and endocrinological findings: Visual impairment and headache are predominant presenting symptoms in sellar AC [6]. Hypogonadism and growth hormone deficiency were described in 50% of previously untreated cases [6] (Figure 2.4.3.1). Diabetes insipidus is not mentioned in the relevant literature as a symptom of sellar AC (Figure 2.4.3.2).

Neuroradiological imaging: MRI shows the typical findings of cystic space-occupying lesion. AC may be localized intrasellar with secondary suprasellar arching or in the suprasellar region. In the differential diagnosis of pituitary cysts, attention must be paid to the signal of the cyst. The signal of the AC contents corresponds to the CSF signal (see online Figure 2.4.3.5). Due to the thin arachnoidal capsule, no peripheral contrast enhancement is found.

Therapy: Surgical treatment of sellar AC is challenging. Most neurosurgeons prefer TSS despite the high risk of postoperative rhinorrhoea. Broad fenestration of the cyst wall toward the suprasellar CSF spaces with meticulous sellar floor reconstruction has been recommended for communicating AC that refill with CSF after cyst evacuation [6]. Alternatively, obliteration of the cyst cavity with graft implantation can be performed [7]. Cyst drainage and broad fenestration of the cyst wall by a small craniotomy is a good alternative to the transsphenoidal evacuation as it avoids the risk of nasal CSF leakage and is particularly recommended in large AC.

Outcome: Visual deficits usually improve postoperatively [6]. Partial hypopituitarism often recovers, whereas panhypopituitarism frequently persists [6].

Perisellar Meningiomas

Meningiomas are by far the most frequent non-pituitary tumours with secondary spread into the pituitary fossa and encroachment upon the pituitary gland or stalk. In meningiomas, the ratio of

women to men is 5:1. Meningiomas arise from arachnoid cover cells of the meninges. Benign World Health Organization (WHO) grade I meningiomas numerically dominate. Mutations in the TERT promoter [8] or in the breast cancer (BRCA)1-associated protein-1 tumour suppressor gene (BAP1) are associated with a worse prognosis.

In addition to the histological classification, meningiomas are also subdivided according to their localization. The most common meningiomas in the pituitary region are the tuberculum sellae meningiomas (TSM) and the cavernous sinus meningiomas (CSM).

Tuberculum Sellae Meningiomas

Symptoms and endocrinological findings: The main symptom of TSM is visual loss, whereby the impairment can be one-sided due to a prechiasmatic lesion or bilateral due to a chiasmal syndrome. Pituitary failure is rare in TSM despite the often considerable size [9]. In some cases, hypogonadism is observed, or hyperprolactinemia due to displacement of the pituitary stalk.

Neuroradiological imaging: A frequent differential-diagnostic error is misdiagnosing a TSM as a pituitary adenoma. Therefore, precise inspection of sagittal MR images is paramount and reveals that the meningioma sits broad-based on the tuberculum sellae and above the pituitary gland, but does not grow in the confines of the pituitary gland (see online Figure 2.4.3.6). In contrast to pituitary adenomas, meningiomas often show dural enhancement (so-called **dural tail**) which is explained by tumour spread but also by increased vascularization of the neighbouring meninges.

Therapy: Traditionally, TSM are operated by a transcranial approach. The extended transsphenoidal approach has become an alternative for removal of TSM [10]. In the majority of cases, TSM can be completely resected.

Outcome: With 10%, the recurrence rate is relatively low [10]. High rates of visual improvement have been reported. However, postoperative visual decline is observed in 10% of the operated patients (8). The risk of postoperative pituitary deficits is low [9].

Cavernous Sinus Meningiomas

Symptoms and endocrinological findings: The main symptom of CSM is double vision due to ocular motor nerve palsies. Often retroorbital pain is reported due to dural involvement and distension of the cavernous sinus. Hormone abnormalities occur when the direction of growth is medial. The most frequent endocrine abnormality is hyperprolactinemia [11].

Neuroradiological findings: Neuroradiologically, the expansive tumour within the cavernous sinus shows strong contrast enhancement. The marked dural tail, which extends to the tentorium, is characteristic in CSM.

Therapy: Meningiomas are usually slow-growing tumours and two-third of meningiomas remained stable in size during observation [12]. Therefore, asymptomatic or oligosymptomatic meningiomas of the cavernous sinus can initially be managed conservatively with regular control MRI. Symptomatic or growing CSM are treated these days primarily by radiotherapy. Single-shot radiosurgery is particularly suitable for meningiomas confined to the cavernous sinus. Hypofractionated radiosurgery or fractionated radiotherapy has advantage in large meningiomas with proximity to the optic system. The radical cavernous sinus surgery of the 1980s has been abandoned due to the high morbidity and tendency to recurrence. Surgical debulking is made in cases of exophytic tumour expansion and in compression of the optic pathways or growth into the optic canal.

Outcome: Tumour control rates of more than 90% can be achieved with radiation modalities [13]. In light of the close proximity to the hypothalamo-pituitary system, radiation of perisellar meningiomas requires close attention to radiation-related endocrinological deficits, which may manifest several years after treatment.

Diaphragma Sellae Meningiomas

The rare variant of diaphragm sellae meningioma originates immediately anterior or posterior to the pituitary stalk and grows above the diaphragm sellae. Visual symptoms but also endocrine deficits prevail at presentation. Due to the location of these meningiomas, operative removal which is performed via a craniotomy is challenging and carries a risk of endocrine deficits.

Intrasellar Meningiomas

Purely intrasellar meningiomas are rare [11]. It is assumed that this entity arises from the lower side of the diaphragma sellae. Thus, it is a special variant of diaphragma sellae meningiomas. Most intrasellar meningiomas have been treated by TSS [11], but resection of these often highly vascular tumours is more difficult than resection of pituitary adenomas.

Pituitary Metastases

Pituitary metastases are reported in published autopsy series of patients with malignant disease with a frequency of 1% to 11.8%. By contrast, metastases in the pituitary and hypothalamus are relatively rare in surgical series. Predilection sites of metastasizing are the bone, in particular the clivus, with secondary spread to the pituitary, but also the pituitary itself and the pituitary stalk. Breast and lung cancer are predominant among metastases in the pituitary region followed by prostate and renal cancer [14].

Symptoms and endocrinological findings: It is important to differentiate from pituitary adenomas since metastases require rapid therapy. About half of the patients presenting with pituitary metastasis have a history of malignoma. Eye muscle paresis up to ophthalmoplegia are characteristic of malignomas in the sellar region. Other frequent complaints are retroorbital pain and visual impairment. In a literature review, the incidence of diabetes insipidus was 33% [14]. Diabetes insipidus is an important endocrinological hint as it points to a pituitary lesion other than a pituitary adenoma. Anterior pituitary insufficiency was found in 40% of the patients [14].

Neuroradiological imaging: MRI usually shows strong, homogeneous contrast enhancement. Other than in pituitary adenomas, pituitary metastases often show tumour expansion along the hypophyseal stalk. Parasellar extension into the cavernous sinus is frequently encountered. Osteodestructive growth is an important differential-diagnostic criterion indicating a malignant tumour.

Therapy: The main indications for surgery are relief of visual deficits and pain, confirmation of diagnosis, and removal of the tumour mass if considered beneficial for the overall outcome. The transsphenoidal approach is most often used [15]. Usually, the indication is given for adjuvant radiation therapy after surgical treatment and confirmation of diagnosis [15]. Administration of

chemotherapy or targeted therapy depends on the underlying malignant disease.

Outcome: While ophthalmological symptoms are likely to improve after surgical decompression, endocrinological deficits are usually not reversible. The overall prognosis for patients with pituitary metastasis is poor. In the individual case, however, it decisively depends on the origin and type of malignoma and the stage of disease.

Clivus Chordomas

Chordomas arise from persisting remnants of the notochord and consist of so-called physaliphorous tumour cells (see online Figure 2.4.3.7). They occur along the neuraxis. The clival region is the second most frequent location after the sacral region [16]. Clivus chordomas (CC) may expand toward the pituitary fossa and suprasellar toward the pituitary stalk and hypothalamus.

Symptoms and endocrinological findings: The patients mean age is 40–50 years [16]. The main symptom of chordoma in the upper clivus area is a one- or two-sided abducens nerve palsy as the nerve runs in Dorello's canal in proximity to the clivus. Hypothalamo-hypophyseal endocrinological deficits are relatively rare (Figures 2.4.3.1 and 2.4.3.2).

Neuroradiological imaging: On MRI, CC are commonly hyperintense on T2-weighted images and enhance inhomogeneously after contrast administration. CC primarily grow in the bone with osteodestructive character. Expansive growth with convex arching of the clivus dura toward the brain stem and pons cerebri is typical.

Therapy and outcome: The primary treatment is surgical. A meta-analysis showed gross total resection (GTR) in 39.9% of CC [16]. In CC near the midline, resection is initially by extended, endoscopic, or microscopic TSS. CC are often soft tumours facilitating resection. Depending on the expansion, other skull base approaches or combined procedures may be necessary. CC are characterized by a high tendency to local recurrence. Metastases are relatively rare, but the risk increases with advanced disease. Radiotherapy is effective and is used as adjuvant therapy. Radiation with heavy particles, such as proton or carbon ion radiation, is preferred [17]. CC are not responsive to cytotoxic chemotherapy. It has recently been shown that patients with advanced CC can benefit from novel targeted therapy. These include tyrosine kinase inhibitors or epidermal growth factor receptor inhibitors. Five-year progression-free survival (PFS) is approximately 50% and 5-year overall survival (OS) is 75% [16].

Petroclival Chondrosarcomas

Petroclival chondrosarcomas (PCS) arise from primitive mesenchymal cells of the chondral matrix originating from the petroclival synchondrosis and can extend to the pituitary fossa.

Symptoms: Similar to CC, patients typically present with cranial nerve palsies among which diplopia due to abducens nerve palsy prevails. Headache is another common symptom.

Therapy: PCS are typically slow-growing with locally invasive behaviour. Initial observation is an option for selected patients [18]. The mainstay of primary therapy is surgical resection. Because of the common eccentric growth, TSS is used less frequently than in CC. Various transcranial approaches have been used depending on the precise location [18]. A good chance of postoperative recovery of abducens nerve palsy exists [18]. Radiation for PCS has been shown to be effective and proton or photon beam radiation has commonly been used [18]. It can be used as adjuvant therapy following initial surgery or can be withheld until recurrence occurs [18].

Outcome: In a large series of skull base chondrosarcomas treated by surgery and consecutive fractionated radiation therapy, the 5- and 10-year PFS was 99% and 98%, respectively [19].

Optico-Hypothalamic Gliomas

Low-grade optic pathway gliomas (OPG) account for approximately 5% of all brain tumours in children and are frequently associated with neurofibromatosis type 1 (NF-1). Gliomas with posterior extension and hypothalamic involvement are named optico-hypothalamic gliomas (OHG).

Symptoms: 20% of children with NF-1 harbour an OPG which is frequently an incidental finding during neuroradiological work-up. In sporadic cases, unsystematic visual impairment is in the foreground.

Endocrinological findings: Preoperative differentiation of OHG from craniopharyngiomas may be difficult. Contrary to craniopharyngiomas, hypopituitarism and diabetes insipidus are rarely seen preoperatively [20] (Figures 2.4.3.1 and 2.4.3.2). In OHG, hypothalamic syndrome occurs in about 20% of the patients. Cachexia prevails among hypothalamic disorders [20].

Neuroradiological imaging: Intrinsic growth within the optic system should be watched for on MRI (Figure 2.4.3.8). Sometimes the tumour has already attained gigantic size at the time of diagnosis and hydrocephalus secondary to Foramen Monroi occlusion may occur. The numerically dominant pilocytic astrocytomas show both cystic portions and solid portions with areas of high contrast uptake.

Therapy: The attitude is most commonly conservative with observation and therapy withheld if tumour size and vision are stable. Spontaneous remissions have been described.

Chemotherapy is currently the predominant treatment modality for OPG [21], and carboplatin and vincristine are mostly used. Surgical tumour debulking has to be considered in large and progressive OPG with exophytic growth.

Radiotherapy is often used as second-line treatment. In very young patients, radiotherapy is avoided where possible due to the adverse long-term sequelae on the developing brain including neuroendocrine deficits, neurodevelopmental delay, and cognitive impairment [21].

Outcome: Sporadic cases of OPG have an inferior prognosis compared to OPG in NF-1. A 10-year OS of patients with OPG of 94% has been reported [22]. Survival rates are inferior if the optic chiasm and the hypothalamus are involved. The rare malignant OHG are an entity of adulthood.

Suprasellar (Hypothalamic) Germinomas

Intracranial germinomas arise from extragonadal primordial germ cells and account for approximately 70% of central nervous system

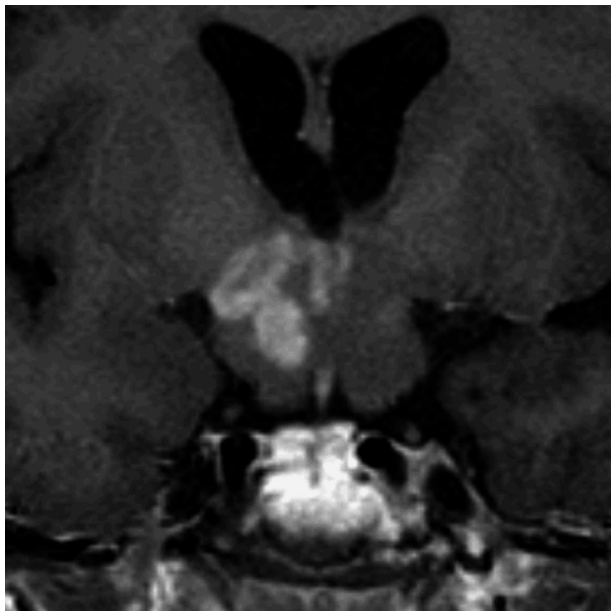


Figure 2.4.3.8 Optico-hypothalamic glioma. The coronal T_1-weighted MRI with contrast shows the intrinsic growth of the glioma within the optic system and hypothalamus.

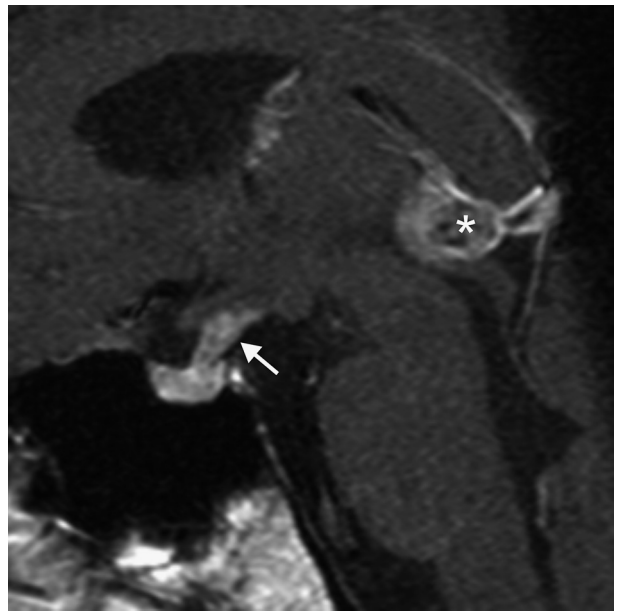


Figure 2.4.3.9 Suprasellar germinoma. The sagittal T_1-weighted MRI with contrast demonstrates the suprasellar germinoma at the pituitary stalk (arrow). A second manifestation is shown in the pineal region (asterisk).

germ cell tumours (CNS-GCT). While most germinomas are located in the pineal region, 10–20% grow in the suprasellar region.

Symptoms and endocrinological findings: The triad of anterior pituitary insufficiency, diabetes insipidus and visual compromise is a characteristic finding in suprasellar germinomas [23]. Often, panhypopituitarism is present. Diabetes insipidus with no imaging proof of a lesion may be evidence of a germinoma in status nascendi and requires close-meshed MRI control examinations.

Neuroradiological imaging: MRI reveals a tumour with marked contrast uptake in the area of the pituitary stalk and infundibulum. In some cases, a second lesion is found in the area of the pineal gland and raises strong suspicion for the diagnosis of a germinoma (Figure 2.4.3.9).

Further diagnostics: The differential-diagnostic delineation from other lesions with contrast uptake in the area of the pituitary stalk, such as infundibulo-hypophysitis, Langerhans cell histiocytosis, or metastases is often not possible by means of MRI. A subtle CSF analysis with examination for tumour cells, inflammatory cells, and tumour markers is demanded in such cases. All patients with suspected CNS-GCT should have the tumour markers α-fetoprotein (AFP) and ß subunit of human chorionic gonadotropin (ß-HCG) measured in CSF and serum [24]. Proof of α-fetoprotein and ß-HCG indicates the presence of a non-germinomatous CNS-GCT. Further tumour markers including placental alkaline phosphatase (PLAP) and soluble c-Kipt can be used to characterize CNS-GCT.

Therapy and outcome: Histo-pathological confirmation of the diagnosis by biopsy is obligatory in patients with suspected suprasellar germinoma. Radical operation is not justified in light of sensitivity to radiotherapy and chemotherapy. Localized intracranial germinomas are treated with whole-ventricle radiotherapy plus a boost to the tumour region [24]. Neoadjuvant chemotherapy followed by radiotherapy is another accepted therapeutic concept [24, 25] and allows a reduction of the radiation dose. Five-year survival rates exceeding 90% have been shown for intracranial germinomas.

In contrast, 5-year survival rates of non-germinomatous CNS-GCT is only 44%. However, they are extremely rare in the pituitary region.

Hypothalamic Hamartomas

Hypothalamic hamartoma (HH) is a non-neoplastic, malformative mass which consists of atypically differentiated glial and neural tissue.

Symptoms: HH usually become symptomatic in early childhood either because of precocious puberty or because of gelastic ('laughing') seizures [26]. The laughing seizures constitute a specific epileptic disorder (so-called **gelastic epilepsy**), which is pathognomonic for the presence of a HH. Gelastic epilepsy is pharmaco-resistant and leads to secondary epileptogenesis with additional types of seizures. In addition, cognitive impairment and behavioural disturbances ranging up to serious psychiatric symptoms occur.

Neuroradiological imaging: MRI reveals a lesion without contrast uptake in the area of the tuber cinereum or the mamillary bodies, which appears isointense to grey matter on T1-weighted images. Pediculated and small hamartomas lead more often to precocious puberty, while broad-based and large hamartomas with intrahypothalamic expansion and involvement of the third ventricle more often elicit gelastic epilepsy [26].

Therapy of precocious puberty: Precocious puberty is usually treated these days with gonadotropin-releasing hormone (GnRH) analogues. The outcome in terms of fertility and final height is favourable.

Therapy of epilepsy: Pharmaco-resistant epilepsy can be treated surgically by transcranial resection or disconnecting of the HH.

From 15% to 67% of the patients have been reported to be seizure-free postoperatively. Postoperative endocrinological and neurological deficits may occur.

Endoscopic disconnection after insertion of the endoscope via the foramen of Monro is used especially for HH with expansion into the third ventricle [27]. Rates of seizure freedom of 50% or greater have been reported in larger series.

Single-session radiosurgery is especially suited for the treatment of small and medium-sized hamartomas. In a prospective trial on Gamma Knife radiosurgery in 48 patients, 39.6% became seizure free (Engel class I) and 29.2% experienced worthwhile improvement (Engel class 2) [28].

Further treatment options are interstitial radiotherapy (brachytherapy), radiofrequency thermocoagulation, and laser thermocoagulation.

Gangliocytomas

Gangliocytomas consist of neural cells. In addition to gangliocytomas in the brain, gangliocytomas are also observed in the hypophysis. The precise aetiology has not finally been uncovered. 65% of sellar gangliocytomas are associated with adjacent pituitary adenomas which are mostly hormone-secreting (Figure 2.4.3.10). Associated growth hormone-secreting pituitary adenomas, leading to acromegaly, prevail [29]. The mainstay of therapy is TSS. A review of the literature found a 63% chance to normalize the associated hypersecretory endocrinopathy by surgery [29].

Tumours of the Posterior Pituitary

Granular cell tumour of the sellar region, pituicytoma, and spindle cell oncocytoma represent a distinct group of low-grade tumours (WHO grade I) arising from pituicytes, specialized glial cells in the infundibulum and neurohypophysis. Expression of TTF-1 which is also found in non-tumorous pituicytes is a clue to the diagnosis of a pituicyte-derived tumour [30] (see online Figure 2.4.3.11). Presumably, they represent three morphological variants of a single entity [30]. Granular cell tumours are more common than pituicytomas. Spindle cell oncocytomas are least frequent [31]. In tumours of the posterior pituitary, the preoperative diagnosis of a pituitary adenoma is a frequent misinterpretation.

Symptoms and endocrinological findings: Visual disturbances are the predominant presenting symptoms for all three types [31]. Partial hypopituitarism is present in one-third of patients with granular cell tumours and one-half of patients with pituicytomas at the time of diagnosis [32]. Panhypopituitarism is particularly frequent in spindle cell oncocytomas. Diabetes insipidus is surprisingly rare despite the infundibular or posterior lobe origin.

Neuroradiological imaging: Strong contrast enhancement is a frequent MRI finding. Purely intrasellar location is only found in pituicytomas. Intrasellar und suprasellar location can be observed in all three entities (Figure 2.4.3.12). Purely suprasellar location is common in granular cell tumours and is found in about 40% of pituicytomas [31, 33]. Spindle cell oncocytomas tend to show infiltrative growth impairing differentiation from the pituitary gland.

Therapy and outcome: TSS is mostly performed. However, transcranial surgery is indicated for purely suprasellar tumours. Tumours of the posterior pituitary tend to be highly vascular impeding radical removal. Massive intraoperative bleeding has been repeatedly reported.

Symptomatic granular cell tumours can progress rapidly and require timely surgical treatment. Gross total removal of pituicytoma has been reported in approximately 50% of the cases and is mostly curative [33]. New endocrine deficit is the most frequent inadvertent sequela of surgery for pituicytomas [33].

Radiotherapy appears to be beneficial in cases with less than total resection [32]. In pituicytomas, the overall outcome is favourable. The outcome in granular cell tumours is inferior. Death

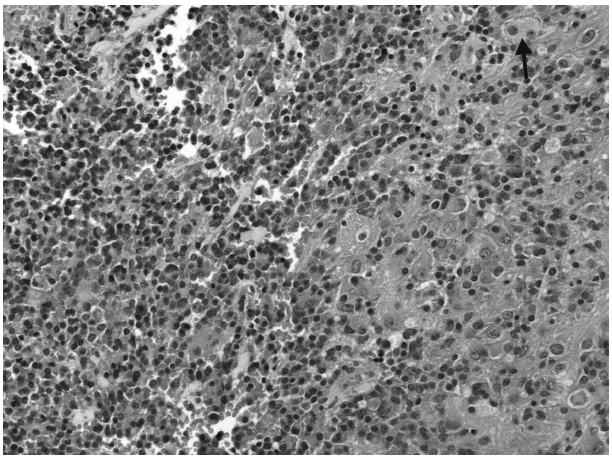


Figure 2.4.3.10 Pituitary gangliocytoma and adenoma. Histology shows the border zone between distinct parts of two tumours. On the right, the tumour consists of mature ganglion cells including single binucleated ganglion cell (arrow). On the left, small epithelial cells are seen as part of a pituitary adenoma. H&E-staining. Original magnification ×200.

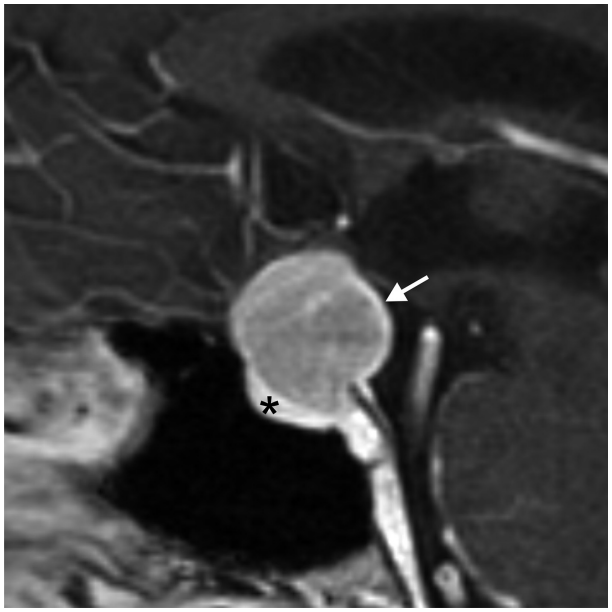


Figure 2.4.3.12 Pituicytoma. The sagittal T1-weighted MRI with contrast shows the pituicytoma delineated from the pituitary gland (asterisk) and the pituitary stalk (arrow).

related to the tumour has been reported in 9% of cases following surgery without radiotherapy and in 16% following surgery with radiotherapy [32]. Unfavourable clinical outcome has also been described in some spindle cell oncocytomas.

Aneurysms

After passing the skull base, the carotid artery has a close proximity to the pituitary over a long distance. Intrasellar (infradiaphragmatic) aneurysms usually originate from the clinoid or cavernous segment of the carotid artery. Suprasellar aneurysms originate from the distal carotid artery after it has entered the intradural space or from the anterior communicating artery [34].

Symptoms: Headache is a frequent complaint of patients with unruptured aneurysms of the perisellar region [34]. Visual deficits are common in supradiaphragmatic aneurysms. Anterior pituitary dysfunction was found in 60% of infradiaphragmatic, intrasellar aneurysms [34]. Diabetes insipidus was not observed.

Neuroradiological imaging: On MRI, aneurysms can be recognized by their direct relationship to the vessels and by their flow signal (see online Figure 2.4.3.13). It is extremely important to recognize the neuroradiological signs, since aneurysms may imitate pituitary tumours. The transnasal operation of a wrongly interpreted aneurysm could have fatal consequences. If an aneurysm is suspected, digital subtraction angiography (DSA) is the gold standard to confirm the diagnosis (see online Figure 2.4.3.13). CT angiography and MR angiography are increasingly used alternative non-invasive methods.

Therapy and outcome: Symptomatic or growing extradural, infradiaphragmatic aneurysms of the carotid artery are usually treated by endovascular coil-embolization (often requiring additional stenting of the parent vessel) or by flow-diversion stents [35]. Supradiaphragmatic, intracranial carotid artery aneurysms require treatment because of the risk of spontaneous aneurysm rupture causing potentially fatal subarachnoid haemorrhage (SAH). Occlusion of the intracranial aneurysm is usually performed by endovascular coil-embolization or by microsurgical aneurysm clipping. Many patients are only diagnosed when SAH has occurred and require emergency treatment.

Beyond the possible deleterious neurological sequelae, SAH can cause endocrinological deficits. A review on pituitary function in the chronic stage after SAH showed a high variability of hypopituitarism between 0% and 55% [36]. Single pituitary hormone deficiency is more frequent than deficiency of multiple hormonal axes. The proximity of the ruptured aneurysm to the hypothalamus and pituitary region increases the risk for developing hypopituitarism.

Haemangiomas

In the perisellar region, there are two rare but well-defined entities of haemangiomas, namely *cavernous sinus haemangiomas* and *cavernous haemangiomas of the optic chiasm*.

Cavernous Sinus Haemangiomas

Cavernous sinus haemangiomas (CSH) are vascular neoplasms that account for 2% of cavernous sinus tumours.

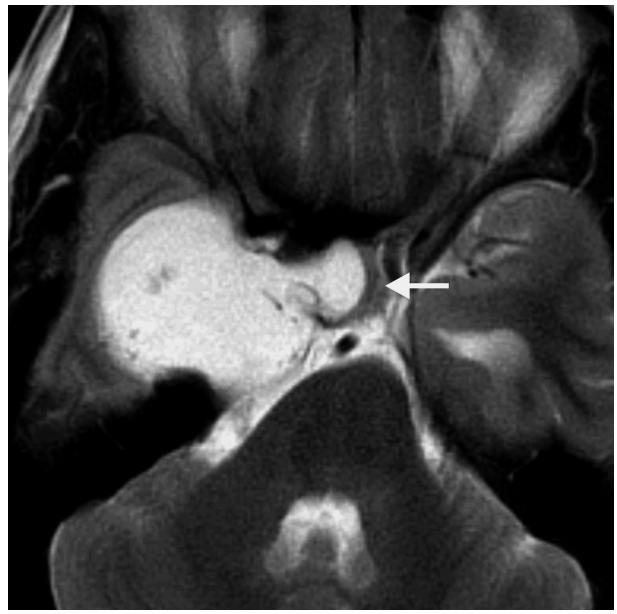


Figure 2.4.3.14 Cavernous haemangioma of the cavernous sinus with lateral expansion toward the temporal lobe and medial expansion with displacement of the pituitary gland (arrow). The lesion appears strongly hyperintense on the axial T_2-weighted image.

Symptoms: Clinically, cranial neuropathies, visual compromise, and headaches are in the foreground [32].

Neuroradiological findings: CSH appear strongly hyperintense on T_2-weighted MRI (Figure 2.4.3.14), and show marked contrast uptake.

Therapy and outcome: In contemporary surgical series, resection of CSH was mostly performed by a transcranial, extradural approach with total resection in more than 80% of the cases [37]. A high blood loss must be anticipated in these highly vascular lesions. A high efficacy and safety have recently been shown with primary radiosurgery or fractionated radiotherapy for CSH. A 100% control rate has been reported with frequent postprocedural recovery of cranial neuropathy [38]. The recent literature indicates that radiotherapy is superior to surgery in the treatment of CSH. No valid data on endocrinological outcome with the different treatment modalities exist.

Cavernous Haemangiomas of the Optic Chiasm

Cavernous haemangiomas of the optic chiasm are vascular malformations composed of closely apposed dilated vascular channels. They are a rare differential diagnosis in suprasellar tumours. Clinically, acute or subacute visual impairment due to acute bleeding occurred in most of the cases described. MRI may reveal bleeding into the optic chiasm, and a berry-shaped 'tumour' arising in the optic chiasm is found. Therapeutically, evacuation of a haematoma and resection of the lesion is performed via a transcranial approach.

REFERENCES

1. Wass JA, Trifanescu R, Ansorge O, Grossman AB, Karavitaki N. Rathke's cleft cysts. *Clin Endocrinol (Oxf)*, 2012; **76**: 151–60.
2. Mendelson ZS, Husainb Q, Elmoursi S, Sviderb PF, Eloy JA, Liu JK. Rathke's cleft cyst recurrence after transsphenoidal surgery: a meta-analysis of 1151 cases. *J Clin Neurosci*, 2014; **21**: 378–85.

3. Lin M, Wedemeyer MA, Bradley D, *et al*. Long-term outcomes following transsphenoidal surgery in patients with Rathke's cleft cysts. *J Neurosurg*, 2018; **18**: 1–7.
4. Trifanescu R, Stavrinides V, Plaha P, *et al*. Outcome in surgically treated Rathke's cleft cysts: long-term monitoring needed. *Eur J Endocrinol*, 2011; **165**: 33–7.
5. Nomikos P, Buchfelder M, Fahlbusch R. Intra- and suprasellar colloid cysts. *Pituitary*, 1999; **2**: 123–6.
6. Dubuisson AS, Stevenaert A, Martin DH, Flandroy PP. Intrasellar arachnoid cysts. *Neurosurgery*, 2007; **61**: 505–13.
7. McLaughlin N, Vandergrift A, Ditzel Filho LF, *et al*. Endonasal management of sellar arachnoid cysts: simple cyst obliteration technique. *J Neurosurg*, 2012; **116**: 728–40.
8. Sahm F, Schrimpf D, Olar A, *et al*. TERT promoter mutations and risk of recurrence in meningioma. *J Natl Cancer Inst*, 2015; **13**: 108.
9. Fujio S, Hirano H, Yamashita M, *et al*. Preoperative and postoperative pituitary function in patients with tuberculum sellae meningioma—based on pituitary provocation tests. *Neurol Med Chir (Tokyo)*, 2017; **57**: 548–56.
10. Magill ST, Morshed RA, Lucas CHG, *et al*. Tuberculum sellae meningiomas: grading scale to assess surgical outcomes using the transcranial versus transsphenoidal approach. *Neurosurg Focus*, 2018; **44**: E9.
11. Honegger J, Fahlbusch R, Buchfelder M, Huk WJ, Thierauf P. The role of transsphenoidal microsurgery in the management of sellar and parasellar meningioma. *Surg Neurol*, 1993; **39**: 18–24.
12. Walsh MT, Couldwell WT. Management options for cavernous sinus meningiomas. *J Neurooncol*, 2009; **92**: 307–16.
13. Lee CC, Trifiletti DM, Sahgal A, *et al*. Stereotactic radiosurgery for benign (World Health Organization Grade I) cavernous sinus meningiomas—International Stereotactic Radiosurgery Society (ISRS) Practice Guideline: a systematic review. *Neurosurgery*, 2018; **83**: 1128–42.
14. He W, Chen F, Dalm B, Kirby PA, Greenlee JDW. Metastatic involvement of the pituitary gland: a systematic review with pooled individual patient data analysis. *Pituitary*, 2015; **18**: 159–68
15. Morita A, Meyer FB, Laws ER. Symptomatic pituitary metastases. *J Neurosurg*, 1998; **89**: 69–73.
16. Labadi M, Watanabe K, Bouazza S, *et al*. Clivus chordomas: a systematic review and meta-analysis of contemporary surgical management. *J Neurosurg*, 2016; **60**: 476–84.
17. Uhl M, Mattke M, Welzel T, *et al*. Highly effective treatment of skull base chordoma with carbon ion irradiation using a raster scan technique in 155 patients: first long-term results. *Cancer*, 2014; **120**: 3410–17.
18. Carlson ML, O'Connell BP, Breen JT, *et al*. Petroclival chondrosarcoma: a multicenter review of 55 cases and new staging system. *Otol Neurotol*, 2016; **37**: 940–50.
19. Rosenberg, AE, Nielsen GP, Keel SB, *et al*. Chondrosarcoma of the base of the skull: a clinicopathologic study of 200 cases with emphasis on its distinction from chordoma. *Am J Surg Pathol*, 1999; **23**: 1370–8.
20. Martinez R, Honegger J, Fahlbusch R, Buchfelder M. Endocrine findings in patients with optico-hypothalamic gliomas. *Exp Clin Endocrinol Diabetes*, 2003; **111**: 162–7.
21. Rasool N, Odel JG, Kazim M. Optic pathway glioma of childhood. *Curr Opin Ophthalmol*, 2017; **28**: 289–95.
22. Gnekow AK, Falkenstein F, von Hornstein S, *et al*. Long-term follow-up of the multicenter, multidisciplinary treatment study HIT-LGG 1996 for low-grade glioma in children and adolescents of the German speaking Society of Pediatric Oncology and Hematology. *Neuro Oncol*, 2012; **14**: 1265–84.
23. Jorsal T, Rørth M. Intracranial germ cell tumours. A review with special reference to endocrine manifestations. *Acta Oncol*, 2012; **51**: 3–9.
24. Murray MJ, Bartels U, Nishikawa R, Fangusaro J, Matsutani M, Nicholson JC. Consensus on the management of intracranial germ-cell tumours. *Lancet Oncol*, 2015; **16**: e470–7.
25. Calaminus G, Kortmann R, Worch J, *et al*. SIOP CNS GCT 96: final report of outcome of a prospective, multinational nonrandomized trial for children and adults with intracranial germinoma, comparing craniospinal irradiation alone with chemotherapy followed by focal primary site irradiation for patients with localized disease. *Neuro Oncol*, 2013; **15**: 788–96.
26. Jung H, Probst EN, Hauffa BP, Partsch CJ, Dammann O. Association of morphological characteristics with precocious puberty and/or gelastic seizures in hypothalamic hamartoma. *J Clin Endocrinol Metab*, 2003; **88**: 4590–5.
27. Procaccini E, Dorfmüller G, Fohlen M, Bulteau C, Delalande O. Surgical management of hypothalamic hamartomas with epilepsy: the stereoendoscopic approach. *Neurosurgery*, 2006; **59**(4 Suppl 2): ONS336–44; discussion ONS344–6.
28. Régis J, Lagmari M, Carron R, *et al*. Safety and efficacy of Gamma Knife radiosurgery in hypothalamic hamartomas with severe epilepsies: a prospective trial in 48 patients and review of the literature. *Epilepsia*, 2017; **58**(Suppl. 2): 60–71.
29. Puchner MJA, Lüdecke DK, Saeger W, Riedel M, Asa SL. Gangliocytomas of the sellar region—a review. *Exp Clin Endocrinol Diabetes*, 1995; **103**: 129–49.
30. Mete O, Lopes MB, Asa SL. Spindle cell oncocytomas and granular cell tumors of the pituitary are variants of pituicytoma. *Am J Surg Pathol*, 2013; **37**: 1694–9.
31. Covington MF, Chin SS, Osborn AG. Pituicytoma, spindle cell oncocytoma, and granular cell tumor: clarification and meta-analysis of the World Literature since 1893. *AJNR Am J Neurorad*, 2011; **32**: 2067–72.
32. Schaller B, Krisch E, Tolnay M, Mindermann T. Symptomatic granular cell tumor of the pituitary gland: case report and review of the literature. *Neurosurgery*, 1998; **42**: 166–71.
33. Feng M, Carmichael JD, Bonert V, Bannykh S, Mamelak AN. Surgical management of pituicytomas: case series and comprehensive literature review. *Pituitary*, 2014; **17**: 399–413.
34. Hanak BW, Zada G, Nayar VV, *et al*. Cerebral aneurysms with intrasellar extension: a systematic review of clinical, anatomical, and treatment characteristics: a review. *J Neurosurg*, 2012; **116**: 164–78.
35. Pierot L, Wakhloo AK. Endovascular treatment of intracranial aneurysms: current status. *Stroke*, 2013; **44**: 2046–54.
36. Khajeh L, Blijdorp K, Neggers SJ, Ribbers GM, Dippel DW, van Kooten F. Hypopituitarism after subarachnoid haemorrhage, do we know enough? *BMC Neurol*, 2014; **81**: 529–37.
37. Li MH, Zhao JL, Li YY, Zeng CH, Xu GS, Hong T. Extradural transcavernous approach to cavernous sinus hemangiomas. *Clin Neurol Neurosurg*, 2015; **136**: 110–15.
38. Lee CC, Sheehan JP, Kano H, *et al*. Gamma Knife radiosurgery for hemangioma of the cavernous sinus. *J Neurosurg*, 2017; **126**: 1498–505.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

2.4.4 Lymphocytic Hypophysitis and other Inflammatory Conditions of the Pituitary

Mark E. Molitch and Jelena Kravarusic

Introduction

Inflammatory conditions of the pituitary are far less common than pituitary adenomas. However, their incidence may be rising due to increased recognition based on symptomatology, laboratory tests, better radiology methods and new immune therapies for cancer whose immune-related adverse effects (IRAE) include hypophysitis.

Although the most common of these, lymphocytic hypophysitis, is limited to the pituitary and pituitary stalk, many of the other lesions are usually part of a systemic process. Nonetheless, even these lesions, such as Langerhans' cell histiocytosis, sarcoidosis and newly recognized immunoglobulin G4-related hypophysitis (IgG4-RH) sometimes are limited to the central nervous system (CNS) and, rarely, present as isolated lesions of the hypothalamic/pituitary area. When lesions are located in the base of the hypothalamus or in the stalk, they commonly present with a combination of diabetes insipidus (DI) and hypopituitarism. In some cases, hypothalamic infiltration may be more widespread, affecting a variety of additional hypothalamic functions, such as satiety, sleep, and temperature regulation. These inflammatory lesions tend to be progressively destructive, resulting ultimately in fibrosis, but the rate of progression is highly variable. When hypopituitarism or DI occur, they rarely recover even if the underlying process is directly treated. Thus, these lesions present more with endocrine hypofunction than with mass effects, although in early stages lymphocytic hypophysitis, may well present with mass effects to the point where it can be confused with a pituitary adenoma.

Although hypophysitis is usually thought of as a primary process, it may occur secondarily in relation to infection (viral, bacterial, fungal, tuberculosis, syphilis) or other processes such as Langerhans cell histiocytosis, sarcoidosis, Wegener's granulomatosis, Crohn-Takayasu disease and ruptured cysts [1, 2–4] and increasingly recognized since 2004, as a part of immunoglobulin G4-related disease (IgG4RD).

Lymphocytic Hypophysitis

Lymphocytic hypophysitis is associated with hypopituitarism and a sellar mass. It is most commonly seen in the peri- or postpartum period but it has also been reported after menopause [1]. The diagnosis may be challenging, as the clinical and radiographic distinction from pituitary adenomas and other sellar masses is often not obvious. The disease is presumably autoimmune in aetiology, although there has never been a specific target antigen identified [1].

Epidemiology

About 500 patients with primary lymphocytic hypophysitis have been described in the literature [5]. It affects women more frequently than men, with a reported ratio of about 5:1; however, the female:male ratio is decreasing in recent years as more men are reported. The mean age at diagnosis is approximately 35 years for women and 45 years for men [5]. Of the 57% of patients developing the disorder in association with pregnancy, most occur during the last month of pregnancy or during the first 2 months postpartum [1, 2]. It has rarely been described in children [6].

Classification and Pathology

Anatomically, the most common form is lymphocytic adenohypophysitis, where anterior pituitary cells and hormones are affected but posterior pituitary involvement is absent or minimal. With the much less common lymphocytic infundibuloneurohypophysitis, the posterior pituitary is primarily involved, causing DI and anterior pituitary function is usually preserved [1, 3, 4]. Lymphocytic infundibulopanhypophysitis is even more rare, with lymphocytic infiltration and destruction present in both the anterior and posterior pituitary. These patients present with a combination of DI and anterior pituitary deficiency [1, 3, 4].

Pathologically, primary hypophysitis has been described in three forms: lymphocytic, granulomatous, and xanthomatous. Lymphocytic hypophysitis is characterized by a dense lymphocytic infiltration of the anterior pituitary with destruction of the normal pituitary architecture and replacement with fibrosis [1, 3, 4, 7] (Figure 2.4.4.1). The lymphocytes are predominantly cytotoxic T lymphocyte (CD8+) cells, suggesting that T-cell-mediated cytotoxicity is critical in the pathogenesis of the disorder [8].

Pathogenesis

The aetiology of lymphocytic hypophysitis is unknown but it has been speculated to have an autoimmune basis [1, 3, 4, 8]. Nearly 30% of patients have a history of coexisting autoimmune diseases such as Hashimoto's thyroiditis, Addison's disease, type 1 diabetes and pernicious anaemia [1, 3, 4, 7], and the condition is now considered a component of the type 1 polyglandular autoimmune syndrome [1, 3, 4]. In one recent prospective study antinuclear and antiextractable nuclear antigens (Ro, La, Sm, RNP, Scl-70, and Jo1) were detected; thus, perhaps all patients with autoimmune hypophysitis should be evaluated for other organ specific and nonspecific antibodies [9].

Although antipituitary antibodies have been demonstrated in some patients with lymphocytic hypophysitis, their specificity for

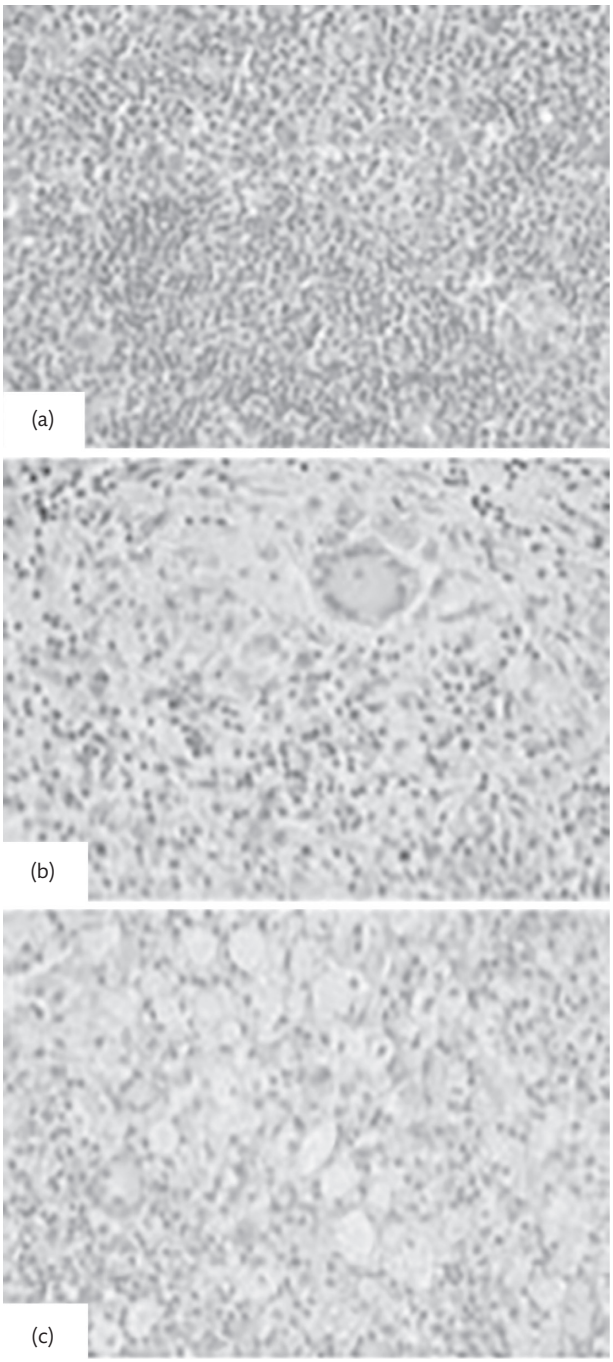


Figure 2.4.4.1 Histologic subtypes of primary hypophysitis. (a) Lymphocytic hypophysitis. Note massive lymphocytic infiltration of pituitary with scattered islands of preserved pituitary cells. (b) Idiopathic granulomatous hypophysitis. Characteristic multinucleated giant cells and granuloma surrounded by fibrosis, sparse infiltration of plasma cells. (c) Xanthomatous hypophysitis. Predominance of foamy macrophages, a few lymphocytes, and single plasma cells (haematoxylin and eosin, original magnification ×40).

hypophysitis is poor, as they are also present in patients with non-autoimmune pituitary disease, other non-pituitary autoimmune disease [10, 11] and in normal postpartum women who do not develop hypophysitis. The pathogenic pituitary autoantigen(s) remain to be elucidated, although several candidates have been proposed [5].

Presentation

When associated with pregnancy, lymphocytic hypophysitis typically presents in the third trimester of pregnancy or within 1 year postpartum, with symptoms usually related to a pituitary mass (headaches or visual symptoms) or hypopituitarism. The disorder often comes to attention due to failure of either lactation or menses following delivery [1, 2–4, 12]. Neurohypophyseal involvement, manifesting as DI, occurs in 15% of cases [1, 2–4, 7, 12]. Other rare presentations of lymphocytic hypophysitis include meningeal irritation, diplopia due to cavernous sinus involvement and occlusion of the internal carotid arteries.

Symptoms resulting from partial or panhypopituitarism occur in approximately 80% of cases, and multiple deficiencies are found in approximately 75% of cases [1, 2–4, 8, 12] as illustrated in the Table 2.4.4.1.

There is an inexplicable unique predilection for the corticotrophs and thyrotrophs to be affected while the gonadotrophs may be spared. Prolactin (PRL) levels range from unmeasurable to elevated; low levels are attributable to destruction of the lactotrophs, while hyperprolactinaemia is expected during pregnancy and the early postpartum period. However, elevated prolactin levels have been reported in cases of lymphocytic hypophysitis in men and non-pregnant women [7, 13, 14], which is likely secondary to compression of the pituitary stalk.

Radiology

Magnetic resonance imaging (MRI) in hypophysitis commonly shows an enlarged pituitary gland, often with suprasellar extension and stalk thickening [1, 2–4] (Figure 2.4.4.2).

The gland is generally symmetrically enlarged and administration of gadolinium homogeneously enhances the gland.

Table 2.4.4.1 Lymphocytic hypophysitis: clinical presentation

Symptoms	Frequency
Mass effects	
Headache	60%
Visual disturbance	40%
Bitemporal hemianopsia	32%
Impaired visual acuity	16%
Diplopia	<5%
Endocrine dysfunction	80%
Adrenal insufficiency	65%
Hypothyroidism	60%
Growth hormone deficiency	54%
Hypogonadism	40%
Hyperprolactinaemia	30%
Diabetes insipidus	15%

Data abstracted from Beressi *et al.* [12] based on analysis of 145 cases of clinically suspected and biopsy-proven lymphocytic hypophysitis.

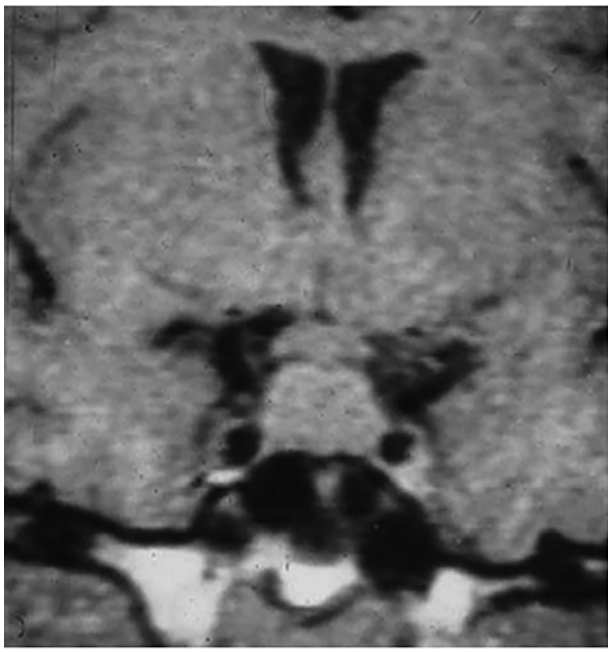


Figure 2.4.4.2 Lymphocytic hypophysitis on coronal section in T1 phase. The pituitary gland is diffusely and symmetrically enlarged extending into the suprasellar region. The floor of the sella is intact.

In contrast, in adenomas gadolinium enhances the gland more focally as described in the Table 2.4.4.2. In lymphocytic hypophysitis, the pituitary displays a relative low signal on T_1- and a relatively high signal on T_2-weighted images. By comparison, in macroadenomas a low signal on T_1-weighted images is uncommon, but a high signal on T_2-weighted images is occasionally seen. Often, the dura mater adjacent to the mass in lymphocytic hypophysitis shows a unique, marked contrast enhancement referred to as a 'dural tail'. In late stages, these MRI findings may be absent due to shrinkage of the mass with resolution of the inflammatory process, and fibrotic changes and an empty sella may be seen [14, 15].

Diagnosis

Lymphocytic hypophysitis should be considered in the differential diagnosis of pituitary masses and/or hypopituitarism in females who are pregnant or in the early postpartum period. This is especially true in cases associated with other autoimmune diseases or unusual patterns of hormone deficiencies. In the past many individuals with postpartum hypopituitarism who lacked a history of hypovolemic shock were inadvertently labelled as having Sheehan's syndrome when, in fact, they had hypophysitis.

A definitive diagnosis of lymphocytic hypophysitis requires tissue biopsy. However, it may be possible to make a presumptive clinical diagnosis in patients who meet the following criteria: (1) a history of gestational or postpartum hypopituitarism, especially after a delivery uncomplicated by haemorrhage or hypotension; (2) a contrast enhancing sellar mass with imaging features characteristic of lymphocytic hypophysitis; (3) a pattern of pituitary hormone deficiency with early loss of adrenocorticotrophic hormone (ACTH) and thyroid stimulating hormone (TSH) unlike that typically found with macroadenomas (i.e. sequential loss of growth hormone [GH], luteinizing hormone [LH]/follicle-stimulating hormone [FSH], ACTH, and TSH); (4) relatively rapid development of hypopituitarism in contrast to the expected slow development of hypopituitarism that would be expected with an adenoma; and (5) a degree of pituitary failure disproportionate to the size of the mass. Nevertheless, biopsy may be required in situations in which a distinction cannot be made between lymphocytic hypophysitis and a non-functioning macroadenoma or prolactinoma and when neurologic signs develop.

Lymphocytic Infundibuloneurohypophysitis

Lymphocytic hypophysitis and infundibuloneurohypophysitis likely represent distinctly separate pathologic entities, as the latter tends to occur in older patients and is less likely to be associated with pregnancy [1, 2–4]. Lymphocytic infundibuloneurohypophysitis causes central DI and spares the anterior pituitary as a result of an inflammatory process confined to the stalk and posterior pituitary [16]. The radiologic features are generally more clearly delineated: thickening of the pituitary stalk or neurohypophysis and homogeneous enhancement of the pituitary stalk or neurohypophysis after the administration of contrast material [17].

Table 2.4.4.2 MRI characteristics of lesions of the hypothalamus/pituitary

Type of lesion	Signal intensity on T_1	Signal intensity on T_2	Contrast enhancement	Pattern of enhancement	Shape	Dural enhancement
Hypophysitis	Relatively low	High	Marked	Homogeneous	Symmetric	Common
Histiocytosis	Isointense	Hyperintense	Moderate	Non-specific	Stalk thickening	Common
Sarcoidosis	Isointense	Hyperintense	Moderate	Non-specific	Stalk thickening	Leptomeningae
Wegener's granulomatosis	Isointense	Hyperintense	Intense	Homogenous	Superior infundibulum thickening	Common, linear
Immune checkpoint related	Isointense to hypointense	Hyperintense	Marked	Non-specific	Nodular stalk thickening	Common
IgG4 related	Iso to hypointense	Hyperintense	Marked	Non-specific	Stalk thickening	Not reported
Pituitary Adenoma	Isointense	Usually isointense	Moderate	Focal	Dumbbell	Rare

Data abstracted from [3, 15, 16, 31, 39, 42, 43].

Management

The natural history of lymphocytic hypophysitis is variable and unpredictable. Typically, the pituitary initially becomes inflamed, oedematous, and enlarged, and the patient develops symptoms secondary to mass effects. Progressive fibrosis develops, causing destruction of the parenchyma leading to hypopituitarism. In some cases, the course is aggressive and neurologic and hormonal deficits progress rapidly [1, 2–4, 7]. However, cases of spontaneous partial or full recovery of pituitary function, as well as resolution of pituitary masses in the absence of any intervention, have been well documented [18–20]. Because the natural history of lymphocytic hypophysitis is so variable, appropriate management remains controversial.

Controlled therapeutic trials are not feasible due to the rarity of hypophysitis, an inability to make definitive diagnosis without histologic proof, and the considerable variability in the natural history of the disorder. Until recently, preoperative suspicion of the diagnosis was rare due to under-recognition, and the traditional diagnostic and therapeutic approach involved transsphenoidal biopsy, exploration, and/or pituitary resection. Consequently, cases illustrating only transient compressive effects and endocrine dysfunction support the concept of conservative management. With a greater knowledge of the course of lymphocytic hypophysitis and the ability to make a presumptive diagnosis in highly suggestive cases, it is possible to avoid routine neurosurgical exploration in many cases.

Corticosteroid therapy has been advocated as a means of attenuating inflammation and, in some patients, has been associated with return of pituitary function and reduction of the mass [21, 22]. Conversely, cases have also been reported in which lymphocytic hypophysitis failed to improve with glucocorticoid therapy [7, 23]. There are also a few documented cases of improvement in symptoms with administration of corticosteroids followed by a relapse when therapy was discontinued [24]. It is unclear, however, whether improvement in the clinical course is directly attributable to corticosteroid treatment or simply reflects the natural course of the disease [18–20]. Given the uncertainty regarding the efficacy of corticosteroid treatment and its known adverse effects, such therapy does not seem justified for most patients.

Patients with a presumed diagnosis of lymphocytic hypophysitis should be observed closely and undergo serial visual field examinations or an MRI if they are managed medically. Surgical decompression of the pituitary mass may be required if the patient fails conservative therapy, as demonstrated by progressive radiologic or neurologic deterioration or by signs of optic nerve compression. However, in this situation, some would argue for a short course of steroids [1, 2–4, 7]. The optimal surgical strategy involves only partial resection of the mass to decompress the surrounding structures via a transsphenoidal approach rather than an attempt at complete resection, because surgery rarely improves endocrine dysfunction. All patients with lymphocytic hypophysitis require appropriate replacement therapy for deficient hormones. Long-term follow-up is mandatory to monitor for the development of other hormonal deficits. Because hypopituitarism is temporary in a subset of patients, a careful attempt should be made to withdraw hormone replacement after resolution of the inflammatory stage if progression to fibrosis does not result in irreversible hypopituitarism.

Langerhans Cell Histiocytosis

Langerhans cell histiocytosis (LCH) is a rare disorder characterized by clonal proliferation of abnormal dendritic antigen-presenting histiocytes, known as Langerhans cells, with an accompanying infiltrate of lymphocytes, eosinophils, and neutrophils resulting in the destruction of a variety of tissues. LCH is also regarded as an inflammatory disease because an altered expression of cytokines and cellular adhesion molecules important for the migration and homing of Langerhans cells has been demonstrated [25, 26].

Epidemiology

LCH is usually considered to be a disease of childhood, with a peak incidence at the ages of 1 to 3 years [27]. Overall, the incidence is 3–5 cases per million per year, with a male to female ratio of 2:1 [27]. In adults, the mean age at diagnosis is 33 years and it is seen even more rarely, the estimated prevalence being 1–2 cases per million [28].

Presentation

In adults, LCH has a predilection for the hypothalamus and pituitary. When only patients with multisystem disease are included, the prevalence of DI can be as high as 40% and DI is considered to be the most common endocrine disease-related permanent consequence [29]. DI can also be the presenting feature, predating the diagnosis of LCH. Established DI is generally permanent and does not respond to any disease-modifying treatment; hence the only treatment is desmopressin [30]. DI associated with structural abnormalities of the hypothalamus and pituitary often heralds the development of anterior pituitary hormone deficiencies and CNS involvement [31]. Anterior pituitary dysfunction is found in up to 20% of patients with LCH, and is almost always associated with DI [30, 32]. Once established, anterior pituitary deficiencies seem to be permanent and are not affected by any form of LCH disease-modifying treatment [31]. The most frequent anterior pituitary hormone deficiency is that of growth hormone, which is found in up to 42% of patients and generally diagnosed with a latency of 1 year from the diagnosis of DI. Deficiencies of LH/FSH are next most common, with a latency of 7 years from the diagnosis of DI [32]. Therefore, if a partial pituitary hormone deficiency is identified in a patient with LCH, regular monitoring for the remaining hormones is advised. In addition to pituitary involvement, up to 40% of patients with DI have hypothalamic infiltration which results in non-endocrine hypothalamic manifestations, including abnormal eating patterns, morbid obesity, and disturbances in social behaviour, temperature, sleep pattern, and thirst. DI may be particularly difficult to manage in patients with impaired memory.

Radiology

Patients with LCH and DI commonly demonstrate a loss of the hyperintense signal of the posterior pituitary on T_1-weighted images ('bright spot') on MRI [30]. Infundibular enlargement is present in up to 71% of patients at the time of diagnosis of DI [33] as illustrated in Figure 2.4.4.3 [34].

Hypothalamic mass lesions have been described in 8–18% of patients exhibiting one or more pituitary hormone deficiencies [33].

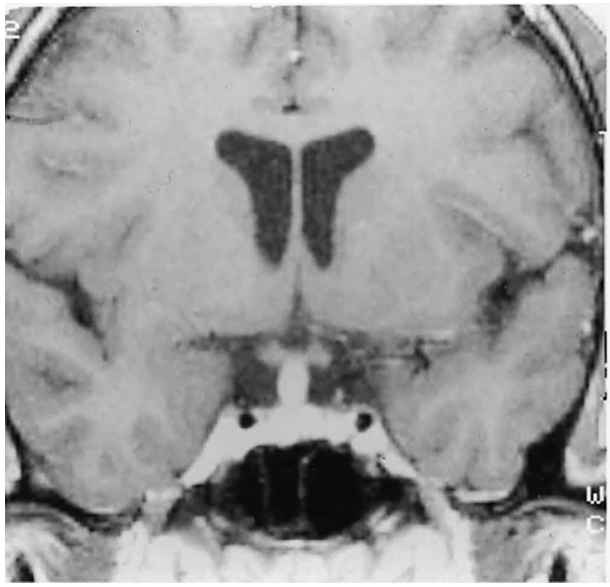


Figure 2.4.4.3 Thickening of pituitary stalk (arrow) due to biopsy-proven Langerhans cell histiocytosis.

Reproduced with permission from Purdy LP, Molitch ME (1998). Sudden onset of diabetes insipidus in an adolescent. *EndoTrends*, 5, 1–7.

The LCH lesions typically are isointense on T_1 images, hyperintense on T_2 images and enhance with gadolinium [33].

Diagnosis

In order to make the diagnosis of LCH, one must search for extracranial manifestations of LCH with a radiographic skeletal survey, skull series, chest X-ray, and bone scan so that these lesions can be biopsied. Osteolytic lesions due to LCH may be present in the jaw or mastoid, so radiographs of the jaw are a worthwhile part of the diagnostic evaluation. When biopsies of other tissues show LCH and the MRI and clinical picture are compatible, biopsy of the hypothalamic/stalk lesion is rarely necessary.

To establish a diagnosis according to the published criteria of the Histiocytosis Society, a tissue biopsy must either show presence of pathognomonic Birbeck granules on electron microscopy or stain positive for CD1a [35]. Birbeck granules are pentalaminar cell inclusions that sometimes have a 'tennis racquet' dilated terminal appearance [36]. Their exact function is unknown, but some studies implicate them in antigen processing.

Management

The course of LCH is often unpredictable, varying from spontaneous regression and resolution to rapid progression and death or repeated recurrence with a considerable risk of permanent sequelae. Patients with disease that is localized to one organ system, usually in the bone, skin, or lymph nodes, have a good prognosis, and need minimal or even no treatment. In contrast, multiple organ involvement carries a risk of a poor outcome, including 10–20% mortality and a 50% risk of life-impairing morbidity. Therefore, an early diagnosis and close follow-up is critical. The mainstays of treatment of LCH have been surgery and radiation over the years. However, vinblastine in combination with steroids is now the most frequently used initial therapy for multisystem disease [32]. Cladribine (2-chlorodeoxyadenosine) can be effective for adults with recurrent and/or disseminated disease [37].

Anterior and posterior pituitary hormonal deficits are replaced as necessary. As with other infiltrative diseases of the hypothalamus, the anterior pituitary hormone deficits may gradually appear. Therefore, periodic testing for many years and then treatment of new deficits may be necessary.

Sarcoidosis

The prevalence of CNS involvement in sarcoidosis is 5–15% and most of these patients are found to have non-caseating granulomas in the hypothalamo–pituitary region in addition to the leptomeninges and cranial nerves [38]. The most commonly found hormonal abnormality is DI (17–90% of patients), followed by hyperprolactinaemia (3–32%) [39]. Hypothalamic involvement may also cause obesity, somnolence with disruption of sleep cycle, alteration in the thirst centre and loss of short-term memory [40].

MRI usually shows pituitary stalk thickening and enhancement as well as pituitary enlargement. Periventricular lesions and leptomeningeal enhancement can be seen in sarcoidosis and this can help distinguish it from lymphocytic hypophysitis. Significant laboratory findings that may aid in diagnosis are elevated levels of serum and cerebrospinal fluid angiotensin converting enzyme (ACE). As with LCH, a search for other systemic tissue involvement is important, so that a biopsy can be obtained.

Management of sarcoidosis frequently involves the use of steroids, but recovery of anterior and posterior pituitary function usually does not occur [39]. Recently, cladribine was also found to reverse DI caused by sarcoidosis [41]. The hypothalamic/pituitary involvement may be gradual and progressive, so that periodic testing is necessary and hormonal deficits treated as they develop.

Granulomatosis with Polyangiitis (GPA, Formerly Known as Wegener's Granulomatosis)

Wegener's granulomatosis is a systemic vasculitis affecting small and medium sized vessels, most commonly in the respiratory tract and kidneys; the pituitary is involved in less than 1% of cases [42]. Involvement of the pituitary can occur via direct extension from nasal, paranasal, or orbital disease, from remote granulomatous involvement, or from vasculitis of the hypothalamus. Patients most frequently present with DI but hyperprolactinaemia and panhypopituitarism have also been reported [3].

The finding of high titres of antineutrophil cytoplasmic antibody (c-ANCA) can be diagnostic but in some cases biopsy of affected tissue is required. When there is hypothalamic/pituitary involvement, MRI reveals an enlarged pituitary with homogeneous enhancement, thickening and enhancement of the pituitary stalk, and enhancement of the optic chiasm [3, 43].

Wegener's granulomatosis is usually treated with glucocorticoids and/or cyclophosphamide. However, such treatment does not usually lead to reversal of the hypopituitarism. Similar to other infiltrative disease, the destruction may be gradual, necessitating repeated testing and treatment of hormonal deficits as they develop.

Immune Checkpoint Therapy-Related Hypophysitis

Ipilimumab is a monoclonal IgG1 antibody against CTLA4 (cytotoxic T lymphocyte antigen 4), an inhibitory molecule expressed on activated and regulatory T cells, ultimately dampening T cell activation. If CTLA4 is blocked, T cells remain active, and this effect is used in cancer destruction. Additionally, there are other immune checkpoint inhibitors, monoclonal antibodies antiprogrammed cell death protein 1 (PD-1) nivolumab and pembrolizumab and antiprogrammed cell death-ligand 1 (PD-L1) atezolizumab, durvalumab, and avelumab, which also cause IRAE, but notably less commonly hypophysitis.

IRAE hypophysitis due to ipilimumab given for therapy of metastatic melanoma was first reported in 2003 [44], and is now observed in approximately 10–15% of patients receiving agents targeting CTLA4 [45–47]. Hypophysitis is significantly less common following use of PD1 antibodies [48]. IRAE hypophysitis seems to favour older men; however, this observation is limited by epidemiology of cancers these medications are used to treat. Some reports show a higher incidence of hypophysitis in combination PD1 plus CTLA4 therapy [47] but others found no difference or a lower rate compared to ipilimumab alone [49].

Pathogenesis

The pathogenesis of cancer immune therapy-related hypophysitis is unclear. One hypothesis is that CTLA-4 blocking antibodies bind pituitary cells expressing CTLA4 antigen and cause hypophysitis through type IV (T-cell dependent) and type II (IgG dependent) immune mechanisms, as evidenced by finding macrophage infiltration, complement fixation, and lymphocyte activation [50]. Antibodies towards PD1 are from the IgG4 class, which cannot activate complement and are not effective mediators for antibody dependent cell mediated cell toxicity [6]. Interestingly, patients treated with PD1 and/or PDL1 IgG4 antibodies instead of IgG1 used in ipilimumab rarely developed pituitary damage [48]. Also, hypophysitis occurs less frequently in patients receiving tremelimumab, an IgG2 antibody which blocks CTLA4 [50].

Presentation

The clinical presentation is similar to other forms of hypophysitis, with a headache, less commonly with visual disturbances and DI and usually presents approximately 2–3 months after starting therapy.

Radiology

Radiological findings are not specific, and may show mild-to-moderate diffuse enlargement of the pituitary gland, with either homogeneous or heterogeneous enhancement after contrast administration on pituitary MRI [45, 51]. Radiological changes can precede the clinical diagnosis by several weeks [45]. Many case reports have highlighted the incidental detection of hypophysitis when positron emission tomography (PET) is used for surveillance. Importantly a normal MRI does not rule out hypophysitis and management should be based on clinical presentation and evaluation of pituitary hormone levels [52].

Management

Corticosteroids have been used for immune therapy-related hypophysitis treatment. High-dose steroids do not seem to improve outcome [53], but can be considered for critically ill patients, either due to hypophysitis or hypopituitarism, significant hyponatraemia, severe headache, or pituitary enlargement with mass effect on the optic chiasm. Hydrocortisone in doses 15–25 mg daily [48] can help with the headache and fatigue, treat adrenal insufficiency and still not prevent the patient from enrolment in a study protocol which may exclude patients treated with high-dose steroids due to their immunomodulatory effect [54].

The clinical response to cancer immunotherapy has been linked with development of IRAEs; in several studies a direct correlation was observed between relapse free survival and development of IRAEs [44, 45]. However, this observation may be biased against patients who don't survive long enough to be observed. This phenomenon needs further investigation, for its potential benefit in prognosis, if any, to be considered.

IGG4-Related Hypophysitis

IgG4-related disease (IgG4RD) is a systemic autoimmune disease characterized by inflammatory pseudotumour (IPT) infiltrates containing IgG4 rich-plasma cells, storiform (cartwheel or whirling pattern) fibrosis, obliterative vasculitis, and often but not always elevated IgG4 serum levels [55]. IgG4 is the least abundant of all IgG antibodies and is associated with several autoimmune and allergic diseases such as Mikulicz disease, retroperitoneal fibrosis, autoimmune pancreatitis, and Riedel thyroiditis [56, 57]. It is unclear what the role is of IgG4, as it is not found in normal pituitaries [58] and has a role in immune downregulation [59]. It has been postulated to possibly be a by-product of an exaggerated immune response [60].

Pathogenesis

Iwata *et al.* showed antipituitary antibodies in 5 out of 17 patients with IgG4-related hypophysitis (IgG4RH) (29%), and in none of the pituitaries of healthy controls. Antibodies were of the IgG1 subclass rather than IgG4 in all five cases, suggesting that IgG4 does not have a direct role in the pathogenesis of the hypophysitis [61].

The prevalence of IgG4RHs unknown, and it depends on the diagnostic criteria and mostly retrospective analyses. As this condition was not previously recognized, some cases previously classified as lymphocytic hypophysitis may have been IgG4RH. Bernreuther *et al.* used histological and immune staining to retrospectively analyse cases previously reported as primary hypophysitis and found that 12 of 29 (41.4%) fulfilled the criteria for IgG4RH [62].

Presentation

It is most commonly seen in older men, in contrast to lymphocytic hypophysitis, which is more common in younger women. Presenting symptoms can mimic any sellar mass with compressive symptoms: headache, visual disturbances, and hypopituitarism [63].

Diagnosis

In an effort to make a correct diagnosis and avoid an invasive tissue biopsy, diagnostic criteria have been proposed that don't require histopathology by Leporati *et al.* [64] and are as follows: (1) mononuclear cell infiltration of the pituitary gland, rich in lymphocytes

and plasma cells, with more than ten IgG4-positive cells per high-power field or IgG4-positive/IgG-positive cells > 40% on pituitary histopathology; (2) sellar mass and/or thickened pituitary stalk on MRI; (3) a biopsy proving the involvement in other organs (association with IgG4-positive lesions in other organs); (4) an increased serum IgG4 level (>140 mg/dl); and (5) a prompt shrinkage of the pituitary mass and symptom improvement with steroids. The diagnosis of IgG4-RH is established when any of the following is fulfilled: (1) or (2) and (3) or (2), (4) and (5). For this reason, these criteria have been widely accepted followed by an increase in reports of IgG4RH. Caution should be exercised in diagnosing IgG4RH solely based on histopathology of IgG4 rich infiltrate because this can be present in other inflammatory, autoimmune, infectious, and neoplastic conditions. FDG-PET scanning may be useful in characterizing systemic involvement of tissues.

Management

The mainstay in therapy of IgG4RH are glucocorticoids, which replace deficient cortisol and lower serum IgG4 levels. There is no consensus for the optimum dose and duration, Yuen *et al.* proposed prednisolone at a dose 0.6 mg/kg for 1 to 2 months and tapered at a rate of 5 mg per week. If the disease relapses after glucocorticoid therapy discontinuation, it can be resumed and addition of rituximab can be considered [58].

Other Rare Forms of Hypophysitis

Granulomas and multinucleated giant cells are not found in lymphocytic hypophysitis and, if observed, suggest an alternative diagnosis of granulomatous hypophysitis. This rare disorder has an incidence of 1/1 000 000. Granulomatous hypophysitis occurs in men and women with equal frequency and is not particularly associated with pregnancy. It may also present as a mass lesion with hypopituitarism and pathologically is characterized by giant cell granulomas [1, 2–4, 7, 8].

Xanthomatous hypophysitis, is exceedingly rare, with only 28 cases reported to date. The pathology of xanthomatous hypophysitis is characterized by a predominance of foamy macrophages, lymphocytes, and single plasma cells [1, 8].

Other rarer types of hypophysitis appear to be part of a more generalized inflammatory process and include Rosai–Dorfman disease and fibrosing inflammatory pseudotumour (also called Tolosa–Hunt syndrome and parasellar chronic inflammatory disease) [65].

REFERENCES

1. Caturegli P, Newschaffer C, Olivi A, Pomper MG, Burger PC, Rose NR. Autoimmune hypophysitis. *Endocr Rev*, 2005; **26**: 599–614.
2. Bellastella A, Bizzarro A, Coronella C, Bellastella G, Sinisi AA, De Bellis A. Lymphocytic hypophysitis: a rare or underestimated disease? *Eur J Endocrinol*, 2003; **149**: 363–76.
3. Lury KM. Inflammatory and infectious processes involving the pituitary gland. *Top Magn Reson Imaging*, 2005; **16**: 301–6.
4. Rivera J-A. Lymphocytic hypophysitis: disease spectrum and approach to diagnosis and therapy. *Pituitary*, 2006; **9**: 35–45.
5. Caturegli P, Lupi I, Landek-Salgado M, Kimura H, Rose NR. Pituitary autoimmunity: 30 years later. *Autoimmun Rev*, 2008; **7**: 631–7.
6. Faje A. Hypophysitis: evaluation and management. *Clin Diabetes Endocrinol*, 2016; **2**: 15.
7. Thodou E, Asa SL, Kontogeorgos G, Kovacs K, Horvath E, Ezzat S. Clinical case seminar: lymphocytic hypophysitis: c linicopathological findings. *J Clin Endocrinol Metab*, 1995; **80**: 2302–11.
8. Gutenberg A, Buslei R, Fahlbusch R, Buchfelder M, Brück W. Immunopathology of primary hypophysitis. Implications for pathogenesis. *Am J Surg Pathol*, 2005; **29** 329–38.
9. Chiloiro S, Tartaglione T, Angelini F, *et al.* An overview of diagnosis of primary autoimmune hypophysitis in a prospective single-center experience. *Neuroendocrinology*, 2017; **104**: 280–90.
10. Crock PA. Cytosolic autoantigens in lymphocytic hypophysitis. *J Clin Endocrinol Metab*, 1998; **83**: 609–18.
11. Komatsu M, Kondo T, Yamauchi K, *et al.* Antipituitary antibodies in patients with the primary empty sella syndrome. *J Clin Endocrinol Metab*, 1988; **67**: 633–8.
12. Beressi N, Beressi JP, Cohen R, Modigliani E. Lymphocytic hypophysitis: a review of 145 cases. *Ann Med Interne*, 1999; **150**: 327–41.
13. Cheung CC, Ezzat S, Smyth HS, Asa SL. The spectrum and significance of primary hypophysitis. *J Clin Endocrinol Metab*, 2001; **86**: 1049–53.
14. Cebelin MS, Velasco ME, de las Mulas JM, Druet RL. Galactorrhea associated with lymphocytic adenohypophysitis. Case report. *Br J Obstet Gynaecol*, 1981; **88**: 675–80.
15. Saiwai S, Inoue Y, Ishihara T, *et al.* Lymphocytic adenohypophysitis: skull radiographs and MRI. *Neuroradiology*, 1998; **40**: 114–20.
16. Shimono T, Yamaoka T, Nishimura K, *et al.* Lymphocytic hypophysitis presenting with diabetes insipidus: MR findings. *Eur Radiol*, 1999; **9**: 1397–400.
17. Abe T. Lymphocytic infundibulo-neurohypophysitis and infundibulo-panhypophysitis regarded as lymphocytic hypophysitis variant. *Brain Tumor Pathol*, 2008; **25**: 59–66.
18. Ishihara T, Hino M, Kurahachi H, *et al.* Long-term clinical course of two cases of lymphocytic adenohypophysitis. *Endocr J*, 1996; **43**: 433–40.
19. Castle D, de Villiers JC, Melvill R. Lymphocytic adenohypophysitis. Report of a case with demonstration of spontaneous tumour regression and a review of the literature. *Br J Neurosurg*, 1988; **2**: 401–5.
20. Leiba S, Schindel B, Weinstein R, Lidor I, Friedman S, Matz S. Spontaneous postpartum regression of pituitary mass with return of function. *JAMA*, 1986; **255**: 230–2.
21. Feigenbaum SL, Martin MC, Wilson CB, *et al.* Lymphocytic adenohypophysitis: a pituitary mass lesion occurring in pregnancy. Proposal for medical treatment. *Am J Obstet Gynecol*, 1986; **164**(6 Pt 1): 1549–55.
22. Kristof RA, Van Roost D, Klingmüller D, Springer W, Schramm J. Lymphocytic hypophysitis: non-invasive diagnosis and treatment by high dose methylprednisolone pulse therapy? *J Neurol Neurosurg Psychiatry*, 1999; **67**: 398–402.
23. Reusch JE, Kleinschmidt-DeMasters BK, Lillehei KO, Rappe D, Gutierrez-Hartmann A. Preoperative diagnosis of lymphocytic hypophysitis (adenohypophysitis) unresponsive to short course dexamethasone: case report. *Neurosurgery*, 1992; **30**: 268–72.

24. Parent AD. The course of lymphocytic hypophysitis [letter; comment]. *Surgical Neurol*, 1992; **37**: 71.
25. Arceci RJ. The histiocytoses: the fall of the Tower of Babel. *Eur J Cancer*, 1999; **35**: 747–67.
26. Aricò M, Egeler RM. Clinical aspects of Langerhans cell histiocytosis. *Hematol Oncol Clin North Am*, 1998; **12**: 247–58.
27. Broadbent V, Egeler RM, Nesbit ME, Jr. Langerhans cell histiocytosis—clinical and epidemiological aspects. *Br J Cancer*, 1994; Suppl. **23**: S11–16.
28. Malpas JS. Langerhans cell histiocytosis in adults. *Hematol Oncol Clin N Amer*, 1998; **12**: 259–68.
29. Haupt R, Nanduri V, Calevo MG, *et al.* Permanent consequences in Langerhans cell histiocytosis patients: a pilot study from the Histiocyte Society–Late Effects Study Group. *Pediatr Blood Cancer*, 2004; **42**: 438–44.
30 Aricò M, Girschikofsky M, Généreau T, *et al.* Langerhans cell histiocytosis in adults. Report from the International Registry of the Histiocyte Society. *Eur J Cancer*, 2003; **39**: 2341–8.
31. Kaltsas GA, Powles TB, Evanson J, *et al.* Hypothalamo–pituitary abnormalities in adult patients with Langerhans cell histiocytosis: clinical, endocrinological, and radiological features and response to treatment. *J Clin Endocrinol Metab*, 2000; **85**: 1370–6.
32. Makras P, Alexandraki KI, Chrousos GP, Grossman AB, Kaltsas GA. Endocrine manifestations in Langerhans cell histiocytosis. *Trends Endocrinol Metab*, 2007; **18**: 252–7.
33. Ouyang DL, Roberts BK, Gibbs IC, Katznelson L. Isolated Langerhans cell histiocytosis in an adult with central diabetes insipidus: case report and review of literature. *Endocr Pract*, 2006; **12**: 660–3.
34. Purdy LP, Molitch ME. Sudden onset of diabetes insipidus in an adolescent. *EndoTrends*, 1998; **5**: 1–7.
35. Favara B, Feller A, Paulli M, *et al.*, for the WHO Committee on Histiocytic/Reticulum cell proliferations and the Reclassification Working Group of the Histiocyte Society. Contemporary classification of histiocytic disorders. *Med Pediatr Oncol*, 1997; **29**: 157–66.
36. Davis SE, Rice DH. Langerhans' cell histiocytosis: current trends and the role of the head and neck surgeon. *Ear Nose Throat J*, 2004; **83**: 340–2.
37. Pardanani A, Phyliky RL, Li CY, Tefferi A. 2-Chlorodeoxyadenosine therapy for disseminated Langerhans cell histiocytosis. *Mayo Clin Proc*, 2003; **78**: 301–6.
38. Agbogu B, Stern BJ, Sewell C, Yang G. Therapeutic considerations in patients with refractory neurosarcoidosis. *Arch Neurol*, 1995; **52**: 875–9.
39. Bullmann C, Faust M, Hoffmann A, *et al.* Five cases with central diabetes insipidus and hypogonadism as first presentation of neurosarcoidosis. *Eur J Endocrinol*, 2000; **142**: 365–72.
40. Stuart CA, Neelon FA, Lebovitz HE. Disordered control of thirst in hypothalamic-pituitary sarcoidosis. *N Engl J Med*, 2000; **303**: 1078–82.
41. Tikoo RK, Kupersmith MJ, Finlay JL. Treatment of refractory neurosarcoidosis with cladribine. *N Engl J Med*, 2004; **350**: 1798–9.
42. Hoffman GS, Kerr GS, Leavitt RY, *et al.* Wegener granulomatosis: an analysis of 158 patients. *Ann Intern Med*, 1992; **116**: 488–98.
43. Murphy JM, Gomez-Anson B, Gillard JH, *et al.* Wegener granulomatosis: MR imaging findings in brain and meninges. *Radiology*, 1999; **213**: 794–9.
44. Phan GQ, Yang JC, Sherry RM, *et al.* Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc Natl Acad Sci U S A*, 2003; **100**: 8372–7.
45. Faje AT, Sullivan R, Lawrence D, *et al.* Ipilimumab-induced hypophysitis: a detailed longitudinal analysis in a large cohort of patients with metastatic melanoma. *J Clin Endocrinol Metab*, 2014; **99**: 4078–85.
46. Albarel F, Gaudy C, Castinetti F, *et al.* Long-term follow-up of ipilimumab-induced hypophysitis, a common adverse event of the anti-CTLA-4 antibody in melanoma. *Eur J Endocrinol*, 2014; **172**: 195–204.
47. Barroso-Sousa R, Barry WT, Garrido-Castro AC, *et al.* Incidence of endocrine dysfunction following the use of different immune checkpoint inhibitor regimens: a systematic review and meta-analysis. *JAMA Oncol*, 2018; **4**: 173–82.
48. Byun DJ, Wolchok JD, Rosenberg LM, Girotra M. Cancer immunotherapy—immune checkpoint blockade and associated endocrinopathies. *Nat Rev Endocrinol*, 2017; **4**: 195–207.
49. Ryder M, Callahan M, Postow MA, Wolchok J, Fagin JA. Endocrine-related adverse events following ipilimumab in patients with advanced melanoma: a comprehensive retrospective review from a single institution. *Endocr Relat Cancer*, 2014; **21**: 371–81.
50. Caturegli P, Di Dalmazi G, Lombardi M, *et al.* Hypophysitis secondary to cytotoxic T-lymphocyte-associated protein 4 blockade: insights into pathogenesis from an autopsy series. *Am J Pathol*, 2016; **186**: 3225–35.
51. Corsello SM, Barnabei A, Marchetti P, De Vecchis L, Salvatori R, Torino F. Endocrine side effects induced by immune checkpoint inhibitors. *J Clin Endocrinol Metab*, 2013; **98**; 1361–75.
52. Voskens CJ, Goldinger SM, Loquai C, *et al.* The price of tumor control: an analysis of rare side effects of anti-CTLA-4 therapy in metastatic melanoma from the ipilimumab network. *PLoS One*, 2013; **8**: e53745.
53. Min L, Hodi FS, Giobbie-Hurder A, *et al.* Systemic high-dose corticosteroid treatment does not improve the outcome of ipilimumab-related hypophysitis: a retrospective cohort study. *Clin Cancer Res*, 2015; **21**: 749–55.
54. Barroso-Sousa R, Ott PA, Hodi FS, Kaiser UB, Tolaney SM, Min L. Endocrine dysfunction induced by immune checkpoint inhibitors: practical recommendations for diagnosis and clinical management. *Cancer*, 2018; **124**: 1111–21.
55. Kamisawa T, Funata N, Hayashi Y, *et al.* A new clinicopathological entity of IgG4-related autoimmune disease. *J Gastroenterol*, 2003; **38**: 982–4.
56. Shikuma J, Kan K, Ito R, *et al.* Critical review of IgG4-related hypophysitis. *Pituitary*, 2017; **20**: 282–91.
57. Stone JH, Khosroshani A, Deshpande V, *et al.* Recommendations of the nomenclature of IgG4-related disease and its individual organ system manifestations. *Arthritis Rheum*, 2012; **64**: 3061–7.
58. Yuen KCJ, Moloney KJ, Mercado JU, *et al.* A case series of atypical features of patients with biopsy-proven isolated IgG4-related hypophysitis and normal serum IgG4 levels. *Pituitary*, 2018; **21**: 238–46.
59. Takano K, Yamamoto M, Takahashi H, Himi T. Recent advances in knowledge regarding the head and neck manifestations of IgG4-related disease. *Auris Nasus Larynx*, 2017; **44**: 7–17.

60. Weindorf SC, Frederiksen JK. IgG4-related disease: a reminder for practicing pathologists. *Arch Pathol Lab Med*, 2017; **141**: 1476–83.
61. Iwata N, Iwama S, Sugimura Y, *et al.* Anti-pituitary antibodies against corticotrophs in IgG4-related hypophysitis. *Pituitary*, 2017; **20**: 301–10.
62. Bernreuther C, Illies C, Flitsch J, *et al.* IgG4-related hypophysitis is highly prevalent among cases of histologically confirmed hypophysitis. *Brain Pathol*, 2017; **27**: 839–45.
63. Sugihara H. Review on recent topics in hypophysitis. *J Nippon Med Sch*, 2017; **84**: 201–8.
64. Leporati P, Landek-Salgado MA, Lupi I, Chiovato L, Caturegli P. IgG4-related hypophysitis: a new addition to the hypophysitis spectrum. *J Clin Endocrinol Metab*, 2011; **96**: 1971–80.
65. Hansen I, Petrossians P, Thiry A, *et al.* Extensive inflammatory pseudotumor of the pituitary. *J Clin Endocrinol Metab*, 2001; **86**: 4603–10.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

2.5

Pineal Physiology and Pathophysiology, Including Pineal Tumours

Susan M. Webb, Anna Aulinas, Cristina Colom, and María-José Barahona

Pineal Physiology

The pineal gland is innervated mainly by sympathetic nerve fibres which inform the gland of the prevailing light-dark cycle and acts as a neuroendocrine transducer; it is located behind the third ventricle in the centre of the brain, is a highly vascular organ formed by neuroglial cells and parenchymal cells or pinealocytes; the latter synthesize melatonin, as well as other indoleamines and peptides.

The main pineal hormone melatonin (N-acetyl 5-methoxytryptamine) exhibits an endogenous circadian rhythm, reflecting signals originating in the suprachiasmatic nucleus; environmental lighting entrains the rhythm, by altering its timing. Independently of sleep, pineal melatonin is inhibited by light and stimulated during darkness, thanks to the neural input by a multisynaptic pathway which connects the retina, through the suprachiasmatic nucleus of the hypothalamus, preganglionic neurons in the upper thoracic spinal cord and postganglionic sympathetic fibres from the superior cervical ganglia, with the pineal gland.

Melatonin deficiency may produce sleeping disorders, behavioural problems, or be associated with precocious or delayed puberty in children, while chronically elevated melatonin has been observed in some cases of hypogonadotropic hypogonadism [1, 2].

Pineal Tumours

The main clinical problem related to the pineal gland is that of pineal tumours (Box 2.5.1). They are rare, ten times more common in children than in adults and mainly derive from the three types of cells: neuroglial, parenchymal, and germinal cells [3–5]. For the first time, the 2016 WHO classification of central nervous system (CNS) tumours uses molecular parameters in addition to histology to define tumour entities, formulating a concept for how CNS tumour diagnoses should be structured in the molecular era [6].

1. **Neuroglial cell tumours (20%)** may affect children and adults of any age. Low-grade, pilocytic astrocytomas usually present before the age of 20 years, with no sex predilection, while other types are more frequent in adults (intermediate diffuse and anaplastic astrocytomas or high-grade malignant glioblastomas; Box 2.5.1).
2. **Parenchymal tumours (15–30%)** secrete melatonin and differ in their degree of malignancy. The 2016 WHO classification grade them from I (pineocytomas, being mostly benign) to

Box 2.5.1 Classification of pineal tumours

Neuroglial cell tumours (20%)
- Low-grade astrocytomas (juvenile pilocytic)
- Intermediate diffuse and anaplastic astrocytomas
- High-grade malignant glioblastomas

Parenchymal tumours (15–30%) (See **Table 2.5.1**)

Germ cell tumours (80% in Japan, 30–50% in Western Europe and USA)
- Germinomas
  - Non-germinomatous germ cell tumours
  - Embryonal carcinoma
  - Yolk sac tumour
  - Choriocarcinoma
  - Teratomas
    - Benign teratomas
      - Immature
      - Mature
    - Teratoma with malignant transformation

Mixed germ cell tumours

Table 2.5.1 Parenchymal pineal tumour classification [3, 4, 7, 8]; the current WHO classification does not provide strict criteria to distinguish grade II and III tumours.

WHO grade	Histological type	Indicators of differentiation (from more to less)	Prognosis
I	Pineocytoma	No mitoses/very positive NF	Good
II III	Intermediate differentiation (20%)	Moderate nuclear atypia/low to moderate mitotic activity/<2 mitoses per HPFs/positive NF protein staining /MIB-1 proliferation indices 3–10% >6 mitoses per HPF/necrosis/negative NF protein staining	↓
II or III	Papillary tumour of the pineal region		
IV	Pineoblastoma	Variable plus positive or negative NF	Bad

WHO, World Health Organization; NF, neurofilaments; HPF, high power field.

IV (pineoblastomas, being highly malignant) (Table 2.5.1). However, most parenchymal tumours are either mixed or show intermediate differentiation (WHO grades II and III, Table 2.5.1) [7]. Histologically, **pineocytomas** characteristically present pineocytomatous rosettes, while pineoblastomas are populated by small, highly undifferentiated cells, often present with haemorrhagic or necrotic components, but rarely calcifications. The major prognostic factor is the extent of surgery. **Pineoblastomas** are highly malignant, aggressive and of rapid growth (similar to other primitive neuroectodermal tumours like neuroblastomas or medulloblastomas). **Papillary tumours** of the pineal region (PTPR) were included as a new entity in the 2007 WHO classification. They are very rare neuroepithelial tumours, macroscopically indistinguishable from pineocytomas, combining papillary and solid areas; however, microscopically, these tumours are easily distinguished. Recently it has been shown that PTPR have ependymal differentiation and are phenotypically more similar to the circumventricular subcommissural organ of the posterior third ventricle than to the pineal gland [8]. The biological and clinical behaviour of these tumours is variable and may correspond to WHO grades II or III, but definite histological grading criteria remain to be defined. Intermediate grades II and III represent different degrees of differentiation and prognosis (Table 2.5.1).

3. **Germ cell tumours (30–50%)**, histologically and biologically homologous to gonadal germ cell neoplasms, will characteristically present positive markers for alpha fetoprotein (AFP) and beta human chorionic gonadotrophin (β-hCG), with more (**teratomas**) or less differentiation (**germinomas**), as well as intermediate degrees (yolk sac tumours).

Very rarely, pineal region tumours may derive from meningothelial, mesenchymal, ependymal, choroid plexus elements and peripheral nerves giving rise to **gangliogliomas, melanocytic neoplasms, atypical teratoid/rhabdoid tumours, meningiomas, cavernous angiomas, hemangiopericytomas** or **neurinomas/neurofibromas**, apart from **lymphomas** or **metastases**.

Clinical Presentation

Clinical presentation of pineal tumours depends on age at onset and histology [9, 10]. Over 90% present with raised intracranial pressure, often with obstructive hydrocephalus; initial symptoms are frequently headache, nausea, vomiting and decreased vision; 50–70% of patients refer visual signs like diplopia, cranial nerve palsies, papilledema and ptosis or Parinaud's syndrome (failure of upward gaze, pupillary dilatation and diminution of pupillary light reflex) due to pressure on the pretectal region. Compression on the brain, cerebellum, hypothalamus, and pituitary may cause paralysis of other cranial nerves, ataxia, diabetes insipidus, and hypopituitarism. Pineal tumours may interfere with puberty, due to either pressure of the tumour on the hypothalamic centres which govern gonadotrophin secretion, excessive melatonin secretion by pinealocyte tumours causing delayed puberty in adolescents, or reduction of the potential antigonadotrophic effect of melatonin, which, together with β-hCG secretion by destructive germ cell tumours could explain precocious puberty in prepubertal children.

The 2016 WHO classification of **parenchymal tumours** (that characteristically secrete melatonin) [11–16] grade them from I to IV (Table 2.5.1) [3, 4, 7]. **Pineocytomas** (grade I) present more often in adults with a mean age of 43 years, with a male to female ratio of 0.6/1, evolve slowly locally (interval between onset of symptoms and surgery may be of several years), do not invade contiguous tissue or seed the cerebrospinal fluid (CSF). Non-specific presenting manifestations reflect compression of neighbouring structures (tectal plate, aqueduct of Sylvius, cerebellum, brainstem, hypothalamus, pituitary) like increased intracranial pressure, changes in mental status, neuroophthalmologic, brain stem, and/or cerebellum dysfunction, hypopituitarism, and hyperprolactinemia. Rarely intratumoural haemorrhage (pineal apoplexy) with subarachnoid extravasation may occur. Concurrent uveoretinitis in occasional patients with pineocytomas probably reflect the common photoreceptor activity of pineal and retinal cells. No metastases and a 5-year survival > 90% have been reported. The major prognostic factor is the extent of surgery.

Pineoblastomas (grade IV) typically appear before the age of 20 years, most often in young children, but there are reports in adults, with a slight male preponderance. Presenting symptoms of this least differentiated and most aggressive pineal parenchymal tumour are more rapidly progressive and of shorter duration (interval between initial symptoms and surgery may be less than a month). Median postsurgical survival varies from 24 to 30 months. They are locally invasive and prone to disseminate through the CSF, often fatal, but may be controlled in some cases by a multimodality combination of aggressive surgery, radiotherapy, and chemotherapy. They may occur in familial bilateral retinoblastoma (due to a germline retinoblastoma gene mutation) known as a trilateral

retinoblastoma syndrome, with a median survival of only 6 months; or in patients with familial adenomatous polyposis. They may also present in patients with *DICER1* germline mutations [7, 17]. Recent studies suggest that in addition to the tumour predisposition syndrome DICER1, the appearance of sporadic pinealoblastomas due to a somatic mutation leading to the inactivation of DICER1 and dysfunction/dysregulation of one of its precursor RNA strands (miRNA) is also possible [18]. Negative prognostic predictors are disseminated disease at diagnosis, the patient's young age and partial surgical resection.

Intermediate grades II and III represent different degrees of differentiation and prognosis (Table 2.5.1).

Papillary tumours present with a variable clinical behaviour, median age at diagnosis of 35 years, but may affect children and adults, with no sex predilection [7, 12–14, 19–22]. Local recurrences are common, especially if resection is incomplete, but spinal dissemination is rare. Gross total resection and younger age are associated with overall survival, while radiotherapy and chemotherapy have no significant impact. Higher Ki-67 proliferation index (>10%) or more than 3 mitoses per 10 high power fields have shown to be related with a shorter progression-free survival [7].

Germ cell tumours arise around the third ventricle, most commonly in the pineal region, but may also be seen in the suprasellar, hypothalamic compartment; 5–10% of patients harbour lesions in both locations. Ninety per cent (90%) appear under the age of 20 years and are more frequent in boys than girls (2.5:1), except suprasellar lesions more common in girls. An increased risk of intracranial germ cell tumours has been associated with Klinefelter's syndrome, Down's syndrome, and neurofibromatosis type 1 [23, 24].

Other tumours, like pineal meningiomas, gangliogliomas, ependymomas, lipomas, and pineal metastases, most frequently of breast or lung origin may occur, often with other brain metastases; symptoms and signs reflect the extent of the disease [20, 25].

Diagnosis

An appropriate tissue specimen for accurate histological diagnosis and determining tumour type is critical to optimize subsequent management. Serum alpha fetoprotein (AFP; synthesized mainly by yolk sac tumours, and teratomas) and β-hCG (in choriocarcinomas or germinomas) concentrations are of diagnostic utility if markedly elevated in serum and/or CSF. Measurement of these markers in CSF for initial staging, and if positive, for follow-up are useful. CSF cytological examination should be delayed at least 2 weeks after surgery to increase the chance of reflecting true dissemination of viable tumour rather than postoperative tumour spillage. If these markers are clearly raised, histological verification may not be required.

Biopsies may be obtained by classical surgical routes (posterior interhemispheric transcallosal, suboccipital transtentorial and infratentorial-supracerebellar routes) or by microsurgical techniques, with significantly reduced perioperative mortality rates (<2%); a neuroendoscopic or stereotactic biopsy are reasonably safe and well tolerated [9, 26], in experimented hands, but the diagnosis of mixed or intermediate tumours may be difficult without extensive tissue sampling, and especially may miss detecting a possible teratoma component. In any case, operative risk should be balanced with the risk of not obtaining an accurate histological diagnosis, with prognostic implications. In cases of non-diagnostic or equivocal biopsies or indicative of a benign tumour (mature teratoma, meningioma), surgery is recommended.

Imaging

MRI is the gold standard imaging technique for evaluating pineal tumours and detecting intradural extramedullary metastases [27]. An MRI will disclose the size and extension of the tumour and possible metastases, but cannot accurately identify the histological nature, which relies on biopsy or serum/CSF tumour markers. In the more malignant tumours (pineoblastomas, germinomas, teratomas) the spine as well as the brain should be imaged, since spread into the subarachnoid space and the spine are frequent.

Neuroimaging of **astrocytic tumours** can vary; MRI usually shows hypodensity on T_1-weighted images and hyperintensity on T_2-weighted images; gadolinium enhancement is uncommon, except if active tumour progression occurs.

Among **parenchymal tumours**, pineocytomas appear as non-invasive, solid masses in the posterior third ventricular region, and tend to be smaller (<3 cm in general), rounder, hypodense, homogeneous masses with dispersed calcifications, particularly peripheral, which enhance heterogeneously or diffusely on computed tomography (CT) and MRI, and present a lesser degree of hydrocephalus. Macrocystic presentation is rare but small cysts may be present. T_1-weighted images are hypointense while T_2 are hyperintense. Haemorrhage and necrosis are exceptional.

Pineoblastomas are larger, lobulated, homogeneous tumours, rarely calcified and present with a greater degree of hydrocephalus and local invasion of contiguous brain or leptomeninges; they may exhibit distant subarachnoid and extracranial metastases, more frequently in young females; they are hyperdense and enhance homogeneously on CT (Figure 2.5.1), while on MRI they appear as hypo-to isointense on T_1-weighted images and enhance diffuse- or heterogeneously with contrast (Figures 2.5.2, 2.5.3, and 2.5.4). Haemorrhage and necrosis are common (Figure 2.5.5).

Germ cell tumours (except teratomas) appear as solid masses on MRI, iso- or hyperdense, which enhance after contrast (Figures 2.5.6 and 2.5.7); small nodular calcifications may be seen on CT scans. Teratomas tend to contain intratumoural cysts next to calcifications and low attenuation signals, typical of fat. Haemorrhages are common in choriocarcinomas and mixed neoplasms.

Treatment

Surgery, chemotherapy, and radiation are used in the treatment of pineal region tumours. Surgery, either open, stereotactic, or endoscopic, is used to obtain a biopsy, mandatory in the majority of cases to obtain a definite histological diagnosis [9]. Morbidity and cure rates have improved over the last years thanks to a greater understanding of the nature of the different tumours, more accurate neurosurgical experience, selective use of chemotherapy, and the introduction of modern irradiation techniques. However, the rarity of pineal tumours, make the obtaining of large prospective multicentre international studies to define their optimal management, difficult.

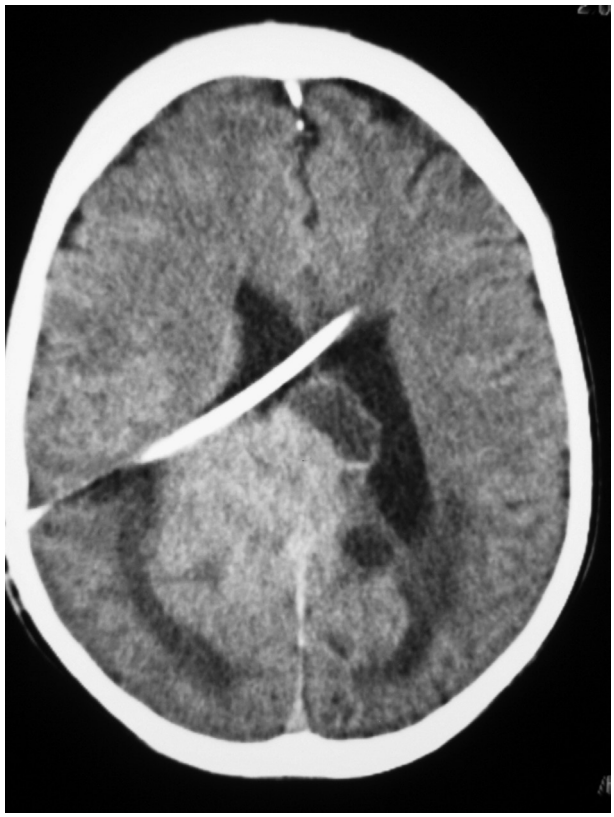


Figure 2.5.1 Contrast-enhanced CT scan of a recurrent pineoblastoma in a 16-year-old boy with ventricular shunt.
Courtesy of Dr. E Guardia.

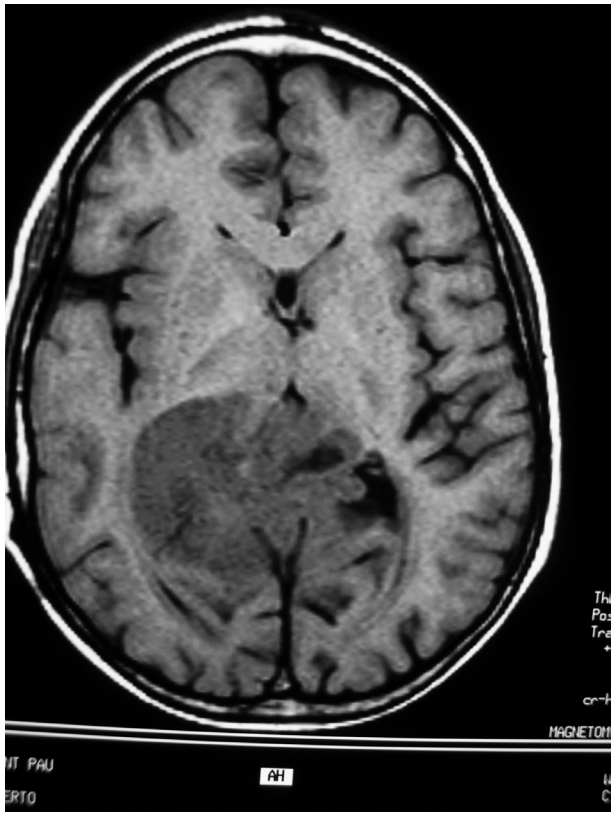


Figure 2.5.2 T_1-weighted MRI of a recurrent pineoblastoma in a 16-year-old boy without contrast.

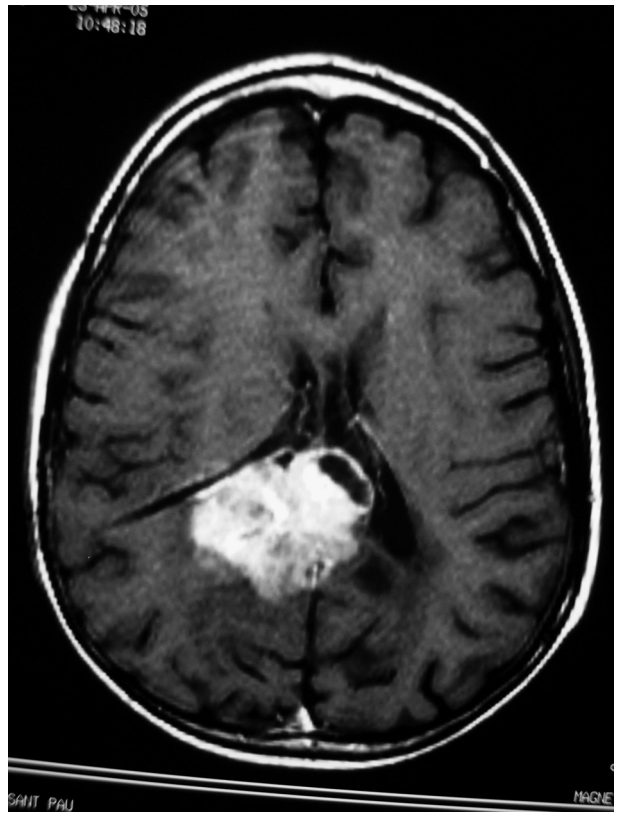


Figure 2.5.3 T_1-weighted gadolinium-enhanced MRI of a recurrent pineoblastoma showing the ventricular shunt.

Treatment depends on histology obtained after surgery, which apart from the biopsy can resolve intracranial hypertension with a ventricular shunt (atrial or peritoneal) and perform partial debulking of the tumour if possible; total resection is rarely possible (Table 2.5.2).

Astrocytomas: Treatment for astroglial derived malignant gliomas is local radiotherapy to the tumour (54 Gy), either conventional or stereotaxic, while surgery may be curative for the more benign pilocytic astrocytomas.

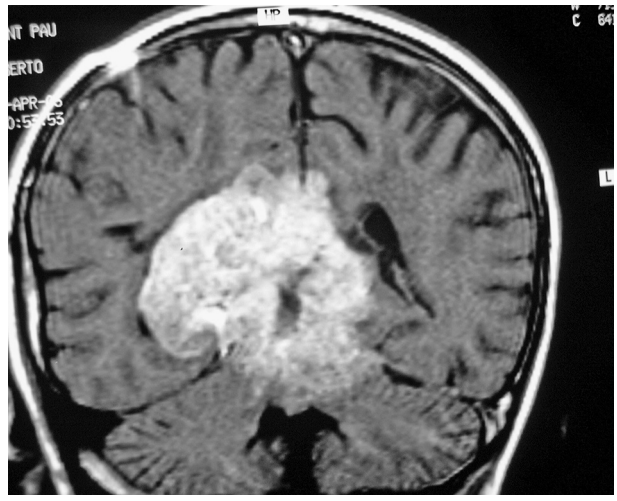


Figure 2.5.4 Coronal T_1-weighted gadolinium-enhanced MRI of a recurrent pineoblastoma.

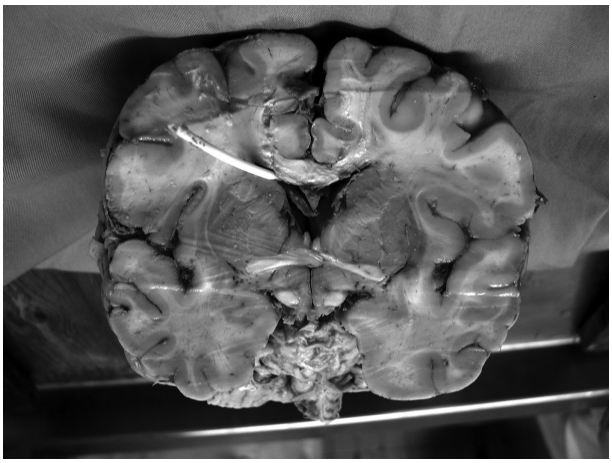


Figure 2.5.5 Coronal slice of the brain corresponding to Figure 2.5.1, showing the pineoblastoma and ventricular shunt.
Courtesy of Dr. E Guardia.

Pineal parenchymal tumours: Pineocytomas only require local radiotherapy to the tumour (54 Gy). In pinealoblastomas, a high probability of spinal seedlings should lead to craniospinal radiotherapy, since they are radiosensitive (25–30 Gy on the neuroaxis with a pineal boost of 40 Gy aimed at more effective local disease control). However, routine craniospinal irradiation has been questioned and may not be necessary in patients with negative staging. Stereotaxic radiosurgery may control local progression and minimize

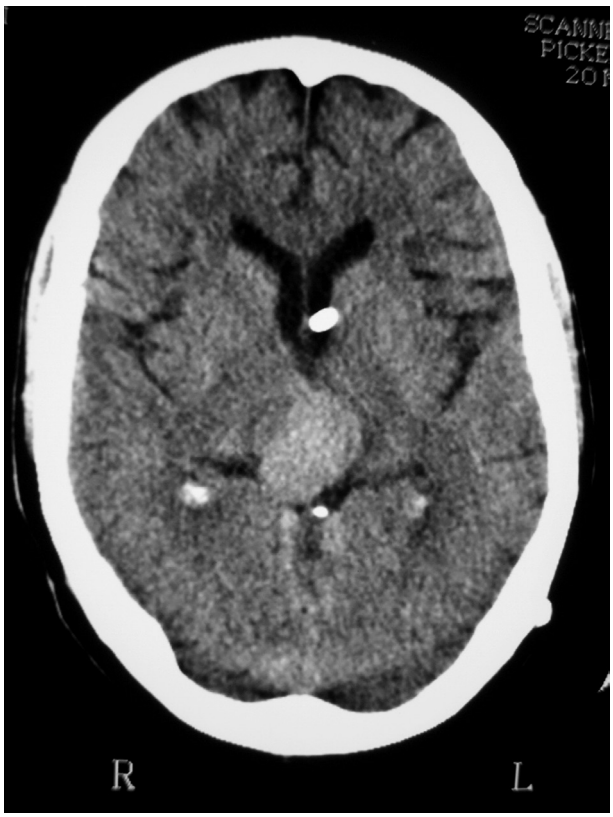


Figure 2.5.6 CT scan of a recurrent pineal germinoma in a 57-year-old man, with a ventriculoperitoneal shunt.
Courtesy of Dr. E Guardia.

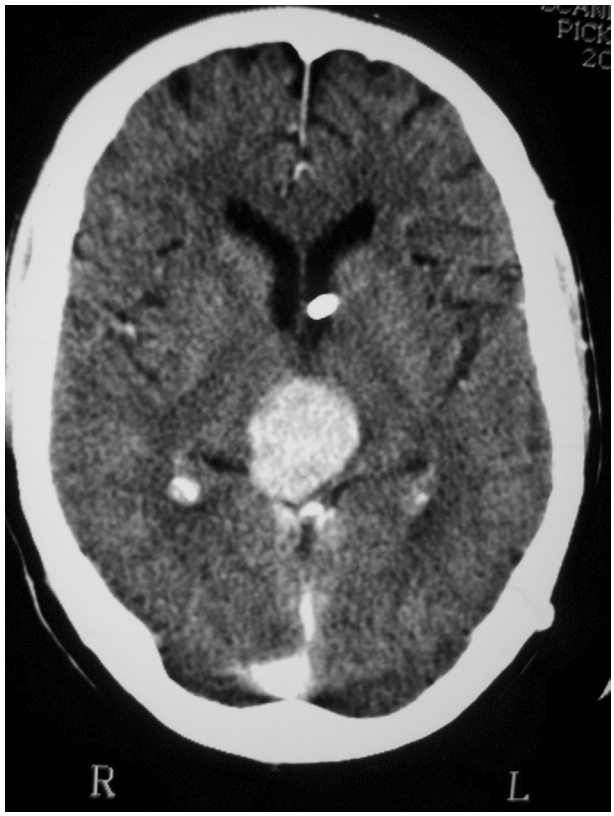


Figure 2.5.7 Contrast-enhanced CT scan of a recurrent pineal germinoma in a 57-year-old man.

damage to the surrounding brain, especially important in prepubertal patients, in whom total brain irradiation is associated with neurocognitive dysfunction, endocrinopathy, second malignancies, vascular complications, and spinal growth impairment. However, it may be associated with a high risk of marginal recurrence and distant metastases, and is not considered the treatment of choice for infiltrative but curable tumours. Furthermore, complications such as ataxic gait and gaze palsy have been reported after radiosurgery. In young children chemotherapy with cisplatinum, etoposide, cyclophosphamide, and vincristine, which alone is not curative, may allow a lower dose of radiotherapy to have similar effects. In older children with pineoblastoma, craniospinal irradiation is followed by chemotherapy (even though its role on final outcome is not fully defined). Age at diagnosis ≥4 years and administration of radiotherapy as initial therapy are considered favourable prognostic factors for progression-free survival [28]. Autologous haematopoietic stem-cell-supported high-dose chemotherapy has been proposed with some initial promising results, although experience is limited [29].

In mixed or intermediate pineal parenchymal cells (Grade II or III, Table 2.5.1), apart from local radiotherapy, craniospinal irradiation and chemotherapy should be considered, when increasing number of mitoses and less differentiation are observed [9, 11–16].

In PTPR, the optimal treatment remains controversial, as no definitive treatment strategy exists for this lesion. It has been described that a minimally invasive strategy (radiotherapy or stereotactic radiosurgery) resulted in a favourable response to treatment, avoiding the risks of aggressive surgical removal. However, incomplete resection tended to be associated with decreased survival and

Table 2.5.2 Treatment guidelines for pineal tumours. Surgery for histologic biopsy and subtyping and if necessary CSF diversion (ventriculoperitoneal shunt or ventriculostomy) should always be performed, with the possible exception of germ cell tumours with diagnostically elevated tumour markers

Tumour type	Radiotherapy	Chemotherapy[1]	Surgery
Glial origin			
Juvenile pilocytic astrocytoma	No	No	Complete resection
Intermediate/diffuse/anaplastic/ Astrocytomas/glioma	Local	No	Debulking
Malignant glioblastoma	Local	No	Debulking
Parenchymal tumours			
Pineocytoma	Local	No	Biopsy
Intermediate or mixed tumour	Local ± craniospinal	Yes, in more undifferentiated tumours	Biopsy
Pineoblastoma	Local Routine craniospinal not always indicated. Age <5 yrs: Lower dose, after initial chemotherapy	Yes (role on final outcome unclear)	Biopsy
Papillary tumours	Local	No	Complete resection
Germ cell tumours			
Germinoma	Local + craniospinal (unless convinced of negative staging)	Yes (alone not curative)	Biopsy
Non-germinatous tumours	Local + craniospinal	Yes (pre- or post-surgery)	Resection as much as possible, without ↑ morbidity

[1] Chemotherapy includes cisplatin, etoposide, and cyclophosphamide or ifosfamide.

with recurrence. In an updated retrospective series of 44 patients, only gross total resection and younger patient age were associated with overall survival; radiotherapy and chemotherapy had no significant impact [7].

Germ cell tumours. Surgery is not considered curative in germinomas, which are radiosensitive and should therefore receive local radiotherapy [9, 23, 24]. Unless firmly confident of negative staging (by negative tumour markers—AFP and β-hCG—in blood and CSF, and negative MRI), craniospinal radiotherapy should be offered given the high probability of spinal seedlings. Germinomas are also highly chemosensitive, and excellent responses to postoperative cisplatin and cyclophosphamide have been reported. Survival is high (>90% at 5 years) in patients with localized pure germinomas, using either chemotherapy or focal radiotherapy or craniospinal irradiation, while focal irradiation alone has a worse outcome. In metastatic germinomas, craniospinal irradiation is the treatment of choice (25–35 Gy to the spine and a local pineal boost of 40 Gy). Lower irradiation doses are currently being considered, especially if adjuvant chemotherapy is offered [23]. Bifocal lesions in the pineal and hypothalamus should be considered localized germinomas rather than metastatic disease, and receive irradiation to both locations.

Other germ cell tumours are less radiosensitive than germinomas, with a poor survival after radiotherapy alone (median survival of under 2 years) and require multimodality treatment [23, 24]. Surgical resection after tumour reduction with initial chemotherapy with cisplatin, etoposide, and ifosfamide is an alternative. Tumour markers are useful for follow-up. Combining chemotherapy with radiotherapy (local up to 54 Gy or craniospinal up to 36 Gy) may increase long-term survival to 80%.

Surgery is the treatment of choice of **pineal meningiomas** and **other localized pineal** tumours if possible; alternatively, localized stereotactic radiosurgery may be offered with good long-term prognosis [9].

Pineal cysts. Masses in the pineal region are most commonly non-neoplastic cysts incidentally discovered at autopsy or on a radiographic work-up for symptoms not reasonably attributed to the cyst. Very rarely they act as a mass lesion and produce signs of increased intracranial pressure, by compressing the aqueduct (obstructive hydrocephalus) or tectal plate (Parinaud's syndrome). On MRI they appear as a 1–3 cm mass, equally or slightly more dense than CSF in T_1-weighted image studies and which brightly enhance in T_2-weighted images, reflecting their fluid nature; evidence of haemorrhage and peripheral calcification may be found. If asymptomatic, pineal cysts do not generally require treatment; if large enough to increase intracranial pressure, resection may be necessary, with an excellent long-term outcome [9, 30]. Patients with growing lesions, contrast enhancement, and haemorrhage are more likely to develop hydrocephalus and should be followed-up with serial MRI. However, no specific management guidelines exist regarding follow-up of these lesions and whether patients need to be monitored over time [31].

REFERENCES

1. Webb SM, Puig-Domingo M. Role of melatonin in health and disease. *Clin Endocrinol (Oxf)*, 1995; **42**: 221–34.
2. Macchi MM, Bruce JN. Human pineal physiology and functional significance of melatonin. *Front Neuroendocrinol*, 2004; **25**: 177–95.

3. Louis DN, Ohgaki H, Wiestler OD, *et al*. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol (Berl)*, 2007; **114**: 97–109.
4. Brat DJ, Parisi JE, Kleinschmidt-DeMasters BK, *et al*. Surgical neuropathology update: a review of changes introduced by the WHO classification of tumours of the central nervous system, 4th edition. *Arch Pathol Lab Med*, 2008; **132**: 993–1007.
5. Osborn AG, Salzman KL, Thurnher MM, Rees JH, Castillo M. The new World Health Organization Classification of Central Nervous System Tumors: what can the neuroradiologist really say? *AJNR Am J Neuroradiol*, 2012; **33**: 795–802.
6. Louis DN, Perry A, Reifenberger G, *et al*. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol (Berl)*, 2016; **131**: 803–20.
7. Nakazato Y, Jouvet A, Vasiljevic A. Tumours of the pineal gland (Chapter 7). In Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, eds. *WHO Classification of Tumours of the Central Nervous System*, 4th edn. Lyon: International Agency for Research on Cancer (IARC), 2016: 3–56.
8. Coy S, Dubuc AM, Dahiya S, Ligon KL, Vasiljevic A, Santagata S. Nuclear CRX and FOXJ1 expression differentiates non-germ cell pineal region tumors and supports the ependymal differentiation of papillary tumor of the pineal region. *Am J Surg Pathol*, 2017; **41**: 1410–21.
9. Moschovi M, Chrousos GP. Pineal gland masses. 2017. UpToDate Version 23.0. Available at: https://www.uptodate.com/contents/pineal-gland-masses (accessed 5 February 2018).
10. Gaillard F, Jones J. Masses of the pineal region: clinical presentation and radiographic features. *Postgrad Med J*, 2010; **86**: 597–607.
11. Chang SM, Lillis-Hearne PK, Larson DA, Wara WM, Bollen AW, Prados MD. Pineoblastoma in adults. *Neurosurgery*, 1995; **37**: 383–90; discussion 390–1.
12. Fauchon F, Jouvet A, Paquis P, *et al*. Parenchymal pineal tumors: a clinicopathological study of 76 cases. *Int J Radiat Oncol Biol Phys*, 2000; **46**: 959–68.
13. Jouvet A, Saint-Pierre G, Fauchon F, *et al*. Pineal parenchymal tumors: a correlation of histological features with prognosis in 66 cases. *Brain Pathol Zurich Switz*, 2000; **10**: 49–60.
14. Schild SE, Scheithauer BW, Schomberg PJ, *et al*. Pineal parenchymal tumors. Clinical, pathologic, and therapeutic aspects. *Cancer*, 1993; **72**: 870–80.
15. Cohen BH, Zeltzer PM, Boyett JM, *et al*. Prognostic factors and treatment results for supratentorial primitive neuroectodermal tumors in children using radiation and chemotherapy: a Childrens Cancer Group randomized trial. *J Clin Oncol*, 1995; **13**: 1687–96.
16. Jakacki RI, Zeltzer PM, Boyett JM, *et al*. Survival and prognostic factors following radiation and/or chemotherapy for primitive neuroectodermal tumors of the pineal region in infants and children: a report of the Childrens Cancer Group. *J Clin Oncol*, 1995; **13**: 1377–83.
17. van Engelen K, Villani A, Wasserman JD, *et al*. DICER1 syndrome: approach to testing and management at a large pediatric tertiary care center. *Pediatr Blood Cancer*, 2018; **65**: PMID: 28960912.
18. Raleigh DR, Solomon DA, Lloyd SA, *et al*. Histopathologic review of pineal parenchymal tumors identifies novel morphologic subtypes and prognostic factors for outcome. *Neuro-Oncol*, 2017; **19**: 78–88.
19. Fèvre-Montange M, Vasiljevic A, Champier J, Jouvet A. Histopathology of tumors of the pineal region. *Future Oncol*, 2010; **6**: 791–809.
20. Mena H, Nakazato Y, Scheithauer BW. Pineal parenchymal tumors. In: Kleihues P, Cavenee WK, eds. *Pathology and Genetics: Tumors of the Nervous System*. Lyon: International Agency for Research on Cancer (IARC); 1997: 115–21.
21. Patel SK, Tomei KL, Christiano LD, Baisre A, Liu JK. Complete regression of papillary tumor of the pineal region after radiation therapy: case report and review of the literature. *J Neurooncol*, 2012; **107**: 427–34.
22. Riis P, van Eck ATCJ, Dunker H, Bergmann M, Börm W. Stereotactic radiosurgery of a papillary tumor of the pineal region: case report and review of the literature. *Stereotact Funct Neurosurg*, 2013; **91**: 186–9.
23. Echevarría ME, Fangusaro J, Goldman S. Pediatric central nervous system germ cell tumors: a review. *The Oncologist*, 2008; **13**: 690–9.
24. Matsutani M, Sano K, Takakura K, *et al*. Primary intracranial germ cell tumors: a clinical analysis of 153 histologically verified cases. *J Neurosurg*, 1997; **86**: 446–55.
25. Bailey S, Skinner R, Lucraft HH, Perry RH, Todd N, Pearson AD. Pineal tumours in the north of England 1968–93. *Arch Dis Child*, 1996; **75**: 181–5.
26. Oi S, Shibata M, Tominaga J, *et al*. Efficacy of neuroendoscopic procedures in minimally invasive preferential management of pineal region tumors: a prospective study. *J Neurosurg*, 2000; **93**: 245–53.
27. Solomou AG. Magnetic resonance imaging of pineal tumors and drop metastases: a review approach. *Rare Tumors*, 2017; **9**: 6715.
28. Mynarek M, Pizer B, Dufour C, *et al*. Evaluation of age-dependent treatment strategies for children and young adults with pineoblastoma: analysis of pooled European Society for Paediatric Oncology (SIOP-E) and US Head Start data. *Neuro-Oncol*, 2017; **19**: 576–85.
29. Gururangan S, McLaughlin C, Quinn J, *et al*. High-dose chemotherapy with autologous stem-cell rescue in children and adults with newly diagnosed pineoblastomas. *J Clin Oncol*, 2003; **21**: 2187–91.
30. Májovský M, Netuka D, Beneš V. Conservative and surgical treatment of patients with pineal cysts: prospective case series of 110 patients. *World Neurosurg*, 2017; **105**: 199–205.
31. Starke RM, Cappuzzo JM, Erickson NJ, Sherman JH. Pineal cysts and other pineal region malignancies: determining factors predictive of hydrocephalus and malignancy. *J Neurosurg*, 2017; **127**: 249–54.

SECTION 3

Thyroid Disease

3.1

Evaluation of the Thyroid Patient

3.1.1 The History and Iconography Relating to the Thyroid Gland

Robert Volpé† *and Clark Sawin*†

This chapter is reproduced from the previous edition.

Introduction

This chapter is a brief summary of the history and art related to the thyroid gland. The reader is referred to other sources for an exhaustive exposition of these matters [1–3].

Early Years

Knowledge of goitre (which was not known to be a thyroid enlargement until about the sixteenth century) goes back into antiquity. In 1600 BC, burnt sponge and seaweed was used for the treatment of goitre in China [2]. In the fourth century BC, the Ayur Veda, a Hindu system of medicine in India, contained a discussion of goitre [1, 2]. In Greece, in the days of Hippocrates, goitre was regarded purely as a deformity, and was attributed to the drinking of snow water [2]. In ancient Greece, swellings in the region of the thyroid gland (and presumably swellings elsewhere in the neck) were referred to as 'bronchocoele' or 'struma'. Galen (AD 130–200), considered the greatest medical practitioner in antiquity after Hippocrates, regarded the thyroid as a lubricant for the larynx [2]. Later, Julius Caesar [2] noted that Gauls had large necks as one of their characteristics. Celsus (25 BC to AD 50) in Rome defined bronchocoele (a tumour in the neck, most likely goitre) and he also described cystic goitre, as well as surgery for these lesions [2]. At the same time, in Egypt, Egyptian coins showed the presence of goitres, [4] and an Egyptian relief of Cleopatra likewise depicted her with what appears to be a goitre.

Even in these early years several writers [1, 2] referred to epidemics of goitre in the Alps, which was a forerunner to a wide literature from this region regarding goitres and their relationship to Alpine culture. The Chinese also were well aware of goitre in those early years, and recommended seaweed for the treatment of goitre as early as AD 340. Much later, the treatment of goitre with desiccated thyroid was advocated as early as AD 1475 by Wang Hei in China [2].

In Switzerland in the sixteenth century, Paracelsus [5] (1493–1541) recognized the connection between cretinism, endemic goitre, and congenital idiocy. He attributed goitre to mineral impurities in the drinking water. Later in that century and the next century, also in Switzerland, many writers described cretins in Swiss cantons and related them to the presence of goitre [3]. In 1656 [6] at St Thomas's Hospital, London, Thomas Wharton named the lobes of the thyroid, 'glandulae thyroidiaeae' because of their anatomical proximity to the thyroid cartilage, and not because of their shape. He felt that the fact that women generally had larger thyroid glands than men was for the purpose of making their necks 'more even and beautiful'.

During the mid and late medieval period in Europe, goitre and cretinism played a significant part in the social history of middle Europe, particularly in Alpine areas where these conditions were quite prevalent. Indeed, there was a connotation that those people with goitres were somehow inherently evil, and this was reflected in the folk art of the period. Indeed, depictions of goitre and cretinism in art and sculpture at that time were commonplace, and form the subject of an entire volume [3].

The discovery of iodine at the beginning of the nineteenth century was a landmark event in relation to the thyroid gland. In 1811, Bernard Courtois (1777–1838), a self-taught chemist and dealer in chemicals and manufacturer of saltpetre in Paris, was using vitriol to clean the vats used for making potash from seaweed, when he noted violet fumes. This violet gas condensed into crystals on cooling; he called the crystals substance X. The substance was soon identified as a new halogen element by Sir Humphrey Davy

† It is with regret that we report the deaths of Dr Robert Volpé and Clark Sawin.

(1778–1829) [7], who happened to be in Paris at the time (despite the ongoing Napoleonic War). As just mentioned, seaweed or burnt sponge had been employed in the treatment of goitre for centuries [2]. However, the credit for using iodine itself in the treatment of goitre appears to go to Coindet (1774–1834) [8] of Geneva in 1820, after he had determined that the substance in burnt sponge that acted against 'bronchocoele' was actually iodine. In his 'tincture of iodine', he used 48 grains of iodine to one ounce of spirit of wine. For adults, he prescribed 10 drops of the tincture in half a tumbler of 'syrup of capillaire' and water three times a day, the dose being increased after a few weeks to 15 or even 20 drops. He noted that bronchocoele would usually subside and be destroyed within the space of 6–10 weeks. A few years later, Lugol (1786–1851) in 1829 [2] also recommended and used (what we now call) Lugol's solution for the treatment of goitre, with considerable success.

With the spread of iodine usage, toxic effects soon appeared, as described by Coindet in 1821, and later by Frederic Rilliet (1814–61) [9]. These ill effects of iodine led to a great deal of anxiety about its use for goitre under any circumstance; for some, it was completely proscribed.

Surgical treatment of goitre had been mentioned by Celsus (25 BC to AD 50) [2]. It was also mentioned in the Turkish manuscript of Charaf Eddin in 1465. Johann A.W. Hedenus (1760–1836) [2] reported in 1822 on six cases of successful excision of a goitre for impending suffocation. Joseph Henry Green (1791–1863) reported the removal of the right lobe of the thyroid gland in St Thomas's Hospital, London, in 1829, but the patient died of sepsis 2 weeks later.

Structure and Function

In a paper by T.W. King (1809–47) [10] of Guy's Hospital, London, there is a description of what was thought to be the secretion of the thyroid as passing into its lymphatics and so into the great veins. King noted that this had been indirectly surmised by Morgagni. He also remarked prophetically that we should be able one day to show that a particular material is slowly formed and partially kept in reserve and that this principle is also supplementary, when poured into the descending inferior vena cava, to important functions in the course of the circulation. Thus, he had a conception of the internal secretion by the thyroid. In notes appended to King's paper, Astley Cooper (1768–1841), a surgeon at the same hospital, agreed with King about this idea.

John Simon (1818–97) [1] in 1844, while assistant surgeon at King's College Hospital and demonstrator of anatomy at King's College, London, published a paper on the comparative anatomy of the thyroid. He stated that in addition to its copious vascular supply, it had the structure of a secreting gland. Several decades later, based on thyroidectomies performed on monkeys and other mammals, Victor Horsley (1857–1916) [11] in 1885 supported the generalization of Felix Semon (1849–1921) in 1883 that myxoedema, cretinism, and operative cachexia strumipriva were all due to thyroid deficiency, and not due to chronic asphyxia (as Theodor Kocher, a surgeon in Bern, Switzerland, originally believed) or due to injury of the sympathetic or other nervous structures. Horsley thought that the thyroid controlled the metabolism of mucus, and that the effects of thyroid insufficiency were due to an accumulation of mucus. Others thought that the function of the thyroid was to neutralize or remove poisons, and so thyroid insufficiency was presumed to produce toxaemia. Horsley, in his report to the Clinical Society's Committee on Myxoedema (1888) divided the effects of complete thyroidectomy on monkeys into (1) the acute effects, which consisted of nervousness, tremor, clonic spasm, contracture, paresis, paralysis, and which came on between the second and the twelfth day (now in retrospect clearly the result of damage to the parathyroid glands); and (2) chronic experimental myxoedema. After Gley's [12] rediscovery in 1891 of the parathyroid glands, it became evident that the effect of complete removal of the thyroid resulted in myxoedema, and that removal of the parathyroids was responsible for what had previously been called tetania thyreopriva.

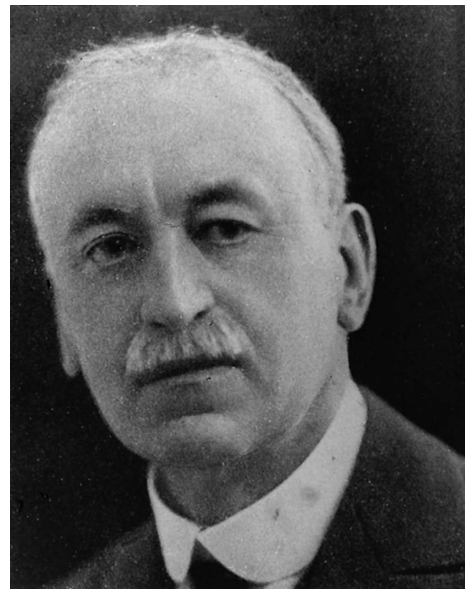


Figure 3.1.1.1 Professor George R. Murray (1865–1939), Newcastle-upon-Tyne, England, who first used extracts of sheep thyroid for the treatment of myxoedema in 1891.

Reproduced with permission from Rolleston HD. *The Endocrine Organs in Health and Disease, with a Historical Review*. London: Oxford University Press, 1936.

The dramatic results of replacement treatment of myxoedema by thyroid preparations by George Murray (1865–1939) [13] (**Figure 3.1.1.1**) in the United Kingdom in 1891, and Magnus-Levy's [14] demonstration in France in 1895 that thyroid medication accelerated metabolism, led to the conclusion that the thyroid and its internal secretion had definite powers other than detoxification.

Theodor Kocher (1841–1917) [2] suggested in 1895 that the thyroid might contain iodine. In the same year, Tschirch [2] was unable to establish this point. However, Eugen Baumann (1846–96) [15] apparently quite independently of Kocher's suggestion, investigated the chemistry of the thyroid, and was much surprised to find iodine. He published his findings in 1896 [15] called the extracted compound that contained iodine 'thyreo-iodin', and considered it to be the active principle of the thyroid. This led to the isolation by Kendall [16] in 1914 of an active principle which was initially called thyroxyindole and later thyroxin. In 1926, Harington (1897–1972) [17] proved that it was derived from tyrosine and not, as Kendall thought, from tryptophane, and was found to be a basic substance, now called thyroxine. Thyroxine was subsequently shown by Harington [2] and Salter [2] to be less powerful than desiccated thyroid. It was of interest that Harington was not able to accept the

possibility that a principle other than thyroxine might exist to explain the metabolic effects of desiccated thyroid. It was not until 1952 that Gross and Pitt-Rivers [18] discovered triiodothyronine, which proved to be that elusive second principle.

Cretinism

The term 'cretin' is thought by some to derive from Christianus, in that the cretinous patients were 'incapable of sin' [2]. They were considered as simple, innocent creatures, 'gens du bon Dieu' [2]. Cretinism was one of the diseases first recognized in early life before it was realized that adults were also affected. The observation that cretinoid conditions occurred in adult life in women occurred much later, and was particularly recognized by Sir William Gull (1816–90) [19] in 1873. He was instrumental in leading the commission (which he initially chaired) to the conclusion that myxoedema was actually due to thyroid deficiency. William Ord (1834–1902) [20] gave the name 'myxoedema' to the adult form in 1877 and was the chairman of the commission at the time of the report (1888).

Significant differences were noticed between the adult and the childhood form. In the latter, there was arrest of development in growth both in body and brain. Because cretinism was found more commonly in the deep mountain valleys, there were many theories as to the relationship of air, water, and food, as noted by Hoefer (1614–81), DeSaussure (1740–99), and Malacarne (1744–1816) [1]. A Royal Commission in Sardinia (1) in 1848 found that the incidence of cretinism was 28% of the population in the District of Aosta but much lower elsewhere.

Causes of Hypothyroidism

The causes of the hypothyroidism were not fully understood throughout the nineteenth century. W.M. Ord [20, 21] described the appearance of the thyroid gland in this condition in 1878, and 13 cases were examined after death in the Report of the Clinical Society of London in 1888. The thyroid showed fibroid atrophy and great diminution in size and weight. There was evidence of chronic inflammation, lymphocytic infiltration, fibrosis, and disappearance of the acini and colloid. However, in 1912, it was Hakaru Hashimoto (1888–1934) [22, 23] a surgeon from Fukuoka, Japan, who described four cases of goitre associated with hypothyroidism in which lymphocytic infiltration of the thyroid gland was an important feature. The most common cause of spontaneous hypothyroidism in areas of the world where there is no iodine deficiency is that of (what is now termed) Hashimoto's thyroiditis or autoimmune thyroiditis. Only much later, in 1956, was autoimmune thyroiditis produced experimentally by Rose and Witebsky in Buffalo, New York [24] and in that same year, thyroid antibodies were found in the circulation of patients with Hashimoto's thyroiditis by Roitt and Doniach [25] in the United Kingdom. These findings helped to usher in the era of autoimmunity.

Toxic Goitre

Few diseases can have more synonyms and none more eponyms. C.P. Howard [26] collected more than 20 such terms.

Looking back into antiquity, the Persian writer, Sayyid Ismail Al-Jurjani in 1136 [1, 2] may have been the first to connect exophthalmos with goitre. Flajani's (1741–1808) [27] description in 1802 in Ascoli, Italy, failed to associate the goitre, exophthalmos, and palpitations as one disease, and his account failed to attract much attention. Indeed, Antonia Testa's (1756–1814) [28] reference in 1800 to the coincidence of prominent eyes and a cardiac disorder likewise did not attract much attention. Testa was the professor of medicine and surgery at Bologna and was said to be a learned theorist, but a mediocre clinician. Caleb Hillier Parry [29, 30] (1755–1822) of Bath, England observed a case in August 1786, but his description of eight cases of 'enlargement of the thyroid gland in connection with enlargement or palpitation of the heart' was not published until 1825, 3 years after his death. His second case, seen in 1802, was of particular interest as it seemed to be precipitated by an acute stress, a factor that still exercises the interest of many observers. He considered that the thyroid was acting as a reservoir for the blood being pumped out by the hyperactive heart. This posthumous report in 1825 still preceded Graves' publication by a decade.

Robert J. Graves [31–33] (1796–1853) of Dublin, Ireland, gave a clinical lecture at the Meath Hospital in Dublin and subsequently published a short article on 'palpitation of the heart with enlargement of the thyroid gland' in 1835. He then included these accounts in textbooks that he later wrote. These texts drew considerable attention to the disorder. This attention was later amplified in 1840 by Karl Adolf von Basedow (1799– 1854) [34] of Merseburg, Germany, who described four patients with exophthalmos, goitre, and palpitations; his description eave rise to the phrase 'the Merseburg triad'. William Stokes (1804–78), a colleague and friend of Graves in Dublin, actually described hyperthyroidism much more fully than Graves in his text, Diseases of the Heart in 1854 [35].

In France, the first description of the disease was provided in 1856 by Jean-Martin Charcot (1825–93) [36]. employing the term 'cachexia exophthalmica' to describe the condition. Charcot's older colleague, Armand Trousseau (1801–67) mentioned in a lecture at the Hotel Dieu in Paris in 1860, that iodine, which had been 'inappropriately' prescribed for hyperthyroidism had actually caused marked amelioration of the disease [37]. Nevertheless, he felt that the use of iodine in toxic goitre was dangerous, and warned against it.

The credit for the precedence for the description of this disease is scarcely resolved to this day. Sir William Osler [38, 39] in his third edition of his famous textbook of medicine belatedly gave the credit for the first important description of this disease to Parry. However, Trousseau [37] had been impressed with the books written by Graves (who was highly regarded as an academic physician) and felt that Graves should be given the credit. In mid-Europe, many observers have given that honour to Basedow. Thus, although the term Graves' disease is in common usage in English-speaking countries, Basedow's disease is generally used in Europe. In Italy, the term, Flajani's disease is sometimes heard. At international meetings, the term Graves' disease is most commonly heard.

The cause of Graves' disease remained unknown, and led to many interesting hypotheses, most notably the importance of psychological factors. Rolleston [1] has summarized the various influences which were thought to be at work in causing this condition. In 1907, Charles Mayo [2] of Rochester, Minnesota, first used the term 'hyperthyroidism', to conform to the idea we hold today,

namely that the disorder represents an excess of thyroid hormone. In 1910, Kocher [1, 2] coined the term 'Jod-Basedow' to describe hyperthyroidism precipitated or aggravated by excess iodine. David Marine [2] also suggested that iodine might actually be a treatment for Graves' disease. A few years later, in 1913, Henry Plummer (1874–1937) [40] was able to separate Graves' disease from hyperthyroidism related to toxic nodular goitre (Plummer's disease). In 1924, Plummer and Boothby [41] showed that the preoperative use of iodine greatly simplified the operative management of Graves' disease.

Treatment of Graves' disease remained mainly surgical until 1942, when Hertz and Roberts [42, 43] introduced radioactive iodine for the diagnosis and treatment of Graves' disease. The following year, Astwood [44] used thiourea and thiouracil in the medical treatment of Graves' disease, thus initiating the era of antithyroid drug therapy. In 1956, the year that thyroid autoantibodies were first identified, Adams and Purves [45] in New Zealand described the presence of an abnormal stimulator of the thyroid in Graves' disease which later proved to be an antibody directed against the thyroid-stimulating hormone receptor (thyroid-stimulating antibody). Thus, Graves' disease, as well as Hashimoto's thyroiditis, proved to be an autoimmune disorder.

Other Conditions

In 1896, Riedel [46, 47] described invasive fibrous thyroiditis, a rare fibrosing condition of the thyroid gland and de Quervain [48] described subacute non-suppurative thyroiditis in 1896.

REFERENCES

1. Rolleston HD. *The Endocrine Organs in Health and Disease, with a Historical Review*. London: Oxford University Press, 1936.
2. Medvei VC. *A History of Endocrinology*. Lancaster, UK: MTP Press, 1982.
3. Merke F. *History and Iconography of Endemic Goitre and Cretinism*. Lancaster, UK: MTP Press, 1984.
4. Hart GD. Even the gods had goitre. *Can Med Assoc J*, 1967; **96**: 1432–6.
5. Paracelsus (Bombastus v. Hohenheim TPA). De Generatione Stultorum. In: *Opera*, 1603; **2**: 174–82.
6. Wharton T. *Adenographia: Sive, Glandularum Totius Corporis Descriptio*. London, 1656: 118.
7. Davy H. *Philosophical Transactions of the Royal Society of London*, 1814; 104: 74.
8. Coindet JF. Decouverte d'un nouveau remede contre le goitre. *Ann Chim Phys*, 1820; **15**: 49–59.
9. Rilliet F. Constitutional iodism. *Bulletin de l'Academie de Medicine de Paris*, 1859; **25**: 382.
10. King TW. Observations on the thyroid gland. *Guys Hosp Rep*, 1836; **1**: 429–46.
11. Horsley V. Functional nervous disorders due to loss of thyroid gland and pituitary body. *Lancet*, 1886; **2**: 5.
12. Gley E. Sur les fonctions des corps thyroide. *Comptes Rendues de la Societe Biologique*, 1891; **43**: 841–7.
13. Murray GR. Notes on the treatment of myxoedema by hypodermic injections of an extract of the thyroid gland of a sheep. *Br Med J*, 1891; **ii**: 796–7.
14. Magnus-Levy A. Ueber den respiratorischen Gaswechel unter dem Einfluss der Thyroidea sowie unter verschiedenen pathologischen Zustaenden. *Berlin Klinischen Wochenschrift*, 1895; **32**: 650–2.
15. Baumann E, Goldman E. Ist das iodothyrin (thyrojodin) der lebenswichtige bestandteil der schilddruse?. *Munch Med Wochensch*, 1896; **43**: 1153.
16. Kendall EC. Collected papers. *Mayo Clinic, Mayo Foundation*, 1917; **9**: 309–36.
17. Harington CR, Barger G. Chemistry of thyroxine. III. Constitution and synthesis of thyroxine. *Biochem J*, 1927; **21**: 169–83.
18. Gross J, Pitt-Rivers RV. Triiodothyronine. I. Isolation from thyroid gland and synthesis. *Biochem J*, 1953; **53**: 645–50.
19. Gull WW. On a cretinoid state supervening in adult life in women. *Trans Clin Soc Lond*, 1873–74; **7**: 180–5.
20. Ord WM. On myxoedema, a term proposed to be applied to an essential condition in the 'cretinoid' affection, occasionally observed in middle-aged women. *Med Chir Trans*, 1878; **61**: 57–78.
21. Ord WM. Report of a committee of the Clinical Society of London nominated December 14, 1883 to investigate the subject of myxoedema. *Trans Clin Soc Lond*, 1888; 8: 15.
22. Hashimoto H. Zur Kenntnis der lymphomatosen veranderung der schilddruse (struma lymphomatosa). *Archiv der Klinische Chirurgie*, 1912; **97**: 219.
23. Volpé R. Historical perspective: the life of Doctor Hakaru Hashimoto. *Autoimmunity*, 1989; **3**: 243–5.
24. Rose NR, Witebsky E. Studies of organ specificity. V. Changes in the thyroid glands of rabbits following active immunization with rabbit thyroid extracts. *J Immunol*, 1956; **76**: 417.
25. Roitt IM, Doniach D, Campbell RN, Hudson RV. Autoantibodies in Hashimoto's disease (lymphadenoid goitre). *Lancet*, 1956; **ii**: 820.
26. Howard CP. Hyperthyroidism. In: Barker LF, Hoskins RG, Mosenthal HO, eds. *Endocrinology and Metabolism*. D. Appleton and Co., 1922; 1: 304.
27. Flajani G. Sopra un tumor freddo nell'anterior parte dell collo detto bronchocele. *Collezione d'osservazioni e riflessioni di chirurgia, Roma*, 1802; **3**: 270–3.
28. Testa A. *Collezione d'osservazioni e reflessioni di chirurgia,* Roma, 1800.
29. Parry CH. *Collections from the Unpublished Medical Writings of the Late Caleb Hillier Parry.* London: Underwoods, 1825; 2: 110.
30. Volpé R. The life of Caleb Hillier Parry. *Endocrinologist*, 1994; **4**: 157–9.
31. Graves RJ. Clinical lectures. *Lond Med Surg J*, 1835; **7**: 513–20.
32. Taylor S. *Robert Graves: The Golden Years of Irish Medicine*. London: Royal Society of Medicine, 1989.
33. Coakley D. *Robert Graves: Evangelist of Clinical Medicine*. Dublin: Irish Endocrine Society, 1996.
34. Basedow CA. Exophthalmos durch hypertrophie des Zellgewebes in der augenhoehle. *Wochenschrift Ges Heilkunde*, 1840; **6**: 197–204, 220–8.
35. Stokes W. *Diseases of the Heart and Aorta*. Dublin, 1854: 278–97.
36. Charcot JM. Memoire sur une affection caractérisé par les palpitation du coeur et les arteres, la tumefavtion de la glande thyroide et une double exophthalmie. *Comptes Rendus desMemoires de la Societe Biologique*, 1857; **3**: 43.
37. Trousseau A. *Lectures on Clinical Medicine*. Paris: New Sydenham Society, 1868: 542.
38. Osler W. *Principles and Practice of Medicine*, 3rd edn. New York: D. Appleton and Co., 1898: 836.

39. Hoffenberg R. The thyroid and Osler. *J Roy Coll Phys Lond*, 1985; **19**: 80–5.
40. Plummer HS. The clinical and pathological relationship of simple and exophthalmic goiter. *Am J Med Sci*, 1913; **146**: 790.
41. Plummer HS, Boothby WM. The value of iodine in exophthalmic goitre. *Iowa Med Soc*, 1924; **14**: 66–73.
42. Hertz S, Roberts A, Evans RD. Radioactive iodine as an indicator in the study of thyroid physiology. *Proc Soc Exp Biol Med*, 1938; **38**: 510–13.
43. Hertz S, Roberts A. Application of radioactive iodine in therapy of Graves' disease. *J Clin Investig*, 1942; **21**: 624.
44. Astwood EB. Treatment of hyperthyroidism with thiourea. *J A Med A*, 1943; **122**: 78–81.
45. Adams DD, Purves HD. Abnormal responses to the assay of thyrotrophin. *Proc University of Otago Medical School*, 1956; **11**: 34.
46. Riedel BM. Die chronische zur bildung eisenharter tumoren fuhrende entzundung der schilddruse. *Verhandlungen der Deutschen Gesellschaft fur Chirurgie*, 1896; **25**: 101–5.
47. Riedel BM. Vorstellung eines kranken mit chronische strumitis. *Verhandlungen der Deutschen Gesellschaft Chirurgie*, 1897; **26**: 127–9.
48. de Quervain F. Die akute nicht eiterige thyreoiditis und die beteiligung der schilddruse und akuten intoxikationen und infectionen uberhaupt. *Mitteilungen aus der Grenzgeheiten der Medizin und Chirurgie*, 1991; **2**(suppl. Bd): 1–165, 6 pt.

3.1.2 Biosynthesis, Transport, Metabolism, and Actions of Thyroid Hormones

W. Edward Visser

Introduction

In healthy humans with a normal iodine intake, the thyroid follicular cells produce predominantly the prohormone thyroxine (3,3′,5,5′-tetraiodothyronine; T_4), which is converted in peripheral tissues to the bioactive hormone 3,3′,5-triiodothyronine (T_3) or to the inactive metabolite 3,3′,5′-triiodothyronine (reverse T_3). The bioavailability of thyroid hormone in target tissues depends to a large extent on the supply of plasma T_4 and T_3, the activity of transporters mediating the cellular uptake and/or efflux of these hormones, as well as the activity of deiodinases and possibly other enzymes catalysing their activation or inactivation. The actions of thyroid hormone are mediated by binding of T3 to the nuclear thyroid hormone (or T3) receptors (TRs). Thyroid function is regulated most importantly by the hypophyseal glycoprotein thyroid-stimulating hormone (TSH), also called thyrotropin. In turn, TSH secretion from the anterior pituitary is stimulated by the hypothalamic factor thyrotropin-releasing hormone (TRH). TSH secretion is downregulated by negative feedback action of thyroid hormone on the hypothalamus and the pituitary. The contribution of locally produced T_3 versus uptake of plasma T_3 is much greater for some tissues such as the brain and the pituitary than for most other tissues. Plasma TSH is an important parameter for the diagnosis of thyroid dysfunction but is not representative for the thyroid state of all tissues. In this chapter various aspects will be discussed of: (a) the neuroendocrine regulation of thyroid function, (b) the biosynthesis of thyroid hormone (i.e. the prohormone T_4), (c) the activation and inactivation of thyroid hormone in peripheral tissues, (d) cellular transport of thyroid hormone and (e) the mechanism by which T_3 exerts it biological activity. A schematic overview of the hypothalamus– pituitary–thyroid–periphery axis is presented in Figure 3.1.2.1.

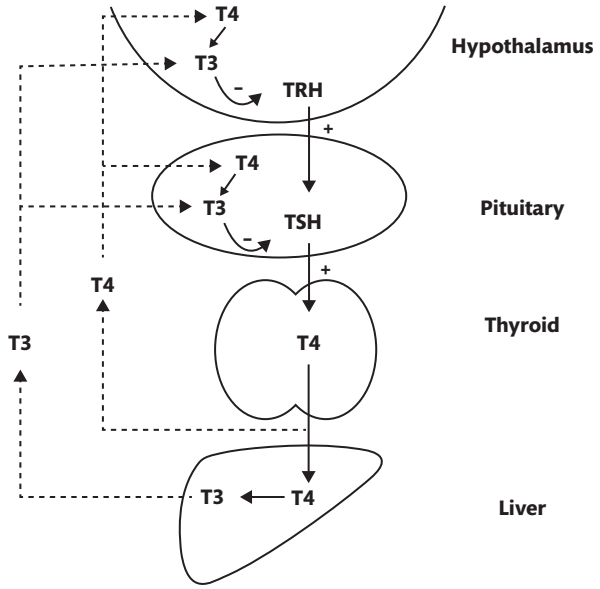


Figure 3.1.2.1 Overview of the regulation of the production and metabolism of thyroid hormone in the hypothalamus–pituitary–thyroid–periphery axis, showing the liver as a major T_3-producing tissue.

Regulation of Thyroid Function

Thyrotropin-Releasing Hormone

TRH is a tripeptide with the structure pyroglutamyl-histidyl-proline amide (pGlu-His-Pro-NH_2) in which the C-terminal carboxyl group is blocked by amidation and the N-terminal α-amino group is blocked by cyclization. Beside stimulating TSH secretion,

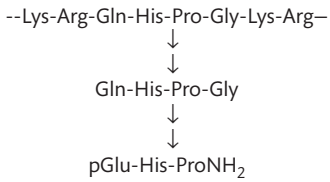


Figure 3.1.2.2 Biosynthesis of TRH. The figure shows the several steps by which the TRH progenitor sequences in proTRH are processed to mature TRH.

TRH also stimulates prolactin secretion. TRH is not only produced in the hypothalamus but is widely distributed through the central nervous system where it functions as a neurotransmitter. TRH is also detected in the posterior pituitary and in different other tissues, but little is known about its role there.

Hypophysiotropic TRH is produced in neurons, the cell bodies of which are located in the paraventricular nucleus of the hypothalamus [1]. The biosynthesis of TRH involves the production of a large precursor protein (proTRH) which, in humans, consists of a sequence of 242 amino acids. This proTRH contains six copies of the TRH progenitor sequence Gln-His-Pro-Gly, flanked at both sides by pairs of the basic amino acids Arg and/or Lys (Figure 3.1.2.2). Cleavage of proTRH at the basic amino acids by prohormone convertases (e.g. PC1 and PC2) and further removal of remaining basic residues by carboxypeptidases results in the liberation of the progenitor sequences. A specific glutaminyl cyclase catalyses the formation of the pGlu ring at the N-terminus and a so-called peptidylglycine α-amidating mono-oxygenase converts Pro-Gly to $ProNH_2$ at the C-terminus [2]. The processing of proTRH takes place in vesicles that transport mature TRH and intervening peptides along the axons of the TRH neurons to the median eminence, where they are released into the portal vessels of the hypophyseal stalk.

TRH is transported over a short distance through the hypophyseal stalk to the anterior lobe of the pituitary, where it stimulates the production and secretion of TSH (and prolactin). These actions of TRH are initiated by its binding to the type 1 TRH receptor (TRHR1), which is expressed on both the thyrotrope (TSH-producing cell) and the lactotroph (prolactin-producing cell) [3]. This receptor belongs to the family of G-protein-coupled receptors, characteristically containing seven transmembrane domains. Human TRHR1 is a protein consisting of 398 amino acids, and binding of TRH induces a change in its interaction with the trimeric G-protein, resulting in the stimulation of phospholipase C activity. The activated phospholipase C catalyses the hydrolysis of phosphatidylinositol-4,5-diphosphate to the second messengers inositol-1,4,5-triphosphate and diacylglycerol, which initiate a cascade of reactions, including an increase in cellular Ca^{2+} levels and protein kinase C activity, that ultimately stimulates the release as well as the synthesis of TSH (and prolactin) [3]. TRH stimulation of TSHβ gene expression is also dependent on the pituitary-specific transcription factor 1.

TRH is subject to rapid degradation in the blood as well as in different tissues. Although multiple enzymes are involved, a very important role is played by the TRH-degrading ectoenzyme TRHDE, which catalyses the cleavage of the pGlu-His bond [4]:

$$pGlu\text{-}His\text{-}ProNH_2 \rightarrow pGlu + His\text{-}ProNH_2.$$

This enzyme has been characterized as a zinc-containing metalloproteinase, which in humans consists of 1024 amino acids. It has a single transmembrane domain and is inserted in the plasma membrane such that most of the protein is exposed on the cell surface (ectopeptidase), in particular in brain, pituitary, liver, and lung. Enzymatic cleavage of the protein close to the cell membrane releases most of the protein in a soluble and enzymatically active form into the circulation, representing the origin of plasma TRHDE. Plasma TRHDE appears to be derived mostly from the liver. In the brain and the pituitary, where the enzyme is probably located in close vicinity of the TRH receptor, TRHDE supposedly plays an important role in the local regulation of TRH bioavailability. TRHDE activity in the pituitary and in plasma is increased in hyperthyroidism and decreased in hypothyroidism, which may contribute to the negative feedback control of TSH secretion by thyroid hormone [4].

Thyroid-Stimulating Hormone

TSH is a glycoprotein produced by the thyrotropic cells of the anterior pituitary. Like the other hypophyseal hormones luteinizing hormone and follicle-stimulating hormone, it is composed of two subunits. The α-subunit is identical and the β-subunit is homologous among the three hormones [5]. Although hormone specificity is conveyed by the β-subunit, dimerization with the α-subunit is required for biological activity. Human TSH consists of 205 amino acids; 92 in the α-subunit and 113 in the β-subunit. It has a molecular weight of 28 kDa, 20% of which is contributed by three complex carbohydrate groups: two on the α-subunit and one on the β-subunit. The structure of these carbohydrate groups is important for the biological activity of TSH and is dependent on the stimulation of the thyrotrope by TRH [5]. Changes in TSH glycosylation underlie the altered TSH bioactivity under certain circumstances, such as pituitary stalk compression.

In addition to the stimulation by TRH and negative feedback by thyroid hormone, TSH production and secretion is also subject to negative regulation by hypothalamic somatostatin and dopamine and by steroids such as cortisol [6].

TSH binds to a specific TSH receptor located in the plasma membrane of the follicular cell. The human TSH receptor is a G-protein-coupled receptor which, is a protein consisting of 764 amino acids with an exceptionally long extracellular N-terminal domain [7]. The TSH receptor is preferentially coupled to a $G_s\alpha$-subunit of the trimeric G-protein. Binding of TSH to its receptor induces the dissociation of the G-protein subunits, resulting in the activation of the membrane-bound adenylate cyclase and, thus, in the stimulation of cAMP formation as second messenger. The increased cAMP levels induce a series of events, including the activation of protein kinase A activity, that ultimately results in the stimulation of the biosynthesis and secretion of thyroid hormone [8]. In particular, the expression of genes coding for key proteins for hormone production (e.g. the iodide transporter, thyroglobulin, and thyroid peroxidase) is increased through mechanisms that also involve different thyroid-specific transcription factors such as TTF1, TTF2, and PAX8.

As discussed elsewhere in this section, hyperthyroidism is often caused by an autoimmune process in which TSH receptor-stimulating antibodies play an important role. Hyperthyroidism may also be caused by a hyperfunctioning adenoma. In most patients with a toxic adenoma, somatic mutations have been identified

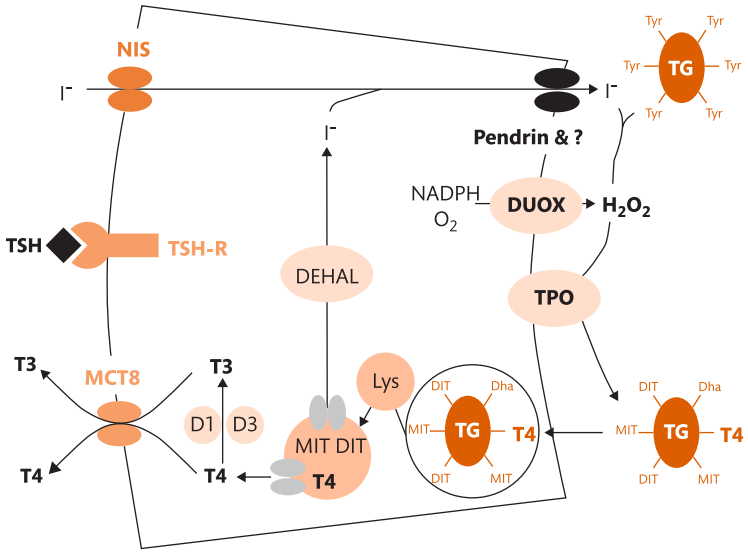


Figure 3.1.2.3 Schematic presentation of a thyroid follicular cell and important steps in the synthesis of thyroid hormone. DIT, 3,5-diiodotyrosine; MIT, monoiodotyrosine.

in the TSH receptor, which result in the constitutive activation of this receptor [9]. In other patients, somatic mutations have been found in the $G_s\alpha$-subunit that result in the constitutive activation of the G-protein in the absence of TSH. Together, mutations in the TSH receptor and the $G_s\alpha$-subunit account for the majority of toxic adenomas. Also, germline, gain-of-function mutations have been identified in patients with congenital, non-autoimmune hyperthyroidism. Conversely, germline, loss-of-function mutations have been described in patients with TSH resistance [9]. However, patients with TSH resistance may be clinically euthyroid because the partial defect in TSH receptor function is compensated by increased plasma TSH concentrations [9].

Biosynthesis of Thyroid Hormone

The functional unit of the thyroid gland is the follicle, composed of a single layer of epithelial cells surrounding a colloidal lumen in which thyroid hormone is stored as an integral part of its precursor protein thyroglobulin. The biosynthesis of thyroid hormone comprises the following steps, which are depicted schematically in Figure 3.1.2.3 [8, 10]:

1. Uptake of iodide through the basolateral membrane and export through the apical membrane.
2. Clustering of thyroglobulin, thyroid peroxidase (TPO) and the dual oxidase DUOX2 in a 'thyroxisome' at the luminal surface of the apical membrane [11].
3. Formation of H_2O_2 by DUOX2.
4. H_2O_2-dependent iodination of tyrosine residues in thyroglobulin by TPO.
5. H_2O_2-dependent coupling of iodotyrosine to iodothyronine residues in thyroglobulin by TPO.
6. Resorption of thyroglobulin from the lumen and hydrolysis in lysosomes.
7. Deiodination of iodotyrosines and reutilization of iodide.
8. Secretion of iodothyronines, predominantly T_4, into the bloodstream.

Iodide Uptake

Iodine is an essential trace element required for the synthesis of thyroid hormone. The basolateral membrane of the follicular cell requires an active transporter that mediates uptake of I^- together with Na^+. This human sodium-iodide symporter (NIS) consists of 618 amino acids and 13 transmembrane domains [12]. Supposedly, these domains form a channel through which I^- and Na^+ are transported in a stoichiometry of 1:2. The surplus of positive charge indicates that I^- transport is electrogenic and further driven by the Na^+ gradient. TSH stimulates the expression of the *NIS* gene to such an extent that the intracellular iodide concentration is about 30–50 times higher than its extracellular level. NIS can also bind other anions, some of which are even transported [12].

An important example is perchlorate (ClO_4^-) which potently inhibits iodide uptake by the NIS, an effect utilized in the perchlorate discharge test used for the diagnosis of an organification defect (i.e. impaired incorporation of iodine in thyroglobulin). Perchlorate inhibits the uptake but not the release of iodide from the thyroid. Therefore, if perchlorate is administered after a dose of radioactive iodide, it will provoke a marked release of radioactivity from the thyroid in case of an organification defect but not from a normal thyroid gland. Pertechnetate (TcO_4^-) is another anion transported by the NIS, and this observation is utilized in the scanning of the thyroid gland using radioactive $^{99m}TcO_4^-$.

Since the iodination of thyroglobulin takes place at the luminal surface of the apical membrane, iodide also has to pass this membrane. A transporter putatively involved in this process has been identified and termed pendrin, since the gene coding for this

protein is mutated in patients with Pendred's syndrome [13]. This is a congenital condition characterized by deafness due to a cochlear defect and hypothyroidism due to an organification defect as indicated by a positive perchlorate discharge test. Pendrin is capable of transporting bicarbonate, chloride, and iodide [13], and is expressed only in the thyroid and the cochlea. The exact function of pendrin in the transport of iodide across the follicular apical membrane is subject to debate. There is likely another protein capable of releasing iodide into the follicular lumen of which anoctamin is a likely candidate [14]. Efflux of iodide from thyroid follicular cells is acutely stimulated by TSH, which may involve recruitment and/or activation of an iodide exporter such as pendrin.

Thyroglobulin, DUOX2, and TPO

Thyroglobulin is an exceptionally large glycoprotein consisting of two identical subunits. Each mature subunit in human thyroglobulin contains 2748 amino acids and has a molecular weight of approximately 330 kDa [15]. The *TG* gene is located on human chromosome 8q24.2-q24.3; it covers about 300 kb of genomic DNA and consists of 48 exons.

DUOX2 is a large glycoprotein embedded in the apical membrane of the thyrocyte. Mature human DUOX2 contains 1527 amino acids and has seven putative transmembrane domains, an NADPH-binding domain, an FAD-binding domain, a haem-binding domain, two calcium-binding EF hands, and a peroxidase domain [16]. It catalyses the oxidation of NADPH from the cytoplasm and delivers its product H_2O_2 to the luminal surface of the membrane. Functional expression of DUOX2 requires the presence of the maturation factor DUOXA2, a protein consisting of 320 amino acids and five putative transmembrane domains [17].

TPO is a glycoprotein consisting of 933 amino acids and featuring a single transmembrane domain. A short C-terminal domain is located in the cytoplasm but most of the protein is exposed on the luminal surface of the apical membrane, which also contains a haem-binding domain, the active centre of the enzyme [18]. The human *TPO* gene covers about 150 kb on chromosome 2p25, distributed over 17 exons.

Formation of Iodothyronines

Thyroid hormone synthesis takes place at the luminal surface of the apical membrane in the scaffold of the thyroglobulin molecule and consists of two important reactions that are both catalysed by TPO (i.e. the iodination of Tyr residues and the subsequent coupling of iodotyrosine to iodothyronine residues) [18]. The structures of these compounds are illustrated in Figure 3.1.2.4. The prosthetic haem group of TPO undergoes a two-electron oxidation by H_2O_2 (supplied by DUOX2) to the intermediate compound 1 (Cpd1). Cpd1 may carry out either a one-electron oxidation reaction, by which it is converted to the intermediate Cpd2, or a two-electron oxidation by which native TPO is regenerated. TPO-catalysed iodination probably involves a two-electron oxidation of I^- to I^+ with subsequent electrophilic substitution of Tyr residues in thyroglobulin, producing 3-iodotyrosine (monoiodotyrosine, MIT). Substitution of MIT residues with a second iodine produces 3,5-diiodotyrosine (DIT).

Coupling of two suitably positioned iodotyrosine residues results in the formation of an iodothyronine residue at the site of the acceptor iodotyrosine, leaving a dehydroalanine residue at the site of the donor iodotyrosine [18]. Coupling of the diiodophenol moiety

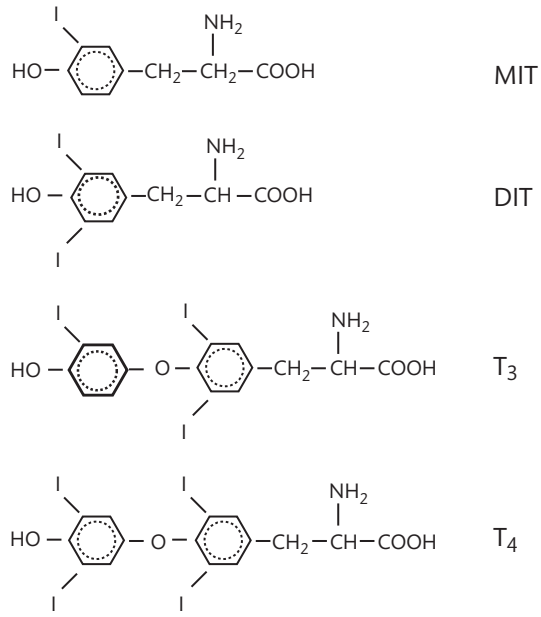


Figure 3.1.2.4 Structures of the iodotyrosines MIT and DIT and the iodothyronines T3 and T4.

of one DIT residue to the phenolic oxygen of a second DIT residue results in the formation of T_4, while coupling of the iodophenol moiety of MIT to a DIT residue yields T_3. Coupling of DIT and MIT to generate reverse T_3 or MIT and MIT to form 3,3′-T_2 are apparently rare events, since thyroidal secretion of reverse T_3 and 3,3′-T_2 are negligible.

Although Tyr is the building block of thyroid hormone, the Tyr content of thyroglobulin is not greater than that of most other proteins. Each thyroglobulin subunit has only four hormonogenic sites, Tyr residues that can ultimately be transformed into iodothyronines. At three sites (positions 5, 1290, and 2553 in the mature protein) T_4 can be formed, while at the fourth site (position 2746) T_3 is preferentially produced. However, at normal levels of iodination the average yield is 1–1.5 molecules of T_4 and approximately 0.1 molecule of T_3 per thyroglobulin subunit. The ratio between T_3 and T_4 formation is under control of TSH. At this stage the iodothyronines are still in peptide linkage with the thyroglobulin backbone and remain stored as such in the lumen until their secretion is required.

Release of Thyroid Hormone

In response to TSH stimulation, thyroglobulin is resorbed from the lumen largely by both macro- and micropinocytosis [8, 10]. The former type of endocytosis is associated with the formation of large pseudopodia that engulf colloid and the thyroglobulin contained therein, resulting in the formation of large cytoplasmic vesicles also known as colloid droplets. The second process concerns the receptor-mediated endocytosis of thyroglobulin, involving the binding of thyroglobulin to apical membrane proteins. Megalin, a very large (*c.*600 kDa) cargo protein located in the apical membrane of different cell types, including thyrocytes, may be involved although it appears to function primarily in the transcellular transport of poorly iodinated thyroglobulin [19].

Both types of endosomes fuse with lysosomes, generating so-called phagolysosomes. In these vesicles thyroglobulin is

hydrolysed by lysosomal proteases, (i.e. cathepsins) [20], resulting in the liberation of T_4, a small amount of T_3, as well as excess MIT and DIT molecules. MIT and DIT are probably exported from the vesicles via a specific transporter [21]. Thus, they have access to the iodotyrosine dehalogenase (DEHAL1), located in the endoplasmic reticulum, which catalyses their deiodination by NADH [22, 23]. The iodide thereby released is reutilized for iodination of thyroglobulin.

Human DEHAL1 is a homodimer of a 289-amino acid protein containing an N-terminal membrane anchor and a conserved nitroreductase domain with an FMN-binding site [22, 23]. The *DEHAL1* gene is located on chromosome 6q24-q25 and consists of five exons. Since DEHAL1 lacks an NADH-binding sequence, iodotyrosine deiodinase activity requires the involvement of a reductase, which has not yet been identified.

The most important mechanism of thyroid hormone secretion involves the secretion from the gland through membrane transporters. This requires iodothyronines to be released via transporters from the vesicles into the cytoplasm, and subsequently secreted through transporters located in the basolateral membrane. In the latter route, some T_4 may be converted before secretion to T_3 by iodothyronine deiodinases present in the thyrocyte (see next). The thyroid hormone transporter MCT8 (see next) is crucial for the excretion of the hormones from the gland into the circulation.

In an average human subject, T_4 and T_3 are secreted in a ratio of about 15:1 (i.e. about 100 μg (130 nmol) T_4 and 6 μg (9 nmol) T_3 per day). The latter represents approximately 20% of daily total T_3 production [24]. Hence, most T_3 is produced by deiodination of T_4 in peripheral tissues.

Inhibitors of Thyroid Hormone Production and/or Secretion

Administration of a large amount of iodide usually results in an acute but transient decrease in thyroid hormone secretion [8, 10]. The mechanism of this inhibition of thyroid hormone secretion by excess iodide is unknown. Excess iodide will also induce an inhibition of the synthesis of thyroid hormone; this phenomenon is known as the Wolff–Chaikoff effect [8, 10]. The mechanism appears to involve, among others, the formation of an iodinated lipid (iodolactone) that inhibits several steps in thyroid hormone synthesis. This includes the inhibition of iodide uptake by the NIS, which results in a decrease in the intracellular iodide concentration and, thus, a decrease in iodolactone formation. This relieves the inhibited hormone synthesis, known as the escape from the Wolff–Chaikoff effect, that occurs despite the continued administration of excess iodide.

Thiourea derivatives have been known since long as potent inhibitors of thyroid hormone synthesis [25]. Two of these, 6-propyl-2-thiouracil and particularly methimazole are widely used in the medical treatment of patients with hyperthyroidism (Figure 3.1.2.5). Their antithyroid activity is based on the potent inhibition of TPO, the mechanism of which depends on the available iodide concentration [18]. In the presence of iodide, the thiourea inhibitors compete with the Tyr residues in thyroglobulin for the TPO–I^+ iodination complex, preventing the formation of thyroid hormone. The thiourea inhibitors are thus converted to the sulfenyl iodide derivatives which undergo further oxidation of the sulphur ultimately to sulphate.

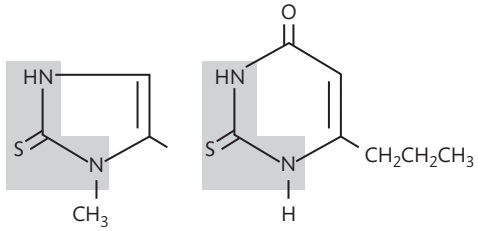


Figure 3.1.2.5 Structures of the TPO inhibitors methimazole and propylthiouracil. The thiourea moiety of the drugs is shaded.

Methimazole is a more potent inhibitor of TPO than propylthiouracil [18], and lower doses of methimazole (or the prodrug carbimazole) are required for the treatment of hyperthyroidism compared with propylthiouracil. Besides inhibiting thyroid hormone (i.e. T_4) synthesis by TPO, propylthiouracil also inhibits conversion of T_4 to T_3 by the type 1 iodothyronine deiodinase located not only in the thyroid but also in liver and kidney (see next).

Transport of Thyroid Hormone

Plasma Transport

In plasma, thyroid hormone is bound to three proteins, thyroxine-binding globulin (TBG), transthyretin (TTR, previously known as thyroxine-binding prealbumin (TBPA)), and albumin (Table 3.1.2.1) [25, 26]. Human TBG is a 54-kDa glycoprotein produced in the liver and consists of 395 amino acids and four carbohydrate residues. The *TBG* gene is located on the human chromosome Xq22.2, spans about 5.5 kb, and contains five exons [27]. Among the different thyroid hormone transport proteins, it shows by far the highest affinity for T_4, with an equilibrium dissociation constant (K_d) of approximately 0.1 nM, but also the lowest plasma concentration (*c.*15 mg/L) [26].

TTR is composed of four identical subunits, each consisting of 127 amino acids. The *TTR* gene is located on human chromosome 18q11.2-q12.1, covers about 7 kb, and contains four exons [28]. TTR can bind two T_4 molecules, with a K_d value of approximately 10 nM of the first T_4 molecule, and the plasma concentration of TTR amounts to approximately 250 mg/L [26]. Plasma TTR is produced in the liver, but the protein is also expressed in the choroid plexus where it is probably involved in T_4 transfer from plasma to the cerebrospinal fluid. Furthermore, TTR is expressed in trophoblasts where it may participate in the transplacental transfer of maternal T_4 to the fetus.

Albumin has multiple low-affinity binding sites for thyroid hormone, with K_d values for T_4 of 1–10 μM, but it has by far the highest plasma concentration (*c.*40 g/L) [26].

The resultant of the concentrations and affinities of the different thyroid hormone-binding proteins is that in normal human subjects approximately 75% of plasma T_4 is bound to TBG, approximately 15% is bound to albumin, and approximately 10% is bound to TTR [26]. The total binding capacity of these proteins is so high that only approximately 0.02% of plasma T_4 is free (non-protein-bound). The affinity of T_3 for the different proteins is roughly 10% of that of T_4. Therefore, plasma T_3 shows a similar distribution to T_4 over the different proteins, and the free T_3 fraction in normal plasma amounts

Table 3.1.2.1 Characteristics of T_4-binding proteins in human plasma

Protein	Concentration in plasma (mg/L)	Concentration in plasma (μmol/L)	Dissociation constant (K_d) (mol/L)	T_4 distribution (%)
TBG	*c.*15	*c.*0.3	*c.*10^{-10}	75
TTR	*c.*250	*c.*5	10^{-8}	10
Albumin	*c.*40 000	*c.*600	10^{-6} to 10^{-5}	15

TBG, T_4-binding globulin; TTR, transthyretin (formerly known as T_4-binding prealbumin, TBPA).

to approximately 0.2%. Thus, while the mean normal plasma total T_4 (*c.*100 nmol/L) and T_3 (*c.*2 nmol/L) levels differ about 50-fold, the difference in the mean normal free T_4 (*c.*20 pmol/L) and free T_3 (*c.*5 pmol/L) is only about fourfold. Reverse T_3 binds with intermediate affinity to the plasma proteins [26].

Since it is the plasma free T_4 and free T_3 concentrations that determine the tissue availability of thyroid hormone, they are more important parameters than the plasma total T_4 and T_3 concentrations in the assessment of thyroid status. Both concentration and thyroid hormone-binding affinity of the different plasma proteins are influenced by a variety of (patho)physiological factors [26]. Beside genetic variation resulting in deficiency or excess, TBG levels are also influenced by various endogenous and exogenous factors [27]. Notably, plasma TBG levels are increased by oestrogens. Different endogenous factors, such as free fatty acids, and drugs, such as salicylates, competitively inhibit T_4 binding to TBG [26].

A large number of mutations have also been identified in the *TTR* gene, some of which are associated with a decrease in T_4 binding affinity, whereas others result in an increased affinity for T_4 [29]. *TTR* mutations often cause neuropathic or cardiomyopathic amyloidosis, resulting from the deposition of insoluble TTR fibrils in nerves or the heart [29]. Finally, binding of thyroid hormone to albumin is subject to genetic variation. In particular, a specific increase in the binding of T_4 to albumin is frequently observed in otherwise healthy subjects, which may lead to the false diagnosis of hyperthyroidism if inadequate methods for analysis of plasma free T_4 are used [26]. This phenomenon of familial dysalbuminaemic hyperthyroxinaemia has been attributed to mutations in the albumin gene (typically R218H mutation), resulting in a marked increase in T_4 affinity [30].

Perturbation of plasma iodothyronine binding provokes an adaptation of the hypothalamus–pituitary–thyroid axis until normal free T_4 and free T_3 concentrations are again obtained. Therefore, measurement of plasma free T_4 rather than total T_4 levels is, together with analysis of plasma TSH, the cornerstone of the diagnosis of thyroid disorders.

Tissue Transport

Because iodothyronines are lipophilic compounds, it was assumed for decades that they readily pass the plasma membrane by simple diffusion. However, the polar nature of the alanine side chain ('zwitterion') is a serious obstacle for passage through the lipid bilayer of the cell membrane. Studies in recent years have established that tissue uptake of thyroid hormone does not take place by diffusion but is mediated by specific plasma membrane transporters [31]. Most early studies have been carried out in isolated rat hepatocytes, but carrier-mediated uptake of iodothyronines has been demonstrated in a variety of cells [31].

The mechanisms of thyroid hormone uptake appear to differ between tissues as studies in rat pituitary cells suggest a common transporter for T_4 and T_3, whereas neonatal rat cardiomyocytes show preferential uptake of T_3 over T_4 [31]. In view of the iodothyronine structure, it is not surprising that thyroid hormone uptake can be mediated by amino acid transporters showing partial (L-type) or complete (T-type) preference for aromatic amino acids [32, 33].

Thyroid Hormone Transporters

Since 2000, a number of thyroid hormone transporters have been identified at the molecular level. These include the Na-taurocholate cotransporting polypeptide (NTCP), different members of the organic anion transporting polypeptide (OATP) family, the L-type amino acid transporters (LATs), members of the monocarboxylate transporter (MCT; SLC16) family and SLC17A4 [34, 35].

Of these transporters, only NTCP (SLC10A1) transports its ligands in a Na^+-dependent manner [36]. It is exclusively expressed in liver and transports primarily bile acids. Human NTCP consists of 349 amino acids and has seven transmembrane domains. The *NTCP* gene is located on chromosome 14q24.1 and has five exons. There are no other thyroid hormone transporters in the SLC10 family. NTCP shows a preference for sulphated over non-sulphated iodothyronines.

Different human OATP members have been shown to transport iodothyronine derivatives [37]. In general, they are multispecific, transporting a variety of ligands. OATPs are glycoproteins containing around 700 amino acids and 12 transmembrane domains. The human OATP1 subfamily contains four members (OATP1A2, 1B1, 1B3, 1C1) with interesting properties. They are encoded by a gene cluster on chromosome 12p12 containing 14–15 exons. OATP1B1 and 1B3 are expressed only in liver and show preference for sulphated over non-sulphated iodothyronines as ligands [38]. The latter also holds for OATP1A2, which is expressed in different tissues. OATP1C1 is an exceptional transporter in this subfamily, showing a high preference for T_4 as the ligand and almost exclusive expression in the brain, especially in choroid plexus and capillaries in mice and in astrocytes in humans [37]. Recently, a young patient with progressive neurodegeneration associated with a defective OATP1C1 was identified [39].

T_4 and T_3 are also transported by two members of the heterodimeric amino acid transporters LAT1 and LAT2 [40]. These transporters are glycoproteins consisting of two subunits, a heavy chain and a light chain. The heavy chains contain a single transmembrane domain, and the light chains contain 12–14 transmembrane domains. LAT1 is composed of the SLC3A2 (4F2hc or CD98hc) heavy chain and the SLC7A5 light chain, and LAT2 is composed of the same heavy chain and the SLC7A8 light chain.

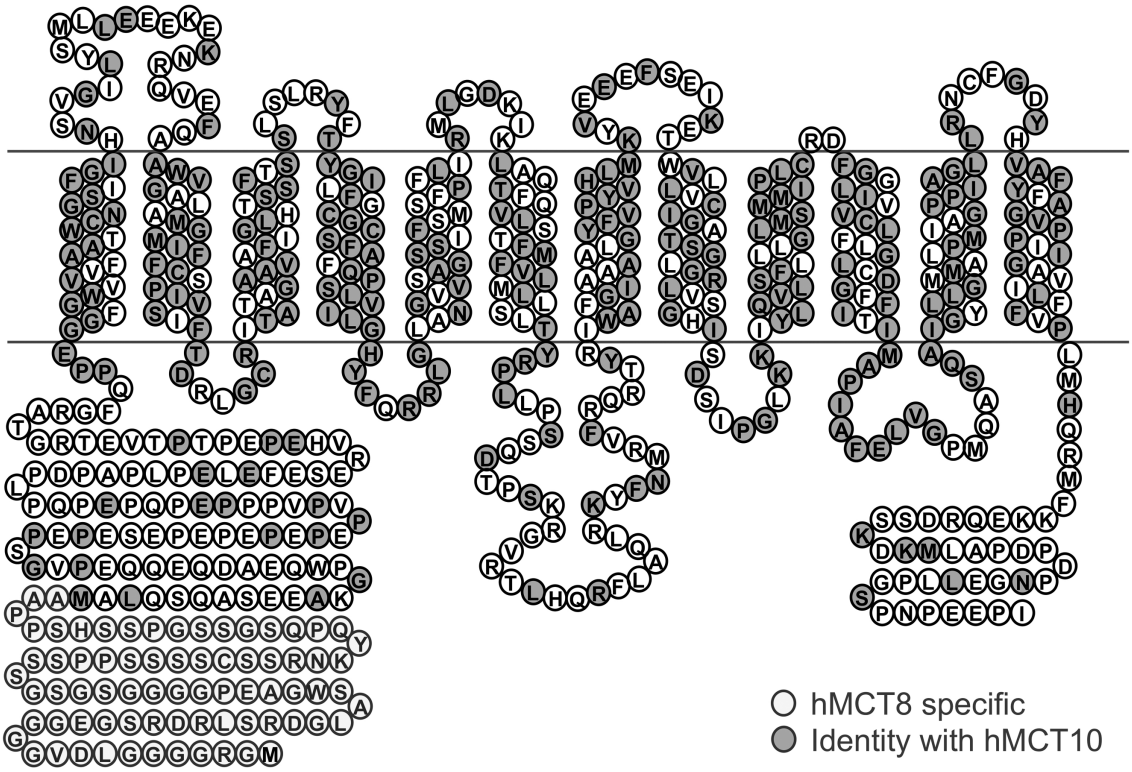


Figure 3.1.2.6 Protein structure of human MCT8 and MCT10.

These transporters are expressed in various tissues and facilitate the bidirectional transport of a variety of aliphatic and aromatic amino acids as well as iodothyronines over the plasma membrane [40].

Two important thyroid hormone transporters belong to the monocarboxylate transporter (MCT) family [41, 42]. MCT8 and MCT10 have been identified as important thyroid hormone transporters. Of these, MCT10 also transports the aromatic amino acids Trp, Tyr, and Phe, but so far only iodothyronines have been identified as ligands for MCT8.

Human MCT8 consists of 613 or 539 amino acids, depending on which of the two possible translation start sites is used, and MCT10 has 515 amino acids. They are homologous proteins with about 50% amino acid identity between 'short' MCT8 and MCT10 (Figure 3.1.2.6). Like the other MCTs, both MCT8 and MCT10 have 12 transmembrane domains. They have identical gene structures; the *MCT8* gene is located on human chromosome Xq13.2, and the *MCT10* gene is located on chromosome 6q21-q22. Both consist of six exons and five introns, with a large approximately 100 kb first intron. MCT8 and MCT10 show wide but different tissue distributions.

MCT8 and MCT10 are the most active and specific thyroid hormone transporters known today [35, 43, 44]. MCT8 is importantly expressed in brain, in particular in the endothelial cells of the blood-brain barrier, but is also expressed in astrocytes, neurons, and in the choroid plexus. MCT8 is essential for thyroid hormone uptake in the brain and into neural cells and, thus, for the crucial action of thyroid hormone during brain development. Mutations in MCT8 have been identified as the cause of Allan–Herndon–Dudley syndrome (AHDS). AHDS or MCT8 deficiency occurs in male patients and comprises a severe intellectual and motor disability and a peripheral thyrotoxicosis resulting from highly elevated serum T_3 levels [35, 43, 44] (see Chapter 3.4.5).

SLC17A4 transports T3 and T4 to a similar extent as MCT8 and is expressed in the human gastrointestinal tract. Although genetic variation is related to fT_4 concentrations, the precise function in physiology is unknown.

Several compounds may affect thyroid hormone uptake through inhibition of transporters (e.g. amiodaron) or changing the expression (e.g. non-thyroidal illness) and as such affect intracellular thyroid hormone bioavailability.

Metabolism of Thyroid Hormone

Deiodination

The thyroid gland of a healthy human adult with an adequate iodine intake produces predominantly the prohormone T_4 and only a small amount of the bioactive hormone T_3. It is generally accepted that, in humans, approximately 80% of circulating T_3 is produced by enzymatic outer ring deiodination (ORD) of T_4 in peripheral tissues [24]. Alternatively, inner ring deiodination (IRD) of T_4 produces the inactive metabolite reverse T_3. Deiodination is also an important pathway by which T_3 and reverse T_3 are further metabolized. T_3 largely undergoes IRD to the inactive compound T_2, which is also the main metabolite produced from reverse T_3 by ORD (Figure 3.1.2.7). Thus, the bioactivity of thyroid hormone is determined to an important extent by the enzyme activities responsible for the ORD (activation) or IRD (inactivation) of iodothyronines.

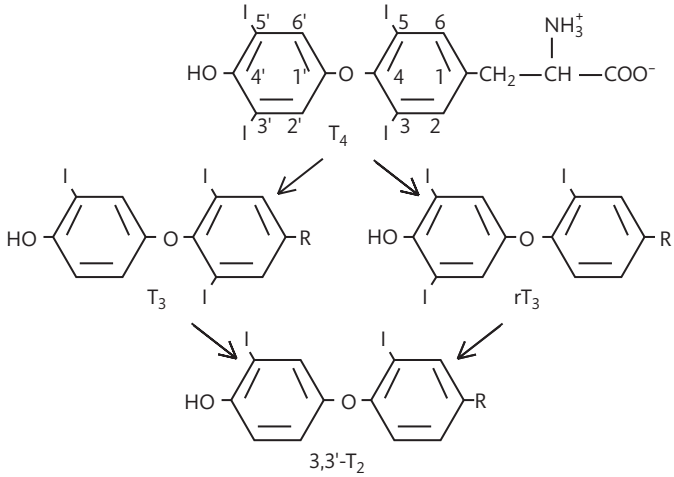


Figure 3.1.2.7 Conversion of the prohormone T_4 by outer ring deiodination (ORD) to the bioactive hormone T_3 or by inner ring deiodination (IRD) to the metabolite reverse T_3, and further conversion of T_3 by IRD and of reverse T_3 by ORD to the common metabolite T_2.

Three iodothyronine deiodinases (D_{1-3}) are involved in the reductive deiodination of thyroid hormone (Figure 3.1.2.8) [45]. They are homologous proteins consisting of 249–278 amino acids, with a single transmembrane domain located at the N-terminus. The deiodinases are inserted in cellular membranes such that the major part of the protein is exposed on the cytoplasmic surface. This is consistent with the reductive nature of the cytoplasmic compartment required for the deiodination process.

The most remarkable feature of all three deiodinases is the presence of a selenocysteine (Sec) residue in the centre of the amino acid sequence. As in other selenoproteins, this Sec residue is encoded by a UGA triplet, which in mRNAs for non-selenoproteins functions as a translation stop codon. The translation of the UGA codon into Sec requires the presence of a particular stem-loop structure in the 3′-untranslated region of the mRNA, termed Sec-insertion sequence (SECIS) element, Sec-tRNA, and several cellular proteins, including SECIS-binding protein (SBP2). A bona fide SECIS element has been identified in the mRNA of all deiodinases [45].

D_1 is a membrane-bound enzyme expressed predominantly in liver, kidneys, and thyroid [45]. It catalyses the ORD and/or IRD of a variety of iodothyronine derivatives, although it is most effective in the ORD of reverse T_3. In the presence of dithiothreitol (DTT) as the cofactor, D_1 displays high K_m and V_{max} values. Hepatic D_1 is probably a major site for the production of plasma T_3 and clearance of plasma reverse T_3. D_1 activity in liver and kidney is increased in hyperthyroidism and decreased in hypothyroidism, representing the regulation of D_1 activity by T_3 at the transcriptional level.

The Sec residue is essential for the function of D_1 since substitution with Cys reduces enzyme activity to 1%, while substitution with Leu yields a completely inactive protein. Rapid inactivation of D_1 by iodoacetate is probably due to modification of the highly reactive Sec residue. Thus, Sec is the catalytic centre of D_1.

The different deiodinases require thiols as cofactor. Although reduced glutathione is the most abundant intracellular thiol, its activity is very low compared with the unnatural thiol DTT, which is often used in *in vitro* studies. Alternative endogenous cofactors include dihydrolipoamide, glutaredoxin, and thioredoxin. D_1 shows ping-pong-type kinetics in catalysing the deiodination of iodothyronines by DTT. D_1 activity is potently inhibited by propylthiouracil, and this inhibition is uncompetitive with substrate and competitive with cofactor. Together, these findings suggest that the catalytic mechanism of D_1 involves the transfer of an iodonium ion (I^+) from the substrate to the selenolate (Se^-) group of the enzyme, generating a selenenyl iodide intermediate which is reduced back to native enzyme by thiols such as DTT or converted into a dead-end complex by propylthiouracil.

D_2 is expressed primarily in brain, anterior pituitary, brown adipose tissue, thyroid, and to some extent also in skeletal muscle [45, 46]. In brain tissue, D_2 mRNA has been localized in astrocytes, in particular also in tanycytes lining the third ventricle in the arcuate

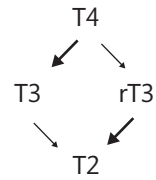
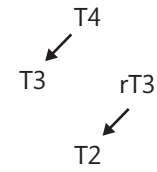
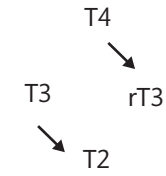
Type	D1	D2	D3
			
Tissues, e.g.	liver, kidney, thyroid	Brain, pituitary, BAT, thyroid, skeletal muscle	Brain, placenta, fetal tissues
Substrates	rT3 >> T4 = T3	T4 > rT3	T3 > T4
Km	~0,1–10uM	~1 nM	~1 nM
Effects of TH	↑	↓	↑
Function	rT3 clearance plasma T3 production	Local T3 production	T3 and T4 clearance

Figure 3.1.2.8 Properties of the three iodothyronine deiodinases.

nucleus–median eminence region. D_2 is a low-K_m, low-capacity enzyme possessing only ORD activity, with a preference for T_4 over reverse T_3 as the substrate. The amount of T_3 in brain, pituitary, and brown adipose tissue is derived to a large extent from local conversion of T_4 by D_2 and to a minor extent from plasma T_3 [24, 45]. The enzyme located in the anterior pituitary and the arcuate nucleus of the hypothalamus appears very important for the negative feedback regulation of TSH and TRH secretion [1].

In general, D_2 activity is increased in hypothyroidism and decreased in hyperthyroidism. This is explained in part by substrate-induced inactivation of the enzyme by T_4 and reverse T_3 involving the ubiquitin-proteasome system [45]. However, inhibition of D_2 activity and mRNA levels by T_3 has also been demonstrated in the brain and pituitary. The substrate (T_4, reverse T_3) and product (T_3)-dependent downregulation of D_2 activity is important to maintain brain T_3 levels in the face of changing plasma thyroid hormone levels.

In mammals, D_2 mRNA contains a second UGA codon just upstream of a UAA stop codon [45]. It remains to be determined to what extent this second TGA codon specifies the incorporation of a second Sec residue or acts as a translation stop codon. The amino acid sequence downstream of this second Sec is not required for enzyme activity.

D_3 activity has been detected in different human tissues, brain, skin, liver, and intestine, where activities are much higher in the fetal stage than in the adult stage [45]. D_3 is also abundantly expressed in placenta and pregnant uterus. D_3 has only IRD activity, catalysing the inactivation of T_4 and T_3 with intermediate K_m and V_{max} values. D_3 in tissues such as the brain is thought to play a role in the regulation of intracellular T_3 levels, while its presence in placenta, pregnant uterus, and fetal tissues may serve to protect developing organs against undue exposure to active thyroid hormone. Indeed, fetal plasma contains low T_3 (and high reverse T_3) concentrations. However, local D_2-mediated T_3 production from T_4 is crucial for brain development. Also in adult subjects, D_3 appears to be an important site for clearance of plasma T_3 and production of plasma reverse T_3. In brain, but not in placenta, D_3 activity is increased in hyperthyroidism and decreased in hypothyroidism, which at least in brain is associated with parallel changes in D_3 mRNA levels [24, 45].

In contrast to the marked decrease in hepatic and renal (but not thyroidal) D_1 activities, there are only minor effects of selenium deficiency on tissue D_2 and D_3 activities [47]. This may be explained by findings that the selenium state of different tissues varies greatly in selenium-deficient animals. In addition, the efficiency of the SECIS element to facilitate read-through of the UGA codon may differ among selenoproteins, which could result in the preferred incorporation of Sec into D_2 or D_3 over other selenoproteins.

The presence of Sec in a strongly conserved region of the proteins suggests the same catalytic mechanism for the different deiodinases. However, D_2 and D_3 are much less susceptible than D_1 to the mechanism-based inhibitors propylthiouracil, iodoacetate, and gold thioglucose [45]. This could be explained if the reactivity of the selenol group in D_2 and D_3 is much lower than that in D_1. Indeed, substitution of Sec in D_1 with the much less reactive Cys is associated with a dramatic decrease in its sensitivity to inhibition by gold thioglucose and propylthiouracil. Interestingly, the amino acid two positions downstream of the catalytic Sec residue (Ser in D_1, Pro in D_2 and D_3) plays an important role in determining the reactivity of the catalytic Sec residue [45].

Mutations in SBP2 cause a multisystem disorder including abnormalities in deiodinase activity [48]. Patients present growth and developmental delay, myopathy, infertility, and metabolic abnormalities. Some of these features are caused by increased accumulation of reactive oxygen species as a result from deficiencies of antioxidant selenoenzymes such as glutathione peroxidase. Thyroid function tests typically show normal TSH concentrations, elevated fT_4 and rT_3, accompanied by decreased T_3 concentrations. The exact mechanisms underlying the abnormal thyroid parameters are not well understood, but the elevated T_4 and rT_3 levels over T_3 levels indicate that ORD is predominantly affected.

Alanine Side Chain Modification

Intriguing metabolites are generated by side chain metabolism of iodothyronines (Figure 3.1.2.9). By action of decarboxylases such as ornithine decarboxylase (ODC) iodothyronines can be converted into iodothyronamines. In particular two of these, 3-iodothyronamine (T_1AM) and thyronamine (T_0AM), can exert acute and dramatic effects on heart rate, body temperature, and physical activity, inducing a torpor-like state [49]. Thus, these thyroid hormone metabolites appear to have neurotransmitter-like properties, adding a novel dimension to the already diverse effects of the conventional thyroid hormone structures.

The iodothyroacetic acid metabolites 3,3′,5,5′-tetraiodothyroacetic acid (Tetrac) and 3,3′,5-triiodothyroacetic acid (Triac) are generated from T_4 and T_3, respectively presumably by further conversion of iodothyronamines by the monoamine oxidases MAO-A or MAO-B, the (Figure 3.1.2.9) [50]. An alternative, iodothyronines can be converted to their aceto-acidic acid derivatives via an α-keto acid intermediate by the α-aminoadipate aminotransferase/kynurenine aminotransferase II (AADT/KAT II or AADAT), which transaminates halogenated tyrosines [51]. Although, in general, oxidative deamination is an inactivating pathway for monoamines, Triac has significant thyromimetic activity and its affinity for the T_3 receptor TRα1 is equal to that of T_3 and for the TRβ receptor it is even higher that of T_3 (see next section). Triac has a role in selective cases of resistance to thyroid hormone due to mutations in TRβ [52]. Preclinical studies suggest the potential of Triac to improve the neurological phenotype in MCT8 deficiency [53]. There may be multiple pathways leading from T_4 and T_3 to T_1AM and T_0AM with different orders for the successive decarboxylation and deiodination steps. Iodothyronamines are deiodinated by the different deiodinases [54], but it is unknown which iodothyronines are substrates for different decarboxylases or which iodothyronamines are converted by MAO-A or MAO-B. Also, the exact biological functions of the iodothyronamine and iodothyroacetic acid metabolites remain to be established.

Sulphation

In addition to deiodination, iodothyronines are metabolized by conjugation of the phenolic hydroxyl group with sulphate or glucuronic acid (Figure 3.1.2.9). Sulphation and glucuronidation are so-called phase II detoxification reactions, which increase the water solubility of substrates and, thus, facilitate their biliary and/or urinary clearance. However, iodothyronine sulphate levels are normally very low in plasma, bile, and urine, as these conjugates

are rapidly degraded by D_1, suggesting that sulphate conjugation is a primary step leading to the irreversible inactivation of thyroid hormone [55, 56]. Thus, the IRD of T_4 sulphate to reverse T_3 sulphate and of T_3 sulphate to T_2 sulphate is orders of magnitude faster than the IRD of non-sulphated T_4 and T_3, whereas the ORD of T_4 sulphate to T_3 sulphate is completely blocked. Plasma levels (and biliary excretion) of iodothyronine sulphates are increased if D_1 activity is inhibited by drugs such as propylthiouracil, and during fetal development, non-thyroidal illness, and fasting. Under these conditions, T_3 sulphate may function as a reservoir of inactive hormone from which active T_3 may be recovered by action of tissue sulfatases and bacterial sulfatases in the intestine.

Sulfotransferases represent a family of enzymes with a monomer molecular weight of approximately 34 kDa, located in the cytoplasm of different tissues, in particular liver, kidney, intestine, and brain. They catalyse the transfer of sulphate from 3′-phosphoadenosine-5′-phosphosulfate to usually a hydroxyl group of the substrate. Different phenol sulfotransferases have been identified with significant activity towards iodothyronines. These include human SULT1A1, 1A2, 1A3, 1B1, and 1C2 [55]. They have a large substrate preference for T_2, which is sulphated orders of magnitude faster than T_3 or reverse T_3, whereas sulphation of T_4 is hardly detectable.

Surprisingly, human oestrogen sulfotransferase (SULT1E1) is an important isoenzyme for sulphation of thyroid hormone. Although human SULT1E1 shows much greater affinity for oestrogens (K_m *c.*nM) than for iodothyronines (K_m *c.*μM), it sulphates T_2 and T_3 as efficiently as other SULTs, and is much more efficient in sulphating reverse T_3 and T_4 [55]. Human tissues expressing SULT1E1 include liver, uterus, and mammary gland [57]. In particular, the enzyme expressed in the endometrium may be a significant source of the high levels of iodothyronine sulphates in human fetal plasma. Different human SULTs have also been shown to catalyse the sulphation of iodothyronamines [58].

Glucuronidation

In contrast to the sulphates, iodothyronine glucuronides are rapidly excreted in the bile. However, this is not an irreversible pathway of hormone disposal since, after hydrolysis of the glucuronides by bacterial β-glucuronidases in the intestine, part of the liberated iodothyronines is reabsorbed, constituting an enterohepatic cycle [56, 59]. Nevertheless, about 20% of daily T_4 production appears in the faeces, probably through biliary excretion of glucuronide conjugates. Glucuronidation is catalysed by UDP-glucanosyltransferases (UGTs) that utilize UDP-glucuronic acid as cofactor. UGTs are localized in the endoplasmic reticulum of predominantly liver, kidney, and intestine. Most UGTs are members of the UGT1A and UGT2B families [60].

Glucuronidation of T_4 and T_3 is catalysed by different members of the UGT1A family, 1A1, 1A3, and 1A7–10. Usually, this involves the glucuronidation of the hydroxyl group (Figure 3.1.2.9), but human UGT1A3 also catalyses the glucuronidation of the side chain carboxyl group, with formation of so-called acyl glucuronides [61]. Interestingly, Tetrac and Triac are much more rapidly glucuronidated in human liver than T_4 and T_3, and this occurs predominantly by acyl glucuronidation [62].

In rodents, metabolism of thyroid hormone is accelerated through induction of T_4-glucuronidating UGTs by different classes of compounds, including barbiturates, fibrates, and polychlorinated biphenyls [63, 64]. This may result in a hypothyroid state as the thyroid gland is not capable of compensating for the increased hormone loss. In humans, thyroid function may be affected by induction of T_4 glucuronidation by antiepileptics, but overt hypothyroidism is rare [65]. Administration of such drugs to T_4-replaced hypothyroid patients may necessitate an increase in the T_4 substitution dose.

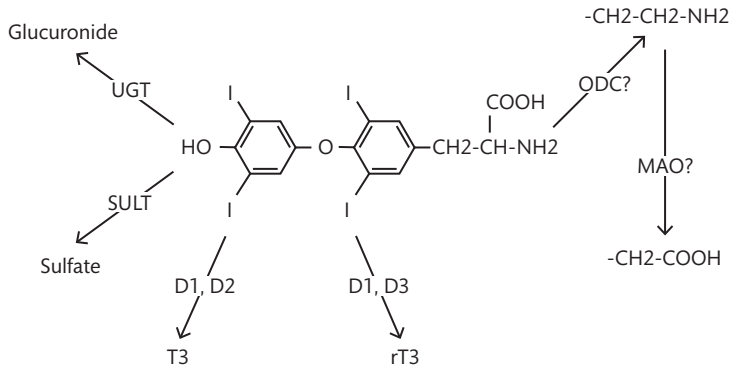


Figure 3.1.2.9 Pathways of thyroid hormone metabolism.

Thyroid Hormone Actions

Role of Thyroid Hormone in Thermogenesis

Thyroid hormone is critical for the development of different tissues, in particular the brain, but it is also essential for an optimal function of most tissues in adult life [66]. It is probably the most important factor regulating thermogenesis, as reflected by the increase in the basal metabolic rate in hyperthyroid subjects and the decrease observed in hypothyroid individuals [67–69]. Thyroid hormone increases the synthesis as well the degradation of proteins, lipids, and carbohydrates, predominantly by stimulating the expression of key enzymes involved in these processes. Examples of these are the lipogenic enzymes, malic enzyme, fatty acid synthase, and glucose-6-phosphate dehydrogenase, and the gluconeogenic enzyme phosphoenolpyruvate carboxykinase.

Special forms of substrate cycling take place between the cytoplasm and the mitochondrion, such as the glycerol-3-phosphate/dihydroxyacetone phosphate shuttle in which cytoplasmic and mitochondrial α-glycerophosphate dehydrogenase (αGPD) isoenzymes participate [67, 68]. This represents one way to enable oxidation of cytoplasmic NADH in the mitochondrion, which is impermeable to this cofactor. Thyroid hormone stimulates the expression of mitochondrial αGPD, and the increased electron flow via this enzyme is associated with an increased heat production relative to adenosine triphosphate (ATP) synthesis.

Thyroid hormone also increases the activity of Na^+,K^+-ATPase, an enzyme located in the plasma membrane of all tissues, in particular kidney, heart, and skeletal muscle, which is responsible for the maintenance of the Na^+ and K^+ gradients across this membrane. In myocytes, the increased Na^+,K^+-ATPase activity accelerates the repolarization of the sarcolemma following a depolarization stimulus that contributes to the tachycardia induced by thyroid hormone.

Another important target for thyroid hormone action is the Ca^{2+}-ATPase located in the sarcoplasmic reticulum of muscle cells [69]. Innervation of the myocyte triggers the release of large amounts of

Ca^{2+} from the sarcoplasmic reticulum into the cytoplasm, where it binds to the actomyosin complex that initiates contraction. There are two Ca^{2+}-ATPase isoenzymes, SERCA1 that is characteristic for fast-type skeletal muscle and SERCA2 that is characteristic for slow-type skeletal muscle and heart. T_3 increases Ca^{2+}-ATPase activity by stimulating the transcription of both *SERCA1* and *SERCA2* genes, which explains the increased relaxation rate of the muscle induced by T_3 [69].

It has been estimated that excess Ca^{2+} cycling in contracting muscle may account for up to 50% of the T_3-dependent energy expenditure during work or shivering [69]. The remainder of the T_3-induced energy turnover in contracting muscle is largely accounted for by the change in the expression of two forms of the myosin heavy chain which are characterized by high (MHCα) and low (MHCβ) ATPase activities and contraction rates. T_3 stimulates the expression of the *MHC*α gene, whereas it inhibits the expression of the *MHC*β gene [69]. A similar T_3-induced shift in MHC expression is also observed in the heart [70].

In addition, T_3 increases the expression of the uncoupling protein UCP1 in brown adipose tissue (BAT) [67, 68]. This is an important mechanism by which T_3 stimulates non-shivering cold-induced thermogenesis. UCP1 is an ion transporter located in the inner mitochondrial membrane which dissipates the proton gradient over this membrane generated by the respiratory chain, producing heat instead of ATP. Significant BAT depots have been demonstrated in the neck and shoulder region of normal adults, especially in cold-adapted subjects and more so in younger females than in older males [71]. Cold exposure leads to a dramatic stimulation of D_2 expression in BAT, and the resultant induction of local T_3 production plays an important role in the stimulation of BAT activity. This includes increased mobilization and burning of lipids as well as stimulated UCP1 expression, together resulting in a major increase in heat production [67, 68].

UCP1 is expressed exclusively in BAT. Other members of the UCP family are expressed in other human tissues, including UCP2 in a variety of tissues including heart and skeletal muscle, UCP3 in skeletal muscle, and UCP4 and UCP5 in brain. The expression of UCP2 and UCP3 is also under positive control of thyroid hormone, but their role in T_3-induced thermogenesis has not been established [72].

The regulation of the mitochondrial proteins UCP1 and αGPD by thyroid hormone is mediated predominantly by interaction of the nuclear T_3 receptor with the promoters of these genes [67, 72]. However, there is also evidence for direct effects of thyroid hormone on the mitochondria, the mechanism of which is incompletely understood but may involve interaction of T_3 and other iodothyronines such as 3,3'-T_2 and 3,5-T_2 with cytochrome *c* oxidase [73]. Many studies have reported effects of thyroid hormone on cellular processes that are not mediated by the nuclear T_3 receptor, including stimulation of transport of glucose, amino acids, and ions over the cell membrane, stimulation of actin polymerization in neurons, and stimulation of mitogen-activated protein kinase activity. The last is mediated by the binding of iodothyronines to integrin, a plasma membrane receptor. The interested reader is referred to an extensive review of these extranuclear actions of thyroid hormone [74].

Specific thyroid hormone-binding sites have also been detected in the cytoplasm in different tissues. A notable example is the NADPH-dependent cytoplasmic thyroid hormone-binding protein present in rat liver, which appears to be important for the trafficking of thyroid hormone to the nucleus or mitochondria [75].

Mechanism of T3 Action

Most biological actions of T_3 are initiated by its binding to nuclear T_3 receptors [76–78]. These proteins are members of the superfamily of ligand-dependent transcription factors, which also includes the receptors for steroids (e.g. cortisol, oestradiol, and testosterone), 1,25-dihydroxyvitamin D_3, retinoic acid, and 9-*cis*-retinoic acid. The last, so-called retinoid X receptor (RXR) is an important member of this gene family, because it forms functional heterodimers with a number of other nuclear receptors, including T_3 receptors. Two T_3 receptor genes have been identified; the *THRA* gene encoding TRα is located on human chromosome 17 and the *THRB* gene encoding TRβ on human chromosome 3. By alternative exon utilization of both genes, four major receptor isoforms, TRα1, TRα2, TRβ1, and TRβ2, are generated, which consist of 410–514 amino acids (Figure 3.1.2.10). Although the *THRB* gene (150 kb) is much larger than the *THRA* gene (*c.*30 kb), they have similar genomic structures, comprising 10 (*THRB*) or 11 (*THRA*) exons, and their coding sequences show a high degree of homology [76–78].

As in the other members of the nuclear receptor family, functional key domains have been recognized in the T_3 receptors, in particular the DNA-binding domain (DBD), which is approximately 100 amino acids long, and the ligand-binding domain (LBD), which

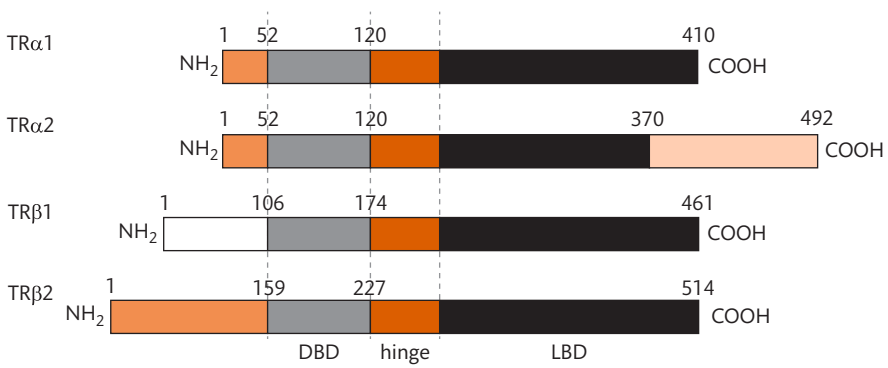


Figure 3.1.2.10 Domain structures of the different T_3 receptor (TR) isoforms. The TRα2 variant is incapable of binding T_3. DBD, DNA-binding domain; LBD ligand-binding domain.

is approximately 250 amino acids in length [76–78]. The amino acid sequences of the TRα and TRβ subtypes are most homologous in their DBD and LBD and least homologous at their N-terminus. The latter contains the ligand-independent AF1 transactivation domain, while an AF2 domain necessary for homo- and heterodimerization and ligand-dependent activation is located at the C-terminus. The short sequence between the DBD and the LBD is usually referred to as the hinge region.

The structural difference between TRα1 and TRα2 is located at the C-terminus of the proteins, where the sequences of the last 40 amino acids in TRα1 and 122 amino acids in TRα2 differ completely due to alternative splicing. The alteration in the LBD of TRα2 is associated with a complete loss of T_3 binding. Therefore, this splice variant is not a *bona fide* T_3 receptor, but for convenience it will still be referred to here as TRα2. TRα2 has a weak negative effect on the action of T_3 through the other T_3 receptors. The N-terminal domains of TRβ1 (106 amino acids) and TRβ2 (159 amino acids) differ almost completely due to utilization of alternative transcription start sites. Apparently, this domain provides TRβ2 with specific properties required for T_3-induced downregulation of *TRH* and *TSH* genes [76–78].

The high homology between the LBDs of TRα1 and TRβ explains their very similar ligand specificity, with affinities much higher for T_3 than for T_4. However, the metabolite Triac also binds to the T_3 receptors with an affinity equal to (TRα1) or even greater than (TRβ1) that of T_3 [79]. Nevertheless, T_3 is the major endogenous iodothyronine occupying the nuclear thyroid hormone receptors, which are thus true T_3 receptors. Several TRβ-specific agonists have been developed with pharmacologically interesting and selective effects on the liver, resulting in lowering of body weight, lipid, and cholesterol without detrimental effects on the heart [78, 79]. Most likely, the tissue-specific effects of these compounds are not only determined by their affinity for the T_3 receptor isoforms but also by the diverse ligand-preference of thyroid hormone transporters in different tissues. Interestingly, non-selective T_3 receptor antagonists have been developed as well [78, 79].

The different T_3 receptor isoforms show distinct tissue distributions [76–78]. The TRα1 is the predominant T_3 receptor expressed in brain, heart, and bone, whereas TRβ1 is the major receptor in other tissues, including liver, skeletal muscle, kidney, and fat. TRβ2 is preferentially expressed in the anterior pituitary and the hypothalamic area of the brain. These locations suggest the particular involvement of TRβ2 in the feedback inhibition of TSH and TRH secretion by thyroid hormone. Exon utilization specifying TRβ2 expression in the anterior pituitary is under the control of pituitary-specific transcription factor 1, response elements for which are located in the TRβ gene promoter [80]. Regulation of the expression of T_3-responsive genes involves the binding of the T_3 receptors to so-called T_3 response elements (TREs) in the promoter region of these genes [76–78]. TREs usually consist of two half-sites arranged as repeats or palindromes. The most prevalent TRE half-site sequence is AGGTCA, and the direct repeat of this half-site spaced by four nucleotides (DR4) is a particularly powerful TRE. However, some TREs show marked deviation from this 'consensus' half-site sequence, which, moreover, is also recognized by other receptors such as RXR and the retinoic acid receptor. This may be the basis for 'cross-talk' between different nuclear receptors and their target genes. Although T_3 receptors may bind as homodimers to the TREs, T_3 effects on gene expression are usually mediated by T_3 receptor/RXR heterodimers.

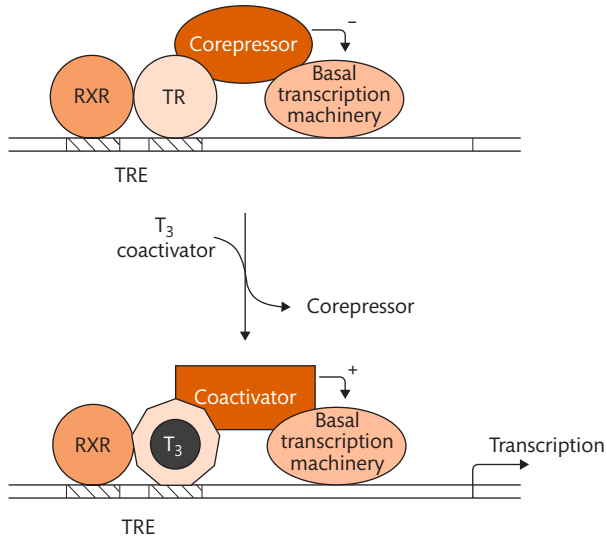


Figure 3.1.2.11 Simplistic model of the regulation of gene transcription by T_3. RXR, retinoid X receptor; TR, T_3 receptor; TRE, T_3 response element in the promoter of a T_3-responsive gene.

Binding of the T_3 receptor/RXR heterodimer to TRE does not require T_3 or 9-*cis*-retinoic acid, the ligand for RXR. The DBDs of these (and other) nuclear receptors contain two 'zinc fingers' (peptide loops that chelate a zinc atom) that fit in the grooves of the DNA and are, thus, very important for the specificity of the receptor-promoter interaction [76–78]. In the absence of T_3 and irrespective of the presence of 9-*cis*-retinoic acid, binding of the T_3 receptor/RXR heterodimer to the TRE results in suppressed gene transcription mediated by the binding of corepressor proteins such as NCoR (nuclear corepressor) or SMRT (silencing mediator of retinoid and thyroid hormone receptors) to a specific region (CoR box) of the unliganded T_3 receptor (Figure 3.1.2.11). These corepressors directly or indirectly inhibit the activity of the basal transcription machinery.

Binding of T_3 induces a conformational change in the T_3 receptor, which results in the release of the corepressors and the recruitment of coactivator proteins such as SRC1 (steroid receptor coactivator-1) and CBP (cAMP response element-binding protein (CREB)-binding protein) [76–78]. The AF2 domain, a highly conserved 9-amino acid sequence located at the C-terminus of the different nuclear receptors, plays an important role in the binding of the coactivators. The latter directly or indirectly stimulate the activity of the basal transcription machinery. One mechanism by which transcription is stimulated involves the histone acetyltransferase activity of the coactivators or of other proteins with which they interact. Acetylation of histones loosens the chromatin structure and thus facilitates interaction of the transcription machinery with the DNA. Conversely, corepressors may recruit proteins with deacetylase activity.

T_3 Inhibition of TSH and TRH Gene Expression

The earlier discussion of the mechanism of action of T_3 concerns the expression of genes which are under positive control of thyroid hormone. However, a roughly equal number of genes are negatively regulated by T_3, in particular those involved in the negative feedback regulation of the hypothalamus–pituitary–thyroid axis (i.e. the *TSH*β and the *TRH* genes). In the promoter regions of these genes

negative TREs have been identified that often consist of only one half-site. In the *TSH*β gene such a negative TRE has been found in close proximity to the AP-1 site which mediates the stimulation of *TSH*β gene transcription by TRH. As just mentioned, there appears to be a specific role for TRβ2 in the regulation of the negative TREs in the *TSH*β and *TRH* genes [76–78]. In contrast to gene regulation through positive TREs, binding of TRβ2 to negative TREs in the absence of T_3 probably results in the activation of gene transcription. In the presence of T_3, transcription is inhibited. The exact mechanism of this negative regulation of gene expression by T_3 and any T_3 receptor is still unclear.

*TSH*β gene transcription is also strongly inhibited by 9-*cis*-retinoic acid, and this effect is mediated by the pituitary-specific RXRγ1 subtype, and involves both TRE-dependent and TRE-independent interactions with the *TSH*β gene promoter. The clinical relevance of this effect is underscored by a recent study showing that treatment of patients with T-cell lymphoma with bexarotene, another RXR-selective ligand, induces central hypothyroidism [81]. It is also interesting to mention that the *TRH* gene promoter contains a glucocorticoid response element. Hypothalamic TRH-producing cells also express the glucocorticoid receptor, and the interaction of this receptor with its response element appears to mediate the inhibition of TRH synthesis by glucocorticoids [82].

In addition to the regulation of TSHβ and α-subunit gene expression, T_3 also acutely inhibits TSH secretion, the exact mechanism of which is still unresolved. Although T_3 is the active hormone exerting the inhibition of TSH production and secretion, serum T_4 appears to be a major player in the negative feedback regulation of the hypothalamus–pituitary–thyroid axis by acting as a precursor for local D_2-mediated generation of T_3 at these central sites [45, 83].

Recent research in two particular areas has led to important advances in our understanding of the mechanism of action of T_3. One type of study has utilized T_3 receptor knockout and mutant mice in which one or more of the different T_3 receptor isoforms is deleted or mutated [78]. These studies reveal which organ functions critically depend on the type of T_3 receptors they express. Similarly, mice mutant for NCoR or SMRT have revealed important insights in the function of corepressors in thyroid hormone signalling [84]. Much knowledge regarding the molecular mechanisms of T_3 receptor/T_3 action has also been gained from studies in patients with thyroid hormone resistance associated with mutations in TRβ (resistance to thyroid hormone; RTH). In recent years, patients with mutations in TRα have been identified. For a thorough discussion of RTHα and RTHβ, the reader is referred to Chapter 3.4.5.

REFERENCES

1. Fliers E, Alkemade A, Wiersinga WM, Swaab DF. Hypothalamic thyroid hormone feedback in health and disease. *Prog Brain Res*, 2006; **153**: 189–207.
2. Perello M, Nillni EA. The biosynthesis and processing of neuropeptides: lessons from prothyrotropin releasing hormone (proTRH). *Front Biosci*, 2007; **12**: 3554–65.
3. Sun Y, Lu X, Gershengorn MC. Thyrotropin-releasing hormone receptors: similarities and differences. *J Mol Endocrinol*, 2003; **30**: 87–97.
4. Heuer H, Schafer MK, Bauer K. The thyrotropin-releasing hormone-degrading ectoenzyme: the third element of the thyrotropin-releasing hormone-signaling system. *Thyroid*, 1998; **8**: 915–20.
5. Grossmann M, Weintraub BD, Szkudlinski MW. Novel insights into the molecular mechanisms of human thyrotropin action: structural, physiological, and therapeutic implications for the glycoprotein hormone family. *Endocr Rev*, 1997; **18**: 476–501.
6. Mariotti S. Normal physiology of the hypothalamic-pituitary-thyroidal system and relation to the neural system and other endocrine glands. Available at: https://www.thyroidmanager.org (accessed: 20 May 2010). South Dartmouth, MA: Endocrine Education Inc.
7. Kleinau G, Krause G. Thyrotropin and homologous glycoprotein hormone receptors: structural and functional aspects of extracellular signaling mechanisms. *Endocr Rev*, 2009; **30**: 133–51.
8. Dumont JE, Opitz R, Christophe D, Vassart G, Roger PP, Maenhaut C. The phylogeny, ontogeny, anatomy and regulation of the iodine metabolizing thyroid. Available at: https://www.thyroidmanager.org (accessed: 20 May 2010). South Dartmouth, MA: Endocrine Education Inc.
9. Davies TF, Ando T, Lin RY, Tomer Y, Latif R. Thyrotropin receptor-associated diseases: from adenomata to Graves' disease. *J Clin Invest*, 2005; **115**: 1972–83.
10. Kopp P. Thyroid hormone synthesis. In: Braverman LE, Utiger RD, eds. *Werner & Ingbar's The Thyroid*. Philadelphia., PA: Lippincott Williams & Wilkins, 2005: 52–77.
11. Song Y, Driessens N, Costa M, *et al*. Roles of hydrogen peroxide in thyroid physiology and disease. *J Clin Endocrinol Metab*, 2007; **92**: 3764–73.
12. Dohan O, De la Vieja A, Paroder V, *et al*. The sodium/iodide symporter (NIS): characterization, regulation, and medical significance. *Endocr Rev*, 2003; **24**: 48–77.
13. Kopp P, Pesce L, Solis SJ. Pendred syndrome and iodide transport in the thyroid. *Trends Endocrinol Metab*, 2008; **19**: 260–8.
14. Wémeau JL, Kopp P. Pendred syndrome. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 213–24.
15. Rivolta CM, Targovnik HM. Molecular advances in thyroglobulin disorders. *Clin Chim Acta*, 2006; **374**: 8–24.
16. Moreno JC, Visser TJ. New phenotypes in thyroid dyshormonogenesis: hypothyroidism due to DUOX2 mutations. *Endocr Dev*, 2007; **10**: 99–117.
17. Grasberger H, Refetoff S. Identification of the maturation factor for dual oxidase. Evolution of an eukaryotic operon equivalent. *J Biol Chem*, 2006; **281**: 18269–72.
18. Taurog A. Hormone synthesis: thyroid iodine metabolism. In: Braverman LE, Utiger RD, eds. *Werner & Ingbar's The Thyroid*. Philadelphia, PA: Lippincott Williams & Wilkins, 2000: 61–85.
19. Lisi S, Pinchera A, McCluskey RT, *et al*. Preferential megalin-mediated transcytosis of low-hormonogenic thyroglobulin: a control mechanism for thyroid hormone release. *Proc Natl Acad Sci U S A*, 2003; **100**: 14858–63.
20. Friedrichs B, Tepel C, Reinheckel T, *et al*. Thyroid functions of mouse cathepsins B, K, and L. *J Clin Invest*, 2003; **111**: 1733–45.
21. Andersson HC, Kohn LD, Bernardini I, Blom HJ, Tietze F, Gahl WA. Characterization of lysosomal monoiodotyrosine transport in rat thyroid cells. Evidence for transport by system h. *J Biol Chem*, 1990; **265**: 10950–4.
22. Gnidehou S, Caillou B, Talbot M, *et al*. Iodotyrosine dehalogenase 1 (DEHAL1) is a transmembrane protein involved

in the recycling of iodide close to the thyroglobulin iodination site. *FASEB J*, 2004; **18**: 1574–6.
23. Moreno JC, Klootwijk W, van Toor H, *et al*. Mutations in the iodotyrosine deiodinase gene and hypothyroidism. *N Engl J Med*, 2008; **358**: 1811–18.
24. Bianco AC, Larsen PR. Intracellular pathways of iodothyronine metabolism. In: Braverman LE, Utiger RD, eds. *Werner & Ingbar's The Thyroid*. Philadelphia, PA: Lippincott Williams & Wilkins, 2005: 109–35.
25. Astwood EB. Landmark article 8 May (1943): treatment of hyperthyroidism with thiourea and thiouracil. *JAMA*, 1984; **251**: 1743–6.
26. Benvenga S. Thyroid hormone transport proteins and the physiology of hormone binding. In: Braverman LE, Utiger RD, eds. *Werner & Ingbar's The Thyroid*. Philadelphia, PA: Lippincott Williams & Wilkins, 2005: 97–109.
27. Refetoff S, Murata Y, Mori Y, Janssen OE, Takeda K, Hayashi Y. Thyroxine-binding globulin: organization of the gene and variants. *Horm Res*, 1996; **45**: 128–38.
28. Richardson SJ. Cell and molecular biology of transthyretin and thyroid hormones. *Int Rev Cytol*, 2007; **258**: 137–93.
29. Saraiva MJ. Transthyretin mutations in hyperthyroxinemia and amyloid diseases. *Hum Mutat*, 2001; **17**: 493–503.
30. Petitpas I, Petersen CE, Ha CE, *et al*. Structural basis of albumin-thyroxine interactions and familial dysalbuminemic hyperthyroxinemia. *Proc Natl Acad Sci U S A*, 2003; **100**: 6440–5.
31. Hennemann G, Docter R, Friesema EC, de Jong M, Krenning EP, Visser TJ. Plasma membrane transport of thyroid hormones and its role in thyroid hormone metabolism and bioavailability. *Endocr Rev*, 2001; **22**: 451–76.
32. Blondeau JP, Beslin A, Chantoux F, Francon J. Triiodothyronine is a high-affinity inhibitor of amino acid transport system L1 in cultured astrocytes. *J Neurochem*, 1993; **60**: 1407–13.
33. Zhou Y, Samson M, Francon J, Blondeau JP. Thyroid hormone concentrative uptake in rat erythrocytes. Involvement of the tryptophan transport system T in countertransport of tri-iodothyronine and aromatic amino acids. *Biochem J*, 1992; **281**: 81–6.
34. Friesema EC, Jansen J, Milici C, Visser TJ. Thyroid hormone transporters. *Vitam Horm*, 2005; **70**: 137–67.
35. Visser WE, Friesema EC, Jansen J, Visser TJ. Thyroid hormone transport in and out of cells. *Trends Endocrinol Metab*, 2008; **19**: 50–6.
36. Geyer J, Wilke T, Petzinger E. The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn Schmiedebergs Arch Pharmacol*, 2006; **372**: 413–31.
37. Hagenbuch B. Cellular entry of thyroid hormones by organic anion transporting polypeptides. *Best Pract Res*, 2007; **21**: 209–21.
38. van der Deure W, Peeters R, Visser T. Molecular aspects of thyroid hormone transporters, including MCT8, MCT10 and OATPs, and the effects of genetic variation in these transporters. *J Mol Endocrinol*, 2010; **44**: 1–11.
39. Strømme P, Groeneweg S, de Souza EL, *et al*. Mutated thyroid hormone transporter OATP1C1 associates with severe brain hypometabolism and juvenile neurodegeneration. *Thyroid*, 2018: DOI: 10.1089/thy.2018.0595 [Epub ahead of print].
40. Taylor PM, Ritchie JW. Tissue uptake of thyroid hormone by amino acid transporters. *Best Pract Res*, 2007; **21**: 237–51.
41. Halestrap AP, Meredith D. The *SLC16* gene family-from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. *Pflugers Arch*, 2004; **447**: 619–28.
42. Meredith D, Christian HC. The *SLC16* monocarboxylate transporter family. *Xenobiotica*, 2008; **38**: 1072–106.
43. Heuer H, Visser TJ. Minireview: pathophysiological importance of thyroid hormone transporters. *Endocrinology*, 2009; **150**: 1078–83.
44. Visser WE, Friesema EC, Jansen J, Visser TJ. Thyroid hormone transport by monocarboxylate transporters. *Best Pract Res*, 2007; **21**: 223–36.
45. Gereben B, Zavacki AM, Ribich S, *et al*. Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. *Endocr Rev*, 2008; **29**: 898–938.
46. Larsen PR. Type 2 iodothyronine deiodinase in human skeletal muscle: new insights into its physiological role and regulation. *J Clin Endocrinol Metab*, 2009; **94**: 1893–5.
47. Kohrle J. Selenium and the control of thyroid hormone metabolism. *Thyroid*, 2005; **15**: 841–53.
48. Schoenmakers E, Agostini M, Michell C, *et al*. Mutations in the selenocysteine insertion sequence-binding protein 2 gene lead to a multisystem selenoprotein deficiency disorder in humans. *J Clin Invest*, 2010; **120**: 4220–35
49. Scanlan TS. Minireview: 3-iodothyronamine (T1AM): a new player on the thyroid endocrine team? *Endocrinology*, 2009; **150**: 1108–11.
50. Wood WJ, Geraci T, Nilsen A, DeBarber AE, Scanlan TS. Iodothyronamines are oxidatively deaminated to iodothyroacetic acids in vivo. *Chembiochem*, 2009; **10**: 361–5.
51. Hallen A, Jamie JF, Cooper AJ. Lysine metabolism in mammalian brain: an update on the importance of recent discoveries. *Amino Acids*, 2013; **45**: 1249–72.
52. Groeneweg S, Peeters RP, Visser TJ, Visser WE. Therapeutic applications of thyroid hormone analogues in resistance to thyroid hormone (RTH) syndromes. *Mol Cell Endocrinol*, 2017; **458**: 82–90.
53. Kersseboom S, Horn S, Visser EW, *et al*. In vitro and mouse studies supporting therapeutic utility of triiodothyroacetic acid in MCT8 deficiency. *Mol Endocrinol*, 2014; **28**: 1961–70.
54. Piehl S, Heberer T, Balizs G, *et al*. Thyronamines are isozyme-specific substrates of deiodinases. *Endocrinology*, 2008; **149**: 3037–45.
55. Kester MHA, Visser TJ. Sulfation of thyroid hormones. In: Pacifici GM, Coughtrie MWH, eds. *Human Cytosolic Sulfotransferases*. Boca Raton, FL: CRC Press, 2005: 121–34.
56. Wu SY, Green WL, Huang WS, Hays MT, Chopra IJ. Alternate pathways of thyroid hormone metabolism. *Thyroid*, 2005; **15**: 943–58.
57. Song WC. Biochemistry and reproductive endocrinology of estrogen sulfotransferase. *Ann N Y Acad Sci*, 2001; **948**: 43–50.
58. Pietsch CA, Scanlan TS, Anderson RJ. Thyronamines are substrates for human liver sulfotransferases. *Endocrinology*, 2007; **148**: 1921–7.
59. Visser TJ. Hormone metabolism. Available at: https://www.thyroidmanager.org (accessed: 20 May 2010). South Dartmouth, MA: Endocrine Education Inc.
60. Mackenzie PI, Bock K, Burchell B, *et al*. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics*, 2005; **15**: 677–85.
61. Kato Y, Ikushiro S, Emi Y, *et al*. Hepatic UDP-glucuronosyltransferases responsible for glucuronidation of thyroxine in humans. *Drug Metab Dispos*, 2008; **36**: 51–5.
62. Moreno M, Kaptein E, Goglia F, Visser TJ. Rapid glucuronidation of tri- and tetraiodothyroacetic acid to ester glucuronides in human liver and to ether glucuronides in rat liver. *Endocrinology*, 1994; **135**: 1004–9.

63. Visser TJ, Kaptein E, Gijzel AL, de Herder WW, Ebner T, Burchell B. Glucuronidation of thyroid hormone by human bilirubin and phenol UDP-glucuronyltransferase isoenzymes. *FEBS Lett*, 1993; **324**: 358–60.
64. Hood A, Allen ML, Liu Y, Liu J, Klaassen CD. Induction of T(4) UDP-GT activity, serum thyroid stimulating hormone, and thyroid follicular cell proliferation in mice treated with microsomal enzyme inducers. *Toxicol Appl Pharmacol*, 2003; **188**: 6–13.
65. Benedetti MS, Whomsley R, Baltes E, Tonner F. Alteration of thyroid hormone homeostasis by antiepileptic drugs in humans: involvement of glucuronosyltransferase induction. *Eur J Clin Pharmacol*, 2005; **61**: 863–72.
66. Hulbert AJ. Thyroid hormones and their effects: a new perspective. *Biol Rev Camb Philos Soc*, 2000; **75**: 519–631.
67. Silva JE. Thyroid hormone control of thermogenesis and energy balance. *Thyroid*, 1995; **5**: 481–92.
68. Silva JE. Thermogenic mechanisms and their hormonal regulation. *Physiol Rev*, 2006; **86**: 435–64.
69. Simonides WS, van Hardeveld C. Thyroid hormone as a determinant of metabolic and contractile phenotype of skeletal muscle. *Thyroid*, 2008; **18**: 205–16.
70. Kahaly GJ, Dillmann WH. Thyroid hormone action in the heart. *Endocr Rev*, 2005; **26**: 704–28.
71. Celi FS. Brown adipose tissue: when it pays to be inefficient. *New Engl J Med*, 2009; **360**: 1553–6.
72. Lanni A, Moreno M, Lombardi A, Goglia F. Thyroid hormone and uncoupling proteins. *FEBS Lett*, 2003; **543**: 5–10.
73. Moreno M, de Lange P, Lombardi A, Silvestri E, Lanni A, Goglia F. Metabolic effects of thyroid hormone derivatives. *Thyroid*, 2008; **18**: 239–53.
74. Davis PJ, Leonard JL, Davis FB. Mechanisms of nongenomic actions of thyroid hormone. *Front Neuroendocrinol*, 2008; **29**: 211–18.
75. Suzuki S, Mori J, Hashizume K. mu-crystallin, a NADPH-dependent T(3)-binding protein in cytosol. *Trends Endocrinol Metab*, 2007; **18**: 286–9.
76. Yen PM. Physiological and molecular basis of thyroid hormone action. *Physiol Rev*, 2001; **81**: 1097–142.
77. Bassett JH, Harvey CB, Williams GR. Mechanisms of thyroid hormone receptor-specific nuclear and extra nuclear actions. *Mol Cell Endocrinol*, 2003; **213**: 1–11.
78. Flamant F, Gauthier K, Samarut J. Thyroid hormones signaling is getting more complex: STORMs are coming. *Mol Endocrinol*, 2007; **21**: 321–33.
79. Brenta G, Danzi S, Klein I. Potential therapeutic applications of thyroid hormone analogs. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 632–40.
80. Wood WM, Dowding JM, Bright TM, *et al.* Thyroid hormone receptor beta2 promoter activity in pituitary cells is regulated by Pit-1. *J Biol Chem*, 1996; **271**: 24213–20.
81. Sharma V, Hays WR, Wood WM, *et al.* Effects of rexinoids on thyrotrope function and the hypothalamic-pituitary-thyroid axis. *Endocrinology*, 2006; **147**: 1438–51.
82. Lee GC, Yang IM, Kim BJ, *et al.* Identification of glucocorticoid response element of the rat TRH gene. *Korean J Intern Med*, 1996; **11**: 138–44.
83. Bianco AC, Larsen PR. Intracellular pathways of iodothyronine metabolism. In: Braverman LE, Utiger. eds. *Werner and Ingbar's The Thyroid*. 9th edn. Philadephia, PA: Lippincott Williams & Wilkins, 2005: 109–35.
84. Vella KR, Hollenberg AN. The actions of thyroid hormone signaling in the nucleus. *Mol Cell Endocrinol*, 2017; **458**: 127–35.

3.1.3 Clinical Assessment of the Thyroid Patient

Inge Bülow Pedersen and Stig Andersen

Introduction

Thyroid disorders are common and their occurrence increase with age. The number of structural abnormalities rise with age reaching 20–40% in women [1, 2] and functional abnormalities are 20–30% in old people [3]. This makes evaluation of thyroid function, size, and structure an important part of any complete history taking and physical examination of a patient.

Deficient or excessive thyroid hormone secretion affects nearly all body systems, and examination of a patient with a proven or suspected thyroid abnormality should include a more general evaluation of the patient. For example, an episode of thyrotoxicosis in an old person may provoke atrial fibrillation and impair cardiac function. The abnormality may persist after treatment of the thyrotoxicosis, and there may be a need for supplementary therapy directed against the atrial fibrillation. Similarly, symptoms of arteriosclerotic heart disease may worsen after initiation of treatment in a patient with hypothyroidism. Both hypothyroidism and heart disease should be diagnosed to make an appropriate plan for therapy.

The three key abnormalities of the thyroid gland are: [1] thyrotoxicosis with excessive thyroid hormone that affects the body; [2] hypothyroidism with thyroid hormone deficiency; [3] and goitre with a general or focal abnormal enlargement of the thyroid gland. A fourth but less common abnormality is the painful thyroid. Examination of the thyroid patient should lead to a conclusion based on symptoms and signs related to these abnormalities.

Clinical symptoms and signs of hyper- and hypothyroidism are many and they are given and detailed in subsequent chapters. However, symptoms and signs of a clinical condition requiring more than usual observation or even acute therapy should be identified during the initial assessment. The risk of thyrotoxic crises should be included in the evaluation of a thyrotoxic patient. The risk of myxoedema coma is very low in the patient with hypothyroidism. Still, the condition certainly should not develop during the period of diagnostic work-up. **Box 3.1.3.1** lists some symptoms and signs that should raise attention to the development of a severe thyroid condition, including factors that signify risk of thyroid malignancy in a patient with goitre. The presence of these symptoms and signs guide the decision to accelerate further diagnostic evaluation.

A number of different thyroid diseases may cause various thyroid abnormalities. Any clinical finding that rise the suspicion of a thyroid functional or structural abnormality should initiate a systematic evaluation of the disease underlying the abnormality (nosological diagnosis). For example, if the patient seems to be thyrotoxic the examination should lead to a provisional conclusion on the

Box 3.1.3.1 Warning symptoms and signs in thyroid patients

- Untreated hyperthyroidism[a]
  - Fever
  - Diarrhoea
  - Severe tachycardia (resting pulse rate >110 beats/min)
  - Complication severe disease
  - Resting dyspnoea
- Untreated hypothyroidism[b]
  - Somnolence
  - Hypothermia
  - Complicating severe disease
- Goitre[c]
  - Hard solitary nodule
  - Growth of nodule
  - Stridor or hoarseness
  - Fixed to surroundings
  - Enlarged lymph nodes
  - Radiation to the neck as a child

[a] In untreated hyperthyroidism imminent thyrotoxic crisis should be looked for. Another severe complication is pulmonary embolism, in part due to dehydration.
[b] Somnolence and hypothermia may be warnings of myxoedema coma.
[c] Symptoms and signs indication a higher risk of malignancy in a goitre.

disease causing the thyrotoxicosis. The four most common causes of thyrotoxicosis are Graves' disease, multinodular toxic goitre, toxic adenoma, and subacute thyroiditis [4]. The diagnosis should be substantiated by further biochemical tests and imaging procedures.

Some of the diseases leading to thyroid abnormalities may have other manifestations. It is important to look out for these. A common example is the orbitopathy that may present even prior to abnormal thyroid function, and some may not develop thyroid dysfunction despite manifest orbitopathy. The pretibial myxoedema of Graves' disease is a rare example, as is the retroperitoneal fibrosis with ureteral obstruction encountered in some patients with Riedel's thyroiditis.

The history and clinical examination may point directly to a specific thyroid disorder and the diagnostic sensitivity and specificity may be nearly 100% in some cases. However, the symptoms and signs of hypo- or hyperthyroidism overlap considerably with complaints and abnormalities commonly seen in other diseases and in apparently healthy people (e.g. fatigue, weight alterations, nervousness, lack of concentration, constipation). Importantly, the diagnostic accuracy of symptoms decreases with age and are no longer valid past the age of 65 years [5]. Biochemical testing of thyroid function is thus essential in the evaluation of the thyroid patient.

Laboratory test results of thyroid function may be influenced by a number of clinical conditions and by medication. Interpretation of test results depend on information on medication and clinical conditions and it is important to obtain this information. One important example is pregnancy [6]. Both total and free thyroid hormones in serum vary during normal pregnancy, and pregnancy-induced modulations of the immune system may modify autoimmune thyroid abnormalities.

Transient hypo- or hyperthyroidism are seen in 5–10 % of women 3–10 months after delivery as part of autoimmune postpartum thyroiditis. Another example is severe general illness [7], which may be accompanied by various alterations in total and free thyroid hormones and thyroid-stimulation hormone (TSH) in serum even if the thyroid gland is not affected.

Medication may alter thyroid function tests [8]. Some important examples are oestrogens (high thyroxin-binding globulin with high total thyroxine (T4) and triiodothyronine (T3)), carbamazepine, and phenytoin (low total and free T4 and T3), and amiodarone (high total and free T4, slightly depressed total and free T3, and TSH in the upper part for the reference range). These are the variations seen in patients without thyroid abnormalities. Amiodarone has high iodine content and is also a frequent cause of thyroid disease.

Excess iodine may induce hypo- or hyperthyroidism in susceptible individuals. The disease is transient in most cases, and the sources may be iodine-containing medications, over-the-counter 'health products' with iodine, intake of seaweed, or iodine-containing radiocontrast agents. The type of dysfunction depends on the general iodine intake of the population. Hypothyroidism is the common abnormality induced in geographical areas with a high iodine intake, while thyrotoxicosis predominates in areas with a low basic iodine intake. This difference in type of abnormality induced be excess iodine intake reflects the underlying difference in the epidemiology of thyroid abnormalities in low and high iodine intake areas. Low iodine intake areas see frequent non-toxic and multinodular toxic goitres, whereas autoimmune diseases with subclinical and clinical hypothyroidism are common in high iodine intake areas. Hence, history taking should include information on excess iodine intake and additional information on the general iodine intake level in the area. This will provide clues to the probability of the various thyroid abnormalities to look out for.

Thyroid diseases cluster in some families and a family history is a valuable contribution to the risk estimation for a patient. Information on more specific genetic defects such as those leading to thyroid hormone resistance syndromes or to alterations of hormone binding proteins in serum is important to avoid diagnostic errors. The presence of autoimmune disorders such as vitiligo, rheumatoid arthritis, type 1 diabetes, Addison's disease and pernicious anaemia in a patient considerably enhances the risk of an autoimmune thyroid disorder in the patient and may guide the focus of the diagnostic work-up.

Previous thyroid disease raises the risk of current thyroid abnormality. For example, both hyper- and hypothyroidism may be transient if induced by excessive iodine intake (see Chapter 3.2.4). However, there may be an underlying subclinical thyroid abnormality such as autonomous thyroid nodules in hyperthyroidism or autoimmune thyroiditis in hypothyroidism and re-exposure to excess iodine commonly causes relapse. Still, spontaneous development of a thyroid function abnormality may also occur.

Patients with postpartum thyroiditis typically harbour an underlying autoimmune thyroiditis. Hence, a new episode of thyroiditis is common following a subsequent pregnancy, and the risk of a permanent thyroid hypofunction increases considerably. The risk of relapse is high (in the order of 50%) in a patient in remission after previous medical treatment for the hyperthyroidism of Graves' disease. Some patients treated with radioiodine or surgery for hyperthyroidism develop immediate or early hypothyroidism, though it may develop later and occur even after decades in others. Patients who have received external radiation of the neck have an increased risk of hypothyroidism. Patients treated with radiation to the neck or exposed to radioactive fallout during childhood have a markedly increased risk of malignant and benign thyroid nodules.

The history of tobacco smoking is pertinent to history taking in the patient with thyroid disease because smoking may aggravate the

orbitopathy of Graves' disease [9], and lead to a high frequency of goitre by interacting with iodine in the thyroid in iodine deficiency [10].

Physical Examination of the Thyroid Gland

Any complete physical examination of the thyroid gland includes inspection and palpation of the anterior region of the neck. Auscultation of the thyroid gland can be used to evaluate bold flow in a goitre and percussion of the upper part of the sternum may be used as an indicator for the presence of a large retrosternal goitre.

The normal thyroid is situated with the upper poles of the lobe at the level of the cricoid cartilage. The lower poles are 1–2 cm above the sternoclavicular junction in young adults, but the thyroid gland tends to be located more caudally on the neck in elderly people.

Inspection of the Thyroid Gland

The patient is examined sitting or standing with light from a window or a lamp at an obliquely angle on the anterior of the neck. The chin of the patient is raised moderately. Inspection of the skin may reveal scars after thyroid surgery and vascular changes suggesting impaired venous flow or pervious radiation of the neck.

Do careful inspection of the thyroid region for signs of thyroid enlargement, nodules, and asymmetry (Figure 3.1.3.1a). The normal thyroid gland is not or only barely visible in most people. The exception is young women with a slender neck, in whom a high and medially situated normal thyroid gland may give the clinical impression of goitre ('pseudo-goitre').

The next step is inspection of the region while the patient is swallowing. In lack of drinking water, the patient may be asked to imagine chewing a piece of lemon. This may induce salivation and facilitate swallowing. The thyroid gland will normally move upwards during swallowing following the trachea (compare the thyroid region before swallowing (Figure 3.1.3.1b) with the region during swallowing (Figure 3.1.3.1c)). Small thyroid enlargements and nodules may be identified in this way. Inspection during swallowing is an important part of characterization of a goitre. If the goitre remains fixed to the surroundings and does not move it may be a sign of malignancy (Box 3.1.3.1). If still uncertain, inspect the

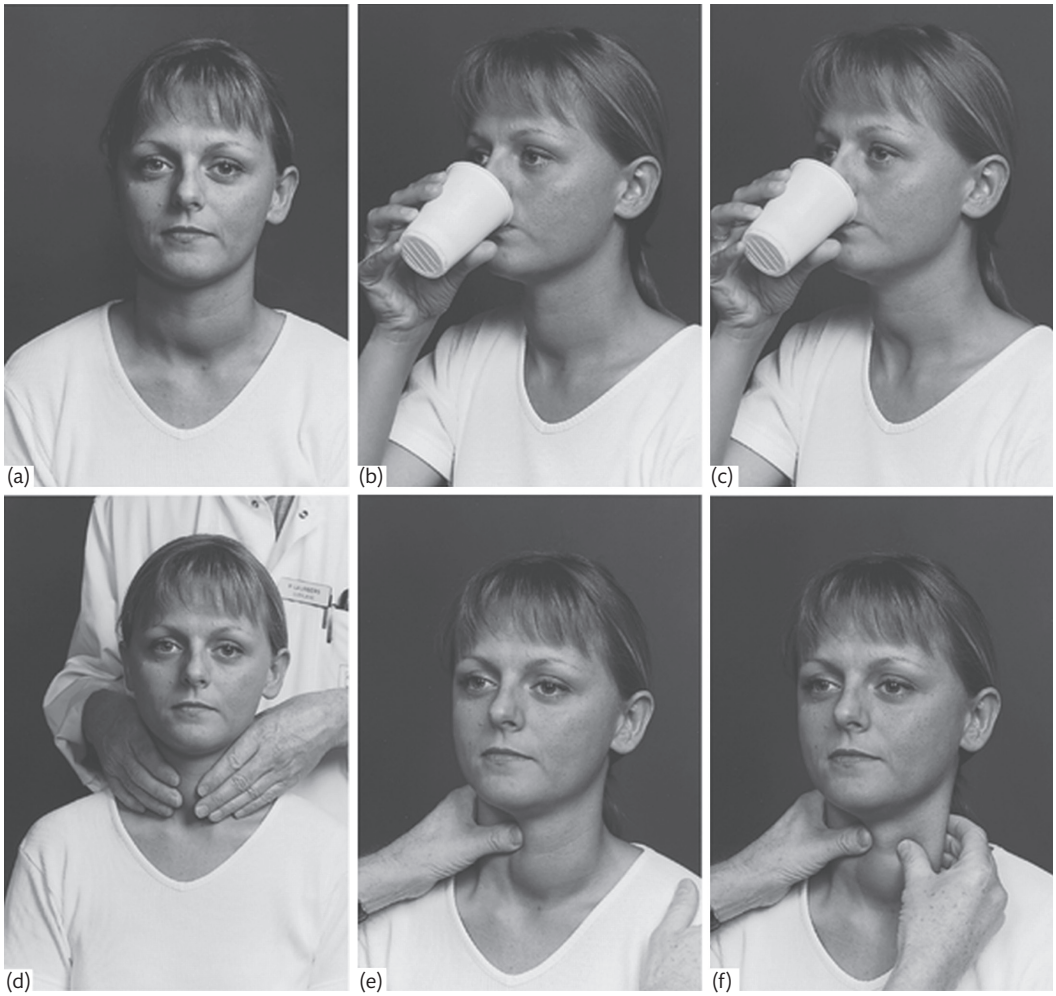


Figure 3.1.3.1 Clinical examination of the thyroid gland in a young woman with a small goitre. (a) Inspection with oblique light, (b) inspection while the patient drinks water before swallowing, and (c) during swallowing. Note the change in position of the small goitre. (d) Palpation of the superficial part of the thyroid gland with flat fingertips. (e) Displacement of the thyroid to the left by pressure on the larynx (pressure on the trachea is more irritant). (f) Bidigital palpation of the deep parts of the left thyroid lobe behind the sternocleidomastoid muscle.

thyroid region while the patient swallows, with light from various angels and with the neck of the patient more or less extended.

Palpation of the Thyroid Gland

Palpation can be performed while the examiner and the patient are sitting or standing in front of each other or while the examiner is standing behind the sitting patient. The patient should hold the head upright but the neck should not be hyperextended. Palpation involves a superficial and a deep examination of the gland. In addition, thorough palpation for enlarged cervical lymph nodes should be performed. The superficial part of the thyroid is examined by moving the flat fingertips systematically across the thyroid region searching for swellings and nodules (Figure 3.1.3.1d). When special care is needed, the examination may be facilitated by using lubricant (e.g. gel for ultrasound examination). It may also be helpful to ask the patient to swallow while palpating the gland softly. Nodules and enlarged lobes may be identified when they are moving.

Palpation of the thyroid lobes between the fingers is achieved by displacing the larynx (and thereby the trachea and the thyroid gland) to one side (Figure 3.1.3.1e) and palpating the thyroid lobe behind the sternocleidomastoid muscle as illustrated in Figure 3.1.3.1f. Nodules in the deeper parts of the thyroid lobes can be detected in this way.

If goitre or one or more nodules are found they should be examined and described with respect to size, hardness, location, mobility, and tenderness. Proper description of location is important as it aids interpretation of the findings on the scintigrams, particularly for support of the suspicion of a cold nodule. Lack of mobility during swallowing may be a sign of malignancy but several other possibilities exist. Fixation could be caused by inflammation surrounding an acute or subacute thyroiditis. Such lesions tend, however, to be painful, which is rare in case of a cancer.

Auscultation of the Thyroid Gland

The goitre of patients with active Graves' disease may have a very high blood flow. The flow can occasionally be heard upon auscultation as a systolic murmur over the gland. When present in the medically treated patient it indicates persistent activity of the disease despite medication, and it is often accompanied by a high T_3/T_4 ratio. If surgery is planned, pretreatment with iodine for 7–10 days prior to surgery is recommended [11].

A similar clinical pattern may occasionally be seen in patients with Graves´ disease grossly overtreated with thyroid blocking drugs. The treatment causes a lowering of serum T_3 and T_4 with subsequent excessive TSH secretion. Such high TSH can induce a 'blocking goitre' with a high blood flow.

A systolic murmur over the thyroid does not always originate from the thyroid gland. Other diagnostic options include referred sound from the heart in a patient with aortic stenosis or sclerosis and a systolic murmur from an arteriosclerotic carotid artery.

Reliability of Clinical Assessment in Thyroid Disease

The typical patient with Graves´ disease is a young or middle-aged woman with a family history of thyroid disease, complains of nervousness, heat intolerance, palpitations, weight loss, high pulse rate, agility, and a diffuse goitre. The diagnosis based on clinical assessment is nearly 100% reliable in this case. The occurrence of eye symptoms further supports the diagnosis. However, both hyper- and hypothyroidism may be difficult to diagnose from clinical findings especially in elderly people where the diseases may be monosymptomatic with, for example, slow cerebration in hypothyroidism and weight loss in thyrotoxicosis. Biochemical evaluation of thyroid function is necessary and TSH measurement should be a first-line test.

The diagnosis of goitre is a separate challenge. Classically, goitre is a thyroid gland, which is palpable or visible due to focal or general enlargement. Occasionally the goitre is not visible or palpable because the growth and extension of the gland occurred behind the sternum causing a retrosternal goitre.

A visible and/or palpable thyroid gland is not a goitre if there is no general or focal enlargement. In young women, this may be seen as a 'pseudo-goitre'. Ultrasound examination of the thyroid gland with measurement of volume and identification or exclusion of thyroid nodules is an important supplement to the clinical evaluation. The interobserver variability of thyroid volume determinations by ultrasonography is around 10% and the reproducibility of identifying nodules is high.

The size of the thyroid gland of apparently healthy people is highly dependent on the iodine intake level of the geographical area of the investigation. Upper normal values of 18 ml for women and 25 ml for men have been suggested [12].

However, there is a marked discrepancy between the 'true' thyroid volume determine by ultrasound examination and the finding of no goitre or a small goitre by clinical examination. This is illustrated in Figure 3.1.3.2, which also shows that clinical goitre may not be present despite thyroid volumes several times the upper normal. Systematic studies of the clinical assessment of goitre have shown that estimation of thyroid size by inspection and palpation is imprecise and large intra- and interobserver variations are present [14, 15].

The estimation of thyroid size by clinical examination is imprecise. Evaluation of nodularity of the thyroid gland faces similar difficulties. Solitary thyroid nodules identified by palpation are often part of multinodular glands when examined by ultrasonography. This is seen in up to 50% of cases [16]. Thyroid nodules are identified more frequently in patients when using ultrasound compared to clinical examination. Only around 50% of nodules larger than 1.5 cm detected by ultrasonography were identified by palpation in a follow-up study of patients who had received neck irradiation as children and therefore had a high risk for developing thyroid cancer later in life [17].

A clinical investigation of the thyroid and surroundings remains first-line investigation of the thyroid gland. Clinical investigation is also the initial investigation used for goitre detection. Increased volume of the thyroid gland by ultrasound examination is of limited clinical importance if thyroid function is normal and the patient has no signs or symptoms of goitre.

The situation differs when it is necessary to add test for thyroid function abnormalities and/or investigations for abnormalities of thyroid structure and size. The first-line test for thyroid function is TSH in serum. This should be performed in many patients presenting various complaints of some duration, even with limited clinical suspicion of thyroid disease. On the other hand, clinical examination of the thyroid gland remains an important first-line evaluation for goitre and thyroid nodules. Sensitive imaging procedures including ultrasound should be used for patients with abnormal clinical and/

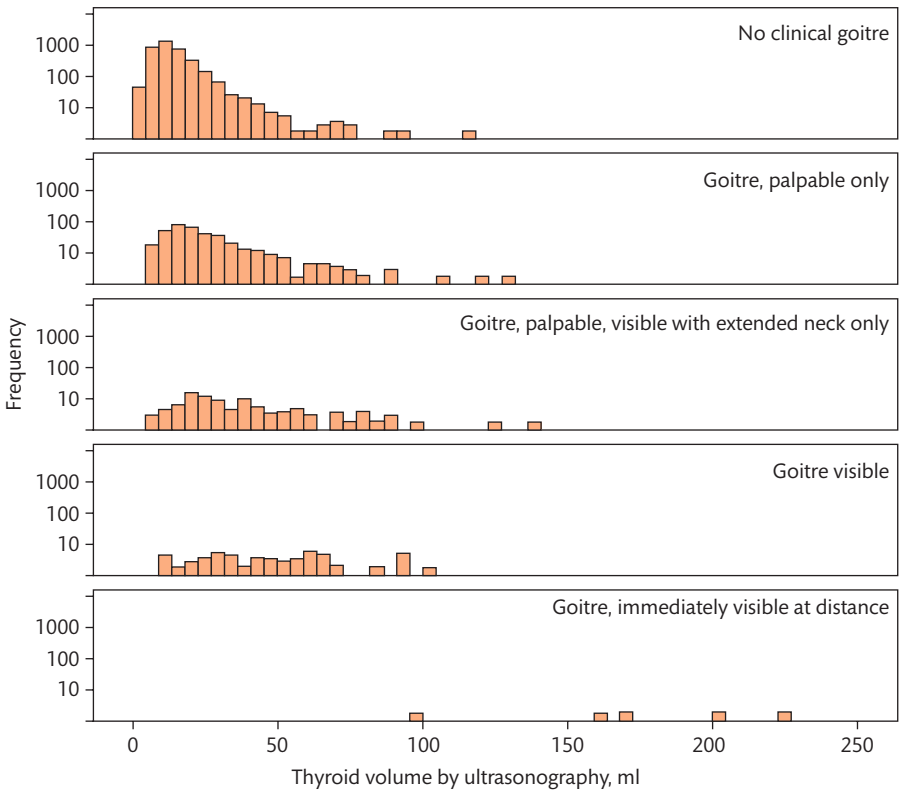


Figure 3.1.3.2 Goitre by clinical examination and thyroid volume measured by ultrasonography in a population study of 4649 people living in an area with mild to moderate iodine deficiency. Data from the DanThyr cross-sectional study performed before the Danish iodine fortification programme [13].

or biochemical findings suggesting thyroid disease, as well as for patients at increased risk of developing thyroid cancer.

REFERENCES

1. Laurberg P, Pedersen KM, Hreidarsson A, Sigfusson N, Iversen E, Knudsen PR. Iodine intake and the pattern of thyroid disorders: a comparative epidemiological study of thyroid abnormalities in the elderly in Iceland and in Jutland, Denmark. *J Clin Endocrinol Metab*, 1998; **83**: 765–9.
2. Knudsen N, Bülow I, Jørgensen T, Laurberg P, Ovesen L, Perrild H. Goitre prevalence and thyroid abnormalities at ultrasonography: a comparative epidemiological study in two regions with slightly different iodine status. *Clin Endocrinol (Oxf)*, 2000; **53**: 479–85.
3. Andersen S, Iversen F, Terpling S, Pedersen KM, Gustenhoff P, Laurberg P. More hypothyroidism and less hyperthyroidism with sufficient iodine nutrition compared to mild iodine deficiency—a comparative population-based study of older people. *Maturitas*, 2009; **64**: 126–31.
4. Carlé A, Pedersen IB, Knudsen N, *et al*. Epidemiology of subtypes of hyperthyroidism in Denmark: a population-based study. *Eur J Endocrinol*, 2011; **164**: 801–9.
5. Carlé A, Pedersen IB, Knudsen N, *et al*. Hypothyroid symptoms fail to predict thyroid insufficiency in old people: a population-based case-control study. *Am J Med*, 2016; **129**: 1082–92.
6. Baloch Z, Carayon P, Conte-Devolx B, *et al*. Guidelines Committee, National Academy of Clinical Biochemistry. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid*, 2003; **13**: 3–126.
7. Chopra IJ. Clinical review 86: euthyroid sick syndrome: is it a misnomer? *J Clin Endocrinol Metab*, 1997; **82**: 329–34.
8. Wenzel KW. Disturbances of thyroid function tests by drugs. *Acta Med Austrica*, 1996; **23**: 57–60.
9. Wiersinga WM. Management of Graves' ophthalmopathy. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 396–404.
10. Vejbjerg P, Knudsen N, Perrild H, *et al*. The impact of smoking on thyroid volume and function in relation to a shift towards iodine sufficiency. *Eur J Epidemiol*, 2008; **23**: 423–9.
11. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid,* 2016; **26**: 1343–421.
12. Gutekunst R, Becker W, Hehrmann R, Olbricht T, Pfannenstiel P. [Ultrasonic diagnosis of the thyroid gland]. *Dtsch Med Wochenschr*, 1988; **113**: 1109–12.
13. Laurberg P, Jørgensen T, Perrild H, *et al*. The Danish investigation on iodine intake and thyroid disease, DanThyr: status and perspectives. *Eur J Endocrinol*, 2006; **155**: 219–28.
14. Berghout A, Wiersinga WM, Smits NJ, Touber JL. The value of thyroid volume measured by ultrasonography in the diagnosis of goitre. *Clin Endocrinol (Oxf)*, 1988; **28**: 409–14.
15. Jarløv EA, Hegedüs L, Gjørup T, Hansen MJ. Inadequacy of the WHO classification of the thyroid gland. *Thyroidology*, 1992; **4**: 107–10.

16. Tan GH, Gharib H, Reading CC. Solitary thyroid nodule. Comparison between palpation and ultrasonography. *Arch Intern Med*, 1995; **155**: 2418–23.
17. Schneider AB, Bekerman C, Leland J, *et al*. Thyroid nodules in the follow-up of irradiated individuals: comparison of thyroid ultrasound with scanning and palpation. *J Clin Endocrinol Metab*, 1997; **82**: 4020–7.

3.1.4 Thyroid Function Tests and the Effects of Drugs

Ulla Feldt-Rasmussen

Introduction

Basal metabolic rate followed by protein-bound iodine were the first laboratory tests of thyroid function until the introduction of radioimmunoassay methods [1], which made it possible to quantify both circulating thyroxine (T_4), triiodothyronine (T_3) and thyroid-stimulating hormone (TSH). However, none of the methods had sufficient sensitivity and precision for diagnosing hypo- or hyperfunction of the thyroid gland. Development of third generation immunometric TSH assays with highly improved functional sensitivity allowed suppressed serum TSH concentrations to be clearly distinguished from normal. Different ingenious techniques to estimate the minute free fraction of total serum T_4 have been developed over the years, but even the best free T_4 methods offer only marginal diagnostic advantage over the measurement of total T_4 when the concentration of thyroxine-binding globulin (TBG) and other binding proteins are abnormal. Current enthusiasm for free T_4 and T_3 estimation needs to be tempered by an understanding of the method-dependent limitations of these techniques, particularly where assessment of thyroid function is most difficult.

While there is little doubt that circulating TSH and T_4 should both be measured when an abnormality of thyroid function is clinically suspected, recent recommendations suggest testing a wide range of patient groups with an increased risk of thyroid dysfunction (**Box 3.1.4.1**). Thus, neonatal screening for congenital hypothyroidism is firmly established. Routine testing of thyroid function with a single measurement of serum TSH in women over 50, a group most likely to have significant thyroid dysfunction [2, 3] has become widely recommended. The recognition that adequate maternal T_4 in the first trimester of pregnancy is a crucial determinant of fetal brain development, has led to increased assessment of thyroid function before pregnancy, especially in women with impaired fertility or risk factors for thyroid dysfunction. The frequency of postpartum thyroid dysfunction places a high priority on assessing thyroid function for any suggestive clinical features in the first year after childbirth [4].

There is a trend for peripheral thyroid hormones to be estimated in primary care only if TSH is abnormal. In case of uncertainty of the diagnosis, other biochemical methods such as circulating TBG, thyroid autoantibodies, thyroglobulin as well as genetic

Box 3.1.4.1 Conditions with an increased likelihood of thyroid dysfunction where thyroid function testing should be considered

- Previous thyroid disease or surgery
- Goitre
- Non-thyroid autoimmune disease(s)
- Other endocrine deficiencies
- Atrial fibrillation
- Down's syndrome, Turner's syndrome
- Type 1 diabetes
- Metabolic syndrome, morbid obesity
- Irradiation of head and neck
- Impaired reproductive function in women
  - Polycystic ovarian syndrome; Endometriosis
  - Premature ovarian failure
  - Recurrent miscarriage
- Postpartum ill health
- Preterm infants
- Pituitary abnormality
- Severe head injury

Drug therapy
- Cytotoxic therapy
- Contrast agent or other iodine exposure
- Amiodarone
- Lithium
- Highly active antiretroviral therapy
- Mitotane
- Sunitinib
- Retinoids
- Biological agents (interferon α, interleukin 2, interferon β 1a or 1b, monoclonal antibody treatment; denileukin diftitox)

testing and imaging techniques can assist in defining the thyroid function disturbance [1]. The value of routine testing needs to be compared with sensitivity and accuracy of clinical assessment, which was in unselected patients assessed by primary care physicians and even specialists inferior to laboratory results in up to one-third of patients evaluated for suspected thyroid disease [5]. Finally, many drugs can influence both thyroid function *per se* while some affect the biochemical thyroid function measurements without function disturbance (i.e. spurious and often uninterpretable. This chapter is an update from the previous edition written by the late Jim Stockigt [5]).

The T_4–TSH Relationship

Regardless of the strategy used for first-line testing, serum TSH and a valid serum T_4 estimate are both necessary for definitive assessment of thyroid status. As shown in Figure 3.1.4.1, the common types of thyroid dysfunction can be identified by diagonal deviations from the normal T_4–TSH relationship, which depends on the negative feedback interaction between target gland secretion and trophic hormone. The figure shows primary hypothyroidism due to target gland failure (high serum TSH with low free T_4: A), failure of TSH secretion (both low: B), autonomous or abnormally stimulated target gland function (high serum free T_4 with low TSH: C), and primary excess of TSH or thyroid hormone resistance (both high: D). Abnormal results that fall outside these areas suggest that some other factor has disturbed this relationship, or that the sample has been collected under non-steady state conditions (e.g. altered trophic-target hormone relationship, medications, TSH receptor mutations, hormone resistance syndromes, inaccurate estimates of active hormone concentrations, inaccurate reference ranges, or low assay sensitivity). The figure shows serum free T_4 rather than T_3 because T_4 is the major circulating determinant of TSH secretion.

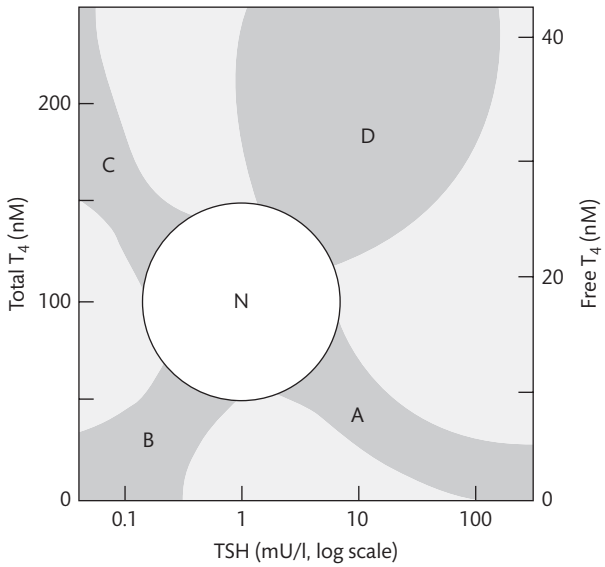


Figure 3.1.4.1 The relationship between serum TSH and total free T_4 concentrations is shown in normal subjects (N) and in various typical abnormalities of thyroid function: primary hypothyroidism (A); central or pituitary-dependent hypothyroidism (B); thyrotoxicosis due to autonomy or abnormal thyroid stimulation (C); and TSH-dependent thyrotoxicosis or generalized thyroid hormone resistance (D). Note that linear free T_4 responses correspond to logarithmic TSH changes. Areas A and C represent primary thyroid abnormalities, while results that fall in areas B and D suggest a primary pituitary abnormality. Abnormal findings that fall in the intermediate areas suggest non-steady state sampling conditions due to the large difference in half-lives of TSH and T_4, an assay artefact, an altered T_4–TSH relationship, or the presence of another agonist (e.g. T_3).

The assumption of steady state condition should always be questioned when associated illness or medications perturb the pituitary–thyroid axis. The large difference between the half-lives of TSH (1 h) and T_4 (1 week) accounts for many transient non-diagnostic abnormalities in the T_4–TSH relationship, such as acute effects of medications, early response to therapy, evolution of disease, and diurnal variations. Optimal assessment of thyroid function depends on collaborative communication across the laboratory–clinical interface [1].

Screening and Case-Finding

In the absence of associated disease, where there are no clinical features to suggest thyroid dysfunction, a normal serum TSH concentration has over 99% negative predictive value in ruling out primary hypothyroidism or thyrotoxicosis [3]. Assessment of untreated subjects with no features of thyroid dysfunction and therefore with low prediagnostic probability now commonly begins with initial measurement of TSH alone, with T_4 and/or T_3 assays added only if TSH is abnormal, or if an abnormality of TSH secretion is suspected. According to this algorithm, free T_4 is measured to distinguish between overt and subclinical hypothyroidism when serum TSH is elevated, while a suppressed or subnormal TSH level should be followed by measuring both free T_4 and T_3 to distinguish subclinical from overt thyrotoxicosis and to identify T_3 toxicosis.

Clinical Suspicion of Thyroid Dysfunction

The use of serum TSH as the sole initial test of thyroid function may lead to incorrect or incomplete assessment of thyroid status in a number of situations [1, 5]. Initial measurement of both T_4 and TSH is appropriate whenever thyroid dysfunction is clinically suspected, because thyroid dysfunction due to pituitary disease, either hypopituitarism, or the less common situation of TSH-dependent hyperthyroidism, may be missed if TSH alone is used for initial assessment [6]. The far-reaching consequences of missing these disorders are not reflected by a small percentage deficit in diagnostic sensitivity.

Evaluating and Adjusting the Response to Treatment

In patients with newly treated thyrotoxicosis, TSH may remain suppressed for months or even longer after normalization of serum T_4 and T_3; serious overtreatment may result if TSH alone is used for adjustment of antithyroid drug dosage. Further, thyrotoxicosis may persist due solely to T_3 excess. Therefore, a reassessment of serum

free T_4 and free T_3 levels is recommended after about 3 weeks of antithyroid drug therapy to allow appropriate dose adjustment.

During standard T_4 replacement therapy a TSH value in the lower normal range usually coincides with an optimal symptomatic response. Serum TSH is the best single index of appropriate replacement of hypothyroidism during long-term T_4-therapy, but during the early phase of treatment, free T_4 should also be measured, because TSH may remain inappropriately elevated long after normalization of T_4. In elderly patients, especially with cardiac ischaemia, dose adjustment is a clinical not a laboratory decision. During T_4 suppressive therapy (e.g. after treatment for thyroid cancer), periodic assessment of free T_4 (but not T_3 [7]), in addition to TSH, is appropriate to identify and avoid thyroid hormone excess that may have adverse effects on the cardiovascular system or bone density.

During treatment of hypothyroidism for hypothalamo-pituitary disease, serum TSH is of limited value in assessing T_4 dosage, which should be based on serum free T_4 and clinical response. However, serum TSH needs to be suppressed to avoid under-replacement [8].

Difficult Diagnostic Situations

Interpretation of thyroid function tests may be compromised by intercurrent illness and medications. There is a high prevalence of abnormal thyroid function tests in patients with acute medical [9] and sometimes acute psychiatric illness [10]. However, when TSH and free T_4 are considered together, as in **Figure 3.1.4.1**, most of these abnormalities do not indicate true thyroid dysfunction. However, the consensus has moved away from routine testing during critical illness without a strict clinical indication [5]. If not due to medications, the combination of low free T_4 and TSH indicates a poor prognosis [5], although these findings cannot usefully influence individual management decisions.

During any severe illness, one or more of the assumptions of hypothalamo–pituitary–thyroid stability may not be valid. Serum TSH values are frequently subnormal in the absence of thyrotoxicosis, but the large majority of thyrotoxic subjects have values below 0.01 mU/L, whereas hospitalized patients with non-thyroidal illness do not show this degree of TSH suppression [11]; serum free T_4 estimates during critical illness are prone to multiple method-dependent interfering influences, due to heparin and other medications. Notably, serum total T_4 measurements are much less prone to such artefacts [12].

In late pregnancy, there are clearly unresolved methodological problems in estimating serum free T_4, with strong negative bias in some methods [4, 13, 14], which has questioned the wisdom of continuing to rely on free thyroxine estimates during pregnancy [13]. In contrast to various free T_4 methods [4, 13] total serum T_4 and its derivative, the free thyroxine index, showed a more robust inverse relationship with serum TSH, with consistent results in numerous reports [13]. Thus, total T_4 measurement may be superior to free T_4 estimates as a guide to therapy during pregnancy, provided modification of the reference values to the normal oestrogen-induced increase in TBG. If free T_4 estimates continue to be used in pregnancy, clinicians should interpret results according to trimester- and method specific reference intervals. Problems inherent in free T_4 measurement during pregnancy and otherwise can possibly be resolved by using isotope dilution liquid chromatography tandem mass spectrometry after ultrafiltration [15].

Measurement of Serum TSH Concentrations

The secretion of TSH, a 24–30 kDa glycoprotein composed of two subunits, from the anterior pituitary is regulated by negative feedback from circulating free T_4 and T_3 concentrations. In normal subjects, serum TSH concentrations show pulsatile and diurnal variation, with approximate mean maximum concentrations of 3 mU/L at 02.00 with nadir values about 1 mU/L at about 16.00 without significant sex difference [16]. This TSH fluctuation with an amplitude of 20–50% makes it difficult to establish the significance of serial changes during follow-up of patients with subclinical hypothyroidism, because a 40% change due to pulsatile secretion can be misread as progression of disease [17].

TSH is measured by immunometric assays using two antibodies against different epitopes on the α- and β-subunits of TSH [11], and serum TSH can be precisely measured at least to 0.03 mU/L. Important factors when clinical decisions are based on values close to the detection limit include between-assay reproducibility or precision profile, composition of assay matrix, possible appearance of non-specific interference during sample storage, as well as possible carryover from one sample to the next during automated sampling [5]. Analytical sensitivity defined as 2 or 3 SD above the zero point is often too optimistic. Functional sensitivity defined as 20% between-assay coefficient of variation has become accepted [11], but assay performance may vary between laboratories despite apparently identical technique. Laboratories should therefore establish their own detection limit from the between-assay precision profile in the subnormal range.

While immunometric TSH assays offer enhanced sensitivity, they may be liable to non-specific interference (e.g. in methods that use mouse monoclonal antibodies). An antimouse immunoglobulin in the test serum allows formation of a false bridge between the solid phase and the signal antibody, thus generating a spuriously high assay value. Inclusion of mouse immunoglobulin in the assay usually blocks this effect, although persistent false-positive serum TSH values are still found in some samples [18].

Most patients with TSH-secreting pituitary tumours have increased serum α -subunit concentrations [19], but values can also be elevated in postmenopausal women and in hypogonadal men.

The TSH Reference Range: Current Controversies and Uncertainties

Between 08.00 and 21.00, reference values for serum TSH are generally in the range 0.3–4 mU/L, following a logarithmic Gaussian distribution and with higher values in the immediate postnatal period. Median values are generally about 1 mU/L with a long tail to the right, so the upper limit of the reference range is contentious. Widespread application of thyroid function testing has identified large numbers of asymptomatic subjects with abnormal TSH and with normal serum T_4, who may merit the designation 'subclinical thyroid dysfunction' [20]. A sustained abnormality should be demonstrated before definite categorization [21]. The

merits and limitations of initiating therapy for these individuals are controversial.

These considerations have become complicated because of lack of consensus on the limits of the TSH range [20, 22], in particular on whether the upper limit of the TSH reference range should be reduced from about 4 mU/L to 3 mU/L or even lower. Similarly, for subclinical hyperthyroidism there is lack of consensus as to classification of subnormal TSH values according to the NHANES III study [11] (TSH values below 0.1 mU/L) or to other guidelines in favour of using values below the lower normal limit of 0.45 mU/L [20]. Since the gradation from normality to severe thyroid dysfunction is a continuum, studies of outcomes from intervention will be critically dependent on uniform cut-off points and terminology.

Until these uncertainties are resolved, it is likely that most clinicians will recommend a period of observation rather than immediate intervention. If a trend towards overt disease is to be the cue to intervention, it is critical to establish what constitutes a significant change in serum TSH value. From an analysis of serial individual variation over 1 year, the difference required for two test results to be convincingly different was 40% for TSH and 15% for free T_4 and free T_3 [17].

Indications for TRH Testing

Thyrotropin-releasing hormone (TRH) testing in clinical practice has almost been eliminated by the highly sensitive TSH assays. However, measurement of serum TSH 20–30 min after intravenous injection of 200–500 µg TRH is still useful for some purposes: (1) when basal serum TSH is out of context (TSH assay artefacts); (2) apparent thyroid hormone resistance or pituitary-dependent thyrotoxicosis (patients with TSH-secreting pituitary tumours show no TSH increase after TRH [19], while those with thyroid hormone resistance usually do); and (3) central hypothyroidism where low serum free T_4 may be associated with normal serum immunoreactive TSH concentration with impaired biological activity [23].

Assays for Serum T_4 and T_3

Concentrations of total serum T_4 and T_3 reflect not only hormone production, but also the number and affinity of plasma protein binding sites. Total concentrations vary in direct relationship to protein binding, while serum free T_4 and free T_3 concentrations should not, if measured by valid methods. Serum total and free T_3 concentrations are higher in children [24]. In pregnancy, reference ranges for free T_4 show marked method-dependent variation; quoted ranges should be both trimester and method specific [25].

Estimation of Serum Free T_4 and Free T_3

There have been many approaches to assay serum free T_4 and T_3 concentrations, with detailed analysis of the validity of various methods [1, 11, 15]. Two-step methods that separate a fraction of the free T_4 pool from the binding proteins before assay are generally least prone to analytical artefacts. No current method conveniently measures the free T_4 concentration in undisturbed, undiluted serum similarly to *in vivo* conditions. Equilibrium dialysis is widely considered the reference method for free T_4 measurement, but is also subject to error. Evaluation of novel serum free T_4 methods should include testing with various protein binding abnormalities, as well as sera that contain substances competing for serum protein binding sites. Unexpected interference may only be noted after methods have been used for some time, as, for example, in the effect of rheumatoid factor [26], or drug competitors for protein binding [27].

Recent reports suggest that methods based on liquid chromatography/tandem mass spectrometry after ultrafiltration, or equilibrium dialysis may improve measurement of free T_4 [15]. Further evaluations, in particular details of long-term reproducibility of these techniques, as well as serial dilution studies to evaluate the effect of circulating inhibitors of T_4 binding are awaited. In many countries the price per sample is prohibitive for routine use, even if the assay quality appears to be superior in terms of sensitivity, precision, and accuracy.

Measurement of Serum T_3

Assays for total or free T_3 have no place in diagnosing hypothyroidism, but should be included in the diagnostic protocol in the following situations:

1. Suspected thyrotoxicosis when serum T_4 is normal and serum TSH is suppressed, to distinguish T_3 thyrotoxicosis from subclinical thyrotoxicosis
2. Antithyroid drug therapy to identify persistent T_3 excess, despite normal or subnormal serum T_4 values
3. Amiodarone-induced thyrotoxicosis, which should not be based on T_4 excess alone
4. Early recurrence of thyrotoxicosis with suppressed TSH, after cessation of antithyroid drug therapy
5. Establish the extent of hormone excess when an intentional T_4 overdose has been taken

The serum T_3 concentration is not useful in assessing T_3 replacement (e.g. because of its short plasma half-life).

Variant Binding Proteins and Antireagent Antibodies

Molecular changes in TBG, transthyretin (TTR), or albumin may result in altered serum concentrations of these binding proteins, or may alter their binding affinity for T_4 and/or T_3. The X-linked structural TBG variants have either normal or reduced affinity for T_4 and T_3. Most X-linked variants cause complete TBG deficiency, while some are associated with subnormal concentrations of immunoreactive serum TBG, often with reduced affinity for T_4 [28]. In complete absence of TBG, total serum T_4 is reduced to 20–40 nmol/L (normal 50–140 nmol/L), whereas in hereditary TBG excess the concentration may increase up to 250 nmol/L [28]; free T_4 should remain normal, but some estimated free T_4 methods do not

correct quantitatively for TBG abnormalities, whether hereditary or acquired.

Familial dysalbuminaemic hyperthyroxinaemia (FDH) [29], shows a selective increase in binding affinity for T_4 resulting in total serum T_4 in the range 180–240 nmol/L. The variant protein has increased affinity for some T_4-analogue tracers, resulting in spuriously high serum free T_4 estimates; equilibrium dialysis, two-step free T_4 methods, and serum TSH and thyroglobulin confirm that people with FDH are euthyroid. The diagnosis is verified by genetic analysis.

Circulating T_3- or T_4-binding autoantibodies can cause methodological artefacts in both total and free measurements of T_4 and T_3 [11, 12]. Tracer bound to the endogenous human antibody will be falsely classified as 'bound' or 'free', leading, respectively, to spuriously low or high serum values. Assay after ethanol extraction establishes the true total hormone concentration.

Interference can also result from antibodies against assay reagents (e.g. antiruthenium, antistreptavidin, or antibiotin) [11]. High dose biotin ingestion has also been shown to result in serious distortion of analyte-specific, platform specific assay results, and is now a frequent cause of false results due to the current popularity of biotin ingestion [11].

Euthyroid Hyperthyroxinaemia and Hypothyroxinaemia

These terms are used when the total or free T_4 concentrations are increased or decreased without evidence of thyroid dysfunction. The effects of medications and alterations in the T_4 binding proteins are the commonest causes. Hypothyroxinaemia is a normal response when TSH secretion is inhibited by another thyromimetic such as T_3 or triiodothyroacetic acid. During critical illness serum T_4 may be subnormal due to inhibition of TSH secretion [30], decreased production of binding proteins, or accelerated T_4 clearance. Hypothyroxinaemia without increase in TSH is also seen in low birthweight premature infants, reflecting hypothalamic–pituitary immaturity [31].

Integration of Tests of Thyroid Function with Other Investigations

When thyroid function is abnormal, additional diagnostic information can be gained from the technologies described here next.

Antibody Measurements

In subclinical hypothyroidism, presence of thyroid peroxidase and/or thyroglobulin antibodies indicates a substantial increased risk of developing overt hypothyroidism [2], and also indicates an increased likelihood of postpartum thyroiditis or amiodarone-induced hypothyroidism. Persistently positive thyrotropin receptor antibody (TRAb) is useful in indicating that apparent remission of Graves' disease is unlikely to be sustained, while negative TRAb are not predictive of permanent remission. TRAb measurement can indicate the possibility of neonatal thyrotoxicosis in the infant of a mother with autoimmune thyroid disease and may also define the aetiology of atypical eye disease.

Thyroid Imaging

The use of isotope imaging techniques in thyrotoxicosis due to Graves' disease varies widely between different centres. While some now regard routine radioisotope imaging as redundant in typical Graves' disease, negligible uptake can be a key feature in confirming thyrotoxicosis due to thyroiditis, iodine contamination, and factitious ingestion of thyroid hormone. Imaging also can confirm a 'hot' nodule as the predominant source of thyroid hormone excess. Computed tomography (CT) is valuable in identifying retrosternal extension, but contrast agents should be avoided. Colour flow Doppler has been reported to differentiate between, for example, type 1 and type 2 amiodarone-induced thyrotoxicosis [32].

Thyroglobulin

In the follow-up of differentiated thyroid cancer, an undetectable serum thyroglobulin concentration (in absence of thyroglobulin antibodies), in presence of high serum TSH indicates effective ablation and may justify less rigorous T_4 suppression of TSH. Thyroglobulin is undetectable in thyrotoxicosis factitia, and generally extremely high in subacute thyroiditis and in amiodarone-induced thyrotoxicosis of the thyroiditis type [5, 11].

Assay of thyroglobulin in the needle wash from suspect neck lymph nodes has a higher sensitivity and specificity than cytology in diagnosing metastatic thyroid tissue [33].

Indices of Thyroid Hormone Action

While there is no reliable laboratory index of peripheral thyroid hormone action, some tests [34], including sex steroid-binding globulin, serum ferritin, serum angiotensin-converting enzyme, as well as oxygen consumption, systolic time interval, and cardiac contractility [35], may be useful in following individual response in situations of suspected thyroid hormone resistance or during long-term suppressive therapy with T_4.

Drug Effects on Serum T_4 and TSH

The multiple effects of medications on the pituitary–thyroid axis have been reviewed elsewhere [1, 28, 36]. Medications that present special problems include amiodarone, heparin, lithium, phenytoin, highly active antiretroviral therapy, immune-mediated therapies [36], and drugs that displace T_4 from TBG. Oestrogen, endogenous or exogenous, most commonly affects tests of thyroid function by increasing total T_4 due to an increase in TBG concentration by increased glycosylation and slower clearance. Free T_4 remains normal. Transdermal oestrogens do not show this effect [28].

Amiodarone

Amiodarone is the most complex and difficult of drugs affecting thyroid status [32]. The clinical entities include two forms of thyrotoxicosis, one due to iodine excess and one attributed to thyroiditis. In iodine-replete regions the predominant amiodarone-induced thyroid abnormality is hypothyroidism, especially in those with associated autoimmune thyroiditis. The drug also causes benign euthyroid hyperthyroxinaemia in up to 25% of treated patients. There is often poor correlation between circulating thyroid

hormone levels and the clinical manifestations of amiodarone-induced thyroid dysfunction, and criteria such as muscle weakness and weight loss are more relevant for severity assessment.

Heparin

In serum obtained from heparin-treated patients, the measured concentration of serum free T_4 may be higher than the true *in vivo* concentration, due to *in vitro* generation of non-esterified fatty acids as a result of heparin-induced lipase activity during sample storage or incubation [27]. Low-molecular-weight heparin preparations have a similar effect [37].

Lithium

Lithium, used in the management of manic-depressive illness, has multiple effects on the pituitary–thyroid axis, the most important being inhibition of thyroglobulin hydrolysis and hormone release [38]. It can exacerbate or initiate autoimmune thyroid disease with development of goitre and hypothyroidism; there are also rare reports of lithium-induced thyrotoxicosis. Serum TSH, T_4, and T_3 assays give a true index of thyroid status during lithium treatment.

Phenytoin and Carbamazepine

The antiepileptic phenytoin and carbamazepine commonly result in subnormal serum total T_4, with an apparent lowering of free T_4, not accompanied by the anticipated increase in TSH [39]. This discrepancy, which is not easily distinguishable from central hypothyroidism due to pituitary deficiency, is a methodological artefact related to underestimation of true free T_4 in diluted serum samples that contain inhibitors of T_4 protein binding [1, 11, 36, 39].

Antiretroviral Therapy

Infection with human immunodeficiency virus may influence tests of thyroid function by various mechanisms, occasionally as result of direct infection of the thyroid gland or alteration of immunological function, but more frequently from effect of medications that alter metabolism of thyroxine or as a non-specific effect of debilitating illness. Although some studies [40] show a higher than expected prevalence of hypothyroidism, predominantly subclinical, during treatment with highly active antiretroviral therapy, a recent study indicates that current treatment is associated with normal thyroid function measurements. There are reports of reduced effectiveness of T_4 replacement or transient over-replacement during treatment with antiviral drugs [41, 42]. There is no consensus as to whether thyroid function should be routinely monitored in infected patients, but testing will frequently be required to assess features that could be due to thyroid dysfunction.

Effects of Drug Competitors for Thyroid Hormone Binding to Plasma Proteins

Both TBG and TTR show extensive cross-reactions with a wide range of drugs [1, 28, 36]. As reviewed elsewhere [12], the failure of current free T_4 and T_3 methodology to reliably reflect the effect of drug competitors that increase free T_4 and T_3 *in vivo* by displacement, remains a major limitation in the general applicability of free hormone assays. These effects are poorly reflected by standard free hormone assays because samples are generally assayed after dilution, which underestimates the free hormone concentration in the presence of competitors. This discrepancy occurs because of dissociation of bound ligand with progressive sample dilution [12, 39]. Important drug competitors have a much smaller proportional reservoir of bound ligand than does T_4, so that their free concentration becomes negligible with progressive dilution while the free T_4 concentration remains unaltered [12]. Since competition is a function of relative free ligand concentrations, the effect of a competitor to increase free T_4 is underestimated, the error being greatest in assays with the highest sample dilution.

Drug Interactions

Drug effects on thyroid function may be especially potent when several agents are given together. For example, infusion of furosemide in high dosage lowers serum T_4, while concurrent dopamine infusion inhibits TSH secretion; together they can result in profound hypothyroxinaemia. Combinations of rifampicin or ritonavir or other medications that accelerate T_4 clearance, with glucocorticoid-induced inhibition of TSH secretion can have a similar effect.

Diagnostic Approach to Anomalous or Discordant Laboratory Results

When there is discordance between laboratory results and clinical findings, a distinction needs to be made between anomalous assay results due to assay interference and those that indicate previously unsuspected or subclinical disease. Consideration of the fundamental assumptions that underlie the diagnostic use of the trophic-target hormone relationship may give a clue to the discrepancy. Anomalous or unexpected assay results can be approached in the following sequence:

1. Clinical re-evaluation with particular attention to medication history and long-term features suggestive of thyroid disease (e.g. weight change, goitre).
2. Measurement of serum TSH by alternative method.
3. Estimation of free T_4 and T_3 by alternative methods with particular attention to method-dependent artefacts and medications.
4. Follow-up sampling to establish whether the abnormality is transient or persistent.
5. Measurement of serum total T_4 to establish whether the free T_4 estimate is disproportionately high or low in relation to total T_4 (e.g. heparin artefact, oestrogen). (Arguably, measurement of total T_4 with correction for variations in TBG, interpreted in conjunction with TSH, could now be regarded as the gold standard where free T_4 estimates are inconclusive).
6. Search for an unusual binding abnormality or hormone resistance syndrome in the propositus and family members.

REFERENCES

1. Stockigt JR. Clinical strategies in the testing of thyroid function, Chapter 6b. Available at: https://www.thyroidmanager.org (accessed: 17 May 2010). South Dartmouth, MA: Endocrine Education Inc.

2. Vanderpump MPJ, Tunbridge WMG, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol*, 1995; **43**: 55–68.
3. Helfand M, Redfern CC. Screening for thyroid disease: an update. *Ann Intern Med*, 1998; **129**: 144–58.
4. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
5. Stockigt J. Thyroid function tests and the effects of drugs. In: Wass JAH, Stewart PM, Amiel SA, Davies MJ, eds. *Oxford Textbook of Endocrinology and Diabetes*, 2nd edn. Oxford, UK: Oxford University Press, 2011: Ch. 3.1.4.
6. Beckett GJ, Toft AD. First-line thyroid function tests: TSH alone is not enough. *Clin Endocrinol*, 2003; **58**: 20–1.
7. Livingston M, Birch K, Guy M, Kane J, Heald AH. No role for tri-iodothyronine (T3) testing in the assessment of levothyroxine (T4) over-replacement in hypothyroid patients. *Br J Biomed Sci*, 2015; **72**: 160–3.
8. Persani L, Brabant G, Dattani M, *et al*. European Thyroid Association (ETA) guidelines on the diagnosis and management of central hypothyroidism (CeH). *Eur Thyroid J*, 2018; **7**: 225–37.
9. DeGroot LJ, Mayor G. Admission screening by thyroid function tests in an acute general care teaching hospital. *Am J Med*, 1992; **93**: 558–64.
10. Ryan WG, Roddam RF, Grizzie WE. Thyroid function screening in newly admitted psychiatric patients. *Ann Clin Psychiatry*, 1994; **6**: 7–12.
11. Spencer CA. Assay of thyroid hormones and related substances. Available at: https://www.thyroidmanager.org (accessed: February 2017). South Dartmouth, MA: Endocrine Education Inc.
12. Stockigt JR, Lim CF. Medications that distort *in vitro* tests of thyroid function, with particular reference to estimates of serum free thyroxine. *Best Prac Res Clin Endocrinol Metab*, 2009; **23**: 753–67.
13. Lee RH, Spencer CA, Mestman JH, *et al*. Free T_4 assays are flawed during pregnancy. *Am J Obstet Gynecol*, 2009; **260**: 260e1–e6.
14. Feldt-Rasmussen U, Bliddal S, Rasmussen AK, Boas M, Hilsted L, Main K. Challenges in interpretation of thyroid function tests in pregnant women with autoimmune thyroid disease. *J Thyroid Res*, 2011; **2011**: 598712.
15. Welsh KJ, Soldin SJ. Diagnosis of endocrine disease: how reliable are free thyroid and total T3 hormone assays? *Eur J Endocrinol*, 2016; **175**: R255–63.
16. Brabant G, Prank K, Ranft U, Schuermeyer T, Wagner TO, Hauser H. Physiological regulation of circadian and pulsatile thyrotropin secretion in normal man and woman. *J Clin Endocrinol Metab*, 1990; **70**: 403–9.
17. Karmisholt J, Andersen S, Laurberg P. Variation in thyroid function tests in patients with stable untreated subclinical hypothyroidism. *Thyroid*, 2007; **18**: 303–8.
18. Ross HA, Menheere PPCA, Thomas CMG, *et al*. Interference from heterophilic antibodies in seven current TSH assays. *Ann Clin Biochem*, 2008; **45**: 616–23.
19. Beck-Peccoz P, Brucker-Davis F, Persani L, Smallridge RC, Weintraub BD. Thyrotropin-secreting pituitary tumors. *Endocr Rev*, 1996; **17**: 610–38.
20. Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet*, 2012; **379**: 1142–54.
21. Meyerovitch J, Rotman-Pikielny S, Sherf M, Battat E, Levy Y, Surks MI. Serum thyrotropin measurements in the community: five-year follow-up in a large network of primary care physicians. *Arch Int Med*, 2007; **167**: 1533–8.
22. Ringel MD, Mazzaferri EL. Subclinical thyroid dysfunction: can there be consensus about a consensus? *J Clin Endocrinol Metab*, 2006; **90**: 588–90.
23. Beck-Peccoz P, Persani L. Variable biological activity of thyroid-stimulating hormone. *Eur J Endocrinol*, 1994; **131**: 331–40.
24. Verheecke P. Free triiodothyronine concentration in serum of 1050 euthyroid children is inversely related to their age. *Clin Chem*, 1997; **43**: 963–7.
25. Bliddal S, Feldt-Rasmussen U, Boas M, *et al*. Gestational age-specific reference ranges from different laboratories misclassify pregnant women's thyroid status: comparison of two longitudinal prospective cohort studies. *Eur J Endocrinol*, 2013; **170**: 329–39.
26. Norden AGW, Jackson RA, Norden LE, *et al*. Misleading results from immunoassays of serum free thyroxine in the presence of rheumatoid factor. *Clin Chem*, 1997; **43**: 957–62.
27. Mendel CM, Frost PH, Kunitake ST, Cavalieri RR. Mechanism of the heparin-induced increase in the concentration of free thyroxine in plasma. *J Clin Endocrinol Metab*, 1987; **65**: 1259–64.
28. Janssen ST, Janssen OE. Directional thyroid hormone distribution via the blood stream to target sites. *Mol Cell Endocrinol*, 2017; **458**: 16–21.
29. Pappa T, Ferrara AM, Refetoff S. Inherited defects of thyroxine-binding proteins. *Best Pract Res Clin Endocrinol Metab*, 2015; **29**: 735–47.
30. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. Thyroid, 2014; **24**: 1456–65.
31. Williams FLR, Mires GJ, Barnett C, *et al*. Transient hypothyroxinemia in preterm infants: the role of cord sera thyroid hormone levels adjusted for prenatal and intrapartum factors. *J Clin Endocrinol Metab*, 2005; **90**: 4599–606.
32. Loy M, Perra E, Melis A, *et al*. Color-flow Doppler sonography in the differential diagnosis and management of amiodarone-induced thyrotoxicosis. *Acta Radiol*, 2007; **48**: 628–34.
33. Kim MJ, Kim E-K, Kim BM, *et al*. Thyroglobulin measurements in fine-needle aspirate washouts: the criteria for neck node dissection for patients with thyroid cancer. *Clin Endocrinol*, 2009; **79**: 145–51.
34. Weiss R, Wu SY, Refetoff S. Diagnostic tests of the thyroid: tests that assess the effects of thyroid hormone on body tissues. In: DeGroot LJ, Jameson LJ, eds. *Endocrinology*, 5th edn. Philadelphia, PA: Saunders, 2006: 1915–6.
35. Fazio S, Biondi B, Carella C, *et al*. Diastolic dysfunction in patients on thyroid-stimulating hormone suppressive therapy with levothyroxine: beneficial effect of β -blockade. *J Clin Endocrinol Metab*, 1995; **80**: 2222–6.
36. Feldt-Rasmussen U, Rasmussen AK. Drug effects and thyroid function. In: Huhtaniemi I, Martini L, eds. *Encyclopedia of Endocrine Diseases*, 2nd edn. Oxford: Elsevier, 2018: Volume 4.
37. Stevenson HP, Archbold GPR, Johnston P, Young IS, Sheridan B. Misleading serum free thyroxine results during low molecular weight heparin treatment. *Clin Chem*, 1998; **44**: 1002–7.
38. Lazarus JH. Lithium and thyroid. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 723–33.
39. Surks MI, DeFesi CR. Normal serum free thyroid hormone concentrations in patients treated with phenytoin or carbamazepine. *JAMA*, 1996; **275**: 1495–8.
40. Madeddu G, Spanu A, Chessa F, *et al*. Thyroid function in HIV patients treated with highly active antiretroviral therapy. *Clin Endocrinol*, 2006; **64**: 575–83.

41. Touzot M, Le Beller C, Touzot F, Louet AL, Piketty C. Dramatic interaction between levothyroxine and lopinavir/ritonavir in a HIV-infected patient. *AIDS*, 2006; **20**: 1210–12.
42. Lanzafame M, Trevenzoli M, Faggian F, *et al*. Interaction between levothyroxine and indinavir in a patient with HIV infection. *Infection*, 2002; **30**: 54–5.

3.1.5 Non-Thyroidal Illness (NTI)

Robin P. Peeters and Anita Boelen

Serum and Tissue Thyroid Hormone Concentrations in NTI

Within 2 hours after the onset of acute illness (and after 24–36 hours of fasting), T_3 levels decrease and rT_3 levels rise [1]. The magnitude of these changes is related to the severity of the disease. T_3 levels decrease progressively with increasing severity of disease without reaching a nadir, whereas rT_3 increases in relation to severity of disease, but reaches a plateau. rT3 is not invariably elevated in all causes of NTI [2, 3].

In mild illness, total and free T_4 (fT_4) levels may rise initially after the onset of disease but in severely ill patients, T_4 levels drop as well. Both low T_4 and T_3, as well as high rT_3 are associated with a worse prognosis [1, 4]. Thyroid-stimulating hormone (TSH) may rise briefly for ~2 hours after the onset of disease, but despite the drop in serum T_3 (and eventually also in T_4), TSH usually remains within the (low) normal range.

Duration of illness is another important determinant of the thyrotropic profile in NTI [5, 6]. Patients requiring intensive care enter a more chronic phase of NTI, and the low levels of thyroid hormone in prolonged illness have a more neuroendocrine origin. Pulsatility and circadian variation in TSH secretion is diminished in prolonged illness and hypothalamic thyrotropin-releasing hormone (TRH) mRNA expression in patients who died from chronic severe illness is low compared to patients who died from an acute trauma [7]. Low TSH secretion and TRH expression correlate with low serum T_3. Reverse T_3 levels remain elevated or may return back to normal with the decrease in serum T_4 (see **Figure 12.1.2.2b** in Chapter 12.1.2 for a simplified overview of the major changes occurring within the thyroid axis during the acute and chronic phase of critical illness.). Type of illness and feeding status are also important [2, 3, 8].

Both low T_4, low T_3, and high rT_3 levels are associated with a worse prognosis. A study in 451 patients who received intensive care for at least 5 days, showed that TSH, T_4, T_3, and the T_3/rT_3 ratio increased in patients who survived, but not in non-survivors [4] (see **Figures 3.1.5.1–3.1.5.3**). Patients who died after critical illness had lower tissue T_3 compared to patients who had died acutely, but the severity varies from one organ to another [9]. Low serum T_4 and T_3 correlated well with local concentrations in liver and skeletal muscle [10].

Neuroendocrine Changes in NTI

Acute Illness

After the onset of acute disease, TSH levels may rise briefly for ~2 hours, despite the ongoing decrease in circulating T_3 (and T_4), TSH usually remains within the (low) normal range. It suggests an altered feedback at the level of the hypothalamus and/or pituitary. The physiological nocturnal TSH surge is absent in the acute phase of illness [5] (see **Figure 3.1.5.3**). These changes cannot be attributed to exogenous glucocorticoids or dopamine, since serum TSH is also in the low-normal range in patients without these drugs.

Different mechanisms have been proposed for this altered feedback. Studies in rodents during fasting and after lipopolysaccharide (LPS) injection, which is a model for acute inflammation, show no compensatory rise in TRH in contrast to hypothyroid animals [11]. Local expression of thyroid hormone-activating type 2 deiodinase (D_2) is increased resulting in higher local concentrations of hypothalamic T_3 [12, 13]. Inflammatory mediators are involved in the inflammation-induced rise in D_2 expression; RelA (the p65 subunit of NF-κB) is able to bind the Dio2 promoter and increases D_2 expression in primary tanycytes after LPS stimulation *in vitro* [14]. Targeting RelA in tanycytes of mice results in reduced hypothalamic Dio2 and TRH mRNA expression after LPS. These changes are not required for the downregulation of TSHβ expression and serum thyroid hormone (TH) [15]. The role of local T_3 in the hypothalamus is supported by the observation that $TR\beta^{0/0}$ mice display an impaired illness-induced TRH decrease and TRβ signalling important for the feedback of T_3 on TRH neurons [16, 17], and in agreement with a downregulation of TRH in fasting and acute illness. Animal data suggest that an altered transmembrane transport of thyroid hormone at the pituitary and/or hypothalamus may be involved in the altered setpoint of the HPT-axis as well as an enhanced occupancy of nuclear T_3 receptors in the pituitary thyrotrophs.

In the acute phase of NTI, serum cytokines are usually high. Injection of cytokines such as IL-1, IL-6, and TNF-α is at least partially able to mimic the thyrotropic alterations of the acute stress response. Studies in IL-12 and IL-18 knockout mice show that these cytokines are also involved. However, cytokine antagonists fail to restore thyroid hormones, both in animals (IL-1, IL-6, TNF, Interferon) and in humans (IL-1) [18, 19].

High endogenous cortisol may also contribute to the blunted TSH response in acute illness.

Chronic Illness

Patients with prolonged NTI have more severe central dysfunction. In addition to the absent nocturnal TSH surge, TSH pulsatility diminishes dramatically, and hypothalamic TRH expression is reduced [5, 6]. Both correlate with low serum T_3 [7]. Patients who die after severe illness have less than 50% of hypothalamic and pituitary T_3 content compared to patients who die acutely [9]. The positive correlation of TSH secretion and TRH expression with serum T_3 levels suggests a more central origin of the low T_3 syndrome in prolonged illness. Vice versa, an increase in TSH is a marker for recovery, suggesting recovery from the low T_3 syndrome is also initiated centrally (see Figure 3.1.5.2 in which the recovery of TSH precedes the recovery of the T_3/rT_3 ratio). In addition, continuous infusion with TRH (especially when combined with a growth hormone secretagogue) is able to (partially) restore serum TSH, T_4, and T_3 in prolonged critical illness, both in humans and in animals [20].

The pathophysiology of the suppressed HPT-axis is incompletely understood. Circulating cytokines are usually low in the chronic phase of NTI and other mechanisms must be involved. Upregulation of hypothalamic D_2, and/or a downregulation of D_3 could suppress TRH expression via relatively high hypothalamic T_3. However, hypothalamic and pituitary T_3 levels are low in patients who die after prolonged illness [9]. This makes an important contribution of hypothalamic and pituitary deiodinases and/or transporters to central suppression less likely.

Other pathways, such as the melanocortin signalling pathway and neuropeptide Y (NPY), might be involved in regulating hypothalamic TRH [6, 21, 22]. However, their exact role is not yet elucidated [22]. Exogenous glucocorticoids and dopamine are known to suppress the hypothalamus–pituitary–thyroid axis, and perhaps prolonged hypercortisolism and/or endogenous dopamine in these patients may also play a role (see Box 3.1.5.1).

Extrathyroidal (Peripheral) Changes in NTI

In the acute phase of critical illness and after starvation, changes in thyroid hormone levels are mainly caused by changes in peripheral metabolism of thyroid hormones and serum binding proteins [1]. In the chronic phase these changes persist but decreased thyroidal T_4 production is superimposed. The serum T_3/rT_3 ratio is the most accurate reflection of peripheral metabolism, since this ratio is independent of variations in binding proteins and independent of decreased thyroidal T_4 production.

Box 3.1.5.1 Essential information

- In critical illness, a decrease in serum T_3 and increase in serum rT_3 are the most characteristic and persistent abnormalities.
- The magnitude of these changes is related to severity of disease and associated with a worse prognosis.
- Feeding status also seems important with regard to the changes in peripheral thyroid hormone levels, an altered feedback setting at the hypothalamus-pituitary level, a decreased activation, and an increased inactivation of thyroid hormone occur in NTI.
- The acute and more chronic phase of NTI should be seen as two separate entities. An altered peripheral metabolism is the major player in the acute situation, whereas central dysfunction is more important in the chronic phase of severe illness.
- There is currently no evidence that NTI (in the acute or in the chronic phase) should be treated with thyroid hormone.
- Possible benefits of treatment with hypothalamic releasing factors should be the subject of future studies.

Deiodination of Thyroid Hormones in NTI

The fall in serum T_3 and increase in serum rT_3 in the acute phase of critical illness and in fasting are largely due to a decreased conversion of T_4 to T_3 and of rT_3 to T_2 [1]. Recent data suggest that the immediate fall in T_3 is related to concomitant fasting rather than to illness *per se*. The randomized controlled trial 'Early versus Late Parenteral Nutrition in Critically Ill Adults (EPaNIC)' showed that not feeding critically ill patients early, (thereby accepting pronounced macronutrient deficit that in these patients), improved clinical outcome [23]. Early feeding prevented a large part of the early drop in plasma TSH, T_4, and T_3, and also prevented the rise in rT_3 [24]. Similar data were obtained in a rabbit model of NTI, in which early feeding also reversed the low liver D_1 and high liver D_3 [25]. Fasting-induced NTI may thus be an adaptive beneficial response [8].

Liver and skeletal muscle biopsies obtained within minutes after death of intensive care unit patients demonstrate low liver D_1 activity compared to healthy individuals, except for patients who died acutely from severe brain damage (Figure 3.1.5.1) [3]. Low D_1 activity clearly correlated with high rT_3 and a low T_3/rT_3 ratio, independent of duration of illness. D_1 activities also correlate with local T_3 and T_3/rT_3 ratios in liver [10].

Altered D_2 activity may not play a role in the pathogenesis of the low T_3 syndrome in prolonged critical illness [26].

Induction of D_3 activity was demonstrated in liver and muscle samples of NTI patients, whereas these tissues normally do not express D_3 (Figure 3.1.5.1). High liver and high muscle D_3 activity was associated with high serum and local tissue rT_3 levels. D_3 induction was independent of duration of illness. Animal studies showed that the decrease of serum T_3 was independent of D_3 activity [27]). Thus, a downregulation of D_1 (and in the acute phase of illness also D_2) is the important factor contributing to low levels of T_3, whereas the high levels of rT_3 seem to be due to a combination of low D_1 and high D_3. D3 is also expressed in granulocytes during infection (see Figure 3.1.5.4) [28–30]. Bacterial meningitis results in a strong increase of T_4 and rT_3 concentrations in cerebrospinal fluid (CSF). This altered thyroid hormone profile is consistent with elevated D_3 activity in infiltrating neutrophils at the site of infection [31].

Tissue deiodinase activities and serum thyroid hormone levels are significantly associated with cause of death [3]. A post-mortem study in over 60 patients demonstrated that liver D_1 activity and serum T_3/rT_3 were highest in patients who died from severe brain damage, intermediate in those who died from sepsis or excessive inflammation and were lowest in patients who died from cardiovascular collapse (see Figure 3.1.5.5). Liver D_3 showed an opposite relationship. There was no relation between deiodinase activities and a marker of inflammation (C-reactive protein) but patients who needed inotropes and/or those requiring dialysis because of acute renal failure had a lower liver D_1 activity and higher liver and muscle D_3 activity. This suggests that poor tissue perfusion and cellular hypoxia may be an important determinant regulating deiodinase activities *in vivo*. D_3 activity and D_3 mRNA are increased

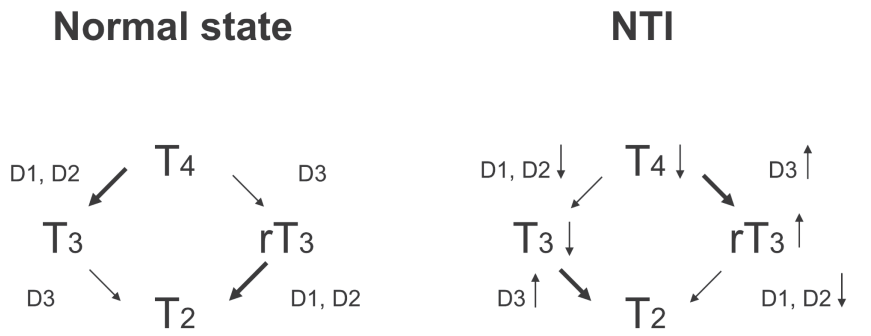


Figure 3.1.5.1 Relationship between the different iodothyronines and the outer (D_1, D_2) and inner ring (D_3) deiodination by the three deiodinases on the left. Like D_3, D_1 has some inner ring deiodination capacity *in vitro*, but there is currently no evidence that this is of any significance *in vivo* (see also Chapter 3.1.2). This is therefore omitted from the figure. Observed changes in deiodinase activities and iodothyronine levels during critical illness, both in the acute and chronic phase of critical illness on the right.

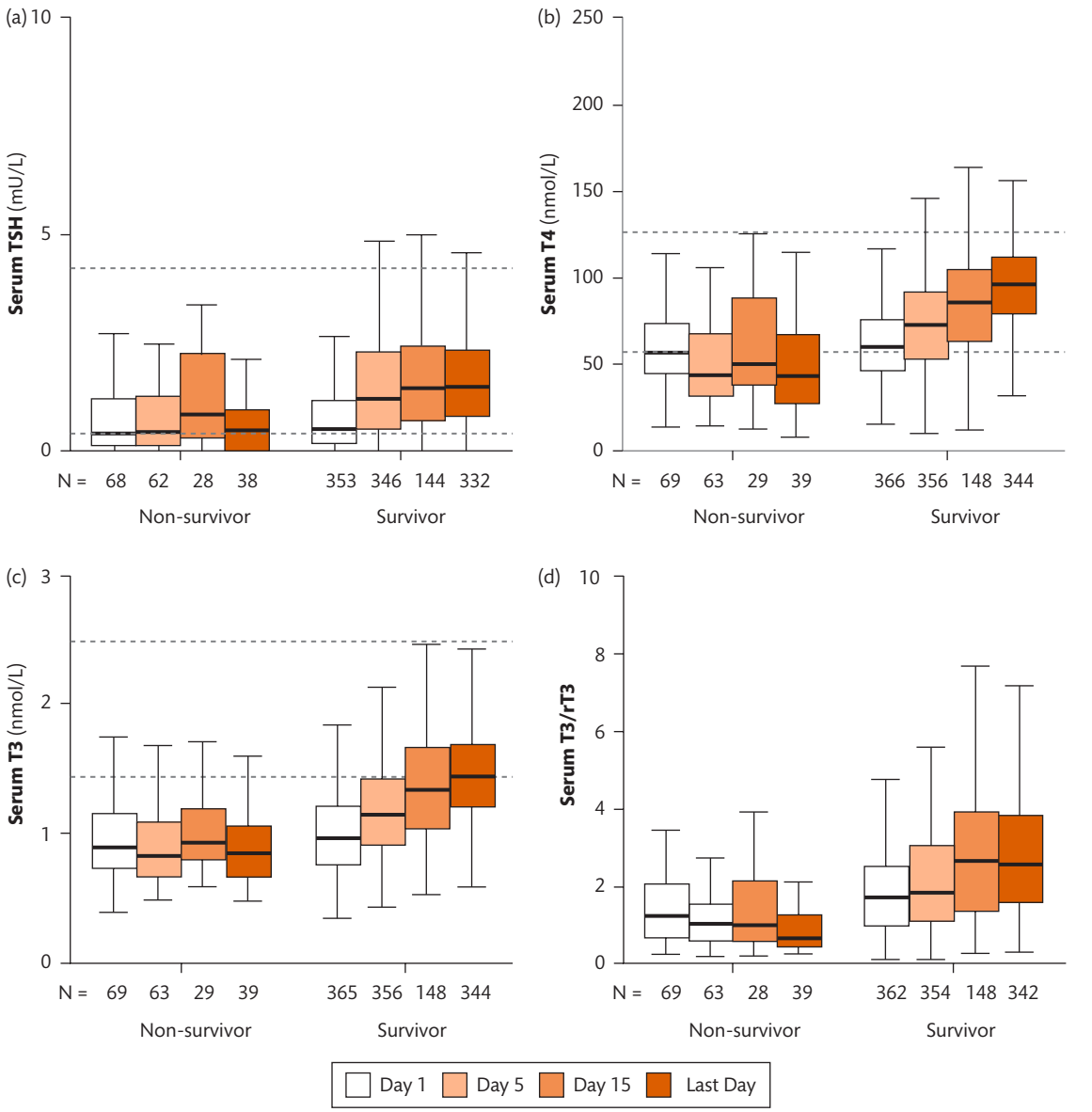


Figure 3.1.5.2 Serum TSH, T_4, and T_3 levels and the T_3 to rT_3 ratio at day 1, 5, 15, and the last day of ICU stay in survivors and non-survivors. From day 5 onward, serum TSH, T_4, and T_3 increased in patients who survived, whereas there was no such pattern in patients who died (a–c). The serum T_3 to rT_3 ratio increased in survivors from day 5 to the last day, whereas it did not alter or even decreased in non-survivors (d). On the last day of ICU stay, the majority of patients had TSH and T4 levels within the normal range, whereas T_3 was still low. The blocked line represents the normal values.

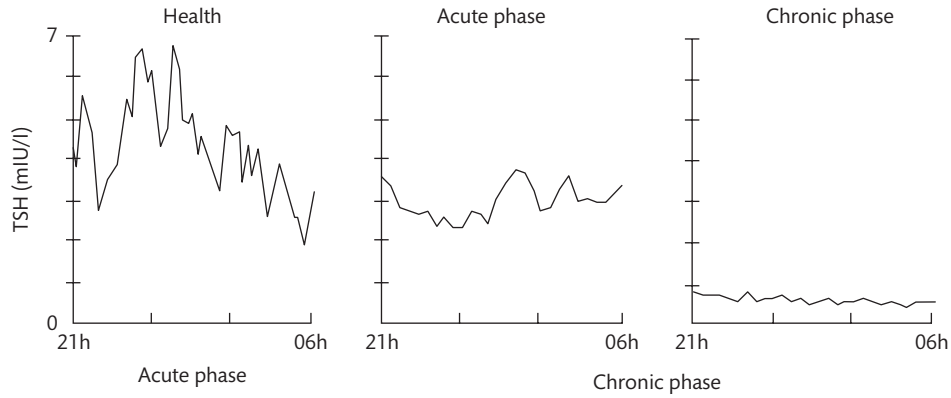


Figure 3.1.5.3 The nocturnal serum concentration profiles of thyrotropin in critical illness are abnormal and differ markedly between the acute and chronic phase of the disease.

Modified with permission from Van den Berghe G, de Zegher F, Bouillon R. Acute and prolonged critical illness as different neuroendocrine paradigms. *J Clin Endocrinol Metab* 1998;83:1827–34. Copyright © 1998, Oxford University Press.

by hypoxia, and hypoxia-inducible factor 1 induces local thyroid hormone inactivation [32].

Transmembrane Transport of Thyroid Hormones in NTI

Cellular uptake of thyroid hormones is rate limiting for subsequent intracellular metabolism and nuclear T_3 binding [33]. Kinetic studies have shown that fasting and non-thyroidal illness result in attenuation of uptake of liver T_4 and rT_3, probably via decreased concentrations of intracellular adenosine triphosphate (ATP). Inhibition of thyroid hormone uptake can also be caused by non-esterified fatty acids and bilirubin, both elevated in critical illness, and certain drugs such as amiodarone.

In critically ill patients, neither liver, nor skeletal muscle MCT8 expression were related to the ratio of the serum over tissue concentration of T_4, T_3, or rT_3 [10]. This suggests MCT8 is not crucial in the transport of these iodothyronines over the plasma membrane in liver and skeletal muscle. Liver TH transporter expression differs between several NTI models; acute illness lowers MCT8 and MCT10 expression whereas chronic inflammation decreases MCT8 mRNA and increases MCT10 mRNA. Lethal bacterial sepsis affects hepatic TH transport marginally without modulating liver TH concentrations [34]. However, this does not exclude an important role of other transporters.

Thyroid Hormone Receptors in NTI

Little is known about the regulation of thyroid hormone receptors (TR) in NTI. In rats, starvation results in a decreased expression and occupancy of hepatic TR [35]. In peripheral mononuclear cells in humans, an increased expression of both TRα and TRβ has been demonstrated in patients with chronic liver and renal disease, whereas in patients in the intensive care unit only TRβ mRNA was increased [36]. In patients with liver disease, both liver TRα (20-fold) and liver TRβ (5-fold) were increased compared to healthy liver controls. A post-mortem study in 58 subjects who had died in the intensive care unit showed an increased expression of the TRα1/TRα2 ratio (active isoform/dominant negative isoform), which was positively related to severity of disease and age [37]. In this study, no relation between severity of disease and TRβ1 expression was observed.

The clinical relevance of these changes is not clear. An increase in the expression of active receptor isoforms might be an adaptive response to decreasing levels of thyroid hormone. On the other hand, a higher TR expression with low levels of T_3 will lead to an increase

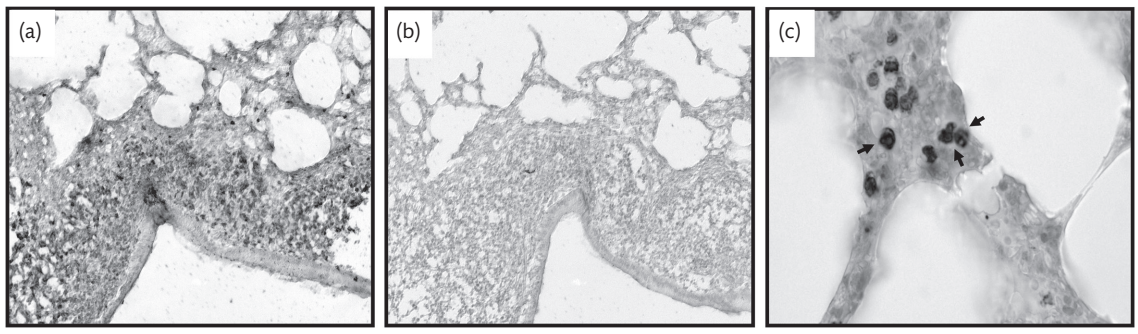


Figure 3.1.5.4 An overview of a lung section showing *S. pneumoniae* induced infiltrate after 48 hours stained both with anti-D_3 antibody (a) and pre-immune serum (b). Note clearly stained D3-positive granulocytes in the infiltrated area that are shown in more detail (indicated by arrows) (c).

Adapted with permission from Boelen *et al.* Type 3 deiodinase is highly expressed in infiltrating neutrophilic granulocytes in response to acute bacterial infection. *Thyroid* 18:1095–1103, 2008, Copyright 2008, Mary Ann Liebert Inc.

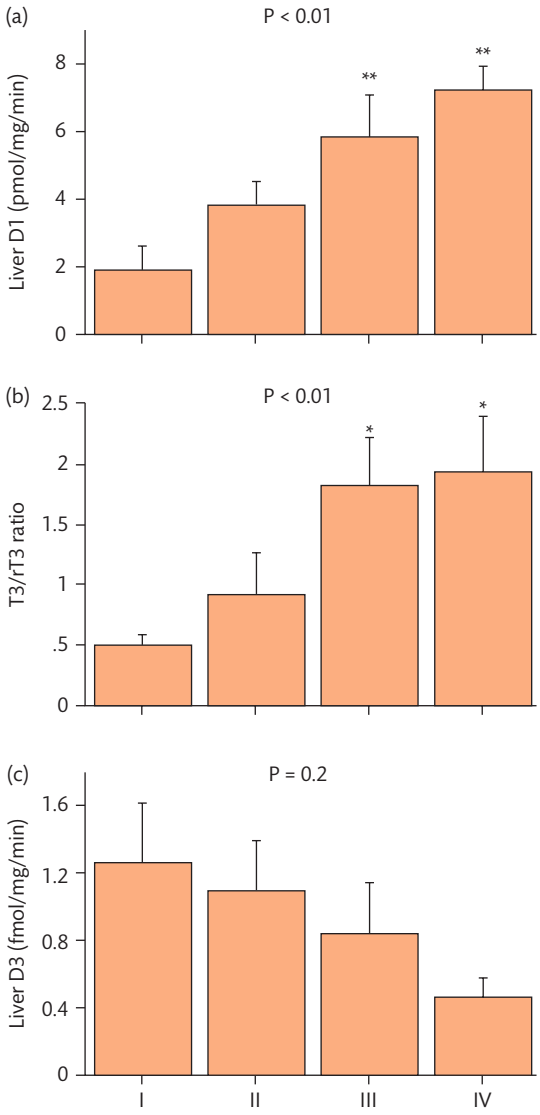


Figure 3.1.5.5 Correlation of liver D_1 (a) and liver D_3 (b) activities, and the T_3/rT_3 ratio (c) with cause of death. Patients are divided into four different groups based on cause of death. I, Cardiovascular collapse (n = 5); II, multiple organ failure with sepsis (n = 21); III, multiple organ failure with systemic inflammatory response syndrome (n = 14); IV, severe brain damage (n = 4). Liver D_1 activity and serum T_3/rT_3 ratio showed a significant relation with cause of death (P <0.01), whereas liver D_3 activity showed an opposite trend. **, P <0.01 vs. group I; *, P <0.05 vs. group I. Data represent means ± SEM and P values were obtained with ANOVA and Fisher's least significant difference for multiple comparisons.

in the proportion of unliganded receptors, which would have an opposite effect. Interestingly, no relation was demonstrated between liver TRβ1 mRNA and serum thyroid hormones in critically ill patients, although D_1 expression is regulated by T_3 via TRβ1 [37].

Other Metabolic Pathways in NTI

Sulphation mediates the rapid and irreversible degradation of iodothyronines by D_1. Inner ring deiodination of T_4 and T_3 by D_1 is markedly facilitated after sulphation, whereas outer ring deiodination of T_4 is completely blocked after sulphation.

Elevated levels of T_4S and T_3S/T_3 ratios have been reported in NTI patients, and post-mortal serum T_4S in critically ill patients are positively correlated with the length of stay in the intensive care unit [38, 39]. Low hepatic D_1 activity in these patients plays an important role in the increased levels of T_4S.

Glucuronidation in NTI may be important with regard to the use of several drugs. Carbamazepine, phenytoin, and rifampicin particularly induce hepatic glucuronidation. This may lower T_4, but T_3 and TSH are usually unaffected.

Ether linked cleavage (ELC) involves the breaking of the ether bridge in between the two tyrosines, yielding diiodotyrosine (DIT) as a main product, and catalysed by peroxidases such as myeloperoxidase (MPO) in leukocytes. Its role is rather limited, but DIT increases during sepsis [40] attributing to TH clearance.

Alterations in Thyroid Hormone Binding

Thyroxine binding globulin (TBG), transthyretin (TTR), and albumin are decreased in NTI reflecting the catabolic state of the patient. However, increased TBG levels can be present in liver disease. Various drugs cause alterations in serum binding, either by decreasing TBG (e.g. glucocorticoids), or by displacing thyroid hormones from binding proteins (e.g. acetyl salicylic acid, furosemide). In heparin-treated patients, the measured concentration of serum fT_4 can be higher than the true *in vivo* concentration. This is the result of heparin-induced lipase activity during sample storage and incubation, resulting in *in vitro* generation of non-esterified fatty acids (NEFA) which displace T_4 and T_3 from TBG (see also Chapter 3.1.3). Low molecular weight heparin preparations have a similar effect.

Low serum binding of thyroid hormone in NTI due to the presence of a circulating binding inhibitor has been proposed in older studies [1]. However, exogenous T_4 administration can easily replenish the T_4 pool in patients with prolonged illness, making it unlikely that such a binding inhibitor is an important cause of low serum T_4 [41].

The Diagnosis of Thyroid Disease in Critical Illness

Evaluation of thyroid status in NTI can be very difficult, especially in patients in the intensive care unit. The value of clinical examination should not be underestimated. Thus, the presence of eye signs (ophthalmic Graves' disease), goitre, and a family history of thyroid disease or autoimmune disease in general, are important points that may support the diagnosis of autoimmune thyroid disease (see Box 3.1.5.2).

Because of the numerous changes in serum thyroid hormones in NTI, and because there is currently no evidence that these changes should be treated, thyroid function should not be tested in critically ill patients, unless there is strong suspicion of thyroid disease. In unselected hospitalized patients, TSH was less than 0.1 mU/L in 3.1%, whereas TSH was above 20 mU/L in 1.6% of patients [42]. When thyroid function is tested, measurement of TSH alone is often not sufficient. Most fT_4 assays are unreliable

Box 3.1.5.2 The diagnosis of thyroid disease in critical illness

- In critical illness, thyroid function should only be tested if there is a strong suspicion of thyroid disease.
- A normal TSH virtually excludes thyrotoxicosis or hypothyroidism. However, when thyroid function is tested, measurement of TSH alone is often not sufficient.
- Most fT_4 assays are unreliable in critical illness.
- Dopamine and glucocorticoids suppress TSH secretion.
- Hyperthyroidism:
  - A low TSH is compatible with both NTI and thyrotoxicosis.
  - Nearly all patients with a low but detectable TSH will have normal thyroid function tests after recovery from illness.
  - Approximately 75% of patients with NTI and a TSH <0.01 mU/L have hyperthyroidism.
  - Serum T_4 and T_3 should be high (or high-normal) in hyperthyroidism, and low (or low-normal) in NTI.
- Hypothyroidism:
  - In patients recovering from NTI, TSH levels may become temporarily elevated.
  - Most patients with an elevated TSH <20 mU/L will have normal thyroid function tests after recovery from illness, especially when thyroid peroxidase and thyrogobulin antibodies are negative.
  - In patients with TSH >20 mU/L, hypothyroidism is permanent in only 50%.
  - Hypothyroid patients have serum T_4 and T_3 (and a relatively higher T_3/T_4 ratio) compared to patients recovering from NTI.
  - Low serum rT_3 suggests hypothyroidism, whereas high rT_3 supports of NTI.
- If no definite diagnosis can be made, thyroid function tests should be repeated after recovery from illness.

in critical illness, due to alterations in binding proteins, increased use of heparin in an ICU setting, and the possible presence of circulating binding inhibitors. Therefore, the total thyroid hormone should be measured as well.

Hyperthyroidism

In an NTI patient, suspected of hyperthyroidism, serum TSH is the most helpful test. If serum TSH is still within the normal range, the presence of thyrotoxicosis is virtually excluded. But when serum TSH is low, it could be a consequence of NTI or of hyperthyroidism. However, NTI almost never results in TSH less than 0.01 mU/L [43, 44]. Nearly all patients with a low but detectable TSH, will have normal thyroid function tests after recovery from illness. On the other hand, approximately 75% of patients with the low T_3 syndrome and a TSH less than 0.01 mU/L have hyperthyroidism [45]. Interpretation of serum TSH becomes more difficult in patients treated with TSH-suppressing agents such as dopamine and corticosteroids. Additional measurement of T_4 and T_3 levels is mandatory, but should be interpreted with care. T_4 levels are low in approximately 50% of critically ill patients, and T_3 levels are low in the majority of patients, which could mask active hyperthyroidism. However, serum T_4 and T_3 levels should be high (or high-normal) in hyperthyroidism, and low (or low-normal) in NTI.

Hypothyroidism

The diagnosis seems straightforward for a critically ill patient, with suspected hypothyroidism if serum TSH is elevated. However, in patients recovering from NTI, TSH levels may become temporarily elevated. Even in patients with TSH of more than 20 mU/L, hypothyroidism is permanent in only ~50% of cases [42]. Patients with permanent hypothyroidism have significantly lower levels of T_4 and T_3. Most patients with an elevated TSH less than 20 mU/L will have normal thyroid function tests after recovery, especially if thyroid peroxidase and thyroglobulin antibodies are negative.

In central hypothyroidism serum TSH is usually low, and differentiation from NTI becomes very difficult. Other pituitary deficiencies and related clinical signs are commonly present in these patients, but prolonged critical illness often leads to suppression of other neuroendocrine axes as well [5]. Serum rT_3 may be helpful in some cases, but rT_3 assays are rarely available. In general, a high T_3/T_4 ratio and a low rT_3 favour the presence of hypothyroidism. If no definite diagnosis can be established, thyroid function test should be repeated after recovery from illness.

Should Patients with NTI Be Treated with Thyroid Hormone?

Both in acute and in prolonged critical illness, low levels of thyroid hormone are associated with a higher mortality rate, but it remains controversial whether NTI is an adaptation protecting against catabolism or a maladaptation. It is important to re-emphasize the teleological differences between the acute and chronic phase of severe illness (see Figure 12.1.2.2b in Chapter 12.1.2). Acute changes within the thyroid axis after the onset of critical illness (low T_3 and elevated rT_3) are similar to the changes observed in starvation. These changes have been interpreted as an attempt to save energy expenditure and protein wasting which does not need intervention. This is supported by recent studies suggesting that (at least part of) the immediate changes in thyroid function tests during acute critical illness are related to the concomitant fasting rather than the illness *per se*. Thyroid hormone replacement in fasting subjects results in an increased nitrogen excretion and negative nitrogen balance, suggesting catabolism. Whether this also applies to the changes in the acute, and especially in the more chronic phase of critical illness, is controversial. However, the improved prognosis of ICU patients that were not fed early suggests that this fasting-induced NTI in the acute phase of critical illness may be an adaptive beneficial response which should not be corrected [8]. In humans, only few studies with T_4 treatment have been performed, with rather low sample sizes. So far, no clear beneficial effect has been demonstrated.

Because of the decreased T_4 to T_3 conversion in NTI, T_3 treatment may be a better choice. However, T_3 treatment may be harmful as well (see [46]). In the chronic phase of critical illness, altered thyroid hormone levels appear to have a more central origin, although peripheral metabolism is also altered. Studies performed in fasting subjects and in patients with acute critical illness should therefore not be extrapolated to the chronic phase of severe illness.

Tissue thyroid hormones were measured in patients who stayed on the intensive care unit for more than 5 days [10]. Some patients were treated with a combination of T_4 and T_3 if they had a serum T_4 of less than 50 nmol/L, a normal TBG, and clinical signs of hypothyroidism. Higher serum T_3 levels in treated patients were associated with higher tissue T_3 contents. However, the fourfold increase in liver T_3 in treated patients was disproportional compared to the

twofold increase in serum and muscle T_3. TSH levels were suppressed in treated patients suggesting overtreatment although their serum T_3 was still in the low or low-normal range. So, if patients are substituted with thyroid hormone, should we aim for thyroid hormone levels within or still below the normal range?

Intervention with hypothalamic releasing factors in patients with chronic critical illness has the advantage that the negative feedback inhibition of thyroid hormone on the pituitary is maintained, thereby providing a safer therapy option. It has been shown in patients with prolonged critical illness and in an animal model, that continuous infusing of TRH in combination with a growth hormone secretagogue is able to restore thyroid hormone levels. In these patients, therapy resulted in a reduction of catabolic markers. Whether this would also result in a beneficial effect on mortality remains to be addressed in future studies.

REFERENCES

1. Docter R, Krenning EP, de Jong M, Hennemann G. The sick euthyroid syndrome: changes in thyroid hormone serum parameters and hormone metabolism. *Clin Endocrinol*, 1993; **39**: 499–518.
2. Kaptein EM. Thyroid hormone metabolism and thyroid diseases in chronic renal failure. *Endocr Rev*, 1996; **17**: 45–63.
3. Peeters RP, Wouters PJ, Kaptein E, van Toor H, Visser TJ, Van den Berghe G. Reduced activation and increased inactivation of thyroid hormone in tissues of critically ill patients. *J Clin Endocrinol Metab*, 2003; **88**: 3202–11.
4. Peeters RP, Wouters PJ, van Toor H, Kaptein E, Visser TJ, Van den Berghe G. Serum 3,3',5'-triiodothyronine (rT_3) and 3,5,3'-triiodothyronine/rT_3 are prognostic markers in critically ill patients and are associated with postmortem tissue deiodinase activities. *J Clin Endocrinol Metab*, 2005; **90**: 4559–65.
5. Van den Berghe G, de Zegher F, Bouillon R. Clinical review 95: acute and prolonged critical illness as different neuroendocrine paradigms. *J Clin Endocrinol Metab*, 1998; **83**: 1827–34.
6. Mebis L, Debaveye Y, Visser TJ, Van den Berghe G. Changes within the thyroid axis during the course of critical illness. *Endocrinol Metab Clin North Am*, 2006; **35**: 807–21, x.
7. Fliers E, Guldenaar SE, Wiersinga WM, Swaab DF. Decreased hypothalamic thyrotropin-releasing hormone gene expression in patients with nonthyroidal illness. *J Clin Endocrinol Metab*, 1997; **82**: 4032–6.
8. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. *Thyroid*, 2014; **24**: 1456–65.
9. Arem R, Wiener GJ, Kaplan SG, Kim HS, Reichlin S, Kaplan MM. Reduced tissue thyroid hormone levels in fatal illness. *Metabolism*, 1993; **42**: 1102–8.
10. Peeters RP, van der Geyten S, Wouters PJ, *et al*. Tissue thyroid hormone levels in critical illness. *J Clin Endocrinol Metab*, 2005; **90**: 6498–507.
11. Boelen A, Kwakkel J, Thijssen-Timmer DC, Alkemade A, Fliers E, Wiersinga WM. Simultaneous changes in central and peripheral components of the hypothalamus–pituitary–thyroid axis in lipopolysaccharide-induced acute illness in mice. *J Endocrinol*, 2004; **182**: 315–23.
12. Diano S, Naftolin F, Goglia F, Horvath TL. Fasting-induced increase in type II iodothyronine deiodinase activity and messenger ribonucleic acid levels is not reversed by thyroxine in the rat hypothalamus. *Endocrinology*, 1998; **139**: 2879–84.
13. Fekete C, Gereben B, Doleschall M, *et al*. Lipopolysaccharide induces type 2 iodothyronine deiodinase in the mediobasal hypothalamus: implications for the nonthyroidal illness syndrome. *Endocrinology*, 2004; **145**: 1649–55.
14. de Vries EM, Kwakkel J, Eggels L, *et al*. NFkappaB signaling is essential for the lipopolysaccharide-induced increase of type 2 deiodinase in tanycytes. *Endocrinology*, 2014; **155**: 2000–8.
15. de Vries EM, Nagel S, Haenold R, *et al*. The role of hypothalamic NF-kappaB signaling in the response of the HPT-axis to acute inflammation in female mice. *Endocrinology*, 2016; **157**: 2947–56.
16. Boelen A, Kwakkel J, Chassande O, Fliers E. Thyroid hormone receptor beta mediates acute illness-induced alterations in central thyroid hormone metabolism. *J Neuroendocrinol*, 2009; **21**: 465–72.
17. Alkemade A, Vuijst CL, Unmehopa UA, *et al*. Thyroid hormone receptor expression in the human hypothalamus and anterior pituitary. *J Clin Endocrinol Metab*, 2005; **90**: 904–12.
18. van der Poll T, Van Zee KJ, Endert E, *et al*. Interleukin-1 receptor blockade does not affect endotoxin-induced changes in plasma thyroid hormone and thyrotropin concentrations in man. *J Clin Endocrinol Metab*, 1995; **80**: 1341–6.
19. Boelen A, Platvoet-ter Schiphorst MC, Wiersinga WM. Immunoneutralization of interleukin-1, tumor necrosis factor, interleukin-6 or interferon does not prevent the LPS-induced sick euthyroid syndrome in mice. *J Endocrinol*, 1997; **153**: 115–22.
20. Van den Berghe G, Wouters P, Weekers F, *et al*. Reactivation of pituitary hormone release and metabolic improvement by infusion of growth hormone-releasing peptide and thyrotropin-releasing hormone in patients with protracted critical illness. *J Clin Endocrinol Metab*, 1999; **84**: 1311–23.
21. Lechan RM, Fekete C. Role of melanocortin signaling in the regulation of the hypothalamic-pituitary-thyroid (HPT) axis. *Peptides*, 2006; **27**: 310–25.
22. Fliers E, Unmehopa UA, Manniesing S, Vuijst CL, Wiersinga WM, Swaab DF. Decreased neuropeptide Y (NPY) expression in the infundibular nucleus of patients with nonthyroidal illness. *Peptides*, 2001; **22**: 459–65.
23. Casaer MP, Mesotten D, Hermans G, *et al*. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*, 2011; **365**: 506–17.
24. Langouche L, Vander Perre S, Marques M, *et al*. Impact of early nutrient restriction during critical illness on the nonthyroidal illness syndrome and its relation with outcome: a randomized, controlled clinical study. *J Clin Endocrinol Metab*, 2013; **98**: 1006–13.
25. Mebis L, Eerdekens A, Guiza F, *et al*. Contribution of nutritional deficit to the pathogenesis of the nonthyroidal illness syndrome in critical illness: a rabbit model study. *Endocrinology*, 2012; **153**: 973–84.
26. Mebis L, Langouche L, Visser TJ, Van den Berghe G. The type II iodothyronine deiodinase is up-regulated in skeletal muscle during prolonged critical illness. *J Clin Endocrinol Metab*, 2007; **92**: 3330–3.
27. Boelen A, Kwakkel J, Wieland CW, St Germain DL, Fliers E, Hernandez A. Impaired bacterial clearance in type 3

deiodinase-deficient mice infected with *Streptococcus pneumoniae*. *Endocrinology*, 2009; **150**: 1984–90.
28. Boelen A, Kwakkel J, Alkemade A, *et al*. Induction of type 3 deiodinase activity in inflammatory cells of mice with chronic local inflammation. *Endocrinology*, 2005; **146**: 5128–34.
29. Boelen A, Boorsma J, Kwakkel J, *et al*. Type 3 deiodinase is highly expressed in infiltrating neutrophilic granulocytes in response to acute bacterial infection. *Thyroid*, 2008; **18**: 1095–103.
30. van der Spek AH, Bloise FF, Tichgelaar W, *et al*. The thyroid hormone inactivating enzyme type 3 deiodinase is present in bactericidal granules and the cytoplasm of human neutrophils. *Endocrinology*, 2016: **157**: 3293–305.
31. van der Spek AH, Jim KK, Karaczyn A, *et al*. The thyroid hormone inactivating type 3 deiodinase is essential for optimal neutrophil function: observations from three species. *Endocrinology*, 2018; **159**: 826–35.
32. Simonides WS, Mulcahey MA, Redout EM, *et al*. Hypoxia-inducible factor induces local thyroid hormone inactivation during hypoxic-ischemic disease in rats. *J Clin Invest*, 2008; **118**: 975–83.
33. Hennemann G, Docter R, Friesema EC, de Jong M, Krenning EP, Visser TJ. Plasma membrane transport of thyroid hormones and its role in thyroid hormone metabolism and bioavailability. *Endocr Rev*, 2001; **22**: 451–76.
34. Boelen A, van der Spek AH, Bloise F, *et al*. Tissue thyroid hormone metabolism is differentially regulated during illness in mice. *J Endocrinol*, 2017; **233**: 25–36.
35. Carr FE, Seelig S, Mariash CN, Schwartz HL, Oppenheimer JH. Starvation and hypothyroidism exert an overlapping influence on rat hepatic messenger RNA activity profiles. *J Clin Invest*, 1983; **72**: 154–63.
36. Williams GR, Franklyn JA, Neuberger JM, Sheppard MC. Thyroid hormone receptor expression in the 'sick euthyroid' syndrome. *Lancet*, 1989; **2**: 1477–81.
37. Thijssen-Timmer DC, Peeters RP, Wouters P, *et al*. Thyroid hormone receptor isoform expression in livers of critically ill patients. *Thyroid*, 2007; **17**: 105–12.
38. Chopra IJ, Santini F, Hurd RE, Chua Teco GN. A radioimmunoassay for measurement of thyroxine sulfate. *J Clin Endocrinol Metab*, 1993; **76**: 145–50.
39. Peeters RP, Kester MHA, Wouters PJ, *et al*. Increased T4S levels in critically ill patients due to a decreased hepatic type I deiodinase activity. *J Clin Endocrinol Metab*, 2005; **90**: 6460–5.
40. Meinhold H, Gramm HJ, Meissner W, *et al*. Elevated serum diiodotyrosine (DIT) in severe infections and sepsis: DIT, a possible new marker of leukocyte activity. *J Clin Endocrinol Metab*, 1991; **72**: 945–53.
41. Brent GA, Hershman JM. Thyroxine therapy in patients with severe nonthyroidal illnesses and low serum thyroxine concentration. *J Clin Endocrinol Metab*, 1986; **63**: 1–8.
42. Spencer C, Eigen A, Shen D, *et al*. Specificity of sensitive assays of thyrotropin (TSH) used to screen for thyroid disease in hospitalized patients. *Clin Chem*, 1987; **33**: 1391–6.
43. Franklyn JA, Black EG, Betteridge J, Sheppard MC. Comparison of second and third generation methods for measurement of serum thyrotropin in patients with overt hyperthyroidism, patients receiving thyroxine therapy, and those with nonthyroidal illness. *J Clin Endocrinol Metab*, 1994; **78**: 1368–71.
44. Spencer CA, LoPresti JS, Patel A, *et al*. Applications of a new chemiluminometric thyrotropin assay to subnormal measurement. *J Clin Endocrinol Metab*, 1990; **70**: 453–60.
45. Stockigt JR. Guidelines for diagnosis and monitoring of thyroid disease: nonthyroidal illness. *Clin Chem*, 1996; **42**: 188–92.
46. Fliers E, Bianco AC, Langouche L, Boelen A. Thyroid function in critically ill patients. *Lancet Diabetes Endocrinol*, 2015; **3**: 816–25.

3.1.6 Thyroid Imaging: Nuclear Medicine Techniques

Steen Joop Bonnema and Laszlo Hegedüs

Introduction

Nuclear medicine techniques, used for many decades in the management of thyroid patients, are based on the principle that a radionuclide or radiopharmaceutical agent ('tracer') is administered orally or intravenously, and then taken up by the target tissue after minutes or hours. By the decay of the isotope, gamma rays are emitted, which can be detected by sensitive equipment located outside the body, mostly by a gamma camera with a pinhole collimator. Detecting sufficient decays ('counts') a pixilated image, referred to as a scintigram, can be produced. This does not have the resolution of images obtained by ultrasonography, computed tomography (CT), or magnetic resonance imaging (MRI), and is therefore not suitable for detailed morphological evaluation of the thyroid. More importantly, nuclear medicine techniques reflect the functional status of the thyroid, sometimes in a quantitative manner, which none of the other methods are able to. The radiation burden imposed by most nuclear medicine techniques is low, and usually less than that imposed by a plain X-ray of the thorax. Nevertheless, administration of radioisotopes during pregnancy or breastfeeding is contraindicated.

Table 3.1.6.1 Commonly used radioactive isotopes for thyroid imaging and therapy

Isotope	Half-life	Administered activity	Clinical use	Comments
^{123}I	13 h	Routine scan: 3.7–11.1 MBq	Routine thyroid imaging	High-quality images Enables uptake measurements
		Whole-body scan: 74–185 MBq	Imaging residual tumour tissue after thyroidectomy in patients with thyroid cancer	High-quality images Enables dosimetry Low radiation to patient
^{124}I	4.18 days	Whole-body PET/CT: 148–185 MBq	Whole-body PET/CT, mostly in patients with thyroid cancer	High resolution PET images Enables dosimetry
^{131}I	8.1 days	Whole-body scan: 74–148 MBq	Pre- or post-therapeutic residual tumour tissue imaging after thyroidectomy in patients with thyroid cancer	Low-resolution images Higher radiation, as compared with ^{123}I
		For therapy: 185–7400 MBq	Used for treatment of hyperthyroid diseases, non-toxic goitre, and differentiated thyroid cancer	Emits primarily β-particles, which act locally in the thyroid tissue
$^{99m}TcO_4$	6 hours	37–370 MBq (i.v.)	Routine thyroid imaging	Produced locally by molybdenum generator. Only evaluates trapping and not organification
FDG	110 min	370–550 MBq (i.v.)	Whole-body PET/CT in patients with thyroid cancer	Most often used for the thyroglobulin-positive iodine-negative patient

The iodine isotopes can be administered intravenously, or more commonly orally.
FDG, 18F-deoxy-glucose.

Radioactive Tracers

The thyroid is unique in its ability to trap and organify iodine. The energy-requiring sodium-iodide symporter (NIS), located in the membrane of the thyrocyte, provides the mechanism for trapping, while the intracellular enzyme thyroid peroxidase is central in incorporating iodine into thyroglobulin (Tg). NIS does not differentiate between various isotopes of iodine, and this enzyme is crucial for thyroid scintigraphy and uptake measurements. Besides iodine, NIS can also trap technetium (^{99m}Tc), as pertechnetate ($^{99m}TcO_4$). This isotope can be produced locally using a molybdenum generator, and is therefore easier to produce and handle than iodine isotopes. However, as $^{99m}TcO_4$ is not organified it leaks out of the thyroid gland. Therefore, the increased concentration gradient of $^{99m}TcO_4$ across the cellular membrane of the thyrocyte is dependent on continuous NIS activity. Consequently, imaging and uptake measurements using $^{99m}TcO_4$ as tracer must take place 10–20 min after i.v. injection of this radiopharmaceutical, rather than after hours or days, as with radioactive iodine.

The unit of radioactivity is expressed in Becquerel (decay/s) or Curie (3.7×10^{10} decays/s). Radioactivity of the nuclide should not be confused with the dose, measured in Gray (Gy), which is the term applied to the absorbed radiation by the exposed tissue. The radioactivity administered for clinical purposes is in the order of megabecquerels (MBq) or gigabecquerels (GBq). The radioactivity used for uptake measurements is up to 100 times lower than that used for therapeutic purposes (i.e. ^{131}I therapy).

^{123}I has a half-life of 13 hours and a γ -photon energy of 159 keV, which is readily detectable by a gamma camera. Almost similar characteristics apply to ^{99m}Tc, which emits γ -photon energy of 140 keV, and with a half-life of 6 hours. However, these characteristics make ^{99m}Tc as well as ^{123}I suitable only for diagnostic and not therapeutic purposes. ^{131}I is an excellent isotope for therapy because it has a half-life of 8 days, and the emitted β -particles act locally due to a path length of only 1–2 mm. These properties are, however, disadvantageous for imaging because the radiation dose is higher while the resolution of the image is poorer. Thus, the γ -photon flux produced by 3.7 GBq ^{131}I can be obtained by merely 148 MBq of ^{123}I. Table 3.1.6.1 lists the clinically useful radioisotopes and some of their properties.

Thyroid Iodine Uptake Measurement

This test is based on oral or intravenous administration of a radioactive iodine isotope to determine the iodine uptake of the thyroid tissue. Once in the body, the tracer is gradually taken up by the thyroid similar to the endogenous iodine pool. At a given time point, the amount of the tracer within the thyroid equilibrates with the rate of iodine uptake and thyroid hormone synthesis, and the release of the isotope into the circulation, incorporated in the thyroid hormones. Importantly, the test does not measure thyroid hormone synthesis and secretion, but merely the avidity of the thyroid gland for iodine, and its clearance rate relative to the renal iodine excretion. Thus, a clear inverse relationship exists between the dietary intake of iodine and the thyroid iodine uptake.

Procedure

An uptake measurement is usually performed 24 hours after oral administration of 3.7–7.4 MBq of ^{123}I. The intrathyroidal amount of tracer can be measured with a gamma scintillation counter, placed at a fixed distance in front of the neck. The number of radioactive counts emitted from the thyroid is compared with the radioactivity administered to the patient. Mostly, the uptake value at 24 hours after radioiodine administration is used for practical reasons, and because most thyroid glands have reached the plateau of isotope accumulation at this time. A high 24 hour thyroid iodine uptake, usually in the range 40–85%, is observed typically in patients with hyperthyroidism, but may be seen after recovery from thyroid failure or severe iodine deficiency leading to elevated thyroid-stimulating hormone (TSH) serum levels and hypothyroidism. In severe hyperthyroidism, seen in some patients with Graves' disease, the thyroid uptake must be measured within 2–6 hours after the radioiodine administration due to a very high iodine turnover rate. ^{99m}Tc instead of iodine isotopes may be used in this situation since

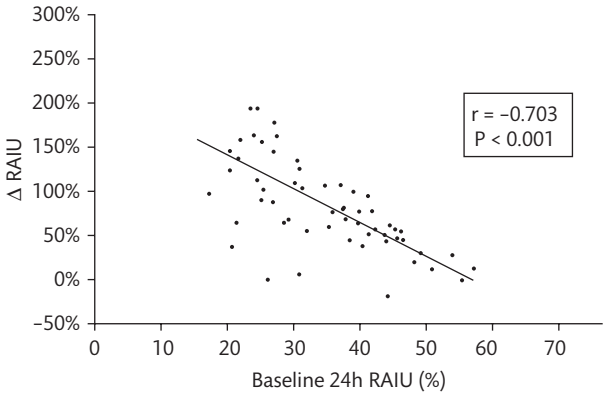


Figure 3.1.6.1 The positive effect of rhTSH stimulation on the thyroid ^{131}I uptake (RAIU) in patients with multinodular non-toxic goitre is clearly inversely related to the initial RAIU.

Reproduced with permission from Fast S, Nielsen VE, Grupe P, *et al*. Optimizing 131I Uptake After rhTSH Stimulation in Patients with Nontoxic Multinodular Goiter: Evidence from a Prospective, Randomized, Double-Blind Study. *J Nucl Med* 2009;50:732–37. Copyright © 2009 the Society of Nuclear Medicine, Inc

the thyroid iodine uptake in the very early period following the nuclide administration mainly reflects the NIS activity [1].

Indication for Thyroid Iodine Uptake Measurement

With accurate thyroid function markers, valid serological tests, and widely accessible ultrasound facilities, most thyroid patients can be diagnosed correctly. In patients with non-toxic benign goitre, an uptake measurement may indicate whether ^{131}I therapy is a therapeutic option, since the 24-hour thyroid iodine uptake should be at least 20%. In hyperthyroid patients treated with ^{131}I, the 24-hour uptake value—and sometimes the ^{131}I half-life—is mandatory for calculation of the administered ^{131}I activity at some centres.

Recombinant Human TSH Stimulation

Stimulation with 0.1 mg recombinant human TSH (rhTSH; doses of 0.005–0.9 mg have been used)—given as a single intramuscular injection 24 hours before ^{131}I administration—increases the 24-hour thyroid iodine uptake by 100% or more, even in iodine-loaded subjects [2]. This effect correlates inversely with the baseline thyroid iodine uptake [3], with the implication that patients with the lowest baseline uptake potentially benefit the most from rhTSH stimulation in relation to ^{131}I therapy (Figure 3.1.6.1). The effect on the 24-hour thyroid iodine uptake from rhTSH stimulation correlates inversely also with serum TSH [3].

Perchlorate Discharge Test

Some ions, like thiocyanate (SCN–) and perchlorate (ClO_4–), inhibit the NIS, resulting in leakage from the thyroid, if the iodine is not intracellularly organified. Thus, the iodine loss from the thyroid gland following NIS inhibition correlates inversely with the thyrocyte's ability to bind iodine.

The test is performed by administering a tracer dose of ^{123}I, and measurement of the thyroid ^{123}I uptake after 4 hours. Thereafter, 500–1000 mg of $NaClO_4$ (or $KClO_4$) is administered orally, and the uptake measurement is repeated after 1 hour. Perchlorate acutely blocks the NIS, allowing non-organified ^{123}I to leak out of the thyrocyte. A thyroid organification defect is very likely to be present if the ^{123}I uptake after perchlorate administration shows a reduction of 15% or more. An ^{123}I uptake value in the range 10–15% is considered borderline, and may be due to a partial organification defect. An abnormal perchlorate discharge test is seen in inborn defects of iodine organification, for example, human thyroid peroxidase (TPO) gene mutations or Pendred's syndrome.

Thyroid Scintigraphy

Thyroid scintigraphy is a rapid and relatively low-cost method. A normal thyroid scintigraphy shows a homogeneous isotope distribution in both lobes. Usually, the isthmus and the pyramidal lobe are not visualized well in a euthyroid individual, but these structures become more apparent in hyperthyroid conditions. A vague isotope uptake in the submandibular glands is often seen due to ability of these structures to actively concentrate iodine intracellularly. Thyroid scintigraphy gives an overall impression of the thyroid volume. However, this technique is inaccurate in comparison with ultrasound, CT, and MRI, which should be used instead [4].

Single Photon Emission Computed Tomography

In line with the principle behind a radiological CT, 3D scintigraphic images can be constructed, if the gamma camera rotates around the region of interest; this technique is termed single photon emission computed tomography (SPECT). By SPECT, the resolution can be enhanced to 6–7 mm, and allows identifying small thyroid nodules. However, SPECT is more time consuming than planar scintigraphy, and the latter method is usually sufficient in a clinical setting. For thyroid volume estimation, SPECT performs better than planar scintigraphy but is less accurate than thyroid ultrasound [4].

The clinical application of thyroid scintigraphy and other nuclide imaging techniques in various thyroid disorders is described in the remaining part of this chapter, and summarized in Table 3.1.6.2.

Thyrotoxicosis

Thyrotoxicosis is defined as elevated serum levels of thyroid hormones, while hyperthyroidism specifically is caused by an active hormone synthesis and secretion from the thyroid gland (or ectopic thyroid tissue). Thyroid scintigraphy can easily differentiate between the two conditions.

In the majority of cases (>90%), the aetiology of hyperthyroidism is either Graves' disease or toxic nodular goitre (solitary or multinodular). Both conditions are characterized by high uptake and turnover rate of iodine in the actively hormone producing thyroid tissue, and the scintigraphic appearance reflects the differences in tissue texture and functionality. There is no need to discontinue antithyroid drugs prior to ^{99m}Tc scintigraphy, as these agents do not affect the NIS function.

Graves' Disease

In Graves' disease, the entire gland, and the pyramidal lobe if present, shows a scintigraphically uniform thyroid isotope uptake. The size of the gland may range from normal to grossly enlarged. In some cases, superimposed activity in the central part of the thyroid lobes due to the greater depth of thyroid tissue may raise suspicion

Table 3.1.6.2 Application of nuclear medicine techniques in the most common thyroid diseases

Thyroid diseases	Indication for nuclear medicine techniques
Graves' disease	• Scintigraphy is not mandatory if TRAB is positive • Thyroid volume assessment for prognostication is more accurately performed by ultrasound • Prior to ^{131}I therapy, measurement of the ^{131}I kinetics for dose calculation is done at some centres
Toxic nodule(s)	• In case of negative TRAB, thyroid scintigraphy is indicated in order to confirm the presence of autonomously functioning nodule(s) suitable for ^{131}I therapy
Non-toxic goitre	• Scintigraphy detects whether a nodule is hypofunctioning (cold), in which case fine-needle aspiration biopsy may be considered • 24 h thyroid iodine uptake measurement can guide whether ^{131}I therapy is a therapeutic option
Thyroiditis (postpartum or subacute)	• Scintigraphically low isotope uptake during the thyrotoxic phase supports the presence of thyroiditis • In the hypothyroid patient, whether permanent or transient, there is normally no indication for scintigraphy
Thyroid cancer	• Whole-body ^{123}I or ^{131}I scintigraphy, performed pre- or post-^{131}I treatment, may detect residual differentiated thyroid cancer tissue • FDG-PET/CT is indicated on suspicion of residual radioiodine non-avid tumour tissue • In medullary thyroid cancer, ^{18}F-DOPA-PET seems the most sensitive method for detection of persistent tumour tissue

TRAB, TSH receptor antibody.

of thyroid nodules, but this can easily be ruled out by an ultrasound examination.

Thyroid scintigraphy can be omitted in patients with a clinical presentation typical of Graves' disease, and in the presence of TSH-receptor antibodies in serum. If a measure for the thyroid volume is needed, thyroid ultrasound is more accurate than scintigraphy.

Toxic Nodular Goitre

In patients with autonomously functioning thyroid nodules, the thyroid scintigraphy shows one or more marked areas with high isotope uptake. Uptake in the normal paranodular thyroid tissue depends on the thyroid status. In patients with moderately hyperactive 'warm' nodules, leading to subnormal serum TSH levels, some isotope uptake may be seen in the paranodular thyroid tissue (Figure 3.1.6.2a). In very hyperfunctioning 'hot' nodules, completely suppressing serum TSH, the remaining part of the thyroid may not be scintigraphically visualized. Thus, a toxic multinodular goitre (Figure 3.1.6.2b), sometimes being disconfigured due to nodular degeneration, may harbour coexisting hyperactive (warm/hot) and hypoactive (cold) nodules, as well as hypofunctioning normal thyroid tissue. If the patient has been treated with antithyroid drugs for some time, achieving a normal serum TSH level, the isotope may be more uniformly distributed within the gland.

Since ^{131}I therapy usually is a first line treatment in patients with one or more thyrotoxic nodules, scintigraphy is always indicated in such cases to confirm the presence of nodular functional autonomy. Importantly, Graves' disease may develop *de novo* in a patient with nodular goitre. In fact, such a coexistence occurs in as many as 10–15% of unselected hyperthyroid patients [5]. Based on thyroid ultrasound only, the clinician may get the impression that the hyperthyroidism is due to a multinodular goitre rather than Graves' disease (Figure 3.1.6.3). However, the presence of TSH-receptor antibodies, and a thyroid scintigraphy showing a more diffuse distribution of the isotope throughout the thyroid gland, will clarify the aetiology of the hyperthyroidism.

Other Thyrotoxic Conditions

Thyrotoxicosis characterized by a low thyroid isotope uptake is most often caused by subacute, silent, or postpartum thyroiditis. These conditions are discussed in section 3.3.

Exposure to excess iodine (e.g. X-ray contrast media) or amiodarone (an antiarrhythmic drug with high iodine content) sometimes leads to thyrotoxicosis. The very high whole-body iodine pool and the iodine load of the gland block for further thyroid iodine uptake. Therefore, the scintigraphy shows low or no thyroid isotope uptake. In fact, knowing that the patient has been exposed to iodine-containing agents, there is rarely indication for thyroid scintigraphy. The clearance of excess iodine, and especially of amiodarone, may take several months, with slow recovery of a normal thyroid function. If the thyrotoxicosis persists, thyroid scintigraphy may be indicated, provided the iodine exposure has ceased.

In patients who factitiously ingest thyroid hormones in excessive amounts, the thyrotoxic state and a faint thyroid scintigram will

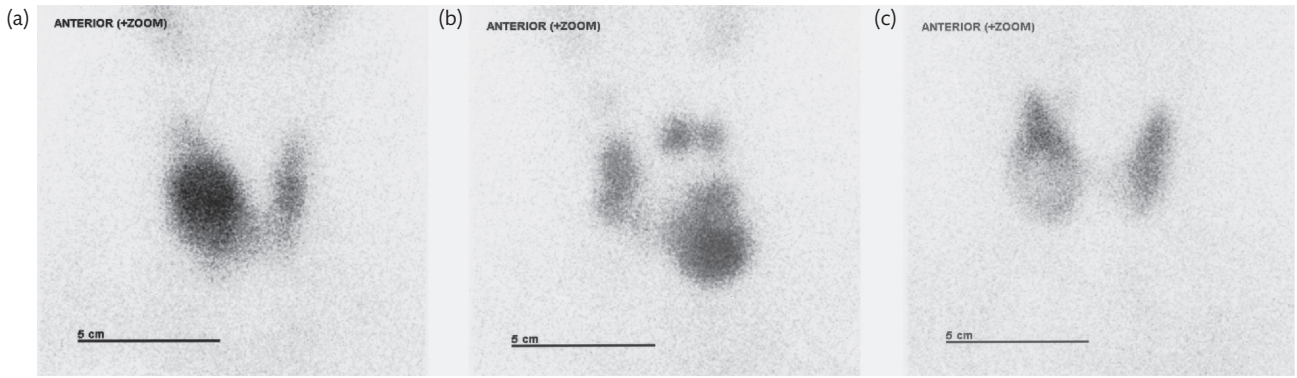


Figure 3.1.6.2 Examples of thyroid scintigrams in various disorders. (a) Solitary warm nodule in the right thyroid lobe. (b) Toxic multinodular goitre with several hot nodules and suppression of the paranodular thyroid tissue. (c) Solitary cold nodule in the right thyroid lobe.

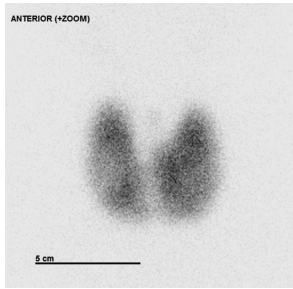


Figure 3.1.6.3 Thyroid scintigraphy in a 57-year-old woman with Graves' disease and coexisting multinodular goitre. From a euthyroid state 6 months earlier the woman became severely hyperthyroid due to Graves' disease, which was confirmed by high serum levels of TSH-receptor autoantibodies. Thyroid ultrasound showed a classical bilateral multinodular configuration. The isotope is irregularly distributed as the immune stimulation affects nodular and paranodular tissue to a different extent.

raise suspicion of thyroiditis. However, excess exogenous thyroid hormone leads to suppression and inactivation of the thyroid gland, reflected by low levels of serum Tg, while this biomarker is usually high in conditions with thyroiditis.

Hypothyroidism

Thyroid imaging is not mandatory and often superfluous in the diagnostic work-up and during follow-up of hypothyroid patients, independent of the aetiology. When goitre is suspected in Hashimoto's thyroiditis, ultrasound examination should be preferred over scintigraphy. In early stages of autoimmune thyroiditis, even with severe thyroid failure, thyroid scintigraphy may show some thyroid uptake in a diffuse or patchy pattern. This is due to preserved function of the NIS, further stimulated by a high serum TSH, whereas the thyroid failure primarily results from defects in iodine organification and thyroid hormone synthesis.

Thyroiditis

The term thyroiditis covers several conditions, the most common being autoimmune (Hashimoto's) thyroiditis. Thyroid scintigraphy is often superfluous in thyroiditis, with a few exceptions as discussed next.

Autoimmune Thyroiditis

Autoimmune thyroiditis may give rise to thyrotoxicosis due to exacerbation of the autoimmune inflammation, which may be clinically apparent as silent thyroiditis and postpartum thyroiditis. In particular, it is important to differentiate between postpartum thyroiditis and Graves' disease. In postpartum thyroiditis, an inflammatory destructive condition with no active thyroid hormone synthesis, thyroid scintigraphy shows universally very low or no isotope uptake. However, the distinction between the two conditions can often be made without thyroid imaging. Nearly all patients with Graves' disease harbour TSH-receptor antibodies, which are absent in typical cases of postpartum thyroiditis. In addition, the dynamics of the two diseases are very different, as the thyrotoxic phase in postpartum thyroiditis is followed by hypothyroidism, before recovery of the thyroid function takes place within some months. With such a course, the diagnosis is usually settled, and thyroid scintigraphy is of no value at this late phase of the disease. Importantly, radioactive isotopes should not be administered to a breastfeeding woman. If necessary, breast milking in advance (with storage in the refrigerator), and no breastfeeding during 24–48 hours after the scintigraphy complies with the safety regulations.

Subacute Thyroiditis (de Quervain's Thyroiditis)

The course of the thyroid hormone serum levels in subacute thyroiditis is similar to that seen in silent or postpartum thyroiditis. Accordingly, thyroid scintigraphy initially shows a very low or no thyroid uptake. Subacute thyroiditis may affect the gland unilaterally, with initially no isotope uptake in a single thyroid lobe, and with subsequent progression to the entire gland within weeks.

Occasionally, typical symptoms like neck pain may be absent in subacute thyroiditis. In the undiagnosed patient, this may prompt an ^{18}F-deoxy-glucose positron emission tomography (FDG-PET). If performed in the thyrotoxic phase, an increased FDG uptake in the thyroid and skeletal muscle is seen, while the uptake in the liver is decreased [6].

Goitre and Thyroid Nodules

Scintigraphy provides a rough estimate of the size and configuration of the gland, and its possible extension into the mediastinum. However, its primary role in patients with thyroid nodules is to discriminate hyperactive from hypoactive nodules (Figure 3.1.6.2c). Fine-needle aspiration biopsy (FNA) should be considered only in cold nodules, since hyperactive nodules with rare exceptions never represent a carcinoma.

The fraction of cold nodules constitutes more than 75% of nodules in consecutive series of thyroid scintigrams, with an á priori risk of 8–25% for these lesions to harbour malignancy [7]. Such risk estimates are most probably influenced by selection bias, since cold nodules ≥10 mm are found in 2–3% of unselected individuals residing in regions with borderline iodine deficiency [8]. The risk of malignancy in a cold nodule is similar, regardless of whether it is solitary or part of a multinodular goitre [7]. Importantly, a toxic multinodular goitre does not exclude the possibility that cold nodules, located in other parts of the goitre, may be malignant. Scintigraphy is of great help in identifying such coexisting lesions.

Choice of Tracer for Scintigraphy of Nodular Goitre

^{99m}Tc used as tracer may result in false-positive uptake in 5% of thyroid nodules ('trapping only nodules'), as compared with ^{123}I scintigraphy [9]. However, comparative studies have been unable to demonstrate any clinically significant differences between the two tracers [7]. In case of substernal goitre, ^{123}I rather than ^{99m}Tc scintigraphy is recommended, since the background activity in the mediastinum attenuates the emission from the isotope (Figure 3.1.6.4).

^{99m}Tc-MIBI is a lipophilic radiopharmaceutical that accumulates in the thyroid, independent of iodine trapping and organification. ^{99m}Tc-MIBI uptake reflects both the number and

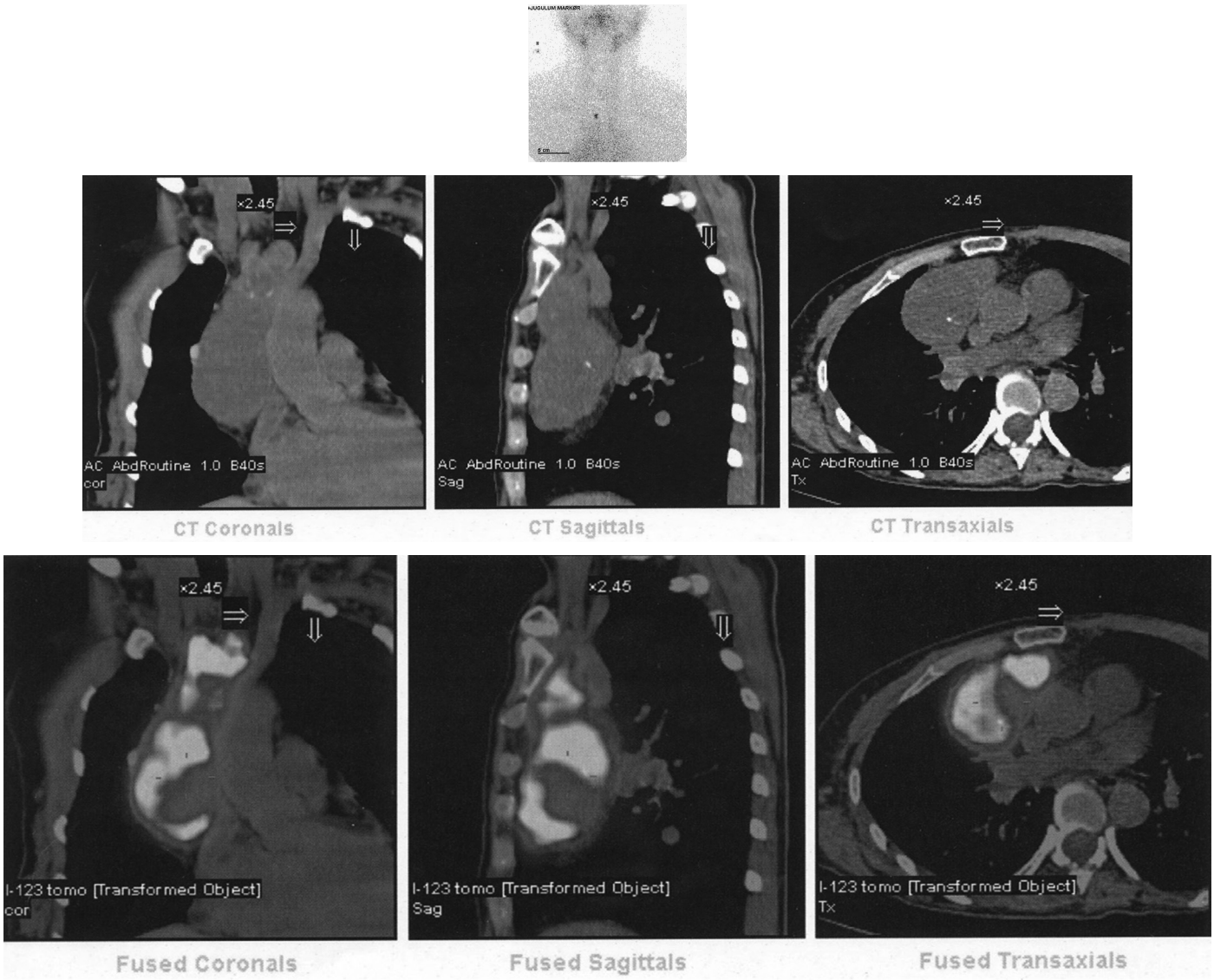


Figure 3.1.6.4 A 51-year-old woman with swallowing difficulties due to a large intrathoracic goitre. A large tumour mass in the mediastinum was detected by CT scan (middle), but no thyroid tissue could be visualized by conventional ^{99m}Tc scintigraphy (top). A ^{123}I SPECT/CT was performed 24 hours after i.m. injection of 0.1 mg of recombinant human TSH. A convincing ^{123}I uptake is seen in the right-anterior part of the mediastinum (bottom), confirming that the mass represents an intrathoracic goitre. The volume was measured to 193 ml, and the 24-hour ^{123}I uptake was 57%.

function of mitochondria, and thus the oxidative burden of the thyroid nodule [10]. Nodules with increased uptake and late retention of ^{99m}Tc-MIBI are suspicious for malignancy, but based on a meta-analysis [11]—demonstrating 82% sensitivity and 63% specificity for detection of malignancy in hypofunctioning thyroid nodules—^{99m}Tc-MIBI scintigraphy as an adjunctive test seems unjustified.

Indication for Thyroid Scintigraphy in Patients with Goitre

Despite the aforementioned potential advantages, thyroid scintigraphy is generally not included in the initial evaluation of goitre patients. Hyperthyroidism is a definite indication for scintigraphy in the setting of nodular goitre, in order to visualize warm or hot nodules if ^{131}I therapy is considered. Even with a normal serum TSH, scintigraphy may detect autonomously functioning nodules in 50% of patients [12]. Thyroid scintigraphy is recommended in cases where FNA has shown follicular neoplasia, particular if serum TSH is below normal. Such a lesion is with great certainty benign, despite the cytological diagnosis, if scintigraphy confirms that the nodule is autonomously functioning.

Positron Emission Tomography

PET scan, combined with CT, has proven very helpful for managing a wide variety of cancers. Different kinds of tracers are used for PET scans, with FDG as the most widely employed. This agent is taken up by the cells, proportional to the rate of glucose metabolism. Because malignant tissue generally is metabolically more active, FDG uptake is a potential indicator of even small areas with malignant tissue. A hindrance for a widespread use of PET is the need of a cyclotron for ^{18}F production, and limitation of transport distances due to a half-life of 110 minutes.

PET in Patients with Thyroid Nodules

^{124}I is transported intracellularly by NIS, like other iodine isotopes. Its half-life is 4.2 days, and 22% of its emission consists of positrons. ^{124}I-PET/CT specifically visualizes the thyroid gland, resulting in better thyroid images than provided by conventional nuclear medicine techniques. In patients with benign thyroid disease, ^{124}I-PET/CT detects significantly more nodular lesions, functional as well as non-functional, than does ^{99m}Tc scintigraphy or ^{99m}Tc-SPECT [13]. The main indication for PET/CT, however, is in situations where thyroid cancer is suspected, or already diagnosed.

FDG-PET may be useful in the evaluation of scintigraphically cold nodules with indeterminate FNA findings. The sensitivity and specificity of FDG-PET/CT for detection of thyroid carcinoma are in the ranges 77–100% and 33–64%, respectively [10]. A meta-analysis found a negative predictive value of 96% in indeterminate thyroid nodules larger than 15 mm [14]. Thus, FDG-PET/CT seems to rule-out malignancy in indeterminate nodules, except in those smaller than 10–15 mm, probably due to limited spatial resolution. The method also has its limitations in Hürthle cell lesions, in which benign as well as malignant lesions can be PET positive [15]. With these exceptions, application of FDG-PET can potentially decrease the number of unnecessary diagnostic hemithyroidectomies, in some studies by 35% [10], and may in fact be cost-effective [16].

Contrasting the generally high sensitivity of FDG-PET, the specificity is consistently poor due to mechanisms not fully understood. The maximum standardized uptake value (SUV_{max}) is significantly higher in malignant than in benign thyroid nodules but with huge overlap, and therefore not useful in a clinical setting [17]. Moreover, SUV_{max} calculations strongly depend on the type of PET-scanner and other external factors.

FDG-PET is currently not routinely recommended in patients with FNA indeterminate thyroid nodules.

Thyroid Incidentalomas Detected by PET

Focal thyroid lesions, detected incidentally, are seen in 1–2% of patients undergoing FDG-PET for non-thyroid reasons. Approximately 35% of these incidentalomas harbour thyroid cancer [17], depending on the selection of patients. The majority of these FDG-PET positive lesions are small papillary carcinomas, often resulting in thyroid cancer surgery and year-long monitoring. It remains to be established whether such an active approach is justified in terms of increased survival, quality of life, and cost. At present, it is recommended that FDG-avid thyroid incidentalomas larger than 10 mm are evaluated by FNA [18].

If FDG-PET shows diffuse uptake of the thyroid, this typically represents an inflammatory condition, like autoimmune thyroiditis, and the patient can be reassured about the benign nature.

Thyroid Cancer

At diagnosis, and before neck surgery, nuclear medicine techniques are not routinely employed. After total thyroidectomy, whole-body scan (WBS) after administration of either ^{123}I or ^{131}I might detect iodine-avid tissue. NIS expression is decreased in more than 90% of thyroid carcinomas due to tumorigenic degeneration, and by as much as 1200-fold [19]. However, NIS remains expressed at a level that allows imaging as well as therapy by radioactive iodine isotopes. Although serum Tg is the primary cancer marker used in the follow-up [18], WBS helps in the stratification of the patient [19]. WBS can be performed either with (post-treatment WBS) or without (diagnostic WBS) prior ^{131}I therapy given for ablative or adjunctive purpose. For a diagnostic WBS, a low dose of ^{123}I (or ^{131}I) is used (Figure 3.1.6.5). A post-treatment WBS is obtained 5–10 days after the therapeutic ^{131}I is administered.

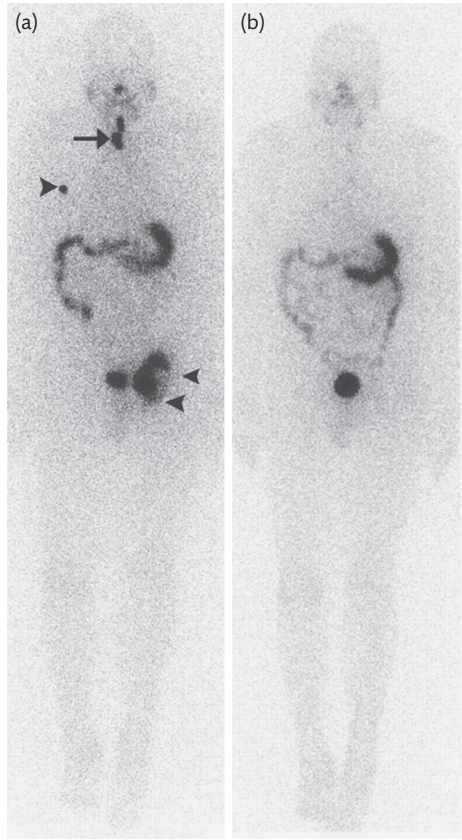


Figure 3.1.6.5 A 67-year-old man with papillary thyroid cancer. (a) Pretreatment whole-body ^{123}I scan shows residual thyroid tissue in the neck (arrow), as well as right rib and left hip metastases (arrowheads). (b) No abnormal uptake is seen on the ^{123}I scan taken at the 1-year follow-up.

A diagnostic WBS is used to detect local and distant metastases, and to tailor the therapeutic ^{131}I dose as well as excluding patients for ^{131}I therapy in case of no uptake suggestive of non-radioiodine avid disease. In fact, diagnostic WBS may provide important information that leads to change in management strategy in up to 50% of cases [20]. Diagnostic WBS is less used at many centres given the fact that ^{131}I therapy is administered according to a risk stratification strategy, or empirically in case of abnormal serum Tg levels during follow-up. According to the American Thyroid Association guidelines, postsurgical diagnostic WBS may be beneficial if the extent of residual tumour tissue cannot be determined by conventional imaging, or the management of the patient may be altered by the additional information [18].

Details of thyroid cancer surveillance and treatment algorithms are given in Chapter 3.5.6.

Optimizing Radioiodine Uptake

Iodine restriction for 1–2 weeks increases the radioiodine uptake during WBS as well as the efficacy of the ^{131}I therapy, but it is unknown whether such iodine restriction has impact on the long-term recurrence rate [21].

A high serum TSH stimulates the radioiodine uptake of residual thyroid cancer tissue. This can be achieved either by thyroid hormone withdrawal—4 weeks for levothyroxine, and two weeks for triiodothyronine—or by the administration of rhTSH. The main advantage of rhTSH stimulation is avoidance of a prolonged hypothyroid state and an improvement of the patient's well-being [22]. The two regimens have proved to be equally effective in terms of the cancer ablation rate. Importantly, rhTSH stimulation is more cost-effective and leads to a lower whole-body irradiation, as compared with the withdrawal regimen [2].

Thyroid Stunning

Thyroid stunning is the phenomenon where the radioiodine uptake is attenuated due to prior irradiation. Stunning, dose-related and detectable even at a very low level of radiation, is associated with decreasing levels of NIS-mRNA and signs of cell cycle arrest but not cell death [2]. It is unclear whether stunning constitutes a significant clinical problem. The discrepancy between a positive diagnostic ^{131}I-WBS and a subsequent negative post-treatment ^{131}I-WBS has been attributed to the impact of stunning. However, what seems to be stunning may in fact be a cell killing effect by the WBS or the ^{131}I therapy [23]. To minimize stunning, ^{123}I should be used for WBS, since the required activity is lower than that needed with ^{131}I. However, the clinical relevance of stunning can be questioned since patients show similar ablation rates, independent of the iodine isotope used prior to the ^{131}I therapy [2]. The potential negative effect of stunning by a diagnostic WBS adds further support to a strategy where ^{131}I ablative therapy is given on an empiric basis.

Interpretation of the Whole-Body Scan

In patients thyroidectomized and devoid of persistent cancer tissue, the 24–72 hours radioiodine uptake in the thyroid region should be no more than 0.2–0.5% of the administered radioactivity. Most iodine is excreted in the urine, and the bladder is easily identified on a WBS. There is almost no radioiodine uptake in muscles, skeleton, brain, heart, and lungs, unless there are iodine-avid metastases. ^{123}I diagnostic and ^{131}I post-treatment WBS concordance rates are approximately 87% and 72% for a positive and negative WBS, respectively [24]. Post-treatment WBS may visualize additional lesions, compared to the diagnostic WBS, in as many as 40% of cases [19]. Before the advent of PET techniques, post-treatment ^{131}I-WBS was thus considered the definite test to visualize iodine-avid thyroid cancer tissue.

A false-positive WBS can result from radioiodine uptake in a non-thyroidal organ expressing NIS (i.e. the stomach, the salivary, and the mammary glands). Conversely, dedifferentiated cancer cells might have lost their ability to trap iodine, resulting in false-negative WBS. Lack of radioiodine uptake in residual thyroid cancer cells may also be due to an erroneous intake of thyroid hormone or high dietary iodine content.

PET in Differentiated Thyroid Cancer

FDG-PET

In patients with abnormally elevated serum Tg, FDG-PET is superior to both diagnostic and post-treatment ^{131}I-WBS for the localization of remnant cancer tissue (Figure 3.1.6.6). Well-differentiated thyroid cancers—with a metabolic rate not significantly different from that in normal thyroid tissue—might fail detection by FDG-PET/CT. In contrast, poorly differentiated and more aggressive cancers demonstrate a high uptake of FDG. Indeed, NIS expression is decreased while the glucose transporter Glut1 is upregulated in poorly differentiated thyroid tumour cells [25]. On functional imaging this is called a "Flip-flop" phenomenon, which can be reversed by some tyrosine kinase inhibitors, inducing re-expression of NIS in the tumour cells [25].

The serum Tg cut-off level for the maximum sensitivity and specificity for FDG-PET/CT is in the range 12–32 ng/ml, but positive lesions can be detected even with Tg levels below 10 ng/ml [19]. RhTSH stimulation prior to FDG-PET/CT might improve the diagnostic sensitivity, but this rarely leads to change in disease management [26].

^{124}I-PET

Results have been variable when ^{124}I-PET is compared to post-treatment ^{131}I-WBS. Differences in the radioactivity and the regimen used for enhancing ^{124}I uptake may explain the lack of agreement across studies. Nevertheless, ^{124}I-PET/CT visualizes residual disease

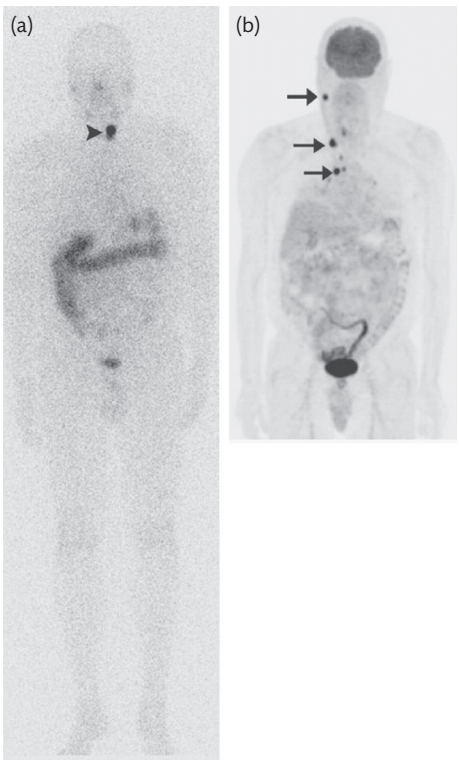


Figure 3.1.6.6 A 74-year-old man with papillary thyroid cancer. (a) Whole-body ^{123}I scan shows thyroid tissue in the neck (arrowhead). (b) FDG-PET/CT (carried out on the same day) shows extensive metastatic disease in cervical and mediastinal lymph nodes (arrows).

with a very high sensitivity, as well as lesions not detected by post-treatment ^{131}I-WBS [27]. The precise role of ^{124}I-PET/CT in the management of patients with differentiated thyroid cancer remains to be established. For dosimetry, which relies on quantification and estimation of the tracer residence time, ^{124}I-PET/CT is superior to SPECT/CT [25].

Imaging in Medullary Thyroid Cancer

As medullary thyroid cancer (MTC) originates from the iodine-independent para-follicular C-cells nuclear imaging techniques are based on non-iodine isotopes. PET/CT is probably the technique mostly used. ^{18}F-DOPA-PET appears to be a more sensitive method than FDG-PET for localizing persistent MTC, but its sensitivity varies from 47% to 83%, depending on the calcitonin level [25]. In particular, ^{18}F-DOPA-PET/CT has high sensitivity for detection of primary MTC and lymph node metastases, as compared to ultrasonography, and lymph node resection during neck surgery may be more successful if guided by ^{18}F-DOPA-PET/CT [28].

MTC expresses one or more types of somatostatin receptors. While ^{68}Ga-labelled somatostatin receptor PET/CT is helpful in the management of patients with neuroendocrine tumours, the diagnostic performance of this method is lower in MTC [25]. Although this method is particularly efficient for detection of bone metastases [29], functional imaging seems to be of limited value in the management of MTC.

New Tracers for Cancer Detection

Prostate-specific membrane antigen (PSMA) is expressed on the endothelial cells of tumour neo-vasculature, not only in prostate tissue but also in many non-prostate cancers. Thus, radioiodine-refractory thyroid cancer tissue expresses PSMA, which might be visualized with ^{68}Ga-PSMA-PET/CT [19]. Another new agent is ^{18}F-tetrafluoroborate that has a bio-distribution characteristic of NIS expression, and may hold promise in the treatment guidance of differentiated thyroid cancer [27].

Despite the potential benefit from new and often expensive nuclear medicine technologies, it remains to be shown that their use improves the prognosis in thyroid cancer patients. Most thyroid malignancies have a good prognosis, and the increased surveillance and application of imaging during the last decades do not seem to have influenced the overall survival [30].

REFERENCES

1. Macauley M, Shawgi M, Ali T, *et al.* Assessment of normal reference values for thyroid uptake of technetium-99m pertechnetate in a single centre UK population. *Nucl Med Commun*, 2018; **39**: 834–8.
2. Bonnema SJ, Hegedüs L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*, 2012; **33**: 920–80.
3. Fast S, Nielsen VE, Grupe P, *et al.* Optimizing 131I uptake after rhTSH stimulation in patients with nontoxic multinodular goiter: evidence from a prospective, randomized, double-blind study. *J Nucl Med*, 2009; **50**: 732–7.
4. Van Isselt JW, De Klerk JM, Van Rijk PP, *et al.* Comparison of methods for thyroid volume estimation in patients with Graves' disease. *Eur J Nucl Med Mol Imaging*, 2003; **30**: 525–31.
5. Pedersen IB, Knudsen N, Perrild H, *et al.* TSH-receptor antibody measurement for differentiation of hyperthyroidism into Graves' disease and multinodular toxic goitre: a comparison of two competitive binding assays. *Clin Endocrinol (Oxf)*, 2001; **55**: 381–90.
6. Kim MH, Kim DW, Park SA, *et al.* Transiently altered distribution of F-18 FDG in a patient with subacute thyroiditis. *Nucl Med Mol Imaging*, 2018; **52**: 82–4.
7. Hegedüs L, Bonnema SJ, Bennedbæk FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev*, 2003; **24**: 102–32.
8. Knudsen N, Perrild H, Christiansen E, *et al.* Thyroid structure and size and two-year follow-up of solitary cold thyroid nodules in an unselected population with borderline iodine deficiency. *Eur J Endocrinol*, 2000; **142**: 224–30.
9. Reschini E, Ferrari C, Castellani M, *et al.* The trapping-only nodules of the thyroid gland: prevalence study. *Thyroid*, 2006; **16**: 757–62.
10. de Koster EJ, de Geus-Oei LF, Dekkers OM, *et al.* Diagnostic utility of molecular and imaging biomarkers in cytological indeterminate thyroid nodules. *Endocr Rev*, 2018; **39**: 154–91.
11. Treglia G, Caldarella C, Saggiorato E, *et al.* Diagnostic performance of (99m)Tc-MIBI scan in predicting the malignancy of thyroid nodules: a meta-analysis. *Endocrine*, 2013; **44**: 70–8.
12. Treglia G, Trimboli P, Verburg FA, *et al.* Prevalence of normal TSH value among patients with autonomously functioning thyroid nodule. *Eur J Clin Invest*, 2015; **45**: 739–44.
13. Darr AM, Opfermann T, Niksch T, *et al.* Low-activity 124I-PET/low-dose CT versus 99mTc-pertechnetate planar scintigraphy or 99mTc-pertechnetate single-photon emission computed tomography of the thyroid: a pilot comparison. *Clin Nucl Med*, 2013; **38**: 770–7.
14. Vriens D, de Wilt JH, van der Wilt GJ, *et al.* The role of [(18)F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: systematic review and meta-analysis of the literature. *Cancer*, 2011; **117**: 4582–94.
15. Pathak KA, Klonisch T, Nason RW, *et al.* FDG-PET characteristics of Hürthle cell and follicular adenomas. *Ann Nucl Med*, 2016; **30**: 506–9.
16. Vriens D, Adang EM, Netea-Maier RT, *et al.* Cost-effectiveness of FDG-PET/CT for cytologically indeterminate thyroid nodules: a decision analytic approach. *J Clin Endocrinol Metab*, 2014; **99**: 3263–74.
17. Soelberg KK, Bonnema SJ, Brix TH, *et al.* Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review. *Thyroid*, 2012; **22**: 918–25.
18. Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2016; **26**: 1–133.
19. Santhanam P, Solnes LB, Rowe SP. Molecular imaging of advanced thyroid cancer: iodinated radiotracers and beyond. *Med Oncol*, 2017; **34**: 189.

20. Chen MK, Yasrebi M, Samii J, *et al*. The utility of I-123 pretherapy scan in I-131 radioiodine therapy for thyroid cancer. *Thyroid*, 2012; **22**: 304–9.
21. Sawka AM, Ibrahim-Zada I, Galacgac P, *et al*. Dietary iodine restriction in preparation for radioactive iodine treatment or scanning in well-differentiated thyroid cancer: a systematic review. *Thyroid*, 2010; **20**: 1129–38.
22. Lee J, Yun MJ, Nam KH, *et al*. Quality of life and effectiveness comparisons of thyroxine withdrawal, triiodothyronine withdrawal, and recombinant thyroid-stimulating hormone administration for low-dose radioiodine remnant ablation of differentiated thyroid carcinoma. *Thyroid*, 2010; **20**: 173–79.
23. Walrand S, Hesse M, Jamar F. Statistical and radiobiological analysis of the so-called thyroid stunning. *EJNMMI Res*, 2015; **5**: 67.
24. Urhan M, Dadparvar S, Mavi A, *et al*. Iodine-123 as a diagnostic imaging agent in differentiated thyroid carcinoma: a comparison with iodine-131 post-treatment scanning and serum thyroglobulin measurement. *Eur J Nucl Med Mol Imaging*, 2007; **34**: 1012–17.
25. Giraudet AL, Taieb D. PET imaging for thyroid cancers: current status and future directions. *Ann Endocrinol (Paris)*, 2017; **78**: 38–42.
26. Leboulleux S, Schroeder PR, Busaidy NL, *et al*. Assessment of the incremental value of recombinant thyrotropin stimulation before 2-[18F]-Fluoro-2-Deoxy-D-glucose positron emission tomography/computed tomography imaging to localize residual differentiated thyroid cancer. *J Clin Endocrinol Metab*, 2009; **94**: 1310–16.
27. Santhanam P, Taieb D, Solnes L, *et al*. Utility of I-124 PET/CT in identifying radioiodine avid lesions in differentiated thyroid cancer: a systematic review and meta-analysis. *Clin Endocrinol (Oxf)*, 2017; **86**: 645–51.
28. Rasul S, Hartenbach S, Rebhan K, *et al*. [(18)F]DOPA PET/ceCT in diagnosis and staging of primary medullary thyroid carcinoma prior to surgery. *Eur J Nucl Med Mol Imaging*, 2018; **45**: 2159–69.
29. Castroneves LA, Coura FG, de Freitas RMC, *et al*. Comparison of 68Ga PET/CT to other imaging studies in medullary thyroid cancer: superiority in detecting bone metastases. *J Clin Endocrinol Metab*, 2018; **103**: 3250–9.
30. Banerjee M, Wiebel JL, Guo C, *et al*. Use of imaging tests after primary treatment of thyroid cancer in the United States: population based retrospective cohort study evaluating death and recurrence. *BMJ*, 2016; **354**: i3839.

3.1.7 Thyroid Imaging: Non-Isotopic Techniques

Laszlo Hegedüs and Finn N. Bennedbæk

Box 3.1.7.1 Clinical settings for the use of thyroid imaging

- Patients with thyroid enlargement or tenderness or with the finding of a cervical lump at physical examination
- Patients with an incidental finding of thyroid nodule(s), e.g. on PET-CT, CT, or MRI not performed for thyroid purposes
- Patients with laboratory abnormalities indicating a thyroid disease
- Patients at high risk for thyroid malignancy (history of familial thyroid carcinoma, previous neck irradiation)
- Stratification of the risk of malignancy of thyroid nodules
- Guidance for thyroid nodule fine-needle aspiration (FNA) or for minimally invasive procedures
- Presurgical staging of patients who undergo surgery for thyroid malignancy or symptomatic goitre
- Postsurgical surveillance of neck recurrences after resection of thyroid carcinoma
- Detection of developmental abnormalities and ectopia varieties of the thyroid gland

Introduction

Clinical examination and evaluation of thyroid function remain fundamental in the evaluation of thyroid disorders, but observer variation leads to a considerable heterogeneity in the evaluation of patients with suspected thyroid disease [1]. Thus, it is not surprising that imaging of the thyroid is often performed. Although it most often cannot distinguish between benign and malignant lesions, and its clinical value is thought to be limited [2], a recent survey confirms the value of ultrasonography (US) in euthyroid patients with a solitary thyroid nodule [3].

The thyroid gland can be evaluated by several non-isotopic imaging techniques. The most commonly used are US, CT, and MRI. Each method has advantages and limitations, and there is no absolute clinical indication for performing any of these imaging procedures in most patients (see Box 3.1.7.1) [4].

Thyroid US examination with high-frequency transducers provides an unmatched spatial and time resolution for the detection of focal lesions, the stratification of the risk of malignancy, and the assessment of diffuse structural abnormalities [5]. Thyroid radioisotope scan is of use for the characterization of the gland's functional status in clinically and subclinically hyperthyroid patients, especially if candidates for radioiodine treatment, and for the detection of hyperfunctioning nodules that do not need fine-needle aspiration biopsy (FNA), in toxic nodular goitre [6]. CT and MRI offer valuable information in patients with a large goitre or an invasive tumour in the presurgical evaluation of whether eligible or not for surgery. These cross-sectional imaging techniques provide a detailed definition of the relationship of the thyroid with the anatomical structures located in the neck or the mediastinum.

This chapter will focus on the clinical use of these methods and, as far as this is possible, compare their advantages and disadvantages.

Techniques

Ultrasonography

Examination of the neck is performed with high-frequency transducers (7–15 MHz), and the patient in the supine position with the

Box 3.1.7.2 Possible applications of US in thyroid disorders

- Size determination
- Morphology (diffuse, uni-, or multinodular, cyst)
- Echogenicity (hypo-, normo-, or hyperechogenic)
- Flow determination
- Determination of tissue elasticity
- Aid in diagnostic biopsies
- Aid in minimally invasive procedures (cyst aspiration, ethanol, and thermal ablation procedures*) [7]
- Evaluation of regional lymph nodes

* Thermal ablation procedures include laser, radiofrequency, high-frequency US, and microwave therapy.

neck hyperextended. The transducer is coupled to the skin with gel since ultrasound does not pass through air. The technique can detect thyroid lesions as small as 2 mm and distinguish solid nodules from simple and complex cysts. It enables the accurate determination of thyroid size, gives a rough estimate of echogenicity, visualizes vascular flow and velocity (colour flow Doppler), provides information on tissue elasticity of lesions, and aids in the accurate placing of needles, be it for diagnostic or therapeutic purposes (see Box 3.1.7.2)

The main drawbacks are the high degree of observer dependency and the inability to visualize retroclavicular or intrathoracic extension of the thyroid [2, 5]. The average investigation rarely takes more than 5–10 min.

The normal thyroid is a bilobed gland, located in the lower anterior neck, with two pear-shaped lobes connected anteriorly by a central rim of tissue (the isthmus). The left and right lobes are located immediately to the left and right of the trachea, respectively. The internal carotid arteries and internal jugular veins are located postero-laterally to the thyroid lobes, whereas the strap muscles of the neck are located anteriorly. Usually, the thyroid gland size is 20 mm or less in both the transverse and the anteroposterior diameter with a length of 40–60 mm.

The normal thyroid parenchyma has a characteristic homogeneous medium-level echogenicity (see Figure 3.1.7.1). The surrounding muscles have a lower echogenicity.

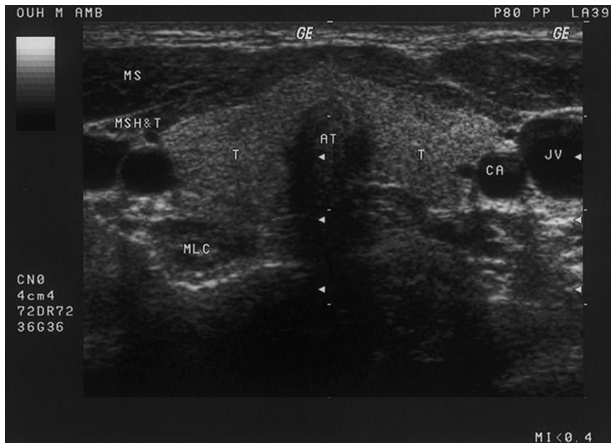


Figure 3.1.7.1 Transverse sonogram of the normal thyroid gland. AT, trachea; CA, common carotid artery; JV, jugular vein; MLC, longus colli muscle, MS, sternocleidomastoid muscle; MSH&T, sternohyoid and thyrohyoid muscle; T, thyroid.

A high proportion of people with a normal thyroid gland have small (1–3 mm) cystic or solid lesions, the frequency being higher in women, increasing with age, and varying between countries [5, 8]. The importance of these abnormalities is unclear, but since incidental sonographic nodules ('incidentalomas') are very common, whereas thyroid cancer is not, a conservative/expectant approach is generally recommended.

Goitre (i.e. an enlarged thyroid gland), remains a clinical diagnosis. But this evaluation carries an inaccuracy of approximately 40% and cannot reliably be used for size determination [1]. Thyroid volume may be estimated with the ellipsoid formula (length × width × thickness × π/6 for each lobe) or, more precisely, with a dedicated tridimensional software. The normal thyroid size (5–20 ml in adults) is positively related to weight and age, increases with decreasing iodine intake, and is influenced by a number of physiological as well as environmental factors [2, 5]. US is the most sensitive technique for screening populations for goitre and is widely used for field studies [2, 5, 9].

Computed Tomography and Magnetic Resonance

US examination is the primary imaging modality for thyroid disease, but other cross-sectional techniques used for the evaluation of neck or chest disease may commonly reveal thyroid abnormalities. The diagnostic accuracy of CT and MRI for thyroid malignancy is low, and these imaging methods should not be used in the routine diagnostic work-up of thyroid nodules, but be reserved for the assessment of intrathoracic goitres as well as for evaluation of the extension of the disease in patients with aggressive thyroid malignancy [2]. Today, multidetector CT allows multiplanar reconstructions (axial, sagittal, coronal planes) with slice thickness of 2 mm or less. Acquisition should include a volume from the skull base to the tracheal bifurcation [2]. At the CT evaluation, when without contrast medium, the thyroid appears as a homogeneous and mildly hyperattenuating gland when compared to the surrounding neck muscles and has an average attenuation of 80–100 Hounsfield Units. After injection of contrast material, the gland shows an intense and homogeneous enhancement. For large goitres, a study without contrast injection may disclose a relevant intrathoracic growth. For invasive thyroid carcinomas, the use of CT contrast media is usually needed to reliably rule out an infiltrative growth or the presence of nodal metastases in cervical regions (e.g. the retrotracheal area) that cannot be reliably visualized by US [10]. In such patients with aggressive disease who are candidates for rapid radioiodine ablation and post-dose whole-body scintigraphy, MRI with gadolinium should be used as the first-choice cross-sectional imaging method because the iodide load of CT contrast media may interfere with thyroid iodide uptake for a few months. MRI examination is commonly performed with axial and coronal T_1-weighted and with fat-saturated T_2-weighted images, followed by postcontrast axial and coronal T_1-weighted images. When compared with the cervical muscles, the thyroid gland is faintly hyperintense on T_1-weighted MR images and iso- to slightly hyperintense on T_2-weighted images. Analogously to contrast-enhanced CT imaging, gadolinium-enhanced MRI shows the thyroid gland characterized by an intense and homogeneous enhancement [10]. Both MRI and CT imaging provide precise anatomic information regarding the position of the thyroid relative to adjacent vascular, respiratory, and muscular structures. Still, despite the better contrast resolution of MR imaging and the elevated spatial resolution of CT, B-mode US examination

remains the best option for the evaluation of thyroid nodules and cervical lesions.

Thyroid Imaging for the Assessment of Diffuse Thyroid Disease: Ultrasound

Non-autoimmune non-toxic diffuse goitre appears diffusely enlarged with a uniform or discretely irregular echo pattern without nodules. Various degrees of hypoechogenicity may be evident, but, when marked, suggest the presence of autoimmunity. In Hashimoto's thyroiditis, hypoechogenicity is always marked but may be inhomogeneous. US cannot reliably differentiate between benign autoimmune thyroiditis and lymphoma or carcinoma. Therefore, goitre growth, especially in the L-thyroxine-treated patient, should raise suspicion of lymphoma and lead to biopsy or operation. In Graves' disease, the thyroid is most often enlarged and the echo pattern homogeneous. Echogenicity can be normal to markedly decreased and the latter suggests a decreased probability of achieving remission on antithyroid drugs. Colour flow Doppler can demonstrate the rich vascularity and increased flow related to the degree of hyperthyroidism. Subacute thyroiditis leads to thyroid enlargement and areas of marked hypoechogenicity probably related to areas that are affected. Remission leads to normalization of size, but areas of hypoechogenicity may remain long after remission is obtained [11].

Computed Tomography and Magnetic Resonance

The normal thyroid gland is easily visualized on CT and its density is always higher than surrounding tissues. Differences in density reported from various countries reflect differences in iodine intake. Disease in the thyroid usually leads to decreased ability to concentrate iodine, therefore, reduced density on CT is the hallmark of thyroid disease [2]. The exact density measurements have not proved useful in distinguishing between various thyroid disorders. Thus, the CT image may be compatible with a certain diagnosis but rarely specific for it.

Non-autoimmune non-toxic diffuse goitre appears to be homogeneously enlarged with various degrees of hypodensity. Graves' disease is characterized by a 50–70% decrease in density and may be slightly inhomogeneous. Hashimoto's thyroiditis typically demonstrates an inhomogeneous iodine distribution and a 50% decrease in CT density which is lowest in hypothyroid individuals. Increasing goitre size is characteristically associated with decreasing density. Asymmetrical hypodense areas should raise the suspicion of lymphoma or carcinoma.

Subacute thyroiditis is also characterized by hypodensity and is focal or diffuse depending on the extent of the disease. In the initial phases, acute suppurative thyroiditis has no characteristic CT image; however, as infection progresses, loculated abscesses with hypodensity may appear.

MRI is highly sensitive but just as non-specific as US and CT in differentiating benign from malignant lesions [2]. It can distinguish solid from simple and complex cysts. It allows thyroid size determination and, as with CT and in contrast to US, it can visualize the retrotracheal area and retroclavicular or intrathoracic goitre and is less operator dependent [12]. The paramagnetic contrast agent gadolinium allows visualization of tumour vascularity. The drawbacks are cost, very limited availability for this purpose, length of the investigation, and cooperability (5% of patients cannot cooperate due to claustrophobia and some, especially children, need to be sedated). Patient and tissue movement (e.g. swallowing) decreases image quality and calcifications are better visualized with CT.

On T_1-weighted images the normal thyroid gland is clearly seen on MRI and shows a nearly homogeneous signal with an intensity as the adjacent neck muscles. On T_2-weighted images, the normal thyroid has a much greater signal intensity than adjacent muscles. Blood vessels, lymph nodes, fat, and muscle are clearly identified and distinguished from the thyroid. In Graves' disease the thyroid has slightly heterogeneous diffusely increased signal on both T_1- and T_2-weighted images. Hashimoto's thyroiditis causes a heterogeneous signal intensity on T_1-weighted images and a diffusely increased signal on T_2-weighted images.

Currently, the only indication for CT or MRI in diffuse, non-neoplastic thyroid disease is the presurgical characterization of

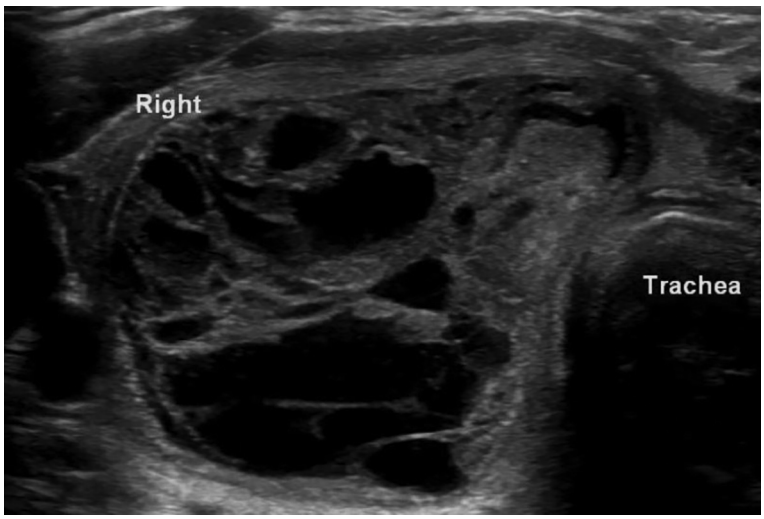


Figure 3.1.7.2 Trans-sectional greyscale US image of the right thyroid lobe showing a cystic–solid nodule with a spongiform appearance.

Box 3.1.7.3 Recommended variables to be included in a US reporting system

- Thyroid volume
- Echogenicity and vascularity of the gland
- Nodules (above 5 mm unless highly suspect)
- Location (side, superior, medial, inferior)
- Size (3 diameters +/– volume), and change in size in case of follow-up US
- Shape, margins, echogenicity, composition
- TIRADS score (see below)
- Retrosternal extension
- Tracheal deviation
- Study of lymph nodes (levels II, III, IV, V, VI)

Based on US features the categorization in five groups correlates to a malignancy risk (see Table 3.1.7.2) [14].

large goitres that are associated with local pressure symptoms. Large goitres may variably affect the trachea and oesophagus, and this may result in stenosis of the upper airway and/or swallowing dysfunction that may be clinically ill-defined unless appropriately evaluated [13]. In these patients presurgical cross-sectional imaging should be performed for a better definition of the indication and timing of surgical resection and for an appropriate planning of the surgical approach [13]. Indeed, both CT and MRI may accurately define the extent of the intrathoracic growth of the goitre, the displacement of the vascular structures of the neck, and the severity of the airway lumen reduction [10, 13].

Thyroid Imaging for the Assessment of Nodular Thyroid Disease: Ultrasound

In case of detection of thyroid nodule(s), a neck US should be performed with the evaluation of: (a) size and number; (b) texture and echogenicity (see Figure 3.1.7.2); (c) shape and margins; (d) presence and type of calcifications; and (e) vascularization and stiffness pattern [2]. Based on this evaluation, a risk stratification rating the probability of malignancy of the nodule can be performed [14]. The US TIRADS (Thyroid Imaging Reporting and Data System) categorization of thyroid nodules is based on an analytical and detailed categorization of US patterns, organized in different classes and subclasses of thyroid nodules [14]. It is recommended that assessment in a standardized reporting system is performed (see Box 3.1.7.3)

Enlarged lymph nodes may represent a benign reactive process or malignancy. Benign lymph nodes typically are oval, with a hyperechoic central stripe and vascular flow being present in the centre of the node. Loss of the hilum is felt to represent interruption of lymphatic flow by tumour invasion. However, the hilum may not always be easily defined by US in a benign lymph node, and the specificity for lack of a hilum has been reported to be only 29% for predicting the presence of cancer [15]. Malignant lymph nodes tend to become rounded in shape, distinguishing them from fusiform benign lymph nodes, but may also become irregular with microcalcifications (see Figure 3.1.7.3). Central and lateral neck lymph nodes are affected by metastatic papillary thyroid carcinoma (PTC) in up to 70% of cases [15]. US is acknowledged to be the first-line imaging modality for node assessment. The sensitivity of US in detecting abnormal lymph nodes ranges from 25% to 60% for the central neck and from 70% to 95% for the lateral neck [15]. US features predictive of malignant lymph node involvement are summarized in Table 3.1.7.1 [15].

Computed Tomography and Magnetic Resonance

CT and MRI are less accurate than US for the evaluation of risk of malignancy in focal thyroid lesions. Small size nodules may be undetectable by CT and/or MRI evaluation, and the presence of microcalcifications may easily be overlooked. Thus, after US characterization, no further imaging is needed for thyroid lesions that are not associated with signs of infiltrative growth of the cervical structures. Primary thyroid lymphoma is usually diagnosed in elderly subjects with a long-standing history of nodular goitre or Hashimoto's thyroiditis and often present with compressive symptoms of the airways and oesophagus. The disease commonly

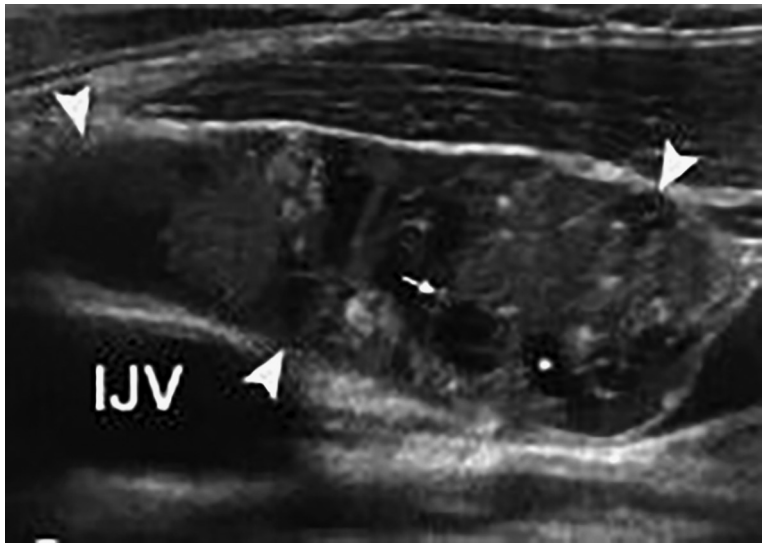


Figure 3.1.7.3 Transverse sonogram of a metastatic lymph node. Irregular heterogeneous echo texture with several microcalcifications. IJV, internal jugular vein.

Table 3.1.7.1 US features predictive of malignant lymph node involvement

Criterion	Sensitivity	Specificity
Size >1 cm	68%	75%
Shape (ratio of long axis to short axis <2.0)	46%	64%
Microcalcifications	46%	100%
Peripheral hypervascularity	86%	82%

Source: Data from Leboulleux S, Girard E, Rose M, Travagli JP, Sabbah N, Caillou B, Hartl DM, Lassau N, Baudin E, Schlumberger M. 2007. Ultrasound criteria of malignancy for cervical lymph nodes in patients followed up for differentiated thyroid cancer. *J Clin Endocrinol Metab* 92:3590–4.

manifests as a spreading mass that is hypoattenuating at CT examination [2], and like in anaplastic carcinoma, abnormal regional lymph nodes are commonly seen.

Multinodular goitre is often an enlarged asymmetrical gland with multiple low-density areas of varying degrees of discreteness (see Figure 3.1.7.4). CT density is decreased but in an inhomogeneous way. After intravenous contrast, enhancement is obtained except for areas containing haemorrhage, necrosis, or cysts. Calcifications are seen in up to 50% of goitres. Compression of the trachea, oesophagus, and great vessels is easily ascertained, and CT has found use especially in patients with monstrous and partly intrathoracic goitres, where it is ideal for the estimation of tracheal compression and quantitation of the intrathoracic extension of the goitre. Anatomical continuity with the cervical thyroid as well as a CT density greater than muscle, provides evidence of its thyroidal origin. Mediastinal lymphoma, lymphadenopathy, or thymus usually has markedly lower CT densities.

On CT imaging simple cysts are hypodense lesions, smooth-walled, and surrounded by normal thyroid tissue. The density of cyst fluid is always less than muscle and contrast injection does not lead to enhancement. Complex cysts are easily distinguished from simple cysts.

Table 3.1.7.2 TIRADS categories and malignancy risk

Category	US features	Malignancy risk, %
EU-TIRADS 1: normal	No nodules	None
EU-TIRADS 2: benign	Pure cyst Entirely spongiform	≅0
EU-TIRADS 3: low risk	Ovoid, smooth isoechoic/ hyperechoic No features of high suspicion	2–4
EU-TIRADS 4: intermediate risk	Ovoid, smooth, mildly hypoechoic No features of high suspicion	6–17
EU-TIRADS 5: high risk	At least one of the following features of high suspicion: • Irregular shape • Irregular margins • Microcalcifications • Marked hypoechogenicity (and solid)	26–87

EU-TIRADS, European Thyroid Imaging Reporting and Data System; US, ultrasound.

Source: Reproduced with permission from Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *Eur Thyroid J*, 2017; 6: 225–37.

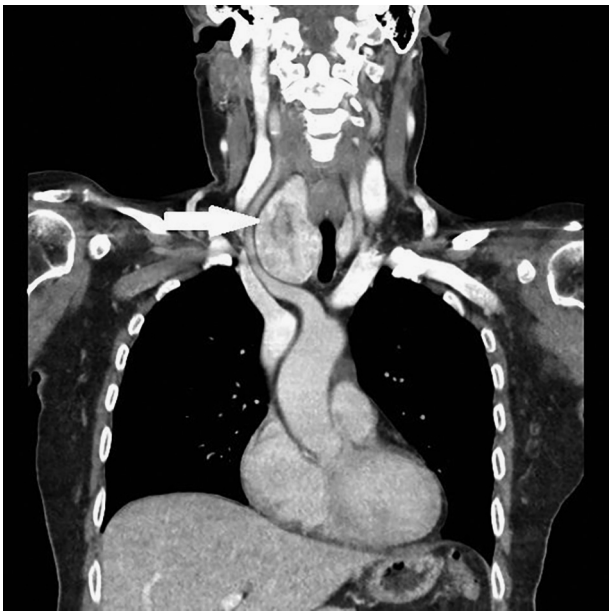


Figure 3.1.7.4 Coronal CT showing a partly intrathoracic multinodular goitre on the right side (the thyroid tissue being inhomogeneous with varying degrees of density) with slight compression of the trachea.

On MRI, simple cysts have a low-intensity signal on both T_1- and T_2-weighted imaging (see Figure 3.1.7.5). The intensity on T_1-weighted images increases with increasing protein and lipid content.

As with US, no CT characteristics will accurately separate benign from malignant lesions [2], although invasive growth into surrounding structures and metastases to cervical lymph nodes are suggestive of carcinoma. Papillary and follicular carcinomas are usually irregular low-density lesions and calcifications are present in the majority. There may be slight enhancement after contrast injection. The CT feature of medullary thyroid carcinomas is a single or multiple low-density lesions of variable size in one or both lobes. Lesions of 1–2 mm in size can be detected. Calcification is less often seen than in papillary carcinoma.

Large irregular masses of low attenuation with central cystic or necrotic areas are suggestive of anaplastic carcinoma especially if calcification is pronounced. Again, these features may also be seen in benign multinodular goitres. Invasion of the trachea, cricoid, or thyroid cartilage, and growth into the tracheal lumen is highly suggestive of carcinoma. Both Hashimoto's thyroiditis and thyroid lymphoma appear as masses of reduced density with little enhancement after contrast injection, and CT alone cannot make a distinction between them.

CT is of value in the follow-up of patients with thyroid cancer. Recurrence is evident as discrete low-density lesions within or outside the thyroid bed. Lymph node metastases typically have a regular rim, a core of central lucency, and no enhancement after intravenous contrast. CT is highly sensitive in detecting extrathyroidal spread of disease and therefore complementary to whole-body scanning with radioactive iodine.

Combined CT and positron emission tomography (PET) with [^{18}F]2-fluoro-2-deoxy-D-glucose (FDG) is a novel multimodality technology that enables a more precise anatomical localization of an area with increased focal uptake (potentially malignant lesion).

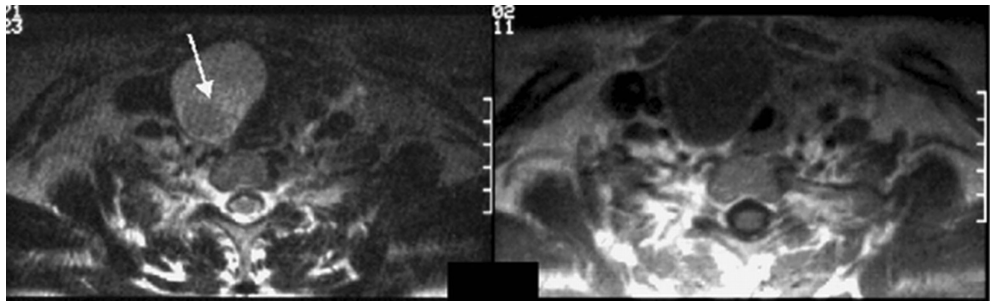


Figure 3.1.7.5 Axial MRI with T_2- (left) and T_1-weighted (right) scans of a cystic–solid thyroid nodule in the right thyroid lobe. A hypointense solid component (arrow) can be seen in comparison with the relatively hyperintense fluid. In the T_1-weighted picture, the lesion cannot be recognized in the hypointense fluid.

Its role in the initial evaluation of a thyroid nodule is limited but can be of value in case of indeterminate cytology [15]. It is increasingly used where there is suspicion of recurrence or spread of thyroid cancer [16].

On MRI follicular adenomas appear round or oval with a heterogeneous signal equal to or greater than that of normal tissue [2]. On T_2-weighted images the nodules have increased signal intensity. No MRI characteristics will accurately separate benign from malignant lesions. Thyroid carcinomas appear as focal or non-focal lesions of variable size; they are isointense or slightly hyperintense on T_1-weighted images and hyperintense on T_2-weighted images. The imaging characteristics of all types of thyroid carcinomas, including medullary carcinoma and lymphoma, are similar.

The extent of thyroid carcinoma can be determined preoperatively and may be useful in the planning of surgery. Extension into adjacent structures is usually evident. Gadolinium may be useful since metastatic nodes are enhanced centrally after gadolinium injection. Furthermore, in the postoperative follow-up recurrent carcinomas enhance with gadolinium, whereas scarring generally does not.

Magnetic Resonance Imaging

Please see Figure 3.1.7.5.

Conclusions and Recommendations

Thyroid imaging offers a relevant diagnostic and therapeutic contribution to the management of patients with thyroid disease. US is a sensitive tool for the evaluation of diffuse and focal thyroid disease and cervical masses. Thyroid US characteristics define the risk of malignancy in thyroid nodules and aids in making the decision of which nodules need FNA. Additionally, US examination confirms the presence of size abnormalities and/or possible changes in thyroid morphology in patients with clinical findings suggestive of thyroid dysfunction and/or inflammation. Thus, all patients suspected of thyroid disease benefit from a dedicated cervical US.

CT and MRI can provide much of the information that is obtained with US [2]. Their greater expense, limited availability, and other drawbacks argue against their use most of the time. CT is valuable in determining the extent of a substernal goitre or in the evaluation of a mediastinal mass. It can give valuable information in the evaluation of thyroid carcinoma and its spread. MRI may be useful in the same setting and is generally superior to CT in the evaluation of recurrent carcinoma, be it in the thyroid bed or in regional lymph nodes [2]. Therefore, a selective use of cross-sectional imaging with CT or MRI is recommended and should be based on clinical and sonographic indicators of extensive disease.

Preoperative PET-CT is not universally performed in patients with thyroid cancer because 18F-FDG uptake depends on the degree of tumour cell differentiation, such that uptake is higher in poorly differentiated than in well-differentiated thyroid tumours. In addition, well-differentiated thyroid tumours with iodine avidity have low glucose metabolism, and therefore PET is more useful for detecting recurrent or metastatic differentiated thyroid cancer in patients with negative radioiodine scans and elevated thyroglobulin levels during follow-up [17, 18].

REFERENCES

1. Jarlov AE, Nygaard B, Hegedüs L, Hartling SG, Hansen JM. Observer variation in the clinical and laboratory evaluation of patients with thyroid dysfunction and goiter. *Thyroid*, 1998; **8**: 393–8.
2. Hegedüs L, Bennedbæk F. Non-isotopic techniques of thyroid imaging. In: Braverman LE, Cooper D, eds. *Werner & Ingbar's The Thyroid: A Fundamental & Clinical Text*, 9th edn. Philadelphia, PA: Lippincott, Williams & Wilkins, 2012: 310–20.
3. Burch HB, Burman KD, Cooper DS, Hennessey JV, Vietor NO. A 2015 survey of clinical practice patterns in the management of thyroid nodules. *J Clin Endocrinol Metab*, 2016; **101**: 2853–62.
4. Gharib H, Papini E, Garber JR, *et al*. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. *Endocr Pract*, 2016; **22**: 622–39.
5. Hegedüs L. Thyroid ultrasound. *Endocrinol Metab Clin North Am*, 2001; **30**: 339–60, viii–ix.
6. Bonnema SJ, Hegedüs L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*, 2012; **33**: 920–80.
7. Gharib H, Hegedüs L, Pacella CM, Baek JH, Papini E. Clinical review: nonsurgical, image-guided, minimally invasive therapy for thyroid nodules. *J Clin Endocrinol Metab*, 2013; **98**: 3949–57.

8. Hegedüs L. Clinical practice. The thyroid nodule. *N Engl J Med*, 2004; **351**: 1764–71.
9. Delange F, Benker G, Caron P, *et al*. Thyroid volume and urinary iodine in European schoolchildren: standardization of values for assessment of iodine deficiency. *Eur J Endocrinol*, 1997; **136**: 180–7.
10. Nachiappan AC, Metwalli ZA, Hailey BS, Patel RA, Ostrowski ML, Wynne DM. The thyroid: review of imaging features and biopsy techniques with radiologic-pathologic correlation. *Radiographics*, 2014; **34**: 276–93.
11. Bennedbaek FN, Hegedüs L. The value of ultrasonography in the diagnosis and follow-up of subacute thyroiditis. *Thyroid*, 1997; **7**: 45–50.
12. Jennings A. Evaluation of substernal goiters using computed tomography and MR imaging. *Endocrinol Metab Clin North Am*, 2001; **30**: 401–14, ix.
13. Sorensen JR, Hegedüs L, Kruse-Andersen S, Godballe C, Bonnema SJ. The impact of goitre and its treatment on the trachea, airflow, oesophagus and swallowing function. A systematic review. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 481–94.
14. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J*, 2017; **6**: 225–37.
15. Yeh MW, Bauer AJ, Bernet VA, *et al*. American Thyroid Association statement on preoperative imaging for thyroid cancer surgery. *Thyroid*, 2015; **25**: 3–14.
16. Sebastianes FM, Cerci JJ, Zanoni PH, *et al*. Role of 18F-fluorodeoxyglucose positron emission tomography in preoperative assessment of cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab*, 2007; **92**: 4485–8.
17. Lind P, Kohlfurst S. Respective roles of thyroglobulin, radioiodine imaging, and positron emission tomography in the assessment of thyroid cancer. *Semin Nucl Med*, 2006; **36**: 194–205.
18. Kim BS, Ryu HS, Kang KH. The value of preoperative PET-CT in papillary thyroid cancer. *J Int Med Res*, 2013; **41**: 445–56.

3.1.8 Epidemiology of Thyroid Disease and Swelling

Mark P.J. Vanderpump

Introduction

Thyroid disorders are among the most prevalent of medical conditions. Their manifestations vary considerably from area to area and are determined principally by the availability of iodine in the diet. The limitations of epidemiological studies of thyroid disorders should therefore be borne in mind when considering the purported frequency of thyroid diseases in different communities [1, 2].

Iodine is an essential component of the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3) produced by the thyroid gland. The ideal dietary allowance of iodine recommended by World Health Organization (WHO) in adults is 150 µg of iodine per day which increases to 250 µg per day in pregnancy and lactation. Iodine deficiency impairs thyroid hormone production and has adverse effects throughout life, particularly early in life as it impairs cognition and growth [3].

Iodine deficiency is a global problem with large populations at risk who are living in an environment where the soil has been deprived of iodine. The mountainous regions of Europe, the Northern Indian Subcontinent, the extensive mountain ranges of China, the Andean region in South America and the lesser ranges of Africa are all iodine deficient. Iodine deficiency remains a significant problem despite major national and international efforts to increase iodine intake, primarily through the voluntary or mandatory iodization of salt. In 2007, the WHO estimated that two billion people, including 285 million school-age children, still had iodine deficiency, defined as a urinary iodine excretion of less than 100 µg/L [4]. Iodine deficiency has emerged as issue in industrialized countries, previously thought of as iodine sufficient. International efforts to control iodine deficiency are slowing and reaching those that remains deficient poses major challenges [5].

In areas where the daily iodine intake is less than 50 µg, goitre is usually endemic, and when the intake falls below 25 µg per day, congenital hypothyroidism is seen. The prevalence of goitre in areas of severe iodine deficiency can be as high as 80%. Goitrogens in the diet, such as thiocyanate in incompletely cooked Cassava or thioglucosides in Brassica vegetables, can explain some of the differences in prevalence of endemic goitre in areas with similar degrees of iodine deficiency. Iodization programmes are of proven value in reducing goitre size and in preventing goitre development and cretinism in children [3].

The effects of mild-to-moderate iodine deficiency on cognition are less well known than those of moderate-to-severe deficiency but there is a continuum of disability with more subtle impairments of the intelligence quotient (IQ) and motor ability associated with less severe deficiency. A systematic review of available published studies from 1980–2011 found that, regardless of study design, iodine deficiency had a substantial impact on mental development of children five years old and under which translated into 6.9 to 10.2 IQ points lower in iodine deficient compared with iodine-replete children [6]. Autonomy can develop in nodular goitres, occasionally leading to thyrotoxicosis, and iodization programmes can also induce thyrotoxicosis, especially in those older than age 40 with nodular goitres [3].

In iodine-replete areas, most persons with thyroid disorders have autoimmune disease, ranging through primary atrophic hypothyroidism, Hashimoto's thyroiditis to hyperthyroidism

caused by Graves' disease. Cross-sectional studies have determined the prevalence of hyperthyroidism and hypothyroidism and the frequency and distribution of thyroid autoantibodies in different, mainly Caucasian, communities [1, 2]. United States (US) data revealed differences in the frequency of thyroid dysfunction and serum antithyroid antibody concentrations in different ethnic groups [7], whereas European studies revealed the influence of dietary iodine intake on the epidemiology of thyroid dysfunction [8]. Studies of incidence of autoimmune thyroid disease have only been conducted in a small number of developed countries [9]. Following a review of the available epidemiological data, the value of screening adult populations for autoimmune thyroid disease will be considered.

Hyperthyroidism

Hyperthyroidism has a significant short-term morbidity and long-term morbidity and mortality [10]. In epidemiological studies, the clinical diagnosis of thyrotoxicosis should be supported by measurements of serum T_4 or T_3 and thyrotropin (TSH) concentrations. Biochemical tests of thyroid function may reveal the diagnosis before it is clinically apparent. A rise in serum T_3 and fall in serum TSH are the earliest measures of thyroid overactivity, followed by a rise in serum T_4. The most common cause of hyperthyroidism in iodine-replete communities is Graves' disease, followed by toxic multinodular goitre, whereas rarer causes include an autonomously functioning thyroid adenoma, autoimmune or viral thyroiditis, excessive levothyroxine (L-T_4) replacement or drugs including amiodarone and lithium carbonate. In epidemiological studies, the aetiology is rarely ascertained.

Prevalence of Hyperthyroidism

The prevalence of hyperthyroidism in women is between 0.5 and 2% and is ten times more common in women than in men in iodine-replete communities [1, 2]. In the Whickham survey in Northeast England, the prevalence of undiagnosed hyperthyroidism was 4.7/1000 women [11]. Hyperthyroidism had been previously diagnosed and treated in 20/1000 women, rising to 27/1000 women when possible but unproven cases were included, as compared with 1.6 to 2.3/1000 men, in whom no new cases were found at the survey. The mean age at diagnosis was 48 years. In the US National Health and Nutrition Examination Survey (NHANES III), of those subjects who were neither taking thyroid medication nor reported a history of thyroid disease, 2/1000 had 'clinically significant' hyperthyroidism, defined as a serum TSH concentration less than 0.1 mU/L and serum total T4 concentration greater than 170 nmol/L [7]. A cross-sectional survey of 25 682 subjects aged over 18 years attending a Health Fair in Colorado, US found that overt hyperthyroidism, defined as serum TSH concentration less than 0.01 mU/L, was present in only 1/1000 of those not taking thyroid medication [12]. A higher prevalence is seen in iodine-deficient areas [8]. Prevalence data in elderly persons show a wide range between 0.4 and 2.0%. The reported prevalence rates for previously undiagnosed hyperthyroidism in hospitalized patients is between 0.3 and 1% and consistent with community surveys [1].

Table 3.1.8.1 The effect of environmental iodine intake on the prevalence of subclinical thyroid disease

Iodine status	Subclinical hypothyroidism	Subclinical hyperthyroidism
Deficient	1–4%	6–10%
Replete	4–9%	1–2%
Excess	18–14%	<1%

Subclinical Hyperthyroidism

The introduction of assays for serum TSH that were sensitive enough to distinguish between normal and low concentrations allowed subjects with subclinical hyperthyroidism to be identified. Subclinical hyperthyroidism is defined as a low serum TSH concentration and normal serum T_4 and T_3 concentrations, in the absence of hypothalamic or pituitary disease, non-thyroidal illness, or ingestion of drugs that inhibit TSH secretion such as glucocorticoids or dopamine [13]. Epidemiological studies differ in the definition of a low serum TSH concentration and whether the subjects included were receiving L-T_4 therapy.

The reported overall prevalence ranges from 0.5 to 6.3%, with men and women over 65 years having the highest prevalence and approximately half taking L-T_4 [1]. The Colorado study of 25 682 healthy volunteers (of whom 88% were white) found 2% had a subnormal serum TSH, with more than half on L-T_4 [12]. In the NHANES III study the prevalence was highest in those subjects aged 20 to 39 years and those aged greater than 79 years [7]. The percentage with serum TSH concentrations less than 0.4 mU/L was significantly higher in women than men, and black subjects had a higher prevalence (0.4%) than whites (0.1%) or Mexican Americans (0.3%). The prevalence is higher in iodine-deficient populations due to functional autonomy from nodular goitres [8] (Table 3.1.8.1).

Among subjects with subclinical hyperthyroidism, those with low but detectable serum TSH values may recover spontaneously when retested [13]. Non-thyroidal illness is an important cause of false-positive serum TSH test results. There are limited data on the risk of progression of subclinical hyperthyroidism to overt hyperthyroidism. Those with an undetectable serum TSH and confirmed aetiology as determined by thyroid scintigraphy due to Graves' disease or nodular disease have an annual incidence of approximately 5–8% [14]. A population study in Tayside, United Kingdom (UK) followed 2024 subjects with at least two serum TSH measurements below the reference range for at least four months for up to seven years [15]. Few subjects developed hyperthyroidism (0.5–0.7%) and the percentage of those reverting to normal increased with time and was more common if the baseline serum TSH was between 0.1 and 0.4 mU/L.

Incidence of Hyperthyroidism

The incidence data available for overt hyperthyroidism in men and women from large population studies are comparable, at 0.4/1000 women and 0.1/1000 men, but the age-specific incidence varies considerably [1, 2]. The peak age-specific incidence of Graves' disease was between 20 and 49 years in two studies but increased with age in Iceland and peaked at 60 to 69 years in Malmö, Sweden. The peak age-specific incidence of hyperthyroidism caused by toxic nodular goitre and autonomously functioning thyroid adenomas in the Malmö study was over 80 years. The only available data in a

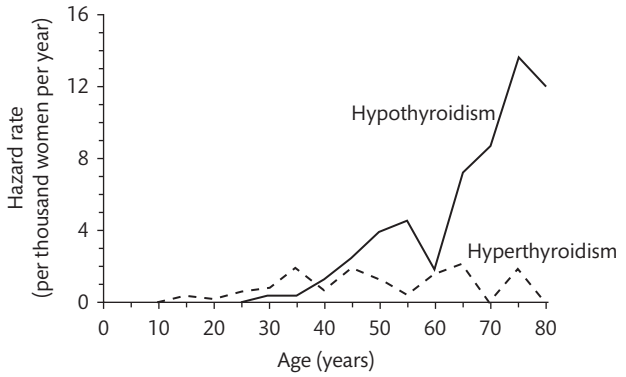


Figure 3.1.8.1 Age-specific hazard rates for the development of overt hyperthyroidism and hypothyroidism in women at 20-year follow-up of the Whickham survey.

Reproduced with permission from Vanderpump MPJ, Tunbridge WMG, French JM, Appleton D, Bates D, Clark F, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol*, 1995; 43: 60. Copyright © 2008, John Wiley and Sons, Blackwell Science.

black population, from Johannesburg, South Africa, suggest a ten-fold lower annual incidence of hyperthyroidism (0.09/1000 women and 0.007/1000 men) than in whites [2].

In the 20-year follow-up of the Whickham cohort, the mean annual incidence of hyperthyroidism in women was 0.8/1000 survivors (95% confidence interval (CI), 0.5–1.4) [16]. The incidence rate was similar in the deceased women. No new cases were detected in men. An estimate of the probability of the development of hyperthyroidism in women at a particular time averaged 1.4/1000 between the ages of 35 and 60 years (Figure 3.1.8.1) [16]. Serum antithyroid antibody status or goitre was not associated with the development of hyperthyroidism at follow-up. Other cohort studies provide comparable incidence data, which suggests that many cases of hyperthyroidism remain undiagnosed in the community unless routine testing is undertaken [9]. In Tayside, UK, 620 incident cases of hyperthyroidism were identified from medical records with an incidence rate of 0.77/1000 per year (95% CI, 0.70–0.84) in women and 0.14/1000 per year (95% CI, 0.12–0.18) in men [17]. The incidence increased with age, and women were affected two to eight times more than men across the age range. The incidence increased in women but not men between 1997 and 2001 [18].

Hypothyroidism

Hypothyroidism is an insidious condition with a significant morbidity, and the subtle and non-specific symptoms and signs may be mistakenly attributed to other illnesses, particularly in postpartum women and older people [19]. The earliest biochemical abnormality is an increase in serum TSH concentration associated with normal serum free T_4 and T_3 concentrations (subclinical hypothyroidism or mild thyroid failure), followed by a decrease in serum free T_4 concentration, at which stage, most patients have symptoms and benefit from treatment (overt hypothyroidism). In iodine-replete communities, the cause is either chronic autoimmune disease (atrophic autoimmune thyroiditis or goitrous autoimmune thyroiditis (Hashimoto's thyroiditis)) or destructive treatment for hyperthyroidism, which may account for up to one-third of cases of hypothyroidism in the community.

Congenital Hypothyroidism

Congenital hypothyroidism affects approximately 1 newborn in 3500 to 4000 births and is the most treatable cause of intellectual disability [20]. There is an inverse relationship between age at diagnosis and IQ in later life. In iodine-replete areas, 85% of the cases are due to sporadic developmental defects of the thyroid gland (thyroid dysgenesis), such as the arrested migration of the embryonic thyroid (ectopic thyroid) or a complete absence of thyroid tissue (athyreosis). The remaining 15% have thyroid dyshormonogenesis defects transmitted by an autosomal recessive mode of inheritance. Iodine deficiency (less than 25 μg per day), particularly in preterm infants, accounts for many cases in Eastern Europe, Asia, and Africa. Clinical diagnosis occurs in less than 5% of newborns with hypothyroidism because symptoms and signs are often minimal, so it is not possible to predict which infants are likely to be affected. Without prompt diagnosis and treatment, most affected children gradually develop growth failure, irreversible intellectual disability, and a variety of neuropsychological deficits. The apparent incidence of congenital hypothyroidism has more than doubled due to more inclusive diagnostic criteria, shifting demographics, and increasing survival of preterm infants [21]. The greatest increase has occurred in mildly affected children. Congenital hypothyroidism may be transient or persistent, but the natural history cannot be predicted by severity at diagnosis. In premature infants, who are especially vulnerable to hypothyroidism, the rise in serum TSH may be delayed and therefore detected only by routine follow-up screening.

Asymptomatic Autoimmune Thyroiditis

Raised serum concentrations of thyroid antibodies (antithyroid peroxidase (microsomal) (TPOAb) and antithyroglobulin (TGAb)) correlate with the presence of focal thyroiditis in thyroid tissue obtained by biopsy and at autopsy from patients with no evidence of hypothyroidism during life. Early post-mortem studies confirmed histological evidence of chronic autoimmune thyroiditis in 27% of adult women, with a rise in frequency over 50 years, and 7% of adult men, and diffuse changes in 5% of women and 1% of men [1]. Patients with hypothyroidism caused by either atrophic or goitrous autoimmune thyroiditis usually have high serum concentrations of these same antibodies. These antibodies also are often detected in serum of patients with Graves' disease and other thyroid diseases, but the concentrations are usually lower. There is considerable variation in the frequency and distribution of antithyroid antibodies because of variations in techniques of detection, definition of abnormal titres, and inherent differences in the populations tested.

A significant proportion of subjects in the community have asymptomatic chronic autoimmune thyroiditis of whom a substantial proportion have subclinical hypothyroidism (Figure 3.1.8.2) [11]. In the NHANES III survey the percentage of subjects with high serum thyroid peroxidase (TPO) and TG antibody concentrations increased with age in both men and women, and high concentrations were more prevalent in women than in men and less prevalent in blacks than in other ethnic groups [7]. Using a competitive immunoassay procedure, the reported prevalence of detectable TGAb and TPOAb levels were 10% and 12% in the healthy population.

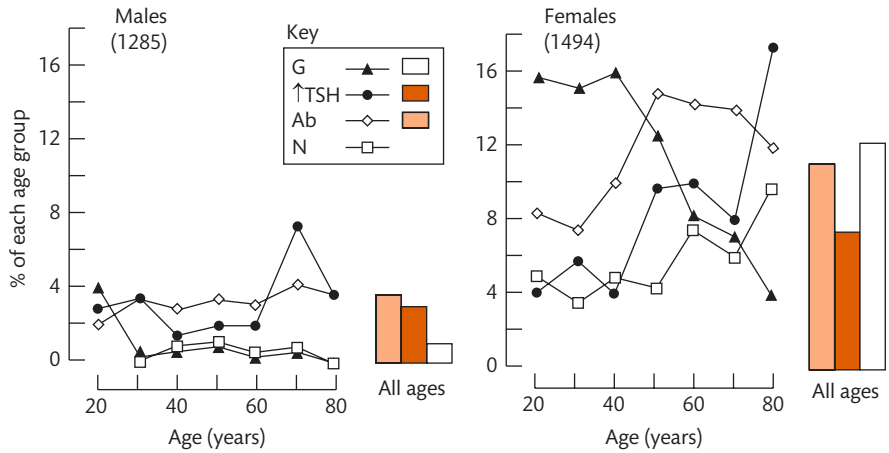


Figure 3.1.8.2 Age and sex distribution of thyroid microsomal antibodies (Ab), raised serum TSH greater than 6 mU/L (↑TSH), visible diffuse and multinodular goitre (G), and nodules (N) in the Whickham survey.

Reproduced with permission from Tunbridge WMG, Evered DC, Hall R, Appleton D, Brewis M, Clark F, *et al*. The spectrum of thyroid disease in the community: the Whickham survey. *Clin Endocrinol*, 1977; 7: 485. Copyright © 2008, John Wiley and Sons, Blackwell Science.

A hypoechoic ultrasound pattern may precede TPOAb positivity in autoimmune thyroid disease, and TPOAb may not be detected in more than 20% of individuals with ultrasound evidence of thyroid autoimmunity [13].

Prevalence of Hypothyroidism

In iodine-replete communities, the prevalence of spontaneous hypothyroidism is between 1% and 2%, and it is more common in older women and ten times more common in women than in men [1]. In the Whickham survey, the prevalence of previously diagnosed and treated hypothyroidism was 14/1000 women, increasing to 19/1000 women when possible, but unproven, cases were included [11]. The overall prevalence in men was less than 1/1000. One-third had been previously treated by surgery or radioiodine for thyrotoxicosis. The mean age at diagnosis was 57 years. The Whickham data are comparable with other studies where the prevalence of newly diagnosed hypothyroidism ranged between 0.6 and 12/1000 women and between 1.3 and 4.0/1000 in men investigated in Northern Europe, Japan, and the United States [1]. In the Colorado and NHANES III studies, the prevalence was 4/1000 and 3/1000, respectively [7, 12]. The prevalence is higher in surveys of older people in the community [1] and lower in areas of iodine deficiency [8]. The testing of hospital inpatients, predominantly elderly women, confirm a prevalence of 2% [1].

Subclinical Hypothyroidism

The term subclinical hypothyroidism represents a compensated state in which increased TSH output is required to maintain normal circulating thyroid hormone levels. An elevated serum TSH is a sensitive indicator of some degree of thyroid failure and, in contrast to below normal serum TSH levels, a clear inverse relationship is found with free T_4 levels. It is found either post radioiodine therapy or post surgery in up to 50% of apparently euthyroid patients. It may be evident for only a few months, but more often it represents a stage in the progression towards overt thyroid failure. Less frequent causes include external beam irradiation of malignant tumours of the head and neck, and drugs including lithium, amiodarone, and undiagnosed Addison's disease. In the community, the most common aetiology is chronic autoimmune thyroiditis [1, 13].

With respect to epidemiological studies, the definition of subclinical hypothyroidism varies from any increase in serum TSH to values more than 10 mU/L or, more stringently, a serum TSH value more than 10 mU/L and a positive test for circulating thyroid antibodies in serum. The term implies that patients should be asymptomatic, although symptoms are difficult to assess, especially in those in whom thyroid function tests have been checked because of non-specific complaints such as tiredness. Spontaneous recovery has also been described in subjects with subclinical hypothyroidism, although the frequency of this phenomenon is unclear. Normalization of serum TSH concentrations is more likely to occur in patients with negative antithyroid antibodies and serum TSH levels less than 10 mU/L, and within the first 2 years after diagnosis [13].

Controversy exists regarding the upper limit of the reference range for serum TSH [22, 23]. Reference ranges are derived from a reference population that comprises a large group of subjects who do not have thyroid disease and are otherwise well. By convention, a reference range usually only comprises 95% of a reference population. Thus, 2·5% of 'normal' individuals will fall above the reference range and 2·5% will fall below the range. For serum TSH, the reference population shows a log normal distribution and has a diurnal variation with the reference range in thyroid disease-free individuals typically cited as between 0.4 and 4.0 mU/L. The serum TSH reference range varies in different ethnic communities, trimesters of pregnancy, and progressively shifts towards higher concentration with age. Analysis of the NHANES III data suggest that the reference range for serum TSH rises with age as the 97.5 centile for those subjects aged more than 80 years was 7.49 mU/L and 70% had a serum TSH more than the population defined upper limit of the reference range of 4.5 mU/L of whom only 40% were antithyroid antibody positive [24].

In the original Whickham survey, 8% of women (10% of women over 55 years of age) and 3% of men had subclinical hypothyroidism [11]. In the Colorado study, 9.4% of the subjects had a high serum TSH concentration, of whom 9.0% had subclinical hypothyroidism

[12]. Among those with a high serum TSH concentration, 74% had a value of between 5.1 and 10 mU/L and 26% had a value greater than 10 mU/L. The percentage of subjects with a high serum TSH concentration was higher for women than men in each decade of age and ranged from 4% to 21% in women and 3% to 16% in men. The NHANES III study found that 11% of those aged 20 to 29 had a serum TSH greater than 2.5 mU/L, increasing to 40% in those aged 80 and over. Serum TSH concentrations were higher in whites than blacks, independent of serum antithyroid antibody concentrations [7]. Approximately 2% of adolescents aged 12 to 19 years had a serum TSH greater than 4.5 mU/L. Community studies of elderly persons have confirmed approximately 10% of subjects over 60 years having serum TSH values above the normal range [1]. Subclinical hypothyroidism is found at higher frequency in areas where iodine intake is high most cases are not of autoimmune origin [8] (Table 3.1.8.1). In surveys of hospital inpatients, the point prevalence rates were similar being between 3% and 6% with most subjects reverting to normal thyroid function three months following the acute illness [1].

Incidence of Hypothyroidism

After destructive treatment of hyperthyroidism with either radioiodine or surgery, the incidence of overt hypothyroidism is greatest in the first year. The incidence of hypothyroidism in patients with Graves' disease was higher than that in patients with nodular goitre (55% vs. 32%) and increased in those given higher doses of radioiodine [1]. If subclinical hypothyroidism is present one year or more after radioiodine or surgical treatment, then the annual rate of progression to overt hypothyroidism after either treatment is 2–6%. Treatment of Graves' disease with antithyroid drugs alone is also associated with the eventual development of hypothyroidism in 5–20% of cases from either autoimmune thyroiditis or the presence of TSH-blocking antibodies. The incidence of hypothyroidism after surgery, external radiation therapy of the neck, or both, in patients with head and neck cancer (including lymphoma) is as high as 50% within the first year after treatment, particularly in patients who underwent surgery and received high doses of radiation. The effect is dose-dependent, the onset is gradual, and subclinical hypothyroidism can be present for many years prior to the development of overt disease.

At the 20-year follow-up of the Whickham cohort the mean annual incidence of spontaneous hypothyroidism in the surviving women during the 20-year follow-up period was 3.5/1000 (95% CI, 2.8–4.5), increasing to 4.1/1000 (95% CI, 3.3–5.0) if all cases including those who had received destructive treatment for thyrotoxicosis were included [16]. The hazard rate increased with age to 13.7/1000 in women 75 to 80 years of age (Figure 3.1.8.1). The mean annual incidence during the 20-year follow-up period in men (all spontaneous except for one case of lithium-induced hypothyroidism) was 0.6/1000 (95% CI, 0.3–1.2). The risk of having developed hypothyroidism was examined with respect to risk factors identified in the first survey. Either raised serum TSH or positive antithyroid antibodies alone or in combination are associated with a significantly increased risk of hypothyroidism in surviving women (Table 3.1.8.2) [16]. The annual risk of spontaneous overt hypothyroidism was 4% in those who had both high serum TSH and antithyroid antibody concentrations, 3% if only their serum TSH concentrations was high, and 2% if only their serum thyroid antibody concentration was high; at the time of follow-up the respective rates of hypothyroidism were 55%, 33%, and 27%. The probability of developing hypothyroidism was higher in those women who had serum TSH concentrations above 2.0 mU/L and high serum titres of antithyroid microsomal antibodies at the first survey (Figure 3.1.8.3) [16]. Neither a positive family history of any thyroid disease, nor the presence of a goitre at either the first or the follow-up survey, or parity at first survey was associated with an increased risk of hypothyroidism.

Table 3.1.8.2 Development of spontaneous hypothyroidism in surviving women and men at 20-year follow-up of Whickham survey: odds ratios (with 95% CI)

Women	
TSH raised, regardless of thyroid antibody status	14 (9–24)
Thyroid antibody +, regardless of TSH status	13 (8–19)
If thyroid antibody -, effect of raised TSH alone	8 (3–20)
If thyroid antibody +, additional effect of raised TSH	5 (2–11)
If TSH normal, effect of thyroid antibody + alone	8 (5–15)
If TSH raised, additional effect of thyroid antibody +	5 (1–15)
TSH raised and thyroid antibody + combined	38 (22–65)
Men	
TSH raised, regardless of thyroid antibody status	44 (19–104)
Thyroid antibody +, regardless of TSH status	25 (10–63)
TSH raised and thyroid antibody + combined	173 (81–370)

Other incidence data for hypothyroidism are from short (and often small) follow-up studies [9] and confirm that serum TSH concentrations in the upper part of the normal range in this study have a predictive value. In Tayside, UK the standardized incidence of primary hypothyroidism remained between 3.90 and 4.89/1000 women per year between 1993 and 2001 [17, 18]. The incidence of hypothyroidism in men however significantly increased from 0.65 to 1.01/1000 per year ($P = 0.0017$) and the mean age at diagnosis of primary hypothyroidism decreased in women from 1994 to 2001.

Thyroid Disease in Pregnancy

Pregnancy has variable effects on thyroid hormone concentrations throughout pregnancy as well as being associated with goitre [25]. The latter is largely preventable by ensuring optimal iodine intake of at least 200 μg per day. Hypothyroidism in pregnancy usually characterized by a high serum TSH value has been found to occur in around 2–3% of otherwise normal pregnancies with the prevalence of overt hypothyroidism estimated to be up to 0.5% [26]. On a worldwide basis the most important cause of thyroid insufficiency remains iodine deficiency, while in iodine-replete communities the cause is usually chronic autoimmune thyroiditis. Untreated hypothyroidism may lead to obstetric complications, such as preterm delivery and fetal loss. Epidemiological data suggest that the children of women with hypothyroxinaemia may have psychoneurological

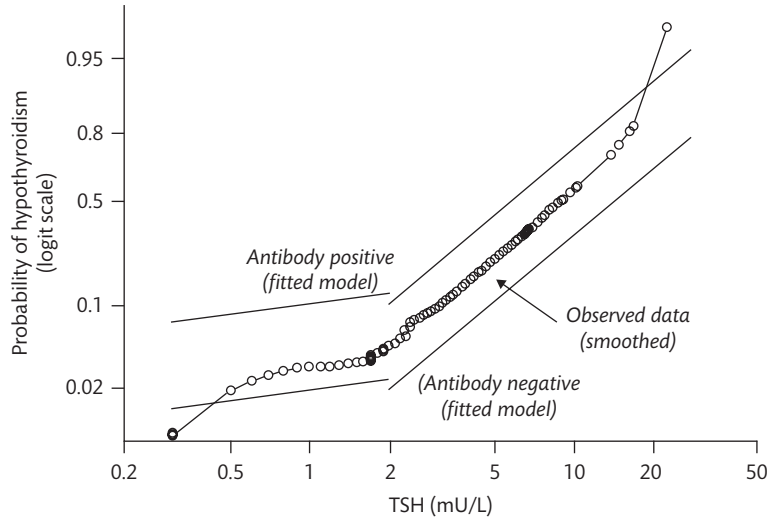


Figure 3.1.8.3 Probability for development of hypothyroidism within 20 years with increasing values of serum TSH at first Whickham survey in 912 survivors.

Reproduced with permission from Vanderpump MPJ, Tunbridge WMG, French JM, Appleton D, Bates D, Clark F, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol*, 1995; 43: 60. Copyright © 2008, John Wiley and Sons, Blackwell Science.

deficits. In classic areas of iodine deficiency, a similar range of deficits in children has been described where maternal hypothyroxinaemia rather than high serum TSH is the main biochemical abnormality [3]. In these areas, maternal iodine intake is often substantially less than the 200 μg per day currently recommended. Even in areas previously thought to be iodine sufficient, there is now evidence of substantial gestational iodine deficiency, which may lead to low maternal circulating thyroxine concentrations.

Hyperthyroidism is found in 0.1–0.4% of all pregnancies and is usually caused by Graves' disease characterized by TSH receptor (TSHR) antibodies, which usually decrease in titre throughout pregnancy. Maternal complications include miscarriage, placenta abruption, pre-eclampsia, and preterm delivery. High titres of TSHR antibodies predict a high risk of neonatal thyrotoxicosis [25].

Antithyroid antibodies, particularly TPOAb, occur in 10% of women at 14 weeks of gestation, which is compatible with the prevalence of antithyroid antibodies in community surveys [1, 13]. A proportion of these women will have subclinical hypothyroidism with a high serum TSH, but most will be euthyroid. However, after delivery a transient, destructive autoimmune thyroiditis that occurs between the twelfth and sixteenth week postpartum will develop in 50% of TPOAb positive women, as ascertained in early gestation, clinically apparent as postpartum thyroiditis (PPT) [26]. It presents as a temporary, usually painless, episode of hypothyroidism, occasionally preceded by a short episode of hyperthyroidism. Up to 25% of women progress to permanent hypothyroidism within approximately five years following an episode of PPT, particularly those with high antibody titres.

Goitre and Thyroid Nodules

The most common thyroid disease is simple (diffuse) goitre. The clinical grading of thyroid size is subjective and imprecise. The WHO grading system recognizes that an enlarged thyroid gland may be palpably but not visibly enlarged. Interexaminer variation may lead to differences in classifying a goitre as diffuse or multinodular. There is considerable overlap between the five WHO grades based on clinical criteria and thyroid volume estimated by ultrasonography. Ultrasonography has been used in epidemiological studies to assess thyroid size, resulting in much higher estimates of goitre prevalence than in studies in which goitre size was assessed by physical examination.

In cross-sectional surveys, the prevalence of diffuse goitre declines with age, the greatest prevalence is in premenopausal women, and the ratio of women to men is at least 4:1. In the Whickham survey, among the women 26% had a goitre; the frequency ranged from 31% in those aged less than 45 years (mostly diffuse) to 12% in those aged over 75 years (who had a higher proportion of nodular goitre) [11] (**Figure 3.1.8.2**). Longitudinal studies confirm the decreasing frequency of diffuse goitre with age [16]. This decline in frequency of diffuse goitres with age is in contrast to the increase in frequency of thyroid nodules and antithyroid antibodies with age. In the Whickham survey, less than 1% of the men but 5% of the women had thyroid nodules detected clinically, and the frequency increased to 9% in women more than age 75 [11]. A higher prevalence of nodular goitre is found in areas of iodine deficiency in Europe, such as Italy, Germany, and Denmark. Longitudinal data suggest an annual incidence for nodules of 1/1000 and that, once formed, they tend to remain present and benign for a long period of time [1].

With the increasing use of sensitive imaging techniques, an increasing proportion of thyroid nodules are detected incidentally. Many nodules are detected because of their size or anterior position in the neck, or the skill of the physician performing the examination but most thyroid nodules will not be clinically recognized. Up to 50% of nodules more than 1 cm detected by ultrasound are undetected by clinical examination. The prevalence of thyroid incidentaloma as an unexpected, asymptomatic thyroid nodule discovered during the investigation of an unrelated condition, is

67% with ultrasonography imaging, 15% with computed tomography or magnetic resonance imaging of the neck, and 1–2% with fluorodeoxyglucose positron emission tomography [27].

Thyroid Cancer

The clinical presentation of thyroid cancer is usually as a solitary thyroid nodule or increasing goitre size. Although thyroid nodules are common, thyroid cancers are rare. The four major histological types are papillary, follicular, medullary, and anaplastic, and each displays a different epidemiology. The annual incidence of all thyroid cancers ranges between 1 and 10/100 000 population in most countries and is two to four times more frequent in women than men [28, 29]. Papillary and follicular tumours, which comprise 60–90% of the total, are rare in children and adolescents but their incidence increases with age in adults.

Papillary thyroid carcinoma (PTC) is the most common thyroid malignancy and worldwide constitutes 50–90% of differentiated follicular cell-derived thyroid cancers. Papillary thyroid microcarcinomas (diameter less than 1 cm) are found in 4–36% of adults post-mortem in population-based studies. Most diagnoses of PTC occur in patients 30–50 years old (median age 44 years), and the majority (60–80%) occur in women. The reported increase in incidence of these carcinomas in recent years has been attributed to an improvement in pathological techniques and earlier and increased detection of small (subclinical) papillary cancers secondary to more widespread use of neck ultrasonography and fine-needle aspiration of very small thyroid nodules. However, a recent analysis of the National Cancer Institute's Surveillance, Epidemiology, and End Results database found an increase in the rates of differentiated thyroid cancer of all sizes, including tumours more than 4 cm and an increase in incidence-based mortality between 1974 and 2013, from 0.40 to 0.46/100 000 person-years [30]. External radiation exposure, particularly in childhood, is a major risk factor for papillary cancer. Four years after the nuclear accident at Chernobyl, a significant increase in the incidence of childhood thyroid cancer (almost exclusively papillary tumours in which a translocation of the *RET* gene occurs) was recorded in the regions most exposed [31]. There is no association between radioiodine therapy for thyrotoxicosis and subsequent development of thyroid cancer in adults.

Follicular thyroid cancer occurs relatively infrequently compared to papillary cancer and accounts for approximately 15% of all thyroid cancers. There is an increased frequency of follicular to papillary carcinoma (5:1) in iodine-deficient endemic goitre areas. It tends to be a malignancy of older persons, with a peak incidence of between ages 40 and 60 years and is approximately three times more common in women than in men. Medullary thyroid cancer (MTC) occurs in both sporadic and hereditary forms. The highest incidence of sporadic disease occurs in the fifth decade. Hereditary MTC can be inherited as an autosomal dominant trait with a high degree of penetrance associated with multiple endocrine neoplasia (MEN) type 2 syndrome or as familial MTC without any other endocrinopathies. It can be diagnosed before clinical presentation by genetic and biochemical screening. Anaplastic thyroid cancer is very rare and is more frequent in populations with endemic goitre. Thyroid lymphoma is also uncommon, constituting approximately 2% of extranodal lymphomas and occurring predominantly in older women. Up to one-third of patients have a history of goitre, whereas some have established autoimmune thyroiditis and may be taking L-T_4 therapy [29].

Screening for Thyroid Disorders

In the 1970s, screening programmes for congenital hypothyroidism were developed in which TSH was measured in heel-prick blood specimens to detect this condition as early as possible. The value of screening for congenital hypothyroidism is unquestioned, but only done routinely in approximately 30% of the world's birth population [20].

Certain groups within the adult population who should have an assessment of thyroid function at least once to detect thyroid dysfunction include those with a goitre or thyroid nodule, atrial fibrillation, dyslipidaemia, subfertility, osteoporosis, and assessing thyroid function in the annual review of people with diabetes mellitus appears cost-effective. There is no consensus on whether healthy pregnant women should be screened for thyroid disorders or PPT although it has been shown to be cost-effective in analytical models [32]. Women with type 1 diabetes are three times more likely to develop postpartum thyroid dysfunction so should be tested in the first trimester for thyroid antibodies.

Any woman with a history of PPT should be offered annual surveillance of thyroid function due to the long-term risk of permanent hypothyroidism. Because of the high prevalence of hypothyroidism in people with Down's syndrome and Turner's syndrome, an annual check of thyroid function is recommended. Thyroid function tests are indicated every six months for those receiving amiodarone, lithium, within three months of initiation of immune checkpoint inhibitor therapy and every 12 months following head and neck irradiation. All patients with hyperthyroidism who receive ablative treatment should be followed indefinitely for the development of hypothyroidism beginning 4–8 weeks after treatment, and then at three-month intervals for one year and annually thereafter. Among patients hospitalized for acute illness, testing should be limited but with a high index of clinical suspicion and with an awareness of the difficulties in interpreting thyroid function tests in the presence of acute illness.

Whether healthy adults living in an area of iodine sufficiency benefit from screening for thyroid disease remains controversial. The benefit from a screening programme must outweigh the physical and psychological harm caused by the test, diagnostic procedures, and treatment [33, 34]. The prevalence of unsuspected overt thyroid disease is low, but a substantial proportion tested will have evidence of thyroid dysfunction, with approximately 10% with subclinical hypothyroidism and 1% with subclinical hyperthyroidism. No appropriately powered prospective, randomized, controlled, double-blinded interventional trial of either L-T_4 therapy for subclinical hypothyroidism or antithyroid therapy for subclinical hyperthyroidism exists [13].

Epidemiological studies have shown an association between subclinical hypothyroidism and coronary heart disease in younger people (less than 65 years) and those with high serum TSH (more

than 10 mU/L) [35]. In older people, a higher serum TSH and lower free T4 concentrations within the euthyroid range are associated with lower risk of multiple adverse events including mortality [36]. Treatment in those who are symptomatic, pregnant or preconception, or aged less than 65 years appears justified [37]. Endogenous subclinical hyperthyroidism is associated with increased risk of total, coronary heart disease mortality and incident atrial fibrillation, with the highest risk being with serum TSH less than 0.10 mU/L [38]. Subclinical hyperthyroidism may be associated with an increased risk for hip and non-spine fractures [39]. Treatment may be indicated in those older than 65 years with serum TSH less than 0.1 mU/L to potentially avoid serious cardiovascular events, fractures, and the risk of progression to overt hyperthyroidism [40]. Any potential benefits of therapy in subclinical hyperthyroidism must be weighed against the morbidity associated with the treatment of hyperthyroidism.

From the available evidence, the following recommendations are suggested for an iodine-replete community:

- Screening for thyroid dysfunction in women younger than age 50 and in men is not warranted in view of the relatively low point prevalence of unsuspected overt thyroid dysfunction.
- Case-finding in women during menopause or during visits to a primary care physician with non-specific symptoms is justified due to the high prevalence of subclinical hypothyroidism.
- If increased serum TSH is found at screening, then measurement should be repeated two months later together with free T_4 measurement after excluding non-thyroidal illness, drugs, and so on.
- Treatment with L-T_4 is recommended if the serum TSH is greater than 10 mU/L, irrespective of whether free T_4 is low.
- Subjects with a serum TSH between 5 and 10 mU/L and normal free T_4 are at increased risk of developing hypothyroidism, and repeat measurement of serum TSH is warranted at least every three years if not annually.
- If a suppressed serum TSH is found at screening, it should be re-measured two months later, and if it is still suppressed, free T_3 should be measured.
- After L-T_4 replacement is initiated, for whatever indication, long-term follow-up with at least an annual measurement of serum TSH is required.

REFERENCES

1. Vanderpump MPJ. The epidemiology of thyroid disease. *Br Med Bull*, 2011; **99**: 39–51.
2. Taylor PN, Albrecht D, Scholz A, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*, 2018; **14**: 301–16.
3. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*, 2015; **3**: 286–95.
4. WHO Global Database on iodine deficiency. Geneva: World Health Organization. Available at: http://www.who.int/vmnis/database/iodine/iodine_data_status_summary/en/index.html (accessed 12 May 2018).
5. Vanderpump MPJ. Epidemiology of iodine deficiency. *Minerva Medica*, 2017; **108**: 116–23.
6. Bougma K, Aboud FE, Harding KB, Marquis GS. Iodine and mental development of children 5 years old and under: a systematic review and meta-analysis. *Nutrients*, 2013; **5**: 1384–416.
7. Hollowell JG, Staehling NW, Flanders WD, *et al*. Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
8. Laurberg P, Cerqueira C, Ovesen L, *et al*. Iodine intake as a determinant of thyroid disorders in populations. *Best Pract Res Clin Endocrinol Metab*, 2010; **24**: 13–27.
9. McGrogan A, Seaman HE, Wright JW, de Vries CS. The incidence of autoimmune thyroid disease: a systematic review of the literature. *Clin Endocrinol*, 2008; **69**: 687–96.
10. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet*, 2016; **388**: 906–18.
11. Tunbridge WMG, Evered DC, Hall R, *et al*. The spectrum of thyroid disease in the community: the Whickham survey. *Clin Endocrinol*, 1977; **7**: 481–93.
12. Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado thyroid disease prevalence study. *Arch Intern Med*, 2000; **160**: 526–34.
13. Biondi B, Cooper DC. Subclinical thyroid disease. *Lancet*, 2012; **379**: 1142–54.
14. Schouten BJ, Brownlie BE, Frampton CM, *et al*. Subclinical thyrotoxicosis in an outpatient population—predictors of outcome. *Clin Endocrinol*, 2011; **74**: 257–61.
15. Vadiveloo T, Donnan PT, Cochrane L, Leese GP. The Thyroid Epidemiology, Audit, and Research Study (TEARS): the natural history of endogenous subclinical hyperthyroidism. *J Clin Endocrinol Metab*, 2011; **96**: 59–61.
16. Vanderpump MPJ, Tunbridge WMG, French JM *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol*, 1995; **43**: 55–69.
17. Flynn RV, MacDonald TM, Morris AD, *et al*. The Thyroid Epidemiology, Audit, and Research Study; thyroid dysfunction in the general population. *J Clin Endocrinol Metab*, 2004; **89**, 3879–84.
18. Leese GP, Flynn RV, Jung RT, *et al*. Increasing prevalence and incidence of thyroid disease in Tayside, Scotland: the Thyroid Epidemiology, Audit, and Research Study (TEARS). *Clin Endocrinol*, 2008; **68**: 311–16.
19. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet*, 2017; **390**: 1550–62.
20. Ford G, LaFranchi SH. Screening for congenital hypothyroidism: a worldwide view of strategies. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 175–87.
21. Cherella CE, Wassner AJ. Congenital hypothyroidism: insights into pathogenesis and treatment. *Int J Pediatr Endocrinol*, 2017; **2017**: 11.
22. Wartofsky L, Dickey RA. The evidence for a narrower thyrotropin reference range is compelling. *J Clin Endocrinol Metab*, 2005; **90**: 5483–8.
23. Surks MI, Goswami G, Daniels GH. The thyrotropin reference range should remain unchanged. *J Clin Endocrinol Metab*, 2005; **90**: 5489–96.
24. Surks MI, Boucai L. Age- and race-based serum thyrotropin reference limits. *J Clin Endocrinol Metab*, 2010; **95**: 496–502.
25. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.

26. De Leo S, Pearce EN. Autoimmune thyroid disease during pregnancy. *Lancet Diabetes Endocrinol*, 2018; 6: 575–86.
27. Russ G, Leboulleux S, Leenhardt L, *et al*. Thyroid incidentalomas: epidemiology, risk stratification with ultrasound and workup. *Eur Thyroid J*, 2014; **3**: 154–63.
28. Perros P, Colley S, Boelaert K, *et al*. Guidelines for the management of thyroid cancer. *Clin Endocrinol*, 2014; **81**: Suppl 1, 1–122.
29. Haugen BR, Alexander EK, Bible KC, *et al*. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 2016; **26**: 1–133.
30. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the United States, 1974–2013. *JAMA*, 2017; **317**: 1338–48.
31. Tuttle RM, Vaisman F, Tronko MD. Clinical presentation and clinical outcomes in Chernobyl-related paediatric thyroid cancers: what do we know now? What can we expect in the future? *Clin Oncol*, 2011; **23**: 268–75.
32. Dosiou C, Barnes J, Schwartz A, *et al*. Cost-effectiveness of universal and risk-based screening for autoimmune thyroid disease in pregnant women. *J Clin Endocrinol Metab*, 2012; **97**: 1536–46.
33. Tunbridge WMG, Vanderpump MPJ. Population screening for autoimmune thyroid disease. *Endocrinol Metab Clin North Am*, 2000; **29**: 239–52.
34. Rugge JB, Bougatsos C, Chou R. Screening and treatment of thyroid dysfunction: an evidence review for the U.S. Preventive services task force. *Ann Intern Med*, 2015; **162**: 35–45.
35. Rodondi N, den Elzen WPJ, Bauer DC, *et al*. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA*, 2010; **304**: 1365–74.
36. Cappola AR, Arnold AM, Wulczyn K, *et al*. Thyroid function in the euthyroid range and adverse outcomes in older adults. *J Clin Endocrinol Metab*, 2015; **100**: 1088–96.
37. Pearce SH, Brabant G, Duntas LH, *et al*. 2013 ETA Guideline: management of subclinical hypothyroidism. *Eur Thyroid J*, 2013; **2**: 215–28.
38. Collet TH, Gussekloo J, Bauer DC, *et al*. Subclinical hyperthyroidism and the risk of coronary heart disease and mortality. *Arch Intern Med*, 2012; **172**: 799–809.
39. Wirth CD, Blum MR, da Costa BR, *et al*. Subclinical thyroid dysfunction and the risk for fractures: a systematic review and meta-analysis. *Ann Intern Med*, 2014; **161**: 189–99.
40. Biondi B, Bartalena L, Cooper DS, *et al*. The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. *Eur Thyroid J*, 2015; **4**: 149–63.

3.2

Aetiology of Thyroid Disorders

3.2.1 The Complex Genetics of Thyroid Disease

Terry F. Davies, Francesca Menconi, and Yaron Tomer

Introduction to Genetics

Genes and Chromosomes

The nucleus of each human cell encodes approximately 30 000 genes. More than 20% of the genes in each individual exist in a form that can vary between individuals. These variable genetic forms are termed polymorphisms among normal individuals, and they account for much of the normal variation in body traits, such as height and hair colour. The genetic information encoded in the DNA is stored on the chromosomes and each somatic cell contains 46 chromosomes (22 autosomes and two sex chromosomes), arranged in 23 pairs, one of each derived from each parent. Since each individual has two copies of each chromosome (for **autosomes**), one from each parent, there are also two copies of each gene; one of which is mostly suppressed (inactivated). The chromosomal location of a gene is termed the **locus** of the gene. The gene in a certain locus usually exists in two forms and these variants of the gene are termed **alleles**. When an individual's two alleles at a locus are identical, that individual is said to be a **homozygous** at that locus, and when the two alleles are different, the individual is a **heterozygous**.

Female somatic cells contain two X chromosomes, whereas male somatic cells contain only one X chromosome. Nevertheless, the activity of genes coded for by the X chromosome is not higher in females than in males. This is due to inactivation of most of the genes on one of the two X chromosomes. Thus, in female somatic cells only one X chromosome gene is usually expressed, and this process of suppression is called X-chromosome inactivation. X-chromosome inactivation occurs early in embryonic life and, thereafter, in each cell either the maternal or paternal chromosome is inactivated in what is often a random mechanism. This results in a tissue mosaic of paternally and maternally expressed X-chromosomal alleles, with an average of 1:1 distribution. As a result, a female who is heterozygous for an X-linked gene will show a mosaic-like distribution of cells expressing either one of the two alleles. Recently X-inactivation has been postulated to play a role in autoimmune diseases and may help explain their female preponderance (see next).

Inheritance

When genetic information is transmitted from parent to offspring the process is called inheritance. Germ cells undergo **meiosis**, a process in which 2 gametes with 23 chromosomes are generated. During meiosis paired chromosomes undergo recombination, a process in which paired chromosomes break at identical points along their length and switch genetic material on opposite sides of the breaking point. Recombination results in an exchange of matching segments of the chromosome between two homologous chromosomes (Figure 3.2.1.1). Since recombination is also a random event, the farther apart two genes are on the same chromosome, the greater the likelihood that a recombination will occur in the space between them. When two genes A and B are very far apart they are transmitted to the offspring independently, as if they were located on different chromosomes (i.e. the probability that a given parental allele of gene A will be transmitted to the offspring with the same parental allele of gene B is 50%, as would be expected to occur at random). On the other hand, if genes A and B are located *close* to each other, the probability that a given parental allele of gene A will be transmitted to the offspring with the same parents allele of gene B is greater than 50%, and the genes are said to be linked (Figure 3.2.1.1). In other words, if two gene loci are linked there is a greater than 50% probability that offspring will inherit the same combination of alleles that are present on the parental chromosome. If the genes were right next to each other then they would have a very high chance of being inherited together. This phenomenon is the basis for the concept of linkage analysis (see next).

Understanding Mutations

A mutation is an alteration in chromosomal DNA that is inherited from one generation to another. Mutations in genes may be silent or if located appropriately can change the structure and/or function of the encoded protein causing differences in enzyme function, receptor signalling, structural proteins, or regulatory proteins. When several mutations in two or more genes produce similar changes then **genetic heterogeneity** is said to exist. For example, in maturity onset diabetes of the young (MODY), similar clinical presentations (**phenotypes**) can be caused by mutations in the glucokinase gene or the hepatic nuclear factor 1∀ gene. However, in some diseases an individual may inherit the mutation but will not develop the disease phenotype. Such diseases are said to have reduced **penetrance**. The penetrance of the disease is defined as the probability that an individual inheriting the mutation will actually develop the disease phenotype. An example of a disease with reduced penetrance is multiple endocrine neoplasia type II (MEN 2) where not every individual inheriting the mutation in the RET proto-oncogene will develop the full phenotype of the disease.

Categories of Genetic Diseases

Genetic diseases can be caused by a variety of changes in our genetic code and for simplicity can be divided into broad categories:

Chromosomal disorders: In these disorders the number of chromosomes in an individual (their **karyotype**), or their DNA structure, is altered producing excessive or deficient genetic material. Normally every individual has 22 pairs of autosomes and one pair of sex chromosomes (XX in females and XY in males). In chromosomal disorders the number of chromosomes may be different (e.g. in Down's syndrome there are three copies of chromosome 21), or large segments of chromosomes may be deleted or exchanged with other chromosomes causing a variety of different disorders and syndromes. One of the most important of these changes is termed copy number variants (CNVs), and they consist of deletions or duplications of DNA segments with a size of 1000 bases to several million bases.

Mendelian disorders: These disorders are caused by a mutation in a single gene. They display specific patterns of inheritance that can be classified as **dominant** (inheritance of one mutant gene, from either parent, will cause disease), **recessive** (only inheritance of both mutant genes, one from each parent will cause disease), or X-linked (see next).

Complex diseases: These disorders are caused by interactions between multiple genes. These interactions may presumably be deficient or excessive or unusual and involve most of the common diseases we know so well including autoimmune thyroid disease. How multiple gene involvement may produce such similar clinical phenotypes is not well understood.

Epigenetics and disease: One explanation to help understand complex genetics has been our increased understanding of epigenetic and environmental factors. Gene expression and suppression is not only controlled by mRNA and its cofactors but by how DNAs folds thus opening and closing genes to the polymerases. Such acetylation and methylation events on the surface of the DNA (epigenetics) may also be influenced by the environment and contribute to the activity of multiple genes impacting the clinical phenotype.

Inheritance of Mendelian Disorders

Dominant Mendelian disorders: These disorders manifest in heterozygotes (i.e. when one mutant allele is present but the second allele of the same gene is normal then the mutant is expressed). The mutant gene in this case is on one of the 22 autosomes. Examples of dominant Mendelian disorders include thyroid hormone resistance syndrome, and MEN 2.

Recessive Mendelian disorders: These disorders are clinically apparent only in homozygotes (i.e. when both alleles at a particular genetic locus are mutated so there is no normal gene at all). The mutation in these disorders is also on one of the 22 autosomes. Examples of recessive Mendelian disorders include familial non-autoimmune hypothyroidism and Pendred's syndrome.

X-linked Mendelian disorders: These disorders occur when the mutant gene is on the X chromosome. Most X-linked Mendelian disorders are recessive rather than dominant. Therefore, females can be affected only if they inherit two mutant genes on both of their X chromosomes, while males are affected if they inherit only one mutant gene on their X chromosome. An endocrine example of an X-linked Mendelian disorder is familial thyroxine binding globulin (TBG) deficiency.

Inheritance of Complex Diseases

As we discussed earlier, the inheritance of complex diseases, such as GD and Hashimoto's thyroiditis (HT), does not follow a simple Mendelian pattern. These diseases are caused, in part, by multiple genes with additive and negative effects. Moreover, the penetrance of complex diseases tends to be low (i.e. not all the individuals inheriting the mutations or polymorphisms will actually develop the clinical phenotype). This causes a non-Mendelian pattern of transmission of the disease in pedigrees and makes mapping their susceptibility genes much more challenging than mapping Mendelian disorder genes. But, of course, all individual genes are inherited in a Mendelian pattern.

Chromosomal and Mendelian Disorders Affecting the Thyroid

Chromosomal Disorders Associated with Thyroid Dysfunction

Several chromosomal disorders are known to be associated with an increased incidence of thyroid disease (Table 3.2.1.1). The association between Down's syndrome and autoimmune thyroid diseases (AITD) is especially intriguing because of the possibility that a gene conferring susceptibility to autoimmune thyroid diseases is located on chromosome 21. Although this has been investigated extensively there is still no evidence of an AITD susceptibility gene on chromosome 21 (see next). Interestingly, on chromosome 21 is the autoimmune regulator (AIRE) gene and, when it is mutated it causes autoimmune polyglandular syndrome type I.

Mendelian Disorders Involving the Thyroid

Many Mendelian disorders of thyroid hormonogenesis and regulation have been described and many of the mutations causing these disorders have been revealed. Table 3.2.1.2 summarizes the major Mendelian disorders affecting the thyroid, their pathophysiology,

Table 3.2.1.2 Mendelian disorders of the thyroid*

Gene	Locus/Chr	Disease	MOI	Gene mutation and pathogenesis	Clinical signs
TRH	3	Isolated TRH deficiency	AR	TRH deficiency leads to low TSH that increases with TRH administration	Central hypothyroidism
TSH-β chain	1p22	Isolated TSH deficiency	AR	Mutations in the TSHβ gene that lead to a mutated TSHβ protein that cannot associate with the α-subunit to produce a functional TSH heterodimer	Central hypothyroidism
Pit-1	3p11	Combined pituitary GH, PRL, TSH deficiency	AD/AR	Mutations in the transcription activating factor Pit-1 that regulates expression of GH, PRL, TSH, and the development of somatotrophs, lactotrophs, and thyrotrophs	Central hypothyroidism and GH and PRL deficiencies
TSH Receptor	14q31	Familial hypothyroidism/ TSH Resistance	AR	Inactivating mutations in the extracellular or transmembrane domains of the TSHR making it unresponsive to TSH	Congenital hypothyroidism, hypoplastic thyroid, or euthyroidism with high TSH
		Familial non-autoimmune hyperthyroidism	AD	Activating mutations in the transmembrane domain of the TSHR leading to constitutive activation of the TSHR	Hyperthyroidism, goitre, no signs of autoimmunity
		Familial gestational hyperthyroidism	AD	Mutation in the extracellular domain of the TSHR leading to hypersensitivity of the TSHR to chorionic gonadotropin	Gestational hyperthyroidism and hyperemesis gravidarum
TTF1	14q12-13	Congenital hypothyroidism	AD	Mutations (missense, nonsense) cause an alteration of the DNA-binding domain resulting in loss of functional activity and reduction in the production of TTF1 levels in heterozygotes (haploinsufficiency). The mechanism responsible for elevated TSH with normal thyroid gland, in several cases, is still unclear	Elevated TSH, normal or low T_4, neurological abnormalities, respiratory distress
TTF-2	9q22	Congenital hypothyroidism	AR	Missense mutations of TTF-2 lead to a protein with impaired DNA binding and total (homozygotes) or partial (heterozygotes) loss of transcriptional function	Congenital hypothyroidism with thyroid agenesis, cleft palate, spiky hair
PAX-8	2q12-q14	Congenital hypothyroidism	AD	Mutations cause a marked reduced DNA binding capacity with loss of transcriptional activation function	Congenital hypothyroidism, hypoplastic, and sometimes ectopic thyroid, renal abnormalities
Na/I Symporter	19p12	Congenital hypothyroidism	AR	Inactivating mutations of the Na/I symporter leading to defective or absent iodine uptake by thyroid cells	Congenital hypothyroidism, goitre, defective iodine uptake in the thyroid
TPO	2p25	Congenital hypothyroidism	AR	Inactivating mutations leading to inactive TPO, or to disturbed integration of TPO in the membrane thus causing defective or absent organification of iodide	Congenital hypothyroidism, goitre, abnormal perchlorate discharge test, sometimes intellectual disability
Pendrin	7q22	Pendred's syndrome	AR	Inactivating mutations in the Pendrin gene, which is a sulphate transporter, cause disruption of iodide transport from thyroid follicular cells to the follicular lumen	Goitre, congenital sensorineural deafness
Tg	8q24	Congenital Tg deficiency	AR	Quantitative abnormalities in Tg, mutations in the Tg gene, defects in glycosylation, or transport of Tg cause impaired coupling of iodotyrosines	Goitre, hypothyroidism, or euthyroidism, low or absent serum Tg, no colloid in the thyroid gland
DUOX2 (THOX2)	15q15.3	Congenital hypothyroidism	AD	Mutations (nonsense, missense) in the *DUOX2* gene result in insufficient or absent production of hydrogen peroxide, needed for TPO action	Permanent/transient congenital hypothyroidism, complete/partial iodide organification defect
DUOXA2	15q15.1	Congenital hypothyroidism	AR	Nonsense mutation of *DUOXA2* result in complete loss of function of the protein. Since *DUOXA2* is essential for DUOX2 activity, it leads to a secondary deficit of *DUOX2* (see above)	Congenital hypothyroidism, goitre, abnormal perchlorate discharge test

(*continued*)

Table 3.2.1.2 *Continued*

Gene	Locus/Chr	Disease	MOI	Gene mutation and pathogenesis	Clinical signs
DEHAL1	6q25	Goitrous hypothyroidism	AR	Missense mutations or deletion in *DEHAL1* lead to a protein with reduced capacity to deiodinate monoiodotyrosine and diiodotyrosine	Hypothyroidism in infanthood or childhood, goitre, elevated serum diiodotyrosine. Mental and psychomotor retardation develop if hypothyroidism is not treated
SECISBP2	9q22.2	Reduced deiodinase activity	AR	Mutations in SECISBP2 gene affect the synthesis of selenoproteins and lead to a reduction in deiodinase 2 activity	Short stature and delayed bone age. High T_4, low T_3, high rT_3, slightly elevated TSH
TBG	Xq23	Congenital TBG deficiency	X-linked	Inactivating (deletions, missense, non-sense) mutations in the TBG gene leading to partial or complete deficiency of TBG	Decreased total T_4, normal free T_4, euthyroidism
		Inherited TBG excess	?	Excess TBG of unknown cause results in elevated levels of TBG and total T_4	Increased total T_4, normal free T_4, euthyroidism
Transthyretin	18	Familial euthyroid hyperthyroxinaemia due to TTR abnormalities	AD	Point mutations in the *TTR* gene cause increased affinity for T_4 and T_3	High total T_4, normal free T_4, euthyroidism, Familial amyloidotic polyneuropathy
Albumin	4q11	Familial dysalbuminaemic hyperthyroxinaemia	AD	Point mutations in the albumin gene lead to increased affinity for T_4 and T_3	High total T_4, normal free T_4, euthyroidism
MCT-8	Xq13.2	Allan–Herndon–Dudley syndrome	X-linked	Mutations (truncating, in-frame deletion, missense) in the *MCT-8* gene alter the intracellular availability of thyroid hormones	Affected males present abnormal thyroid function tests (increased T_3 and decreased T_4 and rT_3) and severe psychomotor and developmental delay. Females carriers have only mild thyroid functions test abnormalities
TRβ	3	Resistance to thyroid hormone	AD	Mutations in the TRβ gene lead to a thyroid hormone receptor with reduced affinity for T_3 or abnormal interaction with a cofactor necessary for T_3 action	Goitre, tachycardia, hyperactivity, developmental delay. Elevated serum T_3 and T_4 with non-suppressed TSH. Serum Tg is often elevated

* Abbreviations used in the table: AD, autosomal dominant; AR, autosomal recessive; Chr, chromosome; DEHAL1, iodotyrosine deiodinase; DUOX2, dual oxidase 2; DUOXA2, dual oxidase maturation factor 2;GH, growth hormone; MOI, mode of inheritance; MCT-8, monocarboxylate transporter 8; Na/I symporter, sodium iodide symporter; PAX-8, paired box 8;PRL, prolactin; SECISBP2, selenocysteine insertion sequence-binding protein 2; TBG, thyroxine binding globulin; Tg, thyroglobulin; THOX2, thyroid oxidase 2; TPO, thyroid peroxidase; TRβ, thyroid hormone receptor β;TRH, thyrotropin releasing hormone; TTF-1, thyroid transcription factor-1; TTF2, thyroid transcription factor-2; TTR, transthyretin; TSH, thyroid stimulating hormone (thyrotropin).

clinical characteristics, and their mode of inheritance. These disorders are described in more detail elsewhere in this book.

Genetic Susceptibility to Autoimmune Thyroid Disease (AITD)

The AITDs are examples of complex genetic diseases affecting the thyroid. Classically, the AITDs encompass a spectrum of related disorders varying from hyperthyroid Graves' disease (GD) to hypothyroid Hashimoto's thyroiditis (HT) and which are characterized by abnormal autoimmune responses to thyroid antigens. Additional variants of AITD include postpartum thyroiditis (reviewed in [1]), drug induced thyroiditis, such as interferon induced thyroiditis (IIT) [2], thyroiditis associated with polyglandular autoimmune syndromes (reviewed in [3]) and just the presence of thyroid antibodies (TAb) with no apparent clinical disease [4] but which simply indicates the presence of an intrathyroidal infiltrate but retention of normal thyroid function. The AITD are among the commonest human autoimmune disorders affecting more than 5% of the general population [5, 6]. The AITD, including Graves' and Hashimoto's diseases, are categorized as complex diseases because they are believed to be caused by an interaction between several genes and a variety of environmental factors. In recent years, sound epidemiologic evidence for a genetic susceptibility to AITD has been established. Box 3.2.1.1 summarizes these main epidemiological data pointing to a genetic susceptibility to AITD.

Geographic and Longitudinal Trends in the Incidence of the AITD

The annual incidence of GD in populations of different geographic locations, excluding extremes of iodine intake, is similar and ranges from 0.22 to 0.27 per 1000 [5, 6]. A similar pattern has been observed for HT. In the UK Whickham survey the prevalence of spontaneous hypothyroidism was 15/1000 in females compared with less than 1/1000 in males [6]. The mean annual incidence of spontaneous hypothyroidism in women was 3.5/1000 and in men was 0.6/1000 [7]. Similar prevalence and incidence data of spontaneous hypothyroidism have been reported by others in the United States, Finland, Japan, and Sweden. The comparable prevalence and incidence of the AITD in geographically different populations has suggested that the genetic contribution to the development of the AITD is more important than the environmental contribution because these populations are exposed to different environmental factors.

A longitudinal study from the Mayo clinic (1935–1967) showed no significant change in the incidence of GD over the 33 years of the study [8]. The stable incidence of GD with time once again points to strong genetic effects because the genetic makeup of a population did not appear to change over several decades, but environmental factors most probably did. The Mayo clinic observations were supported by a more recent study from Sweden [9]. However, the Swedish study found an increased incidence of GD in a subset of the population, demonstrating that environmental effects may still play a crucial role in the aetiology of GD. In the Mayo survey (1935–1967) there was a significant increase in the incidence of HT over the 33 years of the survey [8]. This could reflect a stronger environmental influence on the development of HT rather than GD or more likely a change in the diagnostic criteria over time [10].

Box 3.2.1.1 Epidemiological evidence for a genetic susceptibility to Graves' disease and Hashimoto's thyroiditis (see text)

- I Secular trends in the incidence of AITD:
  - 1) The incidence of GD/HT is similar in different ethnic populations
  - 2) The incidence of GD has not changed over time in the past several decades
- II Variations in the incidence of AITD with age:
  - 1) The incidence of GD/HT peaks in the fifth decade of life
  - 2) After peaking the incidence of GD/HT declines to zero (suggesting that all genetically susceptible individuals have developed the disease)
- III Familial clustering of AITD:
  - 1) AITD develops in 20–30% of siblings of patients with AITD
  - 2) The sibling risk ratio (8_s) for AITD is 16.9
  - 3) Thyroid antibodies are found in up to 50% of siblings of patients with AITD
- IV Twin studies:
  - 1) Concordance rate in MZ twins for GD is 30–35% and in DZ twins it is 3–5%
  - 2) Concordance rate in MZ twins for GD is 55% and in DZ twins it is 0%

Familial Clustering

The familial occurrence of the AITD has been recognized by investigators for many years. More than 50 years ago it was recognized that a familial predisposition can be found in approximately 50% of cases with GD. Later studies have shown a high frequency of thyroid abnormalities in relatives of patients with AITD [11], most commonly the presence of thyroid autoantibodies which were reported in up to 50% of the siblings of patients with GD [12]; 36% of GD patients with ophthalmopathy had a family history of AITD while 32% had a first-degree relative with AITD [13]. Moreover, a study from Holland in which 790 healthy female relatives of AITD patients were followed for up to 5 years, showed that 7.5% of them developed overt hypothyroidism or hyperthyroidism [14].

Such familial clustering of a disease can be due to non-genetic factors including the shared environmental exposures (e.g. infections, diet). Therefore, methods have been developed to determine whether familial clustering of a disease is the result of genetic susceptibility or non-genetic factors. One method is to calculate the sibling risk ratio (λs) which expresses the increased risk of developing the disease in an individual who has a sibling with the disease compared to the risk in the general population, and has been claimed as a quantitative measure of the genetic contribution to the disease. A λs of > 5 has been used to indicate a significant genetic contribution to the pathogenesis of a disease [15]. We calculated the λs in AITD in a cohort of 155 AITD patients. The λs was 16.9 for AITD, 11.6 for GD, and 28.0 for HT. These high λs values indicate a strong familial influence on the development of AITD.

Twins Studies

Twin studies can provide information concerning the inheritance of a disease and may yield certain quantitative evaluations on the role of heredity in relation to exogenous factors. Twin analysis is based upon comparison of the concordance (simultaneous occurrence) of a disease among monozygotic (MZ) twins vs. dizygotic (DZ)

twins. MZ twins have similar genetic makeup, whereas DZ twins share an average of 50% of their genes (like siblings). Therefore, if concordance is higher in the MZ twins when compared to the DZ twins it suggests that the disease has an inherited component. Any discordance among the MZ twin pairs is usually interpreted to mean that the gene or genes concerned show reduced penetrance (i.e. certain epigenetic or environmental factors must be present before the disease becomes manifest). However, this logic has its limits for autoimmune diseases. This is because MZ twins are not immunologically identical. Since the T-cell receptor and immunoglobulin V genes undergo random recombination, twins have different immune repertoires although they may be sculpted by the same environment.

Nevertheless, the concordance rate in MZ twins is taken as an estimate of the penetrance of the disease but this is only up to the age they are examined. For example, if the concordance rate among the MZ twins is 50%, this is taken to mean that the penetrance of the disease genes is approximately 50%. As discussed, however, it must be emphasized again that MZ twins are not identical and their individual immune experiences will influence their immune repertoire. Therefore, part of the observed discordance between MZ twins may also be due to the discordance in their immune repertoire.

The concordance rate for GD in MZ twins was found to be approximately 30–35% while the concordance rate in DZ twins was reported to be about 3–5% [16–18]. These data indicated that there is a substantial inherited susceptibility to Graves' disease, presumably related to both immune and non-immune genes. For HT the concordance rates were 55% and 0% for MZ and DZ twins, respectively [19], again pointing to a strong genetic component in the aetiology of the disease. Finally, for thyroid antibodies, MZ twins had 80% concordance and DZ twins had 40% concordance [19]. Thus, the twin data support a substantial inherited susceptibility to AITD. Indeed, a recent study estimated that 80% of the liability to develop GD was due to genetic factors [17]. A similar strong genetic contribution was estimated for the production of thyroid antibodies.

Thyroid Autoantibodies

Autoantibodies to thyroglobulin (Tg) and thyroid peroxidase ([TPO] the microsomal antigen) have been widely used to show the population at most risk for the development of AITD. An increased prevalence of thyroid autoantibodies has been reported in relatives of patients with AITD [20–30]. Antithyroid autoantibodies (antithyroglobulin and antithyroid peroxidase) have been found in up to 50% of the siblings of patients with the AITD [12], in contrast to a prevalence of 7–20% in the general population [6]. These findings are true in different populations such as the Japanese [20] and British populations [11]. In one study, it was found that thyroid antibodies were almost always present in one of the parents of an affected individual with AITD [11]. These data suggested an inherited influence on the production of antithyroid antibodies compatible with dominant inheritance. Indeed, segregation analyses in a panel of families with thyroid antibodies also suggested a Mendelian dominant pattern of inheritance for the tendency to develop antithyroid antibodies [21]. In keeping with these observations, it was reported that recognition of particular thyroid peroxidase (TPO) epitopes within the autoantibody immunodominant region may be transmitted within families [22].

Graves' Orbitopathy

Depending on the definition of Graves' ophthalmopathy (GO), it would appear to affect about 90% of patients with GD and so is an intrinsic part of the disorder although rarely it may be found without thyroid involvement and also is seen regularly in patients with HT who may or may not have experienced GD. However, the severe form of ophthalmopathy occurs in less than 10% of GD patients. Severe GO manifests by proptosis, conjunctival injection, eye muscle weakness to paralysis, and sometimes optic nerve damage. GO is considered pathognomonic of GD even when the individual is not thyrotoxic. It was speculated that the genetic influence on the development of Graves ophthalmopathy could be more pronounced because GO represents the most severe form of the disease. We, therefore performed a segregation analysis in patients selected for severe GO by studying the first-degree relatives of individuals with severe GO. If the disease is hereditary the first-degree relatives would be expected to be affected by the disease more often than individuals randomly selected from the general population. Our segregation analysis showed that GO did not have a major genetic component [13]. Although a genetic contribution to the susceptibility to develop GO has been claimed [23] such a non-specific gene association could simply enhance susceptibility to more severe GD.

Tools to Map Complex Disease Genes

Based on the abundant epidemiologic evidence for a strong genetic effect on the development of AITD, many searches for the susceptibility genes have been performed. The basic strategies used for finding complex disease genes include association and linkage studies of candidate genes, and whole genome screening. These tools have proven successful in the mapping of many novel complex disease genes such as rheumatoid arthritis but often find genes which provide very small contributions to susceptibility.

Association and Linkage

Association analyses: Association analyses are sensitive tests which can locate even minor susceptibility genes. Population-based association tests compare **marker allele** frequencies (such as an HLA haplotype) in unrelated patients to a set of unrelated, carefully matched, controls. Identification of a significant difference between patients and controls suggests that a genetic locus at, or near, the marker locus influences disease predisposition due to linkage disequilibrium. Linkage disequilibrium exists when chromosomes with the mutant allele at the disease locus carry certain marker alleles more often than expected by random chance. It cannot be emphasized enough that the control population must be appropriate since everything hinges on simple comparisons. Furthermore, the larger the control group the less chance of getting this matching wrong.

Association analysis is very sensitive and may detect genes contributing less than 5% of the total genetic contribution to a disease. Classically, association studies were used for studying candidate genes, and for fine mapping linked loci. However, association studies are also utilized to screen the entire human genome (see next).

As discussed earlier, one of the weaknesses of population-based methods is that they can produce spurious associations if the patients and controls are not accurately matched (so-called **population**

stratification). Therefore, family-based association tests have been developed which use an internal control group from within each family, thus avoiding the necessity to match patients and controls for ethnicity. The most widely used family-based association test is the **transmission disequilibrium test** (TDT). The TDT is based on the comparison of the parental marker alleles which ***are*** transmitted and those which are ***not*** transmitted to affected children (Figure 3.2.1.2). Assuming two heterozygote parents for a certain tested marker, each parent has two sets of chromosomes so there are four parental alleles in each family. These are categorized into two groups: those transmitted to a child with the disease (T alleles), and those not transmitted to any affected child (N alleles). The same allele may belong to the T group or the N group in different families. The frequency of the T alleles versus the N alleles is then compared by a Π^2 test. An association between a certain allele and the disease exists if there is an excess occurrence of this allele in the T group compared to the N group.

Linkage: Genetic linkage techniques are powerful tools for analysing complex disease-related genes because they detect only genes that theoretically have a major effect (>5%) on the aetiology of a disease. The consequence is that linkage studies are less sensitive than association studies since they do not detect minor contributing genes. The principle of linkage analysis is based on the fact that if two genes are close together on a chromosome they will segregate together because the likelihood that a recombination will occur between them is low. Therefore, if a tested marker is close to a disease susceptibility gene, its alleles will cosegregate with the disease in families (Figure 3.2.1.3). The LOD (logarithm of odds) score is the statistical measure of the likelihood of linkage between a disease and a genetic marker. The LOD score is the base-10 logarithm of the odds ratio in favour of linkage. Arbitrarily, a LOD score of greater than 3 (i.e. odds ratio greater than 1000) is considered strong evidence for linkage and a LOD score greater than 2 is suggestive of linkage.

Linkage at a certain locus is established when at least two or more independent datasets give strong evidence for linkage at the same locus. Conversely, a LOD score lower than –2 can be used to exclude linkage. The classical linkage tests are model based (i.e. a mode of inheritance and penetrance have to be assumed when calculating the likelihood of linkage). However, in complex diseases the mode of inheritance is often unknown and, therefore, model-independent methods are sometimes used. One such method, which became very popular, is sib-pair analysis (Figure 3.2.1.4). Here siblings, which are both affected by the disease being studied, are tested for sharing of alleles at a marker locus. By random chance alone the sibs would be expected to share one allele of the marker 50% of the time. If affected sib-pairs share a significantly higher than expected proportion of alleles at the marker locus, this suggests that the marker locus is in linkage disequilibrium with the disease gene. This observed to expected allele sharing can be converted to a LOD score equivalent.

Candidate Gene Analysis

Candidate genes are genes of known sequence and location that by virtue of their known functions may be involved in disease pathogenesis. For example, one can hypothesize that the TSH receptor may be a candidate gene for GD because the hallmark of the disease is the presence of TSH receptor antibodies. If a candidate locus is indeed the cause of a disease, then markers in that locus should segregate with the disease within families giving high LOD scores. Since the basic abnormality in AITD is an immune response against thyroid antigens, possible candidate genes for AITD include genes that control immune responses (e.g. the major histocompatibility complex [MHC or HLA] genes, and costimulatory molecule genes), and genes encoding the target autoantigens in AITD (thyroglobulin, thyroid peroxidase, iodide transporter, TSH receptor). Many of these genes have now been studied for their possible role in the genetic susceptibility to the AITD (see next).

Whole Genome Screening

Another approach is to screen the whole human genome without any assumptions on disease pathogenesis. This method is called whole genome screening. Whole genome scans can be performed using linkage as well as association methods.

Linkage-based genome scans: The two requirements for performing a linkage-based whole genome screen in a complex disease are:

1. The availability of a sizeable and well validated data set of multiplex families (large families with more than one individual affected).
2. The availability of a map of highly polymorphic markers covering the whole genome.

Microsatellite markers: The first useful polymorphic markers for whole genome screening were discovered in 1989 and were called microsatellites. Microsatellites are regions in the genome that are composed of short sequence repeats, most commonly two-base CA-repeats (Figure 3.2.1.5). Microsatellite loci are highly polymorphic (i.e. have many alleles) because the number of repeats in each individual is variable. Moreover, they are extremely abundant and uniformly distributed throughout the genome at distances of less than 1 million base pairs. Therefore, microsatellites serve as excellent markers in whole genome linkage studies.

Single nucleotide polymorphisms (SNPs): SNPs are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals. In humans most SNPs have only two alleles. SNPs are very abundant with a frequency of about one SNP per <500 bp. Since SNPs have only two alleles they are less informative than microsatellites, and a larger number of SNPs is required to screen the human genome by linkage. However, since SNP's are much more abundant and closely spaced than microsatellites, they are ideal for fine mapping genes, in linked regions, using association studies and for association-based genome scans (see next). The importance of SNPs stems from the fact that many have the potential to change the amino acid sequence of a gene product, or other regulatory sequences (e.g. promoter) and be directly involved in the susceptibility to complex diseases because they cause changes in gene function. Thus, if a SNP allele inside a gene is found to be significantly associated with a disease it may be the actual causative allele, increasing susceptibility to the disease.

Using microsatellites or SNPs many linkage-based whole genome screens have been now completed for several complex diseases including the AITD (see next).

Genome-wide association studies: Genome-wide association studies have two major requirements:

1. The availability of a sizeable data set of affected individuals and a sizeable ethnically matched control group.
2. The availability of closely spaced polymorphic markers covering the whole genome.

For genome-wide screening by association analysis one would need to employ more than 300 000 closely spaced markers. The completion of the HapMap project made whole genome scanning by association studies feasible. The HapMap project genotyped more than one million SNPs spanning the entire human genome in four ethnically distinct human populations and tested these SNPs for linkage disequilibrium (LD). This project discovered that the human genome is highly organized into discrete **linkage disequilibrium blocks** (LD blocks) which are flanked by **recombination hot spots**, or areas at which recombinations most often occur. This enabled the utilization of tagged SNPs (each SNP representing an entire LD block) to test the entire human genome for association with disease without having to employ all SNPs in the genome. Moreover, microarray-based genotyping technology enabled the typing of up to 500 000 SNPs in a single experiment. Thus, it became possible to scan the entire human genome by association analysis using selected SNPs on a chip.

Next Generation Sequencing for Rare Variants

Association studies detect common variants with small effect size as shown by numerous GWAS studies. In contrast, linkage studies, which test cosegregation of variants with disease in families, are robust for detecting variants with large effect size as shown by our linkage studies in AITD [24]. However, variants with large effect size may be rare and can only be identified by sequencing. Rare variants are defined as having minor allele frequency of less than 1%. The recent developments in next generation sequencing have made it possible to identify rare variants either by sequencing linked loci which are usually 5–30 Mb in size or by sequencing the entire human genome. Sequencing the entire human genome is usually limited to exons (exome sequencing) but can also be performed for the entire genome including coding and non-coding sequences.

Genetics of Autoimmune Thyroid Disease

The Primary Role of HLA Genes

The MHC region, encoding the HLA glycoproteins, consists of a complex of genes located on chromosome 6p21 (Figure 3.2.1.6). The MHC region also encodes various additional proteins, most of which are associated with immune responsiveness. Since this gene region is highly polymorphic and contains many immune response genes it was the first candidate genetic region to be studied for association and linkage with AITD.

Association of HLA with GD: GD was initially found to be associated with HLA-B8 in Caucasians. This finding was then confirmed in a wide number of studies, mostly examining populations of Caucasian origin (Table 3.2.1.3). In these early studies HLA-B8 was associated with relative risks for GD ranging from 1.5 to 3.5. Subsequently, it was found that GD was more strongly associated with HLA-DR3, which is in LD with HLA-B8. The frequency of DR3 in GD patients was 40–55% and in the general population 20–30% giving a relative risk for people with HLA-DR3 of 2–3. Even though the frequency of HLA-DR3 was increased in Caucasians with GD, there were also HLA-DR3 negative associations with GD, and in addition, the HLA associations were found to be different in other ethnic groups (Table 3.2.1.4). In the Japanese population GD was associated with HLA-B35, and in the Chinese population

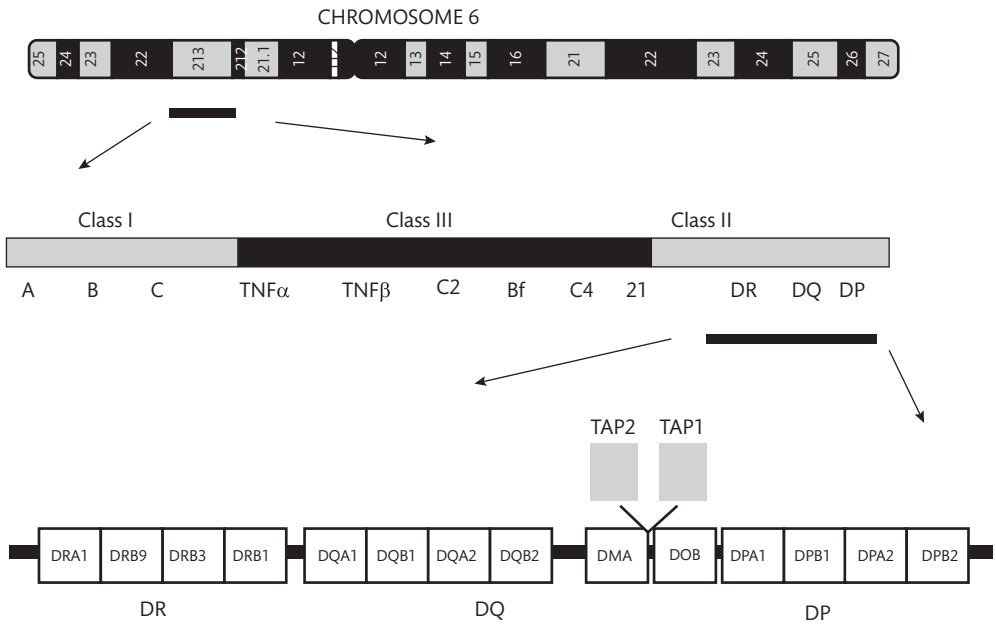


Figure 3.2.1.6 The HLA region is located on chromosome 6p21. It is a complex genetic region which consists of several loci, all of which code for proteins which influence the different arms of the immune system. Depicted are the major loci.

Table 3.2.1.3 Some of the important HLA association studies in Graves' disease performed in Caucasians

Country	Ethnic group	No. of patients	HLA allele	RR
Canada	Caucasians	175	B8 DR3	3.1 5.7
Belgium	Caucasians	101	DRB1*0301 DQA1*0501	
UK	Caucasians	127	B8 DR3	2.8 2.1
France	Caucasians	94	B8 DR3	3.4 4.2
Hungary	Caucasians	256	B8 DR3	3.5 4.8
Ireland	Caucasians	86	B8 DR3	2.5 2.6
Canada	Caucasians	133	DR3	4.6
Sweden	Caucasians	78	B8 DR3	4.4 3.9
USA	Caucasians	65	DR3	3.4
UK	Caucasians	120	DQA1 *0501	3.8
UK[1]	Caucasians	228	DRB1*0304 DQB1*0301 DQA1* 0501	2.7 1.9 3.2
USA	Caucasians	94	DQA1 *0501	3.7

[1]In this study the TDT also showed an association to the extended haplotype DRB1*0304-DQB1*02-DQA1*0501.

an increased frequency of HLA-Bw46 was reported. In African Americans, no overall susceptibility could be associated with any DR allele, although subdivision of the patients revealed that DRw6 was associated with thyroid antibody formation. Among Caucasians, HLA-DQA1*0501 has also been associated with Graves' disease (Table 3.2.1.3), but it appears that the most common susceptibility allele in GD is indeed HLA-DR3 (HLA-DRB1*03) [25].

Table 3.2.1.4 Some of the important HLA association studies in Graves' disease performed in non- Caucasian populations

Country	Ethnic group	No. of patients	HLA allele(s)	RR
Korea	Asian	128	B13 DR5 Drw8	3.8 4.4 2.3
India	Asian Indian	57	B8 Dqw2	4.1 5.4
USA	Black	73	No association	
South Africa	Black	103	DR1 DR3	3.5 2.4
Japan	Japanese	30	DR5 Drw8	8.1 3.1
Japan	Japanese	76	A2 DPB1*0501	2.9 5.3
Hong Kong	Chinese	132	Bw46	4.8
Hong Kong	Chinese	67 (children)	DQB1 *0303	4.2
Hong Kong	Chinese	97	B46 DR9 DQB1*0303	2.3 2.2 3.2

The exact amino-acid sequence in the DRβ1 chain conferring susceptibility to GD was determined by sequencing the HLA-DRB1 locus in a population of GD patients and controls. These studies identified arginine at position 74 of the HLA-DRβ1 chain (DRβ-Arg74) as the most critical DR amino acid conferring susceptibility to GD [25]. Further analysis showed that the presence of glutamine at position 74 was protective for GD. This suggests that position 74 of the DRβ1 chain is critical for GD development in certain patients but not in others. Recent work suggests that certain TSHR and Tg peptides, in particular, have higher affinity for these HLA-DR pocket sequences.

Not only has the association with 'disease' been explored but the role of HLA polymorphisms may be critical for the clinical expression of the different GD phenotypes. Some groups reported an association between the likelihood of relapse of GD with HLA-DR3 while others were unable to confirm this. Studies of HLA associations in GO have also produced conflicting results with some workers reporting increased frequency of HLA-DR3 in patients with GO, and others reporting no difference in the distribution of HLA-DR alleles between GD patients with and without ophthalmopathy. Likewise, no difference in the DR3 frequency was found in GD patients with and without pretibial myxoedema. It appears, therefore, that HLA provides a broad degree of susceptibility rather than for any particular phenotype.

Association of HLA with Hashimoto's thyroiditis: Data on HLA haplotypes in HT have been much less definitive than in GD (see Table 3.2.1.5). Early studies suggested an association of goitrous HT with HLA-DR5 (RR = 3.1) and of atrophic HT with DR3 (RR = 5.1) in Caucasians. Later studies in Caucasians reported weak associations of HT with HLA-DR3 and HLA-DR4. Associations of HT with other HLA haplotypes have also been reported in different ethnic populations (e.g. HLA-DRw53 in Japanese, and HLA-DR9 in Chinese). An HLA-DR pocket amino acid signature was found to confer a strong risk for HT resulting in an odds ratio of 3.7 [26] from a unique pocket structure that could influence the binding of pathogenic peptides to the HLA-DR pocket and their presentation to T cells. In addition, studies suggest that certain Tg peptides, in particular, have affinity for these specific HLA-DR pocket sequences [27].

Table 3.2.1.5 Some of the important HLA association studies in Hashimoto's thyroiditis

Country	Ethnic group	No. of patients	HLA allele	RR[1]
Canada+ England	Caucasians	66	DR4 DR5 DQw7	2.9 3.8 4.7
Canada-Newfoundland	Caucasians	40	DRw3	3.5
Canada-Newfoundland	Caucasians	40	DR5	3.1
England	Caucasians	49	DQB1*0301 DQA1*0301/2	NR[2]
England	Caucasians	36	DR5 DQ7	3.5
England	Caucasians	86	DR3	2.23
Japan	Japanese	99	DRw53	3.33

[1] RR, relative risk; [2]NR, not reported.

Non-HLA Genes in AITD

HLA genes most likely account for only part of the genetic susceptibility to AITD. At least five additional non-HLA genes have been found to confer risk for AITD although each confers much less risk than HLA. Three immunoregulatory genes (*CTLA-4, CD40*, and *PTPN22*) and two thyroid-specific genes (thyroglobulin, and the *TSHR*) have been reproducibly associated with these diseases. There are multiple other reports in the literature but many of these involve low risks and lack confirmation in different study populations. Such reproducibility is the hallmark of significance.

Cytotoxic T-lymphocyte antigen-4 (*CTLA-4*): *CTLA-4* is a costimulatory molecule that participates in the interaction between T cells and antigen presenting cells (APCs). APC's activate T cells by presenting to the T-cell receptor an antigenic peptide bound to an HLA class II protein on the cell surface. For this activation to work, a second signal is required and these **costimulatory signals** are provided by a variety of proteins which are expressed on APCs (e.g. B7-1, B7-2, B7h, CD-40) and interact with receptors (*CD28, CTLA-4*, and *CD-40L*) on the surface of CD 4+ T-lymphocytes during antigen presentation. Whereas, the binding of B7 to CD28 on T cells costimulates T-cell activation, the presence of *CTLA-4*, which has a higher affinity for B7, downregulates T-cell activation by competing for the binding of B7 to CD28. The suppressive effects of *CTLA-4* on T-cell activation may be altered by polymorphisms reducing *CTLA-4* expression and/or function resulting in exaggerated T-cell activation which may lead to the development of autoimmunity in susceptible individuals. Indeed, *CTLA-4* polymorphisms have been consistently shown to be associated with many different autoimmune conditions [28] including both GD and HT [29, 30] and this association has been consistent across populations of different ethnic backgrounds.

Several *CTLA-4* variants have been reported as associated with AITD but three have shown the most consistent association including an (AT)n microsatellite within the 3'UTR region of the *CTLA-4* gene [29], a SNP at position 49 in the *CTLA-4* leader peptide (designated A/G49), resulting in an alanine/threonine substitution [13, 31], and a SNP (designated CT60) located near the 3'UTR of the *CTLA-4* gene [30]. Interestingly the *CTLA-4* gene also appears to confer susceptibility to the production of thyroid antibodies (TAbs) without clinically significant disease [32, 33], thus substantiating its role as a general autoimmunity gene.

CD40: *CD40* is a member of the tumour necrosis factor receptor (TNFR) family of molecules and is expressed primarily on B cells and other APCs including thyroid cells. *CD40* plays a fundamental role in B-cell activation, inducing B-cell proliferation, immunoglobulin class switching, and antibody secretion.

Recently, using a combination of linkage and association studies, *CD40* has been identified as a novel susceptibility gene for GD [34]. A C/T SNP at the 5'UTR of *CD40* was associated with GD, with the CC genotype conferring the risk [35] an observation replicated in several studies [36, 37].

The protein tyrosine phosphatase-22 (*PTPN22*) gene: The lymphoid tyrosine phosphatase (LYP), encoded by the protein tyrosine phosphatase-22 (*PTPN22*) gene, is a 110 kDa protein tyrosine phosphatase that, like *CTLA-4*, is a powerful inhibitor of T-cell activation. A tryptophan/arginine substitution at codon 620 (R620W) of *PTPN22* was found to be associated with AITD mostly with GD [38], as well as with other autoimmune diseases.

Thyroglobulin: Thyroglobulin represents one of the major targets of the immune response in AITD. The Tg gene has been shown consistently to also be an important AITD susceptibility gene [39, 40] with four Tg SNPs significantly associated with AITD [41]. Moreover, three of the associated Tg SNPs were non-synonymous (i.e. they caused amino acid changes in the Tg protein). The association between Tg and AITD has been well replicated [42, 43] although the associated Tg polymorphism has not been consistent in different populations [43, 44].

The thyrotropin receptor (TSHR): The presence of stimulating TSHR autoantibodies is the hallmark of GD indicating that the TSHR is a primary antigen. It is not surprising, therefore, that the *TSHR* gene is also reproducibly associated with GD [45] and the most consistent studies have identified non-coding SNPs in intron 1 of the *TSHR* [46–48].

Whole Genome Screening in AITD

Several early microsatellite and **linkage-based whole genome screens** were performed in AITD with considerable success. In a data set of 102 multiplex Caucasian families (540 individuals) [40] whole genome screening revealed 7 loci that showed evidence for linkage to AITD. Three loci, on chromosomes 6 (named AITD-1 and appearing distinct from the HLA region), chromosome 8 (identified as the thyroglobulin locus), and chromosome 10, showed evidence for linkage with both GD and HT. In contrast, one locus, on chromosome 12 (called HT-2), showed evidence for linkage to HT only while three loci showed evidence for linkage with GD: GD-1 on chromosome 14 (identified as the TSHR locus), GD-2 on chromosome 20 (identified as the *CD40* locus), and a locus on 7q. Another whole genome screen from Japan identified the same AITD locus on 8q (the thyroglobulin gene), as well as a 5q locus [39]. The same 5q locus was also identified in a genome scan performed in the Old Order Amish population in the United States [49]. A large whole genome scan in 1119 European sib-pairs identified three different loci on chromosomes 18p11, 2q36 (distinct from *CTLA-4*), and 11p15 [50]. These studies confirmed that the genetic contribution to the development of AITD involves multiple genes with mostly low-level effects.

More recent **whole genome association studies** (GWAS) using large numbers of SNPs have dominated recent genetic thinking but have added very little to our understanding of AITD genetics. The data have shown primarily HLA as the major contributor to genetic susceptibility and have confirmed the sites previously identified by earlier techniques. Online Table 3.2.1.6 lists the current state of associated sites, not all of which have been replicated but continues the idea that multiple genes contribute small genetic effects.

A Note on Graves' Orbitopathy (GO) and the Importance of Phenotype

Although sensational reports of specific genetic associations with GO have appeared, they have not been replicated, and there remains no conclusive evidence that any of the known AITD susceptibility genes are specific for GO [51]. It appears that the GO phenotype is simply a more severe form of GD rather than a distinct entity and that other factors (such as the environment) may also be important in the development of the more malignant GO pathology.

Mechanisms of Disease Induction by Susceptibility Genes

Mapping susceptibility genes for complex diseases can better our understanding of their pathogenesis. However, even when a

complex disease gene is mapped, unravelling the mechanisms underlying its association with disease is not straightforward. In contrast to classical monogenic diseases where a genetic mutation usually inactivates a gene or causes unchecked activation of a gene, in complex diseases such as AITD the associated genetic variants may cause subtle changes in the function of one or more genes. Therefore, even when a gene associated with a complex disease is mapped it can be challenging to prove that a certain variant changes the function of the gene in a way that will promote the development of the disease. However, considerable progress has been made in dissecting some of the mechanisms by which AITD associated genes predispose to disease.

HLA

The mechanisms by which HLA associations confer disease susceptibility in many autoimmune diseases are now well understood. For T cells to recognize and respond to an antigen they require interaction with a molecular complex consisting pf an antigenic peptide and an HLA class II molecule. It is thought that different HLA alleles have different affinities for peptides derived from the processing of autoantigens (e.g. such as Tg and the TSHR) and which are then recognized by T-cell receptors. Thus, certain alleles may have a preferential affinity for a particular autoantigen because the peptide is able to fit in the antigen binding groove inside the HLA molecule allowing it to be recognized by the T-cell receptor while other HLA molecules may not have the same binding pocket structure. This would determine if an autoimmune response to that antigen peptide will develop.

Studies have demonstrated that DRβ-Arg74 is a critical HLA-DR pocket amino acid associated with GD [52]. Position 74 of the DRβ chain is located in pocket 4 (P4) of the DR peptide-binding cleft. Structural modelling analysis demonstrated that the change at position 74, from the common neutral amino acids (Ala or Gln) to a positively charged hydrophilic amino acid (Arg), significantly modified the three dimensional structure of the P4 peptide-binding pocket [52] thus altering the peptide-binding properties of the pocket and favouring peptides able to induce GD [52, 53]. A similar pocket HLA-DR amino acid signature has been strongly associated with HT [26].

For thyroid autoantigens to be presented by HLA molecules to T cells, a mechanism of autoantigen presentation must exist within the target tissue. One potential mechanism not utilizing professional APCs may be through aberrant expression of HLA class II molecules on the target tissue cells. Indeed, thyroid epithelial cells from patients with AITD have been shown to express HLA class II antigen molecules which are normally expressed only on APCs such as macrophages and dendritic cells [54]. This aberrant expression of HLA molecules on thyroid cells, could initiate thyroid autoimmunity via direct thyroid autoantigen presentation [54, 55].

CTLA-4

CTLA-4 is a negative regulator of T cells and a polymorphism that decreases *CTLA-4* function and/or cell surface expression would cause enhanced T-cell activation potentially contributing to the development of an autoimmune condition [45] as seen when treating patients with anti-*CTLA-4*. Several *CTLA-4* variants have been analysed in detail for their effect on *CTLA-4* function and/or expression and a comprehensive analysis of the *CTLA-4* gene locus reported that the CT60 SNP of *CTLA-4* (rs3087243) showed the strongest association with GD, suggesting that it might be the causative SNP [30].

CD40

CD40 is an important costimulator of T-cell activation associated with a number of autoimmune diseases. The CC genotype of a *CD40* 5'UTR SNP was shown to be associated with GD. This *CD40* SNP resides in a region which can influence the initiation of translation and, therefore, the expression of *CD40*. Indeed, the C-allele of the 5'UTR SNP was shown to increase the translation of *CD40* mRNA transcripts, by 20–30% compared to the T-allele [56, 57]. At least two potential mechanisms can explain how the C-allele of the *CD40* 5'UTR SNP increases the risk for GD: (1) The C-allele may increase *CD40* expression and function on B cells, thereby potentially lowering the threshold of activation of thyroid autoreactive B cells; and/or (2) The C-allele may increase the expression of *CD40* in the thyroid gland itself [58, 59]. *CD40* signalling in thyrocytes can result in cytokine secretion (e.g. IL-6) [58]. Thus, overexpression of *CD40* on thyroid cells may, under certain conditions (e.g. infection), result in increased secretion of cytokines by thyroid cells causing local inflammation and activation of autoreactive T cells that were dormant or suppressed by peripheral regulatory mechanisms. This mechanism is known as a bystander mechanism of induction of autoimmunity and is seen in experimental thyroiditis [60]. Studies with transgenic mice overexpressing thyroid *CD40* indicate that the development of GD after TSHR immunization was more profound confirming the potential role of this costimulator [61].

PTPN22

The lymphoid tyrosine phosphatase (LYP) encoded by the *PTPN22* gene belongs to a family of protein tyrosine phosphatases that are expressed in both immature and mature B- and T-lymphocytes. LYP is a powerful inhibitor of the T-cell antigen receptor signalling pathway. LYP binds to the C terminal of the protein kinase, Csk, restricting the response to antigens by disrupting protein tyrosine phosphorylation events that control cell activation and differentiation. This negative control mechanism prevents spontaneous T-cell activation [62].

The exact mechanism by which the associated R620W variant of the *PTPN22* gene predisposes to autoimmunity is not known. The substitution of arginine with tryptophan at this position interferes with the interaction of LYP with Csk. *In vitro* experiments show that only LYP with arginine at position 620 forms a complex with Csk whereas LYP with tryptophan at this position binds less efficiently [63]. One study suggested that the tryptophan variant is a gain-of-function change that makes the protein an even stronger inhibitor of T cells [64]. Thus, the disease-associated tryptophan variant would be expected to suppress T-cell activation and proliferation. Reduced T-cell receptor signalling could lead to a tendency for self-reactive T cells to escape thymic deletion and thus remain in the periphery.

Thyroglobulin

Thyroglobulin (Tg) is a 660 kDA homodimeric protein that serves as a precursor and storehouse for thyroid hormones. Tg is one of the main targets of the immune response in both HT and GD and by immunization is able to induce experimental autoimmune thyroiditis (EAT) in a variety of animal models, most notably the mouse [65]. EAT, like HT is characterized by a diffuse cellular infiltrate of the thyroid with anti-Tg T-cell responses as well as high titres of Tg autoantibodies. This indicates that Tg is a critical thyroid-specific antigen involved in the aetiology of AITD.

It is not surprising, therefore, that the Tg gene is an AITD susceptibility gene [42, 45]. Three amino acid substitutions in Tg were first reported to be significantly associated with AITD [40] and several mechanisms have been postulated to explain the association. Clearly the presentation of a Tg peptide by APCs to T cells must be involved and since peptide antigens are presented within HLA class II molecules, as discussed earlier, this mechanism would imply that there exist a more potent interaction between certain Tg variant peptides related to the associated SNPs and certain HLA-DR variants in susceptible patients predisposing them to AITD. This was shown for one such peptide which had a strong statistical interaction with the Arg74 polymorphism of HLA-DR, resulting in a high odds ratio of 15 for susceptibility to GD [66]. Supporting this hypothesis was a study which identified specific HLA-DR bound Tg peptides within the thyroid glands from Graves' patients [67]. Moreover, studies in 'humanized' mice expressing HLA-DR3 have shown that certain Tg peptides can induce EAT in these mice [68]. Another variant in the Tg promoter (-1623A/G) was also found to be associated with AITD. Functional analysis showed that the nucleotide substitution introduced by this Tg promoter SNP created a binding site for interferon regulatory factor-1 (IRF-1), causing up-regulating Tg promoter activity upon IRF-1 binding [69]. These data suggest an attractive mechanism for environmental–genetic interaction whereby during viral infections local production of interferon alpha can lead to upregulation of Tg transcription via IRF-1 in individuals carrying the risk (G) allele of the Tg promoter SNP but not in individuals carrying the protective (A) allele.

TSH Receptor (TSHR)

All the TSHR SNPs which are consistently associated with GD are intronic [46–48]. Therefore, the mechanism by which they predispose to GD is more challenging to dissect. It has been postulated that these intronic SNPs may influence the expression of the TSHR through regulatory elements. For example, intrathymic TSHR expression was decreased in individuals homozygous for a particular associated SNP (rs12101261) compared with carriers of the disease-protective allele. Furthermore, such repression was enhanced by interferon alpha acting through an epigenetic mechanism [70]. Alternatively, SNPs may be associated with changes in the alternative splicing of the TSHR although how this alters the immune response is uncertain [71] but changes in thymic expression may be involved [72].

The X Chromosome

There are a number of possible mechanisms whereby the X chromosome could influence the development of AITD. One mechanism is probabilistic. Females have two X chromosomes (one paternal and one maternal) while males have only one X chromosome (maternal). Therefore, females are twice as likely to inherit an X chromosome AITD susceptibility gene as males. Several immune regulatory genes are located on the X chromosome including the *FOXP3* gene which was associated with AITD in Caucasians [73]. FOXP3 is the master regulator gene of T regulatory cell differentiation hinting at a mechanism for the female preponderance of AITD.

Another possible mechanism involves X-inactivation. X-inactivation in females results in the production of two classes of cells that differ in the transcription of X-chromosome encoded genes including genes coding for self-antigens. If these two cell classes extend to the thymic cells responsible for tolerizing T cells in embryonic life, some lymphocytes may not be tolerized to one of the two self-antigens encoded by the X chromosome. Such lymphocytes would be autoreactive to that antigen and could induce an autoimmune response. Supporting this hypothesis are data showing skewing of X-inactivation in females with AITD [74–76].

Conclusions

Genetic susceptibility plays an important role in the development of AITD. While significant progress has been made in identifying AITD susceptibility genes and understanding the mechanisms by which they confer risk for disease, their contribution to the genetic susceptibility is far from overwhelming suggesting other important genetic mechanisms must be involved. Intriguingly, the AITD susceptibility genes identified participate in the immunological synapse and/or the signalling pathways activated by the immunological synapse. The immunological synapse is the interface between antigen presenting cell and T-cell formed during antigen presentation. This finding, alone, suggests that the genetic factors predisposing to AITD may lead to breakdown of tolerance by altering the immunological synapse.

Some of the AITD susceptibility genes identified so far are unique for GD or HT, while other are common to both conditions [40] and even type 1 diabetes [77] (**Figure 3.2.1.7**). Mechanistically, the AITD susceptibility genes can be divided into immune modulating genes (*HLA-DR, CD40, CTLA-4*, and *PTPN22*), and thyroid-specific genes (Tg and TSHR). Interactions between these genes may significantly increase the assigned risk [3], for example with Tg and *HLA-DR* [66]. In addition, some genes, while exerting small effects in all patients, have a much stronger effect in subsets of patients [78]. Most importantly, identifying the AITD susceptibility genes and understanding the mechanisms by which they predispose to disease gives us a better understanding of the molecular mechanisms causing thyroid autoimmunity.

REFERENCES

1. Muller AF, Drexhage HA, Berghout A. Postpartum thyroiditis and autoimmune thyroiditis in women of childbearing age: recent insights and consequences for antenatal and postnatal care. *Endocr Rev*, 2001; **22**: 605–30.
2. Mandac JC, Chaudhry S, Sherman KE, Tomer Y. The clinical and physiological spectrum of interferon-alpha induced thyroiditis: toward a new classification. *Hepatology*, 2006; **43**: 661–72.
3. Huber A, Menconi F, Corathers S, Jacobson EM, Tomer Y. Joint genetic susceptibility to type 1 diabetes and autoimmune thyroiditis: from epidemiology to mechanisms. *Endocr Rev*, 2008; **29**: 697–725.
4. Hall R, Dingle PR, Roberts DF. Thyroid antibodies: a study of first degree relatives. *Clin Genet*, 1972; **3**: 319–24.
5. Hollowell JG, Staehling NW, Flanders WD, *et al*. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
6. Tunbridge WMG, Evered DC, Hall R, *et al*. The spectrum of thyroid disease in a community: the Whickham survey. *Clin Endocrinol Oxf*, 1977; **7**: 481–93.
7. Vanderpump MPJ, Tunbridge WMG, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol (Oxf)*, 1995; **43**: 55–68.

8. Furszyfer J, Kurland LT, McConahey WM, Woolner LB, Elveback LR. Epidemiologic aspects of Hashimoto's thyroiditis and Graves' disease in Rochester, Minnesota (1935–1967), with special reference to temporal trends. *Metabolism*, 1972; **21**: 197–204.
9. Berglund J, Ericsson UB, Hallengren B. Increased incidence of thyrotoxicosis in Malmo during the years 1988–1990 as compared to the years 1970–1974. *J Intern Med*, 1996; **239**: 57–62.
10. Davies TF, Amino N. A new classification for human autoimmune thyroid disease. *Thyroid*, 1993; **3**: 331–3.
11. Hall R, Stanbury JB. Familial studies of autoimmune thyroiditis. *Clin Exp Immunol*, 1967; **2**: 719–25.
12. Burek CL, Hoffman WH, Rose NR. The presence of thyroid autoantibodies in children and adolescents with AITD and in their siblings and parents. *Clin Immunol Immunopathol*, 1982; **25**: 395–404.
13. Villanueva RB, Inzerillo AM, Tomer Y, *et al*. Limited genetic susceptibility to severe graves' ophthalmopathy: no role for *CTLA-4* and evidence for an environmental etiology. *Thyroid*, 2000; **10**: 791–8.
14. Strieder TG, Tijssen JG, Wenzel BE, Endert E, Wiersinga WM. Prediction of progression to overt hypothyroidism or hyperthyroidism in female relatives of patients with autoimmune thyroid disease using the Thyroid Events Amsterdam (THEA) score. *Arch Intern Med*, 2008; **168**: 1657–1663.
15. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. *Cell*, 1996; **85**: 311–18.
16. Brix TH, Christensen K, Holm NV, Harvald B, Hegedus L. A population-based study of Graves' diseases in Danish twins. *Clin Endocrinol*, 1998; **48**: 397–400.
17. Brix TH, Kyvik KO, Christensen K, Hegedus L. Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. *J Clin Endocrinol Metab*, 2001; **86**: 930–4.
18. Ringold DA, Nicoloff JT, Kesler M, Davis H, Hamilton A, Mack T. Further evidence for a strong genetic influence on the development of autoimmune thyroid disease: the California twin study. *Thyroid*, 2002; **12**: 647–53.
19. Brix TH, Kyvik KO, Hegedus L. A population-based study of chronic autoimmune hypothyroidism in Danish twins. *J Clin Endocrinol Metab*, 2000; **85**: 536–9.
20. Aho K, Gordin A, Sievers K, Takala J. Thyroid autoimmunity in siblings: a population study. *Acta Endocrinol*, 1983; Suppl. **251**: 11–15.
21. Phillips DIW, Prentice L, McLachlan SM, Upadhyaya M, Lunt PW, Rees Smith B. Autosomal dominant inheritance of the tendency to develop thyroid autoantibodies. *Exp Clin Endocrinol*, 1991; **97**: 170–2.
22. Jaume JC, Guo J, Pauls DL, *et al*. Evidence for genetic transmission of thyroid peroxidase autoantibody epitopic 'fingerprints'. *J Clin Endocrinol Metab*, 1999; **84**: 1424–31.
23. Huber AK, Jacobson EM, Jazdzewski K, Concepcion ES, Tomer Y. Interleukin (IL)-23 receptor is a major susceptibility gene for Graves' ophthalmopathy: the IL-23/T-helper 17 axis extends to thyroid autoimmunity. *J Clin Endocrinol Metab*, 2008; **93**: 1077–81.
24. Tomer Y, Ban Y, Concepcion E, *et al*. Common and unique susceptibility loci in Graves and Hashimoto diseases: results of whole-genome screening in a data set of 102 multiplex families. *Am J Hum Genet*, 2003; **73**: 736–47.
25. Ban Y, Davies TF, Greenberg DA, *et al*. Arginine at position 74 of the HLA-DRb1 chain is associated with Graves' disease. *Genes Immun*, 2004; **5**: 203–8.
26. Menconi F, Monti MC, Greenberg DA, *et al*. Molecular amino acid signatures in the MHC class II peptide-binding pocket predispose to autoimmune thyroiditis in humans and in mice. *Proc Natl Acad Sci U S A*, 2008; **105**: 14034–9.
27. Jacobson EM, Yang H, Menconi F, *et al*. Employing a recombinant HLA-DR3 expression system to dissect major histocompatibility complex II-thyroglobulin peptide dynamism: a genetic, biochemical, and reverse immunological perspective. *J Biol Chem*, 2009; **284**: 34231–43.
28. Tomer Y. Unraveling the genetic susceptibility to autoimmune thyroid diseases: CTLA-4 takes the stage. *Thyroid*, 2001; **11**: 167–9.
29. Kotsa K, Watson PF, Weetman AP. A *CTLA-4* gene polymorphism is associated with both Graves' disease and autoimmune hypothyroidism. *Clin Endocrinol*, 1997; **46**: 551–4.
30. Ueda H, Howson JM, Esposito L, *et al*. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*, 2003; **423**: 506–11.
31. Nithiyananthan R, Heward JM, Allahabadia A, Franklyn JA, Gough SC. Polymorphism of the *CTLA-4* gene is associated with autoimmune hypothyroidism in the United Kingdom. *Thyroid*, 2002; **12**: 3–6.
32. Tomer Y, Greenberg DA, Barbesino G, Concepcion ES, Davies TF. *CTLA-4* and not CD28 is a susceptibility gene for thyroid autoantibody production. *J Clin Endocrinol Metab*, 2001; **86**: 1687–93.
33. Zaletel K, Krhin B, Gaberscek S, Hojker S. Thyroid autoantibody production is influenced by exon 1 and promoter *CTLA-4* polymorphisms in patients with Hashimoto's thyroiditis. *Int J Immunogenet*, 2006; **33**: 87–91.
34. Tomer Y, Barbesino G, Greenberg DA, Concepcion ES, Davies TF. A new Graves disease-susceptibility locus maps to chromosome 20q11.2. *Am J Hum Genet*, 1998; **63**: 1749–56.
35. Tomer Y, Concepcion E, Greenberg DA. A C/T single nucleotide polymorphism in the region of the *CD40* gene is associated with Graves' disease. *Thyroid*, 2002; **12**: 1129–35.
36. Pearce SH, Vaidya B, Imrie H, *et al*. Further evidence for a susceptibility locus on chromosome 20q13.11 in families with dominant transmission of Graves disease [letter]. *Am J Hum Genet*, 1999; **65**: 1462–5.
37. Ban Y, Tozaki T, Taniyama M, Tomita M, Ban Y. Association of a C/T single-nucleotide polymorphism in the 5' untranslated region of the *CD40* gene with Graves' disease in Japanese. *Thyroid*, 2006; **16**: 443–6.
38. Velaga MR, Wilson V, Jennings CE, *et al*. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. *J Clin Endocrinol Metab*, 2004; **89**: 5862–5.
39. Sakai K, Shirasawa S, Ishikawa N, *et al*. Identification of susceptibility loci for autoimmune thyroid disease to 5q31-q33 and Hashimoto's thyroiditis to 8q23-q24 by multipoint affected sib-pair linkage analysis in Japanese. *Hum Mol Genet*, 2001; **10**: 1379–86.
40. Tomer Y, Ban Y, Concepcion E, *et al*. Common and unique susceptibility loci in Graves and Hashimoto diseases: results of whole-genome screening in a data set of 102 multiplex families. *Am J Hum Genet*, 2003; **73**: 736–47.
41. Ban Y, Greenberg DA, Concepcion E, Skrabanek L, Villanueva R, Tomer Y. Amino acid substitutions in the thyroglobulin gene are associated with susceptibility to human and murine autoimmune thyroid disease. *Proc Natl Acad Sci U S A*, 2003; **100**: 15119–24.
42. Collins JE, Heward JM, Carr-Smith J, Daykin J, Franklyn JA, Gough SCL. Association of a rare thyroglobulin gene

microsatellite variant with autoimmune thyroid disease. *J Clin Endocrinol Metab*, 2003; **88**: 5039–42.
43. Ban Y, Tozaki T, Taniyama M, Tomita M, Ban Y. Association of a thyroglobulin gene polymorphism with Hashimoto's thyroiditis in the Japanese population. *Clin Endocrinol (Oxf)*, 2004; **61**: 263–8.
44. Collins JE, Heward JM, Howson JM, *et al*. Common allelic variants of exons 10, 12, and 33 of the thyroglobulin gene are not associated with autoimmune thyroid disease in the United Kingdom. *J Clin Endocrinol Metab*, 2004; **89**: 6336–9.
45. Jacobson EM, Tomer Y. The *CD40, CTLA-4*, thyroglobulin, TSH receptor, and *PTPN22* gene quintet and its contribution to thyroid autoimmunity: Back to the future. *J Autoimmun*, 2007; **28**: 85–98.
46. Hiratani H, Bowden DW, Ikegami S, *et al*. Multiple SNPs in intron 7 of thyrotropin receptor are associated with Graves' disease. *J Clin Endocrinol Metab*, 2005; **90**: 2898–903.
47. Ho SC, Goh SS, Khoo DH. Association of Graves' disease with intragenic polymorphism of the thyrotropin receptor gene in a cohort of Singapore patients of multi-ethnic origins. *Thyroid*, 2003; **13**: 523–8.
48. Dechairo BM, Zabaneh D, Collins J, *et al*. Association of the TSHR gene with Graves' disease: the first disease specific locus. *Eur J Hum Genet*, 2005; **13**: 1223–30.
49. Allen EM, Hsueh WC, Sabra MM, *et al*. A genome-wide scan for autoimmune thyroiditis in the Old Order Amish: replication of genetic linkage on chromosome 5q11.2-q14.3. *J Clin Endocrinol Metab*, 2003; **88**: 1292–6.
50. Taylor JC, Gough SC, Hunt PJ, *et al*. A genome-wide screen in 1119 relative pairs with autoimmune thyroid disease. *J Clin Endocrinol Metab*, 2006; **91**: 646–53.
51. Yin X, Latif R, Bahn R, Davies TF. Genetic profiling in Graves' disease: further evidence for lack of a distinct genetic contribution to Graves' ophthalmopathy. *Thyroid*, 2012; **22**: 730–6.
52. Ban Y, Davies TF, Greenberg DA, *et al*. Arginine at position 74 of the HLA-DRb1 chain is associated with Graves' disease. *Genes Immunol*, 2004; **5**: 203–8.
53. Sawai Y, DeGroot LJ. Binding of human thyrotropin receptor peptides to a Graves' disease-predisposing human leukocyte antigen class II molecule. *J Clin Endocrinol Metab*, 2000; **85**: 1176–9.
54. Hanafusa T, Pujol Borrell R, Chiovato L, Russell RC, Doniach D, Bottazzo GF. Aberrant expression of HLA-DR antigen on thyrocytes in Graves' disease: relevance for autoimmunity. *Lancet*, 1983; **2**: 1111–15.
55. Piccinini LA, Mackenzie WA, Platzer M, Davies TF. Lymphokine regulation of HLA-DR gene expression in human thyroid cell monolayers. *J Clin Endocrinol Metab*, 1987; **64**: 543–8.
56. Jacobson EM, Concepcion E, Oashi T, Tomer Y. A Graves' disease-associated Kozak sequence single-nucleotide polymorphism enhances the efficiency of *CD40* gene translation: a case for translational pathophysiology. *Endocrinology*, 2005; **146**: 2684–91.
57. Park JH, Chang HS, Park CS, *et al*. Association analysis of *CD40* polymorphisms with asthma and the level of serum total IgE. *Am J Respir Crit Care Med*, 2007; **175**: 775–82.
58. Metcalfe RA, McIntosh RS, Marelli-Berg F, Lombardi G, Lechler R, Weetman AP. Detection of *CD40* on human thyroid follicular cells: analysis of expression and function. *J Clin Endocrinol Metab*, 1998; **83**: 1268–74.
59. Smith TJ, Sciaky D, Phipps RP, Jennings TA. *CD40* expression in human thyroid tissue: evidence for involvement of multiple cell types in autoimmune and neoplastic diseases. *Thyroid*, 1999; **9**: 749–55.
60. Arata N, Ando T, Unger P, Davies TF. By-stander activation in autoimmune thyroiditis: studies on experimental autoimmune thyroiditis in the GFP+ fluorescent mouse. *Clin Immunol*, 2006; **121**: 108–17.
61. Huber AK, Finkelman FD, Li CW, *et al*. Genetically driven target tissue overexpression of *CD40*: a novel mechanism in autoimmune disease. *J Immunol*, 2012; **189**: 3043–53.
62. Vang T, Miletic AV, Bottini N, Mustelin T. Protein tyrosine phosphatase PTPN22 in human autoimmunity. *Autoimmunity*, 2007; **40**: 453–61.
63. Bottini N, Vang T, Cucca F, Mustelin T. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol*, 2006; **18**: 207–13.
64. Vang T, Congia M, Macis MD, *et al*. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet*, 2005; **37**: 1317–19.
65. Stafford EA, Rose NR. Newer insights into the pathogenesis of experimental autoimmune thyroiditis. *Int Rev Immunol*, 2000; **19**: 501–33.
66. Hodge SE, Ban Y, Strug LJ, *et al*. Possible interaction between HLA-DRbeta1 and thyroglobulin variants in Graves' disease. *Thyroid*, 2006; **16**: 351–5.
67. Muixi L, Carrascal M, Alvarez I, *et al*. Thyroglobulin peptides associate *in vivo* to HLA-DR in autoimmune thyroid glands. *J Immunol*, 2008; **181**: 795–807.
68. Flynn JC, McCormick DJ, Brusic V, *et al*. Pathogenic human thyroglobulin peptides in HLA-DR3 transgenic mouse model of autoimmune thyroiditis. *Cell Immunol*, 2004; **229**: 79–85.
69. Stefan M, Jacobson EM, Huber AK, *et al*. Novel variant of thyroglobulin promoter triggers thyroid autoimmunity through an epigenetic interferon alpha-modulated mechanism. *J Biol Chem*, 2011; **286**: 31168–79.
70. Stefan M, Wei C, Lombardi A, *et al*. Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity. *Proc Natl Acad Sci U S A*, 2014; **111**: 12562–7.
71. Graves PN, Tomer Y, Davies TF. Cloning and sequencing of a 1.3 kb variant of human thyrotropin receptor mRNA lacking the transmembrane domain. *Biochem Biophys Res Commun*, 1992; **187**: 1135–43.
72. Colobran R, Armengol MP, Faner R, *et al*. Association of an SNP with intrathymic transcription of TSHR and Graves' disease: a role for defective thymic tolerance. *Hum Mol Genet*, 2011; **20**: 3415–23.
73. Ban Y, Tozaki T, Tobe T, *et al*. The regulatory T cell gene FOXP3 and genetic susceptibility to thyroid autoimmunity: an association analysis in Caucasian and Japanese cohorts. *J Autoimmun*, 2007; **28**: 201–7.
74. Brix TH, Knudsen GP, Kristiansen M, Kyvik KO, Orstavik KH, Hegedus L. High frequency of skewed X-chromosome inactivation in females with autoimmune thyroid disease: a possible explanation for the female predisposition to thyroid autoimmunity. *J Clin Endocrinol Metab*, 2005; **90**: 5949–53.
75. Ozcelik T, Uz E, Akyerli CB, *et al*. Evidence from autoimmune thyroiditis of skewed X-chromosome inactivation in female predisposition to autoimmunity. *Eur J Hum Genet*, 2006; **14**: 791–7.
76. Yin X, Latif R, Tomer Y, Davies TF. Thyroid epigenetics: X chromosome inactivation in patients with autoimmune thyroid disease. *Ann N Y Acad Sci*, 2007; **1110**: 193–200.
77. Tomer Y, Dolan LM, Kahaly G, *et al*. Genome wide identification of new genes and pathways in patients with both autoimmune thyroiditis and type 1 diabetes. *J Autoimmun*, 2015; **60**: 32–9.
78. Tomer Y, Menconi F, Davies TF, *et al*. Dissecting genetic heterogeneity in autoimmune thyroid diseases by subset analysis. *J Autoimmun*, 2007; **29**: 69–77.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

3.2.2 Environmental Factors

Josef Köhrle

Introduction

The hypothalamus–, pituitary–, thyroid–periphery (HPTP) axis has been known as a vulnerable target for environmental factors and nutritional agents since centuries. Goitrogenesis, hypo-, hyperthyroidism, tumorigenesis and autoimmune diseases of this gland have been linked to single or combined deficiencies of several essential trace elements. Normal thyroid function depends on adequate and balanced availability of the essential trace elements iodine, selenium, iron, and the mineral zinc in the daily diet. Evolution of humankind has been suggested to closely follow coast lines and regions with high availability of iodine, the key element required for thyroid hormone (TH) synthesis [1–4]. Involuntary or voluntary environmental or nutritive exposure to adverse factors and agents impairing TH synthesis, secretion, binding, transport, metabolism, and action ('goitrogens') contributes to development and persistence of thyroid disorders [5]. Especially conditions of iodine deficiency, still prevalent in many regions of our world, but also iodine excess [6], both of which might occur already during embryonal and fetal development, in newborns, adolescents, and adults provide the vulnerable platform for action of adverse agents, which might be well tolerated by a normally functioning, 'quiescent' thyroid gland with adequate iodine supply (see Chapters 3.2.3 and 3.2.4). Compounds adversely affecting the HPTP axis belong to several chemical classes of food ingredients and environmental contaminants, but might also represent pharmaceutical drugs acting either directly on biomolecules comprising the HPTP axis or after modification by phase I and/or II drug metabolism (see Table 3.2.2.1). Apart from ingestion several agents reach their targets after inhalation (e.g. occupational exposure or smoking) or by dermal application (e.g. UV sunscreens).

Environmental factors such as temperature, light, altitude, and latitude of living as well as physical, emotional, and acute mental stress, diseases and adverse life events impinge on normal HPTP function (Box 3.2.2.1) [7]. Very recently, it has been suggested that the worldwide pandemic of diseases associated with changes in industrialized and developing countries, such as obesity, diabetes, and metabolic syndrome, is linked to inadequate iodine supply and altered TH homeostasis during development by epigenetic mechanisms [6–13]. Current conditions in industrialized westernized countries are characterized by the permanent availability and overconsumption of energy-rich, fibre-poor, semi-processed, industrialized, enhanced, fortified, or even 'novel food', sedentary lifestyle, lack of sufficient mobility and physical activity, all of which impinge on hormonal homeostasis that is mainly integrated at the hypothalamic level involving thyrotropin-releasing-hormone producing neurons. It is becoming apparent that not only starvation, fasting, and protein-calorie malnutrition in developing countries but also overfeeding with hypercaloric, energy-dense food and obesity in Western-style regions can lead to inadequate intake of micronutrients (minerals, vitamins, and secondary metabolites of plants), so-called hidden hunger. In addition, active and passive smoking, wellness-, neuroenhancement-, lifestyle-, fashion-, and psychodrugs and narcotics have strong impact on TH synthesis, secretion, and action.

Impaired thyroid function has also been observed after consumption of protein-restricted diets, as recommended for patients with phenylketonuria or milder hyperphenylalaninaemias, where adequate iodide supply is essential to compensate for possible adverse effects on TH synthesis and metabolism [14]. Various staple foods, if inadequately processed or preserved, contain efficient goitrogens (e.g. linamarin, goitrin), which release (iso-)thiocyanate, potent inhibitors of NIS-mediated iodide uptake by thyrocytes and—at higher concentrations—also act as effective blockers of thyroperoxidase (TPO), if iodide supply is inadequate.

This chapter will summarize established data for humans, discuss recent findings and possible risk factors identified from exposure data of human subgroups, epidemiology, and selected findings in experimental animal models accepted as relevant for human risk analysis. Figure 3.2.2.1 illustrates currently identified targets of the HPTP axes for environmental agents. Issues of iodine deficiency and excess, pharmaceutical drugs, and radioactive isotopes interfering with the thyroid axis will be discussed elsewhere in this volume. It is quite obvious that adverse effects of nutritional and environmental agents on the human TH axis can only be deduced and extrapolated from occupational exposure data, epidemiological observations, mainly retrospective analysis, but not from interventional studies or very rarely from prospective or blinded trials. However, for many of these observational and epidemiological observations solid interventional studies have been made from appropriate validated animal experimental studies, which clearly provide mechanistic insight and allow conclusions on cause–effect relationships also for adverse human health effects.

Various mechanisms for interference of environment with the HPTP axis have been identified, such as:

- Reversible or irreversible competition with ligand binding sites of the TH axis
- Interference with or alterations of the feedback setpoints, which can already occur in the pre- and early postnatal phase
- Classic 'goitrogenesis' by impaired hormone synthesis
- Disturbance of serum hormone binding, tissue distribution, cellular uptake, metabolism, and action

Many of these disturbances may be initially compensated by the complex regulatory network of the axis, which is characterized by multiply redundant and fail-safe feedback mechanisms and a high degree of plasticity and adaptation to the environment. However,

Table 3.2.2.1 Agents and compounds interfering with the HPTP axis

Compounds	Source and occurrence	Mechanism of action	Effects	Reference(s)
Environmental				
Perchlorate	Solid rocket and missile fuel; air bags	Inhibition of NIS	Goitrogenic	[16–27]
Phthalate esters	Daily life and medical products	?	?	[59, 73, 74]
Polychlorinated (PCB) and polybrominated (PBB) biphenyls	Chemicals; daily life products; flame retardants; fracking	TH transport, uptake, metabolism; T_3 receptor binding		[7, 48, 49, 51–58, 72, 109]
Dioxins (TCDD) and furans	Unintentionally produced by-products, pyrolysis	TH transport, uptake, metabolism		[46, 47, 54–58]
Polycyclic aromatic hydrocarbons PAH)	Chemicals			[104, 109]
(Poly-)Phenols, bisphenol A (BPA)	Plasticizer in daily life products	Multiple actions		[63–71]
Nitrate; nitrite	Fertilizers; food preserver	NIS inhibitors	Goitrogenic	[22, 25, 28, 29]
UV screens (4-MBC, BP2–4)	Sunscreens in cosmetics, daily life products	NIS expression TPO inhibitors	Goitrogenic	[5, 29]
Tobacco and cigar smoke	Goitrogens	Inhibition of TH synthesis, metabolism, and action	Goitrogenic, Strongest risk factor for Graves' disease; might prevent Hashimoto's thyroiditis? Altered maternal and fetal thyroid function	[34–39]
Nutritive				
Various goitrogens	Staple food, vegetables, inappropriately processed food	TPO inhibitor	Goitrogenic	[31, 32]
Linamarin, goitrin, various glycosides	Staple food (Cassava), vegetables (*Cruciferae*)	TPO inhibitor	Goitrogenic	[31]
Flavonoids (polyphenols), silychristin		TPO inhibitors TTR competitors, MCT8 inhibitors	Goitrogenic	[5, 105]
Polyhydroxyphenols and phenol derivatives				[59]
Sulphurated organics, (Iso-)thiocyanate, cyanide, thio-oxazolidone (goitrin)	Nutritive goitrogen Tobacco smoke	NIS inhibitor TPO inhibitor at higher concentration	Goitrogenic	[17, 22, 31, 34, 42]
aliphatic disulphides	Onions, garlic; Water contaminated from coal mining	TPO inhibitors	Goitrogenic	[31]
Humic acids	Ground and drinking water	Interference with TH synthesis	Goitrogenic	[44, 45]
Deficiencies in trace elements and minerals				
Selenium deficiency	Seafood and red meat	Essential trace element for TH synthesis and metabolism	Goitrogenic	[94–96]
Iron deficiency	Protein malnutrition Genetic predisposition	Essential trace element for TH biosynthesis, inadequate function of the haemoprotein TPO for TH biosynthesis		[96, 106]
Zn deficiency				[107]
Pharmaceuticals				
Lithium	Antidepressant	Inhibition of TH secretion	Goitrogenic	[108]
Iodinated agents	Oral bile duct X-ray contrast agents (e.g. iopanoate); antiarrhythmic drug amiodarone	TR antagonist	Hypothyroidism	[80–82]

Box 3.2.2.1 Environmental agents and factors interfering with hypothalamus–pituitary–thyroid–periphery axis

- Light, day–night rhythm (shift workers)
- Latitude of living
- Ambient temperature
- Drinking water, food, nutritional components (voluntarily or involuntarily exposure)
- Inappropriately processed food: goitrogens
- Diets (vegetarian, vegan, or macrobiotics with inadequate iodide, selenium, iron, zinc and retinol content)
- Diet containing constituents with goitrogenic effects under conditions of inadequate iodide intake
- Environmental emissions and exposure: inhalation of aerosol, 'particulate matter' and nanoparticles
- Industrial contaminants
- Agricultural environmental agents
- UV filters and UV screens

long-term exposure to low dose or acute challenge by adverse agents may overstrain the HPTP axis especially under conditions of inadequate iodine supply or during vulnerable phases of the individual's life (development, pregnancy, and lactation, non-thyroidal disease) and thus create harm or disease. There are clear 'windows of susceptibility' to irreversible disruption of the HPTP axis and its setpoints (e.g. during early intrauterine development, adolescence, puberty, and adult life) [13, 15].

Adverse Effects of Various Agents on HPTP Axis

Perchlorate

Perchlorate, similar to pertechnetate, perrhenate, astatine and (iso-)thiocyanate, is a voluminous anion and relevant substrate for the sodium-iodide symporter (NIS) [16], which is located in the basolateral membrane of thyrocytes but also in the lactating mammary gland, salivary gland and several internal epithelial structures (gastric mucosa, lung epithelium, etc.). These anions effectively compete for the essential iodide uptake (Table 3.2.2.2) but are not organified in thyroglobulin (Tg) by the haemoprotein thyroperoxidase (TPO). Thus, perchlorate has been even used as efficient pharmaceutical to treat hyperthyroidism and to block unwanted (radio-)iodide uptake into the gland or for the diagnostic perchlorate discharge test, performed to identify iodide organification defects. Hypothyroidism can be obtained by regular administration of doses of 0.4 mg/kg body weight per day, while reference doses, where no appreciable risk can be observed for human populations, are in the range of 0.7 µg/kg/day. Thiocyanate or nitrate are less potent by a factor of 15 or 240, respectively [17]. Recently, reports have been emerging on increasing contamination of surface land and water by potassium or ammonium perchlorate around areas close to civil or military plants as well as installations producing and handling rockets, missiles, ammunitions, and fireworks. Potassium and ammonium perchlorate are increasingly used and widely distributed over our planet as rocket and missile fuel waste and are extremely stable and poorly degraded in environment. Other perchlorate salts are used as oxidizers, electrolytes and in various technical processes [18]. Concerns have been raised and published whether this increasing contamination of surface soil and drinking water might negatively impact on thyroid function of exposed populations, especially babies, children and adolescents [19–22], who still have limited capacity and reserve to synthesize and store iodinated thyroglobulin. In contrast, adults, whose follicular colloid thyroglobulin stores might last for up to 3 months, if adequately supplied with iodide before interference, might be less vulnerable except during pregnancy or lactation, where iodine demands are increased. This controversial issue is subject of several ongoing surveys by environmental and regulatory authorities, but for the moment no clear evidence for risk assessment is available. However, perchlorate concentrations in drinking water of exposed areas have been determined which are in the range or even exceed recommendations by regulatory authorities. Observations in workers regularly exposed to airborne perchlorate provided no evidence for adverse effects, but individuals with inadequate iodide intake, exposed to other environmental NIS inhibitors or belonging to other susceptible risk groups might experience negative consequences of long-term perchlorate exposure by drinking water with perchlorate concentrations in the range of the discussed US reference doses (US Environmental Protection Agency reference dose 0.7 µg/kg body weight per day) [17, 20].

Perchlorate exposure leads to increased urinary iodine excretion due to the blocking of thyroidal uptake [23] and perchlorate has been found in mothers' and cows' milk, generating a risk for babies and children with inadequate iodide supply. A longitudinal study in pregnant women exposed to different perchlorate concentrations in drinking water during pregnancy and lactation revealed no changes in thyroid status and function or adverse effects in mothers and newborns [24], but several recent studies report on changes in maternal thyroid status and effects on their offspring [19–22, 25].

Exposure of tadpoles and adult African clawed frog Xenopus laevis to perchlorate impairs amphibian metamorphosis and thyroid function of these model organisms, which might serve as a

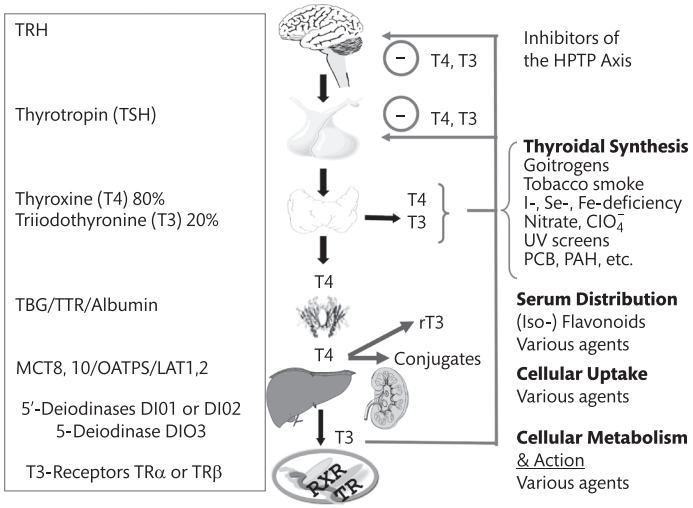


Figure 3.2.2.1 Hypothalamus–pituitary–thyroid–periphery (HPTP) axis, hormonal feedback regulation, and interference by inhibitors.

PAH, polycyclic aromatic hydrocarbons; PCB, polychlorinated biphenyls; RXR, retinoid X receptor.

Table 3.2.2.2 Synthesis of thyroid hormones by follicular thyrocyte epithelial cells, storage of iodinated thyroglobulin in colloidal space and secretion of thyroid hormones is affected by environmental and nutritive agents

Reaction contributing to thyroid hormone biosynthesis	Interfering compound
Basolateral iodide uptake by NIS	Perchlorate, (iso-) thiocyanate; 4-MBC
Apical export by pendrin (PDS)	?
Synthesis and apical secretion of Tg	?
Synthesis and apical insertion of TPO and DUOX	Iron, ?
NADPH-dependent production of H_2O_2 by DUOX	?
Iodide oxidation, iodination of Tg-tyrosyl residues	Goitrogens, goitrin
Coupling of Tg iodotyrosine residues to iodothyronines is catalyzed by TPO using H_2O_2 as cosubstrate	BP-2–4
Polymerization and deposition of iodinated Tg in colloid	?
Micropinocytosis, reduction, and proteolysis of Tg in secondary lysosomes	?
Release of thyroid hormones T_4 and T_3 into the blood by the transporter MCT8	Silychristin
Dehalogenation of DIT and MIT and re-utilization of iodide for thyroid hormone biosynthesis	3-nitro-L-tyrosine (MNT)
Secretion of pGPx (GPx-3) into the colloidal space for degradation of excess H_2O_2	Selenium, ?

DIT, diiodotyrosine; DUOX, dual oxidase; MIT, monoiodotyrosine; NIS, sodium-iodide symporter; PDS, pendrin; Tg, thyroglobulin; TPO, thyroid peroxidase.

very sensitive biomarker for monitoring purposes of several compensatory and also adverse effects caused by environmental exposure to goitrogens such as perchlorate and others; interspecies differences for adverse effects in amphibians, rodents and humans have not been ruled out [26]. These studies indicate that aquatic life forms might already be affected by environmental agents, while humans and terrestrial animals might still be able to compensate or adapt to some extent as long as the concentrations of goitrogens are not excessive. Issues related to extrapolations of rodent studies with perchlorate for human iodide and TH kinetics and possible risk assessments have been extensively studied and discussed [27].

Nitrate

A continuously increased world population requires enhanced efforts for production of sufficient food and this is achieved by enhanced use of nitrogen containing fertilizers in agricultural production. Nitrate and nitrite contamination of ground, surface, and drinking water as well as many food products, especially vegetables, is the downside of this development. Nitrite and nitrate are also widely used as preservatives for fish and meat. Nitrate efficiently interferes with NIS catalysed iodide uptake and represents a relevant goitrogen especially in children exposed to drinking water containing 100 mg/L or more nitrate [28]. In highly contaminated or nutritionally exposed areas, the goitrogenic effects of nitrate/nitrite cannot be neglected, but adequate iodide supply might prevent such an adverse effect and subchronic exposure (15 µg/kg body weight for 28 d in volunteers) appears to be tolerable in humans with respect to thyroid function [29]. Whether nitrate also directly interferes with TPO or thyroid oxidase (DUOX) is unclear.

Nitric oxide (NO, identified as prostacyclin and endothelium-derived hyperpolarizing factor) is a powerful signalling molecule activating guanylate cyclase and cGMP production in the vascular system and thyrocytes and acts as a potent inhibitor of TH synthesis and function [30]. Whether pharmaceuticals generating NO, which is also endogenously produced in the thyroid by NO synthase isoenzymes, have adverse or therapeutic effects on thyroid function in hyperthyroidism remains to be analysed [30]. Adverse nitrate effects on maternal thyroid function during pregnancy as well as on size of newborns has been reported in studies, which included relatively limited numbers of pregnant women (week 12, n = 284) [22] or mother–child pairs (n = 107) exposed to several agents [25].

Goitrin, Thiocyanate, and Smoking

This voluminous anion and its related structural isomer isothiocyanate are both potent iodide competitors for NIS, at higher concentrations thiocyanate also inhibits TPO by acting as pseudosubstrate. Both agents are formed by metabolic pathways from cyanogenic glucosides or thioglucosides from plant origin, respectively. Plant and (bacterial) glucosidases in the gut cleave these glucosides, and release cyanide, which is converted to thiocyanate, or isothiocyanate. 'Goitrin' (L-5-vinyl-2-thio-oxazolidone), isolated from yellow turnips and from *Brassica* seeds, is a potent antithyroid compound and thiocyanate is also endogenously released from linamarin, a cyanogenic glucoside present particularly in the tuberous roots of the staple food cassava [31]. Goitrin, a goitrogen as potent as propylthiouracil (PTU), is not degraded like thioglycosides. Relevant sources for such goitrogens are the *Brassicacea* (e.g. cabbage, broccoli, cauliflower, Brussels sprouts), *Cruciferaceae, Compositae*, and *Umbelliferae*, but food content of these adverse metabolites strongly depends on adequate processing by cooking, hydrolysis, and preservation. Exposure to these goitrogens, monitored by urinary excretion of (iso-) thiocyanate, is a major problem in developing countries, where inadequate economic and social life conditions or energy resources prevent correct processing of these staple foods such as cassava, sweet potatoes, lima beans, sorghum, pearl millet, and corn representing the main source for carbohydrates. Especially, concomitant iodide deficiency

and protein malnutrition leading to inadequate iron supply can exaggerate this problem for risk groups such as pregnant and lactating women, infants, children, and adolescents. In the thyroid thiocyanate is metabolized to sulphate and thus does not accumulate. Probably, thiocyanate and goitrin in plant derived nutrients may play a minor role in impairing thyroid function in societies with adequate iodine intake [32], but overzealous unbalanced human vegetarian diets might cause adverse effects well known also from life stock production.

Thiocyanate and isothiocyanate exposure is also of relevance in western countries as recently documented [33] for nutritional sources but especially also for tobacco smokers, which inhale significant amounts of these goitrogens together with other adverse agents. Up to fourfold increased thiocyanate concentrations were determined in breast milk of breast feeding mothers who smoked and this was associated with up to two fold decreases in iodide content, the combination of which amplified the goitrogenic risk for the baby [34]. Further adverse constellations might be observed if nutritional and environmental exposures to combinations several of these goitrogens add to or amplify the risk and potential damage to the TH homeostasis. Altered maternal and fetal thyroid function has been reported for mothers who smoke [35] and smoking is one of the strongest risk factors for progress and severity of Graves' disease. Conversely smoking reduces anti-TPO and antithyroglobulin antibody titres and might reduce incidence of Hashimoto's thyroiditis by mechanisms not understood so far [36, 37]. Smoking is also a risk factor for goitrogenesis, even under improved iodide supply [36–39].

Tolerable exposure limits recommended by environmental and health authorities and goitrogen contents of nutrients and foods strongly vary even by orders of magnitude in different regions, countries, and continents of our globe. As several of these environmental or nutritive exposures cannot be modified by exposed individuals it is more than necessary to warrant adequate iodide intake for the whole population, and especially for risk groups [17, 19, 22, 32, 40–42].

Gaitan *et al.* have reported on the occurrence of small aliphatic disulphides (R-S-S-R; R = methyl-, ethyl-, *n*-propyl, phenyl-), which are goitrogens inhibiting iodide organification catalysed by TPO. These compounds occur in some vegetables (onion and garlic), well water, sedimentary rocks and as water contaminants in aqueous effluents from coal-conversion processes [43]. Humic acids, another coal or plant origin contaminant of well and drinking water, have also been identified as goitrogens, but again their effect is only observed when there is inadequate iodide supply, at least in animal experimental models [44]. Iodine may be bound to such humic substances in drinking water from natural wells but seems to be bioavailable [45].

Environmental Chemicals

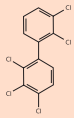
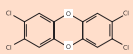
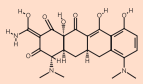
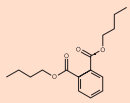
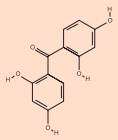
The remarkable progress of worldwide industrialization, expanded and intensified agricultural and (semi-)industrial production of nutrients and food, and the tremendous increase in quality of life associated with longevity had and continues to have a major impact on our environment, which raises several concerns. Especially the synthesis, use, and dissemination of ten-thousands of new chemicals and compounds, some of which are produced at high tonnage worldwide and are persistent chemicals, have introduced new agents into our environment some of which interfere with the hormone system, including the thyroid. Several candidate agents with relevance to the HPTP axis have been identified from effects in wildlife including aquatic life forms and for some compounds adverse effects on thyroid morphology, structure, function, and TH status have already been described. Among those are polychlorinated (polychlorinated biphenyls (PCBs), hexachlorobenzene (HCB), organochlorines) and polybrominated (polybrominated diphenyl ethers (PBDE)) aromatic and phenolic (resorcinol, bisphenol A (BPA)) compounds, chlorinated furans such as dioxin derivatives, polyphenolic hydrocarbons, phthalates, pyridines, and others (see **Table 3.2.2.3** for selected compounds and their main characteristics).

The Seveso accident in 1976 releasing highly persistent 2,3,7,8-tetrachloro-dibenzo-p-dioxin (TCDD) created a first caesura, shortly followed by Bhopal, India in 1984, Basel 1986, and several other accidents, which raised awareness of the issue of endocrine disrupting compounds (EDC) which have not only immediate toxic effects but might act in a transgenerational way including epigenetic mechanisms of action (**Box 3.2.2.2**).

Maternal exposure to TCDD and related compounds in the Seveso area has been linked to markedly elevated neonatal thyroid-stimulating hormone (TSH) blood levels in the inner versus the marginally affected and the control area [46] in the affected offspring after the accident. This still controversial data, although correlated with current plasma concentrations of TCDD and coplanar dioxin-like compounds, suggests long-lasting impact of such contamination both for the immediately exposed population and for the subsequent generation. A follow-up analysis of women in premenarche at the accident later on observed an inverse association between serum tetrachlorodibenzo-p-dioxin (TCCD) and total T_4 concentrations measured shortly after the accident but not the other TH parameters analysed in 2008 [47]. Elevated neonatal TSH is a well-accepted biomarker for fetal hypothyroidism and used in the worldwide highly successful screening programmes for congenital hypothyroidism. Increased thyroid volume, elevated prevalence of antibodies against TPO and the TSH receptor associated with impaired fasting glucose, was also reported for the young adult offspring of mothers exposed and living in a highly polluted area in eastern Slovakia, where a mix of organochlorines (PCBs, dichlorodiphenyldichloroethylene (DDE), HCB) can still be detected in the exposed inhabitants, albeit at lower levels in the young adults compared to their parents, but still higher than in a control region [48, 49]. Again, these data were discussed in the context of a transgenerational adverse effect on the HPTP and other endocrine feedback circuits, and convincing evidence for an altered HPTP axis seems obvious for the total exposed population of that region.

A major human biomonitoring programme was initiated in 2002 in a heavily industrialized and populated area in Flanders, Belgium. Results from this carefully conceived analysis, including internal exposure to various EDC and agents, revealed among other alterations of hormonal parameters, effects on the serum TH levels (e.g. lowered TSH, and elevated free triiodothyronine) [50]. TH serum profiles were also found to be altered in adolescent

Table 3.2.2.3 Representative examples of nutritive and environmental agents interfering with the hypothalamus–pituitary–thyroid-periphery axis (HPTP axis)

Chemical structure	Compound name	IUPAC	MW	Typical use
	Goitrin, DL-goitrin	5-ethenyl-1,3-oxazolidine-2-thione	129	Contained in food, goitrogen
	Arochlor 1254	1,2,3-trichloro-4-(2,3-dichlorophenyl)benzene	326	Antithyroid agent, pesticide
	Tetradioxin; dioxin	2,3,7,8-tetrachlorooxanthrene	322	Insecticide, teratogen
	Minocycline	(2Z,4S,4aS,5aR,12aS)-2-[amino(hydroxy)methylidene]-4,7-bis(dimethylamino)-10,11,12a-trihydroxy-4a,5,5a,6-tetrahydro-4H-tetracene-1,3,12-trione	457	Antibacterial agent
	Bisphenol A; diphenylolpropane	4-[2-(4-hydroxyphenyl)propan-2-yl]phenol	228	'Free radical scavenger'
	Dibutyl phthalate	dibutyl benzene-1,2-dicarboxylate	278	Plasticizer
	Benzophenone-2; Uvinul D-50	bis(2,4-dihydroxyphenyl) methanone	246	UV screen
	Enzacamene; Neo Heliopan MBC; Eusolex 63	(3E)-1,7,7-trimethyl-3-[(4-methylphenyl)methylidene]bicyclo[2.2.1]heptan-2	254	UV screen

from the Akwesasne Mohawk Nation living in a PCB exposed area [12]. Different relationships were observed for different PCB congeners, HCB, and DDE versus TSH and free thyroxine (T_4) levels, with breast milk feeding modulating these interactions. Although only a small group of adolescents had been analysed, the authors interpret their findings as evidence for prenatal impact of exposure to these EDC on long-lasting alterations of the setpoints of the HPTP axis, in agreement with several other recent studies in exposed regions. Apparently small amounts of selected persisting EDC present during fetal, postnatal, and pubertal development might lead to adverse effects on the HPTP axis via epigenetic mechanisms. Whether these mechanisms manifest only in certain subpopulations, susceptible or genetically predisposed subgroups or individuals remain to be studied in more detail. Nevertheless, even subtle alterations of the HPTP axis will have major impact on brain development, IQ, long-term metabolic and age-related disease risks during life-course due to the pleiotropic nature of TH action and feedback regulation (Box 3.2.2.3) [51–53].

Recently, first results from the large Hokkaido Birth Cohort Study on Environment and Children's Health revealed complex associations of prenatal exposure to PCDD/PCDFs and PCBs with maternal and infant TH concentrations in multiple regression analysis with some hints towards elevated neonatal free T_4 in boys [54] but no information on developmental outcome is yet available. Several of these agents are known to activate aryl hydrocarbon receptor-mediated gene expression (e.g. of various cytochrome P450 enzymes).

As there are trends for direct relationships between blood, tissue, adipose tissue, whole body or breast milk contents for TCDD,

Box 3.2.2.2 Mechanisms of adverse action of environmental and nutritive agents interfering with hypothalamus–pituitary–thyroid–periphery axis

Thyroid hormone synthesis, secretion, distribution, transport, uptake, metabolism, action

Known targets:

Thyroid: TSHr, NIS, TPO, DUOX, dehalogenase

Serum: transthyretin, albumin, TBG, lipoproteins

Target cells: transmembrane transporters, deiodinases, conjugating enzymes, T3 receptors

Possible targets: TRH, TRH receptor, TRH degrading ectoenzyme, TSH, dehalogenase, cathepsins, cellular uptake systems: MCT8, MCT10, OATP14, LAT2

Known mechanisms: direct competition (reversible, irreversible) for TH protein binding sites, inactivation of essential protein components of the HPTP axis (e.g. heavy metals, toxins)

Probable mechanisms: epigenetic effects

Age- and life phase-dependent actions: developmental, teratogenic, *in utero*, pregnancy, lactation,

Box 3.2.2.3 Current research concepts and paradigms for analysis of endocrine disruptor-like effects of hypothalamus–pituitary–thyroid–periphery axis

- Convergence of EDC effects on neural and endocrine targets in hypothalamus
- Neuroendocrine regulation and organizational units of hypothalamus
- Timing of EDC exposure is key to its ultimate effects, windows of susceptibility, life-course specific effects
- Transgenerational effects of EDC: both personal and parental exposure is relevant
- Analysis of subpopulations with high accidental or occupational exposure or (genetically) predisposed vulnerability
- Epidemiological analysis of consumer and occupationally exposed groups
- Impact of subtle thyroid axis alterations on pre- and postnatal development and long-term ageing-associated risks

polychlorinated dibenzofurans (PCDF), PCB congeners, and other related EDC to impaired thyroid function especially in babies, children and young adults [54–58], there is not only need for monitoring and further research of potential long-term damage, but again the urgent necessity to warrant adequate iodide intake in these areas. This precaution might reduce the risk for the HPTP axis by EDC exposure and contamination. Even under such conditions breast feeding should be considered, provided the mother adapts her iodine intake not only to pregnancy and lactation but also to her elevated EDC contamination, transferred into milk.

PCBs are environmental persistent, some of their more than 200 congeners show bioaccumulation in adipose tissue and exhibit high structural similarity to TH. This is reflected by their strong competition of TH binding to serum transthyretin but also significant competition for T_3 binding to T_3 receptors as demonstrated by *in vitro*, cellular, and intact animal experimental models. Whether these mechanistically plausible effects, which are however associated with divergent findings on serum TH status of affected humans, will also have impact on functionally relevant readouts and biomarkers, remains to be analysed in long-term studies in larger cohorts.

Recently several highly sensitive, powerful, and sophisticated high-throughput *in vitro* screening systems have been established, validated, and are currently used in research. They are also applied for biomonitoring by environmental authorities and allow for detailed analysis of terrestrial and aquatic environment, food components, nutrition, and occupational exposure with respect to endpoints of interference of EDC with TH synthesis, metabolism and action [26, 51, 59–62].

The 'xenoestrogen' BPA, recently receiving much scientific and public attention among the most controversial compounds, is a relevant antagonistic ligand for T_3 receptor and might affect modulation of T_3-responsive genes. BPA, an agent used as plasticizer in daily life articles from polycarbonate baby bottles, food can inner linings, to cosmetic and dental products, is currently intensively analysed as a potent EDC not only for the thyroid but also the hypothalamus–pituitary–gonadal (HPG) axis [13, 63–65]. Some companies stopped already using the compound in baby products. Animal experimental studies suggest immediate and also transgenerational BPA effects including interference with sex differentiation. As long as only few accepted data exist on critical BPA leakage and human exposure levels during various life phases caution should be taken in further expanding the use of this compound in human daily life. BPA has clear adverse effects on several components of the HPTP axis in experimental *in vitro* and *in vivo* models and is a powerful EDC inhibiting several T_3-regulated pathways in vertebrate development as analysed in the excellent premetamorphic *Xenopus laevis* model [66]. These observations add complexity to the analysis of adverse effects and necessary risk assessment because BPA in the scientific and public discussion was mainly considered to be a 'xenoestrogenic' compound with impact on development, differentiation, and function of the HPG axis. Findings like this and related observations of EDC affecting more than one endocrine axis with rather distinct developmental windows of susceptibility to even very low doses of the compounds led to new initiatives, motions and approaches how to analyse such EDC effects that will probably rarely be detected with classical approaches of toxicology focusing on serum parameters, morphology, linear dose-response relationships and toxicological endpoints [63]. It should not be forgotten that relevant species differences for the HPG axis require careful analysis of potential human impact of findings in non-human *in vitro* and animal experimental models. BPA at human relevant concentrations interferes with TH metabolism in pregnant ewes, which represent a valuable model for interference of EDC with maternal–fetal TH communication and more detailed analysis of mechanism involved [67]. Associations between BPA exposure and altered TH parameters or tissue structure have been reported for Chinese school children [68], adult or pregnant women [69–71] and their offspring in studies analysing EDC exposure at single or repeated time points, but mainly single agents such as BPA and not complex EDC mixtures have been analysed.

Similar considerations apply for the various phthalates in use which have already created a worldwide significant exposure level in humans. Here only very few data have been collected for their interference with the HPTP axis, but most studies indicate relevant interference with TH levels in children, pregnant women, and adult individuals [13, 71–74]. The detailed mode of action remains unclear so far as the wide number of phthalate congeners and their metabolites poses major analytical problems for clear cause–effect analysis.

Also, for another group of persistent organohalogen pollutants, the perfluorinated compounds, which are markedly enriched in aquatic food chain, interference with the HPTP axis has been shown in environmental, nutritional, and occupational exposure analysis. Perfluorinated compounds are structurally related to free fatty acids and thus bind to albumin in the blood, thereby competing with TH and interfering with TH bioavailability to target tissues. While at high occupational exposures altered TH serum parameters were reported so far no clinical evidence for disturbed TH status in humans is evident [53, 75–77]. However, as exposure to perfluorinated agents increases globally this issue will remain on the agenda.

Pharmaceuticals and Drugs

Thyroid is a Sensitive Target of Side Effects of Various Drugs

Many drugs are known to interact with the thyroid gland or with components involved in the function and regulation of the HPTP axis. In toxicology departments of pharmaceutical industries, the thyroid gland is a well-known problematic target for adverse or toxic side effects of new pharmaceuticals. It is estimated that up to one-third of

newly developed compounds, especially aromatic and polycyclic compounds, fail the acute or chronic toxicity screening test batteries due to their side effects leading to alterations of thyroid morphology, goitrogenesis, development of thyroid tumours or merely to changes of serum TSH and/or TH concentrations. The reason for this is not completely understood, but the permanent, lifelong H_2O_2 production by thyroid follicles catalysed by the NADPH-dependent DUOX, as well as peroxide consumption by TPO for iodide oxidation, organification, and TH synthesis on the thyroglobulin scaffold might be the major cause. Compounds accumulating in the thyroid and its luminal colloid might be exposed to H_2O_2, be chemically modified by oxidative processes, and be deposited there or damage the follicles. One illustrative example might be the incidentally rare observation of 'black thyroid' syndrome, which is a tetracycline (especially minocycline)-induced asymptomatic discolouration of the thyroid gland probably related to TPO-induced oxidation of the tetracycline [78].

Benzofurans

The powerful benzofuran drugs amiodarone and dronedarone are widely used for treating resistant tachyarrhythmia. Apart from their target molecules in the heart, these drugs, the active metabolite desethylamiodarone and other derivatives are potent antagonistic ligands for the T_3 receptor and inhibitors for the Dio enzymes (Figure 3.2.2.2). However, dibutyl-dronedarone acts as a selective T_3 receptor α1 antagonist (see Chapter 3.3.12). Therefore, chronic administration leads to impaired thyroid function with a clear cumulative dose associated increase in risk. The drugs are substantially accumulated in the thyroid, which in addition to the high iodine content of amiodarone (between 3 and 20 mg iodide are released per day into the blood) explains their prominent disturbance of thyroid function [79, 80]. Not only the iodide contamination associated with the administration of the drug but also the thyroid accumulation of the drug might lead to the severe structural defects of the gland and follicles seen in some patients treated with amiodarone [80, 81]. Therefore, the new iodine-free alternative dronedarone raises high interest as no comparable thyroid-related effects are reported such as inhibition of deiodinase, binding to T_3 receptor, or thyroid accumulation. The adverse effects with respect to iodide contamination of other iodinated drugs such as the iodinated oral bile X-ray contrast agents iopanoate and its congeners [82], will be discussed elsewhere in this volume.

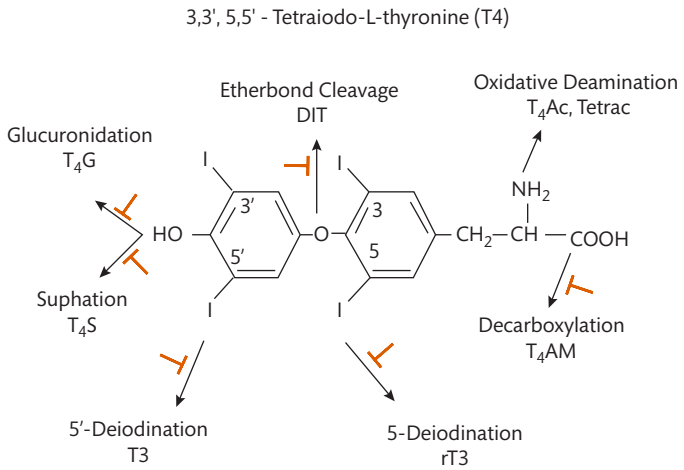


Figure 3.2.2.2 Pathways of thyroxine (T4) metabolism known to be affected by nutritional or environmental agents ⊣.

UV Screens in Cosmetics and Daily Life Products

A further representative example of a group of EDC with relevance for the HPTP axis are widely used UV screens, filters, or absorbers. These are ubiquitous components of various plastic materials of our daily life which have to be protected from UV damage; they are also contained in sunscreens and various cosmetics such as lip sticks or body lotions. Octyl methoxycinnamate (OMC) is one of the most commonly used filters. UV screens may contain up to 10 % (w/w) of typical compounds such as 4-methylbenzylidene camphor (4-MBC) and benzophenones 1–4 and related products. Typical administration of the UV filters leads to measurable serum concentrations in the submicromolar to micromolar range [83]. At these concentrations clear adverse effects have been observed in thyroid-related *in vitro* and *in vivo* animal models, such as rapid and dose-dependent goitrogenesis in rats after 4-MBC administration or efficient inhibition of TPO by benzophenone 2 [5]. Some of these effects might be prevented or at least attenuated by adequate iodide supply which still is not warranted globally. Considering the increased application of these UV screens not only for product protection but also for prevention of human skin cancer due to higher exposure to UV irradiation associated with ozone loss in the atmosphere, some of these UV filters might impose marked risks for the adequate function of the HPTP axis. This might apply especially for babies and children, whose skin is more sensitive to UV light, less protected by endogenous melanocytes and therefore more frequently treated with these dermal UV lotions. Also, these products might be even more easily absorbed by young skin. So far no clear evidence for a goitrogenic action of UV filter ingredients has been described in humans. Therefore, the advice might be to guarantee and adequate iodide supply, and to protect skin from UV irradiation by avoiding too much sun and applying UV screens that have less risk for interference with the HPTP axis especially in babies, children, and individuals with sensitive skin. However, human biomonitoring assessment values, recently updated for such compounds [84] and used as references for safety considerations and risk assessment, remain controversial [85, 86].

Heavy Metals and Thyroid

Environmental contamination by heavy metals and their ions has raised public concern based on their direct effects on several tissues and organs. Both accidentally and occupationally exposed subgroups are affected, and significant adverse effects might result in the central nervous system (CNS) during development and with respect to the pathogenesis of neurodegenerative, ageing-associated diseases. Whether the TH system, known to have major impact on proper brain development and function in children and adults, is directly involved in these processes remains unclear. As thyrocytes exhibit a highly active redox-regulated cellular metabolism [87], impairment of reactive redox centres of enzymes and other thyrocyte proteins such as metallothioneins by heavy metals will create problems for TH synthesis and secretion. Therefore, environmental or occupational exposure to high mercury, lead, and cadmium concentrations has been associated with altered thyroid homeostasis. For example, a gender-specific effect on increased serum TSH, correlated with increased hair and blood mercury concentrations in males, has been reported in lakeside communities of Quebec and is associated with consumption of contaminated lake fish from

exposed environment [88]. Divergent reports have been published on the relationship between cadmium exposure and TSH, positively associated in several studies, but inversely related in a pilot study in cord blood of Japanese newborns [89]. Animal experimental data clearly suggest adverse effects of cadmium exposure, which interferes with both TH synthesis and peripheral Dio1 activity.

Adverse effects of lead on the thyroid axis have been reported. Previous environmental lead sources were lead-enhanced gasoline and lead-based paints, but both of these sources are of decreasing relevance due to bans that have been enforced in most countries, while contaminations by cadmium, mercury, and, recently, platinum leaking into environment from car exhaust catalytic converters are tending to further increase. The evidence that manganese, essential constituent of several redox-relevant enzymes such as manganese-superoxide dismutase, may directly or indirectly affect thyroid function by injuring the thyroid gland or dysregulating dopaminergic modulation of TH synthesis and thus contributing to altered TH homeostasis and neurodegenerative diseases has recently been reviewed [90].

On the other hand, adequate selenium supply can efficiently counteract the adverse effects of several heavy metal cations such as cadmium, mercury, lead, and vanadium and thus avoid their age-related neurotoxicity [91, 92]. It may be plausible that many adverse effects claimed to be caused by heavy metal exposure might be only occurring indirectly because most heavy metal ions very efficiently inactivate selenocysteine-containing proteins, which are pivotal in cellular and systemic redox regulation [93] as well as in TH synthesis and metabolism. Apparently Se leads to their accumulation or deposit in a presumably non-toxic complex in the brain, kidney, and several other tissues.

Many nutritional or environmental contaminants exhibit their goitrogenic potential only under conditions of inadequate maternal, fetal or neonatal iodide (selenium, and iron) supply [94–96]. Therefore, comprehensive nutritional iodide supplementation is one of the most efficient preventive measures to avoid impaired and delayed development of humans and other higher life forms.

Environmental Temperature

Temperature, light, circadian and circannual rhythms, altitude, latitude, and extreme environmental life conditions are well known to influence TH, energy, thermoregulatory and metabolic homeostasis in free living animals (homeotherms, hibernators or aestivators such as bears) but also in humans and livestock adapted to modern housing conditions. Nevertheless, there exist clear circadian and circannual rhythms for TSH and, delayed in phase, for free T_3 in human serum, while T_4, tightly bound to its four serum distributor proteins (thyroxine-binding globulin (TBG), transthyretin, albumin, and lipoproteins) shows no significant circadian or circannual variation [97–99].

Lowest TSH values are observed in spring and summer and increases by 25% are seen in autumn not reflected by T_4 variations and not related to iodine intake. It has to be kept in mind that the TSH response curve is exponentially related to linear changes in TH serum concentration. Whether alterations in food intake, enhanced sympathetic tone, and adrenergic stimulation of thermogenesis, altered contribution of thermogenesis by uncoupling protein(s) activation in skeletal muscles mediated by fatty acids and bile acid metabolites or neuroendocrine hypothalamic adaptations are contributing to these changes remains to be studied. During prolonged stay in arctic environments enhanced TH secretion by the thyroid has been documented, indicated by increased serum thyroglobulin and elevated T_3 production and turnover, reflected by decreased total and free T_3, but accompanied by unchanged total and free T_4 and TBG. This constellation has been termed 'polar T3 syndrome' and related constellations can be found under extreme physical exercise, endurance training, etc. Some of the changes might be prevented by increased calorie intake, sleep adaptation or TH treatment [100].

Increased environmental temperature during summertime but also elevated body temperature during febrile conditions are associated with lower TSH and serum T_3 levels [101, 102].

Various short- and long-term adaptations of TH secretion, turnover, serum levels and feedback set points have been observed in studies examining the HPTP axis in people at high altitude, but results were controversial and might be confounded by other altered factors such as nutritional profiles, physical activity, light, sleep rhythm, and altered time zone adaptations. Animal experimental simulations could dissociate between distinct effects of high altitude and hypoxia and suggest powerful adaptations of the TH axis, characterized by decreased TH synthesis and secretion but elevated serum free TH [7]. Recently, the role of deiodinase 2, locally generating T3 and activating brown adipose tissue or browning 'beige' fat in humans, has attracted major interest in the context of regulation of body weight by stimulation of expression of uncoupling protein(s), thermogenesis and fatty acid oxidation [103].

ACKNOWLEDGEMENTS

This work has been supported by grants of the Deutsche Forschungsgemeinschaft (DFG) to J. Köhrle (DFG Priority Programme ThyroidTransAct Ko 922/17-1/2).

REFERENCES

1. Dobson JE. The iodine factor in health and evolution. *The Geographical Review*, 1998; **88**: 1–28.
2. Kraiem Z. Thoughts on the role of iodine in the emergence of modern humans. *Thyroid*, 2001; **11**: 807–8.
3. Stenzel D, Huttner WB. Role of maternal thyroid hormones in the developing neocortex and during human evolution. *Front Neuroanat*, 2013; **7**: 19.
4. Crockford SJ. Evolutionary roots of iodine and thyroid hormones in cell-cell signaling. *Integr Comp Biol*, 2009; **49**: 155–66.
5. Köhrle J. Environment and endocrinology: the case of thyroidology. *Ann Endocrinol (Paris)*, 2008; **69**: 116–22.
6. Leung AM, Braverman LE. Consequences of excess iodine. *Nat Rev Endocrinol*, 2014; **10**: 136–42.
7. Sarne D, Feingold KR, Anawalt B, *et al*. Effects of the environment, chemicals and drugs on thyroid function. In: Sarne D, ed. *Thyroid Disease Manager.* Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc., 2016: Chapter 5: 1–59.
8. Verheesen RH, Schweitzer CM. Iodine deficiency, more than cretinism and goiter. *Medical Hypotheses*, 2008; **71**: 645–8.
9. Kopp W. Nutrition, evolution and thyroid hormone levels—a link to iodine deficiency disorders? *Med Hypotheses*, 2004; **62**: 871–5.
10. Zhang X, Ho SM. Epigenetics meets endocrinology. *J Mol Endocrinol*, 2011; **46**: R11–32.
11. Martinez ME, Duarte CW, Stohn JP, *et al*. Thyroid hormone influences brain gene expression programs and behaviors in later

generations by altering germ line epigenetic information. *Mol Psychiatry*, 2018. DOI: 10.1038/s41380-018-0281-4.
12. Schell LM, Gallo MV, Ravenscroft J, DeCaprio AP. Persistent organic pollutants and anti-thyroid peroxidase levels in Akwesasne Mohawk young adults. *Environ Res*, 2009; **109**: 86–92.
13. Gore AC, Chappell VA, Fenton SE, *et al*. EDC-2: the Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*, 2015; **36**: E1–E150.
14. van Bakel MM, Printzen G, Wermuth B, Wiesmann UN. Antioxidant and thyroid hormone status in selenium-deficient phenylketonuric and hyperphenylalaninemic patients. *Am J Clin Nutr*, 2000; **72**: 976–81.
15. Schug TT, Janesick A, Blumberg B, Heindel JJ. Endocrine disrupting chemicals and disease susceptibility. *J Steroid Biochem Mol Biol*, 2011; **127**: 204–15.
16. Ravera S, Reyna-Neyra A, Ferrandino G, *et al*. The sodium/iodide symporter (NIS): molecular physiology and preclinical and clinical applications. *Annu Rev Physiol*, 2017; **79**: 261–89.
17. De Groef B, Decallonne BR, Van der Geyten S, *et al*. Perchlorate versus other environmental sodium/iodide symporter inhibitors: potential thyroid-related health effects. *Eur J Endocrinol*, 2006; **155**: 17–25.
18. Srinivasan A, Viraraghavan T. Perchlorate: health effects and technologies for its removal from water resources. *Int J Environ Res Public Health*, 2009; **6**: 1418–42.
19. McMullen J, Ghassabian A, Kohn B, Trasande L. Identifying subpopulations vulnerable to the thyroid-blocking effects of perchlorate and thiocyanate. *J Clin Endocrinol Metab*, 2017; **102**: 2637–45.
20. Steinmaus CM. Perchlorate in water supplies: sources, exposures, and health effects. *Curr Environ Health Rep*, 2016; **3**: 136–43.
21. Rubin R, Pearl M, Kharrazi M, *et al*. Maternal perchlorate exposure in pregnancy and altered birth outcomes. *Environ Res*, 2017; **158**: 72–81.
22. Horton MK, Blount BC, Valentin-Blasini L, *et al*. CO-occurring exposure to perchlorate, nitrate and thiocyanate alters thyroid function in healthy pregnant women. *Environ Res*, 2015; **143**(Pt A): 1–9.
23. Braverman LE. Clinical studies of exposure to perchlorate in the United States. *Thyroid*, 2007; **17**: 819–22.
24. Téllez RT, Chacón PM, Abarca CR, *et al*. Long-term environmental exposure to perchlorate through drinking water and thyroid function during pregnancy and the neonatal period. *Thyroid*, 2005; **15**: 963–75.
25. Evans KA, Rich DQ, Weinberger B, *et al*. Association of prenatal perchlorate, thiocyanate, and nitrate exposure with neonatal size and gestational age. *Reprod Toxicol*, 2015; **57**: 183–9.
26. Opitz R, Schmidt F, Braunbeck T, *et al*. Perchlorate and ethylenethiourea induce different histological and molecular alterations in a non-mammalian vertebrate model of thyroid goitrogenesis. *Mol Cell Endocrinol*, 2009; **298**: 101–14.
27. Merrill EA, Clewell RA, Robinson PJ, *et al*. PBPK model for radioactive iodide and perchlorate kinetics and perchlorate-induced inhibition of iodide uptake in humans. *Toxicol Sci*, 2005; **83**: 25–43.
28. Radikova Z, Tajtakova M, Kocan A, *et al*. Possible effects of environmental nitrates and toxic organochlorines on human thyroid in highly polluted areas in Slovakia. *Thyroid*, 2008; **18**: 353–62.
29. Hunault CC, Lambers AC, Mensinga TT, *et al*. Effects of subchronic nitrate exposure on the thyroidal function in humans. *Toxicology Letters*, 2007; **175**: 64–70.
30. Bazzara LG, Velez ML, Costamagna ME, *et al*. Nitric oxide/cGMP signaling inhibits TSH-stimulated iodide uptake and expression of thyroid peroxidase and thyroglobulin mRNA in FRTL-5 thyroid cells. *Thyroid*, 2007; **17**: 717–27.
31. Ermans A-M, Bourdoux P. Antithyroid sulfurated compounds. In: Gaitan E, ed. *Environmental Goitrogenesis*. Boca Raton, FL: CRC Press, 1989: 15–31.
32. Felker P, Bunch R, Leung AM. Concentrations of thiocyanate and goitrin in human plasma, their precursor concentrations in brassica vegetables, and associated potential risk for hypothyroidism. *Nutr Rev*, 2016; **74**: 248–58.
33. Brauer VFH, Below H, Kramer A, *et al*. The role of thiocyanate in the etiology of goiter in an industrial metropolitan area. *Eur J Endocrinol*, 2006; **154**: 229–35.
34. Laurberg P, Nøhr SB, Pedersen KM, Fuglsang E. Iodine nutrition in breast-fed infants is impaired by maternal smoking. *J Clin Endocrinol Metab*, 2004; **89**: 181–7.
35. Shields B, Hill A, Bilous M, *et al*. Cigarette smoking during pregnancy is associated with alterations in maternal and fetal thyroid function. *J Clin Endocrinol Metab*, 2009; **94**: 570–4.
36. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol (Oxf)*, 2013; **79**: 145–51.
37. Pedersen IB, Laurberg P, Knudsen N, *et al*. Smoking is negatively associated with the presence of thyroglobulin autoantibody and to a lesser degree with thyroid peroxidase autoantibody in serum: a population study. *Eur J Endocrinol*, 2008; **158**: 367–73.
38. Andersen SL, Nøhr SB, Wu CS, *et al*. Thyroglobulin in smoking mothers and their newborns at delivery suggests autoregulation of placental iodide transport overcoming thiocyanate inhibition. *Eur J Endocrinol*, 2013; **168**: 723–31.
39. Ittermann T, Schmidt C, Kramer A, *et al*. Smoking as a risk factor for thyroid volume progression and incident goiter in a region with improved iodine supply. *Eur J Endocrinol*, 2008; **159**: 761–6.
40. Zimmermann MB, Jooste PL, Pandav CS. Iodine-deficiency disorders. *Lancet*, 2008; **372**: 1251–62.
41. Lee SY, McCarthy AM, Stohl H, *et al*. Urinary iodine, perchlorate, and thiocyanate concentrations in U.S. lactating women. *Thyroid*, 2017; **27**: 1574–81.
42. Willemin ME, Lumen A. Thiocyanate: a review and evaluation of the kinetics and the modes of action for thyroid hormone perturbations. *Crit Rev Toxicol*, 2017; **47**: 537–63.
43. Gaitan E, Cooksey RC, Legan J, *et al*. Antithyroid and goitrogenic effects of coal-water extracts from iodine-sufficient goiter areas. *Thyroid*, 1993; **3**: 49–53.
44. Huang TS, Lu FJ, Tsai CW, Chopra IJ. Effect of humic acids on thyroidal function. *J Endocrinol Invest*, 1994; **17**: 787–91.
45. Andersen S, Pedersen KM, Iversen F, *et al*. Naturally occurring iodine in humic substances in drinking water in Denmark is bioavailable and determines population iodine intake. *Br J Nutr*, 2008; **99**: 319–25.
46. Baccarelli A, Giacomini SM, Corbetta C, *et al*. Neonatal thyroid function in Seveso 25 years after maternal exposure to dioxin. *PLoS Med*, 2008; **5**: e161.
47. Chevrier J, Warner M, Gunier RB, *et al*. Serum dioxin concentrations and thyroid hormone levels in the Seveso Women's Health Study. *Am J Epidemiol*, 2014; **180**: 490–8.
48. Langer P, Kocan A, Tajtakova M, *et al*. Multiple adverse thyroid and metabolic health signs in the population from the area heavily polluted by organochlorine cocktail (PCB, DDE, HCB, dioxin). *Thyroid Res*, 2009; **2**: 3.
49. Langer P, Kocan A, Tajtakova M, *et al*. Increased thyroid volume, prevalence of thyroid antibodies and impaired fasting glucose in young adults from organochlorine cocktail polluted area: outcome of transgenerational transmission? *Chemosphere*, 2008; **73**: 1145–50.

50. Croes K, Baeyens W, Bruckers L, *et al*. Hormone levels and sexual development in Flemish adolescents residing in areas differing in pollution pressure. *Int J Hyg Environ Health*, 2009; **212**: 612–25.
51. Mughal BB, Fini JB, Demeneix BA. Thyroid-disrupting chemicals and brain development: an update. *Endocr Connect*, 2018; **7**: R160–86.
52. Fini JB, Mughal BB, Le Mével S, *et al*. Human amniotic fluid contaminants alter thyroid hormone signalling and early brain development in Xenopus embryos. *Sci Rep*, 2017; **7**: 43786.
53. Ghassabian A, Trasande L. Disruption in thyroid signaling pathway: a mechanism for the effect of endocrine-disrupting chemicals on child neurodevelopment. *Front Endocrinol (Lausanne)*, 2018; **9**: 204.
54. Baba T, Ito S, Yuasa M, *et al*. Association of prenatal exposure to PCDD/Fs and PCBs with maternal and infant thyroid hormones: the Hokkaido Study on Environment and Children's Health. *Sci Total Environ*, 2018; **615**: 1239–46.
55. Massart F, Meucci V. Environmental thyroid toxicants and child endocrine health. *Pediatr Endocrinol Rev*, 2007; **5**: 500–9.
56. Jacobson MH, Darrow LA, Barr DB, *et al*. Serum polybrominated biphenyls (PBBs) and polychlorinated biphenyls (PCBs) and thyroid function among michigan adults several decades after the 1973–1974 PBB contamination of livestock feed. *Environ Health Perspect*, 2017; **125**: 097020.
57. Lignell S, Aune M, Darnerud PO, *et al*. Maternal body burdens of PCDD/Fs and PBDEs are associated with maternal serum levels of thyroid hormones in early pregnancy: a cross-sectional study. *Environ Health*, 2016; **15**: 55.
58. Li ZM, Hernandez-Moreno D, Main KM, *et al*. Association of *in utero* persistent organic pollutant exposure with placental thyroid hormones. *Endocrinology*, 2018; **159**: 3473–81.
59. Tanida T, Warita K, Ishihara K, *et al*. Fetal and neonatal exposure to three typical environmental chemicals with different mechanisms of action: mixed exposure to phenol, phthalate, and dioxin cancels the effects of sole exposure on mouse midbrain dopaminergic nuclei. *Toxicol Lett*, 2009; **189**: 40–7.
60. Hofmann PJ, Schomburg L, Köhrle J. Interference of endocrine disrupters with thyroid hormone receptor-dependent transactivation. *Toxicol Sci*, 2009; **110**: 125–37.
61. Zoeller RT. Environmental chemicals impacting the thyroid: targets and consequences. *Thyroid*, 2007; **17**: 811–17.
62. Murk AJ, Rijntjes E, Blaauboer BJ, *et al*. Mechanism-based testing strategy using *in vitro* approaches for identification of thyroid hormone disrupting chemicals. *Toxicol In Vitro*, 2013; **27**: 1320–46.
63. Myers JP, vom Saal FS, Akingbemi BT, *et al*. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A. *Environ Health Perspect*, 2009; **117**: 309–15.
64. Vandenberg LN, Maffini MV, Sonnenschein C, *et al*. Bisphenol-A and the great divide: a review of controversies in the field of endocrine disruption. *Endocr Rev*, 2009; **30**: 75–95.
65. Boas M, Feldt-Rasmussen U, Main KM. Thyroid effects of endocrine disrupting chemicals. *Mol Cell Endocrinol*, 2012; **355**: 240–8.
66. Heimeier RA, Das B, Buchholz DR, Shi YB. The xenoestrogen bisphenol A inhibits postembryonic vertebrate development by antagonizing gene regulation by thyroid hormone. *Endocrinology*, 2009; **150**: 2964–73.
67. Guignard D, Gayrard V, Lacroix MZ, *et al*. Evidence for bisphenol A-induced disruption of maternal thyroid homeostasis in the pregnant ewe at low level representative of human exposure. *Chemosphere*, 2017; **182**: 458–67.
68. Wang N, Zhou Y, Fu C, *et al*. Influence of bisphenol A on thyroid volume and structure independent of iodine in school children. *PLoS One*, 2015; **10**: e0141248.
69. Aung MT, Johns LE, Ferguson KK, *et al*. Thyroid hormone parameters during pregnancy in relation to urinary bisphenol A concentrations: a repeated measures study. *Environ Int*, 2017; **104**: 33–40.
70. Andrianou XD, Gängler S, Piciu A, *et al*. Human exposures to bisphenol A, bisphenol F and chlorinated bisphenol A derivatives and thyroid function. *PLoS One*, 2016; **11**: e0155237.
71. Romano ME, Webster GM, Vuong AM, *et al*. Gestational urinary bisphenol A and maternal and newborn thyroid hormone concentrations: the HOME Study. *Environ Res*, 2015; **138**: 453–60.
72. Meeker JD, Altshul L, Hauser R. Serum PCBs, p,p'-DDE and HCB predict thyroid hormone levels in men. *Environ Res*, 2007; **104**: 296–304.
73. Huang PC, Tsai CH, Liang WY, *et al*. Early phthalates exposure in pregnant women is associated with alteration of thyroid hormones. *PLoS One*, 2016; **11**: e0159398.
74. Huang HB, Kuo PL, Chang JW, *et al*. Longitudinal assessment of prenatal phthalate exposure on serum and cord thyroid hormones homeostasis during pregnancy—Tainan birth cohort study (TBCS). *Sci Total Environ*, 2018; **619–20:** 1058–1065.
75. Dufour P, Pirard C, Seghaye MC, Charlier C. Association between organohalogenated pollutants in cord blood and thyroid function in newborns and mothers from Belgian population. *Environ Pollut*, 2018; **238**: 389–96.
76. Shah-Kulkarni S, Kim BM, Hong YC, *et al*. Prenatal exposure to perfluorinated compounds affects thyroid hormone levels in newborn girls. newborns and mothers from Belgian population. *Environ Pollut*, 2018; **238**: 389–96.
77. Ballesteros V, Costa O, Iñiguez C, *et al*. Exposure to perfluoroalkyl substances and thyroid function in pregnant women and children: a systematic review of epidemiologic studies. *Environ Int*, 2017; **99**: 15–28.
78. Taurog A, Dorris ML, Doerge DR. Minocycline and the thyroid: antithyroid effects of the drug, and the role of thyroid peroxidase in minocycline-induced black pigmentation of the gland. *Thyroid*, 1996; **6**: 211–19.
79. Han TS, Williams GR, Vanderpump MP. Benzofuran derivatives and the thyroid. *Clin Endocrinol (Oxf)*, 2009; **70**: 2–13.
80. Nakazawa T, Murata S, Kondo T, *et al*. Histopathology of the thyroid in amiodarone-induced hypothyroidism. *Pathol Int*, 2008; **58**: 55–8.
81. Elnaggar MN, Jbeili K, Nik-Hussin N, *et al*. Amiodarone-induced thyroid dysfunction: a clinical update. *Exp Clin Endocrinol Diabetes*, 2018; **126**: 333–41.
82. Lee SY, Rhee CM, Leung AM, *et al*. A review: radiographic iodinated contrast media-induced thyroid dysfunction. *J Clin Endocrinol Metab*, 2015; **100**: 376–83.
83. Janjua NR, Kongshoj B, Petersen JH, Wulf HC. Sunscreens and thyroid function in humans after short-term whole-body topical application: a single-blinded study. *Brit J Dermatol*, 2007; **156**: 1080–2.
84. Apel P, Angerer J, Wilhelm M, Kolossa-Gehring M. New HBM values for emerging substances, inventory of reference and HBM values in force, and working principles of the German Human Biomonitoring Commission. *Int J Hyg Environ Health*, 2017; **220**: 152–66.

85. Aker AM, Johns L, McElrath TF, *et al.* Associations between maternal phenol and paraben urinary biomarkers and maternal hormones during pregnancy: a repeated measures study. *Environ Int*, 2018; **113**: 341–9.
86. Wang J, Pan L, Wu S, *et al.* Recent advances on endocrine disrupting effects of UV filters. *Int J Environ Res Public Health*, 2016; **13**: pii: E782.
87. Schweizer U, Chiu J, Köhrle J. Peroxides and peroxide-degrading enzymes in the thyroid. *Antioxid Redox Signal*, 2008; **10**: 1577–92.
88. Abdelouahab N, Mergler D, Takser L, *et al.* Gender differences in the effects of organochlorines, mercury, and lead on thyroid hormone levels in lakeside communities of Quebec (Canada). *Environ Res*, 2008; **107**: 380–92.
89. Iijima K, Otake T, Yoshinaga J, *et al.* Cadmium, lead, and selenium in cord blood and thyroid hormone status of newborns. *Biol Trace Elem Res,* 2007; **119**: 10–18.
90. Soldin OP, Aschner M. Effects of manganese on thyroid hormone homeostasis: potential links. *Neurotoxicology*, 2007; **28**: 951–6.
91. Whanger PD. Selenium and the brain: a review. *Nutr Neurosci*, 2001; **4**: 81–97.
92. Hammouda F, Messaoudi I, El Hani J, *et al.* Reversal of cadmium-induced thyroid dysfunction by selenium, zinc, or their combination in rat. *Biol Trace Elem Res*, 2008; **126**: 194–203.
93. Ralston NVC, Raymond LJ. Mercury's neurotoxicity is characterized by its disruption of selenium biochemistry. *Biochim Biophys Acta Gen Subj*, 2018. DOI: 10.1016/j.bbagen.2018.05.009
94. Schomburg L, Köhrle J. On the importance of selenium and iodine metabolism for thyroid hormone biosynthesis and human health. *Mol Nutr Food Res*, 2008; **52**: 1235–46.
95. Köhrle J. Selenium and the thyroid. *Curr Opin Endocrinol Diabetes Obes*, 2015; **22**: 392–401.
96. Zimmermann MB, Köhrle J. The impact of iron and selenium deficiencies on iodine and thyroid metabolism: biochemistry and relevance to public health. *Thyroid*, 2002; **12**: 867–78.
97. Maes M, Mommen K, Hendrickx D, *et al.* Components of biological variation, including seasonality, in blood concentrations of TSH, TT3, FT4, PRL, cortisol and testosterone in healthy volunteers. *Clin Endocrinol (Oxf)*, 1997; **46**: 587–98.
98. Ehrenkranz J, Bach PR, Snow GL, *et al.* Circadian and circannual rhythms in thyroid hormones: determining the TSH and free T4 reference intervals based upon time of day, age, and sex. *Thyroid*, 2015; **25**: 954–61.
99. Russell W, Harrison RF, Smith N, *et al.* Free triiodothyronine has a distinct circadian rhythm that is delayed but parallels thyrotropin levels. *J Clin Endocrinol Metab*, 2008; **93**: 2300–6.
100. Do NV, Mino L, Merriam GR, *et al.* Elevation in serum thyroglobulin during prolonged Antarctic residence: effect of thyroxine supplement in the polar 3,5,3'-triiodothyronine syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 1529–33.
101. Ljunggren JG, Kallner G, Tryselius M. The effect of body temperature on thyroid hormone levels in patients with non-thyroidal illness. *Acta Med Scand*, 1977; **202**: 459–62.
102. Epstein Y, Udassin R, Sack J. Serum 3,5,3'-triiodothyronine and 3,3',5'-triiodothyronine concentrations during acute heat load. *J Clin Endocrinol Metab*, 1979; **49**: 677–8.
103. Broeders EP, Vijgen GH, Havekes B, *et al.* Thyroid hormone activates brown adipose tissue and increases non-shivering thermogenesis—a cohort study in a group of thyroid carcinoma patients. *PLoS One*, 2016; **11**: e0145049. Erratum in: *PLoS One*, 2018; 13: e0209225.
104. Builee TL, Hatherill JR. The role of polyhalogenated aromatic hydrocarbons on thyroid hormone disruption and cognitive function: a review. *Drug Chem Toxicol*, 2004; **27**: 405–24.
105. Johannes J, Jayarama Naidu R, Meyer F, *et al.* Silychristin, a flavonolignan derived from the milk thistle, is a potent inhibitor of the thyroid hormone transporter MCT8. *Endocrinology*, 2016; **157**: 1694–701.
106. Teng X, Shan Z, Li C, Yu X, *et al.* Iron deficiency may predict greater risk for hypothyroxinemia: a retrospective cohort study of pregnant women in China. *Thyroid*, 2018; **28**: 968–75.
107. Ertek S, Cicero AF, Caglar O, Erdogan G. Relationship between serum zinc levels, thyroid hormones and thyroid volume following successful iodine supplementation. *Hormones (Athens)*, 2010; **9**: 263–8.
108. Lazarus JH. Lithium and thyroid. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 723–33.
109. Sapouckey SA, Kassotis CD, Nagel SC, Vandenberg LN. Prenatal exposure to unconventional oil and gas operation chemical mixtures altered mammary gland development in adult female mice. *Endocrinology*, 2018; **159**: 1277–89.

3.2.3 Iodine Deficiency Disorders

Michael B. Zimmermann

Dietary Sources and Metabolism

Iodine (atomic weight 126.9 g/atom) is an essential component of the hormones produced by the thyroid gland. Thyroid hormones, and therefore iodine, are essential for mammalian life [1]. The native iodine content of most foods and beverages is low, and most commonly consumed foods provide 3–80 µg per serving [1]. Major dietary sources of iodine in the US and Europe are bread and milk [2]. Boiling, baking, and canning of foods containing iodized salt cause only small losses (≤10%) of iodine content. Iodine content in foods is also influenced by iodine-containing compounds used in irrigation, fertilizers, livestock feed, dairy industry disinfectants, and bakery dough conditioners. Recommendations for iodine intake by age and population group [3] are shown in Table 3.2.3.1.

Table 3.2.3.1 Recommendations for iodine intake (μg/day) by age or population group

Age or population group[a]	U.S. Institute of Medicine [4]	Age or population group[c]	World Health Organization [3]
Infants 0–12 months[b]	110–130	Children 0–5 years	90
Children 1–8 years	90	Children 6–12 years	120
Children 9–13 years	120		
Adults ≥14 years	150	Adults >12 years	150
Pregnancy	220	Pregnancy	250
Lactation	290	Lactation	250

[a] Recommended daily allowance. [b] Adequate intake. [c] Recommended nutrient intake.

Iodide is rapidly and nearly completely absorbed (>90%) in the stomach and duodenum [4]. Iodate, widely used in salt iodization, is reduced in the gut and absorbed as iodide. Thyroid clearance of circulating iodine varies with iodine intake: in conditions of adequate iodine supply, ≤10% of absorbed iodine is taken up by the thyroid. In chronic iodine deficiency, this fraction can exceed 80% [1]. Under normal circumstances, plasma iodine has a half-life of ≈10 hours, but this is reduced in iodine deficiency. During lactation, the mammary gland concentrates iodine and secretes it into breast milk to provide for the newborn. The body of a healthy adult contains 15–20 mg of iodine, of which 70–80% is in the thyroid. In chronic iodine deficiency, the iodine content of the thyroid may fall to less than 20 μg. In iodine-sufficient areas, the adult thyroid traps ≈60 μg of iodine/day to balance losses and maintain thyroid hormone synthesis; the sodium/iodide symporter (NIS), transfers iodide into the thyroid at a concentration gradient 20–50 times that of plasma [5]. Iodine comprises 65% and 59% of the weights of thyroxine (T_4) and triiodothyronine (T_3), respectively. Turnover is relatively slow: the half-life of T_4 is ≈5 days and for T_3, 1.5–3 days. The released iodine enters the plasma iodine pool and can be taken up again by the thyroid or excreted by the kidney. More than 90% of ingested iodine is ultimately excreted in the urine.

Historical Perspective

In 1811, Courtois noted a violet vapour arising from burning seaweed ash, and Gay-Lussac subsequently identified the vapour as iodine, a new element. The Swiss physician Coindet, in 1813, hypothesized the traditional treatment of goitre with seaweed was effective because of its iodine content, and successfully treated goitrous patients with iodine [6]. Two decades later, the French chemist Boussingault, working in the Andes Mountains, was the first to advocate prophylaxis with iodine-rich salt to prevent goitre. The French chemist Chatin was the first to publish, in 1851, the hypothesis that iodine deficiency was the cause of goitre. In 1883, Semon suggested myxoedema was due to thyroid insufficiency, and the link between goitre, myxoedema and iodine was established when, in 1896, Baumann and Roos discovered iodine in the thyroid. In the first two decades of the twentieth century, pioneering studies by Swiss and American physicians demonstrated the efficacy of iodine prophylaxis in the prevention of goitre and cretinism [6].

Epidemiology

Only a few countries, including Switzerland, some of the Scandinavian countries, Australia, the United States, and Canada, were completely iodine sufficient before 1990 [3]. Since then, globally, the number of households using iodized salt has risen from <20% to >80%, dramatically reducing iodine deficiency. The Iodine Global Network (IGN) Global Scorecard currently tracks global and national progress towards iodine sufficiency [7]. In 2020, based on recent national or subnational data, only 24 countries remain iodine deficient, 115 have optimal iodine intake, and 14 have excessive iodine intake [7]. This remarkable progress has been spurred by a coalition of international organizations, including the Iodine Global Network (formerly ICCIDD), WHO, and UNICEF, working closely with national iodine deficiency disorders (IDD) control committees and the salt industry; this informal partnership was established after the World Summit for Children in 1990.

In the US and the UK, iodine intakes have fallen over the past 2–3 decades, likely because of decreased iodine intake from dairy products. Pregnant women in the US are now mildly iodine deficient [8]. Other countries, because of over-iodized salt or high iodine in groundwater (e.g. Somalia), have excessive iodine intakes [8]. Vietnam, a country in Southeast Asia with a previously effective iodized salt programme, has experienced backsliding, and has relapsed to iodine deficiency [7]. These changes emphasize the importance of regular and systematic monitoring of iodine status in countries, to detect both low and excessive intakes of iodine.

Pathogenesis and Pathology

Iodine deficiency has multiple adverse effects on growth and development in animals and humans. These are collectively termed the IDD (Table 3.2.3.2) and result from inadequate thyroid hormone production due to lack of sufficient iodine [1].

Goitre

Thyroid enlargement (goitre) is the classic sign of iodine deficiency, and can occur at any age, even in the newborn. It is a physiologic adaptation to chronic iodine deficiency. As iodine intake falls, secretion of thyroid-stimulating hormone (TSH) increases in an effort to maximize uptake of available iodine, and TSH stimulates thyroid hypertrophy and hyperplasia. Initially, goitres are characterized by diffuse, homogeneous enlargement, but over time, nodules often

Table 3.2.3.2 The iodine deficiency disorders, by age group

Physiological groups [1]	Health consequences of iodine deficiency
All ages	Goitre, including toxic nodular goitre
	Increased occurrence of hypothyroidism in moderate-to-severe iodine deficiency; decreased occurrence of hypothyroidism in mild-to-moderate iodine deficiency
	Increased susceptibility of the thyroid gland to nuclear radiation
Fetus	Abortion
	Stillbirth
	Congenital anomalies
	Perinatal mortality
Neonate	Infant mortality
	Endemic cretinism
Child and adolescent	Impaired mental function
	Delayed physical development
Adults	Impaired mental function
	Iodine-induced hyperthyroidism
	Overall, moderate-to-severe iodine deficiency causes subtle but widespread adverse effects in a population secondary to hypothyroidism, including decreased educability, apathy, and reduced work productivity, resulting in impaired social and economic development

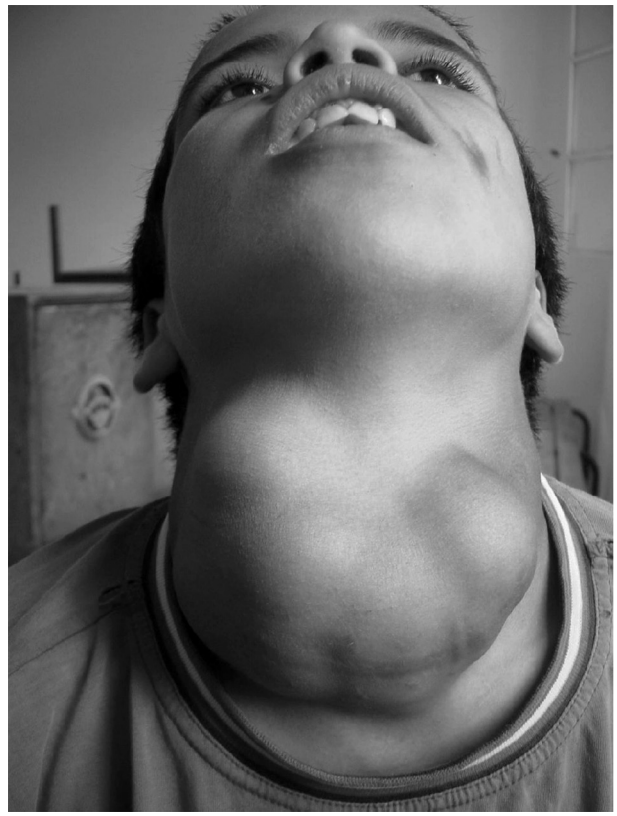


Figure 3.2.3.1 Large nodular goitre in a 14-year-old boy photographed in 2004 in an area of severe IDD in northern Morocco, with tracheal and oesophageal compression and hoarseness, likely due to damage to the recurrent laryngeal nerves.

develop (Figure 3.2.3.1). Many thyroid nodules derive from a somatic mutation and are of monoclonal origin; the mutations appear to be more likely to result in nodules under the influence of a growth promoter, such as iodine deficiency. Iodine deficiency is associated with a high occurrence of multinodular toxic goitre, mainly seen in women older than 50 years. Large goitres may be cosmetically unattractive, can obstruct the trachea and oesophagus, and may damage the recurrent laryngeal nerves and cause hoarseness. Surgery to reduce goitre has significant risks, including bleeding and nerve damage, and hypothyroidism may develop after removal of thyroid tissue.

Severe Iodine Deficiency in Pregnancy: Cretinism and Increased Fetal and Perinatal Mortality

The most serious adverse effect of iodine deficiency is damage to the fetus. Maternal thyroxine crosses the placenta before onset of fetal thyroid function at 10–12 weeks and represents up to 20–40% of T_4 measured in cord blood at birth. Normal levels of thyroid hormones are required for neuronal migration and myelination of the fetal brain, and lack of iodine irreversibly impairs brain development [9]. Severe iodine deficiency during pregnancy increases risk for stillbirths, abortions, and congenital abnormalities [1]. Iodine treatment of pregnant women in areas of severe deficiency reduces fetal and perinatal mortality and improves motor and cognitive performance of the offspring [10]. Severe iodine deficiency *in utero* causes a condition characterized by gross intellectual disability along with varying degrees of short stature, deaf mutism, and spasticity that is termed cretinism [1]. Two distinct types—neurological and myxedematous— have been described, but it may also present as a mixed form (Figure 3.2.3.2). The more common type, neurologic cretinism, has specific neurologic deficits that include spastic quadriplegia with sparing of the distal extremities. The myxedematous form is seen most frequently in central Africa, and has the predominant finding of profound hypothyroidism, with thyroid atrophy and fibrosis. In areas of severe iodine deficiency, cretinism can affect 5–15% of the population. Iodine prophylaxis has completely eliminated the appearance of new cases of cretinism in previously iodine-deficient Switzerland and many other countries; however, new cases have been recently reported from Papua New Guinea.

Mild-to-Moderate Deficiency in Pregnancy

The potential adverse effects of mild-to-moderate iodine deficiency during pregnancy are unclear. Maternal subclinical hypothyroidism (an increased TSH in the second trimester) and maternal hypothyroxinaemia (a free T_4 concentration <10 percentile at 12-week gestation) are associated with impaired mental and psychomotor development of the offspring [11]. However, in these studies, the maternal thyroid abnormalities were unlikely due to iodine deficiency. Observational studies have found that mild-to-moderate iodine deficiency during pregnancy is associated with impaired cognitive development in the offspring [12]. However, observational studies are often confounded by other factors that affect child development. In Europe, several randomized controlled trials of iodine

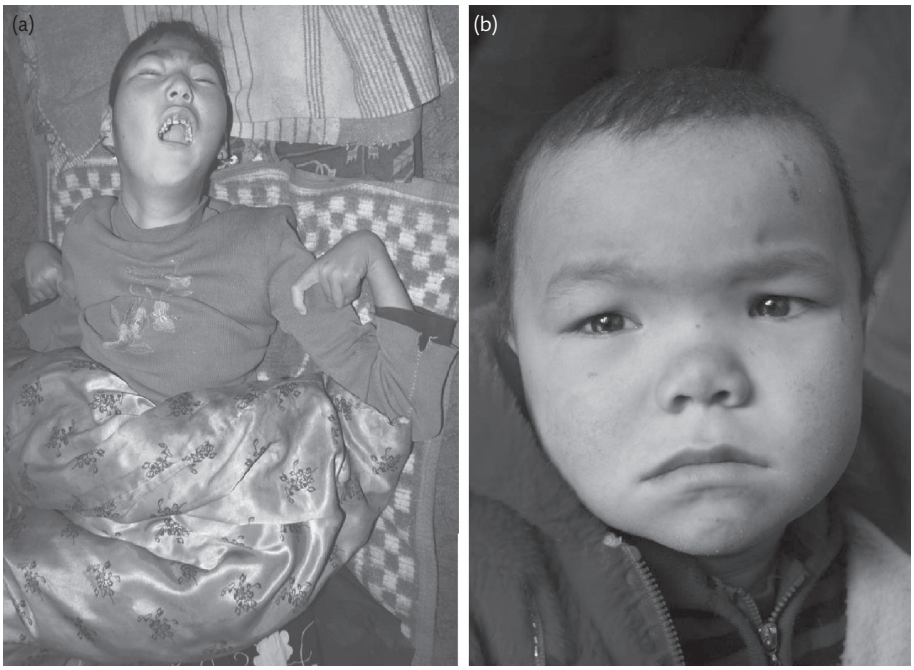


Figure 3.2.3.2 (a) Neurological cretinism. This 2007 photograph of a 9-year-old girl from Western China demonstrates the three characteristic features: severe mental deficiency together with squint, deaf mutism, and motor spasticity of the arms and legs. The thyroid is present, and frequency of goitre and thyroid dysfunction is similar to that observed in the general population. (b) Myxedematous cretinism. This 2007 photograph of a 5-year-old boy from Western China demonstrates the characteristic findings: profound hypothyroidism, severe growth retardation, incomplete maturation of the features including the naso-orbital configuration, atrophy of the mandible, puffy features, umbilical hernia, myxedematous, thickened, dry skin, and dry hair, eyelashes, and eyebrows. The thyroid typically shows atrophic fibrosis.

supplementation in mild-to-moderately iodine-deficient pregnant women have been done [13]. Iodine reduced maternal and newborn thyroid size, and, in some, decreased maternal TSH. However, none of the trials showed an effect on maternal and newborn total or free thyroid hormone concentrations, the most important outcome, and none measured long-term clinical outcomes, such as maternal goitre, thyroid autoimmunity, or child development [13]. A recent randomized multicentre intervention trial in Thailand and India that provided mildly iodine-deficient pregnant women 200 µg iodine supplements or placebo from the first trimester to delivery did not find benefits on maternal thyroid function, or offspring development at age 5–6 years [14].

Growth and Cognition in Childhood

Although iodine deficiency *in utero* impairs fetal growth and brain development, its postnatal effects on growth and cognition are less clear. Cross-sectional studies of moderate-to-severely iodine-deficient children have generally reported impaired intellectual function and fine motor skills; meta-analyses suggest populations with chronic iodine deficiency experience a reduction in IQ of 12.5–13.5 points [15]. Two randomized controlled trials in mild-to-moderately deficient school-aged children have shown clear benefits of iodine on cognitive and motor function [16, 17]. Moderately iodine-deficient 10–12-year-old children (n = 310) in Albania were randomized to receive either 400 mg of iodine as oral iodized oil or placebo [16]. Compared to placebo, iodine treatment significantly improved performance on tests of information processing, fine motor skills, and visual problem solving. The second placebo-controlled, double-blind trial was conducted in mildly iodine-deficient New Zealand children (n = 184) randomly assigned to receive 150 µg I daily or placebo for 28 weeks [17]. The overall cognitive score of the iodine-supplemented group was 0.19 SDs higher than that of the placebo group ($P = 0.011$) [17]. Thus, in children born and raised in areas of iodine deficiency, cognitive impairment is at least partially reversible by iodine repletion.

Data from cross-sectional studies on iodine intake and child growth are mixed, with most studies finding modest positive correlations. In iodine-deficient children, impaired thyroid function and goitre were inversely correlated with IGF-1 and IGFBP-3 concentrations. Iodine repletion in school-age children increased insulin-like growth factor (IGF)-1 and insulin-like growth factor binding protein (IGFBP)-3 and improved somatic growth [18]. However, other controlled intervention studies of iodized oil alone and iodine given with other micronutrients have generally not found effects on child growth. In a recent systematic review [19], iodine supplementation of severely iodine-deficient pregnant women increased mean birthweight (mean difference: 200 g; 95% CI: 183, 217 g) compared to controls, but iodine repletion in infants and children showed no effects on somatic growth.

A recent systematic review by WHO summarized the efficacy of iodized salt for preventing IDD [20]. Comparisons were made between the consumption of iodized salt and a placebo, non-iodized salt, or no intervention. The participants included members of the general population of any age and sex. The results of this review showed that iodized salt has substantial benefits on cognition including reducing the risk of: cretinism (odds ratio [OR] = 0.13

(95% CI, 0.08–0.20)), and low IQ (RR = 0.28 (95% CI, 0.21–0.36) [20]. The International Child Development Steering Group identified iodine deficiency as one of four key global risk factors for impaired child development where the need for intervention is urgent [21].

Assessment and Diagnosis

Four methods are generally recommended for assessment of iodine nutrition: urinary iodine concentration (UI), the goitre rate, serum TSH, and serum thyroglobulin (Tg) [3, 22]. These indicators are complementary, in that UI is a sensitive indicator of recent iodine intake (days), Tg shows an intermediate response (weeks to months), while changes in the goitre rate reflect long-term iodine nutrition (months to years).

Thyroid Size

Two methods are available for measuring goitre: neck inspection and palpation, and thyroid ultrasonography. By palpation, a thyroid is considered goitrous when each lateral lobe has a volume greater than the terminal phalanx of the thumbs of the subject being examined [3]. However, palpation of goitre in mild iodine deficiency has poor sensitivity and specificity and measurement of thyroid volume (Tvol) by ultrasound is preferable [22]. Thyroid ultrasound is non-invasive, quickly done (2–3 minutes per subject) and feasible even in remote areas using portable equipment. However, interpretation of Tvol data requires valid reference criteria, and age- and gender-specific references are available for 6–12-year-old children [3], but there are no established reference values for adults. Goitre can be classified by thyroid ultrasonography only if Tvol is determined by a standard method. Thyroid ultrasound is subjective; differences in technique can produce interobserver errors in Tvol as high as 26% [22].

Urinary Iodine Concentration

Because more than 90% of ingested iodine is excreted in the urine, UI is an excellent indicator of recent iodine intake. UI can be expressed as a concentration (μg/L), in relationship to creatinine excretion (μg iodine/g creatinine), or as 24-hour excretion (μg/day). For populations, because it is impractical to collect 24-hour samples in field studies, UI can be measured in spot urine specimens from a representative sample of the target group, and expressed as the median, in μg/L (3) (Table 3.2.3.3). However, the median UI is often misinterpreted. Individual iodine intakes, and, therefore, spot UI concentrations are highly variable from day-to-day and a common mistake is to assume that all subjects with a spot UI <100 μg/L are iodine deficient. To estimate iodine intakes in individuals, 24-hour collections are preferable, but difficult to obtain. An alternative is to use the age-and sex-adjusted iodine:creatinine ratio in adults, but this also has limitations [22]. Creatinine may be unreliable for estimating daily iodine excretion from spot samples, especially in malnourished subjects where creatinine concentration is low. Daily iodine intake in adults can be extrapolated from the median UI in populations using estimates of mean 24-hour urine volume and assuming an average iodine bioavailability of 92% using the

Table 3.2.3.3 Epidemiological criteria for assessing iodine nutrition in a population based on median and/or range of urinary iodine concentrations

Median urinary iodine (μg/L) [3]	Iodine intake	Iodine nutrition
School-aged children		
<20	Insufficient	Severe iodine deficiency
20–49	Insufficient	Moderate iodine deficiency
50–99	Insufficient	Mild iodine deficiency
100–199	Adequate	Optimal
200–299	More than adequate	Risk of iodine-induced hyperthyroidism in susceptible groups
>300	Excessive	Risk of adverse health consequences (iodine-induced hyperthyroidism, autoimmune thyroid disease)
Pregnant women		
<150	Insufficient	
150–249	Adequate	
250–499	More than adequate	
≥500	Excessive[a]	
Lactating women[b]		
<100	Insufficient	
≥100	Adequate	
Children less than 2 years old		
<100	Insufficient	
≥100	Adequate	

[a] The term 'excessive' means in excess of the amount required to prevent and control iodine deficiency.

[b] In lactating women, the figures for median urinary iodine are lower than the iodine requirements because of the iodine excreted in breast milk.

formula: Urinary iodine (μg/L) × 0.0235 × body weight (kg) = daily iodine intake [4]. Using this formula, a median UI of 100 μg/L in adults corresponds roughly to an average daily intake of 150 μg.

Thyroid-Stimulating Hormone

Because serum TSH is determined mainly by the level of circulating thyroid hormone, which in turn reflects iodine intake, TSH can be used as an indicator of iodine nutrition. However, in older children and adults, although serum TSH may be slightly increased by iodine deficiency, values often remain within the normal range. TSH is therefore a relatively insensitive indicator of iodine nutrition in adults [3]. In contrast, TSH is a sensitive indicator of iodine status in the newborn period. Compared to the adult, the newborn thyroid contains less iodine but has higher rates of iodine turnover. Particularly when iodine supply is low, maintaining high iodine turnover requires increased TSH stimulation. Serum TSH concentrations are therefore increased in iodine-deficient infants for the first few weeks of life, a condition termed transient newborn hypothyroidism. In areas of iodine deficiency, an increase in transient newborn hypothyroidism, indicated by more than 3% of newborn TSH values above the threshold of 5 mU/L whole blood collected 3 to 4 days after birth, suggests iodine deficiency in the population [3]. Newborn TSH is an important measure because it reflects iodine status during a period when the developing brain is particularly sensitive to iodine deficiency.

Thyroglobulin

Thyroglobulin (Tg) is synthesized only in the thyroid, and is the most abundant intrathyroidal protein. In iodine sufficiency, small amounts of Tg are secreted into the circulation, and serum Tg is normally less than 10 μg/L [22]. In iodine deficiency, serum Tg increases due to greater thyroid cell mass and TSH stimulation. Serum Tg is well correlated with the severity of iodine deficiency as measured by UI. Tg falls rapidly with iodine repletion, and Tg is a more sensitive indicator of iodine repletion than TSH or T_4 [22].

New assays for Tg have been developed for dried blood spots taken by a finger prick, simplifying collection and transport [23]. In prospective studies, dried blood spot Tg has been shown to be a sensitive measure of iodine status and reflects improved thyroid function within several months after iodine repletion [23]. International reference ranges for DBS Tg have been proposed for school-age children [3, 23] and pregnant women [24].

Thyroid Hormone Concentrations

Thyroid hormone concentrations (T_4 and T_3) are poor indicators of iodine intake. In iodine-deficient individuals, serum T_3 increases or remains unchanged, and serum T_4 usually decreases. However, these changes are often within the normal range, and make thyroid hormone levels an insensitive measure of iodine nutrition [3].

Prevention and Treatment

Salt Fortification with Iodine

In nearly all regions affected by iodine deficiency, the most effective way to control iodine deficiency is through salt iodization [3]. Universal salt iodization (USI) is a term used to describe the iodization of all salt for human (food industry and household) and livestock consumption. Although the ideal, even in countries with successful salt iodization programmes, USI is rarely achieved, as food industries are often reluctant to use iodized salt, and many countries do not iodize salt for livestock.

WHO/UNICEF/ICCIDD recommends that iodine is added at a level of 20–40 mg iodine/kg salt, depending on local salt intake [3]. Iodine can be added to salt in the form of potassium iodide (KI) or potassium iodate (KIO_3). Because KIO_3 has higher stability than KI in the presence of salt impurities, humidity, and porous packaging, it is the recommended form in tropical countries and those with low-grade salt. Iodine is usually added after the salt has been dried. Two techniques are used: (1) the wet method, where a solution of KIO_3 is dripped or sprayed at a regular rate on to salt passing by on a conveyor belt; (2) the dry method, where KI or KIO_3 powder is sprinkled over the dry salt. Optimally, packaging should be in low-density polyethylene bags, as high humidity combined with porous packing may result in up to 90% losses of iodine after 1 year of storage in high-density polyethylene bags.

Health Economics of Salt Iodization

Salt iodization remains the most cost-effective way of delivering iodine and of improving cognition in iodine-deficient populations [25, 26]. Worldwide, the annual costs of salt iodization are estimated at 0.02–0.05 US$ per child covered, and the costs per child death averted are US$1000 and per disability-adjusted life year (DALY) gained are US$34–36 US [25]. Looked at in another way, prior to widespread salt iodization, the annual potential losses attributable to iodine deficiency in the developing world have been estimated to be US$35.7 billion as compared with an estimated US$0.5 billion annual cost for salt iodization (i.e. a 70:1 benefit:cost ratio) [1]. The World Bank [26] strongly recommends that governments invest in micronutrient programmes, including salt iodization, to promote development, concluding: 'Probably no other technology offers as large an opportunity to improve lives at such low cost and in such a short time.'

Supplementation

In some regions, iodization of salt may not be practical for control of iodine deficiency, at least in the short term. This may occur in remote areas where communications are poor or where there are numerous small-scale salt producers. In these areas, iodized oil supplements can be used [3]. Iodized oil is prepared by esterification of the unsaturated fatty acids in seed or vegetable oils, and addition of iodine to the double bonds. It can be given orally or by intramuscular injection [3]. The intramuscular route has a longer duration of action, but oral administration is more common because it is simpler. Usual doses are 200–400 mg iodine/year and it is often targeted to women of childbearing age, pregnant women, and children [3] (Table 3.2.3.4). Its disadvantages are an uneven level of iodine in the body over time and the need for direct contact with individuals with the accompanying increased programme costs.

Iodine can also be given as KI or KIO_3 as drops or tablets. Single oral doses of potassium iodide monthly (30 mg) or biweekly (8 mg) can provide adequate iodine for school-age children [27]. Lugol's iodine, containing ≈6 mg iodine per drop, and similar preparations are often available as antiseptics in rural dispensaries in developing

Table 3.2.3.4 Recommendations for iodine supplementation in pregnancy and infancy in areas where <90% of households are using iodized salt and the median UI is <100 µg/L in schoolchildren

Women of childbearing age	A single annual oral dose of 400 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the reference nutrient intake (RNI) of 150 µg/d of iodine
Women who are pregnant or lactating	A single annual oral dose of 400 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the RNI of 250 µg/d iodine. • Iodine supplements should not be given to a woman who has already been given iodized oil during her current pregnancy or up to 3 months before her current pregnancy started
Children aged 0–6 months	A single oral dose of 100 mg of iodine as iodized oil OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the of 90 µg/d of iodine • Should be given iodine supplements only if the mother was not supplemented during pregnancy or if the child is not being breast-fed
Children aged 7–24 months old	A single annual oral dose of 200 mg of iodine as iodized oil as soon as possible after reaching 7 months of age OR A daily oral dose of iodine as potassium iodide should be given so that the total iodine intake meets the RNI of 90 µg/d of iodine

countries and offer another simple way to deliver iodine locally. In countries or regions where a salt iodization programme covers ≥90% of households and has been sustained for ≥2 years, and the median UI indicates iodine sufficiency (Table 3.2.3.4), pregnant and lactating women do not need iodine supplementation [3]. In iodine-deficient countries or regions that have weak iodized salt distribution, supplements should be given to pregnant women, lactating women, and infants, according to the guidelines in Table 3.2.3.4 [3]. In a recent controlled trial, lactating women who received one dose of 400 mg iodine as oral iodized oil soon after delivery provided adequate iodine to their infants through breast milk for at least 6 months, enabling the infants to achieve euthyroidism [28].

Clinical Nutrition

Preterm Infants

Balance studies in healthy preterm infants have suggested iodine intakes of at least 30 µg/kg body weight/day are required to maintain positive balance, and experts generally recommend iodine intakes of 30–60 µg/kg/day for this group [29]. Formula milks for preterm infants contain 20–170 µg iodine/L, and, depending on the dietary iodine intake of the mother, breast milk generally contains 50–150 µg/L. Thus, particularly during the first postnatal weeks when feed volumes are often low, enterally fed preterm infants may not achieve the recommended intake of iodine [29]. US and European clinical nutrition societies recommend parenteral iodine intakes of 1 µg/kg body weight/day [30], far below fetal accretion rates. This conservative recommendation assumes parenterally fed preterm infants will absorb iodine through the skin from topical iodinated disinfectants, and also receive small amounts of adventitious iodine in other infusions.

Because of concerns over possible iodine excess, use of iodinated antiseptics in infants may be decreasing, putting infants at risk of iodine deficiency [29]. If parentally fed preterm infants are not exposed to adventitious sources of iodine, they may receive only 1–3 µg iodine/kg body weight/day, and be in negative iodine balance during the first few postnatal weeks [29]. Several authors have argued that iodine deficiency should be avoided during this period because it may transiently lower thyroid hormone levels in the first weeks of life, and transient hypothyroxinaemia in preterm infants has been linked to impaired neurodevelopment [31].

Children and Adults

A daily dose of 1 µg iodine/kg body weight is recommended for children receiving parenteral nutrition [29]. For adults, commercially available products for enteral nutrition generally supply 75–110 µg iodine/serving. A technical review recommended iodine intakes of 70–140 µg/day during parenteral nutrition [32]. Although most parenteral nutrition formulations do not contain iodine, deficiency is not likely to occur because of cutaneous absorption from iodine-containing disinfectants and other adventitious sources of iodine. It is likely that thyroidal iodine stores are often adequate to meet the needs of patients requiring total parenteral nutrition for less than 3 months; in iodine-sufficient adults, thyroidal iodine content is 15–20 mg [1]. For these reasons, supplemental iodine is not routinely recommended for subjects receiving total parenteral nutrition [33].

Correction of Iodine Deficiency and Thyroid Disorders in Populations

Several reviews have examined the relationship between correction of iodine deficiency and the pattern of thyroid diseases in populations [34, 35]. In summary, as a population moves from severe iodine deficiency to mild iodine deficiency and then to iodine sufficiency, there is a modest shift from excess hypothyroidism to excess hyperthyroidism, which is transient, and then a small shift back towards excess mild hypothyroidism. Severe iodine deficiency causes more hypothyroidism because, despite an increase in thyroid activity to maximize iodine uptake and recycling, there is simply not enough iodine to maintain thyroid hormone production. In mild-to-moderate iodine deficiency, the thyroid gland is able to compensate for deficient dietary intake by increasing thyroid activity and this maintains thyroid hormone production, but at a price: in some individuals, chronic stimulation of the thyroid will lead to thyroid nodularity and autonomy [34]. This increase in nodularity

subsequently increases risk of hyperthyroidism if iodine intakes are raised by supplementation or fortification. However, this is transient and since iodine sufficiency normalizes thyroid activity this results, in the long-term, in reduced nodularity and autonomy [35]. The small increase in mild hypothyroidism that occurs with optimal or excessive iodine intakes may be linked to thyroid autoimmunity and may also transient, but more long-term studies are needed [35]. The overall incidence of thyroid carcinoma in populations does not appear to be influenced by iodine intake. The distribution of the subtypes of thyroid carcinoma is related to iodine intake [36]; in areas of higher iodine intake, there appear to be fewer of the more aggressive follicular and anaplastic carcinomas, but more papillary carcinomas. When iodine prophylaxis is introduced in populations, there may be an increase in the ratio of papillary to follicular carcinoma [36], and this shift towards less malignant types of thyroid cancer, as well as a lower radiation dose to the thyroid in case of nuclear fallout, are benefits of the correction of mild-to-moderate iodine deficiency.

REFERENCES

1. Zimmermann MB, Jooste PL, Pandav CS. The iodine deficiency disorders. *Lancet*, 2008; **372**: 1251–62.
2. Haldimann M, Alt A, Blanc A, Blondeau K. Iodine content of food groups. *J Food Comp and Anal*, 2005; **18**: 461–71.
3. World Health Organization/International Council for the Control of the Iodine Deficiency Disorders/United Nations Childrens Fund. *Assessment of the Iodine Deficiency Disorders and Monitoring their Elimination*, 2nd edn. Geneva: WHO, 2007.
4. Institute of Medicine (IOM), Academy of Sciences, USA. *Dietary reference intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium and Zinc*. Washington DC: National Academy Press, 2001.
5. Eskandari S, Loo DD, Dai G, Levy O, Wright EM, Carrasco N. Thyroid Na+/I– symporter. Mechanism, stoichiometry, and specificity. *J Biol Chem*, 1997; **272**: 27230–8.
6. Zimmermann M. Research on iodine deficiency and goiter in the 19th and early 20th centuries. *J Nutr*, 2008; **138**: 2060–3.
7. Iodine Global Network. Global Scorecard 2020. Available at: http://www.ign.org (accessed 10 June 2022).
8. Caldwell KL, Pan Y, Mortensen ME, Makhmudov A, Merrill L, Moye J. Iodine status in pregnant women in the National Children's Study and in U.S. women (15–44 years), National Health and Nutrition Examination Survey 2005–2010. *Thyroid*, 2013; **23**: 927–37.
9. Morreale de Escobar G, Obregon MJ, Escobar del Rey F. Role of thyroid hormone during early brain development. *Eur J Endocrinol*, 2004; **151**(S3): 25–37.
10. Cao XY, Jiang XM, Dou ZH, *et al*. Timing of vulnerability of the brain to iodine deficiency in endemic cretinism. *N Engl J Med*, 1994; **331**: 1739–44
11. Haddow JE, Palomaki GE, Allan WC, *et al*. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*, 1999; **341**, 549–55.
12. Bath SC, Steer CD, Golding J, Emmett P, Rayman MP. Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: results from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Lancet*, 2013; **382**: 331–7.
13. Zimmermann MB. The adverse effects of mild-to-moderate iodine deficiency during pregnancy and childhood: a review. *Thyroid*, 2007; **17**: 829–35.
14. Gowachirapant S, Jaiswal N, Melse-Boonstra A, *et al*. Effect of iodine supplementation in pregnant women on child neurodevelopment: a randomised, double-blind, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 853–63.
15. Bleichrodt N, Garcia I, Rubio C, Morreale de Escobar G, Escobar del Rey F. Developmental disorders associated with severe iodine deficiency. In: Hetzel B, Dunn J, Stanbury J, eds. *The Prevention and Control of Iodine Deficiency Disorders*. Amsterdam: Elsevier, 1987: 65–84.
16. Zimmermann MB, Connolly K, Bozo M, *et al*. Iodine supplementation improves cognition in iodine-deficient schoolchildren in Albania: a randomized, controlled, double-blind study. *Am J Clin Nutr*, 2006; **83**: 108–14.
17. Gordon RC, Rose MC, Skeaff SA, Gray AR, Morgan KM, Ruffman T. Iodine supplementation improves cognition in mildly iodine-deficient children. *Am J Clin Nutr*, 2009; **90**: 1264–71.
18. Zimmermann MB, Jooste PL, Mabapa NS, *et al*. Treatment of iodine deficiency in school-age children increases insulin-like growth factor (IGF)-I and IGF binding protein-3 concentrations and improves somatic growth. *J Clin Endocrinol Metab*, 2007; **92**: 437–42.
19. Farebrother J, Naude CE, Nicol L, *et al*. Effects of iodized salt and iodine supplements on prenatal and postnatal growth: a systematic review. *Adv Nutr*, 2018; **9**: 219–37.
20. World Health Organization. *Guideline: Fortification of Food-Grade Salt with Iodine for the Prevention and Control of Iodine Deficiency Disorders*. Geneva: World Health Organization, 2014.
21. Walker SP, Wachs TD, Gardner JM, *et al*. Child development: risk factors for adverse outcomes in developing countries. *Lancet*, 2007; **369**: 145–57.
22. Zimmermann MB. Methods to assess iron and iodine status. *Br J Nutr*, 2008; **99**(S3): 2–9.
23. Zimmermann MB, de Benoist B, Corigliano S, *et al*. Assessment of iodine status using dried blood spot thyroglobulin: development of reference material and establishment of an international reference range in iodine-sufficient children. *J Clin Endocrinol Metab*, 2006; **91**: 4881–7.
24. Stinca S, Andersson M, Weibel S, *et al*. Dried blood spot thyroglobulin as a biomarker of iodine status in pregnant women. *J Clin Endocrinol Metab*, 2017; **102**: 23–32.
25. Caulfield LE, Richard SA, Rivera JA, Musgrove P, Black RE. Stunting, wasting, and micronutrient deficiency disorders. In: Jamison DT, Breman JG, Measham AR, *et al*., eds. *Disease Control Priorities in Developing Countries*, 2nd edn. New York: Oxford University Press, 2006: 551–68.
26. McGuire J, Galloway R. Enriching lives. *Overcoming Vitamin and Mineral Malnutrition in Developing Countries*. Washington, DC: World Bank, 1994.
27. Todd CH, Dunn JT. Intermittent oral administration of potassium iodide solution for the correction of iodine deficiency. *Am J Clin Nutr*, 1998; **67**: 1279–83.
28. Bouhouch RR, Bouhouch S, Cherkaoui M, *et al*. Direct iodine supplementation of infants versus supplementation of their breastfeeding mothers: a double-blind, randomised, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2014; **2**:197–209.
29. Zimmermann MB, Crill C. Iodine in enteral and parenteral nutrition. *Best Pract Res Clin Endocrinol Metab*, 2010; **24**: 143–58.

30. Koletzko B, Goulet O, Hunt J, Krohn K, Shamir R; Parenteral Nutrition Guidelines Working Group. Guidelines on Paediatric Parenteral Nutrition of the European Society of Paediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) and the European Society for Clinical Nutrition and Metabolism (ESPEN), supported by the European Society of Paediatric Research (ESPR). *J Pediatr Gastroenterol Nutr*, 2005; **41**(S2): 1–87.
31. Ibrahim M, Sinn J, McGuire W. Iodine supplementation for the prevention of mortality and adverse neurodevelopmental outcomes in preterm infants. *Cochrane Database Syst Rev*, 2006; **2**: CD005253.
32. Koretz RL, Lipman TO, Klein S, American Gastroenterological Association. AGA technical review on parenteral nutrition. *Gastroenterology*, 2001; **121**: 970–1001.
33. Atkinson M, Worthley LI. Nutrition in the critically ill patient: part II. Parenteral nutrition. *Crit Care Resusc*, 2003; **5**: 121–36.
34. Laurberg P, Nøhr SB, Pedersen KM, *et al*. Thyroid disorders in mild iodine deficiency. *Thyroid*, 2000; **10**: 951–63.
35. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*, 2015; **3**: 286–95.
36. Zimmermann MB, Galetti V. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. *Thyroid Res*, 2015; **18**: 8.

3.2.4 Radiation-Induced Thyroid Disease

Shunichi Yamashita, Furio Pacini, and Rossella Elisei

Introduction

Radiation is a mitogen which may cause damage to the cell DNA. When sufficiently severe, the damage may result in cell death. When the damage is less severe, the consequences to the cell depend upon the gene and cell system that are affected. The thyroid gland is particularly sensitive to the effects of radiation and the evidence that radiation may damage the thyroid gland is overwhelming. Both external and internal radiation have been associated with thyroid diseases (cancer and hypothyroidism, with or without thyroid autoimmunity) both *in vitro* and *in vivo*. External radiation to the thyroid was first recognized as a cause of thyroid carcinoma in the 1950s, when cases were found in individuals who had been given radiotherapy during childhood for an enlarged thymus [1]. Since then, numerous studies have confirmed and extended this initial observation.

Radioactive isotopes are used in several situations in humans. They are given in very large doses in the treatment of thyroid cancer, when the dose used is intended to kill all thyroid cancer cells, and in smaller doses in the treatment of thyrotoxicosis, with the intent to produce hypothyroidism. In these conditions the radiation doses are sufficiently high to kill the cells, thus no unwanted secondary thyroid disease occurs. Low doses of iodine isotopes are also used as tracers for diagnostic evaluation of the thyroid gland. In this situation, no cell killing is observed and there is the theoretical possibility for thyroid cell damage. However, no convincing evidence of subsequent thyroid disorders has so far been provided.

Many animal studies have shown that radioiodine is carcinogenic to the thyroid. Some of the earlier data suggested that internal radiation by radioiodine was less effective than external radiation, but according to one more recent study in rats [2], the carcinogenic potential of ^{131}I and X-rays appears to be the same. In both cases, the dose–response relationship seems to be linear, indicating that low doses also carry a risk. Iodine-131 is 20–30% as effective as external X- or χ-rays.

Thyroid Carcinoma After External Irradiation

Methodology in Epidemiological Studies

The relationship between radiation and thyroid carcinoma was first recognized in 1950 [1], and thyroid carcinoma was the first solid malignant tumour found to be increased among Japanese atomic bomb survivors [3]. This relationship was confirmed subsequently by many epidemiological studies [4].

Two major limitations should be taken into account in studies of the relationship between radiation and thyroid carcinoma. One is due to the fact that many patients are unaware of, or uncertain about, prior radiation exposure, especially when therapeutic irradiation was administered at a young age (recall bias), taking into account that radiation-induced thyroid carcinoma occurs several years later. The second and perhaps more relevant limitation is related to the frequent occurrence of thyroid nodules in the general population (4–7% by palpation and up to 50% by ultrasonography in people over 60 years). Moreover, most thyroid tumours are indolent and frequently not recognized clinically. Thus, the diagnosis of thyroid tumours depends on the extent of the diagnostic procedures used (diagnostic bias).

In case-control studies, the cases are patients with thyroid cancer identified by entry into a tumour registry. The controls are matched subjects free from thyroid carcinoma. Information on risk factors, such as radiation exposure, is obtained and the distribution in the two groups is compared. In such studies diagnostic bias is minimized, but recall bias may be important. In cohort studies, exposure to radiation is generally well documented, and recall bias is minimized. The frequency of thyroid carcinoma in the radiation-exposed group is compared with a group of similar subjects not exposed to radiation. In this case, diagnostic bias may be important. A final additional caveat is due to the fact that retrospective estimates of doses delivered to the thyroid are necessary to prove the

Box 3.2.4.1 Most common indices of cancer risk from radiation exposure

- RR (relative risk) = *O/E* (*n* observed in radiation-exposed/*n* expected in non-exposed)
- SIR (standardized incidence ratio) = *O/E* when *E* is derived from a registry
- ERR (excess relative risk) = *O/E*-1
- ERR/gray = ERR/mean dose (Gy) in the group
- EAR (excess absolute risk)/gray = *O-E*/PYG (person-year-gray of exposure to the risk)

aetiological weight of radiation in thyroid cancer (dose–effect relationship). These estimates may be difficult to obtain and are subject to error.

Most epidemiological studies dealing with the risk of developing radiation-induced thyroid cancer use the relative risk (RR) as an index, i.e. the ratio between the observed (*O*) number of cancers in the radiation-exposed group and the expected (*E*) number of cancers in the non-exposed group (RR = *O*/*E*). When the expected number is obtained from a registry, the RR is called the standardized incidence ratio. The most frequently used indices of risk estimates are reported in Box 3.2.4.1.

Risk estimates for radiation-induced thyroid cancer have been calculated in people exposed to external radiation. According to the National Council of Radiation Protection (NCRP) [5], the excess absolute risk is 2.5×10^{-4}/Gy per year for persons exposed under the age of 18. For adults, the risk per year is assumed to be half this value. Because of their smaller number of years at risk, the lifetime risk for adults is about one-quarter the risk for children.

In a pooled analysis [6] the excess absolute risk was 4.4×10^{-4}/Gy per year for persons exposed before the age of 15, confirming that the RR is largely dependent upon age at exposure with young children carrying the highest risk. As shown in Tables 3.2.4.1 and 3.2.4.2, little risk is carried after the age of 20 and almost none after the age of 40, as demonstrated in the study of atomic bomb survivors in Hiroshima and Nagasaki [7].

External Irradiation to the Head and Neck for Benign Diseases

Irradiation to the head and neck has been performed in children since 1920 for the treatment of benign conditions such as enlargement of the thymus, tonsils, adenoids, or neck lymph nodes, skin angioma, acne, otitis, or tinea capitis [4, 8–10]. This modality of treatment was particularly popular in the United States, where in 1970 as many as 76% of children with thyroid carcinoma had a history of radiation exposure [11]. In Europe, it was less frequently used: in two large referral centres for thyroid cancer, the Institut Gustave-Roussy in Villejuif, France, and the Department of Endocrinology in Pisa, Italy, the incidence of radiation-induced thyroid carcinoma in children or adolescents was 10% and 7%, respectively.

Ron and colleagues [6] reported an analysis of radiation exposure and thyroid cancer from seven large studies of a total of 58 000 children exposed to external radiation, in whom individual doses to the thyroid were known. About 700 thyroid carcinomas were observed. The excess relative risk per gray (ERR/Gy) was 7.7 (95% CI 2.1–28.7). The authors concluded that 88% of thyroid carcinomas that were found in children exposed to 1 Gy were attributable to radiation. The excess absolute risk was 4.4/10 000 population-year-gray of exposure (95% CI 1.9–10.1). The risk of thyroid cancer significantly increased after a mean dose as low as 100 mGy to the thyroid. There was no evidence for a threshold dose below which the effect disappeared. At higher doses (up to 1500 cGy), there was a linear relationship between dose and risk of cancer. At doses higher than 1500 cGy, the risk per gray decreased, probably because of cell killing, but the overall risk remained elevated.

In a study of 2634 patients from the Michael Reese Hospital in Chicago [12], whose thyroids received a mean dose of 590 mGy for benign disorders during childhood, about 60% developed thyroid nodules and 15% developed thyroid carcinoma within 40 years after radiation. These studies indicate clearly that the risk of thyroid cancer after exposure to external radiation is indeed very high, suggesting that the thyroid gland is very sensitive to radiation, especially during childhood (Table 3.2.4.2).

In most studies, the latency period between the time of radiation exposure and the appearance of the thyroid nodule ranges between 5 and 15 years. In the pooled analysis of seven studies mentioned earlier [6], only two cases were observed within the first 5 years after exposure; the ERR clearly increased between 5 and 9 years after exposure, with a peak at 15–19 years, an excess risk still being apparent at 40 years [12].

Since 1970, external radiation for benign disorders has been virtually abandoned in most countries.

Table 3.2.4.1 Thyroid cancer excess relative risk (ERR) from exposure to external radiation in adults

Study	Irradiated subjects	Mean does (cGy)	ERR/Gy
Atomic bomb	11 000	26	0.8
Neck cancer therapy	82 816	11	3.1
Tuberculous adenitis	124	820	1.2

Data taken from Shore RE, *Radiation Research*, 1992; 131: 98–111.

From Shore RE. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res*, 1992; 131: 98–111.

Table 3.2.4.2 Thyroid cancer excess relative risk (ERR) from exposure to external radiation before the age of 20 years

Study	Irradiated subjects	Mean does (cGy)	ERR/Gy
Atomic bomb	13 000	23	4.7
Thymus	2475	136	9.1
Tinea capitis	10 384	9	32.5
Tonsils	2634	59	2.5
Skin haemangioma	14 351	26	4.9
Skin haemangioma	11 807	12	7.5
Lymphoid hyperplasia	1195	24	20
Childhood cancer	9170	1250	1.1

Data taken from Shore RE, *Radiation Research*, 1992; 131: 98–111.

From Shore RE. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res*, 1992; 131: 98–111.

External Irradiation to the Head and Neck for Malignant Diseases

In case of external radiation to the head and neck for malignant disease, the dose delivered to the thyroid may be very high, and usually greater than that delivered for the treatment of benign conditions. Animal experiments have shown that for doses larger than 15–20 Gy, the risk of thyroid tumour is increased but the risk per gray decreases. This finding has been attributed to cell killing, which decreases the number of cells that may become neoplastic, and explains the high frequency of hypothyroidism observed in those animals.

In humans, high-dose radiation therapy to the neck (more than 20 Gy), as used for Hodgkin's disease, results in a high rate of hypothyroidism, but also in an increased risk of thyroid cancer [13]. The final outcome of the thyroid damage is probably related to the distance between the thyroid gland and the radiation field. If this is far from the thyroid, as in case of thoracic or abdominal radiation fields in children, the thyroid gland may receive radiation doses of some hundred milligray, not enough to produce hypothyroidism but sufficient to trigger thyroid cancer [14, 15]. Recently, it has been reported that all survivors of human cancer treated with craniospinal external radiotherapy during childhood require long-term observation, up to 25 years after the exposure, since their risk of developing thyroid cancer and other thyroid dysfunctions is increased with respect to the general population without a specific age-related plateau [16].

Atomic Bombs in Hiroshima and Nagasaki

After the atomic bombing in Hiroshima and Nagasaki in 1945, the body dose was mainly due to external irradiation (X-rays and neutrons). Contamination by radioactive isotopes of iodine is poorly known. The health consequences were studied in a cohort of 94 000 survivors and of 26 000 individuals who resided in Hiroshima and Nagasaki shortly after the bombing. A total of 225 thyroid cancers were diagnosed between 1958 and 1987 among the 79 972 survivors who were alive and free of cancer as of January 1958 and who had radiation dose estimates [7]. From a histological point of view, these tumours are very similar to conventional sporadic papillary thyroid cancer and, at variance with the post-Chernobyl thyroid tumours, the solid variant is very rare among atomic bomb survivors. However, molecular oncology analysis of 50 adult-onset papillary thyroid cancer exposed to A-bomb radiation showed that the prevalence of RET/PTC rearrangements was significantly correlated with the radiation dose and that other unknown gene alterations tended to be more frequent with increased radiation dose [17]. These findings suggest that radiation-associated gene alterations, mainly chromosomal rearrangements, other than RET/PTC might be involved in the adult-onset thyroid cancer of subjects who were exposed to high radiation dose.

Factors Affecting Sensitivity to Radiation-Induced Thyroid Cancer

Age and Sex

A major risk factor is a young age at the time of irradiation. The risk of thyroid cancer after external irradiation in children less than 5 years of age is two times higher than in children treated between 5 and 9 years and five times higher than in children treated between 10 and 14 years [6]. From the Lifespan Study of atomic bomb survivors in Hiroshima and Nagasaki [16], it is known that little risk is carried for exposures after the age of 20 and almost none after the age of 40. The excess risk of thyroid cancer was 9.5, 3.0, 0.3, and 0.2 in the age categories 0–9, 10–19, 20–39, and over 40 years, respectively, at the time of bombing. The excess risk was not significant for subjects exposed above the age of 15–20 years. This increased risk of very young children to develop thyroid cancer after radiation exposure can be explained, at least in part, by the higher proliferative activity of thyroid cells during intrauterine development and childhood [18]. The high susceptibility of young children to radiation has been confirmed in the thyroid cancer studies after the Chernobyl nuclear reactor accident (see next), supporting the concept that the radiation effect is maximal during periods of rapid cell proliferation, as in the case of the developing thyroid of very young children. Data on irradiation in adults are scarce, but estimates of the ERR/Gy are largely below those of individuals exposed during childhood, and it is likely the risk is negligible.

Gender does not seem to influence the risk of developing radiation-induced thyroid cancer. Although females are 2–3 times more likely to develop both benign and malignant thyroid nodules after irradiation, this finding reflects the higher natural incidence of thyroid nodules and cancer in the female general population. Very recently a new study on the association between radiation dose and thyroid cancer incidence among Japanese survivors who were adults at the time of the atomic bombings of Hiroshima and Nagasaki has shown that the exposure to ionizing radiation in adults was positively associated with thyroid cancer among women atomic bomb survivors. However, this association was lower than that observed in those who were exposed during childhood [19].

Fractionation and Dose Rate

External radiation therapy for benign and malignant diseases is given at a high dose rate. Lower dose rates or fractionation of the dose may theoretically allow radiation-induced DNA lesions to be repaired, thus decreasing the carcinogenic effects of radiation. In the pooled analysis of seven studies, fractionation of the dose was associated with a 30% reduction of the ERR/Gy [6]. However, in a recent update of thyroid cancer after radiation therapy for malignant disorders in childhood, no reduction in ERR/Gy was observed with fractionation.

The importance of the dose rate is suggested by several observations. In children treated for skin angioma of the neck, a dose–effect relationship was observed after external radiation at a high dose rate, but no such relation was found after brachytherapy at a low dose rate. The incidence of thyroid nodules is similar in two regions of China where natural radiation is different (i.e. 140 mGy and 50 mGy/lifetime, respectively) [20]. In contrast, an increased RR (1.7) of thyroid cancer was found among 27 000 medical diagnostic radiographers in China, who probably received more than 1 Gy to the thyroid during their working life [21]. No such increase was observed in similar workers in industrialized countries.

Genetic Predisposition

Several clinical observations suggest that genetic predisposition, such as defects in the DNA repair mechanisms, may affect the risk of developing radiation-induced thyroid cancer [22, 23]. Patients who experience one radiation-related cancer are more likely to

develop a second radiation-related cancer. Sibling pairs, exposed to radiation, develop thyroid tumours more often than would be expected by chance [22, 23]. The risk of thyroid cancer in patients treated with radiotherapy during childhood for a cancer (other than neuroblastoma) is 3–10 times higher than in children treated for benign conditions. Those treated for neuroblastoma have a fivefold risk of thyroid cancer with respect to patients treated for other cancers, suggesting a common predisposition for neuroblastoma and thyroid cancer.

The search for the gene(s) predisposing to radiation-induced thyroid cancer is currently in progress in pedigrees showing recurrence of thyroid cancer. No linkage has been found as yet with genes known to be involved in thyroid tumorigenesis, such as *ras, p53, BRaf*, or *RET/PTC*. A distinct genome-wide gene expression profiling has been reported in post-Chernobyl papillary thyroid cancer when compared with that occurring naturally, suggesting a greater susceptibility to thyroid cell radiation damage [24].

Thyroid Carcinoma after Exposure to Radioactive Iodine, and the Chernobyl Experience

Iodine-131, being physiologically accumulated in the thyroid by an active mechanism, has been widely used for several decades in the diagnostic evaluation of the thyroid gland and in the treatment of patients with hyperthyroidism and differentiated thyroid cancer. The radiation dose delivered by ^{131}I to the thyroid is 1000- to 10 000-fold higher than that delivered to other tissues. Thus, even a relatively low amount of ^{131}I may deliver a significant, potentially carcinogenic radiation dose to the thyroid gland. Increasing the radiation dose beyond a few hundred megabecquerels, increases the likelihood of obtaining cell killing and decreases the possibility of tumoural changes.

The role of radioactive iodine for medical use in the development of thyroid cancer has been addressed in several studies, which showed no significant risk and led to the conclusion that ^{131}I is sufficiently safe both as a diagnostic and a therapeutic tool. However, most patients included in these studies were treated as adults, whereas the post-Chernobyl epidemic of thyroid cancer occurred mainly in children and adolescents, supporting evidence that the young thyroid is particularly sensitive to the effect of radiation. This event has renewed concern about the carcinogenic risk of medical use of ^{131}I, at least in young patients.

Exposure to ^{131}I for Diagnostic Purposes

The most informative analysis in this setting was performed in Sweden on 34 104 patients exposed to diagnostic doses of ^{131}I between 1950 and 1969, for a mean thyroid dose estimate of 110 cGy [25]. A small increase in the number of observed thyroid cancers (n = 67) was found with respect to the expected number (n = 50). However, the increase was confined to patients undergoing thyroid scan for suspicion of thyroid cancer. When the analysis was limited to patients tested for reasons other than thyroid cancer, no increase was observed.

In the same Swedish cohort, the incidence of thyroid nodules was compared in a subset of 1005 women and 248 matched controls. The average length of follow-up was 26 years and the average age at exposure was 26 years. No difference was found in the two groups; the incidence of nodules was 10.6 and 11.7, respectively. Similar findings have been reported in other surveys in the United States and in Germany.

Table 3.2.4.3 Thyroid cancer excess relative risk (ERR)/Gy after exposure to ^{131}I before the age of 20 years

Study	Irradiated subjects	Mean does (cGy)	ERR/Gy
Swedish diagnostic ^{131}I	2408	150	0.25
Food and Drug Administration diagnostic ^{131}I	3503	80	0.10
Utah ^{131}I fallout	2473	17	7.9
Marshall islanders	127	1240	0.32
Juvenile hyperthyroidism	602	8800	0.3

Data taken from Shore RE, *Radiation Research*, 1992; 131: 98–111.
From Shore RE. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res*, 1992; 131: 98–111.

The conclusion drawn from this study is that diagnostic use of ^{131}I has no untoward health effect on the thyroid. However, a note of caution is needed because only a minority of the exposed patients were children. The ERR of thyroid cancer after exposure to ^{131}I before age 20 and in adults is reported in Tables 3.2.4.3 and 3.2.4.4, respectively.

Exposure to ^{131}I for Therapeutic Purposes

Radioiodine is used widely to treat hyperthyroidism caused either by Graves' disease, toxic nodular goitre, or metastatic thyroid cancer. No evidence of an increased risk of thyroid cancer after treatment of hyperthyroidism with ^{131}I has been reported. In a Swedish study, including 10 552 adult patients (mean age 57 years) followed for a mean period of 15 years, the relative risk of thyroid cancer was not significantly increased (RR 1.29; 95% CI 0.76–2.03). The average estimated radiation dose to the thyroid was 100 Gy [26].

In a study carried out in the United States [27] in hyperthyroid patients, after a mean follow-up of 21 years, ^{131}I treatment was not found to be linked to total cancer deaths (standardized mortality ratio (SMR) 1.02), or to the development of any cancer other than thyroid cancer (SMR 3.94; 95% CI 2.52–5.86). The SMR was 2.08 in patients with Graves' disease and 6.53 in those with toxic nodular goitre. The excess number of deaths was small (observed/expected 27/10), and the underlying disease, rather than radiation, seemed to play the major role. This result is not surprising. The large dose

Table 3.2.4.4 Thyroid cancer excess relative risk (ERR)/Gy after exposure to ^{131}I in adult life

Study	Irradiated subjects	Mean does (cGy)	ERR/Gy
Swedish diagnostic ^{131}I	24 200	42	<0
German diagnostic ^{131}I	13 896	100	0.3
Marshall islanders	126	466	0.5

Data taken from Shore RE, *Radiation Research*, 1992; 131: 98–111.
From Shore RE. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res*, 1992; 131: 98–111.

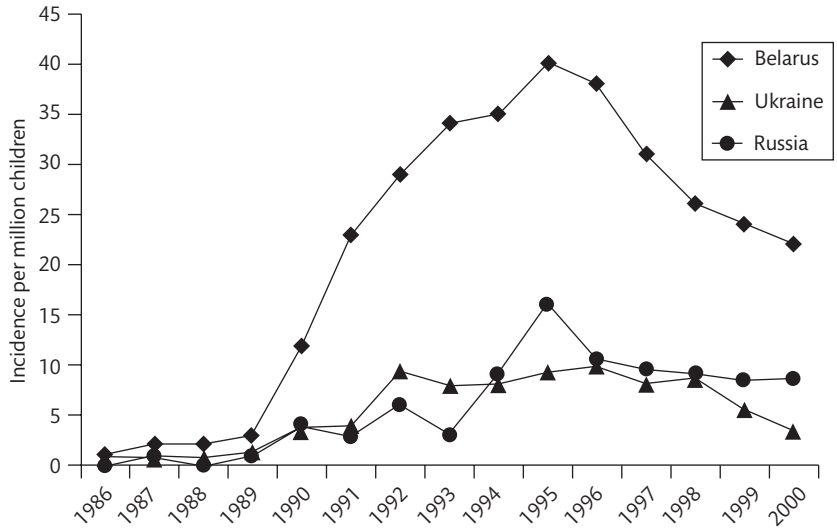


Figure 3.2.4.1 New cases of thyroid carcinoma per year diagnosed in Belarus, Ukraine, and Russia in children and adolescents exposed to radiation fallout after the Chernobyl accident.

delivered for the treatment of hyperthyroidism is frequently sufficient to produce hypothyroidism, through cell killing. Indeed, the risk of hypothyroidism at 2 years increases linearly with the thyroid dose, for radioactive concentrations ranging from 0.9 to 8.3 MBq/g.

As for the diagnostic use of ^{131}I, the hyperthyroid patients treated with radioiodine were adults. In a few hundred children treated with ^{131}I, no significant increase in the incidence of thyroid cancer has been observed [26]. However, in view of the high sensitivity of the young thyroid gland to radiation, it is probably advisable to avoid treating hyperthyroidism in young children and adolescents with ^{131}I.

As far as the treatment of differentiated thyroid cancer with ^{131}I is concerned, the only theoretical risk is the possibility that ^{131}I might act as an additional mutagen, inducing the progression of thyroid cancer to a more aggressive and less well-differentiated phenotype. At the present moment, no evidence supports this possibility. The other potential hazard of ^{131}I therapy of thyroid cancer is the occurrence of secondary effects on other organs when accumulating high radiation doses during several courses of treatment. Controversial data on this issue have been reported so far [28, 29].

Post-Chernobyl Thyroid Cancer

Circumstances of the Accident

In April 1986, the explosion of one of the reactors at the nuclear power plant in Chernobyl, Ukraine released large amounts of radioactive particles into the atmosphere, including ^{131}I (32–46 MCi), ^{132}I (27 MCi; resulting from the decay of ^{132}Tc), and ^{133}I (68 MCi). Most likely, radioiodines were released intermittently over a period of 10 days or more after the explosion. The time and place of deposition varied, depending on the direction of the wind and other meteorological conditions. The most contaminated territories were southern Belarus, northern Ukraine, and to a lesser extent the Bryansk and Kaluga regions of southern Russia.

As a result of the accident, a tremendous increase in the number of childhood papillary thyroid cancers occurred in the following years [30] (Figure 3.2.4.1). The magnitude of this increase and the geographical and temporal distribution of the cases strongly suggest that thyroid cancer was due to the reactor explosion and, in particular, to the huge amount of iodine radioisotopes released. The initial scepticism, allowing the possibility that the increased incidence of thyroid cancer might be due to ascertainment bias following intensive screening, has been totally discouraged by subsequent compelling evidence. Many of the tumours diagnosed in the first years were relatively large, invasive, and associated with lymph node metastases, unlike those detected during screening programmes, which are minimal, limited to the thyroid, and not aggressive [31]. The prevalence of childhood thyroid cancer exceeded that of any other country in the world and, very importantly, decreased dramatically in children conceived and born after the accident.

Being volatile, radioactive isotopes could be first inhaled and, after they were deposited on the ground, ingested. The time at which ingestion occurred varied considerably, but the milk chain, particularly in children, was the major route of ingestion: at this time, short-lived isotopes of iodine were no longer present. Several factors contributed to the high radiation exposure of the population. Immediate protective countermeasures, such as advising and evacuating the people at risk and distributing iodine prophylaxis, were not undertaken. Furthermore, the most contaminated regions were in a state of moderate iodine deficiency, which is responsible for increased iodine uptake. All these factors combined give enough explanation of why the most serious health consequence of the disaster was thyroid cancer, and why mostly children were affected [32].

In the case of radioactive contamination, the thyroid gland is a critical organ at risk. Its contamination depends upon the magnitude of contamination, the amount of radioactive iodine taken up by the gland, and the thyroid mass itself. Whatever the level of contamination, the thyroid dose is always higher in children than in adults. The thyroid dose is dependent on the final concentration, namely the ratio between radioiodine uptake and thyroid mass. In children, the uptake is similar to adults but, the thyroid mass being smaller, the dose per gram of tissue is greater, and extremely high in newborn and very young children.

In children who remained in the contaminated territories and drank locally produced milk, most of the radiation dose to the thyroid was due to ^{131}I and only a small amount to short-lived isotopes. The thyroid dose in children evacuated soon after the accident was lower, and mainly due to short-lived isotopes. Although dosimetric data are imprecise, the mean thyroid dose has been estimated to be nearly 700 mSv in Belarus. In Ukraine, 79% of the children received a thyroid dose below or equal to 300 mSv, 10.5% received from 300 mSv to 1 Sv, and 10.5% received more than 1 Sv [33]. As a term of comparison, in children exposed to external irradiation [6] the risk of thyroid cancer was significant even for thyroid doses as low as 100 mSv.

In most of the children that developed thyroid cancer the estimated thyroid dose was equal to or less than 300 mGy. An excess thyroid cancer incidence has been observed even in areas where the mean thyroid dose in children was estimated at 50–100 mGy.

Clinical Features of Post-Chernobyl Thyroid Cancer

The increase in the number of thyroid carcinomas in children and adolescents has been observed since 1990, only 4 years after the Chernobyl accident, in southern Belarus [34] and northern Ukraine, and from 1994 in southern Russia [35]. In the Gomel region, the most contaminated area of Belarus, the incidence between 1986 and 1996 was 13/100 000 children/year, compared to a baseline incidence of less than 1 per year. To date, more than 5000 cases of thyroid cancer have been reported among those who were children or adolescents at the time of the accident and living in the three most contaminated countries, Belarus, Ukraine, and Russia [30].

As shown in Figure 3.2.4.2, most of the cases were registered in children below age 10 at the time of the accident, and nearly two-thirds in those younger than 5 years. Thyroid cancer cases have also been registered up to 20 years after the nuclear accident in children who were already conceived but still *in utero*, at that time [36]. With respect to the 12 years before the accident, in the 12 years after the accident the increase of thyroid cancer in Belarus was 75-fold in children aged 3–14 years at the time of diagnosis, 10.1-fold in adolescents (15–18 years at diagnosis), 3.7-fold in young adults (19–29 years), and 3.4-fold in adults (Table 3.2.4.5). This increase in adults is much less important than that observed in children and it is likely to be due to greater attention to thyroid diseases after the nuclear accident.

Table 3.2.4.5 Thyroid cancer in Belarus before and after the Chernobyl accident

Age	1971–1985	1986–2000	Fold of increase
0–14	8	703	87.8
15–18	21	267	12.7
≥19	1465	6719	4.6
Total	1494	7689	5.1

Over 90% of the cancers were papillary. In the years following the accident most cancers were classified as a solid or follicular variant of papillary thyroid cancer (Figure 3.2.4.3), that is, the less frequently observed variant among naturally occurring papillary carcinomas. The clinical and pathological features were those of an aggressive tumour, as demonstrated by the histological appearance, the large size, the frequent multifocality and extracapsular invasion, and the frequency of node and lung metastases early in the course of the disease [31]. However, later studies showed a decline over time in the proportion of the solid variant and an increase in the proportion of the classic variant. These changes correlated both with the increasing age and increasing latency period and it is not yet clear which of these two variables could be mostly responsible for these changing patterns, which are also associated with a change in the molecular features of these tumours [37].

The comparison between post-Chernobyl thyroid carcinomas diagnosed in Belarus and naturally occurring cases diagnosed in age-matched patients in Italy and France showed different clinical and epidemiological features [38]. Post-Chernobyl tumours were much less influenced by gender (female:male ratio 1.6:1 versus 2.5:1 in Italy and France), were more advanced at presentation, were more frequently papillary, and were mainly diagnosed before age 15, while in Italy and France the majority were diagnosed after age 14.

As far as treatment and outcome are concerned, the available follow-up data indicate that post-Chernobyl thyroid carcinoma, when appropriately treated with a combination of total thyroidectomy, radioiodine, and hormone suppressive therapy,

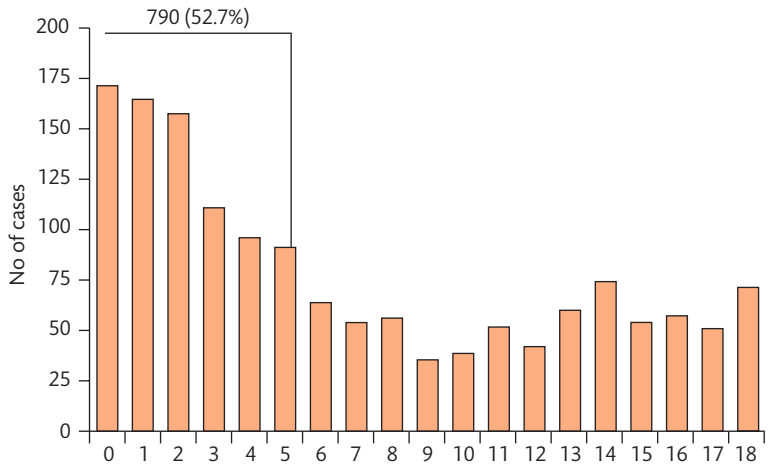


Figure 3.2.4.2 Children and adolescents with post-Chernobyl thyroid cancer in Belarus (1500 cases diagnosed from 1986 to 2002).

Table 3.2.4.6 RET activation in spontaneous and post-Chernobyl childhood papillary thyroid carcinoma

Spontaneous	Post-Chernobyl	References
12/17 (71%) USA	33/38 (87.0%) Belarus	[43]
n.d.	4/6 (66.6%) Belarus	[41]
n.d.	17/28 (60.7%) Ukraine	Thomas G.A. *et al.*, 1999
n.d.	20/39 (51.3%) Belarus	Thomas G.A. *et al.*, 1999
n.d.	9/15 (60.0%) Belarus	[42]
n.d.	25/51 (49.0%) Belarus	Smida J. *et al.*, 1999
10/21 (48.0%) UK	n.d.	Williams G.H. *et al.*, 1996
6/9 (67.0%) Italy	n.d.	[48]
3/10 (30.0%) Japan	n.d.	Motomura T. *et al.*, 1998
10/25 (40%) Italy	19/25 (76%) Belarus	[44]
41/82 (50%)	127/202 (63%)	Meta-analysis (P = 0.045)

n.d., not determined

has the same favourable outcome as naturally occurring papillary cancer. Definitive cure is achieved in many patients, even in those with node and lung involvement, the quality of life is good, and the death rate does not exceed the usual 1–2% reported in many series of paediatric thyroid cancer [39].

A similar observation has been recently reported in a study focused on external radiation-induced thyroid carcinoma. Although these tumours showed generally more aggressive features, the similar prognostic factors for their outcome indicate that they should be treated and followed in the same way as naturally occurring thyroid cancer [40].

Genetics of Post-Chernobyl Thyroid Cancer

Post-Chernobyl tumours show interesting genetic peculiarities when investigated by molecular biology. Molecular studies of the early post-Chernobyl thyroid cancer showed that a very high proportion harboured a RET/PTC rearrangement with a higher prevalence of RET/PTC3 [41–44]. Also the subtype of *RET/PTC* rearrangement showed a peculiar pattern. Several authors reported that *RET/PTC3* (and more rare variants of *RET/PTC3*) was the form more frequently expressed in radiation-induced tumours, thus suggesting that RET/PTC3 might represent a marker for these tumours. A correlation was also established between the solid variant of papillary tumours and the activation of *RET/PTC3* [43]. Interestingly, over the years and with the elongation of the latency period, the prevalence of RET/rearrangements became lower and more similar to that of naturally occurring thyroid carcinoma (Table 3.2.4.6). Also, the relative prevalence of RET/PTC1 and 3 subtypes changed in favour of RET/PTC1 [37].

The presence of RET/PTC rearrangements in radiation-induced cancer is in keeping with the *in vitro* findings that RET/PTC rearrangements can be induced in human thyroid cells after exposure to 0.1–10 Gy γ-radiation [45] and that *RET* gene fragmentation induced by ionizing radiation exposure is significantly higher than fragmentation of any other DNA region [46]. The generation of a RET/PTC3 rearrangement seems to be particularly facilitated by the alignment of ELE1 (i.e. the partner of RET/PTC3 rearrangement) and RET introns in opposite orientation [47]. Since *RET/PTC* is also frequently found in paediatric papillary thyroid cancer without known exposure to radiation [43, 48], it is also possible that age *per se* may play an important role. Alternatively, one can speculate that virtually all paediatric papillary thyroid cancers are radiation-induced cancers, developing in children with an increased susceptibility to spontaneous background radiation.

It has been noted that no point mutations of BRAF oncogene have been found in post-Chernobyl childhood thyroid carcinomas [49]. BRAF V600E activating mutation is present in about 40% of naturally occurring thyroid cancers in adults but is almost absent in children. The question of whether the very low frequency of BRAF mutations in post-Chernobyl childhood carcinoma is related to the young age of patients rather than the inability of ionizing radiation to induce oncogene point mutations is still not clarified. However, the finding of BRAF rearrangements in radiation-induced but not in naturally occurring thyroid cancer suggests that the oncogene alterations determined by ionizing radiations are mainly chromosomal rearrangements more than single point mutations [50]. In recent years, studies on genomic profiling have suggested distinct patterns in radiation-induced and sporadic thyroid cancer and in particular the expression of seven genes was found to be completely different in the two groups [24, 51].

Thyroid Cancer after the Fukushima Nuclear Power Plant Accident

Following the Fukushima accident in March 2011, the large-scale sophisticated thyroid ultrasound screening under the strict examination protocol launched from October 2011 in Fukushima Prefecture of Japan demonstrated a high detection rate of thyroid cancer in young individuals revealing 116 and 71 cases in the first and second rounds, respectively, among the same cohort of approximately 300 000 subjects, aged at the time of accident from 0 to 18 years old [52]. The postoperative pathological diagnosis revealed 121 cases (98.6%) of papillary thyroid adenocarcinoma (PTC), which included 110 cases of the classical type, 4 cases of the follicular variant, 3 cases of the diffuse sclerosing variant, and 4 cases of the cribriform-morular variant [53]. Intrathyroidal spread was observed in 61.6%, and calcifications, such as psammoma bodies, in 78.4%. The rates of lymph node metastasis and extrathyroidal tumour extension were also high, especially the rate of lymph node metastasis, which exceeded 70%. These findings, as well as a high rate of intrathyroidal spread and of calcifications, however, are common for childhood PTC.

Molecular characteristics of young thyroid cancers in Fukushima demonstrated that the frequency of *BRAF* gene mutations among all genetic alterations in cancer tissue is high, nearly the same as in adult thyroid cancers, and the frequency of gene rearrangement is substantially lower than that in Chernobyl [54].

These findings raise concerns among residents and the public that it might be due to putative exposure to radiation from the accident at Fukushima Daiichi Nuclear Power Plant. However, it is now apparent that effective doses to the whole body in the general population after the accident have been estimated to be less than several mSv for the majority of people, including infants and children [55]. No radiation level corresponding to thyroid dose exceeding 50 mSv was detected among 1080 children directly examined with scintillation survey meter at the end of March 2011 [56]. Especially in case of low-dose exposures, relevance to confounding

factors, such as hereditary predisposition, iodine-intake condition, hyperthyrotropinaemia, obesity, smoking, various hormone levels, and possible exposure to carcinogenic chemicals need to be addressed. So far, the international agencies and scientific experts support the notion that in view of low doses to the public, the increase of radiation-related risk for thyroid cancer is highly unlikely. The reasons are as follows. Exposure doses in Fukushima are significantly lower than those in Chernobyl. There is no apparent regional difference in the rate of thyroid cancer cases among the residential areas at the time of the earthquake and, hence, exposure status of the patients. The mean age of the subjects diagnosed with thyroid cancer was 10–15 at the time of accident, while there were no cases in younger children (aged 0–5 for the first 4 years), who are more vulnerable to radiation exposure. Although it is expected that thyroid cancer would also be detected at certain frequencies in those who were infants at the time of accident after some years, its causes need to be specially investigated.

Finally the lessons from Fukushima emphasize the importance of correct and sound understanding of radiation risk on the thyroid cancers and of difference of Chernobyl [57], and clearly demonstrate the necessity of a special attention on its overdiagnosis of PTC in children as well as in adults, following the long-term strategies for thyroid health monitoring after nuclear accidents in future [58].

Radiation-Induced Thyroid Diseases Other Than Thyroid Tumours

Thyroid cancer and benign thyroid nodules after thyroid radiation exposure occur as stochastic effects. Depending on the radiation dose, deterministic effects resulting in hypothyroidism and acute thyroiditis may also occur. Another documented consequence of radiation is the possibility of developing chronic autoimmune thyroid disorders.

Hypothyroidism is caused by radiation doses of the order of more than several gray to the thyroid. Such doses are used in the treatment of Graves' disease and toxic nodular goitre, and in these conditions hypothyroidism should be considered the aim rather than an untoward effect of treatment. Primary 'spontaneous' hypothyroidism (or subclinical hypothyroidism) was reported in survivors of the atomic bomb in Nagasaki [59]. In a study of 2587 survivors, 43 were diagnosed with hypothyroidism, 27 of whom were thyroid antibody positive and 16 were thyroid antibody negative, with no gender differences. Since an association was observed between thyroid dose and prevalence of antibody positivity, but not antibody negativity, primary hypothyroidism could conceivably have stemmed from an underlying autoimmune thyroid disorder. However, more recently, the same group reported that 55–58 years after radiation exposure, autoimmune thyroid disorders were not found to be significantly associated with radiation exposure while, in the same study, the authors confirmed a significant linear dose–response relationship in the prevalence of both thyroid cancer and benign thyroid nodules and that the relationship was higher in individuals who were exposed at younger ages [60].

The occurrence of thyroid autoimmunity after external irradiation to the head and neck has been reported in several studies. An increased incidence of thyroid antibodies was found by De Groot *et al.* [61] in individuals who received radiation during childhood for benign disorders. Variable degrees of thyroid lymphocytic infiltration have been reported in more than two-thirds of individuals who received radiation several years before thyroidectomy for nodular thyroid lesions. In patients who received radiation of the neck for Hodgkin's disease, 3% or more developed Graves' disease (a 7- to 20-fold excess risk) and 1% thyroiditis.

Hypothyroidism has also been reported after exposure to internal radiation (radioactive iodine). In the people exposed to the fallout of the Marshall Islands accident [62], hypothyroidism was noted within 10 years after the accident. On this occasion most of the cases were not associated with an autoimmune thyroid reaction.

In contrast, an increased prevalence of antithyroid antibodies (19.5%), without hypothyroidism, has been reported in children living in a Belarus village heavily contaminated by the post-Chernobyl radioactive fallout, as opposed to children living in a non-contaminated village (3.8% prevalence) [63]. The susceptibility to develop thyroid autoimmunity increased with age at the time of exposure and, in girls, reached its maximum at puberty, suggesting that puberty (oestrogen) and radiation have a cumulative effect in the development of thyroid antibodies in girls. However, a more recent study demonstrated that the increased prevalence of thyroid antibodies in exposed children was a real but transient phenomenon not accompanied by the development of 'overt' hypothyroidism or other thyroid dysfunction 13–15 years after the Chernobyl accident [64]. A relationship between prevalence of subclinical hypothyroidism and individual ^{131}I thyroid doses due to environmental exposure has been reported in a very large cohort of people exposed to the post-Chernobyl radioactive fallout during childhood. However, the same authors suggest further prospective studies since the radiation increase in hypothyroidism was quite small (10% per gray) [65]. In this regard it is worth noting that autoimmune hypothyroidism can naturally take place over decades [66] and, consequently, an unexposed age- and sex-matched control group should be analysed, especially as the cohort mean age increases. Furthermore, it should also be taken into account that differences in other environmental factors, such as iodine deficiency, may play some roles in favouring the development of autoimmune phenomena [67].

REFERENCES

1. Duffy BJ, Fitzgerald PJ. Thyroid cancer in childhood and adolescents: a report on twenty-eight cases. *Cancer*, 1950; **3**: 1018–32.
2. Lee W, Chiacchierini RP, Shleien B, Telles NC. Thyroid tumors following 131I or localized X irradiation to the thyroid and pituitary glands in rats. *Radiat Res*, 1982; **92**: 307–19.
3. Socolow EL, Hashizume A, Neriishi S, Niitani R. Thyroid carcinoma in man after exposure to ionizing radiation. A summary of the findings in Hiroshima and Nagasaki. *N Engl J Med*, 1963; **268**: 406–10.
4. Shore RE. Issues and epidemiological evidence regarding radiation-induced thyroid cancer. *Radiat Res*, 1992; **131**: 98–111.

5. National Council of Radiation Protection and Measurements. *Induction of Thyroid Cancer by Ionizing Radiation*. Bethesda: NCRP Publications, 1985: (NCRP Report N. 80).
6. Ron E, Lubin JH, Shore RE, *et al*. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res*, 1995; **141**: 259–77.
7. Thompson DE, Mabuchi K, Ron E, *et al*. Cancer incidence in atomic bomb survivors. Part II: solid tumors, 1958–1987. *Radiat Res*, 1994; **137** (Suppl 2): S17–67.
8. Lundell M, Hakulinen T, Holm LE. Thyroid cancer after radiotherapy for skin hemangioma in infancy. *Radiat Res*, 1994; **40**: 334–9.
9. Pottern LM, Kaplan MM, Larsen PR, *et al*. Thyroid nodularity after childhood irradiation for lymphoid hyperplasia: a comparison of questionnaire and clinical findings. *J Clin Epidemiol*, 1990; **43**: 449–60.
10. Shore RE, Labert RE, Pasternack BS. Follow-up study of patients treated by X-ray epilation for tinea capitis. *Arch Environ Health*, 1976; **31**: 17–24.
11. Winship T, Rosvoll RV. Thyroid carcinoma in childhood: final report on a 20 year study. *Clin Proc Child Hosp Washington DC*, 1970; **26**: 327–48.
12. Schneider AB, Ron E, Lubin J, Stovall M, Gierlowski TC. Dose–response relationship for radiation-induced thyroid cancer and thyroid nodules: evidence for prolonged effects of radiation on the thyroid. *J Clin Endocrinol Metab*, 1993; **77**: 362–9.
13. Hancock SL, Cox RS, McDougall IR. Thyroid disease after treatment of Hodgkin's disease. *N Engl J Med*, 1991; **325**: 599–605.
14. Hawkins MM, Draper GJ, Kingston JE. Incidence of second primary tumours among childhood cancer survivors. *Br J Cancer*, 1987; **56**: 339–47.
15. Tucker MA, Jones PH, Boice JD Jr, *et al*. Therapeutic radiation at a young age is linked to secondary thyroid cancer. *Cancer Res*, 1991; **51**: 2885–8.
16. Chow EJ, Friedman DL, Stovall M, *et al*. Risk of thyroid dysfunction and subsequent thyroid cancer among survivors of acute lymphoblastic leukemia: a report from the Childhood Cancer Survivor Study. *Pediatr Blood Cancer*, 2009; **53**: 432–7.
17. Nakachi K, Hayashi T, Hamatani K, Eguchi H, Kusunoki Y. Sixty years of follow-up of Hiroshima and Nagasaki survivors: current progress in molecular epidemiology studies. *Mutation Research*, 2008; **659**: 109–17.
18. Saad AG, Kumar S, Ron E, *et al*. Proliferative activity of human thyroid cells in various age groups and its correlation with the risk of thyroid cancer after radiation exposure. *J Clin Endocrinol Metab*, 2006; **91**: 2672–7.
19. Richardson DB. Exposure to ionizing radiation in adulthood and thyroid cancer incidence. *Epidemiology*, 2009; **20**: 181–7.
20. Wang Z, Boice JD Jr, Wei LX, *et al*. Thyroid nodularity and chromosome aberrations among women in areas of high background radiation in China. *J Natl Cancer Inst*, 1990; **82**: 478–85.
21. Wang JX, Inskip PD, Boice JD, Li BX, Zhang JY, Fraumeni JF. Cancer incidence among medical diagnostic X-ray workers in China, 1950 to 1985. *Int J Cancer*, 1990; **45**: 889–95.
22. Perkel VS, Gail MH, Lubin J, *et al*. Radiation-induced thyroid neoplasms: evidence for familial susceptibility factors. *J Clin Endocrinol Metab*, 1988; **66**: 1316–22.
23. Schenider AB, Shore-Freedman E, Weinstein RA. Radiation-induced thyroid and other head and neck tumors: occurrence of multiple tumors and analysis of risk factors. *J Clin Endocrinol Metab*, 1986; **63**: 107–12.
24. Detours V, Delys L, Libert F, *et al*. Genome-wide gene expression profiling suggests distinct radiation susceptibilities in sporadic and post-Chernobyl papillary thyroid cancers. *Br J Cancer*, 2007; **97**: 818–25.
25. Hall P, Mattsson A, Boice JD. Thyroid cancer after diagnostic administration of iodine-131. *Radiat Res*, 1996; **145**: 86–92.
26. Holm LE, Hall P, Wiklund K, *et al*. Cancer risk after iodine-131 therapy for hyperthyroidism. *J Natl Cancer Inst*, 1991; **83**: 1072–7.
27. Dobyns BM, Sheline GE, Workman JB, Tompkins EA, McConahey WM, Becker DV. Malignant and benign neoplasms of the thyroid in patients treated for hyperthyroidism: a report of the cooperative thyrotoxicosis therapy follow-up study. *J Clin Endocrinol Metab*, 1974; **38**: 976–98.
28. Bhattacharyya N, Chien W. Risk of second primary malignancy after radioactive iodine treatment for differentiated thyroid carcinoma. *Ann Otol Rhinol Laryngol*, 2006; **115**: 607–10.
29. Rubino C, de Vathaire F, Dottorini ME, *et al*. Second primary malignancies in thyroid cancer patients. *Br J Cancer*, 2003; **89**: 1638–44.
30. Cardis E, Howe G, Ron E, *et al*. Cancer consequences of the Chernobyl accident: 20 years on. *J Radiol Prot*, 2006; **26**: 127–40.
31. Nikiforov YE, Heffess CS, Korzenko AV, Fagin JA, Gnepp DR. Characteristics of follicular tumors and nonneoplastic thyroid lesions in children and adolescents exposed to radiation as a result of the Chernobyl disaster. *Cancer*, 1995; **76**: 900–9.
32. Cardis E, Kesminiene A, Ivanov V, *et al*. Risk of thyroid cancer after exposure to 131I in childhood. *J Natl Cancer Inst*, 2005; **97**: 724–32.
33. Karaoglou A, Desmet G, Kelly GN, Menzel HG. *The Radiological Consequences of the Chernobyl Accident.* Brussels-Luxembourg: Commission of the European Communities, 1996: (Publication EUR 16544 EN).
34. Kazakov VS, Demidchik EP, Astakhova LN. Thyroid cancer after Chernobyl. *Nature*, 1992; **359**: 21.
35. Demidchik YE, Saenko VA, Yamashita S. Childhood thyroid cancer in Belarus, Russia and Ukraine after Chernobyl and at present. *Arq Bras Endocrinol Metab*. 2007; **51**: 748–62.
36. Hatch M, Brenner A, Bogdanova T, *et al*. A screening study of thyroid cancer and other thyroid diseases among individuals exposed *in utero* to iodine-131 from Chernobyl fallout. *J Clin Endocrinol Metab*, 2009; **94**: 899–906.
37. Williams D. Twenty years' experience with post-Chernobyl thyroid cancer. *Best Pract Res Clin Endocrinol Metab*, 2008; **22**: 1061–73.
38. Pacini F, Vorontsova T, Demidchik EP, *et al*. Post-Chernobyl thyroid carcinoma in Belarus children and adolescents: comparison with naturally occurring thyroid carcinoma in Italy and France. *J Clin Endocrinol Metab*, 1997; **82**: 3563–9.
39. Ceccarelli C, Pacini F, Lippi F, *et al*. Thyroid cancer in children and adolescents. *Surgery*, 1988; **104**: 1143–7.
40. Naing S, Collins BJ, Schneider AB. Clinical behavior of radiation-induced thyroid cancer: factors related to recurrence. *Thyroid*, 2009; **19**: 479–85.
41. Fugazzola L, Pilotti S, Pinchera A, *et al*. Oncogenic rearrangements of the RET proto-oncogene in papillary thyroid carcinomas from children exposed to the Chernobyl nuclear accident. *Cancer Res*, 1995; **55**: 5617–20.
42. Klugbauer S, Lengfelder E, Demidchik EP, Rabes HM. High prevalence of RET rearrangement in thyroid tumors of children from Belarus after the Chernobyl reactor accident. *Oncogene*, 1995; **11**: 2459–67.

43. Nikiforov YE, Rowland JM, Bove KE, Monforte-Munoz H, Fagin JA. Distinct pattern of RET oncogene rearrangements in morphological variants of radiation-induced and sporadic thyroid papillary carcinomas in children. *Cancer Res*, 1997; **57**: 1690–4.
44. Elisei R, Romei C, Vorontsova T, *et al.* RET/PTC rearrangements in thyroid nodules: studies in irradiated and not irradiated, malignant and benign thyroid lesions in children and adults. *J Clin Endocrinol Metab*, 2001; **86**: 3211–16.
45. Caudill CM, Zhu Z, Ciampi R, Stringer JR, Nikiforov YE. Dose-dependent generation of RET/PTC in human thyroid cells after *in vitro* exposure to gamma-radiation: a model of carcinogenic chromosomal rearrangement induced by ionizing radiation. *J Clin Endocrinol Metab*, 2005; **90**: 2364–9.
46. Volpato CB, Martínez-Alfaro M, Corvi R, *et al.* Enhanced sensitivity of the RET proto-oncogene to ionizing radiation *in vitro*. *Cancer Res*, 2008; **68**: 8986–92.
47. Nikiforov YE, Koshoffer A, Nikiforova M, Stringer J, Fagin JA. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene*, 1999; **18**: 6330–4.
48. Bongarzone I, Fugazzola L, Vigneri P, *et al.* Age-related activation of the tyrosine kinase receptor protooncogenes ret and NTRK1 in papillary thyroid carcinoma. *J Clin Endocrinol Metab*, 1996; **81**: 2006–9.
49. Lima J, Trovisco V, Soares P, *et al.* BRAF mutations are not a major event in post-Chernobyl childhood thyroid carcinomas. *J Clin Endocrinol Metab*, 2004; **89**: 4267–71.
50. Ciampi R, Knauf JA, Kerler R, *et al.* Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest*, 2005; **115**: 94–101.
51. Port M, Boltze C, Wang Y, Röper B, Meineke V, Abend M. A radiation-induced gene signature distinguishes post-Chernobyl from sporadic papillary thyroid cancers. *Radiat Res*, 2007; **168**: 639–49.
52. Yamashita S, Suzuki S, Suzuki S, Shimura H, Saenko V. Lessons from Fukushima: latest findings of thyroid cancer after the Fukushima Nuclear Power Plant accident. *Thyroid*, 2018; **28**: 11–22.
53. Suzuki S, Bogdanova T, Saenko VA, *et al.* Histopathological analysis of papillary thyroid carcinoma detected during ultrasound screening examinations in Fukushima. *Cancer Sci*, 2019; **110**: 817–26.
54. Mitsutake N, Fukushima T, Matsuse M, *et al.* BRAF(V600E) mutation is highly prevalent in thyroid carcinomas in the young population in Fukushima: a different oncogenic profile from Chernobyl. *Sci Rep*, 2015; **5**: 16976.
55. Ishikawa T, Yasumura S, Ozasa K, *et al.* The Fukushima Health Management Survey: estimation of external doses to residents in Fukushima Prefecture. *Sci Rep*, 2015; **5**: 12712.
56. Nagataki S, Takamura N. Radioactive doses—predicted and actual—and likely health effects. *Clin Oncol*, 2016; **28**: 245–54.
57. Yamashita S, Thomas G, eds. *Thyroid Cancer and Nuclear Accidents. Long-term After Effects of Chernobyl and Fukushima*. London/San Diego/Cambridge/Oxford: Academic Press, Elsevier Inc, 2017: 221.
58. Togawa K, Ahn HS, Auvinen A, *et al.* Long-term strategies for thyroid monitoring after nuclear accidents: recommendations from an Expert Group convened by IARC. *Lancet Oncol*, 2018; **19**: 1280–3.
59. Nagataki S, Shibata Y, Inoue S, Yokoyama N, Izumi M, Shimaoka K. Thyroid disease among atomic bomb survivors in Nagasaki. *JAMA*, 1994; **272**: 364–70.
60. Imaizumi M, Usa T, Tominaga T, *et al.* Radiation dose–response relationships for thyroid nodules and autoimmune thyroid diseases in Hiroshima and Nagasaki atomic bomb survivors 55–58 years after radiation exposure. *JAMA*, 2006; **295**: 1011–22.
61. De Groot LJ, Reilly M, Pinnamaneni K, Refetoff S. Retrospective and prospective study of radiation-induced thyroid disease. *Am J Med*, 1983; **74**: 852–6.
62. Larsen PR, Conard RA, Knudsen KD, *et al.* Thyroid hypofunction after exposure to fallout from a hydrogen bomb explosion. *JAMA*, 1982; **247**: 1571–5.
63. Pacini F, Vorontsova T, Molinaro E, *et al.* Prevalence of thyroid autoantibodies in children and adolescents from Belarus exposed to the Chernobyl radioactive fallout. *Lancet*, 1998; **352**: 763–6.
64. Agate L, Mariotti S, Elisei R, *et al.* Thyroid autoantibodies and thyroid function in subjects exposed to Chernobyl fallout during childhood: evidence for a transient radiation-induced elevation of serum thyroid antibodies without an increase in thyroid autoimmune disease. *J Clin Endocrinol Metab*, 2008; **93**: 2729–36.
65. Ostroumova E, Brenner A, Oliynyk V, *et al.* Subclinical hypothyroidism after radioiodine exposure: Ukrainian-American cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident (1998–2000). *Environ Health Perspect*, 2009; **117**: 745–50.
66. Vanderpump MP, Tunbridge WM, French JM, *et al.* The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)*, 1995; **43**: 55–68.
67. Tronko MD, Brenner AV, Olijnyk VA, *et al.* Autoimmune thyroiditis and exposure to iodine 131 in the Ukrainian cohort study of thyroid cancer and other thyroid diseases after the Chernobyl accident: results from the first screening cycle (1998–2000). *J Clin Endocrinol Metab*, 2006; **91**: 4344–51.

3.2.5 Autoimmune Thyroid Disease

Anthony P. Weetman

The Spectrum of Thyroid Autoimmunity

This chapter assumes a basic knowledge of immunology; readers unfamiliar with this topic can obtain further details about the fundamental processes involved in self/non-self-discrimination by the immune system elsewhere [1]. The range of thyroid autoimmunity is shown in Table 3.2.5.1. The most frequent manifestation is probably the presence of focal thyroiditis, which can be found in around 40% of Caucasian women at autopsy, and is half as frequent in men [2]. Focal thyroiditis is often accompanied by the formation of thyroid antibodies, discussed later, but it is presently unclear whether all examples of focal thyroiditis have a truly autoimmune basis, especially if negative for thyroid antibodies. Careful longitudinal community studies have shown that individuals with positive thyroid antibodies (and presumably an underlying focal thyroiditis) have an increased risk of developing overt or clinical autoimmune hypothyroidism, which in women might be expected to occur in around 2% per year over a 20-year follow-up period [3]. In men, the risk is threefold greater. Individuals who have a sustained elevated thyroid-stimulating hormone (TSH) but normal free thyroxine (fT_4) levels, a state termed subclinical hypothyroidism, have a similar risk of progression to clinical hypothyroidism, and it may be assumed that these patients initially had focal autoimmune thyroiditis which progressed, albeit without the autoimmune response giving rise to detectable thyroid antibodies. When individuals have both subclinical hypothyroidism and positive thyroid antibodies, the relative risk of progression to clinical hypothyroidism is substantially increased, especially for men. Juvenile thyroiditis may be self-limiting.

Postpartum thyroiditis, discussed in detail in Chapter 9.5, arises from subclinical autoimmune hypothyroidism. The underlying autoimmune process is enhanced 3–6 months postpartum, for reasons which remain obscure, and at this point biochemically or clinically evident thyroid dysfunction occurs, only to remit months later as the postpartum exacerbation subsides. The occurrence of permanent clinical hypothyroidism over subsequent years in 10–20% of women presumably results from a continued and worsening autoimmune injury, as found in any type of subclinical hypothyroidism. Like postpartum thyroiditis, silent (or painless) thyroiditis causes a transient disturbance of thyroid function, most often presenting with mild destructive thyrotoxicosis followed by hypothyroidism, and indeed in the early literature, postpartum and silent thyroiditis were not distinguished. Excess iodide is an inciting factor in some cases, and others are due to inadvertent exposure to thyroid hormone (e.g. thyroid tissue contamination of meat products), but in most cases the condition seems to be a spontaneous exacerbation of an underlying autoimmune process and goitre, permanent hypothyroidism, or thyroid antibodies are present in half of such individuals several years after presentation.

The term 'Hashimoto's thyroiditis' is strictly a histological definition, with the features as described next. Clinically, patients present with a painless, lymphocytic goitre of variable size, with or without hypothyroidism, hence the alternative name, goitrous thyroiditis. Thyroid antibodies are strongly positive in almost all cases. Primary myxoedema, or atrophic thyroiditis, presents with clinical hypothyroidism, because the thyroid has usually been severely damaged by the autoimmune process, as the name implies. It seems that there is usually a continuum from one to the other, with fibrosis and follicular destruction gradually dominating in a previously lymphocytic goitre.

At first sight, Graves' disease appears as a distinct autoimmune disorder, characterized by the presence of stimulating antibodies against the TSH receptor, but it is now clear that such antibodies can also occur in occasional patients with autoimmune hypothyroidism, in whom their effects are masked by a more dominant autoimmune process leading to hypothyroidism. Moreover, up to 20% of Graves' patients treated successfully with antithyroid drugs develop spontaneous hypothyroidism over the subsequent 20 years, most likely due to supervening destructive autoimmunity. In some unusual incidences with patients, fluctuation between hyper- and hypothyroidism occurs over weeks or months, and alterations in the relative levels of TSH-receptor antibodies with stimulating and blocking capabilities may explain this phenomenon. The term 'Hashitoxicosis' is used to describe occasional patients with clinical Graves' disease but a histological picture of Hashimoto's thyroiditis, again demonstrating the close relationship between these disorders, and their sharing of common pathogenetic features.

Table 3.2.5.1 The range of thyroid autoimmunity

	Goitre	Thyroid function	Features
Focal thyroiditis	No	Normal or subclinical hypothyroidism (elevated TSH; normal free T_4)	May progress to overt hypothyroidism; associated with positive thyroid antibodies* and with thyroid cancer
Hashimoto's (or goitrous) thyroiditis	Variable size	Normal or hypothyroid (clinical or subclinical)	Almost always thyroid antibody-positive
Atrophic thyroiditis (or primary myxoedema)	No	Hypothyroid	May evolve from goitrous thyroiditis; usually thyroid antibody-positive
Silent thyroiditis	Small or absent	Transient thyrotoxicosis and/or hypothyroidism	May progress to permanent hypothyroidism; often thyroid antibody-positive
Postpartum thyroiditis	Small	Transient thyrotoxicosis and/or hypothyroidism	Occurs within a year of delivery; thyroid antibody-positive in most cases
Graves' disease	Variable size	Hyperthyroid	Associated with ophthalmopathy; positive for TSH-receptor stimulating antibodies and usually for other thyroid antibodies

*Thyroglobulin and/or thyroid peroxidase antibodies.

Pathological Features

Autoimmune Thyroiditis

In focal thyroiditis the thyroid is usually normal in size and contains foci of lymphocytes which are predominantly T cells, although lymphoid follicles can also occur. Thyroid cells adjacent to these foci are usually atrophic and deficient in colloid, but away from the foci, thyroid follicular architecture is normal [4]. Focal thyroiditis may also be prominent adjacent to a papillary carcinoma or other neoplasm. By contrast, the whole thyroid is usually involved in Hashimoto's thyroiditis. The lymphocytic infiltrate is more extensive, diffuse, and composed mainly of T cells, with prominent germinal centres containing B cells scattered through the gland (Figure 3.2.5.1). Macrophages, dendritic cells, and sometimes giant cells may be prominent. The thyroid follicles suffer variable degrees of destruction, depending largely on chronicity, and in the process undergo hyperplasia and oxyphil metaplasia, giving rise to so-called Hürthle or Askanazy cells. These cells are generally absent in juvenile autoimmune thyroiditis.

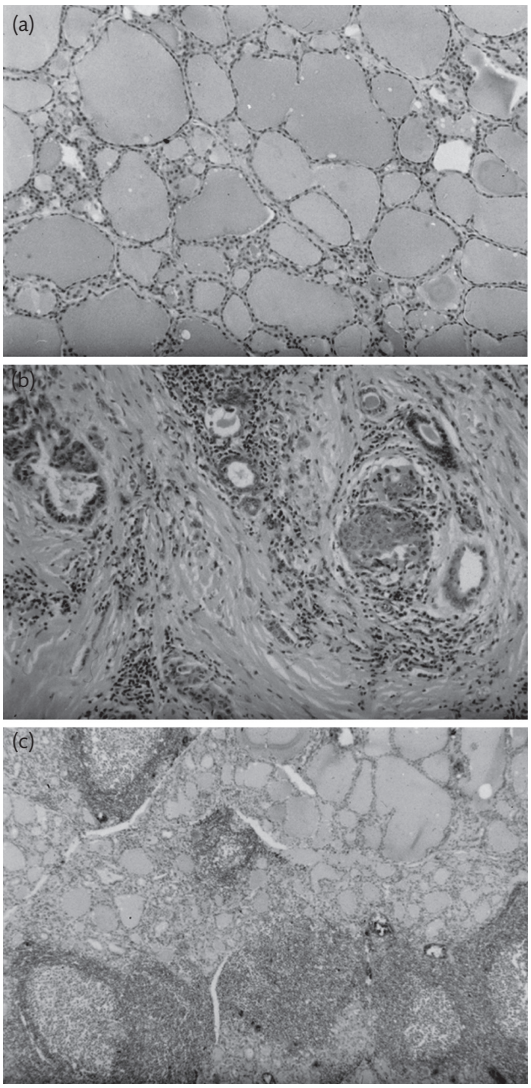


Figure 3.2.5.1 Histological features of (a) normal thyroid, (b) atrophic thyroiditis, and (c) Hashimoto's thyroiditis.
Original magnification ×100; photomicrographs courtesy of Dr K. Suvarna.

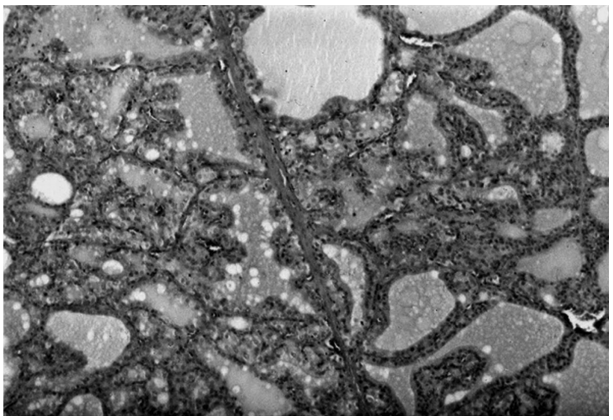


Figure 3.2.5.2 Histological features of Graves' disease.
Original magnification ×100; photomicrographs courtesy of Dr K. Suvarna.

However, the relative proportion of lymphocytic infiltrate, thyroid follicular cell change, and fibrosis varies greatly, in keeping with the suggestion made previously that there is a broad spectrum of changes which may ultimately result in atrophic thyroiditis. In this condition, the thyroid is small, has extensive fibrosis mixed with a scattered lymphocytic infiltrate, and there is a marked reduction in thyroid follicular cells. A distinct subset of patients with Hashimoto thyroiditis has been delineated recently with high circulating IgG4 levels and IgG4-positive plasma cells in the thyroid, accompanied by extensive fibrosis, and a high frequency of hypothyroidism [5]. The pathology in postpartum and silent thyroiditis generally resembles mild to moderate Hashimoto's thyroiditis, although without the oxyphil metaplasia. Germinal centres are usually absent.

Graves' Disease

It is now unusual to see the full histological picture of Graves' disease as patients are almost all treated with antithyroid drugs which diminish the lymphocytic infiltrate.(4) Even after such treatment, however, there is often a diffuse or focal lymphocytic thyroiditis, predominantly of T cells, sometimes with germinal centre formation. As an aside, lymphoid hyperplasia may also involve the lymph nodes, thymus, and spleen in Graves' disease, once again being reversed by antithyroid drugs. The thyroid follicles are both hypertrophied and hyperplastic, with scalloping and reduction in colloid (Figure 3.2.5.2). The epithelial cells are columnar and extend as papillae into the lumen. These changes are also attenuated by antithyroid drugs, so that after prolonged treatment, the colloid re-accumulates, the papillae regress, and the epithelium becomes cuboidal.

Factors Determining Susceptibility

A complex combination of genetic, environmental, and endogenous factors determines susceptibility (Figure 3.2.5.3). These factors operate differently in individuals, so that the factors leading to disease in one patient will differ from the next, which makes analysis of the importance of each factor difficult with present tools. Genetic effects are seen most clearly in children and adolescents,

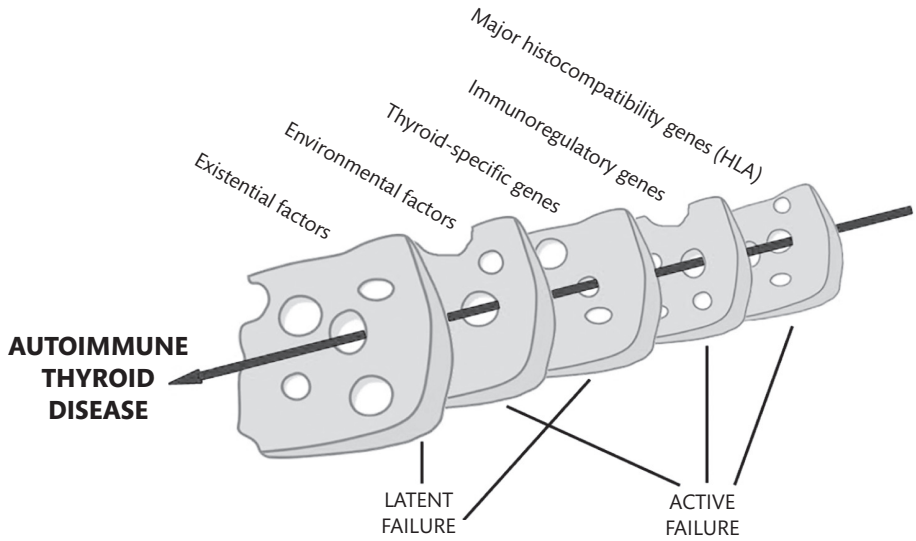


Figure 3.2.5.3 Interaction of genetic, exogenous, and endogenous or 'existential' factors in the pathogenesis of autoimmune thyroid disease. Individual factors are frequent in the general population and by themselves do not cause disease; it is their concatenation which creates the necessary conditions, as in the Swiss cheese model of accident causation.

with environmental factors having an increasing chance to operate with age.

Genetic Factors

These are dealt with extensively in Chapter 3.2.1. However, a brief discussion is given here, in relation to genetic effects on the autoimmune process. It is obvious clinically that thyroid diseases cluster in families more often than expected by chance, although the association of Graves' disease and autoimmune hypothyroidism in such families, and their coassociation with autoimmune polyglandular syndrome type 2, indicates that at least some of the susceptibility is determined by genes that control a generalized tendency to organ-specific autoimmunity. One such determinant in Caucasians is the HLA-DR3 specificity, which is associated with all of the major autoimmune endocrinopathies [6, 7].

The reason why the highly polymorphic alleles of HLA class II genes (also called major histocompatibility complex or MHC class II genes) are associated with autoimmunity is that their products are expressed by antigen-presenting cells and are crucial in initiating any immune response (Figure 3.2.5.4). Autoimmune disease may arise because a certain class II allele is able to bind and present a crucial fragment of an autoantigen, called an epitope, to a $CD4_+$ T cell. Alternatively, the effect of class II alleles in determining immune responsiveness may be exerted in the thymus during development, at which stage future autoreactive T cells may be deleted (negative selection) or allowed to develop (positive selection). Finally, some class II molecules may determine selection of regulatory T cells, and deficiencies in these cells have been postulated as a cause of autoimmunity. It still remains unclear whether other genes in linkage disequilibium with HLA-D region genes confer additional susceptibility and it is possible that class I genes have a role independent of those in the class II region.

The existence of non-HLA susceptibility genes is shown by the higher frequency of thyroid autoimmunity in monozygotic twins than in HLA-identical siblings, which in turn is higher than non-HLA-identical siblings. The critical role of CTLA-4 in costimulation of T-cell responses is discussed next and this has made it an excellent candidate to test as a susceptibility gene. It is now clear that polymorphisms in this gene have a significant role in autoimmune thyroid disease as well as several other autoimmune disorders that are associated with thyroid disease clinically. Polymorphisms in other T-cell regulatory genes, including *PTPN22* and *interleukin-2 receptor/CD25*, can similarly increase susceptibility to autoimmune thyroid disease and related disorders. Overall it is now clear that many other genes, including those that encode autoantigens such as thyroglobulin and the TSH receptor, exert small effects which contribute to these diseases and their influence varies between individuals, which in turn may explain the diverse clinical presentations of thyroid autoimmunity.

Environmental Factors

The lack of complete concordance for thyroid autoimmunity in monozygotic twins, and the clinically obvious lack of a family history in many patients with autoimmune thyroid disease, point to a role for environmental factors in determining susceptibility. Furthermore, at least part of the family clustering of disease could be the result of shared exposure to environmental triggers. Some of the best evidence for the involvement of the environmental factors comes from epidemiological changes [8] and from animal models of experimental autoimmune thyroiditis (Table 3.2.5.2) which resemble Hashimoto's thyroiditis [9]. Excess dietary iodide exacerbates the severity of the lymphocytic thyroiditis in rats with experimental autoimmune thyroiditis, and leads to enhanced production of thyroglobulin antibodies, and similar observations have

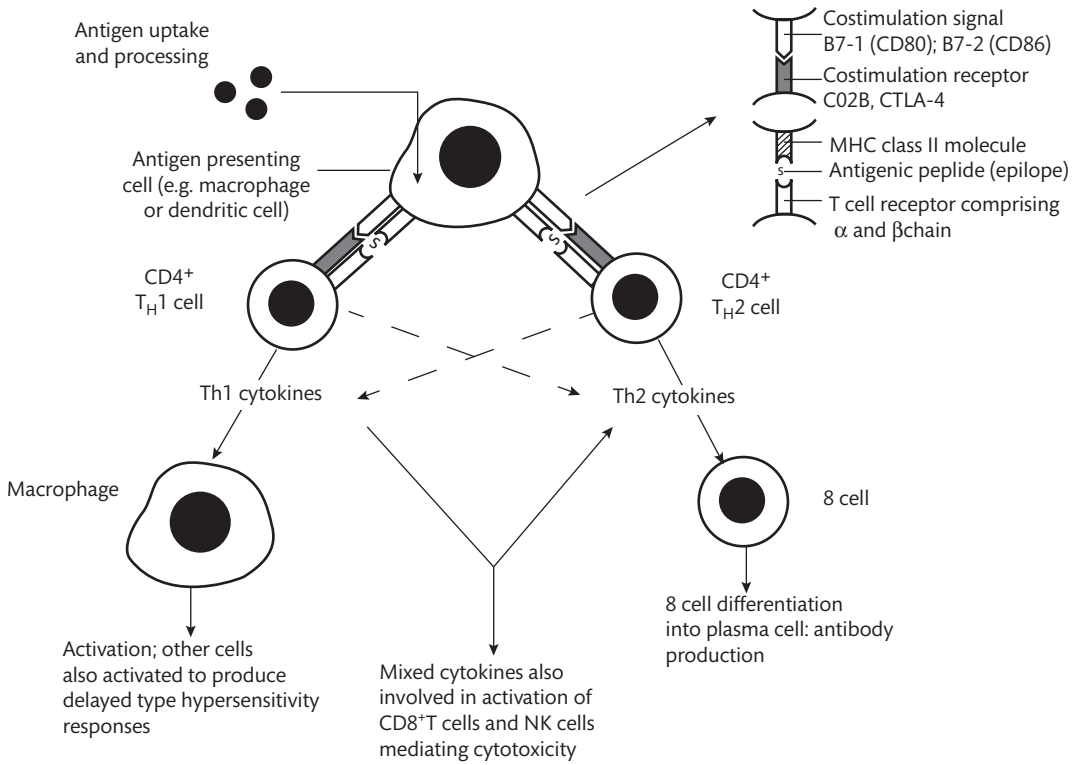


Figure 3.2.5.4 Key steps in antigen presentation and T-cell activation. The dotted line represents an inhibitory pathway.
Reproduced with permission from Weetman AP. Recent progress in autoimmune thyroid disease: an overview for the clinician. *Thyroid Today*, 1996; 19(2): 1Ð9.

been made in the obese strain chicken and non-obese diabetic (NOD) mouse which develop spontaneous autoimmune hypothyroidism. Excess iodide may act directly on the immune system, the formation of an important part of a major T-cell epitope on the iodinated thyroid antigen thyroglobulin, or the generation of toxic metabolites within the thyroid which damage thyroid cells. There is epidemiological evidence to support a similar effect of excess iodide on human autoimmune thyroiditis. Selenium deficiency may predispose to thyroid autoimmunity although trials of selenium supplementation have little obvious effect on established thyroiditis [10].

Infections could precipitate an autoimmune response by target cell damage, leading to release of autoantigens, by altering target cell expression of autoantigen or immunoregulatory molecules, such as HLA, or by molecular mimicry, in which an immune response against micro-organism antigens that resemble host autoantigens triggers an autoimmune response. Despite the appeal of the notion, and the success of animal models, there is surprisingly little evidence

Table 3.2.5.2 The main experimental models of autoimmune thyroiditis

Model	Species	Antigen	
Immunization	Mouse, rat, rabbit, guinea-pig	TG, TPO, TSH-R	Strain-dependent, transient, and transferable using T cells
Thymectomy-induced	Mouse, rat	TG	May need additional sublethal irradiation
T-cell manipulation	Mouse	TG	Transfer of specific T cells induces thyroiditis
Spontaneous	Chicken, dog, rat	Mainly TG	Thyroiditis occurs in OS chickens, beagles, NOD mice, and BB and Buffalo strain rats (NOD and BB animals have autoimmune diabetes)
Transplantation	Severe combined immunodeficiency mouse or nude mouse	TG, TPO, TSH-R	Transplanted thyroid tissues from Graves' and Hashimoto patients survive but the animal does not develop disease
cDNA immunization	Mouse, hamster	TSH-R	TSH-R antibodies and altered thyroid function produced; some recent models include orbital changes resembling ophthalmopathy
Immunization with fibroblasts transfected with TSH-R and MHC class II	Mouse	TSH-R	TSH-R antibodies and altered thyroid function produced

MHC, major histocompatibility complex; TG, thyroglobulin; TPO, thyroid peroxidase; TSH-R, thyroid-stimulating hormone receptor.

linking infection to human autoimmune disease. Autoimmune hypothyroidism occurs with increased frequency after congenital rubella infection, and some, but not all, epidemiological as well as serological studies have suggested a role for *Yersinia* infection in Graves' disease. On the other hand, studies showing a lower frequency of thyroid and other types of autoimmunity in areas with a poor standard of hygiene suggest that in some settings infections may enhance immune responses in a way that avoids the emergence of autoimmunity, perhaps through skewing of the cytokine secretion of T helper cells, discussed next [11]. Despite many attempts, no convincing role for retroviruses in autoimmune thyroid disease has been proven. Taking the opposite view, subacute thyroiditis is caused by a wide variety of viruses, and gives rise to thyroid destruction, yet rarely (if ever) triggers autoimmune thyroid disease. Only low and infrequent levels of thyroid antibodies occur in the course of infection and then disappear, although subacute thyroiditis may lead to permanent hypothyroidism in individuals who have coincidental subclinical autoimmune thyroiditis.

Stress appears to be an important precipitant of Graves' disease, possibly mediated via its neuroendocrine effects. The obese strain chicken, which develops spontaneous autoimmune thyroiditis, has an abnormal corticosteroid secretion profile that may be one of the genetic determinants of this disease [9].

As the therapeutic armamentarium expands, an increasing number of iatrogenic factors have been found to precipitate autoimmune thyroid disease. Mantle irradiation for lymphoma and other conditions is associated with an increased frequency of Graves' disease and autoimmune thyroiditis, and rare cases of Graves' disease have been reported following radioiodine treatment of nodular thyroid disease. While these examples could be the result of thyroid injury, leading to autoantigen release, the lack of a parallel response in the wake of virally induced thyroid damage suggests additional mechanisms, such as a differentially suppressive effect of radiation on critical immunoregulatory T cells. An increased prevalence of thyroid autoantibodies has also been reported in children exposed to fallout from the Chernobyl nuclear reactor explosion [12]. Lithium treatment is also associated with an increased prevalence of thyroid autoantibodies, hypothyroidism, and probably Graves' disease.

Therapeutic doses of cytokines precipitate autoimmune hypothyroidism, but rarely Graves' disease [13]. The major culprit is α-interferon, probably because it is the most extensively used, but granulocyte–macrophage colony stimulating factor, interleukin-2 (IL-2), and IL-4 have also been implicated. How these effects relate to the role of the same cytokines, at far lower endogenous concentrations, in untreated patients is unknown. However, an association exists between attacks of allergic rhinitis and the time of relapse of Graves' disease, which may well depend on the nonspecific enhancing effects of cytokines released during the allergic response. A variety of new cancer therapies, in particular kinase inhibitors (like sunitinib) and those that modulate CTLA-4 and the programmed death receptor protein PD1, are associated with destructive thyroiditis, sometimes with autoimmune features [14].

Environmental pollutants and toxins are theoretically important factors but remain underinvestigated. Administration of anthracene derivatives to genetically predisposed rats can precipitate experimental autoimmune thyroiditis. The potential of pollutants to operate in this way in man is illustrated by the association between cigarette smoking and thyroid-associated ophthalmopathy (Chapter 3.3.8) as well as, to a lesser extent, Graves' disease, whereas smoking appears to decrease the risk of Hashimoto's thyroiditis [14]. Alcohol has a modest protective effect against the development of thyroid autoimmunity [15].

Endogenous Factors

The most impressive of these is pregnancy, which can lead to postpartum thyroiditis in around 5% of ostensibly healthy women (Chapter 9.5). However, the frequency of Graves' disease is also increased in the 2 years postpartum and permanent hypothyroidism is frequently encountered after an episode of transient autoimmune hypothyroidism, indicating that pregnancy can produce a longer-lasting bias of the autoimmune response. Hyperprolactinaemia has only been inconsistently associated with an increased frequency of autoimmune thyroiditis, but clear evidence for a role of sex hormones has come from work on experimental autoimmune thyroiditis. Female animals given testosterone have a reduced frequency of thyroiditis, while castrated males, or those given oestrogen, have an increased frequency, which approaches that of females [9]. These effects explain in large part the much higher rates of autoimmune thyroid disease in women, although it remains to be seen whether any other effects are encoded on the sex chromosomes to explain this dichotomy. Alterations in T-cell regulation during pregnancy, fetal microchimerism or skewed inactivation of the X chromosome are alternative possibilities [16]. Ageing is associated with an increase in thyroid autoimmunity, although healthy centenarians may be relatively protected.

Autoantigens

There are three major autoantigens in autoimmune thyroid disease, detailed next, but there are also a number of specific and nonspecific autoantigens whose involvement is suggested by molecular cloning of candidates or by the demonstration of antibodies to cytoskeletal or nuclear components. Thyroid hormones are occasionally the target of autoantibody formation. These antibodies have no physiological consequences but can interfere in some assays for thyroid hormones, although this is now less of a problem with improved methods. The significance of autoantibodies against pendrin and the sodium iodide symporter, found in up to 10% of patients, is not yet known [17].

Thyroglobulin

Thyroglobulin is a homodimeric 660 kDa glycosylated iodoprotein which is secreted by thyroid follicular cells and stored in the luminal colloid: thyroglobulin also circulates. There are around 100 tyrosine molecules in each molecule and around 25 are normally iodinated, but this varies greatly depending on iodine uptake and thyroid activity. The iodination reaction depends on thyroid peroxidase and occurs at the apical border of the thyroid cells. Four thyroglobulin domains, termed A to D, have been identified from analysis of internal homology, and contain between them four to eight hormonogenic sites, two of which, at residues 5 and 2746, correspond to sites of preferential T_4 and T_3 synthesis, respectively. When stimulated by TSH or thyroid-stimulating antibodies, thyroglobulin is endocytosed and hydrolysed in lysosomes to release T_3 and T_4.

Although iodination of thyroglobulin plays a major role in the antigenicity of the molecule in animal models of autoimmune thyroiditis, the place of iodination in human autoimmune thyroid disease is less clear, with continuing uncertainty over whether the hormonogenic sites are part of T- or B-cell epitopes. As the immune response diversifies with time, an increasing number of epitopes are recognized, especially by sera with high levels of thyroglobulin antibodies, but patients with autoimmune thyroid disease show greater restriction of epitope recognition by autoantibodies than those who have autoantibodies but remain clinically euthyroid. These epitopes are largely conformational, although certain Hashimoto sera recognize linear determinants; all thyroglobulin antibodies cloned from patients so far recognize native but not denatured thyroglobulin. The immunopathogenic non-dominant nature of thyroid autoantibody epitopes suggests that the disease may arise from unmasking of cryptic epitopes which leads to a loss of tolerance [17].

The antibody response to thyroglobulin is relatively restricted, with a predominance of IgG1 and IgG4 subclasses and over-representation of certain immunoglobulin variable (V) genes. However, thyroglobulin antibodies, even of the IgG1 subclass, do not fix complement due to the wide spacing of epitopes, which prevents cross-linking [18]. The potential role of these antibodies in pathogenesis is considered next. Less is known about T-cell epitopes on thyroglobulin, information about which could lead to important insights regarding molecular mimicry with other self-determinants or microbial antigens.

Thyroid Peroxidase

Thyroid peroxidase is a glycosylated haemoprotein which exists in two alternatively spliced forms of 100 to 105 kDa. The predominant form, TPO-1, is responsible for tyrosine iodination and coupling to form thyroid hormones and is predominantly located at the apical border of the thyroid cell, anchored by a transmembrane segment near the carboxyl terminus, with the catalytic domain facing the follicular lumen. TPO-2 has no enzymatic activity and is restricted to the endoplasmic reticulum: its role in autoimmunity is unknown.

Initial studies of B-cell epitopes on thyroid peroxidase found two sequences, C2 (amino acids 590–622) and C21 (710–722), which are linear epitopes recognized by the majority of Hashimoto sera and a smaller proportion of Graves' sera [19]. It is likely that these and other linear determinants identified subsequently are only the target of antibodies late in disease when degradation of thyroid peroxidase allows spreading of the immune response. In the initial stages, however, conformational epitopes are probably involved in antibody binding, and these have been identified by human and mouse monoclonal antibodies. There are two large overlapping domains, A and B, which are the target of more than 80% of thyroid peroxidase antibodies in Graves' disease and Hashimoto's thyroiditis and, in the absence of thyroid peroxidase crystals, modelling has allowed prediction of the structure of these. Furthermore the immunoglobulin V gene usage of thyroid peroxidase antibodies is remarkably restricted, with domain B-binding antibodies using a particular light-chain sequence (Vκ 012), irrespective of heavy chain, although heavy chain V gene usage is also relatively restricted. Relative binding of thyroid peroxidase antibodies to the individual domains varies little over time, indicating a genetic component to the control of thyroid peroxidase antibody formation. Thyroid peroxidase antibodies in general show the same type of IgG subclass restriction as those against thyroglobulin but are able to fix complement.

T-cell epitopes are multiple and individual patients respond to different combinations of epitopes without any apparent correlation with disease type or chronicity [9]. As the T-cell response is likely to have had many months to diversify or 'spread' by the time of diagnosis, this observation is not surprising, but it does emphasize how difficult identification of any dominant epitope (which might cross-react with a microbial epitope) will be.

TSH Receptor

The TSH receptor is a typical G-protein-coupled receptor, with an extracellular domain of 398 amino acids, a transmembrane region of 266 amino acids organized in seven loops, and an intracellular domain of 93 amino acids. There are two subunits, A (55 kDa) and B (40 kDa), which correspond to the extracellular and transmembrane domains and are joined by disulphide bonds. The A-subunit, uniquely for this type of receptor, can be shed from the cell surface, which may well have immunological consequences by allowing greater access of the autoantigen to the immune system [20].

Although clearly highly expressed in the thyroid, where the receptor is fundamental for cell activation, there is now considerable evidence that the TSH receptor is expressed in fat, particularly preadipocytes, where it may make a contribution to thyroid-associated ophthalmopathy (see Chapter 3.3.10). The main physiological regulator of the TSH receptor is obviously TSH, which causes a rise in intracellular cyclic adenosine monophosphate (AMP) and these actions are mimicked by thyroid-stimulating antibodies in Graves' disease. The interaction of TSH-receptor antibodies with the receptor is even more complex, with additional antibodies blocking the effect of TSH, contributing to hypothyroidism in around 10% of patients [21], and others (neutral antibodies) do not stimulate function but may induce apoptosis, at least *in vitro* [22]. The terminology of these antibodies has been obscure, and **Table 3.2.5.3** gives an overview.

As would be predicted from the heterogeneous nature of TSH-receptor antibodies, multiple B-cell epitopes have been identified. In summary, the majority are conformational and comprise discontinuous sequences. Both stimulating and blocking antibodies bind to sites on the receptor which overlap with, but are distinct from, the TSH biding site. The greatest separation between the binding of these three entities occurs at the N-terminal region of the receptor. Much is still to be determined, including whether receptor desensitization might explain the poor correlation between circulating TSH-receptor-stimulating antibody levels and the degree of abnormal thyroid function in patients. TSH-receptor-stimulating antibodies show restriction immunoglobulin of heavy and light chain usage, implying oligoclonality of the B-cell repertoire.

TSH-receptor T-cell epitopes have been identified and, as with thyroid peroxidase, there is considerable heterogeneity both within and between patients in the regions recognized, with no clear dominant epitope. Certain TSH-receptor sequences are recognized by 10–20% of healthy individuals, but it is not known whether these represent potentially pathogenic T cells kept in check by regulatory mechanisms or low-affinity, non-specific interactions of unlikely relevance to the initiation of Graves' disease.

Table 3.2.5.3 Nomenclature and assay of the major types of TSH-receptor antibodies

Antibody	Assay
Long-acting thyroid stimulator (LATS)	The original assay for TSAb which measured the effects of TSH-R antibodies on radioiodine release in the intact mouse
LATS-protector (LATS-P)	Assayed by measuring inhibition (protection) of LATS interaction with thyroid; now superseded by new assays
Thyroid-stimulating antibodies (TSAb)	Usually measurement of cAMP production by primary cultures of thyroid cells, thyroid cell lines (for example, FRTL-5) or cells transfected with TSH-R. Other functions such as iodide uptake can be used as endpoints instead
Thyroid-blocking antibodies	Measurement of inhibition of cAMP production after TSH-mediated stimulation of thyroid cells or TSH-R transfected cells
TSH-binding inhibiting immunoglobulins (TBII)	Measurement of the inhibition of labelled TSH (or a labelled-thyroid-stimulating monoclonal antibody) binding to purified or recombinant TSH-R by serum antibodies

TSH-R, thyroid-stimulating hormone receptor.

T-cell Function in Autoimmune Thyroid Disease

Animal Models

T cells play a vital role in the pathogenesis of experimental autoimmune thyroiditis. Disease is easily transferable with T cells, whereas attempts to transfer disease using serum or antibodies produce only weak or inconsistent effects at best. Full-blown disease requires the transfer of both $CD4^+$ and $CD8^+$ cells from an animal with experimental autoimmune thyroiditis to a naive recipient (disease being established in the donor by immunization with thyroglobulin in adjuvant). However, a subpopulation within the $CD4^+$ cells also has an important regulatory function, being capable of preventing the action of thyroglobulin-specific, disease-inducing T cells [9]. In essence, these findings are consistent with a model in which autoreactive T cells are largely, but not completely, deleted or rendered anergic in the thymus during development. These T cells are normally kept in check, either because they are controlled by a regulatory T-cell subset or through clonal ignorance in which the T cells fail to react to antigen in the absence of an appropriate costimulatory signal (Figure 3.2.5.5). Animal strains particularly prone to experimental autoimmune thyroiditis have genetic defects either in

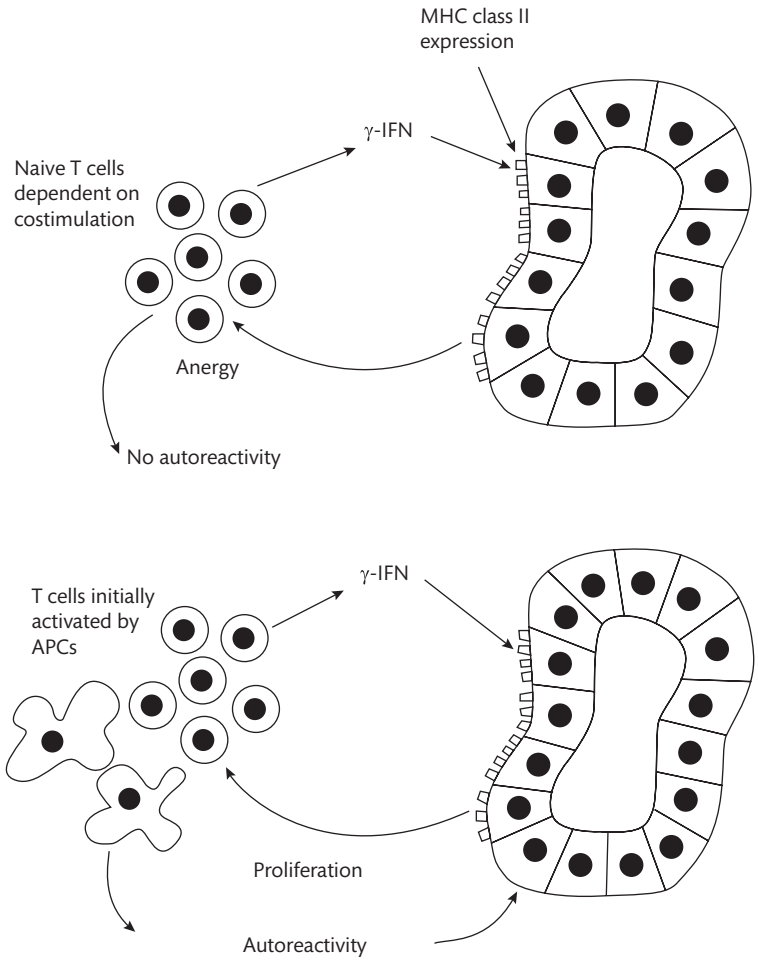


Figure 3.2.5.5 Alternative outcomes of major histocompatibility complex (MHC) class II molecule expression by thyroid cells, depending on the provision of costimulatory signals from antigen-presenting cells (APCs).

Table 3.2.5.4 Interaction of experimental manipulations in animal models of autoimmune thyroiditis

Factor	Probable site of action
Genetic background Thymectomy (irradiation) Intrathymic antigen	Thymic selection of T cells
Infection Sex hormones Adjuvant	Peripheral autoreactive T cells escaping intrathymic tolerance
Genetic background	Recognition of autoantigen
Iodide uptake	
Genetic background	T-cell-mediated regulation
Soluble autoantigen	
Thymectomy	
T-cell subset depletion	
Cytokines	
Toxins	

positive/negative selection of T cells (which make it more likely that the adult animal has sufficient autoreactive T cells to develop disease), or in the regulatory T-cell subsets, and these defects interact with environmental factors to result in disease (Table 3.2.5.4).

An appropriate balance of Th1 and Th2 cells (Table 3.2.5.5) is needed for full expression of disease, and the reciprocal inhibition between these two subsets (Figure 3.2.5.4) may be one of the most important regulatory pathways controlling the activity of autoreactive T cells. For instance, blocking IL-2 receptor activation or removing γ-interferon leads to a granulomatous rather than lymphocytic thyroiditis, and the production of high levels of thyroglobulin antibodies, due to preferential Th2 activation [23]. Typical experimental autoimmune thyroiditis is most likely Th1-dependent.

Table 3.2.5.5 Features of $CD4^+$ T-cell helper (Th) cell subsets in the mouse; similar but not identical profiles are found in humans. Further subsets are also recognized, especially Th-17 cells which secrete IL-17, TNF, and IL-6, and are highly pro-inflammatory

	Th1	Th2
Cytokine profile		
IL-2	++	–
IL-3	++	++
IL-4	–	++
IL-5	–	++
IL-6	–	++
IL-10	–	++
γ-interferon (-IFN)	++	+
Tumour necrosis factor (TNF)	++	–
Lymphotoxin	++	+
Function		
Delayed-type hypersensitivity; host defence against intracellular pathogens	++	+
B-cell help (for antibody synthesis)	+	++
Eosinophil/mast cell production; host defence against parasites and a pathological role in allergy and asthma	–	++

Further support for this T-cell-dependent mode of pathogenesis comes from the induction of experimental autoimmune thyroiditis by modulation of the T-cell repertoire alone, without the need to immunize animals with thyroid antigen (Table 3.2.5.2). Certain strains of rat or mice develop experimental autoimmune thyroiditis after thymectomy, sometimes coupled with sublethal irradiation, when performed at a critical stage of postnatal development [9], and T-cell depletion/reconstitution or cyclosporin A can have similar effects. More recently knockout mouse models have shown a crucial role both for Th1 cytokines and for the pro-inflammatory Th-17 pathway in generating thyroiditis [24].

Disease in these models is reversed by a subset of $CD4^+$ T cells from untreated donors. One major regulatory $CD4^+$ T-cell population can be identified because it expresses CD25 and Foxp3. Depletion of this T-cell subset causes severe thyroiditis in certain mouse strains and this subset also appears to be reduced when thymectomy is performed, knockout models have also confirmed the importance of T-regulatory cells in maintaining freedom from thyroiditis [24]. From these studies it is clear that thyroid-reactive T cells are present early after birth and that depletion of a critical, regulatory subset of $CD4^+$ T cells can induce organ-specific autoimmune disease. Transgenic mice have been used to confirm that tolerance, imposed in the thymus or periphery, is a major step in the production of thyroid reactivity: by contrast, B cells were not tolerized in animals overexpressing a membrane-bound antigen specifically on thyroid cells, presumably because the antigen is sequestered from B but not T cells [25]. These B cells are harmless (or 'ignorant') unless specific T cells are available in a non-tolerized state, in which case help in the form of B-cell stimulation might be provided, leading to thyroid antibody formation. The frequency of thyroid antibodies (and focal thyroiditis) in the healthy population may be due to the existence of such untolerized B cells, which can be partially activated if T-cell tolerance is disrupted or bypassed, for example by the provision of B-cell-stimulatory cytokines by non-thyroid-specific T cells.

Human Studies

The methods used to examine thyroid-reactive T cells in humans are shown in Table 3.2.5.6 and, despite their limitations, have provided important insights into the pathogenesis of autoimmune thyroid disease. A major problem has been the difficulty of access to critical thyroid-infiltrating T cells in untreated patients: blood-borne lymphocytes contain only a small proportion of thyroid-specific T cells which happen to be trafficking at the time of sampling, and although Graves' thyroid tissue is often available for study, such patients have usually received treatment with antithyroid drugs which reduce the severity of the lymphocytic infiltrate, making the remaining T cells unrepresentative. Furthermore, it is obvious that any immune response, initially directed against a single epitope on a single antigen, rapidly diversifies to involve other epitopes and antigens, and this phenomenon of determinant spreading makes any analysis of T-cell reactivity in autoimmune diseases as chronic as those affecting the thyroid very difficult to interpret.

T-cell Phenotypes

Perhaps the simplest type of analysis, but giving easily misunderstood information, is the definition of T-cell phenotypes using

Table 3.2.5.6 Methods used for examining T-cell responses to thyroid antigens

Assay	Comment
Phenotypic analysis	Measures expression of T-cell surface molecules or cytokines but provides only indirect evidence of function
Proliferation	Measures [^{3}H]thymidine incorporation after *in vitro* stimulation with antigen
Migration inhibition factor (MIF) assay	Measures production of MIF, a poorly characterized cytokine, in response to antigen; no longer in use
ELISpot assay	Measures production of cytokines (for example, IL-1, γ-IFN) by individual T cells, usually after stimulation *in vitro*
Flow cytometry after activation by antigen *in vitro*	Measures cell surface expression of markers of activation (for example, CD69)
Immunoglobulin or antibody production	An indirect assay of Th2-type responses by T cells cultured with autologous B cells
Cytotoxicity	Measures killing of target cells incubated with cytotoxic T cells

monoclonal antibodies against an array of surface molecules. From such studies on peripheral blood, it is now fairly clear that CD8$^+$ T-cell numbers are decreased in Graves' disease, active Hashimoto's thyroiditis, and postpartum thyroiditis, giving a rise in the ratio of CD4 to CD8 cells, and so-called activated T cells, expressing HLA-DR and other activation molecules, are also increased. However, the cause and meaning of these changes remain unclear, and their original interpretation as showing a defect in so-called T-suppressor cells is now regarded as naive. It should also be noted that similar changes are found in other autoimmune diseases.

Thyroid-infiltrating T cells are a mix of CD4$^+$ and CD8$^+$ cells, many expressing activation markers, and CD4$^+$ cells often predominate in Hashimoto's thyroiditis. Most of the T cells express the αβ T-cell receptor, but a minor population of uncertain significance expresses the γδ receptor. Analysis of clonality within the T-cell population expressing the αβ receptor families by the unfractionated thyroid-infiltrating T-cell population in Hashimoto's thyroiditis and Graves' disease shows no evidence of restriction, even in the activated T-cell population which might be predicted to contain the most disease-specific cells. Although it is likely that the autoimmune response begins with a clonally restricted response, this response rapidly diversifies, particularly when multiple thyroid autoantigens are known to be involved. Detailed analysis of the T-cell infiltrate in autoimmune thyroiditis shows that there is an influx of recent thymic emigrants early on in the disease process, which in turn implies that there may be some disturbance of central tolerance, in addition to a problem with peripheral tolerance, in these patients [26].

Functional Responses

Thyroglobulin-, thyroid peroxidase-, and TSH-receptor-reactive T cells can be identified in the circulating and thyroid lymphocyte populations of patients with thyroid autoimmunity but such responses tend to be weak and epitope mapping studies have generally revealed a remarkably heterogeneous response.

Recent work has identified the importance of regulatory T cells, now known to have a central role in maintaining tolerance to autoantigens [27]. Perhaps the clearest evidence for their importance comes from the rare, lethal disorder IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked) syndrome in which there are mutations in the *FOXP3* gene that result in a defect in immunoregulatory T cells which express CD25 and Foxp3. Babies with this syndrome have very early onset autoimmune disorders including thyroid disease. Further possible examples of thyroid autoimmunity appearing in the wake of a disturbance of T-cell-mediated immunoregulation occur during reconstitution of the immune system after monoclonal antibody treatment directed against lymphocytes, or after antiretroviral treatment for HIV [28].

Analysis of cytokine production in thyroid autoimmunity, either *in situ* or by cultured T cells, has shown a complex picture, with both Th1 and Th2 cytokines being present [29]. There is also an increase in pro-inflammatory Th17 production by intrathyroidal lymphocytes [30]. It is likely that a Th1 (and Th17) pattern predominates in autoimmune hypothyroidism, but the expected Th2 predominance in Graves' disease, shown by IL-4 production, is not apparent, either because the disease has been studied too late or because other cytokines known to be produced in the thyroid, such as IL-6, IL-10, and IL-13, are able to sustain antibody production. Besides CD4$^+$ T cells, CD8$^+$ T cells, macrophages, and the thyroid follicular cells themselves all contribute to the intrathyroidal cytokine profile, and the pathogenic implications of such cytokines are discussed next.

Antigen Presentation to T Cells

Antigen presentation is the fundamental first step in any immune response (**Figure 3.2.5.4**) and in most cases is believed to be a function of specialized antigen-presenting cells, such as dendritic cells, macrophages, or B cells. These have the ability to take up antigen, process it into the form of epitopes, and present the epitope, bound to a MHC class II molecule, to a CD4$^+$ T cell which recognizes this bimolecular complex through a specific T-cell receptor. In addition, a number of other molecules on the antigen-presenting cell interact with the T cell, either to stabilize this interaction or deliver additional or costimulatory signals. T cells vary in their requirement for costimulatory signalling to achieve activation; broadly speaking, naive T cells depend more on such signals than memory or activated T cells. Some antigen-presenting cell-derived signals may also mediate T-cell inhibition. For instance, the B7 surface proteins, CD80 and CD86, cause T-cell activation when they bind to CD28 on a T cell, but if they bind instead to CTLA-4, T-cell anergy ensues. Moreover, T cells dependent on costimulatory signals are rendered anergic if antigen presentation occurs in the absence of the signal. This alternative outcome from antigen presentation is an important mechanism for determining peripheral tolerance, although much remains to be learned about what determines T-cell requirements for costimulatory signals.

Against this background, the identification of MHC class II molecule expression by thyroid cells in Hashimoto's thyroiditis and Graves' disease, but not under normal conditions, was initially taken as evidence that such expression could initiate or perpetuate the autoimmune response through the presentation of thyroid antigens by thyrocytes which, in effect, had been converted to antigen-presenting cells [31]. Such class II expression is not an intrinsic property of thyroid cells in the disease state, but depends instead on the cytokine γ-interferon released by the infiltrating T cells, and therefore cannot be the initiating step in thyroid autoimmunity: in experimental autoimmune thyroiditis, when class II

molecules are expressed *de novo* on thyroid cells in transgenic mice, thyroiditis does not appear [32].

Thyroid-specific T cells can be stimulated to proliferate in response to antigen presented by class II-positive thyroid cells, but using cloned T cells it is apparent that this is not a universal property, as T cells requiring B7 costimulation cannot be stimulated by thyroid cells, which do not express B7 proteins [33]. Moreover, the T cells that fail to respond are rendered anergic, as subsequent attempts at stimulation using conventional antigen-presenting cells fail, and this is achieved by at least two mechanisms, one partially reversible by addition of appropriate cytokines (especially IL-2) and the other dependent on Fas-mediated signalling (see next). Therefore the peripheral tolerance induced by thyroid cells is complex and appears, teleologically, to be an appropriate mechanism for inducing peripheral tolerance in potentially autoreactive T cells, which could otherwise respond to released autoantigen, for instance, after viral thyroiditis (**Figure 3.2.5.5**). The local production of γ-interferon during the infection may ensure sufficient MHC class II expression by thyroid cells to ensure that autoimmune responses are not initiated, but this backfires in the setting of an already ongoing autoimmune response. In this case, conventional antigen-presenting cells provide initial costimulatory signals and the resulting T cells, no longer dependent on costimulatory signals, will be further stimulated by class II-positive thyroid cells.

B-cell Function in Autoimmune Thyroid Disease

As already discussed, B cells specific for certain thyroid antigens are not deleted during development in transgenic animal models [25]. Such ignorant but potentially autoaggressive populations of B cells may become activated non-specifically in response to the right combination of cytokines, leading to autoantibody production. It is unknown in humans which thyroid autoantigens, if any, can actually induce B-cell tolerance, either through deletion or anergy mechanisms. Judging by the frequent appearance of low levels of low-affinity IgM class thyroglobulin antibodies in healthy individuals, B cells specific for thyroglobulin are frequently not tolerized, but whether such natural autoantibodies have a pathogenic role is uncertain. Maturation of the B-cell response, leading to the production of high levels of high-affinity, IgG class thyroglobulin antibodies, requires $CD4^+$ T-cell help, and it is these antibodies that characterize autoimmune thyroid disease.

TSH receptor and thyroid peroxidase are much more localized to the thyroid than thyroglobulin and therefore might be expected to impose even less tolerance on B cells than thyroglobulin, which circulates at relatively high levels. However, little is known about the frequency of B cells with these specificities in normal individuals. *A priori,* it would seem that TSH-receptor-specific B cells are uncommon, particularly those capable of producing thyroid-stimulating antibodies, and there is even the possibility that such antibodies are the product of only a very small number of B-cell clones.

Circulating B-cell numbers are largely normal in autoimmune thyroid disease, although increases in the $CD5^+$ B-cell subset, responsible for synthesis of polyreactive natural autoantibodies, can occur. Such increases in $CD5^+$ B cells occur in other autoimmune diseases and have no known pathogenic role in thyroiditis. B cells and plasma cells are found in varying numbers in the thyroid, and may be organized in germinal follicles, especially in Hashimoto's thyroiditis. Rarely, these follicles can show light-chain restriction, from which a single dominant clone may emerge to produce non-Hodgkin's lymphoma, a recognized complication of Hashimoto's thyroiditis.

Although both blood-borne and thyroid lymphocytes can produce thyroid antibodies *in vitro* after mitogen stimulation, only the thyroid lymphocytes produce antibody spontaneously, so that the thyroid seems likely to be a major source of antibodies *in vivo*. In addition, however, the bone marrow and lymph nodes draining the thyroid are sites of thyroid antibody production. The decline in thyroid antibody production which occurs after thyroid ablation is explicable as the result of either removal of thyroidal B cells or removal of thyroid antigen and thyroid-specific helper T cells [9]. In simple terms of B-cell population size, it would seem that the thyroid is not the major site of antibody synthesis, but the real importance of this compartment may lie in the ability of B cells to take up specific autoantigen. B cells are uniquely able to amplify the T-cell response to any given autoantigen and may even break T-cell tolerance by presentation of cryptic self-epitopes generated by processing within the B cell. Thus, within the thyroid, the autoimmune response will be sustained and increased by B-cell-mediated presentation of locally derived thyroglobulin, thyroid peroxidase, and TSH receptor, and supported by the intrathyroidal production of cytokines which cause B-cell proliferation and differentiation (**Figure 3.2.5.6**). This information has been behind attempts to treat Graves' disease and ophthalmopathy with rituximab, a monoclonal antibody that depletes B cells but not plasma cells, although results so far remain rather equivocal [34].

Mechanisms Altering Thyroid Function

It is now clear that TSH-receptor-stimulating antibodies cause Graves' disease, but there is no clear correlation between the circulating level of these antibodies and the severity of hyperthyroidism. The most likely reason for this discrepancy is that humoral and cellular factors, identical to those operating in autoimmune hypothyroidism, are also active in Graves' disease, and it is the balance between the level of stimulatory antibodies and these conflicting processes, including antibodies which block the TSH receptor, that determines the degree of hyperthyroidism. As already noted, the natural history of Graves' disease tends to thyroid destruction over 20 years in a small proportion of patients [9].

The mechanisms mediating hypothyroidism are less clear, in particular with regard to the relative importance of each in the pathogenesis of thyroid cell dysfunction and destruction, and these processes are considered next.

Humoral Immunity

The role of thyroglobulin antibodies is uncertain, as they do not fix complement, but these antibodies may be involved in mediating antibody-dependent cell-mediated cytotoxicity. In this, the effector cell is a natural killer cell which binds to the antibody via Fc receptors on the natural killer cell surface. This allows the natural killer cell to destroy a specific target cell, in this case a thyroid cell, as otherwise natural killer cell-mediated destruction is not

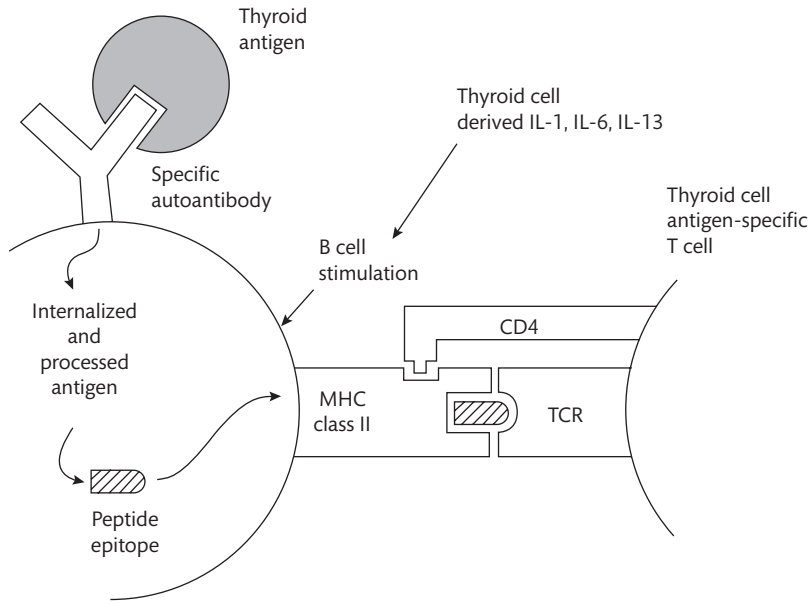


Figure 3.2.5.6 Cognate interaction of B cells, capturing specific thyroid antigens by surface autoantibody, and T cells.

restricted by recognition of specific antigen. Antibody-dependent cell-mediated cytotoxicity is demonstrable *in vitro* with both thyroglobulin and thyroid peroxidase antibodies, small numbers of natural killer cells appear in the thyroid infiltrate, and monocytes may also be involved in this destructive pathway [35]. However, transplacental transfer of thyroglobulin antibodies is not accompanied by thyroid dysfunction, and similar considerations apply to the frequent presence of thyroglobulin antibodies in euthyroid individuals. Thyroid peroxidase antibodies can fix complement but, for similar reasons, would seem to be of minor importance as primary mediators of thyroid cell destruction. Thyroid peroxidase may well be sequestered from access by autoantibodies until late in the disease process, when cell-mediated injury will permit access and antibody binding.

A second reason for the failure of complement-fixing thyroid peroxidase antibodies to destroy thyroid cells is that, in common with all nucleated cells, thyroid cells express complement regulatory proteins which prevent lethal injury by interfering with C3 convertase activity or by impairing terminal complement component formation. The most important of these regulatory proteins functionally is CD59, and its expression is upregulated by IL-1, γ-interferon, and tumour necrosis factor, all of which are produced by the lymphocytic infiltrate, thus enhancing the ability of thyroid cells to defend themselves from complement attack [36]. There is good evidence that complement is activated in thyroid autoimmunity, with elevated serum levels of terminal complement complexes, and local deposition of such complexes around the thyroid follicles in both Graves' disease and Hashimoto's thyroiditis. Unless formed in overwhelming amounts, complement membrane attack complexes do not overcome the thyroid cell's defences, but none the less, sublethal effects of complement attack are demonstrable *in vitro*, and include impaired responses to TSH stimulation and the release of cytokines, reactive oxygen metabolites, and prostaglandins, which will contribute to the local inflammatory response.[36] Antithyroid drugs block this pro-inflammatory response to complement attack, which may explain the selective immunomodulatory effects of these drugs.

A final mechanism by which antibodies can cause hypothyroidism is through their direct effects on cell function, most clearly illustrated by TSH-receptor antibodies. Although all patients with Graves' disease must, by definition, have TSH-receptor stimulating antibodies, these may be absent in the serum of occasional patients [37]. As well as assay insensitivity as an explanation, it is possible in these cases that there is exclusively intrathyroidal production of autoantibody which is sufficient to sustain disease.

Cell-Mediated Immunity

Cytokines released locally by the infiltrating lymphocytes and macrophages may have a number of effects that exacerbate thyroid injury. Some of these effects are related to the metabolic activity of the thyroid cells, such as decreased synthesis of thyroglobulin or thyroid peroxidase, which will ultimately impair thyroid hormone production (Table 3.2.5.7), while others evoke responses by

Table 3.2.5.7 Functional effects of cytokines on human thyroid cells

Cytokine	Growth	Iodide uptake	cAMP production	Expression of TG or TPO
IL-1	↑ but can also ↑ PGE_2, causing ↓ growth	↓	↓	Biphasic: ↑ at low concentration and ↓ at high concentration
IL-6	↑ (with TSH)	0	↓/0	↓/0
	↓ (with EGF)			
γ-IFN	↓ (with TNF)/0	↓	Variable	↓
TNF-α	0 (alone)	↓	↓/0	↓

↑, increase, ↓, decrease, 0, no effect. TG, thyroglobulin; TPO, thyroid peroxidase; PGE_2, prostaglandin E_2; EGF, epidermal growth factor; γ-IFN: γ-interferon; TNF-α: tumour necrosis factor-α.

thyroid cells which have direct immunological relevance. One of these has already been discussed, namely the expression of MHC class II molecules induced by γ-interferon, but many other effects are being uncovered. Adhesion molecules allow cytotoxic T cells and natural killer cells to bind initially to their targets, and the upregulation of thyroid cell adhesion molecule expression by cytokines will enhance the susceptibility of thyroid cells to such attack.[9] Nitric oxide and reactive oxygen species may play a key role in thyroid injury and their production by thyroid cells is initiated by the intrathyroidal pro-inflammatory environment which exists in autoimmune thyroiditis [9, 38]. Finally, IL-1, IL-6, IL-8, IL-12, IL-13, IL-14, IL-15, and IL-16 are all produced by thyroid cells themselves in response to inflammatory cytokines, especially IL-1[29], and this may set up a mutually reinforcing pathway of cytokine interactions which results in escalation and perpetuation of the autoimmune process (Figure 3.2.5.7).

As well as thyroid cells, vascular endothelial cells in the thyroid are exposed to cytokines which upregulate expression of selectins and other molecules essential to the egress of inflammatory cells from the blood. Thyroid cells can also produce an array of chemokines, molecules which are able to enhance the recruitment of lymphocytes to the gland in disease. Chemokine synthesis may also be critical in the formation of lymphoid germinal centres in chronically affected thyroid tissue [39]. The process of adhesion molecule expression and chemokine synthesis must be essential to the recruitment of lymphocytes to the infiltrate, although it is unknown what proportion of these are blood-derived and what proportion result from local expansion.

Specific cytotoxic T cells have long been thought to be key mediators of thyroid cell destruction in autoimmune thyroiditis, but evidence for their existence is surprisingly sparse and best documented in experimental autoimmune thyroiditis [9]. As well as releasing cytokines, cytotoxic T cells kill either by insertion of perforin into the target cell membrane, or by interaction of Fas ligand on the T-cell surface with the widely expressed Fas molecule on the target cell. Perforin-expressing T cells are present in the thyroid infiltrate in both Hashimoto's thyroiditis and Graves' disease, with slightly differing phenotypes in the two conditions [40]. This certainly indicates the potential for perforin-mediated cell destruction, although attention has also focused on Fas-mediated apoptosis as a major mechanism for thyroid cell death [41]. This interest has been sparked by the demonstration of Fas ligand expression by thyroid cells in Hashimoto's thyroiditis, but not other conditions. Fas ligand expression was enhanced *in vitro* by IL-1β but not by other cytokines, suggesting that, in addition to the classical pathway of apoptosis mediated by T cells, Fas and Fas ligand on thyroid cells could interact and lead to cell suicide. Normally Fas ligand expression is limited to sites of immunological privilege, such as the trophoblast and Sertoli cells, where it is clear that suicide is not an outcome: instead, Fas ligand expression at these sites ensures tolerance by deleting any autoaggressive Fas-expressing lymphocytes specific for these tissues. Thus, a major effect of thyroid cell Fas ligand expression *in vivo* may be the evasion of thyroid cell recognition by T cells.

In summary, thyroid cell dysfunction and destruction result from a wide array of insults (Figure 3.2.5.8) and, in the initial stage at least, seem dependent on cell-mediated autoimmune processes. It is likely that within the same clinically identified disease there are interindividual differences in the relative contributions from each type of injury. This variation would account for the diversity

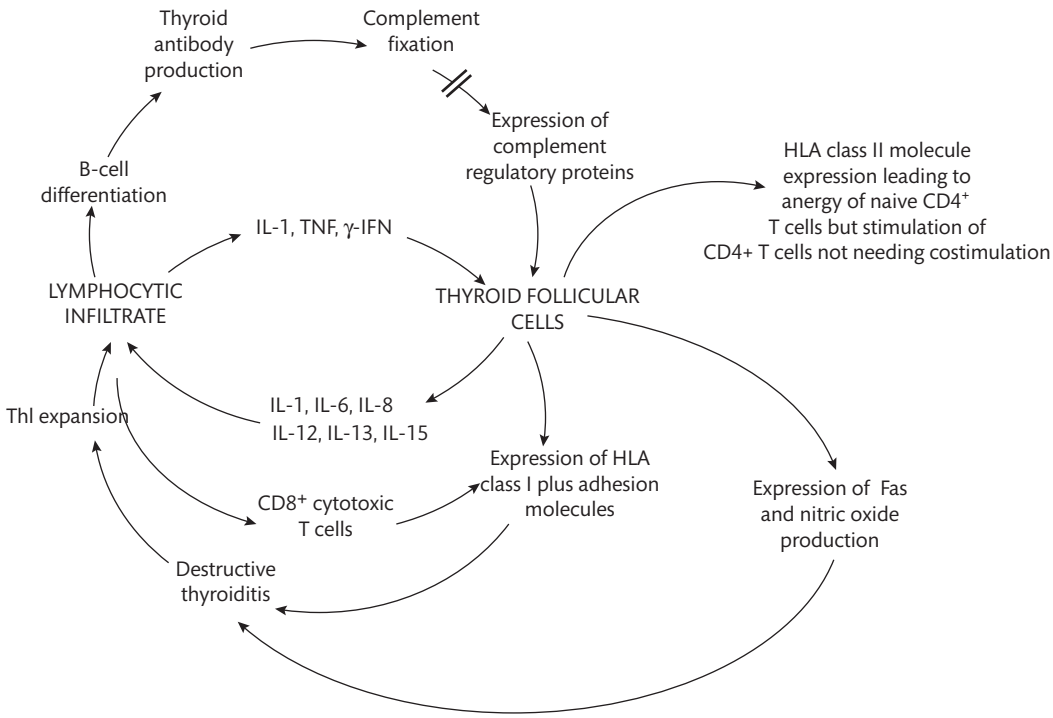


Figure 3.2.5.7 Cytokine interactions between the immune system and thyroid cells in autoimmune thyroid disease.

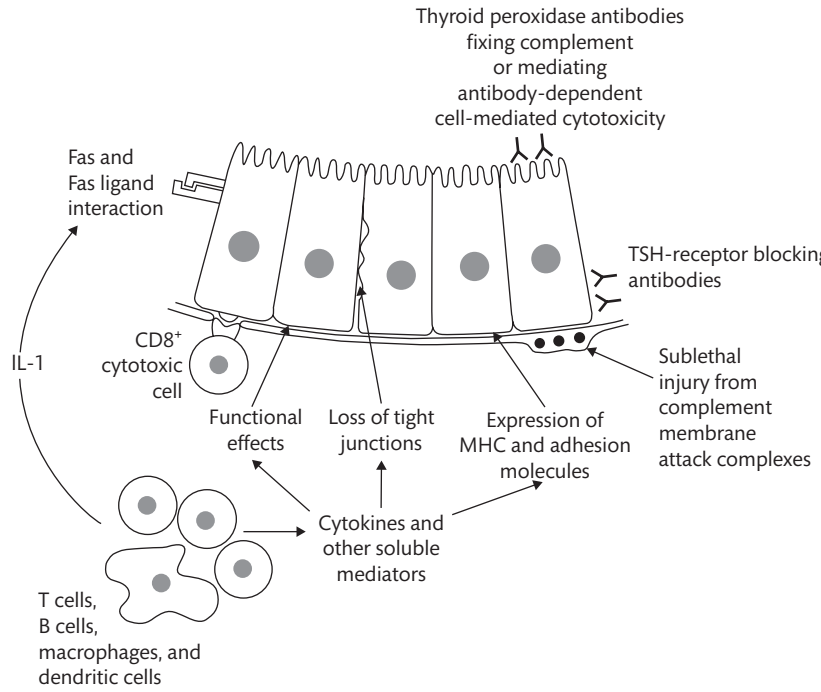


Figure 3.2.5.8 Main mechanisms involved in thyroid cell dysfunction in autoimmune hypothyroidism.

of pathological processes previously described, and because of this complexity, it is highly improbable that only two types of mechanism predominate, one resulting in atrophic thyroiditis and the other in goitrous thyroiditis.

Use of Thyroid Autoantibodies in Diagnosis

Although thyroglobulin and thyroid peroxidase antibodies appear to have a secondary rather than a primary role in disease pathogenesis, none the less they are invaluable markers of the presence of autoimmune thyroid disease. After considering the assays available, this section will review the results from antibody testing and then consider the use of TSH-receptor antibodies in diagnosis.

Thyroglobulin and Thyroid Peroxidase Antibodies

There are essentially four methods for assaying thyroglobulin and thyroid peroxidase antibodies. The two oldest are haemagglutination and indirect immunofluorescence, which depend on dilution of the test serum to determine the level of antibodies. Although robust and providing reasonably sensitive and specific results, the more modern methods of enzyme-linked immunosorbent assay (ELISA) and radioimmunoassay allow truly quantitative determination of antibodies and, in the case of assays for thyroid peroxidase antibodies, can use antigen of high purity, if necessary for research purposes. Thyroid peroxidase was previously called the microsomal antigen, and assays for these antibodies have relied on positive immunofluorescence staining with an appropriate pattern or, in the case of haemagglutination, have used an excess of thyroglobulin to absorb out thyroglobulin antibody activity when testing crude microsomal extracts of thyroid homogenate. Comparison of assays based on haemagglutination with microsomal antigen and more modern methods with purified or even recombinant thyroid peroxidase has shown a good correlation between the two, although those assays based on thyroid peroxidase are more sensitive. Further improvement in assay standardization has come from the use of reference positive serum samples, and increased sensitivity has been achieved with the use of immunometric assays, which are now automated, utilize chemiluminescence, and are widely available. Nonetheless there are still significant differences in the cut-offs for positivity between assay kits from different manufacturers [42].

With the most sensitive assays, up to 20% of healthy women have thyroglobulin and/or thyroid peroxidase antibodies, although in the majority the levels are very low. Using more conventional assays, 11% of women and 3% of men were positive in a large community-based survey in the United Kingdom [3] and similar results have been reported elsewhere. Antibodies are not entirely stable, appearing or disappearing in 17% and 2% of women, respectively, over a 20-year period [3]. In healthy individuals, the presence of such antibodies is a marker of future thyroid dysfunction, especially if coupled with subclinical hypothyroidism, and all patients with positive thyroid antibodies should be offered annual screening to detect early thyroid failure, while patients with subclinical hypothyroidism should have antibodies measured to stratify their risk.

Thyroid peroxidase antibodies are found in 80–90% of Graves' sera and 95–100% of Hashimoto sera, with thyroglobulin antibodies in up to 70% of Graves' and 90–100% of Hashimoto sera, using sensitive assays. Occasional patients with Hashimoto's thyroiditis are negative for serum thyroid antibodies, although synthesis can usually be detected locally within the thyroid, presumably at too low a level to be detectable in serum. In most patients, thyroglobulin antibodies are accompanied by thyroid peroxidase antibodies, but some patients only have one or the other type.

Box 3.2.5.1 Conditions associated with an increased prevalence of thyroglobulin and thyroid peroxidase antibodies

- Polyglandular disease
  - Insulin-dependent diabetes mellitus
  - Addison's disease
  - Premature ovarian failure
  - Lymphocytic hypophysitis
- Conditions associated with autoimmune polyglandular syndromes
  - Vitiligo
  - Pernicious anaemia
  - Alopecia
  - Myasthenia gravis
  - Coeliac disease and dermatitis herpetiformis
  - Autoimmune serositis
- Rheumatological disorders
  - Rheumatoid arthritis
  - Systemic lupus erythematosus
  - Sjögren's syndrome
  - Polymyalgia rheumatica/temporal arteritis
  - Relapsing polychondritis
  - Systemic sclerosis
- Other disorders
  - Chronic active hepatitis
  - Primary biliary cirrhosis

Thyroglobulin and thyroid peroxidase antibodies are found in a variety of other conditions at higher frequency than would be expected by chance (Box 3.2.5.1).

Antibody testing is certainly useful in patients with Addison's disease, as around 25% may develop thyroid dysfunction due to associated autoimmune polyglandular syndrome type 2. Similar considerations apply to pernicious anaemia, coeliac disease and, more debatably, other autoimmune disorders which are associated with a high frequency of thyroid autoimmunity [43]. Another situation where prospective thyroid antibody testing is particularly worthwhile is in patients starting amiodarone, as those with antibodies are more likely to develop amiodarone-induced hypothyroidism. The presence of thyroid antibodies is also useful in predicting the risk of thyroid dysfunction after treatment with γ-interferon.

On the other hand, measurement of thyroid antibodies can be misleading in goitre, as patients with multinodular goitre can have thyroid antibodies in association with focal thyroiditis, although the antibodies are usually only at low or moderate levels. Similarly, around 25% of patients with papillary or follicular thyroid cancer have thyroglobulin and/or thyroid peroxidase antibodies. There does appear to be an association between Hashimoto's thyroiditis and differentiated thyroid cancer, as well as with lymphoma: there is controversy regarding the prognostic value of thyroglobulin antibodies in thyroid cancer, possibly related to epitope recognition differences and the assays used [44]. Such antibodies can interfere with the assay of thyroglobulin in thyroid cancer follow-up.

Thyroid peroxidase antibody positivity is strongly related to postpartum thyroiditis, giving rise to the suggestion that it may be worthwhile screening all pregnant women antepartum, but the positive predictive value of thyroid peroxidase antibodies is quite low and some cases have been reported in women who are thyroid peroxidase antibody-negative. Universal screening is not currently recommended, although the TSH should be checked in all pregnant women already known to be thyroid antibody-positive [45]. Because of the high frequency of postpartum thyroiditis in type 1 diabetes mellitus and other autoimmune disorders, there is a strong case for TSH measurement (and thyroid peroxidase antibody screening if the TSH exceeds 2.5 mU/L) in this group of women antepartum. Women with positive thyroid antibodies, even without clinical thyroid dysfunction, are at risk of recurrent first trimester miscarriage, but it seems that early thyroxine treatment may not have the benefits that were originally predicted in this group [46]. There is also evidence that the presence of thyroid antibodies, in the absence of an elevated TSH, may be associated with a lack of well-being, depression, and even rare cases of encephalopathy [47]. Whether these associations are due to a direct effect of the antibodies, or the underlying effects of an autoimmune response, is not known.

TSH-Receptor Antibodies

The terminology of TSH-receptor antibodies (Table 3.2.5.3) has evolved from the methods used for their measurement. In essence there are two current methods: the binding assay, which measures the capacity of immunoglobulins to inhibit the binding of labelled TSH (or a labelled monoclonal TSH-receptor antibody) to purified or recombinant TSH receptor; and bioassays which measure the stimulatory or inhibitory effects of immunoglobulins on some aspect of thyroid cell function [48]. Generally, cyclic AMP production is used as the endpoint in bioassays, but there has been an irreversible move away from using primary cultures of animal or human thyroid cells in these assays, with their attendant problems of supply and standardization, to using either cell lines, such as rat FRTL-5 cells, or Chinese hamster ovary cells transfected with TSH receptor. With the most sensitive bioassays for TSH-receptor-stimulating antibodies almost all patients with Graves' disease are positive, but these antibodies are rarely found in the absence of Graves' disease, and then are associated with a greatly increased risk of future hyperthyroidism. This is shown most clearly by the finding that 30–50% of euthyroid patients with thyroid-associated ophthalmopathy have TSH-receptor antibodies, and this proportion increases if the most sensitive assays are used; the majority of these patients subsequently develop hyperthyroidism.

As would be predicted, there is only a weak correlation between levels of TSH-receptor antibodies measured in the binding and stimulatory bioassays. Current binding assays have a greater than 95% sensitivity and specificity for the diagnosis of Graves' disease but these antibodies must be used with knowledge of the clinical context. For instance, 10% of patients with autoimmune hypothyroidism will have TSH-receptor blocking antibodies that will be detected as positive in the binding assay, but the thyroid status makes the interpretation of the result clear [21]. Neutral antibodies with binding but not biological activity may also be detected. Antibodies against the TSH receptor are present at much lower concentrations than thyroid peroxidase antibodies and this makes the development of robust and simple solid-phase assays very difficult, compounded by problems in expressing the TSH receptor in its native form.

TSH-receptor antibody testing is not necessary for the diagnosis of Graves' disease when this is clinically obvious, for instance because there is coincident ophthalmopathy, or when such information will not influence management, for instance if the decision has already been made to proceed with radioiodine treatment. However, as the cost and precision of TSH-receptor binding assays

are now reasonable, there is an increasing tendency to use these assays to diagnose Graves' disease in cases where determining the aetiology of hyperthyroidism is required, and to rule out destructive thyroiditis. Other diagnostic tools (such as measurement of thyroid peroxidase antibodies or thyroid iodine uptake, or performing a thyroid scan) are either less convenient to undertake or less specific and sensitive. Prediction of outcome after antithyroid drugs has been another frequently suggested use for these assays, but although there is no doubt that the presence of high levels of TSH-receptor antibodies before treatment, or detectable levels after treatment, is associated with a higher rate of relapse, the sensitivity and specificity of these measurements are too poor to be used alone routinely in clinical practice [49]. However, when pretreatment serum TSH-binding inhibiting immunoglobulins (TBII) is combined with other measurements (age, goitre size, and HLA/*PTPN22* genotype), a score can be generated which has better predictive value [50].

The one clear situation where measurement of TSH-receptor antibodies is definitely indicated is during pregnancy in Graves' disease: a high level of maternal antibodies is a strong predictor of neonatal thyrotoxicosis, which occurs in 1–5% of women with Graves' disease who become pregnant. Current recommendations are that TSH-receptor antibody measurement should be made in early pregnancy in (i) women taking an antithyroid drug for Graves' disease, and (ii) women with a history of Graves' disease successfully treated by radioiodine or surgery. Measurement should be repeated at 18–22 weeks in (i) those still requiring an antithyroid drug to maintain euthyroidism, and (ii) those who were antibody-positive after surgery or radioiodine on initial screening; further tests in later pregnancy as well as fetal monitoring are required in those women who are positive for TSH-receptor antibodies at this stage [45].

REFERENCES

1. Theofilopoulos AN, Kono DH, Baccala R. The multiple pathways to autoimmunity. *Nat Immunol*, 2017; **18**: 716–24.
2. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. *N Engl J Med*, 2003; **348**: 2646–55.
3. Vanderpump MPJ, Tunbridge WM, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham survey. *Clin Endocrinol*, 1995; **43**: 55–68.
4. LiVolsi VA. The pathology of autoimmune disease: a review. *Thyroid*, 1994; **4**: 333–9.
5. Kottahachchi D, Topliss, DJ. Immunoglobulin G4-related thyroid diseases. *Eur Thyroid J*, 2016; **5**: 231–9.
6. Anderson MS. Update in endocrine autoimmunity. *J Clin Endocrinol Metab*, 2008; **93**: 3663–70.
7. McLachlan SM, Rapoport B. Breaking tolerance to thyroid antigens: changing concepts in thyroid autoimmunity. *Endocr Rev*, 2014; **35**: 59–105.
8. Caturegli P, De Remigis A, Chuang K, *et al*. Hashimoto's thyroiditis: celebrating the centennial through the lens of the Johns Hopkins hospital surgical pathology records. *Thyroid*, 2013; **23**: 142–50.
9. Weetman AP, DeGroot L. Autoimmunity to the thyroid gland, 2016. Available at: http://www.thyroidmanager.org/chapter/autoimmunity-to-the-thyroid-gland/
10. Winther KH, Bonnema SJ, Hegedüs L. Is selenium supplementation in autoimmune thyroid diseases justified? *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 348–55.
11. Kondrashova A, Viskari H, Haapala A-M, *et al*. Serological evidence of thyroid autoimmunity among schoolchildren in two different socioeconomic environments. *J Clin Endocrinol Metab*, 2008; **93**: 729–34.
12. Agate L, Mariotti S, Elisei R, *et al*. Thyroid autoantibodies and thyroid function in subjects exposed to Chernobyl fallout during childhood: evidence for a transient radiation-induced elevation of serum thyroid antibodies without an increase in thyroid autoimmune disease. *J Clin Endocrinol Metab*, 2008; **93**: 2729–36.
13. Byun DJ, Wolchok JD, Rosenberg LM, Girotra M. Cancer immunotherapy—immune checkpoint blockade and associated endocrinopathies. *Nat Rev Endocrinol*, 2017; **13**: 195–207.
14. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol*, 2013; **79**: 145–51.
15. Carle A *et al*. Moderate alcohol consumption may protect against overt autoimmune hypothyroidism: a population-based case-control study. *Eur J Endocrinol*, 2012; **167**: 483–90.
16. Weetman AP. Immunity, thyroid function and pregnancy: molecular mechanisms. *Nat Rev Endocrinol*, 2010; **6**: 311–18.
17. Brix TH, Hegedüs L, Weetman AP, Kemp HE. Pendrin and NIS antibodies are absent in healthy individuals and are rare in autoimmune thyroid disease: evidence from a Danish twin study. *Clin Endocrinol*, 2014; **81**: 440–4.
18. Gentile F, Conte M, Formisano S. Thyroglobulin as an autoantigen: what can we learn about immunopathogenicity from the correlation of antigenic properties with protein structure? *Immunology*, 2004; **112**: 13–25.
19. McLachlan SM, Rapoport B. Thyroid peroxidase as an autoantigen. *Thyroid*, 2007; **17**: 939–48.
20. Rapoport B, McLachlan SM. TSH receptor cleavage into subunits and shedding of the A-subunit; a molecular and clinical perspective. *Endocr Rev*, 2016; **37**: 114–34.
21. Diana T, Krause J, Olivo PD, *et al*. Prevalence and clinical relevance of thyroid stimulating hormone receptor-blocking antibodies in autoimmune thyroid disease. *Clin Exp Immunol*, 2017; **189**: 304–9.
22. Morshed SA, Ando T, Latif R, Davies TF. Neutral antibodies to the TSH receptor are present in Graves' disease and regulate selective signaling cascades. *Endocrinology*, 2010; **151**: 5537–49.
23. Tang H, Sharp GC, Peterson KP, Braley-Mullen H. IFN-α-deficient mice develop severe granulomatous experimental autoimmune thyroiditis in eosinophil infiltration in thyroids. *J Immunol*, 1998; **160**: 5105–12.
24. Horie I1, Abiru N, Sakamoto H, Iwakura Y, Nagayama Y. Induction of autoimmune thyroiditis by depletion of CD4+CD25+ regulatory T cells in thyroiditis-resistant IL-17, but not interferon-gamma receptor, knockout nonobese diabetic-H2h4 mice. *Endocrinology*, 2011; **152**: 4448–54.
25. Akkaraju S, Canaan K, Goodnow CC. Self-reactive B cells are not eliminated or inactivated by autoantigen expressed on thyroid epithelial cells. *J Exp Med*, 1997; **186**: 2005–12.
26. Armengol MP, Sabater L, Fernández M, *et al*. Influx of recent thymic emigrants into autoimmune thyroid disease glands in humans. *Clin Exp Immunol*, 2008; **153**: 338–50.
27. Glick AB, Wodzinski A, Fu P, Levine AD, Wald DN. Impairment of regulatory T cell function in autoimmune thyroid disease. *Thyroid*, 2013; **23**: 871–8.

28. Weetman A. Immune reconstitution syndrome and the thyroid. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 693–702.
29. Ajjan RA, Weetman AP. Cytokines in thyroid autoimmunity. *Autoimmunity*, 2003; **36**: 351–59.
30. Figueroa-Vega N, Alfonso-Pérez M, Benedicto I, *et al*. Increased circulating pro-inflammatory cytokines and Th17 lymphocytes in Hashimoto's thyroiditis. *J Clin Endocrinol Metab*, 2010; **95**, 953–62.
31. Bottazzo GF, Pujol-Borrell R, Hanafusa T, Feldmann M. Role of aberrant HLA-DR expression and antigen presentation in induction of endocrine autoimmunity. *Lancet*, 1983; **2**: 1115–19.
32. Kimura H, Kimura M, Tzou S-C, *et al*. Expression of class II major histocompatibility complex molecules on thyrocytes does not cause spontaneous thyroiditis but mildly increases its severity after immunization. *Endocrinology*, 2005; **146**: 1154–62.
33. Marelli-Berg F, Weetman A, Frasca L, *et al*. Antigen presentation by epithelial cells induces anergic immunoregulatory $CD45R0_+$ T cells and deletion of $CD45RA_+$ T cells. *J Immunol*, 1997; **159**: 5853–61.
34. Stan MN, Salvi M. Rituximab therapy for Graves' orbitopathy—lessons from randomized control trials. *Eur J Endocrinol*, 2017; **176**: R101–9.
35. Rebuffat SA, Nguyen B, Robert B, Castex F, Peraldi-Roux S. Antithyroperoxidase antibody-dependent cytotoxicity in autoimmune thyroid disease. *J Clin Endocrinol Metab*, 2008; **93**: 929–34.
36. Weetman AP, Tandon N, Morgan BP. Antithyroid drugs and release of inflammatory mediators by complement-attacked thyroid cells. *Lancet*, 1992; **340**: 633–6.
37. Vos XG, Smit N, Endert E, Tijssen JG, Wiersinga WM. Frequency and characteristics of TBII-seronegative patients in a population with untreated Graves' hyperthyroidism: a prospective study. *Clin Endocrinol*, 2008; **69**: 311–17.
38. Burek CL, Rose NR. Autoimmune thyroiditis and ROS. *Autoimmun Rev*, 2008; **7**: 530–7.
39. Liu C, Papewalis C, Domberg J, Scherbaum WA, Schott M. Chemokines and autoimmune thyroid diseases. *Horm Metab Res*, 2008: **40**: 361–8.
40. Wu Z, Podack ER, McKenzie JM, Olsen KJ, Zakarija M. Perforin expression by thyroid-infiltrating T cells in autoimmune thyroid disease. *Clin Exp Immunol*, 1994; **98**: 470–7.
41. Stassi G, De Maria R. Autoimmune thyroid disease: new models of cell death in autoimmunity. *Nat Rev Immunol*, 2002; **2**: 195–204.
42. Tozzoli R, D'Aurizio F, Ferrari A, *et al*. The upper reference limit for thyroid peroxidase autoantibodies is method-dependent: a collaborative study with biomedical industries. *Clinica Chimica Acta*, 2016: **452**: 61–5.
43. Boelaert K, Newby PR, Simmonds MJ, *et al*. Prevalence and relative risk of other autoimmune diseases in subjects with autoimmune thyroid disease. *Am J Med*, 2010; **123**: 183.e1–9.
44. McLeod DS, Cooper DS, Ladenson PW, *et al*. Prognosis of differentiated thyroid cancer in relation to serum thyrotropin and thyroglobulin antibody status at time of diagnosis. *Thyroid*, 2014; **24**: 35–42.
45. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
46. Negro R, Schwartz A, Stagnaro-Green A. Impact of levothyroxine in miscarriage and preterm delivery rates in first trimester thyroid antibody-positive women with TSH less than 2.5 mIU/L. *J Clin Endocrinol Metab*, 2016; **101**: 3685–90.
47. Leyhe T, Müssig K. Cognitive and affective dysfunctions in autoimmune thyroiditis. *Brain Behav Immun*, 2014; **41**: 261–6.
48. Ajjan RA, Weetman AP. Techniques to quantify TSH receptor antibodies. *Nat Clin Pract Endocrinol Metab*, 2008; **4**: 461–8.
49. Hesarghatta Shyamasunder A, Abraham P. Measuring TSH receptor antibody to influence treatment choices in Graves' disease. *Clin Endocrinol*, 2017; **86**: 652–7.
50. Vos XG, Endert E, Zwinderman AH, Tijssen JGP, Wiersinga WM. Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with Graves' hyperthyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 1381–9.

3.2.6 Thyroiditis

Elizabeth N. Pearce and Alan P. Farwell

Introduction

Thyroiditis comprises a diverse group of disorders that are among the most common endocrine abnormalities encountered in clinical practice. These disorders range from the extremely common chronic lymphocytic thyroiditis (Hashimoto's thyroiditis) to the extremely rare invasive fibrous thyroiditis (Riedel's thyroiditis) (Box 3.2.6.1). Clinical presentations are also diverse, ranging from

Box 3.2.6.1 Types of thyroiditis (most common to least common)

- Chronic autoimmune thyroiditis (Hashimoto's thyroiditis)
- Postpartum thyroiditis
- Sporadic silent thyroiditis
- Subacute granulomatous thyroiditis (De Quervain's thyroiditis)
- Radiation thyroiditis
- Infectious thyroiditis
- Invasive fibrous thyroiditis (Riedel's thyroiditis)
- Miscellaneous *(all are rare apart from drug-associated)*
  - Sarcoid
  - Amyloid
  - Drug-associated
  - Traumatic
  - Palpation-induced

an incidental finding of a goitre to potentially life-threatening illness, from hypothyroidism to thyrotoxicosis. The term 'thyroiditis' implies that the disorders described in this section are inflammatory processes involving the thyroid gland. However, some of the lesions are not inflammatory, but are included in the thyroiditis category largely for convenience. A rational approach to such patients, including history, physical examination, laboratory evaluation, radionuclide or ultrasonographic imaging, and fine-needle aspiration biopsy, will allow the appropriate diagnosis to be made in the vast majority of cases.

This chapter will review the following forms of thyroiditis: Hashimoto's, subacute, infectious, and Riedel's. Other forms of thyroiditis are discussed within other chapters, as follows: postpartum and painless sporadic thyroiditis (Chapter 9.4), radiation thyroiditis (Chapter 3.2.4), drug-induced thyroiditis (Chapter 3.1.4), thyroiditis associated with neoplasms (Chapter 3.5.5), and focal thyroiditis associated with non-toxic nodular goitre (Chapter 3.5.1).

Chronic Autoimmune Thyroiditis (Hashimoto's)

Chronic autoimmune thyroiditis, also known as chronic lymphocytic thyroiditis and Hashimoto's thyroiditis, was first described by Hashimoto in 1912 (Table 3.2.6.1). He described four patients with goitre, the thyroid histology of which were all characterized by diffuse lymphocytic infiltration, atrophy of parenchymal cells, fibrosis, and eosinophilic change in some of the parenchymal cells. While this condition is common, there are several variants that differ somewhat from the one initially described by Hashimoto [1]. Classically, the disorder occurs as a painless diffuse goitre (goitrous form) in a young or middle-aged woman, with or without concomitant hypothyroidism. The atrophic form of Hashimoto's thyroiditis is less common and is usually diagnosed by serology in the hypothyroid patient with a normal-sized or atrophic thyroid. The hallmarks of this disorder are high circulating titres of antibodies to thyroid peroxidase (primarily) and thyroglobulin (less often).

In iodine-sufficient countries, the most common cause of goitre, hypothyroidism, and elevated thyroid antibody levels is Hashimoto's thyroiditis. The incidence of autoimmune thyroiditis has increased over the past three generations, perhaps due to the increase in iodine intake that has occurred in many regions globally. Elevated serum thyroid antibody concentrations are found in approximately 10% of the US population and in up to 25% of US women over the age of 60 [3]. About 45% of older women will have lymphocytic infiltration within the thyroid gland. Autoimmune thyroiditis has a female predominance, with reported female to male ratios ranging between 5:1 and 9:1.

Aetiology and Pathogenesis

Hashimoto's thyroiditis is an autoimmune disease. Thyroid autoimmunity tends to aggregate in families, suggesting a genetic component. This likely results from the interplay of genes which confer susceptibility and environmental triggers. Genes regulating the immune system as well as thyroid-specific genes have been implicated

Table 3.2.6.1 Comparison between the syndromes of thyroiditis

	Hashimoto's thyroiditis	Painless sporadic/postpartum thyroiditis	Subacute thyroiditis	Infectious thyroiditis	Riedel's thyroiditis
Age of onset (years)	All ages, peak 30–50	Sporadic: all ages, peak 30–40 Postpartum: childbearing years	Peak onset 40–50	Children, 20–40	30–60
Sex ratio (F:M)	8–9:1	Sporadic—2:1	5:1	1:1	3–4:1
Frequency	General population 10%, elderly women 25%	Postpartum: 1–17% Sporadic: unknown	12.1 cases per 100 000 person-years	Rare	Extremely rare
Aetiology	Autoimmune	Autoimmune	Viral (?)	Infectious organisms	Unknown
Genetic predisposition	Moderate	Low	Moderate	Low	Low
Pathology	Lymphocytic infiltration, germinal centres, fibrosis	Lymphocytic infiltration	Giant cells, granulomas	Abscess formation	Dense fibrosis
Prodrome	None	Postpartum: pregnancy Sporadic: none	Viral illness	None	None
Goitre	Usually non-painful, persistent	Non-painful, persistent	Painful, transient	Painful, transient	Non-painful, persistent
Fever and malaise	No	No	Yes	Yes	No
Thyroid antibodies	High titre, persistent	High titre, persistent	Low titre/absent, transient	Absent	Present in most patients
Thyroid function	Euthyroid or hypothyroid	Thyrotoxicosis followed by hypothyroidism	Thyrotoxicosis followed by hypothyroidism	Usually euthyroid	Usually euthyroid
ESR	Normal	Normal	High	High	Normal
Radioactive iodine uptake (24 h)	Increased, normal, or decreased	<5%	<5%	Normal	Low/normal
Relapse	Persistent	Postpartum: common with subsequent pregnancies Sporadic: 2% [2]	1–4%	Common only with pyriform sinus fistula	Persistent
Permanent hypothyroidism	Frequent	Common	5%	Rare	Occasionally

[4]. Smoking decreases risk for hypothyroidism and thyroid autoimmunity [5]. Environmental factors which increase the risk for Hashimoto's thyroiditis include excessive iodine intake, selenium deficiency, and exposure to ionizing radiation [6]. It has been speculated that viral infection with human parvovirus B19 or hepatitis C virus may serve as a trigger for Hashimoto's thyroiditis [7].

In Hashimoto's thyroiditis T cells attack the thyroid gland, leading to exposure of thyroid antigens such as thyroid peroxidase (TPO) and thyroglobulin, against which antibodies are produced. TPO antibodies are detectable in about 90% of patients with Hashimoto's thyroiditis. Thyroglobulin antibodies are in about 35–60% of patients with Hashimoto's thyroiditis [8]. Thyroid-stimulating hormone (TSH) receptor antibodies that block TSH binding but do not stimulate thyroid cell function may play a role in the clinical presentation of Hashimoto's thyroiditis, producing or exacerbating hypothyroidism in the absence of significant thyroid gland destruction [9]. Such antibodies have been reported to bind to epitopes near the carboxyl end of the TSH receptor extracellular domain, in contrast to thyroid-stimulating antibodies, which bind to epitopes near the N-terminus [10]. The prevalence of TSH receptor blocking antibodies in adult hypothyroid patients has been reported to be as high as 10% [11] and a decrease in the titre of these antibodies is likely to be responsible for 'remission' of hypothyroidism in occasional patients with Hashimoto's thyroiditis [12]. Antibodies to colloid antigen, other thyroid autoantigens, thyroxine (T_4), and triiodothyronine (T_3), as well as other growth-promoting and inhibiting antibodies may also be present.

Pathologically, there is lymphocytic infiltration of equal proportions of T and B cells and the formation of germinal centres (Figure 3.2.6.1a). The follicular cells undergo metaplasia into larger, eosinophilic cells known as Hürthle or Askanazy cells which are packed with mitochondria. These cells exhibit high metabolic activity but ineffective hormonogenesis. There is ongoing cellular destruction and progressive fibrosis, which may be extensive. The quantity of parenchymal tissue left in the thyroid is variable, as the pathological involvement ranges from focal regions to an entire lobe to the entire gland.

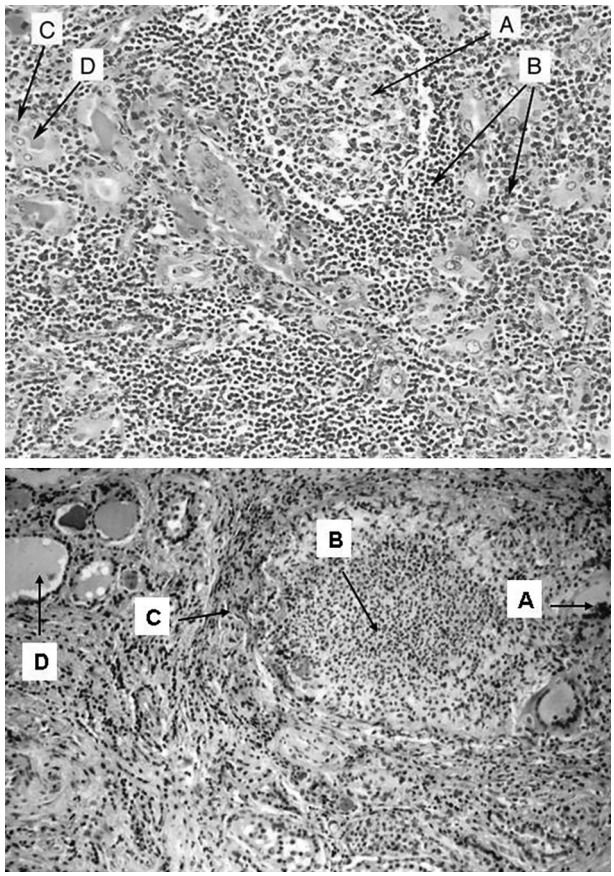


Figure 3.2.6.1 Typical pathological changes of Hashimoto's thyroiditis and subacute thyroiditis. (a) Hashimoto's thyroiditis. A, lymphoid follicle with germinal centres; B, small lymphocytes and plasma cells; C, thyroid follicles with Hürthle cell metaplasia; D, minimal colloid material. (b) Subacute thyroiditis. A, multinucleate giant cell; B, mixed inflammatory infiltrate; C, fibrous band; D, residual follicles. Haematoxylin and eosin, ×200.

Clinical Features

Hashimoto's thyroiditis occurs most frequently in middle-aged women but can occur at any age. The usual presentation is with symptoms of, or biochemical evidence for, hypothyroidism, although patients occasionally present with simple goitre. The usual course is for slow enlargement of the thyroid over years; however, the thyroid occasionally may enlarge rapidly and can produce compressive symptoms of dyspnoea and/or dysphagia. Rarely, Hashimoto's thyroiditis may be painful [13] and must be distinguished from subacute thyroiditis (see next). Systemic symptoms of hypothyroidism will be present in up to 20% of patients at the time of diagnosis [14], although this incidence is a little higher with the atrophic form of the disorder.

Physical examination typically reveals a firm bumpy non-tender goitre, which is generally symmetrical and often has a palpable pyramidal lobe. Regional lymph node enlargement may be observed. While nodular thyroid disease can, and frequently does, occur in Hashimoto's thyroiditis, suspicious nodules should be evaluated with a fine-needle aspiration biopsy to rule out a coexistent malignancy [15]. Ophthalmopathy is present in a small subset of patients with Hashimoto's thyroiditis and may be severe [16].

Laboratory Evaluation and Diagnosis

The hallmark of Hashimoto's thyroiditis is elevated TPO antibody levels. The majority of individuals with elevated TPO antibody levels are biochemically euthyroid. Up to 10% of postmenopausal women with an elevated TPO antibody level will have an increased TSH but a minority of these (about 0.5%) will have overt hypothyroidism [1]. Individuals with elevated TPO antibody levels have been reported to develop overt hypothyroidism at a rate of 2–4% per year [17] (see Chapter 3.1.8). Mild thyrotoxicosis ('Hashitoxicosis') has rarely been reported to be the initial manifestation in some patients with Hashimoto's thyroiditis [18]. The clinical course in these patients follows a pattern similar to that observed in sporadic silent or postpartum thyroiditis (Chapter 9.4), suggesting that differentiation between these disorders may be largely semantic. Finally, while the hallmark of the disease is a painless goitre, occasionally Hashimoto's thyroiditis can be quite painful.

While the diagnosis of Hashimoto's thyroiditis is confirmed by the presence of TPO antibodies, the titre of antibodies does not necessarily indicate the severity of the disease or even if hypothyroidism is present. Serum T4 and TSH concentrations depend solely on the level of thyroidal dysfunction that is present. Serum T3 concentrations are often preserved in all but the most severely hypothyroid patients and, thus, are of little clinical value. Similarly, the radioactive iodine uptake is usually not helpful, as it may be elevated, normal, or depressed. Thyroid isotope scanning usually reveals patchy uptake and, in general, provides little useful information unless a dominant thyroid nodule is present. Ultrasound examination of the thyroid frequently reveals marked hypoechogenicity with pseudonodules [19].

When imaged, an enlarged thymus gland is frequently found in Hashimoto's thyroiditis and may be important in the pathogenesis of the condition. In both affected patients and their relatives, there is an association with other autoimmune diseases including type 1 diabetes mellitus, pernicious anaemia, Addison's disease, and vitiligo. Thyroid lymphoma is rare; however, the risk is increased in those individuals with Hashimoto's thyroiditis by a factor of 67 [20]. In patients in whom a fine-needle aspiration biopsy is performed, lymphocyte subsets should be determined on the biopsy specimen if the more typical pathological features of Hashimoto's thyroiditis are not present.

Treatment

Treatment of Hashimoto's thyroiditis consists of thyroid hormone replacement if hypothyroidism is present. L-thyroxine is the hormone of choice for thyroid hormone replacement therapy because of its consistent potency and prolonged duration of action. The average daily adult replacement dose of L-thyroxine sodium is 1.6 μg/kg body weight. Institution of therapy in healthy younger individuals can begin at full replacement doses. Because of the prolonged half-life of thyroxine (7 days), new steady-state concentrations of the hormone will not be achieved until 4–6 weeks after a change in dose. Thus, re-evaluation with determination of serum TSH concentration need not be performed at intervals of less than 4–6 weeks. The goal of thyroxine replacement therapy is to achieve a TSH value in the normal range, as over-replacement of thyroxine suppressing TSH values to the subnormal range may induce osteoporosis and cause cardiac dysfunction [21]. In non-compliant young patients, the cumulative weekly doses of L-thyroxine may be given as a single weekly dose which is safe, effective, and well tolerated. In older individuals or those with known cardiovascular disease, institution of therapy at a lower daily dose of L-thyroxine (25 μg/day) is indicated to avoid exacerbation of underlying and undiagnosed cardiac disease. Daily doses of thyroxine may be interrupted periodically because of intercurrent medical or surgical illnesses that prohibit taking medications by mouth. A lapse of several days of hormone replacement is unlikely to have any significant metabolic consequences. However, if more prolonged interruption in oral therapy is necessary, L-thyroxine may be given intravenously at a dose 25–50% less than the patient's daily oral requirements. Euthyroid asymptomatic patients require monitoring, but no treatment. Recommendations for treatment of subclinical hypothyroidism (increased TSH without a corresponding low free T4 concentration) have varied, but the general consensus is that levothyroxine treatment should be initiated when the serum TSH is >10 mIU/L and treatment considered for TSH values 5–10 mIU/L, particularly in the setting of hypothyroid symptoms or detectable TPO antibodies [22].

In addition to replacement therapy, thyroid hormone therapy may be considered in patients with a serum TSH in the normal range in an attempt to decrease the size of a goitre. While goitre suppression with L-thyroxine is frequently not fruitful, in the subset of patients with Hashimoto's thyroiditis early in the course of the disease and before fibrosis develops such therapy may be useful. However, goitre suppression with L-thyroxine is unlikely to be successful if the initial TSH is less than 1 mIU/L. The goal of L-thyroxine suppression therapy is to decrease the serum TSH into the subnormal range. Patients on L-thyroxine suppression therapy should be re-evaluated periodically and the suppressive hormone dose should be reduced or discontinued if significant goitre reduction is not achieved. Surgery is occasionally indicated for compressive goitres with local obstructive symptoms.

Subacute Thyroiditis

Subacute thyroiditis, like painless sporadic and postpartum thyroiditis, is a spontaneously remitting inflammatory disorder of the thyroid that lasts for weeks to months [1, 23] (Table 3.2.6.1). This disorder has a number of eponyms, including De Quervain's thyroiditis, giant cell thyroiditis, pseudogranulomatous thyroiditis, subacute painful thyroiditis, and subacute granulomatous thyroiditis. Mygind first described subacute thyroiditis in 1895 by reporting 18 cases of 'thyroiditis akuta simplex' [23]. The pathology of subacute thyroiditis was initially described by Fritz De Quervain in 1904, when he reported giant cells and granulomatous-type changes in the thyroids of affected patients. Subacute thyroiditis is the most common cause of a painful thyroid and may account for up to 5% of clinical thyroid abnormalities [1, 23, 24]. Women are more frequently affected than men. The peak incidence is in the fourth and fifth decades [25]. This disorder is rarely observed in children and older people. Although 'subacute thyroiditis' implies a temporal quality that could apply to any inflammation of the thyroid, this term is specifically refers to the granulomatous appearance of the thyroid found on pathological examination.

Aetiology and Pathogenesis

Infectious Association

Although there is no clear evidence for a specific aetiology, indirect evidence suggests that subacute thyroiditis may be triggered by a viral infection [26]. The condition is often preceded by a prodrome consisting of myalgias, malaise, low-grade fevers, fatigue, and often by an upper respiratory tract infection. It has been found to occur seasonally with the highest incidence in the summer months, which coincide with the peak incidence of enterovirus (echovirus, Coxsackie virus A and B) infection [26]. During mumps epidemics the incidence of subacute thyroiditis has reported to increase. Interestingly, antibodies to the mumps virus have been detected in patients with subacute thyroiditis who do not have clinical evidence of mumps. Subacute thyroiditis has also been associated with measles, influenza, the common cold, adenovirus, infectious

mononucleosis, Coxsackie virus, myocarditis, cat-scratch fever, St Louis encephalitis, hepatitis A, and the parvovirus B19 infection. Antibodies to Coxsackie virus, adenovirus, influenza, and mumps have been detected in the convalescent phase of this disease [27]. Coxsackie virus is most commonly associated with subacute thyroiditis and, in fact, Coxsackie virus antibody titres have been shown to directly follow the course of the thyroid disease [26].

Autoimmune Association

Unlike painless sporadic or postpartum thyroiditis, there is no clear association between subacute thyroiditis and thyroid autoimmunity. Serum thyroid peroxidase and thyroglobulin antibody levels are usually normal. When present, antithyroid antibodies are thought to represent a transient, non-specific response to the release of thyroid antigens rather than being causal [28].

Antibodies to the thyrotropin (TSH) receptor have been detected in some patients during the course of subacute thyroiditis [29]. Although most studies have reported no correlation between the presence of thyrotropin-receptor-binding inhibitory or thyrotropin-receptor-stimulating antibodies and the thyrotoxic phase of the thyroiditis, thyroid-blocking antibodies may be associated with the development of hypothyroidism. It is thought that the appearance of the TSH receptor antibodies results from an immune response that occurs after there is damage to the thyrocytes, specifically membrane desquamation. Following recovery from the inflammatory process of subacute thyroiditis, all immunological phenomena disappear [28].

Genetic Association

There is likely a genetic predisposition for subacute thyroiditis, an association with HLA-Bw35 having been reported in multiple ethnic groups [30, 31]. A case of simultaneous development of subacute thyroiditis in identical twins, both heterozygous for the HLA-Bw35 haplotype, has been described [32]. Other genes may also be implicated, as for example, in the case of an epidemic of 'atypical' subacute thyroiditis described in Netherlands in which HLA-B15/62 was found in five of the 11 patients tested, while only one patient tested positive for HLA-Bw35 [33]. In addition, a weak association of subacute thyroiditis with HLA-DRw8 has been reported in Japanese patients [34].

Pathology

Destruction of the follicular epithelium and loss of follicular integrity are typically seen on pathological examination. The histopathological changes differ from those seen in Hashimoto's thyroiditis (**Figure 3.2.6.1b**). Lesions may be of varying stages of development and are patchy in distribution, with infiltration of mononuclear cells in affected regions, partial or complete loss of colloid, and fragmentation and duplication of the basement membrane [35]. Macrophages congregate around masses of colloid, both within the follicles and in the interstitial tissues, producing so-called giant cells that are the hallmark of subacute thyroiditis. True giant cells and granulomas do appear in this disorder as well.

During recovery, the inflammation regresses and there is a variable amount of fibrosis and fibrotic band formation. Follicular regeneration occurs without caseation, haemorrhage, or calcification. Recovery is generally complete. Only in very rare instances does complete destruction of the thyroid parenchyma lead to permanent hypothyroidism. In the few electron microscope studies reported, viral inclusion bodies have not been demonstrated [35].

Clinical Features

The manifestations may be preceded by an upper respiratory tract infection, or a prodromal phase of malaise, generalized myalgias, pharyngitis, and low-grade fevers [36]. Pain or swelling in the thyroid region usually develops 2–4 weeks later, accompanied by higher fever; approximately 50% of patients will have symptoms of thyrotoxicosis [1]. Pain and tenderness may be moderate or severe (or even exquisite). Involvement may in occur in both lobes simultaneously, or may move from one lobe to the other. Pain may radiate from the thyroid up to the angle of the jaw, to the ear, or to the anterior chest. Pain is often worsened by head movements, swallowing, or coughing. Although some patients have no systemic symptoms, many complain of myalgias, fatigue, extreme malaise, and diffuse arthralgia. The systemic reaction may be minimal or severe, and fever may be as high as 40°C. Rarely, subacute thyroiditis presents as a non-tender solitary nodule.

On physical examination, patients appear uncomfortable. The thyroid will be exquisitely tender to palpation over one or both lobes, will feel firm, and may contain palpable nodules. Overlying skin is occasionally warm and erythematous. Cervical lymphadenopathy is not typically present. While most patients are only mildly to moderately ill, subacute thyroiditis occasionally presents with marked fever, severe thyrotoxicosis, and obstructive symptoms due to pronounced thyroid inflammation and oedema.

Laboratory Evaluation

During the painful phase, the hallmark of subacute thyroiditis is an elevated erythrocyte sedimentation rate (ESR), often markedly so. In fact, a normal ESR makes the diagnosis of subacute thyroiditis unlikely. The white blood cell count is normal or mildly increased. There may be a normochromic normocytic anaemia. There are also increases in serum ferritin, soluble intercellular adhesion molecule-1, selectin, interleukin-6 levels, and C-reactive protein during the inflammatory phase [37–40]. Alkaline phosphatase and other hepatic enzymes may be elevated in the early phase [41]. It has been suggested that subacute thyroiditis may actually represent a multisystem disease also affecting the thyroid.

The thyrotoxic phase is caused by leakage of preformed thyroid hormone from the damaged thyroid gland into the circulation, a stage typically lasting from 2 to 4 months. In about one-third of patients, the initial phase is followed by a subsequent hypothyroid phase during which stored intrathyroidal hormone is depleted, but the gland has not recovered sufficiently to resume normal thyroid hormone synthesis [36]. In the thyrotoxic phase, the serum T_4 concentration will be disproportionately elevated relative to the serum T_3 concentration, with a ratio >20:1, reflecting the intrathyroidal T_4 and T_3 stores [42]. The acute illness likely also decreases the peripheral deiodination of T_4 to T_3, resulting in lower-than-expected serum T_3 concentrations. Serum TSH concentrations will be low to undetectable. Antibodies directed against thyroglobulin and TPO are either absent or present at low levels; these develop several weeks after disease onset and tend to disappear thereafter.

The radioactive iodine uptake during the thyrotoxic phase is low, usually <2% at 24 h. A normal radioactive iodine uptake during the thyrotoxic phase of the illness essentially rules out subacute thyroiditis as a diagnosis [43]. Ultrasound examination may show generalized, multiple, or single regions of hypoechogenicity [44] that can masquerade as nodules

Diagnosis

Subacute thyroiditis must be differentiated from the other aetiologies for anterior neck pain, including acute haemorrhage into a nodule or cyst, infectious thyroiditis, and rapidly enlarging thyroid carcinoma. Painful Hashimoto's thyroiditis usually involves the entire gland, with high titres of TPO and thyroglobulin antibodies. Infectious thyroiditis typically presents with a much greater leukocytosis, more inflammation in tissues surrounding the thyroid, and often evidence for extrathyroidal systemic infection. The radioactive iodine uptake is usually normal in infectious thyroiditis and the scan will reveal decreased uptake in the region of suppuration.

Rarely, rapidly infiltrating cancer of the thyroid can present with a clinical and laboratory evidence indistinguishable from subacute thyroiditis, requiring fine-needle aspiration biopsy for the diagnosis. Amiodarone may cause thyroiditis which is occasionally painful. Both sporadic silent and postpartum thyroiditis follow a similar clinical course as subacute thyroiditis but lack the clinical feature of a painful goitre. In addition, patients with painless or postpartum thyroiditis often exhibit high titres of TPO antibodies and the ESR is normal to only slightly elevated. Fine-needle aspiration biopsies may occasionally be useful to exclude other diagnoses; typical findings include follicular epithelial cells showing degenerative changes on a background of cellular debris, lymphocytes, and large numbers of macrophages [45, 46]. However, caution is occasionally needed in interpreting cytology, since the atypical cells produced by the acute inflammation may be read as papillary carcinoma [47]. Since papillary cancer is rarely an emergency, waiting 1–3 months and repeating an ultrasound may demonstrate regression of the nodule that is consistent with subacute thyroiditis rather than papillary cancer.

Course and Management

Despite the differing aetiologies, the clinical course of subacute thyroiditis is similar to that of painless sporadic and postpartum thyroiditis (see Chapter 9.4). The initial phase is characterized by pain and thyrotoxicosis in most patients and may last up to 3–4 months. Symptoms of thyrotoxicosis are usually mild, likely due to the relatively low levels of serum T3. If thyrotoxic symptoms are present, β -adrenergic blocking drugs such as propranolol or atenolol are may be used. Antithyroid drugs are contraindicated in the management of subacute thyroiditis because the gland is not hyperfunctioning.

Salicylates and non-steroidal anti-inflammatory drugs can be used for the management of thyroid pain in mild to moderate cases. In more severe cases, oral glucocorticoids provide relief of pain and swelling, usually within 24–48 hours [48]. A starting dose of 40 mg of prednisone has typically been recommended, although a recent study suggested that doses as low as 15 mg/day prednisolone may be effective for most patients, as long as the dose is tapered slowly [49]. In fact, if pain fails to begin to improve after the first 24 hours of glucocorticoid therapy, the diagnosis of subacute thyroiditis should be questioned. Despite the clinical response to corticosteroids, the underlying inflammatory process may persist, and symptoms are likely to recur if the dose is tapered too rapidly. Up to one-third of patients will have recurrent thyroid pain upon discontinuation of prednisone, which responds to restarting the corticosteroid. Full-dose corticosteroids are given for a week, followed by tapering of the dose over at least 4 weeks.

Determining the radioactive iodine uptake before discontinuation of glucocorticoids may be helpful in identifying patients at high risk for relapse. If the radioactive iodine uptake is still low, the inflammatory process is ongoing and corticosteroids should not be stopped. However, once the radioactive iodine uptake has normalized, the corticosteroid can be withdrawn safely. Patients with exacerbations of symptoms after withdrawal or tapering of corticosteroids usually respond to restarting or increasing the corticosteroids for an additional few weeks. While subacute thyroiditis is a self-limited disease and the vast majority of patients respond to the measures just discussed, occasional patients have repeated exacerbations of pain and inflammation.

The acute phase is frequently followed by a period of transient (1–2 months) asymptomatic euthyroidism. Hypothyroidism, lasting for up to 6–9 months, may be the final phase before full recovery. Hypothyroidism may be permanent in up to 5% of patients. Relapse of subacute thyroiditis is rare, occurring in less than 1–4% of patients [50]. However, some patients with a history of subacute thyroiditis were found to be particularly sensitive to the inhibitory effects of exogenously administered iodides, suggesting a persistent thyroid abnormality [51]. Thus, long-term follow-up of patients after an episode of subacute thyroiditis is recommended.

Infectious Thyroiditis

Infectious thyroiditis is also known as acute thyroiditis, suppurative thyroiditis, bacterial thyroiditis, and pyogenic thyroiditis (Table 3.2.6.1). Bacterial infections of the thyroid are rare, with only 224 cases having been reported in the literature from 1900 to 1980 [52] and only 60 cases reported in the paediatric literature [53]. Bacterial infections are the aetiology of most cases of infectious thyroiditis and the infections are generally suppurative and acute. Infectious thyroiditis caused by fungal and parasitic infections are more frequently chronic and indolent. In this section, emphasis will be placed on bacterial infections.

Aetiology and Pathogenesis

The thyroid gland's high iodine content, significant vascularity, lymphatic drainage, as well as its protective capsule provide the thyroid gland with notable resistance to infection [54]. The most common predisposing factor to infections of the thyroid appears to be pre-existing thyroid disease. Simple goitre, nodular goitre, Hashimoto's thyroiditis, or thyroid carcinoma has been observed in up to two-thirds of women and one-half of men with infectious thyroiditis [52]. Patients with HIV infection are a population particularly at risk for bacterial thyroiditis [55]. As with other opportunistic infections in patients with HIV infection, infections of the thyroid gland are often chronic and insidious in onset.

Table 3.2.6.2 Pathogenesis of acute suppurative thyroiditis

Organism	Frequency (%)	
	Prior to 1983 [52]	Current [54]
Bacterial	68	76
Parasitic	15	15
Mycobacterial	9	9
Fungal	5	<1
Syphilitic	3	0

In the adult, *Staphylococcus aureus* and *Streptococcus pyogenes* are the offending pathogens in about 40% of patients [54] (Table 3.2.6.2). In children, α - and β -haemolytic streptococcus and a variety of anaerobes account for about 70% of cases, while mixed pathogens are identified in over 50% of cases [53]. In comparison to an earlier review examining all reported cases of suppurative thyroiditis up until 1983 [56], the most recent review [53] reports that fungal infections are exceedingly rare and syphilitic infections have not been recently reported. Cases of thyroidal infections with methicillin-resistant *Staphylococcus aureus* (MRSA) have been reported [57]. Other thyroidal bacterial pathogens that have been shown to cause infectious thyroiditis include *Salmonella Actinomyces, Actinobacillus actinomycetemcomitans, Brucella melitensis, Clostridium septicum, Eikenella corrodens, Enterobacter, Escherichia coli, Haemophilus influenzae, Klebsiella, Pseudomonas aeruginosa, Serratia marcescens*, and *Acinetobacter baumannii* [54].

Infection and suppuration may result from direct spread from a nearby infection, or via the bloodstream or lymphatics. The seminal observation regarding the pathogenesis of bacterial thyroiditis was made in 1979 when Takai *et al.* reported seven cases of infectious thyroiditis due to a fistula originating from the left pyriform sinus [58]. Subsequently, multiple case series of patients with infectious thyroiditis have described pyriform sinus fistulae, primarily left-sided, especially in those with recurrent episodes. Additional reports identified infected embryonic cysts from the third and fourth brachial pouches and thyroglossal duct cysts as routes of thyroidal infection. On pathological examination, characteristic changes of acute bacterial inflammation, including necrosis and abscess formation, are frequently seen.

Clinical Manifestations

Bacterial thyroiditis is often preceded by an upper respiratory tract infection, which may induce inflammation of the fistula and promote the transmission of pathogens to the thyroid. Consistent with these observations, bacterial thyroiditis is more common in the late autumn and late spring. Over 90% of patients will present with thyroidal pain, tenderness, fever, and local compression resulting in dysphagia and dysphonia; the pain is often referred diffusely to adjacent structures. Systemic symptoms such as fever, chills, tachycardia, and malaise may be present.

Laboratory and Radiologic Findings

Thyroid function tests are usually normal [59]; however, cases of hypothyroidism and thyrotoxicosis have been reported and thyroid function should be assessed. A nuclear medicine thyroid scan may show the suppurative region as a 'cold' area, whereas an ultrasound examination may demonstrate a hypoechoic lesion [60]. The polymorphonuclear leucocyte count and the sedimentation rate are usually elevated. The organism frequently can be identified by Gram's stain and culture.

Diagnosis

The diagnosis is made with a fine-needle aspiration, Gram's stain, and culture. Symptomatically, infectious thyroiditis may be difficult to differentiate from subacute thyroiditis in the early phases, although the characteristic thyroid function changes in the latter disease should be helpful in discriminating between the two. Leucocytosis and an elevated ESR are not discriminatory tests as they are commonly observed in both subacute thyroiditis and infectious thyroiditis. In general, patients with bacterial thyroiditis have a greater febrile response than those with subacute thyroiditis. Once abscess formation has occurred, the local redness, lymphadenopathy, hyperpyrexia, and leucocytosis should lead to the correct diagnosis. Malignant neoplasms and haemorrhages into cysts may sometimes present with manifestations that mimic this disorder.

Course and Management

The prognosis of bacterial thyroiditis is often dependent on the prompt recognition and treatment of this disorder, as mortality may approach 100% if the diagnosis is delayed and appropriate antimicrobial therapy is not instituted. Much depends upon the identification of the microorganism either from needle aspirate, incision, and drainage, or occasionally from blood culture. If no organisms are seen on the Gram's stain, nafcillin and gentamicin or a third-generation cephalosporin is appropriate initial therapy in adults while clindamycin or a penicillin with a β-lactamase inhibitor is reasonable in children. If an abscess develops and prompt response to antibiotics does not occur, incision and drainage is necessary. Sometimes partial lobectomy must be performed, especially if the disease is recurrent. Usually the lesions heal with reasonable speed after initiation of the correct antimicrobial agent, and recurrences are uncommon. Mortality from acute bacterial thyroiditis has markedly improved from the 20–25% reported in the early 1900s, with an extensive review by Berger estimating an overall mortality of 8.6% [52]. A ten-year review involving over 100 patients failed to list mortality as a complication of acute bacterial thyroiditis [61].

Sclerosing Thyroiditis (Riedel's Thyroiditis)

Sclerosing thyroiditis is also historically known as invasive fibrous thyroiditis, Riedel's thyroiditis, struma fibrosa, and chronic fibrous thyroiditis. In 2011 it was linked to the newly-described entity immunoglobulin G4 (IgG4)-related disease, a systemic disorder that also includes other manifestations such as hypophysitis, parotid and lacrimal gland enlargement, interstitial pneumonitis, interstitial nephritis, retroperitoneal fibrosis, prostatitis, autoimmune pancreatitis, lymphadenopathy, inflammatory aortic aneurysm, and inflammatory pseudotumor [62]. Sclerosing thyroiditis is a rare disorder of unknown cause, characterized pathologically by dense fibrous tissue which replaces the normal thyroid parenchyma and extends into adjacent tissues, such as muscles, parathyroid glands, blood vessels, and nerves [63] (Table 3.2.6.1). The first report by Riedel in 1896 described cases of chronic sclerosing thyroiditis, primarily

affecting women, which frequently caused pressure symptoms in the neck and tended to progress ultimately to complete destruction of the thyroid gland. Riedel's interesting description was that of a 'specific inflammation of mysterious nature producing an iron-hard tumefaction of the thyroid' [64].

This condition is quite rare. Among thyroidectomies performed for all disorders, an incidence between 0.03 and 0.98% has been reported. At the Mayo Clinic, the operative incidence over 64 years was 0.06%, and the incidence in outpatients was 1.06/100 000 [65]. Because the manifestations frequently lead to surgery, the incidence of invasive fibrous thyroiditis among patients undergoing thyroidectomy is much greater than the incidence in patients with goitres in general.

Aetiology

The cause of this disorder remains unknown. Thyroid antibodies have been reported in 67–90% of patients [63, 66]. This observation, in addition to the presence of both B and T cells in the inflammatory infiltrate, suggests a possible autoimmune mechanism, although no direct relationship has been shown. Patients with concomitant invasive fibrous thyroiditis and autoimmune diseases such as type 1 diabetes mellitus, pernicious anaemia, and Addison's disease have been reported [67–69]. The expression of HLA-DR, heat-shock protein (HSP72), and soluble intercell adhesion molecule-1 (ICAM-1) receptor in invasive fibrous thyroiditis tissue suggests a role for an active cell-mediated immune response early in the evolution of this condition [70].

Marked tissue eosinophilia and eosinophil degranulation have been observed in Riedel's struma [71]. These findings may suggest that the release of eosinophil-derived products may play a role in the fibrogenic stimulus. The nature of these products is not yet known.

Whatever the ultimate aetiology is, it will have to account for the extrathyroidal fibrosclerosis and other manifestations of IgG4-related disease as well. Areas of extrathyroidal fibrosclerosis include salivary gland fibrosis, sclerosing cholangitis, pseudo tumours of the orbits, fibrous mediastinitis, retroperitoneal fibrosis, and lachrymal gland fibrosis. Long-term follow-up of patients with invasive fibrous thyroiditis has shown that about one-third develop fibrosing disorders of the retroperitoneal space (often with ureteral obstruction), chest, or orbit, almost always with a single extracervical site involved [72]. Conversely, less than 1% of patients with retroperitoneal fibrosis have invasive fibrous thyroiditis. The association of certain drugs with retroperitoneal fibrosis has not been observed with invasive fibrous thyroiditis. There does not seem to be a genetic predisposition for this condition.

Clinical Features

The age of onset varies between 25 and 81 years, although most cases are diagnosed in the fourth to sixth decades. The female to male ratio is about 4:1 [72, 73].

The clinical presentation is of a painless goitre that is gradually or rapidly enlarging; constitutional symptoms of inflammation are rare. The extensive fibrosis is progressive and may eventually cause compression of adjacent structures, particularly the trachea and oesophagus. Local compressive symptoms include a marked sense of pressure or severe dyspnoea, with symptoms out of proportion to the size of the goitre. Hoarseness may result if there is involvement of the recurrent laryngeal nerve. In some patients, the fibrotic process affects the entire gland causing hypothyroidism; the prevalence of hypothyroidism in this population is between 25 and 80% [63]. Hypoparathyroidism can develop when parathyroid gland infiltration occurs [72].

On examination, the thyroid gland is stony hard, often described as 'woody' in texture, densely adherent to adjacent cervical structures (such as muscles, blood vessels, and nerves), and may move poorly on swallowing. The lesion may be limited to one lobe. It has a harder consistency than a carcinoma and is usually non-tender. Although adjacent lymph nodes are only occasionally enlarged, when they are present a diagnosis of carcinoma is often suspected.

Laboratory and Radiologic Findings

At presentation, most patients with Riedel's thyroiditis are euthyroid; however, as noted earlier, some patients do develop hypothyroidism. Thyroid antibodies may be detected in the majority of these patients. Calcium and phosphorus levels should be evaluated at presentation to identify those patients who also have concurrent hypoparathyroidism. Thyroid radionuclide imaging can show either a heterogeneous pattern or low isotope uptake; the 'cold' areas reflect the fibrosis. The extent of the fibrosis can best be determined on either CT or MRI; the affected regions appear homogeneous and hypointense on T_1- and T_2-weighted MRI images [74]. Ultrasound examinations can be helpful as the areas affected appear hypoechoic; on colour flow Doppler, the fibrotic areas are avascular. 18F-FDG PET-CT will show intense uptake in areas of fibrotic inflammation and can be used to evaluate for the presence of extrathyroidal disease manifestations [75]. The white blood cell count and sedimentation rate are usually normal, but can be elevated. In IgG4-related disease serum IgG4 levels are typically >135 mg/dL, and levels correlate with the number of involved organs [62].

Pathology consists of an exuberant fibrosis involving part of or the entire thyroid [63]. Fibrotic extension beyond the capsule of the thyroid into adjacent structures such as nerves, blood vessels, muscles, parathyroid glands, trachea, and oesophagus is characteristic. Pathological diagnostic criteria for this condition includes complete destruction of involved thyroid tissue with absence of normal lobulation, lack of a granulomatous reaction, and extension of the fibrosis beyond the thyroid into adjacent muscle, nerves, blood vessels, and adipose. Histological examination reveals almost no thyroid follicles and few plasma cells, eosinophils, and Hürthle cells. Lymphocytes are also sparse, in contrast to the findings in Hashimoto's thyroiditis, although occasionally a few foci of lymphocytes may be observed. An associated arteritis and phlebitis with intimal proliferation, medial destruction, adventitial inflammation, and thrombosis may also occur. Similar features are observed in the extracervical fibrosclerotic lesions, retroperitoneal and mediastinal regions, orbit, and lachrymal glands, and in sclerosing cholangitis.

Diagnosis and Treatment

The diagnosis is made by biopsy of the goitre in order to differentiate this disorder from carcinoma. However, a fine-needle aspiration biopsy is usually inadequate due to the extreme hardness of the gland and, thus, an open biopsy is often required.

Treatment of Riedel's thyroiditis is surgical to relieve compressive symptoms. Extensive resection is often impossible due to fibrosis of surrounding structures, but wedge resection, especially over the isthmus to relieve tracheal compression, is often extremely effective.

Despite its invasive nature, recurrences of obstruction after resection are rare. Thyroid hormone therapy is indicated only if hypothyroidism is present, as suppression therapy is ineffective. Calcium and vitamin D therapy is indicated in those patients with associated hypoparathyroidism. There have been several reports of disease improvement with glucocorticoid therapy, and relapses have reversed with the reinstitution of steroids; however, it has not been helpful in all instances [63]. Tamoxifen has been reported to cause disease regression in a few case reports. Its mechanism of action is unclear; however, it may play a role in fibroblastic proliferation inhibition [76]. A small open-label trial demonstrated that rituximab treatment for patients with IgG4-related disease is associated with clinical improvement in the majority and remission in up to half [77].

Prognosis

Riedel's thyroiditis is usually slowly progressive; however, it may stabilize or remit spontaneously. Mortality rates range from 6% to 10%, with deaths usually attributed to asphyxia secondary to tracheal compression or laryngospasm [66]. However, these mortality rates are derived from older literature, and may not reflect (the presumably lower) current rates. In many instances, the condition is self-limiting, and improvement often persists after isthmic wedge resection. Following surgery, the disease can remit or be self-limiting. Repeat surgery is only rarely required.

REFERENCES

1. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. *N Engl J Med*, 2003; **348**: 2646–55.
2. Nikolai TF, Coombs GJ, McKenzie AK. Lymphocytic thyroiditis with spontaneously resolving hyperthyroidism and subacute thyroiditis. Long-term follow-up. *Arch Intern Med*, 1981; **141**: 1455–8.
3. Hollowell JG, Staehling NW, Flanders WD, *et al*. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
4. Lee HJ, Li CW, Hammerstad SS, Stefan M, Tomer Y. Immunogenetics of autoimmune thyroid diseases: a comprehensive review. *J Autoimmun*, 2015; **64**: 82–90.
5. Pedersen IB, Laurberg P, Knudsen N, *et al*. Smoking is negatively associated with the presence of thyroglobulin autoantibody and to a lesser degree with thyroid peroxidase autoantibody in serum: a population study. *Eur J Endocrinol*, 2008; **158**: 367–73.
6. Ferrari SM, Fallahi P, Antonelli A, Benvenga S. Environmental issues in thyroid diseases. *Front Endocrinol (Lausanne)*, 2017; **8**: 50.
7. Morohoshi K, Takahashi Y, Mori K. Viral infection and innate pattern recognition receptors in induction of Hashimoto's thyroiditis. *Discov Med*, 2011; **12**: 505–11.
8. Saravanan P, Dayan CM. Thyroid autoantibodies. *Endocrinol Metab Clin N Am*, 2001; **30**: 315–37.
9. Botero D, Brown RS. Bioassay of thyrotropin receptor antibodies with Chinese hamster ovary cells transfected with recombinant human thyrotropin receptor: clinical utility in children and adolescents with Graves' disease. *J Pediatr*, 1998; **132**: 612–18.
10. Kosugi S, Ban T, Akamizu T, Kohn LD. Identification of separate determinants on the thyrotropin receptor reactive with Graves' thyroid-stimulating antibodies and with thyroid-stimulating blocking antibodies in idiopathic myxedema: these determinants have no homologous sequence on gonadotropin receptors. *Mol Endocrinol*, 1992; **6**: 168–80.
11. Tamaki H, Amino N, Kimura M, Hidaka Y, Takeoka K, Miyai K. Low prevalence of thyrotropin receptor antibody in primary hypothyroidism in Japan. *J Clin Endocrinol Metab*, 1990; **71**: 1382–6.
12. Takasu N, Yamada T, Takasu M, *et al*. Disappearance of thyrotropin-blocking antibodies and spontaneous recovery from hypothyroidism in autoimmune thyroiditis. *N Engl J Med*, 1992; **326**: 513–18.
13. Rotondi M, Capelli V, Locantore P, Pontecorvi A, Chiovato L. Painful Hashimoto's thyroiditis: myth or reality? *J Endocrinol Invest*, 2017; **40**: 815–18.
14. Tunbridge WMG, Brewis M, French JM, *et al*. Natural history of autoimmune thyroiditis. *Br Med J*, 1981; **282**: 258–62.
15. Haugen BR, Alexander EK, Bible KC, *et al*. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2016; **26**: 1–133.
16. Bartalena L, Baldeschi L, Dickinson AJ, *et al*. Consensus statement of the European group on Graves' orbitopathy (EUGOGO) on management of Graves' orbitopathy. *Thyroid*, 2008; 18: 333–46.
17. Vanderpump MP, Tunbridge WM, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)*, 1995; **43**: 55–68.
18. Nabhan ZM, Kreher NC, Eugster EA. Hashitoxicosis in children: clinical features and natural history. *J Pediatr*, 2005; **146**: 533–6.
19. Rago T, Chiovato L, Grasso L, Pinchera A, Vitti P. Thyroid ultrasonography as a tool for detecting thyroid autoimmune diseases and predicting thyroid dysfunction in apparently healthy subjects. *J Endocrinol Invest*, 2001; **24**: 763–9.
20. Resende de Paiva C, Grønhøj C, Feldt-Rasmussen U, von Buchwald C. Association between Hashimoto's thyroiditis and thyroid cancer in 64,628 patients. *Front Oncol*, 2017; **7**: 53.
21. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet*, 2017; **390**: 1550–62.
22. Peeters RP. Subclinical Hypothyroidism. *N Engl J Med*, 2017; **376**: 2556–65.
23. Farwell AP. Sporadic painless, painful subacute and acute infectious thryoiditis. In: Braverman LE, Utiger RD, eds. *The Thyroid*, 10th edn. Philadelphia, PA: Lippincott-William & Wilkins, 2013: 414–29.
24. Golden SH, Robinson KA, Saldanha I, Anton B, Ladenson PW. Clinical review: prevalence and incidence of endocrine and metabolic disorders in the United States: a comprehensive review. *J Clin Endocrinol Metab*, 2009; **94**: 1853–78.
25. Nishihara E, Ohye H, Amino N, *et al*. Clinical characteristics of 852 patients with subacute thyroiditis before treatment. *Intern Med*, 2008; **47**: 725–9.
26. Desailloud R, Hober D. Viruses and thyroiditis: an update. *Virol J*, 2009; **6**: 5.
27. Volpe R, Row VV, Ezrin C. Circulating viral and thyroid antibodies in subacute thyroiditis. *J Clin Endocrinol Metab*, 1967; **27**: 1275–84.

28. Latrofa F, Ricci D, Montanelli L, *et al*. Thyroglobulin autoantibodies of patients with subacute thyroiditis are restricted to a major B cell epitope. *J Endocrinol Invest*, 2012; **35**: 712–14.
29. Iitaka M, Momotani N, Hisaoka T, *et al*. TSH receptor antibody-associated thyroid dysfunction following subacute thyroiditis. *Clin Endocrinol (Oxf)*, 1998; **48**: 445–53.
30. Nyulassy S, Hnilica P, Buc M, Guman M, Hirschová V, Stefanovic J. Subacute (de Quervain's) thyroiditis: association with HLA-Bw35 antigen and abnormalities of the complement system, immunoglobulins and other serum proteins. *J Clin Endocrinol Metab*, 1977; **45**: 270–4.
31. Zein EF, Karaa SE, Megarbane A. Familial occurrence of painful subacute thyroiditis associated with human leukocyte antigen-B35. *Presse Med*, 2007; **36**: 808–9.
32. Hamaguchi E, Nishimura Y, Kaneko S, Takamura T. Subacute thyroiditis developed in identical twins two years apart. *Endocr J*, 2005; **52**: 559–62.
33. de Bruin TW, Riekhoff FP, de Boer JJ. An outbreak of thyrotoxicosis due to atypical subacute thyroiditis. *J Clin Endocrinol Metab*, 1990; **70**: 396–402.
34. Goto H, Uno H, Tamai H, *et al*. Genetic analysis of subacute (de Quervain's) thyroiditis. *Tissue Antigens*, 1985; **26**: 110–13.
35. Mizukami Y, Michigishi T, Kawato M, Matsubara F. Immunohistochemical and ultrastructural study of subacute thyroiditis, with special reference to multinucleated giant cells. *Hum Pathol*, 1987; **18**: 929–35.
36. Fatourechi V, Aniszewski JP, Fatourechi GZ, Atkinson EJ, Jacobsen SJ. Clinical features and outcome of subacute thyroiditis in an incidence cohort: Olmsted County, Minnesota, study. *J Clin Endocrinol Metab*, 2003; **88**: 2100–5.
37. Pearce EN, Bogazzi F, Martino E, *et al*. The prevalence of elevated serum C-reactive protein levels in inflammatory and noninflammatory thyroid disease. *Thyroid*, 2003; **13**: 643–8.
38. Ozata M, Bolu E, Sengül A, et al. Soluble intercellular adhesion molecule-1 concentrations in patients with subacute thyroiditis and in patients with Graves' disease with or without ophthalmopathy. *Endocr J*, 1996; **43**: 517–25.
39. Sakata S, Nagai K, Maekawa H, et al. Serum ferritin concentration in subacute thyroiditis. *Metabolism*, 1991; **40**: 683–8.
40. Bartalena L, Brogioni S, Grasso L, Martino E. Increased serum interleukin-6 concentration in patients with subacute thyroiditis: relationship with concomitant changes in serum T4-binding globulin concentration. *J Endocrinol Invest*, 1993; **16**: 213–18.
41. Matsumoto Y, Amino N, Kubota S, *et al*. Serial changes in liver function tests in patients with subacute thyroiditis. *Thyroid*, 2008; **18**: 815–16.
42. Amino N, Yabu Y, Miki T, *et al*. Serum ratio of triiodothyronine to thyroxine, and thyroxine-binding globulin and calcitonin concentrations in Graves' disease and destruction-induced thyrotoxicosis. *J Clin Endocrinol Metab*, 1981; **53**: 113–16.
43. Espinoza PG, Guendelman CL, Quevedo Limón LN, Fernández RJ. A comparison between two imaging techniques for the diagnosis of subacute thyroiditis (de Quervain thyroiditis): brief communication. *Clin Nucl Med*, 2010; **35**: 862–4.
44. Omori N, Omori K, Takano K. Association of the ultrasonographic findings of subacute thyroiditis with thyroid pain and laboratory findings. *Endocr J*, 2008; **55**: 583–8.
45. Lamichaney R, Sherpa M, Das D, Bhutia CT, Laishram S. Fine-needle aspiration of De Quervain's thyroiditis (subacute granulomatous thyroiditis): a cytological review of 20 cases. *J Clin Diagn Res*, 2017; **11**: EC09–11.
46. Lu CP, Chang TC, Wang CY, Hsiao YL. Serial changes in ultrasound-guided fine needle aspiration cytology in subacute thyroiditis. *Acta Cytol*, 1997; **41**: 238–43.
47. Park HK, Kim DW, Lee YJ, *et al*. Suspicious sonographic and cytological findings in patients with subacute thyroiditis: two case reports. *Diagn Cytopathol*, 2015; **43**: 399–402.
48. Sato J, Uchida T, Komiya K, *et al*. Comparison of the therapeutic effects of prednisolone and nonsteroidal anti-inflammatory drugs in patients with subacute thyroiditis. *Endocrine*, 2017; **55**: 209–14.
49. Kubota S, Nishihara E, Kudo T, Ito M, Amino N, Miyauchi A. Initial treatment with 15 mg of prednisolone daily is sufficient for most patients with subacute thyroiditis in Japan. *Thyroid*, 2013; **23**: 269–72.
50. Iitaka M, Momotani N, Ishii J, Ito K. Incidence of subacute thyroiditis recurrences after a prolonged latency: 24-year survey. *J Clin Endocrinol Metab*, 1996; **81**: 466–9.
51. Roti E, Minelli R, Gardini E, Bianconi L, Braverman LE. Iodine-induced hypothyroidism in euthyroid subjects with a previous episode of subacute thyroiditis. *J Clin Endocrinol Metab*, 1990; **70**: 1581–5.
52. Berger SA, Zonszein J, Villamena P, Mittman N. Infectious diseases of the thyroid gland. [Review]. *Rev Infect Dis*, 1983; **5**: 108–22.
53. Rich EJ, Mendelman PM. Acute suppurative thyroiditis in pediatric patients. [Review]. *Pediatr Infect Dis J*, 1987; **6**: 936–40.
54. Paes JE, Burman KD, Cohen J, *et al*. Acute bacterial suppurative thyroiditis: a clinical review and expert opinion. *Thyroid*, 2010; **20**: 247–55.
55. N'gattia KV, Kacouchia NB, Vroh BT, Kouassi-Ndjeundo J. Suppurative thyroiditis and HIV infection. *Eur Ann Otorhinolaryngol Head Neck Dis*, 2015; **132**: 371–2.
56. Berger SA, Zonszein J, Villamena P, Mittman N. Infectious diseases of the thyroid gland. *Rev Infect Dis*, 1983; **5**: 108–22.
57. Brook I. Role of methicillin-resistant *Staphylococcus aureus* in head and neck infections. *J Laryngol Otol*, 2009; **123**: 1301–7.
58. Takai S-I, Miyauchi A, Matsuzuka F, Kuma K, Kosaki G. Internal fistula as a route of infection in acute suppurative thyroiditis. *Lancet*, 1979; **i**: 751–2.
59. Chi H, Lee YJ, Chiu NC, *et al*. Acute suppurative thyroiditis in children. *Pediatr Infect Dis J*, 2002; **21**: 384–7.
60. Masuoka H, Miyauchi A, Tomoda C, *et al*. Imaging studies in sixty patients with acute suppurative thyroiditis. *Thyroid*, 2011; **21**: 1075–80.
61. Jeng LB, Lin JD, Chen MF. Acute suppurative thyroiditis: a ten-year review in a Taiwanese hospital. *Scand J Infect Dis*, 1994; **26**: 297–300.
62. Kubo K, Yamamoto K. IgG4-related disease. *Int J Rheum Dis*, 2016; **19**: 747–62.
63. Hennessey JV. Clinical review: Riedel's thyroiditis: a clinical review. *J Clin Endocrinol Metab*, 2011; **96**: 3031–41.
64. Riedel BM. Die chronische zur Bildung eisenharter Tumoren fvehrende Entzuendung der Shilddruese. *Verh Ges Chir*, 1896; **25**: 101–5.
65. Hay ID. Thyroiditis: a clinical update. *Mayo Clin Proc*, 1985; **60**: 836–43.
66. Schwaegerle SM, Bauer TW, Esselstyn C, Jr. Riedel's thyroiditis. [Review]. *Am J Clin Pathol*, 1988; **90**: 715–22.
67. Heufelder AE HI, Carney JA, Gorman CA. Coexistence of Graves' disease and Riedel's (invasive fibrous) thyroiditis: further evidence of a link between Riedel's thyroiditis and organ-specific autoimmunity. *Clin Invest*, 1994; **72**: 788–93.

68. Zimmermann-Belsing T, Feldt-Rasmussen U. Riedel's thyroiditis: an autoimmune or primary fibrotic disease? *J Intern Med*, 1994; **235**: 271–4.
69. Heufelder AE, Hay ID. Further evidence for autoimmune mechanisms in the pathogenesis of Riedel's invasive fibrous thyroiditis. *J Intern Med*, 1995; **238**: 85–6.
70. Heufelder AE, Bahn RS. Soluble intercellular adhesion molecule-1 (sICAM-1) in sera of patients with Graves' ophthalmopathy and thyroid diseases. *Clin Exp Immunol*, 1993; **92**: 296–302.
71. Heufelder AE, Goellner JR, Bahn RS, Gleich GJ, Hay ID. Tissue eosinophilia and eosinophil degranulation in Riedel's invasive fibrous thyroiditis. *J Clin Endocrinol Metab*, 1996; **81**: 977–84.
72. Fatourechi MM, Hay ID, McIver B, Sebo TJ, Fatourechi V. Invasive fibrous thyroiditis (Riedel thyroiditis): the Mayo Clinic experience, 1976–2008. *Thyroid*, 2011; **21**: 765–72.
73. Falhammar H, Juhlin CC, Barner C, Catrina SB, Karefylakis C, Calissendorff J. Riedel's thyroiditis: clinical presentation, treatment and outcomes. *Endocrine*, 2018; **60**: 185–92.
74. Papi G, Corrado S, Cesinaro AM, Novelli L, Smerieri A, Carapezzi C. Riedel's thyroiditis: clinical, pathological and imaging features. *Int J Clin Pract*, 2002; **56**: 65–7.
75. Cárdenas-Perilla R, Monturiol-Duran J, Simó-Perdigó M, Barios-Profitós M, Castell-Conesa J. ^{18}F-FDG PET-CT usefulness in extra-pancreatic involvement in IgG4 related diseases. *Rev Esp Med Nucl Imagen Mol*, 2014; **33**: 366–9.
76. Few J, Thompson NW, Angelos P, Simeone D, Giordano T, Reeve T. Riedel's thyroiditis: treatment with tamoxifen. *Surgery*, 1996; **120**: 993–9.
77. Carruthers MN, Topazian MD, Khosroshahi A, *et al*. Rituximab for IgG4-related disease: a prospective, open-label trial. *Ann Rheum Dis*, 2015; **74**: 1171–7.

3.3

Thyrotoxicosis and Related Disorders

3.3.1 Clinical Assessment and Systemic Manifestations of Thyrotoxicosis

Claudio Marcocci and Filomena Cetani

Introduction

The term thyrotoxicosis refers to the clinical syndrome due to excessive serum concentrations of free thyroxine, free triiodothyronine, or both. The term hyperthyroidism is used to mean sustained increases in hormone biosynthesis and secretion by the thyroid gland, being Graves' disease the most common. Other causes responsible for thyrotoxicosis are destructive thyroiditis, excessive ingestion of thyroid hormones, or secretion of thyroid hormones from ectopic sites. The various causes of thyrotoxicosis are discussed in Chapter 3.3.5. The clinical features depend on the severity and duration of the disease, the age of the patient, the presence or absence of extrathyroidal manifestations, and the specific thyrotoxic disorder. Older patients have fewer symptoms and signs of sympathetic activation and more symptoms and signs of cardiovascular dysfunction. Rarely a patient with 'apathetic' hyperthyroidism will lack almost all usual clinical manifestations of thyrotoxicosis [1].

All organ systems are affected by thyroid hormone excess (Table 3.3.1.1). In Graves' disease some signs and symptoms are due to extrathyroidal immunological processes rather than the excessive levels of thyroid hormones (Table 3.3.1.2).

Table 3.3.1.1 Systemic effects of thyrotoxicosis

System	Effects
General	Heat intolerance, weight loss, fatigue, insomnia, nervousness, tremulousness
Skin	Fine, warm, and moist, hyperpigmentation, hyperhidrosis, onycholysis, fine and often straight hair, urticaria, pruritus
Eye	Exophthalmos, lid oedema, lid lag, globe lag, chemosis, ophthalmoplegia, optic nerve involvement
Mental	Irritability, restlessness, anxiety, inability to concentrate, lability, depression, psychiatric reactions
Neurological	Syncope, delirium, stupor, coma, choreoathetosis
Cardiovascular	Tachycardia, widened pulse pressure, and bounding pulse. Occasionally cardiomegaly, congestive heart failure, angina pectoris, and paroxysmal tachycardia or atrial fibrillation
Respiratory	Dyspnoea
Gastrointestinal	Hyperphagia, increased thirst, diarrhoea, or increased frequency of stools, hepatomegaly
Neuromuscular	Tremulousness, quickened, and hypermetric reflexes, weakness of proximal muscles, muscle atrophy, myopathy, periodic paralysis
Metabolic	Hypercalcaemia, decreased serum magnesium, increased bone alkaline phosphatase, hypercalciuria
Osseous	Osteopaenia or osteoporosis
Reproductive	Irregular menses or amenorrhoea, gynaecomastia, decreased fertility
Haematopoietic	Anaemia (usually normochromic, normocytic), lymphocytosis, splenomegaly, lymphadenopathy, enlarged thymus

Table 3.3.1.2 Clinical findings in patients with Graves' hyperthyroidism and controls[a]

	Hyperthyroid				Controls			
	Total	Age decades			Total	Age decades		
		2nd	3rd to 5th	6th to 8th		2nd	3rd to 5th	6th to 8th
Number	880	74	635	171	880	79	636	165
Symptoms (%)								
Palpitations	65	58	57	56	13	6	14	10
Increased perspiration	45	39	49	30	7	1	9	3
Heat intolerance	55	49	60	36	8	6	8	8
Weight loss	61	29	60	74	13	6	13	13
Weight gain	12	29	12	5	21	26	21	16
Increased appetite	42	61	12	5	5	9	21	16
Decreased appetite	11	5	10	16	6	6	7	4
Increased number of bowel movements	22	19	22	21	2	6	2	1
Increased appetite with weight loss	24	19	24	20	0	0	0	0
Tiredness	69	62	70	69	41	32	43	37
Irritability	45	47	35	33	18	16	21	10
Nervousness	69	59	71	64	15	11	17	12
Signs (%)								
Fine finger tremor	69	69	70	59	6	5	5	4
Pulse rate 90 beats/min	80	84	80	78	18	21	18	19
Atrial fibrillation[b]	3	0	1	9	–	–	–	–
Thyroid size (× normal)	1.9 ± 0.6	2.4 ± 0.6	2.0 ± 0.6	1.4 ± 0.4	1.3 ± 0.4	1.4 ± 0.6	1.3 ± 0.4	1.3 ± 0.4

[a] Source data from Nordyke RA, Gilbert FI Jr, Harada AS. Graves' disease. Influence of age on clinical findings. *Arch Intern Med*. 1988 Mar;148(3):626–31.
[b] The presence of atrial fibrillation was not assessed in control subjects.

Skin, Hair, and Nails

The skin is usually moist and warm because of vasodilatation, which represents a homeostatic mechanism for dissipating the increased heat generated in the body [2]. The patient may complain of cutaneous flushing, resting perspiration, and sweaty palms. Hyperpigmentation particularly at the knuckles and skin creases may be present. Vitiligo occurs in some patients with Graves' disease and Hashimoto's thyroiditis [3]. Dermographism, urticaria, purpura, and generalized erythematous eruptions may be less frequently observed. Pruritus is rare. The skin changes are reversed after restoration of euthyroidism.

The hair may be fine and soft, and hair loss can be excessive. Alopecia aerate and loss of axillary, pubic, body, and eyebrow hairs are uncommon.

Localizing non-pitting oedema on extensor surfaces, the so-called pretibial myxoedema, may be the hallmark to establish the diagnosis of Graves' disease (Figure 3.3.1.1).

The nails become shiny and may be soft and friable. The rate of nail growth is increased, and longitudinal striations associated with a flattening of the surface contour result in a scoop-shovel appearance. In many patients onycholysis may be present.

Eyes

Upper eyelid retraction, evident as the presence of a rim of sclera between the lid and the limbus, is frequent in all forms of thyrotoxicosis, and is responsible for the bright-eyed 'stare' or 'fish eyes' of the patient (Figure 3.3.1.2) [4]. Lid lag may be present because the upper lid lags behind the globe when the patient is asked to gaze downward; globe lag occurs when the globe lags behind the upper lid when the patient gazes slowly upward. In severe cases the movements of the lids are jerky and spasmodic, and a fine tremor of the lightly closed lids can be observed. It is important to differentiate these manifestations from those of Graves' disease [5] (see Chapters 3.3.9 and 3.3.10).

Thyroid Gland

Thyroid enlargement is usually associated with nodular goitre or Graves' disease (Figure 3.3.1.3a) and may be absent when thyrotoxicosis is due to excessive ingestion of thyroid hormones, unless goitre was present before treatment. An asymmetrical gland is generally found in patients with toxic adenoma or multinodular goitre (Figure 3.3.1.3b), but may occur also in Graves' disease. The pyramidal lobe should always be searched for, since enlargement indicates a diffuse involvement of the thyroid. The marked increase in the blood flow to the thyroid in Graves' disease is reflected clinically by the presence of a bruit or a thrill. Colour flow Doppler sonography shows hypervascularity and increased peak systolic velocity [6].

Respiratory System

Respiratory features are reported in Box 3.3.1.1. The frequency and the relative relevance of these changes is uncertain because available

Box 3.3.1.1 Respiratory changes in thyrotoxicosis

- Dyspnoea
- Respiratory muscle weakness
- Decreased vital capacity
- Decreased pulmonary compliance
- Increased ventilation
- Increased oxygen uptake
- Pulmonary artery dilatation and hypertension
- Increased carbon dioxide production
- Increased ventilatory response to hypercapnia

data are scarce and often conflicting. The increased metabolic rate stresses the lung, requiring a more rapid net rate of gas exchange to accommodate the increased oxygen consumption and carbon dioxide production [7]. Dyspnoea is present in patients with severe thyrotoxicosis [8] and is due to respiratory muscle weakness, reduction of vital capacity, decreased pulmonary compliance, and increase in respiratory dead space ventilation.

An atopic background, with elevated serum immunoglobulin E levels has been found in 30% of patients with Graves' hyperthyroidism [9].

A decrease of residual volume and vital and total lung capacities was early reported in a quarter of patients [8]. More recently, no significant differences in the mean baseline vital and total lung capacities, residual volume, static compliance, or pressure–volume curves between patients and controls have been reported [9].

Lung compliance may be altered by changes in the elastic properties or by vascular engorgement. Manifestations of respiratory muscle dysfunction include rapid, shallow respirations, respiratory dyskinesia, hypoventilation, respiratory acidosis, and easy fatigability [10]. Most patients with overt thyrotoxicosis have diminished proximal muscle strength. Chronic thyrotoxic myopathy affects the diaphragm and other respiratory muscles in up to 50% of severely affected patients, causing loss of maximal respiratory muscle power.

Thyrotoxicosis may affect the central regulatory response to a blood gas perturbation. This abnormality can be revealed by evaluating the increase of ventilation while breathing either a hyperoxic hypercapnic or a hypoxic isocapnic gas mixture. The mechanisms of these changes are not completely understood. Thyrotoxicosis, by increasing the ventilatory drive superimposed on underlying lung disease, may worsen dyspnoea and cause respiratory failure.

Resting heart rate, cardiac output, respiratory rate, and minute ventilation are increased [9]. The amount of oxygen required to perform any workload is increased. Pulmonary artery pressures of patients may rise more than usual with exercise, but this has not been evaluated carefully. Exercise normally decreases the mixed venous oxygen saturation and the dead space/tidal volume ratio; the opposite occurs in thyrotoxicosis.

Cardiac changes may affect the lungs either by pulmonary artery dilatation or by high-output cardiac failure [11]. The findings of an accentuated pulmonary second heart sound and a right ventricular heave suggest pulmonary hypertension with an increased risk of heart failure in 6–16% of thyrotoxic patients [11]. The exact pathogenesis remains unknown. Some patients may have symptomatic pulmonary hypertension that normalizes by restoring euthyroidism [11].

Mild increases of resting pulmonary artery pressure are common, and the pressure frequently rises significantly during exercise. A physical sign of thyrotoxicosis is the Means–Lerman sign, a scratchy coarse systolic ejection rub or murmur that is heard best along the left sternal border at the base of the heart.

Renal System

Most of the renal effects of thyrotoxicosis produce no symptoms except mild polyuria [12].

Renal plasma flow and glomerular filtration rate (GFR) are increased, probably because of the increase in cardiac output and decrease in peripheral resistance [12]. Intrarenal vasodilatation also occurs. The 24-h urine creatinine is significantly lower in thyrotoxic patients compared to normal subjects, likely because of loss of muscle mass and it occurs despite an increase in urea clearance. These changes normalized when the euthyroidism is restored.

Renal tubular mass is increased, and the morphological changes are accompanied by an increased renal tubular transport capacity. An activation of the renin–angiotensin system contributes to cardiac hypertrophy in these patients [13]. Serum cystatin C (CysC) levels are higher despite an increase in GFR. Thyroid hormone acting on general metabolism may influence the production rate of CysC, therefore thyroid status must be considered in the measurement of renal function using CysC equation [14].

Thyrotoxic patients rarely have abnormalities in water metabolism. Serum electrolytes are usually normal. Some patients have polydipsia and polyuria (up to 3–4 litres daily) that revert to normal after achieving euthyroidism.

Plasma atrial natriuretic hormone levels and renin activity are increased with no clinical consequences except for mild oedema. The total amount of exchangeable potassium is decreased, and the amount of exchangeable sodium tends to be increased. The level of exchangeable magnesium concentration is often decreased, and urinary magnesium is increased.

Renal tubular acidosis occasionally occurs and there is a failure to achieve maximal urinary acidification. It may occur in Graves' disease, also in the absence of nephrocalcinosis, and may persist after restoration of the euthyroid state. This condition may have an autoimmune basis [15].

Pitting oedema, involving legs, hands, ankles, and sacrum, may occur in thyrotoxic patients, because of renal salt and water retention in response to the reduction in effective arterial volume. Renal and salt retention contributes to an increase in blood volume and venous pressure.

Gastrointestinal System

The classic manifestations are rapid intestinal transit, increased frequency of semiformed stools, and weight loss from increased caloric requirement or malabsorption [16]. The mechanism underlying the gastrointestinal hypermotility has not been elucidated, but it disappears when euthyroidism is restored. An increase in appetite is common, but rarely seen in patients with mild disease. In severe disease, the increased food intake is usually inadequate to meet the increased caloric requirements, and variable weight loss occurs. Anorexia may occur in patients with severe thyrotoxicosis,

and in about one-third of elderly patients, where it contributes to the 'apathetic' picture.

Graves' disease may be associated with other autoimmune diseases, namely coeliac disease, atrophic gastritis, with or without pernicious anaemia.

The combination of biochemical thyrotoxicosis and persistent vomiting is also common in the context of hyperemesis gravidarum, often occurring in the first trimester of pregnancy. This self-limiting disorder, termed transient gestational hyperthyroidism, results from stimulation of thyroid-stimulating hormone (TSH) receptors by high levels of human chorionic gonadotrophin (hCG) produced by the placenta. This effect is due to the α-subunit homology between hCG and TSH [16].

Hepatic function may be altered, particularly when the disease is severe [16]; hypoproteinaemia and increased serum alkaline phosphatase and transaminase levels may be present.

Hepatic damage is usually self-limiting, but fulminant hepatitis has previously been described, precipitated by congestive cardiac failure and arrhythmia [16]. Histological hepatic changes are usually non-specific, but progressive liver damage with centrizonal necrosis and perivenular fibrosis and intrahepatic cholestasis have also been described.

In severe cases hepatomegaly and jaundice may occur. Graves' disease and autoimmune hepatitis coexist more often than can be expected by chance.

Liver disease is reversible in the vast majority of cases with restoration of euthyroidism [16].

Nervous System

Hyperactivity, emotional lability, distractibility, and anxiety may reflect changes in the nervous system, but the pathogenic mechanisms remain obscure [17]. Examination reveals a fine rhythmic tremor of the hands, tongue, and eyelids when slightly closed. Emotional lability causes patients to lose their tempers easily and to have episodes of crying without apparent reason. Rarely mental disturbance may be severe. Fatigue is due either to muscle weakness or to insomnia, which is frequently present.

Persistent fine tremor is the most prominent finding and may involves the hands, but also the feet, chin, lips, and tongue. Chorea seldom appears as a manifestation of thyrotoxicosis. Thyrotoxic crisis may rarely lead to coma and status epilepticus.

The electroencephalogram of most patients reveals increased fast-wave activity. The basal metabolic rate tends to correlate with the frequency of brain waves, but at the extremes of thyroid abnormality the correlation is frequently poor.

Muscle

Muscle weakness and fatigue are frequent [18]. Weakness is often most prominent in the proximal muscles of the limbs [18]. Occasionally, in severe untreated cases, muscle wasting occurs as a predominant feature (thyrotoxic myopathy). In extreme forms, the patient may be unable to rise from a sitting or lying position and may be unable to walk.

Muscle manifestations affect men more commonly than women. The involvement of ocular muscles may mimic myasthenia gravis. Graves' disease occurs in about 3–5% of patients with myasthenia gravis, and about 1% of patients with Graves' disease develop myasthenia gravis [19]. Hypokalaemic periodic paralysis may be associated with thyrotoxicosis and is characterized by sporadic attacks, most commonly involving flaccidity and paralysis of either legs, arms, or trunk, even though any muscle can be involved [19]. Hypokalaemic periodic paralysis is most frequent in Asian populations (see Chapter 3.3.2).

Whether patients with thyrotoxic periodic paralysis have an underlying genetic predisposition, compared with patients with non-thyrotoxic periodic paralysis is unknown. Familial hypokalaemic periodic paralysis has been associated with mutations of skeletal muscle calcium channel (CACNA1S), sodium channel (SCN4A), and voltage-gated potassium channel (KCNE3) [20]. However, such associations have not been found in patients with thyrotoxic periodic paralysis, except for one patient harbouring a KCNE3 mutation [21].

Skeletal System: Calcium and Phosphorus Metabolism

Thyrotoxicosis is associated with an increase of bone turnover and eventually bone loss, especially in postmenopausal women [22]. Patients with a long-standing history of thyrotoxicosis may have overt osteoporosis and an increased risk of fractures [23].

Bone turnover is increased, but the increase in bone resorption is relatively greater than that of bone formation; the urinary excretion of calcium, phosphorus, and hydroxyproline is increased [22]. Total and particularly ionized serum calcium may be slightly increased in as many as 27% and 47% of patients. Alkaline phosphatase and osteocalcin levels are also frequently increased [24]. Parathyroid hormone (PTH) and 1,25 dihydroxyvitamin D levels tend to be low as a result of the increased calcium released from bone. The alterations in bone metabolism are reversed when euthyroidism is restored [24, 25].

Thyrotoxicosis is a well-known risk factors for osteoporosis [22]. Thyroid hormone action on osteoblasts accounts for the increased circulating levels of alkaline phosphatase and osteocalcin. Despite the increased mineralization rate and osteoblastic activity, the increased bone formation cannot compensate for increments in bone resorption, and bone mass may be decreased. The pathological changes may include osteoporosis, osteomalacia, and osteitis fibrosa. A recent meta-analysis has shown that subclinical hyperthyroidism is also associated with femoral neck bone loss and potentially increased fracture risk [26]. Restoration of euthyroidism may reverse bone density in premenopausal [27] but not in postmenopausal women.

The skeletal effects of thyroid hormone replacement are unclear. Recently, some reports suggested that patients receiving chronic L-thyroxine treatment, particularly in doses that suppress TSH secretion may have a reduced bone mass [28]. Other studies [29, 30] suggested that L-thyroxine suppressive therapy, if carried out carefully and monitored, using the smallest suppressive dose, has no significant effect on bone mass, at least in premenopausal women

and in men, whereas in postmenopausal women some degree of bone loss can be observed. A recent cross-sectional study evaluated the prevalence and determinants of radiological vertebral fractures in women receiving L-thyroxine therapy for differentiated thyroid cancer. Vertebral fractures were found in 51 patients (28.5%), with significantly higher prevalence in patients with undetectable TSH ((44.6%) as compared with patients with TSH between 0.5 and 1.0 mU/L (24.0%) and >1 mU/L (4.3%) [31].

Thyroid acropachy, namely clubbing, periostitis, and swelling at peripheral sites occurs in approximately 1% of patients with Graves' disease, and is always associated with exophthalmos and pretibial myxoedema [32].

Haematopoietic System

In most patients the red blood cell mass is increased either for a direct effect of thyroid hormones on the erythroid marrow or an increased production of erythropoietin [33]. A parallel increase in plasma volume also occurs, and therefore the haematocrit value is normal.

Microcytosis may occur in about 37% of patients and usually resolves with the restoration of euthyroidism. Patients with severe thyrotoxicosis may develop a normocytic anaemia and less frequently iron deficiency anaemia.

Approximately 3% of patients with Graves' disease have pernicious anaemia, and a further 3% have antibodies to intrinsic factor but normal absorption of vitamin B_{12}. Autoantibodies against gastric parietal cells are present in about one-third of patients with Graves' disease, and the requirements for vitamin B_{12} and folic acid appear to be increased.

The total white blood cell count is often low because of a decrease in the number of neutrophils. The absolute lymphocyte count is normal or increased, leading to a relative lymphocytosis. A generalized lymphadenopathy may be present, and the spleen may be enlarged in 10% of patients with Graves' disease.

Blood platelets and the intrinsic clotting mechanism are normal. However, the concentration of factor VIII is often increased and returns to normal when thyrotoxicosis is treated. Furthermore, there is an enhanced sensitivity to coumarin anticoagulants because of an accelerated clearance of vitamin K-dependent clotting factors [33].

Cardiovascular System

The cardiovascular manifestations are among the most characteristic symptoms and signs of the disorder (Box 3.3.1.2) [34]. Tissue blood flow is increased in response to accelerated metabolism and increased oxygen consumption. Haemodynamic changes are characterized by an elevated cardiac output and a decreased peripheral vascular resistance. Thyroid hormone itself may be involved directly through its action on the smooth muscle of blood vessels [34]. Moreover, the finding in thyrotoxic patients of elevated levels of plasma adrenomedullin and proadrenomedullin-N-terminal 20-peptide, which have a potent vasodilatory activity, raises the possibility that both might also decrease the vascular resistance [35].

Box 3.3.1.2 Cardiovascular symptoms and signs of thyrotoxicosis

- Palpitations
- Paroxysmal tachycardia
- Orthopnoea
- Exercise intolerance
- Hyperdynamic precordium
- Third heart sound
- Atrial fibrillation
- Widened pulse pressure
- Cardiac flow murmurs

Nearly all patients have tachycardia and a bounding pulse. The common complaint of palpitations usually indicates a resting tachycardia. The heart rate is also elevated during sleep. The heart may be enlarged, but echocardiography is usually normal. In elderly patients the cardiovascular manifestations may be limited to resting tachycardia [36]; other classic thyrotoxic symptoms may be absent.

Thyrotoxic patients may have chest pain similar to that angina pectoris, probably caused by either relative myocardial ischaemia or coronary artery spasm. In elderly patients, however, the increased myocardial oxygen demand due to thyrotoxicosis may unmask coronary artery disease. The plasma level of homocysteine in thyrotoxic patients did not differ significantly from that of controls [34]. Conversely, hyperhomocystinaemia has been found in hypothyroid patients and, combined with lipid abnormalities, may increase the risk of coronary disease [34].

At physical examination tachycardia is the most common findings. Systolic blood pressure is elevated and diastolic blood pressure is decreased [37]; the mean blood pressure is usually normal. An exaggerated increase in systolic blood pressure may be present in older patients due to the loss of elasticity of the larger arteries. Auscultation may reveal a systolic ejection murmur and a gallop rhythm caused by rapid flow of blood through the aortic outflow tract. Systolic murmurs may arise from valve prolapse, left ventricular dilatation, or dysfunction of the mitral valve. A systolic 'scratch' is heard in the pulmonary area corresponding to contact between the pleural and pericardial surfaces during cardiac contraction. Heart failure rarely occurs, unless an underlying cardiac disease is also present [34].

Sinus tachycardia is present in the majority of patients. Cardiac arrhythmias are almost invariably supraventricular. Approximately 10% of patients with thyrotoxicosis have atrial fibrillation, and a similar percentage of patients with unexplained atrial fibrillation are thyrotoxic [34]. Tachycardia may be the presenting symptom of thyrotoxicosis, particularly in older people. In elderly patients with subclinical thyrotoxicosis the risk of persistent atrial fibrillation is approximately three times that of normal subjects [34]. Ventricular premature contractions are rare. Angina pectoris and myocardial infarction may rarely occur in the absence of coronary artery disease.

Thyrotoxicosis itself may cause heart failure in elderly and, much less often, in young patients. Thyrotoxic patients with heart failure are generally old and, therefore, at risk of underlying heart disease. Elderly patients with rhythm disturbances have the greatest risk of heart failure [38]. In young patients, or in the absence of underlying heart disease, the heart failure is thought to be 'high output'

likely due to a circulatory congestion caused by fluid retention. In thyrotoxicosis, cardiac output is near to maximal at rest and cannot increase in response to exercise, stress, surgery, or pregnancy [38]. As a consequence, atrial filling pressures rise, leading to pulmonary and peripheral oedema. This situation may worsen if atrial fibrillation is present. Left ventricular function is impaired because by the persistent tachyarrhythmia. Sustained tachycardia causes abnormal ventricular systolic and diastolic function, which resolves when arrhythmia is treated. β-adrenergic receptor blockade-mediated slowing of the heart rate can rapidly reverse even severe degrees of left ventricular dysfunction in thyrotoxic patients.

Endocrine System

Thyrotoxicosis affects the secretion of most pituitary hormones. Children with thyrotoxicosis grow more rapidly than normal children [39]. The height and bone ages are accelerated, but their relationship remains normal. Growth acceleration suggests that growth hormone (GH) secretion might be greater than normal. Serum GH concentrations, however, are lower in thyrotoxic patients than normal subjects probably due to the increased metabolic clearance rate. Serum IgF1 concentration is higher in thyrotoxic patients and returns to normal after restoration of euthyroidism. Basal secretion of prolactin (PRL) and its response to thyrotropin-releasing hormone (TRH) may also be decreased. No physiological or clinical consequences of these abnormalities are known.

The half-life of cortisol is shortened, but both the number of bursts of adrenocorticotrophic hormone (ACTH) and those of cortisol secretion are increased and maintain serum cortisol levels [40]. A subtle impairment of adrenocortical reserve has been reported in thyrotoxicosis [40]. The plasma concentration of corticosteroid-binding globulin is normal. The urinary excretion of the free cortisol and 17-hydroxycorticosteroids is normal or slightly increased, whereas the urinary excretion of 17-ketosteroids may be reduced [40]. The turnover rate of aldosterone is increased, but its plasma concentration is normal. Plasma renin activity is increased, and sensitivity to angiotensin II is reduced.

β-adrenergic receptor blockade ameliorates most of the cardiovascular manifestations of thyrotoxicosis suggesting that catecholamines play a causative role, but the secretion rate and plasma levels of adrenaline and noradrenaline are normal [41]. The sympathetic hyperactivity appears to be the consequence of a direct effect of thyroid hormones on peripheral tissues. Thyrotoxicosis in early life may cause delayed sexual maturation, although physical development is normal and skeletal growth may be accelerated.

Thyrotoxicosis, after puberty, influences the reproductive function [42], especially in women. An increase in libido occurs in both genders. The intermenstrual interval may be prolonged or shortened, and menstrual flow initially diminishes and ultimately ceases. Fertility may be reduced. In some women, menstrual cycles are predominantly anovulatory with oligomenorrhoea, but in most, ovulation occurs. It is unclear whether these changes are due to a direct action of thyroid hormones on either the ovary and uterus or the pituitary and hypothalamus, or both. With treatment, menstrual cycles return to their regular pattern. Thyrotoxicosis in prepubertal girls may result in slightly delayed menarche. In premenopausal women, basal plasma concentrations of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are normal but may display an enhanced responsiveness to LH [42].

An increase in libido has also been reported in men [43]. An increase in sex hormone-binding globulin is present and because of its increase, testosterone levels are elevated, but free testosterone levels are normal. The metabolic clearance rates of testosterone and, to some extent, of oestradiol are decreased. Conversion rates of androstenedione to testosterone, oestrone, and oestradiol, and of testosterone to dihydrotestosterone are increased. Extragonadal conversion of androgens to oestrogens is increased and this may account for finding gynaecomastia in a minority of thyrotoxic men. A reduction of sperm motility (asthenozoospermia) has been reported in about 60% of adult thyrotoxic patients; it normalizes after treatment with antithyroid drugs [44]. Other seminal fluid alterations, namely low sperm count and an increased number of spermatozoa with altered morphology, has been found in about 42% and 40%, respectively. One limit of these studies is that semen parameters have been evaluated in infertile couples in whom fertility of the partner has not always been studied. Premature ejaculation, erectile dysfunction, and hypoactive sexual desire have also been reported.

Energy Metabolism: Protein, Carbohydrate, and Lipid Metabolism

One of the most prominent symptoms is heat intolerance reflecting an increase in the basal metabolism of many substrates [45] leading to an increased consumption of adenosine triphosphate (ATP) and oxygen. The consequent thermogenesis is responsible for heat intolerance. Despite the increased food intake, a state of chronic caloric inadequacy often ensues, depending on the degree of increased metabolism, and becomes more pronounced with age. In addition to losing fat stores, there is often a loss of muscle mass, making weakness a common complaint. Both synthesis and degradation of proteins are increased.

Both glucose absorption and glucose production are increased. The oral glucose tolerance test is often abnormal [46]. Pre-existing diabetes mellitus is aggravated by thyrotoxicosis, one cause being increased degradation of insulin.

Both synthesis and clearance of cholesterol and triglycerides are increased, but the latter effect predominates [47]. Plasma phospholipid and low-density lipoprotein (LDL) cholesterol concentrations fall, while high-density lipoprotein (HDL) cholesterol levels increase. Thyroid hormones also may influence cholesterol metabolism by increasing its conversion to bile acid and its clearance through the membrane surface LDL receptors. In this regard, *in vitro* studies indicate that triiodothyronine increases LDL receptor promoter activity and surface LDL receptor protein [48].

Although fatty acid synthesis is increased in both adipose tissue and liver, degradation of most lipids appears to be stimulated more than synthesis; body lipid deposits consequently become depleted and plasma concentrations of various lipid components fall. Several studies have investigated the relationship between leptin level and thyroid status. Two studies reported a relative hypoleptinaemia, but most have found no effect of thyrotoxicosis on leptin levels [49].

Vitamin Metabolism in Thyrotoxicosis

Vitamin A concentrations tend to be low and a minor impairment of dark adaptation has been detected in some patients. Serum PTH levels are low and the conversion of 25-hydroxyvitamin D to 1,25-hydroxyvitamin D is diminished, resulting in lowered serum concentrations of the latter [50]. Calcium balance is negative as a result of decreased intestinal absorption and increased urinary loss. The serum concentration of vitamin E tends to be reduced because serum concentrations of HDL and LDL, in which vitamin E is incorporated, are decreased.

Differential Diagnosis of Thyrotoxicosis

The condition that most frequently simulates thyrotoxicosis is an anxiety state characterized by nervous irritability, fatigue, and insomnia. Fatigue is pronounced and differs from that in thyrotoxicosis because it is not accompanied by a desire to be active. Tachycardia is common during examination but, in contrast to thyrotoxicosis, the sleeping pulse rate is normal. Hyperreflexia is present in both disorders.

Phaeochromocytoma may closely resemble thyrotoxicosis. Tachycardia and hypermetabolism are common to both conditions. The patient may have weight loss despite a good appetite and hyperglycaemia with glycosuria.

Myeloproliferative disorders may mimic thyrotoxicosis because of increased sweating, weight loss, and tachycardia, especially if anaemia is present. In diabetes mellitus, weight loss despite a good appetite, muscle wasting, and occasionally diarrhoea may suggest thyrotoxicosis.

REFERENCES

1. Chiovato L, Mariotti S, Pinchera A. Thyroid disease in the elderly. *Baillieres Clin Endocrinol Metab*, 1997; **11**: 251–70.
2. Rosen T, Kleman WR. Cutaneous manifestation of thyroid disease. *J Am Acad Dermatol*, 1992, **26**: 885–7.
3. Baldini E, Odorisio T, Sorrenti S, *et al.* Vitiligo and autoimmune thyroid disorders. *Front Endocrinol (Lausanne)*, 2017, **8**: 290–4.
4. Dickinson AJ, Perros P. Controversies in the clinical evaluation of active thyroid-associated orbitopathy: use of detailed protocol with comparative photographs for objective assessment. *Clin Endocrinol*, 2001; **55**: 283–303.
5. European Group on Graves' Orbitopathy (EUGOGO), Wiersinga WM, Perros P, *et al.* Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. *Eur J Endocrinol*, 2006; **55**: 387–9.
6. Vitti P, Rago T, Mazzeo S, *et al.* Thyroid blood flow evaluation by color-flow Doppler sonography distinguishes Graves' disease from Hashimoto's thyroiditis. *J Endocrinol Invest*, 1995; **18**: 857–61.
7. Kendric AH, O'Reilly JR, Laszlo G. Lung function and exercise performance in hyperthyroidism before and after treatment. *QJM*, 1988; **68**: 615–18.
8. Small D, Gibbons W, Levy RD, de Lucas P, Gregory W, Cosio MG. Exertional dyspnea and ventilation in hyperthyroidism. *Chest*, 1992; **101**: 1268–73.
9. Zimmerman L. Pulmonary complications of endocrine diseases. In: Broaddus VC, Mason RJ, Ernst JD, *et al.*, eds. *Murray and Nadel's Textbook of Respiratory Medicine*, 6th edn. Philadelphia, PA: Elsevier Saunders, 2016; Chapter 95: 1671–8.
10. Siafakas NM, Alexopoulou C, Bouros D. Respiratory muscle function in endocrine disease. *Monaldi Arch Chest Dis*, 1999; **54**: 154–9.
11. Armigliato M, Paolini R, Aggio S, *et al.* Hyperthyroidism as a cause of pulmonary arterial hypertension: a prospective study. *Angiology*, 2006; **57**: 600–6.
12. Iglesias P, María Auxiliadora Bajo MA, Rafael Selgas R, Juan José Díez JJ. Thyroid dysfunction and kidney disease: an update. *Rev Endocr Metab Disord*, 2017; **18** :131–44.
13. Basset A, Blanc J, Messas E, Hagège A, Elgozi JL. Renin–angiotensin system contribution to cardiac hypertrophy in experimental hyperthyroidism: an echocardiographic study. *J Cardiovasc Pharmacol*, 2001; **37**: 163–72.
14. Barreto-Chaves ML, Carrillo-Sepúlveda MA, Carneiro-Ramos MS, Gomes DA, Diniz GP. The crosstalk between thyroid hormones and the renin–angiotensin system. *Vascul Pharmacol*, 2010; **52**: 166–70.
15. Konishi K, Hayashi M, Saruta T. Renal tubular acidosis with autoantibody directed to renal collecting-duct cells. *N Engl J Med*, 1994; **331**: 1593–6.
16. Kyriacou A, McLaughlin J, Syed AA. Thyroid disorders and gastrointestinal and liver dysfunction: a state of the art review. *Eur J Int Med*, 2015, **26**: 563–71
17. Swee du S, Chng CL, Lim A. Clinical characteristics and outcome of thyroid storm: a case series and review of neuropsychiatric derangements in thyrotoxicosis. *Endocr Pract*, 2015; **21**: 182–9.
18. Cakir M, Samanchi N, Balci N, Balci MK. Musculoskeletal manifestations in patients with thyroid disease. *Clin Endocrinol (Oxf)*, 2003; **59**: 162–7.
19. Kung AW. Neuromuscular complications of thyrotoxicosis. *Clin Endocrinol (Oxf)*, 2007; **67**: 645–50.
20. Ng WY, Lui KF, Thai AC, *et al.* Absence of ion channels CACN1AS and SCN4A mutations in thyrotoxic hypokalemic periodic paralysis. *Thyroid*, 2004; **14**: 187–90.
21. Kung AWC, Lau KS, Fong GCY, *et al.* Association of novel single nucleotide polymorphisms in the calcium channel alpha 1 subunit gene (Ca(v)1.1) and thyrotoxic periodic paralysis. *J Clin Endocrinol Metab*, 2004, **89**:1340–45.
22. Mundy GR, Shapiro JL, Bandelin JG, Canalis EM, Raisz LG. Direct stimulation of bone resorption by thyroid hormones. *J Clin Invest*, 1976; **58**: 529–32.
23. Cummings SR, Nevitt MC, Browner WS, *et al.* Risk factors for hip fractures in white women. Study of Osteoporotic fractures research group. *N Engl J Med*, 1995; **332**: 767–73.
24. Garnero P, Vassy V, Bertholin A, Riou JP, Delmas PD. Markers of bone turnover in hyperthyroidism and the effects of treatment. *J Clin Endocrinol Metab*, 1994; **78**: 955–9.
25. Diamond T, Vine J, Smart R, Butler P. Thyrotoxic bone disease in women: a potentially reversible disorder. *Ann Intern Med*, 1994; **120**: 8–12.
26. Segna D, Bauer DC, Feller M, *et al.* Thyroid Studies Collaboration. Association between subclinical thyroid dysfunction and change in bone mineral density in prospective cohorts. *J Intern Med*, 2018, **283**: 56–72.
27. Rosen C, Adler RA. Longitudinal changes in lumbar bone density among thyrotoxic patients after attainment of euthyroidism. *J Clin Endocrinol Metab*, 1992; **75**: 1531–4.

28. Taeelman P, Kaufman JM, Janssens X, Vandecauter H, Vermeulen A. Reduced forearm bone mineral content and biochemical evidence of increased bone turnover in women with euthyroid goiter treated with thyroid hormone. *Clin Endocrinol*, 1990; **33**: 107–17.
29. Marcocci C, Golia F, Bruno-Bossio G, Vignali E, Pinchera A. Carefully monitored levothyroxine suppressive therapy is not associated with bone loss in premenopausal women. *J Clin Endocrinol Metab*, 1994; **78**: 818–23.
30. Marcocci C, Golia F, Vignali E, Pinchera A. Skeletal integrity in men chronically treated with suppressive doses of L-thyroxine. *J Bone Miner Res*, 1997; **12**: 72–7.
31. Mazziotti G, Formenti AM, Frara S, *et al*. High prevalence of radiological vertebral fractures in women on thyroid-stimulating hormone-suppressive therapy for thyroid carcinoma. *J Clin Endocrinol Metab*, 2018, **103**: 956–64.
32. Fatourechi V. Thyroid dermopathy and acropachy. *Best Pract Res Clin Endocrinol Metab*, 2012; **26**: 553–65.
33. Franchini M. Hemostatic changes in thyroid diseases: haemostasis and thrombosis. *Hematology*, 2006; **11**: 203–8.
34. Dahl P, Dansi S, Klein I. Thyrotoxic cardiac disease. *Curr Heart Fail Rep*, 2008; **5**: 170–6.
35. Tuniyama M, Kitamura K, Ban Y, Sugita E, Ito K, Katagiri T. Elevation of circulating proadrenomedullin-N terminal 20-peptide in thyrotoxicosis. *Clin Endocrinol*, 1997; **46**: 271–4.
36. Kahaly GJ, Nieswandt J, Mohr-Kahays S. Cardiac risk of hyperthyroidism in the elderly. *Thyroid*, 1998; **8**: 1165–9.
37. Dansi S, Klein I. Thyroid hormone and blood pressure regulation. *Curr Hypertens Rep*, 2003; **5**: 513–20.
38. Biondi B, Palmieri EA, Lombardi G, Fazio S. Effects of thyroid hormone on cardiac function: the relative importance of heart rate, loading conditions, and myocardial contractility in the regulation of cardiac performance in human hyperthyroidism. *J Clin Endocrinol Metabol*, 2002; **87**: 986–74.
39. Wong GW, Lai J, Cheng PS. Growth in childhood thyrotoxicosis. *Eur J Pediatr*, 1999; **158**: 776–9.
40. Tsotsoulis A, Johnson EO, Kalogera CH, Seferiadis K, Tsolas O. The effect of thyrotoxicosis on adrenocortical reserve. *Eur J Endocrinol*, 2000; **142**: 231–5.
41. Coulombe P, Dussault JH, Walker P. Catecholamine metabolism in thyroid disease. II. Norepinephrine secretion rate in hyperthyroidism and hypothyroidism. *J Clin Endocrinol Metab*, 1977; **44**: 1185–9.
42. Ridgway EC, Maloof F, Longcope C. Androgen and oestrogen dynamics in hyperthyroidism. *J Clin Endocrinol Metab*, 1990; **20**: 250–4.
43. Krassas GR, Pontikides N, Deligianni V, Miras K. A prospective controlled study of the impact of hyperthyroidism on reproductive function in males. *J Clin Endocrinol Metab*, 2002; **87**: 3667–71.
44. La Vignera S, Vita R, Condorelli RA, *et al*. Impact of thyroid disease on testicular function. *Endocrine,* 2017, **58**:397–407.
45. Silva JE. The thermogenic effect of thyroid hormone and its clinical implications. *Ann Intern Med*, 2003; **139**: 205–13.
46. Woeber KA, Arky R, Braverman LE. Reversal by guanethidine of abnormal oral glucose tolerance in thyrotoxicosis. *J Clin Endocrinol Metab*, 1998; **80**: 102–5.
47. Bech K, Damsbo P, Eldrup E, *et al*. Beta-cell function and glucose and lipid oxidation in Graves' disease. *Clin Endocrinol*, 1996; **44**: 59–66.
48. Bakker O, Hudig F, Meijessen S, Wiersinga WM. Effects of triiodothyronine and amiodarone on the promoter of the human LDL receptor gene. *Biochem Biophys Res Commun*, 1998; **249**: 517–21.
49. Korbonits M. Leptin and the thyroid: a puzzle with missing pieces. *Clin Endocrinol*, 1998; **49**: 569–72.
50. Bouillon R, Muls E, DeMoor P. Influence of thyroid function on the serum concentrations of 1,25-dihydroxyvitamin D. *J Clin Endocrinol Metab*, 1980; **51**: 793–5.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

3.3.2 Thyrotoxic Periodic Paralysis

Annie W.C. Kung and C.L. Cheung

Epidemiology

The association of thyrotoxicosis and periodic paralysis was first described in 1902 in a Caucasian patient. However, it soon became evident that thyrotoxic periodic paralysis (TPP) affects mainly Asian populations, in particular the Chinese and Japanese, although isolated cases have also been reported in other ethnic groups such as Caucasians, Hispanics, African Americans, and American Indians. The incidence of TPP in non-Asian thyrotoxic patients is around 0.1%, whereas in the Chinese and Japanese, TPP affects 1.8% and 1.9% respectively, of the thyrotoxic patients [1–3]. Despite a higher incidence of thyrotoxicosis in women, TPP affects mainly men, with a male to female ratio ranging from 17: 1 to 70: 1, according to different series. In the Chinese population, TPP affects 13% of male and 0.17% of female thyrotoxic patients. In the Japanese population, TPP was reported to occur in 8.2% of male and 0.4% of female thyrotoxic patients in the 1970s, but in 1991, the reported incidence decreased to 4.3% and 0.04% respectively [4].

Clinical Features

TPP patients are usually between 20 and 40 years of age, similar to the age distribution for thyrotoxicosis. Isolated cases were reported in paediatric age group [5]. The paralytic attacks are characterized by transient, recurrent episodes of muscle weakness. Attacks involve proximal more than the distal muscles, with an initial involvement of the lower limbs and subsequently the truncal muscles, and finally all four limbs. The degree of weakness varies from mild weakness to total flaccid paralysis and hyporeflexia. Some patients may experience prodromal symptoms of aches, cramps, or stiffness

in the affected muscles. Weakness usually affects skeletal muscles only. However, total paralysis of respiratory, bulbar, and ocular muscles has been reported in severe cases [6–8]. Recovery is usually complete, but the duration of paralysis can vary from a few hours in a mild attack to 36–72 h in a severe attack. Electromyographic studies have confirmed the myopathic changes with intact peripheral nerve function. The presentation of TPP may be confused with Guillain-Barre syndrome, acute spinal cord compression, myelitis, and hysteria. The attacks of weakness are similar to those of familial hypokalaemic periodic paralysis (FHPP) except for the presence of hyperthyroidism. While FHPP is an autosomal dominant condition affecting mainly Caucasians, TPP is a sporadic disease found mainly in Asian males, and familial cases of TPP are extremely rare.

High carbohydrate loads and strenuous exercise are well recognized precipitating factors for TPP [9]. The paralytic attacks do not occur during exercise but occur during the resting period that follow strenuous exercise, and the attacks may be aborted by continuation of exercise. In subtropical cities such as Hong Kong, attacks are most common during the seasons of summer and fall. This seasonal variation is probably associated with an increased intake of sugary drinks as well as outdoor activities and exercise in these weather conditions. In tropical cities such as Singapore, seasonal variation is not seen. Attacks usually occur in the middle of the night or early morning, which coincides with a period of rest following a heavy meal or exercise. Paralysis can be induced in these patients with high carbohydrate loads with or without insulin infusion, strenuous exercise, or even thyroxine therapy. However, attacks cannot be induced once the patient has become euthyroid. Although high carbohydrate loads and strenuous exercise are well recognized precipitating factors for TPP, they were implicated in only 16 and 10 out of 135 cases in a prospective observational study, respectively [10]. Other precipitation factors include trauma, stress, alcohol consumption (especially binge drinking), acute urinary tract infection, and drugs (such as steroids and non-steroidal anti-inflammatory drugs) [10].

Hypokalaemia is the hallmark of TPP. Plasma potassium concentrations have been reported to be as low as 1.1 mmol/L. Some patients may have a near to normal plasma potassium concentration if they are admitted during the recovery phase of the attack. Mortality due to cardiac arrhythmia associated with the hypokalaemia has been reported. The complication of rhabdomyolysis may occur in a severe attack. Potassium concentration returns to normal when the patient recovers spontaneously from the weakness. The degree of hypokalaemia and the severity of weakness have no correlation with the severity of hyperthyroidism and the serum thyroid hormone concentration. Indeed, many patients have relatively few symptoms of hyperthyroidism and TPP may be their only manifestation of thyrotoxicosis. Apart from hypokalaemia, patients may also experience mild to moderate hypophosphataemia and hypomagnesaemia. These are also a result of intracellular shift as these electrolyte abnormalities would return to normal spontaneously when the patient recovers from the paralysis.

The underlying cause of hyperthyroidism in the majority of TPP patients is Graves' disease. However, TPP can also be associated with thyroiditis (either spontaneous or induced by interferon therapy), toxic nodular goitre, toxic adenoma, TSH-secreting pituitary tumour, and even overdosage of thyroid hormone. TPP is usually the early presentation of the underlying thyroid disease. In the case of Graves' disease, TPP can also be a presenting feature of relapse of the disease. Paralysis only occurs when the patient is thyrotoxic but not euthyroid.

Muscle biopsies from patients with TPP have revealed a variety of abnormalities. The most consistent finding is proliferation and focal dilation of the sarcoplasmic reticulum and transverse tubular system, with prominent vacuoles arising from the sarcoplasmic reticulum [11]. It is uncertain whether these vacuoles represent coalescence of dilated sarcoplasmic reticulum or sequestrated areas of focal myofibrial necrosis.

Genetics

A number of genetic association studies on TPP had been reported. Associations with the human leukocyte antigen (HLA) genotypes HLA B46, DR9, and DQBl*0303 were reported in Hong Kong Chinese, A2 Bw22, AW19, and B17 in Singapore Chinese, and DRW8 in Japanese [12]. However, it is uncertain whether these associations were related to the genetic predisposition to Graves' disease rather than to TPP, especially when the majority of these TPP patients had an underlying autoimmune thyroid disease.

In view of the similar presentations between TPP and FHPP, the role of the voltage-dependent calcium channel or dihydropyridine-sensitive L-type calcium channel receptor ($Ca_v1.1$), which is associated with FHPP-1, was studied in TPP patients. None of the few mutation hot spots associated with FHPP was present in Asian or non-Asian patients with TPP [13, 14]. However, certain single nucleotide polymorphisms (SNPs) of $Ca_v1.1$, including nucleotide (nt)—476, intron 2 nt 57, and intron 26 nt 67 were associated with TPP in southern Chinese [13]. The location of these SNPs lies at or close to the TRE of the gene, and it is likely that they affect the binding affinity of TRE and modulate the stimulation of thyroid hormone on the $Ca_v1.1$ gene. Similarly, isolated case reports with mutations in other skeletal muscle ionic channels were reported in Caucasian subjects but were not identified in other populations.

In view of the insulin resistance and increased Na/K-ATPase activity and increased adrenergic response observed in TPP patients, the genes encoding for the 1-, 2-, ß1-, ß2-, and ß4-subunits of Na/K-ATPase and ß-adrenergic receptor were examined. So far, no mutations or polymorphisms in these genes have been identified to be associated with TPP.

Since TPP is considered a channelopathy, with hypokalaemic being the hallmark of the disease, an international collaborative sequencing project of candidate genes of potassium voltage-gated channel subfamily J (KCNJ) gene family discovered a novel unreported gene, KCNJ18 that encodes inwardly rectifying potassium (Kir) channel 2.6, as a novel susceptibility gene of TPP. Mutations in KCNJ18 were found to be highly prevalent in TPP patients from Singapore (25.9%), and Brazil, France, and the United States (33.3%), but not in patients from Hong Kong (1.2%) or Thailand (0%). Subsequently, genome-wide association studies conducted in Hong Kong and Thailand found a common variant on chromosome 17q24.3 near KCNJ2, that encodes Kir2.1, was significantly associated with TPP. Notably, mutation in KCNJ2 has been implicated in Andersen–Tawil syndrome (ATS), a rare genetic disease that is characterized by periodic paralysis and arrhythmia. Thyrotoxicosis has been reported to severely exacerbate the periodic paralysis in

ATS patients [15]. Altogether, these genetic studies uncovered the importance of both common and rare variant in members of KCNJ gene family in TPP susceptibility.

Pathogenesis

The pathogenesis of TPP remains unclear. Hypokalaemia is due to a rapid and massive shift of plasma potassium from the extracellular into the intracellular compartment, mainly into the muscles, and is not due to depletion through losses in urine or faeces. This massive shift of potassium is believed to be due to increased sodium-potassium adenosine triphosphatase (Na^+, K^+-ATPase) pump activity in these patients. It is known that thyroid hormone can increase Na^+, K^+-ATPase activity in skeletal muscle, liver, and kidney, and also induce influx of plasma potassium into the intracellular space [16]. Thyroid hormone responsive element has been described in the promotor region of the α1 and β1 subunits of the Na^+, K^+-ATPase pump. The action of thyroid hormone on Na^+, K^+-ATPase activity is believed to be mediated through both transcriptional and post-transcriptional levels. Thyroid hormone also increases the number and sensitivity of β-adrenergic receptors. The increased β-adrenergic stimulation further increases Na^+, K^+-ATPase activity, which may explain why non-selective β-blockers can prevent attacks of TPP. The finding that selective β_1 antagonists do not protect patients from paralytic attacks is consistent with the specific role of β_2 receptor in mediating the catecholamine-induced increase in Na^+, K^+-ATPase activity in skeletal muscle [17].

As it is difficult to determine potassium transport in intact skeletal muscles during TPP and in between attacks, most studies have resorted to measurement of the potassium flux and sodium pump activity in peripheral tissues such as the red blood cells, leucocytes, and platelets. Various groups have shown that the number of Na^+, K^+-ATPase pumps, as well as Na^+, K^+-ATPase mediated cation influx, were increased in leucocytes [18] and platelets [19] in thyrotoxic patients with or without TPP when compared to healthy controls. However, TPP patients have significantly higher pump capacity and activity than those with plain thyrotoxicosis. When thyrotoxicosis is controlled, the Na^+, K^+-ATPase activity in TPP patients returns to levels similar to those of healthy subjects.

Insulin stimulates Na^+, K^+-ATPase and plays a permissive role for the potassium shift in TPP. Serum insulin levels vary widely in spontaneous attacks or during induction of paralysis, but hyperinsulinaemia during the attack or after glucose challenge has been reported in TPP [20]. The hyperinsulinaemic response may explain the association of the paralysis with heavy meals or sweet snacks. Exercise releases potassium from muscle while rest promotes influx of potassium, which may explain why mild exercise may abort an attack. It would thus appear that TPP subjects have an underlying predisposition for activation of Na^+, K^+-ATPase activity, and that thyroid hormone and insulin enhance the exaggerated response of the pump activity in these subjects. It is of interest to note that Na^+, K^+-ATPase activity is possibly increased by androgens and inhibited by oestrogens, and this may explain the male predilection for TPP [21].

As aforementioned, KCNJ gene family plays an important role in TPP susceptibility, thus it is expected that the Kir channels that encoded by KCNJs are involved in the pathogenesis of TPP. The inward and outward K^+ movement is controlled by not only Na^+, K^+-ATPase, but also Kir channels. TPP susceptibility genes *KCNJ2* and *KCNJ18* encode Kir2.1 and Kir2.6, which form functional homotetramer and heterotetramer in skeletal muscle and regulate excitability of muscles. Mutations in Kir2.6 exert dominant negative effect and reduce the whole cell current [22] predisposing the sarcolemma to hypokalaemia-induced depolarization that in turn leads to Na^+ channel inactivation and inexcitability of muscles. Notably, thyroid response element is present in the promoter region of *KCNJ18* [23], therefore this potentially explains why the attack is induced by hyperthyroidism and resolved once euthyroid state is achieved. On the other hand, the genetic variant rs312691 is located in a long non-coding RNA that regulates KCNJ2 expression during hyperthyroid state [24]. This explains the underlying mechanism how genetic variation in KCNJ2 and KCNJ18 affect TPP susceptibility.

Treatment

Treatment of TPP consists of two components: the acute management of the paralytic attack and the definitive treatment of hyperthyroidism. During the paralysis associated with marked hypokalaemia, treatment with intravenous potassium can hasten the recovery of muscle function and prevent cardiac arrhythmia. However, the serum potassium level has to be monitored closely, as rebound hyperkalaemia may occur when the potassium is being shifted back into the extracellular compartment. The use of oral potassium supplements during the early phase of weakness can sometimes help to prevent further progression to complete paralysis. Whereas potassium replacement is most effective during paralysis, regular potassium supplements are not effective for prophylaxis against further paralytic attack. Further attacks of paralysis can be prevented by the administration of spironolactone or propranolol. The most effective agent is propranolol, a non-selective ß-blocker. At a dose of 40 mg four times a day, propranolol can prevent paralysis induced by high carbohydrate load in about two-thirds of those with a history of TPP [25]. The selective $ß_1$-antagonist, metoprolol, does not protect patients from paralytic attacks. Thyroxine and acetazolamide have been reported to reduce the frequency of attacks in FHPP whereas the reverse is the case with TPP.

Patients should be advised to avoid the factors that may precipitate the attack, including heavy carbohydrate intake, alcohol ingestion, and excessive exertion. However, since patients will not have further paralytic attacks when they are euthyroid, adequate control of hyperthyroidism is necessary. Definitive treatment of the hyperthyroidism with radioactive iodine or thyroidectomy is indicated. It has to be noted that TPP may occur after radioactive iodine therapy when the patient is still toxic, and addition of antithyroid drugs for several weeks after radioactive iodine therapy may be necessary to establish a euthyroid state. When treatment leads to hypothyroidism, careful monitoring of the thyroxine replacement therapy is essential to avoid overtreatment, which may lead to a recurrence of paralytic attack.

REFERENCES

1. Kelley DE, Gharib H, Kennedy FP, Duda RJ, Jr., McManis PG. Thyrotoxic periodic paralysis. Report of 10 cases and review of electromyographic findings. *Arch Int Med*, 1989; **149**: 2597–600.

2. McFadzean AJ, Yeung R. Periodic paralysis complicating thyrotoxicosis in Chinese. *BMJ*, 1967; **1**: 451–5.
3. Tinker TD, Vannatta JB. Thyrotoxic hypokalemic periodic paralysis: report of four cases and review of the literature. *J Okla State Med Assoc*, 1987; **80**: 76–83.
4. Shizume K, Shishiba Y, Kuma K, Noguchi S, Tajiri J, Ito K, Noh JY. Comparison of the incidence of association of periodic paralysis and hyperthyroidism in Japan in 1957 and 1991. *Endocrinol Japon*, 1992; **39**: 315–18.
5. Jones P, 3rd, Papadimitropoulos L, Tessaro MO. Thyrotoxic periodic paralysis in a pediatric patient. *Pediatr Emerg Care*, 2014; **30**: 35–7.
6. Ahlawat SK, Sachdev A. Hypokalaemic paralysis. *Postgrad Med J*, 1999; **75**: 193–7.
7. Crane MG. Periodic paralysis associated with hyperthyroidism. *Calif Med*, 1960; **92**: 285–8.
8. Liu YC, Tsai WS, Chau T, Lin SH. Acute hypercapnic respiratory failure due to thyrotoxic periodic paralysis. *Am J Med Sci*, 2004; **327**: 264–7.
9. Yeo PPB, Lee KO, Cheah JS. Hyperthyroidism and periodic paralysis. In: Imura H, Shizume K, Yoshida S, eds. *Progress in Endocrinology*, vol. 2. Amsterdam: Excerpta Medica, 1988: 1341–6.
10. Chang CC, Cheng CJ, Sung CC, Chiueh TS, Lee CH, Chau T, Lin SH. A 10-year analysis of thyrotoxic periodic paralysis in 135 patients: focus on symptomatology and precipitants. *Eur J Endocrinol*, 2013; **169**: 529–36.
11. Cheah JS, Tock EP, Kan SP. The light and electron microscopic changes in the skeletal muscles during paralysis in thyrotoxic periodic paralysis. *Am J Med Sci*, 1975; **269**: 365–74.
12. Ober KP. Thyrotoxic periodic paralysis in the United States. Report of 7 cases and review of the literature. *Medicine*, 1992; **71**: 109–20.
13. Kung AW, Lau KS, Fong GC, Chan V. Association of novel single nucleotide polymorphisms in the calcium channel alpha 1 subunit gene (Ca(v)1.1) and thyrotoxic periodic paralysis. *J Clin Endocrinol Metab*, 2004; **89**: 1340–5.
14. Dias da Silva MR, Cerutti JM, Tengan CH, *et al*. Mutations linked to familial hypokalaemic periodic paralysis in the calcium channel alpha1 subunit gene (Cav1.1) are not associated with thyrotoxic hypokalaemic periodic paralysis. *Clin Endocrinol*, 2002; **56**: 367–75.
15. Diaz-Manera J, Querol L, Alejaldre A, *et al*. Severe exacerbation of Andersen-Tawil syndrome secondary to thyrotoxicosis. *J Human Genet*, 2014; **59**: 465–6.
16. Curfman GD, Crowley TJ, Smith TW. Thyroid-induced alterations in myocardial sodium-potassium-activated adenosine triphosphatase, monovalent cation active transport, and cardiac glycoside binding. *J Clin Invest*, 1977; **59**: 586–90.
17. Layzer RB. Periodic paralysis and the sodium-potassium pump. *Ann Neurol*, 1982; **11**: 547–52.
18. Khan FA, Baron DN. Ion flux and Na+, K+-ATPase activity of erythrocytes and leucocytes in thyroid disease. *Clin Sci*, 1987; **72**: 171–9.
19. Chan A, Shinde R, Chow CC, Cockram CS, Swaminathan R. *In vivo* and *in vitro* sodium pump activity in subjects with thyrotoxic periodic paralysis. *BMJ*, 1991; **303**: 1096–9.
20. Lee KO, Taylor EA, Oh VM, Cheah JS, Aw SE. Hyperinsulinaemia in thyrotoxic hypokalaemic periodic paralysis. *Lancet*, 1991; **337**: 1063–4.
21. Fraser CL, Sarnacki P. Na+- K+-ATPase pump function in rat brain synaptosomes is different in males and females. *Am J Physiol*, 1989; **257**(2 Pt 1): E284–9.
22. Cheng CJ, Lin SH, Lo YF, *et al*. Identification and functional characterization of Kir2.6 mutations associated with non-familial hypokalemic periodic paralysis. *J Biol Chem*, 2011; **286**: 27425–35.
23. Ryan DP, da Silva MR, Soong TW, *et al*. Mutations in potassium channel Kir2.6 cause susceptibility to thyrotoxic hypokalemic periodic paralysis. *Cell*, 2010; **140**: 88–98.
24. Melo MCC, de Souza JS, Kizys MML, *et al*. Novel lincRNA susceptibility gene and its role in etiopathogenesis of thyrotoxic periodic paralysis. *J Endocr Soc*, 2017; **1**: 809–15.
25. Yeung RT, Tse TF. Thyrotoxic periodic paralysis. Effect of propranolol. *Am J Med*, 1974; **57**: 584–90.

3.3.3 Thyrotoxic Storm

Leonard Wartofsky, Dorina Ylli, and Joanna Klubo-Gwiezdzinska

Introduction

Thyrotoxic storm is arguably the most serious complication of hyperthyroidism and occurs in 1–2% of hospital admissions for thyrotoxicosis with an incidence of 0.2 per 100 000 hospitalized patient per year [1] and mortality rate ranging from 10% to 75% in hospitalized patients [2–4]. It is difficult to distinguish between thyrotoxic storm and uncomplicated thyrotoxicosis simply on the basis of routine function tests. Rather, the clinical diagnosis is based on the identification of signs and symptoms typically seen in thyrotoxic storm suggesting organ decompensation. Some of these typical manifestations include fever (temperature usually above 38.5°C), tachycardia out of proportion to the fever, central nervous system signs varying

Table 3.3.3.1 Diagnostic criteria for thyroid storm

Thermoregulatory dysfunction	Score	Cardiovascular dysfunction	Score
Temperature		**Tachycardia**	
99–99.9 °F (37.2–37.7°C)	5	90–109 beats/min	5
100–100.9 °F (37.8–38.2°C)	10	110–119 beats/min	10
101–101.9 °F (38.3–38.8°C)	15	120–129 beats/min	15
102–102.9 °F (38.9–39.3°C)	20	130–139 beats/min	20
103–103.9 °F (39.4–39.9°C)	25	≥140 beats/min	25
≥104 °F (40°C) or higher	30	**Congestive heart failure**	
Central nervous system effects		Absent	0
Absent	0	Mild (oedema)	5
Mild agitation	10	Moderate (bibasilar rales)	10
Delirium, psychosis, lethargy	20	Severe (pulmonary oedema)	15
Seizure or coma	30	**Atrial fibrillation**	
Gastrointestinal dysfunction		Absent	0
Absent	0	Present	10
Diarrhoea, nausea, vomiting, or abdominal pain	10	**History of precipitating event (surgery, infection, etc.)**	
Unexplained jaundice	20	Absent	0
		Present	10
Based upon the total score, the likelihood of the diagnosis of thyrotoxic storm is: unlikely <25; impending 25–44; highly likely >45.			

Reproduced with permission from Burch HB, Wartofsky L. Life-threatening thyrotoxicosis. Thyroid storm. *Endocrinol Metab Clin North Am*, 1993; 22: 263–77.

from confusion to apathy and even coma, and gastrointestinal dysfunction, which can include nausea, vomiting, diarrhoea, and, in severe cases, jaundice. A scale has been developed to aid in diagnosis (Table 3.3.3.1) [2]. The earliest possible diagnosis and treatment initiation are required to avoid a fatal outcome.

Clinical Features

The patient's history may include hyperthyroidism but decompensation into thyrotoxic crisis usually follows some precipitating event, as indicated in Box 3.3.3.1. Most patients will have obvious signs and symptoms of thyrotoxicosis, including goitre and perhaps Graves' ophthalmopathy. However, in older patients, particularly those who may have an underlying toxic multinodular goitre rather than Graves' disease, the thyrotoxic storm may atypically present as masked or apathetic thyrotoxicosis [5]. Even more rarely, thyrotoxic storm may occur with subacute thyroiditis or factitious thyrotoxicosis due to thyroxine overdose [6, 7].

In hospitalized patients, infection is the most common precipitating event associated with thyrotoxic storm [1]. The differential diagnosis between true storm and uncomplicated infection in a thyrotoxic patient may be quite difficult, because of the likely presence of signs of tachycardia and fever in both. In this regard, very high fever seemingly out of proportion to an apparent infection along with dramatic diaphoresis could be a strong clinical clue to impending thyrotoxic storm. As the storm progresses, symptoms of central nervous system dysfunction will appear in various forms as increasing agitation, confusion, paranoia, psychosis, and even coma [8]. Cases of status epilepticus and stroke have also been reported [9]. Less frequent events associated with thyroid storm include pregnancy, during labour, placenta praevia, hydatidiform mole [10], cytotoxic chemotherapy for acute leukaemia, aspirin overdose [11–13], organophosphate intoxication [14], new anticancer medication (sorafenib) [15], nivolumab, ipilimumab [16, 17] and thyroid trauma such as repetitive examination of a large Graves' gland and surgery.

Box 3.3.3.1 Events associated with precipitation of thyrotoxic storm

- More common:
  - Withdrawal of antithyroid drug treatment
  - Iodine-131 treatment
  - Sepsis, infection
  - Surgery, trauma
  - Iodinated contrast dyes
  - Parturition
  - Vigorous palpation of thyroid
  - Burn injury
  - Diabetic ketoacidosis
  - Pulmonary thromboembolism
- Less common
  - Hypoglycaemia
  - Emotional stress
  - Subacute thyroiditis
  - Thyroxine overdosage
  - Cytotoxic chemotherapy
  - Aspirin overdosage
  - Organophosphates
  - Seizure disorder

Treatment delay increases the likelihood of irreversible progression and ultimate demise. Hence, when the diagnosis is likely but indefinite, treatment for thyrotoxic storm should be initiated.

Cardiovascular Manifestations

Cardiovascular manifestations are present due to influence of thyroid hormones on the heart, arteries, and veins. Rhythm disturbances commonly seen include sinus tachycardia, atrial fibrillation, or other supraventricular tachyarrhythmias. The signs and symptoms of congestive heart failure may be present at any age but prevail in older people. Most patients will have systolic hypertension with widened pulse pressure. Due to high oxygen demand and coronary artery spasm, myocardial infarction can be observed, even in young patients [18, 19]. A relatively rare complication is pulmonary hypertension, presumed to be on an autoimmune basis when associated with Graves' disease but may also be secondary to an augmented blood volume, cardiac output, and sympathetic tone, leading to pulmonary vasoconstriction and increased pulmonary arterial pressure, a condition usually reversible after treatment with antithyroid drugs. The other possible reason for pulmonary hypertension is pulmonary embolism due to the thrombotic or hypercoagulable state observed in severe hyperthyroidism.

Gastrointestinal Manifestations

The most common symptoms are diffuse abdominal pain, vomiting, or diarrhoea which can cause volume depletion, postural hypotension, and shock with vascular collapse. The pathophysiological mechanisms are complex, but impaired neurohormonal regulation of gastric myoelectrical activity with delayed gastric emptying plays an important role [20]. Other gastrointestinal manifestations could include presentation as an acute abdomen [21], intestinal obstruction [22], and/or hepatomegaly, splenomegaly, and various abnormalities in liver function tests. The presence of jaundice is another poor prognostic sign and warrants immediate and vigorous therapy.

Acid–Base Balance and Renal and Electrolyte Manifestations

Due to augmented lipolysis and ketogenesis, ketoacidosis may occur with lactic acidosis due to basal metabolic demands exceeding oxygen delivery and/or reduced hepatic clearance of lactic acid. Renal failure is not uncommon and can progress due to glomerulosclerosis, proteinuria, and oxidative stress. Also, a postrenal urinary retention due to dyssynergy of the detrusor muscle is possible [23]. Moreover, Graves' disease can be accompanied by autoimmune complex-mediated nephritis [24].

Neuropsychiatric Manifestations

Presentation of a wide range of central nervous system signs and symptoms in the hyperthyroid patient has been described as a key component for the differential diagnosis of thyroid storm to compensated thyrotoxicosis [25]. Moreover, the presence of central nervous system derangement was associated with increased risk of mortality [26]. In patients with neurological symptoms, a high index of suspicion for cerebral thrombosis should be considered [27].

Hyperthermia

Hyperthermia in thyroid crisis can represent both as defective thermoregulation by the hypothalamus and/or increased basal metabolic rate. However, sometimes pyrexia is not observed in elderly patients as part of the complex of so-called apathetic thyrotoxicosis with storm [7].

Haematological Manifestations

Hyperthyroidism may be associated with hypercoagulability due to increased concentrations of fibrinogen, factors VIII and IX, tissue plasminogen activator inhibitor 1, von Willebrand's factor, and tendency to augmented platelet plug formation [28] and increase erythropoietin secretion. Major thromboembolic complications can be responsible for 18% of deaths caused by thyrotoxicosis [28–33]. Therapeutic initiatives should be undertaken in thyroid storm to prevent thromboembolic complications. Optimal treatment requires a balance between anticoagulant dosage and the effect of vitamin K antagonists that can be potentiated by thyrotoxicosis [32].

Laboratory Findings

Relatively similar estimates of serum total T_4 and T_3, T_3 resin uptake, and the 24-h radioiodine uptake will be found in thyrotoxic storm as in uncomplicated thyrotoxicosis. Indeed, serum total T_3 levels may be within normal or low limits, as in the case of diabetic ketoacidosis or underlying systemic illness [34, 35] congruent with the 'euthyroid sick syndrome'. Thus, a low serum T_3 may obscure the diagnosis of thyrotoxicosis.

Other laboratory abnormalities may include a modest hyperglycaemia in the absence of diabetes mellitus, probably as a result of augmented glycogenolysis and catecholamines. However, when thyrotoxicosis is prolonged, hypoglycaemia may occur, particularly in older people [36]. Although most haematology values tend to be normal, a moderate leucocytosis with a mild shift to the left is common even in the absence of infection. Increased serum calcium levels may be seen perhaps due to both haemoconcentration and the effects of thyroid hormone on bone resorption, but serum sodium, potassium, and chloride are usually normal. Hepatic dysfunction will result in elevated levels of serum lactate dehydrogenase, glutamic oxaloacetate transaminase, and bilirubin. Because serum cortisol levels should be elevated as in any other acute stressful situation, a normal value may be interpreted as being inappropriately low. In view of the known, albeit rare coincidence of adrenal insufficiency with Graves' disease, one should maintain a reasonably high index of suspicion for this disorder, particularly if there is hypotension and suggestive electrolyte abnormalities.

Pathogenesis

The precise pathogenesis underlying the precipitation of thyroid storm remains incompletely understood. The serum hormone levels themselves do not appear to be critical. One illustrative model is provided by children with extraordinarily high serum T_4 and T_3 concentrations after accidental ingestion of T_4 in whom storm is not seen. However, the dramatic clinical improvement seen in storm after an abrupt decrease in serum T_4 or T_3 by peritoneal dialysis or plasmapheresis suggests that hormone elevation does play a role [37, 38]. In general, serum total T_4 and T_3 values do not differ significantly from those in uncomplicated thyrotoxicosis, although the levels in an affected person could be higher than the values before the precipitating event.

We believe that the critical factor relates to the actual 'free' concentration of thyroid hormone and not the 'total' measured hormone in blood. This former concentration is directly associated with the relationship of the hormone to its circulating binding proteins. Thus, any perturbation of hormone binding could increase the absolute concentration of free hormone. Conditions associated with decrease in binding affinity include surgery, anaesthesia, stress, infection, burns [39], and ketoacidosis [34, 40]. While this explanation may apply in thyroid storm [41], the pathogenesis may involve more than one factor. For example, because older patients with a systemic illness would have decreased binding and higher free T_4, a more cautious approach to radioiodine therapy in such patients is recommended to avoid thyrotoxic crisis [42].

A possible interaction between the effects of thyroid hormone and catecholamines has been a subject of both research and clinical interest for decades. Although normal serum catecholamine levels and urinary excretion rates mitigate against the idea of augmented adrenergic activity, dramatic clinical improvement follows the use of agents that either deplete tissue catecholamines, such as reserpine, or block β -adrenergic receptors, such as propranolol.

Treatment of Thyroid Storm

A four-part approach is recommended for the treatment of thyroid storm. First, antithyroid drugs are used to reduce the increased thyroid hormone production and release of T_4 and T_3. Second, treatment to block the effects of the remaining but excessive circulating concentrations of free T_4 and T_3. The third arm involves treatments against the underlying systemic decompensation (e.g. fever, congestive failure, and shock). The fourth addresses any underlying precipitating illness such as infection or ketoacidosis.

Therapy Directed to the Thyroid Gland

Inhibition of new synthesis of the thyroid hormones is achieved by administration of thionamide antithyroid drugs, such as carbimazole, propylthiouracil, or methimazole (tapazole). These drugs are given by mouth or if necessary per rectum. There is an intravenous form of thiamazole used in European countries. Propylthiouracil can be started at a loading dose of 500–1000 mg, then 250 mg every 4 h [43]. In the case of methimazole, the daily dose is approximately one-tenth of that of propylthiouracil or 60–80 mg/d [43]. Some experienced clinicians believe that propylthiouracil will provide more rapid clinical improvement because of the additional advantage of inhibiting conversion of T_4 to T_3, a property not shared by methimazole. However recent studies have not observed significant differences in disease severity or mortality between patients with thyroid storm treated with methimazole (MMI) or propylthiouracil (PTU) [44].

Because thionamides reduce new hormone synthesis but not thyroidal secretion of preformed glandular stores of hormone, either inorganic iodine or lithium carbonate may be used for this purpose. Iodides may be given either orally as Lugol's solution or as a saturated solution of potassium iodide (5 drops every 6 h) [43]. The sequence of administration of iodine and antithyroid drugs to thyrotoxic patients is very important. Use of iodine without prior thionamide dosage is to be avoided because the iodine will enhance thyroid hormone synthesis, enrich hormone stores within the gland, and thereby permit further exaggeration of thyrotoxicosis. However, when iodine is administered subsequent to, and in conjunction with full doses of antithyroid drugs, dramatic rapid decreases in serum T_4 are seen, with values approaching the normal range within 4 or 5 days [45].

Other agents that may be used in this manner are the radiographic contrast dyes ipodate (Oragrafin) and iopanoic acid (Telepaque). After a loading dose of 3 g, ipodate may be administered as 1 g orally on a daily basis and, like iodine, should only be employed with simultaneous thionamide.

In patients who may be allergic to iodine, lithium carbonate may be used [46], although some caution has been raised in regard to its use in the setting of storm [47]. Lithium should be administered initially as 300 mg every 6 h, with subsequent adjustment of dosage as necessary to maintain serum lithium levels at about 1 mmol/L.

Therapy Directed at the Continuing Effects of Thyroid Hormone in the Periphery

Therapies to reduce high levels of circulating T_4 and T_3 should be applied. Peritoneal dialysis or plasmapheresis [37, 38] have been used, as has experimental haemoperfusion through a resin bed [48] or charcoal columns [49, 50]. Such aggressive management should be considered in a severe case. Oral administration of cholestyramine resin [51] provides a less aggressive means of removing T_4 and T_3, by binding thyroid hormone entering the gut via enterohepatic recirculation; the resin–hormone complex is then excreted. A highly aggressive approach has been employed in which plasmapheresis for rapid lowering of serum T_3 and T_4 was followed by immediate total thyroidectomy [52].

Hughes [53] was the first to treat a patient with thyrotoxic storm using a β -adrenergic blocker to ameliorate the manifestations of thyroid hormone excess. Propranolol is used at an oral dosage of 60–80 mg every 4 h [43]. Indeed, because of the more rapid metabolism of the drug in severe thyrotoxicosis, even larger oral doses, or preferably intravenous doses, should be given. Initial intravenous doses of 0.5–1 mg should be given cautiously while the patient's cardiac rhythm is continuously monitored, with subsequent doses of 2–3 mg given intravenously over 10–15 min every several hours,

while awaiting clinical improvement from the effect of orally administered drug.

One caution in patients with significant underlying intrinsic cardiac disease relates to the adverse effect of adrenergic blockade in neutralizing the little remaining sympathetic drive to the myocardium. Although use of propranolol might be contraindicated in patients with moderate to severe congestive heart failure, it may be used judiciously in patients with minor cardiac compromise related to their thyrotoxicosis [54].

Patients with thyrotoxic storm, particularly those with cardiac decompensation, should be managed in an intensive care setting. Careful monitoring of fluids, volume status, and central haemodynamics is essential in these patients. In patients refractory or with contraindications to the use of propranolol, reserpine [55], guanethidine, or selective β1-blocking agents can be considered instead. A very short acting β-adrenergic blocker, esmolol, has also been used in thyroid storm with success. An initial loading dose of 0.25–0.5 mg/kg is followed by continuous infusion of 0.05–0.1 mg/kg per min [56, 57].

In spite of a theoretical risk of hypercalcaemia in thyrotoxicosis, supplementation with vitamin D_3 may actually suppress calcium release from bones [58]. In a recent study, the administration of L-carnitine 2 g/day [59] in thyrotoxic storm facilitated a dose reduction of methimazole. The mechanism appears to be related to an inhibition by L-carnitine of T_3 and T_4 entry into cell nuclei [60, 61] however, the utility of this adjunct to therapy requires confirmation.

Therapy Directed at Systemic Decompensation

Fluid depletion caused by the hyperpyrexia and diaphoresis, as well as by vomiting or diarrhoea if present, must be vigorously replaced to avoid vascular collapse and help correct hypercalcaemia. Intravenous fluids containing 10% dextrose in addition to electrolytes will better restore depleted hepatic glycogen. For fever, acetaminophen rather than salicylates is the preferred antipyretic, because salicylates inhibit thyroid hormone binding and could increase free hormone blood levels. External cooling with alcohol sponging, ice packs, or a hypothermia blanket also may be used. Vitamin supplements may be added to the intravenous fluids to replace probable coexistent deficiency. When present, congestive heart failure should be treated with the usual measures, including digoxin and diuretics, although somewhat greater than usual doses of digoxin may be required.

Hypotension not readily reversed by adequate hydration may temporarily require pressor therapy and glucocorticoids given on the basis of both postulated relative adrenal insufficiency and the ability to inhibit conversion of T_4 to T_3. An initial dose of 300 mg hydrocortisone followed by 100 mg every 8 h during the first 24–36 h should be adequate. Thyroid storm has been reported to recur when steroids had been discontinued after initial clinical improvement [62].

Therapy Directed at the Precipitating Illness

Thyrotoxicosis may have existed in many patients presenting in thyroid storm until some precipitating event, such as infection, led to increments in free T_4. Thus, therapy is not complete unless a diagnosis of the precipitating event is made, and early treatment for the underlying illness is implemented. This is not a problem in obvious cases, such as trauma, surgery, or labour, however, patients with enhanced thyroid secretion caused by withdrawal of thiourea treatment, iodine administration, iodinated contrast dyes, or ^{131}I may require some specific attention. Since the premature withdrawal of treatment could result in an exacerbation of thyrotoxicosis, an effective blockade of hormone biosynthesis and release must be continued beyond the period of immediate improvement.

Conditions such as ketoacidosis, pulmonary thromboembolism, or stroke may underlie thyrotoxic crisis and require the same vigorous management ordinarily indicated. In the patient with thyrotoxic crisis in whom none of the latter precipitating factors is apparent, a diligent search for some focus of infection must be carried out. Broad-spectrum antibiotic coverage on an empirical basis may be required initially while awaiting results of cultures. In most patients who survive thyrotoxic crisis, clinical improvement is dramatic within 12 or 24 h. The subsequent 72–96 h will be marked by continued progressive recovery. During this recovery period, supportive therapy such as corticosteroids, antipyretics, and intravenous fluids, may be tapered and gradually withdrawn on the basis of patient status.

After the crisis, attention may be turned to consideration of future short- and long-term management of the patient's thyrotoxicosis. Radioiodine treatment is often precluded by the recent use of inorganic iodine in virtually all cases of storm, but it could be considered at a later date. Continuing treatment with antithyroid drugs alone in the hope of the patient's sustaining a spontaneous remission is also possible. Surgery may be chosen, as many physicians favour a more definitive form of therapy. Should thyroidectomy be considered, thyrotoxicosis must have been adequately treated beforehand, to obviate any likelihood of another episode of crisis during or immediately following the surgery. When a surgical approach is selected, a total thyroidectomy is the procedure of choice [63].

REFERENCES

1. Akamizu T, Satoh T, Isozaki O, *et al.* Diagnostic criteria, clinical features, and incidence of thyroid storm based on nationwide surveys. *Thyroid*, 2012; **22**: 661–79.
2. Burch HB, Wartofsky L. Life-threatening thyrotoxicosis. Thyroid storm. *Endocrinol Metab Clin North Am*, 1993; **22**: 263–77.
3. Dillman WH. Thyroid storm. *Curr Ther Endocrinol Metab*, 1997; **6**: 81–5.
4. Tietgens ST, Leinung MC. Thyroid storm. *Med Clin North Am*, 1995; **79**: 169–84.
5. Feroze M, May H. Apathetic thyrotoxicosis. *Int J Clin Pract*, 1997; **51**: 332–3.
6. Swinburne JL, Kreisman SH. A rare case of subacute thyroiditis causing thyroid storm. *Thyroid*, 2007; **17**: 73–6.
7. Yoon SJ, Kim DM, Kim JU, *et al.* A case of thyroid storm due to thyrotoxicosis factitia. *Yonsei Med J*, 2003; **44**: 351–4.
8. Aiello DP, DuPlessis AJ, Pattishall EG III, Kulin HE. Thyroid storm presenting with coma and seizures. *Clin Pediatr (Phila)*, 1989; **28**: 571–4.
9. Lee TG, Ha CK, Lim BH. Thyroid storm presenting as status epilepticus and stroke. *Postgrad Med J*, 1997; **73**: 61.

10. Chiniwala NU, Woolf PD, Bruno CP, Kaur S, Spector H, Yacono K. Thyroid storm caused by a partial hydatidiform mole. *Thyroid*, 2008; **18**: 479–81.
11. Al-Anazi KA, Inam S, Jeha MT, Judzewitch R. Thyrotoxic crisis induced by cytotoxic chemotherapy. *Support Care Cancer*, 2005; **13**: 196–8.
12. Sebe A, Satar S, Sari A. Thyroid storm induced by aspirin intoxication and the effect of hemodialysis: a case report. *Adv Ther*, 2004; **21**: 173–7.
13. Hirvonen EA, Niskanen LK, Niskanen MM. Thyroid storm prior to induction of anaesthesia. *Anaesthesia*, 2004; **59**: 1020–2.
14. Yuan YD, Seak CJ, Lin CC, Lin LJ. Thyroid storm precipitated by organophosphate intoxication. *Am J Emerg Med*, 2007; **25**: 861.
15. Haraldsdottir S, Li Q, Villalona-Calero MA, Olencki TE, Kendra K, Ing SW. Case of sorafenib-induced thyroid storm. *J Clin Oncol*, 2013; **31**: e262–4.
16. McMillen B, Dhillon MS, Yong-Yow S. A rare case of thyroid storm. *BMJ Case Rep*, 2016; 10.1136/bcr-2016-214603.
17. Yu C, Chopra IJ, Ha E. A novel melanoma therapy stirs up a storm: ipilimumab-induced thyrotoxicosis. *Endocrinol Diabetes Metab Case Rep*, 2015; **2015**: 140092.
18. Opdahl H, Eritsland J, Sovik E. Acute myocardial infarction and thyrotoxic storm—a difficult and dangerous combination. *ACTA Anaesthesiol Scand*, 2005; **49**: 707–11
19. Lee SM, Jung TS, Hahm JR, *et al*. Thyrotoxicosis with coronary spasm that required coronary artery bypass surgery. *Intern Med*, 2007; **46**: 1915–18.
20. Barczynski M, Thor P. Reversible autonomic dysfunction in hyperthyroid patients affects gastric myoelectrical activity and emptying. *Clin Auton Res*, 2001; **114**: 243–9.
21. Bhattacharyya A, Wiles PG. Thyrotoxic crisis presenting as acute abdomen. *J R Soc Med*, 1997; **90**: 681–2.
22. Cansler CL, Latham JA, Brown PM Jr, Chapman WH, Magner JA. Duodenal obstruction in thyroid storm. *South Med J*, 1997; **90**: 1143–6.
23. Goswami R, Seth A, Goswami AK, Kochupillai N. Prevalence of enuresis and other bladder symptoms in patients with active Graves' disease. *Br J Urol*, 1997; **804**: 563–6.
24. Kahara T, Yoshizawa M, Nakaya I, *et al*. Thyroid crisis following interstitial nephritis. *Intern Med*, 2008; **47**: 1237–40.
25. Angell TE, Lechner MG, Nguyen CT, Salvato VL, Nicoloff JT, LoPresti JS. Clinical features and hospital outcomes in thyroid storm: a retrospective cohort study. *J Clin Endocrinol Metab*, 2015; **100**: 451–9.
26. Swee du S, Chng CL, Lim A. Clinical characteristics and outcome of thyroid storm: a case series and review of neuropsychiatric derangements in thyrotoxicosis. *Endocr Pract*, 2015; **21**: 182–9.
27. Dai A, Wasay M, Dubey N, Giglio P, Bakshi R. Superior sagittal sinus thrombosis secondary to hyperthyroidism. *J Stroke Cerebrovasc Dis*, 2000; **9**: 89–90.
28. Homoncik M, Gessl A, Ferlitsch A, Jilma B, Vierhapper H. Altered platelet plug formation in hyperthyroidism and hypothyroidism. *J Clin Endocrinol Metab*, 2007; **92**: 3006–12.
29. Romualdi E, Squizzato A, Ageno W. Venous thrombosis: a possible complication of overt hyperthyroidism. *Eur J Intern Med*, 2008; **19**: 386–7.
30. Pekdemir M, Yilmaz S, Ersel M, Sarisoy HT. A rare cause of headache; cerebral venous sinus thrombosis due to hyperthyroidism. *Am J Emerg Med*, 2008; **26**: 383.
31. Osman F, Gamma MD, Sheppard MC, Franklyn JA. Clinical review 142: cardiac dysrhythmias and thyroid dysfunction: the hidden menace? *J Clin Endocrinol Metab*, 2002; **87**: 963–7.
32. Squizzato A, Gerdes VE. Thyroid disease and haemostasis: a relationship with clinical implications? *Thromb Haemost*, 2008; **100**: 727–8.
33. Lippi G, Franchini M, Targher G, *et al*. Hyperthyroidism is associated with shortened APTT and increased fibrinogen values in a general population of unselected outpatients. *J Thromb Thrombolysis*, 2008; **28**: 362–5.
34. Wartofsky L, Burman KD. Alterations in thyroid function in patients with systemic illness: the 'euthyroid sick syndrome'. *Endocr Rev*, 1982; **3**: 164–217.
35. Wartofsky L. The low T_3 or 'sick euthyroid syndrome': update 1994. In: Braverman LE, Refetoff S, eds. *Endocrine Reviews Monographs. Clinical and Molecular Aspects of Diseases of the Thyroid*. Bethesda: The Endocrine Society, 1994: 248–51.
36. Kobayashi C, Sasaki H, Kosuge K, *et al*. Severe starvation hypoglycemia and congestive heart failure induced by thyroid crisis, with accidentally induced severe liver dysfunction and disseminated intravascular coagulation. *Intern Med*, 2005; **44**: 234–9.
37. Ashkar FS, Katims RB, Smoak WM, Gilson AJ. Thyroid storm treatment with blood exchange and plasmapheresis. *JAMA*, 1970; **214**: 1275–9.
38. Herrmann J, Hilger P, Kruskemper HL. Plasmapheresis in the treatment of thyrotoxic crisis (measurement of half-concentration tissues for free and total T_3 and T_4). *Acta Endocrinol Suppl (Copenh)*, 1973; **173**: 22.
39. Naito Y, Sone T, Kataoka K, Sawada M, Yamazaki K. Thyroid storm due to functioning metastatic thyroid carcinoma in a burn patient. *Anesthesiology*, 1997; **87**: 433–5.
40. Ahmad N, Conen MP. Thyroid storm with normal serum triiodothyronine level during diabetic ketoacidosis. *JAMA*, 1981; **245**: 2516–17.
41. Brooks MH, Waldstein SS. Free thyroxine concentrations in thyroid storm. *Ann Intern Med*, 1980; **93**: 694–7.
42. McDermott MT, Kidd GS, Dodson LE, Hofeldt FD. Radioiodine-induced thyroid storm. *Am J Med*, 1983; **75**: 353–9.
43. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. *Thyroid*, 2016; **26**: 1343–421.
44. Isozaki O, Satoh T, Wakino S, *et al*. Treatment and management of thyroid storm: analysis of the nationwide surveys. *Clin Endocrinol (Oxf)*, 2016; **84**: 912–18.
45. Wartofsky L, Ransil BJ, Ingbar SH. Inhibition by iodine of the release of thyroxine from the thyroid glands of patients with thyrotoxicosis. *J Clin Invest*, 1970; **49**: 78–86.
46. Boehm TM, Burman KD, Barnes S, Wartofsky L. Lithium and iodine combination therapy for thyrotoxicosis. *Acta Endocrinol (Copenh)*, 1980; **94**: 174–83.
47. Reed J, Bradley EL III. Postoperative thyroid storm after lithium preparation. *Surgery*, 1985; **98**: 983–6.
48. Burman KD, Yeager HC, Briggs WA, Earll JM, Wartofsky L. Resin hemoperfusion: a method of removing circulating thyroid hormones. *J Clin Endocrinol Metab*, 1976; **42**: 70–8.
49. Kreisner E, Lutzky M, Gross JL. Charcoal hemoperfusion in the treatment of levothyroxine intoxication. *Thyroid*, 2010; **20**: 209–12.
50. Candrina R, DiStefano O, Spandrio S, Giustina G. Treatment of thyrotoxic storm by charcoal plasma perfusion. *J Endocrinol Invest*, 1989; **12**: 133–4.
51. Solomon BL, Wartofsky L, Burman KD. Adjunctive cholestyramine therapy for thyrotoxicosis. *Clin Endocrinol (Oxf)*, 1993; **38**: 39–43.

52. Enghofer M, Badenhoop K, Zeuzem S, *et al*. Fulminant hepatitis A in a patient with severe hyperthyroidism; rapid recovery from hepatic coma after plasmapheresis and total thyroidectomy. *J Clin Endocrinol Metab,* 2000; **85**: 1765–9.
53. Hughes G. Management of thyrotoxic crisis with a beta-adrenergic blocking agent (pronethalol). *Br J Clin Pract*, 1966; **20**: 579–81.
54. Ikram H. Haemodynamic effects of beta-adrenergic blockade in hyperthyroid patients with and without heart failure. *BMJ*, 1977; **1**: 1505–7.
55. Anaissie E, Tohme JF. Reserpine in propranolol resistant thyroid storm. *Arch Intern Med*, 1985; **145**: 2248–9.
56. Knighton JD, Crosse MM. Anaesthetic management of childhood thyrotoxicosis and the use of esmolol. *Anaesthesia*, 1997; **52**: 67–70.
57. Brunette DD, Rothong C. Emergency department management of thyrotoxic crisis with esmolol. *Am J Emerg Med*, 1991; **9**: 232–4.
58. Kawakami-Tani T, Fukawa E, Tanaka H, Abe Y, Makino I. Effect of alpha-hydroxyvitamin D_3 on serum levels of thyroid hormones in hyperthyroid patients with untreated Graves' disease. *Metabolism*, 1997; **46**: 1184–8.
59. Kimmoun A, Munagamage G, Dessalles N, Gerard A, Feillet F, Levy B. Unexpected awakening from comatose thyroid storm after a single intravenous injection of L-carnitine. *Intensive Care Med*, 2011; **37**: 1716–17.
60. Benvenga S, Lapa D, Cannavo S, Trimarchi F. Successive thyroid storms treated with L-carnitine and low doses of methimazole. *Am J Med*, 2003; **115**: 417–18.
61. Benvenga S, Ruggeri RM, Russo A, Lapa D, Campenni A, Trimarchi F. Usefulness of L-carnitine, a naturally occurring peripheral antagonist of thyroid hormone action, in iatrogenic hyperthyroidism: a randomized, double-blind, placebo-controlled clinical trial. *J Clin Endocrinol Metab*, 2001; **86**: 3579–94.
62. Kidess AJ, Caplan RH, Reynertson MD, Wickus G. Recurrence of ^{131}I induced thyroid storm after discontinuing glucocorticoid therapy. *Wis Med J*, 1991; **90**: 463–5.
63. Leow MK, Loh KC. Fatal thyroid crisis years after two thyroidectomies for Graves' disease: is thyroid tissue auto transplantation for postthyroidectomy hypothyroidism worthwhile? *J Am Coll Surg*, 2002; **195**: 434–5.

3.3.4 Subclinical Hyperthyroidism

Simon H.S. Pearce

Introduction

Subclinical hyperthyroidism (SH) is characterized by a serum thyroid stimulating hormone (TSH, thyrotropin) concentration that is below the reference range, with circulating thyroid hormone levels (both thyroxine and triiodothyronine) that are within normal limits. In a minority of patients SH represents a state of mild thyroid autonomy or borderline hyperthyroidism. However, it is frequently an asymptomatic state and in many people it is also a transient biochemical finding that does not persist. In addition, a number of medications can cause SH or a low TSH: most obviously levothyroxine, but also glucocorticoids, opiates, and compounds containing iodine. The reference range for TSH in young adults is generally between 0.4 mU/L and 4.0 mU/L. With overt hyperthyroidism, the serum TSH concentration becomes undetectable in most assays, at levels of 0.05 mU/L or less, which is often referred to as 'fully suppressed'. This allows SH to be divided into two different grades: those patients with a low but detectable serum TSH in the 0.05 to 0.4 mU/L range (grade 1 SH), and those with a fully suppressed TSH <0.05 mU/L (grade 2 SH) (Table 3.3.4.1) [1]. The former, less severe form is more frequently transient and will resolve over 1 year in around 70% of cases [2]. The latter, grade 2 SH more closely resembles hyperthyroidism with a fully suppressed TSH and is more likely to be persistent, and in the absence of external precipitating factors is still present in 80% or more of patients after one year. It is therefore key to distinguish individuals who have mild or transient SH, as a consequence of medication, intercurrent illness or advanced age who will require no treatment, from those who have intrinsic thyroid disease causing mild thyroid autonomy who may benefit from specific management of their thyroid condition.

Epidemiology and Biochemical Prognosis

SH prevalence increases with advancing age and is more frequent in women. Around 1% of a US population up to the age of 80 years who had no known thyroid disease and who were free of interfering medications were found to have a serum TSH below 0.4 mU/L [3]. The prevalence increased to 3% in participants over 80 years old. As SH is frequently a transient state, information based on a single measurement of TSH may overestimate the prevalence. In a UK study from Tayside in which TSH was measured twice, SH persisting for more than 4 months was found in 0.6% of adults, with a yearly incidence of 30 cases per 100 000 population [4]. The peak prevalence of both grades of SH was in subjects between the ages of 75 and 85 years, and 77% occurred in women. The severity of SH could be divided, with approximately a quarter having grade 2 SH (TSH <0.1 mU/L) and three-quarters with a serum TSH between 0.1 and 0.4 mU/L [4]. Despite the careful ascertainment of these cases with a persistent abnormality of serum TSH, there was still a significant spontaneous remission rate, with 40% of subjects with grade 1 SH and 30% of grade 2 SH subjects reverting to euthyroidism over 7 years of follow-up. Conversely, over the first year of follow-up progression to overt hyperthyroidism was found in 5% and 10% of patients with grades 1 and 2 SH, respectively [4]. The rate of progression to overt hyperthyroidism is also dependent on aetiology, with 34% of SH patients who had positive TRAb antibodies diagnostic of Graves' disease developing overt hyperthyroidism over a 3-year

Table 3.3.4.1 Subclinical hyperthyroidism in comparison to other thyroid states

Test	Non-thyroidal illness	Euthyroid	Subclinical hyperthyroidism grade 1	Subclinical hyperthyroidism grade 2	T3 thyrotoxicosis	Overt hyperthyroidism
TSH	N or low	N	0.05–0.4 mU/L	≤0.05 mU/L	≤0.05 mU/L	≤0.05 mU/L
fT_3	Low	N	N	N	High	High
fT_4	N	N	N	N	N	High

N, normal; fT_3, serum-free triiodothyronine; fT_4, serum-free thyroxine.

follow-up period [5]. Thus, SH has a significant spontaneous remission rate, even after repeated 'confirmatory' measurements. The rate of progression to overt hyperthyroidism is low and dependent both on the grade of SH and any underlying intrinsic thyroid condition.

Thyroid Function: Physiology and Ageing

Even in healthy people, thyroid function as assessed with serum free thyroid hormones and TSH is not static over time. For instance, there is a circadian rhythm of TSH secretion, which is greater during the hours of darkness, with around a 25% variation in an individual's serum TSH from the nocturnal peak to the midday trough [6]. Furthermore, there is a seasonal variation of TSH among the population, with higher levels in winter months (7). With rising serum chorionic gonadotropin (hCG) levels in early pregnancy, there is also a physiological stimulation of the thyroid through the direct action of hCG on the TSH receptor. This leads to a low or even fully suppressed TSH during the first trimester of many pregnancies. The degree of TSH suppression is proportional to hCG concentration leading to lower serum TSH levels in twin pregnancies [8]. In addition, there are complex changes in serum TSH with advancing age, such that the reference interval expands both at the upper and lower limits. An Italian study of healthy community-dwelling elderly individuals showed that around one-fifth of Centenarians had a serum TSH below what would be considered the lower limit of the reference range for younger individuals [9]. Thus, many apparently healthy older individuals might be considered to have subclinical thyroid disease, both SH and subclinical hypothyroidism based on TSH reference ranges derived from younger people.

There are several different changes in the thyroid axis with age. A conventional analysis is that TSH secretion by the pituitary gland is reduced with advancing age [10]. A lower amplitude of nocturnal TSH pulses has been observed in healthy elderly volunteers, which may be as a result of decreased hypothalamic thyrotropin-releasing hormone (TRH) secretion. Indeed, a blunted TSH response to TRH stimulation has been demonstrated in elderly subjects [11, 12], suggesting increased pituitary sensitivity to circulating thyroid hormone levels. However, in common with hepatic drug clearance, thyroid hormone clearance is also known to decrease with increasing age [12]. In parallel, thyroxine secretion may also be reduced, resulting in unchanged total and fT_4 concentrations. In contrast to this, total and fT_3 levels decrease in older people, which is proposed to be due to reduced peripheral conversion of T_4 to T_3. Thus, either a central change in sensitivity to thyroid hormone feedback, reduced pituitary sensitivity to TRH or reduced peripheral hormone clearance leading to reduced axis tone and less throughput of thyroid hormone (i.e. less production, less clearance), or a combination of these mechanisms, contribute to the changes observed in the hypothalamic–pituitary–thyroid (HPT) axis with ageing [10–13].

Serum thyroid hormones, including TSH also change during acute or significant illnesses. In particular, any serious condition or one with an inflammatory or infective component can lead to the changes known as 'non-thyroidal illness' or sick-euthyroid syndrome [14] (Table 3.3.4.1). A low serum fT_3, or a reduction in fT_3 that remains within the lower limit of the reference range, is the typical first change. However, with chronicity there is frequently a transient reduction in serum TSH, leading to the pattern of SH. This 'non-thyroidal illness' pattern of blood tests is the likely explanation for the significant spontaneous recovery rate seen in epidemiological studies of SH. Lastly, in patients who are hypopituitary, central hypothyroidism may be accompanied by a low serum TSH, which may precede any drop in serum free thyroxine concentration and mimic a state of SH. Thus, an understanding of these normal physiological and pathophysiological changes is needed to interpret low serum TSH levels, particularly in older people where intercurrent illness is common and low TSH is less likely to signify intrinsic thyroid disease.

Medications

Several drugs may cause a low serum TSH through different mechanisms (Table 3.3.4.2). Minor levothyroxine or liothyronine (T_3) overtreatment may cause a reduction or complete suppression of serum TSH without elevating serum fT_4 or fT_3. Several other drugs directly suppress TSH secretion at the hypothalamo-pituitary level including opiates, glucocorticoids, levodopa, and the retinoic acid derivative, etretinate. Whether these drugs actually cause a state of mild central hypothyroidism is unclear, but on occasions a low fT_4 may be found during high-dose opiate treatment in conjunction with hypogonadotropic hypogonadism and/or secondary adrenal insufficiency, suggesting that this might be the case. Metformin may also reduce serum TSH in some patient groups such as those with an underlying thyroid condition, type 2 diabetes, or polycystic ovarian syndrome [15]. This effect is likely to be owing to adenosine monophosphate (AMP) kinase activation leading to a central inhibition of TSH secretion, but has not been robustly demonstrated in otherwise healthy people. Drugs containing iodine, including amiodarone and radiographic contrast media (iopaque) may also cause a transient reduction in serum TSH, or even frank thyrotoxicosis by augmenting the production of thyroid hormone in patients with pre-existing mild thyroid autonomy. This typically resolves spontaneously within 12 weeks of a dose of contrast media, but amiodarone has a much larger volume of distribution and its effects may linger for more than a year.

Table 3.3.4.2 Factors that can cause low serum TSH

Factor	Mechanism
Physiological/Pathological	
Pregnancy	hCG stimulation of thyroid
Trophoblastic tumours	
Germ cell tumours	
Ageing	Reduced HPT axis activity
Non-thyroidal illness	
Pituitary disease	
Drugs	
Levothyroxine (T_4)	Central suppression of pituitary TSH secretion
Liothyronine (T_3)	
Opiates	
Glucocorticoids	
Levodopa	
Dopamine	
Dobutamine	
Octreotide	
Etretinate	
Metformin	
Amiodarone	Stimulation of thyroid hormone secretion
Iodinated contrast media	

Aetiology

Once medication use and intercurrent illness are excluded, low TSH and SH can be the manifestation of mild hyperthyroidism. This is most typically due to a single or multiple autonomous follicular thyroid adenomas, which when thyrotoxicosis is more severe would be known as solitary toxic nodule or toxic multinodular goitre. However, in many individuals the degree of autonomy is mild and may commonly be fluctuant, with sometimes a serum TSH at the lower end of reference range, sometimes reduced but detectable (grade 1 SH) and sometimes a fully suppressed TSH (grade 2 SH). These fluctuations may depend upon changes in dietary iodine intake, and a transient period of grade 2 SH may also be seen in such a patient following iodinated radiographic contrast media, which later recovers. It is clear though that there is not inevitable progression to overt hyperthyroidism even in cases of persistent grade 2 SH. Therefore, it is incorrect to conceptualize SH as an 'early' manifestation of nodular thyrotoxicosis; even though there will be progression in a 2–5% of cases each year, spontaneous remission is generally more probable than progression to overt disease [4]. As well as autonomous thyroid nodules, SH is less commonly associated with Graves' disease, demonstrable by its coexistence with detectable TSH receptor antibodies (TRAbs) [5]. Indeed, SH may be discovered when an apparently euthyroid patient is investigated for Graves' orbitopathy. In these circumstances there is a higher risk of progression to overt hyperthyroidism, but a decline into overt hypothyroidism is also possible in the presence of high TRAb titres. As well as these potential thyroid diseases, as mentioned just now, other factors related to the fitness, chronological age and the ageing status of the patient need to be considered before making a judgement about the aetiology of SH. If the patient is frail, with multiple comorbidities or prematurely aged, it is more likely that SH is due to their 'stage of life' rather than any intrinsic thyroid disease.

Investigations

The most helpful investigation for SH is the passage of time, as many cases will resolve and some will worsen, unveiling an underlying intrinsic thyroid condition. Unless there are unusual circumstances following an initial low TSH measurement, it is satisfactory to repeat thyroid function testing including fT_3 in 3 and 6 months' time [1]. If grade 2 hyperthyroidism persists, or there are other clues to the presence of intrinsic thyroid disease, then it is worth measuring TRAb antibodies and obtaining a ^{99}-Tc nuclide thyroid uptake scan to identify any potential 'hot' autonomous nodules. Pregnancy should be excluded where appropriate and interfering medications and coexisting non-thyroidal illnesses should be considered. With regards complications of SH, physical examination will reveal the presence of atrial fibrillation, but an electrocardiograph (ECG) or 24-hour Holter monitor record should be obtained if there is doubt or a suggestive history of paroxysmal palpitations. Similarly, measurement of bone mineral content by dual X-ray absorptiometry (DXA) may be helpful, although densitometry-defined osteoporosis is highly prevalent in older individuals rendering this test rather non-discriminatory.

Prognosis and Complications

Several large epidemiological studies have shown that community-dwelling older people who have SH have a higher incidence of atrial fibrillation than euthyroid people during 10 or more years of follow-up [16, 17]. Furthermore, some but not all epidemiological studies have shown an increased mortality in SH people compared to euthyroid individuals [17–19] (**Figure 3.3.4.1**). Indeed, meta-analysis confirms an increased risk of atrial fibrillation, heart failure, cardiovascular mortality, and all-cause mortality in populations with a TSH <0.1 mU/L [20, 21]. Pertinently, there are numerous minor abnormalities of cardiac function that can be detected in younger patients with either SH due to endogenous thyroid autonomy or levothyroxine overtreatment [22]. Similarly, bone mineral density, fractures, dementia, and cognitive impairment have all also been associated with low TSH in large epidemiological studies [23–26]. A plausible case can therefore be made that there are clear adverse associations of SH at a population level and thus it should be aggressively treated. However, there are numerous caveats to this approach, not least that there is no good quality-controlled trial showing clinical benefit from treatment of SH. Therefore, we are a long way from being able to causally implicate SH in the adverse patient outcomes that it is epidemiologically associated with.

One fundamental problem in extrapolating the results of epidemiological studies to an individual is that we know SH is commonly a transient state, so that after the 10 years of follow-up in an observational study, only a minority of people who initially had SH are likely to still have a low TSH [2, 4]. Furthermore, the majority of people who were categorized as SH in these studies had a low but not

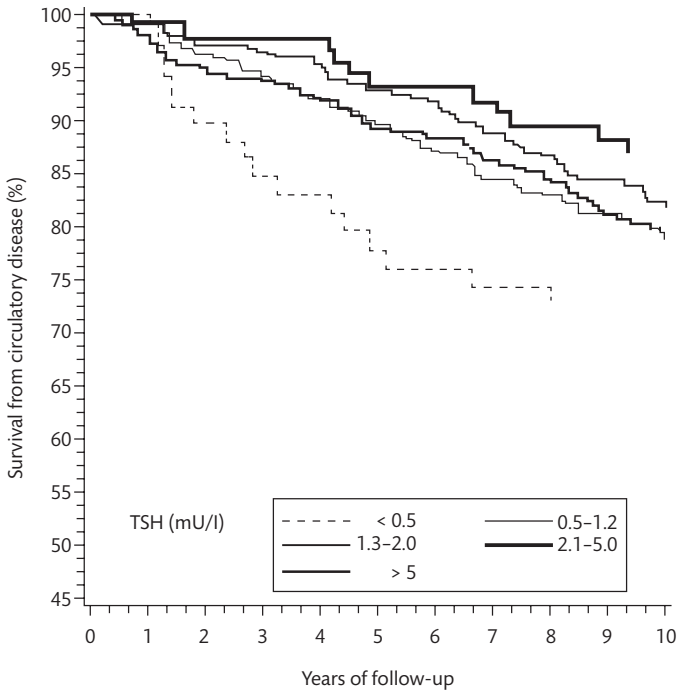


Figure 3.3.4.1 Kaplan–Meier survival curves showing the relation between survival from circulatory disease and serum thyrotropin (TSH) concentration. The lowest serum TSH group (<0.5 mU/L) has an excess mortality, but many people in this groups will have TSH at the lower end of the normal reference range; few will have grade 2 subclinical hyperthyroidism.

Reproduced from Parle JV, Maisonneuve P, Sheppard MC, Boyle P, Franklyn JA. Prediction of all-cause and cardiovascular mortality in elderly people from one low serum thyrotropin result: a 10-year cohort study. *Lancet*, 2001; 358: 861–5. Copyright © 2001 Elsevier.

suppressed TSH (grade 1 SH), which is more compatible with a 'non-thyroidal illness' pattern of thyroid function than with true thyroid autonomy. Indeed, while cardiovascular disease and osteoporosis could be aetiologically linked to thyroid autonomy, the pathophysiological mechanism for dementia and Alzheimer's disease is less clear. It therefore seems likely that in many patients low serum TSH is a marker for advanced biological age. When we compare these low TSH people to their euthyroid peers, there is an excess burden of many diseases associated with ageing, which include all the aforementioned conditions, and which might wrongly be regarded as the direct complications of SH. Nevertheless, some patients with grade 2 SH will have genuine thyroid autonomy and identifying these individuals in order to treat their underlying thyroid condition and prevent complications is the key to appropriate management.

Management

Most patients with grade 1 SH do not have thyroid autonomy and do not require specific treatment. Intermittent follow-up with once- or twice-yearly monitoring of serum TSH, fT_4 and fT_3 is sufficient to determine if there is remission of the biochemical abnormalities or if there is progression [1]. A key indicator of likely thyroid autonomy in patients with grade 2 SH is the serum fT_3 value. There is an age-related decline in the serum fT_3 reference range, such that an fT_3 value around 6.0 pmol/L is at the upper limit of normal for an 85-year-old, as compared to 6.5 or 7.0 in younger people [27]. In contrast, with frailty and comorbidity fT_3 tends to be subnormal or at the lower limits of the reference range (≤4.0 pmol/L). Thus, in the presence of persistent grade 2 SH and an fT_3 ≥5.5pmol/L, then thyroid autonomy is a serious consideration. The decision about

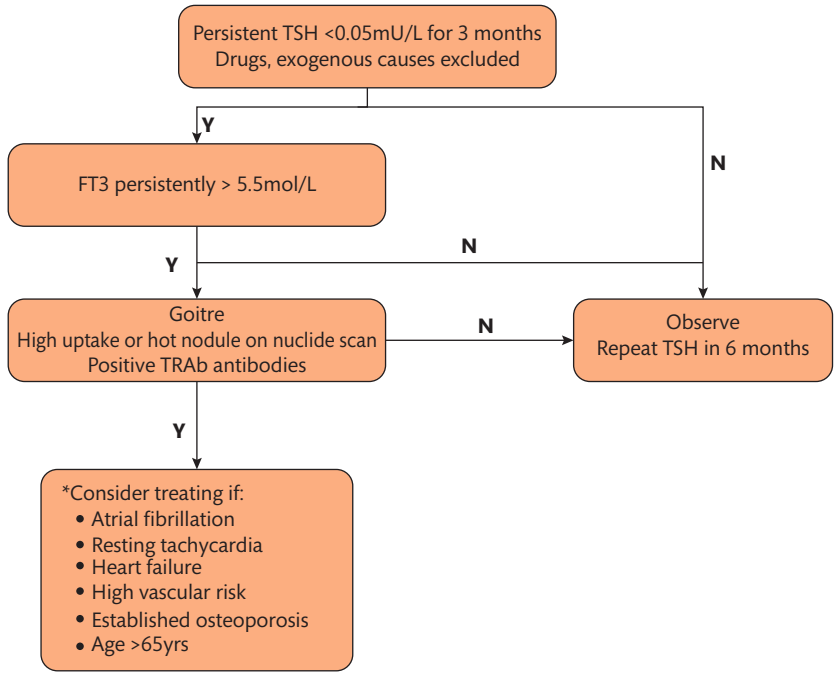


Figure 3.3.4.2 Algorithm for management of subclinical hyperthyroidism.

*First-line treatment should be antithyroid drugs in most cases.

whether to treat a patient will depend on the patient's age, the presence of potential complications (atrial fibrillation in particular) and the degree of certainty about the diagnosis of their thyroid condition (Figure 3.3.4.2). If a decision is to treat, the choice is between antithyroid drug therapy (e.g. carbimazole or methimazole) and ^{131}I radioiodine therapy. In addition, β-blockers should be prescribed for people with atrial fibrillation or other tachyarrhythmias. If investigations suggest nodular thyroid disease, then radioiodine should be considered as a curative treatment. The downside to radioiodine is that it may precipitate a period of overt hyperthyroidism (in 5–10%), or occasionally thyroid eye disease; both of which may be more troublesome to the patient than their original problem. In addition, 50% of people who receive radioiodine will ultimately become hypothyroid and dependent on daily levothyroxine. These are significant drawbacks and long-term low dose antithyroid drugs may be a superior choice, particularly for older patients. Furthermore, in the presence of TRAb antibodies, indicating Graves' disease, then antithyroid drugs are the treatment of choice, as there is an excellent chance of long-term remission follow a 12-month course. Decision-making concerning the optimal choice of therapy is hindered by lack of randomized studies. However, two uncontrolled studies have shown improvement in cardiovascular abnormalities [28, 29] and two small studies have also shown improvement in bone mineral content, both following radioiodine and antithyroid drugs [30, 31].

Summary Points

- SH is characterized by a low serum TSH with normal circulating fT_4 and fT_3.
- The commonest causes of SH are medication use and non-thyroidal illness.
- Large population studies of elderly people show that SH is associated with several adverse outcomes including atrial fibrillation (AF), heart failure, osteoporosis, and death.
- There is no randomized controlled trial (RCT) evidence that demonstrates treatment of low TSH will improve outcome.
- Treatment should be considered in patients with a fully suppressed serum TSH (<0.05 mU/L), with evidence of thyroid autonomy (fT_3 >5.5 pmol/L) and complications that may be attributed to SH.

REFERENCES

1. Mitchell AL, Pearce SH. How should we treat patients with low serum thyrotropin concentrations? *Clin Endocrinol*, 2010; **72**: 292–6.
2. Parle J, Franklyn JA, Cross K, Jones C. Prevalence and follow up of abnormal thyrotropin concentrations in the elderly in the United Kingdom. *Clin Endocrinol*, 1991; **34**: 77–83.
3. Hollowell JG, Staehling NW, Flanders WD, *et al*. Serum TSH, T4 and thyroid antibodies in the United States Population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
4. Vadiveloo T, Donnan PT, Cochrane L, Leese GP. The Thyroid Epidemiology, Audit, and Research Study (TEARS): the natural history of endogenous subclinical hyperthyroidism. *J Clin Endocrinol Metab*, 2011; **96**: E1–8.
5. Zhyzhneuskaya S, Addison C, Tsatlidis V, Weaver JU, Razvi S. The natural history of subclinical hyperthyroidism in Graves' disease: the rule of thirds. *Thyroid*, 2016; **26**: 765–9.
6. Brabant G, Prank K, Ranft U, *et al*. Physiological regulation of circadian and pulsatile thyrotropin secretion in normal man and woman. *J Clin Endocrinol Metab*, 1990; **70**: 403–9.
7. Yoshihara A, Noh JY, Watanabe N, *et al*. Seasonal changes in serum thyrotropin concentrations observed from big data obtained during six consecutive years from 2010 to 2015 at a single hospital in Japan. *Thyroid*, 2018; **28**: 429–36.
8. Grün JP, Meuris S, De Nayer P, Glinoer D. The thyrotrophic role of human chorionic gonadotrophin (hCG) in the early stages of twin (versus single) pregnancies. *Clin Endocrinol (Oxf)*, 1997; **46**: 719–25.
9. Mariotti S, Barbesino G, Caturegli P, *et al*. Complex alteration of thyroid function in healthy centenarians. *JCEM*, 1993; **77**: 1130–4.
10. Lewis GF, Alessi CA, Imperial JG, Refetoff S. Low serum free thyroxine index in ambulating elderly is due to a resetting of the threshold of thyrotropin feedback suppression. *J Clin Endocrinol Metab*, 1991; **73**: 843–9.
11. Monzani F, Del Guerra P, Caraccio N, *et al*. Age-related modifications in the regulation of the hypothalamic-pituitary-thyroid axis. *Horm Res* 1996; **46**: 107–12.
12. Gregerman RI, Gaffney GW, Shock NW, *et al*. Thyroxine turnover in euthyroid man with special reference to changes with age. *J Clin Invest*, 1962; **41**: 2065–74.
13. Mariotti S, Franceschi C, Cossarizza A, *et al*. The aging thyroid. *Endocrine Rev*, 1995, **16**: 686–715.
14. Pappa TA, Vagenakis AG, Alevizaki M. The nonthyroidal illness syndrome in the non-critically ill patient. *Eur J Clin Invest*, 2011; **41**: 212–20.
15. Lupoli R, Di Minno A, Tortora A, Ambrosino P, Lupoli GA, Di Minno MN. Effects of treatment with metformin on TSH levels: a meta-analysis of literature studies. *J Clin Endocrinol Metab*, 2014; **99**: E143–8.
16. Sawin CT, Geller A, Wolf PA, *et al*. Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N Engl J Med*, 1994; **331**: 1249–52.
17. Cappola AR, Fried LP, Arnold AM, *et al*. Thyroid status, cardiovascular risk and mortality in older adults. *JAMA*, 2006; **295**: 1033–41.
18. Parle J, Maisonneuve P, Sheppard M, Boyle P, Franklyn J. A single low thyrotrophin (TSH) concentration predicts increased all-cause and cardiovascular mortality in older persons in the community. A 10 year cohort study. *Lancet*, 2001; **358**: 861–5.
19. Gussekloo J, van Exel E, de Craen AJM, Meinders AE, Frolich M, Westendorp RGJ. Thyroid status, disability and cognitive function and survival in old age. *JAMA*, 2004; **292**: 2591–9.
20. Collet TH, Gussekloo J, Bauer DC, *et al*. Subclinical hyperthyroidism and the risk of coronary heart disease and mortality. *Arch Intern Med*, 2012; **172**: 799–809.
21. Gencer B, Collet TH, Virgini V, *et al*. Subclinical thyroid dysfunction and the risk of heart failure events: an individual participant data analysis from 6 prospective cohorts. *Circulation*, 2012; **126**: 1040–9.
22. Fazio S, Palmieri EA, Lombardi G, Biondi B. Effects of thyroid hormone on the cardiovascular system. *Recent Prog Horm Res*, 2004; **59**: 31–50.

23. Bauer DC, Ettinger B, Nevitt MC, *et al*. Risk for fracture in women with low serum levels of thyroid stimulating hormone. *Ann Intern Med*, 2001; **134**: 561–8.
24. Blum MR, Bauer DC, Collet TH, *et al*. Subclinical thyroid dysfunction and fracture risk: a meta-analysis. *JAMA*, 2015; **313**: 2055–65.
25. Kalmijn S, Mehta KM, Pols HAP, *et al*. Subclinical hyperthyroidism and the risk of dementia. The Rotterdam Study. *Clin Endocrinol*, 2000; **53**: 733–7.
26. Gan EH, Pearce SH. The thyroid in mind: cognitive function and low thyrotropin in older people. *J Clin Endocrinol Metab*, 2012; **97**: 3438–49.
27. Pearce SH, Razvi S, Yadegarfar ME, *et al*. Serum thyroid function, mortality and disability in advanced old age: the Newcastle 85+ study. *J Clin Endocrinol Metab*, 2016; **101**: 4385–94.
28. Sgarbi JA, Villaça FG, Garbeline B, Villar HE, Romaldini JH. The effects of early antithyroid therapy for endogenous subclinical hyperthyroidism in clinical and heart abnormalities. *J Clin Endocrinol Metab*, 2003; **88**: 1672–7.
29. Mark PD, Andreassen M, Petersen CL, Kjaer A, Faber J. Treatment of subclinical hyperthyroidism: effect on left ventricular mass and function of the heart using magnetic resonance imaging technique. *Endocr Connect*, 2015; **4**: 37–42.
30. Faber J, Jensen IW, Petersen L, Nygaard B, Hegedüs L, Siersbaek-Nielsen K. Normalization of serum thyrotrophin by means of radioiodine treatment in subclinical hyperthyroidism: effect on bone loss in postmenopausal women. *Clin Endocrinol (Oxf)*, 1998; **48**: 285–90.
31. Mudde AH, Houben AJ, Nieuwenhuijzen Kruseman AC. Bone metabolism during anti-thyroid drug treatment of endogenous subclinical hyperthyroidism. *Clin Endocrinol (Oxf)*, 1994; **41**: 421–4.
32. Biondi B, Cooper DS. Subclinical hypothyroidism. *N Engl J Med*, 2018; **378**: 2411–19.

3.3.5 Causes and Laboratory Investigations of Thyrotoxicosis

Francesco Latrofa and Paolo Vitti

Introduction

The term thyrotoxicosis identifies the clinical syndrome caused by elevated circulating thyroid hormones of all sources, while hyperthyroidism includes only the disorders due to an increased secretion of hormones by the thyroid gland. Hyperthyroidism is the most frequent cause of thyrotoxicosis. Unregulated discharge of preformed thyroid hormones due to destructive processes of the gland (destructive thyrotoxicosis) and exogenous ingestion or extrathyroidal production of thyroid hormones (thyrotoxicosis of non-thyroidal origin) are less common causes of thyrotoxicosis. Although careful history taking and physical examination often direct towards the diagnosis of thyrotoxicosis, laboratory confirmation by measurement of thyroid-stimulating hormone (TSH) and thyroid hormones is always needed. Once thyrotoxicosis is confirmed, laboratory testing and thyroid imaging are required to identify the cause of thyrotoxicosis.

Causes of Thyrotoxicosis

Classification

From a clinical standpoint it is useful to classify the different causes of thyrotoxicosis according to their pathogenic mechanisms. A practical classification distinguishes the forms of thyrotoxicosis into the two broad syndromes of thyroidal and non-thyroidal origin (Table 3.3.5.1). The first group can be further divided in forms associated with thyroid hormone hypersecretion (hyperthyroidism) and in forms induced by the release of preformed hormones, as consequence of destructive processes (destructive thyrotoxicosis). The second group includes a heterogeneous group of disorders, in which the thyroid is not the source of thyroid hormone. The most useful test in differentiating hyperthyroidism from the other causes of thyrotoxicosis is thyroidal radioiodine uptake (RAIU), which is high or high–normal in hyperthyroidism and low in destructive thyrotoxicosis and in thyrotoxicosis of non-thyroidal origin.

Causes of Hyperthyroidism

Graves' Disease

Graves' disease is the most frequent cause of hyperthyroidism, accounting for more than 70% of cases in iodine-sufficient areas, where its prevalence may be as high as 2% in women [1]. In Graves' disease, hyperthyroidism is caused by an autoimmune reaction to the thyroid, leading to the production of autoantibodies to the TSH receptor (TSHR autoantibodies) [2]. These antibodies mimic the action of TSH in stimulating the TSH receptor on thyroid follicular cells (TSHR-S autoantibodies).

As the effect of TSHR-S autoantibodies is exerted on all follicular cells, a diffusely enlarged thyroid is the hallmark of the disease, but in some cases thyroid nodules can develop. Graves' ophthalmopathy and, rarely, pretibial myxoedema are other typical physical findings. On careful physical exam, 30–45% of patients with Graves' disease have some signs of Graves' ophthalmopathy. When obviously present, Graves' ophthalmopathy is extremely useful in supporting Graves' disease as the cause of thyrotoxicosis. Pretibial myxoedema is only rarely observed in Graves' disease, but almost never observed without it.

Toxic Adenoma and Multinodular Toxic Goitre

Toxic adenoma and multinodular toxic goitre are frequent causes of hyperthyroidism, especially in iodine-deficient areas. Toxic adenomas are monoclonal benign encapsulated tumours

Table 3.3.5.1 Classification of known causes of thyrotoxicosis, with their distinctive diagnostic features and radioiodine uptake (RAIU) findings

Disease	Distinctive features	Neck RAIU
Thyrotoxicosis of thyroidal origin associated with hyperthyroidism		
Graves' disease	Diffuse goitre Ophthalmopathy Positive TSHR autoantibodies	High
Toxic adenoma	Single 'hot' nodule at thyroid scan	High
Multinodular toxic goitre	Multiple 'hot' nodules at thyroid scan	High
Iodine-induced thyrotoxicosis	High urinary iodine	Low to high
TSH-secreting adenomas	Inappropriately high TSH level	High
Familial gestational hyperthyroidism	Pregnancy-associated DNA analysis	Presumably high
Trophoblastic tumours	High chorionic gonadotropin	High
Neonatal transfer thyrotoxicosis	Positive TSHR autoantibodies	High
Non-autoimmune congenital and familial hyperthyroidism	TSH receptor gene mutations by DNA analysis	High
Type I amiodarone-induced thyrotoxicosis	High urinary iodine	Normal, high, low
Other drugs-induced hyperthyroidism	Positive TSHR autoantibodies	High
Thyrotoxicosis of thyroidal origin associated with thyroid destruction		
Subacute thyroiditis	Neck pain High ESR	Low
Silent thyroiditis	Positive thyroid autoantibodies	Low
Type II amiodarone-induced thyrotoxicosis	High urinary iodine	Low
Other drugs-induced thyrotoxicosis	Negative TSHR autoantibodies	Low
Thyrotoxicosis of non-thyroidal origin		
Factitious thyrotoxicosis	History Low serum thyroglobulin	Low
Dermoid tumours (struma ovarii)	Abdominal RAIU	Low
Metastatic differentiated thyroid cancer	Bone RAIU	Low

that synthesize thyroid hormones independently of TSH stimulation. They are characterized by heterozygous gain-of-function mutations involving the TSHR or the Gsα protein genes which induce a permanent and TSH-independent activation of the adenylate-cyclase pathway. Somatic mutations of the *TSHR* gene, which cause amino acid changes leading to constitutive activation of the *TSHR*, are the cause of 20–80% of toxic adenomas, while the rate of mutations of the Gsα protein range from 8% to 75% [3]. The natural history of toxic adenoma is characterized by one autonomous adenoma developing in an otherwise normal thyroid, slowly growing over many years. The coexistence or two or more toxic adenomas (multiple adenomatosis) is uncommon. In the early phases, the amount of secreted thyroid hormones is not sufficient to completely suppress TSH secretion (partial autonomy) and the function of the extranodular tissue. With further growth of the nodule, TSH suppression becomes complete, while circulating thyroid hormones are in the upper range of normal values (complete autonomy). Eventually, overt thyrotoxicosis ensues, with frankly elevated thyroid hormone levels (hyperthyroidism). The rate of progression is quite slow. The risk of overt hyperthyroidism is higher for adenomas greater than 3 cm in size.

Multinodular toxic goitre is also often detected in iodine-deficient countries, in which accounts for up to 60% of cases of thyrotoxicosis. The prevalence of multinodular toxic goitre in these iodine-deficient areas has been reduced by iodine prophylaxis [4]. The same somatic activating mutations of the TSHR demonstrated in toxic adenoma have been observed in toxic multinodular goitre as well [5]. However, in many nodules neither TSHR nor Gsα protein mutations have been observed [3]. The clonal development of follicles with high replicative capacity will induce the onset of non-functioning (cold) or hyperfunctioning (hot) nodules, in which the uptake and thyroid hormone synthesis is independent from TSH stimulation. The natural history of multinodular toxic goitre is similar to that of toxic adenoma, with the slow formation of multiple autonomously functioning nodular areas in the setting of an overall nodular goitre. The only known mechanism inducing subclinical hyperthyroidism and overt thyrotoxicosis is the administration of excessive amount of iodine. Because of the slow progression through several degrees of thyrotoxicosis and of their advanced age, patients with multinodular toxic goitre may report few symptoms.

Thyroid-Stimulating Hormone-Secreting Adenoma

TSH secretion by a benign pituitary adenoma, which is characterized by a partial or complete loss of the feedback regulation by thyroid hormones (central hyperthyroidism), causes a sustained stimulation of the thyroid gland, with the subsequent development of goitre and hyperthyroidism (see Chapter 3.3.11). An increased prevalence has been reported, probably as a consequence of the introduction of ultra-sensitive assays for TSH measurement, that enables an earlier detection of an inappropriate secretion of TSH.

hCG-Dependent Hyperthyroidism

Due to its partial homology with TSH, hCG can act as a weak (about 1/10 00th less potent) TSH agonist. When present in large amounts in the bloodstream, it can overstimulate the thyroid gland, inducing hyperthyroidism. Human chorionic gonadotropin (hCG) is secreted in large amounts by placental tissue in normal pregnancy and also by trophoblastic tumours.

Trophoblastic Tumours

Hyperthyroidism may ensue in patients with a hydatidiform mola or a choriocarcinoma as well as with chorionic tumour of the testes, as a consequence of the large quantities of hCG produced by the tumour. Thyrotoxicosis is common in trophoblastic tumours and 200–400 IU/ml of hCG are associated with high fT_4 and fT_3. However, clinical overt thyrotoxicosis is observed in a minority (10%) of patients, when hCG is extraordinarily high (>3000 IU/ml). The routine use of ultrasonography during pregnancy has led to earlier diagnosis of hydatidiform mola, when the tumour mass is smaller and the thyrotoxicosis less likely.

Hyperemesis Gravidarum

Hyperemesis gravidarum is characterized by prominent nausea and vomiting, weight loss, ketosis, and electrolyte abnormalities. It occurs in 1.5% of pregnancy and is more common in twin pregnancies because of the higher serum hCG concentrations. In 25% to 75% of cases have been reported increased levels of fT_4 and fT_3, which correlates with serum hCG concentrations. A minority of women with hyperemesis gravidarum experience a clinically evident thyrotoxicosis (gestational hyperthyroidism) [6]. It resolves spontaneously within the first 3 to 4 months of gestation.

Familial Gestational Hyperthyroidism

Two families with recurrent hyperemesis gravidarum and gestational hyperthyroidism due to a mutation in the TSHR gene causing increased responsiveness to hCG have been reported [7, 8]. Hyperthyroidism only manifests during pregnancy and recurs every time an affected woman becomes pregnant.

Fetal and Neonatal Autoimmune Hyperthyroidism

TSHR-S autoantibodies in the serum of mothers with Graves' disease can cross the placenta and cause fetal and neonatal hyperthyroidism, through direct stimulation of the fetal and neonatal thyroid. Usually the mother is affected by Graves' disease, in the majority of cases presenting with hyperthyroidism, but sometimes with hypothyroidism after thyroidectomy or ^{131}I treatment. Transplacental passage of TSHR-S autoantibodies from maternal to fetal circulation increases from low levels at 15 weeks to reach maternal levels by 30 weeks of gestation. Hyperthyroid fetus presents with tachycardia, hyperactivity, poor growth, and occasionally acceleration of skeletal maturation. Fetal goitre may occur. In severe and untreated hyperthyroidism intrauterine death has been reported.

Neonatal hyperthyroidism can be very severe and is characterized by tachycardia, jaundice, heart failure, failure to thrive. A goitre is usually present. The disease is transient and resolves within 3 to 12 weeks after birth, when TSHR-S autoantibodies disappear.

Non-Autoimmune Congenital and Familial Hyperthyroidism

After the first report of congenital hyperthyroidism caused by a germline activating mutation of the *TSHR* gene [9], only few cases of non-autoimmune neonatal hyperthyroidism have been described. The diagnosis should be suspected when a neonate presents with severe hyperthyroidism and goitre and the mother has no history of Graves' disease.

Familial hyperthyroidism due to autosomal dominant activating germline mutations of the TSHR has been described [10]. In the few cases reported, hyperthyroidism and goitre developed in adulthood because the effect of the mutation is mild. Germline mutations of the TSHR gene are uncommon in juvenile thyrotoxicosis.

Cases of congenital hyperthyroidism from mutations of the Gsα protein have been also reported, associated with McCune–Albright syndrome.

Causes of Destructive (Low RAIU) Thyrotoxicosis

Subacute Thyroiditis

Subacute thyroiditis (or granulomatous or giant cells or de Quervain's thyroiditis) is reported in Chapter 3.2.6.

Painless Thyroiditis

Painless thyroiditis (or sporadic or silent thyroiditis) is an autoimmune thyroid disorder, characterized by a transient phase of thyrotoxicosis, similar to subacute thyroiditis, in absence of neck pain and general symptoms (see Chapter 3.2.5).

Postpartum Thyroiditis

Postpartum thyroiditis (PPT) is the painless thyroiditis that occurs in susceptible women within 12 months after delivery (see Chapter 9.4).

Other Forms of Destructive Thyrotoxicosis

Rarely, destructive thyrotoxicosis can be precipitated by anterior neck injuries. Thyrotoxic crises following thyroid surgery have now become extremely rare with the optimal use of antithyroid drugs and with the refinement of surgical procedures. Thyrotoxicosis may transiently worsen or recur in patients with Graves' disease, toxic adenomas, and multinodular toxic goitre who are treated with radioiodine. Two mechanisms are responsible for this phenomenon: ongoing thyroid hyperfunction before radioiodine fully takes effect and radiation-induced thyroid destruction.

Iodine-Induced Thyrotoxicosis

Iodine deficiency increases thyrocyte proliferation and mutation rates, inducing the development of multifocal autonomous growth and of cell clones harbouring activating mutations of the TSHR.

Some of these nodules maintain the ability to store iodine and can become autonomous causing thyrotoxicosis after iodine load or even iodine supplementation. Iodine-induced thyrotoxicosis is by far more prevalent in aged patients and in areas of iodine deficiency (Table 3.3.5.2). Another predisposing condition is euthyroid or latent Graves' disease or Graves' disease in remission. In individual thyrotoxic patients, iodine contamination may be caused by a variety of medications and diagnostics, including lipid-soluble contrast media, disinfectants and drugs and some foods containing large amounts of iodine (Table 3.3.5.3).

Drugs

Amiodarone

The antiarrhythmic amiodarone, which contains a large amount of iodine, can cause thyrotoxicosis by two mechanisms. In iodine-deficient regions, where several elder people have nodular thyroid autonomy, and in patients with euthyroid Graves' disease, it can precipitate hyperthyroidism (type I amiodarone-induced thyrotoxicosis) [11]. In patients with no underlying thyroid disease, it can cause a thyroiditis with the release of preformed hormones (type II amiodarone-induced thyrotoxicosis). Distinction between the two forms is useful for the appropriate treatment but some patients present with a mixed form.

Lithium

Treatment with lithium is more frequently associated with hypothyroidism than thyrotoxicosis. Thyrotoxicosis due to destructive thyroiditis is more commonly reported.

Interferon

For a long period, the mainstay in the treatment of hepatitis C (HCV) has been the combination of IFN-α with ribavirin (RBV). Monotherapy with IFN-α has been shown to induce the development of thyroid autoantibodies in 20.6% and thyroid dysfunction in 2.7% of patients, while the combined treatment (IFN-α plus RBV) caused development of thyroid autoantibodies in 5.0% and thyroid dysfunction in 12.8% of patients [12]. Fifty per cent (50%) of patients with positive TPOAb before treatment developed thyroid dysfunction in comparison with 5.4% of autoantibody-negative patients. Thyrotoxicosis is usually a destructive process, presents with a mild or subclinical course, is transient, lasting just a few weeks or months while Graves' disease is uncommon. Destructive thyroiditis occurs in 5% of patients treated with IFN-α, and can lead to permanent hypothyroidism. Serum TSH should be measured every 2–3 months during IFN-α treatment, and 6 months after its discontinuation.

Highly Active Antiretroviral Therapy (HAART)

Patients infected with HIV have a higher prevalence of thyroid dysfunction when compared with the general population, euthyroid sick syndrome, Graves' disease and subclinical hypothyroidism being the most frequently reported [13].

Interleukin-2

Thyroid diseases have been reported in 10–50% of patients treated with interleukin-2, alone or in combination with other immunotherapies [14]. Hypothyroidism, thyrotoxicosis, and hypothyroidism after a phase of thyrotoxicosis have been reported.

Denileukin Diftitox

It can induce thyrotoxicosis by destructive thyroiditis or triggering of autoimmunity in predisposed individuals [15].

Thalidomide and Lenalidomide

Both hypothyroidism and thyrotoxicosis have been reported [16]. Interference with thyroid hormone secretion, reduction of iodine uptake, destructive thyroiditis by ischaemia, or immune-mediated mechanisms have been proposed as the potential causes of thyroid dysfunction.

Alemtuzumab

Alemtuzumab causes thyroid dysfunction in 30% of patients, with onset ranging from 6 to 61 month [17]. About half of the cases have been Graves' disease with or without ophthalmopathy.

Immune Checkpoint Inhibitors

Blocking of immune checkpoints, such as cytotoxic T-lymphocyte antigen-4 (CTLA4) and programmed death-1 (PD1), two coinhibitor receptors that are expressed on activated T cells, has emerged as an option for treatment of cancer.

Anti-CTLA4 Monoclonal Antibodies

Ipilimumab and tremelimumab are mAbs directed against CTLA4. Hypophysitis, which can cause central hypothyroidism, is the most severe and dose-limiting endocrine adverse effect. Thyroid disorders have been reported in 0–7.4% of patients treated with ipilimumab, with an incidence of hyperthyroidism of 0–2.8% [18]. Thyroid disorders occur in 0.5–5.2% of patients treated with tremelimumab

Anti-PD-1 Monoclonal Antibodies

PD-1 is a negative regulatory receptor expressed on T and B lymphocytes and natural killer cells which limits their response. Nivolumab and pembrolizumab are anti-PD-1 mAb approved for treatment of advanced malignant tumours. Thyroid dysfunction has been observed in 9% of treated patients, with 3% developing hyperthyroidism [19].

Thyrotoxicosis of Extrathyroidal Origin

Thyrotoxicosis Factitia

The term thyrotoxicosis factitia describes the voluntary excessive ingestion of thyroid hormone preparations with the purpose of mimicking thyrotoxicosis (from the Latin *factitius* = fake). However, the term has been widely applied to all forms of thyrotoxicosis due to the ingestion of thyroid hormones. True thyrotoxicosis factitia is most often observed in women with psychiatric disturbances, with access to the thyroid medication. Very often thyroid hormone is taken for weight reduction or to receive medical attention. The diagnosis is rarely obtained at history taking and denial of thyroid hormone assumption may be extreme. Accidental or suicidal ingestion of large amounts of thyroid hormone has been also described, but this can usually be diagnosed by history alone. Sometimes, thyroid hormone is inadvertently taken as a component of 'herbal' or 'alternative' medications, usually for weight reduction. Finally, accidental

grinding of cattle thyroids in hamburger meat has been reported as the cause of an outbreak of thyrotoxicosis among hamburger consumers in the US [20].

Struma Ovarii

Struma ovarii is a on ovarian tumour, usually a teratoma composed of at least 50% thyroid cells. It comprises about 3% of ovarian teratomas, is bilateral in 10% and malignant in 5% of cases. Thyrotoxicosis occurs in 10% of cases.

Functional Metastatic Thyroid Carcinoma

Differentiated thyroid carcinoma, even when metastatic and with large tumour burdens, does not usually produce relevant amounts of thyroid hormones. Very rarely however, thyroid carcinomas of the follicular histotype, extensively metastatic to the bone may cause thyrotoxicosis. Coexistent TSHR-S autoantibodies are an extremely rare cause of thyrotoxicosis in patients with metastatic thyroid cancer.

Laboratory Diagnosis of Thyrotoxicosis

Thyroid-Stimulating Hormone

The mainstay of the diagnosis of thyrotoxicosis is measurement of serum TSH and thyroid hormones. In fact, many laboratories measure only TSH in the initial assessment and measurement of thyroid hormones is automatically added only if TSH is abnormal ('TSH reflex'). Indeed, TSH concentration is inversely log-linearly proportional to fT_4 level [21]. The current immunoassays are very sensitive and can measure TSH levels well below the normal range, with a functional sensitivity of less than 0.02 mU/L. Since pituitary TSH secretion is tightly downregulated by thyroid hormone level, TSH is undetectable in most cases of thyrotoxicosis. The only remarkable exception are TSH-secreting adenomas, in which a high or inappropriately normal TSH level is found in spite of overt thyrotoxicosis. Because of the sensitivity of the assay, low (less than 0.4 mU/L) but detectable TSH levels can be found. These levels are encountered in subclinical thyrotoxicosis and in other conditions, such as non-thyroidal illnesses, endogenous or exogenous corticosteroid excess. TSH is a heterogeneous molecule and different TSH isoforms circulate in the blood and are present in pituitary extracts used for assay standardization [22]. Although current methods have eliminated cross-reactivity with other glycoprotein hormones, they may detect different epitopes of abnormal TSH isoforms, secreted by some euthyroid individuals and some patients with pituitary diseases. Rarely, the presence in the serum of antimouse immunoglobulin antibodies may interfere in the TSH assay, causing falsely elevated TSH levels.

Thyroid Hormone

Measurement of serum thyroid hormone levels is mandatory in all patients with suspected thyrotoxicosis, for a proper evaluation of a low TSH level and for an estimation of the severity of the disease. The active form of the hormones in serum is the very small amount of freely circulating T_4 (free T_4–fT_4) and T_3 (free T_3–fT_4), which can enter cells, interacting with the specific receptors. Total T_4 (tT_4) and total T_3 (tT_3) can be easily and inexpensively measured by radioimmunoassay, but their levels are influenced by the levels of binding protein, which vary in healthy subjects and may change in several conditions [23]. Thus, total thyroid hormone levels may not parallel those of free thyroid hormones, and their measurement is nowadays considered less useful in the evaluation of thyrotoxicosis. Free thyroid hormone levels measurements, although not completely exempt from flaws, are therefore more satisfactory, since they provide a more accurate measurement of the active hormone [24].

In iodine-sufficient countries a single measurement of fT_4 is sufficient to confirm or reject the suspicion of thyrotoxicosis and, after TSH measurement, this is the test most often used in North America for thyroid function screening [25]. In contrast, in iodine-deficient countries, a significant proportion of hyperthyroid patients (up to 12%) may have elevated fT_3 and normal fT_4 levels (T_3-toxicosis). Conversely, fT_4 can be falsely elevated in conditions causing reduced peripheral conversion of T_4 to T_3. In our practice, when thyrotoxicosis is suspected, we initially assess both fT_4 and fT_3 levels along with TSH with little additional expense in order to obtain a complete assessment of the thyroid function status.

Laboratory Investigations in the Differential Diagnosis of Thyrotoxicosis

In many cases, history and physical examination can readily identify the cause of thyrotoxicosis. However, in many other situations, a careful differential diagnosis is needed in order to establish an etiological diagnosis. Classically, RAIU has represented a mainstay of the differential diagnosis of thyrotoxicosis. RAIU is easily performed by administering a minimal (tracer) dose of radioactive iodine and then measuring the percent of administered radioactivity accumulated in the neck. In iodine-sufficient countries the upper limit of RAIU, 24 hours after the administration of the tracer, is around 25%, while it may reach 40% in areas with mild to moderate iodine deficiency. Whenever excessive active formation of thyroid hormone takes place in the thyroid gland, RAIU is increased, since the thyroidal machinery for iodine trapping and organification is activated. Therefore, a high RAIU readily identifies true hyperthyroidism (e.g. with thyroid hyperfunction). In contrast, thyrotoxicosis with a low RAIU indicates either thyroidal destruction, with release of preformed hormone, or an extrathyroidal source of thyroid hormone. In thyroid destruction, the damaged follicular cells transiently lose their capability of iodine trapping, while when exogenous hormones are administered in excess, the suppression of the pituitary secretion of TSH causes shutting-off of the trapping capacity of follicular cells. The only exception to this rule is iodine-induced thyrotoxicosis, in which a low RAIU can be observed because of dilution of the tracer dose in the large body pool of iodine, in spite of true hyperthyroidism.

Nowadays, RAIU is not performed in the initial assessment of a thyrotoxic patient and a vast array of laboratory and imaging techniques have provided excellent tools for accurately identifying the cause of thyrotoxicosis without the information provided by RAIU. RAIU is still useful in difficult cases, to broadly define forms of thyrotoxicosis according to their pathogenesis and orderly proceed with adjunctive diagnostic tools.

Graves' Disease

TSH Receptor Autoantibodies

Since the cause of Graves' disease hyperthyroidism is uncontrolled thyroid gland stimulation by circulating TSHR autoantibodies, their detection in the serum of thyrotoxic patients is particularly useful in establishing the diagnosis of Graves' disease. Serum TSHR-S autoantibodies can be measured by different methods [2, 26]. They were originally detected with *in vivo* bioassays, and thereafter by *in vitro* systems. The most common tests assess the displacement of labelled TSH or TSHR autoantibodies from the TSHR by the immunoglobulin fraction of patients' sera. These methods are termed TSH binding-inhibition (TBI) tests and do not distinguish between TSHR-S autoantibodies and TSHR blocking (TSHR-B) autoantibodies, which can be also detected in thyroid autoimmune disorders [2]. TSHR-S autoantibodies can be tested in cellular systems carrying a functional TSH receptor, detecting the release of cAMP in the culture medium upon challenge with serum or purified immunoglobulins (TSHR-S autoantibodies assay) [2]. In a modification of the assay, TSHR-B autoantibodies can be detected as well. Since TSHR-S autoantibodies are properly the cause of hyperthyroidism in Graves' disease, their assay should be considered the gold standard in the diagnosis of Graves' disease. Unfortunately, the assay is quite expensive and requires cell-culture capabilities, making it available only to research centres. For clinical purposes, the TBI assays are most often used. By last generation assays, positive TBI tests are found in 75–95% of patients, with a high specificity (99%) [27]. TSHR autoantibodies showing inhibition of TSH binding to TSHR but not TSHR stimulating activity have been demonstrated in serum of subjects with no evidence of autoimmune thyroid diseases [28]. A TBI test is strictly needed in the minority of cases of Graves' disease, in which the clinical picture is unclear, for example in the differential diagnosis of hyperemesis gravidarum, in the nodular variants of Graves' disease that must be differentiated from toxic nodular goitre, in patients with exophthalmos without thyrotoxicosis (euthyroid Graves' disease). The presence of high levels of TSHR autoantibodies at the end of antithyroid drug therapy has a high positive predictive value and specificity for relapse of hyperthyroidism but a low negative predictive value and sensitivity [29].

Whereas TSHR-S autoantibodies interact mainly with the N-terminal components of the ectodomain, TSHR-B autoantibodies interact to a greater extent with the C terminus and to a lower degree with the N terminus and the mid-region of the TSHR [2, 30]. An assay based on a chimeric receptor in which residues 261–370 of TSHR have been substituted by the homologous portion of the LH/hCG receptor (Mc4) has been proposed to be more accurate in the detection of TSAb as well as in detecting both TSAb and TBAb autoantibodies [31, 32].

Thyroid Peroxidase (TPO) and Thyroglobulin (Tg) Autoantibodies

TPO autoantibodies can be found by commercial assays in up to 90% of patients with untreated Graves' disease, while Tg autoantibodies are less frequently positive, in 50–80% of cases. Both autoantibodies however are also present in other forms of thyroid autoimmune disorders, some of which may cause thyrotoxicosis, such as postpartum thyroiditis and silent thyroiditis. A relatively high percentage (up to 25%) of positive thyroid autoantibodies tests is also found in normal subjects, especially women [33]. Thus, TPO and Tg autoantibodies tests do not establish the diagnosis of Graves' disease as the cause of thyrotoxicosis, but may be useful as complementary tests in confirming the presence of thyroid autoimmunity.

The finding of autoantibodies cross-reacting with Tg and TPO (TgPO Ab) in patients with autoimmune thyroid diseases, which suggested a role for cross-reactivity of B cell response to Tg and TPO, has not been confirmed [34].

Total thyroid ablation (thyroidectomy plus 131-I) induces the disappearance of thyroid autoantibodies [35].

Other Thyroid Autoantibodies

Autoantibodies to megalin were detected in 50% of patients with chronic autoimmune thyroiditis and in some patients with Graves' disease and thyroid carcinoma [36]. A role of Na/I symporter as autoantigen in thyroid autoimmunity has been proposed by some authors but excluded by others. Neither megalin nor Na/I symporter autoantibodies are routinely used for diagnosing autoimmune thyroid diseases.

Thyroid RAIU and Scan

In untreated hyperthyroid Graves' disease patients, a high value of RAIU at the 24th hour is always found. As a distinctive feature, in some cases the 3rd or 6th hour value can be even higher than at the 24th hour, as an expression of an extremely high iodine turnover. The test is very useful to rule out transient thyrotoxicosis due to Hashitoxicosis or painless or subacute thyroiditis, factitious thyrotoxicosis, and type II amiodarone-induced thyrotoxicosis [11].

Thyroid imaging with radioisotopes can be performed with radioiodine at the time RAIU is done or using ^{99m}pertechnetate. Thyroid scanning in Graves' disease is useful only when coexisting nodules are detected and their functional status needs to be evaluated.

Thyroid Ultrasound

The ultrasound appearance of the thyroid gland undergoes typical changes during Graves' disease hyperthyroidism. Because of the reduction in the colloid content and of the lymphocytic infiltrate, the gland becomes diffusely hypoechoic. A similar pattern is also observed in chronic goitrous thyroiditis and, when diffuse, indicates the presence of thyroid autoimmunity. Therefore, thyroid ultrasound can be useful in confirming the suspicion of thyroid autoimmunity, during the evaluation of thyrotoxicosis. Marked hypoechogenicity at the end of antithyroid drug therapy may predict recurrence of thyrotoxicosis [37]. As an adjunctive value, thyroid ultrasound scanning also allows an accurate measurement of the goitre size, an information that is important in the choice of the most appropriate treatment. Thyroid ultrasound accurately distinguishes true thyroid nodules from the lobulations that can be occasionally felt at palpation in Graves' disease glands.

The measurement of blood flow to the thyroid gland by colour Doppler ultrasonography has been also used in Graves' disease patients. In untreated Graves' disease the colour Doppler pattern is characterized by markedly increased signals with a patchy distribution [38]. Colour Doppler studies of the thyroid gland can therefore be used similarly to RAIU in distinguishing Graves' disease from other forms of thyrotoxicosis, for example, amiodarone-induced

destructive thyrotoxicosis [39] or subacute thyroiditis and, possibly, painless thyroiditis.

Toxic Adenoma

When a single nodule is palpated in the thyroid of a patient being evaluated for thyrotoxicosis, the presence of a toxic adenoma must be always suspected. In confirming the diagnosis of thyrotoxicosis, it is important to measure both fT_4 and fT_3 levels, since T_3-toxicosis is distinctly frequent in toxic adenomas.

^{99m}Technetium or radioiodine thyroid scanning are extremely helpful in confirming the diagnosis, yielding typical findings (Figure 3.3.5.1). The nodule will appear 'warm', with the extranodular thyroid tissue clearly visible, when partial autonomy is present. In this case, parallel thyroid function tests will show a low but detectable TSH, and thyroid hormone levels in the upper part of normal range. Only the nodule is visible on the scan when TSH is completely suppressed (e.g. in case of complete autonomy or overt thyrotoxicosis).

Ultrasound scanning of the neck provides no direct diagnostic information on the functional property of the nodule, but it is useful in detecting coexisting cold nodules and accurately defines the size of the nodule. Preliminary reports have shown a distinctive colour Doppler pattern in autonomously functioning thyroid nodules, characterized by an increased blood flow in the nodular tissue, in good correlation with radionuclide scans. However, the technique is not able to distinguish benign from malignant nodules and is therefore of limited value.

Ultrasound elastography has showed high sensitivity and specificity in differentiating benign from malignant thyroid nodules [40]. The risk of malignancy in hot nodules is extremely low, although occasionally reported. Therefore, in presence of a low TSH, fine needle aspiration is only needed when coexisting nodules detected by palpation or ultrasound are cold at radionuclide scanning.

Toxic Multinodular Goitre

The same range of thyroid function test alterations described in toxic adenomas can be observed in toxic multinodular goitre, from a subnormal TSH level to an undetectable TSH level with frankly elevated thyroid hormone levels. The diagnosis of toxic multinodular goitre can often be suspected on history and physical findings.

Thyroid radionuclide scanning is quite useful in identifying and mapping autonomous nodules and distinguishing them from other coexistent, cold nodules. Scanning is also useful in adjunct to TSHR autoantibodies measurement in distinguishing true toxic multinodular goitre from Graves' hyperthyroidism which develops on a pre-existing non-toxic multinodular goitre. RAIU is always elevated, unless iodine overload is present.

Thyroid ultrasonography is also useful to measure goitre size and, in association with radionuclide scanning images, to identify cold nodules, amenable to fine needle aspiration biopsy.

TSH-Secreting Adenomas

The presence of a TSH-secreting adenoma should be suspected when a detectable TSH level in presence of clearly elevated circulating thyroid hormone levels (inappropriate secretion of TSH) is found. The first step in the evaluation of inappropriate secretion of TSH is making sure that artefacts in the measurement of TSH or thyroid hormone levels are not the cause of the laboratory findings. Falsely elevated TSH levels can be observed occasionally when heterophilic antibodies are present in the patient's serum. These antibodies are antimouse immunoglobulins that bind both the solid-phase and the labelled mouse antibodies employed in most TSH immunoradiometric assays, causing bridging between the two and therefore mimicking the presence of TSH [22]. The problem can be overcome by incubating the patient's serum with mouse immunoglobulins prior to TSH testing, thus precipitating the heterophilic antibody [41]. The most recent TSH commercial assays contain these antibodies in their incubation buffers, making this problem quite rare nowadays. Macro TSH is composed by anti-TSH autoantibodies which complex monomeric TSH. It is considered to be inactive and therefore a cause of falsely high TSH value. Falsely elevated fT_4 and fT_3 levels must also be excluded in the preliminary evaluation of suspected inappropriate secretion of TSH.

Mild spurious elevations of fT_4 and fT_3 can occasionally be found in presence of thyroid hormone autoantibodies, genetic or drug-induced alterations of thyroxine-binding globulin and in non-thyroidal illnesses [24]. The two-step methods for measurement of free thyroid hormones may be useful to rule out these conditions. About 50% of immunoassays use the biotin-streptavidin interaction as an immobilizing system and can be influenced by the ingestion of biotin (vitamin B_7). A low to medium dose of biotin is commonly present in multivitamin preparations while patients with progressive multiple sclerosis are treated with high doses of biotin (10 000 times the recommended daily intake of approximately 30 μg). In some assays, serum biotin can induce both falsely low TSH and falsely high fT_4 and fT_3 levels. A washout period ranging from 8 hours to more than 3 days has been suggested to eliminate biotin interference [42].

Once these artefacts have been excluded, extensive laboratory testing is required to clarify the cause of inappropriate secretion of TSH. True inappropriate secretion of TSH is observed in two conditions: TSH-secreting pituitary adenoma and resistance to thyroid hormone. In theory, only TSH-secreting adenomas cause true and symptomatic hyperthyroidism (see next).

The syndrome of resistance to thyroid hormone is caused by a relative insensitivity of the thyroid hormone receptor to the action of its ligand. Therefore, higher thyroid hormone concentrations are needed to down regulate TSH secretion. In the majority of patients, the defect is due to mutations of thyroid hormone β-receptor gene [43]. As a consequence of the defect, the pituitary set-point for TSH suppression is set at a higher level of circulating thyroid hormone (i.e. a higher level of thyroid hormone is required for TSH suppression). After its first identification [44] a few cases of mutations of TRα have been reported. In these patients TSH is normal, fT_4 low/low-normal and T_3 high/high-normal.

Differential Diagnosis

When the suspicion of inappropriate secretion of TSH is confirmed, the presence of a TSH-secreting adenoma must be differentiated from resistance to thyroid hormone. Because of the overlapping clinical presentation and because no single test accurately allows clear-cut differentiation between the two conditions, extensive baseline and dynamic laboratory testing is usually required (Table 3.3.5.4).

A number of tests have been used to confirm the presence of thyrotoxicosis at the tissue level, including measurement of biochemical markers of thyrotoxicosis such as sex hormone binding globulin,

Table 3.3.5.4 Laboratory investigations in the differential diagnosis of the syndrome of inappropriate secretion of TSH

Test	TSH-secreting adenomas	RTH	Comment
Peripheral markers of thyroid hormone action	High	Normal-high	Non-specific
α-subunit/TSH molar ratio	>1	1	High in menopause
TSH after T_3-suppression test	Unchanged or slightly reduced	Frankly reduced or suppressed	Hazardous in elderly and cardiopathic patient
TSH after TRH	Unchanged	Increased	
Pituitary imaging	Positive	Negative	Confirmatory

alkaline phosphatase, cholesterol, and creatine phosphokinase. Unfortunately, these parameters are quite non-specific and may be elevated (or reduced) in a number of other conditions. At variance with normal pituitary, TSH-secreting adenomas secrete the α-subunit of TSH in molar excess with respect to TSH. A serum α-subunit/TSH ratio higher than 1 is observed in approximately 90% of patients with TSH-secreting adenoma. High ratios can be also observed in postmenopausal women and even in normal subjects, making this test alone unable to establish the diagnosis. Growth hormone (GH), IGF-I, and prolactin serum measurements are useful, since about 30% of TSH-secreting adenomas cosecrete GH and prolactin.

Dynamic testing aims at demonstrating the unresponsiveness of TSH-secreting adenomas to normal stimuli. In most (92%) TSH-secreting tumours the TSH level fails to increase in response to a standard thyrotropin-releasing hormone (TRH) stimulation test, while a normal or increased response is observed in resistance to thyroid hormone [45]. Diagnostic protocols to test the response of pituitary TSH to exogenous triiodothyronine are also used. Triiodothyronine is administered orally and the dose is increased every three days, starting from 50 up to 200 mcg/daily. Before every increase basal and TRH-stimulated TSH is measured, together with peripheral markers of thyroid hormone action [46]. In TSH-secreting adenomas, only partial or no suppression of TSH secretion is observed, while complete or partial suppression is observed in resistance to thyroid hormone. Alternatively, 80–100 μg triiodothyronine can be administered for 8–10 days. Using this protocol, complete TSH suppression is obtained in normal subjects, while no changes or slight reduction in TSH levels are observed in all patients with TSH-secreting adenomas. On the other hand, clear-cut reductions of TSH levels are observed in patients with resistance to thyroid hormone [45]. The test is contraindicated in elderly patients and in patients with arrhythmias and/or coronary artery disease. The demonstration of a mutation of TRβ gene is sufficient to establish the diagnosis of syndrome of resistance to thyroid hormone.

Pituitary imaging is very important in confirming the diagnosis of TSH-secreting adenomas. Ninety per cent (90%) of TSH-secreting adenomas are larger than 1 cm at diagnosis and therefore easily detected at pituitary MRI scanning. As a complement, radiolabelled octreotide pituitary scintigraphy can be useful in detecting small tumours, although it can be positive in other types of pituitary tumours.

hCG-Dependent Thyrotoxicosis

The presence of a trophoblastic tumour should be suspected when thyrotoxicosis is found in an amenorrhoeic woman, especially when a palpable abdominal mass is found. The diagnosis is readily confirmed by the finding of extremely high circulating hCG levels and of a pelvic mass at ultrasonography. In trophoblastic tumours serum hCG levels usually exceed 200 IU/ml, whereas the peak concentration of normal pregnant women is 100 IU/ml.

The diagnosis of thyrotoxicosis during hyperemesis gravidarum can be particularly challenging and it is one of exclusion. Because of weight loss and malnutrition, fT_3 levels may be disproportionately low or even normal in comparison with fT_4 levels, due to a reduced peripheral conversion of T_4 to T_3. The TSH level is often low during early normal pregnancy, but seldom undetectable, as it is in true thyrotoxicosis. The only distinctive laboratory feature is an inappropriately high hCG level, but large overlap with normal pregnancies exists. Therefore, the diagnosis of thyrotoxicosis in hyperemesis gravidarum relies mainly on the clinical picture and on appropriate exclusion of other more common forms of hyperthyroidism by appropriate testing. RAIU, as any other *in vivo* radioisotopic procedure, is contraindicated in pregnancy.

Fetal and Neonatal Hyperthyroidism

Women with a past or present history of Graves' disease should be carefully monitored throughout pregnancy. Fetuses and neonates of pregnant women who have been previously treated with radioiodine or surgery are at high risk because they lack protective effect of antithyroid drugs administered to the mother. The presence of a fetal heart rate over 160 bpm, in the absence of other fetal abnormalities is suggestive of fetal hyperthyroidism. TSHR autoantibodies should be measured starting at 22 weeks of gestation in mother with Graves' disease; only high levels can cause fetal or neonatal autoimmune hyperthyroidism.

The persistence of high levels of TSHR autoantibodies in the maternal serum by the end of pregnancy, when the transplacental passage is maximal, is a predictor of hyperthyroidism in the neonate. Fetal cord blood sampling is a risky procedure, and it is not generally recommended. Conversely, it is very useful to test neonatal cord blood at the time of delivery for thyroid function tests and TSHR autoantibodies. When the mother has been treated with high-dose antithyroid drugs, the neonate should be re-tested 10 days after birth, since transplacental passage of methimazole or propylthiouracil may initially mask hyperthyroidism.

Neonatal hyperthyroidism in the absence of a maternal history of Graves' disease and with negative TSHR autoantibodies should raise the suspicion of non-autoimmune congenital hyperthyroidism. Familiar hyperthyroidism should be suspected when relatives are affected and serum TSHR autoantibodies are absent. In

both types of hyperthyroidism sequencing of the TSHR gene is required to confirm the diagnosis.

Iodine-Induced Thyrotoxicosis

Excessive iodine assumption should be always suspected when hyperthyroidism abruptly appears in a patient with a history of nodular thyroid disease. A careful history taking often identifies the source of iodine and all patients should be asked about recent assumption of any of the compounds listed in Table 3.3.5.3 (online). With the exception of type II amiodarone-induced thyrotoxicosis, RAIU is usually low in thyrotoxic patients with heavy iodine contamination, but it is almost never suppressed, a feature that allows distinction from subacute and painless thyroiditis. The iodine/creatinine urinary ratio is the gold standard in confirming iodine contamination and is high in all cases.

Amiodarone-Induced Thyrotoxicosis

When a history of amiodarone assumption is elicited in a thyrotoxic patient, further testing is required to distinguish between type I and type II forms of amiodarone-induced thyrotoxicosis [11]. Type I (non-destructive) amiodarone-induced thyrotoxicosis differs little from other forms of iodine-induced thyrotoxicosis and an underlying thyroid disease such as Graves' disease or nodular thyroid disease is usually detected with the appropriate diagnostic tools. Accordingly, RAIU is usually low, but not suppressed and may be normal or increased. In contrast, in type II (destructive) amiodarone-induced thyrotoxicosis, RAIU is always low or suppressed and often no clear underlying thyroid disorder can be identified. Flow by colour Doppler sonography is a useful tool in differentiating the two forms of amiodarone-induced thyrotoxicosis, vascularity being increased in type I and absent in type II [39].

Drug-Induced Thyrotoxicosis

As just reported, destructive thyroiditis or triggering of autoimmunity in predisposed individuals are the mechanisms of drug-induced thyrotoxicosis. Evaluation to differentiate Graves' hyperthyroidism from low RAIU thyrotoxicosis is recommended, in order to advice the proper treatment.

Subacute, Painless, and Postpartum Thyroiditis

Classically, subacute, painless, and postpartum thyroiditis are characterized by a low (<1%) RAIU during the thyrotoxic phase. This test alone, in the presence of a suggestive clinical presentation allows the diagnosis in almost all cases. Serum T_4 concentration is disproportionately elevated compared with T_3 concentration, reflecting the preferential release of T_4 from the injured thyroid. In subacute thyroiditis, a very high (always >50 and often >100 mm/hr) erythrocyte sedimentation rate (ESR) is a distinctive diagnostic feature. C-reactive protein is also elevated and a mild leukocytosis is often observed.

High levels of TPO and Tg autoantibodies are usually found in postpartum and painless thyroiditis, as a marker of prominent thyroid autoimmunity, while only weakly and transiently positive tests are occasionally found in subacute thyroiditis. Ultrasound findings are generally characterized by patchy areas of hypoechogenicity in subacute thyroiditis, while a more diffuse hypoechoic pattern, closely resembling Hashimoto's thyroiditis is found in postpartum and painless thyroiditis. The colour Doppler pattern shows reduced vascularity in the three disorders. Occasionally, and especially when patients are first seen in the recovery or hypothyroid phase, a more subtle picture can emerge from testing with a low but not nil RAIU, and with only mild elevations of ESR, making the differential diagnosis more difficult.

Thyrotoxicosis of Extrathyroidal Origin

An extrathyroidal source of thyroid hormone should be suspected when the other, more frequent causes of low RAIU thyrotoxicosis have been ruled out. When thyrotoxicosis factitia is suspected, a serum Tg measurement can be extremely useful in confirming the diagnosis, since this disorder represents the only condition in which thyrotoxicosis is associated with an undetectable Tg level [47]. At the time of Tg measurement, however, it is important to test the patient's serum for Tg autoantibodies, since these may cause falsely low Tg levels [48]. Colour Doppler shows hypovascularity in thyrotoxicosis factitia.

The suspicion of struma ovarii can be confirmed at the time of RAIU, simply by scanning the pelvic area with the probe. The presence of functional thyroid tissue is demonstrated by a significantly increased uptake of iodine in the ovarian region. Further imaging (CT or ultrasound scan) will confirm the presence of an ovarian mass. The levels of CA 125 are elevated in both malignant and benign tumours.

When the source of thyroid hormone is metastatic thyroid follicular cancer, the presence of the latter is usually evident from the history. Since all patients with differentiated thyroid cancer after thyroidectomy take levothyroxine, thyroid function tests should be repeated after withdrawal of the medication, in order to rule out iatrogenic thyrotoxicosis. Whole body radioiodine scanning will show multiple foci of uptake in several skeletal regions.

REFERENCES

1. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, Braverman LE. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
2. Rapoport B, Chazenbalk GD, Jaume JC, McLachlan SM. The thyrotropin (TSH) receptor: interaction with TSH and autoantibodies. *Endocr Rev*, 1998; **19**: 673–716.
3. Krohn K, Fuhrer D, Bayer Y, *et al*. Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev*, 2005; **26**: 504–24.
4. Aghini LF, Fiore E, Tonacchera M, *et al*. The effect of voluntary iodine prophylaxis in a small rural community: the Pescopagano survey 15 years later. *J Clin Endocrinol Metab*, 2013; **98**: 1031–9.
5. Tonacchera M, Agretti P, Chiovato L, *et al*. Activating thyrotropin receptor mutations are present in nonadenomatous hyperfunctioning nodules of toxic or autonomous multinodular goiter. *J Clin Endocrinol Metab*, 2000; **85**: 2270–4.
6. Goodwin TM, Montoro M, Mestman JH, Pekary AE, Hershman JM. The role of chorionic gonadotropin in transient hyperthyroidism of hyperemesis gravidarum. *J Clin Endocrinol Metab*, 1992; **75**: 1333–7.
7. Rodien P, Bremont C, Sanson ML, *et al*. Familial gestational hyperthyroidism caused by a mutant thyrotropin receptor

hypersensitive to human chorionic gonadotropin. *N Engl J Med*, 1998; **339**: 1823–6.
8. Coulon AL, Savagner F, Briet C, *et al*. Prolonged and severe gestational thyrotoxicosis due to enhanced hCG sensitivity of a mutant thyrotropin receptor. *J Clin Endocrinol Metab*, 2016; **101**: 10–11.
9. Kopp P, Van Sande J, Parma J, *et al*. Brief report: congenital hyperthyroidism caused by a mutation in the thyrotropin-receptor gene. *N Engl J Med*, 1995; **332**: 150–4.
10. Duprez L, Parma J, Van Sande J, *et al*. Germline mutations in the thyrotropin receptor gene cause non-autoimmune autosomal dominant hyperthyroidism. *Nat Genet*, 1994; **7**: 396–401.
11. Martino E, Bartalena L, Bogazzi F, Braverman LE. The effects of amiodarone on the thyroid. *Endocr Rev*, 2001; **22**: 240–54.
12. Nair KC, Haamann F, Nienhaus A. Frequency of thyroid dysfunctions during interferon alpha treatment of single and combination therapy in hepatitis C virus-infected patients: a systematic review based analysis. *PLoS One*, 2013; **8**: e55364.
13. Abelleira E, De Cross GA, Pitoia F. [Thyroid dysfunction in adults infected by human immunodeficiency virus]. *Medicina (B Aires)*, 2014; **74**: 315–20.
14. Krouse RS, Royal RE, Heywood G, *et al*. Thyroid dysfunction in 281 patients with metastatic melanoma or renal carcinoma treated with interleukin-2 alone. *J Immunother Emphasis Tumor Immunol*, 1995; **18**: 272–8.
15. Ghori F, Polder KD, Pinter-Brown LC, *et al*. Thyrotoxicosis after denileukin diftitox therapy in patients with mycosis fungoides. *J Clin Endocrinol Metab*, 2006; **91**: 2205–8.
16. Figaro MK, Clayton W, Jr., Usoh C, *et al*. Thyroid abnormalities in patients treated with lenalidomide for hematological malignancies: results of a retrospective case review. *Am J Hematol*, 2011; **86**: 467–70.
17. Cossburn M, Pace AA, Jones J, *et al*. Autoimmune disease after alemtuzumab treatment for multiple sclerosis in a multicenter cohort. *Neurology*, 2011; **77**: 573–9.
18. Torino F, Barnabei A, Paragliola R, Baldelli R, Appetecchia M, Corsello SM. Thyroid dysfunction as an unintended side effect of anticancer drugs. *Thyroid*, 2013; **23**: 1345–66.
19. Costa R, Carneiro BA, Agulnik M, *et al*. Toxicity profile of approved anti-PD-1 monoclonal antibodies in solid tumors: a systematic review and meta-analysis of randomized clinical trials. *Oncotarget*, 2017; **8**: 8910–20.
20. Hedberg CW, Fishbein DB, Janssen RS, *et al*. An outbreak of thyrotoxicosis caused by the consumption of bovine thyroid gland in ground beef. *N Engl J Med*, 1987; **316**: 993–8.
21. Spencer CA, LoPresti JS, Patel A, *et al*. Applications of a new chemiluminometric thyrotropin assay to subnormal measurement. *J Clin Endocrinol Metab*, 1990; **70**: 453–60.
22. Baloch Z, Carayon P, Conte-Devolx B, *et al*. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid*, 2003; **13**: 3–126.
23. Refetoff S. Inherited thyroxine-binding globulin abnormalities in man. *Endocr Rev*, 1989; **10**: 275–93.
24. Ekins R. Measurement of free hormones in blood. *Endocr Rev*, 1990; **11**: 5–46.
25. Singer PA, Cooper DS, Levy EG, *et al*. Treatment guidelines for patients with hyperthyroidism and hypothyroidism. Standards of Care Committee, American Thyroid Association. *JAMA*, 1995; **273**: 808–12.
26. Vitti P, Elisei R, Tonacchera M, *et al*. Detection of thyroid-stimulating antibody using Chinese hamster ovary cells transfected with cloned human thyrotropin receptor. *J Clin Endocrinol Metab*, 1993; **76**: 499–503.
27. Costagliola S, Morgenthaler NG, Hoermann R, *et al*. Second generation assay for thyrotropin receptor antibodies has superior diagnostic sensitivity for Graves' disease. *J Clin Endocrinol Metab*, 1999; **84**: 90–7.
28. Latrofa F, Chazenbalk GD, Pichurin P, Chen CR, McLachlan SM, Rapoport B. Affinity-enrichment of thyrotropin receptor autoantibodies from Graves' patients and normal individuals provides insight into their properties and possible origin from natural antibodies. *J Clin Endocrinol Metab*, 2004; **89**: 4734–45.
29. Vitti P, Rago T, Chiovato L, *et al*. Clinical features of patients with Graves' disease undergoing remission after antithyroid drug treatment. *Thyroid*, 1997; **7**: 369–75.
30. Schwarz-Lauer L, Chazenbalk GD, McLachlan SM, Ochi Y, Nagayama Y, Rapoport B. Evidence for a simplified view of autoantibody interactions with the thyrotropin receptor. *Thyroid*, 2002; **12**: 115–20.
31. Giuliani C, Cerrone D, Harii N, *et al*. A TSHr-LH/CGr chimera that measures functional TSAb in Graves' disease. *J Clin Endocrinol Metab*, 2012; **97**: E1106–15.
32. Li Y, Kim J, Diana T, Klasen R, Olivo PD, Kahaly GJ. A novel bioassay for anti-thyrotrophin receptor autoantibodies detects both thyroid-blocking and stimulating activity. *Clin Exp Immunol*, 2013; **173**: 390–7.
33. Mariotti S, Sansoni P, Barbesino G, *et al*. Thyroid and other organ-specific autoantibodies in healthy centenarians. *Lancet*, 1992; **339**: 1506–8.
34. Latrofa F, Pichurin P, Guo J, Rapoport B, McLachlan SM. Thyroglobulin-thyroperoxidase autoantibodies are polyreactive, not bispecific: analysis using human monoclonal autoantibodies. *J Clin Endocrinol Metab*, 2003; **88**: 371–8.
35. Chiovato L, Latrofa F, Braverman LE, *et al*. Disappearance of humoral thyroid autoimmunity after complete removal of thyroid antigens. *Ann Intern Med*, 2003; **139**: 346–51.
36. Marino M, Chiovato L, Friedlander JA, Latrofa F, Pinchera A, McCluskey RT. Serum antibodies against megalin (GP330) in patients with autoimmune thyroiditis. *J Clin Endocrinol Metab*, 1999; **84**: 2468–74.
37. Vitti P, Rago T, Mancusi F, *et al*. Thyroid hypoechogenic pattern at ultrasonography as a tool for predicting recurrence of hyperthyroidism after medical treatment in patients with Graves' disease. *Acta Endocrinol (Copenh)*, 1992; **126**: 128–31.
38. Vitti P, Rago T, Mazzeo S, *et al*. Thyroid blood flow evaluation by color-flow Doppler sonography distinguishes Graves' disease from Hashimoto's thyroiditis. *J Endocrinol Invest*, 1995; **18**: 857–61.
39. Bogazzi F, Bartalena L, Brogioni S, *et al*. Color flow Doppler sonography rapidly differentiates type I and type II amiodarone-induced thyrotoxicosis. *Thyroid*, 1997; **7**: 541–5.
40. Rago T, Santini F, Scutari M, Pinchera A, Vitti P. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. *J Clin Endocrinol Metab*, 2007; **92**: 2917–22.
41. Zweig MH, Csako G, Spero M. Escape from blockade of interfering heterophile antibodies in a two-site immunoradiometric assay for thyrotropin. *Clin Chem*, 1988; **34**: 2589–91.
42. Favresse J, Burlacu MC, Maiter D, Gruson D. Interferences with thyroid function immunoassays: clinical implications and detection algorithm. *Endocr Rev*, 2018; **39**: 830–50.
43. Refetoff S, Weiss RE, Usala SJ. The syndromes of resistance to thyroid hormone. *Endocr Rev*, 1993; **14**: 348–99.

44. Bochukova E, Schoenmakers N, Agostini M, *et al*. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*, 2012; **366**: 243–9.
45. Beck-Peccoz P, Brucker-Davis F, Persani L, Smallridge RC, Weintraub BD Thyrotropin-secreting pituitary tumors. *Endocr Rev*, 1996; **17**: 610–38.
46. Sarne DH, Sobieszczyk S, Ain KB, Refetoff S. Serum thyrotropin and prolactin in the syndrome of generalized resistance to thyroid hormone: responses to thyrotropin-releasing hormone stimulation and short term triiodothyronine suppression. *J Clin Endocrinol Metab*, 1990; **70**: 1305–11.
47. Mariotti S, Martino E, Cupini C, et al. Low serum thyroglobulin as a clue to the diagnosis of thyrotoxicosis factitia. *N Engl J Med*, 1982; **307**: 410–12.
48. Latrofa F, Ricci D, Sisti E, et al. Significance of low levels of thyroglobulin autoantibodies associated with undetectable thyroglobulin after thyroidectomy for differentiated thyroid carcinoma. *Thyroid*, 2016; **26**: 798–806.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

3.3.6 Antithyroid Drugs for Thyrotoxicosis

Luigi Bartalena

Introduction

Hyperthyroidism due to Graves' disease can be treated either conservatively (antithyroid drugs, ATDs) or by reduction/ablation of thyroid tissue by radioiodine (RAI) or thyroidectomy [1, 2]. ATDs are the first-line treatment for newly diagnosed Graves' hyperthyroidism in Europe, Latin America, Asia, and Oceania, whereas RAI is preferred in North America [1, 3, 4] (Figure 3.3.6.1). When hyperthyroidism is caused by toxic adenoma or multinodular goitre, ATDs are only a tool to control hyperthyroidism prior to definitive treatment, because ATD discontinuation is inevitably followed by a recurrence of hyperthyroidism [5].

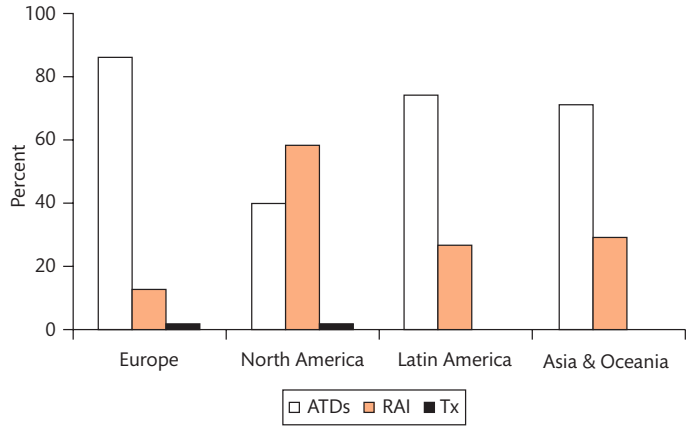


Figure 3.3.6.1 First-line treatment for newly diagnosed hyperthyroidism due to Graves' disease in different continents. Data are derived from Burch *et al*. [3] and Bartalena *et al*. [4].

ATDs belong to the family of thionamides and include propylthiouracil (PTU), carbimazole (CBZ), and its active metabolite, methimazole (MMI). Carbimazole is mainly used in the United Kingdom, MMI in the other European countries and Asia, while PTU has traditionally been the most commonly used ATD in North and Latin America [1]. However, recent data suggest that MMI has become the more common ATD in North America as well [6, 7].

Pharmacological Characteristics

ATDs are oral drugs completely absorbed from the gastrointestinal tract. Because of its longer half-life, MMI is given as a single daily dose, whereas PTU often requires fractionated doses. CBZ is rapidly converted to MMI. Therapeutic potency of MMI is 10–50-fold higher than that of PTU [2] (Table 3.3.6.1). PTU is not available in several countries.

Table 3.3.6.1 Pharmacological features of methimazole (MMI) and propylthiouracil (PTU)

	MMI	PTU*
Relative potency	10–50	1
Route of administration	oral	oral
Absorption from gastrointestinal (GI) tract	complete	complete
Half-life in serum (h)	6–8	1–2
Duration of action (h)	>24	8–12
Transplacental transfer	yes	yes
Passage into milk	yes	yes
Excretion	renal	renal
Administration schedule	once daily	2–3 times daily
Cost	low	moderate
Mechanism of action	TPO inhibition	TPO inhibition
		Type 1 deiodinase inhibition

*not available in several countries; TPO, thyroid peroxidase.

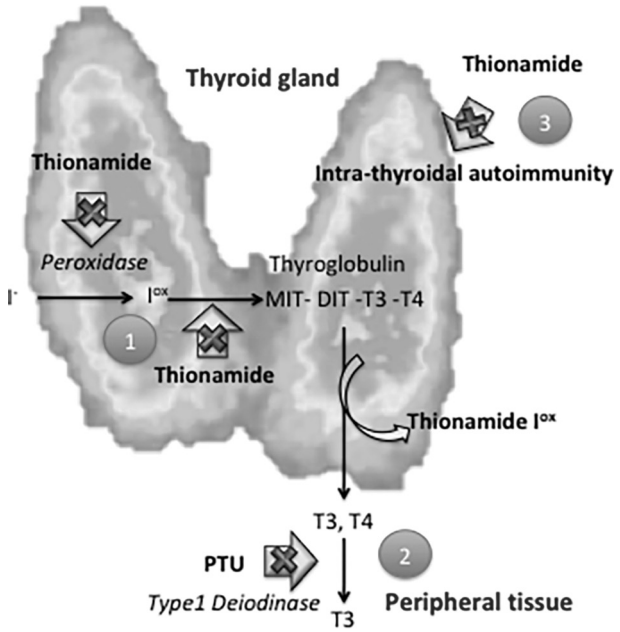


Figure 3.3.6.2 Mechanism of action of thionamides. The main action (**1**) is inhibition of thyroid peroxidase, leading to inhibition of thyroid hormone synthesis. In addition, propylthiouracil (PTU) (**2**) inhibits type 1 deiodinase activity, causing a decrease in the peripheral conversion of thyroxine (T4) to the metabolically active hormone, triiodothyronine (T_3). Finally, thionamides exert some immunosuppressive actions (**3**) that might be relevant.

ATDs mainly act by blocking thyroid hormone synthesis through inhibition of thyroid peroxidase (Figure 3.3.6.2). In addition, PTU inhibits type 1 deiodinase, the enzyme that peripherally converts thyroxine (T_4) to the metabolically active hormone, triiodothyronine (T_3). ATDs also exerts some immunosuppressive effects that may contribute to their therapeutic action in Graves' hyperthyroidism [8]. Whether the latter are direct or mainly indirect (i.e. bound to normalization of thyroid status) is still unsettled.

Adverse Events

In general, ATD treatments is safe. Side effects of thionamides are usually mild and self-limiting; major adverse events occur in less than 5% of cases, more frequently, particularly with MMI, during the initial phases of treatment, when high doses are often employed [9] (Table 3.3.6.2). If one thionamide causes an adverse event, the other thionamide can substitute, but cross-reactivity is frequent (about 50%) [10]. Accordingly, this strategy should not in principle be recommended.

Frequent side effects, such as itching or skin reactions, do not need ATD withdrawal, are usually transient, and can be controlled by antihistamines.

Agranulocytosis (neutrophil count <500 per mm^3) is one of the most feared adverse events of thionamides, with a reported incidence of about 0.3–0.6% [9], although a recent report suggested it may be as high as 1.2% [11]. Immediate drug discontinuation, prompt antibacterial treatment, and supportive measures are imperative; the efficacy of granulocyte colony-stimulating factor in accelerating recovery has recently been questioned [11]. There seems to be a dose-dependency of this serious complication in the case of MMI [12], which usually occurs within the first 3 months of treatment, while this relation is lacking for PTU [12]. Agranulocytosis seems to be more common in older people [13]. Recent studies have shown that genetic variants of NOX3 [14], HLA-B*27:05 [15], HLA-B*38:02, and HLA-DRB1*08:02 [16] confer increased susceptibility to ATD-induced agranulocytosis. Onset of agranulocytosis is generally abrupt. Accordingly, the usefulness of regular monitoring of white blood cell count is questionable, but patients should be advised that in case of fever, sore throat, or other signs of infections, a leukocyte count should be promptly obtained [17].

Aplasia cutis, the congenital (localized or disseminated) absence of skin, has been rarely reported in infants born to hyperthyroid mothers receiving MMI during pregnancy, but not in those treated with PTU [9]. Very rare cases of **embryopathy** (particularly,

Table 3.3.6.2 Adverse events of antithyroid drugs

	Adverse event	Frequency
Blood		
	Mild leucopoenia	Relatively frequent*
	Agranulocytosis	Uncommon (0.3–0.6%)
	Aplastic anaemia	Very rare
	Thrombocytopenia	Very rare
	Pancytopenia	Very rare
Skin		
	Skin rash	Very common*
	Urticaria	Very common*
	Itching	Very common*
	Generalized dermatitis	Very rare
	Alopecia	Very rare
Liver		
	Hepatocellular necrosis	Rare (PTU)
	Cholestasis	Very rare (MMI)
Collagen		
	Arthralgias	Common
	SLE-like syndrome	Very rare (PTU>MMI)
	Vasculitis	Very rare (PTU)
Embryopathy		
	Choanal atresia, oesophageal atresia, aplasia cutis	Very rare (MMI)
	Situs inversus ± dextrocardia, unilateral kidney a/dysgenesis, cardiac outflow tract defect	Very rare (PTU)
	Unilateral kidney a/dysgenesis, cardiac outflow tract defect	Very rare (PTU)
Miscellaneous		
	Loss of taste or smell	Rare (MMI)
	Hypothrombinaemia	Rare (PTU)
	Insulin autoantibodies	Very rare
	Pancreatitis	Very rare

choanal, and oesophageal atresia) were reported in infants whose mothers were given CBZ/MMI in early pregnancy [18]. It should be noted that although usually rarer and usually milder, cases of malformations, involving face and neck and urinary tract, have been observed also in the offspring of mothers treated with PTU [19] (see later section in this chapter on pregnancy and lactation).

Because PTU can be associated, although rarely, with severe **hepatoxicity** potentially lethal or requiring liver transplant, MMI rather than PTU should be prescribed as first-line ATD both in adults and children [8, 17, 20]. Hepatic abnormalities, with cholestatic features (but occasionally with necrotic features as well), may rarely occur also in patients treated with MMI [21], but they are usually milder and reversible after MMI withdrawal [21].

A rare but severe adverse event of ATDs is **vasculitis,** more common in PTU-treated patients [21]. Antineutrophil cytoplasmic antibodies (ANCA) are more frequently found in treated Graves' patients than in untreated patients or controls [22]. Vasculitis may manifest with arthritis, vasculitic rash, respiratory symptoms, or acute renal dysfunction. This rare complication develops early during MMI treatment, while the risk of vasculitis increases with time in PTU-treated patients [23]. Vasculitis requires prompt thionamide withdrawal; treatment with glucocorticoids or other immunosuppressive drugs may be necessary.

Because of potential thionamide adverse events, patients should be carefully informed and advised to get in touch promptly with their caregiver [17].

Dose Regimens

ATDs can be administered with two different regimens: the **Titration method,** in which the lowest thionamide dose maintaining euthyroidism is used, and the **Block-and-Replace method,** in which high doses of thionamides are persistently given combined with levothyroxine replacement to avoid hypothyroidism [24].

In the Titration method, treatment is usually initiated with 15–40 mg/day of MMI (in a single dose or divided doses) or 300–400 mg of PTU (always in divided doses). The choice of the initial daily dose depends on the severity of hyperthyroidism [17]. Thyroid function should be periodically assessed by measuring serum-free T_4 and T_3 concentrations, while serum thyroid-stimulating hormone (TSH) remains suppressed for several weeks after restoration of euthyroidism. The daily thionamide dose is tapered down to the lowest effective dose, generally 2.5–5 mg of MMI or 50–100 mg of PTU. Normalization of thyroid function usually occurs within 4–12 weeks. Evaluation of thyroid status should be made every 4–6 weeks for the first 4–6 months, and then every 3–4 months.

In the Block-and-Replace method, high doses of thionamides are administered (with no adjustment) together with levothyroxine. The rationale is that high doses of ATDs might have a greater immunosuppressive effect abating the autoimmune process and leading to permanent remission, but this putative effect remains to be demonstrated. With the Block-and-Replace method avoidance of hypothyroidism seems easier than with the Titration method, and the number of visits lower [24]. On the other hand, the prolonged use of higher doses of thionamides (e.g. 30–40 mg/day of MMI) might expose the patient to a potentially higher risk of adverse events [24]. This conclusion might not be tenable any longer in view of the lower than in the past doses of ATDs used also in the Block-and-Replace method [25]. The high number of daily tablets (unless high-dose tablets are available) may, however, decrease compliance with therapy. The Block-and-Replace regimen does not offer any advantage in terms of permanent remission of hyperthyroidism, while it may bear a slightly higher risk of side effects [24].

Duration of Treatment

Using the Titration method, a 18-month treatment is associated with a significantly lower relapse rate than a 6-month treatment [24]. There seems to be little advantage to prolong treatment up to 24 or 42 months [24], although the use of high doses of thionamides may result in a longer relapse-free period following thionamide withdrawal [26]. Using the Block-and-Replace regimen, extension of ATD treatment from 6 months to 12 months does not enhance the chance of a permanent remission of hyperthyroidism [27]. This view is not shared by some authors who have advocated the efficacy of very long treatment with low doses of thionamides [28].

Outcome of Treatment

The major limitation of ATD treatment is the high relapse rate after drug withdrawal. As many as 30–70% of patients experience a relapse of hyperthyroidism [9], in the majority of cases within one year after ATD discontinuation [29]. Hyperthyroidism may however relapse even years after ATD withdrawal [25]. A substantial proportion of thionamide-treated patients become spontaneously hypothyroid with time [30], indicating that a lifelong follow-up is needed after ATD treatment.

There is no reliable single predictor of ATD treatment outcome. However, patients with large goitres that do not shrink, severe hyperthyroidism difficult to control even with high ATD doses, detectable TRAb levels at the end of the treatment course, are indeed very good candidates to have a relapse of hyperthyroidism [31] (Table 3.3.6.3). Failure of TRAb to become undetectable almost invariably predicts relapse of hyperthyroidism following thionamide withdrawal, but also patients whose TRAb levels become undetectable during

Table 3.3.6.3 Risk factors for relapse of hyperthyroidism after antithyroid drug treatment

Factor	Impact on relapse
Large goitre	Strong
Age <40 years	Likely (controversial)
Male gender	Uncertain
Difficult control of hyperthyroidism with high ATD doses	Strong
Smoking	Strong
Detectable TRAb at the end of the ATD course*	Strong
Postpartum period	Strong

TRAb, TSH-receptor antibody. *Relapses are frequent also in patients whose TRAb tests are negative at the end of ATD treatment.

treatment have a relatively high chance of experiencing a relapse of Graves' hyperthyroidism [31]. Persistence of high TRAb levels during ATD treatment might represent an indication to switch to a definitive treatment of hyperthyroidism by RAI or thyroidectomy even before completion of the ATD course [32].

Men have a lower remission rate than women, while young age (less than 40 years) seems to be associated with a higher relapse rate [31], possibly because hyperthyroidism tends to be less severe in older people. Smoking might reduce the likelihood of a permanent remission after ATD treatment [33] (Table 3.3.6.3).

Recently, a predictive score (GREAT, Graves' Recurrent Events After Therapy), calculated as the sum of scores given to four clinical markers (age, serum-free thyroxine levels, serum TRAb levels, goitre size by palpation) assessed at diagnosis, has been developed [34] and has received an initial external validation as a good predictor of relapse after a course of ATDs [35–37]. Another (clinical severity score, CSS), based on the Merseburg triad (hyperthyroidism, goitre, and orbitopathy) has been shown to predict control of hyperthyroidism at 6 and 12 months of ATD treatment, but no information is insofar available on its predictivity of recurrences after ATD withdrawal. Predictive scores, by grouping single risk factors at diagnosis, might contribute to selecting the most appropriate treatment (conservative vs. ablative) in an individual patient at diagnosis

While in children and adolescents a second course of ATD treatment may be proposed if hyperthyroidism recurs, in adults relapse of hyperthyroidism after ATD therapy commonly is an indication to switch to definitive therapy (RAI or thyroidectomy), although a shared decision-making process is advised, keeping patient choice in due consideration.

ATD Treatment During Pregnancy and Lactation

Untreated hyperthyroidism is risky for the mother (pre-eclampsia, congestive heart failure), the fetus (accelerated bone maturation, growth retardation, prematurity, stillbirth, small risk of malformations), and pregnancy outcome (miscarriage, preterm delivery) [38]. Thus, restoration and stable maintenance of euthyroidism are warranted, although, in general, Graves' hyperthyroidism tends spontaneously to improve during pregnancy [39]. Conversely, Graves' hyperthyroidism may develop or be exacerbated in the postpartum period [40], although the relapse associated with the postpartum period appears to be more prone to undergo permanent remission and should probably be treated conservatively [41]. ATDs are the treatment of choice for Graves' pregnant women, because RAI therapy is contraindicated in pregnancy (as well as during lactation), and surgery is only exceptionally indicated, in this case in the second trimester [9, 17]. As mentioned in the paragraph on adverse events of thionamides, rare cases of aplasia cutis and 'CBZ/MMI embryopathy' have been reported in the offspring of mothers given MMI [18]. However, milder and less frequent malformations were also associated with prenatal exposure to PTU [19]. Thus, if PTU may still be preferred, for its lower risk, during the first trimester of pregnancy when organogenesis occurs, the current recommendation, in view of the possible severe PTU-related hepatoxicity, is that the pregnant woman be switched to MMI after the sixteenth week of pregnancy [17, 42]. In women whose preconception hyperthyroidism is controlled with very low doses of thionamides, an attempt might be made to stop treatment during pregnancy [19]. The goal of ATD treatment during pregnancy is to use the lowest dose of thionamides sufficient to maintain free thyroid hormones in the high-normal range, avoiding the risk of fetal hypothyroidism and goitre on one hand, and the risk that even mild maternal hypothyroidism may affect subsequent neuropsychological development in the offspring [17].

The use of thionamides during lactation appears to be safe, and doses of MMI up to 10 mg/day do not affect newborns' thyroid function. PTU is also safe for breastfeeding. In view of thionamide pharmacokinetics, it seems wise to suggest taking thionamide pill shortly after breastfeeding.

ATD Treatment of Hyperthyroidism in Children and Adolescents

Graves' disease is the most common cause of hyperthyroidism also in children and adolescents, and ATD treatment is the first-line treatment in the majority of them [20]. Remission rate after ATD treatment is lower in children than in adults. Thus, many if not most children/adolescents eventually require either RAI therapy or thyroidectomy [20]. Sometimes switching to definitive treatments is mandatory because of thionamide adverse events, which are more common in young patients. Although MMI may be associated also in children with a low but real risk of side effects, the risk is much higher using PTU [20]. In particular, PTU-induced liver failure may develop in 1:2000–1:4000 children treated with this drug [43], and reversible hepatotoxicity is likely much more frequent [43]. For these reasons, only MMI should be used in children/adolescents.

Neonatal Hyperthyroidism

Neonatal hyperthyroidism is due to the transplacental passage of maternal TSH-receptor autoantibodies. It is therefore a transient condition, which however requires a vigorous, although short-term, treatment with thionamides, in association with propranolol (or other β-blockers) and iodine [44].

ATD Treatment Before, During, and After RAI Therapy

Whether ATD pretreatment may decrease the effectiveness of subsequent RAI therapy has been long discussed. Two prospective, randomized studies demonstrated that MMI pretreatment does not decrease the effectiveness of RAI [45]. On the other hand, PTU pretreatment can cause a reduced cure rate [46]. Thus, it seems that only PTU pretreatment may cause a decrease in the efficacy of RAI, probably due to its longer radioprotective effect. This problem can however be overcome using higher therapeutic doses of RAI [47]. Thionamide pretreatment has the obvious beneficial effect of controlling hyperthyroidism: this is warranted in older people and in patients with relevant comorbidities [48]. ATDs should be withdrawn 5–7 days prior to RAI therapy. A short course of lithium carbonate for 2 weeks after RAI administration prevents effectively any

increase in serum thyroid hormone levels due either to ATD discontinuation or RAI cytotoxicity and is associated with a prompter control of hyperthyroidism and goitre shrinkage after RAI therapy [49]. There is convincing evidence that continuation of ATD treatment during RAI therapy or its reinstitution shortly after [50] is bound to a higher rate of treatment failure. Thus, ATD treatment should be restarted only in at-risk patients (old patients with associated comorbidities) and at least 7–10 days after RAI therapy.

ATD Treatment Before Thyroidectomy

Control of hyperthyroidism prior to thyroidectomy is mandatory. ATDs must be administered until euthyroidism is restored, and iodide (Lugol's solution or saturated solution of potassium iodide) or inorganic iodine should be added for 1–2 weeks prior to thyroid surgery. In case of urgent and emergent surgery that cannot be postponed, a combination of radiographic contrast agents, β-blockers and glucocorticoids should be added, because of the delayed effect of thionamides [51].

REFERENCES

1. Bartalena L. Diagnosis and management of Graves' disease: a global overview. *Nature Rev Endocrinol*, 2013; **9**: 724–34.
2. Smith TJ, Hagedus L. Graves' disease. *N Engl J Med*, 2016; **375**: 1552–65.
3. Burch HB, Burman KD, Cooper DS. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab*, 2012; **97**: 4549–58.
4. Bartalena L, Burch HB, Burman KD, Kahaly GJ. A 2013 European survey of clinical practice patterns in the management of Graves' disease. *Clin Endocrinol (Oxf)*, 2016; **84**: 115–20.
5. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet*, 2016; **388**: 906–18.
6. Emiliano AB, Governale L, Parks M, Cooper DS. Shifts in propylthiouracil and methimazole prescribing practices: antithyroid drug use in the United States from 1991 to 2008. *J Clin Endocrinol Metab*, 2010; **95**: 2227–33.
7. Brito JP, Schilz S, Singh Ospina N, *et al*. Antithyroid drugs—the most common treatment for Graves' disease in the United States: a nationwide population-based study. *Thyroid*, 2016, **26**: 1144–5.
8. Weetman AP. How antithyroid drugs works in Graves' disease. *Clin Endocrinol (Oxf)*, 1992; **37**: 317–18.
9. Bartalena L, Chiovato L, Vitti P. Management of hyperthyroidism due to Graves' disease: frequently asked questions and answers (if any). *J Endocrinol Invest*, 2016; **39**: 1105–14.
10. Ahmed K, Rao S, Simha V. ANCA-positive vasculitis in a patient with Graves' disease: cross-reaction between propylthiouracil and methimazole. *Endocr Pract*, 2010; **9**: 1–11.
11. Yang J, Zhu Y-J, Zhong J-J, *et al*. Characteristics of antithyroid-drug-induced agranulocytosis in patients with hyperthyroidism: a retrospective analysis of 114 cases in a single institution in China involving 9690 patients referred for radiopiodine treatment over 15 years. *Thyroid*, 2016; **26**: 627–33.
12. Takata K, Kubota S, Fukata S, *et al*. Methimazole-induced agranulocytosis in patients with Graves' disease is more frequent with an initial dose of 30 mg daily than with 15 mg daily. *Thyroid*, 2009; **19**: 559–63.
13. Watanabe N, Narimatsu H, Yoshimura Noh J, *et al*. Antithyroid drug-induced hematopoietic damage: a retrospective cohort study of agranulocytosis and pancytopenia involving 50,385 patients with Graves' disease. *J Clin Endocrinol Metab*, 2012; **97**: E49–53.
14. Plantinga TS, Arts P, Knarren GH, *et al*. Rare NOX3 variants confer susceptibility to agranulocytosis during thyrostatic treatment of Graves' disease. *Clin Pharmacol Ther*, 2017; **102**: 1017–24.
15. Hallberg P, Eriksson N, Ibanez L, *et al*. Genetic variants associated with antithyroid drug-induced agranulocytosis: a genome-wide association study in a European population. *Lancet Diabetes Endocrinol*, 2016; **4**: 507–16.
16. Chen PL, Shih SR, Wang P-W, *et al*. Genetic determinants of antithyroid drug-induced agranulocytosis by human leukocyte antigen genotyping and genome-wide association study. *Nature Commun*, 2015; **6**: 7633.
17. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*, 2016; **26**: 1343–416.
18. Bowman P, Osborne NJ, Sturley R, Vaidya B. Carbimazole embryopathy: implications for the choice of antithyroid drugs in pregnancy. *Q J Med*, 2012; **105**: 189–93.
19. Laurberg P, Linding Andersen S. Antithyroid drug use in early pregnancy and birth defects: time windows of relative safety and high risk? *Eur J Endocrinol*, 2014; **171**: R13–20.
20. Rivkees SA. Controversies in the management of Graves' disease in children. *J Endocrinol Invest*, 2016; **39**: 1247–57.
21. Cooper DS. Antithyroid drugs. *N Engl J Med*, 2005; **352**: 905–17.
22. Harper L, Chin L, Daykin J, *et al*. Propylthiouracil and carbimazole-associated antineutrophil cytoplasmic antibodies (ANCA) in patients with Graves' disease. *Clin Endocrinol (Oxf)*, 2004; **60**: 671–5.
23. Cin MO, Gursoy A, Morris Y, Aydintug OT, Kamel N, Gullu S. Prevalence and clinical significance of antineutrophil cytoplasmic antibody in Graves' patients treated with propylthiouracil. *Int J Clin Pract*, 2009; **63**: 299–302.
24. Abraham P, Avenell A, McGeosch SC, Clark LF, Bevan JS. Antithyroid drug regimen for treating Graves' hyperthyroidism. *Cochrane Database Syst Rev*, 2010; **1**: CD003420.
25. Razvi S, Vaidya B, Perros P, Pearce SHS. What is the evidence behind the evidence-base? The premature death of block-replace antithyroid drug regimens for Graves' disease. *Eur J Endocrinol*, 2006; **154**: 783–6.
26. Grebe SKG, Feek CM, Ford HC, *et al*. A randomized trial of short-term treatment of Graves' disease with high-dose carbimazole plus thyroxine versus low-dose carbimazole. *Clin. Endocrinol (Oxf)*, 1998; **48**: 585–92.
27. Weetman AP, Pickerill AP, Watson P, Chatterjee VK, Edwards OM. Treatment of Graves' disease with the block-replace regimen of antithyroid drugs: the effect of treatment duration and immunogenetic susceptibility on relapse. *Q J Med* 1994; **87**: 337–41.
28. Azizi F, Malboosbaf R. Long-term antithyroid drug treatment: a systematic review and meta-analysis. *Thyroid*, 2017; **27**: 1223–31.
29. Vitti P, Rago T, Chiovato L, *et al*. Clinical features of patients with Graves' disease undergoing remission after antithyroid drug treatment. *Thyroid*, 1997; **7**: 369–75.
30. Wood LC, Ingbar SH. Hypothyroidism as a late sequel in patients with Graves' disease treated with antithyroid agents. *J Clin Invest*, 1979; **64**: 1429–36.

31. Piantanida E, Lai A, Sassi L, *et al*. Outcome prediction of treatment of Graves' hyperthyroidism with antithyroid drugs. *Horm Metab Res*, 2015: **47**: 767–72.
32. Bartalena L, Marcocci C, Lai A, Tanda ML. Graves' hyperthyroidism of recent onset and Graves' orbitopathy: to ablate or not to ablate the thyroid? *J Endocrinol Invest*, 2008; **31**: 578–81.
33. Kimball LE, Kulinskaya E, Brown B, Johnston C, Farid NR. Does smoking increase relapse rates in Graves' disease? *J Endocrinol Invest*, 2002; **25**: 152–7.
34. Vos XG, Endert E, Zwinderman AH, Tijssen JGP, Wiersinga WM. Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with Graves' hyperthyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 1381–9.
35. Struja T, Kaeslin M, Boesiger F, *et al*. External validation of the GREAT score to predict relapse risk in Graves' disease: results from a multicentre, retrospective study with 741 patients. *Eur J Endocrinol*, 2017; **176**: 413–19.
36. Struja T, Fehlberg H, Kutz A, *et al*. Can we predict relapse in Graves' disease? Results from a systematic review and meta-analysis. *Eur J Endocrinol*, 2017; **176**: 87–97.
37. Bartalena L, Masiello E, Magri F, *et al*. The phenotype of newly diagnosed Graves' disease in Italy in recent years is milder than in the past: results of a large observational longitudinal study. *J Endocrinol Invest*, 2016; **39**: 1445–51.
38. Laurberg P, Linding Andersen S. Pregnancy and the incidence, diagnosing and therapy of Graves' disease. *Eur J Endocrinol*, 2016; **175**: R291–30.
39. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
40. Rotondi M, Cappelli C, Pirali B, *et al*. The effect of pregnancy on subsequent relapse from Graves' disease after a successful course of antithyroid drug therapy. *J Clin Endocrinol Metab*, 2008; **93**: 3985–8.
41. Rotondi M, Capelli V, Coperchini F, *et al*. Post-partum and non-post-partum relapsing Graves' hyperthyroidism display different response to anti-thyroid drugs. *Eur J Endocrinol*, 2018; **178**: 589–94.
42. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
43. Rivkees SA, Mattison DR. Ending propylthiouracil-induced liver failure in children. *N Engl J Med*, 2009; **360**: 1574–5 (letter).
44. Samuels SL, Namoc SM, Bauer AJ. Neonatal thyrotoxicosis. *Clin Perinatol*, 2018: **45**: 31–40.
45. Braga M, Walpert N, Burch HB, Solomon BL, Cooper DS. The effect of methimazole on cure rates after radioiodine treatment for Graves' hyperthyroidism: a randomized clinical trial. *Thyroid*, 2002; **12**: 135–9.
46. Bonnema SJ, Bennedbaek FN, Veje A, Marving J, Hegedus L. Propylthiouracil before ^{131}I therapy of hyperthyroid diseases: effect on cure rate evaluated by a randomized clinical trial. *J Clin Endocrinol Metab*, 2004; **89**: 4439–44.
47. Alexander EK, Larsen PR. High dose of ^{131}I for the treatment of hyperthyroidism caused by Graves' disease. *J Clin Endocrinol Metab*, 2002; **87**: 1073–7.
48. Bartalena L, Bogazzi F, Pinchera A, Martino E. Treatment with thionamides before radioiodine therapy for hyperthyroidism: yes or no? *J Clin Endocrinol Metab*, 2005; **90**: 256.
49. Bogazzi F, Giovannetti C, Fessehatsion R, *et al*. Impact of lithium on radioactive iodine efficacy for Graves' disease: a cohort study on cure rate, time to cure, and frequency of increased serum thyroxine after antithyroid drug withdrawal. *J Clin Endocrinol Metab*, 2010; **95**: 201–8.
50. Sabri O, Zimny M, Schulz G, *et al*. Success rate of radioiodine therapy in Graves' disease: the influence of thyrostatic medication. *J Clin Endocrinol Metab*, 1999; **84**: 1229–33.
51. Panzer C, Beazley R, Braverman L. Rapid preoperative preparation for severe hyperthyroid Graves' disease. *J Clin Endocrinol Metab*, 2004; **89**: 2142–4.

3.3.7 Radioiodine Treatment of Hyperthyroidism

Markus Luster and Michael Lassmann

Introduction

For more than 70 years radioactive iodine has been used successfully in the treatment of thyroid diseases in millions of patients. After the first therapy took place at the Massachusetts General Hospital in Boston in 1941 it was however, not until after the Second World War that I-131 became generally available for clinical applications [1].

The radioactive iodine isotope is chemically identical to 'stable' iodine (I-127) and thus becomes a part of the intrathyroidal organification and metabolism. Its principle of action is based on the emission of β-rays with a range of 0.5–2 mm in the tissue leading to high local radiation absorbed doses while sparing surrounding structures. The additional gamma-ray component of I-131 allows for scintigraphic imaging of the distribution in the gland and can also be used for pre- and post-therapeutic individual dosimetry (see next).

Several therapeutic options are available for the treatment of benign thyroid disorders, namely hyperthyroidism: surgical resection (hemithyroidectomy, near-total or total thyroidectomy), long-term antithyroid drug medication (ATD), radioiodine therapy (RAIT) [2–4] and more recently locally ablative procedures [5]. These different treatment modalities are used in varying frequencies depending on geographical location (e.g. iodine supply, availability and logistics, 'cultural' background, local experience), and patient specific features like goitre size, presence of local symptoms, age, and hormonal status.

The diversity of approaches on an international scale still remains impressive and is reflected by a great heterogeneity throughout Europe and also as compared to the United States where RAIT is still being applied more frequently than in most European countries [4, 6–9].

Radioiodine therapy was originally aimed at eliminating hyperthyroidism and thus leaving the patient euthyroid. Current strategies however established postradioiodine induction of hypothyroidism as treatment objective, thus included in the category of 'cure'. This definition holds especially true for the management of Graves' disease (GD) when long-term hypothyroidism was the rule and stabilization of euthyroidism failed in the majority of cases. In fact, the term 'ablation', meaning removal or destruction, has been increasingly used to characterize the outcome of RAIT and administration of larger amounts of radioiodine have tended to make this a self-fulfilling prophecy. Although many clinicians prefer that the end result of treatment should be the more easily managed hypothyroidism, others are still reluctant to give up the therapeutic ideal of euthyroidism as the preferred result of radioiodine therapy and continue their efforts to solve the enigma of thyroid radiosensitivity.

Indications

The causes of hyperthyroidism include the following: (a) autoimmune hyperthyroidism called previously toxic diffuse goitre (GD); (b) toxic adenoma; (c) toxic multinodular goitre (Plummer's disease); (d) silent thyroiditis; and (e) painful subacute thyroiditis. The first three entities constitute a clear indication for radioiodine treatment, while silent thyroiditis and subacute thyroiditis for reasons of the underlying pathophysiology are not treated with radioiodine.

Recently, there has been an emerging role for I-131 in the treatment of subclinical hyperthyroidism caused by any of the three first thyroid dysfunctions. Another category that is frequently regarded as a suitable entity for this kind of treatment is *non-toxic goitre (NTG)* which encompass a group of patients who are euthyroid but may clinically benefit from a reduction of the organ volume [10–12]. The available treatment options in NTG patients in whom the risk of malignancy is considered low (i.e. was ruled out by fine needle aspiration cytology are 'wait and see' policy, surgery, thyroid hormone/iodine therapy, and radioiodine treatment). The main indications for radioiodine treatment of NTG are to reduce the size of goitre to relieve compressive signs or symptoms and secondly alleviate potential cosmetic problems for the patient. Surgery remains the fastest and most effective way to reduce goitre size and relieve any acute compressive symptoms and is mandatory if there are any doubts about malignancy, but in elderly, frail, and patients with comorbidities RAIT represents a feasible alternative [13]. Preparation with recombinant human thyroid-stimulating hormone (rhTSH) for the augmentation of the radioiodine uptake in hypofunctional areas of goitre was evaluated in prospective studies but is not formally approved for this setting [14–16].

Radioiodine is in most cases the first line treatment for GD and toxic adenoma, or it can be administered if hyperthyroidism is not controlled or recurs after antithyroid drug treatment [17]. Surgery needs to be considered if there are contraindications to radioiodine therapy.

Contraindications

Pregnancy and breast feeding are absolute contraindications to radioiodine treatment; all females of reproductive age should have a pregnancy test immediately before administration. It is recommended that women not attempt conception for 6–12 months after radioiodine treatment. I-131 is generally not indicated for patients who are incontinent of urine, whereas concomitant haemodialysis for renal failure represents no exclusion criterion and is routinely performed in experienced centres.

Technical Aspects and Response to Radioiodine

The effect of radioiodine therapy is gradual and varies substantially among individuals resulting in the necessity for repeated testing after the treatment to rule out persistent hyperthyroidism or short-term development of a hypothyroid state. After 4 to 6 weeks a follow-up visit may be sensible to evaluate the effect of the procedure. In case of pre-existing marked hyperthyroidism symptom relief should be achieved peritherapeutically by the administration of β-blocking agents and resumption of antithyroid drugs should be considered when tachycardia and/or palpitations are present. The potential influence of ATD on radioiodine kinetics and on the therapeutic outcome has led to the recommendation, that coadministration of methimazole and propylthiouracil should be avoided [18–22]. If tolerated restarting ATD should preferably be initiated approximately 1 week after the radioiodine has been administered.

Potential Side Effects

Acute Side Effects

Clinical exacerbation of hyperthyroidism after radioiodine treatment appears to be relatively uncommon and is usually of minor clinical significance. It presumably is related to radiation thyroiditis, with destruction of thyroid follicles and release of thyroglobulin and stored hormone into the circulation. There may be a transient rise in fT_4 and fT_3 levels several days following administration, and patients suffering from poorly controlled symptoms before radioiodine therapy may encounter an exacerbation of cardiac arrhythmia and heart failure. In some patients a 'thyroid storm' may develop. Intravenous infusion of antithyroid drugs, corticosteroids and β-blockers is the treatment of choice, but prophylactic measures and thorough initial work-up are crucial.

Patients with large goitres may notice transient swelling and dyspnoea in some cases and some discomfort may be associated with it. Slight irritation of the salivary gland function may be noted, but in contrast to thyroid cancer, the risk of permanent injury is negligible due to the much lower activities applied for therapy of thyrotoxicosis [4].

Hypothyroidism

The main (side-)effect of radioiodine treatment is permanent hypothyroidism. Its rates vary and incidence continues to increase over time, so that lifelong follow-up is essential. Pretreatment prediction is hardly possible using current variables; however, the appearance

is higher in GD than in toxic goitre and relatively uncommon in solitary hyperfunctioning nodules.

The most prominent radiobiological factor for the determination of overall outcome besides radiation sensitivity of the thyroid follicular cells remains to be the radiation absorbed dose to the thyroid tissue, its exact calculation however being one of the obstacles in therapeutic nuclear medicine.

Ophthalmopathy

GD is frequently accompanied by ophthalmopathy; the reported incidences largely depending on the diagnostic criteria employed [23–26]. Prospective, randomized, controlled trials have shown that radioiodine treatment is associated with a greater risk of the appearance or worsening of ophthalmopathy in Graves' patients than antithyroid treatment. The risk is especially increased in patients who smoke cigarettes in keeping with the importance of smoking as a susceptibility factor in the development of ophthalmopathy, thus patients should be strongly advised to quit smoking. Oral or intravenous administration of steroids helps preventing exacerbation of thyroid eye disease, and this approach has to be considered the standard of care in patients who have clinically active ophthalmopathy at the time of treatment [22–24, 26, 27]. A radiation absorbed dose below 200 Gy, a thyroid volume of more than 60–80 ml and use of radioiodine without steroid medication have been shown to be associated with a higher risk of worsening of eye symptoms. Despite the controversy regarding adequate management of patients with Graves' hyperthyroidism and thyroid eye disease, most authors agree that in the presence of predisposing risk factors such as large goitres or heavy smoking, ablative therapy should be recommended [27–33].

Radiation-Induced Cancers

A small excess of mortality from malignancy was reported in one investigation but the study was biased by the increased surveillance. In other large series, no effects of radioiodine therapy on survival have been observed, whereas some reports suggested an increased relative risk for the development of certain types of cancer (thyroid, stomach, bladder, kidney, and haematological malignancies). However, these observations still remain to be confirmed by monitoring larger patient samples, so that currently no definite conclusion with respect to risk for subsequent malignancies can be drawn [34–36].

Dosimetry

For the treatment of GD or Plummer's disease (toxic nodular goitre) I-131 is normally administered orally using activities between 100 MBq and 1500 MBq.

The rationale behind dosimetry for this kind of treatment is that the incidence of long-term hypothyroidism is higher with an earlier onset for patients treated with higher activities [37] resulting in an attempt to individualize and thus optimize therapy. A large variation exists in the literature on the value of target absorbed dose to be delivered to the hyperthyroid tissue to become euthyroid. Most authors indicate 70 Gy but absorbed doses as high as 200 Gy are reported [37].

For a pretherapeutic dosimetric assessment of the activity needed to achieve a certain prescribed absorbed dose to the target volume the formalism described in the EANM SOP on dosimetry prior to radioiodine therapy of benign thyroid diseases is recommended for use [38]:

$$A[MBq] = \frac{1}{\bar{E}} \cdot \frac{M[g] \cdot D[Gy]}{\int_0^\infty RIU(t)dt}.$$

The activity A to be administered is calculated from

M: Mass of the target volume,
D: Absorbed dose to be achieved in the target volume,

RIU(t): Relative radioiodine uptake (unit: %) as a function of time:

$$\bar{E} : \bar{E} = 2.808 \frac{Gy.g}{MBq.d}$$

Constant (valid for a thyroid with M = 20 g) which will typically introduce ≤5% error for masses M ≤90 g and produce results with adequate accuracy for most patients [38]. A simple mass-dependent expression for the factor 1/Ē is provided in the supplement of the EANM SOP and may be used for an improved estimate in large goitres [38].

Guidance to the assessment and details of the calculation procedures can be found in the EANM SOP [38]. In short, a determination of the mass of the target volume and of the pretherapeutic iodine biokinetics are needed. For measuring the biokinetics either repeated scans of the neck or probe measurements of the patient's thyroid are needed [39]. Care of the appropriate calibration of the measuring system should be taken.

The thyroid or the target volume mass is generally determined by ultrasound [40], by pretherapeutic scintigraphy [41], by computed tomography [42], by magnetic resonance imaging [43], or by I-124 PET [44–46].

This dosimetric approach assumes that the iodine kinetics of a tracer and of a therapeutic amount of administered activity are similar. For a confirmation of the absorbed dose achieved after therapy a post-therapeutic dose assessment is recommended as, according to some authors, a pretherapeutic tracer 'dose' may induce 'stunning'. This effect might limit the uptake of the therapeutic activity in the thyroid gland [47].

Due to the uncertainties related to all of these procedures described here, an overall systematic uncertainty of the dose assessment process of 30–50% must be assumed [48].

Special Considerations in Children

Hyperthyroidism in children is mostly caused by GD and the risk of relapse in this age group is much higher than in adults. There is good evidence that the fetal and young thyroid is particularly sensitive to radiation and it is therefore appropriate to avoid treating hyperthyroid children with radioiodine if reasonable and safe alternatives are available. This can be a challenging decision since surgical thyroidectomy in young children has been accompanied by a

relatively high morbidity and antithyroid drugs have a certain incidence of compliance problems and drug complications [49–51]. At the very least, an extended trial of antithyroid drugs is advisable, although occasionally drug toxicity makes this strategy impractical.

However, reports of radioiodine therapy in young children have shown that it is effective and late follow-up has shown no deleterious effects [52].

REFERENCES

1. Becker DV, Sawin CT. Radioiodine and thyroid disease: the beginning. *Semin Nucl Med*, 1996; **26**: 155–64.
2. Hegedüs L. Treatment of Graves' hyperthyroidism: evidence-based and emerging modalities. *Endocrinol Metab Clin North Am*, 2009; **38**: 355–71.
3. Brent GA. Clinical practice. Graves' disease. *N Engl J Med*, 2008; **358**: 2594–605.
4. Silberstein EB, Alavi A, Balon HR, *et al*. The SNMMI practice guideline for therapy of thyroid disease with I-131 3.0. *J Nucl Med*, 2012; **53**: 1633–51.
5. Nixon IJ, Angelos P, Shaha AR, Rinaldo A, Williams MD, Ferlito A. Image-guided chemical and thermal ablations for thyroid disease: review of efficacy and complications. *Head Neck*, 2018; **40**: 2103–15.
6. Surks MI, Ortiz E, Daniels GH, *et al*. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *JAMA*, 2004; **291**: 228–38.
7. Royal College of Physicians. *Radioiodine in the Management of Benign Thyroid Disease: Clinical Guidelines: Report of a Working Party*. London: Royal College of Physicians, 2007.
8. Vaidya B, Williams GR, Abraham P, Pearce SH. Radioiodine treatment for benign thyroid disorders: results of a nationwide survey of UK endocrinologists. *Clin Endocrinol (Oxf)*, 2008; **68**: 814–20.
9. Bonnema SJ, Bennedbaek FN, Ladenson PW, Hegedüs L. Management of the nontoxic multinodular goiter: a North American survey. *J Clin Endocrinol Metab*, 2002; **87**: 112–17.
10. Bachmann J, Kobe C, Bor S, *et al*. Radioiodine therapy for thyroid volume reduction of large goitres. *Nucl Med Commun*, 2009; **30**: 466–71.
11. Fast S, Nielsen VE, Bonnema SJ, Hegedüs L. Time to reconsider nonsurgical therapy of benign non-toxic multinodular goitre: focus on recombinant human TSH augmented radioiodine therapy. *Eur J Endocrinol*, 2009; **160**: 517–28.
12. Knobel M. Etiopathology, clinical features, and treatment of diffuse and multinodular nontoxic goiters. *J Endocrinol Invest*, 2016; **39**: 357–73.
13. Cramon P, Bonnema SJ, Bjorner JB, *et al*. Quality of life in patients with benign nontoxic goiter: impact of disease and treatment response, and comparison with the general population. *Thyroid*, 2015; **25**: 284–91.
14. Fast S, Nielsen VE, Grupe P, Bonnema SJ, Hegedüs L. Optimizing ^{131}I uptake after rhTSH stimulation in patients with nontoxic multinodular goiter: evidence from a prospective, randomized, double-blind study. *J Nucl Med*, 2009; **50**: 732–7.
15. Graf H. Recombinant human TSH and radioactive iodine therapy in the management of benign multinodular goiter. *Eur J Endocrinol*, 2015; **172**: R47–52.
16. Lee YY, Tam KW, Lin YM, *et al*. Recombinant human thyrotropin before I-131 therapy in patients with nodular goitre: a meta-analysis of randomized controlled trials. *Clin Endocrinol (Oxf)*, 2015; **83**: 702–10.
17. Weetman AP. Radioiodine treatment for benign thyroid diseases. *Clin Endocrinol (Oxf)*, 2007; **66**: 757–64.
18. Bonnema SJ, Hegedüs L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*, 2012; **33**: 920–80.
19. Walter MA, Briel M, Christ-Crain M, *et al*. Effects of antithyroid drugs on radioiodine treatment: systematic review and meta-analysis of randomised controlled trials. *BMJ*, 2007; **334**: 514–17.
20. Lind P. Strategies of radioiodine therapy for Graves' disease. *Eur J Nucl Med Mol Imaging*, 2002; **29** Suppl 2: S453–7.
21. Sabri O, Zimny M, Schulz G, *et al*. Success rate of radioiodine therapy in Graves' disease: the influence of thyrostatic medication. *J Clin Endocrinol Metab*, 1999; **84**: 1229–33.
22. Imseis RE, Vanmiddlesworth L, Massie JD, Bush AJ, Vanmiddlesworth NR. Pretreatment with propylthiouracil but not methimazole reduces the therapeutic efficacy of iodine-131 in hyperthyroidism. *J Clin Endocrinol Metab*, 1998; **83**: 685–7.
23. Acharya SH, Avenell A, Philip S, Burr J, Beva JS, Abraham P. Radioiodine therapy for Graves' disease and the effect on ophthalmopathy—a systematic review. *Clin Endocrinol*, 2008; **69**: 943–50.
24. Sisson JC, Schipper MJ, Nelson CC, Freitas JE, Frueh BR. Radioiodine therapy and Graves' ophthalmopathy. *J Nucl Med*, 2008; **49**: 923–30.
25. Tanda ML, Lai A, Bartalena L. Relation between Graves' orbitopathy and radioiodine therapy for hyperthyroidism: facts and unsolved questions. *Clin Endocrinol (Oxf)*, 2008; **69**: 845–7.
26. Laurberg P, Wallin G, Tallstedt L, Abraham-Nordling M, Lundell G, Tørring O. TSH receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*, 2008; **158**: 69–75.
27. Bartalena L, Baldeschi L, Boboridis K, *et al*. The 2016 European Thyroid Association/European Group on Graves' Orbitopathy Guidelines for the Management of Graves' Orbitopathy. *Eur Thyroid J*, 2016; **5**: 9–26.
28. Bartalena L, Baldeschi L, Dickinson AJ, *et al*. Consensus statement of the European Group on Graves' orbitopathy (EUGOGO) on management of GO. *Eur J Endocrinol*, 2008; **158**: 273–85.
29. Perros P, Kendall-Taylor P, Neoh C, Frewin S, Dickinson J. A prospective study of the effects of radioiodine therapy for hyperthyroidism in patients with minimally active graves' ophthalmopathy. *J Clin Endocrinol Metab*, 2005; **90**: 5321–23.
30. Bartalena L, Marcocci C, Bogazzi F, *et al*. Relation between therapy for hyperthyroidism and the course of Graves' ophthalmopathy. *N Engl J Med*, 1998; **338**: 73–8.
31. Wiersinga WM. Preventing Graves' ophthalmopathy. *N Engl J Med*, 1998; **338**: 121–2.
32. Read CH Jr, Tansey MJ, Menda Y. A 36-year retrospective analysis of the efficacy and safety of radioactive iodine in treating young Graves' patients. *J Clin Endocrinol Metab*, 2004; **89**: 4229–33.
33. Li HX, Xiang N, Hu WK, Jiao XL. Relation between therapy options for Graves' disease and the course of Graves'

ophthalmopathy: a systematic review and meta-analysis. *J Endocrinol Invest*, 2016; **39**: 1225–33.
34. Saenger EL, Thoma GE, Tompkins EA. Incidence of leukemia following treatment of hyperthyroidism. Preliminary report of the Cooperative Thyrotoxicosis Therapy Follow-Up Study. *J Am Med Assoc*, 1968; **205**: 855–62.
35. Ron E, Doody MM, Becker DV *et al.* Cancer mortality following treatment of adult hyperthyroidism. *JAMA*, 1998; **280**: 347–55.
36. Hieu TT, Russell AW, Cuneo R, *et al.* Cancer risk after medical exposure to radioactive iodine in benign thyroid diseases: a meta-analysis. *Endocr Relat Cancer*, 2012; **19**: 645–55.
37. Bockisch A, Jamitzky T, Derwanz R, Biersack HJ. Optimized dose planning of radioiodine therapy of benign thyroidal diseases. *J Nucl Med*, 1993; **34**: 1632–8.
38. Hänscheid H, Canzi C, Eschner W, *et al.* EANM Dosimetry Committee series on standard operational procedures for pre-therapeutic dosimetry II. Dosimetry prior to radioiodine therapy of benign thyroid diseases. *Eur J Nucl Med Mol Imaging*, 2013; **40**: 1126–34.
39. Giovanella L, Verburg F, Ceriani L. One-stop-shop radioiodine dosimetry in patients with Graves' disease. *Endocrine*, 2017; **56**: 220–1.
40. Hegedüs L, Perrild H, Poulsen LR, *et al.* The determination of thyroid volume by ultrasound and its relation to body weight, age, and sex in normal subjects. *J Clin Endocrinol Metab*, 1983; **56**: 260–3.
41. van Isselt JW, de Klerk JMH, van Rijk PP, *et al.* Comparison of methods for thyroid volume estimation in patients with Graves' disease. *Eur J Nucl Med*, 2003; **30**: 525–31.
42. Hermans R, Bouillon R, Laga K, *et al.* Estimation of thyroid gland volume by spiral computed tomography. *Eur Radiol*, 1997; **7**: 214–16.
43. Bonnema SJ, Andersen PB, Knudsen DU, Hegedüs L. MR imaging of large multinodular goiters: observer agreement on volume versus observer disagreement on dimensions of the involved trachea. *Am J Roentgenol*, 2002; **179**: 259–66.
44. Crawford DC, Flower MA, Pratt BE, *et al.* Thyroid volume measurement in thyrotoxic patients: comparison between ultrasonography and iodine-124 positron emission tomography. *Eur J Nucl Med*, 1997; **24**: 1470–8.
45. Westphal JG, Winkens T, Kühnel C, Freesmeyer M. Low-activity I-124-PET/low-dose CT versus I-131 probe measurements in pretherapy assessment of radioiodine uptake in benign thyroid diseases. *J Clin Endocrinol Metab*, 2014; **99**: 2138–45.
46. Gabler AS, Kühnel C, Winkens T, Freesmeyer M. Assessment of minimum I-124 activity required in uptake measurements before radioiodine therapy for benign thyroid diseases. *J Nucl Med*, 2016; **57**: 1201–6.
47. Traino AC, Di Martino F, Grosso M, *et al.* A predictive mathematical model for the calculation of the final mass of Graves' disease thyroids treated with ^{131}I. *Phys Med Biol*, 2005; **50**: 2181–91.
48. Sabri O, Zimny M, Schreckenberger M, Meyer-Oelmann A, Reinartz P, Buell U. Does thyroid stunning exist? A model with benign thyroid disease. *Eur J Nucl Med*, 2000; **27**: 1591–7.
49. Traino AC, Xhafa B. Accuracy of two simple methods for estimation of thyroidal I-131 kinetics for dosimetry-based treatment of Graves' disease. *Med Phys*, 2009; **36**: 1212–18.
50. Rivkees SA, Dinauer C. An optimal treatment for pediatric Graves' disease is radioiodine. *J Clin Endocrinol Metab*, 2007; **92**: 797–800.
51. Lee JA, Grumbach MM, Clark OH. The optimal treatment for pediatric Graves' disease is surgery. *J Clin Endocrinol Metab*, 2007; **92**: 801–3.
52. Cheetham T, Bliss R. Treatment options in the young patient with Graves' disease. *Clin Endocrinol (Oxf)*, 2016; **85**: 161–4.

3.3.8 Surgery for Thyrotoxicosis

Nancy D. Perrier, Orlo H. Clark, and Sarah B. Fisher

Introduction

Thyrotoxicosis, or the constellation of symptoms resulting from hyperthyroidism, is most often due to a single hyperfunctioning nodule (toxic adenoma), toxic multinodular goitre (Plummer's disease), or diffuse toxic goitre (Graves' disease). Symptoms may include anxiety, fatigue, palpitations, weight loss, heat intolerance, diarrhoea, and/or menstrual irregularities (in women), and patients may present with tremor, tachycardia or atrial fibrillation (particularly in older patients), thyromegaly with an audible bruit, and/or proptosis (Graves' orbitopathy or thyroid eye disease (TED)). Thyrotoxicosis is managed in one of three ways: use of antithyroid medications (ATDs) to decrease thyroid hormone production, destruction of thyroid tissue with radioactive iodine (RAI), or surgical resection of the thyroid (thyroidectomy). Of the three, surgery is the most invasive but also the most definitive, and may be favoured over medical management or RAI depending on specific clinical factors, patient preferences, and healthcare infrastructure (Table 3.3.8.1 [1–3]).

Definitive indications for thyroidectomy in the setting of thyrotoxicosis include symptomatic compression, concomitant documented or suspected malignancy (including thyroid nodule(s) suspicious for malignancy or greater than 3–4 cm in size), or coexisting hyperparathyroidism requiring surgical intervention. Surgery may be the most timely and definitive therapy for women who are pregnant or lactating, or who are planning pregnancy (ideally upon return to normal thyroid levels), and is favoured for patients with moderate to severe TED. Surgery is also appropriate in situations in which immediate control of symptoms is necessary (concomitant cardiac comorbidities with exacerbation) or in thyroid storm refractory to medical therapy. Other considerations for surgery include medication intolerance or non-adherence, failure to achieve a euthyroid state with medication, and/or patient preference to avoid exposure

Table 3.3.8.1 Comparison of thyroidectomy, radioactive iodine ablation, and medical therapy for thyrotoxicosis

	Thyroidectomy	Radioactive iodine ablation (RAI)	Medical therapy
Time of onset	• Immediate	• 4–6 months	• 4–6 weeks
Advantages	• Immediately curative with quick onset • May be beneficial for patients with limited access to care • Cost-effective	• Less invasive than surgery • May be more cost-effective for patients above age 60	• Noninvasive • Non-ablative (retain native thyroid function) • Achieves remission in 40–50% of patients after 12–18 months
Disadvantages	• Typically requires medical therapy preoperatively • Invasive—risk of surgical injury • Results in iatrogenic hypothyroidism	• Requires medical therapy preprocedure in selected patients due to the risk of transient worsening of thyrotoxicosis • ~80% become hypothyroid within 2–3 months of treatment • 5–15% require repeat treatment • Requires isolation • Avoid radioactivity in women who are pregnant or lactating • 15–20% risk of inducing/aggravating Graves' orbitopathy*	• Medication non-adherence • Side effects Rare but serious: • Agranulocytosis • Hepatotoxicity • Birth defects (methimazole) • Vasculitis More common but less serious: • Rash • Arthralgias
Patient preferences[2]	Patients choosing surgery likely place a higher relative value on prompt and definitive control of symptoms and avoidance of radioactivity and/or side effects from medications, with less relative concern for potential surgical risks and need for lifelong hormone replacement	Patients choosing RAI likely place a higher relative value on definitive control of symptoms with avoidance of surgery and/or side effects from medications, with less relative concern for the rapid resolution of symptoms, need for lifelong hormone replacement, and worsening of Graves' orbitopathy	Patients choosing medication likely place a higher relative value on the possibility of remission and avoidance of lifelong hormone replacement, and avoidance of surgery or radioactive exposure, with less relative concern for medication side effects and/or disease recurrence

*Exacerbation of orbitopathy may be ameliorated by concomitant steroid use.

to radioactivity or prolonged medication use. Contraindications to thyroidectomy include severe medical comorbidity precluding operation or presence of terminal disease states substantially limiting life expectancy. Relative contraindications to thyroidectomy include previously operated or irradiated necks or concurrent pregnancy in the first or third trimesters due to the increased risk of fetal loss or the risk of preterm labour, respectively [1, 2].

The decision to proceed with thyroidectomy for thyrotoxicosis also depends on the pathogenesis of the disease. Patients with toxic adenomas or Plummer's disease are less likely to gain long-term control of hyperthyroidism with ATDs, and as such either RAI or thyroidectomy is favoured over medication. For patients with Plummer's disease, thyroidectomy offers near immediate resolution of hyperthyroidism with a low risk (<1%) of disease recurrence, as compared to 20% risk with RAI; treatment failure rates may be even higher in patients with very large goitres undergoing RAI. These benefits associated with surgery are balanced by the risk of hypothyroidism—100% post-thyroidectomy as compared to 3% at 1 year after RAI, and increasing thereafter [2]. As with all aetiologies for thyrotoxicosis, symptomatic compression, or concern for concurrent malignancy favours thyroidectomy.

Choices of Thyroid Surgery

Thyroid lobectomy, while inappropriate for patients with global glandular dysfunction as in Graves' or Plummer's disease, is advantageous for the patient with a toxic adenoma. In this setting the remainder of the gland is functionally suppressed but physiologically capable of normal function. Thyroid lobectomy for toxic adenoma is associated with a faster return to the euthyroid state (2–3 days versus 75% response rate at 3 months) and a lower risk of disease recurrence (<1% vs. 3–5.5%) as compared to RAI. Additionally, although rates of hypothyroidism after lobectomy for non-toxic nodules are as high as 20%, hypothyroidism after lobectomy for toxic adenoma is uncommon (2–3%) [2].

For the patient with global glandular dysfunction, a subtotal or total thyroidectomy is appropriate. Historically subtotal thyroidectomy (thyroidectomy leaving approximately 2.5 gm of tissue) was utilized to avoid postoperative hypothyroidism, and remains particularly useful in geographic areas in which hormone replacement is not readily available. The Hartley-Dunhill procedure, which also attempts to preserve thyroid function by performing a thyroid lobectomy with a subtotal lobectomy on the contralateral side, has the theoretical advantages of making it easier to tailor the size of the remnant and limiting reoperative dissection to one side in the case of disease recurrence, thus decreasing the risk of recurrent laryngeal nerve injury. In reality there are no observed differences in the rates of recurrent laryngeal nerve injury, permanent hypocalcaemia, wound complications, operative mortality, or disease recurrence between the two subtotal approaches [4, 5]. When choosing a subtotal approach, the size of the remnant can be tailored to the exact patient situation. Younger patients, those with TED, and those with high thyroid-stimulating immunoglobulin titres are at higher risk for disease recurrence and thus may benefit from a smaller remnant, whereas patients with concurrent Hashimoto's disease or toxic multinodular goitre may have a lower risk of recurrence and can be left with a larger functional thyroid remnant.

With the availability of synthetic hormone replacement, total or near-total thyroidectomy is now the favoured operative approach as compared to subtotal thyroidectomy, with a lower risk of disease recurrence (0% vs. 7.6% in one study of 346 patients; 0% vs.

30% in another study of 136 patients) and no major differences in the rates of permanent hypoparathyroidism, recurrent laryngeal nerve injury, or postoperative hematoma formation [6–9]. Total thyroidectomy carries a higher rate of temporary postoperative hypocalcaemia (20–30% vs. 4–13%) as compared to subtotal approaches, and it is important to counsel the patient preoperatively about this risk.

Preoperative Evaluation

Patients who present with dysphonia, change in vocal perception, or with a history of previous neck operation should undergo evaluation of the vocal cords preoperatively (most often with direct laryngoscopy or videostroboscopy). Once a decision has been made to pursue thyroidectomy patients should be rendered euthyroid with ATDs, β-blockade, and/or iodinated solutions such as Lugol's solution (LS) [2]. Antithyroid drugs (methimazole, propylthiouracil) inhibit thyroid hormone synthesis (with propylthiouracil also inhibiting T_4 to T_3 conversion), with onset of action over a 2- to 6-week period. Propylthiouracil is considered second line except in patients who are allergic or intolerant to methimazole. Adverse effects are rare but can be serious, with agranulocytosis occurring in 1 in 500 patients and most often presenting with fever and pharyngitis, usually within the first 90 days of drug use. Hepatoxicity occurs between 0.3 to 0.7 per 1000 patients and is usually manifested by cholestatic injury with methimazole and hepatocellular injury and/or fulminant liver failure with propylthiouracil; hepatotoxicity from both drugs typically presents within the first three months of use [2]. Skin rashes with pruritis can also occur. B-blockade is particularly useful in controlling symptoms during the initiation of ATDs, with relief of tachycardia, palpitations, tremors, anxiety, and heat sensitivity. Additionally, at high doses, some β-blockers may also inhibit peripheral conversion of T_4 to T_3 (propranolol, atenolol, and metoprolol) [1].

LS, a combination of elemental iodine and potassium iodide in distilled water, or its equivalents (saturated solution of potassium iodide—SSKI, or potassium iodide—KI, hereafter used interchangeably for the purposes of this chapter) are often used either alone or as adjuncts to β-blockade and/or ATDs in the preoperative setting. Currently the American Thyroid Association recommends preoperative treatment with LS in most patients with Graves' disease to decrease thyroid blood flow, vascularity, and intraoperative blood loss during thyroidectomy [2]. Although initially used as preoperative therapy prior to thyroidectomy in 1929, and recommended by the ATA [10], controversy remains regarding the utility of preoperative treatment with LS [11].

The strongest data supporting the use of LS prior to thyroidectomy comes from several small randomized trials. Erbil and colleagues [12] demonstrated decreased thyroid blood flow on Doppler ultrasound, decreased microvessel density on pathologic examination, and lower operative blood loss (54.4 ml vs. 108.7 ml, P <0.001) in 17 patients with Graves' disease treated preoperatively with LS versus 19 patients treated without. Similarly, Yilmaz and colleagues [13] randomized 40 patients with either Graves' disease or toxic multinodular goitre to preoperative treatment with LS versus without; the patients treated with LS had lower mean thyroid blood flow as measured on Doppler ultrasound and experienced significantly less intraoperative blood loss than those not treated (76 ml vs. 172 ml, P = 0.0001). Another group randomized patients with euthyroid Graves' disease undergoing thyroidectomy to pretreatment with SSKI versus none and demonstrated a trend towards less blood loss in the SSKI treated group (61.7 vs. 161.7 ml, P = 0.054) [14]. None of these studies included data on differences in postoperative complication rates between the groups, and some critics have questioned whether the observed differences translate to a clinically meaningful difference [15].

In a retrospective study of 162 patients with Graves' disease undergoing thyroidectomy without preoperative KI, Shinall and colleagues [15] demonstrated similar outcomes (operative blood loss, duration of operation, and rates of complications including nerve injury, permanent hypocalcaemia, and hematoma formation) as compared to 102 patients with toxic multinodular goitre undergoing thyroidectomy in whom preoperative KI is not indicated. They concluded that the equivalent results between the two groups suggests that medical therapy exclusive of iodinated solutions combined with expert surgical technique renders the utility of iodinated solutions obsolete. Another retrospective study concluded that postoperative morbidity rates among 382 patients with Graves' disease undergoing thyroidectomy without preoperative LS were similar to that of historical controls, with 1.1% (n = 4) requiring reoperation for haematoma, 3.7% (n = 14) developing permanent hypocalcaemia, and 2.4% (n = 9) sustaining permanent nerve injury [16].

While meaningful for the patient who experiences them, the rate of complications after thyroidectomy is generally low, and as such using complications as a measure of the success or failure of therapy with iodinated solutions such as Lugol's may not be realistic. The side effects incurred from a short course of LS or other iodinated solution are rare and generally well tolerated, and it is the authors' practice to use this adjunct therapy.

In patients in whom the need for thyroidectomy is urgent, or in whom ATDs fail to render a euthyroid state, cholestyramine and/or glucocorticoids (hydrocortisone 100 mg every 8 hours) may be considered as adjuncts in the perioperative period [2]. Additionally, iodinated contrast dyes such as iopanoat can be effective, and in the event of thyroid storm can be administered via a nasogastric tube. Iodinated preparations can be halted after the patient is rendered afebrile and the cardiac and neurologic status improve. These patients are at high risk of thyroid storm and the care team should prepare for the possibility with adequate IV access, vasoactive drugs, and appropriate preoperative counselling.

Surgical Procedure

The patient is positioned with the neck hyperextended and a rolled drape placed along the patient's spine to assist in opening the surgical field. A horizontal incision is made in a natural skin crease approximately 1 cm below the cricoid cartilage. The platysma is divided and superior and inferior subplatysmal skin flaps may be elevated. The midline raphe is separated longitudinally and the strap muscles are dissected laterally taking care to avoid or ligate anterior jugular veins. The plane between the sternothyroid and sternohyoid muscle is separated gently to the ansa cervicalis. The middle thyroid veins are divided close to the gland to allow for additional lateral

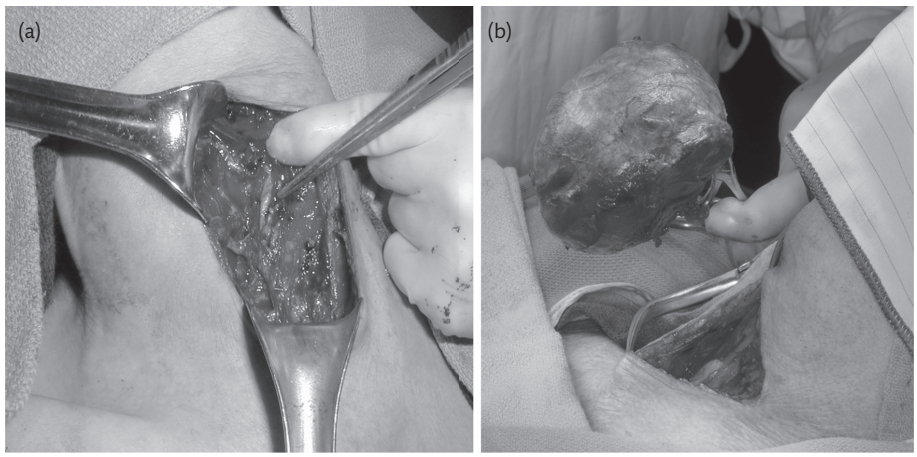


Figure 3.3.8.1 (a) The right recurrent laryngeal nerve; when dissecting around the nerve care is taken to avoid thermal injury from electrocautery, as demonstrated by the silk ligatures around small vascular bundles on either side of the nerve. (b) The noticeably enlarged thyroid gland after thyroidectomy in a symptomatic male.

retraction of the carotid sheath. As the thyroid is freed from surrounding tissues and rotated medially, careful dissection along the capsule of the gland should allow for identification of the recurrent laryngeal nerve (Figure 3.3.8.1). The superior thyroid artery and vein should be ligated as close to the thyroid as possible. Care should be taken to avoid injury to the external branch of the recurrent laryngeal nerve (just lateral to the cricothyroid muscle) and the superior parathyroid gland (typically at the level of the cricoid cartilage). The inferior thyroid artery and vein are approached similarly, with ligation close to the gland to avoid compromise of the parathyroid vascular supply. In the case of uncertainty regarding parathyroid location, frozen section of a portion of the suspected gland can be useful for parathyroid identification. Parathyroid autotransplantation should be considered for devascularized glands. Berry's ligament is encountered as the thyroid is separated from the anterior surface of the trachea, and represents a common site of nerve injury as small vessels in this area may bleed and obscure visualization. After completion of the initial side the procedure is repeated on the contralateral side, with the understanding that if the function of the recurrent laryngeal nerve is in question it is better to defer additional operation until postoperative vocal cord function is assessed.

Complications

Thyroidectomy can be performed safely with very low rates of morbidity and mortality. The most common serious complications following thyroidectomy are hypocalcaemia (transient or permanent), recurrent or superior laryngeal nerve injury (temporary or permanent), and postoperative hematoma. Surgeon experience (as measured by case volume) is well correlated with outcome [17, 18] and patients who undergo surgery by low-volume surgeons (<25 cases/year) are at an increased risk for any complication (odds ratio 1.51, $P = 0.002$). Among high volume thyroid surgeons the rates of permanent hypocalcaemia are <2%, permanent recurrent laryngeal nerve injury <1%, reoperation for hematoma 0.3–0.7%, and mortality less than 1 in 10 000 [2]. Transient hypocalcaemia is more common, with variable rates depending on study definition and ranging between 6 and 72% [19], and a median incidence of 27% in one meta-analysis [20].

When compared to patients undergoing total thyroidectomy for other indications, patients with Graves' disease have a higher rate of symptomatic postoperative hypocalcaemia [19, 21, 22] (OR 1.75, 95% CI 1.34–2.28 in one large meta-analysis [20]). Most often transient, the presumed mechanisms for the increased risk of hypocalcaemia are multifactorial and include parathyroid damage as well as other non-parathyroid mediated causes. From a technical perspective, the increased vascularity and friability of the gland makes visualization more difficult, and may place the parathyroid glands at a higher risk of inadvertent injury during operation. Additionally, the hyperthyroid biochemical profile may increase release of calcitonin during gland manipulation. The hyperthyroid state preoperatively stimulates bone turnover (thyrotoxic osteodystrophy) and urinary calcium loss resulting in depletion of whole body stores of calcium, with postoperative bone hunger noted after the hyperthyroid stimulus is effectively treated [19, 23] Technical factors, such as identifying and protecting all four parathyroid glands and leaving them on a well vascularized lateral pedicle during the operation is critical. Selective calcium and vitamin D supplementation both preoperatively and postoperatively have been proposed to decrease the rate of postoperative hypocalcaemia. One prospective study of 45 patients with Graves' disease treated with oral calcium supplementation for two weeks preoperatively demonstrated a lower rate of symptomatic hypocalcaemia (9% vs. 26%, $P < 0.05$) and a higher average postoperative calcium level (8.6 mg/dl vs. 8.3 mg/dl or 2.15 vs. 2.08 mmol/L, $P = 0.05$) as compared to patients with Graves' disease who did not receive preoperative supplementation [19]. It is the authors practice to correct vitamin D deficiency preoperatively, to selectively give preoperative calcium supplementation to patients with laboratory studies consistent with high bone turnover (elevated alkaline phosphatase and/or osteocalcin), and to monitor postoperative parathyroid hormone (PTH) levels [24] combined with serum calcium values and the presence or absence of symptoms to guide postoperative calcium supplementation.

Criteria to initiate supplementation include serum calcium <7.2 mg/dl, tingling or numbness in the postoperative period or serum PTH level <3.0 pg/ml on lab draws on the evening of the operation or on the morning prior to discharge after thyroidectomy.

Special Situations

Pregnancy

Pregnancy results in substantial physiologic changes in thyroid hormone production, which is often well tolerated in the healthy individual, but leads to unique considerations in the setting of thyroid dysfunction. Graves' disease is the most common cause of hyperthyroidism in pregnancy, occurring in 0.2% of women during pregnancy [25]. Graves' disease should be differentiated from gestational transient thyrotoxicosis, which is mediated by human chorionic gonadotropin (hCG) stimulation of the thyroid stimulating hormone (TSH) receptor, typically occurs in the first trimester, often is associated with hyperemesis, and is self-limited with typically milder symptoms. Pregnancy loss or stillbirth, prematurity, low birth weight, intrauterine growth restriction, thyroid storm, pregnancy induced hypertension, and maternal congestive heart failure are associated with poor control of maternal hyperthyroidism [26].

Treatment of thyrotoxicosis during pregnancy and after pregnancy/during lactation has unique considerations. Antithyroid drugs have been associated with birth defects (3–4% with methimazole, 2–3% with propylthiouracil) and their use should be avoided in the first trimester when possible, or used at the lowest necessary dose [25]. Radioactive iodine is contraindicated during pregnancy or lactation, and, when given to a woman of childbearing age, pregnancy should be avoided within 6 months [1]. Thyroidectomy can be considered for pregnant women with Graves' disease. If surgery is pursued and disease stability allows elective scheduling, thyroidectomy should be performed during the second trimester when the risk of preterm labour or pregnancy loss is lowest.

Paediatrics

Hyperthyroidism in childhood is rare, with only 1–5% of all Graves' disease presenting during this time, although Graves' disease remains the most common cause [27]. Antithyroid drugs are often the first line of therapy, but the rate of remission is lower than that seen in adults (20–30% after 2 years of therapy) [28]. Ablation therapy with RAI is acceptable in the paediatric population, and data suggest that paediatric glands may be more sensitive to RAI than adults, although there are similar limitations in efficacy for patients with large glands (>80 gm). In contrast to external radiation, which has a dose-dependent association with risk of thyroid cancer most pronounced at younger ages (<5 years), RAI has not been associated with later development of thyroid cancer. The effects of RAI on very young patients (<5 years) are not well described and as such thyroidectomy is generally favoured in this group. Thyroidectomy should be performed by an experienced endocrine surgeon [29]. The results are immediate and largely definitive, but there is a small risk of hypoparathyroidism, postoperative bleeding, and nerve injury. As with adults, children should be treated with ATDs, β-blockade, and/or iodinated solutions to achieve a euthyroid state prior to thyroidectomy.

Thyroid Eye Disease

Radioactive iodine ablation therapy is associated with new or worsening orbitopathy, or active TED, making these patients more appropriate for either medical management or thyroidectomy. Smoking cessation, corticosteroids at the time of therapy, and euthyroidism help prevent exacerbation and decrease the duration of TED [1]. Patients who smoke are more likely to develop TED when compared to non-smokers, and smoking reduces the effectiveness of treatments such as corticosteroids and RAI. Biologic drugs such as rituximab have had variable results when studied. Teprotumumab, a human monoclonal antibody against the insulin-like growth factor I receptor, has been suggested to be effective in reducing proptosis [30]. Care for the patient with TED should be coordinated by a multidisciplinary team of endocrinologists, surgeons, ophthalmologists, and radiation oncologists.

Summary

Thyrotoxicosis is managed with antithyroid medications (ATDs), radioactive iodine (RAI), or thyroidectomy, with patient comorbidities and preferences guiding therapy. Of the three modalities, surgery is the most invasive but also the most definitive. Thyroidectomy can be performed with minimal morbidity, particularly in experienced hands, and should be considered in the patient who places more relative emphasis on prompt and definitive control of symptoms with avoidance of radioactive therapy and/or medications, with less concern regarding operative risks and/or need for lifelong thyroid hormone replacement. Preoperative treatment with ATDs, β-blockers, and/or iodinated solutions to achieve a euthyroid state is recommended.

REFERENCES

1. Burch HB, Cooper DS. Management of Graves' disease: a review. *JAMA*, 2015; **314**: 2544–54.
2. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*, 2016; **26**: 1343–421.
3. Vidal-Trecan GM, Stahl JE, Eckman MH. Radioiodine or surgery for toxic thyroid adenoma: dissecting an important decision. A cost-effectiveness analysis. *Thyroid*, 2004; **14**: 933–45.
4. Andaker L, Johansson K, Smeds S, Lennquist S. Surgery for hyperthyroidism: hemithyroidectomy plus contralateral resection or bilateral resection? A prospective randomized study of postoperative complications and long-term results. *World J Surg*, 1992; **16**: 765–9.
5. Muller PE, Bein B, Robens E, Bein HS, Spelsberg F. Thyroid surgery according to Enderlen-Hotz or Dunhill: a comparison of two surgical methods for the treatment of Graves' disease. *Int Surg*, 2001; **86**: 112–16.
6. Barakate MS, Agarwal G, Reeve TS, Barraclough B, Robinson B, Delbridge LW. Total thyroidectomy is now the preferred option for the surgical management of Graves' disease. *ANZ J Surg*, 2002; **72**: 321–4.

7. Mittendorf EA, McHenry CR. Thyroidectomy for selected patients with thyrotoxicosis. *Arch Otolaryngol Head Neck Surg*, 2001; **127**: 61–5.
8. Wilhelm SM, McHenry CR. Total thyroidectomy is superior to subtotal thyroidectomy for management of Graves' disease in the United States. *World J Surg*, 2010; **34**: 1261–4.
9. Yip J, Lang BH, Lo CY. Changing trend in surgical indication and management for Graves' disease. *Am J Surg*, 2012; **203**: 162–7.
10. Calissendorff J, Falhammar H. Lugol's solution and other iodide preparations: perspectives and research directions in Graves' disease. *Endocrine*, 2017; **58**: 467–73.
11. Hope N, Kelly A. Pre-operative Lugol's iodine treatment in the management of patients undergoing thyroidectomy for Graves' disease: a review of the literature. *Eur Thyroid J*, 2017; **6**: 20–5.
12. Erbil Y, Ozluk Y, Giris M, *et al*. Effect of Lugol solution on thyroid gland blood flow and microvessel density in the patients with Graves' disease. *J Clin Endocrinol Metab*, 2007; **92**: 2182–9.
13. Yilmaz Y, Kamer KE, Ureyen O, Sari E, Acar T, Karahalli O. The effect of preoperative Lugol's iodine on intraoperative bleeding in patients with hyperthyroidism. *Ann Med Surg (2012)*, 2016; **9**: 53–7.
14. Whalen G, Sullivan M, Maranda L, Quinlan R, Larkin A. Randomized trial of a short course of preoperative potassium iodide in patients undergoing thyroidectomy for Graves' disease. *Am J Surg*, 2017; **213**: 805–9.
15. Shinall MC, Jr., Broome JT, Baker A, Solorzano CC. Is potassium iodide solution necessary before total thyroidectomy for Graves' disease? *Ann Surg Oncol*, 2013; **20**: 2964–7.
16. Mercier F, Bonal M, Fanget F, *et al*. Does surgery without Lugol's solution pretreatment for Graves' disease increase surgical morbidity? *World J Surg*, 2018; **42**: 2123–6.
17. Adam MA, Thomas S, Youngwirth L, *et al*. Is there a minimum number of thyroidectomies a surgeon should perform to optimize patient outcomes? *Ann Surg*, 2017; **265**: 402–7.
18. Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. *Ann Surg*, 1998; **228**: 320–30.
19. Oltmann SC, Brekke AV, Schneider DF, Schaefer SC, Chen H, Sippel RS. Preventing postoperative hypocalcemia in patients with Graves' disease: a prospective study. *Ann Surg Oncol*, 2015; **22**: 952–8.
20. Edafe O, Antakia R, Laskar N, Uttley L, Balasubramanian SP. Systematic review and meta-analysis of predictors of post-thyroidectomy hypocalcaemia. *Br J Surg*, 2014; **101**: 307–20.
21. Hallgrimsson P, Nordenstrom E, Bergenfelz A, Almquist M. Hypocalcaemia after total thyroidectomy for Graves' disease and for benign atoxic multinodular goitre. *Langenbecks Arch Surg*, 2012; **397**: 1133–7.
22. Pesce CE, Shiue Z, Tsai HL, *et al*. Postoperative hypocalcemia after thyroidectomy for Graves' disease. *Thyroid*, 2010; **20**: 1279–83.
23. Lal G, Ituarte P, Kebebew E, Siperstein A, Duh QY, Clark OH. Should total thyroidectomy become the preferred procedure for surgical management of Graves' disease? *Thyroid*, 2005; **15**: 569–74.
24. Carr AA, Yen TW, Fareau GG, *et al*. A single parathyroid hormone level obtained 4 hours after total thyroidectomy predicts the need for postoperative calcium supplementation. *J Am Coll Surg*, 2014; **219**: 757–64.
25. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
26. Millar LK, Wing DA, Leung AS, Koonings PP, Montoro MN, Mestman JH. Low birth weight and preeclampsia in pregnancies complicated by hyperthyroidism. *Obstet Gynecol*, 1994; **84**: 946–9.
27. Azizi F, Amouzegar A. Management of thyrotoxicosis in children and adolescents: 35 years' experience in 304 patients. *J Pediatr Endocrinol Metab*, 2018; **31**: 159–65.
28. Rivkees SA. Pediatric Graves' disease: management in the post-propylthiouracil Era. *Int J Pediatr Endocrinol*, 2014; **2014**: 10.
29. Breuer C, Tuggle C, Solomon D, Sosa JA. Pediatric thyroid disease: when is surgery necessary, and who should be operating on our children? *J Clin Res Pediatr Endocrinol*, 2013; **5** Suppl 1: 79–85.
30. Smith TJ, Kahaly GJ, Ezra DG, *et al*. Teprotumumab for Thyroid-Associated Ophthalmopathy. *N Engl J Med*, 2017; **376**: 1748–61.

3.3.9 Management of Graves' Hyperthyroidism

Jacques Orgiazzi

Advantages and Disadvantages of the Various Treatment Modalities of Graves' Disease

For decades, Graves' disease therapeutic approaches have been debated and opposed [1] up to the present time when a large consensus has been reached in favour of the antithyroid drug treatment being the first-line option, even in the United States [2]. Table 3.3.9.1 summarizes the main characteristics of each type of treatment. The advantage-disadvantage balance of each modality should be discussed with the patients in order to reach as much as possible a personalized option based not only on disease presentation but also on patient's preference or professional and social

Table 3.3.9.1 Advantages and disadvantages of the various modalities of treatment of Graves' hyperthyroidism

Treatment modality	Advantages	Disadvantages
Antithyroid drug	• Possibility of long-term remission • No subsequent irreversible hypothyroidism	• Long duration (1–2 years) of treatment and repeated consultations • High relapse rate
Radio-iodine[1]	• Simplicity (depending on local regulation) • Low cost • Rarity of recurrence (depending in the dose)	• Delay of action • Transient expansion of anti-TSH-R autoimmunity • Occurrence or exacerbation of Graves' orbitopathy • Lifelong hypothyroidism
Surgery[2,3]	• Highest restoration rate of euthyroidism • Recurrence uncommon	• Highly experienced surgical-medical team mandatory • Low but unavoidable morbidity • Lifelong hypothyroidism • High cost

[1] Prior restoration of euthyroidism with antithyroid drugs is appropriate in severe cases.
[2] Prior restoration of euthyroidism requires a several weeks antithyroid drugs treatment,
[2, 3] Iatrogenic definitive hypothyroidism is no longer considered a complication but, on the contrary, a marker of no relapse risk.

obligations. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve full recovery [3]. In contrast, it has also been considered that hypothyroidism should no longer be considered as a complication of the 'radical' treatment options but as the ultimate goal to prevent the relapse of the disease [4]. The cost-effectiveness of the various treatment modalities has been evaluated [5]. Radioiodine is the more cost-effective primary treatment modality while surgery is the more costly. The cost of the medical treatment modality is intermediate, but inclusion of the relapse costs tends to level off the differences between the three modalities.

Treatment Modalities

Therapeutic modalities are discussed according to the various clinical presentations, including the usual case, the presence of a large goitre, children and adolescents, presence of orbitopathy, and pregnancy. Severe or acute forms (thyrotoxic storm) are discussed in Chapter 3.3.3.

Overview of Graves' Disease Management

Attention must be given to the quality and conditions of the communication with the patients who must be involved in the selection of the therapeutic strategy [6]. One could suggest considering three successive phases in the management of the patients:

a) Initial clinical and biological evaluation, assessment of disease severity, determination of thyroid function tests and antithyroid receptor antibodies (TRAb) serum level, blood count, and transaminase assay.

b) Restoration of euthyroidism with antithyroid drug treatment. This initial 1–3-month antithyroid drug treatment phase may be considered in every patient, even before subsequent radioactive iodine administration, especially in the more severe cases. This is the period during which the patient may be informed on medical prescriptions as well as antithyroid drug side effects and agreement to medications and biological controls.

c) Re-evaluation of patient's opinion, feelings, and compliance. This is the appropriate time for the selection of the further treatment modality.

The Usual Form of Hyperthyroid Graves' Disease

This corresponds to the 'average 40–50-year-old female patient' without a large goitre, without severe orbitopathy.

Antithyroid Drug Treatment

The Initial Antithyroid Drug (ATD) Treatment

The objective is to reach normal or near-normal serum free T_4 (fT_4) and free T_3 (fT_3) concentrations by 6–8 weeks. The initial dose of carbimazole (CBZ) or of its active derivative methyl-mercaptoimidazole (MMI), 30–40 or 15–30 mg daily, respectively, depends on the severity of the hyperthyroidism assessed clinically and on the serum levels of fT_4 and fT_3. Clinical and biological serum fT_4 and fT_3, but not thyroid-stimulating hormone (TSH) levels which may remain suppressed for several weeks, must re-evaluated after 3–4 weeks of treatment. As the serum concentrations of fT_4 and fT_3 decrease, the daily dose of MMI is reduced to 10–20 mg and that of CBZ to 20–30 mg. Further adjustment is necessary in order to prevent iatrogenic hypothyroidism. Patients should be instructed on the risk of the side effects of antithyroid drugs. At this stage, β-blocking medications (e.g. propranolol, 40–160 mg/d) are useful to alleviate cardiovascular symptoms, nervousness, and anxiety.

In newly diagnosed patients, the completion of this first phase is the point in time when to discuss and select, or confirm the subsequent treatment strategy.

The 'Medical' 12–18 Month Antithyroid Drug Treatment

Since the main drawback of the medical option is an overall 50% relapse rate after the end of the antithyroid drug course, it is important to try and select the appropriate treatment strategy. Male gender, young age, severity of hyperthyroidism, large thyroid volume, and high level of TRAb are the initial factors statistically correlated with a greater risk of relapse. A recent study has confirmed lower age (<40 years), higher serum fT_4 (≥ 40 pmol/L), higher serum TRAb (>20 IU/L), and larger goitre size at diagnosis as independent relapse risk factors which, combined with genetic markers, PTPN22 C/T polymorphism and human leukocyte antigen (HLA) subtypes

DQB1*02, DQA1*05, and DRB1*03, provide the basis for a continuous score, the Graves' Recurrent Events After Therapy (GREAT +) score, predictive of remission or relapse, with the potential to orient the allocation of a given patient to the medical or radical option [7]. Interestingly enough, even the clinical GREAT score, not including the genetic markers, may assess quantitatively the recurrence risk, as recently independently confirmed [8]. Importantly, remission rate is greater after prolonged antithyroid drug treatment (12–18 months) but not with a higher dose of antithyroid drug or the combination of levothyroxine and antithyroid drug, the so called 'block–replace' regimen. However, as the block–replace regimen prevents iatrogenic hypothyroidism, it allows for less frequent biological controls.

In practical terms, in the medical option, antithyroid drug treatment is then continued at 5 mg/day of MMI, or 5–15 mg/day of CBZ in the titration regimen, or maintained at higher dose in the block–replace regimen, associated with L-thyroxine started at 25 µg/day and progressively increased to reach 100–125 µg/day after 6–10 weeks. Clinical and biological follow-up is mandatory. Except in the occurrence of non-compliance or of untoward effects that the patient should be instructed to report without delay, the course of the treatment is usually uneventful, and in most cases correction of hyperthyroidism is sufficient to allow resumption of a normal life.

There is no single good marker to determine when to stop the antithyroid drug treatment after the 12–18-month course. After about 18 months of treatment, antithyroid drug is usually withdrawn, or tapered then stopped. In the case of the 'block–replace' modality it is advisable to shift to the 'titration' modality after 10–14 months of treatment so that the intrinsic functional thyroid status can be indirectly assessed through the required daily dose of antithyroid drug. Lack of goitre volume reduction, persistence of thyroid hypervascularity, persistence of suppressed level of TSH or the requirement for a full dose of antithyroid drug (titration regimen) are indicative of persistent disease activity and, consequently, of recurrence at drug withdrawal. Similar information is provided by persistently elevated serum levels of TRAb, while low or negative TRAb is not predictive of remission. Prolonged follow-up is necessary to avoid misinterpreting the period of euthyroidism that follows drug withdrawal as a true remission. Most relapses occur within months of cessation of antithyroid drug and 90% within 3 years.

There is no specific treatment strategy for relapsing Graves' disease. However, a radical approach appears more appropriate and acceptable to the patient. Nevertheless, after a second course of antithyroid drug treatment chances of relapse or remission are similar to the ones after the first course [9].

Ultra-Long Antithyroid Drug Treatment

In some patients, continuation of antithyroid drug treatment for 2 to 3 years, or even longer, appears an appropriate option. The necessary conditions for ultra-long antithyroid drug course are: (a) requirement for a low antithyroid drug dose (carbimazole 5 mg/day) or less; (b) patient's dependability and drug tolerance; and (c) moderate intensity of the disease, with absent or small goitre [10]. That such a low-dose maintenance treatment is efficient is suggested by the possible occurrence of a relapse at drug withdrawal or after iodine contamination.

Radioactive Iodine

Treatment with 131I radioactive iodine is now considered, as thyroidectomy, as a radical treatment for Graves' disease aiming at eradicating hyperthyroidism at the expense of definitive hypothyroidism. Some patients have concern about potential radiation danger to self and others. Also, in some countries local legislation restrict the ambulatory use of radioactive iodine treatment. And, in many places radioactive iodine is not administered to patients under 30–35 years of age, a rule which tends to be less rigid in specific cases, even in adolescents and children. Radioiodine as the first-line treatment should be considered for patients more likely to relapse after an antithyroid drug course, if monitoring of antithyroid drug treatment appears impractical, in case of intolerance to antithyroid drugs, and at relapse after an antithyroid drug course. In order to eradicate hyperthyroidism, ablative 131I doses range from 5.9 to 6.5 MBq/gram (160–176 µCi/g) of thyroid tissue for adjusted doses, or from 400 to 600 MBq (10.8–16.2 mCi) when 'fixed' doses are administered. Whether routine measurement of the effective half-life of intrathyroidal 131I is cost-effective has to be evaluated according to local iodine intake status. This range of doses eradicates hyperthyroidism in about two-thirds of patients and more than 60% are hypothyroid at 1 year [11]. Failure can be treated 6 months after initial irradiation with a second dose of 131I.

Although not required, except in older patients or in case of cardiovascular comorbidities to prevent a possible exacerbation of hyperthyroidism, antithyroid drug treatment prior to radioactive iodine administration would be prescribed by 20–40% of thyroidologists. Antithyroid drugs may also be prescribed after 131I administration to shorten the hyperthyroid lag phase until the irradiation effect [12]. Antithyroid drugs should be stopped 2 to 3 days prior to radioactive iodine therapy and resumed 3 to 5 days later [3]. They decrease the efficiency of radioactive iodine, but this is no longer an issue with the relatively high doses currently in use.

Prevention of post-irradiation hypothyroidism requires close follow-up or systematic L-thyroxine treatment, 100 µg/day, to be started at day 15. In any case, plasma fT_4 level should be tested 4 to 6 weeks post-irradiation, then monthly. The efficiency of the radioactive treatment is evaluated at 4 months. TSH should be followed-up until normalization in order not to miss post-irradiation subclinical hyperthyroidism or moderate T_3 hyperthyroidism. The risk of late-onset hypothyroidism should be anticipated with annual follow-up. Serum TRAb may increase transiently after radioiodine administration; peaking 3 to 5 months after irradiation, an effect to take into consideration in women of childbearing age as well as in adolescents [13].

Thyroid Surgery

Thyroidectomy, total or near-total according to the recent recommendations, is to be considered in the case of suspicion of malignant nodule, radioiodine treatment refusal or inapplicability, in the perspective of a future pregnancy, and in severe intolerance to antithyroid drug as well as in the case of severe hyperthyroidism. Thyroidectomy in hyperthyroid patients with Graves' disease requires the prior restoration of euthyroidism with antithyroid drug

to prevent postoperative thyroid crisis and suffocating hematoma. If the 6–12 week antithyroid drug treatment is inapplicable, alternate medical preparation includes the use of intravenous or oral propranolol, iodine as Lugol's solution, and, according to the severity of hyperthyroidism, oral or intravenous methylprednisolone, oral cholestyramine or oral lithium carbonate, and if necessary extracorporeal plasmapheresis [14, 15]. Thyroidectomy requires a highly and specifically trained and experienced anaesthetic, surgical, and nursing team. The two main complications of thyroidectomy, permanent recurrent laryngeal nerve palsy and permanent hypocalcaemia/hypoparathyroidism occur, on the average, in 1.8% and 5.9% of the patients, respectively [16].

Other Clinical Conditions

Presence of a Large Goitre

The schedule of the therapeutic management programme must be precisely defined. It includes: (a) restoration of euthyroidism with antithyroid drugs; (b) if required by the surgical team, preoperative iodine (50–120 mg/day: Lugol's solution or saturated solution of potassium iodide (SSKI), the duration of which should not exceed 10 days; (c) the surgical procedure itself; (d) the immediate postoperative management. Alternative preoperative programmes have been advocated, using only β-blockers with or without iodide. They require tight monitoring and highly experienced medical–surgical teams. Whatever the preparation, surgery on patients with hypervascular goitre has to be carefully planned and is best performed in selected centres.

Children and Adolescents

Medical Treatment

Initial treatment, in this age group, is antithyroid drugs. However, because compliance is difficult to maintain in the long-term, the relapse rate is high so that thyroidectomy or radioiodine administration is often an alternative. Even more than in adults, only MMI or carbimazole should be used in children because of the high risk of propylthiouracil (PTU)-induced severe hepatitis [17]. The initial starting dose of MMI or carbimazole is 0.5–1 mg/kg/day, with a maximal dose of 30 mg per day, dose to be reduced after 2–4 weeks as thyroid hormone levels normalize. TSH becomes detectable usually within 2–4 months. Thyroid function should be assessed every 3–6 weeks in order to prevent hypothyroidism, then every 3 months at lower antithyroid drug dose. Of note, in this age group, 10% of the patients may elicit T_3-predominant Graves' disease with a notably enlarged thyroid gland and elevated TRAb levels [18]. As to the feasibility of the titration and block–replace regimens, in this age range, compliance may be lower with the latter. Side-effect prevalence is very low at MMI dose <10 mg/day, and may be dose-dependent. Less than 30% of children treated with antithyroid drug treatment for 2 years achieve remission [18]. No initial individual predictive factors have been identified. However, the remission rate increases from 20% to 49% as the duration of the antithyroid drug treatment increases from 4 to 10 years [18]. Therefore, long-term 8–10-year antithyroid drug treatment might be carried on in children and adolescents in case of good compliance and drug tolerance before radical treatment is considered.

Radical Treatment

- Radioactive iodine treatment is safe in children and adolescents. Observational data indicate (a) no case of thyroid cancer developed in paediatric patients treated with dose of radioiodine iodine >150 μCi/g; (b) no evidence for an increased risk of non-thyroid cancer; however, it is prudent to avoid radioactive iodine treatment in children below 5 years of age; and (c) no evidence of adverse effects to offspring of children treated with 131I [19]. Radioiodine should be properly used with large enough doses (120–200 μCi/g) to interfere with the replication potential of residual cells. Strict and prolonged follow-up is mandatory to detect subclinical hypothyroidism.
- Surgical treatment is less controversial, especially in the case of a significant goitre, and in children <5 years, provided an experienced paediatric thyroid surgeon is available as the risk of complication is dependent on the surgeon's experience [20]. The preparation to the thyroidectomy is similar to that in adults.

Graves' Orbitopathy

Incidence of orbitopathy, including that of the more severe forms, has decreased in recent years in Western Europe. In recent studies, while it was initially present in 25% of the patients at diagnosis of Graves' disease, Graves' orbitopathy occurred in 13–15% during antithyroid drug treatment, as a mild form in 80% of the cases [21, 22]. In addition to smoking and to long-standing hyperthyroidism, iatrogenic hypothyroidism, and radioactive iodine treatment may trigger or worsen Graves' orbitopathy in patients treated for Graves' disease. In patients with mild or even moderate-to-severe orbitopathy, glucocorticoids prevent the radioactive iodine exacerbation of the orbitopathy [23]. Whether medical or radical treatment of hyperthyroidism is more appropriate in cases of severe or malignant ophthalmopathy remains unsettled [22]. The major point, in this occurrence, is to ensure and maintain a state of stable euthyroidism, even on prolonged antithyroid drug treatment [24].

Graves' Disease and Pregnancy

During pregnancy, hyperthyroidism is deleterious to the mother, the evolution of pregnancy, and the fetus. Graves' disease, after a transient exacerbation of clinical symptoms likely related to the first-trimester peak of hCG, usually tends to spontaneously improve or even remit in the second half of pregnancy. The primary therapeutic objectives are (a) restoration of maternal euthyroidism; (b) avoidance of fetal hypothyroidism; and (c) evaluation of the risk of fetal hyperthyroidism resulting from the transplacental transfer of maternal stimulating TRAb. Radioactive iodine is contraindicated in pregnant women. Contrarily to T_4, antithyroid drugs cross the placenta without limitation, a combination which prohibits the block–replace regimen and requires restricting antithyroid drug dose to the minimum to maintain maternal fT_4 in the upper normal range in order to prevent fetal hypothyroidism. Antithyroid drug treatment can be withdrawn in the majority of the cases near mid-pregnancy. In contrast, in severe forms of Graves' disease, usually with large hypervascular goitre and very high TRAb concentrations, restoration of euthyroidism requires prolonged full-dose antithyroid drug treatment. Tight management usually allows avoiding thyroidectomy during pregnancy, an option which does not protect from the risk of fetal hyperthyroidism [25]. Antithyroid drugs may be teratogenic and birth defects of all types are observed in 2%–3%

of exposed children. The high-risk period encompasses gestational weeks 6–10, the major period of organogenesis. Birth defects are more severe in newborns of CBZ/MMI than PTU treated mothers. Consequently, PTU should be preferred to CBZ/MMI as early as possible in the first trimester of pregnancy, or even prior to conception and patients on CBZ/MMI should be switched to PTU [26].

Fetal/neonatal hyperthyroidism occur in 2–10% of pregnancies in women with active Graves' disease. TRAb level at the beginning of the third trimester is predictive of fetal/neonatal hyperthyroidism. Women who must be screened for TRAb include, in addition to patients with ongoing Graves 'disease, patients previously treated for Graves' disease either by surgery or radioiodine whatever their current thyroid status, and patients with a previous child with neonatal transient hyper- or hypothyroidism. Fetal thyroid status may be assessed indirectly through clinical signs of hyperthyroidism, of low sensitivity/specificity, and by thyroid ultrasonography after the 23–25th week of pregnancy [27]. However, depending independently on the serum concentration of stimulating TRAb and on the dose of antithyroid drug in the mother, fetal thyroid enlargement may reflect either fetal hyper or hypothyroidism the differentiation of which may require, especially when the goitre is large, to determine TSH and thyroid hormone levels in fetal blood through cordocentesis [27]. A team approach, including obstetrician, experienced ultrasonographer, neonatologist, and endocrinologist, is mandatory to monitor fetal development properly in the last trimester, parturition, the neonate thyroid status, and breastfeeding, as well as the risk of postpartum maternal exacerbation of hyperthyroidism. Neonatal hyperthyroidism, although self-limited, may be immediately fatal if unrecognized or poorly managed. Antithyroid drugs, β-blocker, and supportive measures should be started even before the exacerbation of thyrotoxicosis that follows clearance of the maternally transferred antithyroid drugs.

Final Note: Patients' Satisfaction Appraisal

In the context of the availability of three therapeutic strategies, and of the increasing patients' involvement into treatment selection [6], current studies are also devoted to the patients' perception of medical management, satisfaction with the treatment, and overall quality of life. One of the more recent studies in the domain, performed in patients encouraged to participate, following physician recommendation, to the choice of the method of treatment, has shown that the majority of them were satisfied with the treatment received, that the modality of treatment had little impact and, significantly, that patient's preference for a specific treatment may be more important than the treatment type *per se* [28].

Conclusion

It may seem that no significant improvement has been achieved in recent years in the treatment of Graves' disease. Currently, however, optimization of existing therapeutic strategies may offer every patient the most appropriate management strategy. The cost-effectiveness of the treatment and the socioprofessional and psychological aspects of the disease are of increasing importance and must also be taken into account. Ongoing clinical investigations, as well as immunological research, aim at improving the understanding and management efficiency of Graves' disease. There is no doubt that, in the future, new immuno-specific therapeutic approaches will become available.

REFERENCES

1. Burch HB, Burman KD, Cooper DS. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab*, 2012; **97**: 4549–58.
2. Brito JP, Schilz S, Singh Ospina N, *et al.* Antithyroid drugs-the most common treatment for Graves' disease in the United States: a nationwide population-based study. *Thyroid*, 2016; **26**: 1144–5.
3. Kahaly G, Bartalena L, Hegedüs, Leenhardt L, Poppe K, Pearce SH. European Thyroid Association Guideline for the management of Graves' hyperthyroidism. *Eur Thyroid J*, 2018; **7**: 1–20.
4. Burch HB, Cooper DS. Management of Graves' disease: a review. *JAMA*, 2015; **314**: 2544–54.
5. Donovan PJ, McLeod DS, Little R, Gordon L. Cost-utility analysis comparing radioactive iodine, anti-thyroid drugs and total thyroidectomy for primary treatment of Graves' disease. *Eur J Endocrinol*, 2016; **175**: 595–603.
6. Brito JP, Castaneda-Guarderas A, Gionfriddo MR *et al.* Development and pilot testing of an encounter tool for shared decision making about the treatment of Graves' disease. *Thyroid*, 2016; **25**: 1191–8.
7. Vos XG, Endert E, Zwinderman AH, Tijssen JG, Wiersinga W. Predicting the risk of recurrence before the start of antithyroid drug therapy in patients with Graves' hyperthyroidism. *J Clin Endocrinol Metab*, 2016; **101**, 1381–9.
8. Struja T, Kaeslin M, Boesiger F, *et al.* External validation of the GREAT score to predict relapse risk 1 in Graves' disease. Results from a multicentre, retrospective study with 741 patients. *Eur J Endocrinol*, 2017; **176**: 413–19.
9. Kim YA1, Cho SW, Choi HS, *et al.* The second antithyroid drug treatment is effective in relapsed Graves' disease patients: a median 11-year follow-up study. *Thyroid*, 2017; **27**: 491–6.
10. Azizi F, Ataie L, Hedayati M, Mehrabi Y, Sheikholeslami F. Effect of long-term continuous methimazole treatment of hyperthyroidism: comparison with radioiodine. *Eur J Endocrinol*, 2005; **152**: 695–701.
11. Boelaert K, Syed AA, Manji N, *et al.* Prediction of cure and risk of hypothyroidism in patients receiving 131I for hyperthyroidism. *Clin Endocrinol (Oxf)*, 2009; **70**: 129–38.
12. Walter MA, Briel M, Christ-Crain M, *et al.* Effects of antithyroid drugs on radioiodine treatment: systematic review and meta-analysis of randomised controlled trials. *BMJ*, 2007; **334**: 514.
13. Laurberg P, Wallin G, Tallstedt L, Abraham-Nordling M, Lundell G, Tørring O. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*, 2008; **158**: 69–75.
14. Papi G, Corsello SM, Pontecorvi A. Clinical concepts on thyroid emergencies. *Front Endocrinol (Lausanne)*, 2014; **5**: 102.
15. Warnock AL, Cooper DS, Burch HB. Life-threatening emergencies: thyroid storm and effects adverse effects of antithyroid drugs. In: Matfin G, ed. *Endocrine and Metabolic Medical Emergencies: Thyroid Disorders*. Washington, DC: Endocrine Society/Endocrine Press, 2014: 110–26.
16. Liu ZW, Masterson L, Fish B, Jani P, Chatterjee K. Thyroid surgery for Graves' disease and Graves' ophthalmopathy. *Cochrane Database Syst Rev*, 2015; **25**: CD010576.

17. Rivkees SA, Szarfman A. Dissimilar hepatotoxicity profiles of propylthiouracil and methimazole in children. *J Clin Endocrinol Metab*, 2010; **95**: 3260–7.
18. Léger J1, Gelwane G, Kaguelidou F, *et al*. Positive impact of long-term antithyroid drug treatment on the outcome of children with Graves' disease: national long-term cohort study. *J Clin Endocrinol Metab*, 2012; **97**: 110–19.
19. Rivkees SA. Pediatric Graves' disease: management in the post propylthiouracil era. *Int J Pediatr Endocrinol*, 2014; **2014**: 10.
20. Breuer CK, Solomon D, Donovan P, Rivkees SA, Udelsman R. Effect of patient age on surgical outcomes for Graves' disease: a case–control study of 100 consecutive patients at a high volume thyroid surgical center. *Int J Pediatr Endocrinol*, 2013; **2013**: 1.
21. Tanda ML, Piantanida E, Liparulo L, *et al*. Prevalence and natural history of Graves' orbitopathy in a large series of patients with newly diagnosed Graves' hyperthyroidism seen at a single center. *J Clin Endocrinol Metab*, 2013; **98**: 1443–9.
22. Wiersinga WM, Žarković M, Bartalena L, *et al*. Predictive score for the development or progression of Graves' Orbitopathy in patients with newly diagnosed Graves' Hyperthyroidism. *Eur J Endocrinol*, 2018; **178**: 635–43.
23. Bartalena L. The dilemma of how to manage Graves' hyperthyroidism in patients with associated orbitopathy. *J Clin Endocrinol Metab*, 2011; **96**: 592–9.
24. Laurberg P, Berman DC, Andersen S, Bülow Pedersen I. Sustained control of Graves' hyperthyroidism during long-term low-dose antithyroid drug therapy of patients with severe Graves' orbitopathy. *Thyroid*, 2011; **21**: 951–6.
25. Laurberg P, Bournaud C, Karmisholt J, Orgiazzi J. Management of Graves' hyperthyroidism in pregnancy: focus on both maternal and foetal thyroid function, and caution against surgical thyroidectomy in pregnancy. *Eur J Endocrinol*, 2009; **160**: 1–8.
26. Andersen SL1, Olsen J, Wu CS, Laurberg P. Birth defects after early pregnancy use of antithyroid drugs: a Danish nationwide study. *J Clin Endocrinol Metab*, 2013; **98**: 4373–81.
27. Abalovich M, Amino N, Barbour LA, *et al*. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2007; **92** (Suppl): S1–47.
28. Conaglen HM, Tamatea JAU, Conaglen JV, Elston MS. Treatment choice, satisfaction and quality of life in patients with Graves' disease. *Clin Endocrinol*, 2018; **88**: 977–84.

3.3.10 Graves' Orbitopathy and Dermopathy

Wilmar M. Wiersinga

Clinical Presentation

The many and often disfiguring features of a typical patient with Graves' ophthalmopathy (GO) are obvious at first glance (**Figure 3.3.10.1**). The changed appearance has a profound effect on the emotional and social status of the patient. The various signs and symptoms can be described according to the NO SPECS classification [1] (**Table 3.3.10.1**). Class 1 signs can be present in any patient with thyrotoxicosis regardless of its cause. Upper eyelid retraction causes stare and lid lag on downward gaze (the latter is the well-known von Graefe's sign). Soft tissue involvement (class 2) comprises swelling and redness of eyelids, conjunctiva, and caruncle. Symptoms are a gritty sandy sensation in the eyes, retrobulbar pressure, lacrimation, photophobia, and blurring of vision. Proptosis (class 3) can be quite marked. Upper eyelid retraction by itself may already give the impression of exophthalmos. Extraocular muscle involvement (class 4) may result in aberrant position of the globe, or fixation of the globe in extreme cases. More common is limitation of eye muscle movements in certain directions of gaze, especially in upward gaze; it is usually associated with diplopia. Elevation is most

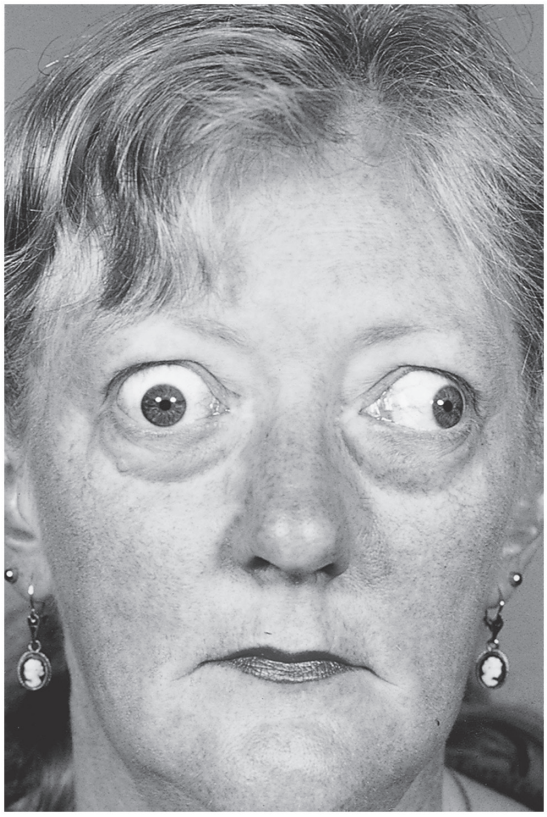


Figure 3.3.10.1 Bilateral eye disease due to Graves' orbitopathy. Note lid retraction, stare, periorbital swelling, marked proptosis, and exotropia of the left globe (see also Plate 15).

Table 3.3.10.1 Assessment of disease severity in the patient with Graves' orbitopathy (GO)

Severity measures of GO (using the mnemonic NO SPECS)		
NO SPECS class	**Item**	**Method**
0. **N**o signs or symptoms		
1. **O**nly signs, no symptoms	Lid aperture	With ruler in midline in mm
2. **S**oft tissue involvement	Eyelid and conjunctiva swelling and redness	Inspection, comparison with colour pictures[#]
3. **P**roptosis	Exophthalmos	Hertel in mm[‡]
4. **E**xtraocular muscle involvement	Eye muscle motility Diplopia	Impaired elevation, abduction, Depression, and adduction Subjective grading*
5. **C**orneal involvement	Keratitis, ulcer	Fluorescein dye
6. **S**ight loss due to optic nerve involvement	Dysthyroid optic neuropathy (DON)	Visual acuity, colour vision, visual fields, optic disc

[#] colour atlas Dickinson *et al. Clin Endocrinol*, 2001; 55: 283–303
* intermittent diplopia = at awakening or when tired; inconstant diplopia = at extremes of gaze; constant diplopia = in primary or reading position
[‡] upper normal limit 16/17 mm in Asians, 19/21 mm in whites, and 23/24 mm in African Americans (females/males, respectively).

frequently impaired, followed by abduction and depression; adduction is rarely affected. Diplopia will not occur if the vision of one eye is very low (e.g. in amblyopia), or if the impairment of eye muscle motility is strictly symmetrical. Patients may correct for double vision by tilting the head, usually backwards and sideways; the ocular torticollis often leads to neck pain and headache. Positioning of the head in the default position will unmask diplopia. Corneal involvement (class 5) occurs through overexposure of the cornea due to lid lag, lid retraction, and exophthalmos, easily leading to dry eyes and keratitis. Lagophthalmos is often noted first by the patient's partner because of incomplete closure of the eyelids during sleep. Sight loss (class 6) due to optic nerve involvement is the most serious feature, often referred to as dysthyroid optic neuropathy (DON) [2]. Besides the decrease of visual acuity, there may be loss of colour vision and visual field defects. Visual blurring may disappear after blinking (caused by alteration of the tear film on the surface of the cornea due to lacrimation or dry eyes) or after closing one eye (attributable to eye muscle imbalance). Visual blurring that persists is of great concern as it may indicate optic neuropathy.

The frequency of the various eye changes among GO patients is as follows [3]: von Graefe's sign 46%, upper eyelid swelling 62%, proptosis of 21 mm or higher 53%, diplopia 32%, corneal involvement 14%, and optic nerve involvement 3.9%. Predisposing factors for DON are male sex, old age, smoking, and diabetes mellitus [2] (Figure 3.3.10.2). Unilateral GO is observed in about 10%. Eye changes are similar to those in bilateral GO, but unilateral cases are more often euthyroid [4]. Extraocular muscle enlargement in the fellow eye can be detected by CT in about one-half of cases. Progression to bilateral GO is common. Unilateral GO may thus represent an early stage of the disease that already is or will develop shortly into a bilateral disease. Unknown local factors must be involved in the unilateral expression of GO, which essentially is a bilateral and fairly symmetrical eye disease.

Mean age at GO presentation is 48 years, 23% are males, and 40% are current smokers [3]. Diabetes is present in 4%, and glaucoma in 5.6%. Male sex, older age, and smoking are associated with more severe ophthalmopathy. Childhood GO is rare [5]. Clinical manifestations are less severe in paediatric patients: exophthalmos is seen in 75%, but impaired muscle motility only in 11% [6].

Figure 3.3.10.2 Unusual presentation of Graves' orbitopathy as unilateral eye disease. Male sex, advanced age, and heavy smoking all predisposed this patient to the development of severe eye disease; note the absence of exophthalmos in this case of dysthyroid optic neuropathy (see also Plate 16).

Epidemiology

A population-based study in Sweden observed an incidence rate of Graves' hyperthyroidism of 210/million/year and of GO of 42/million/year [7]. Thus, 20% of Graves' hyperthyroid patients have also GO, which is mild in 15% and severe in 5%. Recent

studies in Denmark and Italy confirm that moderate-to-severe GO is present in about 5% of Graves' hyperthyroid patients [8, 9]. GO develops in a few patients before the onset of Graves' hyperthyroidism, but simultaneous onset of GO and Graves' hyperthyroidism occurs in the majority of GO patients. GO sometimes develops after Graves' hyperthyroidism [9, 10]; risk factors are minimal redness or swelling of eyelids or conjunctiva, high TSH-receptor antibodies, long duration of hyperthyroid symptoms, and current smoking [10]. GO occurs in 3% of cases in patients with autoimmune hypothyroidism and in 7–10% in euthyroid patients [3]. The euthyroid and primarily hypothyroid patients have milder and more asymmetrical GO than the hyperthyroid GO patients [11, 12]. It is not unusual that hypothyroid GO patients proceed to Graves' hyperthyroidism, linked to a shift from TSH receptor blocking to stimulating antibodies. Euthyroid GO patients develop hyperthyroidism in due time in about 20%, but it is unknown why others remain euthyroid despite the presence of TSH receptor antibodies.

Smoking greatly increases the risk for GO (odds ratio 7.7, 95% CI 4.3 to 13.7) [13]. A secular trend to a lower incidence rate of GO is reported. The proportion of patients with GO among all referred patients with Graves' hyperthyroidism decreased from 57% in 1960 to 35% in 1990, and to 29% in 2010 [9, 14]. In a European questionnaire study in 1998, 43% of respondents thought GO was decreasing in frequency, in rough agreement with the decline in smoking [15]. It looks the prevalence of severe GO is also declining: DON was present in 30% in 1960 and in 21% in 1990 in a single centre [14], and in 21% in 2000 and in 4% in 2012 in European Group on Graves' Orbitopathy (EUGOGO) centres [3]. GO patients referred in 2012 had less severe and less active GO than those referred in 2000 [3]. Earlier diagnosis and treatment of hyperthyroidism, identification of risks conferred by ^{131}I therapy and postradioiodine hypothyroidism, and focus on the detrimental effects of smoking have likely contributed to this secular decline in the incidence and severity of GO.

Pathogenesis

Mechanistic Explanation of Eye Changes

GO is characterized by enlarged extraocular muscles and increased orbital fat. The increase in tissue volume and the associated rise of retrobulbar pressure can explain the various signs and symptoms. The swollen retrobulbar tissues impair the venous drainage of eyelids and conjunctiva, resulting in oedematous swelling of eyelids and chemosis. Upper and lower eyelid swelling can also be caused by herniation of retrobulbar fat through openings in the orbital septum. The only other outlet for the increased orbital content, in view of the confinement within the bony walls of the orbit, is pushing forward the eyeball, resulting in exophthalmos. The average volume of the orbital cavity is approximately 27 and 24 ml in healthy males and females respectively; corresponding values of extraocular muscle volume are 4.2 and 3.7 ml, and of orbital fat volume 16.1 ml and 14.0 ml [16]. It has been calculated that an increase of 4 ml in muscle or fat will cause a proptosis of 6 mm.

The enlargement of extraocular muscles impairs muscle relaxation, not the ability for muscle contraction. Limited motion of eye muscles is due to impaired relaxation of the antagonist upon contraction of the agonist. Impaired elevation is thus primarily the result of insufficient relaxation of the rectus inferior muscle. Restricted eye muscle motility may cause diplopia. Upper eyelid retraction (due either to an increased adrenergic activity in hyperthyroidism or to swelling of the levator muscle) and proptosis contribute to overexposure of the cornea, which may become dry and inflamed.

Marked swelling of the extraocular muscles in the apex of the orbit (known as apical crowding), close to the entrance of the optic nerve in the optic canal, may damage the optic nerve either via direct pressure or via impairment of the blood supply to the nerve [17]. The resulting DON causes loss of visual functions. The degree of proptosis in patients with optic neuropathy, despite a greater mass of extraocular muscles, is less than in patients without optic neuropathy [2]. This is remarkable because the retrobulbar pressure (in the normal orbit 3.0–4.5 mmHg) is greatly elevated in patients with DON up to values between 17 and 40 mmHg, much higher than the pressure of 9–11 mmHg measured in orbits of patients with exophthalmos but no optic neuropathy [18]. A well-developed tight orbital septum might preclude proptosis, resulting in a very high retrobulbar pressure and DON.

Immunopathogenesis

Orbital fibroblasts (OF) are considered the target cells of the autoimmune attack in GO. For, retrobulbar T cells from GO patients recognize autologous OF (but not eye muscle extracts) in a major histocompatibility complex (MHC) class I restricted manner, and proliferate in response to autologous proteins from OF (but not from orbital myoblasts). Conversely, OF proliferate in response to autologous T cells dependent on MHC class II and CD40-CD40L signalling [19]. OF are stimulated to produce excessive amounts of glycosaminoglycans (GAG), notably hyaluronic acid. GAGs are very hydrophilic compounds and thus attract much water, resulting in oedematous swelling. GAGs accumulate in the endomysial space between muscle fibres. There is no increase in the number of muscle fibres and no ultrastructural damage to the muscle cells themselves (except in very advanced cases when some damage may be seen). An increased number of OF is found in the endomysial space and in the orbital fat. OF not expressing Thy-1 (present in orbital fat but not in extraocular muscles) may differentiate into mature adipocytes, thereby further contributing to volume expansion. Volumetric studies have shown that most GO patients have increased muscle volume at diagnosis, whereas fat expansion seems a rather late phenomenon [20, 21]. Why some patients present with exophthalmos and enlarged orbital fat volume but do not develop increased muscle volume, remains uncomprehended.

Orbital immunocompetent cells consist of T helper ($T_{h)}$ cells, suppressor/cytotoxic T cells, many macrophages, and relatively few B cells. Many of these cells are activated memory cells ($CD45RO^+$), frequently located adjacent to blood vessels. The infiltrating immunocompetent cells produce cytokines capable of remodelling orbital tissues. The cytokine profile in the early stages is predominantly derived from T_h1 cells, whereas cytokines are mostly derived from T_h2 cells in patients with a duration of GO >2 years [22]. The data suggest GO is primarily a T-cell mediated disease.

The *cytokines* induce expression of immunomodulatory proteins on orbital endothelial cells and fibroblasts (such as HLA-DR,

heat shock protein 72 and several adhesion molecules), generating T-cell migration. Cytokine-activated OF synthesize chemoattractants IL-16 and RANTES, perpetuating the immune attack. Macrophages may present antigen to T cells (CD40L) through provision of costimulatory signals and proinflammatory cytokines. Activated T cells bind to $CD40^+$ fibroblasts, inducing proinflammatory compounds (like cytokines, COX2, and PGE2) and excessive production of GAGs. A subset of OF may differentiate into mature adipocytes, which is associated with increased expression of TSH receptors [23]. This brings us to the nature of the autoantigen in the orbit.

The TSH receptor is presently viewed as the major autoantigen in GO. Full-length functional TSH receptors are expressed on OF, in active stages of the disease to a greater extent than in the inactive stages, and directly related to IL-1β [24]. Graves' immunoglobulins and M22 (a TSH receptor stimulating monoclonal antibody) recognize TSH receptors on OF as evident from increased cAMP and hyaluronan production in cell cultures of differentiated human OF [19]. Clinical studies support the role of TSH receptors. The serum concentrations of TSH receptor antibodies (TSHR-Ab) are higher in Graves' patients with GO than in Graves' patients without GO [25], and are directly related to the severity and activity of GO (whereas thyroid peroxidase and thyroglobulin antibodies are not) [26, 27]. The higher the level of TSHR-Ab, the higher the risk of an unfavourable course of the eye changes [28]. The level of TSHR-Ab fall after thyroidectomy or treatment with antithyroid drugs, but increase by 100% in the first six months after radioactive iodine therapy coinciding with development or worsening of GO after ^{131}I therapy in about 15% [29, 30]. Finally, genetic immunization against the TSH receptor A-subunit plasmid have produced the first experimental animal model of GO [31–35]; within the inherent restrictions of the model (the mouse has no bony orbit), the model mimics rather well the human condition.

Another autoantigen might be the IGF-1 receptor (IGF-1R). IGF-1R are indeed upregulated on OF GO patients, but serum IGF-1 and IGF-binding proteins in GO are normal. The stimulating effect of Graves' IgG on hyaluronan secretion by cultured OF is attenuated by IGF-1R blocking antibodies, suggesting Graves IgG could contain IGF-1R stimulating antibodies [36]. IGF-1R-Ab were found in 14% of GO patients and in 11% of controls [37]. The presence of IGF-1R-Ab was not related to activity or severity of GO nor to TSHR-Ab. IGF-1R-Ab failed to stimulate IGF-1R autophosphorylation, and inhibited IGF-1 induced signalling. Thus, there is no evidence for the existence of IGF-1R stimulating antibodies in GO patients. Further studies suggest crosstalk between TSH receptors and IGF-1 receptors [38, 39]. TSH and IGF-1 synergistically increase hyaluronan secretion by OF; the effect of TSHR stimulating antibodies can be blocked only partially by IGF-1R antagonists but completely by TSHR antagonists [40]. Serum immunoglobulins isolated from GO patients are unable to autophosphorylate IGF-1R, and do not directly activate IGF-1R [41, 42]. The observed crosstalks might be explained from overlap between post-receptor pathways: TSHR signalling via PLC/PI-3K/Akt and IGF-1R signalling via PI-3K/Akt result in the same end-product (**Figure 3.3.10.3**) [43]. Crosstalk and the putative role of

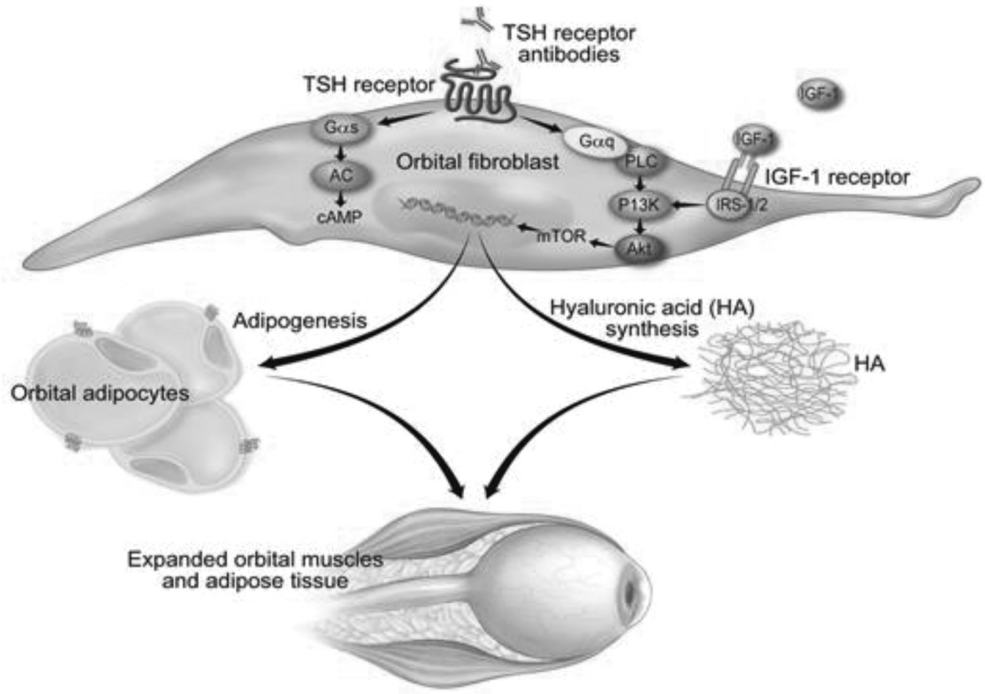


Figure 3.3.10.3 Role of the TSH receptor in the immunopathogenesis of GO. Ligation of the TSH receptor on orbital fibroblasts with TSHR-Ab activates post-receptor signalling via the adenylyl cyclase/cAMP pathway and the phospholipase C/ phosphoinositide-3-kinase/Akt pathway. The latter overlaps with the post-receptor signalling pathway of the IGF-1 receptor. The end result is excessive secretion of hyaluronic acid by OF, and differentiation of a subset of into mature adipocytes (adipogenesis), causing enlarged extraocular muscles and increased orbital fat.

various classes of TSHR-Ab is an area of active research. The IGF-1R is thus not a primary autoantigen in GO, as evident also from the failure of genetic immunization with IGF-1R to produce an animal model of GO [31].

Fibrocytes are bone-marrow derived cells from the monocyte lineage, expressing CD45, CD34, CXCR4, collagen 1, functional TSHR, and Tg. Circulating fibrocytes are highly abundant in GO patients, and seem to infiltrate orbital connective tissue where they might transition to CD34+ OF [44]. Teprotumumab, a human monoclonal IGF-1R blocking antibody, attenuates the actions of both IGF-1 and TSH in fibrocytes; specifically, it blocks the induction of proinflammatory cytokines by TSH [45].

Many questions remain unanswered. It is difficult to explain why most patients with Graves' hyperthyroidism, despite high titres of TSHR-Ab, do not develop GO. This could be related to genetic and environmental factors. GO is more prevalent in white patients than in Asian patients [46]. There is, however, no difference in the frequency of particular polymorphisms in susceptibility genes for Graves' disease (*HLA, CTLA4, PTPN22, CD40, FRCL3, TSHR*) between Graves' hyperthyroid patients with and without GO [47]. Smoking increases greatly the risk of GO [13]. Exposure of OF *in vitro* to cigarette smoke extract dose-dependently increases adipogenesis and hyaluronan production [48].

Natural History

GO has a tendency towards spontaneous improvement. There have been few studies on the natural history of the eye disease. The most extensive ones were carried out in the 1940s and 1950s by Rundle [49] (Figure 3.3.10.4). He describes a stage of ingravescence,

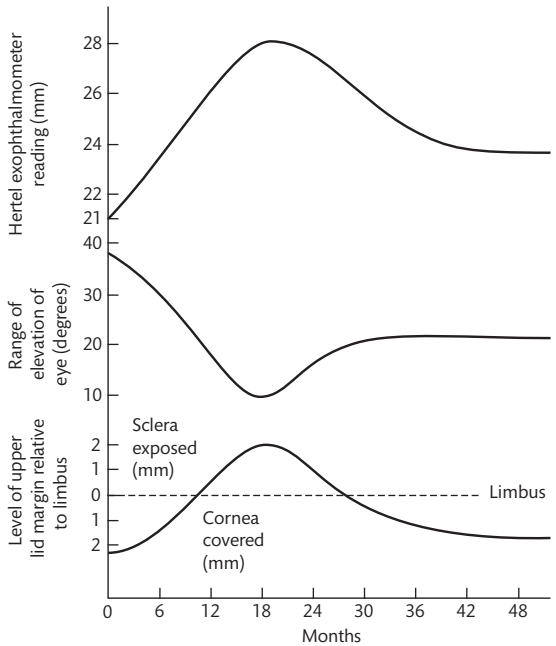


Figure 3.3.10.4 Rundle's curves depicting the natural history of GO, characterized by an initial dynamic phase of ingravescence and remission, followed by the static endstage.

Reproduced with permission from Rundle FF. Management of exophthalmos and related ocular changes in Graves' disease. *Metabolism*, 1957; 6: 36–47. with permission from Elsevier.

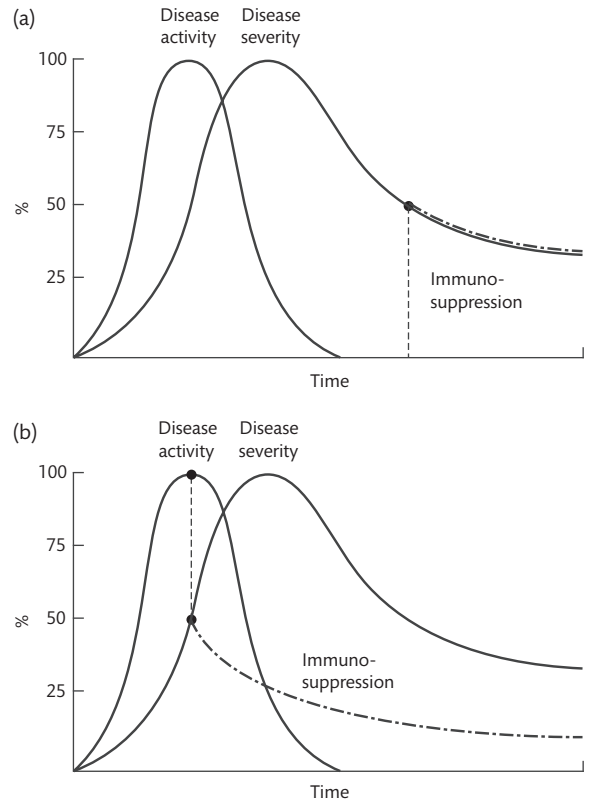


Figure 3.3.10.5 Schematic outcome of immunosuppressive treatment in GO patients as a function of disease activity. The natural history of GO activity and severity is depicted by two separate curves. Immunosuppression is administered at one-half maximal disease severity. The response to treatment is zero when given at zero disease activity like in (a), but a substantial response can be expected when given at peak disease activity like in panel (b).

Reproduced with permission from Wiersinga WM. Advances in medical therapy of thyroid-associated ophthalmopathy. *Orbit*, 1996; 15: 177–86. Copyright 1996 Taylor & Francis.

characterized by the development of exophthalmos (by 0.5 mm monthly, up to an average extent of 2–5 mm) and limitation of elevation; 4–5 degrees elevation is lost for each millimetre of protrusion. Thereafter a stage of remission occurs, which is slower and less complete than ingravescence. Recovery from restricted eye muscle motility precedes that from proptosis. This dynamic phase is succeeded by a static phase, in which exophthalmos and eye muscle disturbance remain unchanged in 75% of patients. The time period in which the stable endstage is reached varies considerably between patients, ranging from several months up to 5 years. Recent studies confirm these early observations: during a 1-year follow-up in patients whose ophthalmopathy did not require immediate treatment, substantial improvement occurred in 22%, slight improvement occurred in 42%, the disease remained stable in 22%, and the disease progressed in 14% [50].

The few histological studies support Rundle's observations. In the early active stage of the disease there is a lymphocytic infiltrate, oedema, and activated fibroblasts; in the end stages there is fibrosis. The data imply that the natural history of GO can also be described according to the activity of the eye disease, next to Rundle's curves on the severity of the eye disease [51] (Figure 3.3.10.5). Assessment of the activity of GO may influence the management plan. Immunosuppression is unlikely to be effective when given

in the fibrotic inactive endstage of the disease, but might be of much benefit in the early active stage with ongoing inflammation. Likewise, the results of eye muscle and lid surgery might be lost when performed in the active stage.

Diagnosis and Differential Diagnosis

The initial work-up is aimed not only to reach a correct diagnosis of GO but also to allow the delineation of a personalized treatment plan [52]. It should contain the following four assessments:

Assessment of Thyroid Function

Restoration and maintenance of a normal thyroid function is relevant because the eye changes are more severe in patients who still have an abnormal thyroid function [53]. Serum concentrations of TSHR-Ab are usually high in GO, and their presence support the diagnosis of GO (quite relevant in patients presenting without hyperthyroidism).

Assessment of GO Severity

The NO SPECS system is a useful memory aid in the examination of disease severity (Table 3.3.10.1). Quantitative measurements are preferred whenever possible. Lid aperture is measured with the ruler centred on the pupil while the facial muscles are relaxed and gaze is directed straight ahead. The degree of soft tissue swelling is difficult to assess objectively, but using a comparative photographic colour atlas is very helpful in this respect [54]. Significant exophthalmos may develop without proptosis readings exceeding the upper normal limit. Comparison with pictures of the patient before the onset of GO are helpful in this respect. Motility disturbances are recognized by asking the patient to move the eyes upwards, downwards, and from side to side, observing impaired elevation, depression, abduction, or adduction. Diplopia is also graded subjectively. More objective measurements can be done by ophthalmologists; most relevant is the field of single binocular vision [55]. Stippling of the cornea can be seen by split lamp and application of dyes; ulceration, clouding, necrosis, and perforation occur in severe cases. Visual acuity is measured using the Snellen chart; the use of a pinhole corrects for refraction disorders and sight loss due to keratitis. Further investigation is warranted if optic nerve involvement is suspected (e.g. in patients complaining of persistent blurred vision or greyish vision). In patients with DON, decreased visual acuity is present in 80%, reduced colour vision in 77%, visual field defects in 71%, afferent pupillary defect (Marcus Gunn's phenomenon) in 45%, optic disc oedema in 56%, and disc pallor in 4% [2]. Choroidal folds, caused by impression of the globe by enlarged retrobulbar tissues, are rare.

Assessment of GO Activity

In some patients, such as the one depicted in Figure 3.3.10.2, one glance is sufficient to conclude the eye disease is in the active phase. However, if it is not self-evident whether GO is active or inactive, an assessment of disease activity by the CAS (clinical activity score) provides guidance in selecting the most appropriate treatment (Figure 3.3.10.6). The CAS is based on the classical signs of inflammation (Table 3.3.10.2). A CAS of <3 is compatible with inactive GO, and CAS ≥3 has fair predictive value for improvement of GO upon immunosuppressive treatment [56]. There have been other

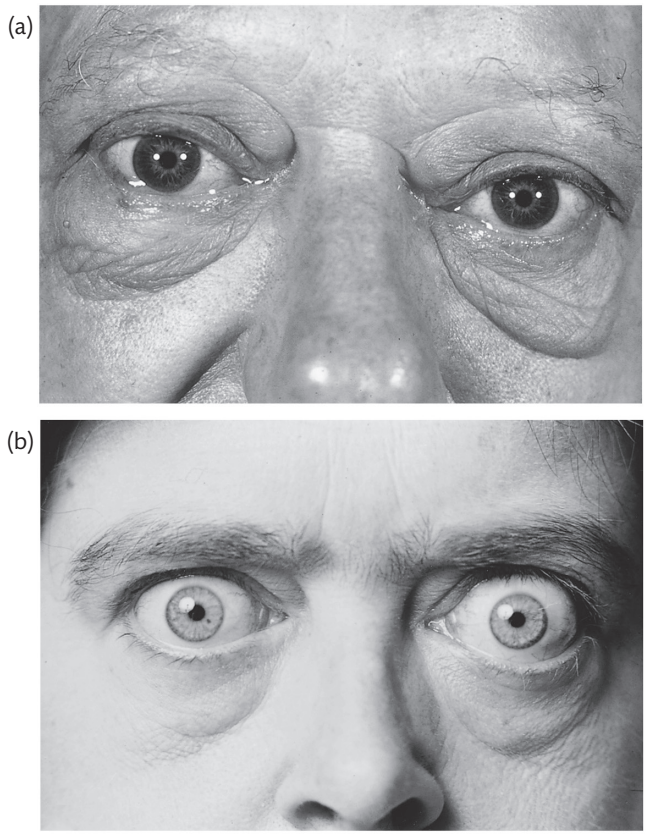


Figure 3.3.10.6 Active versus inactive GO. (a) Note periorbital swelling caused by oedema, redness of eyelids, redness of conjunctiva, and chemosis in a patient with active eye disease. (b) Periorbital swelling in a patient with inactive eye disease due to fat prolapse through the orbital septum and/or fibrotic degeneration; redness and chemosis are absent.

methods for assessing GO activity, like orbital A-mode echography measuring echogenicity in extraocular muscles, orbital ^{111}In-labelled octreotide scintigraphy measuring orbital/occipital skull uptake ratio, and orbital MRI measuring (e.g. T_2-relaxation time). Each of these methods have their own specific drawbacks, and have not become popular although their predictive values for the outcome of immunosuppression are reasonable. A short duration of GO (less than 18 months) is likely to be associated with still active disease [57]. Maybe the combination of GO duration, CAS, and orbital MRI provides the most accurate method for assessing disease activity.

Assessment of Quality of Life

Eye changes adversely affect a patient's self-image and daily functioning. The overall health-related quality of life (QoL) of patients with moderate-to-severe GO is lower than for patients with other chronic conditions such as diabetes mellitus, emphysema, or heart failure [58]. A disease-specific QoL questionnaire has been developed, the GO-QoL, which has been thoroughly validated [59, 60]. The GO-QoL contains eight questions about problems with visual functioning and eight questions about the psychosocial consequences of changed appearance (Table 3.3.10.3). The answers are summarized to one score for visual functioning and one score for appearance. The GO-QoL might be useful not only in evaluation of treatment, but also in reconciling priorities of the patient and of the physician in delineating a management plan. The psychosocial burden imposed by GO is considerable, and support by mental

Table 3.3.10.2 Assessment of disease activity in the patient with Graves' orbitopathy (GO)

Activity measures of GO (using the clinical activity score, CAS)		
Inflammatory sign	**Item**	**Score**
Pain (dolor)	Spontaneous retrobulbar pain Pain on up-, side-, or downgaze	1 1
Redness (rubor)	Redness of the eyelids Redness of the conjunctiva	1 1
Swelling (tumour)	Swelling of the eyelids Swelling of the caruncle and/or plica Chemosis	1 1 1
Maximum CAS score (assessed momently)		7
Impaired function (functio leasa)	Increase in proptosis ≥2 mm in 1–3 months Decrease of ≥8^0 in eye muscle motility in any direction in 1–3 months Decrease in visual acuity of >1 line on the Snellen chart (using pinhole) in 1–3 months	1 1 1
Maximum CAS score (assessed over time)		10

healthcare professionals might be needed. The sequelae of GO are indeed substantial: 45% of patients are restricted in their daily activities, 36% are on sick leave, 28% are disabled, 5% went into early retirement, and 3% lost their jobs, all because of GO [61]. The recent ETA/EUGOGO guidelines for the management of GO state: 'We recommend a patient-focussed approach of patients with GO, which encompasses the effects of disease and its treatment on QoL and psychosocial well-being. Use of GO-QoL, a disease-specific and well validated tool, is recommended in routine clinical practice' [62]. The GO-QoL is available in 18 languages and can be downloaded for free from the EUGOGO website (https://www.eugogo.eu).

Orbital Imaging

The degree of swelling of extraocular muscles and orbital fat can be evaluated using CT scans or MRI. Coronal sections are preferred in view of the pear-shaped orbit with its axis directing backwards and medially (Figure 3.3.10.7) [2]. The bony structures are best evaluated by CT scan and are of relevance in case of surgical decompression. MRI has the advantage of providing an activity parameter in addition to imaging. The muscles in GO are swollen typically at the belly, leaving the tendons uninvolved. For unknown reasons the inferior and medial rectal muscles are most frequently enlarged, followed by the superior rectus; the lateral rectus muscle is least affected. The extraocular muscles originate in Zinn's annulus, which surrounds the optic canal and thus the optic nerve. Muscle swelling at this location (apical crowding) and intracranial fat prolapse are risk factors for optic neuropathy [2]. Orbital imaging is not necessary in most GO patients. Indications for orbital imaging are: (1) suspicion on DON (apical crowding?); (2) unilateral GO (GO is unilateral in about 10% of cases, but GO is the most prevalent cause (15%) of unilateral exophthalmos); (3) euthyroid or hypothyroid GO; (4) before surgical orbital

Table 3.3.10.3 Assessment of quality of life in the patient with Graves' orbitopathy (GO)

Quality of life measures of GO (using the disease-specific GO-QoL)	
Visual functioning During the past week, because of GO, to what extent were you limited in …?	*Appearance* During the past week, because of GO, do you feel that …?
1. bicycling (cannot bicycle ☐)?	1. your appearance has changed?
2. driving (no driver's licence ☐)?	2. you are stared at in the streets?
3. moving around the house?	3. people react unpleasantly?
4. walking outdoors?	4. it influences your self-confidence?
5. reading?	5. you are socially isolated?
6. watching television?	6. it influences on making friends?
7. hobby or pastime, i.e. ……..?	7. you appear less often in photos?
8. something you wanted to do?	8. you mask changes in appearance?
Tick the box that matches each answer ☐ yes, seriously limited → score 1 ☐ yes, a little limited → score 2 ☐ no, not at all limited → score 3 Range raw score 8–24 Total score = (raw score−8)/16 × 100 Higher score = better QoL	Tick the box that matches each answer ☐ yes, very much so → score 1 ☐ yes, a little → score 2 ☐ no, not at all → score 3 Range raw score 8–24 Total score = (raw score−8)/16 × 100 Higher score = better QoL

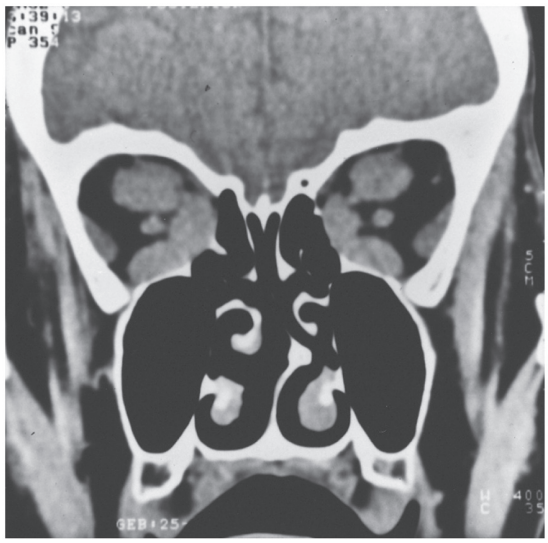


Figure 3.3.10.7 Coronal section of an orbital CT scan, showing enlarged inferior, medial, and superior rectus muscles but no apical crowding. Effacement of the perineural fat surrounding the optic nerve over more than 50% of its circumference puts the patient at risk for optic neuropathy.

decompression (bony structures are best evaluated by CT); (5) uncertainty about the diagnosis of GO.

Differential Diagnosis

None of the eye signs is pathognomonic for GO. Lid retraction can be due to non-Graves' thyrotoxicosis, contralateral ptosis, or cocaine use. Diplopia is common in myasthenia gravis, and bilateral proptosis can be caused by orbital fat accumulation (Cushing's syndrome, obesity), lithium therapy, liver cirrhosis, Wegner's granulomatosis, arteriovenous malformations, lymphoma, metastatic tumours, or severe myopia (pseudoproptosis).

Proptosis, motility disturbances, and optic nerve compression can be caused by the ill-defined disease entity of orbital pseudotumour, an idiopathic non-specific unilateral focal or diffuse fibroinflammatory orbital lesion [63]. The clinical presentation is characterized by acute or subacute signs of inflammation (pain, redness, oedema) and mass effects; imaging shows a focal or diffuse mass which is poorly demarcated. Orbital myositis is, after GO, the second most common cause of extraocular muscle enlargement, due to non-specific inflammation. The cardinal clinical feature is acute orbital pain exacerbating on eye movements. The disease is bilateral in 50% of patients. One or more muscles may be enlarged and, in contrast to GO, involve the muscle tendons. Orbital pseudotumour and orbital myositis respond quickly to corticosteroids, but recurrences occur in 50%. IgG4-related orbitopathy is a recently described condition characterized by lymphoplasmacytic infiltration and tissue fibrosis, which may mimic GO [64].

Management

Management of GO requires close consultation between the patient, the endocrinologist, and the eye physician, but in this respect notable differences in the delivery of care exist. Combined thyroid–eye clinics are most appropriate to delineate the treatment plan best suited for a particular patient. The timing and mode of thyroid treatment, immunosuppressive treatment, and surgical treatment should be coordinated in a multidisciplinary approach. The patient should be reassured of the possibilities for improvement of eye changes, both functionally and cosmetically, but at the same time be informed that it may require 1–2 years before full rehabilitation is reached. In this respect the experience of other patients with GO can be quite informative, and contact with patient self-help groups is very useful. It is recommended that all GO patients except mildest cases, are referred to specialized clinics or to combined thyroid/eye clinics [62]. Such a policy is likely to improve the quality of care delivered to GO patients [65]. Specific guidelines for the management of GO have been published [62] (Figure 3.3.10.8).

General Measures

About 40% of GO patients are current smokers [3]. The advice to stop smoking should be given repeatedly, and smoking cessation programmes should be offered. Patients should be confronted with the evidence that a/ smokers have more severe GO than non-smokers, b/ risk of worsening of GO after ^{131}I therapy is 4× greater in smokers than in non-smokers, c/ outcome of immunosuppression is less good in smokers than in non-smokers. Passive smoking may also be a risk for developing GO [66].

Simple measures that may be helpful in any stage of the disease are: (1) artificial tears to reduce surface symptoms and protect the epithelium; (2) sunglasses to reduce photophobia but also to comfort patients who are self-conscious of their appearance; (3) lubricant ointments to protect against exposure keratopathy during sleep; (4) prisms to improve diplopia. Botulinum toxin A injections may provide temporary control of eyelid retraction.

Thyroid Treatment

GO is more severe in patients who still have an abnormal thyroid function despite antithyroid treatment; it improves slightly after thyroid function has returned to normal [67]. Restoration and maintenance of euthyroidism is thus relevant for the eyes. Prolonged treatment with **antithyroid drugs** (ATD) (preferably in combination with thyroxine, the so-called block and replace regimen) until full rehabilitation of GO is obtained, is a feasible option. When GO has become inactive and needs no further treatment, ATD are discontinued; if Graves' hyperthyroidism recurs, it can be treated with ^{131}I without adverse effects on the eyes [68, 69]. A large longitudinal cohort study from Australia suggests preference for ATD: the odds for GO are 86% less in hyperthyroid patients using ATD than in those not using ATD (OR 0.14, CI 0.06–0.34) [70].

Iodine-131 therapy of Graves' hyperthyroidism is associated with a relative risk of 4.23 (CI 2.04–8.77) for progression of GO when compared to treatment with antithyroid drugs [30]. This conclusion is based on three large randomized clinical trials. Risk factors are pretreatment serum T3 >5 nmol/L, pre-existent active GO, high TSHR-Ab, and smoking. High TSH levels after ^{131}I therapy also constitute a risk for GO and should be avoided. The eye changes usually occur within 6 months after radioactive iodine, and are mostly transient and mild in nature. They can be prevented by steroids which should be applied in patients with one or more risk factors [71]. Uncertainty exists about the most appropriate dosage schedule: oral prednisone 0.2 mg/kg per day given for 6 weeks might be as effective as 0.3–0.5 mg/kg per day given for 3 months [62]. A causal relationship between ^{131}I therapy and the eye changes is plausible in view of

the radiation-induced release of thyroid antigens, T-cell activation, and immediate and significant rise in serum TSHR-Ab [29, 30].

Total thyroidectomy has been advocated in GO in order to remove all thyroid antigens and thyroid-directed T lymphocytes. This approach is logical assuming cross-reactivity between thyroid and orbital antigens is involved in the immunopathogenesis of GO, although orbital autoimmunity may proceed independently once the eye disease is well established. The implication is that thyroidectomy should be done early in the course of the disease. In a case–control study, development or worsening of eye changes after near-total thyroidectomy occurred in 3.3%, exactly the same as in carefully matched patients treated with ATD [72]. Another study found no difference in the course of GO between subtotal and total thyroidectomy [73]. A randomized trial in GO found total thyroid ablation (total thyroidectomy + ^{131}I ablation) associated with a slightly better outcome at 9 months than total thyroidectomy alone, but improvement was limited to lid aperture and proptosis and was of minor clinical relevance [74]. A longitudinal study using a large database of a US managed care network, observed that surgical thyroidectomy was associated with a 74% decreased hazard for GO (HR 0.26, CI 0.12–0.51) compared with ^{131}I therapy alone [75]. If confirmed, it would make total thyroidectomy an attractive option for treatment of Graves' hyperthyroidism when GO is very mild or still absent.

Eye Treatment

Management of eye changes in GO depends on disease severity and disease activity, as clearly shown in the algorithm of Figure 3.3.10.8 [62]. Mild GO is characterized by lid retraction <2 mm, mild soft tissue involvement, proptosis <3 mm above upper normal limit, absent or intermittent diplopia, absent corneal exposure, and normal optic nerve status. Moderate-to-severe GO is defined by lid retraction ≥2 mm, moderate-to-severe soft tissue involvement, proptosis ≥3 mm above upper normal limit, inconstant or constant diplopia, mild punctuate keratopathy, and normal optic nerve status, and very severe sight-threatening GO by corneal breakdown or DON. When evaluating GO severity and activity, it can be helpful to have a look at the QO-QoL results. It could facilitate discussion with the patient on the most disturbing features, and to identify patients in need of further counselling. Counselling may reduce anxiety, provide reassurance, and help in developing better coping strategies.

Treatment of Mild GO

Spontaneous improvement occurs in about 30% of patients with mild GO after 1 year. A 'wait-and-see' policy is therefore justified [62]. An alternative is treatment with sodium selenite 100 μg twice daily for 6 months: it improves eye manifestations (in 61% of Se and 36% of placebo treated patients) and QoL, and prevents progression to more severe GO (which occurred in 7% upon Se and in 26% upon placebo treatment) (Figure 3.3.10.9) [76]. Antioxidant actions of selenium in OF provide a basis for these favourable effects [77]. The results were obtained in European countries with relatively low selenium intake. It is unknown if selenium would have the same beneficial effect in countries like the United States which are selenium sufficient. Sometimes the eye changes despite being labelled as mild, have such a negative impact on a patient's life that more invasive intervention (e.g. steroids) is warranted. Retrobulbar irradiation in mild GO has a response rate of 50–60%, twice as high as that of sham irradiation, but it does not prevent progress to more severe ophthalmopathy that occurs in about 15% [78].

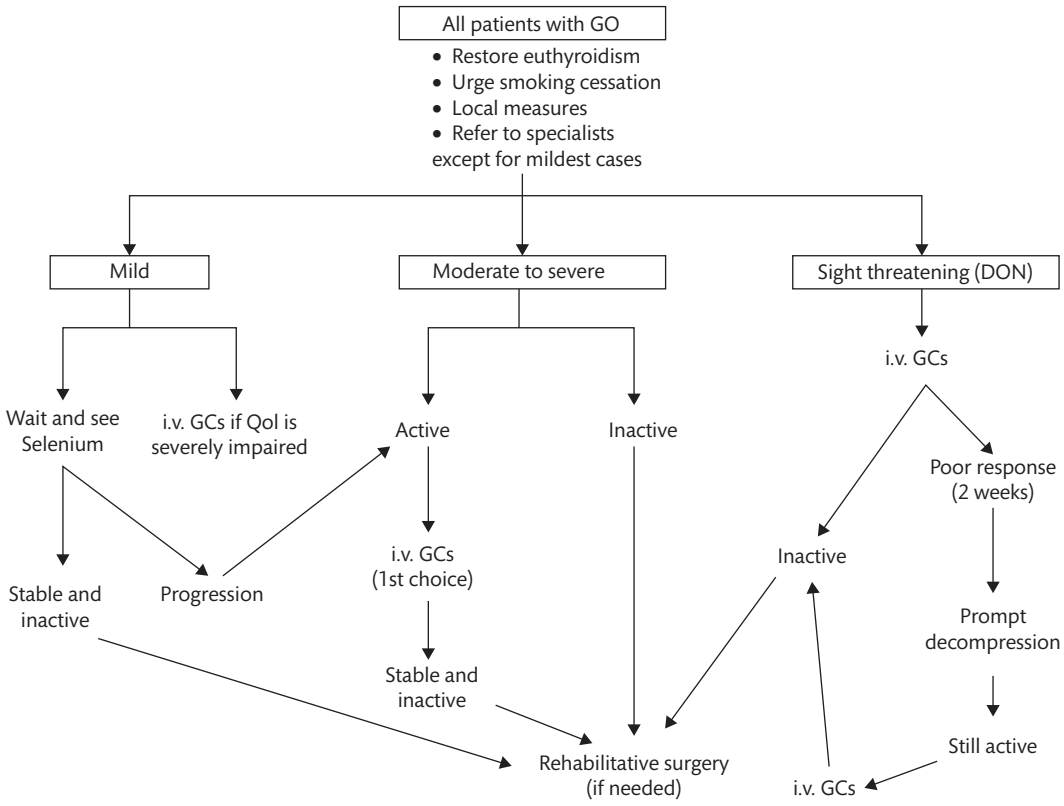


Figure 3.3.10.8 The ETA/EUGOGO guidelines for the management of Graves' orbitopathy.

Reproduced with permission from Bartalena L, Baldeschi L, Boboridis K *et al*. The 2016 ETA/EUGOGO guidelines for the management of Graves' orbitopathy. *Eur Thyroid J*, 2016; 5: 9–26. Copyright © 2016 European Thyroid Association. Published by S. Karger AG, Basel.

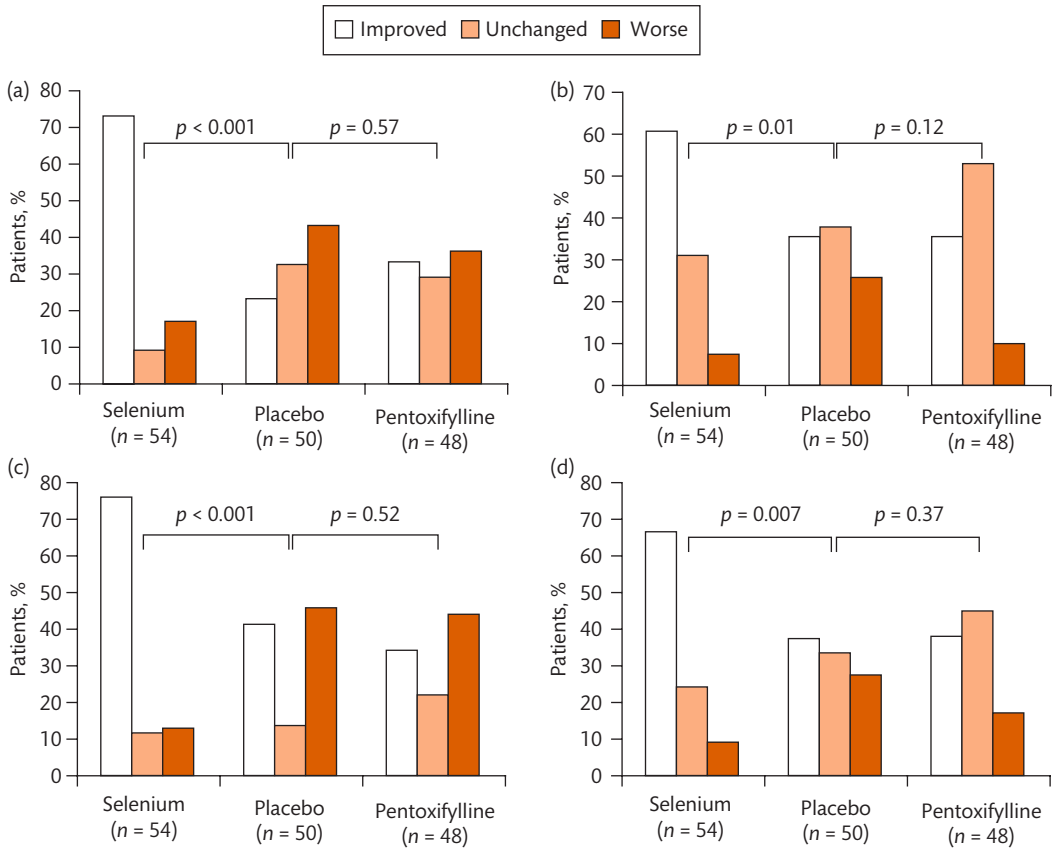


Figure 3.3.10.9 Results of a RCT in patients with mild Graves' orbitopathy comparing a six-month course of selenium or pentoxifylline with placebo on GO-QoL scores (a and b) and a composite score reflecting overall eye evaluation (c and d) at 6 months (a and c) and at 12 months (b and d).

Treatment of Moderate-to-Severe GO

Immunosuppressive treatment modalities are indicated in active moderate-to-severe GO. Glucocorticoids are better than placebo [79], and for long considered the mainstay of immunosuppression in GO. Steroids are effective for the relief of orbital pain and restoring visual acuity; their effectiveness is moderate for improvement of soft tissue involvement and extraocular muscle dysfunction, and limited for reduction of proptosis. Many patients still require rehabilitative surgery after a course of steroids. The main accomplishment of immunosuppression thus seems to be inactivation of the eye disease, thereby permitting earlier corrective surgery.

Intravenous methylprednisolone pulses (IVMP) have a higher response rate than oral prednisone (74–88% vs. 51–63%) and fewer side effects (17–56% vs. 51–85%), rendering IVMP presently the treatment of choice [62, 80]. Dose-finding studies indicate a cumulative dose of 4.5 g IVMP is appropriate for most patients, administered as follows: 500 mg i.v. once weekly for 6 weeks, followed by 250 mg i.v. once weekly for another 6 weeks [81]. Lower doses (2.25 g) are less effective, whereas higher doses (750 mg once weekly for 6 weeks followed by 500 mg once weekly for another 6 weeks, cumulative dose 7.5 g) are more effective at the expense of more side effects, and should be reserved for worst cases. IVMP have been associated with significant cardiovascular and cerebrovascular morbidity and hepatic toxicity, leading to a few sporadic fatalities. These serious events occurred if IVMP was administered in high single doses of ≥1000 mg, in cumulative doses of >8 gram, and/or as repeat infusions on consecutive days [82]. It is therefore recommended that IVMP should not exceed a cumulative dose of >8 g, and should not be given in patients with recent viral hepatitis, significant liver dysfunction, and severe cardiovascular morbidity or psychiatric disorders [62]. Severe hypertension, inadequately managed diabetes, and glaucoma are other contraindications. IVMP therapy should be monitored by regular measurements of blood pressure, blood glucose, and liver function tests. Efficacy of IVMP can be slightly enhanced by coadministration of mycophenolate (one 360 mg tablet twice daily for 24 weeks) [83]. Whereas the combination of oral prednisone and radiotherapy is more effective than oral prednisone alone [62], it is unsettled if addition of radiotherapy to IVMP provides extra benefit.

If eye changes worsen after discontinuation of glucocorticoids, shared decision-making is recommended to select a second-line therapy. This could be another course of IVMP if the cumulative dose of 8 g is not exceeded, oral prednisone (20 mg/day) in combination with either cyclosporine or orbital irradiation, or rituximab [62]. The combination of oral prednisone and cyclosporine is effective in about 60% of patients who had not responded adequately to steroids [84, 85]. The rationale of retrobulbar irradiation is the radiosensitivity of lymphocytes and fibroblasts. A dose of 20 Gy per

orbit fractionated in ten daily doses of 2 Gy over a 2-week period is commonly used [86]. Improvement is seen in about 50–60% of patients, predominantly in soft tissue involvement and eye muscle ductions [78, 87, 88]. A transient increase of conjunctival irritation is seen in 15% of patients. Radiation-induced retinopathy is extremely rare. Long-term follow-up studies did not detect serious complications after 21 years [89], but there exists a theoretical risk of about 0.5% for radiation-induced cancer. Therefore it is prudent to avoid radiotherapy in patients under 35 years of age.

Rituximab (RTX) is an anti-CD20 monoclonal antibody which effectively causes B-cell depletion. In a dose of 1 gram repeated once after two weeks, it is remarkable effective in open case series [90]. However, two randomized controlled trials (RCT) in active moderate-to-severe GO had discrepant outcomes. The first, conducted in the Mayo Clinics, found RTX not better than placebo [91]. The second, performed in Italy, found RTX slightly better than IVMP [92]. Sample size in both trials was very modest, and some differences in patient characteristics and trial execution are noted [93]. Nevertheless, it remains difficult to reconcile the difference in outcomes. It seems too early to accept RTX as an alternative for IVMP, but also too early to dismiss RTX as a disease-modifying drug. Side effects of RTX occur in about 30%; notable is the risk of the cytokine-release syndrome with development of DON.

Other treatment modalities (like azathioprine, ciamexone, pentoxifylline, or acupuncture) have shown no benefit in RCT, but intravenous immunoglobulins can be effective. Somatostatin analogues (octreotide and lanreotide) have little effect. A few case reports mention favourable effects of anti-TNFα antibodies (etanercept and infliximab) and especially of anti-IL6-receptor antibodies (tocilizumab), but these agents have not been tested in RCT [94]. Of considerable interest is teprotumumab, a monoclonal antibody inhibitor of IGF-1 receptors. In a placebo-controlled RCT it had great efficacy in reducing CAS, reducing exophthalmos, and improving diplopia and GO-QoL scores [95]. Its potential to reduce exophthalmos is remarkable, and could be greater than that of IVMP. However, to displace IVMP as first-line treatment, teprotumumab should be compared directly to IVMP in a RCT.

Treatment of Very Severe (Sight-Threatening) GO

There is only one RCT in DON, which, although its sample size is limited, indicates IVMP have a better outcome than immediate surgical decompression [96]. The applied IVMP scheme was 1000 mg daily, given as a 60-min intravenous infusion, on three successive days in week 1 and repeated in week 2. If visual functions have not improved after 2 weeks, we do an urgent surgical decompression. If improved, we continue with oral prednisone (40 mg/day for 2 weeks, 30 mg/day for 4 weeks, 20 mg/day for 4 weeks, and then tapering to zero dose by 2.5 mg/week).

Rehabilitative Surgery

Once GO has become inactive, rehabilitative surgery can be carried out to improve visual functions and appearance. Most orbital surgeons require stable eye disease for 6 months before surgery. Orbital decompression, achieved by removal of part of the bony orbital wall, is very effective in reducing exophthalmos. The more walls are removed, the greater the reduction in proptosis. Careful fat removal during bony decompression is increasingly done. Photographs of how the patient looked before the onset of GO are helpful in determining the required extent of proptosis reduction. Transeyelid or transconjunctival incisions leave a barely visible scar. Postoperative *de novo* or worsening of diplopia occurs in 10–30% of patients. Corrective eye muscle surgery for diplopia should therefore be done after decompressive surgery. Single binocular vision can be obtained in about 80% of patients, but more than one surgical session is often required to reach this goal. Lastly, eyelid surgery can be performed in order to correct upper or lower eyelid retraction; blepharoplasty can reduce any remaining eyelid swelling.

Graves' Dermopathy

Graves' dermopathy or pretibial myxoedema occurs usually in the pretibial area, but can occasionally occur at other sites subject to local mechanical pressure. The most frequent form (49%) is non-pitting oedema with violet discolouration and induration of the skin and prominent hair follicles, so that the lesions have the appearance and texture of orange peel (*peau d'orange*). Other clinical forms are plaques (27%), nodules (18%), and elephantiasis (5%) [97]. Graves' dermopathy occurs almost always in conjunction with GO and high serum concentrations of TSHR-Ab. It is observed in 4% of GO patients, and develops mostly about 1 year after the onset of GO. The postulated pathogenesis is remarkably similar to that of GO: cytokine-induced glycosaminoglycan production by dermal fibroblasts, upregulated TSH receptor expression on dermal fibroblasts, and a contributory role of local trauma and mechanical pressure [98].

The natural course of Graves' dermopathy is not well known. One-half of patients do not require any therapy; such cases may go slowly into complete remission (50% within 17 years). If treatment is necessary because of functional or cosmetic complaints, nighttime occlusive dressings of 0.05–0.1% triamcinolone acetonide in a cream base induce partial remissions in one-third of patients. Usually a trial of 4–10 weeks is needed, followed by intermittent maintenance therapy. The use of compressive bandages or stockings during the day provides additional benefit. Treatment of coexistent GO with systemic glucocorticoids or biologicals can cause regression of the skin lesions as well.

REFERENCES

1. Werner SC. Modification of the classification of the eye changes of Graves' disease: recommendations of the ad hoc committee of the American Thyroid Association. *J Clin Endocrinol Metab*, 1997; **44**: 203–4.
2. McKeag D, Lane CM, Lazarus JH, *et al*. Clinical features of dysthyroid optic neuropathy: a European Group on Graves' Orbitopathy (EUGOGO) survey. *Br J Ophthalmol*, 2007; **91**: 455–8.
3. Perros P, Zarkovic M, Azzolini C, *et al*. PREGO (presentation of Graves' orbitopathy) study: changes in referral patterns to European Group on Graves' Orbitopathy (EUGOGO) centres over the period from 2000 to 2012. *Br J Ophthalmol*, 2015; **99**: 1531–5.
4. Daumerie C, Duprez T, Boschi A. Long-term multidisciplinary follow-up of unilateral thyroid-associated orbitopathy. *Eur J Intern Med*, 2008; **19**: 531–6.

5. Krassas GE, Segni M, Wiersinga WM. Childhood Graves' ophthalmopathy: results of a European questionnaire study. *Eur J Endocrinol*, 2005; **153**: 515–20.
6. Durairaj VD, Bartley GB, Garrity JA. Clinical features and treatment of Graves' ophthalmopathy in pediatric patients. *Ophthal Plast Reconstr Surg*, 2006; **22**: 7–12.
7. Abraham-Nordling M, Bystrom K, Torring O, *et al*. Incidence of hyperthyroidism in Sweden. *Eur J Endocrinol*, 2011; **165**: 899–905.
8. Laurberg P, Berman DC, Bulow-Pedersen I, Anderson S, Carle A. Incidence and clinical presentation of moderate to severe Graves' orbitopathy in a Danish population before and after iodine fortification of salt. *J Clin Endocrinol Metab*, 2012; **97**: 2325–32.
9. Tanda ML, Piantanida E, Liparulo L, *et al*. Prevalence of natural history of Graves' orbitopathy in a large series of patients with newly diagnosed Graves' hyperthyroidism seen in a single centre. *J Clin Endocrinol Metab*, 2013; **98**: 1443–9.
10. Wiersinga WM, Zarkovic M, Bartalena L, *et al*. Predictive score for the development or progression of Graves' orbitopathy in patients with newly diagnosed Graves' hyperthyroidism. *Eur J Endocrinol*, 2018; **178**: 635–43.
11. Eckstein AK, Lösch C, Glowacka D, *et al*. Euthyroid and primarily hypothyroid patients develop milder and significantly more asymmetrical Graves' ophthalmopathy. *Br J Ophthalmol*, 2009; **93**: 1052–6.
12. Khoo DH, Eng PH, Ho SC, *et al*. Graves' ophthalmopathy in the absence of elevated free thyroxine and triiodothyronine levels: prevalence, natural history, and thyrotropin receptor antibody levels. *Thyroid*, 2000; **10**: 1093–100.
13. Prummel MF, Wiersinga WM. Smoking and risk of Graves' disease. *JAMA*, 1993; **269**: 479–82.
14. Perros P, Kendall-Taylor P. Natural history of thyroid eye disease. *Thyroid*, 1998; **8**: 423–5.
15. Weetman AP, Wiersinga WM. Current management of thyroid-associated ophthalmopathy in Europe: results of an international survey. *Clin Endocrinol*, 1998; **49**: 21–8.
16. Regensburg NI, Wiersinga WM, van Velthoven MEJ, *et al*. Age- and gender specific reference values of orbital fat- and muscle volumes in Caucasians. *Br J Ophthalmol*, 2011; **95**: 1660–3.
17. Neigel JM, Rootman J, Belkin RI, *et al*. Dysthyroid optic neuropathy. The crowded orbital apex syndrome. *Ophthalmology*, 1988; **95**: 1515–21.
18. Otto AJ, Koornneef L, Mourits MP, Deen-van Leeuwen L. Retrobulbar pressures measured during surgical decompression of the orbit. *Br J Ophthalmol*, 1996; **80**: 1042–5.
19. Prabhakar BS, Bahn RS, Smith TJ. Current perspective on the pathogenesis of Graves' disease and ophthalmopathy. *Endocr Rev*, 2003; **24**: 802–35.
20. Regensburg NI, Wiersinga WM, Berendschot TTJM, Potgieser P, Mourits MP. Do subtypes of Graves' orbitopathy exist? *Ophthalmology*, 2011; **118**: 191–6.
21. Potgieser PW, Wioersinga WM, Regensburg NI, Mourits MP. Some studies on the natural history of Graves' orbitopathy: increase in orbital fat is a rather late phenomenon. *Eur J Endocrinol*, 2015; **173**: 149–53.
22. Aniszewski JP, Valyasevi RW, Bahn RS. Relationship between disease duration and predominant orbital T cell subset in Graves' ophthalmopathy. *J Clin Endocrinol Metab*, 2000; **85**: 776–80.
23. Valyasevi RW, Erickson DZ, Harteneck DA, *et al*. Differentiation of human orbital preadipocyte fibroblasts induces expression of functional thyrotropin receptor. *J Clin Endocrinol Metab*, 1999; **84**: 2257–62.
24. Wakellamp IMMJ, Bakker O, Baldeschi L, Wiersinga WM, Prummel MF. TSH-R expression and cytokine profile in orbital tissue of active vs. inactive Graves' ophthalmopathy patients. *Clin Endocrinol*, 2003; **58**: 280–7.
25. Vos XG, Smit N, Endert E, Tijssen JG, Wiersinga WM. Frequency and characteristics of TBII-seronegative patients in a population with untreated Graves' hyperthyroidism: a prospective study. *Clin Endocrinol*, 2008; **69**: 311–17.
26. Lytton SD, Ponto KA, Kanitz M, Matheis N, Kohn LD, Kahaly GJ. A novel thyroid stimulating bioassay is a functional indicator of activity and severity of Graves' orbitopathy. *J Clin Endocrinol Metab*, 2010; **95**: 2123–31.
27. Eckstein AK, Plicht M, Lax H, *et al*. Clinical results of anti-inflammatory therapy in Graves' ophthalmopathy and association with thyroidal autoantibodies. *Clin Endocrinol*, 2004; **61**: 612–18.
28. Eckstein AK, Plicht M, Lax H, *et al*. Thyrotropin receptor autoantibodies are independent risk factors for Graves' ophthalmopathy and help to predict severity and outcome of the disease. *J Clin Endocrinol Metab*, 2006; **91**: 3464–70.
29. Laurberg P, Wallin G, Tallstedt L, Abraham-Nordling M, Lundell G, Torring O. TSH-receptor autoimmunity in Graves' disease after therapy with antithyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*, 2008; **158**: 69–75.
30. Acharya SH, Avenell A, Philip S, Burr J, Bevan JS, Abraham P. Radioiodine therapy (RAI) for Graves'disease (GD) and the effect on ophthalmopathy: a systematic review. Clin *Endocrinol*, 2008; **69**: 943–50.
31. Moshkelgosha S, So PW, Deasy N, Diaz-Cano S, Banga P. Cutting edge: retrobulbar inflammation, adipogenesis, and acute orbital congestion in a preclinical female mouse model of Graves' orbitopathy induced by thyrotropin receptor plasmid-*in vivo* electroporation. *Endocrinology*, 2013; **154**: 3008–15.
32. Nagayama Y, Nakahara M, Abiru N. Animal models of Graves' disease and Graves' orbitopathy. *Curr Opin Endocrinol Diabetes Obes*, 2015; **22**: 381–6.
33. Berchner-Pfannschmid U, Moshkelgosha S, Diaz-Cano S, *et al*. Comparative assessment of female mouse model of Graves' orbitopathy under different environments accompanied by proinflammatory cytokine and T-cell responses to thyrotropin hormone. *Endocrinology*, 2016; **157**: 1673–82.
34. Görtz GE, Moshkelgosha S, Jessenek C, *et al*. Pathogenic phenotype of adipogenesis and hyaluronan in orbital fibroblasts from female Graves' orbitopathy mouse model. *Endocrinology*, 2016: **157**: 3771–8.
35. Schlüter A, Horstmann M, Diaz-Cano S, *et al*. Genetic immunization with mouse thyrotropin hormone receptor plasmid breaks self-tolerance for a murine model of autoimmune thyroid disease and Graves' orbitopathy. *Clin Exp Immunol*, 2018; **191**: 255–67.
36. Smith TJ, Hoa N. Immunoglobulins from patients with Graves' disease induce hyaluronan synthesis in their orbital fibroblasts through the self-antigen, insulin-like growth factor-1 receptor. *J Clin Endocrinol Metab*, 2004; **89**: 5076–80.
37. Minich WB, Dehina N, Welsink T, *et al*. Autoantibodies to the IGF1 receptor in Graves' orbitopathy. *J Clin Endocrinol Metab*, 2013; **98**: 752–60.
38. Tsui S, Naik V, Hoa N, *et al*. Evidence for an association between thyroid-stimulating hormone and insulin-like growth factor 1

receptors: a tale of two antigens implicated in Graves' disease. *J Immunol*, 2008; **181**: 4397–405.
39. Wiersinga WM. Autoimmunity in Graves' ophthalmopathy: the result of an unfortunate marriage between TSH receptors and IGF-1 receptors? *J Clin Endocrinol Metab*, 2011; **96**: 2386–94.
40. Krieger CC, Neumann S, Place RF, Marcus-Samuels B, Gershengorn MC. Bidirectional TSH and IGF-1 receptor cross talk mediates stimulation of hyaluronan secretion by Graves' disease immunoglobulins. *J Clin Endocrinol Metab*, 2015; **100**: 1071–7.
41. Krieger CC, Place RF, Bevilacqua C, *et al*. TSH/IGF-1 receptor cross talk in Graves' ophthalmopathy pathogenesis. *J Clin Endocrinol Metab*, 2016; **101**: 2340–7.
42. Marcus-Samuels B, Krieger CC, Boutin A, Kahaly GJ, Neumann S, Gershengorn MC. Evidence that Graves' ophthalmopathy immunoglobulins do not directly activate IGF-1 receptors. *Thyroid*, 2018; **28**: 650–5.
43. Iyer S, Bahn RS. Immunopathogenesis of Graves' ophthalmopathy: the role of the TSH receptor. *Best Pract Res Clin Endocrinol Metab*, 2012; **26**: 281–9.
44. Smith TJ. TSH-receptor expressing fibrocytes and thyroid-associated ophthalmopathy. *Nat Rev Endocrinol*, 2015; **11**: 171–81.
45. Chen H, Mester T, Raychaudhuri N, *et al*. Teprotumumab, an IGF-1R blocking monoclonal antibody inhibits TSH and IGF-1 action in fibrocytes. *J Clin Endocrinol Metab*, 2014; **99**: E1635–40.
46. Tellez M, Cooper J, Edmonds C. Graves' ophthalmopathy in relation to cigarette smoking and ethnic origin. *Clin Endocrinol*, 1992; **36**: 291–4.
47. Bednarczuk T, Gopinath B, Ploski R, Wall JR. Susceptibility genes in Graves' ophthalmopathy: searching for a needle in a haystack? *Clin Endocrinol*, 2007; **67**: 3–19.
48. Cawood TJ, Moriarty P, O'Farrelly C, O'Shea D. Smoking and thyroid-associated ophthalmopathy: a novel explanation of the biological link. *J Clin Endocrinol Metab*, 2007; **92**: 59–64.
49. Rundle FF. Management of exophthalmos and related ocular changes in Graves' disease. *Metabolism*, 1957; **6**: 36–47.
50. Perros P, Crombie AL, Kendall-Taylor P. Natural history of thyroid associated ophthalmopathy. *Clin Endocrinol*, 1995; **42**: 45–50.
51. Wiersinga WM. Advances in medical therapy of thyroid-associated ophthalmopathy. *Orbit*, 1996; **15**: 177–86.
52. European Group on Graves' Orbitopathy (EUGOGO), Wiersinga WM, Perros P, *et al*. Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. *Eur J Endocrinol*, 2006; **155**: 387–9.
53. Prummel MF, Wiersinga WM, Mourits MP, Koornneef L, Berghout A, van der Gaag RD. Effect of abnormal thyroid function on the severity of Graves' ophthalmopathy. *Arch Intern Med*, 1990; **150**: 1098–101.
54. Dickinson AJ, Perros P. Controversies in the clinical evaluation of active thyroid-associated orbitopathy: use of a detailed protocol with comparative photographs for objective assessment. *Clin Endocrinol*, 2001; **55**: 283–303.
55. Mourits MP, Prummel MF, Wiersinga WM, Koornneef L. Measuring eye movements in Graves' ophthalmopathy. *Ophthalmology*, 1994; **101**: 1341–6.
56. Mourits MP, Prummel MF, Wiersinga WM, Koornneef L. Clinical activity score as a guide in the management of patients with Graves' ophthalmopathy. *Clin Endocrinol*, 1997; **47**: 9–14.
57. Gerding MN, Prummel MF, Wiersinga WM. Assessment of disease activity in Graves' ophthalmopathy by orbital ultrasonography and clinical parameters. *Clin Endocrinol*, 2000; **52**: 641–6.
58. Gerding MN, Terwee CB, Dekker FW, Koornneef L, Prummel MF, Wiersinga WM. Quality of life in patients with Graves' ophthalmopathy is markedly decreased: measurement by the Medical Outcomes Study instrument. *Thyroid*, 1997; **7**: 885–9.
59. Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QoL. *Br J Ophthalmol*, 1998; **82**: 773–9.
60. Terwee CB, Dekker FW, Mourits MP, *et al*. Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire (GO-QoL) after different treatments. *Clin Endocrinol*, 2001; **54**: 391–8.
61. Ponto KA, Merkesdal S, Hommel G, Pitz S, Pfeiffer N, Kahaly GJ. Public health relevance of Graves' orbitopathy. *J Clin Endocrinol Metab*, 2013; **98**: 145–52.
62. Bartalena L, Baldeschi L, Boboridis K *et al*. The 2016 ETA/EUGOGO guidelines for the management of Graves' orbitopathy. *Eur Thyroid J*, 2016; **5**: 9–26.
63. Mombaerts I, Goldschmeding R, Schlingemann RO, Koornneef L. What is orbital pseudotumor? A clinical pathological review. *Surv Ophthalmol*, 1996; **41**: 66–78.
64. Erdei A, Steiber Z, Molnar C, Berenyi E, Nagy EV. Exophthalmos in a young woman with no Graves' disease—a case report of IgG4-related orbitopathy. *BMC Ophthalmol*, 2018; **18**: 5. Doi: 10.1186/s12886–018-0672-y.
65. Estcourt S, Hickey J, Perros P, Dayan C, Vaidya B. The patient experience of services for thyroid eye disease in the United Kingdom: results of a nationwide survey. *Eur J Endocrinol*, 2009; **161**: 483–7.
66. Krassas GE, Wiersinga WM. Smoking and autoimmune thyroid disease: the plot thickens. *Eur J Endocrinol*, 2006; **154**: 777–80.
67. Prummel MF, Wiersinga WM, Mourits MPh, Koornneef L, Berghout A, van der Gaag RD. Amelioration of eye changes of Graves' ophthalmopathy by achieving euthyroidism. *Acta Endocrinol*, 1989; **121**(Suppl 2): 185–9.
68. Perros P, Kendall-Taylor P, Neoh C, Frewin S, Dickinson J. A prospective study of the effects of radioiodine therapy for hyperthyroidism in patients with minimally active Graves' ophthalmopathy. *J Clin Endocrinol Metab*, 2005; **90**: 5321–3.
69. Elbers L, Mourits M, Wiersinga W. Outcome of very long-term treatment with antithyroid drugs in Graves' hyperthyroidism associated with Graves' orbitopathy. *Thyroid*, 2011; **21**: 279–83.
70. Khong JJ, Finch S, De Silva C, *et al*. Risk factors for Graves' orbitopathy: the Australian thyroid-associated orbitopathy research (ATOR) study. *J Clin Endocrinol Metab*, 2016; **101**: 2711–20.
71. Wiersinga WM. Preventing Graves' ophthalmopathy. *N Engl J Med*, 1998; **338**: 121–2.
72. Marcocci C, Bruno-Bossio G, Manetti L, *et al*. The course of Graves' ophthalmopathy is not influenced by near-total thyroidectomy: a case-control study. *Clin Endocrinol*, 1999; **51**: 503–8.
73. Jarhult J, Rudberg C, Larsson E, *et al*. Graves' disease with moderate-severe endocrine ophthalmopathy: long term results of a prospective, randomized study of total or subtotal thyroid resection. *Thyroid*, 2005; **15**: 1157–64.
74. Menconi F, Marinò M, Pinchera A, *et al*. Effects of total thyroid ablation versus near-total thyroidectomy alone on mild to

moderate Graves' orbitopathy treated with intravenous glucocorticoids. *J Clin Endocrinol Metab*, 2007; **92**: 1653–8.
75. Stein JD, Childers D, Gupta S, *et al.* Risk factors for developing thyroid-associated ophthalmopathy among individuals with Graves' disease. *JAMA Ophthalmol*, 2015; **133**: 290–6.
76. Marcocci C, Kahaly GJ, Krassas GE, *et al.* Selenium and the course of mild Graves' orbitopathy. *N Engl J Med*, 2011; **364**: 1920–31.
77. Rotondo G, Leo M, Casini G, *et al.* Anti-oxidant actions of selenium in orbital fibroblasts: a basis for the effects of selenium in in Graves' orbitopathy. *Thyroid*, 2017; **27**: 271–8.
78. Prummel MF, Terwee CB, Gerding MN, *et al.* A randomized controlled trial of orbital radiotherapy versus sham irradiation in patients with mild Graves' ophthalmopathy. *J Clin Endocrinol Metab*, 2004; **89**: 15–20.
79. van Geest RJ, Sasim IV, Koppeschaar HP, *et al.* Methylprednisolone pulse therapy for patients with moderately severe Graves' orbitopathy: a prospective, randomized, placebo-controlled study. *Eur J Endocrinol*, 2008; **158**: 229–37.
80. Stiebel-Kalish H, Robenshtok E, Hasanreisoglu M, Ezrachi D, Shimon I, Leibovici L. Treatment modalities for Graves' ophthalmopathy: systematic review and metaanalysis. *J Clin Endocrinol Metab*, 2009; **94**: 2708–16.
81. Bartalena L, Krassas GE, Wiersinga WM, *et al.* Efficacy and safety of three different cumulative doses of intravenous methylprednisolone for moderate to severe and active Graves' orbitopathy. *J Clin Endocrinol Metab*, 2012; **97**; 4454–63.
82. Marcocci C, Watt T, Altea MA, *et al.* Fatal and non-fatal adverse events of glucocorticoid therapy for Graves' orbitopathy: a questionnaire survey among members of the European Thyroid Association. *Eur J Endocrinol*, 2012; **166**: 247–53.
83. Kahaly GJ, Riedl M, Kónig M, *et al.* Mycophenolate plus methylprenisolone versus methylprednisolone alone in active, moderate-to-severe Graves' orbitopathy (MINGO): a randomized, observer-masked, multicentre trial. *Lancet Diabetes Endocrinol*, 2018; **6**: 287–98.
84. Prummel MF, Mourits MP, Berghout A, *et al.* Prednisone and cyclosporine in the treatment of severe Graves' ophthalmopathy. *N Engl J Med*, 1989; **321**: 949–54.
85. Kahaly G, Schrezenmeir J, Krause U, Schweikert B, Meuer S, Muller W. Cyclosporin and prednisone versus prednisone in treatment of Graves' ophthalmopathy: a controlled, randomized and prospective study. *Eur J Clin Invest*, 1986; **16**: 415–22.
86. Kahaly GJ, Rösler HP, Pitz S, Hommel G. Low- versus high-dose radiotherapy for Graves' ophthalmopathy: a randomized, single blind trial. *J Clin Endocrinol Metab*, 2000; **85**: 102–8.
87. Prummel MF, Mourits MP, Blank L, Berghout A, Koornneef L, Wiersinga WM. Randomized double-blind trial of prednisone versus radiotherapy in Graves' ophthalmopathy. *Lancet*, 1993; **342**: 949–954.
88. Mourits MP, Van Kempen-Harteveld ML, Begonia Garcia M, Koppeschaar HPF, Tick L, Terwee CB. Randomized placebo-controlled study of radiotherapy for Graves' orbitopathy. *Lancet*, 2000; **355**: 1505–9.
89. Wakelkamp IM, Tan H, Saeed P, *et al.* Orbital irradiation for Graves' ophthalmopathy. Is it safe? A long-term follow-up study. *Ophthalmology*, 2004; **111**: 1557–62.
90. Salvi M, Vannucchi G, Beck Peccoz P. Potential utility of rituximab for Graves' orbitopathy. *J Clin Endocrinol Metab*, 2013; **98**: 4291–9.
91. Stan MN, Garrity JA, Carranza Leon BG, Prabin T, Bradley EA, Bahn RS. Randomized controlled trial of rituximab in patients with Graves' orbitopathy. *J Clin Endocrinol Metab*, 2015; **100**: 432–41.
92. Salvi M, Vannucchi G, Curro N, *et al.* Efficacy of B-cell targeted therapy with rituximab in patients with active moderate to severe Graves' orbitopathy: a randomized controlled study. *J Clin Endocrinol Metab*, 2015; **100**: 422–31.
93. Stan MN, Salvi M. Rituximab therapy for Graves' orbitopathy—lessons from randomized control trials. *Eur J Endocrinol*, 2017; **176**: R101–9.
94. Perez-Moreiras JV, Alvarez-Lopez A, Gomez EC. Treatment of active corticosteroid-resistant Graves' orbitopathy. *Ophthalmic Plast Reconstr Surg*, 2014; **30**: 162–7.
95. Smith TJ, Kahaly GJ, Ezra DG, *et al.* Teprotumumab for thyroid-associated ophthalmopathy. *N Engl J Med*, 2017; **376**: 1748–61.
96. Wakelkamp IMMJ, Baldeschi L, Saeed P, Mourits MP, Prummel MF, Wiersinga WM. Surgical or medical decompression as a first-line treatment of optic neuropathy in Graves' ophthalmopathy? A randomized controlled trial. *Clin Endocrinol*, 2005; **63**: 323–8.
97. Fatourechi V. Pretibial myxedema. Pathophysiology and treatment options. *Am J Clin Dermatol*, 2005; **6**: 295–309.
98. Rapoport B, Alsabeh R, Aftergood D, McLachlan SM. Elephantiasic pretibial myxedema: insight into and a hypothesis regarding the pathogenesis of the extrathyroidal manifestations of Graves' disease. *Thyroid*, 2000; **10**: 685–92.

FURTHER READING

Wiersinga WM, Kahaly GJ (eds). Graves' Orbitopathy. A Multidisciplinary Approach. Basel: Karger, 2017 (3rd edition).

3.3.11 Management of Toxic Multinodular Goitre and Toxic Adenoma

Dagmar Führer and Holger Jäschke

Definition

Toxic adenoma (TA) and toxic multinodular goitre (TMNG) represent the clinically important presentations of thyroid autonomy. Thyroid autonomy is a condition where thyrocytes produce thyroid

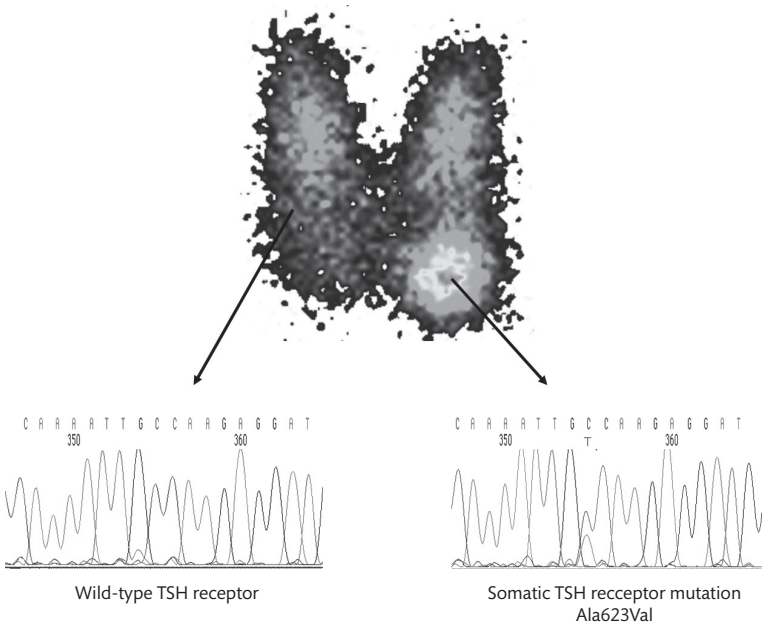


Figure 3.3.11.1 Scintiscan of uninodular goitre shows a circumscribed area of increased ^{99m}Tc-uptake in the left lobe ('hot nodule'). DNA was extracted from the toxic adenoma and surrounding normal thyroid tissue and exon 10 of the TSHR was amplified by polymerase chain reaction (PCR). Sequencing of the PCR products showed the presence of a heterozygous point mutation (GCC→GTC) resulting in an amino acid exchange (Val→Ala) in the toxic adenoma (right panel), whereas only the wild-type TSHR was present in the normal thyroid tissue (left panel). The mutation causes a constitutive activation of the TSHR, which leads to thyrotoxicosis and thyroid growth.

hormones independently of thyrotropin (TSH) and in absence of TSH receptor stimulating antibodies (TSAb).

Toxic adenoma (TA) is a clinical term referring to a solitary autonomously functioning thyroid nodule. The autonomous properties of TA are best shown by radioiodine or ^{99m}Tc imaging. The classic appearance of TA is a local and focal increase in tracer uptake with suppressed uptake in the surrounding extranodular thyroid tissue ('hot' nodule, Figure 3.3.11.1).

Toxic multinodular goitre (TMNG) is a heterogeneous disorder characterized by the presence of autonomously functioning thyroid nodules in an enlarged thyroid gland with or without additional nodules. These additional nodules can show normal or decreased tracer uptake (cold nodules) on scintiscan. TMNG constitutes the most frequent form of thyroid autonomy.

Epidemiology and Pathogenesis

The prevalence of thyroid autonomy is inversely correlated with iodine intake. Thus, thyroid autonomy is a common finding in iodine deficient areas, where it accounts for up to 60% of cases of thyrotoxicosis (TMNG: ~50%; TA: ~10%), but is rare (5–10%) in regions with iodine sufficiency [1, 2]. Specifically, in areas with iodine deficiencies the incidence of TMNG is 18 cases per 100 000 per year. In contrast, an incidence of 1.5 cases per 100 000 per year was determined in regions with high iodine supply [2, 3]. For TA the prevalence is 3.6 vs. 1.6 cases per 100 000 per year in regions with low and high iodine supply, respectively [2, 3]. In a Swedish area with stable iodine supply incidence rates for TMNG of 4.3 and for TA of 1.8 per 100 000 cases per year have been reported [2, 4]. Several studies have suggested that TMNG originates from euthyroid goitres and microscopic autonomous foci have been demonstrated in up to 40% of euthyroid goitres in iodine deficient areas [1]. Correction of iodine deficiency in a population results in decrease of thyroid autonomy and this has been impressively shown, for example, in Switzerland where a doubling in iodine salt content resulted in a 73% reduction of TMNG [5, 6]. Moreover, the prevalence of thyroid autonomy correlates with thyroid nodularity and increases with age as shown in a Danish population study evaluating the epidemiology of subtypes of hyperthyroidism [7]. In this study the most prevalent form of hyperthyroidism in individuals <50 years was Graves' disease. With advancing age, the incidence of TMNG and TA increased and outnumbered the cases of Graves' disease.

Thyrotoxicosis is rare disorder in children. In a Danish nationwide study, the incidence was 1.58/100 000 persons per year and about 96% of cases account for Graves' disease and less than 3% of cases were toxic nodular goitre or TA [8].

Somatic mutations of the G-protein coupled TSH receptor (TSHR; 50–75% http://www.uni-leipzig.de/innere/TSHR) or less frequently the Gs alpha protein subunit (5–10%) have been identified in TA and TMNG and represent the predominant molecular cause of thyroid autonomy. These mutations cause constitutive activation of the cAMP pathway, which stimulates TSH-independent thyroid hormone production and thyroid growth (Krohn 2005 PMID: 15615818, Parma 1993 PMID: 8413627). Recently, a specific mutation in the gene for a histone methyltransferase called **enhancer of zest homolog 1** (*EZH1*) (*c*.1712A>G; p.Gln571Arg) was identified in 27% (25 of 94) of adult TA. Surprisingly, the mutation was neither found in papillary or follicular thyroid carcinoma nor in 29 TA from children. This suggests that paediatric TA constitute a different group

or that the EZH1 variant may be acquired later during progression of the tumour. In contrast to TSHR or Gs alpha mutations, which constitutively activate the cAMP pathway, the EZH1 mutation exerts its effect via hypermethylation of histone H3 and consequently to the repression of gene transcription [9].

Clinical Features and Diagnosis

Clinical features of thyroid autonomy may be related to hyperthyroidism and/or compression signs due to the nodule and TMNG [10].

Clinical presentation of overt hyperthyroidism, defined by suppressed TSH with elevated fT_4 and/or fT_3, varies with age. While classic hyperthyroid features such as tremor, sweating, and hyperkinesis can be found in the younger patients, thyrotoxicosis is often oligosymptomatic in older people. In this population, atrial fibrillation, congestive heart failure, and anorexia may prevail. Subclinical hyperthyroidism is based exclusively on the biochemical determination of thyroid function and is defined by low or suppressed TSH with normal fT_4 and fT_3 levels. It is more commonly observed in older patients and is more than 'just' a low TSH status, since it confers increased risk for atrial fibrillation and contributes in postmenopausal women to reduced bone density [11].

In addition, a history of possible iodine contamination (contrast media, amiodarone) should be obtained. In the European Study Group of Hyperthyroidism iodine contamination was found in 36.8% of patients from iodine deficient areas with first diagnosis of hyperthyroidism. Severity of iodine deficiency, autonomous thyroid cell mass, and older age have been proposed as risk factors for the development of iodine induced hyperthyroidism, which responds less well to antithyroid drug treatment and poses the patient at risk for a life-threatening thyroid storm [12].

Alternatively, a patient may present with a lump or disfigurement of the neck, intolerance of tight necklaces, or increase in collar size. Moreover, dysphagia or breathing difficulties due to local oesophageal or tracheal compression may be present, particularly with TMNG.

Unusually, in some patients there may be a family history of thyroid autonomy and a characteristic course of frequent relapses of hyperthyroidism following thyrostatic therapy or partial thyroidectomy. Depending on the age of onset (neonatal to adulthood) these patients may present with a diffusely enlarged goitre or a TMNG. The underlying cause of this condition is an activating germline mutation in the *TSHR* gene, which can be confirmed through molecular diagnostics from a peripheral blood sample. Patients with an activating TSHR germline mutation require definite treatment in form of total thyroidectomy or an ablative dose of radioiodine to prevent further relapses. Genetic counselling is also mandatory as the condition is autosomal dominantly inherited [13].

Diagnosis of TA and TMNG is based on clinical examination, thyroid function tests, thyroid ultrasound and scintiscanning [10]:

- Examination of the neck will reveal the degree of thyroid enlargement and nodularity of the gland. A history of a recently enlarging nodule and cervical lymph node enlargement should be noted because of concern of a developing malignancy at this site. In addition, clinical evidence of lymph node enlargement and tracheal deviation and/or compression should be sought.
- Standard thyroid function tests (TSH and free thyroid hormones fT_4 and fT_3) will confirm overt or subclinical hyperthyroidism, but depending on the autonomous cell mass, euthyroidism may still prevail.
- Localization, size, and number of thyroid nodule(s) as well as goitre volume can be determined by ultrasound using a 7.5 or 9 MHz linear scanner. In addition, presence or absence of cervical lymph node enlargement should be noted.
- Increased ^{99m}Tc or radionucleotide uptake in the nodule(s) concomitant with a decreased uptake in the surrounding extranodular thyroid tissue is the typical finding on scintiscanning (**Figure 3.3.11.1**). If thyroid autonomy is suspected in a patient with a (still) euthyroid nodule, a 'suppression' scan can be performed, after administration of thyroid hormones (e.g. 75 µg levothyroxine/d for 2 weeks followed by 150 µg/d for 2 weeks). Thereby non-autonomous tissue will be suppressed and thyroid autonomy unmasked.

Measurement of thyroid autoantibodies is not routinely performed in thyroid autonomy. However, in iodine deficient areas distinction between Graves' disease and TMNG can be difficult if extrathyroidal manifestations of autoimmune thyroid disease are absent and ultrasound shows presence of thyroid nodules (~27–34%). In this scenario measurement of TSAb is helpful to establish the correct diagnosis. Urinary iodine excretion can be measured in case of suspected iodine contamination. Computer tomography or magnetic resonance imaging are not routinely indicated for diagnosis of thyroid autonomy but may be used for presurgical planning in cases of large and partly intrathoracic TMNG.

Treatment of TMNG and TA

The management of patients with thyroid autonomy (TMNG and TA) will to some extent depend on the patient's age, the severity of hyperthyroidism, the size of the thyroid gland and concomitant other medical illness [11, 14–16].

Antithyroid medication (ATD) is the first-line of treatment in all patients with overt hyperthyroidism. Depending on the type of antithyroid drug, an initial dosage of 10–20 mg/d methimazole, 15–30 mg/d carbimazole is recommended. Higher dosages are associated with more frequent adverse effects and will only result in marginally faster resolution of thyrotoxicosis. β-blockers are helpful for symptom relief, until the patient is euthyroid.

A trial of low dose antithyroid medication (5–10 mg methimazole/d) may be justified in selected patients with symptomatic subclinical hyperthyroidism (i.e. atrial fibrillation; alternatively, β-blocking agents can be used) [11]. Monitoring of thyroid function and ATD side effects, in particular full blood count and liver function tests are mandatory [11, 15].

Due to the underlying molecular defect there is no spontaneous resolution of thyroid autonomy and definite treatment is indicated once thyroid autonomy becomes clinically manifest. Elderly patients with severe non-thyroidal illness may be an exception from this rule.

Table 3.3.11.1 Treatment of TMNG and TA

Modality	Advantages	Disadvantages
Surgery	Effective	Hospitalization
	Simple operation	Anaesthesia
	Rapid euthyroidism	Side effects (vocal cords, Hypoparathyroidism)
Radioiodine	Outpatient therapy	Time to cure possible hypothyroidism in long term
Antithyroid drugs	Rapid euthyroidism	Relapse on stopping
	Not destructive	Side effects (skin, liver, bone marrow)
Local ablative treatment practised	Outpatient procedure	Not widely
(e.g. Laser ablation, radiofrequency, percutaneous ethanol injections)	Effective	Several treatment sessions required Side effects (vocal cord, pain)

However, benefits and risks of such 'long-term' ATD have to be considered against the nowadays very low risk of definite treatment.

Two different ablative treatment options are widely available for TA and TMNG and comprise thyroid surgery and radioiodine treatment.

The purpose of thyroid surgery is to cure hyperthyroidism by removing all autonomously functioning thyroid tissue and other macroscopically visible nodular thyroid tissue [10, 17]. Thus the extent of surgery will vary depending on preoperative ultrasound findings and intraoperative morphological inspection. Hemithyroidectomy is usually adequate for TA, if no further nodules are detectable, while in case of TMNG a subtotal, near-total, or total thyroidectomy is performed. Advantages of surgery are a fast ablation of hyperthyroidism and the immediate relief of compression symptoms. Disadvantages of surgery are thyroid-specific side effects (i.e. vocal cord paralysis and permanent hypoparathyroidism), which should be less than 1% with an experienced endocrine surgeon. Evidently, the rate of postoperative hypothyroidism will vary with the extent of thyroid resection and in case of a near-total or total thyroidectomy requires start of efficient thyroid hormone replacement therapy (1.6–1.8 μg levothyroxine/kg body weight), aiming for a normal range TSH value after surgery.

Surgery is usually recommended in large TMNG (>100 ml) and is indicated in case of suspicion of thyroid cancer. In addition, surgery is also advocated in patients with overt hyperthyroidism and adverse side effects of ATD, or as an early emergency procedure in patients with a thyroid storm [12, 18].

Radioiodine therapy is also highly effective for ablation of hyperthyroidism and reduction of TA or TMNG volume [10, 14, 15, 19]. Different protocols have been suggested for ^{131}I therapy in benign thyroid disease. Some investigators prefer to administer a standard dose (e.g. 370–740 MBq/thyroid gland) while others apply a certain ^{131}I activity/gram thyroid tissue. The success rate of an individually dosed ^{131}I therapy has been reported to range between 85% and 100% in TA and reaches up to 90% in TMNG. An average thyroid and/or nodule volume reduction of ~40% can be anticipated. Advantages of radioiodine are its simple and outpatient-based applicability. Disadvantages are the 'time to euthyroidism' period (6 weeks to >3 months), during which ATD has to be continued and thyroid function monitored at 3- to 6-week intervals. Radioiodine treatment is contraindicated in pregnancy and contraception is advocated for at least 6 months after receiving ^{131}I therapy.

Postradioiodine hypothyroidism in TMNG and TA usually develops insidiously and depends on the extent of TSH suppression prior to ^{131}I therapy and the protocol applied. In one study the incidence of hypothyroidism was 3% at 1 year, 31% at 8 years, and 64% at 24 years follow-up after radioiodine treatment. These data emphasize the requirement of long-term monitoring of thyroid function in all patients receiving ^{131}I therapy.

For the treatment of TMNG and TA non-surgical ultrasound-guided ablation techniques have been explored. These ablation modalities include ethanol injection therapy (PEIT), radiofrequency ablation (RFA) and laser therapy and were studied by a limited number of centres mainly in Europe and South Korea [20–23]. In experienced hands, these methods have been shown to eliminate autonomy in a tissue and to decrease the size of a nodule with very few complications, except for pain and local discomfort. Since these techniques are not widely available they might be reserved for further trials and for patients for whom no other treatment option is feasible or for those who decline traditional option. However, treatment of hyperfunctioning nodules remains best treated with radioiodine due to better control of hyperthyroidism long-term and with fewer side effects [24].

A summary of the advantages and disadvantages of the different treatment modalities for TA and TMNG is shown in Table 3.3.11.1.

Follow-up

The long-term management of patients with TA and TMNG is directed at the detection and adequate treatment of thyroid dysfunction (TSH-level), detection of novel nodular thyroid disease (palpation and ultrasonography) and in case of surgery, detection and treatment of postsurgical hypoparathyroidism (serum calcium). In case of ^{131}I therapy, long-term follow-up for development of hypothyroidism is mandatory (e.g. at an annual rate). Thyroxine with or without iodine is often administered after thyroid surgery to prevent recurrent goitre/thyroid nodules. Although large randomized trials are lacking to provide definite evidence that postoperative thyroxine administration is beneficial, unless the patient is hypothyroid, this treatment strategy is inferred by studies treating goitre/nodules with levothyroxine. In addition, in iodine deficient areas iodine supplementation may be appropriate to prevent further nodular thyroid disease.

REFERENCES

1. Krohn K, Führer D, Bayer Y, *et al*. Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev*, 2005; **26**: 504–24.
2. Taylor PN, Albrecht D, Scholz A, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*, 2018; **14**: 301–16.
3. Laurberg P, Pedersen KM, Vestergaard H, Sigurdsson G. High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. *J Int Med*, 1991; **229**: 415–20.
4. Nystrom HF, Jansson S, Berg G. Incidence rate and clinical features of hyperthyroidism in a long-term iodine sufficient area of Sweden (Gothenburg) 2003–2005. *Clin Endocrinol (Oxf)*, 2013; **78**: 768–76.
5. Baltisberger BL, Minder CE, Burgi H. Decrease of incidence of toxic nodular goitre in a region of Switzerland after full correction of mild iodine deficiency. *Eur J Endocrinol*, 1995; **132**: 546–9.
6. Burgi H, Kohler M, Morselli B. Thyrotoxicosis incidence in Switzerland and benefit of improved iodine supply. *Lancet*, 1998; **352**: 1034.
7. Carle A, Pedersen IB, Knudsen N, *et al*. Epidemiology of subtypes of hyperthyroidism in Denmark: a population-based study. *Eur J Endocrinol*, 2011; **164**: 801–9.
8. Havgaard Kjaer R, Smedegard Andersen M, Hansen D. Increasing incidence of juvenile thyrotoxicosis in Denmark: a nationwide study, 1998–2012. *Horm Res Paediatr*, 2015; **84**: 102–7.
9. Calebiro D, Grassi ES, Eszlinger M, *et al*. Recurrent EZH1 mutations are a second hit in autonomous thyroid adenomas. *J Clin Invest*, 2016; **126**: 3383–8.
10. Hegedus L. Clinical practice. The thyroid nodule. N Engl J Med, 2004; **351**: 1764–71.
11. Biondi B, Bartalena L, Cooper DS, Hegedus L, Laurberg P, Kahaly GJ. The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. *Eur Thyroid J*, 2015; **4**: 149–63.
12. Roti E, Uberti ED. Iodine excess and hyperthyroidism. Thyroid, 2001; **11**: 493–500.
13. Ferraz C, Paschke R. Inheritable and sporadic non-autoimmune hyperthyroidism. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 265–75.
14. Hegedus L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev*, 2003; **24**: 102–32.
15. Ross DS, Burch HB, Cooper DS, *et al*. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*, 2016; **26**: 1343–421.
16. Gharib H, Papini E, Garber JR, *et al*. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules—2016 Update. *Endocrine Pract*, 2016; **22**: 622–39.
17. Porterfield JR, Thompson GB, Young WF, Jr., *et al*. Surgery for Cushing's syndrome: an historical review and recent ten-year experience. *World J Surg*, 2008; **32**: 659–77.
18. Sarlis NJ, Gourgiotis L. Thyroid emergencies. *Rev Endocr Metab Disord*, 2003; **4**: 129–36.
19. Reiners C, Schneider P. Radioiodine therapy of thyroid autonomy. *Eur J Nucl Med Mol Imaging*, 2002; **29** Suppl 2: S471–8.
20. Gharib H, Hegedus L, Pacella CM, Baek JH, Papini E. Clinical review: nonsurgical, image-guided, minimally invasive therapy for thyroid nodules. *J Clin Endocrinol Metab*, 2013; **98**: 3949–57.
21. Ha EJ, Baek JH, Kim KW, *et al*. Comparative efficacy of radiofrequency and laser ablation for the treatment of benign thyroid nodules: systematic review including traditional pooling and Bayesian network meta-analysis. *J Clin Endocrinol Metab*, 2015; **100**: 1903–911.
22. Ha SM, Sung JY, Baek JH, *et al*. Radiofrequency ablation of small follicular neoplasms: initial clinical outcomes. *Int J Hyperthermia*, 2017; **33**: 931–7.
23. Tarantino L, Francica G, Sordelli I, *et al*. Percutaneous ethanol injection of hyperfunctioning thyroid nodules: long-term follow-up in 125 patients. *AJR Am J Roentgenol*, 2008; **190**: 800–8.
24. Papini E, Pacella CM, Misischi I, *et al*. The advent of ultrasound-guided ablation techniques in nodular thyroid disease: towards a patient-tailored approach. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 601–18.

3.3.12 Management of Thyrotoxicosis Without Hyperthyroidism

Wilmar M. Wiersinga

Introduction

Thyrotoxicosis without hyperthyroidism originates either from the thyroid gland as a result of destructive lesions or from extrathyroidal sources (Box 3.3.12.1). In contrast, thyrotoxicosis with hyperthyroidism refers to conditions like Graves' hyperthyroidism and toxic nodular goitre in which there is increased biosynthesis of thyroid hormones in the thyroid gland. Radioiodine uptake in the neck and thyroid vascularity at ultrasonography are low in thyrotoxicosis without hyperthyroidism, but high in thyrotoxicosis with hyperthyroidism. Cytokine-, iodine-, and amiodarone-induced thyrotoxicosis manifest themselves either as thyrotoxicosis without hyperthyroidism, or as thyrotoxicosis with hyperthyroidism in which thyroidal radioiodine uptake is preserved. Both types are discussed in this chapter.

Box 3.3.12.1 Thyrotoxicosis without hyperthyroidism

1 Destructive thyroiditis
- Subacute thyroiditis of De Quervain
- Painless or silent thyroiditis
- Postpartum thyroiditis
- Radiation-induced thyroiditis
- Cytokine-induced thyroiditis
- Checkpoint inhibitor-induced thyroiditis

2 Iodine-induced thyrotoxicosis
3 Amiodarone-induced thyrotoxicosis
4 Thyrotoxicosis of extrathyroidal origin
- Factitious thyrotoxicosis (excess of exogenous thyroid hormone, Hamburger thyrotoxicosis)
- Struma ovarii
- Functional metastases of differentiated thyroid carcinoma

Thyrotoxicosis Due to Destructive Thyroiditis

The inflammatory reaction in destructive thyroiditis disrupts the normal architecture of the thyroid gland, causing release of preformed T_4 and T_3 stored in the colloid. The colloid contains more T_4 than T_3, explaining the lower serum T_3/T_4 ratio in thyrotoxic patients without hyperthyroidism as compared to thyrotoxic patients with hyperthyroidism. Leakage of the iodine-rich contents of the colloid into the bloodstream expands the iodide pool in the circulation; administered radioiodine will be diluted in the expanded iodide pool, which in conjunction with damage to the thyrocytes causes a low thyroidal radioiodine uptake. The efficacy of treatment with radioiodine or antithyroid drugs is consequently very low.

Subacute, painless, and postpartum thyroiditis run a self-limited course, in which the inflammation gradually subsides under restoration of the normal thyroid architecture (see Chapter 3.2.6). Thyrotoxicosis associated with thyroiditis is usually mild, lasting for only a few weeks, and either no treatment or treatment with a β-adrenoceptor antagonist is sufficient. Short-term salicylates, or glucocorticoids in more resistant cases, may be required to relieve neck pain in subacute thyroiditis. The thyrotoxic phase may be followed by a hypothyroid phase in any type of thyroiditis and can last for 1–4 months. Thyroxine treatment may be warranted in symptomatic patients, but should be withdrawn after 6 months because most patients will spontaneously regain euthyroidism. Postpartum thyroiditis is very likely to recur after a subsequent pregnancy. Permanent hypothyroidism develops in about one-third of patients with silent or postpartum thyroiditis.

Radiation-induced thyroiditis may occur after treatment with large doses of ^{131}I. It develops in the first 2 weeks, and is characterized by neck and ear pain, painful swallowing, thyroid swelling, and tenderness. Thyrotoxicosis is mild and transient, resolving spontaneously within a week or two, and not requiring any specific treatment besides salicylates for mild pain. Glucocorticoids (e.g. 30 mg prednisone daily) may be given for severe pain or swelling, tapering the dose when the complaints have disappeared.

Cytokine-induced thyroiditis has been observed after the administration of interleukin-2, interferon-α (IFNα), or granulocyte-macrophage colony stimulating factor. The clinical picture may resemble that of destructive thyroiditis: transient thyrotoxicosis developing after a few weeks to months, followed by a hypothyroid phase which may be transient as well. Radioiodine thyroid uptake is low and a small non-tender goitre is sometimes present. However, it is not uncommon that the hypothyroid phase is permanent, or develops much later without preceding thyrotoxicosis. In patients treated with IFNα for chronic hepatitis C, *de novo* occurrence of thyroid antibodies is observed in 10–14% [1–3]. Thyroid function disorders develop in 15%, due to autoimmune hypothyroidism (*c.*50%), Graves' hyperthyroidism (*c.*25%), or destructive thyroiditis (*c.*25%); however, only a small minority needs treatment as spontaneous recovery is the rule rather than the exception. These abnormalities have a median date of onset 17 weeks after starting treatment, but can occur at any time. Risk factors are female sex and pre-existent thyroid antibodies, but IFNα dosage or efficacy are not. It is recommended to measure thyroid-stimulating hormone (TSH) and TPO-Ab before treatment, and to monitor TSH during treatment (e.g. every 3 months). IFNα-induced Graves' disease does not resolve spontaneously after discontinuation of IFNα, and thyroid ablation with ^{131}I or surgery is preferred. IFNα-induced destructive thyroiditis is mostly mild or subclinical and resolves spontaneously, although permanent hypothyroidism develops in less than 5% of cases. IFNα treatment can usually be continued, except in very severe cases. Corticosteroids are generally contraindicated in patients with hepatitis C. Rechallenge with IFNα may result in recurrent destructive thyroiditis.

Immunotherapy with **checkpoint inhibitors** (like the CTLA4-inhibitor ipilimumab, and the PD-1 inhibitors nivolumab and pembrolizumab) improves survival in a number of cancers but may induce various endocrinopathies, including thyroid function disorders and destructive thyroiditis resembling those induced by interferon-α [4].

Trauma of the thyroid gland may rarely result in transient thyrotoxicosis associated with a low radioiodine thyroid uptake. Infiltration of the thyroid gland with malignant lymphoma or cancer metastases may cause thyrotoxicosis associated with low radioiodine thyroid uptake in exceptional cases, due to invasion and disruption of thyroid follicles.

Iodide-Induced Thyrotoxicosis

Exposure to pharmacological quantities of iodine may cause occasionally iodide-induced thyrotoxicosis (IIT), also called Jod-Basedow [5–8]. IIT is more prevalent in iodine-deficient areas than in iodine-sufficient areas. Risk factors are old age, nodular goitre, and low TSH. In case of underlying thyroid disease (like in patients with multinodular goitre or latent Graves' disease and subclinical hyperthyroidism), thyroidal radioiodine uptake may be normal or even high. In contrast, in patients without underlying thyroid disease, ITT is caused by the cytotoxic effect of iodine excess on thyrocytes resulting in destructive thyroiditis, and radioiodine uptake will be low. IIT usually resolves spontaneously within 6 months, often after a hypothyroid phase. The efficacy of antithyroid drugs is impaired by the presence of iodine excess, and ß-blockers or no treatment at all may suffice.

Many drugs contain huge quantities of iodine, such as kelp tablets, certain expectorants, and iodinated contrast agents. The risk

for developing IIT is 0.3% in unselected people after coronary angiography in iodine-deficient areas [9]. In patients with an autonomous thyroid function before coronary angiography, treatment with 20 mg thiamazole and/or 900 mg sodium perchlorate, starting the day before angiography and continued for 2 weeks, was not very effective in preventing IIT [10, 11]. Close monitoring of high-risk patients rather than prophylaxis is thus recommended, with prescription of β-adrenoceptor antagonists if thyrotoxicosis occurs [12].

Amiodarone-Induced Thyrotoxicosis

One tablet of 200 mg amiodarone contains 74.4 mg iodine of which 10% is released *in vivo* during biotransformation of the drug. A maintenance dose of 300 mg amiodarone daily results in a 40-fold rise of plasma inorganic iodide and urinary iodide excretion [13]. The iodine excess may induce thyrotoxicosis in patients with underlying thyroid disease, referred to as amiodarone-induced thyrotoxicosis (AIT) type 1. Amiodarone and especially its main metabolite desethylamiodarone have a potent cytotoxic effect on thyrocytes causing destructive thyroiditis [14, 15]; it may give rise to a destructive type of thyrotoxicosis, referred to as AIT type 2. The analogy with IIT is obvious, but the molar concentrations required for the cytotoxic effect of amiodarone are about three times lower than for potassium iodide [14].

Epidemiology and screening. Whereas amiodarone-induced hypothyroidism (AIH) is most prevalent in iodine-sufficient regions, AIT is more prevalent in iodine-deficient areas. A prospective study in the Netherlands, a country with an average daily iodine intake of 150 μg, demonstrated development of AIH in 6% and of AIT in 8% of patients starting amiodarone [16]. It means one out of every five to six patients taking amiodarone will develop overt thyroid dysfunction. It is striking that no new cases of AIH develop after 2 years on amiodarone therapy, whereas new cases of AIT continue to occur with continuation of amiodarone. The incidence rate/100 person-years in France and the Netherlands is 4.6 and 1.9 for AIH, respectively, and 1.6 and 1.6 for AIT [16, 17]. The high incidence and the potential danger of worsening of heart disease upon occurrence of AIH or AIT [18] call for thyroid monitoring. It is therefore recommended to measure TSH and TPO-Ab at baseline, and TSH every 6 months thereafter [19, 20]. A normal serum TSH during follow-up, however, does not guarantee that AIT will not develop in the interval to the next visit in view of the unpredictable and often sudden onset of AIT [21]. Furthermore, the finding of a suppressed TSH during follow-up does not necessarily mean AIT that has to be treated, because in one-half of these cases TSH returns spontaneously to normal values [21, 22]. AIT can develop up to 12 months after discontinuation of amiodarone, related to the very long terminal half-life of the drug.

Diagnosis. AIT can be asymptomatic. Recurrence of cardiac arrhythmias, which previously had been controlled, may suggest the diagnosis. Symptoms can be unexplained weight loss (50%), heavy sweating (42%), palpitations (37%), hyperkinesia (29%), muscle weakness (27%), heat intolerance (24%), overall weakness (12%), and diarrhoea (12%) [23]. The biochemical diagnosis of AIT is based on a suppressed TSH in combination with an elevated free T_4. T_3 can be elevated or normal (related to the strong inhibition of T_4 deiodination into T_3 by amiodarone), and cases of T_4 toxicosis do occur. The fT3 to fT_4 ratio in AIT (like in IIT and subacute thyroiditis) is thus much lower than in Graves' hyperthyroidism.

Distinction between AIT subtypes is considered to be useful because management of types 1 and 2 is different (Table 3.3.12.1) [24]. However, none of the proposed methods accurately discriminates between both subtypes. Serum interleukin-6 was originally advocated as a good discriminator (being much higher in type 2 than in type 1), but subsequent studies have been unable to confirm its value. Thyroidal radioiodine uptake is low or absent in type 2, but can also be low in type 1, and in one study did not differ at all between both types. Colour flow Doppler sonography can be useful, revealing a patchy pattern of thyroid vascularity to a markedly increased blood flow in type 1 and an absent blood flow in type 2 [25]. The latest tool has been the ^{99m}Tc-sestamibi scan, showing mostly increased MIBI retention in type 1 and no uptake in type 2 [26]. It should be admitted that sometimes distinction between type 1 and type 2 AIT fails, and mixed cases do occur in about 10–15% [22].

Treatment. Guidelines for the management of AIT have been published by the European Thyroid Association (Figure 3.3.12.1)

Table 3.3.12.1 Characteristics of amiodarone-induced thyrotoxicosis types 1 and 2

	Type 1	Type 2
Underlying thyroid abnormality	Yes	No
Pathogenesis	Iodide-induced thyrotoxicosis	Destructive thyrotoxicosis
Physical examination	Usually nodular or diffuse goitre	Occasionally small diffuse firm goitre
Thyroid antibodies	Can be present	Mostly absent
Thyroid ultrasound	Diffuse or nodular goitre	Heterogeneous pattern
Doppler sonography	Normal or increased blood flow	Decreased blood flow
Thyroidal radioiodine uptake	Low or normal	Low or absent
^{99m}Tc-sestamibi scan	Clear thyroid retention	No thyroid uptake
Spontaneous remission	Unlikely	Likely
Preferred drug treatment	$NaClO_4$ + methimazole	Glucocorticoids
Amiodarone medication	Discontinue if possible	Can be continued
Subsequent hypothyroidism	Unlikely	Possible

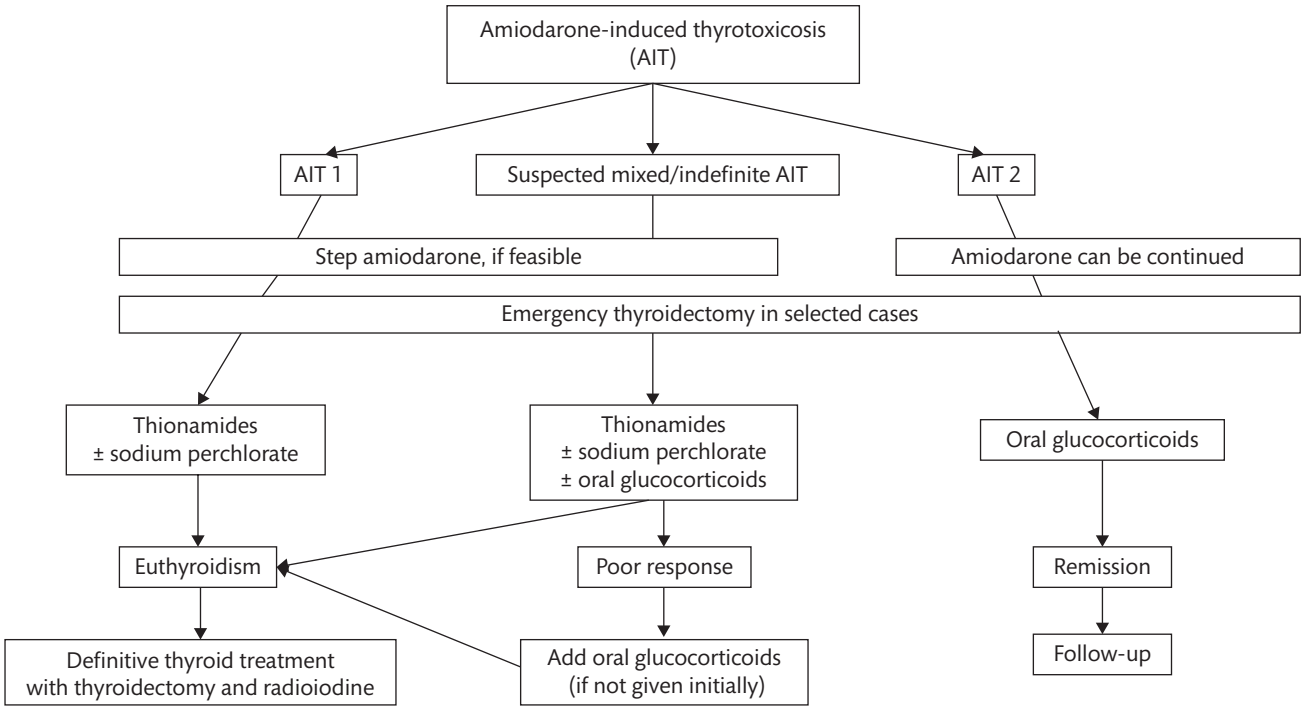


Figure 3.3.12.1 Algorithm for the management of amiodarone-induced thyrotoxicosis.
Reproduced with permission from Bartalena L, Bogazzi F, Chiovato L, Hubalewska-Dydejczyk A, Links TP, Vanderpump M. 2018 European Thyroid Association (ETA) guidelines for the management of amiodarone-associated thyroid dysfunction. *Eur Thyroid J*, 2018; 7: 55–66. Copyright © 2018 S. Karger AG, Basel.

[24]. Amiodarone should be stopped in AIT type 1. Older studies indicate most patients with AIT type 1 are still thyrotoxic 6–9 months after discontinuation of amiodarone; in contrast, many patients with AIT type 2 are cured spontaneously in 3–5 months, in line with the self-limiting nature of AIT type 2 [22]. A recent prospective controlled trial has indeed demonstrated the feasibility of continuation of amiodarone in AIT type 2 [27].

AIT type 1 should be treated with rather high doses of antithyroid drugs (40–60 mg methimazole daily) because the ambient iodine excess renders the thyroid gland less sensitive to methimazole. One may opt to co-administer sodium perchlorate, which by acute inhibition of thyroidal iodide uptake decreases intrathyroidal iodide thereby increasing the efficacy of methimazole [24]. The recommended dose is 500 mg sodium perchlorate twice daily for a restricted period of time (12 weeks). Using this schedule, dreaded side effects of $NaClO_4$ like agranulocytosis have not been encountered.

AIT type 2 should be treated with oral prednisone in a starting dose of 30 mg daily, and tapering when fT_4 and/or TSH have been normalized (which occurs after a median time of 30 and 90 days, respectively) [28]. Time to restore euthyroidism was much longer after iopanoic acid therapy compared to prednisone in an RCT (221 vs. 40 days, respectively) [29]. Time to euthyroidism is little affected if amiodarone is continued [27]. $KClO_4$ inhibits the cytotoxic effect of amiodarone on thyrocytes *in vitro*, but adding the drug to steroids did not improve outcome [15, 27]. It may not always be necessary to treat mild cases of AIT type 2 in view of its self-limiting nature. In this context one should be reminded that slightly increased fT_4 levels up to 25 pmol/L in the presence of a normal TSH are not unusual during amiodarone treatment, caused by a decreased T_4 metabolic clearance rate induced by amiodarone. Smaller thyroid volumes and modest increases of fT_4 are predictors of a fast response to steroids in AIT type 2 [28], and may help in choosing an expectant or active policy.

In mixed/indefinite AIT, one can start with methimazole ± perchlorate, and add glucocorticoids if there is no improvement in 4–6 weeks. Alternatively, one can start immediately with triple therapy. AIT sometimes has a rather severe course, and fatal cases do occur. In very severe cases and in patients resistant to medical therapy, one should not wait too long to perform total thyroidectomy. Despite compromised cardiac function, the surgery has a rather low mortality and morbidity [30–33]. Total thyroidectomy improves cardiac function and mortality, especially in those with ejection fraction <40% preoperatively [32]. ^{131}I therapy may sometimes be feasible despite the iodine excess, by applying high doses [34, 35]. rhTSH prior to ^{131}I therapy is not recommended in AIT type 1 [24].

When euthyroidism has been restored in AIT type 1, definitive thyroid treatment with thyroidectomy or radioiodine is recommended [24, 36]. When AIT type 2 has been cured and amiodarone needs to be restarted, prophylactic thyroid ablation is not recommended. If amiodarone had been continued in AIT type 2, recurrent AIT may occur in 6–18%; recurrences are apparently less severe and more easy to handle [27, 37, 38].

Prognosis. Patients with AIT have more major adverse cardiovascular events than patients who remain euthyroid (32% vs. 11%, P <0.001) [18]. Mortality is also higher in AIT patients with severe left ventricle dysfunction (31% at ejection fraction <50% vs. 14% at ejection fraction >50%) [39]. Permanent hypothyroidism develops in 17% of cured AIT type 2 patients, occurring at 10 months (range 6–24) after reaching euthyroidism [27, 40].

Thyrotoxicosis of Extrathyroidal Origin

Factitious thyrotoxicosis is due to ingestion of excess thyroid hormone [41]. The hormone can be ingested unintentionally in diet pills and in ground beef contaminated with bovine thyroid tissue, or intentionally by people (mostly women) with psychiatric disturbances and by children as an accident [42, 43]. Characteristic features are thyrotoxicosis associated with a low thyroidal radioiodine uptake, normal urinary iodine excretion, no goitre, and no thyroid antibodies. Strong evidence for the existence of factitious thyrotoxicosis is the finding of a low serum thyroglobulin. The disease will resolve spontaneously after the ingestion of excess thyroid hormone is stopped. Thyrotoxicosis due to the ingestion of a well-cooked 227-g hamburger prepared from contaminated ground beef, disappears within 1 month. Symptomatic treatment with propranolol may be necessary. Acute thyroxine intoxication may benefit from gastric lavage; plasmapheresis has been recommended in life-threatening situations. Discrepancy is often noted between modest clinical toxicity and very high thyroid hormone concentrations in serum.

Struma ovarii is an ovarian tumour with thyroid tissue as an important constituent. These tumours are often part of a mature teratoma, and harbour thyroid cancer in under 10% of instances [44]. They are mostly unilateral, occurring in less than 1% of all ovarian tumours. The highest incidence is in the fourth to sixth decade. Thyrotoxicosis develops in 5–15%, especially in tumours bigger than 30 mm in size. Radioiodine uptake is low in the neck and high in the abdomen at the side of the lesion in classic cases, but the presence of a goitre (with some uptake in the neck) is not uncommon. Treatment is by surgical removal of the tumour [45].

Functional metastases of differentiated thyroid cancer are a rare cause of thyrotoxicosis due to a large bulk of tumour. Treatment is primarily by radioiodine [46].

REFERENCES

1. Dalgard O, Bjøro K, Hellum K, *et al*. Thyroid dysfunction during treatment of chronic hepatitis C with interferon-alpha: no association with either interferon dosage or efficacy of therapy. *J Intern Med*, 2002; **251**: 400–6.
2. Wong V, Fu AX, George J, Cheung NW. Thyrotoxicosis induced by alpha-interferon therapy in chronic viral hepatitis. *Clin Endocrinol*, 2002; **56**: 793–8.
3. Hwang Y, Kim W, Kwon SY, Yu HM, Kim JH, Choe WH. Incidence of and risk factors for thyroid dysfunction during peginterferon α and ribavirin treatment in patients with chronic hepatitis C. *Korean J Intern Med,* 2015; **30**: 792–800.
4. Higham CE, Olsson-Brown A, Carroll P, *et al*. Acute management of the endocrine complications of checkpoint inhibitor therapy. *Endocr Conn,* 2018; **7**: G1–7.
5. Fradkin JE, Wolff J. Iodide-induced thyrotoxicosis. *Medicine*, 1983; **62**: 1–20.
6. Leger AF, Massin JP, Laurent MF, *et al*. Iodine-induced thyrotoxicosis: analysis of eighty-five consecutive cases. *Eur J Clin Invest*, 1984; **14**: 449–55.
7. Stanbury JB, Ermans AE, Bourdoux P, *et al*. Iodine-induced hyperthyroidism: occurrence and epidemiology. *Thyroid*, 1998; **8**: 83–100.
8. Roti E, Uberti ED. Iodine excess and hyperthyroidism. *Thyroid,* 2001; **11**: 493–500.
9. Hintze G, Blombach O, Fink H, Burkhardt U, Köbberling J. Risk of iodine-induced thyrotoxicosis after coronary angiography: an investigation in 788 unselected subjects. *Eur J Endocrinol*, 1999; **140**: 264–7.
10. Nolte W, Müller R, Siggelkow H, Emrich D, Hüfner M. Prophylactic application of thyrostatic drugs during excessive iodine exposure in euthyroid patients with thyroid autonomy: a randomized study. *Eur J Endocrinol*, 1996; **134**: 337–41.
11. Fricke E, Fricke H, Esdorn E, *et al*. Scintigraphy for risk stratification of iodine-induced thyrotoxicosis in patients receiving contrast agent for coronary angiography: a prospective study of patients with low thyrotropin. *J Clin Endocrinol Metab*, 2004; **89**: 6092–6.
12. van der Molen AJ, Thomsen HS, Morcos SK. Effect of iodinated contrast media on thyroid function in adults. *Eur Radiol*, 2004; **14**: 902–7.
13. Rao RH, McReady VR, Spathis GS. Iodine kinetic studies during amiodarone treatment. *J Endocrinol Metab*, 1986; **62**: 563–7.
14. Chiovato L, Martino E, Tonacchera M, *et al*. Studies on the *in vitro* cytotoxic effect of amiodarone. *Endocrinology*, 1994; **134**: 2277–82.
15. Brennan MD, Erickson DR, Carney JA, Bahn RS. Nongoitrous (type I) amiodarone-associated thyrotoxicosis: evidence of follicular disruption *in vitro* and *in vivo*. *Thyroid*, 1995; **5**: 177–83.
16. Ahmed S, van Gelder IC, Wiesfeld AC, van Veldhuisen DJ, Links TP. Determinants and outcome of amiodarone-associated thyrotoxicosis and hypothyroidism. *Clin Endocrinol*, 2011; **75**: 388–94.
17. Bongard V, Marc D, Philippe V, Jean-Louis M, Maryse LM. Incidence rate of adverse drug reactions during long-term follow-up of patients newly treated with amiodarone. *Am J Ther*, 2006; **13**: 315–19.
18. Yiu KH, Jim MH, Siu C-W, *et al*. Amiodarone-induced thyrotoxicosis is a predictor of adverse cardiovascular outcome. *J Clin Endocrinol Metab*, 2009; **94**: 109–14.
19. Bahn RS, Burch HB, Cooper DS, *et al*. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid*, 2011; **21**: 593–646.
20. Benjamins S, Dullaart RPF, Sluiter WJ, Rienstra M, van Gelder IC, Links TP. The clinical value of regular thyroid function tests during amiodarone treatment. *Eur J Endocrinol*, 2017; **177**: 9–14.
21. Trip MD, Wiersinga W, Plomp TA. Incidence, predictability, and pathogenesis of amiodarone-induced thyrotoxicosis and hypothyroidism. *Am J Med*, 1991; **91**: 507–11.
22. Eskes SA, Wiersinga WM. Amiodarone and thyroid. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 735–51.
23. Conen D, Melly L, Kaufmann C, *et al*. Amiodarone-induced thyrotoxicosis: clinical course and predictors of outcome. *J Am Coll Cardiol*, 2007; **49**: 2350–5.
24. Bartalena L, Bogazzi F, Chiovato L, Hubalewska-Dydejczyk A, Links TP, Vanderpump M. 2018 European Thyroid Association (ETA) guidelines for the management of amiodarone-associated thyroid dysfunction. *Eur Thyroid J*, 2018; **7**: 55–66.
25. Bogazzi F, Bartalena L, Brogioni S, *et al*. Color flow Doppler sonography rapidly differentiates type I and type II amiodarone-induced thyrotoxicosis. *Thyroid*, 1997; **7**: 541–5.
26. Piga M, Cocco MC, Serra A, Boi F, Loy M, Mariotti S. The usefulness of 99mTc-sestaMIBI thyroid scan in the differential diagnosis and management of amiodarone-induced thyrotoxicosis. *Eur J Endocrinol*, 2008; **159**: 423–9.

27. Eskes SA, Endert E, Fliers E, *et al*. Treatment of amiodarone-induced thyrotoxicosis type 2: a randomized clinical trial. *J Clin Endocrinol Metab*, 2012; **87**: 499–506.
28. Bogazzi F, Bartalena L, Tomisti L, *et al*. Glucocorticoid response in amiodarone-induced thyrotoxicosis resulting from destructive thyroiditis is predicted by thyroid volume and serum free thyroid hormone concentrations. *J Clin Endocrinol Metabol*, 2007; **92**: 556–62.
29. Bogazzi F, Bartalena L, Cosci C, *et al*. Treatment of type II amiodarone-induced thyrotoxicosis by either iopanoic acid or glucocorticoids: a prospective, randomized study. *J Clin Endocrinol Metab*, 2003; **88**: 1999–2002.
30. Houghton SG, Farley DR, Brennan MD, van Heerden JA, Thompson GB, Grant CS. Surgical management of amiodarone-induced thyrotoxicosis: Mayo Clinic experience. *World J Surg*, 2004; **24**: 1083–7.
31. Gough J, Gough IR. Total thyroidectomy for amiodarone-associated thyrotoxicosis in patients with severe cardiac disease. *World J Surg,* 2006; **30**: 1957–61.
32. Tomisti L, Materazzi G, Bartalena L, *et al*. Total thyroidectomy in patients with amiodarone-induced thyrotoxicosis and severe left ventricular systolic dysfunction. *J Clin Endocrinol Metab*, 2012; **97**: 3515–21.
33. Kaderli RM, Fahrner R, Christ Er, *et al*. Total thyroidectomy for amiodarone-induced tyhrotoxicosis in the hyperthyroid state. *Exp Clin Endocrinol Diab,* 2016; **124**: 45–8.
34. Albino CC, Paz-Filho G, Graf H. Recombinant human TSH as an adjuvant to radioiodine for the treatment of type 1 amiodarone-induced thyrotoxicosis (AIT). *Clin Endocrinol*, 2009; **70**: 810–11.
35. Gursoy A, Tutuncu NB, Gencoglu A, Anil C, Demirer AN, Demirag NG. Radioactive iodine in the treatment of type 2 amiodarone-induced thyrotoxicosis. *J Natl Med Assoc*, 2008; **100**: 706–19.
36. Maqdasy S, Batisse-Lignier M, Auclair C, *et al*. Amiodarone-induced thyrotoxicosis recurrence after amiodarone reintroduction. *Am J Cardiol*, 2016; **117**: 1112–16.
37. Sato K, Shiga T, Matsuda N, *et al*. Mild and short recurrence of type II amiodarone-induced thyrotoxicosis in three patients receiving amiodarone continuously for more than 10 years. *Endocrine J*, 2006; **53**: 531–8.
38. Uzan L, Guignat L, Meune C, *et al*. Continuation of amiodarone therapy despite type II amiodarone-induced thyrotoxicosis. *Drug Saf*, 2006; **29**: 231–6.
39. O'Sullivan AJ, Lewis M, Diamond T. Amiodarone-induced thyrotoxicosis: left ventricular dysfunction is associated with increased mortality. *Eur J Endocrinol,* 2006; **154**: 533–6.
40. Bogazzi F, Dell'Unto E, Tanda MI, *et al*. Long-term outcome of thyroid function after amiodarone-induced thyrotoxicosis, as compared to subacute thyroiditis. *J Endocrinol Invest*, 2006; **29**: 694–9.
41. Cohen JH, Ingbar SH, Braverman LE. Thyrotoxicosis due to ingestion of excess thyroid hormone. *Endocr Rev*, 1989; **10**: 113–24.
42. Kaiserman I, Avni M, Sack J. Kinetics of the pituitary–thyroid axis and the peripheral thyroid hormones in 2 children with thyroxine intoxication. *Horm Res*, 1995; **44**: 229–37.
43. Hedberg CW, Fishbein DB, Janssen RS, *et al*. An outbreak of thyrotoxicosis caused by the consumption of bovine thyroid gland in ground beef. *N Engl J Med*, 1987; **316**: 993–8.
44. Wei S, Baloch ZW, LiVolsi VA. Pathology of struma ovarii: a report of 96 cases. *Endocr Pathol,* 2015; **26**: 342–8.
45. Wee JY, Li X, Chem BS, Chua IS. Struma ovarii: management and follow-up of a rare ovarian tumor. *Singapore Med J*, 2015; **56**: 35–9.
46. Guglielmi R, Pacella CM, Dottorini ME, *et al*. Severe thyrotoxicosis due to hyperfunctioning liver metastasis from follicular carcinoma: treatment with ^{131}I and interstitial laser ablation. *Thyroid*, 1999; **9**: 173–7.

3.4

Hypothyroidism

3.4.1 Clinical Assessment and Systemic Manifestations of Hypothyroidism

Massimo Tonacchera and Luca Chiovato

Introduction

Hypothyroidism may affect people of both sexes and all ages. The clinical expression of thyroid hormone deficiency varies considerably between individuals. It is influenced mainly by the age of the patient and the rate at which hypothyroidism develops although being largely independent of its cause. Most adult patients complain of a slowing of physical and mental activity.

Hypothyroidism is a graded phenomenon, ranging from very mild cases, in which biochemical abnormalities (see Chapter 3.4.4, 'Subclinical Hypothyroidism') are present but the individual hardly notices symptoms and signs of thyroid hormone deficiency, to very severe cases in which the danger exists of sliding down into a life-threatening myxoedema coma.

Organ System Manifestations of Hypothyroidism

Cutaneous Manifestations and Changes in the Connective Tissues

The cutaneous changes observed in hypothyroidism belong to the most classic and frequent findings of the disease (Table 3.4.1.1) [1]. Although other important symptoms and signs of hypothyroidism may be present, changes in the skin may be the most important factor for seeking medical attention [1]. In over 80% of patients with primary hypothyroidism, the epidermis is dry, rough, cool, and covered with fine superficial scales. This is an expression of decreased cutaneous metabolism, reduced secretion of sweat and sebaceous glands, vasoconstriction, thinning of the epidermis, and hyperkeratosis of the stratum corneum. The skin may have a finely wrinkled, parchment-like character. Unusual coldness of the arms and legs is sometimes a subject of complaint. The palms are cool and dry. Subcutaneous fat may be increased, with the formation of definite fat pads, especially above the clavicles, but is conspicuously absent in the more advanced form of the disease (myxoedematous cachexia). The hands and feet have a broad appearance, due to thickening of subcutaneous tissue.

Table 3.4.1.1 Cutaneous signs and symptoms of hypothyroidism

Cutaneous manifestation	Frequency (%)
Cold intolerance	50–95
Nail abnormality	90
Thickening and dryness of hair and skin	80–90
Oedema of hands, face, and eyelids	70–85
Change in shape of face	70
Malar flush	50
Non-pitting oedema	30
Alopecia	30–40
Pallor	25–60
Decreased sweat secretion	10–70

The diffuse pallor and pale waxy surface colour can be attributed to two mechanisms. First vasoconstriction occurs and second the excess fluid and mucopolysaccharides in the dermis may compress small vessels to create blanching as well as interference with the transmission of colour from the deeper vessels. Anaemia may also contribute to pallor. Yellowish discolouration of the skin, most notably of the palms, soles, and nasolabial folds, occurs in patients with long-standing hypothyroidism and is caused by elevation of serum and tissue carotene concentrations. The face is puffy, pale, and expressionless at rest (Figure 3.4.1.1). The skin of the face is also parchment-like. In spite of the swelling, it may be traced with fine wrinkles, particularly in pituitary myxoedema. The swelling sometimes gives it a round or moonlike appearance. The palpebral fissure maybe narrowed because of blepharoptosis, due to diminished tone of the sympathetic nervous fibres to Müller's elevator palpebral

Figure 3.4.1.1 A patient with hypothyroidism.

superior muscle. The modest measurable exophthalmos seen in some patients with myxoedema is presumably related to accumulation of the same mucous oedema in the orbit as is seen elsewhere.

The tongue is usually large, and some patients will complain of this problem. The tongue is smooth if pernicious anaemia coexists. The voice is husky, low-pitched, and coarse due to the enlargement of the tongue and thickening of the pharyngeal and laryngeal mucous membranes. The speech is deliberate and slow, and there may be difficulty in articulation.

There are other, less common, cutaneous findings seen in adult hypothyroid patients. Six patients have been reported in literature of an acquired palmoplantar keratoderma, verrucous in character, and predominantly affecting the plantar surface [2]. An additional reported cutaneous finding specifically linked to atrophic thyroiditis is dermatitis herpetiformis, a gluten-sensitive skin disease characterized by blisters on the elbows, buttock, and knees [3].

Hair Follicles and Nails

The hair is dry, dull, and coarse, growing slowly, becoming sparse, and falling out readily. Loss of scalp, genital, and beard hair may also occur. Hair may be lost from the temporal aspects of the eyebrows (Queen Anne's sign). However, this sign is not uncommon in elderly euthyroid women and occurs in association with several types of cutaneous disease, including atopic dermatitis, seborrhoeic dermatitis, and lupus erythematosus. In men, the beard becomes sparse, and its rate of growth becomes greatly retarded. The scalp is dry and scaly. The nails, through retardation of growth, become thickened and brittle, striated both in transverse and longitudinal grooves, and show frequent deformities.

Dermal Changes

The dermal pathological findings in patients with hypothyroidism are clinically manifested by the non-pitting swelling, most marked around the eyes and hands, that is myxoedema. This is due to an abnormal accumulation of salts, mucopolysaccharides, and protein in the interstitial spaces of the skin [4]. Histopathological examination of the skin reveals that the connective tissue fibres are separated by an increased amount of metachromatically staining, periodic acid-Schiff-positive mucinous material [1]. This material consists of protein complexed with two mucopolysaccharides, hyaluronic acid and chondroitin sulphate B [1]. An increase in the synthesis and accumulation of glycosaminoglycans leads to an excess of these normal intercellular substances.

The glycosaminoglycans are polymers of D-glucuronic acid and *N*-acetyl-D-glucosamine, forming hyaluronic acid, or of L-hyaluronic acid and *N*-acetyl-D-galactosamine sulphate, forming chondroitin sulphate B. They exist free and in ionic or covalent linkage to proteins. These mucoproteins comprise part of the normal non-fibrillar intercellular matrix, the ground substance holding cells together. Due to its strong water binding capacity, accumulated hyaluronic acid may also contribute to the peculiar non-pitting quality of myxoedema. Capillary permeability is augmented in hypothyroidism with increased accumulation of sodium, water, and proteins.

Cardiovascular Changes

Lack of thyroid hormones causes multiple alterations in the cardiovascular system [5]. The most frequent changes in hypothyroid patients are increased systemic vascular resistance, diastolic dysfunction, reduced systolic function, and decreased cardiac preload [5]. Bradycardia, cardiomegaly, and low voltage complexes on the electrocardiogram (ECG) are well-known features (**Box 3.4.1.1**). The decrease in pulse rate approximately parallels the decrease in the body's metabolic rate. Myocardial contractility is reduced. The cardiac output at rest is decreased because of reduction in both stroke volume and heart rate, reflecting loss of the inotropic and chronotropic effects of thyroid hormones [5]. The mechanism responsible for the impaired ventricular performance is multifactorial. The actions of thyroid hormone on the hearth are mediated by genomic and non-genomic mechanisms targeted to membrane proteins, cytoskeletal components, and organelles. The genomic actions are mediated by nuclear receptors alpha and beta [5]. In animal models, low thyroid hormone concentrations alter the expression of myocyte-specific genes and the distribution of the heavy-chain isoforms of sarcomeric myosin and of the calcium-regulating proteins [6]. Alterations in myocyte calcium uptake and release are responsible for the change in the inotropic state [6]. Peripheral vascular resistance at rest is increased, and blood volume is reduced. These haemodynamic alterations cause narrowing of pulse pressure, prolongation of circulation time, and decreased blood flow to the tissues. In most tissues the decrease in blood flow is proportional to the decrease in oxygen consumption, so the arteriovenous oxygen difference remains normal or may be slightly increased. Slow peripheral circulation, and therefore more complete extraction of

Box 3.4.1.1 Cardiovascular signs and symptoms in hypothyroidism

- Symptoms
  - Dyspnoea
  - Decreased exercise tolerance
  - Angina
- Signs
  - Low pulse rate
  - Increased systemic vascular resistance
  - Diastolic hypertension
  - Cardiomegaly
  - Pericardial effusion
  - Peripheral non-pitting oedema
  - Low voltage ECG, non-specific ST-T changes

oxygen, as well as anaemia, may be responsible for the increased arteriovenous oxygen difference. Myocardial oxygen consumption is decreased, usually more than blood supply to the myocardium, so that angina is infrequent [5]. In some patients a reduction in cardiac output greater than the decline in oxygen consumption indicates specific cardiac damage from the myxoedema [5].

The haemodynamic alterations at rest resemble those of congestive heart failure, but cardiac output increases and peripheral vascular resistance decreases normally in response to exercise unless the hypothyroid state is severe [5]. The non-pitting oedema observed in hypothyroid patients is due to an increase in protein distribution in the extravascular extracellular space resulting from increased capillary permeability.

Venous pressure is normal, but peripheral resistance is increased. The mechanism responsible for the increase in systemic vascular resistance is not known. Triiodothyronine (T_3) may act as a vasodilator and in its absence vascular resistance may rise [5, 7]. Arterial blood pressure is often mildly increased. Hypertension is present in 10–20% of patients with hypothyroidism [5, 7]. Diastolic hypertension is usually restored to normal after treatment [7]. Three factors can contribute to systemic hypertension, increased peripheral resistance, increased arterial stiffness, and endothelial dysfunction [7–9].

Few symptoms referable to the cardiovascular system are referred in patients with hypothyroidism. Exertional dyspnoea and exercise intolerance are probably due to skeletal muscle dysfunction. There has been much discussion as to whether the hypercholesterolaemia that accompanies primary hypothyroidism accelerates the development of coronary atherosclerosis. An increased risk for atherosclerosis is supported by autopsy and epidemiological studies in patients with thyroid hormone deficiency and may be, in part, explained by the hypercholesterolaemia and marked increase in low-density lipoprotein (LDL) [8]. Moreover, diastolic hypertension, increased arterial stiffness and endothelial dysfunction, altered coagulability, and increased levels of C-reactive protein may further contribute to the increased cardiovascular risk (8, 9). Most autopsied myxoedematous individuals have severe atherosclerosis, but they are also usually 60 years of age or more. Occasionally angina pectoris is encountered in myxoedema [5]. Sometimes angina or angina-like pain is present before treatment. This generally indicates the presence of significant coronary artery disease since there is inadequate myocardial oxygenation despite reduced cardiac output and oxygen utilization. Angina may also appear for the first time after treatment has been initiated, indicating that coronary flow is inadequate for resumption of normal cardiac function [7]. The presence of a structural lesion must be strongly suspected.

On physical examination certain findings can suggest hypothyroidism. The heart rate is lowered, the pulse pressure is narrowed, and the carotid upstroke and left ventricular apical impulse are diminished [5]. The heart sounds are diminished in intensity; this finding is due largely to effusion into the pericardial sac of fluid rich in protein and glycosaminoglycans.

The combination of a large heart, associated with typical haemodynamic and electrocardiographic alterations, and the serum enzyme changes (creatine kinase, aspartate aminotransferase, and lactate dehydrogenase may be increased) has been termed myxoedema heart. This term was introduced by Zondek in 1918 [10]. It embraced dilatation of the left and right sides of the heart, a slow indolent heart action with normal blood pressure, and lowering of the P and T waves of the electrocardiogram. Zondek found that after treatment with thyroid hormone there was a return of the dilated heart to near normal size, a more rapid pulse without change in blood pressure, and gradual return of the P and T waves to normal. Microscopic examination discloses myxoedematous changes of the myocardial fibres. The myocardium is pale and flabby. Histopathological examination of the myocardium reveals interstitial oedema and swelling of the muscle fibres with loss of striations. The cause of the cardiac enlargement has been disputed. It is not due to hypertrophy alone, since it would not disappear so rapidly with treatment. One factor may be a decrease in contractility of the heart muscle; this would require a lengthening of muscle fibres in order to perform the required work.

In myxoedema, when the heart does not return to a normal size under thyroid hormone administration, hypertrophy due to some other disease is present as a complication. The slow and progressive return to normal size under treatment requires between 3 weeks and 10 months for completion. This decrease in size, like the progressive elevation of the T waves, is of diagnostic value.

Electrocardiographic Changes

Electrocardiographic changes include sinus bradycardia, prolongation of the P–R interval, low amplitude of the P wave and QRS complex, alterations of the ST segment, and flattened or inverted T waves. Although suggestive of myocardial ischaemia, these waveform changes often disappear during thyroxine (T_4) treatment. Pericardial effusion is probably responsible for the low amplitude. Rarely, complete heart block may be present, but this disappears when the hypothyroidism is treated. In hypothyroidism, the atrial pacemaker function is normal and atrial ectopy is rare, but ventricular premature beats and occasionally ventricular tachycardia may occur. The syndrome of *torsades de pointes* with a long Q–T interval and ventricular tachycardia can occur with hypothyroidism, and resolve with T_4 treatment alone.

Systolic Time Intervals and Echographic Findings

Systolic time intervals are altered, the pre-ejection period is prolonged, and the ratio of pre-ejection period to left ventricular ejection time is increased. Some patients have been reported to have asymmetrical hypertrophy of the intraventricular septum by echocardiography that resolves with T_4 treatment [11], but a recent study failed to show septal hypertrophy in any hypothyroid patient studied. Pericardial effusion occurs in one-third to one-half of patients with overt hypothyroidism. The effusion is more common and their volume is greater in patients with long-standing severe disease. Cardiac tamponade is very rare. More sophisticated techniques have been used to assess systolic and diastolic function and myocardial texture, such as cardiac MRI, tissue Doppler imaging [12], and ultrasonic myocardial textural analysis [12]. By using magnetic resonance spectroscopy, an early cardiac bioenergetics impairment was demonstrated in patients with subclinical hypothyroidism which was reversible after L-thyroxine therapy [13].

Laboratory Tests

The serum levels of creatine kinase, aspartate aminotransferase, and lactate dehydrogenase may be increased. Serum creatine kinase activity is high in as many as 30% of patients. Whereas the increase may reflect myocardial necrosis, in most patients the isoenzyme

distribution indicates its origin from the skeletal rather than cardiac muscle. Prolongation of the half-life of creatine kinase in the circulation contributes to the elevated serum concentration.

Respiratory Changes

Respiratory troubles are rarely a serious complaint in hypothyroid patients. However, hypothyroidism may cause respiratory problems through: (1) depression of the respiratory centre in the brain; (2) disturbed neural conduction and/or neuromuscular transmission to the respiratory muscles (due to hypothyroid neuropathy); (3) diseased respiratory muscle function (due to hypothyroid myopathy); and (4) changes in the alveolar-capillary membranes and the surfactant lining the alveoli, leading to impaired gas exchange. Fatigue and dyspnoea on exertion are frequent symptoms. Dyspnoea is a frequent complaint of myxoedematous patients, but is also a common symptom among well people. Congestive heart failure of separate origin, pleural effusion, anaemia, obesity, or pulmonary disease may be responsible.

Some information on pulmonary function in hypothyroidism is available [14]. Wilson and Bedell [14] found a normal vital capacity and arterial PCO_2 and $_pO_2$ in 16 hypothyroid patients. They also found a decreased maximal breathing capacity, decreased diffusion capacity, and decreased ventilatory response to carbon dioxide. Decreased ventilatory drive is present in about one-third of hypothyroid patients, and the response to hypoxia returns rapidly within a week after beginning therapy. Summarizing the few studies, there is little abnormality of resting pulmonary function in most non-obese patients with hypothyroidism [15–16]. Some patients may exhibit a decreased vital capacity, probably due to muscular weakness. Overall oxygen transfer may be slightly decreased, as evidenced by a decreased $_pO_2$, possibly due to a decreased diffusing capacity for carbon monoxide. An increase in ventilation perfusion mismatching or an opening of anatomical shunts may also contribute to these modifications.

The severity of hypothyroidism parallels the incidence of impaired ventilatory drive [17, 18]. Patients with myxoedema may develop carbon dioxide retention, and carbon dioxide narcosis may be a cause of myxoedema coma. Hypothyroidism-induced breathing disorders during sleep, particularly sleep apnoea syndromes *(OSAS)*, have been described [19]. Obstructive sleep apnoea has been documented in hypothyroidism in about 30% of patients and is reversible with treatment [19, 20]. Hypothyroidism may predispose to upper airway obstruction by several mechanisms: increased size of the tongue and other pharyngeal skeletal muscles; a slow and sustained pharyngeal muscle contraction pattern; or diminished neural output of the respiratory centre. After LT4 replacement therapy, apnoea periods, oxygen desaturation events, and snoring usually improve [16]. Myxoedematous patients are more subject to respiratory infections. Pleural effusions usually are evident only on radiological examination.

Gastrointestinal Changes

The gastrointestinal manifestations of hypothyroidism are listed in Box 3.4.1.2. Poor appetite can be a leading symptom in hypothyroid patients. Anorexia can be interpreted as the reflection of a lowered food requirement. Although two-thirds of patients have reported weight gain, it is of modest degree and due largely to retention of fluid by hydrophilic glycoprotein deposits in the tissues.

Box 3.4.1.2 Gastrointestinal manifestations of hypothyroidism

- Symptoms
  - Anorexia
  - Gaseous distension
  - Constipation
  - Prolonged gastric emptying
- Signs
  - Prolonged intestinal transit time
  - Ascites
  - Elevated liver enzymes
  - Gallbladder hypotonia

True obesity is not a feature of hypothyroidism. Younger patients with iatrogenic hypothyroidism secondary to treatment for thyrotoxicosis commonly gain weight because of decreased physical activity coupled with unchanged food intake.

Constipation is commonly present and is the result of a lowered food intake and decreased peristaltic activity. The latter may lead to faecal impaction and may mimic mechanical ileus when accompanied by colicky pains. Spontaneous hypothyroidism most often afflicts older people, who may discount the significance of an insidious decrease of bowel movements [21]. Severe constipation that is unresponsive to treatment may, therefore, be a prominent finding at the time of diagnosis. Gastric emptying and intestinal transit time are prolonged [21, 22]. Gaseous distension may be a troublesome symptom; it responds slowly to thyroid treatment. In most patients, intestinal absorption is normal. Although the rates of absorption for many substances are decreased [21, 22], the total amount absorbed may be normal or even increased because the decreased bowel motility may allow more time for absorption. Occasional malabsorption has been attributed to myxoedema of the intestinal mucosa or altered intestinal motility. Galactose and glucose tolerance curves show a delayed rise to a lower peak than normal and a delayed return to baseline.

Ascites in the absence of another cause is unusual in hypothyroidism, but it can occur in association with pleural and pericardial effusion [22]. Myxoedema ascites consists of a yellow and gelatinous peritoneal exudate. It has been related to congestive heart failure, enhanced capillary permeability, or inappropriate secretion of antidiuretic hormone.

Atrophy of the gastric [23] and intestinal mucosa and myxoedematous infiltration of the bowel wall may be present at histological examination. Immune gastritis is often observed in hypothyroid patients [23] with autoimmune thyroiditis. As many as 50% of patients with autoimmune hypothyroidism have achlorhydria, 25% have circulating antibodies directed against the gastric parietal cells or intrinsic factor, and 10% have pernicious anaemia caused by impaired absorption of vitamin B_{12}.

A history of overt hypothyroidism has been associated with small intestinal bacterial overgrowth (SIBO), which is a clinical condition caused by an increased level of microorganisms exceeding the presence of more than 106 colony-forming units/ml within the small intestine. SIBO is considered a malabsorption syndrome [24].

Symptoms or signs of disturbed liver or exocrine pancreatic function are usually not encountered, but biochemical tests may suggest disease. Non-alcoholic fatty liver disease (NAFLD) is also frequently observed in hypothyroidism [25]. The association of liver disease and hypothyroidism is suggestive of a multisystem autoimmune

disease affecting both the liver (e.g. chronic active hepatitis or primary biliary cirrhosis) and the thyroid. Structural liver damage is unusual in hypothyroidism. Serum glutamic-oxaloacetic transaminase, lactate dehydrogenase, and creatine phosphokinase levels are elevated in patients with hypothyroidism [26]. The enzymes return to normal over 2–4 weeks during treatment. Urinary amylase levels may be increased. Gallbladder motility is decreased, and the gallbladder may appear distended on radiographic examination; hypothyroidism increases the risk of gallstones due to a reduced gallbladder motility and bilirubin excretion and increased serum cholesterol [22].

Cerebral and Neurological Changes

Thyroid hormone is essential for the development of the central nervous system. Deficiency in fetal life or at birth causes hypoplasia of cortical neurons with poor development of cellular processes, retarded myelination, and reduced vascularity. Deficiency of thyroid hormone beginning in adult life causes less severe manifestations that usually respond to treatment with thyroid hormone. Recent studies using ^{32}P nuclear magnetic resonance spectroscopy of the frontal lobe of adult hypothyroid patients report reversible alterations in phosphate metabolism, suggesting impairment of mitochondrial metabolism [27]. Cerebral blood flow is reduced in hypothyroidism, but cerebral oxygen consumption is usually normal [28]. This finding is in accord with the observation that the oxygen consumption of isolated brain tissue *in vitro*, unlike that of most other tissues, is not stimulated by the administration of thyroid hormones. In severe cases, decreased cerebral blood flow may lead to cerebral hypoxia. These and other findings indicate that the adult human brain is a thyroid hormone-responsive organ. Hypothyroidism may induce neurological abnormalities at an early stage of the disease [29]. It has been shown [29] that peripheral and central neuropathy develops in patients of hypothyroidism at an early stage of disease and the electrophysiological investigations of such patients can help in timely detection and treatment of neurological disorders that occur due to thyroid hormone deficiency.

Box 3.4.1.3 lists the numerous symptoms suggesting either neurological or psychiatric disorders in patients with moderate to severe hypothyroidism. In adult and elderly patients, mental changes may go unrecognized because of their slow development and because they may mimic cerebral atherosclerosis. However, an unusual complacency, fatigue, and pronounced somnolence or even lethargy together with a prolonged reaction time should suggest the possibility of hypothyroidism. Special attention is required for patients who need an increasing amount of sleep (over 12–14 h/day). They may lapse into stupor or even coma, and develop convulsions. This may be the beginning of myxoedema coma, a rare but very serious condition, the extreme expression of severe hypothyroidism. All intellectual functions, including speech, are slowed. There is loss of initiative, and slow wittedness and memory defects are common; in a study [30] working memory was impaired in hypothyroidism. Dementia in elderly patients may be mistaken for senile dementia. Memory is undoubtedly impaired, and attention and the desire to think are reduced. The emotional level seems definitely low, and irritability is decreased. Except in the terminal stage, reasoning power is preserved. Cognitive tests of patients with moderate to severe hypothyroidism indicate difficulties in performing calculations, recent memory loss, reduced attention span, and slow reaction time [31, 32]. In a metanalysis the association between subclinical hypothyroidism (sHT) and cognitive impairment has been shown only in individuals younger than 75 years of age and those with higher thyroid-stimulating hormone (TSH) concentrations [33].

Headaches are frequent.

In a minority of patients, nervousness and apprehension are present. Psychiatric disorders are common and are usually of the paranoid or depressive type and may induce agitation (myxoedema madness). Depression is so often associated with hypothyroidism [31] that thyroid function tests should be performed in the evaluation of any patient presenting with this symptom [33]. Central 5-hydroxytryptamine activity is reduced in hypothyroid patients, and T_3 supplementation might increase the efficacy of antidepressant drugs. At times, this manifestation of hypothyroidism is more severe than are many of the other clinical manifestations of the disease. Because hypothyroidism is so readily treated, it is an especially important cause to eliminate.

In rare cases of long-standing hypothyroidism cerebellar ataxia with or without intention tremor has been found. Jellinek and Kelly [34] described a series of myxoedematous patients with ataxia, intention tremor, nystagmus, and dysdiadochokinesia. Ataxia has been noted in 8% of a large series of hypothyroid patients [35]. Patients may have intention tremor, nystagmus, and an inability to make rapid alternating movements. The cause of this syndrome is not apparent, but myxoedematous infiltrates of glycogen and mucinous material have been found in the cerebellum. There may be foci of degeneration and an increase in glial tissue. These symptoms show a prompt and definite decrease after replacement therapy with thyroid hormone.

Sensory phenomena are common. Numbness and tingling of the extremities are frequent. Mononeuropathies occur in hypothyroidism, as attested to by the high incidence of carpal tunnel syndrome (compression of the median nerve at the wrist) [36]. Nocturnal paraesthesia and pain in the median nerve distribution in one or both hands is a common manifestation of this condition. Paraesthesia or lancing pain in the legs are manifestations of lower extremity peripheral neuropathy. A study of 39 patients with primary hypothyroidism found complaints of polyneuropathy in 64%, findings of polyneuropathy in 33%, and a definite diagnosis by electrophysiological criteria in 72% [36]. A metachromatic infiltrate has

Box 3.4.1.3 Neurologic and psychiatric manifestations in hypothyroidism

- Neurologic symptoms or signs
  - Somnolence, lethargy
  - Slow speech
  - Impaired cognitive functions
  - Headache
  - Paraesthesias
  - Cerebellar ataxia
  - Deafness
  - Vertigo
  - Delayed relaxation of deep tendon reflexes
- Psychiatric syndromes
  - Depression
  - Bipolar disorders
  - Affective psychosis

been found in the lateral femoral cutaneous nerve and sural nerve, together with axon cylinder degeneration.

The tendon reflexes are slow, especially during the relaxation phase, producing the characteristic 'hung-up' reflexes: this phenomenon is due to a decrease in the rate of muscle contraction and relaxation, rather than a delay in nerve conduction. The presence of extensor plantar responses or diminished vibration sense should alert the physician to the possibility of coexisting pernicious anaemia with combined system disease.

Electroencephalographic changes include slow α-wave activity and general loss of amplitude. The concentration of protein in the cerebrospinal fluid is often increased, but cerebrospinal pressure is normal.

Deafness is a very characteristic and troublesome symptom of hypothyroidism. It may be due to both conduction or nerve impairment and usually responds very well to treatment. Vestibular abnormalities have also been demonstrated. Serous otitis media is not uncommon. Two-thirds of patients complain of dizziness, vertigo, or occasionally tinnitus: these problems suggest damage to the eighth nerve or labyrinth, or possibly to the cerebellum. Whatever type of deafness is present, there is marked improvement after thyroid treatment. Acquired hearing loss in association with adult-onset hypothyroidism should be distinguished from the sensorineural deafness of Pendred's syndrome. In the latter, treatment of hypothyroidism does not correct the hearing defect.

Night blindness is not uncommon. It is caused by a deficiency in the pigment retinene, which is required for the adaptation to dark.

Musculoskeletal Changes

Muscles

In patients with hypothyroidism, disordered muscle function often is the predominating feature of the clinical syndrome. Generalized muscular hypertrophy, accompanied by easy fatigue and slowness of movements, occurs in some myxoedematous children or adults. It has been referred to as the Kocher–Debré–Sémélaigne syndrome in children [37] and as Hoffmann's syndrome in adults. These patients do not have the classic electromyography findings of myotonia. The myopathy of hypothyroidism in some patients is associated with weakness even though the muscles are hypertrophied. The typical patient presents with firm large well-developed muscles, like those of an athlete. The entire musculature is affected to some extent, but the most obvious enlargement is in the arms and legs.

Muscle symptoms such as myalgia, muscle weakness, stiffness, cramps, and easy fatigability are very prevalent in hypothyroid patients [38, 39]. The symptoms are aggravated by exposure to cold. They are also prominent during the rapid onset of hypothyroidism after surgery or ^{131}I treatment. Impairment of mitochondrial oxidative metabolism provides a biochemical substrate for these complaints.

Reflex contraction and relaxation time is prolonged, mainly because of the intrinsic alterations in muscle contractility. Nerve conduction time may also be prolonged. Delayed reflex relaxation is characteristic and has been developed into a diagnostic test of thyroid function. The rate-limiting step in muscle relaxation is the reuptake of calcium by the sarcoplasmic reticulum. In skeletal muscle, this process is dependent on the content of Ca^{2+}-ATPase. Recent studies have indicated that Ca^{2+}-ATPase activity of the fast twitch variety (SERCA1) is markedly reduced in hypothyroidism [40] with impairment of calcium reuptake as a consequence. This occurs at a transcriptional level, since thyroid hormone response elements have been identified in the 5′ flanking region of the SERCA1 Ca^{2+}-ATPase gene. The reduction in Ca^{2+}-ATPase would explain the delayed relaxation of the deep tendon reflexes. On histopathological examination the muscles appear pale and swollen. The muscle fibres may show swelling, loss of normal striation, and mucinous deposits.

Skeletal System: Calcium and Phosphorus Metabolism

In the adult skeleton, thyroid hormone deficiency decreases recruitment, maturation, and activity of bone cells, leading to decreased remodelling, which is especially reflected in the impaired function of the osteoclasts [41]. Bone mass may be normal or slightly increased; however hypothyroid patients show a two-to threefold increased risk fracture [42]. In some patients, fracture may occur before bone mass reaches levels compatible with osteoporosis, reflecting an impairment of bone quality [41]. Recent data suggest that high TSH levels may be linked to the risk of osteoporotic fractures only in young and middle-aged men, while in postmenopausal women, the long term risk of hip and other osteoporotic fractures is strongly related to the cumulative duration of LT4 overtreatment [43].

Urinary excretion of calcium is decreased as is the glomerular filtration rate, whereas faecal excretion of calcium and both urinary and faecal excretion of phosphorus are variable. The concentrations of calcium and phosphorus in serum are usually normal, but calcium may be slightly elevated. Serum alkaline phosphatase levels are often decreased, as are serum osteocalcin levels. Because the levels of parathyroid hormone are often slightly increased, some degree of resistance to its action may be present. Serum concentrations of 1,25-dihydroxycholecalciferol are also increased.

Joints

At the clinical level, patients with hypothyroidism often complain of articular and muscular pain and stiffness of the extremities. These symptoms may suggest rheumatoid arthritis or also polymyalgia rheumatica or primary myositis. Patients may exhibit joint effusions involving the knees and small joints of the hands and feet. In 5–10% of patients with carpal tunnel syndrome, primary hypothyroidism may be the cause due to the accumulation of the hygroscopic glycosaminoglycan in the interstitial space with compression of the median nerve.

Changes in Kidney Function

Clinically significant disturbances of kidney function, and hence of water and electrolyte metabolism, are uncommon in hypothyroidism. Renal blood flow and glomerular filtration rate can be reduced [44]. Because of the moderate extent of these reductions and the hypothyroidism-induced decreased metabolism, renal failure does not usually occur. Factors contributing to the decrease in renal blood flow are a decrease in cardiac output, a decrease in plasma volume, and a narrowing of renal blood vessels through enlargement of endothelial and mesangial cells and thickening of the glomerular basement membrane.

Laboratory examinations may reveal a slight increase of serum creatinine and uric acid [44]. Urine flow is reduced, and delay in the excretion of a water load may result in reversal of the normal diurnal pattern of urine excretion. The delay in water excretion appears to be due to decreased volume delivery to the distal diluting

segment of the nephron resulting from diminished renal perfusion and inappropriate secretion of vasopressin [44]. Since urinary hydroxycorticoid excretion is decreased, the adrenals might be responsible for delayed water excretion. Other evidence suggests that the tissue supply of adrenal cortical hormones is usually normal in myxoedema. The ability to concentrate urine may be slightly impaired. Occasionally, minimal proteinuria is seen. This condition could be due to congestive heart failure or to the increased capillary transudation of protein typical of hypothyroidism.

The total body sodium content is increased [45]. The excessive sodium is presumably bound to extracellular mucopolysaccharides. In spite of reduced renal blood flow and blood volume, the sodium retention is probably not a reflection of altered renal function. In fact, salt loads are usually excreted readily and serum sodium concentrations tend to be low [45], in contrast to other clinical situations associated with sodium retention, such as congestive heart failure. No consistent changes in plasma potassium levels have been reported. Total magnesium levels may be elevated and the bound fraction and urinary excretion are reduced. Plasma homocysteine concentrations are increased in hypothyroidism, related to lower folate levels and a lower creatinine clearance in thyroid hormone deficiency [46].

Haematological Changes

Erythrocytes

Anaemia is present in up to two-thirds of hypothyroid children and adolescents, and in about one-third of adults with hypothyroidism [47]. Anaemia is usually mild. Anaemia in hypothyroidism may be a normochromic and normocytic anaemia due to the diminished oxygen requirements and decreased production of erythropoietin [47] or may result from a specific depression of marrow that lacks thyroid hormone [47]. The bone marrow generally shows mild hypoplasia with an increase in fatty marrow. The anaemia may be macrocytic, sometimes from deficiency of vitamin B_{12}. Folate deficiency from malabsorption or dietary inadequacy may also cause macrocytic anaemia. The frequent menorrhagia and the defective absorption of iron resulting from achlorhydria may contribute to a microcytic hypochromic anaemia [47].

Leucocytes and Thrombocytes

Granulocyte, lymphocyte, and platelet counts are usually normal in hypothyroidism. Leucopenia might indicate associated vitamin B_{12} or folic acid deficiency. Mean platelet volume can be decreased. The erythrocyte sedimentation rate (ESR) may be elevated in uncomplicated hypothyroidism.

Haemostasis

Hypothyroid patients may have bleeding symptoms such as easy bruising, menorrhagia, or prolonged bleeding after tooth extraction. The most frequent defects in haemostasis are prolonged bleeding time, decreased platelet adhesiveness, and low plasma concentrations of factor VIII and von Willebrand's factor (acquired von Willebrand syndrome type 1) [48]. This coagulopathy should be taken into account when planning invasive procedures in hypothyroid patients. The mild or moderate acquired von Willebrand syndrome is reversible after restoration of euthyroidism. There are also data suggesting existence of a hypercoagulation and hypofibrinolytic state, particularly in subclinical hypothyroidism [48].

Changes in the Reproductive Tract

In both sexes thyroid hormone influences sexual development and reproductive function. Infantile hypothyroidism leads to sexual immaturity and juvenile hypothyroidism causes a delay in the onset of puberty followed by anovulatory cycles. Paradoxically, primary hypothyroidism may also cause precocious sexual development and galactorrhoea [49].

In adult men, hypothyroidism may lead to impotence, lack of libido, and, rarely, to testicular tubular involution. The testicles are histologically immature if hypothyroidism preceded puberty and show tubular involution if its onset was after puberty [50]. In adult hypothyroid men, semen analysis is usually normal. Delayed ejaculation, hypoactive sexual desire, and erectile dysfunction have been described [51]. In adult women, severe hypothyroidism may be associated with diminished libido and failure of ovulation [52, 53]. In general, hypothyroid women complain of menorrhagia and, occasionally, oligo- and amenorrhoea. Plasma gonadotropins are usually in the normal range in primary hypothyroidism and the pulsatile gonadotropin release in the follicular phase is normal [52], but the ovulatory surge may not happen. Secretion of progesterone is inadequate, and endometrial proliferation persists, resulting in excessive and irregular breakthrough menstrual bleeding. The anovulation is reflected in the frequent finding of a proliferative endometrium. These changes may be due to a deficient secretion of luteinizing hormone. Mild to moderate hyperprolactinaemia is a frequent finding in hypothyroid women, with or without galactorrhoea. It is attributed to the stimulatory effect of increased thyrotropin-releasing hormone on prolactin secretion. Fertility is reduced, and spontaneous abortion may result, although many pregnancies are successful.

The total concentrations of both testosterone and oestradiol in serum are decreased, predominantly due to a diminution in the concentration of the carrier sex hormone-binding globulin. Because of the concomitant increase in the unbound fraction of sex steroids, their absolute free concentration remains normal. The metabolism of testosterone is shifted towards aetiocholanolone rather than androsterone. With respect to oestradiol and oestrone, hypothyroidism favours metabolism of these steroids via 16α-hydroxylation with the result that formation of oestriol is increased.

The literature contains many reports of pregnancy in untreated hypothyroid women [53, 54]. Euthyroid neonates born to hypothyroid mothers during pregnancy have been reported to achieve a lower IQ later in life [55]. When treatment has been started during pregnancy, generally a normal child is produced [53], but abortion is frequent in women with myxoedema. Pregnancy-induced hypertension is 2–3 times more common in hypothyroid women. Low birthweight is secondary to premature delivery for gestational hypertension. The incidence of various congenital abnormalities may be increased, but recent studies do not report an increased risk of fetal death or congenital anomalies with proper treatment [53, 56].

Other Endocrine Glands

Pituitary Function

Hypothyroidism can affect the secretion of all pituitary hormones. The effect of hypothyroidism on the secretion of vasopressin, follicle-stimulating hormone, and luteinizing hormone are discussed

in other sections. Hypothyroidism decreases growth hormone secretion and hypothyroid children have a dramatic retardation of growth [57]. Retarded growth caused by hypothyroidism appears to result from deficient secretion of growth hormone as well as from impaired action of growth hormone. Many hypothyroid children have subnormal serum growth hormone response to insulin-induced hypoglycaemia. Growth hormone secretion is decreased in hypothyroidism related to an increase in hypothalamic somatostatinergic tone [57], and results in low serum insulin-like growth factor (IGF)-1 concentrations [57]. Serum IGF-2, IGFBP1, and IGFBP3 also fall, whereas IGFBP2 rises; these changes are reversible upon treatment [57].

Thyrotropic hyperplasia caused by primary hypothyroidism may result in sellar enlargement, particularly when the condition has remained untreated for a long time [58]. Rarely, such hyperplasia may give rise to a pituitary macroadenoma that shrinks after thyroxine replacement [59]. Patients with severe hypothyroidism may have an increase in serum prolactin level that correlates with the level of serum TSH, and some patients develop galactorrhoea. Since thyroid hormone decreases the mRNA for preprothyrotropin-releasing hormone in the paraventricular nuclei, hypothyroidism may lead to increased thyrotropin-releasing hormone secretion, unopposed by thyroid hormones, with consequent hyperprolactinaemia.

Adrenal Cortex

Patients with primary hypothyroidism have subtle abnormalities of pituitary–adrenal function that may be correlated with the severity and duration of hypothyroidism [60]. Cortisol secretion and the rate of turnover are decreased in patients with hypothyroidism. The net result is that serum cortisol concentrations and urinary cortisol excretion are normal. Hepatic clearance of cortisol and adrenal androgens 17-hydroxycorticosteroids and 17-ketosteroids are decreased. This slowing is principally due to a decrease in the rate of cortisol oxidation as a result of reduced 11-hydroxysteroid dehydrogenase activity [60]. Conjugation with glucuronic acid in the liver is normal. The turnover rate of aldosterone is also decreased in hypothyroidism [61]. The decrease in hepatic clearance causes an increase in its plasma half-life. The reduced rate of clearance of aldosterone is balanced by a lower secretion rate. The serum concentration of aldosterone is normal and there is no clinical evidence of hyperaldosteronism. Angiotensinogen production in the liver is reduced, as is plasma renin activity. These subtle modifications are not responsible for alterations in sodium and potassium homeostasis.

The adrenal response to adrenocorticotropic hormone (ACTH) is normal or reduced. In long-standing hypothyroidism, a significant reduction in cortisol secretion after ACTH has been documented in primary adrenal cell culture and it was recently confirmed also in humans [62]. Pituitary–adrenal responses to the metyrapone test have been variable. Normal but delayed peak response, impaired response, or even lack of response has been reported. Grossly impaired responses to the stimulation with lysine-8-vasopressin and a delayed increase in serum cortisol levels after insulin-induced hypoglycaemia have also been observed. Whether steroid production can be augmented sufficiently in times of stress is not clear, but the provocative test results suggest that these patients usually have a mildly impaired hypothalamic–pituitary–adrenal axis.

Metabolic Changes

Energy Metabolism

The decrease in energy metabolism and heat production is reflected in the low basal metabolic rate, decreased appetite, cold intolerance, and slightly low basal body temperature. Measurement of the resting energy expenditure is rarely performed nowadays. In patients with complete athyreosis it falls to between 35% and 45% below normal. In Addison's disease, the basal metabolic rate may fall to 25% or 30% below normal, and in hypopituitarism to 50% below normal.

Thyroid hormones are involved in the regulation of energy balance in mammals, acting not only in the periphery but also in brain areas involved in the regulation of food intake and energy expenditure [63].

The effect of hypothyroidism on appetite and energy intake is not precisely known, but energy expenditure decreases leading to a slight net gain in energy stores. An increase in adipose tissue mass results in an increase of serum leptin, which mediates a decrease in energy intake while energy disposal increases, eventually leading to a reduction in adipose tissue mass. In adult rats methimazole induced hypothyroidism impairs hypothalamic leptin signalling and the anorectic response to the peripheral administration of leptin. These findings suggest that thyroid hormones are essential for the effect of leptin on food intake. A large number of studies have investigated the relationship between thyroid dysfunction and circulating levels of leptin, but results have been conflicting [64].

Protein Metabolism

Both the synthesis and the degradation of proteins are decreased, the latter especially so, with the result that nitrogen balance is usually slightly positive. Despite both a decrease in the rate of albumin synthesis and degradation, the total exchangeable albumin pool increases in myxoedema [65]. The albumin is distributed in a much larger volume, suggesting enhanced permeability of capillary walls. The synthesis of thyroid hormone-responsive proteins is clearly reduced in the hypothyroid state, whereas that of proteins such as TSH or glycosaminoglycans may be increased under the same circumstances.

Comparative studies of protein translation by hepatic ribosomes from T_3-treated hypothyroid rats show that the mRNAs from some proteins are increased and others are decreased. Most of these proteins have not been identified. Treatment of myxoedema is accompanied by a marked but temporary negative nitrogen balance, reflecting the mobilization of extracellular protein [65]. In a later phase there is an increase in urinary potassium and phosphorus together with nitrogen in amounts suggesting that cellular protein is also being metabolized.

Carbohydrate Metabolism

In hypothyroidism, absorption of glucose from the gastrointestinal tract is slowed and peripheral glucose assimilation is retarded for impairment of insulin-mediated translocation of glucose transporter 4. At the same time, glycerol release from adipose tissue is slowed, and the availability of amino acids and glycerol for gluconeogenesis is decreased. The oral glucose tolerance curve is characteristically flat, and the insulin response to glucose is delayed. Degradation of insulin is slow, so the sensitivity to exogenous insulin may be increased. Despite the easily demonstrable abnormalities

Box 3.4.1.4 Changes in serum lipids in hypothyroidism

- Total cholesterol—increase
- LDL cholesterol—increase
- HDL2 cholesterol—modest increase
- HDL3 cholesterol—no change
- Triglycerides—no change or modest increase

in carbohydrate metabolism in hypothyroidism, clinical manifestations of these abnormalities are seldom conspicuous. Although hypoglycaemia is sometimes listed as a manifestation of hypothyroidism, it is rarely a sign of isolated hormone deficiency, and the presence of hypoglycaemia in a patient with hypothyroidism should suggest the presence of hypopituitarism. The occurrence of hypothyroidism in a patient with insulin-dependent diabetes mellitus may result in a diminution in exogenous insulin requirement and a greater risk of developing hypoglycaemia.

Lipid Metabolism

Thyroid hormones have multiple effects on the regulation of lipid synthesis, absorption, and metabolism. Therefore, overt and subclinical hypothyroidism significantly affects the lipids profile and promotes cardiovascular disease [25].

A variety of abnormalities in plasma lipid concentrations occur in hypothyroidism (Box 3.4.1.4). Plasma free fatty acid concentrations are normal, plasma concentrations of triglycerides, phospholipids, and LDL cholesterol are well elevated [25, 66]. Biosynthesis of fatty acids and lipolysis are reduced. The changes bear, in general, a reciprocal relationship to the level of thyroid activity.

The increased serum cholesterol may represent an alteration in the substrate steady-state level caused by a transient proportionally greater retardation in degradation than in synthesis [25, 66]. The increase of serum cholesterol is largely accounted for by an increase of LDL cholesterol, which is cleared less efficiently from the circulation due to a decreased T_3-dependent gene expressing the hepatic LDL receptor [25, 66]. Interestingly, the LDL particles of hypothyroid patients are also susceptible to increased oxidizability [25, 66]. The increase of high-density lipoprotein (HDL) 2 but not of HDL3 cholesterol is due to a diminished activity of cholesterol ester transfer protein [66] and hepatic lipase (which is involved in the conversion of HDL2 to HDL3). The modest increase of serum triglycerides seen in certain cases has been related to a decreased lipoprotein lipase activity in post-heparin plasma. Lipoprotein(a) is increased in hypothyroidism in some but not all studies. A recent study evaluating lipid profile and TSH levels in a large Spanish population including euthyroid and hypothyroid individuals identified a serum TSH cut-off of 2.57 mU/L for detecting significant differences in circulating lipid levels [67].

Clinical Aspects of Hypothyroidism at Different Ages

Infantile and Juvenile Hypothyroidism

Hypothyroidism in newborn infants results in mental and physical impairment unless treatment is initiated within weeks after birth. Hypothyroidism in children is mainly characterized by retarded growth and impaired mental performances. Infantile hypothyroidism leads to sexual immaturity; juvenile hypothyroidism causes a delay in the onset of puberty followed by anovulatory cycles in girls. Rarely precocious puberty may occur [49].

Thyroid hormone is essential for normal growth and maturation of the skeleton [41]. Deficient thyroid hormone production *in utero* and in the neonate retards growth and delays skeletal maturation. Deficiency in early life leads both to a delay in the development of and an abnormal stippled appearance of the epiphyseal centres of ossification (epiphyseal dysgenesis). Before puberty, thyroid hormones also play an important role in the maturation of bone. Impairment of linear growth leads to dwarfism in which the limbs are disproportionately short in relation to the trunk [41]. Bone age is retarded in hypothyroid children [41].

Hypothyroidism in Adults

In adults, the clinical manifestations of hypothyroidism, though they may be profound, are reversible [68]. The development of spontaneous hypothyroidism is usually slow and many patients seek medical attention for variable and non-specific symptoms [68]. In contrast, patients who develop hypothyroidism rapidly (when replacement therapy is discontinued in a patient with primary hypothyroidism, or after surgical removal of the gland) have more symptoms. In such patients, manifestations of overt hypothyroidism are present by 6 weeks. Older patients tend to have fewer symptoms and signs of hypothyroidism than do young adults.

In adults, common features of hypothyroidism include easy fatigability, tiredness, coldness, weight gain, constipation, menstrual irregularities, and muscle cramps. Drowsiness and slowing of intellectual and motor activity is often referred. Sensitivity to cold is suggested by the use of more blankets on the bed. Women frequently complain of hair loss, brittle nails, and dry skin. Periorbital puffiness may be present. Stiffness and aching of muscles may be attributed to rheumatism. Constipation may occur. Numbness and tingling of the extremities are frequent. Physical findings include a cool, dry skin, puffy face and hands, hoarse husky voice, and slow reflexes.

In a recent study [68] in overt autoimmune hypothyroidism, 94% of women and 91% of men reported at least one of the hypothyroidism-associated symptoms, with tiredness as the most common symptom followed by dry skin and shortness of breath. In contrast, women free of thyroid disease self-reported at least one hypothyroidism-associated symptom considerably more often than men [68].

Hypothyroidism in Older People

Hypothyroidism in older people is often atypical and elusive and lacks the classic clinical features present in younger patients [69]. This is due to a combination of factors including the insidious onset, the ambiguity of several signs and symptoms (fatigue, weakness, cold intolerance, dry skin, hair loss, constipation, poor appetite, depression and/or mental deterioration, hearing loss, cardiomegaly, congestive heart failure) which may be attributed to normal ageing, and to the frequent coexistence of several age-associated diseases.

The most relevant clinical findings that lead one to suspect hypothyroidism in older people are an unexplained increase in serum cholesterol, constipation, congestive heart failure (particularly when it presents as restrictive cardiomyopathy), and macrocytic anaemia (as a consequence of folate deficiency or coexistent autoimmune

gastritis and pernicious anaemia). Other common clinical features encountered in elderly hypothyroid patients include neurological signs (syncope, seizures, impaired cerebellar function, carpal tunnel syndrome) and vague arthritic complaints. Due to the frequent involvement of the cardiovascular system, the presenting symptoms of hypothyroidism in elderly patients include dyspnoea in more than 50% and chest pain in up to one-quarter. A significant minority of elderly hypothyroid patients may paradoxically lose weight as a consequence of reduced appetite. Neuropsychiatric symptoms are often prominent and depression occurs in up to 60% of patients; psychoses are rare. Dementia may be found in elderly hypothyroid patients but it is rarely the direct consequence of thyroid failure, although a few patients show marked improvement of intellectual function after correction of hypothyroidism. In a recent population-based study of hypothyroidism, the power of symptom presence in predicting overt hypothyroidism in both young and older subjects was investigated [69]. Hypothyroid symptom score is a good discriminating tool to identify hypothyroidism in young patients but fails to identify hypothyroidism in older people [69].

In a multicentre, double-blind, randomized, placebo-controlled, parallel-group trial involving old hypothyroid patients with persisting subclinical hypothyroidism [70], treatment with L-thyroxine had no consistent beneficial effect on thyroid related symptoms. Besides L-thyroxine treatment yielded no significant beneficial effects on a range of secondary outcome measures [70].

Elderly patients are more susceptible to myxoedema coma, a rare but serious complication of hypothyroidism. It generally occurs in the winter months, in hospitalized patients, and can be precipitated by intercurrent non-thyroidal illness, use of drugs, exposure to cold, and stress. Progressive deterioration of mental status to stupor and coma, localized neurological signs, marked hypothermia (which may not be present in patients with systemic infections), hyponatraemia, and hypoglycaemia are the hallmarks of myxoedema coma. The mortality of clearly hypothermic myxoedema coma is very high (over 80%), unless vigorous supportive therapy and thyroid hormone replacement are given immediately.

Clinical Aspects of Hypothyroidism due to Different Aetiologies

Primary Hypothyroidism

Primary hypothyroidism in adults results mainly from autoimmune thyroiditis, it is more common in women than in men, and occurs between the ages of 40 and 60 years [71]. In these patients, clinical features of hypothyroidism may be accompanied by the typical goitre of Hashimoto's thyroiditis. When present, the goitre is usually firm in consistency, generally moderate in size, and often lobulated; well-defined nodules are unusual. Both lobes are enlarged, but the gland may be asymmetrical. Adjacent structures, such as the trachea, oesophagus, and recurrent laryngeal nerves may be compressed but this is a rare occurrence. Goitre develops gradually over many years. Rarely, the thyroid enlarges rapidly and may be accompanied by pain and tenderness. In other cases of hypothyroidism due to autoimmune thyroiditis the gland is atrophied. Infiltrative ophthalmopathy similar to that of Graves' disease occurs in a small proportion of patients.

Other organ-specific autoimmune diseases such as insulin-dependent diabetes mellitus, Addison's disease, premature ovarian failure, hypoparathyroidism, myasthenia gravis, and coeliac disease may coexist [72]. Patients with primary hypothyroidism may also complain of vitiligo and alopecia. Primary autoimmune hypothyroidism may be present as a component of either the type I or type II polyglandular autoimmune syndrome. The specific association of primary hypothyroidism and primary adrenal cortical insufficiency is known as Schmidt's syndrome [72]. The type I syndrome consists of at least two of the triad of Addison's disease, hypoparathyroidism, and chronic mucocutaneous candidiasis; other autoimmune disorders, such as alopecia, chronic autoimmune thyroiditis, and malabsorption syndrome, may also be present. Autoimmune thyroid disease is reported in 10–12% of these patients. Type I polyglandular autoimmune syndrome generally presents in childhood, whereas the type II syndrome is more common and usually presents in adult life. Addison's disease, Hashimoto's thyroiditis, and type 1 diabetes are the most common endocrine deficiencies found in these patients, although gonadal failure, pernicious anaemia, and vitiligo are observed in a significant percentage.

Rarely a combination of primary and pituitary hypothyroidism with or without ACTH deficiency occurs, presumably also on an autoimmune basis. Thus, other glands may be affected with increased frequency in patients with autoimmune hypothyroidism.

Postablative Hypothyroidism

A common cause of hypothyroidism in adults is the type following total thyroidectomy for thyroid carcinoma or near-total thyroidectomy for euthyroid or toxic multinodular goitre or Graves' disease. Hypothyroidism following radioiodine treatment for Graves' hyperthyroidism is also frequent, and is currently regarded as a common outcome of ^{131} I treatment rather than a complication [73].

Overt hypothyroidism in patients who have received ^{131} I is often preceded by subclinical hypothyroidism, which may become apparent within 2–4 months after ^{131} I therapy. The early onset of hypothyroidism may cause distinct symptoms in the previously thyrotoxic patient who received ^{131} I or surgery. These patients may develop muscle cramps, often in large muscle groups (trapezius, latissimus dorsi, or the proximal muscles of the extremities).

Central Hypothyroidism

The clinical picture of central hypothyroidism varies depending on the severity of thyroid failure, the extent of the associated hormone deficiencies, the age of the patients, and the nature of the underlying lesion. Central hypothyroidism is due to TSH deficiency caused by either hypothalamic or pituitary disease [74]. The differentiation of secondary from primary hypothyroidism is important for the institution of the proper therapy. The clinical features of central hypothyroidism are similar to those of primary hypothyroidism, although generally less pronounced. The skin is pale and cool, but not as coarse and dry as in primary hypothyroidism. Periorbital and peripheral oedema are uncommon in patients with central hypothyroidism. Loss of axillary, pubic, and facial hair and thinning of the lateral eyebrows are more pronounced. The tongue is not enlarged, and hoarseness of the voice is not prominent as in primary hypothyroidism. The heart tends to be small, and blood pressure is low. Atrophic breasts and amenorrhoea are found in women.

Table 3.4.1.2 Sensitivity and specificity of the 14 symptoms and signs of hypothyroidism and analysis of their positive and negative predictive values

Symptoms and signs	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Ankle reflex	77	93	92	80
Dry skin	76	64	68	73
Cold intolerance	64	65	65	64
Coarse skin	60	81	76	67
Puffiness	60	96	94	71
Pulse rate	58	42	50	50
Sweating	54	86	79	65
Weight increase	54	77	70	63
Paraesthesia	52	82	75	63
Cold skin	50	80	71	61
Constipation	48	85	76	62
Slow movements	36	99	96	61
Hoarseness	34	87	73	57
Hearing	22	97	90	53

Body weight is more likely to be reduced than increased. Defects in growth hormone and gonadotropin secretion usually precede TSH insufficiency, and in most cases ACTH secretion is the last to be affected. Growth failure with delayed skeletal maturation results from growth hormone deficiency in children. Hypoglycaemia may occur. Gonadotropin insufficiency results in impotence, loss of libido, diminished beard growth, amenorrhoea, infertility, and atrophy of the breasts in women. ACTH deficiency leads to weakness, postural hypotension, and depigmentation of the areole and of other normally pigmented areas of the skin. Symptoms and signs that arise directly from the hypothalamic or pituitary lesion may precede, accompany, and even obscure manifestations of pituitary failure. The manifestations of a sellar mass include headache and symptoms secondary to compression of adjacent structures with visual field disturbances and ophthalmoplegia.

Diagnostic Accuracy

Several attempts have been made to develop a clinical score system, based on the most frequent symptoms and signs of hypothyroidism, that could accurately predict the diagnosis of thyroid

Table 3.4.1.3 Scoring of symptoms and signs of hypothyroidism

	On the basis of	Score	
		Present	Absent
Symptoms			
Diminished sweating	Sweating in a warm room or on a hot summer day	1	0
Hoarseness	Speaking voice, singing voice	1	0
Paraesthesia	Subjective sensation	1	0
Dry skin	Dryness of skin, noticed spontaneously, requiring treatment	1	0
Constipation	Bowel habit, use of laxative	1	
Impairment of hearing	Progressive impairment of hearing	1	0
Weight increase	Recorded weight increase, tightness of clothes	1	0
Physical signs			
Slow movements	Observe patient removing his/her clothes	1	0
Delayed ankle reflex	Observe the relaxation of the reflex	1	0
Coarse skin	Examine hands, forearms, elbows for roughness and thickening of skin	1	0
Periorbital puffiness	This should obscure the curve of the malar bone	1	0
Cold skin	Compare temperature of patient's hands with examiner's	1	0

For clinical judgement, add 1 point to the sum of symptoms and signs present in women younger than 55 years. Hypothyroid, more than 5 points; euthyroid, less than 3 points; intermediate, 3–5 points.

failure in individual patients. In the 1960s, Billewicz *et al.* [75] described a diagnostic index that scored the presence or absence of various signs and symptoms of hypothyroidism. However, at that time, modern laboratory thyroid function tests were not available to validate the diagnostic accuracy of such a score [65]. Recently a convenient clinical score has been proposed by Zulewski *et al.* [76] that is both easy to perform and sensitive for individual assessment of the severity of thyroid failure. The frequencies of the 14 more common symptoms and signs of overt hypothyroidism are shown in Table 3.4.1.2. The most common features in hypothyroid patients were prolonged ankle reflex (77%) and complaints about dry skin (76%). A reduced pulse rate and cold intolerance were recorded with a high frequency in euthyroid controls and were, therefore, excluded from this score. The sensitivity and specificity of each symptom and sign of hypothyroidism and the analysis of their positive and negative predictive values are shown in Table 3.4.1.2. Table 3.4.1.3 shows the scoring system of symptoms and signs of hypothyroidism. Because a correlation analysis revealed a significant correlation of these scores with age, a simple age correcting factor was defined by adding 1 point to the sum of symptoms and signs in women younger than 55 years. According to this analysis, the following diagnostic ranges for the clinical judgement with the age-corrected score were defined: hypothyroid, more than 5 points; euthyroid, 0–2; intermediate, 3–5 points [76].

REFERENCES

1. Rosen T, Kleman GA. Thyroid and skin. In: Callen JP, Jorizzo J, Bolognia J, Piette W, eds. *Dermatological Signs of Internal Disease*, 2nd edn. Philadelphia, PA: WB Saunders, 1994: 189–204.
2. Miller J, Roling D, Spiers E, Davies A, Rawlings A, Leyden J. Palmoplantar keratoderma associated with hypothyroidism. *Br J Dermatol*, 1998; **139**: 738–59.
3. Zetting G, Weissel M, Flores J, Dudczak R, Volgensang H. Dermatitis herpetiformis is associated with atrophic but not with goitrous variant of Hashimoto's thyroiditis. *Eur J Clin Invest*, 2000; **30**: 53–7.
4. Safer JD. Thyroid hormone action on skin. *Curr Opin Endocrinol Diabetes Obes*, 2012; **19**: 388–93.
5. Klein I, Danzi S. Thyroid disease and the hearth. *Curr Probl Cardiol*, 2016; **41**: 65–92.
6. Ojamaa K, Klein I. *In vivo* regulation of recombinant cardiac myosin heavy chain gene expression by thyroid hormone. *Endocrinology*, 1993; **132**: 1002–6.
7. Kahly GJ, Dillmann WH. Thyroid hormone action in the heart. *Endocr Rev*, 2005; **26**: 704–28.
8. Coppola AR, Landenson PW. Hypothyroidism and atherosclerosis. *J Clin Endocrinol Metab*, 2003; **88**: 2438–44.
9. Mattace-Raso SU, van der Cammen TJ, Hofman A, *et al.* Arterial stiffness and risk of coronary heart disease and stroke: the Rotterdam study. *Circulation*, 2006; **113**: 657–63.
10. Zondek H. Das Myxödemherz. *Muench Med Wochenschr*, 1918; **65**: 1180–3.
11. Santos AD, Miller RP, Mathew PK, Wallace WA, Cave WT Jr, Hinojosa L. Echocardiographic characterization of the reversible cardiomyopathy of hypothyroidism. *Am J Med*, 1980; **68**: 675–9.
12. Galderisi M, Vitale G, D'Errico A, Lupoli GA, Ciccarelli A, Cicala S. Usefulness of pulsed tissue Doppler for the assessment of left ventricular myocardial function in overt hypothyroidism. *Ital Heart J*, 2004; **5**: 257–64.
13. Mathadil A, Hollingsworth KG, Blamire AM, *et al.* Levothyroxine improves abnormal cardiac bioenergetics in subclinical hypothyroidism: a cardiac magnetic resonance spectroscopic study. *J Clin Endocrinol Metab*, 2015; **100**: E607–10.
14. Wilson WR, Bedell GN. The pulmonary abnormalities in myxedema. *J Clin Invest*, 1960; **39**: 42–5.
15. Duranti R, Gheri RG, Gorini M, *et al.* Control of breathing in patients with severe hypothyroidism. *Am J Med*, 1993; **95**: 29–33.
16. Sorensen JR, Winther KH, Bonnema SJ, Godballe C, Hegedüs L. Respiratory manifestations of hypothyroidism: a systematic review. Thyroid, 2016; **26**: 1519–27.
17. Ambrosino N, Pacini F, Paggiaro PL, *et al.* Impaired ventilatory drive in short-term primary hypothyroidism and its reversal by L-triiodothyronine. *J Endocrinol Invest*, 1985; **8**: 533–8.
18. Ladenson PW, Goldenheim PD, Ridgway EC. Prediction and reversal of blunted ventilatory responsiveness in patients with hypothyroidism. *Am J Med*, 1988; **84**: 877–83.
19. Pelttari L, Rauhala E, Polo O, *et al.* Upper airway obstruction in hypothyroidism. *J Intern Med*, 1994; **236**: 177–81.
20. Zhang M, Zhang W, Tan J, Zhao M, Zhang Q, Lei P. Role of hypothyroidism in obstructive sleep apnea: a metaanalysis. *Curr Med Res Opin*, 2016; **32**: 1059–64.
21. Shafer RB, Prentiss RA, Bond JH. Gastrointestinal transit in thyroid disease. *Gastroenterology*, 1994; **86**: 852–6.
22. Ebert EC. The Thyroid and the gut. *J Clin Gastrenterol*, 2010; **44**: 402–6.
23. Counsell CE, Taha A, Rudell WJJ. Coeliac disease and autoimmune thyroid disease. *Gut*, 1994; **35**: 844–50.
24. Lauritano EC, Bilotta AL, Gabrielli M, *et al.* Association between hypothyroidism and small intestinal bacterial overgrowth. *J Clin Endocrinol Metab*, 2007; **92**: 4180–4.
25. Sigha RA, Singh BK, Yen PM. Direct effects of thyroid hormones on hepatic lipid metabolism. *Nat Rev Endocrinol*, 2018; **14**: 259–69.
26. Saha B, Maity C. Alterations of serum enzymes in primary hypothyroidism. *Clin Chem Lab Med*, 2002; **40**: 609–11.
27. Smith CD, Ain KB. Brain metabolism in hypothyroidism studied with 31P magnetic-resonance spectroscopy. *Lancet*, 1995; **345**: 619–20.
28. Constant EL, Volder AG, Ivanoiu A, *et al.* Cerebral blood flow and glucose metabolism in hypothyroidism: a positron emission tomography study. *J Clin Endocrinol Metab*, 2001; **86**: 3864–70.
29. Gupta N, Arora M, Sharma R, Arora KS. Peripheral and central nervous system involvement in recently diagnosed cases of hypothyroidism: an electrophysiological study. *Ann Med Health Sci Res*, 2016; **6**: 261–6.
30. Zhu DF, Wang ZX, Zhang DR, *et al.* fMRI revealed neural substrate for reversible working memory dysfunction in subclinical hypothyroidism. *Brain*, 2006; **129**: 2923–30.
31. Grabe HI, Völzke H, Lüdemann J, *et al.* Mental and physical complaints in thyroid disorders in the general population. *Acta Psychiatr Scand*, 2005; **112**: 286–93.
32. Bartalena L, Placidi GF, Martino E, *et al.* Nocturnal serum thyrotropin (TSH) surge and the TSH response to TSH-releasing hormone: dissociated behavior in untreated depressives. *J Clin Endocrinol Metab*, 1990; **71**: 650–5.
33. Pasqualetti G, Pagano G, Rengo G, Ferrara N, Monzani F. Subclinical hypothyroidism and cognitive

impairment: systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2015; **100**: 4240–8.
34. Jellinek EH, Kelly RE. Cerebellar syndrome in myxoedema. *Lancet*, 1960; **ii**: 225–8.
35. Sanders V. Neurological manifestations of myxedema. *N Engl J Med*, 1962; **266**: 599–604.
36. Torres CF, Moxley RT. Hypothyroid neuropathy and myopathy: clinical and electrodiagnostic longitudinal findings. *J Neurol*, 1990; **237**: 271–5.
37. Debré R, Sémélaigne G. Syndrome of diffuse muscular hypertrophy in infants causing athletic appearance: its connection with congenital myxedema. *Am J Dis Child*, 1935; **50**: 1351–4.
38. Duyff RF, Bosch J, Laman DF, van Loon BJ, Linssen WH. Neuromuscular findings in thyroid dysfunction: a prospective clinical and electrodiagnostic study. *J Neurol Neurosurg Psychiatry*, 2000; **68**: 750–5.
39. Salvatore D, Simonides WS, Dentice M, Zavacki AM, Larsen PR. Thyroid hormones and skeletal muscle; new insights and potential implications. *Nat Rev Endocrinol*, 2014; **10**: 206–14.
40. Famulski KS, Pilarska M, Wrzosek A, Sarzala MG. ATPase activity and protein phosphorylation in rabbit fast skeletal muscle sarcolemma. *Eur J Biochem*, 2018; **41**: 99–109.
41. Williams GR, Basset JHD. Thyroid diseases and bone health. *J Endocrinol Invest*, 1986; **7**: 379–84.
42. Basset JH, Boyde A, Howell PG, *et al*. Optimal bone strength and mineralization requires the type 2 iodothyronine deiodinase in osteoblasts. *Proc Natl Acad Sci USA*, 2010; **107**: 764–9.
43. Abrahamsen B, Jorgensen HL, Laulund AS, *et al*. The excess risk of major osteoporotic fractures in hypothyroidism is driven by cumulative hyperthyroid as opposed to hypothyroid time: an observational register-based time-resolved cohort analysis. *J Bone Miner Res*, 2015; **30**: 898–905.
44. Iglesias P, Diez JJ. Thyroid dysfunction and kidney disease. *Eur J Endocrinol*, 2009; **160**: 503–15.
45. Hanna FWF, Scanlon MF. Hyponatraemia, hypothyroidism and role of arginine-vasopressin. *Lancet*, 1997; **350**: 755–6.
46. Diekman MJM, Put NM, Blom HJ, Tijssen JGP, Wiersinga WM. Determinants of changes in plasma homocysteine in hyperthyroidism and hypothyroidism. *Clin Endocrinol*, 2001; **54**: 197–204.
47. M' Rabet-Bensalah, Aubert CE, Coslovsky M, *et al*. Thyroid dysfunction and anaemia in a large population-based study. *Clin Endocrinol*, 2016; **84**: 627–31.
48. Kyriakakis, Lynch J, Ajjan R, Murray RD. The effects of pituitary and thyroid disorders on hemeostasis: potential clinical implications. *Clin Endocrinol*, 2016; **84**: 473–84.
49. Anasti JN, Flack MR, Froehlich J, Nelson LM, Nisula BC. A potential novel mechanism for precocious puberty in juvenile hypothyroidism. *J Clin Endocrinol Metab*, 1995; **80**: 276–9.
50. Van Haaster LH, de Jong FH, Docter R, de Rooij DG. The effect of hypothyroidism on Sertoli cell proliferation and differentiation and hormone levels during testicular development in the rat. *Endocrinology*, 1992; **131**: 1574–81.
51. Carani C, Isidori MI, Granata A, *et al*. Multicenter study on the prevalence of sexual symptoms in male hypo- and hyperthyroid patients. *J Clin Endocrinol Metab*, 2005; **90**: 6472–9.
52. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 102–39.
53. Chiovato L, Lapi P, Fiore E, Tonacchera M, Pinchera A. Thyroid autoimmunity and female gender. *J Endocrinol Invest*, 1993; **16**: 384–91.
54. Burrow GN, Fisher DA, Larsen PR. Maternal and fetal thyroid function. *N Engl J Med*, 1994; **331**: 1072–9.
55. Haddow JE, Palomaki GE, Allan WC, *et al*. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*, 1999; **341**: 549–55.
56. Liu H, Momotani N, Noh JY, Ishikawa N, Takebe K, Ito K. Maternal hypothyroidism during early pregnancy and intellectual development of progeny. *Arch Intern Med*, 1994; **154**: 785–9.
57. Miell JP, Taylor AM, Zini M, Maheshwari HG, Ross RJ, Valcavi R. Effects of hypothyroidism and hyperthyroidism on insulin-like growth factors and growth hormone and IGF-binding proteins. *J Clin Endocrinol Metab*, 1993; **76**: 950–5.
58. Yamamoto K, Saito K, Takai T, Naito M, Yoshida S. Visual field defects and pituitary enlargement in primary hypothyroidism. *J Clin Endocrinol Metab*, 1983; **57**: 283–6.
59. Sarlis NJ, Brucker-Davis F, Doppman JL, Skarulis MC. MRI-demonstrable regression of a pituitary mass in a case of primary hypothyroidism after a week of acute thyroid hormone therapy. *J Clin Endocrinol Metab*, 1997; **82**: 808–11.
60. Iranmanesh A, Lizarralde G, Johnson ML, Veldhuis JD. Dynamics of 24 hour endogenous cortisol secretion and clearance in primary hypothyroidism assessed before and after partial hormone replacement. *J Clin Endocrinol Metab*, 1990; **70**: 155–60.
61. Deschepper CF, Hong-Brown LQ. Hormonal regulation of the angiotensinogen gene in liver and other tissues. In: Raizada MK, Phillips MI, Sumners C, eds. *Cellular and Molecular Biology of Renin-Angiotensin System*. Boca Raton, FL: CRC Press, 1993: 152–65.
62. Rodrigueaz-Gutierrez R, Gonzalez-Velazquez C, Gonzalez-Saldivar G, Villareal-Perez JZ, Gonzlalez-Gonzalez JG. Glucocorticoid functional reserve in full-spectrum intensity of primary hypothyroidism. *Int J Endocrinol*, 2014; **31**: 355–60.
63. Santini F, Marzullo P, Rotondi M, Ceccarin G, Pagano L, Ippolito S, Chiovato L, Biondi B. Mechanisms in endocrinology: the crosstalk between thyroid gland and adipose tissue: signal integration in health and disease. *Eur J Endocrinol*, 2014; **8**: 171–87.
64. Calvino C, Imperio GE, Wiliemam M, *et al*. Hypothyroidism induces hypophagia associated with alterations in protein expression of neuropeptide Y and proopiomelanocortin in the arcuate nucleus, indipenbdently of hypothalamic nuclei-specific changes in leptin signaling. *Thyoid*, 2016, **26**: 134–41.
65. Marchesini G, Fabbri A, Bianchi GP, *et al*. Hepatic conversion of amino nitrogen to urea nitrogen in hypothyroid patients and upon L-thyroxine therapy. *Metabolism*, 1993; **42**: 1263–8.
66. Duntas LH, Brenta G. The effect of thyroid disorders on lipid levels and metabolism. *Med Clin North Am*, 2012; **96**: 269–73.
67. Santos Palacios S, Brugos-Larumbe A, Guillen-Grima F, Galofre JC. A cross-sectional study of the association between circulating TSH level and lipid profile in a large Spanish population. *Clin Endocrinol*, 2013, **79**: 874–81
68. Carle A, Pedersen IB, Knudsen N, Perrild H, Oversen L, Laurberg P. Gender differences in symptoms of hypothyroidism: a population-based DanThyr study. *Clin Endocrinol*, 2015: **83**: 717–25.
69. Hypothyroid symptoms fail to predict thyroid insufficiency in old people: a population-based case control study. *Am J Med*, 2016; **129**: 1082–92.
70. Stott DJ, Rodondi N, Kearney PM, *et al*. Thyroid hormone therapy for older adults with subclinical hypothyroidism. *N Engl J Med*, 2017; **376**: 2534–44.

71. Cogni G, Chiovato L. An overview of the pathogenesis of thyroid autoimmunity. *Hormones*, 2013; **12**: 19–29.
72. Husebye ES, Anderson MS, Kampe O. Autoimmune polyendocrine syndromes. *N Engl J Med*, 2018; **378**: 1132–41.
73. Bartalena L, Marcocci C, Bogazzi F, Panicucci M, Lepri A, Pinchera A. Use of corticosteroids to prevent progression of Graves' ophthalmopathy after radioiodine therapy for hyperthyroidism. *N Engl J Med*, 1989; **321**: 1349–52.
74. Persani L. Central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*, 2012; **97**: 3068–18.
75. Billewicz WZ, Chapman RS, Crooks J, *et al*. Statistical methods applied to the diagnosis of hypothyroidism. *QJM*, 1969; **38**: 255–66.
76. Zulewski H, Müller B, Exer P, Miserez AR, Staub JJ. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. *J Clin Endocrinol Metab*, 1997; **82**: 771–6.

3.4.2 Causes and Laboratory Investigation of Hypothyroidism

Ferruccio Santini

Introduction

Hypothyroidism is the clinical state that develops as a result of the lack of action of thyroid hormones on target tissues [1]. Hypothyroidism is usually due to impaired hormone secretion by the thyroid, resulting in reduced concentrations of serum thyroxine (T_4) and triiodothyronine (T_3). The term primary hypothyroidism is applied to define the thyroid failure deriving from inherited or acquired causes that act directly on the thyroid gland by reducing the amount of functioning thyroid tissue or by inhibiting thyroid hormone production. The term central hypothyroidism is used when pituitary or hypothalamic abnormalities result in an insufficient stimulation of an otherwise normal thyroid gland. Both primary and central hypothyroidism may be transient, depending on the nature and the extent of the causal agent. Hypothyroidism following a minor loss of thyroid tissue can be recovered by compensatory hyperplasia of the residual gland. Similarly, hypothyroidism subsides when an exogenous inhibitor of thyroid function is removed.

Peripheral hypothyroidism may arise as a result of excessive thyroid hormone degradation due to anomalous overexpression of type 3 iodothyronine deiodinase in some neoplastic tissues, while loss of protein-bound thyroid hormones in patients with nephrotic syndrome may unravel or deteriorate a pre-existing thyroid dysfunction. Peripheral hypothyroidism may also occur as a consequence of tissue resistance to thyroid hormones due to a mutation in the thyroid hormone receptor or in the molecules involved in membrane transport and intracellular metabolism of thyroid hormones. Reduced sensitivity to thyroid hormones is a heterogeneous clinical entity with most patients appearing to be clinically euthyroid while some of them have symptoms of thyrotoxicosis and others display selected signs of hypothyroidism or neurodevelopment abnormalities, depending on the specific gene responsible for the disease. The variability in clinical manifestations depends on the severity of the hormonal resistance, the relative degree of hyposensitivity to thyroid hormone in different tissues, and the coexistence of associated genetic defects (see Chapter 3.4.8).

Primary Hypothyroidism

A list of the causes of primary hypothyroidism is given in Box 3.4.2.1. Autoimmune thyroiditis is the most common cause of spontaneous hypothyroidism in areas with adequate iodine intake. Iatrogenic hypothyroidism is responsible for many hypothyroid patients in these regions and inborn errors of thyroid hormone synthesis, goitrogens, and other destructive processes of the thyroid gland account for a few cases. Iodine deficiency is crucial in the pathogenesis of endemic cretinism and of adult hypothyroidism in areas in which an efficient iodine prophylaxis has not been undertaken.

Autoimmune Thyroiditis

Autoimmune thyroiditis includes a spectrum of diseases that are distinguished for their clinical course, the degree of thyroid dysfunction, and the changes of thyroid size. All these variants recognize an immune-mediated pathogenesis and usually present with high titres of circulating antithyroid antibodies. As in most organ-specific autoimmune reactions, the aetiology of autoimmune thyroiditis is still unknown but is somehow linked to genetic and environmental factors, and is influenced by the gender and the age (see Chapter 3.2.6).

Chronic Thyroiditis

Chronic thyroiditis is the most common among the autoimmune thyroiditis. Historically, two clinical variants of the disease are described. A goitrous variant (Hashimoto's thyroiditis), characterized by heavy lymphocytic infiltration and thyroid enlargement, and an atrophic variant (primary myxoedema) with progressive fibrosis and reduction of thyroid size. In clinical practice, a clear distinction between the two forms is not always possible. In the initial stage of the disease the two variants commonly do not present distinctive features. Moreover, atrophy may be a destructive end result of goitrous thyroiditis with the thyroid gland showing near complete replacement with fibrosis. When overt hypothyroidism has occurred, thyroid volume shows a unimodal distribution, with thyroid atrophy and goitre being extremes within the distribution

Box 3.4.2.1 Causes of primary hypothyroidism

- Autoimmune thyroiditis
  - Chronic thyroiditis
    - Hashimoto's thyroiditis
    - Atrophic thyroiditis
  - Postpartum thyroiditis
  - Graves' disease (spontaneous late evolution)
- Subacute thyroiditis
- Riedel's thyroiditis
- Iatrogenic
  - Thyroidectomy
  - ^{131}I therapy for hyperthyroidism
  - External radiotherapy
  - Excessive iodine
  - Drugs
    - Thionamides
    - Amiodarone
    - Lithium
    - Tyrosine kinase inhibitors
    - Immune checkpoint inhibitors
    - Others
- Severe iodine deficiency
- Natural goitrogens
- Thalassaemia major
- Congenital abnormalities
  - Thyroid dysgenesis
    - Agenesis
    - Ectopic gland
    - Hypoplasia
- Inherited defects in thyroid hormone biosynthesis
  - Iodide transport defect
  - Organification defect
  - Pendred's syndrome
  - Iodotyrosine deiodinase defect
  - Thyroglobulin defect
  - TSH-receptor defect
  - Gs-protein defects
- Transient neonatal hypothyroidism
  - Iodine deficiency or excess
  - Administration of antithyroid agents to the mother
  - Maternal TSH-blocking antibody.

[2]. Overall, these observations suggest that the two variants do not represent separate disorders.

Thyroid failure usually develops very slowly and, as thyroid function fades, the resulting increase in serum thyroid-stimulating hormone (TSH) limits the decline in thyroid secretion. Thus overt hypothyroidism is commonly preceded by a variable period of time in which elevated TSH is the only hormonal abnormality (subclinical hypothyroidism) [3, 4]. The transition from euthyroidism to hypothyroidism may pass unrecognized and initial symptoms may be attributed to ageing, menopause, or other chronic concomitant diseases [5]. Thus, it is not uncommon that chronic thyroiditis is diagnosed when clinical manifestations of thyroid failure become severe or complications of hypothyroidism have occurred. The circumstances leading to early diagnosis of the disease include family history for autoimmune thyroid diseases, appearance of goitre, blood testing for screening of autoimmune diseases in patients with polyglandular autoimmunity, routine diagnostic protocols for patients with menstrual dysfunction, or hyperlipidaemia.

Occasionally hypothyroidism may be due to TSH-receptor blocking antibodies preventing thyroid cell stimulation by TSH. TSH-receptor blocking antibodies are more frequent in atrophic thyroiditis than in goitrous thyroiditis [6]. Hypothyroidism may be reversible if the TSH-receptor blocking antibody titre declines and enough thyroid tissue remains for thyroid hormone synthesis. Graves' hyperthyroidism may develop in hypothyroid patients with chronic thyroiditis because of a change in the nature of TSH-receptor antibodies from blocking to stimulating [7].

In some instances, the disease may be preceded by a transient phase of thyrotoxicosis (hashitoxicosis) due to the discharge of preformed thyroid hormones, as a result of an unusually intense inflammatory process. The gland is tender and sometimes painful, resembling subacute thyroiditis. Hypothyroidism usually develops in a short time and may be permanent, especially in patients with elevated thyroid peroxidase antibody.

Features of thyroid-associated ophthalmopathy may occur in patients with chronic thyroiditis and hypothyroidism. This condition is termed 'hypothyroid Graves' disease' and may be the endstage of Graves' disease after spontaneous remission of hyperthyroidism or may represent a distinct entity with pathogenetic mechanisms common to Graves' disease, involving thyroid-stimulating antibody [8].

Focal thyroiditis is characterized by spotty collections of mononuclear cells within thyroid tissue, and minimal changes in follicular epithelium or stromal fibrosis. Most patients with focal thyroiditis are euthyroid and only 10–20% have subclinical hypothyroidism [9]. The disease may be suspected at ultrasound examination in patients with circulating thyroid autoantibodies, or may be a histological occurrence in surgical or autopsy specimens. In the presence of circulating thyroid autoantibodies, focal thyroiditis may represent the earliest stage of chronic autoimmune thyroiditis, whereas the clinical significance of non-specific isolated lymphocytic infiltration in patients without circulating autoantibodies has still to be clarified.

Juvenile thyroiditis (autoimmune thyroiditis in childhood and adolescence) is described as a separate entity because follicular oxyphilia is usually mild or absent, goitre is soft, and thyroid antibody titres are not as high as in adults. Fine-needle aspiration biopsy is sometimes required to establish the diagnosis. Spontaneous resolution is relatively common, but hypothyroidism may develop during the course of the disease [10].

Postpartum Thyroiditis

Pregnancy is known to influence the clinical course of various autoimmune disorders, including autoimmune thyroid disease. Typically, amelioration during pregnancy is followed by aggravation after delivery. This phenomenon is thought to depend on the physiological need of inhibiting maternal immune reactions that might cause rejection of the fetus. Thus, thyroid peroxidase antibodies, thyroglobulin antibodies, and TSH-receptor antibody titres decrease or may even disappear during pregnancy. Following delivery, a rebound of autoimmune processes occurs and may result in destructive thyroiditis with release of preformed thyroid hormones and transient thyrotoxicosis, followed by a transient hypothyroid phase. Persistent hypothyroidism may develop in a minority of women. This clinical entity is named postpartum thyroiditis and occurs in 5–9% of unselected postpartum women [11] (see Chapter 3.4.6).

Graves' Disease

Spontaneous hypothyroidism may develop during the course of Graves' disease whenever destructive processes of thyroiditis predominate over thyroid-stimulating events (burnt out Graves' disease). This may occur after long-term remission of hyperthyroidism associated with disappearance of TSH-receptor stimulating antibodies, or following prolonged therapy with antithyroid drugs [12]. TSH-receptor blocking antibodies may also appear and neutralize TSH-receptor stimulating antibodies, leading to hypothyroidism.

Subacute Thyroiditis

Subacute thyroiditis is an inflammatory disease of viral origin [13]. Although the disease is relatively uncommon it must be suspected any time a patient presents with anterior neck pain. Recovery is complete in most patients but in rare cases (1–5%) epithelial loss is severe, resulting in permanent hypothyroidism (see Chapter 3.2.7).

Riedel's Thyroiditis

Riedel's thyroiditis is an extremely rare chronic disease of unknown aetiology, characterized by progressive fibrosis of the thyroid and surrounding tissues [14]. Hypothyroidism develops when fibrosclerosis has involved most of the gland (see Chapter 3.2.7).

Iatrogenic Hypothyroidism

Thyroid ablation for therapeutic purposes is a common cause of primary hypothyroidism in the adult. Thyroid failure is an obvious consequence of total or subtotal thyroidectomy for thyroid cancer, goitre, or Graves' disease, but clinical hypothyroidism does not develop as long as substitutive therapy is started shortly after surgery. Similarly, ^{131}I therapy for Graves' disease is directed to destroy thyroid tissue. However, the success rate of radioiodine therapy and the time of onset of hypothyroidism are not fully predictable; they depend on several factors including the dose of radiation delivered, the size of the goitre, and the underlying autoimmune phenomena [15]. Drug-induced hypothyroidism is also common. Excessive inhibition of thyroid hormone synthesis commonly occurs during therapy for hyperthyroidism with antithyroid agents. Furthermore, primary hypothyroidism may develop as a side effect of several drugs administered for different purposes.

Postoperative Hypothyroidism

Total thyroidectomy is performed for thyroid cancer, Graves' disease, and large diffuse or multinodular goitres, occasionally also harbouring Hashimoto's thyroiditis. However, hypothyroidism does not develop as long as L-thyroxine replacement therapy is started soon after thyroidectomy. In the past, patients with thyroid cancer had to discontinue thyroid hormone therapy before ^{131}I scanning and therapy. The availability of recombinant human TSH has greatly reduced the need for thyroid hormone withdrawal before radioiodine administration (see Chapter 3.5.6).

The frequency of hypothyroidism after subtotal thyroidectomy varies depending on the mass of remaining tissue and the degree of its autonomous function. A small thyroid residue may be sufficient for maintenance of the euthyroid state in Graves' disease. On the other hand, a large residue of a multinodular or Hashimoto's goitre may not be enough for adequate thyroid hormone secretion. Partial thyroidectomy or lobectomy for multinodular goitres or solitary nodules are usually not associated with permanent hypothyroidism, although L-thyroxine is usually administered to prevent relapse of the goitre.

Post-Irradiation Hypothyroidism

Among different radioactive isotopes of iodine, ^{131}I is the agent of choice in the treatment of thyroid hyperfunction. After oral administration, radioiodine is completely absorbed, rapidly concentrated, oxidized, and organified by thyroid follicular cells. The biological effects of radioiodine include necrosis of follicular cells, shorter survival and impaired replication of undestroyed cells, and vascular occlusion, leading to atrophy and fibrosis of thyroidal tissue.

The goal of radioiodine therapy for hyperthyroidism is to destroy sufficient thyroid tissue to cure the hyperthyroidism with one dose of ^{131}I. This dose is calculated on the basis of thyroid size and uptake of ^{131}I. Because of radiation safety restrictions, in some centres small repeated doses of radioiodine are administered. In other centres standard fixed doses are given. Small glands are destroyed more readily by radioiodine than larger ones, and toxic adenoma or toxic multinodular goitre are usually more radioresistant than Graves' glands. Radioiodine has a delayed effect and several months may be required for the complete control of hyperthyroidism.

In the case of Graves' disease, the goal of radioiodine should be to destroy as much thyroid tissue as possible [16]. This strategy has been adopted because residual tissue, necessary to ensure euthyroidism, is responsible for the relapse of hyperthyroidism in a large proportion of patients. A strict control of thyroid function is required during the first 6–12 months following ^{131}I therapy for Graves' disease to avoid the appearance of symptoms of hypothyroidism, which may be rapidly progressive and severe. Early post-radioiodine hypothyroidism may be transient, and hyperthyroidism may relapse during L-thyroxine replacement therapy.

Radioiodine-induced hypothyroidism is less frequent after treatment for toxic adenoma or multinodular goitre because non-functioning thyroid tissue should not receive the radioisotope. Yet, hypothyroidism may develop whenever TSH is not completely suppressed at the time ^{131}I is administered. Furthermore, a small degree of iodine uptake is maintained in normal thyroid cells even in the absence of TSH stimulation, and this may be the cause of hypothyroidism many years after radioiodine administration.

External irradiation to the neck for non-thyroidal neoplasias (lymphomas, tumours of the head and neck, spinal tumours, or metastases) may produce hypothyroidism in up to 50% of patients [17]. Thyroid failure may develop after a variable interval, depending on the dose of radiation that has been administered. Hypothyroidism after total body irradiation for acute leukaemia or aplastic anaemia has also been reported [18]. An increased risk of hypothyroidism has been found in breast cancer patients treated with radiation, since a portion of the thyroid gland may be included in the treatment fields [19].

Drug-Induced Hypothyroidism

Transient hypothyroidism is common in the course of medical treatment for hyperthyroidism with thionamides, and quickly subsides with adjustment of the dose. Excess iodide, such as in disinfectants, radiographic contrast agents, and seaweed-containing preparations, may precipitate hypothyroidism in autoimmune chronic thyroiditis, due to failure of the thyroid to escape from the

Wolff–Chaikoff effect. Animal studies suggest that excessive iodide increases the incidence of thyroid autoimmunity but evidence in humans is controversial. Amiodarone is an antiarrhythmic agent containing about 37 mg iodine per 100 mg drug. Amiodarone may produce hypothyroidism by the excess iodine released with metabolism of the drug. As in other cases of excess iodine administration, an underlying autoimmune thyroid disease is a prerequisite. Amiodarone may also induce destructive thyroiditis in an otherwise normal thyroid gland. The prevalence of overt hypothyroidism may be as high as 5% of amiodarone-treated patients, with no clear association between the occurrence of hypothyroidism and the dose of the drug. Substitutive L-T_4 can be administered with no need to discontinue amiodarone, if the antiarrhythmic drug is essential for the underlying cardiac disease. Mild (subclinical) hypothyroidism is more common and does not necessarily progress to overt hypothyroidism. Treatment of mild hypothyroidism may be avoided in patients at high risk of cardiovascular events [20].

Lithium inhibits thyroid hormone synthesis and secretion, and lithium therapy for psychiatric disorders is associated with an increased risk of hypothyroidism, with a hazard ratio of 2.31 [21]. The risk is increased in patients with positive antithyroid antibodies or with minor thyroid abnormalities, which reduce the ability of the thyroid gland to override the inhibitory effects of lithium. Goitre is also common in lithium-treated patients, even when serum thyroid hormones and TSH are within normal limits. If hypothyroidism appears, LT4 therapy should be initiated and lithium therapy may be continued.

Tyrosine kinase inhibitors are newly developed drugs approved for the treatment of several tumours. The first observation of hypothyroidism after sunitinib treatment has been reported in 2006 [22]. Since then several studies have been published and have confirmed that various tyrosine kinase inhibitors can affect thyroid function tests through different physiopathological mechanisms impairing thyroid function or thyroid hormone metabolism [23].

Targeted manipulation of immune checkpoints by monoclonal antibodies against molecules that are critical for immune regulation, has been recently introduced as an effective anticancer therapy for its ability to enhance the immune responses against malignant cells. Triggering of autoimmunity due to disruption of immunological tolerance to self-antigens may induce immune-related adverse events, with primary hypothyroidism scoring among the most frequent, usually associated with appearance of thyroid autoantibodies [24].

Several other drugs have been reported to be capable of inducing primary hypothyroidism [25]. Treatment with interferon-α or interleukin-2 may produce hypothyroidism, thyrotoxicosis, or the biphasic pattern of silent thyroiditis. Pre-existent thyroid autoimmunity increases the risk of thyroid dysfunction during treatment with these agents. Other medications occasionally reported to induce hypothyroidism include tricyclic antidepressants, selective serotonin reuptake inhibitors, rifampin, sulphonamides, sulphonylureas, ethionamide, *p*-aminosalicylic acid, phenylbutazone, and nicardipine, but the antithyroid potential of these drugs is weak and an underlying thyroid abnormality or concurrent iodine deficiency are usually associated.

Severe Iodine Deficiency and Natural Goitrogens

Environmental iodine deficiency is common in many areas throughout the world, particularly in inland mountainous areas. Goitre is the most common disorder due to iodine deficiency and its prevalence is inversely related to the median iodine intake of the population. Endemic goitre is usually not associated with hypothyroidism. However, the pattern of circulating thyroid hormones in the population from areas of severe iodine deficiency differs from that found in iodine-sufficient areas [26]. The mean serum T_4 is reduced while serum T_3 is unchanged or increased and an inverse correlation between serum TSH and T_4 is found. The low iodine content within the thyroid gland and the increased TSH stimulation lead to preferential secretion of T_3, which is far more potent than T_4 in terms of metabolic responses. Thus, the relative increase in T_3 secretion enables a patient to maintain the euthyroid status in spite of reduced availability of iodide.

Cretinism is the result of an insufficient supply of thyroid hormones to fetal tissues and is due to severe iodine deficiency in both the mother and the fetus during early stages of gestation [27]. Fetal hypothyroidism is not compensated by transplacental passage of maternal T_4 and is responsible for severe physical and neurological damage.

Adult hypothyroidism may occur in rural populations living in areas of severe iodine deficiency where isolation prevents access to iodine-rich foodstuff. In this case, hypothyroidism is rapidly reversed by iodine supplementation. Consumption of food containing antithyroid agents, such as thiocyanate in cassava meal and flavonoids in a variety of plants, may aggravate the effects of dietary iodine deficiency and add to the development of goitre and hypothyroidism. Phloroglucinol, a potent antithyroid compound contained in some species of seaweeds, may play an additional role to that of iodine excess in the development of iodine-induced hypothyroidism.

Acquired hypothyroidism has been described in children on chronic parenteral nutrition as a consequence of the lack of iodine in delivered nutrients [28]. More recently, attention has been focused on environmental endocrine disrupters (pesticides and industrial pollutants) as a possible cause of thyroid imbalance, but their effects on human thyroid function have not been fully elucidated (see also Chapter 3.2.2) [29].

Thalassaemia Major

A high prevalence of primary hypothyroidism has been described in patients with thalassaemia major. The incidence and severity of thyroidal dysfunction appears related to the degree of iron overload. Hypothyroidism may be reversible if intensive chelation therapy is timely started [30].

Congenital Abnormalities

Congenital hypothyroidism is detected in about 1/2000 neonates, its prevalence being increased over the years mainly due to lowering of TSH cut-offs for primary screening of the disease [31]. Primary congenital hypothyroidism accounts for most affected children, whereas central congenital hypothyroidism is rare. In a small proportion of detected cases a transient dysfunction occurs, which recover in the following months. Screening programmes for early detection of congenital hypothyroidism are active in developed countries, but many affected newborns worldwide remain unrecognized and the burden of the disease is still a relevant public health challenge [32] (see Chapter 3.4.7).

Both the fetus and the neonate are particularly sensitive to the block of thyroid function induced by excess iodide since the immature gland is not able to escape from the Wolff–Chaikoff effect [33]. Iodide-induced transient hypothyroidism is most common in

premature infants and in low-birthweight babies, and has occurred more in relatively iodine-deficient areas of Europe [34], than in iodine-sufficient North America.

Transient fetal–neonatal hypothyroidism and goitre may develop in babies born to hyperthyroid mothers with Graves' disease treated with excessive doses of propylthiouracil or methimazole. Both hypothyroidism and goitre resolve spontaneously with the clearance of the drug from the circulation of the neonate. TSH-receptor blocking antibodies may be present in patients with autoimmune hypothyroidism; the antibodies compete with TSH and inhibit the biological effects of TSH on thyroid cell function and growth [6]. These antibodies have been found mainly in patients with autoimmune atrophic thyroiditis, and contribute to the development of thyroid failure and atrophy. The maternal TSH-receptor antibody responsible for thyroid failure in the neonate inhibits TSH binding to its receptor and therefore blocks the effect of TSH on adenylate cyclase stimulation, iodine uptake, and thyroid cell growth [35]. Delayed thyroid development due to the transplacental transfer of potent TSH-receptor blocking antibodies has been described [36]. TSH-receptor blocking antibody may also occur in women with Graves' disease and be transmitted to the fetus. Although thyroid-stimulating antibodies usually predominate in these patients, transient hypothyroidism is possible in the offspring of women with Graves' disease due to very high TSH-blocking antibody titres and relatively low concentrations of thyroid-stimulating antibody. Because TSH-induced growth is blocked, these infants do not have a goitre.

Central Hypothyroidism

Central hypothyroidism is the consequence of anatomical or functional disorders of the pituitary or the hypothalamus [37]. Several of the causes reported in **Box 3.4.2.2** may affect both the pituitary and the hypothalamus, and in many instances the main anatomical site of the dysfunction cannot be identified. Thus, the former terms of secondary hypothyroidism (of pituitary origin) and tertiary hypothyroidism (of hypothalamic origin) are no longer recommended. Central hypothyroidism is rarely isolated, being part of a generalized disorder involving the secretion of other pituitary hormones. Permanent central hypothyroidism is rare, its prevalence ranging from 1:20 000 to 1:80 000 in the general population. However, transient functional abnormalities of TSH secretion are relatively common, and often pass unrecognized due to rapid recovery of the normal thyroid hormone balance.

Pituitary adenomas represent the most common cause of central hypothyroidism. Reduced secretion of TSH is usually a consequence of mechanical compression of non-tumorous cells and of adenohypophyseal blood vessels by the adenoma. The pituitary stalk and the hypothalamus may also be involved by suprasellar extension of the tumour. The tumour may be non-functioning or secrete other hormones. Thus, the resulting syndrome will depend on the extent of hypopituitarism and on the particular hormone secreted by the adenoma. A sudden enlargement of pituitary adenomas may occur as a result of haemorrhage within the tumour, leading to pituitary apoplexy.

Several other causes may produce central hypothyroidism, by acting at the hypothalamic or pituitary level. Primary extrasellar brain tumours or metastatic tumours originating from other sites may produce a variable degree of hypopituitarism, depending on the location and the extension of their mass. Among brain tumours, craniopharyngiomas should be suspected when central hypothyroidism is diagnosed in young people. Craniopharyngiomas are usually extrasellar but they may extend inferiorly causing destruction of the bony margins of the sella. Pituitary infarction may develop postpartum following excessive blood loss during delivery (Sheehan's syndrome), or in patients with severe shock or during systemic anticoagulation therapy. Various degrees of pituitary insufficiency may be observed in these cases. Traumatic head injuries can lead to central hypothyroidism because of hypothalamic or pituitary infarction or haemorrhage. Iatrogenic causes of central hypothyroidism include external radiation and surgery for pituitary or brain tumours. The empty sella syndrome is caused by a defect of the sellar diaphragm leading to cisternal herniation within the pituitary fossa and flattening of the pituitary. Hypopituitarism develops

Box 3.4.2.2 Causes of central hypothyroidism

- Tumours
  - Pituitary adenomas
  - Craniopharyngioma
  - Meningioma
  - Dysgerminoma
  - Other brain tumours
  - Metastatic tumours
- Ischaemic necrosis
  - Postpartum (Sheehan's syndrome)
  - Severe shock
  - Diabetes mellitus
- Aneurysm of internal carotid artery
- Iatrogenic
  - External radiation
  - Surgery
- Infectious diseases
  - Abscesses
  - Tuberculosis
  - Syphilis
  - Toxoplasmosis
- Sarcoidosis
- Histiocytosis
- Haemosiderosis
- Chronic lymphocytic hypophysitis
- Empty sella
- Traumatic brain injury
- Subarachnoid haemorrhage
- Pituitary dysplasia
- Congenital malformations of the hypothalamus
- Inherited diseases
  - Combined pituitary hormone abnormalities: pituitary transcription factor defects; immunoglobulin superfamily gene 1 mutations
  - Isolated CH: TSHβ-subunit; thyrotropin-releasing hormone receptor (TRHR) mutations
- Transient central hypothyroidism
  - Recovery from prolonged thyrotoxicosis
  - Severe non-thyroidal illnesses
  - Drugs
    - Somatostatin analogues
    - Glucocorticoids
    - Dopamine
    - Bexarotene

slowly along with expansion of the cisternal herniation caused by transmission of cerebrospinal fluid pressure. Hypothalamic or pituitary lesions may derive from any of the infectious or granulomatous diseases listed in Box 3.4.2.2.

Autoimmune (primary) hypophysitis is an increasingly recognized disease that can cause central hypothyroidism, isolated or associated with other tropin defects, in up to 45% of the cases. However, when an appropriate treatment with immunosuppressive drugs is started, TSH deficiency can considerably improve with no need of hormone replacement therapy with L-thyroxine in the long term [38]. Another emerging disorder, first described in 2003, is hypophysitis secondary to checkpoint inhibitors immunotherapy [39]. In this condition, observed in approximately 10% of cancer patients treated, various degrees of hypopituitarism, including isolated central hypothyroidism, have been described.

A high prevalence of hypothyroidism following traumatic brain injury or subarachnoid haemorrhage has been demonstrated, although TSH deficiency is less common than growth hormone, luteinizing hormone/follicle-stimulating hormone, and adrenocorticotropic hormone deficiencies [40]. Pituitary aplasia or hypoplasia is a rare congenital defect, usually associated with other severe malformations. In most instances these patients die shortly after birth.

Genetic abnormalities in TSH synthesis may cause central hypothyroidism characterized by inherited isolated TSH deficiency. A number of mutations in genes involved in pituitary function have been described, leading to isolated central hypothyroidism or combined pituitary hormone defects, with variable clinical phenotypes and degrees of severity [41]. In some patients no demonstrable pathology can be found to explain TSH deficiency, and the term idiopathic central hypothyroidism is still applied.

Transient impairment of TSH secretion is commonly observed and may depend on a variety of causes, including the use of several drugs (see Box 3.4.2.2). The recognition of these conditions is essential to avoid unnecessary and expensive diagnostic procedures. In most instances, replacement therapy is not necessary or is contraindicated.

Laboratory Investigation of Hypothyroidism

The diagnosis of hypothyroidism and of its cause requires the evaluation of several clinical, laboratory, and instrumental parameters to manage the patient properly (Table 3.4.2.1).

Hormonal Evaluation

A small decrease in thyroid secretion may produce only minor changes in serum concentrations of thyroid hormones that remain within the normal range. The most sensitive index of a reduction in serum thyroid hormone concentration is serum TSH because of a decrease in feedback inhibition of pituitary TSH secretion. Thus, elevated serum TSH is the earliest laboratory abnormality in patients with primary hypothyroidism.

The combination of normal thyroid hormones and elevated TSH has been long defined as subclinical hypothyroidism. However, this term may be misleading because it suggests the absence of symptoms and signs of hypothyroidism, whereas they may be present if thyroid hormone deficiency does actually exist [42]. Indeed, an isolated increase in serum TSH may arise as a consequence of mild thyroid failure due to a recognized pathogenic event in an otherwise normal thyroid gland (most commonly autoimmune thyroiditis), and in this case grading of hypothyroidism based on biochemical criteria should be preferred. Differently, a slight TSH elevation may depend on inherited defects that produce a minor impairment of TSH action on thyroid cells, leading to isolated hyperthyrotropinaemia and an apparent euthyroid state [43]. A mild increase in serum TSH can also be observed in obese subjects, which is secondary to the obesity state and is reversible after weight loss [44]. Distinction among these conditions is important to decide on the opportunity of starting the replacement therapy.

With the progression of thyroid dysfunction, serum levels of T_4 fall below the normal limit while serum T_3 may still be normal. This is because high TSH levels induce preferential secretion of T_3 by residual thyroid tissue.

The lack of TSH response to reduced thyroid hormone levels complicates the diagnosis of central hypothyroidism, and the finding of low serum T_4 is a prerequisite for the diagnosis of this condition. Usually in central hypothyroidism, basal serum TSH concentrations are inappropriately low with respect to reduced serum thyroid hormones. Yet, in some instances serum TSH may be slightly elevated due to secretion of immunoreactive but biologically inactive TSH [45].

Assays for measurement of total thyroid hormones in serum are gradually being replaced by methods that determine the free (unbound) fraction of T_4 and T_3 [46]. Although measurement of free T_4 and free T_3 concentrations is more cumbersome as compared to that for total T_4 and T_3, free T_4 and free T_3 determinations are preferred because free thyroid hormones are those capable of entering

Table 3.4.2.1 Differential diagnosis of hypothyroidism

	Primary	Central	Resistance to thyroid hormone	Non-thyroidal illness
Symptoms of hypothyroidism	Present	Present	Occasionally present	Absent
Thyroid volume	↑, N,↓	N,↓	↑	N
TSH	↑	N, ↓, (↑)	N, ↑	N, ↓, (↑)
Free T_4	↓	↓	↑	N, ↓, (↑)
Free T_3	N, ↓	N, ↓	↑	↓
Radioiodine uptake	↑, N, ↓	↓	↑	N, ↓
TSH response to TRH	↑	N, ↓	↑	N, ↓, (↑)

↑, increased; N, normal; ↓, decreased; (·), slight changes.

the cell and therefore represent the biologically active hormone. Indeed, the concentrations of total thyroid hormones may be elevated or reduced in spite of normal free fractions, due to changes in the concentrations of serum transport protein (see Chapter 3.1.2). Current free thyroid hormone immunoassays are valuable in most clinical instances but more accurate methodologies may be required in selected cases [47].

The initial assessment of thyroid function may be accomplished by testing of TSH first, followed by reflex testing for free T_4 if TSH is outside the reference range. The implementation of appropriate TSH reference ranges is essential to achieve optimal effects on case detection [48]. However, it must be emphasized that TSH alone is not reliable for the detection of central hypothyroidism and may be misleading in other clinical settings, including treated hyperthyroidism and hypothyroidism before stabilization of the hormonal equilibrium, combined therapy with T_4 and T_3, and non-thyroidal illness.

Measurement of the serum TSH response to thyrotropin-releasing hormone (TRH) (200–500 μg intravenously) may be useful in selected patients with a borderline to low value of T_4 and borderline to high or borderline to low values of basal TSH, to identify subclinical primary or central hypothyroidism, respectively [49]. An exaggerated response will be observed in primary hypothyroidism whereas in central hypothyroidism the serum response of TSH may be reduced or abnormally prolonged. The TRH test may be useful also to measure the increase in serum T_3 levels following the rise in serum TSH. In people with normal thyroid function, serum T_3 increases 30–100% above the baseline value 120–180 min after the injection of 200 μg TRH. In central hypothyroidism, the T_3 response may be impaired or absent in spite of a normal peak of TSH, indicating secretion of biologically inactive TSH. Evaluation of the nocturnal surge of TSH in samples taken every 30 min from 11.00 p.m. to 2.00 a.m. may be useful to confirm the diagnosis of central hypothyroidism. At variance with people with normal thyroid function, the TSH surge is blunted or absent in central hypothyroid patients [50].

A transient phase of central hypothyroidism may occur in patients with non-thyroidal illness, particularly hospitalized patients with medical or psychiatric illnesses. In these cases, repeated hormonal measurements are useful since values usually become normal as patients recover from that illness.

Other *In Vitro* Tests

Antithyroglobulin and antithyroperoxidase antibodies are sensitive markers of thyroid autoimmunity [51]. Thus, if present, they may contribute to the diagnosis of autoimmune thyroiditis, often preceding the appearance of thyroid dysfunction, represent a prognostic index for the development of postpartum thyroiditis, and help to predict the outcome of iodine- or drug-induced hypothyroidism. Antithyroperoxidase antibodies and, less frequently, antithyroglobulin antibodies are observed in most patients with chronic autoimmune thyroiditis but hypothyroid patients with serum negative Hashimoto's disease have also been reported, who display a milder clinical picture compared with classic Hashimoto's thyroiditis [52]. Screening for antithyroperoxidase and antithyroglobulin antibodies in patients with non-thyroidal autoimmune diseases may be helpful for early recognition of concurrent autoimmune thyroid diseases, in the context of polyglandular autoimmune syndromes [53].

TSH-receptor antibodies can either have stimulating activity (thyroid-stimulating antibody), as in Graves' disease, or block the receptor (TSH-receptor blocking antibody) preventing TSH stimulation of the follicular cell. TSH-receptor blocking antibodies are highly specific for autoimmune thyroiditis. They are found in up to 30% of patients with chronic autoimmune thyroiditis and can produce or add to the development of hypothyroidism by blocking the thyroid response to TSH [6]. Hypothyroidism produced by TSH-receptor blocking antibodies can spontaneously remit following disappearance of antibody from serum. Assays for TSH-receptor antibodies measure the ability of a patient's IgG to inhibit the binding of ^{125}I-TSH to its receptor in thyroid membrane preparations. Radioreceptor assays are now easy to perform, inexpensive, and provide reliable results but do not distinguish thyroid-stimulating antibodies from TSH-receptor blocking antibodies. For this purpose, methods that assess the capacity of IgG to stimulate or to prevent TSH-induced cAMP production in thyroid preparations are necessary.

Endogenous antibodies against thyroid hormones may develop in patients with autoimmune thyroiditis [54]. These antibodies usually have no clinical relevance, but may interfere on assays for serum total and free T_4 and T_3, producing artefactual results depending on the technique used to measure the hormones. The presence of T_4 or T_3 antibodies should always be suspected in autoimmune patients with unexpected results of thyroid hormone assays. These antibodies can be detected easily by immunoprecipitation of radiolabelled T_4 or T_3 with the patient's serum.

Thyroglobulin is present at low concentrations in serum of people with normal thyroid function, and is elevated in all states associated with enlargement, hyperfunction, or injury of the thyroid gland. Measurement of serum thyroglobulin has no meaning for the diagnosis or the management of hypothyroidism, but may be useful to estimate the amount of residual thyroid tissue after surgery or other thyroid destructive events. Furthermore, detectable serum thyroglobulin in congenital hypothyroidism excludes thyroid agenesis. Increased levels of serum thyroglobulin in iodine deficiency have also been observed [55]. Antithyroglobulin antibodies in serum interfere with measurement of thyroglobulin and therefore this test should not be performed in such patients.

Measurement of urinary iodide provides information about the daily iodide intake in epidemiological studies [56]. The demonstration of elevated concentrations of urinary iodide in a hypothyroid patient may be useful if exposure to excessive iodide is suspected.

Thyroid Imaging in Hypothyroidism

Ultrasonography

Thyroid ultrasonography may be helpful in determining the cause of hypothyroidism by providing important information on location, size, structure, and vascularity of the gland. In autoimmune thyroiditis a gross inhomogeneity and low echogenicity characterize the echo pattern of the gland. Areas of apparently normal tissue of variable size may be observed, whereas true nodules reflect a different aetiology and should raise the possibility of coexisting nodular goitre, adenomas, or malignancies. A diffuse low thyroid echogenicity is indicative of diffuse autoimmune involvement of

the gland and is associated with or may predict the development of hypothyroidism [57]. Studies using colour flow Doppler show a variable degree of vascularity in goitrous autoimmune thyroiditis, whereas vascularity is decreased in the atrophic variant of the disease. In subacute thyroiditis the gland is usually enlarged and presents large hypoechoic areas with poorly defined boundaries, mainly within the painful lobe. A large diffuse or multinodular goitre can be documented by ultrasonography in hypothyroidism with inherited defects in thyroid hormone biosynthesis. No evidence of thyroidal tissue in its appropriate location and the demonstration of an ectopic gland are helpful in the diagnosis of congenital hypothyroidism due to thyroid dysgenesis.

In Vivo Isotopic Tests

Thyroid scintiscan may be helpful in the evaluation of hypothyroid patients to indicate the location of functioning thyroid tissue and to provide an estimation of overall thyroid size, although in this regard better evidence is usually obtained by thyroid ultrasonography. Occasionally scintiscan may reveal ectopic thyroid tissue not discernible by other means (e.g. lingual thyroid). Thyroid scintiscan can also be used to reveal substernal thyroid tissue when hypothyroidism is associated with a large goitre.

Radioiodine uptake is expressed as the percentage of radioactivity that is trapped by the thyroid at a given time after administration of a tracer quantity of inorganic radioiodine. Early radioiodine uptake measurements (3–6 h) provide information on the rates of transport and organification of iodide within the gland, whereas 24- and 48-h radioiodine uptake measurement reflects the rate of release of radioiodine from thyroidal tissue. It is also a way of estimating the extrathyroidal pool of iodide, being low to absent after intake of excess iodide but increased in iodine deficiency. An exception is represented by amiodarone-induced hypothyroidism in which radioiodine uptake is preserved despite iodine excess [58]. Radioiodine uptake is increased if hypothyroidism is caused by defective synthesis of thyroid hormones since TSH stimulates all steps in hormone synthesis capable of response. In chronic autoimmune thyroiditis values of the radioiodine uptake depend on the amount of residual functioning thyroid tissue and the serum concentration of TSH. Radioiodine uptake may be normal or even increased during the initial phase of chronic thyroiditis, whereas it tends to decrease as the disease progresses. Very low values of the radioiodine uptake are characteristic of the early phase of destructive thyroiditis (e.g. subacute thyroiditis) which is usually associated with thyrotoxicosis caused by follicular disruption. In these cases, return of radioiodine uptake to within the normal range may be helpful to indicate recovery of thyroid function. Radioiodine uptake measurement, which is obviously reduced in postablative hypothyroidism, may be used occasionally to estimate the amount of residual thyroid tissue after thyroidectomy or radioactive treatment.

REFERENCES

1. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet*, 2017; **390**: 1550–62.
2. Carlè A, Pedersen IB, Knudsen N, *et al*. Thyroid volume in hypothyroidism due to autoimmune disease follows a unimodal distribution: evidence against primary thyroid atrophy and autoimmune thyroiditis being distinct diseases. *J Clin Endocrinol Metab*, 2009; **94**: 833–9.
3. Tunbridge WM, Brewis M, French JM, *et al*. Natural history of autoimmune thyroiditis. *BMJ*, 1981; **282**: 258–62.
4. Effraimidis G, Strieder TG, Tijssen JG, Wiersinga WM. Natural history of the transition from euthyroidism to overt autoimmune hypo- or hyperthyroidism: a prospective study. *Eur J Endocrinol*, 2011;**164**: 107–13.
5. Mariotti S, Chiovato L, Franceschi C, Pinchera A. Thyroid autoimmunity and aging. *Exp Gerontol*, 1998; **33**: 535–41.
6. Chiovato L, Vitti P, Santini F, *et al*. Incidence of antibodies blocking thyrotropin effect *in vitro* in patients with euthyroid or hypothyroid autoimmune thyroiditis. *J Clin Endocrinol Metab*, 1990; **71**: 40–5.
7. McLachlan SM, Rapoport B. Thyrotropin-blocking autoantibodies and thyroid-stimulating autoantibodies: potential mechanisms involved in the pendulum swinging from hypothyroidism to hyperthyroidism or vice versa. *Thyroid*, 2013; **23**: 14–24.
8. Kahaly GJ, Diana T, Glang J, Kanitz M, Pitz S, König J. Thyroid stimulating antibodies are highly prevalent in Hashimoto's thyroiditis and associated orbitopathy. *J Clin Endocrinol Metab*, 2016; **101**: 1998–2004.
9. Mizukami Y, Michigishi T, Nonomura A, Nakamura S, Ishizaki T. Pathology of chronic thyroiditis: a new clinically relevant classification. *Pathol Annu*, 1994; **29**: 135–58.
10. Gopalakrishnan S, Marwaha RK. Juvenile autoimmune thyroiditis. *J Pediatr Endocrinol Metab*, 2007; **20**: 961–70.
11. Lazarus JH. Clinical manifestations of postpartum thyroid disease. *Thyroid*, 1999; **9**: 685–9.
12. Wood LC, Ingbar SH. Hypothyroidism as a late sequela in patient with Graves' disease treated with antithyroid agents. *J Clin Invest*, 1979; **64**: 1429–36.
13. Volpe R. The management of subacute (DeQuervain's) thyroiditis. *Thyroid*, 1993; **3**: 253–5.
14. Falhammar H, Juhlin CC, Barner C, Catrina SB, Karefylakis C, Calissendorff J. Riedel's thyroiditis: clinical presentation, treatment and outcomes. *Endocrine*, 2018; **60**: 185–92.
15. Chiovato L, Santini F, Pinchera A. Treatment of hyperthyroidism. *Thyroid Int*, 1995; **2**: 1–16.
16. Chiovato L, Fiore E, Vitti P, *et al*. Outcome of thyroid function in Graves' patients treated with radioiodine: role of thyroid-stimulating and thyrotropin-blocking antibodies and of radioiodine-induced thyroid damage. *J Clin Endocrinol Metab*, 1998; **83**: 40–6.
17. Boomsma MJ, Bijl HP, Langendijk JA. Radiation-induced hypothyroidism in head and neck cancer patients: a systematic review. *Radiother Oncol*, 2011; **99**: 1–5.
18. Farhadfar N, Stan MN, Shah P, *et al*. Thyroid dysfunction in adult hematopoietic cell transplant survivors: risks and outcomes. *Bone Marrow Transplant*, 2018; Feb 6; 1–6. https://doi.org/10.1038/s41409-018-0109-5.
19. Kikawa Y, Kosaka Y, Hashimoto K, *et al*. Prevalence of hypothyroidism among patients with breast cancer treated with radiation to the supraclavicular field: a single-centre survey. *ESMO Open*, 2017; **2**: e000161.
20. Bartalena L, Bogazzi F, Chiovato L, Hubalewska-Dydejczyk A, Links TP, Vanderpump M. 2018 European Thyroid Association (ETA) guidelines for the management of amiodarone-associated thyroid dysfunction. *Eur Thyroid J*, 2018; **7**: 55–66.

21. Shine B, McKnight RF, Leaver L, Geddes JR. Long-term effects of lithium on renal, thyroid, and parathyroid function: a retrospective analysis of laboratory data. *Lancet*, 2015; **386**: 461–8.
22. Desai J, Yassa L, Marqusee E, *et al*. Hypothyroidism after sunitinib treatment for patients with gastrointestinal stromal tumors. *Ann Intern Med*, 2006; **145**: 660–4.
23. Illouz F, Braun D, Briet C, Schweizer U, Rodien P. Endocrine side-effects of anti-cancer drugs: thyroid effects of tyrosine kinase inhibitors. *Eur J Endocrinol*, 2014; **171**: R91–9.
24. Chalan P, Di Dalmazi G, Pani F, De Remigis A, Corsello A, Caturegli P. Thyroid dysfunctions secondary to cancer immunotherapy. *J Endocrinol Invest*, 2018; **41**: 625–38.
25. Montanelli L, Benvenga S, Hegedus L, Vitti P, Latrofa F, Duntas LH. Drugs and other substances interfering with thyroid function. In: Vitti P, Hegedus L, eds. *Thyroid Diseases: Endocrinology*. Cham: Springer, 2018: 733–61.
26. Delange F, Camus M, Ermans AM. Circulating thyroid hormones in endemic goiter. *J Clin Endocrinol Metab*, 1972; **34**: 891–5.
27. Utiger RD. Maternal hypothyroidism and fetal development. *N Engl J Med*, 1999; **341**: 601–2.
28. Ikomi C, Cole CR, Vale E, Golekoh M, Khoury JC, Jones NY. Hypothyroidism and iodine deficiency in children on chronic parenteral nutrition. *Pediatrics*, 2018; **141**: DOI: 10.1542/peds.2017-3046.
29. Duntas LH. Chemical contamination and the thyroid. *Endocrine*, 2015; **48**: 53–64.
30. Delvecchio M, Cavallo L. Growth and endocrine function in thalassemia major in childhood and adolescence. *J Endocrinol Invest*, 2010; **33**: 61–8.
31. Lain S, Trumpff C, Grosse SD, Olivieri A, Van Vliet G. Are lower TSH cutoffs in neonatal screening for congenital hypothyroidism warranted? *Eur J Endocrinol*, 2017; **177**: D1–12.
32. Ford G, LaFranchi SH. Screening for congenital hypothyroidism: a worldwide view of strategies. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 175–87.
33. Burrow GN, Fisher DA, Larsen PR. Maternal and fetal thyroid function. *N Engl J Med*, 1994; **331**: 1072–8.
34. Montanelli L, Pinchera A, Santini F, Cavaliere R, Vitti P, Chiovato L. Transient congenital hypothyroidism: physiopathology and clinica. *Ann Ist Super Sanita*, 1998; **34**: 321–9.
35. van der Gaag RD, Drexhage HA, Dussault JH. Role of maternal immunoglobulins blocking TSH-induced thyroid growth in sporadic forms of congenital hypothyroidism. *Lancet*, 1985; **i**: 246–50.
36. Evans C, Jordan NJ, Owens G, Bradley D, Ludgate M, John R. Potent thyrotrophin receptor-blocking antibodies: a cause of transient congenital hypothyroidism and delayed thyroid development. *Eur J Endocrinol*, 2004; **150**: 265–8.
37. Persani L. Clinical review: central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*, 2012; **97**: 3068–78.
38. Lupi I, Cosottini M, Caturegli P, *et al*. Diabetes insipidus is an unfavorable prognostic factor for response to glucocorticoids in patients with autoimmune hypophysitis. *Eur J Endocrinol*, 2017; **177**: 127–35.
39. Phan GQ, Yang JC, Sherry RM, *et al*. Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc Natl Acad Sci U S A*, 2003; **100**: 8372–7.
40. Yamada M, Mori M. Mechanisms related to the pathophysiology and management of central hypothyroidism. *Nat Clin Pract Endocrinol Metab*, 2008; **4**: 683–94.
41. Persani L, Bonomi M. The multiple genetic causes of central hypothyroidism. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 255–63.
42. Wiersinga WM. Guidance in subclinical hyperthyroidism and subclinical hypothyroidism: are we making progress? *Eur Thyroid J*, 2015; **4**: 143–8.
43. Persani L, Calebiro D, Cordella D, *et al*. Genetics and phenomics of hypothyroidism due to TSH resistance. *Mol Cell Endocrinol*, 2010; **322**: 72–82.
44. Santini F, Marzullo P, Rotondi M, *et al*. Mechanisms in endocrinology: the crosstalk between thyroid gland and adipose tissue: signal integration in health and disease. *Eur J Endocrinol*, 2014; **171**: R137–52.
45. Persani L. Hypothalamic thyrotropin-releasing hormone and thyrotropin biological activity. *Thyroid*, 1998; **8**: 941–6.
46. Ceccarini G, Santini F, Vitti P. Tests of thyroid function. In: Vitti P, Hegedus L, eds. *Thyroid Diseases. Endocrinology*. Springer, Cham, 2017: 1–23.
47. Thienpont LM, Van Uytfanghe K, Poppe K, Velkeniers B. Determination of free thyroid hormones. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 689–700.
48. Henze M, Brown SJ, Hadlow NC, Walsh JP. Rationalizing thyroid function testing: which TSH cutoffs are optimal for testing free T4? *J Clin Endocrinol Metab*, 2017; **102**: 4235–41.
49. Faglia G. The clinical impact of the thyrotropin-releasing hormone test. *Thyroid*, 1998; **8**: 903–8.
50. Bartalena L, Martino E, Falcone M, *et al*. Evaluation of the nocturnal serum thyrotropin (TSH) surge, as assessed by TSH ultrasensitive assay, in patients receiving long term L-thyroxine suppression therapy and in patients with various thyroid disorders. *J Clin Endocrinol Metab*, 1987; **65**: 1265–71.
51. McLachlan SM, Rapoport B. Thyroid autoantibodies display both 'original antigenic sin' and epitope spreading. *Front Immunol*, 2017; **8**: 1845.
52. Rotondi M, de Martinis L, Coperchini F, *et al*. Serum negative autoimmune thyroiditis displays a milder clinical picture compared with classic Hashimoto's thyroiditis. *Eur J Endocrinol*, 2014; **171**: 31–6.
53. Kahaly GJ, Frommer L. Polyglandular autoimmune syndromes. *J Endocrinol Invest*, 2018; **41**: 91–8.
54. Pietras SM, Safer JD. Diagnostic confusion attributable to spurious elevation of both total thyroid hormone and thyroid hormone uptake measurements in the setting of autoantibodies: case report and review of related literature. *Endocr Pract*, 2008; **14**: 738–42.
55. Ma ZF, Skeaff SA. Thyroglobulin as a biomarker of iodine deficiency: a review. *Thyroid*, 2014; **24**: 1195–209.
56. Ittermann T, Johner S, Below H, *et al*. Interlaboratory variability of urinary iodine measurements. *Clin Chem Lab Med*, 2018; **56**: 441–7
57. Marcocci C, Vitti P, Cetani F, Catalano F, Concetti R, Pinchera A. Thyroid ultrasonography helps to identify patients with diffuse lymphocytic thyroiditis who are prone to develop hypothyroidism. *J Clin Endocrinol Metab*, 1991; **72**: 209–13.
58. Wiersinga WM, Touber JL, Trip MD, van Royen EA. Uninhibited thyroidal uptake of radioiodine despite iodine excess in amiodarone-induced hypothyroidism. *J Clin Endocrinol Metab*, 1986; **63**: 485–91.

3.4.3 Myxoedema Coma

Leonard Wartofsky, Dorina Ylli, and Joanna Klubo-Gwiezdzinska

Introduction

Myxoedema coma is the extreme expression of severe hypothyroidism and fortunately is quite rare. The first reported case appears to have been in 1879 by Ord from St Thomas's Hospital, London. Two other patients who died in a hypothyroid coma were reported in 1888 in the proceedings of the Clinical Society of London [1]. The next cases in the literature appeared in 1953 [2, 3] and some 500 cases have since been reported. Epidemiological data indicate an incidence rate of 0.22–1.08/1 000 000 per year [4, 5]. The most common presentation of the syndrome is in hospitalized elderly women with long-standing hypothyroidism, with 80% of cases occurring in women over 60 years of age. However, myxoedema coma occurs in younger patients as well, with 36 documented cases occurring during pregnancy [6, 7] and a few cases occurring in the paediatric age group [8–11]. In spite of early diagnosis and treatment, the mortality rate may be as high as 30–60% [5].

Clinical Presentation and Precipitating Events

Patients with myxoedema coma generally present in the winter months, suggesting that external cold may be an aggravating factor. The precipitating events vary from the most frequent including pulmonary infections, congestive heart failure, and cerebrovascular accidents (**Box 3.4.3.1**) to the more unusual such as in subclinical hypothyroidism or consumption of large amounts of raw bok choy [12, 13]. The comatose and hypoventilating patient is also at risk for pulmonary infection or aspiration pneumonia as a secondary event. Similarly, other abnormalities frequently accompanying myxoedema coma, such as hypoglycaemia, hypercalcaemia, hyponatraemia, hypercapnia, and hypoxaemia, may be either precipitating factors or secondary consequences. In hospitalized patients, drugs such as anaesthetics, narcotics, sedatives, antidepressants, and tranquillizers may depress respiratory drive and thereby either cause or compound the deterioration of the hypothyroid patient into coma. Also, drugs that could impair thyroid function such as amiodarone, lithium and tyrosine kinase inhibitors might precipitate myxoedema coma [14, 15].

Box 3.4.3.1 Myxoedema coma: precipitating factors

- Cerebrovascular accidents
- Drugs
- Anaesthetics
  - Sedatives
  - Tranquillizers
  - Narcotics
  - Amiodarone
  - Lithium carbonate
  - Tyrosine kinase inhibitors
- Hypothermia
- Congestive heart failure
- Infections
- Trauma
- Gastrointestinal bleeding
- Metabolic disturbances compounding obtundation
  - Acidosis
  - Hypoglycaemia
  - Hyponatraemia
  - Hypercapnia

Hypothermia (often profound to 80 °F (26.7 °C)) and unconsciousness constitute two of the cardinal features of myxoedema coma. The syndrome will typically present in a patient who develops an infection or other systemic disease superimposed upon previously undiagnosed hypothyroidism. Sometimes a history of antecedent thyroid disease, thyroidectomy, treatment with radioactive iodine, or thyroxine (T_4) replacement therapy that was discontinued for no apparent reason can be elicited. Other clues to the presence of underlying thyroid disease may be seen on examination of the neck, such as a surgical thyroidectomy scar, goitre, or even the absence of palpable thyroid tissue as may occur in chronic Hashimoto's thyroiditis. A pituitary or hypothalamic basis for hypothyroidism is encountered in less than 10–15% of patients. In two large series [16, 17] that identified 12–14 patients with myxoedema coma, the findings on presentation included hypoxaemia in 36–80%, hypotension 50%, hypercapnia in 36–54%, bradycardia in 36%, and hypothermia with a temperature below 94°F (34.4°C) in 50–88%. Fifty per cent (50%) of patients died despite treatment with thyroid hormone. A dreaded aspect of the usual clinical course is progression into respiratory failure and CO_2 retention which is heralded by hypoventilation with lethargy progressing to stupor and then coma. Because of the delayed metabolism of drugs in

hypothyroidism, the deterioration may be hastened by the use of sedative, hypnotics, or narcotics.

Cardiovascular Manifestations

In myxoedema coma, typical findings of hypothyroid heart disease may include bradycardia, decreased quality and intensity of the heart sounds, enlarged cardiac silhouette, and minor electrocardiographic (ECG) abnormalities such as varying degrees of block, low voltage, flattened or inverted T waves, and prolonged Q–T interval which can result in *torsades de pointe* ventricular tachycardia [18]. Myocardial infarction should be ruled out by the usual diagnostic procedures. The lactate dehydrogenase isoenzyme pattern in severe hypothyroidism may mimic that of myocardial infarction [19], and creatine kinase levels also are elevated [20, 21]. Moreover, aggressive or injudicious triiodothyronine (T_3) or T_4 replacement may increase the risk of myocardial infarction (see next). The enlarged cardiac silhouette may be due, in part, to ventricular dilatation or a pericardial effusion which can be confirmed by echocardiography. This fluid is rich in mucopolysaccharide and tends to accumulate slowly over time, only rarely causing cardiac tamponade.

Cardiac contractility is impaired, leading to reduced stroke volume and cardiac output, but congestive heart failure is rare. T_4 replacement therapy will slowly reverse the abnormalities in left ventricular function; although the pericardial effusion may also gradually diminish, reduced cardiac output with hypotension secondary to the effusion must be borne in mind. Patients should be admitted to an intensive care unit because of the propensity for shock and potentially fatal arrhythmias. Hypotension may occur in spite of increases in total body water and extracellular fluid volume because of reduction in intravascular volume. Although blood pressure may be normalized with T_4 replacement, severe hypotension or shock may supervene acutely before the T_4 effect is seen, necessitating the use of pressor drugs.

Respiratory System

The mechanism for hypoventilation in profound myxoedema is a combination of a depressed hypoxic respiratory drive and a depressed ventilatory response to hypercapnia [22]. CO_2 narcosis results from the reduction in alveolar ventilation with the hypoventilation compounded by impairment in respiratory muscle function ultimately leading to coma. The central factor in the pathophysiology of coma appears to be a depressed ventilatory response to CO_2 [23–25]. When present, obesity may impair the bellows action of the chest. Improvement in the response to CO_2 after T_4 therapy has been seen in some [23, 25, 26] but not all [22] studies. Irrespective of the underlying pathophysiology, the mechanical function of the chest in myxoedema coma usually is reduced sufficiently to require mechanically assisted ventilation. Tidal volume may be reduced by other factors such as pleural effusion or ascites. Upper airway partial obstruction may also play a role, caused by oedema or swelling of the tongue, or laryngeal obstruction due to marked oedema of the vocal cords. Hypothyroid patients may be predisposed to increased airway hyper-responsiveness and chronic inflammation [27]. Even with appropriate and adequate therapy, the complexity of the pathophysiology of respiratory failure means that ultimate recovery may be prolonged.

Gastrointestinal Manifestations

The gastrointestinal tract in myxoedema may be marked by mucopolysaccharide infiltration and oedema of the muscularis and neuropathic changes leading to impaired peristalsis, obstipation, and potential paralytic ileus [28]. Given the risks of anaesthesia in the profoundly hypothyroid patient, surgical intervention can be temporized for apparent obstruction by conservative management with decompression until the therapeutic response to thyroid hormone might occur. Initially, parenteral administration of T_4 or triiodothyronine (T_3) may be preferable because absorption of oral medications could be impaired due to the gastric atony often present in myxoedema coma. Ascites has been documented in 51 cases [29] and gastrointestinal bleeding can occur secondary to a coagulopathy.

Renal and Electrolyte Manifestations

Alterations in mineral metabolism and renal clearance in severe hypothyroidism may include decreases in plasma volume, serum sodium, and osmolality, glomerular filtration rate, and renal plasma flow, and increases in total body water, urine sodium, and urine osmolality. Atony of the urinary bladder with retention of large residual urine volumes is commonly seen. High creatine kinase levels are typical of hypothyroidism, but unusually high values may be a clue to underlying rhabdomyolysis. Increased serum antidiuretic hormone levels [30] and impaired water diuresis caused by reduced delivery of water to the distal nephron [31] are likely to account for the hyponatraemia. Depending upon its duration and severity, hyponatraemia will add to altered mental status, and when severe may be largely responsible for precipitating the comatose state. T_4 treatment promotes water diuresis resulting in an increase in serum sodium and a decrease in oedema and total body water.

Neuropsychiatric Manifestations

Although coma is the predominant clinical presentation in myxoedema coma, a history of disorientation, depression, paranoia, or hallucinations ('myxoedema madness') may often be elicited. Other findings present either just before entering the comatose state or early during recovery include cerebellar signs, such as poorly coordinated purposeful movements of the hands and feet, ataxia, adiadochokinesia, poor memory and recall, or even frank amnesia. Abnormal findings on electroencephalography are few and include low amplitude and a decreased rate of α -wave activity. Status epilepticus has been described [32] and up to 25% of patients with myxoedema coma may experience minor to major seizures possibly related to hyponatraemia, hypoglycaemia, or hypoxaemia due to reduced cerebrovascular perfusion from low cardiac output and atherosclerotic vessels in elderly patients. T_4 treatment will generally lead to improved perfusion.

Haematological Manifestations

A microcytic anaemia may be seen secondary to gastrointestinal haemorrhage, or a macrocytic anaemia due to vitamin B_{12}deficiency which may also worsen the neurological state. Granulocytopaenia with a decreased cell-mediated immunological response may contribute to a higher risk of severe infection. In contrast to the tendency to thrombosis seen in mild hypothyroidism, severe hypothyroidism is associated with a higher risk of bleeding due to coagulopathy related to an acquired von Willebrand's syndrome (type 1) and decreases in factors V, VII, VIII, IX, and X [33]. The von Willebrand syndrome is reversible with T_4 therapy [34]. Another cause of bleeding may be disseminated intravascular coagulation associated with sepsis.

Hypothermia

The first clinical clue to the diagnosis of myxoedema coma may be hypothermia which occurs in approximately 75% of patients and may be dramatic (below 80°F (26.7°C)), with temperatures of less than 90°F (32.2°C) being associated with the worst prognosis. Because patients with myxoedema and infection may not mount a febrile response, a diagnosis of profound hypothyroidism should be entertained in any unconscious patient with a known infection but no fever. In view of the latter and because undiscovered infection might lead inexorably to vascular collapse and death, some authors have advocated the routine use of antibiotics in patients with myxoedema coma. Underlying hypoglycaemia may further compound the decrement in body temperature. With T_4 therapy, the hypothermia gradually improves in parallel with the fall in serum thyroid-stimulating hormone (TSH) and increments in serum T_4 and T_3 levels.

Diagnosis

The typical patient presenting with myxoedema coma is a woman in the later decades of her life who may have a history of thyroid disease and who is admitted to hospital during the winter months, possibly with pneumonitis. Physical findings could include bradycardia, macroglossia, hoarseness, delayed reflexes, dry skin, general cachexia, hypotension, hypoventilation, and hypothermia, commonly without shivering. Even though specific diagnostic criteria are not available a diagnostic scoring system has been proposed. Based upon the presence of thermoregulatory and central nervous system dysfunction, gastrointestinal findings, precipitating events, cardiovascular dysfunction, and metabolic alterations, the scoring system has reached a sensitivity of 100% and specificity of 80% (Table 3.4.3.1) [17]. According to the score achieved, the patient would be classified as either highly suggestive, suggestive, or unlikely to have myxoedema coma. Another objective screening tool was proposed based on heart rate, temperature, Glasgow coma scale, TSH, free thyroxine, and precipitating events which was associated with a sensitivity and specificity of approximately 80% [35]. Despite the importance that both scoring systems could have in allowing

Table 3.4.3.1 Diagnostic scoring system for myxoedema coma

Thermoregulatory dysfunction (temperature, °C)		**Cardiovascular dysfunction**	
>35	0	Bradycardia	
32–35	10	Absent	0
<32	20	50–59	10
Central nervous system effects		40–49	20
Absent	0	<40	30
Somnolent/lethargic	10	Other EKG changes[a]	10
Obtunded	15	Pericardial/pleural effusions	10
Stupor	20	Pulmonary oedema	15
Coma/seizures	30	Cardiomegaly	15
Gastrointestinal findings		Hypotension	20
Anorexia/abdominal pain/constipation	5	**Metabolic disturbances**	
Decreased intestinal motility	15	Hyponatraemia	10
Paralytic ileus	20	Hypoglycaemia	10
Precipitating event		Hypoxemia	10
Absent	0	Hypercarbia	10
Present	10	Decrease in GFR	10

Abbreviations: EKG, electrocardiogram; GFR, glomerular filtration rate.

A score of 60 or higher is highly suggestive/diagnostic of myxoedema coma; a score of 25 to 59 is suggestive of risk for myxoedema coma, and a score below 25 is unlikely to indicate myxoedema coma.

[a] Other EKG changes: QT prolongation, or low voltage complexes, or bundle branch blocks, or non-specific ST-T changes, or heart blocks.

Reproduced with permission from Popoveniuc G, Chandra T, Sud A, Sharma M, Blackman MR, Burman KD, Mete M, Desale S, Wartofsky L. A diagnostic scoring system for myxedema coma. *Endocr Pract.* 2014 Aug;20(8):808–17.

early diagnosis and initiation of treatment, the limited number of patients in these studies requires additional validation.

Laboratory evaluation may indicate anaemia, hyponatraemia, hypercholesterolaemia, and increased serum lactate dehydrogenase and creatine kinase. On lumbar puncture there is increased pressure and the cerebrospinal fluid will have a high protein content. The electrocardiogram and chest radiograph may demonstrate the characteristic findings described earlier. If hypothyroidism is suspected in a comatose patient, blood should be obtained for thyroid function testing, but treatment should not be delayed awaiting laboratory confirmation of the diagnosis. On the other hand, a correct diagnosis is particularly important because the unnecessary administration of large doses of T_4 or T_3 to an elderly euthyroid patient could induce a fatal arrhythmia or coronary event. In addition to routine thyroid function tests, ancillary studies should be performed to determine whether CO_2 retention, hypoxia, hyponatraemia, or infection are present. Indeed, in many patients the clinical features may be so notable as to render the measurement of thyroid function tests necessary only for confirmation of the diagnosis. The urgency of the diagnosis should be stressed to the laboratory, which often can perform a serum T_4 and TSH determination in 3–4 h. Although an elevated serum TSH concentration is the most important laboratory evidence of the diagnosis, the presence of severe complicating systemic illness or treatment with drugs such as dopamine, dobutamine, or corticosteroids may serve to reduce the elevation in TSH levels [36, 37]. Furthermore, an association between the use of catecholamines and higher mortality has been reported [5]. There may also be a pituitary cause for the hypothyroidism, in which case an increased TSH would not be found. Until the presence of pituitary disease is ruled out, corticosteroid therapy is recommended in addition to T_4.

Treatment

Myxoedema coma is a true medical emergency, and treatment must be instituted in a critical care setting with modern electronic monitoring equipment as soon as the diagnosis is made in view of the extremely high mortality anticipated in these patients when treatment is delayed. As outlined here, a multifaceted approach is required because of the multiple metabolic derangements derived from, or affecting, several organ systems which may be contributing to the comatose state.

Ventilatory Support

The patient's comatose state is perpetuated by hypoventilation with CO_2 retention and respiratory acidosis. Appropriate diagnostic and therapeutic measures must be instituted for any suspicious infiltrate seen on chest radiographs. The high mortality rate is often related to inexorable respiratory failure, and hence maintenance of an adequate airway and prevention of hypoxaemia is the single most important supportive measure required to avoid a disastrous outcome. Mechanical ventilation is usually required during the first 36–48 h, particularly if the hypoventilation is related in part to drug-related respiratory depression. Although the patient may become alert by the second or third day of treatment, it may be necessary to continue assisted ventilation for as long as 2–3 weeks in some patients.

Intubation may be necessary initially or with worsening of hypoxaemia or hypercarbia, and arterial blood gases need to be monitored regularly until the patient is fully recovered. The hypercapnia may be rapidly relieved with mechanical ventilation, but the hypoxia may tend to persist possibly due to shunting in non-aerated lung areas [38]. Moreover, the physician should guard against extubating the patient prematurely; some case reports have cited the danger of relapse, and it should not be attempted until the patient is fully conscious.

Hyponatraemia

Total body sodium is believed to be normal to increased, but it is the impaired excretion of water that causes hyponatraemia. Low serum sodium may cause a semicomatose state or seizures even in euthyroid patients, and the very severe hyponatraemia (105–120 mmol/L) in profound myxoedema is likely to contribute substantially to the coma in these patients. With such severe hyponatraemia, it may be appropriate to administer a small amount of hypertonic saline (50–100 ml 3% sodium chloride), enough to increase sodium concentration by about 2 mmol/L early in the course of treatment, and this can be followed by an intravenous bolus dose of 40–120 mg furosemide to promote a water diuresis [39]. A small quick increase in the serum sodium concentration (2–4 mmol/L) is effective in acute hyponatraemia because even a slight reduction in brain swelling results in a substantial decrease in intracerebral pressure [40]. On the other hand, too rapid correction of hyponatraemia can cause a very dangerous complication, the osmotic demyelination syndrome. In patients with chronic hyponatraemia this complication is avoided by limiting the sodium correction to less than 10–12 mmol/L in 24 h and to less than 18 mmol/L in 48 h.

After achieving a sodium level of more than 120 mmol/L, no further hypertonic saline infusion should be required, and restriction of fluids may be all that is necessary to correct hyponatraemia, especially if it is mild (120–130 mmol/L). Because of the likelihood of decreased cardiac reserve, therapy with saline or other intravenous fluids must be approached cautiously. If hypoglycaemia is present, dextrose in 0.5 N sodium chloride may be used to correct the low blood glucose. With regard to fluid or saline therapy, careful monitoring of volume status based on clinical parameters and central venous pressure measurements is essential in patients with significant cardiovascular decompensation.

New vasopressin antagonists named vaptans (conivaptan and tolvaptan), have been approved by the Food and Drug Administration (FDA) in the United States for the treatment of patients with euvolaemic and hypervolaemic hyponatraemia. These treatments could be attempted in this clinical setting in view of the high vasopressin levels observed in myxoedema coma. Conivaptan current dosing recommendations are a 20-mg loading dose to be infused over 30 min followed by a 20 mg/day continuous infusion for up to 4 days. Tolvaptan is administered orally in a starting dose of 15 mg the first day followed by titration up to 30 mg and 60 mg at 24 hours if necessary. During the active phase of correction with tolvaptan, fluid restriction is not recommended in order to reduce the risk of overcorrection. With both therapies, serum [Na+]

concentration should be measured frequently during the active phase of correction of the hyponatraemia at a minimum of every 6–8 hours. No data are available on the use of vaptans in severe hyponatraemia (<115 mmol/L) in hypothyroid patients, or whether sole therapy with vaptans without hypertonic saline would be effective [40, 41].

Hypothermia

Restoration of body temperature to normal will require administration of T_4 or T_3. Blankets or increasing room temperature can be used to keep the patient warm until the thyroid hormone effect is achieved, but caution must be exercised in the use of more vigorous electric warming blankets. Too aggressive warming may cause peripheral vasodilatation, a precipitous fall in peripheral vascular resistance with increased peripheral blood flow, and increased oxygen consumption, which may then lead to hypotension or shock.

Hypotension

Hypotension should also be corrected after treatment with T_4; this may take several days, and the hypotensive patient may require additional therapy. Fluids may be administered cautiously as 5–10% glucose in 0.5 N sodium chloride initially, or as isotonic normal saline if hyponatraemia is present. It is wise to administer hydrocortisone (100 mg intravenously every 8 h) until the hypotension is corrected. Pressors are only very rarely required, and the possibility of an adverse cardiac event needs to be kept in mind, especially in patients with suspected underlying ischaemic heart disease. An agent such as dopamine might be employed to maintain coronary blood flow, but patients should be weaned off the pressor as soon as possible. The physician must weigh the risk of a pressor-induced ischaemic event against the known high mortality of poorly managed hypotension in myxoedema coma.

Corticosteroids

A rising urea nitrogen, hypotension, hypothermia, hypoglycaemia, hyponatraemia, and hyperkalaemia may signal the coexistence of adrenal insufficiency. Indeed, decreased adrenal reserve has been found in 5–10% of patients on the basis of either hypopituitarism or primary adrenal failure accompanying Hashimoto's disease (Schmidt's syndrome). Otherwise, plasma total and free cortisol levels and the adrenal response to adrenocorticotropic hormone (ACTH) infusion should be normal in hypothyroidism or myxoedema coma. However, ACTH reserve or the ACTH response to stress may be impaired in myxoedema coma. There should be no reluctance to administer short-term corticosteroids until the patient is stable and the integrity of the pituitary–adrenal axis can be determined. On theoretical grounds, one should also administer corticosteroids when first instituting thyroid hormone therapy, in view of the potential risk of precipitating acute adrenal insufficiency due to the accelerated metabolism of cortisol that follows T_4 therapy. The typical dosage of hydrocortisone is 50–100 mg every 6–8 h during the first 7–10 days with tapering of the dosage thereafter based upon clinical response and any plans for further diagnostic evaluation.

Thyroid Hormone Therapy

One of the most controversial aspects of the management of myxoedema coma is which thyroid hormone medication to give and how to give it (dose, frequency, and route of administration). Because of the relative rarity of this condition, the paucity of reported treatment results, and the difficulties inherent in performing a controlled investigation, the optimum treatment remains uncertain, and several approaches will be discussed. Some of the differences of opinion relate to whether to administer T_4 and rely on the patient to convert it to the more active T_3, or to give T_3 itself. One must balance the need for quickly attaining physiologically effective thyroid hormone levels against the risk of precipitating a fatal tachyarrhythmia or myocardial infarction. T_4 provides a steady smooth onset of action with a lower risk of adverse effects.

Parenteral preparations of either T_4 or T_3 are available for intravenous administration. Although oral forms of either T_3 or T_4 can be given by nasogastric tube in the comatose patient, this route is fraught with risks of aspiration and uncertain absorption, particularly in the presence of gastric atony or ileus.

Parenteral preparations of T_4 may be available in ampoules of 100 and 500 μg. A loading dose of 200–400 μg of levothyroxine may be given intravenously. After this initial 'loading' dose, a maintenance dose of 1.6 μg/kg body weight, reduced to 75% if being intravenously administered, can be given thereafter. This method may be attended by increases in serum T_4 to within the normal range within 24 h and by significant decrements in serum TSH. After clinical improvement the patient may be switched to oral therapy. Larger doses of T_4 probably have no advantage and may, in fact, be more dangerous [42]. Due to its conversion from T_4, a progressive increase in serum T_3 is seen after 300–600 μg doses of T_4, as has been described by Ridgway *et al.* [42].

The approach to therapy employing an initial large intravenous bolus dose of T_4 followed by maintenance therapy has been considered optimal [41, 42], but other evidence suggests improved outcomes with lower doses of thyroid hormone [43]. This was also indicated in a prospective trial in which patients were randomized to receive either a 500-μg loading dose of intravenous T_4 followed by a 100-μg daily maintenance dose, or only the maintenance dose [44, 45]. The overall mortality rate was 36.4% with a lower mortality rate in the high-dose group (17%) versus the low dose group (60%). Although suggestive, the difference was not statistically significant. Factors associated with a worse outcome included a decreased level of consciousness, lower Glasgow coma score, and increased severity of illness on entry as determined by an APACHE II score of more than 20.

T_4 treatment has been generally considered effective, but there is one important drawback to total reliance on T_3 generation from T_4. The rate of conversion of T_4 to T_3 is reduced in many systemic illnesses (the euthyroid sick or low T_3 syndrome) [37] and hence T_3 generation may be reduced in myxoedema coma as a consequence of any associated illness [39]. Theoretically then, one might administer

T_4, see increases in serum T_4 levels confirming adequate absorption, but fail to witness any significant fall in TSH or dramatic clinical improvement. As a consequence, small supplements of T_3 should be given along with T_4 during the initial few days of treatment, especially if obvious associated illness is present. Irrespective of the type of treatment selected, all patients should have continuous ECG monitoring with reduction in thyroid hormone dosage should arrhythmias or ischaemic changes be detected.

T_3 is available for intravenous use (Triostat) in 1 ml vials containing 10 μg/ml. When therapy is approached with T_3 alone, it may be given as a 10–20 μg bolus followed by 5–10 μg every 4 h for the first 24 h, dropping to 5–10 μg every 6 h for days 2–3, by which time oral administration should be feasible. T_3 has a much quicker onset of action than T_4 and increases in body temperature and oxygen consumption may occur 2–3 h after intravenous T_3, compared to 8–14 h after intravenous T_4. A patient with profound secondary myxoedema believed due to postpartum pituitary necrosis has been reported who presented with cardiogenic shock which responded to T_3 but not T_4 therapy [44]. Because of the high mortality rate in myxoedema coma, advocates for T_3 therapy argue that the more rapid onset of action could make the difference between life and death. But the benefits of the more rapid onset of action need to be weighed against the greater risk of complications. As a consequence, it is difficult to justify the high risk/benefit ratio of a regimen that uses rapid replacement with relatively large doses of intravenous T_3 alone. Such treatment would be marked by large and unpredictable fluctuations in serum T_3 levels, and high serum T_3 levels during treatment with thyroid hormone have been associated with fatal outcomes [45].

A more conservative but seemingly rational course of management is to provide combined therapy with both T_4 and T_3. Rather than administer 300–500 μg T_4 intravenously initially, a dose of 4 μg/kg lean body weight (or about 200–300 μg) is given, and an additional 100 μg is given 24 hours later. By the third day, the dose is reduced to a daily maintenance dose of 50 μg, which can be given by mouth as soon as the patient is conscious. Simultaneously with the initial dose of T_4, a bolus of 5–20 μg T_3 is given and intravenous T_3 is continued at a dosage of 2.5–10 μg every 8–12 hours until the patient is conscious and taking maintenance T_4 [46]. Sensitivity to thyroid hormone in terms of cardiac risk varies, depending on age, cardiac medications, and the presence of underlying hypoxaemia, coronary artery disease, congestive failure, and electrolyte imbalance. Clinical improvement has been seen with even a single dose of only 2.5 μg T_3 [47]. It is wise to monitor the patient for any untoward effects of therapy before administering each dose of thyroid hormone.

Myxoedema Coma and Emergent Surgery

Clearly, given their fragile clinical state, non-emergent surgery should be deferred in a patient with myxoedema coma. However, in the patient with myxoedema coma requiring emergent surgery, the same general management principles prevail [48] with particular attention to careful monitoring of intraoperative and postoperative respiratory and cardiovascular status. Postoperatively, close monitoring for maintenance of the airway is essential.

General Supportive Measures

In addition to the specific therapies outlined, other treatments will be indicated as in the management of any other elderly patient with multisystemic problems. This might include the treatment of underlying problems such as infectious processes, congestive heart failure, diabetes, or hypertension. The dosage of specific medications (e.g. digoxin for congestive heart failure) may need to be modified based on their altered distribution and slowed metabolism in myxoedema. Even with this vigorous therapy, the prognosis for myxoedema coma remains grim, and patients with severe hypothermia and hypotension seem to do the worst. Several prognostic factors may be associated with a fatal outcome [43, 45, 49, 50] and include: older age, persistent hypothermia or bradycardia, lower degree of consciousness by Glasgow coma scale, multiorgan impairment indicated by an APACHE II score of more than 20, or SOFA score ≥6, and use of catecholamines [5]. The most common causes of death are respiratory failure, sepsis, and gastrointestinal bleeding. Early diagnosis and prompt treatment, with meticulous attention to the details of management during the first 48 hours, remain critical for the avoidance of a fatal outcome.

REFERENCES

1. Report of a committee of the Clinical Society of London to investigate the subject of myxedema. *Trans Clin Soc (Lond) (Suppl)*, 1888: **21**.
2. LeMarquand HS, Hausmann W, Hemstead EH. Myxedema as a cause of death. *BMJ (Clin Res)*, 1953; **1**: 704–6.
3. Summers VK. Myxedema coma. *BMJ (Clin Res)*, 1953; **2**: 336–8.
4. Rodríguez I, Fluiters E, Pérez-Méndez LF, Luna R, Páramo C, García-Mayor RV. Factors associated with mortality of patients with myxoedema coma: prospective study in 11 cases treated in a single institution. *J Endocrinol*, 2004; **180**: 347–50.
5. Ono Y, Ono S, Yasunaga H, Matsui H, Fushimi K, Tanaka Y. Clinical characteristics and outcomes of myxedema coma: analysis of a national inpatient database in Japan. *J Epidemiol*, 2017; **27**: 117–22.
6. Blignault EJ. Advanced pregnancy in severely myxedematous patient. A case report and review of the literature. *S Afr Med J*, 1980; **57**: 1050–1.
7. Patel S, Robinson S, Bidgood RJ, Edmonds CJ. A pre-eclamptic-like syndrome associated with hypothyroidism during pregnancy. *Q J Med*, 1991; **79**: 435–41.
8. Schutt-Aine JC. Hypothyroid myxedema and hyponatremia in an eight year—old child: a case report. *J Natl Med Assoc*, 1980; **72**: 705–8.
9. Thompson MD, Henry RK. Myxedema coma secondary to central hypothyroidism: a rare but real cause of altered mental status in pediatrics. *Horm Res Paediatr*, 2016; **87**: 350–3.
10. Root JM, Vargas M, Garibaldi LR, *et al*. Pediatric patient with altered mental status and hypoxemia: case report. *Pediatr Emerg Care*, 2017; **33**: 486–8.
11. Zhu Y, Qiu W, Deng M, Zhu X. Myxedema coma: a case report of pediatric emergency care. *Medicine (Baltimore)*, 2017; **96**: e6952.
12. Mallipedhi A, Vali H, Okosieme O. Myxedema coma in a patient with subclinical hypothyroidism. *Thyroid*, 2011; **21**: 87–9.

13. Chu M, Seltzer TF. Myxedema coma induced by ingestion of raw bok choy. *N Engl J Med*, 2010; **362**: 1945–6.
14. Lele AV, Clutter S, Price E, De Ruyter ML. Severe hypothyroidism presenting as myxedema coma in the postoperative period in a patient taking sunitinib: case report and review of literature. *J Clin Anesth*, 2013; **25**: 47–51.
15. Chen SY, Kao PC, Lin ZZ, Chiang WC, Fang CC. Sunitinib-induced myxedema coma. *Am J Emerg Med*, 2009; **27**: 370.e1–370.e3.
16. Reinhardt W, Mann K. Incidence, clinical picture and treatment of hypothyroid coma. Results of a survey. *Med Klin*, 1997; **92**: 521–4.
17. Popoveniuc G, Chandra T, Sud A, Sharma M, Blackman MR, Burman KD, Mete M, Desale S, Wartofsky L. A diagnostic scoring system for myxedema coma. *Endocr Pract*, 2014; **20**: 808–17.
18. Schenck JB, Rizvi AA, Lin T. Severe primary hypothyroidism manifesting with torsades de pointes. *Am J Med Sci*, 2006; **331**: 154–6.
19. Aber CP, Noble RL, Thomson GS, Jones EW. Serum lactic dehydrogenase isoenzymes in 'myxedema heart disease'. *Br Heart J*, 1966; **28**: 663–73.
20. Nee PA, Scane SC, Lavelle PH, Fellows IW, Hill PG. Hypothermic myxedema coma erroneously diagnosed as myocardial infarction because of increased creatine kinase MB. *Clin Chem*, 1987; **33**: 1083–4.
21. Hickman PE, Silvester W, Musk AA, McLellan GH, Harris A. Cardiac enzyme change in myxedema coma. *Clin Chem*, 1987; **33**: 622–4.
22. Zwillich CW, Pierson DJ, Hofeldt FD, Lufkin EG, Weil JV. Ventilatory control in myxedema and hypothyroidism. *N Engl J Med*, 1975; **292**: 662–5.
23. Massumi RA, Winnacker JL. Severe depression of the respiratory center in myxedema. *Am J Med*, 1964; **36**: 876–82.
24. Ladenson PW, Goldenheim PD, Ridgway EC. Prediction of reversal of blunted respiratory responsiveness in patients with hypothyroidism. *Am J Med*, 1988; **84**: 877–83.
25. Domm BB, Vassallo CL. Myxedema coma with respiratory failure. *Am Rev Respir Dis*, 1973; **107**: 842–5.
26. Wilson WR, Bedell GN. The pulmonary abnormalities in myxedema. *J Clin Invest*, 1960; **39**: 42–55.
27. Birring SS, Patel RB, Parker D, *et al*. Airway function and markers of airway inflammation in patients with treated hypothyroidism. *Thorax*, 2005; **60**: 249–53.
28. Garrahy A, Agha A. Dementia, cardiomyopathy and pseudo-obstruction in a 63-year-old female. *Eur J Intern Med*, 2018; **55**: e5–6.
29. Ji JS, Chae HS, Cho YS, *et al*. Myxedema ascites: case report and literature review. *J Korean Med Sci*, 2006; **21**: 761–4.
30. Skowsky WR, Kikuchi TA. The role of vasopressin in the impaired water excretion of myxedema. *Am J Med*, 1978; **64**: 613–21.
31. DeRubertis FR Jr, Michelis MF, Bloom ME, et al. Impaired water excretion in myxedema. *Am J Med*, 1971; **51**: 41–53.
32. Jansen HJ, Doebé SR, Louwerse ES, van der Linden JC, Netten PM. Status epilepticus caused by a myxoedema coma. *Neth J Med*, 2006; **64**: 202–5.
33. Manfredi E, van Zaane B, Gerdes VE, Brandjes DP, Squizzato A. Hypothyroidism and acquired von Willebrand's syndrome: a systematic review. *Haemophilia*, 2008; **14**: 423–33.
34. Michiels JJ, Schroyens W, Berneman Z, van der Planken M. Acquired von Willebrand syndrome type 1 in hypothyroidism: reversal after treatment with thyroxine. *Clin Appl Thromb Hemost*, 2001; **7**: 113–15.
35. Chiong YV, Bammerlin E, Mariash CN. Development of an objective tool for the diagnosis of myxedema coma. *Transl Res*, 2015; **166**: 233–43.
36. Hooper MJ. Diminished TSH secretion during acute non-thyroidal illness in untreated primary hypothyroidism. *Lancet*, 1976; **i**: 48–9.
37. Wartofsky L, Burman KD. Alterations in thyroid function in patients with systemic illness: the 'euthyroid sick syndrome'. *Endocr Rev*, 1982; **3**: 164–217.
38. Nicoloff JT. Thyroid storm and myxedema coma. *Med Clin North Am*, 1985; **69**: 1005–17.
39. Pereira VG, Haron ES, Lima-Neto N, Medeiros-Neto GA. Management of myxedema coma: report on three successfully treated cases with nasogastric or intravenous administration of triiodothyronine. *J Endocrinol Invest*, 1982; **5**: 331–4.
40. Verbalis JG, Goldsmith SR, Greenberg A, *et al*. Diagnosis, evaluation, and treatment of hyponatremia: expert panel recommendations. *Am J Med*, 2013; **126**(10 Suppl 1): S1–42.
41. Holvey DN, Goodner CJ, Nicoloff JT, Dowling JT. Treatment of myxedema coma with intravenous thyroxine. *Arch Intern Med*, 1964; **113**: 139–46.
42. Ridgway EC, McCammon JA, Benotti J, Maloof F. Acute metabolic responses in myxedema to large doses of intravenous L-thyroxine. *Ann Intern Med*, 1972; **77**: 549–55.
43. Yamamoto T, Fukuyama J, Fujiyoshi A. Factors associated with mortality of myxedema coma: report of eight cases and literature survey. *Thyroid*, 1999; **9**: 1167–74.
44. McKerrow SD, Osborn LA, Levy H, Eaton RP, Economou P. Myxedema-associated cardiogenic shock treated with triiodothyronine. *Ann Intern Med*, 1992; **117**: 1014–15.
45. Hylander B, Rosenqvist U. Treatment of myxoedema coma: factors associated with fatal outcome. *Acta Endocrinol (Copenh)*, 1985; **108**: 65–71.
46. Jonklaas J, Bianco AC, Bauer AJ, *et al*.; Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association Task Force on Thyroid Hormone Replacement. *Thyroid*, 2014; **24**: 1670–751.
47. McCulloch W, Price P, Hinds CJ, Wass JA. Effects of low dose oral triiodothyronine in myxoedema coma. *Intensive Care Med*, 1985; **11**: 259–62.
48. Mathes DD. Treatment of myxedema coma for emergency surgery. *Anesth Analg*, 1998; **86**: 450–1.
49. Jordan RM. Myxedema coma. Pathophysiology, therapy, and factors affecting prognosis. *Med Clin North Am*, 1995; **79**: 185–94.
50. Dutta P, Bhansali A, Masoodi SR, Bhadada S, Sharma N, Rajput R. Predictors of outcome in myxoedema coma: a study from a tertiary care centre. *Crit Care*, 2008; **12**: 1–8.

FURTHER READING

Fliers E, Wiersinga WM. Myxedema coma. *Rev Endocr Metab Disord*, 2003; **4**: 137–41.

Kwaku MP, Burman KD. Myxedema coma. *J Intensive Care Med*, 2007; **22**: 224–31.

Ringel MD. Management of hypothyroidism and hyperthyroidism in the intensive care unit. *Crit Care Clin*, 2001; **17**: 59–74.

Sanders V. Neurologic manifestations of myxedema. *N Engl J Med*, 1962; **266**: 547–52.

3.4.4 Subclinical Hypothyroidism

Bijay Vaidya and Chantal Daumerie

Introduction

Subclinical hypothyroidism is biochemically defined as a raised serum thyrotropin (TSH) level together with serum free thyroxine (fT_4) and free tri-iodothyronine (fT_3) levels within the population reference range [1]. It is also referred to as compensated hypothyroidism or mild thyroid failure. Although the term 'subclinical' suggests the absence of symptoms, some patients with subclinical hypothyroidism have vague non-specific symptoms of hypothyroidism. However, the diagnosis of subclinical hypothyroidism cannot be made based on clinical features alone and laboratory thyroid function tests are required for the diagnosis. The risk of long-term complications of subclinical hypothyroidism depends upon the degree of TSH elevation and the age of the patient, with younger adults showing an increased risk of cardiovascular and cerebrovascular morbidities. The paucity of well conducted large randomized controlled trials in subclinical hypothyroidism means its management remains controversial, and there is a wide variation in the routine clinical practice [2].

This chapter reviews the diagnosis, long-term complications, and management of subclinical hypothyroidism in non-pregnant adults.

How Common is Subclinical Hypothyroidism?

In population-based studies, the prevalence of subclinical hypothyroidism ranges from 4 to 15%, with higher prevalence associated with advancing age, female sex, and suboptimal iodine status of the population [3, 4]. It affects about 10% women in perimenopausal age, and more than 20% women aged 75 years or above [5]. The apparent variation in the prevalence rate of subclinical hypothyroidism in different studies reflects the differences in the demographic characteristics of the study populations and the upper limits of the reference range used for TSH measurements.

Causes of Subclinical Hypothyroidism

The commonest cause of subclinical hypothyroidism, like overt hypothyroidism, is autoimmune thyroiditis (Hashimoto's disease or autoimmune atrophic thyroiditis). Other causes include previous thyroid surgery, radioiodine treatment, drugs, inadequate dose of levothyroxine for overt hypothyroidism, thyroiditis (subacute, painless, or postpartum), and radiotherapy involving head and neck (Box 3.4.4.1). Rarely, it is associated with loss-of-function mutations in the TSH receptor gene [6]. Although severe iodine deficiency can cause both subclinical and overt hypothyroidism, chronic excess iodine intake, for example, following inadequately monitored universal salt iodization, is a recognized risk factor for subclinical hypothyroidism [7]. In addition, long-term occupational exposure to pesticides has also been shown to be associated with subclinical hypothyroidism [8]. In contrast, smoking is associated with a reduced risk of subclinical hypothyroidism [9].

Box 3.4.4.1 Causes of subclinical hypothyroidism

- Chronic autoimmune thyroiditis (Hashimoto's disease or autoimmune atrophic thyroiditis)
- Iatrogenic (partial thyroidectomy, radioiodine therapy, external radiotherapy involving head and neck)
- Thyroiditis (subacute, painless, or postpartum)
- Drugs (iodine, amiodarone, lithium, interferon-alpha, interleukin 2, tyrosine kinase inhibitors, checkpoint inhibitor immunotherapy)
- Suboptimal levothyroxine replacement for overt hypothyroidism
- Chronic excessive iodine intake
- Occupational exposure to pesticides
- Thyroid infiltration (sarcoidosis, lymphoma, haemochromatosis, amyloidosis)
- Loss-of-function mutations in the TSH receptor gene

Diagnosis of Subclinical Hypothyroidism

Symptoms and Signs

Few patients with subclinical hypothyroidism present with classical symptoms and signs of hypothyroidism. In the Colorado thyroid disease prevalence study, more patients with subclinical hypothyroidism than euthyroid individuals reported multiple symptoms, such as dry skin, poor memory, slow thinking, muscle weakness, constipation, and tiredness [5]. However, the association was weak with low sensitivities (3–28%) and positive predictive values (8–12%) for the individual symptoms, such that only a minority of patients with subclinical hypothyroidism had the symptoms while many of the individuals with these symptoms had normal thyroid function. Similarly, another case–control study found that patients with subclinical hypothyroidism and euthyroid controls have similar frequencies of symptoms of hypothyroidism, except tiredness which was more prevalent in patients with subclinical hypothyroidism [10].

Laboratory Investigations

The diagnosis of subclinical hypothyroidism is based on laboratory thyroid function tests: a serum TSH level above the laboratory reference range together with a normal serum fT_4 level is suggestive of the diagnosis. A serum TSH level of 4.5–5 mU/L is widely accepted as the upper limit of the reference range, although there is a considerable debate whether this should be set lower at 2.5 mU/L [11].

Several physiological and demographic factors affect serum TSH levels that may influence the interpretation of laboratory thyroid function tests. First, the presence of a diurnal variation in TSH secretion with peak levels at night and trough levels in the afternoon means that the timing of the blood test may affect the TSH result [12]. Second, TSH levels progressively rise with age, possibly reflecting an adaption of the hypothalamus-pituitary-thyroid axis to ageing, and therefore some experts advocate the use of age specific TSH reference ranges [13]. Third, ethnicity influences serum TSH; for example, the levels tend to be higher in Caucasians than in Blacks or Hispanics [4]. Finally, an individual's TSH levels over time fluctuate within a much narrower range than the population reference range [14]. This suggests that each individual has their own TSH set point, and any deviation from that set point may be abnormal for the individual even if the level is within the population reference range.

It is also important to distinguish subclinical hypothyroidism from other conditions associated with a raised serum TSH and normal fT_4 (Table 3.4.4.1). The presence of heterophile or human anti-animal antibodies can result in spuriously high serum TSH by interfering TSH assay [15]. Macro-TSH (a biologically inert large molecular complex of TSH and immunoglobulin G) is another cause of spurious elevation of serum TSH, and accounts for about 1.5% cases of apparent subclinical hypothyroidism [16]. Patients recovering from non-thyroidal illness and thyroiditis (subacute, painless, or postpartum) often show transient rise in TSH levels. Likewise, patients with untreated Addison's disease may have a raised TSH, which normalizes after glucocorticoid replacement. Patients on levothyroxine replacement for overt hypothyroidism frequently have raised TSH due to inadequate dose, non-concordance, malabsorption, and concomitant use of drugs impairing levothyroxine absorption (e.g. iron, calcium, and cholestyramine) or increasing levothyroxine clearance (e.g. phenytoin, phenobarbital, and rifampicin). Resistance to thyroid hormone due to mutations in the thyroid hormone receptor β gene is often associated with raised TSH with variable levels of fT_4. However, most patients with this condition will have fT_4 above or towards the upper end of the reference range; in contrast, fT_4 levels tend to be towards the lower half of the reference range in subclinical hypothyroidism. Furthermore, as resistance to thyroid hormone is an autosomal dominant condition, first degree relatives of the patient may show similar thyroid function test abnormalities.

In patients with subclinical hypothyroidism, the presence of thyroid peroxidase antibodies (TPO-Ab) not only confirms autoimmunity as the aetiology, but also provides a prognostic value for predicting progression to overt hypothyroidism. Thyroid ultrasound may show hypoechogenicity or heterogeneous echogenicity in autoimmune thyroid disease leading to subclinical hypothyroidism; however, a routine thyroid ultrasound is unnecessary in the diagnostic work-up of subclinical hypothyroidism.

Natural History of Subclinical Hypothyroidism

Without intervention, a small percentage of patients with subclinical hypothyroidism will progress to overt hypothyroidism, and the risk of progression is higher in patients with TPO-Ab. In the Whickham follow-up study, 2.6% TPO-Ab negative patients with subclinical hypothyroidism progressed to overt hypothyroidism each year as compared to 4.3% TPO-Ab positive patients [3]. Apart from TPO-Ab, TSH level also predicts the likelihood of progression. In a longitudinal study, patients with TSH 10.0–14.9 mU/L had a 10-fold higher incidence of progression to overt hypothyroidism, as compared to those with TSH 5.0–9.9 mU/L [17]. However, it is noteworthy that thyroid function recovers spontaneously over time in about 40% patients with subclinical hypothyroidism [17, 18]. Patients with lower baseline TSH levels are more likely to recover spontaneously.

Long-Term Complications of Subclinical Hypothyroidism

Cardiovascular Disease

Several epidemiological studies have examined for an association between subclinical hypothyroidism and the risk of cardiovascular outcomes. They have shown conflicting results due to differences in the demographic characteristics of patient populations, the study designs, and the TSH cut-off levels to define subclinical hypothyroidism. However, there is emerging evidence that patient age and TSH level have a major influence on the association. A meta-analysis of observational studies found a significant association between subclinical hypothyroidism and cardiovascular mortality in younger adult patients aged below 65 years (odds ratio (OR) 1.37; 95% confidence interval (CI) 1.04–1.79) but the association was not seen in older patients [19]. In fact, in very elderly patients above the age of 85 years, subclinical hypothyroidism has been shown to be associated

Table 3.4.4.1 Conditions simulating subclinical hypothyroidism with a raised TSH and normal free T_4

Conditions	Characteristic features
Assay interference (including heterophile antibodies and macro-TSH)	Laboratory results inconsistent with clinical status
Non-thyroidal illness	Recent acute illness
Thyroiditis (subacute, painless, or postpartum)	Recent viral illness, neck tenderness (subacute thyroiditis); childbirth in the last 6 months (postpartum thyroiditis)
Non-concordance with levothyroxine treatment	History of intermittently missing levothyroxine tablets
Addison's disease	Clinical features of adrenal insufficiency (e.g. fatigue, low blood pressure, postural hypotension, weight loss), skin pigmentation, worsening of symptoms after starting levothyroxine
Resistance to thyroid hormone (due to mutations in the thyroid hormone receptor β gene)	Often asymptomatic, presence of similar thyroid function test results in first degree relatives

with a decreased risk of cardiovascular and all-cause mortality [20]. In another meta-analysis, which included more than 50 000 participants from 11 prospective cohorts, increasing TSH level in subclinical hypothyroidism was associated with a higher risk of fatal and non-fatal coronary heart disease events [21]. This association was most striking in patients with TSH levels of 10 mU/L or above, with hazard ratios of 1.89 (95% CI 1.28–2.8) for non-fatal coronary heart disease events and 1.58 (95% CI 1.1–2.27) for coronary heart disease deaths. In a meta-analysis of six prospective observational studies with over 25 000 participants, subclinical hypothyroidism was associated with a significantly higher risk of heart failure in patients with TSH levels of 10 mU/L or above (hazard ratio 1.86; 95% CI 1.27–2.72) [22]. In addition, patients with pre-existing heart failure may have a poorer prognosis in the presence of coexisting subclinical hypothyroidism [23]. Finally, subclinical hypothyroidism has also been shown to be associated with several functional cardiovascular anomalies, including left ventricular diastolic dysfunction, reduced resting and exertional systolic function, increased vascular resistance, arterial stiffness, and endothelial dysfunction [24–26].

Cerebrovascular Disease

Like cardiovascular disease, the risk of cerebrovascular disease is increased in younger adult patients with subclinical hypothyroidism. A meta-analysis of 17 observational studies, which included over 47 000 participants, failed to demonstrate a significant increase in the risk of cerebrovascular events or fatal stroke in patients with subclinical hypothyroidism in the whole cohort [27]. However, in younger patients under the age of 65 years, subclinical hypothyroidism was significantly associated with an increased risk of fatal stroke (HR 2.29; 95% CI 1.41–3.74).

Adverse Metabolic Characteristics

Many observational studies have shown associations between subclinical hypothyroidism and several adverse metabolic parameters, which may partly explain the increased cardiovascular and cerebrovascular risk in younger patients with subclinical hypothyroidism. A meta-analysis of 16 observational studies, which included over 40 000 participants, found significantly increased levels of serum total cholesterol, low-density lipoprotein cholesterol and total triglyceride in patients with subclinical hypothyroidism as compared to euthyroid individuals although there was no difference in serum high-density lipoprotein cholesterol levels between the two groups [28]. The risk of adverse lipid profile increases with an increasing TSH level. In addition, there appears to be a gender-related difference in the relationship between subclinical hypothyroidism and adverse lipid profile, with a more pronounced effect in women [29]. Apart from dyslipidaemia, a meta-analysis found an association between subclinical hypothyroidism and increased risks of type 2 diabetes (OR 1.93, 95% CI 1.66–2.24) and its complications, including diabetic nephropathy (OR 1.74, 95% CI 1.34–2.28), diabetic retinopathy (OR 1.42, 95% CI 1.21–1.67), peripheral vascular disease (OR 1.85, 95% CI 1.35–2.54) and peripheral neuropathy (OR 1.87, 95% CI 1.06–3.28) [30]. Furthermore, subclinical hypothyroidism has been shown to be associated with a higher levels of plasma homocysteine (a risk factor for atherosclerosis) and insulin resistance as measured by homeostatic index of insulin resistance (HOMA-IR) [31]. Finally, a meta-analysis has also shown an association between subclinical hypothyroidism and non-alcoholic steatohepatitis (OR 1.63, 95% CI 1.19–2.24) [32].

Impaired Cognitive Function

Several studies have assessed for an association between subclinical hypothyroidism and impaired cognitive function showing inconsistent results. Two meta-analyses found no significant association between subclinical hypothyroidism and decline in cognitive function, as measured by Mini-Mental State Examination (MMSE) score [33, 34]. However, another meta-analysis has suggested that subclinical hypothyroidism is associated with an increased risk of impaired cognitive function and dementia in younger patients aged less than 75 years, particularly in those with higher TSH levels [35]. These observations suggest more research is needed to conclude whether subclinical hypothyroidism increases the risk of cognitive decline.

Bone Loss, Osteoporotic Fractures, and Frailty

Although optimum thyroid function is important for bone health, subclinical hypothyroidism has not been shown to be associated with an increased risk of bone loss [36] or osteoporotic fractures [37, 38]. Likewise, subclinical hypothyroidism is not associated with an increased risk of frailty in elderly populations [39, 40].

Mood, Mental Health, and Well-Being

Some studies have suggested that subclinical hypothyroidism is associated with poor neuropsychological function, mood, and quality of life [41, 42]. However, the association remains uncertain as other studies have failed to confirm these observations [43, 44].

In summary, subclinical hypothyroidism is associated with a higher risk of cardiovascular and cerebrovascular disease in younger adult patients (below the age of 65 years), particularly in those with TSH levels of 10 mU/L or above. This association may be partly explained by an increased prevalence of dyslipidaemia, type 2 diabetes, and other adverse metabolic characteristics in patients with subclinical hypothyroidism. In contrast, there is no convincing evidence to support the association between subclinical hypothyroidism and impaired cognitive function, osteoporosis, frailty, poor neuropsychological function, or reduced quality of life.

Management of Subclinical Hypothyroidism

Efficacy of Levothyroxine Treatment

The potential benefits of treating subclinical hypothyroidism include improvement in symptoms and quality of life, stopping progression to overt hypothyroidism and prevention of long-term complications. A previous systematic review of randomized controlled trials failed to demonstrate a benefit of levothyroxine treatment over placebo to improve symptoms or health-related quality of life in subclinical hypothyroidism [45]. However, a subsequent randomized controlled trial found an improvement in tiredness with levothyroxine treatment [46]. A recent randomized double-blind placebo-controlled trial of levothyroxine treatment in elderly patients (aged 65 years or above) failed to show an improvement in thyroid specific quality of life with the treatment in subclinical hypothyroidism [47]. Furthermore, another randomized controlled trial in elderly patients (above the age of 65 years) found no benefit of levothyroxine in subclinical hypothyroidism to improve cognitive function [48].

There is some evidence to show improvement in lipid profile, features on echocardiography and cardiac magnetic resonance imaging, and arterial stiffness with levothyroxine replacement in subclinical hypothyroidism [45, 46, 49]. However, at present, there are no large scale randomized controlled trials to show that levothyroxine treatment improves cardiovascular or cerebrovascular morbidity and mortality in subclinical hypothyroidism. Interestingly, a large observational study based on the UK general practice research database has found a reduction in fatal and non-fatal coronary heart disease events with levothyroxine in younger patients (aged between 40 and 70 years) with subclinical hypothyroidism, but not in older patients above the age of 70 years [50]. In contrast, a large cohort study based on the Danish national registries showed no evidence of reduction in myocardial infarction, cardiovascular deaths, or all-cause mortality with levothyroxine treatment in patients with subclinical hypothyroidism, although all-cause mortality was reduced in patients under the age of 65 years [51].

Potential Risks of Levothyroxine Treatment

Levothyroxine is considered a safe treatment in subclinical hypothyroidism provided that thyroid function is monitored regularly to avoid both overtreatment and undertreatment. Patients with subclinical hypothyroidism are prone to overtreatment with levothyroxine resulting in TSH suppression [52], which is associated with an increased risk of atrial fibrillation and osteoporosis, particularly in elderly patients [53, 54].

Recommendations from the International Guidelines

In the absence of randomized controlled trial evidence to show benefit of levothyroxine replacement in improving quality of life and long-term cardiovascular and cerebrovascular complications, the management of subclinical hypothyroidism remains controversial. However, because of the association between subclinical hypothyroidism and long-term cardiovascular morbidity and mortality, particularly in young patients and those with high TSH levels, the current guidelines from the American Thyroid Association/the American Association of Clinical Endocrinologists [55] and the European Thyroid Association [56] recommend levothyroxine treatment in subclinical hypothyroidism if TSH is 10 mU/L or above. The European guidelines take a more cautious approach in treating subclinical hypothyroidism in older patients above the age of 70 years and recommend treatment if TSH is 10 mU/L or above in the presence of clear symptoms of hypothyroidism or in the presence of high vascular risk [56].

Managing Subclinical Hypothyroidism in Routine Clinical Practice

An algorithm for management of non-pregnant adult patients under the age of 65 years with subclinical hypothyroidism is shown in Figure 3.4.4.1. In new patients with elevated serum TSH and normal fT_4, other potential causes (such as, assay interference) should be considered and excluded (Table 3.4.4.1). Thyroid function tests

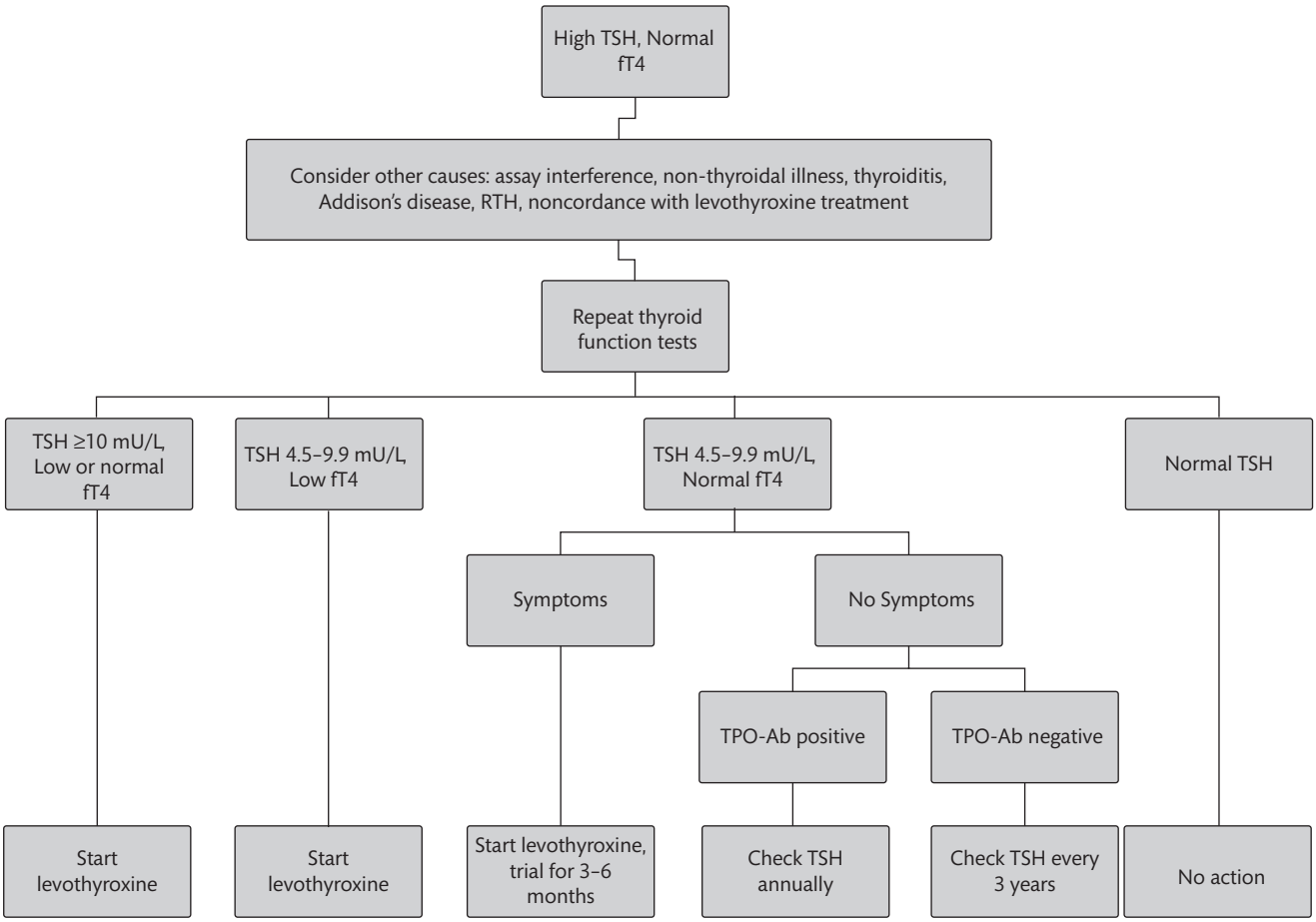


Figure 3.4.4.1 Algorithm for management of subclinical hypothyroidism in non-pregnant young adults. Please see the text for the management in elderly patients (>65 years age).

Abbreviations: fT_4, free thyroxine; RTH, resistance to thyroid hormone; TPO-Ab, thyroid peroxidase antibodies; TSH, thyroid-stimulating hormone.

should be repeated in 2–3 months, as thyroid function in patients with transient thyroiditis and non-thyroidal illness usually returns to normal without intervention. In patients with persistent subclinical hypothyroidism, treatment with levothyroxine should be considered if TSH is 10 mU/L or higher. In young adult patients with TSH <10 mU/L, levothyroxine treatment could be prescribed after evaluation case by case and discussion with the patient. For example, a trial of levothyroxine for 3–6 months should be considered in the presence of non-specific symptoms to assess symptomatic benefit, especially in the presence of TPO-Ab. In elderly patients (above the age of 65 years) with subclinical hypothyroidism, levothyroxine replacement is indicated in the presence of TSH above 10 mU/L and clear symptoms [56]. In cases of non-treatment, thyroid function should be monitored annually in presence of TPO-Ab and every 3 years in absence of TPO-Ab.

If the decision is made to treat subclinical hypothyroidism, levothyroxine is the treatment of choice. In young patients without known coronary heart disease, levothyroxine should be started on a dose based on bodyweight, at around 1.5 μg/kg/day [56]. In patients starting levothyroxine, TSH should be checked every 6–8 weeks adjusting the dose of levothyroxine until TSH is in the reference range. After that, thyroid function tests should be performed annually to keep thyroid function normal. In elderly patients and those with coronary heart disease, levothyroxine should be started on a smaller dose (e.g. 25 μg/day) and the dose increment should be more gradual.

There is no evidence to support the use of liothyronine, combination of liothyronine and levothyroxine, or desiccated thyroid extract in subclinical hypothyroidism [55, 56]. In autoimmune thyroid disease, selenium supplement can reduce titres of thyroid antibodies but does not affect thyroid hormone levels [57]. Therefore, it is not recommended as a treatment for subclinical hypothyroidism due to autoimmune thyroid disease [55, 56].

Conclusions

The current uncertainties about the management of subclinical hypothyroidism and the associated variability in the clinical practice reflect the lack of high-quality evidence. Clearly, there is an urgent need for well-designed large randomized controlled trials to investigate whether levothyroxine reduces the risk of cardiovascular and cerebrovascular events and improves quality of life in young adults with subclinical hypothyroidism. Even when a decision is made to treat subclinical hypothyroidism with levothyroxine, the optimal TSH target remains uncertain, particularly for elderly patients. It is hoped that future studies will inform optimum TSH targets for patients of different ages treated with levothyroxine.

REFERENCES

1. Peeters RP. Subclinical hypothyroidism. *N Engl J Med*, 2017; **376**: 2556–65.
2. den Elzen WP, Lefebre-van de Fliert AA, Virgini V, *et al*. International variation in GP treatment strategies for subclinical hypothyroidism in older adults: a case-based survey. *Br J Gen Pract*, 2015; **65**: e121–32.
3. Vanderpump MP, Tunbridge WM, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol*, 1995; **43**: 55–68.
4. Hollowell JG, Staehling NW, Flanders WD, *et al*. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*, 2002; **87**: 489–99.
5. Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado thyroid disease prevalence study. *Arch Int Med*, 2000; **160**: 526–34.
6. Alberti L, Proverbio MC, Costagliola S, *et al*. Germline mutations of TSH receptor gene as cause of nonautoimmune subclinical hypothyroidism. *J Clin Endocrinol Metab*, 2002; **87**: 2549–55.
7. Katagiri R, Yuan X, Kobayashi S, Sasaki S. Effect of excess iodine intake on thyroid diseases in different populations: a systematic review and meta-analyses including observational studies. *PloS One*, 2017; **12**: e0173722.
8. Lerro CC, Beane Freeman LE, DellaValle CT, *et al*. Occupational pesticide exposure and subclinical hypothyroidism among male pesticide applicators. *Occup Environ Med*, 2018; **75**: 79–89.
9. Wiersinga WM. Smoking and thyroid. *Clin Endocrinol*, 2013; **79**: 145–51.
10. Jorde R, Waterloo K, Storhaug H, Nyrnes A, Sundsfjord J, Jenssen TG. Neuropsychological function and symptoms in subjects with subclinical hypothyroidism and the effect of thyroxine treatment. *J Clin Endocrinol Metab*, 2006; **91**: 145–53.
11. Wartofsky L, Dickey RA. The evidence for a narrower thyrotropin reference range is compelling. *J Clin Endocrinol Metab*, 2005; **90**: 5483–8.
12. Roelfsema F, Pereira AM, Adriaanse R, *et al*. Thyrotropin secretion in mild and severe primary hypothyroidism is distinguished by amplified burst mass and Basal secretion with increased spikiness and approximate entropy. *J Clin Endocrinol Metab*, 2010; **95**: 928–34.
13. Surks MI, Boucai L. Age- and race-based serum thyrotropin reference limits. *J Clin Endocrinol Metab*, 2010; **95**: 496–502.
14. Andersen S, Pedersen KM, Bruun NH, Laurberg P. Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab*, 2002; **87**: 1068–72.
15. Koulouri O, Moran C, Halsall D, Chatterjee K, Gurnell M. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 745–62.
16. Hattori N, Ishihara T, Yamagami K, Shimatsu A. Macro TSH in patients with subclinical hypothyroidism. *Clin Endocrinol*, 2015; **83**: 923–30.
17. Diez JJ, Iglesias P. Spontaneous subclinical hypothyroidism in patients older than 55 years: an analysis of natural course and risk factors for the development of overt thyroid failure. *J Clin Endocrinol Metab*, 2004; **89**: 4890–7.
18. Somwaru LL, Rariy CM, Arnold AM, Cappola AR. The natural history of subclinical hypothyroidism in the elderly: the cardiovascular health study. *J Clin Endocrinol Metab*, 2012; **97**: 1962–9.
19. Razvi S, Shakoor A, Vanderpump M, Weaver JU, Pearce SH. The influence of age on the relationship between subclinical hypothyroidism and ischemic heart disease: a metaanalysis. *J Clin Endocrinol Metab*, 2008; **93**: 2998–3007.

20. Gussekloo J, van Exel E, de Craen AJ, Meinders AE, Frolich M, Westendorp RG. Thyroid status, disability and cognitive function, and survival in old age. *JAMA*, 2004; **292**: 2591–9.
21. Rodondi N, den Elzen WP, Bauer DC, *et al.* Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA*, 2010; **304**: 1365–74.
22. Gencer B, Collet TH, Virgini V, *et al.* Subclinical thyroid dysfunction and the risk of heart failure events: an individual participant data analysis from 6 prospective cohorts. *Circulation*, 2012; **126**: 1040–9.
23. Sato Y, Yoshihisa A, Kimishima Y, *et al.* Subclinical hypothyroidism is associated with adverse prognosis in heart failure patients. *Can J Cardiol*, 2018; **34**: 80–7.
24. Monzani F, Di Bello V, Caraccio N, *et al.* Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2001; **86**: 1110–15.
25. Owen PJ, Rajiv C, Vinereanu D, Mathew T, Fraser AG, Lazarus JH. Subclinical hypothyroidism, arterial stiffness, and myocardial reserve. *J Clin Endocrinol Metab*, 2006; **91**: 2126–32.
26. Biondi B, Galderisi M, Pagano L, *et al.* Endothelial-mediated coronary flow reserve in patients with mild thyroid hormone deficiency. *Eur J Endocrinol*, 2009; **161**: 323–9.
27. Chaker L, Baumgartner C, den Elzen WP, *et al.* Subclinical hypothyroidism and the risk of stroke events and fatal stroke: an individual participant data analysis. *J Clin Endocrinol Metab*, 2015; **100**: 2181–91.
28. Liu XL, He S, Zhang SF, *et al.* Alteration of lipid profile in subclinical hypothyroidism: a meta-analysis. *Med Sci Monit*, 2014; **20**: 1432–41.
29. Tognini S, Polini A, Pasqualetti G, *et al.* Age and gender substantially influence the relationship between thyroid status and the lipoprotein profile: results from a large cross-sectional study. *Thyroid*, 2012; **22**: 1096–103.
30. Han C, He X, Xia X, *et al.* Subclinical hypothyroidism and type 2 diabetes: a systematic review and meta-analysis. *PloS One*, 2015; **10**: e0135233.
31. Yang N, Yao Z, Miao L, *et al.* Novel clinical evidence of an association between homocysteine and insulin resistance in patients with hypothyroidism or subclinical hypothyroidism. *PloS One*, 2015; **10**: e0125922.
32. He W, An X, Li L, *et al.* Relationship between hypothyroidism and non-alcoholic fatty liver disease: a systematic review and meta-analysis. *Front Endocrinol*, 2017; 8: 335.
33. Akintola AA, Jansen SW, van Bodegom D, *et al.* Subclinical hypothyroidism and cognitive function in people over 60 years: a systematic review and meta-analysis. *Front Aging Neurosci*, 2015; **7**: 150.
34. Rieben C, Segna D, da Costa BR, *et al.* Subclinical thyroid dysfunction and the risk of cognitive decline: a meta-analysis of prospective cohort studies. *J Clin Endocrinol Metab*, 2016; **101**: 4945–54.
35. Pasqualetti G, Pagano G, Rengo G, Ferrara N, Monzani F. Subclinical hypothyroidism and cognitive impairment: systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2015; **100**: 4240–8.
36. Segna D, Bauer DC, Feller M, *et al.* Association between subclinical thyroid dysfunction and change in bone mineral density in prospective cohorts. *J Int Med*, 2018; **283**: 56–72.
37. Blum MR, Bauer DC, Collet TH, *et al.* Subclinical thyroid dysfunction and fracture risk: a meta-analysis. *JAMA*, 2015; **313**: 2055–65.
38. Yan Z, Huang H, Li J, Wang J. Relationship between subclinical thyroid dysfunction and the risk of fracture: a meta-analysis of prospective cohort studies. *Osteoporos Int*, 2016; **27**: 115–25.
39. Yeap BB, Alfonso H, Chubb SA, *et al.* Higher free thyroxine levels are associated with frailty in older men: the Health In Men Study. *Clin Endocrinol*, 2012; **76**: 741–8.
40. Virgini VS, Rodondi N, Cawthon PM, *et al.* Subclinical thyroid dysfunction and frailty among older men. *J Clin Endocrinol Metab*, 2015; **100**: 4524–32.
41. Sait Gonen M, Kisakol G, Savas Cilli A, *et al.* Assessment of anxiety in subclinical thyroid disorders. *Endocr J*, 2004; **51**: 311–15.
42. Gulseren S, Gulseren L, Hekimsoy Z, Cetinay P, Ozen C, Tokatlioglu B. Depression, anxiety, health-related quality of life, and disability in patients with overt and subclinical thyroid dysfunction. *Arch Med Res*, 2006; **37**: 133–9.
43. Bell RJ, Rivera-Woll L, Davison SL, Topliss DJ, Donath S, Davis SR. Well-being, health-related quality of life and cardiovascular disease risk profile in women with subclinical thyroid disease—a community-based study. *Clin Endocrinol*, 2007; **66**: 548–56.
44. Kim JS, Zhang Y, Chang Y, *et al.* Subclinical hypothyroidism and incident depression in young and middle-age adults. *J Clin Endocrinol Metab*, 2018; **103**: 1827–33.
45. Villar HC, Saconato H, Valente O, Atallah AN. Thyroid hormone replacement for subclinical hypothyroidism. *Cochrane Database Syst Rev*, 2007; **3**: CD003419.
46. Razvi S, Ingoe L, Keeka G, Oates C, McMillan C, Weaver JU. The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial. *J Clin Endocrinol Metab*, 2007; **92**: 1715–23.
47. Stott DJ, Rodondi N, Kearney PM, *et al.* Thyroid hormone therapy for older adults with subclinical hypothyroidism. *N Engl J Med*, 2017; **376**: 2534–44.
48. Parle J, Roberts L, Wilson S, *et al.* A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: the Birmingham Elderly Thyroid study. *J Clin Endocrinol Metab*, 2010; **95**: 3623–32.
49. Madathil A, Hollingsworth KG, Blamire AM, *et al.* Levothyroxine improves abnormal cardiac bioenergetics in subclinical hypothyroidism: a cardiac magnetic resonance spectroscopic study. *J Clin Endocrinol Metab*, 2015; **100**: E607–10.
50. Razvi S, Weaver JU, Butler TJ, Pearce SH. Levothyroxine treatment of subclinical hypothyroidism, fatal and nonfatal cardiovascular events, and mortality. *Arch Int Med*, 2012; **172**: 811–17.
51. Andersen MN, Olsen AM, Madsen JC, *et al.* Levothyroxine substitution in patients with subclinical hypothyroidism and the risk of myocardial infarction and mortality. *PloS One*, 2015; **10**: e0129793.
52. Taylor PN, Iqbal A, Minassian C, *et al.* Falling threshold for treatment of borderline elevated thyrotropin levels-balancing benefits and risks: evidence from a large community-based study. *JAMA Int Med*, 2014; **174**: 32–9.
53. Sawin CT, Geller A, Wolf PA, *et al.* Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N Engl J Med*, 1994; **331**: 1249–52.
54. Turner MR, Camacho X, Fischer HD, *et al.* Levothyroxine dose and risk of fractures in older adults: nested case–control study. *BMJ*, 2011; **342**: d2238.

55. Garber JR, Cobin RH, Gharib H, *et al.* Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Thyroid*, 2012; **22**: 1200–35.
56. Pearce SH, Brabant G, Duntas LH, *et al.* 2013 ETA Guideline: management of subclinical hypothyroidism. *Eur Thyroid J*, 2013; **2**: 215–28.
57. Winther KH, Wichman JE, Bonnema SJ, Hegedus L. Insufficient documentation for clinical efficacy of selenium supplementation in chronic autoimmune thyroiditis, based on a systematic review and meta-analysis. *Endocrine*, 2017; **55**: 376–85.

3.4.5 Syndromes of Resistance to Thyroid Hormone

Carla Moran, Mark Gurnell, and Krishna Chatterjee

Introduction

Thyroid hormones (TH—thyroxine, T_4; triiodothyronine, T_3) regulate many physiological processes in virtually every type of tissue. The diverse effects of TH include regulation of growth, control of basal metabolic rate, cardiac chronotropic and inotropic effects, and functional differentiation of the central nervous system. The synthesis of TH is controlled by hypothalamic thyrotropin-releasing hormone (TRH) and pituitary thyroid-stimulating hormone (TSH), and in turn, T_4 and T_3 regulate TRH and TSH production as part of a negative feedback loop.

The regulation of physiological processes by TH are mediated by changes in expression of specific target genes in different tissues. Thus, the feedback effects of TH on TSH production are mediated by inhibition of hypothalamic TRH and pituitary TSHα and β subunit gene expression. Nuclear thyroid hormone receptors (TRα1, TRβ1, TRβ2), encoded by different genes (*THRA, THRB*), mediate canonical TH action to regulate transcription of target genes in a hormone-dependent manner. TRα1 is most highly expressed in the central nervous system, bone, intestine, skeletal & cardiac muscle; TRβ1 is the predominant receptor subtype in liver and kidney; the TRβ2 isoform is most highly expressed in the pituitary and hypothalamus but is also found in the inner ear and retina. TRα2, a variant protein which is unable to bind thyroid hormone, is widely expressed but its biological function is not understood [1].

Unliganded TRs recruit a multiprotein complex, containing corepressors (CoR, e.g. nuclear receptor corepressor, NCoR; silencing mediator of retinoic acid and thyroid hormone receptor, SMRT) and histone deacetylase (HDAC), to inhibit basal gene transcription; hormone (T_3) occupancy of TRs promotes dissociation of the corepressor complex and relief of target gene repression, followed by recruitment of coactivators (e.g. steroid receptor coactivator 1 (SRC-1), cAMP response element-binding protein (CREB)-binding protein (CBP)), which mediate target gene activation [2] (Figure 3.4.5.1).

Thyroid hormone action is also regulated at other levels. Monocarboxylate transporter 8 (MCT8), a membrane protein, mediates cellular thyroid hormone uptake, particularly in the central nervous system (CNS). Intracellularly, a family of deiodinase enzymes (DIOs) mediate hormone metabolism: type 1 deiodinase (DIO1) in peripheral tissues is responsible for T_3 generation; type 2 deiodinase (DIO2) mediates T_4 to T_3 conversion in the CNS, including pituitary and hypothalamus; type 3 deiodinase (DIO3) catabolizes T_4 and T_3 to inactive metabolites (Figure 3.4.5.1).

The first syndrome of resistance to thyroid hormone (RTH), described in 1967 and characterized by elevated circulating TH, non-suppressed TSH and variable resistance to hormone action in target tissues, is mediated by defects in TRβ. However, recognition that other disorders, due to defective TH transport, intracellular metabolism of TH, or hormone action via TRα, are also characterized by tissue resistance, has led to a broader definition of RTH, encompassing all disorders that can interfere with the biological action of TH secreted in normal or excessive amounts [3].

Allan-Herndon-Dudley Syndrome—A Disorder of Thyroid Hormone Transport

Clinical Features

In 1944, a large family with X-linked intellectual disability was described by Allan, Herndon, and Dudley, with this and subsequent similar cases being designated an eponymous (Allan-Herndon-Dudley) syndrome (AHDS). Many decades later AHDS patients were noted to have abnormal thyroid function tests.

Patients with AHDS are usually born at term from uncomplicated pregnancy with normal birthweight. Generalized muscular hypotonia is a feature in early neonatal life, with persistent truncal hypotonia being associated with poor head control; peripherally, hypotonia progresses into spasticity with hyperreflexia, clonus, and extensor plantar responses. Poor motor development usually precludes ability of patients to sit, stand, or walk; cognitive impairment (IQ <40) limits speech severely, but patients retain awareness of surroundings with smiling and crying; hearing and vision are typically unaffected. Epilepsy, ranging from absence to tonic-clonic seizures, occurs in 25% of cases. Involuntary movements (dystonic, choreoathetoid, ballismus), sometimes occurring in paroxysms triggered by external stimuli, are a further characteristic. Progressive microcephaly is associated with reduced head circumference. Brain MRI in early life (<2 years) shows reduced myelination but this normalizes by 4 years of age, distinguishing this disorder from other hypomyelinating leukodystrophies (e.g.

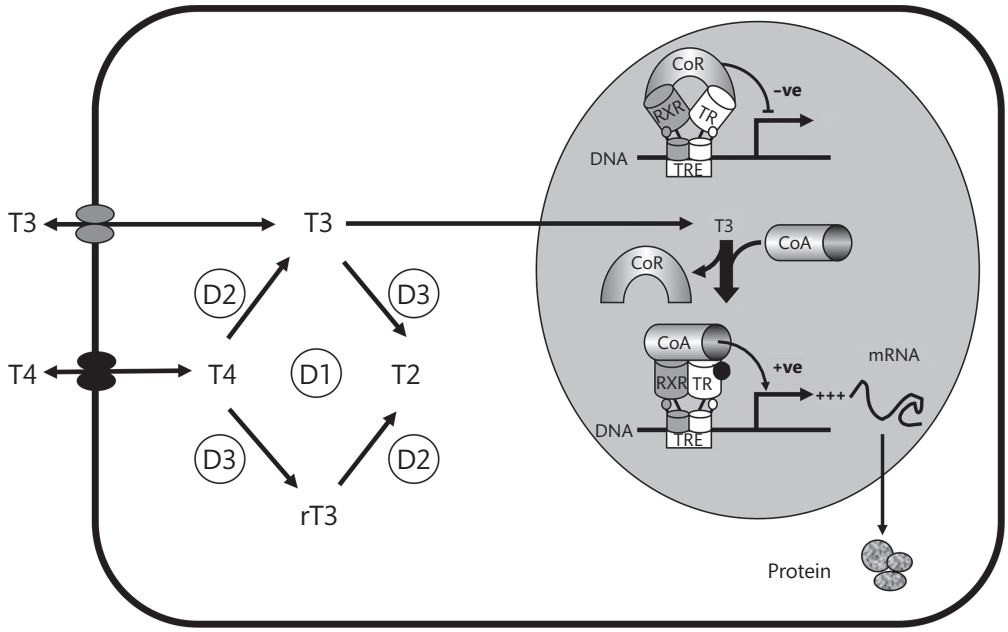


Figure 3.4.5.1 Schematic outline of thyroid hormone uptake, metabolism, and regulation of target gene transcription via binding to the nuclear receptor TR (positively regulated target gene shown). Transporters are required for the passage of T_3 and T_4 across the plasma membrane. The deiodinases (D1–3) catalyse conversion of T_4 to T_3 (D1, D2) or inactivation of T_4 to reverse T_3 (rT_3) and T_3 to T_2 (D_3). In the absence of ligand, TR binds to target gene response elements (TREs) as either a homodimer (not shown) or heterodimer with the retinoid X receptor (RXR). Basal gene transcription is inhibited by recruitment of a corepressor complex (CoR). The deacetylation of core histones in chromatin reduces access to general transcription factors resulting in transcriptional repression. Following addition of T_3, TR homodimers dissociate, whereas the heterodimer-DNA complex is stable. The corepressor complex is released, enabling recruitment of coactivator proteins (CoA). The intrinsic histone acetylase activity of the latter results in remodelling of chromatin leading to transcriptional activation.

Pelizaeus–Merzbacher disease) [4]. Although linear growth is preserved, body weight is usually extremely low in AHDS patients despite gastrostomy feeding to overcome swallowing difficulties; proximal muscle wasting is an added contributory factor [5–8]. Somatic features include elongated myopathic facies and a tented upper lip due to hypotonia. Some patients exhibit a milder neurological phenotype, walking with ataxic gait and being capable of dysarthric speech [9, 10].

AHDS patients exhibit reduced circulating free T_4, raised free T_3, normal or mildly elevated TSH, and subnormal reverse T_3 levels; consequently, circulating T_3/rT_3 ratios are raised. Serum levels of sex hormone-binding globulin (SHBG), a hepatic marker of thyroid action, are markedly elevated; in addition, ammonemia and raised lactic acid levels, signifying a catabolic state, has been documented [8, 11]. These biochemical features support the notion of a relative hyperthyroid state in peripheral tissues of MCT8-deficient patients.

Molecular Genetics and Pathogenesis

Consistent with the male-limited nature of the disorder, AHDS is due to defects in the monocarboxylate transporter 8 (*MCT8, SLC16A2*) gene located on chromosome Xq13.2 [12, 13]. Over 70 different *MCT8* mutations (deletions, frameshift/premature stop, missense) have been identified in ~100 families from diverse ethnic backgrounds, with complete penetrance in males and occurring *de novo* in ~25% of cases. When expressed in cells the majority of MCT8 mutants exhibit complete loss-of-function, with failure of TH transport; however, some missense mutants show residual transport capacity, which may correlate with a milder clinical phenotype in patients [10, 14]. Female carriers are fertile with normal neurocognitive and thyroid function, but in rare cases it is possible that skewed X-inactivation leading to MCT8 deficiency could mediate a phenotype.

MCT8 is expressed in many tissues, including brain, liver, kidney, heart, thyroid and placenta; in murine brain MCT8 is found predominantly in neurons and coexpressed with DIO_3 which catabolizes T_3, whereas DIO_2 mediating T_4 to T_3 conversion is largely expressed in adjacent astrocytes and glial cells; accordingly, one theory is that MCT8-mediated neuronal uptake of T_3, generated by surrounding, DIO_2-containing, astrocytes, maintains euthyroid status of the CNS [5]. Alternatively, a study using patient-derived stem cells has suggested that the key defect is loss of MCT8-mediated T3 transport across the endothelial blood-brain barrier [15].

Management

General supportive measures include use of braces or baclofen to prevent joint contractures; gastrostomy feeding to prevent aspiration and catabolic weight loss; counteracting dystonia with anticholinergic agents or L-DOPA; reducing hypersalivation with glycopyrrolate or scopolamine; and anticonvulsants to treat seizure activity.

Specific treatment approaches include normalization of TH levels with a 'block and replace' combination of propylthiouracil and T4. Treatment of older AHDS patients has targeted increased body weight and reduced heart rate [10].

An alternative approach involves administration of a thyroid hormone analogue whose CNS uptake is MCT8-independent. Diiodothyropropionic acid (DITPA) has been used in four, younger, AHDS patients, with normalization of circulating TH levels, weight gain, and reduction in heart rate [16]. 3,5,3'-triiodothyroacetic acid (TRIAC), another TH analogue, has been trialled in AHDS patients (ClinicalTrials.gov NCT02060474), with preliminary results showing normalization of fT_3 levels and alleviation of peripheral hyperthyroidism [17].

Selenoprotein Deficiency–A Disorder of Thyroid Hormone Metabolism

Selenium, a trace element that is essential for human health, is incorporated as selenocysteine (Sec) into 25 different human proteins, including the deiodinase enzymes. A complex molecular mechanism mediates Sec incorporation into proteins during their biosynthesis (Figure 3.4.5.2); consequently, genetic defects in this pathway result in a multisystem disorder due to deficiency of diverse selenoproteins, with lack of deiodinase activities mediating abnormal thyroid function.

Clinical Features

Many cases have presented in childhood due to growth retardation with raised circulating fT_4, normal to low fT_3, and raised reverse T_3 levels, reflecting deficiency of all three deiodinase enzymes [13, 18].

Muscle weakness due to progressive rigid spine muscular dystrophy, affecting axial and proximal limb muscles with raised creatine kinase (CK) levels and fatty infiltration on MRI scan, is attributable to deficiency of selenoprotein N which is enriched in skeletal muscle. Male azoospermic infertility reflects loss of specific selenoproteins (e.g. mitochondrial GPx4, thioredoxin-glutathione reductase, selenoprotein V) required for spermatogenesis [18].

Lack of antioxidant selenoenzymes results in cellular accumulation of hydrogen peroxide and reactive oxygen species (ROS), mediating with oxidative DNA damage and peroxidation of membrane lipids. Clinical consequences of raised cellular ROS include skin photosensitivity, sensorineural hearing loss which worsens with age and increased total body adipose tissue mass, but associated paradoxically with preserved or enhanced systemic insulin sensitivity [18]. We have documented significant, progressive dilatation of the ascending aorta due to cystic medial degeneration requiring cardiac surgery in older patients [19].

Additional age-dependent phenotypes such as neurodegeneration, premature ageing or neoplasia may emerge, but have not been described hitherto.

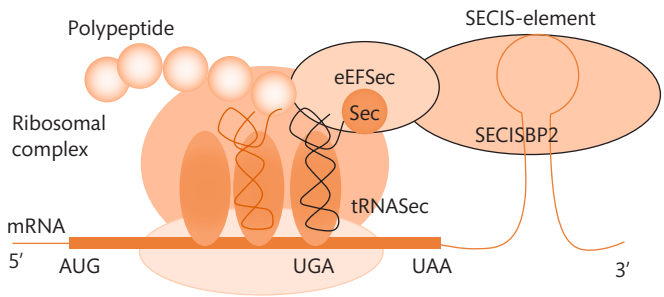


Figure 3.4.5.2 Mechanism of selenoprotein biosynthesis. The 3′-untranslated region of selenoprotein mRNAs contains a stem-loop RNA structure (SECIS element) which interacts with a protein complex that includes SECIS binding protein 2 (SECISBP2) and Sec-specific elongation factor (eEFSec), to enable ribosomal recruitment of selenocysteyl-transfer RNA (tRNASec) to the UGA codon and incorporation of selenocysteine (Sec) into the nascent polypeptide. Failure of this mechanism results in miscoding of the UGA as a stop codon, terminating protein synthesis.

Molecular Genetics

To date, 13 individuals from 11 families, with compound heterozygous or homozygous *SECISBP2* defects have been described; residual production of functional SBP2 protein likely preserves some selenoprotein synthesis and precludes the disorder being embryonic lethal, as occurs in *secisbp2* knockout mice [18, 20, 21]. As SBP2 is differentially rate-limiting for Sec incorporation into different proteins, a hierarchy of selenoprotein synthesis preserves production of essential, cellular housekeeping selenoproteins (e.g. thioredoxin reductases) at the expense of other factors (e.g. deiodinases) that are dispensable for survival.

In one patient, a homozygous mutation in *TRU-TCA1-1*, affecting synthesis and posttranscriptional modification of selenocysteine transfer ribonucleic acid (RNA), has been described [22].

Mutations in *SEPSECS*, an enzyme mediating synthesis of tRNAsec, are associated with cerebrocerebellar atrophy and a profound neurological phenotype due to disruption of selenoprotein synthesis in the CNS, but are not associated with a thyroid biochemical phenotype [21].

Management

Oral selenium supplementation raises circulating selenium levels but with no effect on circulating selenoprotein or TH levels [23]. T_3 treatment can improve growth in childhood [18]; antioxidant (vitamin E) administration reduces cellular lipid peroxidation but other clinical effects have not been determined [24].

Resistance to Thyroid Hormone Beta

Resistance to thyroid hormone action in the hypothalamic–pituitary–thyroid axis is the hallmark of RTHβ, with inappropriate pituitary TSH secretion driving T_4 and T_3 production, to establish a new equilibrium with high circulating TH, together with non-suppressed TSH levels. RTHβ was first described in 1967 in two siblings with high circulating TH levels who were clinically euthyroid and exhibited other features (deaf-mutism, stippled femoral epiphyses, winging of scapulae, pectus carinatum, short stature, and dysmorphic facies) which are unique to this kindred, in which the disorder was recessively inherited [25, 26].

The prevalence of RTHβ is approximately 1 in 40 000 and several hundred cases (from more than 250 families) have been described [27]. The disorder is usually dominantly inherited and associated with variable clinical features. Many patients are either asymptomatic or have non-specific symptoms and a goitre, prompting testing of thyroid function, which suggests the diagnosis. In such individuals, classified as exhibiting generalized resistance (GRTH), the high thyroid hormone levels are thought to compensate for ubiquitous

tissue resistance, resulting in a euthyroid state. In contrast, a subset of individuals with the same TH abnormalities exhibit thyrotoxic clinical features: in adults these can include weight loss, tremor, palpitations, insomnia, and heat intolerance; in children failure to thrive, accelerated growth, and hyperkinetic behaviour have also been noted. When the latter clinical entity was first described, patients were thought to have 'selective' or predominant pituitary resistance to thyroid hormone action (PRTH), with preservation of normal responses to TH in peripheral tissues [28].

However, comparison of characteristics of individuals classified as GRTH or PRTH indicates significant overlap between these entities, with no differences in age, sex ratio, frequency of goitre, or levels of free T_4, free T_3, or TSH between patients with the two types of RTHβ. Significantly, features such as tachycardia, hyperkinetic behaviour, and anxiety have been documented in individuals with GRTH. Conversely, serum SHBG—a hepatic marker of thyroid hormone action—is normal in patients with PRTH, suggesting that tissue resistance is not solely confined to the pituitary–thyroid axis in this group [29]. Indeed, in some RTHβ cases, hypothyroid features such as growth retardation, delayed dentition, or bone age in childhood or hypercholesterolaemia in adults, may coexist with thyrotoxic symptoms in the same individual. Nevertheless, the absence or presence of overt thyrotoxic symptoms, signifying either GRTH or PRTH phenotypes, respectively, is a clinical distinction which remains useful in guiding management of the disorder [27].

Clinical Features

Goitre

Palpable goitre is present in up to 65% of individuals—especially adult women. Fewer children with RTHβ born to affected mothers exhibit thyroid enlargement (35%) compared to offspring born of unaffected mothers (87%), suggesting that maternal hyperthyroxinaemia may protect against goitre formation [30]. Increased biological activity of circulating TSH may account for goitre and marked hyperthyroxinaemia in some RTHβ patients with normal immunoreactive TSH levels [31]. A multinodular thyroid gland can develop, particularly following previous goitre surgery, but thyroid cancer is rare and most likely represents coincidental micropapillary carcinoma [32].

Cardiovascular System

In a large cohort of children and adults with RTHβ, resting heart rate was significantly raised with some indices of cardiac systolic and diastolic function (e.g. stroke volume, cardiac output), suggesting a 'partially hyperthyroid' cardiac phenotype. Atrial fibrillation is commoner in older RTHβ patients [33], but they do not exhibit the hypercoagulable state associated with conventional hyperthyroidism.

Skeletal Abnormalities

Childhood short stature (height <5th centile) has been noted in 18% and delayed bone age (more than two standard deviations) in 29% in RTHβ, but final adult height is not usually affected [30, 34]. In adults, we have measured bone mineral density in approximately 80 subjects with RTHβ and documented a reduction at both femoral neck (mean Z score –0.71) and lumbar spine (mean Z score –0.73), but with normal bone turnover markers (Gurnell, Chatterjee and Beck-Peccoz, unpublished observations).

Metabolic

Resting energy expenditure (REE) is substantially increased in adults and children with RTHβ, due to mitochondrial uncoupling in skeletal muscle [35]. Increased energy expenditure is accompanied by hunger, reduced satiety, and raised energy intake, with hyperphagia being particularly evident in childhood [35].

Central Nervous System

A history of attention-deficit hyperactivity disorder (ADHD) in childhood was more frequent (75%) in RTHβ patients compared to their unaffected relatives (15%) [36], but screening for RTHβ in ADHD cohorts is negative. Children and adults with RTHβ exhibit problems with language development, manifested by poor reading skills and problems with articulation [37]. Frank intellectual disability (IQ less than 60) is quite uncommon but 30% of patients show mild learning disability (IQ less than 85) [30].

Hearing and Vision

Significant hearing loss has been documented in 21% of RTHβ cases [38]: in most, audiometry indicated a conductive defect, probably related to an increased recurrent ear infections in childhood; abnormal otoacoustic emissions, suggestive of cochlear dysfunction, were also documented in those with hearing deficit [38]. Defective colour vision has been documented in homozygous RTHβ [25]; abnormal photoreceptor electroretinography in heterozygous RTHβ [39] is not associated with overt colour visual dysfunction.

Other Associated Disorders

RTHβ cases with coexistent autoimmune hypothyroidism or Graves' disease have been recorded and the increased prevalence of thyroid autoantibodies found in a large RTHβ cohort suggests possible predisposition to thyroid autoimmunity [40]. Inappropriate thyroid ablation in RTHβ results in pituitary thyrotrope hyperplasia [41], and a small number of RTHβ cases with coexisting pituitary adenomas have also been described. Recurrent otitis media and upper respiratory tract infections are more frequent in RTHβ (30). A higher miscarriage rate and neonatal growth retardation has been documented in unaffected offspring of mothers with RTHβ, suggesting that intrauterine exposure to elevated TH could be detrimental [42].

Differential Diagnosis

Concordance of elevated free thyroid hormone levels when measured in different immunoassays or by equilibrium dialysis makes artefactual hyperthyroxinaemia due to assay interference from familial dysalbuminaemic hyperthyroxinaemia (FDH) or anti-iodothyronine antibodies less likely. Likewise, linear fall in TSH levels following serial dilution of serum or its adequate recovery following polyethylene glycol (PEG) precipitation can exclude interference with TSH measurement. Other causes of hyperthyroxinaemia (neonatal period, systemic illness, drugs) are excluded by recognition of the abnormal clinical context or documenting subsequent normalization of thyroid function following recovery or drug withdrawal [43].

The main differential diagnosis of RTHβ is from a TSH-secreting pituitary tumour (TSHoma), but many factors can make this

distinction difficult. There are no differences in age, gender, fT_4, fT_3, or TSH levels between the two disorders. MRI scan may show an obvious macroadenoma but fail to visualize small microadenomas; the recognized occurrence of pituitary 'incidentalomas' in RTHβ can confound diagnosis. Circulating pituitary α-subunit levels can be normal in micro TSHomas; serum SHBG levels are elevated in TSHoma but can be normal with tumours cosecreting growth hormone. Abnormal thyroid function tests (TFTs) in first-degree relatives are highly suggestive of RTHβ but are also a feature of FDH which is dominantly inherited; conversely, normal TFTs in family members does not exclude RTHβ as the disorder does occur sporadically [43]. With most TSHomas, administration of long-acting somatostatin analogue normalizes TH levels whereas fT_4 and fT_3 levels remain unchanged in RTHβ cases [44]; inhibition of tracer uptake with functional pituitary imaging following somatostatin analogue administration can be invaluable in diagnosis of TSHoma [45].

Molecular Genetics

Following cloning of TRs, RTHβ was shown to be tightly linked to the *THRB* locus. In keeping with the dominant inheritance of RTHβ, affected individuals are heterozygous for *THRB* mutations, which occur *de novo* in approximately 10% of sporadic cases. Over 150 different defects (missense, frameshift/premature stop, in-frame deletions) have been recorded, localizing to three mutation clusters within the hormone-binding domain of the receptor [27] (Figure 3.4.5.3).

THRB mutations have been identified in both GRTH and PRTH, indicating that these clinical entities represent different phenotypic manifestations of a single genetic disorder. Nevertheless, certain TRβ mutations (e.g. R338W/L, R383C/H, R429Q) may be particularly associated with a PRTH phenotype [27].

Nine cases of homozygous or hemizygous RTHβ have been described, associated with a severe clinical phenotype encompassing marked elevation of TH levels, dysmorphic features, audiovisual abnormalities with a propensity to developing life-threatening cardiac failure [46, 47].

In ~10% of cases, biochemical evidence of RTHβ is not associated with a *THRB* defect—so-called non-TRβ RTH. Possible explanations include somatic mosaicism with occurrence of TRβ mutations whose expression is restricted so as to be undetectable in peripheral blood leucocyte DNA [48] or non-receptor mechanisms (e.g. cofactor gene defects) whereby thyroid hormone action is disrupted to produce the RTHβ phenotype [27].

Properties of Mutant Receptors

In keeping with their location in the hormone-binding domain, the ability of TRβ mutants to bind T_3 and recruit coactivator is reduced, impairing their ability to regulate target gene transcription.

In the first documented family with RTHβ, both affected siblings were homozygous for complete deletion of *THRB*. Significantly, their heterozygous parents, with deletion of one *THRB* allele, were completely normal with no evidence of thyroid dysfunction [27]. Thus, TRβ haploinsufficiency is insufficient to mediate RTH; furthermore, TRβ mutants in heterozygous RTHβ are not simply functionally impaired but also capable of inhibiting the action of their wild-type counterparts in a 'dominant negative' manner [49] (Figure 3.4.5.4).

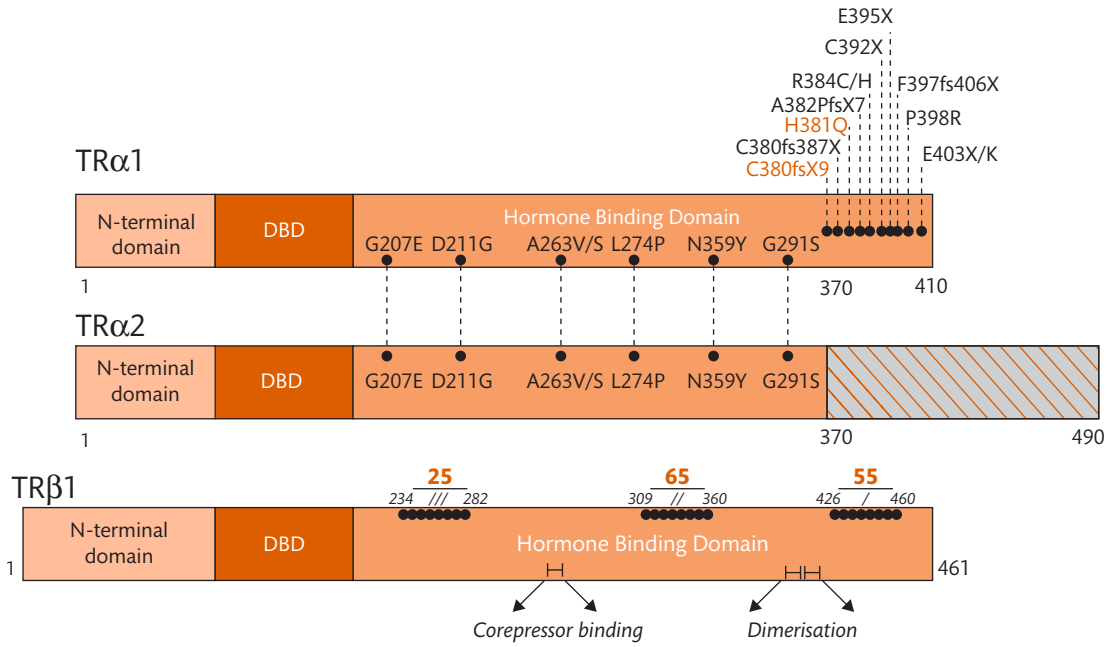


Figure 3.4.5.3 Schematic alignment of TRα1, TRα2 and TRβ1 showing functional regions (N-terminal, DNA-binding domain (DBD), hormone-binding domains); the divergent carboxyterminus of TRα2 is cross-hatched. The location of published (black) and unpublished (orange) RTHα mutations, involving either TRα1 alone or both TRα1 and α2 proteins, is superimposed. Three regions (aminoacids 234 to 282, 309 to 360 and 426 to 460) of TRβ1 within which RTHβ mutations cluster are depicted. The number of different mutations described hitherto within each cluster (orange) is superimposed. No RTHβ mutations have been recorded in regions of the hormone-binding domain that are important for corepressor binding or dimerization with RXR.

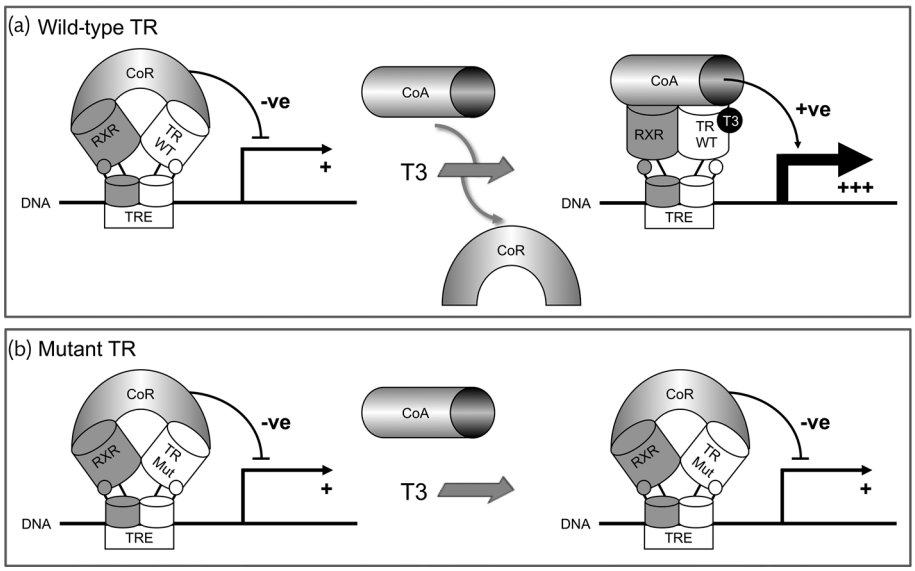


Figure 3.4.5.4 Possible mechanism for dominant negative inhibition by TRβ or TRα mutants. The upper panel (a) depicts wild-type (WT) TR action on target genes. The unliganded RXR-TR heterodimer recruits a corepressor complex (CoR) to silence basal gene transcription. Receptor occupancy by ligand (T_3), promotes corepressor dissociation followed by binding of a coactivator complex (CoA) which leads to target gene activation. The lower panel (b) shows mutant receptor action. In comparison to wild-type TR, the primary defect in mutant receptors is impaired hormone-dependent corepressor release or coactivator recruitment. For most receptor mutants, this functional alteration is a consequence of reduced hormone binding. However, a subset exhibits intrinsic enhanced binding to corepressor or impaired recruitment of coactivator, with preserved hormone binding. Constitutive occupancy of thyroid response elements (TREs) by mutant receptor-corepressor complexes results in inhibition of target gene expression.

Pathogenesis of Variable Tissue Resistance

The ability to exert a dominant negative effect within the hypothalamic–pituitary–thyroid axis is a key property of mutant receptors in RTHβ, generating abnormal TFTs that are the hallmark of the disorder. For some RTHβ mutants, their functional impairment *in vitro* correlates with the degree of axis resistance *in vivo*, quantified by the magnitude of elevation in circulating free T_4 levels.

Variable tissue resistance is partly mediated by differing tissue distributions of TR subtypes. Thus, TRβ-expressing tissues (hypothalamus, pituitary, liver) exhibit hormone resistance, exemplified by non-suppressed TSH and normal SHBG levels in patients; conversely, cardiac hyperthyroidism and raised metabolic rate seen in RTHβ represent features of sensitivity of TRα-expressing myocardium and skeletal muscle to elevated TH. Differences in relative expression of wild-type versus mutant TRβ in tissues or variable dominant negative inhibitory potency of TRβ mutants in different target gene contexts are further variables which may influence the degree of hormone resistance [27].

Management

The management of RTHβ is complex, as variable resistance makes it difficult to maintain euthyroidism in all tissues. In most individuals, the receptor defect is compensated by high circulating thyroid hormone levels, leading to a clinically euthyroid state. Inappropriate thyroid ablation with surgery or radioiodine is commonly unsuccessful, with recrudescence of goitre and disruption of the pituitary–thyroid axis, rendering the RTHβ patient hypothyroid unless levothyroxine (L-T_4) is administered in supraphysiologic dosage.

Conversely, reduction in thyroid hormone levels may be of benefit in the management of patients with thyrotoxic symptoms. However, conventional antithyroid drug administration results in marked elevation of TSH levels with consequent further thyroid enlargement and pituitary thyrotrope hyperplasia, with a risk of autonomous tumour formation. 3,5,3'-triiodothyroacetic acid (TRIAC), a TH analogue which acts centrally to inhibit TSH secretion thereby reducing thyroid hormone levels, yet is devoid of peripheral thyromimetic activity, has been shown to be beneficial in both childhood and adult cases [50]. A daily dose of 1.4–4.5 mg is used, with one study suggesting that twice-daily administration inhibits TSH secretion more effectively [51]. Given spontaneous variation in thyrotoxic symptoms in RTHβ, periodic cessation of TRIAC therapy and re-evaluation of the clinical status of the patient is advisable. The use of TRIAC in one pregnancy controlled maternal thyrotoxic symptoms but may have induced fetal goitre [52]. Use of antithyroid drugs in combination with TRIAC may be of value in homozygous RTHβ associated with cardiac hyperthyroidism [47]. Thyroid ablation followed by thyroxine replacement in subphysiological dosage could also be used in RTHβ associated with life-threatening, thyrotoxic cardiac failure.

TRIAC treatment is not always successful and dextro-thyroxine (D-T4) is another agent which has been shown to be effective in some cases [53]. Bromocriptine or octreotide have been used in RTHβ but, unlike TSHomas, pituitary TSH secretion escapes from their inhibitory effects [54, 55].

The treatment of RTHβ with thyrotoxic manifestations (for example, failure to thrive) in childhood also requires careful monitoring to ensure that any reduction in thyroid hormone levels is not associated with growth retardation or adverse neurological sequelae. Indeed, control of cardiac and peripheral sympathetic overactivity with β-blockade may be the safest course in this context. One study showed that L-T3 therapy improved hyperactivity in nine children with ADHD, including three individuals who were unresponsive to methylphenidate [56].

Dyslipidaemia should be managed with statin therapy if associated cardiac risk factors suggest benefit; in future, TRβ-selective thyromimetics may prove a more targeted therapeutic option [57]. If reduced bone mineral density is identified, usual lifestyle measures (weight bearing exercise, calcium and vitamin D supplementation) can be advised. Whether bisphosphonate therapy is beneficial in RTHβ remains unknown.

During pregnancy, normal fetal growth and heart rate in the third trimester may provide reassurance that significant fetal thyrotoxicosis is not present. In women with RTHβ and a history of recurrent miscarriage, antithyroid drug treatment in early pregnancy could be considered [58].

Resistance to Thyroid Hormone Alpha

TRβ and TRα proteins are highly homologous, with 80% aminoacid identity within their hormone-binding domains favouring the existence of individuals with resistance to thyroid hormone due to defective TRα (RTHα). Indeed, transgenic mice harbouring mutant TRα are viable and exhibit recognizable phenotypes but associated with normal TH levels [59], likely accounting for delayed identification of the first human RTHα case in 2012 [60].

Clinical Features

Interestingly, some features of congenital hypothyroidism, such as umbilical hernia, macroglossia, poor feeding and hoarse cry, are reported in RTHα patients at birth [61, 62]. The majority of patients exhibit delayed growth, which can be predominantly lower segmental; many have macrocephaly and facial dysmorphic features including broad facies, hypertelorism, a flattened nasal bridge, enlarged tongue, thick lips [63, 64]. Older children and adults have an excessive number of skin tags and moles [60–62].

Neurocognitive

Developmental milestones (motor, speech) in childhood are almost uniformly delayed. Dyspraxia, slow and dysarthric speech, and ataxic gait are also common. IQ is variably reduced, being markedly subnormal in a few cases, but with other patients achieving third level qualifications [64–66]. Autism spectrum disorder is associated with one case [67]. In severe RTHα cases, MRI has documented cortical microencephaly, reduced cerebellar volume and white matter tract density [65].

Gastrointestinal

Constipation is common and variable, but was sufficiently severe in one case to warrant consideration of surgical intervention prior to diagnosis of RTHα [60].

Cardiovascular

Many patients are bradycardic with some indices of cardiac function being within the hypothyroid range [61].

Metabolic and Endocrine

Resting energy expenditure is reduced, likely reflecting relative hypothyroidism of skeletal muscle and myocardium [60–62, 66]. No reproductive abnormalities have been identified to date, with both maternal and paternal transmission of the disorder being documented [62, 64, 68].

Biochemical Features

TFTs are marginally deranged, with the most consistent abnormality being low or low-normal free T_4 with high or high-normal T_3, resulting in a low T_4/T_3 ratio. Reverse T_3 levels are subnormal in some cases [63]. Mild anaemia and raised muscle CK levels are observed frequently [60–62, 64, 66, 69]. Dyslipidaemia has been recorded in both children and adults [61, 69].

Skeletal

Radiographic features in childhood include delayed fusion of fontanelles and excessively serpiginous (Wormian bone) skull sutures, femoral epiphyseal dysgenesis, delayed dentition, and bone maturation [60, 64, 70]. In adulthood, features include a thickened skull vault, cortical hyperostosis, and increased bone mineral density [60, 68].

Differential Diagnosis

RTHα should be considered in any child with growth retardation or delayed development; the presence of features such as macrocephaly, constipation, anaemia, skeletal dysplasia, or raised muscle CK levels further raises suspicion, reducing the threshold for genetic testing. Low free T_4 and high free T_3 levels are also features of MCT8 deficiency or dyshormonogenetic hypothyroidism (due to iodine deficiency and/or a genetic cause), but these conditions can be distinguished clinically from RTHα (Table 3.4.5.1).

Molecular Genetics

Nineteen different heterozygous *THRA* mutations (frameshift/premature stop, missense) inherited from either parent or occurring *de novo* have been recorded in 30 cases (Figure 3.4.5.3). Frameshift/premature stop TRα1 mutants exhibit either markedly reduced hormone binding and impaired regulation of target genes [60, 61, 68] and are associated with a severe clinical phenotype [60, 61, 69]. Analogous to TRβ mutants in RTHβ, TRα1 mutants inhibit the function of their wild-type receptor counterparts in a dominant negative manner [60, 68, 69], via a mechanism involving constitutive mutant TRα1-corepressor interaction, with failure of normal dissociation of CoR from TRs in the presence of thyroid hormone [60, 61] (Figure 3.4.5.3).

Several RTHα cases involve *THRA* mutations which affect both TRα1 and TRα2 subtypes [62, 66, 70–72] (Figure 3.4.5.3). When studied in the TRα2 subtype background, these mutations exert no gain or loss-of-function, which correlates with absence of any discernible additional clinical phenotype attributable to mutant TRα2, in these patients [62, 66, 70, 71]. In a single case, unusual phenotypic

Table 3.4.5.1 Comparison of biochemical and clinical features in syndromes of resistance to thyroid hormone

Disorder	Resistance to thyroid hormone beta (RTHβ)	Resistance to thyroid hormone alpha (RTHα)	Allan-Herndon-Dudley syndrome	Selenoprotein deficiency*
Gene (S)	*THRB*	*THRA*	*MCT8*	*SBP2; TRU-TCA 1-1*
Free T4	Raised	Low-normal or Low	Normal or low	Raised
Free T3	Raised	High-normal or High	Raised	Normal or Low
TSH	Normal (or raised)	Normal (or raised)	Normal (or raised)	Normal (or raised)
Reverse T3	Raised	Low	Low	Raised
Clinical Features	Goitre, hyperthyroid symptoms	Neurodevelopmental retardation, skeletal dysplasia, constipation	Severe mental and psychomotor retardation	Growth retardation, muscular dystrophy, hearing loss, male infertility, aortic dilatation

* Low plasma selenium levels are also characteristic of these disorders.

features (micrognathia, clavicular agenesis, syndactyly of digits, chronic diarrhoea) may not be related to the *THRA* defect [73].

The clinical phenotype in patients with missense TRα1 mutations is variable and likely correlates with the severity of the underlying receptor defect. Thus, a mild phenotype (mild developmental and pubertal delay, subtle facial abnormalities, normal stature, normal schooling) and favourable response to thyroxine therapy was recorded in a patient with a missense TRα1 mutation (A263V) exhibiting partial, T3-reversible, loss-of-function *in vitro*; in contrast, a severe phenotype (marked developmental and growth retardation, dysmorphic facial features, special schooling), refractory to thyroxine therapy, was seen in another patient harbouring a missense TRα1 mutation (L274P) with severe, virtually irreversible, dysfunction *in vitro* [66] with similar correlations of phenotype with genotype-phenotype being reported in other cases [70].

Given the prevalence (~1 in 40 000) of RTHβ due to over 150 different *THRB* mutations and marked homology of TRα and TRβ subtypes (with equivalent receptor defects being recorded in RTHα and RTHβ), it is highly likely that RTHα is more common than currently ascertained. Such underdiagnosis may reflect lack of a clear-cut, diagnostic biochemical signature, or a variable phenotype in this disorder.

Pathogenesis

Skeletal abnormalities (delayed fontanelle closure and dentition, femoral epiphyseal dysgenesis, Wormian skull sutures) and dysmorphic features (coarse facies, macroglossia, flattened nasal bridge) in RTHα are recognized features of untreated congenital or childhood hypothyroidism. Likewise, constipation due to reduced colonic motility is a recognized symptom of hypothyroidism and can be complicated by colonic dilatation or ileus [63].

Thyroid hormone abnormalities in RTHα are attributable to altered TH metabolism, rather than biosynthesis: increased hepatic DIO1 expression (seen in transgenic mice harbouring mutant TRα1), likely mediates enhanced T_4 to T_3 conversion, resulting in elevated free T_3 and low free T_4 levels; diminished DIO3 activity (a TRα1-regulated deiodinase) may contribute to reduced conversion of T_4 to rT_3 and also predispose patients to developing skin tags, as inhibition of this enzyme enhances keratinocyte proliferation in mice [63]. TRα mutations affect the balance between proliferation and differentiation of erythroid progenitor cells *in vitro*, possibly accounting for the mild anaemia seen in RTHα patients [74]. Idiopathic epilepsy seen in RTHα cases [61, 62] may not be coincidental; increased susceptibility to seizures following photic or audiogenic stimulation and aberrant GABAergic inhibitory interneuron development have been recorded in TRα mutant mice [63].

Treatment

Thyroxine therapy of RTHα is beneficial, improving growth, well-being, basal metabolic rate, and lowering elevated low-density lipoprotein (LDL) cholesterol and CK levels [60–62, 66, 69, 71, 72]; in turn, these metabolic changes may limit excessive weight gain in patients. In the childhood patient first described [60], 7 years of thyroxine therapy have been clearly beneficial, improving lower segment length and total height, alleviating constipation (with improved colonic contractile activity), and enhancing well-being (Moran & Chatterjee, unpublished observations). Addition of growth hormone to thyroxine therapy in childhood does not result in further improvement in growth [69]. Commencement of thyroxine in early childhood may have ameliorated the phenotype in cases harbouring mutant TRα1 whose dysfunction is reversible at higher TH levels [62]. In adult life, these individuals report that thyroxine therapy improves dyspraxia, constipation and enhances social interaction (Moran & Chatterjee, unpublished observations). In virtually all cases thyroxine treatment does not improve anaemia.

Following thyroxine therapy TSH levels fall quickly and remain suppressed, with elevation of fT_3 to supraphysiologic levels; serum SHBG rises further from high-normal baseline levels [60, 72] and in one case, biochemical markers of bone turnover became progressively elevated [61]; conversely, cardiac parameters, REE and muscle CK levels are less responsive to thyroxine treatment [61, 62]. These observations suggest preserved sensitivity and possible propensity to TH toxicity of TRβ-containing tissues (e.g. liver) but relative thyroid hormone resistance in TRα1-expressing organs (e.g. heart, skeletal muscle). Accordingly, ideal future therapies would selectively activate mutant or residual normal TRα1 function to overcome resistance in TRα-expressing tissues.

With current circulating abnormal parameters in RTHα being non-specific and insensitive, future identification of RTHα-specific biomarkers would be invaluable: better diagnosis of the disorder with earlier commencement of thyroxine therapy may improve outcome; specific biomarkers may also guide treatment, correlating with overcoming hormone resistance in TRα-expressing tissues and helping avoid TH toxicity in TRβ-containing organs.

REFERENCES

1. Lazar MA. Thyroid hormone receptors: multiple forms, multiple possibilities. *Endocr Rev*, 1993; **14**: 184–93.
2. Horlein AJ, Heinzel T, Rosenfeld MG. Gene regulation by thyroid hormone receptors. *Curr Opin Endocrinol Diabetes*, 1996; **3**: 412–16.
3. Refetoff S, Bassett JHD, Beck-Peccoz P, *et al*. Classification and proposed nomenclature for inherited defects of thyroid hormone action, cell transport, and metabolism. *J Clin Endocrinol Metab*, 2014; **99**: 768–70.
4. Vaurs-Barriere C, Deville M, Sarret C, *et al*. Pelizaeus-Merzbacher-Like disease presentation of MCT8 mutated male subjects. *Ann Neurol*, 2009; **65**: 114–18.
5. Friesema ECH, Jansen J, Heuer H, Trajkovic M, Bauer K, Visser TJ. Mechanisms of disease: psychomotor retardation and high T3 levels caused by mutations in monocarboxylate transporter 8. *Nat Clin Pract Endocrinol Metab*, 2006; **2**: 512–23.
6. Brockmann K, Dumitrescu AM, Best TT, Hanefeld F, Refetoff S. X-linked paroxysmal dyskinesia and severe global retardation caused by defective MCT8 gene. *J Neurol*, 2005; **252**: 663–6.
7. Holden KR, Zuniga OF, May MM, *et al*. X-linked MCT8 gene mutations: characterization of the pediatric neurologic phenotype. *J Child Neurol*, 2005; **20**: 852–7.
8. Namba N, Etani Y, Kitaoka, *et al*. Clinical phenotype and endocrinological investigations in a patient with a mutation in the MCT8 thyroid hormone transporter. *Eur J Pediatr*, 2008; **167**: 785–91.
9. Schwartz CE, May MM, Carpenter NJ, *et al*. Allan-Herndon-Dudley syndrome and the monocarboxylate transporter 8 (MCT8) gene. *Am J Hum Genet*, 2005; **77**: 41–53.
10. Visser WE, Vrijmoeth P, Visser FE, Arts WFM, van Toor H, Visser TJ. Identification, functional analysis, prevalence and treatment of monocarboxylate transporter 8 (MCT8) mutations in a cohort of adult patients with mental retardation. *Clin Endocrinol (Oxf)*, 2013; **78**: 310–15.
11. Herzovich V, Vaiani E, Marino R, *et al*. Unexpected peripheral markers of thyroid function in a patient with a novel mutation of the MCT8 thyroid hormone transporter gene. *Horm Res*, 2007; **67**: 1–6.
12. Friesema EC, Grueters A, Biebermann H, *et al*. Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet*, 2004; **364**: 1435–7.
13. Dumitrescu AM, Liao XH, Best TB, Brockmann K, Refetoff S. A novel syndrome combining thyroid and neurological abnormalities is associated with mutations in a monocarboxylate transporter gene. *Am J Hum Genet*, 2004; **74**: 168–75.
14. Jansen J, Friesema EC, Kester MH, Schwartz CE, Visser TJ. Genotype-phenotype relationship in patients with mutations in thyroid hormone transporter MCT8. *Endocrinology* 2008; **149**: 2184–90.
15. Vatine GD, Al-Ahmad A, Barriga BK, *et al*. Modeling psychomotor retardation using ipscs from MCT8-deficient patients indicates a prominent role for the blood-brain barrier. *Cell Stem Cell*, 2017; **20**: 831–43.
16. Verge CF, Konrad D, Cohen M, *et al*. Diiodothyropropionic acid (DITPA) in the treatment of MCT8 deficiency. *J Clin Endocrinol Metab*, 2012; **97**: 4515–23.
17. Groeneweg S, de Coo R, van Beynum I, *et al*. Thyroid hormone analog therapy in patients with MCT8 deficiency: the TRIAC trial. *Eur Thyroid J*, 2018; **7**(suppl 1): 3.
18. Schoenmakers E, Agostini M, Mitchell C, *et al*. Mutations in the selenocysteine insertion sequence-binding protein 2 gene lead to a multisystem selenoprotein deficiency disorder in humans. *J Clin Invest*, 2010; **120**: 4220–35.
19. Visser WE, Schoenmakers E, Moran C, *et al*. Thoracic aortic degeneration and dilatation: a newly-recognised complication of human selenoprotein. *Se2017—200 Years of Selenium Research Abstract Book*. Stockholm Aug 2017; abstract P97: 243.
20. Dumitrescu AM, Liao X, Abdullah MSY, *et al*. Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat Genet*, 2005; **37**: 1247–52.
21. Schoenmakers E, Schoenmakers N, Chatterjee K. Mutations in humans that adversely affect the selenoprotein synthesis pathway. In: Hatfield DL, Schweizer U, Tsuji PA, Gladyshev VN, eds. *Selenium. Molecular Biology and Its Role in Human Health*, 4th edn. New York, NY: Springer, 2016: 523–38.
22. Schoenmakers E, Carlson B, Agostini M, *et al*. Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *J Clin Invest*, 2016; **126**: 992–6.
23. Schomburg L, Dumitrescu AM, Liao X, *et al*. Selenium supplementation fails to correct the selenoprotein synthesis defect in subjects with SBP2 gene mutations. *Thyroid*, 2009; **19**: 277–81.
24. Saito Y, Shichiri M, Hamajima T, *et al*. Enhancement of lipid peroxidation and its amelioration by vitamin E in a subject with mutations in the *SBP2* gene. *J Lipid Res*, 2015; **56**: 2172.
25. Refetoff S, De Wind LT, De Groot LJ. Familial syndrome combining deaf-mutism, stippled epiphyses, goitre and abnormally high PBI: possible target organ refractoriness to thyroid hormone. *J Clin Endocrinol Metab*, 1967; **27**: 279–94.
26. Refetoff S, Weiss RE, Usala SJ. The syndromes of resistance to thyroid hormone. *Endocr Rev*, 1993; **14**: 348–99.
27. Gurnell M, Visser TJ, Beck-Peccoz P, Chatterjee VK. Resistance to thyroid hormone. In: Jameson JL, DeGroot LJ, eds. *Endocrinology*, 7th edn. Philadelphia, PA: Elsevier Saunders, 2015: 1648–65.
28. Gershengorn MC, Weintraub BD. Thyrotropin-induced hyperthyroidism caused by selective pituitary resistance to thyroid hormone. A new syndrome of inappropriate secretion of TSH. *J Clin Invest*, 1975; **56**: 633–42.
29. Beck-Peccoz P, Chatterjee VKK. The variable clinical phenotype in thyroid hormone resistance syndrome. *Thyroid*. 1994; **4**: 225–32.
30. Brucker-Davis F, Skarulis MC, Grace MB, *et al*. Genetic and clinical features of 42 kindreds with resistance to thyroid hormone. *Ann Intern Med*, 1995; **123**: 572–83.
31. Persani L, Asteria C, Tonacchera M, Vitti P, Chatterjee VKK, Beck-Peccoz P. Evidence for the secretion of thyrotropin with enhanced bioactivity in syndromes of thyroid hormone resistance. *J Clin Endocrinol Metab*, 1994; **78**: 1034–9.
32. Vinagre J, Borges F, Costa A, *et al*. Differentiated thyroid cancer in patients with resistance to thyroid hormone syndrome. A novel case and a review of the literature. *Front Mol Biosci*, 2014; **1**: 1–5.
33. Kahaly GJ, Matthews CH, Mohr-Kahaly S, Richards CA, Chatterjee VK. Cardiac involvement in thyroid hormone resistance. *J Clin Endocrinol Metab*, 2002; **87**: 204–12.
34. Weiss RE, Refetoff S. Effect of thyroid hormone on growth: lessons from the syndrome of resistance to thyroid hormone. *Endocrinol Metab Clin North Am*, 1996; **25**: 719–30.
35. Mitchell CS, Savage DB, Dufour S, *et al*. Resistance to thyroid hormone is associated with raised energy expenditure, muscle

mitochondrial uncoupling, and hyperphagia. *J Clin Invest*, 2010; **120**: 1345–54.
36. Hauser P, Zametkin AJ, Martinez P, *et al*. Attention deficit-hyperactivity disorder in people with generalized resistance to thyroid hormone. *N Engl J Med*, 1993; **328**: 997–1001.
37. Mixson AJ, Parrilla R, Ransom SC, *et al*. Correlation of language abnormalities with localization of mutations in the thyroid hormone receptor in 13 kindreds with generalized resistance to thyroid hormone: identification of four new mutations. *J Clin Endocrinol Metab*, 1992; **75**: 1039–45.
38. Brucker-Davis F, Skarulis MC, Pikus A, *et al*. Prevalence and mechanisms of hearing loss in patients with resistance to thyroid hormone. *J Clin Endocrinol Metab*, 1996; **81**: 2768–72.
39. Campi I, Cammarata G, Bianchi Marzoli S, *et al*. Retinal photoreceptor functions are compromised in patients with resistance to thyroid hormone syndrome (RTHβ). *J Clin Endocrinol Metab*, 2017; **102**: 2620–7.
40. Barkoff MS, Kocherginsky M, Anselmo J, Weiss RE, Refetoff S. Autoimmunity in patients with resistance to thyroid hormone. *J Clin Endocrinol Metab*, 2010; **95**: 3189–93.
41. Gurnell M, Rajanayagam O, Barbar I, Jones MK, Chatterjee VK. Reversible pituitary enlargement in the syndrome of resistance to thyroid hormone. *Thyroid*, 1998; **8**: 679–82.
42. Anselmo J, Cao D, Karrison T, Weiss RE, Refetoff S. Fetal loss associated with excess thyroid hormone exposure. *JAMA*, 2004; **292**: 691–5.
43. Koulouri O, Moran C, Halsall DJ, Chatterjee VK, Gurnell M. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 745–62.
44. Mannavola D, Persani L, Vannucchi G, *et al*. Different responses to chronic somatostatin analogues in patients with central hyperthyroidism. *Clin Endocrinol*, 2005; **62**: 176–81.
45. Koulouri O, Hoole AC, English P, *et al*. Localisation of an occult thyrotropinoma with 11C-methionine PET-CT before and after somatostatin analogue therapy. *Lancet Diabetes Endocrinol*, 2016; **4**: 1050.
46. Ferrara AM, Onigata K, Ercan O, Woodhead H, Weiss RE, Refetoff S. Homozygous thyroid hormone receptor β-gene mutations in resistance to thyroid hormone: three new cases and review of the literature. *J Clin Endocrinol Metab*, 2012; **97**: 1328–36.
47. Moran C, Habeb AM, Kahaly GJ, *et al*. Homozygous resistance to thyroid hormone β: can combined antithyroid drug and triiodothyroacetic acid treatment prevent cardiac failure. *J Endocr Soc*, 2017; **1**: 1203–12.
48. Mamanasiri S, Yesil S, Dumitrescu AM, *et al*. Mosaicism of a thyroid hormone receptor-beta gene mutation in resistance to thyroid hormone. *J Clin Endocrinol Metab*, 2006; **91**: 3471–7.
49. Collingwood TN, Adams M, Tone Y, Chatterjee VKK. Spectrum of transcriptional dimerization and dominant negative properties of twenty different mutant thyroid hormone receptors in thyroid hormone resistance syndrome. *Mol Endocrinol*, 1994; **8**: 1262–77.
50. Anzai R, Adachi M, Sho N, Muroya K, Asakura Y, Onigata K. Long-term 3,5,3'-triiodothyroacetic acid therapy in a child with hyperthyroidism caused by thyroid hormone resistance: pharmacological study and therapeutic recommendations. *Thyroid* 2012; **22**: 1069–75.
51. Ueda S, Takamatsu J, Fukata S, et al. Differences in response of thyrotropin to 3,5,3'triiodothyronine and 3,5,3'-triiodothyroacetic acid in patients with resistance to thyroid hormone. *Thyroid*, 1996; **6**: 563–70.
52. Asteria C, Rajanayagam O, Collingwood TN, *et al*. Prenatal diagnosis of thyroid hormone resistance. *J Clin Endocrinol Metab*, 1999; **84**: 405–10.
53. Hamon P, Bovier-LaPierre M, Robert M, Peynaud D, Pugeat M, Orgiazzi J. Hyperthyroidism due to selective pituitary resistance to thyroid hormones in 15-month-old boy: efficacy of D-thyroxine therapy. *J Clin Endocrinol Metab*, 1988; **67**: 1089–93.
54. Dulgeroff AJ, Geffner ME, Koyal SN, Wong M, Hershman JM. Bromocriptine and TRIAC therapy for hyperthyroidism due to pituitary resistance to thyroid hormone. *J Clin Endocrinol Metab*, 1992; **75**: 1071–5.
55. Beck-Peccoz P, Mariotti S, Guillausseau PJ, *et al*. Treatment of hyperthyroidism due to inappropriate secretion of thyrotropin with the somatostatin analog SMS 201–995. *J Clin Endocrinol Metab*, 1989; **68**: 208–14.
56. Weiss RE, Stein MA, Refetoff S. Behavioral effects of liothyronine (L-T3) in children with attention deficit hyperactivity disorder in the presence and absence of resistance to thyroid hormone. *Thyroid*, 1997; **7**: 389–93.
57. Taub R, Chiang E, Chabot-Blanchet M, *et al*. Lipid lowering in healthy volunteers treated with multiple doses of MGL-3196, a liver-targeted thyroid hormone receptor b agonist. *Atherosclerosis*, 2013; **230**: 373–80.
58. Pappa T, Anselmo J, Mamanasiri S, Dumitrescu AM, Weiss RE, Refetoff S. Prenatal diagnosis of resistance to thyroid hormone and its clinical implications. *J Clin Endocrinol Metab*, 2017; **102**: 3775–82.
59. Vennstrom B, Mittag J, Wallis K. Severe psychomotor and metabolic damages caused by a mutant thyroid hormone receptor alpha 1 in mice: can patients with a similar mutation be found and treated? *Acta Paed*, 2008; **97**: 1605–10.
60. Bochukova E, Schoenmakers N, Agostini M, *et al*. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*, 2012; **366**: 243–9.
61. Moran C, Schoenmakers N, Agostini M, *et al*. An adult female with resistance to thyroid hormone mediated by defective thyroid hormone receptor α. *J Clin Endocrinol Metab*, 2013; **98**: 4254–61.
62. Moran C, Agostini M, Visser E, *et al*. Resistance to thyroid hormone caused by a mutation in thyroid hormone receptor (TR) alpha1 and alpha2: clinical, biochemical and genetic analyses of three related patients. *Lancet Diabetes Endocrinol*, 2014; **2**: 619–26.
63. Moran C, Chatterjee VK. Resistance to Thyroid Hormone due to defective thyroid receptor alpha. *Best Pract Res Clin Endocrinol Metab*, 2015; **29**: 647–57.
64. Tylki-Szymanska A, Acuna-Hidalgo R, Krajewska-Walasek M, *et al*. Thyroid hormone resistance syndrome due to mutations in the thyroid hormone receptor α gene (THRA). *J Med Gen*, 2015; **52**: 312–16.
65. Krieger T, Moran C, Frangini A, *et al*. Mutations in thyroid hormone receptor a1 cause premature neurogenesis and progenitor cell depletion in human cortical development. BioRxiv 2019; http://dx.doi.org/10.1101/529677
66. Moran C, Agostini M, McGowan A, *et al*. Contrasting phenotypes in resistance to thyroid hormone α correlate with divergent properties of thyroid hormone receptor α1 mutant proteins. *Thyroid*, 2017; **27**: 973–82.

67. Yuen RKC, Thiruvahindrapuram B, Merico D, *et al*. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med*, 2015; **21**: 185–91.
68. van Mullem A, van Heerebeek R, Chrysis D *et al*. Clinical phenotype and mutant TRα1. *N Engl J Med*, 2012; **366**: 1451–3.
69. van Mullem AA, Chrysis D, Eythimiadou A, *et al*. Clinical phenotype of a new type of thyroid hormone resistance caused by a mutation of the TRα1 receptor: consequences of LT4 treatment. *J Clin Endocrinol Metab*, 2013; **98**: 3029–38.
70. Demir K, van Gucht A, Buyukinan M, *et al*. Diverse genotypes and phenotypes of three novel thyroid hormone receptor alpha mutations. *J Clin Endocrinol Metab*, 2016; **101**: 2945–54.
71. Van Gucht AL, Meima ME, Zwaveling-Soonawala N, *et al*. Resistance to thyroid hormone alpha in an 18-month-old girl: clinical, therapeutic and molecular characteristics. *Thyroid*, 2016; **26**: 338–46.
72. Korkmaz O, Ozen S, Ozdemir TR, Goksen D, Darcan S. A novel thyroid hormone receptor alpha gene mutation, clinic characteristics, and follow-up findings in a patient with thyroid hormone resistance. *Hormones (Athens)*, 2019; **18**: 223–7.
73. Espiard S, Savagner F, Flamant F *et al*. A novel mutation in THRA gene associated with an atypical phenotype of resistance to thyroid hormone. *J Clin Endocrinol Metab*, 2015; **100**: 2841–8.
74. van Gucht ALM, Meima ME, Moran C, *et al*. Anemia in patients with resistance to thyroid hormone α: a role for thyroid hormone receptor α in human erythropoiesis. *J Clin Endocrinol Metab*, 2017; **102**: 3517–25.

3.4.6 Treatment of Hypothyroidism

Birte Nygaard

Indications of Initiating Therapy

Hypothyroidism can be divided into clinical/overt hypothyroidism and subclinical/mild hypothyroidism (SCH) [1] (see Chapter 3.4.4.)

Overt hypothyroidism is defined as having increased serum thyroid-stimulating hormone (TSH) combined with serum-free thyroxine (fT_4) less than lower reference value. Serum TSH of more than 10 mU/L will, in most patients, represent overt hypothyroidism.

Biochemical Measurements

Measurement of serum fT_4 and not just serum TSH is necessary to discriminate between SCH and overt hypothyroidism due to individual TSH setpoints [2], and repeated measurement must be applied, as a single slight increased serum TSH could represent transient fluctuation in serum TSH [3].

Initiating Therapy

L-T_4 is available in most countries as tablets of 25, 50, and 100 μg. It has a half-life of 7 days, and should be given as a single daily dose.

Practice varies slightly from centre to centre and between countries. A dose of L-T_4 app 1.6 μg/kg/day is recommended as a final dose. Some centres recommend in younger to middle-aged patients with no history of cardiac disease to start directly on this dose while other centres recommend a stepwise increase (i.e. 50 μg, increased to 100 μg, and then to 125–150 μg at intervals of 2–3 weeks). After approximately 3 months of therapy, minor adjustment can be made to stabilize the serum concentrations of TSH within reference values.

In older patients and patients with a known cardiovascular disease a stepwise increase is recommended to avoid a sudden increase in metabolic rate in a patient with long-standing severe hypothyroidism which may unmask previously unrecognized ischaemic heart disease risking angina, myocardial infarction, dysrhythmia, or even sudden death. In these patients, the recommendations are to begin with 25 μg of L-T_4 daily with increments of 25 μg daily every 3–4 weeks.

Patients begin to feel better within weeks of L-T_4 treatment with a minor reduction in body weight, rarely more than 10% caused by fluid loss, whereas maximal improvement can't be evaluated before the thyroid parameters have been stable for several months.

Not Reaching the Treatment Goal

Most hypothyroid patients treated according to the aforementioned guidelines feel they return to normal health without complaints. However, persistent symptoms such as tiredness, depression, lack of energy, decreased cognitive function, weight increase, and musculoskeletal symptoms are described in 25% of L-T_4 treated hypothyroid patients compared to 15–20% in controls without thyroid disease [4, 5].

At the same time, an increased risk of getting/having a diagnose of psychiatric disease *hazard ratio* (HR) 1.5 (95% confidence-interval (CI): 1.12–2.04) [6] and an increased need for a disability pension HR 2.24 (95% CI: 1.73–2.89) is described [7].

Several explanations for this have been suggested [8]: autoimmunity by itself; comorbidity (known as well as unknown); dysregulation /low compliance; non-specific symptoms—not related to thyroid disease; *and* a possible subgroup needing a supplement of T_3.

Ensuring Compliance

Once the correct dose of L-T_4 has been established, the tradition of good practice has been to evaluate the patient and measure serum

TSH concentrations annually. However, several studies have demonstrated [9, 10] that almost 50% of patients treated with long-term L-T_4 have serum TSH out of reference range and compliance to L-T_4 treatment seems to be a major determinant factor to reach the target TSH levels in hypothyroid patients [9]. A combination of both raised serum fT_4 and TSH concentrations is most likely due to overenthusiastic tablet-taking for a few days before a clinic visit by a patient who was previously taking L-T_4 sporadically.

If compliance is suspected to be the problem, a week dosage box is recommended, and the patients can be instructed to take a missing L-T_4 tablet, when the patients realize that it has been forgotten. If access to the pharmacy databases is possible, an evaluation of the number of tablets, bought by the patients, can give a reasonable overview of the exact taken dose.

It is recommended to take L-T_4 before breakfast on an empty stomach. However, L-T_4 can be taken at *any* time of day and in patients with variations in thyroid function parameters better compliance and by this more stable parameter can often be obtained by changing the schedule as mornings often are busy, especially for women having small children. Taken at bedtime, is associated with higher thyroid hormone concentrations and lower serum TSH concentrations compared to the same dose taken in the morning, probably due to greater gastrointestinal uptake of L-T_4 during the night [11].

Weekly administration of seven times the daily dose of L-T_4 may be of benefit in *very* poorly compliant patients [12, 13], but there is little or no experience of its efficacy and safety.

Other Reasons for Variation in Thyroid Parameters

Malabsorption due to coeliac disease, small intestinal bacterial overgrowth with Helicobacter pylori or other bacteria's can be the explanation for the need of unexpected high doses of L-T_4 [14].

Pseudo malabsorption due to drug and food interference could also be the reason for dysregulation (see Table 3.4.6.1). If feasible L-T_4 intake should be separated from potential interfering drugs if these induce changes in serum TSH. This can, however, lead to decreased compliance—therefore a simpler approach should be recommended as standard and there is no need to for taking L-T_4 separately from food in most patients.

Different Formulations of L-T4

There are numerous manufacturers of L-T_4 preparations and divergence in bioequivalence can be present. An overall recommendation is to prescribe the same preparation for each prescription refill in each patient. When changes from one brand to another are made, it is important to check the thyroid function parameters to ensure that the dose is correct [15].

Liquid L-T_4 and soft gel capsules have higher bioavailability than the tablet L-T_4 in in hypothyroid patients with gastric disorders, malabsorption, or drug interference [16].

Effect of Altering L-T4 Dose on Quality of Life (QoL)

It has been discussed whether small changes within the normal thyroid function reference range could optimize QoL. In a new randomized controlled trial (RCT) on 138 L-T_4 treated patients with stable serum TSH were randomized to have either lower, unchanged, or higher L-T_4 dose. The mean TSH levels were 1.85 ± 0.25, 3.93 ± 0.38, and 9.49 ± 0.80 mU/L (P <0.001), respectively, in the three arms. However, these changes did not affect QoL significantly [17].

Other Treatments of Hypothyroidism

Combinations of L-T_4 and L-T_3

It has been assumed that all the necessary T_3 could be derived from peripheral mono-deiodination of orally administered L-T_4. Nevertheless, data has pointed towards a possible need for supplementary T_3 to fully restore the balance between TSH, TRH, thyroid hormones, deiodinase, and metabolism. Data show a 15–20% lower serum fT_3/fT_4 ratio given a stabile serum TSH after thyroidectomy than before in a prospective study on 50 euthyroid patients before and after total thyroidectomy for a benign nodular goitre, or thyroid cancer [18], and when comparing thyroidectomized patients to euthyroid control having similar serum TSH [19] (see Figure 3.4.6.1).

In accordance with this a study in rats have demonstrated a downregulation of the deiodinase enzyme leading to a downregulation of serum T3 as a respond to higher levels of serum fT_4 [20].

In thyroidectomized rats it has been demonstrated, that it is only possible to restore normal concentrations of fT_3 in all tissues, when normal serum concentrations of fT_3, fT_4, and TSH are maintained, by giving a combination of L-T_4 as well as L-T_3, and not just L-T_4 [21]. Although the relative importance of thyroidal T_3 secretion versus extrathyroidal T_3 generation from T_4 is different in rats compared to humans with a molar ratio of T_4 to T_3 in thyroidal secretion in 5.7:1 in rats and 14:1 in humans [22] these findings could suggest that the treatment modality with L-T_4 alone might be inadequate.

This theory was supported in a small human study in 1999 showing that a combination of L-T_4 and L-T_3 induced significant improvements in mood and neuropsychological function, compared to giving a higher dose of L-T_4 alone [23]. Hereafter several

Table 3.4.6.1 Interaction between fT_4 and drugs

Decreased intestinal absorption of fT_4	Increased clearance	Increased binding to TBG	fT_4 changes effect of medicine
PPI Ferrous sulphate Calcium carbonate Cholestyramine Aluminium hydroxide Cimetidine Sucralfate Magnesium Zinc fibres Caffeine Iodine Selenium	Phenytoin Carbamazepine Phenobarbital Rifampicin Sertraline (exact mechanism is not clear)	Oestrogen	*Increased effect* Warfarin Amitriptyline *Decreased effect* Propranolol

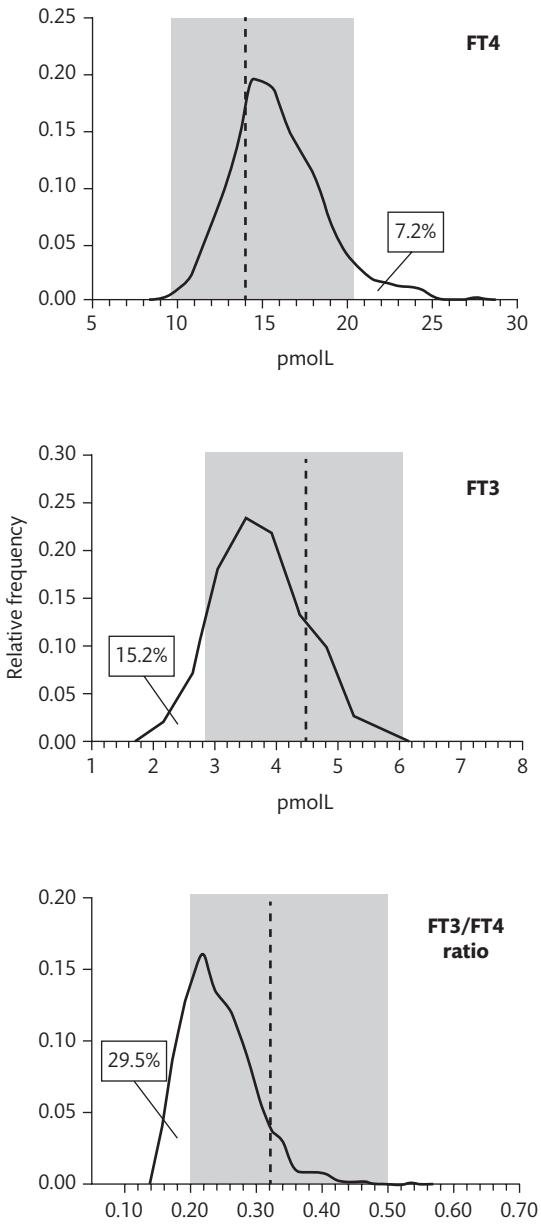


Figure 3.4.6.1 Free thyroid hormones and fT_3/ fT_4 ratio frequency distribution. fT_3 and fT_4 serum levels and fT_3/fT_4 ratio distribution in 1811 athyreotic patients under levothyroxine (L-T_4) monotherapy. Shaded areas indicate the normal range (2.5–97.5 percentiles) calculated in 3875 euthyroid controls. Vertical dotted lines indicate the median of the normal values. Percentages indicate the patients with values under or above the normal values.

Reproduced with permission from Gullo D, Latina A, Frasca F, Le Moli R, Pellegriti G, Vigneri R. Levothyroxine monotherapy cannot guarantee euthyroidism in all athyreotic patients. *PLoS One*. 2011;6(8):e22552. Copyright © 2011 Gullo D et al.

studies were initiated to confirm these results. A meta-analysis in 2006 of data from more than 1000 unselected patients from 11 trials could not confirm these results [24]. However, the included studies in the meta-analysis were heterogeneous, including unselected patients for example patients with SCH.

A minor study in a highly selected patient group with Hashimoto' disease and overt hypothyroidism (serum TSH>20 mU/L at the time of diagnosis) as well as persistent symptoms despite a stable L-T_4 dose and serum TSH with in normal range confirmed the original data with significant effects in 7 out of 11 quality of life scores and preference for L-T_3/L-T_4 combination therapy in 49% compared to L-T_4 monotherapy in 15% ($P = 0.002$) [25].

A calculation of patient preferences was published in the European Thyroid Association (ETA) guidelines from 2012. In 5 RCT crossover studies (228 patients), patients were asked which medication they preferred: 27% preferred L-T_4, 25% had no preference, and 48% preferred L-T_4/L-T_3 combination [8].

A possible decrease in bodyweight has been suggested at being the explanation for preference of combination therapy [26], however a new study on 23 patients shifted from L-T_4 monotherapy to L-T_4/L-T_3 combination therapy did not show any correlation between changes in bodyweight and changes in QoL [27].

A suggested explanation for a possible need for a T_3 substitution as well as T_4 has been related to a change in the feedback in the hypothalamic–pituitary–thyroid axis induced by high dose of T_4 and a possible relative inactivation of deiodinase 2 and hereby lower values of T_3. Based on this theory a suggestion of using serum T_3 as a simple maker of decreased QoL has been brought up [20, 28]. However clinical data on patients changes from L-T_4 monotherapy to L-T_4/L-T_3 combination therapy have not showed a correlation between changes in serum values of T_3 and increase in QoL [25, 27, 29].

Another possible suggested explanation is polymorphisms in the deiodinase 2 gene (D2–92 Ala) or to the cellular membrane transporter (MCT10) [30, 31] and by this a potential decrease in the transportation of thyroid hormone from plasma to the intracellular space and a decreased deiodination of T_4 to T_3. However, extended research is needed to explore the exact effect of these polymorphisms.

A request of L-T_4/L-T_3 combination therapy from patients with persistent symptoms is often brought forward with reference to patient's platforms on the social medias on the internet [32, 33].

It has been described that 5–10% of patients treated with hypothyroidism do have decreased QoL [4, 5], however, it is important to recognize that in these studies the same symptoms were present in app 15% of the controls. Therefore, it is pertinent to evaluate the specific symptoms in these patients and to discuss with the patient if any other explanations could be present; such as symptoms of menopause, stress, or depression, and to eliminate other specific disease that potentially could explain the symptoms.

The European Thyroid Association recommends in a guideline from 2012 **not** to use L-T_4/L-T_3 combination therapy as a standard treatment but suggest the following algorithm [8]:

- L-T_4 + L-T_3 combination therapy might be considered as an experimental approach in compliant L-T_4-treated hypothyroid patients who have persistent complaints despite serum TSH values within the reference range, provided they have previously given support to deal with the chronic nature of their disease and associated autoimmune diseases have been ruled out.
- T_4 + T_3 combination therapy is not recommended in pregnant women and in patients with cardiac arrhythmias.
- It is suggested that L-T_4 + L-T_3 combination therapy is discontinued if no improvement is experienced after 3 months. The L-T_4/L-T_3 ratio should be between 13:1 and 20:1 by weight (a suggestion for a calculation of the T_3 dose were made in the ETA guideline Table 3.4.6.2 [8]).

Table 3.4.6.2 A method for calculating L-T_4 and L-T_3 dosages for T_4 + T_3 combination therapy

T4 monotherapy = dose x #	Dose *x* = 100 µg L-T_4	Dose *x* = 150 µg L-T_4	Dose *x* = 200 µg L-T_4
T4 + T3 combination therapy			
L-T_3 dose *y* = *x*: 20	L-T_3 dose 5 µg	L-T_3 dose 7.5 µg	L-T_3 dose 10 µg
L-T_4 dose *z* = *x*–3y	L-T_4 dose 85 µg *	L-T_4 dose 127.5 µg *	L-T_4 dose 170 µg *
L-T_4: L-T_3 dose ratio	17: 1	17: 1	17: 1

Dose *x* = daily L-T4 dose that normalized serum TSH during T4 monotherapy; * dose could be round off to 87.5 µg, 125 µg and 175 µg, respectively, for convenience reasons (i.e. availability of L-T4 tablets of 25 µg but not of 10 µg).

Animal Thyroid Extracts

Extracts from animal thyroid glands were first used successfully in the treatment of hypothyroidism in 1892 and this was the only drug until the 1960s. The product had standardization problems and when the synthetic L-T_4 became available in the 1960s, as a more stable drug with the ability to normalize serum values of fT_4 as well as fT_3, this became the drug of choice. From time to time, thyroid extracts have enjoyed a renaissance as a 'more natural thyroid hormone product', usually among practitioners on the more 'alternative' part of medicine, and often used in a high dose inducing SCH to overt hyperthyroidism.

A single RCT is present comparing L-T_4 treatment to thyroid extract from pigs. A preference for thyroid extract was described in 49% vs. 19% in L-T_4, but no difference was found in symptom scores or neurocognitive function [34].

Risks Related to Thyroid Hormone Replacement Treatment

The potential risk when treating with thyroid hormones is overtreatment inducing overt or subclinical hyperthyroidism. Population studies have verified an association between suppressed serum TSH and the following:

- Atrial fibrillation, relative risk (RR): 1.16 (95% CI: 0.99–1.36) increasing with age to 1.41 (95% KI: 1.25–1.59) [35].
- Overall dead, RR 1.23 (95% KI: 1.16–1.30) [35].
- Osteoporosis with a hip fracture hazard ratio (HR) 1.16 (95% CI: 1.07–1.26) [36].
- Dementia, RR: 2–3 [37].

There are no studies comparing risks of suppressed serum TSH when treating with L-T_4 monotherapy, L-T_4+ L-T_3 combination therapy, animal thyroid extract, or when suppressed serum TSH are caused by Graves' disease or nodular goitre, but the effect must be expected to be a class effect of thyroid hormones and hereby a suppressed serum TSH is a potential risk whether it is caused by any thyroid hormone given or by endogenous produced thyroid hormone as in Graves' or nodular goitre.

Treatment During Pregnancy

Most patients require an increase of 30–50% daily in L-T_4 dosage within the first months of pregnancy. The principal reason for this change in L-T_4 requirement is the increase in the serum concentration of thyroxine-binding globulin in pregnancy, which results in decreased serum concentrations of fT_3 and fT_4. These decreases cannot be compensated for by increased thyroidal secretion due to lack of functioning thyroid tissue.

In hypothyroid pregnant—or patients planning to become pregnant a serum TSH <2.5 mU/L is recommended. As minor studies have described an increased risk of abortion and preterm delivery in patients with serum TSH is between 2.5 and 4 mU/L and presence of TPO-Ab treatment is also recommended in these patients [38].

Temporary Hypothyroidism

Most patients with primary hypothyroidism require lifelong L-T_4 therapy. However, hypothyroidism may be transient and even short-term treatment may be unnecessary.

Transient hypothyroidism is seen in the recovery phase of subacute (De Quervain's), painless, and postpartum thyroiditis, if excess iodine or iodine-containing drugs such as amiodarone or lithium have been applied; presence of TSH-receptor blocking antibodies; inhibition of synthesis or release of thyroid hormone by immune modulating therapy such as interferon and other cytokines; thyroiditis induces by tyrosine kinase inhibitors; blockage of receptors for vascular endothelial growth factor; and central hypothyroidism by anti-CTLA4 or anti-PD-1 antibody drugs [39].

Unnecessary Thyroid Hormone Therapy

Obesity often link to a question of a need of thyroid hormone. Weight loss after initiating L-T_4, in even severe hypothyroidism are modest, and weight reduction can only be obtained if supraphysiological doses of L-T_4 or L-T_3 are given [15].

Depression: Although a relationship between thyroid function at depression has been assumed for many years, the exact relationship is still poorly defined, and a historical suggested effect of thyroid hormone (primarily T_3) in euthyroid depressed patients are not recommended due to low evidence and risk of subclinical hyperthyroidism [40].

Non-thyroid illness: Thyroid hormone therapy in patients with low levels of thyroid hormones due to non-thyroid illness has been investigated in patients in intensive care units, and no evidence for treatment with T_4 or T_3 are present in general. A question on a possible effect in a subgroup of patients with severe prolonged illness is still questioned [41].

Heart failure: In a small study in patients New York Heart Association (NYHA) class III or heart failure demonstrated a beneficial effect from short-term intravenous L-T_3 therapy [42], however, this effect could not be confirmed in a trial of long-term T_3 therapy [43].

Unspecific symptoms: Occasionally, patients query treatment with thyroid hormone for non-specific symptoms such as tiredness—although normal values of serum TSH as well as T_4 are present. In an RCT in patients with non-specific symptoms L-T_4 100 µg did not improve cognitive function and psychological well-being [44]. In these cases, other conditions or diseases must be considered and thyroid hormone therapy cannot be recommended.

Patients with previously initiated thyroid hormone therapy based on unclear indications or with a single measurement of a slightly elevated serum TSH is recommended to assess the continued need for L-T_4, treatment by stopped for 4 weeks if the serum TSH concentration is normal, or for 6 weeks if it is undetectable, to allow recovery of the suppressed TSH. Measurement of serum T_4 and TSH concentrations at this stage will determine whether the patient is truly hypothyroid.

REFERENCES

1. Pearce SH, Brabant G, Duntas LH, *et al*. 2013 ETA Guideline: management of subclinical hypothyroidism. *Eur Thyroid J*, 2013; **2**: 215–28.
2. Andersen S, Pedersen KM, Bruun NH, Laurberg P. Narrow individual variations in serum T_4 and T_3 in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab*, 2002; **87**: 1068–72
3. Karmisholt J, Andersen S, Laurberg P. Variation in thyroid function tests in patients with stable untreated subclinical hypothyroidism. *Thyroid*, 2008; **18**: 303–8.
4. Saravanan P, Chau W-F, Roberts N, Vedhara K, Greenwood R, Dayan CM. Psychological well-being in patients on 'adequate' doses of L-thyroxine: results of a large, controlled community-based questionnaire study. *Clin Endocrinol*, 2002; **57**: 577–85.
5. Panicker V, Evans J, Bjoro T, Asvold BO, Dayan CM, Bjerkeset O. A paradoxical difference in relationship between anxiety, depression and thyroid function in subjects on and not on T4: findings from the HUNT study. *Clin Endocrinol*, 2009; **71**: 574–80.
6. Thvilum M, Brandt F, Brix TH, Hegedüs L. Hypothyroidism is a predictor of disability pension and loss of labor market income: a Danish register-based study. *J Clin Endocrinol Metab*, 2014; **99**: 3129–35.
7. Thvilum M, Brandt F, Almind D, Christensen K, Brix TH, Hegedüs L. Increased psychiatric morbidity before and after the diagnosis of hypothyroidism: a nationwide register study. *Thyroid*, 2014; **24**: 802–8.
8. Wiersinga WM, Duntas L, Fadeyev V, Nygaard B, Vanderpump MP: 2012 ETA guidelines: the use of L-T4 + L-T3 in the treatment of hypothyroidism. *Eur Thyroid J*, 2012; **1**: 55–71.
9. Yavuz DG, Yazıcı D, Keskin L, *et al*. Out-of-reference range thyroid-stimulating hormone levels in levothyroxine-treated primary hypothyroid patients: a multicenter observational study. *Front Endocrinol (Lausanne)*, 2017; **8**: 215.
10. Wouters HJ, van Loon HC, van der Klauw MM, *et al*. No effect of the Thr92Ala polymorphism of deiodinase-2 on thyroid hormone parameters, health-related quality of life, and cognitive functioning in a large population-based cohort study. *Thyroid*, 2017; **27**: 147–55.
11. Bolk N, Visser TJ, Kalsbeck A, van Domburg RT, Berghout A. Effects of evening vs morning thyroxine ingestion on serum thyroid hormone profiles in hypothyroid patients. *Clin Endocrinol*, 2007; **66**: 43–8.
12. Grebe SK, Cooke RR, Ford HC, *et al*. Treatment of hypothyroidism with once weekly thyroxine. *J Clin Endocrinol Metab*, 1997; **82**: 870–5.
13. Rajput R, Pathak V. The effect of daily versus weekly levothyroxine replacement on thyroid function test in hypothyroid patients at a tertiary care centre in Haryana. *Eur Thyroid J*, 2017; **6**: 250–4
14. Brachman T, Sperlbaum A, Schmiegel W. Levothyroxine therapy and impaired clearance are the strongest contributors to small intestinal bacterial overgrowth: results of a retrospective cohort study. *World J Gastroenterol*, 2017; **23**: 842–52.
15. Jonklaas J, Bianco AC, Bauer AJ, *et al*. Guidelines for the treatment of hypothyroidism: prepared by the American thyroid association task force on thyroid hormone replacement. *Thyroid*, 2014; **24**: 1670–751.
16. Fallahi P, Ferrari SM, Ruffilli I, *et al*. Advancements in the treatment of hypothyroidism with L-T4 liquid formulation or soft gel capsule: an update. *Expert Opin Drug Deliv*, 2017; **14**: 647–55.
17. Samuels MH, Kolobova I, Niederhausen M, Janowsky JS, Schuff KG. Effects of altering levothyroxine (L-T4) doses on quality of life, mood, and cognition in L-T4 treated subjects. *J Clin Endocrinol Metab*, 2018; **103**: 1997–2008.
18. Bhagat S, Soldin SJ. Triiodothyronine levels in athyreotic individuals during levothyroxine therapy. *JAMA*, 2008; **299**: 817–19.
19. Gullo D, Latina A, Frasca F, Le Moli R, Pellegriti G, Vigneri R. Levothyroxine monotherapy cannot guarantee euthyroidism in all athyreotic patients. *PLoS One*, 2011; **6**: e22552.
20. Werneck de Castro JP, Fonseca TL, Ueta CB, *et al*. Differences in hypothalamic type 2 deiodinase ubiquitination explain localized sensitivity to thyroxine. *J Clin Invest*, 2015; **125**: 769–81.
21. Escobar-Morreale HF, Escobar del Ray F, Obregon MJ, Morreale de Escobar G. Only the combined treatment with thyroxine and triiodothyronine ensures euthyroidism in all tissues of the thyroidectomized rat. *Endocrinology*, 1996; **137**: 2490–502.
22. Pilo A, Iervasi G, Vitek F, Ferdeghini M, Cazzuola F, Bianchi R. Thyroidal and peripheral production of 3,5,3'-triiodothyronine in humans by multicompartmental analysis. *Am J Physiol*, 1990; **258**(4 Pt 1): E715–26.
23. Bunevicius R, Kazanavicius G, Zalinkevicius R, Prange AJ Jr. Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *N Engl J Med*, 1999; **340**: 424–9.
24. Grozinsky-Glasberg S, Fraser A, Nahshoni E, Weizman A, Leibovici L. Thyroxine-triiodothyronine combination therapy versus thyroxine monotherapy for clinical hypothyroidism: meta-analysis of randomized controlled trials. *J Clin Endocrinol Metab*, 2006; **91**: 2592–9.
25. Nygaard B, Jensen EW, Kvetny J, Jarlov A, Faber J. Effect of combination therapy with thyroxine (T4) and 3,5,3'-triiodothyronine versus T4 monotherapy in patients with hypothyroidism, a double-blind, randomised cross-over study. *Eur J Endocrinol*, 2009; **161**: 895–902.

26. Wiersinga WM. Therapy of endocrine disease: T4 + T3 combination therapy: is there a true effect? *Eur J Endocrinol*, 2017; **177**: R287–96.
27. Michaelsson LF, la Cour JL, Medici BB, Watt T, Faber J, Nygaard B. Levothyroxine/liothyronine combination therapy and quality of life: is it all about weight loss? *Eur Thyroid J*, 2018; **7**: 243–50.
28. Hoermann R, Midgley JEM, Larisch R, Dietrich JW. Relational stability in the expression of normality, variation, and control of thyroid function. *Front Endocrinol (Lausanne)*, 2016; **7**: 142.
29. Medici BB, la Cour JL, Michaelsson LF, Faber JO, Nygaard B. Neither baseline nor changes in serum triiodothyronine during levothyroxine/liothyronine combination therapy predict a positive response to this treatment modality in hypothyroid patients with persistent symptoms. *Eur Thyroid J*, 2017; **6**: 89–93.
30. Panicker V, Saravanan P, Vaidya B, *et al*. Common variation in the DIO2 gene predicts baseline psychological well-being and response to combination thyroxine plus triiodothyronine therapy in hypothyroid patients. *J Clin Endocrinol Metab*, 2009; **94**: 1623–9.
31. Carlé A, Laurberg P, Faber J, Steffensen R, Nygaard B. Hypothyroid patients encoding combined MCT10 and DIO2 gene polymorphisms may prefer T3+T4 combination treatment—a blinded randomized clinical study. *Eur Thyroid J*, 2017; **6**: 143–51.
32. Michaelsson LF, Medici BB, la Cour JL, *et al*. Treating hypothyroidism with thyroxine/triiodothyronine combination therapy in Denmark: following guidelines or following trends? *Eur Thyroid J*, 2015; **4**: 174–80.
33. Dayan C, Panicker V. Management of hypothyroidism with combination thyroxine (T_4) and triiodothyronine (T_3) hormone replacement in clinical practice: a review of suggested guidance. *Thyroid Res*, 2018: **11**: 1.
34. Hoang TD, Olsen CH, Mai VQ, Clyde PW, Shakir MK. Desiccated thyroid extract compared with levothyroxine in the treatment of hypothyroidism: a randomized, double-blind, crossover study. *J Clin Endocrinol Metab*, 2013; **98**: 1982–90.
35. Selmer C1, Olesen JB, von Kappelgaard LM, *et al*. Subclinical and overt thyroid dysfunction and risk of all-cause mortality and cardiovascular events: a large population study. *J Clin Endocrinol Metab*, 2014; **99**: 2372–82.
36. Abrahamsen B, Jørgensen HL, Laulund AS, Nybo M, Brix TH, Hegedüs L. Low serum thyrotropin level and duration of suppression as a predictor of major osteoporotic fractures-the OPENTHYRO register cohort. *J Bone Miner Res*, 2014; **29**: 2040–50.
37. Kalmijn S, Mehta KM, Pols HA, Hofman A, Drexhage HA, Breteler MM. Subclinical hyperthyroidism and the risk of dementia. The Rotterdam Study. *Clin Endocrinol (Oxf)*, 2000; **53**: 733–7.
38. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
39. Rizzo LFL, Mana DL, Serra HA. Drug-induced hypothyroidism. *Medicina (B Aires)*, 2017; **77**: 394–404.
40. Dayan CM, Panicker V. Hypothyroidism and depression. *Eur Thyroid J*, 2013; **2**: 168–79.
41. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. *Thyroid*, 2014; **24**: 1456–65.
42. Pingitore A, Galli E, Barison A, *et al*. Acute effects of triiodothyronine (T3) replacement therapy in patients with chronic heart failure and low-T3 syndrome: a randomized, placebo-controlled study. *J Clin Endocrinol Metab*, 2008; **93**: 1351–8.
43. Holmager P, Schmidt U, Mark P, *et al*. Long-term L-Triiodothyronine (T3) treatment in stable systolic heart failure patients: a randomized, double-blind, cross-over, placebo-controlled intervention study. *Clin Endocrinol (Oxf)*, 2015; **83**: 931–7.
44. Pollock MA, Sturrock A, Marshall K, *et al*. Thyroxine treatment in patients with symptoms of hypothyroidism but thyroid function tests within the reference range: randomized double-blind placebo-controlled crossover trial. *BMJ*, 2001; **323**: 891–5.

3.5

Thyroid Lumps

3.5.1 Pathogenesis of Non-Toxic Goitre

Dagmar Führer and Holger Jäschke

Definition

Nodular goitre can be divided into solitary nodular and multinodular thyroid disease and constitutes a complex thyroid disorder with heterogeneous, morphological, functional and pathogenetic properties [1]. Histologically, thyroid nodules are distinguished by morphological criteria according to the World Health Organization (WHO) classification [2]. On functional grounds, nodules are classified as either 'cold', 'normal', or 'hot' depending whether they show decreased, normal, or increased uptake of radioactive nuclides such as technetium on scintiscan. In contrast to solitary nodular thyroid disease, which has a more uniform clinical, pathological, and molecular picture, multinodular goitre (MNG) usually comprises a mixed group of nodular entities (i.e. one usually finds a combination of hyperfunctional, hypofunctional, or normally functioning thyroid lesions within the same thyroid gland). The overall balance of functional properties of individual thyroid nodules within an MNG ultimately determines the functional status in the individual patient, which may be euthyroid, subclinical hyperthyroid, or overt hyperthyroid. On the molecular level, thyroid nodules within goitre may represent polyclonal lesions or true monoclonal thyroid neoplasia.

The Role of Environmental Factors

The development of nodular goitre is influenced by extrinsic factors interacting with intrinsic factors and constitutional parameters of gender and age [1, 2].

The most important trigger for nodular (and diffuse) goitre is iodine deficiency [3]. There is a direct correlation between goitre prevalence and iodine deficiency and *vice versa* between correction of iodine deficiency and regression of goitre incidence. This indicates that iodine supply or intake has to be adjusted adequately to reduce the burden of thyroid goitre development in a population.

Iodine deficiency was common in Germany until the early 1990s. Introduction of iodized salt into food industries resulted in a marked improvement in nutritional iodine supply as reflected by an increased urinary iodine excretion (median 72 µg iodine/L urine in 1994 to 125 µg iodine/L urine in 2003). Also, the percentage of private households covered with sufficient iodized salt increased from less than 10% to 66% between 1990 and 2002 [4].

In 1994 the prevalence of diffuse goitre was 21% in the age group of 18–30 years and 33% in the age group of 46–65 years. In the Papillon study of 2002 investigating 96 000 German employees an impressive reduction in goitre frequency to 6% in the group of 18–30-year-olds and to 26% in the group of 46–65-year-old participants was found [5]. The epidemiological study SHIP-0, has underscored that with a decrease in overall-goitre prevalence due to improved iodine supply, thyroid nodules now tend to occur in normal size rather than enlarged thyroid glands [6]. This may be explained by the thyroid inherent disposition to develop focal hyperplasia as discussed next.

Table 3.5.1.1 Gender and age dependent upper reference values for normal thyroid volume

Men	25 ml
Women	18 ml
13–14 yr	8–10 ml
3–4 yr	3 ml
Newborn	0.8–1.5 ml

Table 3.5.1.2 Histological diagnosis of goitre

Pathology	Thyroid function	Comment
Diffuse goitre	Euthyroid	
Nodular goitre (uni- or multinodular)	Euthyroid-hyperthyroid	May concur with autoimmune disease in regions with endemic goitre
Thyroid cancer	Euthyroid	May be present in an NTG and/or autoimmune disease
Autoimmune thyroid disease	Euthyroid, hypo-, or hyperthyroid	
Thyroiditis	Variable	
Defect in thyroid hormone synthesis	Hypothyroid-euthyroid	
Thyroid hormone resistance	Euthyroid	End organ dependent function
Acromegaly	Euthyroid	IGF-1 dependent
Drugs (Lithium, antithyroid drugs)	Euthyroid-hypothyroid	
TSHoma	Hyperthyroid	TSH dependent

Recently, two German studies in children (KiGGS) [7] and adults (DEGS) [8] showed results that the iodine status may not be sufficient. In both surveys about 25–30% of the individuals were below the estimated average requirement for iodine. This was also supported by comparing results from the SHIP-0 (1997–2001) [6] and SHIP-TREND (2008–2012) [9] study in an adult population of Northeast Germany. The median urinary excretion levels significantly decreased from 123µg/L to 112µg/L between 2000 and 2010. These data clearly indicated that informative advertising of iodine prophylaxis needs to be intensified to maintain and still to improve iodine status.

Various other goitrogenic factors are known and are relevant to thyroid disease in situations with coexisting iodine deficiency. First, metabolites of various nutrients (e.g. cabbage, cauliflower, and broccoli) may interfere with iodine uptake. Second, industrial pollutants, including resorcinol and phthalic acid, are known to be goitrogenic. Third, deficiencies of selenium, iron, and vitamin A may exacerbate the pathogenic effects of iodine deficiency [10].

Other risk factors for nodular goitre have been suggested, whereby a possibly distinct impact on the prevalence of thyroid nodules occurring in a normal sized or enlarged thyroid gland is less clear [1, 2]. Smoking has been proposed as a risk factor for goitre and nodules were also found with higher prevalence in goitres of smokers compared with non-smokers. The impact of smoking on thyroid disease is most likely due to increased thiocyanate levels in smokers exerting a competitive inhibitory effect on iodide uptake. In line with this, the association is more pronounced in areas with iodine deficiency [11].

Radiation is another environmental risk factor not only for thyroid malignancy but also for benign nodular thyroid disease. An increased prevalence of thyroid nodules disease has been associated with exposure to radio nuclear fallouts and therapeutic external radiation.

Nodular thyroid disease and goitre are more frequent (2.5- to 7-fold) in women and it seems that this prevalence appears after puberty. The determined gender difference is more pronounced in areas of long-standing iodine deficiency compared with regions of iodine sufficiency. However, the precise reasons for this observation remain to be clarified [12]. A growth promoting effect of oestrogens has been described *in vitro* and oestradiol has been suggested to amplify growth factor-dependent signalling in normal thyroid cells and thyroid tumours [11]. However, pregnancy-related thyroid enlargement appears to be mostly related to iodine deficiency and in one German study [13] increased MNG prevalence with parity was only observed in women who had not taken iodine supplementation during an earlier pregnancy.

Several studies suggest that thyroid volume is also significantly correlated with body mass index. In agreement with this, a recent study has shown that in obese women, weight loss of more than 10% may result in a significant decrease in thyroid volume [14]. Lastly, because of the cumulative impact of external risk factors on the thyroid gland, the prevalence of thyroid nodular disease increases with age.

For example, a study (DanThyr) from a borderline iodine-deficient area in Denmark demonstrated the relationship of nodular goitre development in ageing women [15]. The survey of 4242 women showed a goitre incidence of 18% with participants aged 20–39 years exhibiting higher prevalence for solitary thyroid nodules than for MNG. The prevalence of these two entities is similar in the fourth decade of life but shifts to MNG after 50 years of age.

Genetic Disposition

Thyroid nodules (and goitre) also occur in individuals without exposure to iodine deficiency, and not all individuals in an iodine-deficient region develop goitre. A familial clustering for nodular goitre is well documented and family and twin pair studies in endemic and non-endemic goitre regions have underscored a genetic predisposition for goitre development [1, 2]. For example, twin studies show a concordance rate of 80% for monozygotic twins and of 42% for dizygotic twins in endemic and of 40–50% and 13% in non-endemic regions, respectively, strongly suggesting interplay between genetic and environmental factors. On the basis of twin studies, the contribution of genetic susceptibility to goitre

development has been calculated to be 39% in endemic regions and 82% in a non-endemic area [10].

Genetic defects in enzymes involved in thyroid hormone synthesis (e.g. thyroglobulin (TG), thyroperoxidase (TPO), sodium-iodide symporter (NIS) typically result in hypothyroid goitres, but in some rare cases genetic variations in the TG, TPO, and NIS gene have also been reported in association with (diffuse or) nodular euthyroid goitre. Furthermore, alterations in the pendrin gene account for the syndromic occurrence of euthyroid nodular goitre and congenital sensorineural hearing loss.

Since these monogenetic defects are exceptionally rare, linkage studies have been performed to identify susceptibility loci for non-toxic goitre (NTG) on a broader scale [16]. A locus on chromosome 14 (termed MNG-1 locus) has been identified in a Canadian and a German study and was found to cosegregate with familial NTG. In an Italian pedigree with euthyroid goitre an X-linked autosomal pattern of inheritance with a putative genetic defect in the Xp22 region was suggested (termed MNG-2 locus) [17]. Moreover, in a study by the European Thyroid Association working group on the 'Genetics of euthyroid goitre' 18 extended Danish, German, and Slovakian families were analysed in a genome-wide scan. Further putative candidate loci for NTG were identified on chromosomes 3p, 2q, 7q, and 8p emphasizing the genetic heterogeneity of euthyroid goitre.

The first association of germline *DICER1* mutations with NTG was shown 2011. DICER1 is an important enzyme in the maturation of microribonucleic acid (RNA), which have substantial impact in gene expression at the posttranscriptional level. The *DICER1* gene is located in the MNG-1 locus and loss-of-function mutations were identified in 33 members of five families [18]. Besides NTG, mutations in *DICER1* are associated with several benign and malignant tumours (e.g. pleuropulmonary blastoma, cystic nephroma, and ovarian Sertoli-Leydig cell tumour) [19]. The incidence of multinodular goitre in individuals with *DICER1* mutations by age 20 years is 32% in females and 13% in males and also the risk of developing thyroid cancer is increased [20]. Thus, for the majority of euthyroid goitres, a complex, multifactorial pathogenesis including interactions between various environmental factors, gender-specific components, and the genetic background has to be assumed.

Molecular Processes Involved in Nodule Formation

Development of nodular goitre most likely proceeds in two phases, that involve global activation of thyroid epithelial cell proliferation (e.g. as the result of iodine deficiency or other goitrogenic stimuli) leading to hyperplasia and a focal increase of thyroid epithelial cell proliferation causing thyroid nodules. So far, the most common stimulus for focal proliferation is a somatic mutation.

Two driving pathogenetic events have to be considered (Figure 3.5.1.1) [1]: first, iodine deficiency causing an increase in thyroid cell numbers (true hyperplasia) as observed in animal models. Second, H_2O_2 production and free radical formation, which occurs physiologically during thyroid hormone synthesis, may damage genomic DNA. Thus in a mouse model, the spontaneous mutation rate in the naive thyroid gland has been found to be almost ten times higher than in other organs [21, 22].

Both processes provide a mutagenic milieu, in which the likelihood of somatic mutations is increased. Whether these somatic mutations lead to thyroid nodular disease critically depends on the affected gene and most likely the environmental selection factors (e.g. iodine deficiency; Figure 3.5.1.2). A proof of principle for this concept is the evolution of a toxic adenoma from a somatic thyroid-stimulating hormone (TSH) receptor mutation ([1]; also see Chapter 3.3.11). Other examples include the origin of papillary thyroid cancer based on *BRAF* mutations or RET/PTC rearrangements. These somatic mutations have been found already in microscopic lesions of thyroid autonomy and papillary microcarcinoma,

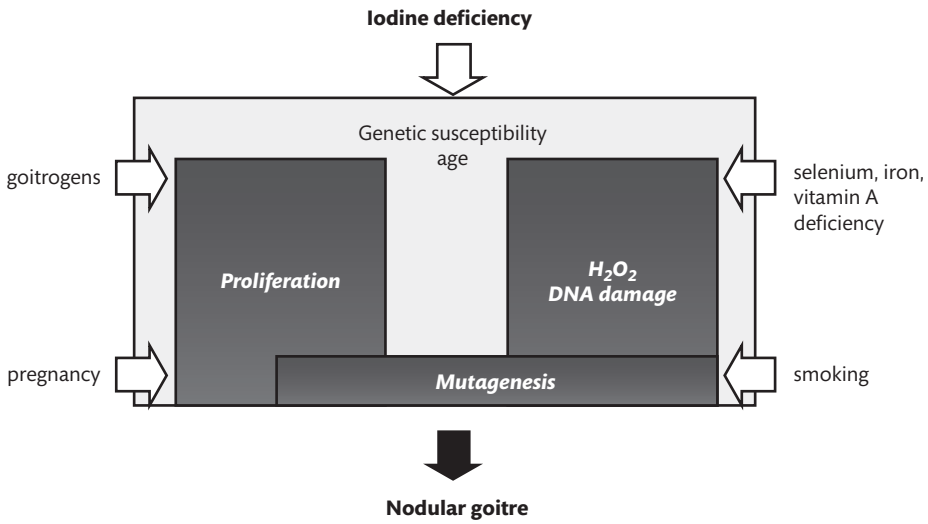


Figure 3.5.1.1 Interaction of extrinsic and intrinsic factors contributing to the development of nodular goitre.

Note that the pathogenic influence of several goitrogenous components (e.g. selenium deficiency, pregnancy) will be aggravated with coexisting iodine deficiency. The two elementary molecular pathomechanism are increased cell proliferation, leading to hyperplasia/goitre and in addition with oxidative stress leading to increased mutagenesis and nodule formation.

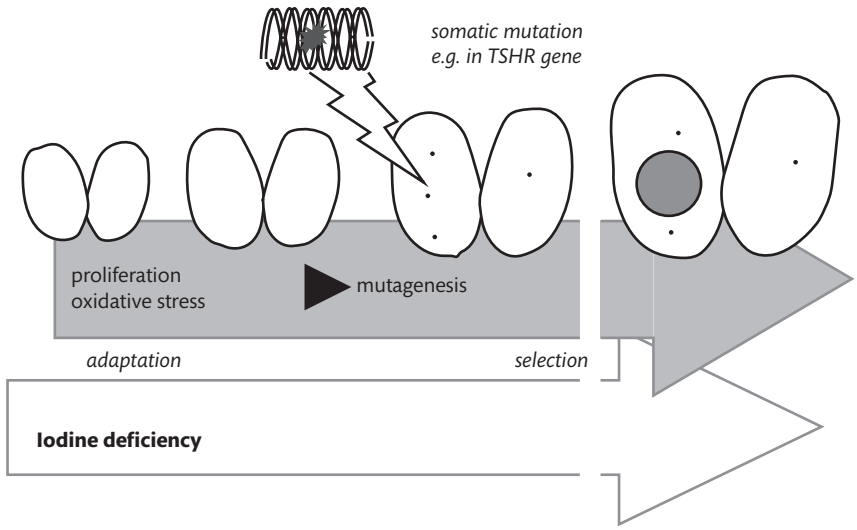


Figure 3.5.1.2 Pathogenesis of nodular goitre in an iodide deficiency environment. According to current concepts the development of nodular goitre proceeds in two phases, that involve: (1) adaptive increase in thyroid epithelial cell proliferation and function, providing a mutagenic milieu with increased likelihood for occurrence of somatic mutations; and (2) clone expansion to a macroscopic thyroid nodule by growth advantage of cell clone with somatic mutation and propagation in persisting iodine deficiency.

respectively. Besides the driving mutation, increased growth factor production and auto- and paracrine action of secreted growth factors (e.g. IGF-1) has been found in monoclonal thyroid tumours and may further propel nodule development.

The development of polyclonal thyroid lesions in an NTG is less clear and putatively is linked to exogenous factors (e.g. intrathyroidal production of growth factors such as IGF-1), which act on the naturally functional and morphological heterogeneous thyroid follicles [11].

Natural Course of Disease

From the epidemiological data discussed earlier, one might expect an inherent progressive course of nodular thyroid disease. Studies aimed at accurate assessment of the nodules by ultrasonography differ in terms of follow-up period, definition of growth, type of thyroid lesion and the background, in which they are conducted. Moreover, the interobserver variability of long-term studies of nodule volumes is not known. With these caveats in mind, the following observations have been reported [1, 2]: in iodine-sufficient areas, nodule 'growth' has been reported in 35% of US patients over a follow-up period of 4.9 to 5.6 years. On long-term follow-up over 15 years in an area of iodine sufficiency, only one-third of benign nodules showed growth as assessed by palpation and ultrasonography compared with the majority of nodules, which remained unchanged or even showed a decrease in size. In Germany, a mean 3-year follow-up of 109 consecutive patients showed a steady and significant (30% volume) increase in nodular size in 50% of patients. In a Danish study, only four (8%) of 45 cold nodules in an area of borderline iodine deficiency showed a change in size (5 mm in diameter), of which only one nodule actually increased and three nodules shrank over a follow-up period of 2 years [23]. Nodule growth was also determined in a 5-year follow-up study from two areas in Italy with mild to moderate iodine deficiency [24]. Significant nodule growth occurred in 153 of 992 patients (15.3%), shrinkage was observed in 13.1% of cases whereas the majority of nodules (69%) exhibited no change in size over the investigated time period. However, nodule growth was significantly linked to patients with multinodular glands [24]. Thus, in an iodine-deficient and -sufficient setting a variable portion, but most likely not all, of nodules will grow, and the speed of growth is highly heterogeneous.

REFERENCES

1. Hegedus L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev*, 2003; **24**: 102–32.
2. Krohn K, Führer D, Bayer Y, *et al*. Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev*, 2005; **26**: 504–24.
3. Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol*, 201;5 **3**: 286–95.
4. Andersson M, de Benoist B, Rogers L. Epidemiology of iodine deficiency: salt iodisation and iodine status. *Best Pract Res Clin Endocrinol Metab*, 2010; **24**: 1–11.
5. Reiners C, Wegscheider K, Schicha H, *et al*. Prevalence of thyroid disorders in the working population of Germany: ultrasonography screening in 96,278 unselected employees. *Thyroid*, 2004; **14**: 926–32.
6. Volzke H, Ludemann J, Robinson DM, *et al*. The prevalence of undiagnosed thyroid disorders in a previously iodine-deficient area. *Thyroid*, 2003; **13**: 803–10.
7. Johner SA, Thamm M, Stehle P, *et al*. Interrelations between thyrotropin levels and iodine status in thyroid-healthy children. *Thyroid*, 2014; **24**: 1071–9.

8. Johner SA, Thamm M, Schmitz R, Remer T. Examination of iodine status in the German population: an example for methodological pitfalls of the current approach of iodine status assessment. *Eur J Nutr*, 2016; **55**: 1275–82.
9. Khattak RM, Ittermann T, Nauck M, Below H, Volzke H. Monitoring the prevalence of thyroid disorders in the adult population of Northeast Germany. *Population Health Metrics*, 2016; **14**: 39.
10. Hansen PS, Brix TH, Bennedbaek FN, Bonnema SJ, Kyvik KO, Hegedus L. Genetic and environmental causes of individual differences in thyroid size: a study of healthy Danish twins. *J Clin Endocrinol Metab*, 2004; **89**: 2071–7.
11. Knudsen N, Laurberg P, Perrild H, Bulow I, Ovesen L, Jorgensen T. Risk factors for goiter and thyroid nodules. Thyroid, 2002; **12**: 879–88.
12. Malboosbaf R, Hosseinpanah F, Mojarrad M, Jambarsang S, Azizi F. Relationship between goiter and gender: a systematic review and meta-analysis. *Endocrine*, 2013; **43**: 539–47.
13. Struve CW, Haupt S, Ohlen S. Influence of frequency of previous pregnancies on the prevalence of thyroid nodules in women without clinical evidence of thyroid disease. *Thyroid*, 1993; **3**: 7–9.
14. Sari R, Balci MK, Altunbas H, Karayalcin U. The effect of body weight and weight loss on thyroid volume and function in obese women. *Clin Endocrinol (Oxf)*, 2003; **59**: 258–62.
15. Carle A, Krejbjerg A, Laurberg P. Epidemiology of nodular goitre. Influence of iodine intake. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 465–79.
16. Bayer Y, Neumann S, Meyer B, *et al*. Genome-wide linkage analysis reveals evidence for four new susceptibility loci for familial euthyroid goiter. *J Clin Endocrinol Metab*, 2004; **89**: 4044–52.
17. Capon F, Tacconelli A, Giardina E, *et al*. Mapping a dominant form of multinodular goiter to chromosome Xp22. *Am J Human Genet*, 2000; **67**: 1004–7.
18. Rio Frio T, Bahubeshi A, Kanellopoulou C, *et al*. DICER1 mutations in familial multinodular goiter with and without ovarian Sertoli-Leydig cell tumors. *JAMA*, 2011; **305**: 68–77.
19. Schultz KAP, Rednam SP, Kamihara J, *et al*. PTEN, DICER1, FH, and their associated tumor susceptibility syndromes: clinical features, genetics, and surveillance recommendations in childhood. *Clin Cancer Res*, 2017; **23**: e76–82.
20. Khan NE, Bauer AJ, Schultz KAP, *et al*. Quantification of thyroid cancer and multinodular goiter risk in the DICER1 syndrome: a family-based cohort study. *J Clin Endocrinol Metab*, 2017; **102**: 1614–22.
21. Krohn K, Maier J, Paschke R. Mechanisms of disease: hydrogen peroxide, DNA damage and mutagenesis in the development of thyroid tumors. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 713–20.
22. Song Y, Driessens N, Costa M, *et al*. Roles of hydrogen peroxide in thyroid physiology and disease. *J Clin Endocrinol Metab*, 2007; **92**: 3764–73.
23. Knudsen N, Perrild H, Christiansen E, Rasmussen S, Dige-Petersen H, Jorgensen T. Thyroid structure and size and two-year follow-up of solitary cold thyroid nodules in an unselected population with borderline iodine deficiency. *Eur J Endocrinol*, 2000; **142**: 224–30.
24. Durante C, Costante G, Lucisano G, *et al*. The natural history of benign thyroid nodules. *JAMA*, 2015; **313**: 926–35.

3.5.2 Management of Non-Toxic Multinodular Goitre

Hans Graf and Gilberto Paz-Filho

Introduction

Multinodular goitre (MNG) is characterized by the progressive nodular enlargement of the thyroid gland, without underlying inflammation, autoimmune thyroid disease, and malignancy. It is the result of the interaction between genetic susceptibility, hormonal, and environmental aspects, being iodine deficiency the most important factor. Patients with MNG usually present multinodular enlarged thyroid glands, and its natural history is characterized initially with a diffuse thyroid growth, followed by nodule formation, and frequently with the progression to nodule autonomy, subclinical hyperthyroidism, and eventually overt hyperthyroidism denominated toxic MNG (TMNG). In TMNG, hyperfunctioning nodules secrete thyroid hormones independent of TSH stimulation [1].

The clinical features of MNG varies from an asymptomatic patient to a critical patient with upper airway compression and respiratory insufficiency. The diagnostic work-up includes thyroid function tests, imaging studies such as thyroid ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI), cytological evaluation of nodular samples obtained through US guided fine-needle aspiration (US-FNAB), and assessment of radioactive iodine uptake (RAIU) [2].

The management of a patient with MNG is based upon the clinical presentation and the patient´s preference, which involves surgery or radioiodine therapy in most cases. Radioactive iodine (^{131}I) is a safe and effective therapy introduced more than three decades ago for the treatment of MNG, leading to significant thyroid volume

(TV) reduction. Its efficacy can be enhanced by strategies leading to the elevation of TSH levels [3].

Clinical Manifestations

The clinical manifestations observed in patients with MNG are very heterogeneous. Patients with smaller goitres and with normal thyroid function are usually asymptomatic. However, the gradual increase in size may determine compressive symptoms, such as cough, respiratory distress, difficulty in swallowing, and the feeling of a lump in the throat. Asymptomatic subclinical hyperthyroidism or overt hyperthyroidism may develop, particularly in older patients, as a result from the overproduction of thyroid hormones due to the onset of thyroid autonomy [4].

In young patients with small and diffuse goitres, with normal thyroid function, observation alone can be recommended. Conversely, patients with large goitres may develop compressive symptoms, determined not only by the size of the goitre, but also by its possible substernal extension. The most widely used classification defines substernal goitre as the one that has >50% of the total bulk of the thyroid gland into the mediastinum. The prevalence of intrathoracic goitres, with substernal or mediastinal extension, ranges between 2.6% and 30.4% [5, 6]. A maximal tracheal diameter of <8–10 mm can cause exertional dyspnoea, and when the diameter is smaller than 5 mm, dyspnoea at rest and stridor may occur [7].

Patients with large goitres may present the Pemberton's sign, the reversible facial congestion triggered by the elevation of both arms. In those patients, that movement leads to dislocation of the goitre into the upper thoracic inlet, compressing the subclavian and jugular vein [8]. Airway compression may occur if there is a sudden increase in TV due to thyroid haemorrhage or upper respiratory tract infections contributing to a decrease in tracheal lumen [9]. The close relationship of the thyroid gland with the recurrent laryngeal nerves may determine unilateral or bilateral vocal cord paralysis. Similarly, compression of the cervical sympathetic chain may elicit paralysis of the phrenic nerve, which can be asymptomatic or present with dyspnoea, or Horner's syndrome, characterized by ptosis, miosis and decreased sweating of the face on the same side [10].

As the MNG progresses, the increase of TV and nodularity leads to subclinical or overt hyperthyroidism in many cases, particularly in older patients, as a consequence of the autonomous overproduction of thyroid hormones [11].

Diagnostic Evaluation

Anamnesis and Physical Examination

Detailed anamnesis and physical examination are essential in the diagnostic evaluation of a patient with MNG. One source of concern is the presence of malignancy, which incidence is the same in uninodular goitre [12]. However, a systematic review and meta-analysis suggests a slightly higher risk of malignancy in solitary nodules, compared to MNG [13]. Features suggesting benign disease include factors such as family history of benign thyroid nodule or goitre and symptoms of hypothyroidism or hyperthyroidism [1].

In the physical examination, smaller goitres may be visible only when the patient is asked to swallow or drink a glass of water. Younger patients tend to have smaller goitres, with few nodules, and without intrathoracic extension. Malignant nodules are usually firm, irregular, and fixed to adjacent tissues. Fortunately, thyroid nodules are generally benign colloid nodules, just around 5% of nodules are carcinomas [12]. Goitres with intrathoracic extension cannot be fully evaluated by palpation and their presence can be identified through the Pemberton's manoeuvre.

Laboratorial Investigation

The initial laboratory investigation of MNG is based on the measurement of serum thyrotropin (TSH). Normal TSH levels indicate euthyroidism, and further laboratory investigation is not necessary [11]. Abnormally low TSH levels suggest thyroid hyperfunction, and must be followed by measurements of serum free T_4 to assess the presence of subclinical or overt hyperthyroidism. T_3 measurements may be useful in situations of difficult interpretation, such as T_3 thyrotoxicosis. A recent panel of the American Thyroid Association (ATA) could not recommend either for or against routine measurement of serum calcitonin in patients with thyroid nodules [12]. In MNG, neither calcitonin or thyroglobulin measurements are recommended, and if measured, calcitonin levels must be interpreted with caution since it has low positive predictive value and can lead to unnecessary surgery. To exclude autoimmune hyperthyroidism or hypothyroidism it is valuable to measure thyroid antibodies as thyroid-stimulating antibodies, antithyroglobulin (ATG) and antithyroid peroxidase antibodies (ATPO).

Imaging

Thyroid US should be the first fundamental imaging test [12, 14]. It is relatively inexpensive, easy to perform, and can be used to guide fine-needle aspiration biopsies (FNAB-US). It provides an estimative of TV and characterizes suspicious thyroid nodules. US may assess not only the baseline characteristics of the thyroid nodules, but is also used to follow those nodules [1, 14].

However, the usefulness of US is very limited in cases of intrathoracic MNG, since the US waves cannot penetrate bone. In these cases, CT or MRI are invaluable tools that fully characterize TV, extension, and compressive effects over the trachea. In addition, CT and MRI can determine the cross-sectional area of the trachea, a useful measure of tracheal compression. When performing these diagnostic imaging tests, it is crucial not to administer iodinated contrast agents, which impair future therapeutic or diagnostic approaches with radioiodine, and can also induce thyrotoxicosis.

CT scan is the preoperative gold-standard exam for intrathoracic goitres, and can be used to categorize goitres into three grades of retrosternal extension: above the aortic arch, between the aortic arch and the pericardium, and below the right atrium. Both CT and MRI allow the estimation of planimetric volumes, especially useful in cases of irregularly enlarged goitres [2, 15] (Figure 3.5.2.1).

Positron emission tomography (PET) with 18-fluoro-deoxyglucose (FDG) is not useful in for the diagnostic work-up, as it does not

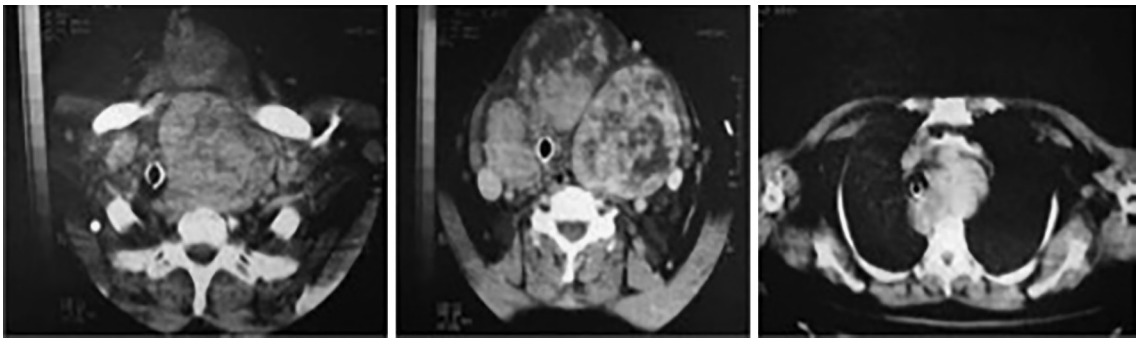


Figure 3.5.2.1 CT scan of a 13 × 15 × 19 cm substernal goitre causing tracheal deviation and acute respiratory failure. Morbid obesity and upper respiratory tract infection were contributing factors for the acute respiratory failure.

Adapted with permission from Ioannidis O, Dalampini E, Chatzopoulos S, *et al*. Acute respiratory failure caused by neglected giant substernal nontoxic goiter. *Arq Bras Endocrinol Metabol* 2011;55:229–32. Copyright © 2011 Ioannidis O, et al.

allow the visualization of the normal thyroid gland, and only shows hypermetabolic thyroid uptake that may be malignant.

Thyroid scintigraphy and RAIU measurements are not recommended in the initial diagnostic evaluation, but they are important for the differential diagnosis of hyperthyroidism in the presence of nodularity. Thyroid scintigraphy also documents the presence of intrathoracic extension of large goitres, and confirms whether intrathoracic masses identified by CT or MRI are of thyroid origin. Goitres with low and heterogeneous uptake are less responsive to radioiodine therapy, and require higher activities of the radioisotope. The use of thyroid scintigraphy and uptake studies was questioned in a study that showed that those tests do not influence diagnosis or treatment outcomes in most cases of hyperthyroidism [16], but it is very important in defining radioactive iodine treatment, especially in patients receiving recombinant human TSH or methimazole for RAIU augmentation [3].

Fine-Needle Aspiration Biopsy (FNAB)

Patients with toxic and non-toxic MNG are at the same risk of malignancy as in patients with single nodules [17]. The recommendations for FNAB in patients with MNG are the same as those for patients with single nodules: the dominant nodule, as well as any sonographically suspicious nodules, should be biopsied [12, 18]. Nodules that are suspicious for malignancy are hypoechoic, have irregular margins, do not have a sonolucent halo, and have intranodular vascularity and microcalcifications [19]. Since the fine-needle capillary (FNC) and the fine-needle aspiration (FNA) techniques provide similar sample adequacy and diagnostic accuracy, either technique can be employed, depending on the operator's personal preferences and experience [20].

Cytopathologic reports should follow the classification proposed at the National Cancer Institute Thyroid Fine-Needle Aspiration State-of-the-Science Conference in Bethesda [21].

Pulmonary Function Tests

Large MNG can lead to tracheal compression and pulmonary function impairment, but this problem is probably overlooked by many clinicians. Pulmonary function studies must include flow-volume loop tracings, which show a reduction of inspiratory capacity. Abnormal results are found even in asymptomatic patients, who can benefit from this evaluation. Improvements in pulmonary function after surgery or radioiodine therapy can be objectively quantified with pulmonary function tests [2].

Management

The management of patients with multinodular goitre is guided by the clinical presentation, professional experience and by the patients' preference: expectant clinical observation, surgery, or radioiodine therapy. Iodine supplementation and suppressive therapy with levothyroxine are not recommended therapeutic options. Nevertheless, surveys involving European [22], American [23], and Latin-American [24] endocrinologists showed that there is no consensus in the management of non-toxic multinodular goitre.

Clinical Observation

Clinical observation is a reasonable approach in the management of patients with non-toxic and asymptomatic benign MNG that do not raise any cosmetic concerns from the patient. In the natural history of many multinodular goitres, thyroid growth is extremely slow, with no significant change in a mean TV over 5 years [25].

For thyroid nodules, the American Thyroid Association (ATA) suggest that monitoring should be conducted according to risk stratification: nodules with high suspicion, low to intermediate suspicion, and low suspicion sonographic pattern should be evaluated every 12, 12–24, and ≥24 months, respectively [12]. If a nodule has been submitted to a repeated US-FNAB with a second benign cytology result, US surveillance for malignancy is no longer indicated [12]. In terms of thyroid dysfunction, a retrospective population study conducted in Scotland found that the percentages of patients with untreated subclinical hyperthyroid patients who remained subclinically hyperthyroid was 81.8% at 2 years, 67.5% at 5 years, and 63.0% at 7 years [26]. In another study evaluating 94 patients with subclinical hyperthyroidism (69 had MNG), the

progression rate from subclinical to overt hyperthyroidism was 8% at 1 year, 16% at 2 years, 21% at 3 years, and 26% at 5 years [27]. Therefore, it is reasonable to adopt a conservative approach based on periodic assessment of the thyroid hormonal status (by measuring serum TSH and free T_4), and on the monitoring of goitre and thyroid nodule sizes by US and CT scans. It is recommended by the ATA that this evaluation be conducted every 6 to 18 months, and the frequency can be reduced if no significant changes are identified over 3–5 years [12].

When clinical observation is chosen as a therapeutic approach, malignancy should first be excluded by FNAB-US. Periodic US surveillance is done to identify potentially malignant nodules that may have been missed by previous US or FNAB-US [28]. If a euthyroid patient with MNG develops subclinical or overt hyperthyroidism, additional tests are recommended, including T_3 or free T_3 measurements and thyroid scintigraphy. If the patient is asymptomatic and has no cosmetic complaints, but is hyperthyroid, therapy with antithyroid drugs, or preferably radioiodine should be employed.

Iodine Supplementation

Despite the knowledge that iodine deficiency is the most important aetiological factor in goitre development, iodine supplementation does not have sufficient therapeutic effect on an established MNG. In a multicentre German study, a 1-year combination of iodine and levothyroxine leading to incomplete TSH suppression reduced thyroid nodule volume compared to either component alone or placebo. In this trial patients were iodine deficient (mean urinary iodine excretion 49.7–59.5 μg/L) and this finding can partially explain the reduction in goitre size [29]. Due to the risk of inducing hyperthyroidism (Jod-Basedow effect), iodine supplementation is not used a therapeutic option in patients with MNG.

Suppressive Therapy with Levothyroxine

As TSH is the most important thyroid growth factor and its suppression by pharmacological doses of levothyroxine (L-T_4) supposedly could inhibit thyroid growth, or even reduce its volume. Several studies have evaluated L-T_4 in patients with diffuse goitre, but few studies have involved euthyroid patients with MNG in a randomized placebo-controlled design [30, 31]. Suppressive therapy is greatly inferior than radioactive iodine (^{131}I) therapy; goitre shrinkage occurred in 7% of patients given suppressive treatment compared to a reduction of 35% in volume in the first year after ^{131}I treatment [32].

Suppressive therapy with LT4 is not recommended as a therapeutic option for MNG due to its low efficacy, the need for continuous treatment, and its adverse effects [33]. Low serum TSH in individuals aged 60 or older is associated with increased mortality from all causes, and, in particular, with increased mortality due to circulatory and cardiovascular diseases [1–4].

Surgery

For most patients with symptomatic multinodular goitres, surgery is the treatment of choice. The optimal treatment is total thyroidectomy, because the recurrence rate is more common (2.5–42%) when a more conservative approach is chosen compared to total thyroidectomy (0–0.3%) [34]. Total thyroidectomy eliminates the risk of recurrence, cures coexisting hyperthyroidism if present, and it is not associated with an increase in surgical risk when performed by experienced surgeons [35]. The higher surgical risks associated with reoperation must be considered—reoperation results in a 3- to 10-fold risk for hypoparathyroidism or permanent vocal cord paralysis [36]. Surgery should be considered in patients whose goitre progressively enlarges or who have evidence of upper airway obstruction despite absence of any symptoms. More recently, minimally invasive thyroidectomy has been evaluated as a technique that is associated with less surgical risks [37]. The complication rate is higher in patients with goitres that extend substernally compared to patients whose goitre is entirely cervical [38].

Total thyroidectomy leads to rapid decompression of vital structures, with resolution of compressive symptoms in 100% of the patients, versus 46% of the patients undergoing radioiodine therapy [39]. However, it is associated with higher costs, requires hospitalization, and poses risks inherent to surgery in general and risks that are specific to thyroidectomy, such as hypoparathyroidism, vocal cord paralysis due to trauma of the laryngeal recurrent nerve, tracheal obstruction due to haemorrhage and tracheomalacia. Hypothyroidism is rather an objective than a complication and occurs virtually in all patients who undergo total thyroidectomy.

After total or near-total thyroidectomy, thyroid hormone replacement can be promptly initiated at 1.5–1.7 μg/kg/day, or 10–15% less in older people. Levothyroxine (L-T_4) replacement should not be used as a prophylactic agent against thyroid regrowth if subtotal thyroidectomy is chosen, since evidence for this is lacking. During the postoperative period, calcium and parathyroid hormone levels should be monitored, and supplementation with calcium and vitamin D should be initiated if necessary. Hypoparathyroidism is observed in up to 4–10% of non-specialized centres [40]. Permanent hypoparathyroidism is defined when calcium and/or calcitriol supplementation is required for more than 6 months after the surgery. Some patients may have permanent hypoparathyroidism despite having normal parathyroid hormone levels [41]. The risk for laryngeal recurrent nerve damage in thyroid surgery is higher in cases of recurrent thyroid cancer, recurrent benign goitre, thyroid cancer, extensive resection, and low surgical experience [42]. More than 50% of patients with vocal cord palsy are asymptomatic, and other diagnostic tools may be useful, such as fibreoptic laryngoscopy, subjective questionnaires, or computerized voice analysis [43]. Surgery is relatively contraindicated in patients with significant comorbidities, when a high-volume thyroid surgeon is not available or during pregnancy. If necessary, it should be performed in the latter portion of the second trimester [44].

Radioactive Iodine

For more than seven decades, radioactive iodine (^{131}I) therapy has been used to treat thyroid diseases, mainly Graves' disease [45]. Radioactive iodine is not only effective for curing hyperthyroid states, but also leads to shrinkage of the thyroid gland. Owing to this effect on the gland volume, ^{131}I has been used for a long time in the treatment of compressive non-toxic nodular goitres. In 1988,

Hegedus *et al.*, using US, demonstrated that ^{131}I treatment of non-toxic MNG leads to significant goitre volume reduction after 1 year of ^{131} administration [46].

Treatment with ^{131}I is an option for patients with contraindications to surgery, for those who reject surgical procedures, and for patients who have had previous surgery or radiation to the neck (making further surgical procedures more difficult) [3]. Pretreatment with methimazole before ^{131}I therapy is indicated to patients with subclinical hyperthyroidism that are at an increased risk for complications due to worsening of hyperthyroidism, including elderly and those with cardiovascular disease or severe hyperthyroidism [11]. Studies employing images such as US, CT, or MRI (for accurate measurements of TV) have confirmed that ^{131}I therapy reduces the volume of MNG by 35–50% within 1 year with further reduction observed after 3–5 years and improvement in obstructive symptoms in most patients [1–3, 47].

The RAIU determines the therapeutic efficacy of radioiodine itself. Low isotope accumulation in inactive and partially suppressed areas around the nodule is a limitation for radioiodine treatment in patients with MNG. Multinodular goitres with low and heterogeneous RAIU requires higher activities, and sometimes repeated administrations of ^{131}I. The improvement in compressive symptoms after therapeutic activities of ^{131}I is accompanied by significant tracheal widening, as measured by CT or MRI [3]. Treatment is usually accomplished by the administration of a single oral dose of radioiodine. An effective administered activity is calculated to deliver 100–150 μCi per gram of thyroid tissue, corrected for 24-hour RAIU [1]. The calculated activity is directly proportional to the TV, and inversely proportional to the radioiodine uptake, aiming at an absorbed thyroid dose of 100 Gy [1, 3].

Activities may range from 15 mCi for small goitres with normal/high RAIU, to 100–150 mCi for large glands with low RAIU and heterogeneous scintigraphic tracer distribution [1]. Higher ^{131}I activities cause considerable irradiation of extrathyroidal organs and tissues [48]. In most cases, patients require hospitalization and isolation [3]. A survey about safety practices among members of major societies of physicians and allied specialists who treat patients with thyroid disorders showed a diversity of responses related to ^{131}I administration, suggesting the importance of a multispecialty collaboration in defining more uniform recommendations for patients receiving ^{131}I treatment [49].

The adverse effects of ^{131}I therapy can be classified as either acute or late effects. Acutely, ^{131}I can cause radiation thyroiditis, characterized by thyroid pain that can present with fever, an increased erythrocyte sedimentation rate, and thyrotoxicosis. This effect is generally mild and self-limited to the first week of treatment. It can be treated with non-steroidal anti-inflammatory drugs or, less frequently, corticosteroids (e.g. 40 mg of prednisone daily) [50]. Radiation-related thyrotoxicosis can be treated with β-blockers. Antithyroid drugs and β-blockers should also be used in patients with TMNG before radioiodine therapy in older people or those with cardiovascular disease, but should be discontinued 3–5 days before radioiodine therapy in order to prevent treatment failure [11].

An acute increase of TV following ^{131}I administration is a theoretical complication, caused by the inflammatory response to ^{131}I. In one study the increase was minimal (approximately 4% after 1 week) [51] and no instances of acute respiratory decompensation are reported in the literature. Prophylactic administration of glucocorticoids may be considered in cases of patients with large goitres and/or very small cross-sectional tracheal areas [3]. The most common late effect of ^{131}I therapy is the development of hypothyroidism with a prevalence of 20% to 60% after 5 to 8 years [52]. Younger patients with small goitres and with circulating ATPO antibodies are more likely to become hypothyroid [3].

On the other hand, autoimmune hyperthyroidism (i.e. Graves' disease) can be triggered by ^{131}I therapy. This complication usually appears 1 to 3 months after therapy, and is characterized by the development of hyperthyroidism with increased RAIU and elevated titres of TSH receptor antibodies (TRAb). It is likely to be caused by the exposure to the immune system to thyroidal antigens following follicular cell destruction. One study showed that there is an estimated 1.1% risk of developing postradioiodine Graves' disease in patients undergoing radioiodine therapy for autonomous thyroid disease, and this rate increases approximately 10-fold when TPO antibody levels are elevated before radioiodine therapy. Approximately 1–2% of patients may also develop increase titres of TRAb, without hyperthyroidism [53].

The risk of cancer development after the administration of ^{131}I should be taken in account. There are no systematic data evaluating the effects of radioiodine treatment for MNG on the development of secondary malignancies; all of the available data are derived from patients with either thyroid cancer or hyperthyroidism. In the Cooperative Thyrotoxicosis Therapy Follow-Up Study, almost 36 000 hyperthyroid patients treated with ^{131}I and other therapeutic alternatives were followed for an average of 8.2 years. There was no evidence of increased risk of total mortality due to cancer or leukaemia in patients treated with ^{131}I. While there was an elevated risk of thyroid cancer mortality in patients with TMNG, the excess number of deaths was small, and may represent an association of thyroid cancer with MNG, rather than a causal relationship, since no excess thyroid cancer mortality was seen in patients with Graves' disease [54]. Absolute contraindications to ^{131}I therapy are pregnancy, lactation, coexisting thyroid cancer on FNAB, the inability to comply with radiation safety requirements, and the desire to become pregnant within 4–6 months.

Due to the usually low and heterogeneous RAIU seen in MNG, many strategies to enhance uptake have been evaluated, such as the use of recombinant human TSH [55] or after the use of methimazole to elevate endogenous TSH [56].

Radioiodine and Recombinant Human TSH (RHTSH)

Recombinant human TSH (rhTSH) significantly increases RAIU, TSH, T_3, and T_4 in patients with MNG [55]. There are important interindividual variations regarding the RAIU increase after rhTSH administration as reported in a study, in which the administration of 0.9 mg rhTSH increased the 24-h thyroid RAIU from 23.0% at baseline to 41.0% [57]. Recombinant human TSH induces thyroid swelling in healthy individuals, in a dose-dependent manner (0.1, 0.3, and 0.9 mg of rhTSH). Fast *et al.* suggest that these adverse effects are probably without clinical significance following doses of rhTSH that are equal to or lower than 0.1 mg [58].

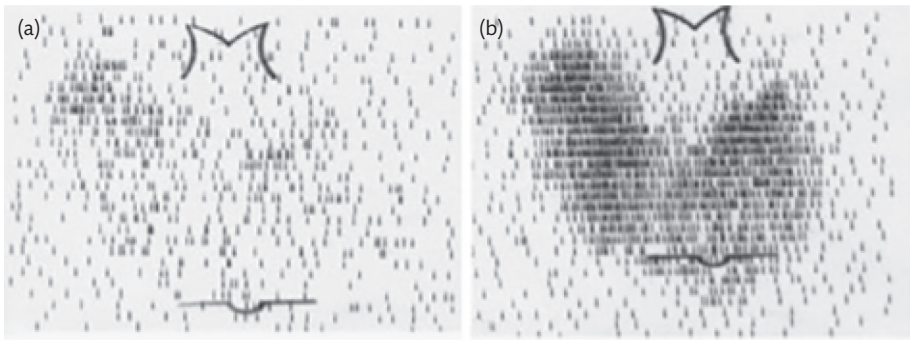


Figure 3.5.2.2 Scintigraphy before (a) and 24hs after the administration of a single dose of 0.1 mg rhTSH (b). Besides making the uptake of ^{131}Imore homogeneous, rhTSH increased the 24-hour uptake from 4.3% to 39%.

Recombinant human TSH is routinely used for diagnostic and therapeutic purposes in patients with differentiated thyroid cancer (DTC). After thyroidectomy, patients with DTC considered with intermediary and high risk by the ATA guidelines [12] receive routinely adjunct ^{131}I for thyroid remnant ablation, and it has been demonstrated that the two existing regimens (rhTSH stimulation vs. thyroid hormone withdrawal) are equally effective for TSH stimulation and thyroid remnant ablation [59].

The administration of a single low dose of rhTSH in patients with MNG significantly enhances and homogenizes thyroid radioiodine uptake [3, 60] (Figure 3.5.2.2).

There is a large interindividual variation in the RAIU increase after the administration of rhTSH. Usually the lower the basal RAIU, the higher the post-rhTSH RAIU [60]. This indicates that individual factors, most of which are yet unidentified, are involved. Much of the variation is explained by differences in the baseline thyroid RAIU, since the effect is highly negatively correlated with this variable [3]. Baseline serum TSH may be a confounding factor, since the increase in RAIU correlates negatively with serum TSH [4]. Since patients with MNG frequently present low serum TSH, radioiodine is only taken up by some 'hot' areas encircled by suppressed thyroid tissue that is inactive on scintigraphy. After RAIU stimulation with rhTSH, these inactive thyroid areas concentrate and amplify the effect of ^{131}I in the gland, with further thyroid reduction. During the last two decades, different rhTSH doses have been utilized: 0.2 mg or more in some studies while 0.1 mg or less in others. Few safety concerns have been observed with the latter doses [61].

In a multicentric randomized controlled study, modified release rhTSH (MRrhTSH) was used to treat patients with MNG [62, 63]. MRrhTSH is an analogue of rhTSH that has the same potency to increase thyroid RAIU, and that determines a lower peak plasma TSH concentration. Potentially, MRrhTSH could reduce the side effects of rhTSH due to its altered pharmacokinetics, with a slightly delayed serum TSH peak after injection, compared to aqueous rhTSH. In this study, the objective was to compare the efficacy and safety of 0.01 and 0.03 mg MRrhTSH as an adjuvant to ^{131}I therapy, vs. ^{131}I alone. TV decreased significantly in all groups after six months: by 23% in patients prestimulated with either placebo or 0.01 mg MRrhTSH, and by 33% in patients prestimulated with 0.03 mg. The smallest cross-sectional area of the trachea increased more in the latter group, without significant difference from the two other groups. The long-term (36 months) results of the same trial demonstrated that patients who received 0.03 mg of MRrhTSH with baseline RAIU <20% achieved a greater reduction in goitre size [63].

Side Effects of rhTSH Use In MNG

Some side effects with the use of rhTSH in MNG can occur [4, 64]. Nielsen *et al.* demonstrated that patients with MNG can report a sensation of thyroid swelling after administration of 0.3 mg rhTSH, but no acute compressive effects have been observed [65]. In this study, patients with a rather small MNG (median volume of 40.0 ml) presented a 24% transient goitre enlargement. Patients with TMNG presented higher increases in thyroid hormone levels after 0.1 mg rhTSH plus ^{131}I, with more side effects [66]. In these cases, safety measures such as the use of β-blockers should be considered.

Currently, the rhTSH adjunct therapy is not indicated for patients with TMNG [4, 11]. Glucocorticoid therapy rarely is necessary in patients receiving ^{131}I for the treatment of MNG [1]. On the other hand, prophylactic glucocorticoid therapy should be considered in patients with critical tracheal narrowing, to prevent thyroid swelling and further respiratory compromise [11]. Painful transient thyroiditis may occur within the first month after treatment, and the development of Graves' disease (with high levels of TSH receptor antibodies) is reported in euthyroid MNG patients with pre-existing TPOAb [67].

The occurrence of hypothyroidism after the use of ^{131}I with or without rhTSH preparation should not be viewed as a complication, rather more as a normal consequence of TV reduction and consequent thyroid hypofunction. The occurrence of hypothyroidism after the use of ^{131}I for MNG usually indicates that goitre shrinkage is occurring. The development of hypothyroidism is common, and depends on the size of the treated goitre, and on the administered rhTSH dose and ^{131}I activity [3, 4]. The optimal rhTSH dose for enhancing ^{131}I therapy in MNG with non-significant side effects is most likely in the range of 0.05–0.1 mg [4].

A recent meta-analysis demonstrated that the administration of rhTSH before radioiodine therapy resulted in a greater reduction in TV than radioiodine therapy alone, and in an increased incidence of hypothyroidism in patients receiving high-dose rhTSH. The authors concluded that low-dose rhTSH before radioiodine therapy was more efficacious than radioiodine therapy alone, when used for treating non-toxic benign thyroid nodules [68].

Use of Methimazole to Enhance RAIU

Recent studies suggest that TSH elevation after the use of methimazole in patients with MNG leads to a rise in the RAIU and an excellent result in goitre volume reduction after ^{131}I activities [56, 69–71]. Albino *et al.* treated nine female patients with MNG with methimazole for 2.8 +/– 0.8 months (10–20 mg, with monthly adjusted doses based on thyroid hormone levels), leading to increases in mean serum TSH to 11.7 +/– 5.4 mU/L, and to increases in mean 24-h RAIU, from median values of 21.3–78.3%. After 1 year of a fixed activity of 1.11 GBq (30 mCi) of ^{131}I, median TV decreased from 97 ml to 56 ml (mean reduction of 46.2%). Eight patients (89%) had initially subclinical hyperthyroidism, which was reversed in all patients after one year. Five patients (56%) developed overt hypothyroidism, and no clinical adverse events were observed [56]. Kyrilli evaluated 22 patients with MNG, with subclinical hyperthyroidism and RAIU <50%. Patients were randomized to receive either a low iodine diet (LID) or methimazole (MTZ, 30 mg/day) [70]. Thiamazole treatment prior to ^{131}I therapy resulted in an average twofold increase in thyroid RAIU and enhanced the efficiency of radioiodine therapy assessed at 12 months.

It is important to mention that the adjunct therapy of MNG with rhTSH and ^{131}I is not approved by the US Food and Drug Administration (FDA) or European Medicines Agency (EMA). Moreover, the cost-effectiveness of the combined rhTSH therapy has not been demonstrated [4].

REFERENCES

1. Hegedus L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev*, 2003; **24**: 102–32.
2. Graf H. Multinodular goiter: pathogenesis and management. In: Braverman LE, Cooper DS, eds. *The Thyroid: A Fundamental and Clinical Text*, 10th edn. Philadelphia, PA: Lippincott Williams & Wilkins, Wolters Kluver; 2013: 635–49.
3. Graf Recombinant human TSH and radioactive iodine therapy in the management of benign multinodular goiter. *Eur J Endocrinol*, 2015; **172**: R47–52.
4. Bahn RS, Castro MR. Approach to the patient with nontoxic multinodular goiter. *J Clin Endocrinol Metab*, 2011; **96**: 1202–1.
5. Shaha AR. Substernal goiter: what is in a definition? *Surgery*, 2010; **147**: 239–40.
6. Mercante G, Gabrielli E, Pedroni C, *et al.* CT cross-sectional imaging classification system for substernal goiter based on risk factors for an extracervical surgical approach. *Head Neck*, 2011; **33**: 792–9.
7. Al-Bazzaz F, Grillo H, Kazemi H. Response to exercise in upper airway obstruction. *Am Rev Respir Dis*, 1975; **111**: 631–40.
8. Pemberton HS. Sign of submerged goitre. *Lancet*, 1946; **248**: 509.
9. Ioannidis O, Dalampini E, Chatzopoulos S, *et al.* Acute respiratory failure caused by neglected giant substernal nontoxic goiter. *Arq Bras Endocrinol Metabol*, 2011; **55**: 229–32.
10. Levin R, Newman SA, Login IS. Bilateral Horner's syndrome secondary to multinodular goiter. *Ann Intern Med*, 1986; **105**: 550–1.
11. Bahn RS, Burch HB, Cooper DS, *et al.* Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid*, 2011; **21**: 593–646.
12. Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 2016; **26**: 1–133.
13. Brito JP, Yarur AJ, Prokop LJ, McIver B, Murad MH, Montori VM. Prevalence of thyroid cancer in multinodular goiter versus single nodule: a systematic review and meta-analysis. *Thyroid*, 2013; **23**: 449–55.
14. Frates MC, Benson CB, Charboneau JW, *et al.* Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Radiology*, 2005; **237**: 794–800.
15. Grainger J, Saravanappa N, D'Souza A, *et al.* The surgical approach to retrosternal goiters: the role of computerized tomography. *Otolaryngol Head Neck Surg*, 2005; **132**: 849–85.
16. Okosieme OE, Chan D, Price SA, *et al.* The utility of radioiodine uptake and thyroid scintigraphy in the diagnosis and management of hyperthyroidism. *Clin Endocrinol (Oxf)*, 2010; **72**: 122–27.
17. Cerci C, Cerci SS, Eroglu E, *et al.* Thyroid cancer in toxic and non-toxic multinodular goiter. *J Postgrad Med*, 2007; **53**: 157–60.
18. Gharib H, Papini E, Paschke R, *et al.* American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *Endocr Pract*, 2010; **16**: 468–47.
19. Fish SA, Langer JE, Mandel SJ. Sonographic imaging of thyroid nodules and cervical lymph nodes. *Endocrinol Metab Clin North Am*, 2008; **37**: 401–17.
20. de Carvalho GA, Paz-Filho G, Cavalcanti TC, *et al.* Adequacy and diagnostic accuracy of aspiration vs. capillary fine needle thyroid biopsies. *Endocr Pathol*, 2009; **20**: 204–8.
21. Cibas ES, Alo SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. 2017; **27**: 1341–46.
22. Bonnema SJ, Bennedbaek FN, Wiersinga WM, *et al.* Management of the nontoxic multinodular goitre: a European questionnaire study. *Clin Endocrinol (Oxf)*, 2000; **53**: 5–12.
23. Bonnema SJ, Bennedbaek FN, Ladenson PW, *et al.* Management of the nontoxic multinodular goiter: a North American survey. *J Clin Endocrinol Metab*, 2002; **87**: 112–17.
24. Diehl LA, Garcia V, Bonnema SJ, *et al.* Management of the nontoxic multinodular goiter in Latin America: comparison with North America and Europe, an electronic survey. *J Clin Endocrinol Metab*, 2005; **90**: 117–23.
25. Costante G, Crocetti U, Schifino E, *et al.* Slow growth of benign thyroid nodules after menopause: no need for long-term thyroxine suppressive therapy in post-menopausal women. *J Endocrinol Invest*, 2004; **27**: 31–6.
26. Vadiveloo T, Donnan PT, Cochrane L, *et al.* The Thyroid Epidemiology, Audit, and Research Study (TEARS): morbidity in patients with endogenous subclinical hyperthyroidism. *J Clin Endocrinol Metab*, 2011; **96**: 1344–135
27. Schouten BJ, Brownlie BE, Frampton CM, *et al.* Subclinical thyrotoxicosis in an outpatient population—predictors of outcome. *Clin Endocrinol (Oxf)*, 2011; **74**: 257–61.
28. Gabalec F, Cap J, Ryska A, *et al.* Benign fine-needle aspiration cytology of thyroid nodule: to repeat or not to repeat? *Eur J Endocrinol*, 2009; **161**: 933–7.
29. Grussendorf M, Reiners C, Paschke R, Wegscheider K, Investigators L. Reduction of thyroid nodule volume by

levothyroxine and iodine alone and in combination: a randomized, placebo-controlled trial. *J Clin Endocrinol Metab*, 2011; **96**: 2786–95.
30. Berghout A, Wiersinga WM, Drexhage HA, Smits NJ, Touber JL. Comparison of placebo with L-thyroxine alone or with carbimazole for treatment of sporadic non-toxic goitre. *Lancet*, 1990; **336**: 193–7.
31. Cesareo R, Iozzino M, Isgro MA, Annunziata F, Di Stasio E. Short term effects of levothyroxine treatment in thyroid multinodular disease. *Endocr J*, 2010; **57**: 803–9.
32. Wesche MF, Tiel VBMM, Lips P, Smits NJ, Wiersinga WM. A randomized trial comparing levothyroxine with radioactive iodine in the treatment of sporadic nontoxic goiter. *J Clin Endocrinol Metab*, 2001; **86**: 998–1005.
33. Fast S, Bonnema SJ, Hegedus L. The majority of Danish non-toxic goitre patients are ineligible for Levothyroxine suppressive therapy. *Clin Endocrinol (Oxf)*, 2008; **69**: 653–8.
34. Moalem J, Suh I, Duh QY. Treatment and prevention of recurrence of multinodular goiter: an evidence-based review of the literature. *World J Surg*, 2008; **32**: 1301–12.
35. Agarwal G, Aggarwal V. Is total thyroidectomy the surgical procedure of choice for benign multinodular goiter? An evidence-based review. *World J Surg*, 2008; **32**: 1313–24.
36. Thomusch O, Machens A, Sekulla C, *et al.* Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. *World J Surg*, 2000; **24**: 1335–41.
37. Perigli G, Cortesini C, Qirici E, *et al.* Clinical benefits of minimally invasive techniques in thyroid surgery. *World J Surg*, 2008; **32**: 45–50.
38. Pieracci FM, Fahey TJ, 3rd. Substernal thyroidectomy is associated with increased morbidity and mortality as compared with conventional cervical thyroidectomy. *J Am Coll Surg*, 2007; **205**: 1–7.
39. Porterfield JR, Jr., Thompson GB, Farley DR, *et al.* Evidence-based management of toxic multinodular goiter (Plummer's Disease). *World J Surg*, 2008; **32**: 1278–84.
40. Bergenfelz A, Jansson S, Kristoffersson A, *et al.* Complications to thyroid surgery: results as reported in a database from a multicenter audit comprising 3,660 patients. *Langenbecks Arch Surg*, 2008; **393**: 667–73.
41. Promberger R, Ott J, Kober F, *et al.* Normal parathyroid hormone levels do not exclude permanent hypoparathyroidism after thyroidectomy. *Thyroid*, 2011; **21**: 145–50.
42. Dralle H, Lorenz K, Machens A. State of the art: surgery for endemic goiter—a plea for individualizing the extent of resection instead of heading for routine total thyroidectomy. *Langenbecks Arch Surg*, 2011; **396**: 1137–43.
43. Ortega J, Cassinello N, Dorcaratto D, *et al.* Computerized acoustic voice analysis and subjective scaled evaluation of the voice can avoid the need for laryngoscopy after thyroid surgery. *Surgery*, 2009; **145**: 265–71.
44. Abalovich M, Amino N, Barbour LA, *et al.* Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2007; **92**: S1–47.
45. Bonnema SJ, Hegedus L. A 30-year perspective on radioiodine therapy of benign nontoxic multinodular goiter. *Curr Opin Endocrinol Diabetes Obes*, 2009; **16**: 379–84.
46. Hegedus L, Hansen BM, Knudsen N, Hansen JM. Reduction of size of thyroid with radioactive iodine in multinodular non-toxic goitre. *Br Med J*, 1988; **297**: 661–2.
47. Bonnema SJ, Bertelsen H, Mortensen J, *et al.* The feasibility of high dose iodine 131 treatment as an alternative to surgery in patients with a very large goiter: effect on thyroid function and size and pulmonary function. *J Clin Endocrinol Metab*, 1999; **84**: 3636–41.
48. Huysmans DA, Buijs WC, van de Ven MT, *et al.* Dosimetry and risk estimates of radioiodine therapy for large, multinodular goiters. *J Nucl Med*, 1996; **37**: 2072–9.
49. Greenlee C, Burmeister LA, Butler RS, *et al.* Current safety practices relating to I-131 administration for diseases of the thyroid: a survey of physicians and allied practitioners. *Thyroid*, 2011; **21**: 151–60.
50. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. *N Engl J Med*, 2003; **348**: 2646–55.
51. Nygaard B, Faber J, Hegedus L. Acute changes in thyroid volume and function following 131I therapy of multinodular goitre. *Clin Endocrinol (Oxf)*, 1994; **41**: 715–18.
52. Fast S, Nielsen VE, Bonnema SJ, *et al.* Time to reconsider non-surgical therapy of benign non-toxic multinodular goitre: focus on recombinant human TSH augmented radioiodine therapy. *Eur J Endocrinol*, 2009; **160**: 517–25.
53. Schmidt M, Gorbauch E, Dietlein M, *et al.* Incidence of postradioiodine immunogenic hyperthyroidism/Graves' disease in relation to a temporary increase in thyrotropin receptor antibodies after radioiodine therapy for autonomous thyroid disease. *Thyroid*, 2006; **16**: 281–8.
54. Ron E, Doody MM, Becker DV, *et al.* Cancer mortality following treatment for adult hyperthyroidism. Cooperative Thyrotoxicosis Therapy Follow-up Study Group. *JAMA*, 1998; **280**: 347–55.
55. Paz-Filho GJ, Graf H. Recombinant human thyrotropin in the management of thyroid disorders. *Expert Opin Biol Ther*, 2008; **8**: 1721–32.
56. Albino CC, Graf H, Sampaio AP, Vigario A, Paz-Filho GJ. Thiamazole as an adjuvant to radioiodine for volume reduction of multinodular goiter. *Expert Opin Investig Drugs*, 2008; **17**: 1781–6.
57. Torres MST, Ramirez L, Simkin PH, Braverman LE, Emerson CH. Effect of various doses of recombinant human thyrotropin on the thyroid radioactive iodine uptake and serum levels of thyroid hormones and thyroglobulin in normal subjects. *J Clin Endocrinol Metab*, 2001; **86**: 1660–4.
58. Fast S, Nielsen VE, Bonnema SJ, Hegedus L. Dose-dependent acute effects of recombinant human TSH (rhTSH) on thyroid size and function: comparison of 0.1, 0.3 and 0.9 mg of rhTSH. *Clin Endocrinol (Oxf)*, 2010; **72**: 411–16.
59. Schlumberger M, Catargi B, Borget I, *et al.* Strategies of radioiodine ablation in patients with low-risk thyroid cancer. *N Engl J Med*, 2012; **366**: 1663–73.
60. Huysmans DA, Nieuwlaat WA, Erdtsieck RJ, *et al.* Administration of a single low dose of recombinant human thyrotropin significantly enhances thyroid radioiodide uptake in nontoxic nodular goiter. *J Clin Endocrinol Metab*, 2000; **85**: 3592–6
61. Knobel M. Which is the ideal treatment for benign diffuse and multinodular non-toxic goiters? *Front Endocrinol (Lausanne)*, 2016; **7**: 48.
62. Graf H, Fast S, Pacini F, *et al.* Modified-release recombinant human TSH (MRrhTSH) augments the effect of (131)I therapy

in benign multinodular goiter: results from a multicenter international, randomized, placebo-controlled study. *J Clin Endocrinol Metab*, 2011; **96**: 1368–76.
63. Fast S, Hegedus L, Pacini F, *et al*. Long-term efficacy of modified-release recombinant human thyrotropin augmented radioiodine therapy for benign multinodular goiter: results from a multicenter, international, randomized, placebo-controlled, dose-selection study. *Thyroid*, 2014; **24**: 727–35.
64. Bonnema SJ, Hegedus L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*, 2012; **33**: 920–80.
65. Nielsen VE, Bonnema SJ, Hegedus L. Transient goiter enlargement after administration of 0.3 mg of recombinant human thyrotropin in patients with benign nontoxic nodular goiter: a randomized, double-blind, crossover trial. *J Clin Endocrinol Metab*, 2006; **91**: 1317–22.
66. Romao R, Rubio IG, Tomimori EK, Camargo RY, Knobel M, Medeiros-Neto G. High prevalence of side effects after recombinant human thyrotropin-stimulated radioiodine treatment with 30 mCi in patients with multinodular goiter and subclinical/clinical hyperthyroidism. *Thyroid*, 2009; **19**: 945–51.
67. Nygaard B, Knudsen JH, Hegedus L, Scient AV, Hansen JE. Thyrotropin receptor antibodies and Graves' disease, a side-effect of 131I treatment in patients with nontoxic goiter. *J Clin Endocrinol Metab*, 1997; **82**: 2926–30.
68. Lee YY, Tam KW, Lin YM, *et al*. Recombinant human thyrotropin before (131)I therapy in patients with nodular goitre: a meta-analysis of randomized controlled trials. *Clin Endocrinol (Oxf)*, 2015; **83**: 702–10.
69. Flores-Rebollar A, Ruiz-Juvera A, Lopez-Carrasco G, Gonzalez-Trevino O. Effect of 1110 MBq radioiodine in reducing thyroid volume in multinodular goiter: a new protocol. *J Clin Med Res*, 2013; **5**: 234–8.
70. Kyrilli A, Tang BN, Huyge V, *et al*. Thiamazole pretreatment lowers the (131)I activity needed to cure hyperthyroidism in patients with nodular goiter. *J Clin Endocrinol Metab*, 2015; **100**: 2261–7.
71. Spaulding SW. Could pretreatment with methimazole be as effective as rhTSH in treating subclinical toxic multinodular goiter with 131I? *Clin Thyroidol*, 2015; **27**: 144–6.

3.5.3 Management of the Single Thyroid Nodule

Laszlo Hegedüs and Finn N. Bennedbæk

Introduction

The main concern of patients and physicians alike, when dealing with the solitary thyroid nodule, is to diagnose the few cancers (approximately 5%) as rapidly and cost-effectively as possible and to reduce superfluous thyroid surgery. This chapter focuses on the palpably discrete swelling within an otherwise normal gland in the clinically and biochemically euthyroid patient [1, 2]. The toxic nodule is dealt with in Chapter 3.3.11, and thyroid malignancy in Chapters 3.5.4–3.5.7.

Occurrence

The estimated life-time risk of developing a thyroid nodule is between 5% and 10% [1, 2], but factors such as sex (four times more common in women), age (frequency increases with age), regional iodine intake (more prevalent in iodine-deficient areas), and whether the diagnosis is made clinically (palpation), by ultrasonography (US), or at autopsy (5–10 times more prevalent using the last two) are of importance when estimating prevalence [1, 2]. The incidence of clinical disease has been estimated at 0.1% by palpation [1, 2].

Natural History

Very little is known regarding the natural history of thyroid nodules, since data are highly selected and generally concern patients with small nodules without suspicion of malignancy and not causing pressure symptoms or cosmetic complaints. With these restrictions, most nodules appear not to change appreciably over time. The nodules that increase in size are predominantly solid and carry a higher risk of harbouring thyroid carcinoma than those predominantly cystic, being more prone to decrease in size or even disappear.

In most patients, US will identify nodules not evident clinically and, given time, most of these patients will be classified as having multinodular goitre. Therefore, the risk of thyroid malignancy, at the patient level, is independent of whether the nodule is solitary or appears in a multinodular gland [1, 3]. Furthermore, non-palpable nodules have the same risk of malignancy as do US confirmed palpable nodules of the same size [4].

Diagnosis

Clinical Examination

History and physical examination are important and patients with a risk of thyroid carcinoma can be identified (**Box 3.5.3.1**). A positive family history of benign goitre suggests a benign disorder, whereas multiple endocrine neoplasia, medullary thyroid carcinoma, or even papillary or follicular thyroid carcinoma in the family should raise suspicion. Nodules occurring in the young or in the old are especially likely to be cancerous, the risk being higher in men than in women. Head or neck irradiation in childhood leads to clinically evident thyroid abnormality in 10–40% of patients 5–40 years later. Rapid tumour growth (weeks to months)

Box 3.5.3.1 Clinical factors increasing the likelihood of thyroid malignancy in a euthyroid patient with a solitary nodule

- Family history of thyroid malignancy
- Age less than 20 or more than 60 years
- Male sex
- History of head and neck irradiation in infancy, childhood, or adolescence
- Large nodule (greater than 4 cm in diameter) and partially cystic
- Rapid nodule growth
- Pain
- Firm or hard nodule
- Fixation to adjacent structures
- Compression symptoms: dysphagia, dyspnoea, vocal cord paralysis
- Regional lymphadenopathy
- Growth during L-thyroxine therapy

and symptoms of local invasion, such as pain, dysphagia, hoarseness, or dyspnoea, suggests a carcinoma, but only a minority of patients have these symptoms.

The physical examination is important in the work-up and certain signs and symptoms are highly suspicious of thyroid malignancy (Box 3.5.3.1), but inter- and intraobserver variation is alarmingly high [5] and the specificity and sensitivity of the diagnosis of a solitary thyroid nodule is low. Thus, nodules of 10 mm or more can usually be palpated depending on their localization in the neck. However, one-half of the nodules found by US examination escape clinical detection, one-third of which are more than 20 mm in diameter. A hard nodule is not necessarily a carcinoma (chronic thyroiditis), whereas a soft nodule may well be a cystic papillary cancer. In view of this, current guidelines rely on a US-based risk stratification and fine-needle aspiration biopsy (FNA) rather than a physical examination [6, 7].

Laboratory Investigation

The only relevant biochemical test that is routinely needed is serum thyroid-stimulating hormone (TSH) measured with a sensitive assay. Subnormal serum TSH values should lead to determination of free thyroxine (T_4) and free triiodothyronine (T_3). In the presence of normal thyroid hormone levels, a suppressed serum TSH on repeat examination should lead to treatment, especially in older patients (Chapter 3.3.4) [8]. Scintigraphy is recommended and will most likely demonstrate a hot or a toxic nodule (Chapters 3.1.6 and 3.3.11) in such patients, most of whom are euthyroid, including those with thyroid malignancy. It seems that the risk of malignancy in a thyroid nodule increases with serum TSH concentration, within the normal range, at presentation, and thus serves as an additional and independent predictor of malignancy [9]. Hypothyroidism indicates that the patient may have Hashimoto's thyroiditis. Thyroglobulin in serum is positively correlated with thyroid size but has no place in the routine investigation or in the follow-up of benign nodules. Calcitonin is the only clinically relevant biochemical marker of medullary thyroid carcinoma. Routine determination has been suggested by European guidelines [10, 11]. It allows the detection of unsuspected medullary thyroid carcinoma with a frequency of 1 in 200–300 thyroid nodules, with better sensitivity than FNA (11). However, there remains unresolved issues of sensitivity, specificity, assay performance, and cost-effectiveness [10]. Thyroid autoantibodies against thyroid peroxidase cannot differentiate between malignant and benign disease. In our opinion, they should be determined routinely in the work-up to identify patients with possible Hashimoto's thyroiditis. These antibodies are also markers of an increased risk of developing hypo- or hyperthyroidism (Graves' disease) spontaneously or secondary to surgery or radioiodine treatment [12]. TSH-receptor antibodies are rarely present and should not be determined routinely.

Diagnostic Imaging

No method of imaging can differentiate benign from malignant nodules accurately. However, 88% of European thyroidologists use either scintigraphy (66%), ultrasonography (80%), or both (58%) in the evaluation of patients with a clinically solitary thyroid nodule, illustrating that the diagnosis is not always straightforward and that they believe diagnostic imaging gives valuable information (see also Chapters 3.1.6 and 3.1.6.1) [13].

There is no consensus on the use of scintigraphy in the euthyroid patient. Since most have a cold nodule—increasing the risk of thyroid malignancy at least 10-fold—many investigators, mainly in the United States, advocate using FNA as the first step [1, 2]. Imaging, if performed, can be with ^{123}I, ^{131}I, or ^{99m}Tc pertechnetate, the latter being preferred (86%) among European thyroidologists [13], although iodine isotopes should be used if the aim is also to reduce the risk of overlooking malignancy. A clinically dominant nodule that is cold on scintigraphy, should be treated as a solitary cold nodule, the risk of malignancy being the same [1, 3], and FNA should be performed. In case of suppressed serum TSH or overt hyperthyroidism, the risk of malignancy is thought to be much lower as is the need for FNA.

US, often used in Europe (80%) and less so in the United States, allows determination of total thyroid volume, individual nodule evaluation for size, echogenicity, internal content, margin, shape, calcifications, nodule vascularity, elasticity, and regional lymph nodes [7, 13]. Elastography, however, should not replace grey-scale US-investigation, but it may be used as a complementary tool for assessing nodules for FNA, especially due to its high NPV.

Classification of thyroid nodules according to a thyroid imaging reporting and data system (TIRADS) is based on the fact that certain US-features of thyroid nodules are consistently predictive of malignancy and are used as criteria for FNA (see Chapter 3.1.7) [7]. About 50% of the nodules can be classified as benign by US criteria with a risk of false negatives of only 3%. US aids in performing accurate biopsies and is of great help in therapeutic procedures such as cyst punctures, percutaneous ethanol injection and thermal (laser, radiofrequency, microwaves and high-frequency US) therapy of solid as well as cystic nodules [1, 14, 15]. For the objective determination of thyroid or nodule size, whether initially or during follow-up, it is the technique of choice.

Focal ^{18}FDG-PET uptake within a US confirmed thyroid nodule conveys an increased risk of thyroid cancer, and FNA is recommended for those nodules [16]. CT and MRI are generally of little value except in the evaluation of the intrathoracic goitre or in the evaluation and follow-up of malignant thyroid disease.

Ultrasonography and Fine-Needle Aspiration Biopsy

FNA provides the most direct and specific information about a thyroid nodule, and it is used by the majority of clinicians on an outpatient basis [13, 17]. As the cornerstone in the evaluation it is

virtually without complications, inexpensive, and easy to learn to perform (see Figure 3.5.3.1).

Current guidelines for thyroid nodule FNA are recommended for the intermediate and high-risk nodules and is based on the structured and standardized US reporting system (TIRADS) (see Chapter 3.1.7, Box 3.1.7.4) [7]. Pure cysts and entirely spongiform nodules (TIRADS 2) should be considered as benign and therefore FNA is not indicated (unless for therapeutic purposes or in case of compressive symptoms). Oval-shaped, isoechoic, or hyperechoic nodules with smooth margins and no high-risk features (TIRADS 3) should be considered at low risk of malignancy (estimated risk of malignancy: 2–4%) and FNA should usually be performed only for nodules larger than 20 mm. Oval-shaped, mildly hypoechoic nodules with smooth margins and no high-risk features (TIRADS 4) should be considered at intermediate risk of malignancy (estimated risk of malignancy: 6–17%). FNA should usually be performed for nodules larger than 15 mm. Nodules having at least one suspicious US feature (i.e. a non-oval shape, irregular margins, microcalcifications, or marked hypoechogenicity) should be considered at high risk of malignancy, increasing with the number of suspicious features, and FNA should be performed for nodules larger than 10 mm. In this group (TIRADS 5) the risk of malignancy is significantly higher: 26–87%. However, in case of subcentimetric nodules with high-risk US features, FNA or at least active surveillance is recommended [7].

Diagnostically useful FNA specimens are obtained in about 80% of cases. The number of insufficient samples depends on operator experience, number of aspirations, the character of the nodule (cystic/solid), the experience of the cytopathologist, and especially the criteria used for adequacy of a sample. The number of sufficient samples increases if FNA is guided by US. Rebiopsy will typically halve the number of insufficient biopsies. Large-core needle biopsies can be considered when cytopathology expertise is not available. Its limitations include the need for local anaesthesia, local discomfort, and decreasing patient acceptance of repeat biopsies [18].

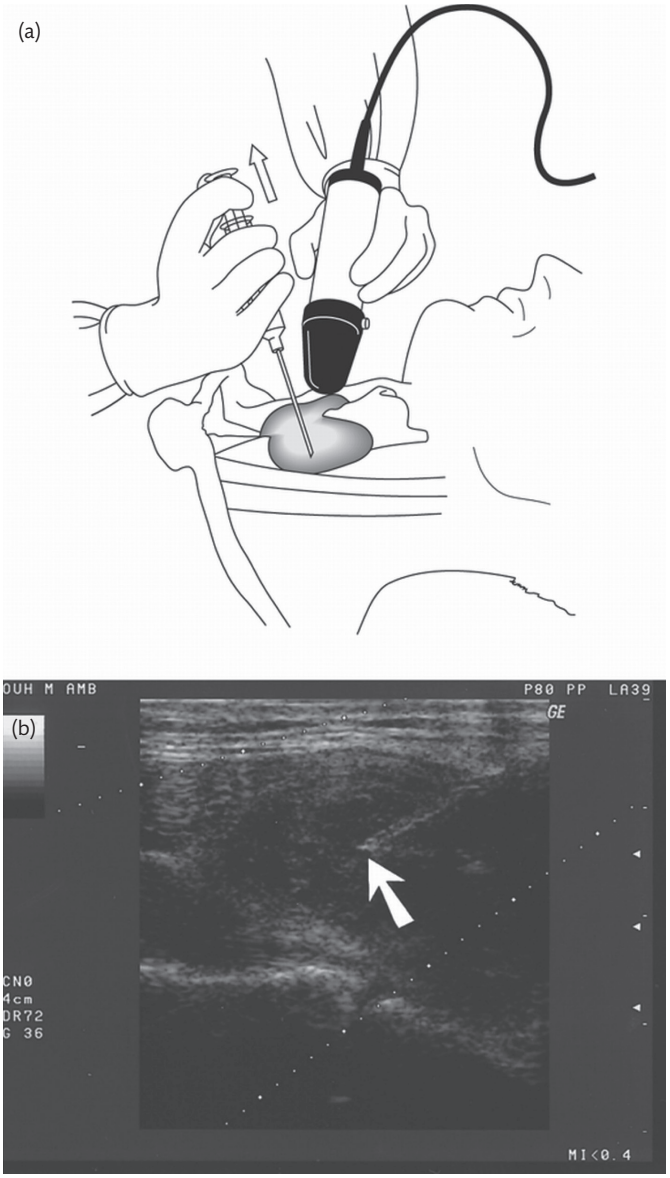


Figure 3.5.3.1 (a) Ultrasound-guided fine-needle aspiration biopsy of a thyroid nodule. The needle is inserted into the nodule. Some use a free-hand technique; others use needle-steering devices. (b) Longitudinal ultrasound scan of the neck showing the needle tract and needle tip (arrow) inside a solid hypoechoic thyroid nodule.

Fine-Needle Aspiration Cytology

There are several cytology classification systems. Here we focus on the Bethesda classification, which is based on five cytological diagnostic categories with a subdivision of the indeterminate category into atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) and follicular neoplasm/suspicious for a follicular neoplasm (FN/sSFN) (see Table 3.5.3.1) [19].

While non-diagnostic and benign cytologies are characterized by a low estimated risk of malignancy (1–4% and 0–3%, respectively), the risk for the indeterminate group is considerably higher (5–15% and 15–30%, respectively) and as expected very high in the suspicious and malignant group (60–75% and 97–99%, respectively) [6]. In unselected patients undergoing FNA, cytologically benign results are expected in 39–74% of the nodules and 2–16% are cytologically malignant [20]. AUS/FLUS and FN/SFN results were observed in 1–27% and 1–25%, respectively, and cytologically non-diagnostic in 2–24% of FNAs [20]. An intrinsic limitation of thyroid FNA cytology will inevitably result in a number of indeterminate results. Vascular or capsular invasion, which are the criteria distinguishing follicular adenomas or adenomatous nodules from follicular carcinomas (FTCs) and follicular variant papillary thyroid carcinomas (fvPTC), cannot be detected in cytology samples. Thus, a high proportion of patients with FNAs classified as indeterminate will undergo diagnostic, potentially unnecessary, thyroid surgery with possible complications with the ultimate histology revealing only a 20% malignancy rate [6, 11]. Repeat FNA for AUS/FLUS FNA results are thus recommended to increase the chance for a definitive cytological diagnosis.

Accumulating evidence suggests that the limitation of AUS/FLUS and FN/SFN cytology can be compensated for by add-on molecular diagnostic approaches [21–23]. The different molecular diagnostic methodologies can be broadly classified into two categories: 'rule out' malignancy or 'rule in' malignancy approaches (see Chapter 3.5.1). While the former aims to reduce the overtreatment of benign nodules, the latter aims to optimize surgical planning (primary total thyroidectomy vs. a two-staged approach with 'diagnostic' lobectomy and subsequent completion surgery).

Table 3.5.3.1 The Bethesda system for reporting thyroid cytology and corresponding risk of malignancy

Diagnostic category	Estimated risk of malignancy % (range)	Actual risk of malignancy in nodules surgically excised % median (range)
Non-diagnostic or unsatisfactory	1–4	20 (9–32)
Benign	0–3	2.5 (1–10)
Undetermined		14 (6–48)
AUS/FLUS	5–15	
FN/SFN	15–30	25 (14–34)
Suspicious for malignancy	60–75	70 (53–97)
Malignant	97–99	99 (94–100)

AUS/FLUS, atypia of undetermined significance/follicular lesion of undetermined significance; FN/SFN, follicular neoplasm/suspicious for a follicular neoplasm.

Approach to the Patient with a Single Thyroid Nodule

- With the discovery of a thyroid nodule, a complete history and physical examination focusing on the thyroid gland and adjacent cervical lymph nodes should be performed.
- Serum TSH. If the serum TSH is subnormal, a thyroid scintigraphy should be obtained to confirm tracer uptake in the nodule. Functioning nodules rarely harbour malignancy.
- Diagnostic US should be performed unless the serum TSH is suppressed and standardized reporting should include:
  - Thyroid volume
  - Echogenicity and vascularity of the gland
  - Nodules (above 5 mm unless highly suspect)
  - Location (side, superior, medial, inferior)
  - Size (three diameters +/– volume)
  - Shape, margins, echogenicity, composition
  - TIRADS score
  - Change of size in case of follow-up US
  - Retrosternal extension and tracheal deviation
  - Study of lymph nodes (levels II, III, IV, V, VI)
  - FNA guided by US is recommended for TIRADS 3–5 (dependent on nodule size)
  - Most impalpable thyroid nodules (incidentalomas) require observation alone, but US-guided FNA is recommended based on TIRADS score [24]

If the nodule is benign on cytology, further immediate diagnostic studies or treatment are not required. Furthermore, an initially benign FNA confers negligible mortality risk during long-term follow-up despite a low but real risk of false negatives [1–2%] in this cytology category. Nodule growth cannot accurately discriminate benign from malignant lesions but cytologically benign nodules in TIRADS 3–5 should be considered for repeat US after 12–24 months [6]. Based on evidence, it is still unclear if patients with thyroid nodules larger than 4 cm and benign cytology carry a higher risk of malignancy and should be managed differently than those with smaller nodules.

Cytology diagnostic for a primary thyroid malignancy or suspicious for malignancy will almost always lead to thyroid surgery. However, an active surveillance management approach can be considered as an alternative to immediate surgery in

× Patients with very low risk tumours (e.g. papillary microcarcinomas without clinically evident metastases or local invasion and no evidence of aggressive disease)
× Patients at high surgical risk because of comorbidity
× Patients with an expected short remaining lifespan

Treatment

Most nodules are asymptomatic and benign and are usually managed by observation only and follow-up. However, some nodules do grow, cause symptoms, and require treatment, but still the optimal therapy for patients with thyroid nodules varies with the lesion and whether it is functioning (see Table 3.5.3.2).

Table 3.5.3.2 Treatment of the single thyroid nodule: comparison of various methods of treatment

Treatment type	Advantages	Disadvantages
Surgery	Nodule ablation, complete relief of symptoms, definite histological diagnosis	Inpatient, high cost, risks associated with surgery, vocal cord paralysis (approximately 3% of patients), hypoparathyroidism (<1%), hypothyroidism (1–10% in case of lobectomy) [25]
L-thyroxine	Outpatient, low cost, may slow nodule growth, may prevent new nodule formation	Low efficacy, need for lifelong treatment, regrowth after cessation of treatment, cardiac tachyarrhythmias, reduced bone density, not feasible when thyrotropin level is suppressed
Radioiodine[a]	Outpatient, low cost, high success rate (normalization of s-TSH in >95% and nodule reduced by 40% within 1 year)	Hypothyroidism (10% in 5 years), risk of radiation thyroiditis and thyrotoxicosis, only gradual reduction of the nodule, use of contraceptives in fertile women
Ethanol injection	Outpatient, relatively low cost, no hypothyroidism, nodule reduced by >40% within 6 months, recurrence rate reduced >80% in cystic nodules	Limited experience with treatment, decreasing efficacy with increasing nodule size, operator dependency, painful (reducing compliance), risk of thyrotoxicosis and vocal cord paralysis (1–2%), seepage of ethanol[b], cytological/histological interpretation impeded in treated nodules, repeat injections often needed. Abandoned by most in solid thyroid nodules
Thermal ablation[c]	Outpatient, relatively low cost, no hypothyroidism, nodule reduced by >40% in 6 months	Limited availability and experience with treatment, operator dependency, cytological/histological interpretation impeded in treated nodules, risk of vocal cord paralysis (1%) following radiofrequency ablation

[a] Treatment of the autonomous thyroid nodule.

[b] Side effects due to ethanol escaping outside the nodule or drainage of ethanol are rare (<1%) and comprise nerve damage, perinodular/periglandular fibrosis jeopardizing subsequent surgery, thrombosis of the jugular vein, and neck haematomas.

[c] Thermal techniques are still experimental and have only been introduced in a limited number of centres. The advantages are similar to those of ethanol injection, but side effects are fewer due to the higher degree of control which limits the risk of extranodular damage.

Surgery

Surgical therapy has so far constituted the mainstay of therapy. The main indications for surgery are malignant or suspicious cytological features and symptoms due to the nodule itself. Certain clinical features raising the suspicion of thyroid malignancy (**Box 3.5.3.1**) lowers the threshold for suggesting surgery despite a benign cytology as recommended by most European thyroidologists [13]. The frequency of complications due to surgery decreases with increasing experience and specialized training. Results from a Danish national thyroid surgery database (6859 patients treated with thyroid surgery) indicate very low complication rates. Thus, recurrent laryngeal nerve injury is seen in 2%, risk factors being malignant histology, neck dissection and extent of the procedure, and favouring the use of intraoperative nerve monitoring [26]. Similarly, the risk of hypocalcaemia, wound haematomas and infections are less than 2–3% [26]. Patients with benign cytology, in whom clinical suspicion results in referral for surgery, may generally be managed with lobectomy (hemithyroidectomy). L-thyroxine postoperatively to prevent regenerative hyperplasia is not recommended routinely in the euthyroid patient [1, 27].

Thyroid Hormone Suppressive Therapy

TSH suppression is intended to shrink or slow the growth of thyroid nodules, and also to prevent the occurrence of new nodules. However, most evidence suggests that changes in nodule size are similar in TSH suppressed and control groups, and treatment seems at best beneficial in a subgroup of patients with smaller solid nodules [12]. Twenty per cent (20%) or less of solitary nodules will regress as a result of L-thyroxine treatment, and regrowth is seen after cessation of therapy [12]. Long-term results confirm that the nodule-reducing effect of L-thyroxine is insignificant [28]. TSH suppression may, however, have adverse effects. Because suppressive treatment (by definition) produces subclinical hyperthyroidism, treated patients are at increased risk of, e.g. atrial fibrillation, other cardiac abnormalities, reduced bone density, and also excess mortality [8]. These side effects, combined with the questionable efficacy, have led to recommendations that vary depending upon the age, sex, and menopausal status of the patient. TSH suppressive thyroxine therapy is least tempting in elderly patients and in postmenopausal women. It should be reserved for small nodules—where treatment is least necessary—in younger patients living in borderline iodine-deficient areas. Based on management guidelines, routine suppression therapy of benign thyroid nodules has been discarded [6, 11].

Percutaneous Tissue Ablation with Ethanol

Ethanol (70–100%) can cause permanent tissue ablation due to coagulative necrosis and local small vessel thrombosis. It has proved useful in the treatment of autonomously functioning thyroid nodules, cystic thyroid nodules, and solid cold thyroid nodules [12]. Using multiple ethanol injections, complete cure (normalization of scintigraphy and serum TSH) can be achieved in two-thirds of patients with toxic nodules and three out of four patients with pretoxic nodules. A single ethanol instillation in recurrent thyroid cysts leads to remission in 80% of patients [29].

Limitations are the need to repeat ethanol injections to achieve complete cure in toxic and pretoxic nodules and to prevent renewed growth in solid cold nodules. Furthermore, the procedure is often painful despite local anaesthesia. Therefore, this procedure is only indicated in benign recurrent cystic nodules [30].

Percutaneous Thermal Procedures

Following an approximately 10-minute session of US-guided laser thermal ablation, a 50% reduction (very similar to that of ethanol injection) in nodule volume can be achieved [31]. This effect seems independent of whether the nodule is hot or cold. One or two additional sessions augment this effect by up to 30% [31]. The main indications for laser ablation are partly cystic nodules resulting in a successful outcome in 70% of patients [15]. Side effects are mainly related to various degrees of local and irradiating pain, which is much milder than with ethanol injection.

Radiofrequency ablation (RFA) is another technique that induces thermal injury to thyroid nodules through the deposition of

electromagnetic energy. Clinical outcome is similar to that obtained by laser ablation, but at a higher cost. Overall complication rate is 3%, with pain being the most common complaint. The major risks include injury to the laryngeal and vagus nerves, although this is rare (1%) [32].

The US-guided thermal procedures also include microwave and high-frequency US therapy with which experience is limited. The techniques are only available in a limited number of centres. And remain experimental to most patients [30].

Radioactive Iodine (Hot Nodule)

In the clinically euthyroid patient, autonomous thyroid nodules may present as a hot lesion on scintigraphy with varying degrees of extranodular suppression. Most of these patients have suppressed serum TSH (see also Chapters 3.3.4 and 3.3.11). Treatment may be dictated either by the nodule size, causing compression of the adjacent structures, or cosmetic disturbances. Additionally, treatment is given to prevent thyrotoxicosis (annual risk about 5%), particularly in older people and those with heart disease [12, 33]. A cure rate (normalization of scintigraphy and serum TSH) of 75% and volume reduction of 40% following a single dose of radioiodine can be anticipated [12, 33]. Side effects are few and consist of hypothyroidism in about 10% after 5 years and seem unrelated to any type of dose planning [33]. Treatment must be individualized based on patient preference and risk factors for adverse effects. Table 3.5.3.2 summarizes the advantages and drawbacks of the treatment options.

The Future

While FNA cytology is currently the most sensitive and specific tool to select thyroid nodules for surgery, after prioritization by assessment of standardized US malignancy criteria, it is characterized by inherent limitations resulting in many 'indeterminate' cytologies. The recent application of molecular analyses has provided a novel opportunity to better 'rule in' or 'rule out' malignancy [22]. Commercially available gene expression classifiers to exclude malignancy as well as gene panels (detection of cancer-specific mutations) to detect malignancy are currently introduced in the daily clinic in the United States but in general the use is limited due to the high cost and limited validation [6, 11].

REFERENCES

1. Hegedüs L. Clinical practice. The thyroid nodule. *N Engl J Med*, 2004; **351**: 1764–71.
2. Burman KD, Wartofsky L. Clinical practice. Thyroid nodules. *N Engl J Med*, 2015; **373**: 2347–56.
3. Frates MC, Benson CB, Doubilet PM, *et al*. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab*, 2006; **91**: 3411–17.
4. Hagag P, Strauss S, Weiss M. Role of ultrasound-guided fine-needle aspiration biopsy in evaluation of nonpalpable thyroid nodules. *Thyroid*, 1998; **8**: 989–95.
5. Jarlov AE, Nygaard B, Hegedüs L, Hartling SG, Hansen JM. Observer variation in the clinical and laboratory evaluation of patients with thyroid dysfunction and goiter. *Thyroid*, 1998; **8**: 393–98.
6. Haugen BR, Alexander EK, Bible KC, *et al*. 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 2016; **26**: 1–133.
7. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J*, 2017; **6**: 225–37.
8. Lillevang-Johansen M, Abrahamsen B, Jorgensen HL, Brix TH, Hegedüs L. Excess mortality in treated and untreated hyperthyroidism is related to cumulative periods of low serum TSH. *J Clin Endocrinol Metab*, 2017; **102**: 2301–9.
9. McLeod DS, Watters KF, Carpenter AD, Ladenson PW, Cooper DS, Ding EL. Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. *J Clin Endocrinol Metab,* 2012; **97**: 2682–92.
10. Paschke R, Hegedüs L, Alexander E, Valcavi R, Papini E, Gharib H. Thyroid nodule guidelines: agreement, disagreement and need for future research. *Nat Rev Endocrinol*, 2011; **7**: 354–61.
11. Gharib H, Papini E, Garber JR, *et al*. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. *Endocr Pract*, 2016; **22**: 622–39.
12. Hegedüs L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. *Endocr Rev*, 2003; **24**: 102–32.
13. Bennedbaek FN, Perrild H, Hegedüs L. Diagnosis and treatment of the solitary thyroid nodule. Results of a European survey. *Clin Endocrinol (Oxf)*,1999; **50**: 357–63.
14. Bennedbaek FN, Nielsen LK, Hegedüs L. Effect of percutaneous ethanol injection therapy versus suppressive doses of L-thyroxine on benign solitary solid cold thyroid nodules: a randomized trial. *J Clin Endocrinol Metab*, 1998; **83**: 830–5.
15. Dossing H, Bennedbaek FN, Hegedüs L. Interstitial laser photocoagulation (ILP) of benign cystic thyroid nodules-a prospective randomized trial. *J Clin Endocrinol Metab*, 2013; **98**: E1213–17.
16. Soelberg KK, Bonnema SJ, Brix TH, Hegedüs L. Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review. *Thyroid*, 2012; **22**: 918–25.
17. Burch HB, Burman KD, Cooper DS, Hennessey JV, Vietor NO. A 2015 survey of clinical practice patterns in the management of thyroid nodules. *J Clin Endocrinol Metab*, 2016; **101**: 2853–62.
18. Pitman MB, Abele J, Ali SZ, *et al*. Techniques for thyroid FNA: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diagn Cytopathol*, 2008; **36**: 407–24.
19. Pusztaszeri M, Rossi ED, Auger M *et al*. The Bethesda system for reporting thyroid cytopathology: proposed modifications and updates for the second edition from an international panel. *Acta Cytol*, 2016; **60**: 399–405.
20. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol*, 2012; **56**: 333–9.
21. Yoo SK, Lee S, Kim SJ, *et al*. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet*, 2016; **12**: e1006239.

22. Paschke R, Cantara S, Crescenzi A, Jarzab B, Musholt TJ, Sobrinho SM. European Thyroid Association guidelines regarding thyroid nodule molecular fine-needle aspiration cytology diagnostics. *Eur Thyroid J*, 2017; **6**: 115–29.
23. Eszlinger M, Hegedüs L, Paschke R. Ruling in or ruling out thyroid malignancy by molecular diagnostics of thyroid nodules. *Best Pract Res Clin Endocrinol Metab,* 2014; **28**: 545–57.
24. Russ G, Leboulleux S, Leenhardt L, Hegedüs L. Thyroid incidentalomas: epidemiology, risk stratification with ultrasound and workup. *Eur Thyroid J*, 2014; **3**: 154–63.
25. Pustelnik FS, Gronbek C, Dossing H, *et al*. The compensatory enlargement of the remaining thyroid lobe following hemithyroidectomy is small and without impact on symptom relief. *Eur Arch Otorhinolaryngol*, 2018; **275**: 161–7.
26. Godballe C, Madsen AR, Sorensen CH, *et al*. Risk factors for recurrent nerve palsy after thyroid surgery: a national study of patients treated at Danish departments of ENT Head and Neck Surgery. *Eur Arch Otorhinolaryngol*, 2014; **271**: 2267–76.
27. Hegedüs L, Nygaard B, Hansen JM. Is routine thyroxine treatment to hinder postoperative recurrence of nontoxic goiter justified? *J Clin Endocrinol Metab*, 1999; **84**: 756–60.
28. Papini E, Petrucci L, Guglielmi R, *et al*. Long-term changes in nodular goiter: a 5-year prospective randomized trial of levothyroxine suppressive therapy for benign cold thyroid nodules. *J Clin Endocrinol Metab*, 1998; **83**: 780–3.
29. Bennedbaek FN, Hegedüs L. Treatment of recurrent thyroid cysts with ethanol: a randomized double-blind controlled trial. *J Clin Endocrinol Metab*, 2003; **88**: 5773–7.
30. Gharib H, Hegedüs L, Pacella CM, Baek JH, Papini E. Clinical review: nonsurgical, image-guided, minimally invasive therapy for thyroid nodules. *J Clin Endocrinol Metab*, 2013; **98**: 3949–57.
31. Dossing H, Bennedbaek FN, Hegedüs L. Long-term outcome following interstitial laser photocoagulation of benign cold thyroid nodules. *Eur J Endocrinol*, 2011; **165**: 123–8.
32. Baek JH, Lee JH, Sung JY, *et al*. Complications encountered in the treatment of benign thyroid nodules with US-guided radiofrequency ablation: a multicenter study. *Radiology*, 2012; **262**: 335–42.
33. Bonnema SJ, Hegedüs L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*, 2012; **33**: 920–80.

3.5.4 Pathogenesis of Thyroid Cancer

Massimo Santoro, Barbara Jarzab, Jolanta Krajewska, and Dagmara Rusinek

Introduction

Thyroid cancer (TC), mainly papillary TC (PTC), demonstrates one of the most rapid increases among all malignant neoplasms. Although the widespread use of diagnostic tools (ultrasound, fine needle aspiration biopsy) is considered as a main reason of the growing TC incidence, the role of environmental factors is also emphasized as suggested by the altered relative prevalence of TC-associated genetic lesions, such as *BRAF* and *RAS* point mutations and RET/PTC rearrangements observed during recent years [1, 2].

Numerous studies resulted in a remarkably better understanding of the biology and molecular pathogenesis of TC. Among the most documented factors predisposing to TC are radiation exposure, iodine diet content, or pre-existing thyroid disease. A potential role of hormones, chemicals, lifestyle, and obesity has also been envisaged [3]. The molecular analyses of TCs, in turn, showed a number of genetic alterations, concerning mainly oncogenes, tumour suppressor genes, epigenetic changes, or alterations in miRNA expression, which altogether are responsible for altered cellular pathways, loss of cell-cycle control, and immortalization of cancer cells. Although many of these alterations are common in different TC types, advanced genome and transcriptome studies revealed heterogeneous nature of these tumours. This chapter summarizes the current concepts regarding the role of the environment and other factors in TC development as well as its molecular pathogenesis (Figure 3.5.4.1).

Differentiated TC (DTC), arising from follicular cells, is the most common. DTC includes PTC, follicular (FTC), and Hürthle cell (HCC) carcinoma. PTC is characterized by typical nuclear features, including enlargement, elongation, overlapping, and clearing, the presence of pseudoinclusions and nuclear grooves with rather rare mitotic figures [4]. Numerous PTC variants have been recognized.

FTC is the second most common DTC type. Its diagnosis requires the presence of capsular and/or vascular invasion and the lack of PTC features. However, in some cases, unequivocal differentiation between FTC and benign follicular adenomas (FTA) is not possible even at histopathological evaluation.

Significant changes in the classification of thyroid tumours have been proposed in a new 2017 WHO classification. A new entity, named 'non-invasive follicular thyroid neoplasm with papillary-like nuclear features' (NIFTP), was separated from PTC [5, 6]. Another important change is the introduction of a group of borderline thyroid tumours placed between FTA and FTC or FV-PTC. This new entity includes follicular tumours and well-differentiated tumours of uncertain malignant potential (FT-UMP and WDT-UMP, respectively), that display questionable capsular or vascular invasion, irrespective of the presence or absence of the nuclear features typical for PTC [7].

Hürthle cell (oncocytic) carcinoma (HCC) is characterized by large cells with granular cytoplasm and large nuclei, often with prominent nucleoli [7].

Poorly differentiated (PDTC) and anaplastic (ATC) thyroid carcinomas are rare and aggressive thyroid malignancies that, differently from DTC, are refractory to radioiodine. PDTC is behaviourally intermediate between differentiated (FTC and PTC) carcinomas and ATC [7]. Morphologically, PDTC may display

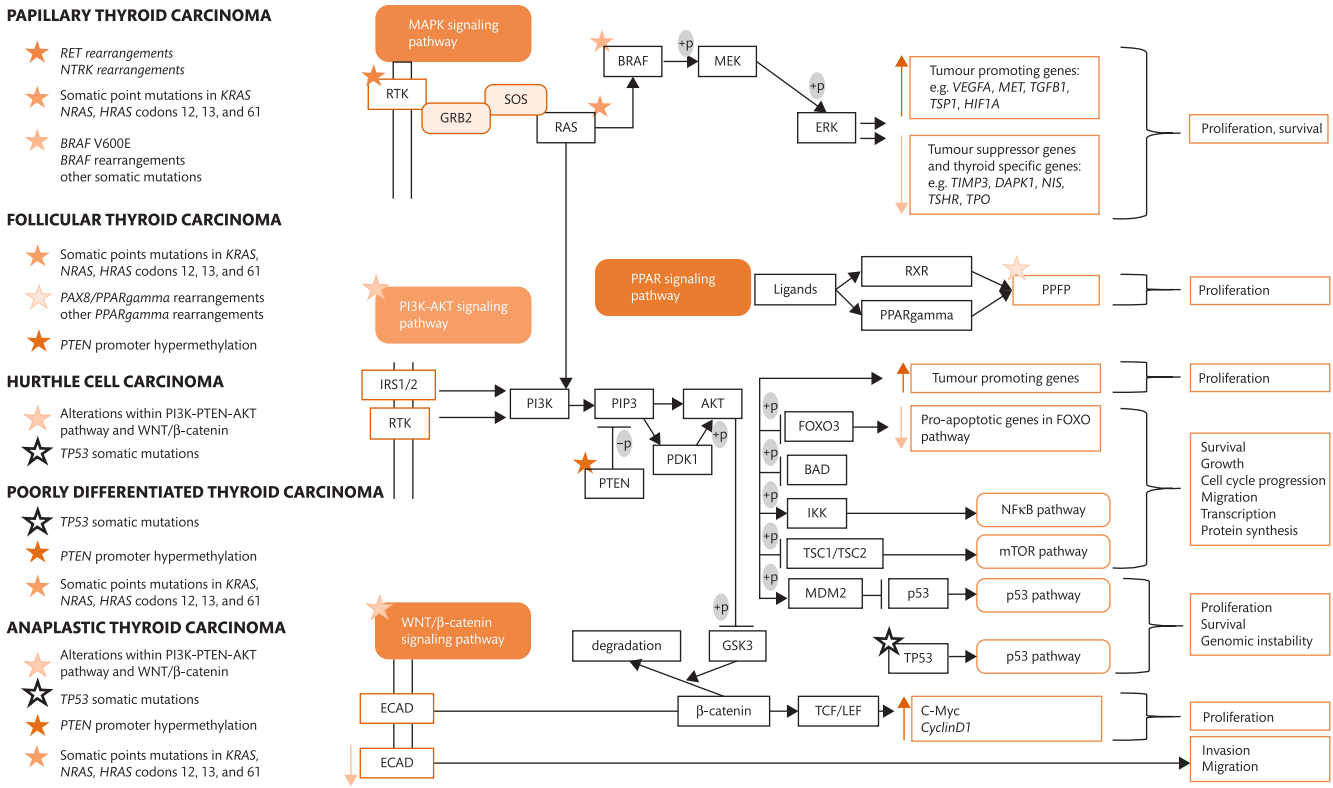


Figure 3.5.4.1 The overview of main signaling pathways altered in thyroid carcinomas. The scheme demonstrates key signal transduction pathways and mechanisms, including major genetic alterations, involved in the regulation of gene expression, the progression of the cell cycle, proliferation, cell migration, and apoptosis that are disrupted in thyroid cancers. Black arrows indicate activation and truncated lines inactivation; light orange arrows = gene downregulation; dark orange arrows = gene upregulation; stars = genetic alterations listed on the left side of the scheme.

insular growth and usually exhibit severe nuclear atypia, mitotic features, necrosis, and vascular invasion, which in ATC are much more intense [7].

Factors Predisposing to Thyroid Cancer

Radiation exposure of the thyroid gland is the most established PTC risk factor. The thyroid is particularly vulnerable to ionizing radiation, with the paediatric population being the most sensitive. Similar to many organs of the body, ionizing radiation may lead to thyroid DNA damage. Depending on the modality and severity of irradiation, it may cause cell death or, if less severe, it may result in specific genetic abnormalities. Accordingly, radiation exposure has been associated with increased formation of oncogenic rearrangements in TC [8].

Iodine intake demonstrates a well-known impact on the incidence and prevalence of thyroid disease and TC. It is believed that iodine prophylaxis changed the epidemiology of DTC. However, there is still a lack of evidence-based data supporting this hypothesis. Iodine deficiency is associated with an increased FTC risk, whereas PTC is more prevalent in regions with sufficient iodine supplementation. Also, the question about the potential role of iodine intake in the increased TC incidence and mechanisms linking the iodine intake and TC remains open. Chronic thyroid-stimulating hormone (TSH) stimulation and *BRAF* mutations are among considered pathways [2, 9].

Some studies demonstrated an increased TC risk in obese individuals [10, 11], probably related to insulin resistance [3].

Information regarding food factors is as yet not conclusive. Some analyses demonstrated decreased TC risk in individuals on fruit, raw, or mixed vegetable diet. Conversely, cruciferous vegetables may increase TC risk [12]. The data concerning the impact of meat or dairy consumption are also inconclusive [12].

One of the highest TC incidences worldwide is noticed in volcanic regions. Active volcanoes produce numerous toxic compounds that might contaminate cultivated fields and affect the animal and food chain that could be involved in thyroid cancerogenesis [3].

Xenobiotics—exogenous compounds and chemicals—may competitively inhibit sodium/iodide symporter, thyroid peroxidase, deiodinase, binding of thyroid hormones to transport proteins, or to thyroid hormone receptors or induce hepatic metabolism of thyroid hormones leading to decreased serum thyroxine concentration and increased TSH level. The question of whether xenobiotics promote or not TC has not been addressed yet [3].

There is a little evidence of a potential association of herpes viruses and DTC, and of a possible role of Epstein–Barr virus in thyroid tumorigenesis [3].

Another important issue is the autoimmune Hashimoto's thyroiditis (HT). Elevated TSH levels, stimulating follicular cells proliferation and promoting TC development, secretion of proinflammatory cytokines, and oxidative stress are considered as plausible mechanisms linking TC risk and HT [9, 13]. Interestingly, while studies using fine-needle aspiration biopsy (FNAB) specimens did not demonstrate an

increased PTC risk in HT, surgical series with histopathological assessment found the coexistence of PTC and HT [9]. However, biased results should be considered.

As TC is more frequent in women, oestrogens are considered a possible risk factor. Oestrogens and cognate receptors are important factors influencing proliferation, migration, and invasion of TC cells *in vitro* [14, 15]. A comparison between 40 TC women and 40 age-matched controls showed that oestrogen metabolism was unbalanced in TC and suggested a possible role of oestrogen-DNA adducts in TC initiation [16]. A significantly higher TC risk was observed in postmenopausal women with hysterectomy comparing to women without hysterectomy, regardless of the ovarian status. Interestingly, when a hysterectomy group was considered alone, hormone therapy was related to a lower TC risk [17]. According to the recently published review, there was a weak association between DTC risk and menstrual or menopausal factors, oral contraceptives, and hormone replacement therapy [18]. Regarding pregnancy, the question of a possible association between the parity and DTC risk remains unanswered. Some reports suggested a possible impact of pregnancy-related factors on DTC progression and short-term outcome in comparison with non-pregnant controls. On the other hand, there is probably no impact of pregnancy on DTC-related death or overall survival [18]. In conclusion, further studies are necessary to provide more direct evidence of the oestrogenic impact on thyroid tumorigenesis.

Medullary thyroid cancer (MTC) in up to 25% of cases is related to a genetic predisposition in the frame of MEN 2A and MEN 2B syndromes caused by germline mutations of the *RET* gene. Non-medullary TC may also develop in some cases in the frame of complex genetic syndromes, including familial adenomatous polyposis, Cowden syndrome, or Carney's complex. Recently, a familial *DICER1* tumour syndrome caused by germline inactivating *DICER1* mutations has been defined. *DICER1* is involved in microRNAs maturation by cleaving ribonucleic acid (RNA) precursors into siRNA and miRNA [19]. Unfortunately, besides these particular conditions, genetic heterogeneity makes it difficult to identify the TC cause in most individual families.

Molecular Pathogenesis of TC

Two models of thyroid carcinogenesis have been proposed, the fetal and the multistep theory [20, 21]. The first one assumes that cancer cells are derived from remnants of the fetal thyroid cells: thyroblasts are PTC precursor cells, prothyrocytes are FTC precursor cells, whereas thyroid stem cells are ATC precursor cells [21]. While the multistep theory is based on genomic alterations, which are indispensable in the carcinogenesis process. This theory hypothesizes that most TCs arise from a single abnormal cell and their progression is a consequence of the accumulation of further genetic alterations. According to this theory, FTC and HCC arise from thyroid adenomas following specific molecular events, while PTCs originate from a thyrocyte carrying a particular mutation. PDTC and ATC, in turn, develop from PTC and FTC through their dedifferentiation [22–24]. Each step of carcinogenesis in this theory is related to the acquisition of subsequent molecular events, which make altered cells more and more aggressive and enable them to escape from defence mechanisms, senescence, and apoptosis. The multistep theory is well supported by experimental data; it is, however, important to highlight that not every adenoma will transform into TC, so these lesions should not be treated as unequivocally precancerous ones. In this frame, the previously mentioned NIFTP may be a precursor lesion of FV-PTC. On the other hand, there is no recognized precursor of CV-PTC, though microcarcinomas are regarded as possible precursors of full-blown CV-PTCs [5].

The majority of genetic changes found in TC concern two types of alterations: gain-of-function mutations in proto-oncogenes, responsible for cell growth, proliferation, and survival stimulation and loss-of-function mutations in tumour suppressor genes that normally negatively regulate cellular growth, cell cycle, and promote DNA repair. Conversion of a proto-oncogene into an active oncogene may result from a point mutation leading to a constitutively acting protein product. This situation is observed in the case of *RAS* genes or the *BRAF* gene, which encode proteins of the MAPK pathway, a critical pathway in cell proliferation, differentiation, and survival. Alternatively, oncogene formation in TC can be mediated by translocations of DNA fragments within or between chromosomes, leading to structural gene rearrangements. These alterations lead to fusions of 5' end of a normally expressed gene with 3' end of a silent gene, usually having kinase activity, which results in constitutive activation of the latter one. In PTC this mechanism leads, e.g. to the *RET* receptor tyrosine kinase gene (another MAPK pathway component) activation. The loss of tumour suppressor function, in turn, usually requires a 'two-hit' ('Knudson') mechanism that leads to impaired function of both alleles coding a particular protein (e.g., loss of genetic material from one chromosome with additional loss or mutation on the other one). There are exceptions from this 'two-hit' mechanism, the most characteristic being represented by the *TP53* gene. Accordingly, specific mutations in the *TP53* gene result in a mutated p53 protein that can prevent the function (dominant negative) of the normal protein encoded by the wild-type allele [25]. Oncogenes and tumour suppressor genes may also be regulated through epigenetic modifications [26]. DNA hypomethylation of CpG-rich regions (CpG islands) in regulatory promoter regions of specific proto-oncogenes leads to their upregulated transcription, whereas DNA hypermethylation may lead to tumour suppressor gene silencing. Epigenetic modifications are also considered as one of the potential mechanisms responsible for thyroid-specific gene silencing and dedifferentiation. Some thyroid-specific genes, like *NIS* or *TSHR*, undergo DNA hypermethylation, which supports the recognition of their aberrant methylation as TC progression marker [26]. Epigenetic changes also include aberrant miRNAs expression. miRNA are small non-coding RNAs that regulate gene expression by binding to the 3' untranslated region of the mRNA targets or, more rarely, to the coding sequence of the target gene [27]. Alterations in miRNAs expression were demonstrated to play a crucial role in cancer cell proliferation, migration, invasion, differentiation, and survival in different cancer types, including TCs.

Molecular Events in PTC

According to the Cancer Genome Atlas (TCGA) study, PTC demonstrates a relatively low mutation rate (11 non-synonymous mutations per tumour; 0.41 mutations per Mb on average), compared to other solid tumours [28]. This may explain the indolent behaviour of most PTCs, although it should be recognized that as aggressive malignancies as leukaemias also carry a reduced mutation burden.

The TCGA study revealed a number of new somatic point mutations and other genetic alterations, which reduced the fraction of PTCs with unknown genetic background approximately to 3.5% [28]. Among commonly mutated genes are those encoding proteins of the MAPK pathway: *BRAF, KRAS, HRAS*, and *NRAS*, whose mutations were present in 74.6% of PTCs, but also *EIF1AX, PPM1D*, and *CHEK2*. The most frequent molecular event in PTC is the $BRAF^{V600E}$ (valine-to-glutamate) point mutation (36–83% PTCs) [29]. The *BRAF* gene encodes a serine-threonine kinase located downstream to RAS in the MAPK pathway, and its activation leads to the phosphorylation of MEK, which, in turn, mediates phosphorylation of MAPKs (ERK) kinases and finally transcription of specific genes. The $BRAF^{V600E}$ mutation is responsible for the constitutive RAS-independent and dimerization-independent activation of BRAF kinase and consequently of the MAPK pathway. This genetic alteration is capable to initiate PTC, as it was demonstrated *in vivo* in mouse models [30, 31], and was described as an early molecular event [32]. However, some studies showed the presence of $BRAF^{V600E}$ mutation in metastatic lymph nodes but not in matched primary thyroid tumour, thus suggesting that the mutation may have arisen *de novo* during progression [33, 34]. Some studies have reported the presence of subclonal *BRAF* mutations in PTC, thus raising the possibility they can be a progression rather than initiating events [35, 36]. However, the TCGA and other studies indicated the clonal nature of $BRAF^{V600E}$ mutation in PTC [28, 37]. Theoretically, there can be a third possibility that $BRAF^{V600E}$ mutation initiates PTC, however, powerful secondary alterations occur during the tumorigenesis and take over cancer progression as drivers, and the *BRAF* mutation is removed by DNA repair mechanisms [38]. $BRAF^{V600E}$ shows a strong association with PTC phenotype being most frequent in its CV-PTC [29] and TCV-PTC (more than 90% of cases) variants [39, 40]. Moreover, many studies demonstrated that the presence of $BRAF^{V600E}$ was associated with poorer PTC outcomes, including aggressive pathological features, loss of sensitivity to radioiodine, or increased recurrence risk [41]. Other *BRAF* point mutations (including K601E) are rather uncommon, less potent MAPK drivers, and usually present in encapsulated FV-PTC [42].

RAS somatic point mutations are mainly restricted to FV-PTC with a frequency of 10% to 21% [43, 44]. *RAS* genes (*NRAS, HRAS, KRAS*) encode members of small GTPase superfamily, which act as molecular transmitters of signals from tyrosine kinase and non-tyrosine kinase receptors to MAPK and PI3K-AKT pathways. *NRAS* is predominantly mutated in TC, mostly in codons 12 and 61.

The *EIF1AX* gene, encoding an essential eukaryotic translation initiation factor, was among the mutated genes in PTC in the TCGA study [28]. Although the frequency of its mutations did not exceed 2%, it was classified as a cancer gene associated with PTC. Further studies indicated the presence of *EIF1AX* mutations, mainly in encapsulated FV-PTC, and also in ATC [45, 46]. The presence of *EIF1AX* mutations in FTA and their absence in FTC may suggest that *EIF1AX*-mutated adenomas may progress to FV-PTC rather than to FTC [45].

Recently *TERT* promoter mutations were detected in a number of cancers, including TCs [47, 48]. The *TERT* gene encodes the reverse transcriptase component of telomerase with the main function, together with its RNA component, of telomere elongation by adding TTAGGG at the end of chromosomes [49]. In addition to unlimited replication, telomerase activation leads to increased proliferation, angiogenesis, resistance to apoptosis, and metastatic potential of cancer cells [50]. Two *TERT* promoter mutations, 1,295,228 C>T and 1,295,250 C>T (–124 and –146 bp from the ATG; termed C228T and C250T, respectively) are common. Both alterations create the 11-nucleotide fragment (5'-CCCCTTCCGGG-3'), which contains a consensus binding site, GGAA, for ETS (E26 transformation specific) transcription factors [48]. It has been demonstrated that both C228T and C250T increase the transcriptional activity of the *TERT* promoter [47]. *TERT* promoter mutations are present in 11.3% of PTCs, with C228T being more frequent than C250T. The frequency is correlated with aggressiveness of the tumour, and it is higher in TCV-PTC, showing an aggressive nature [48, 49]. Recent studies indicated the association of *TERT* promoter mutations with the presence of $BRAF^{V600E}$ mutation and their coexistence with poorer outcomes compared to the impact of each of these alterations alone. Similar consequences were observed in the coexistence of *TERT* promoter and *RAS* mutations [49, 51].

RET rearrangements are characteristic of radiation-induced PTCs. These chromosomal aberrations concern mainly young patients since thyroid cells of children are particularly sensitive to radiation. *RET* encodes a transmembrane receptor with tyrosine kinase activity, and normally, in contrast to C cells, it is not expressed in follicular thyroid cells. Activation of *RET* in these cells results from a fusion of *RET* fragment encoding domain with tyrosine kinase activity with 5' end of other normally expressed genes [52]. More than 16 different *RET* rearrangements have been described so far with *RET/PTC1* (*CCDC6-RET*), *RET/PTC3* (*NCOA4-RET*), and *RET/PTC2* (*PRKAR1A-RET*) being the most frequent and accounting for 60%, 30% and 5% of all *RET/PTC* translocations, respectively [53]. The overall frequency of *RET* rearrangements varies between studies, from 3% to 85% in the adult population, an average of 35% [54]. Although *RET* rearrangements are the hallmark of radiation-induced PTCs, these alterations are also present in benign lesions, like in Hashimoto's disease [55, 56]. Currently, it is obvious that *RET* rearrangements are characteristic of childhood PTC, also sporadic one. In addition to *RET/PTC* rearrangements, other chromosomal aberrations are observed in PTCs, however, in a significantly lower number of cases. *NTRK* rearrangements concern mainly *NTRK1* and *NTRK3* genes with *ETV6/NTRK3* being the most frequent one in radiation-induced TC [57]. There are also *BRAF* gene rearrangements (*AKAP9/BRAF; AGK/BRAF*), detected in about 10% of radiation-induced PTC [58, 59] and *ALK* rearrangements described in 1–5% cases [60–62].

Depending on the underlying driver mutation, two molecular subtypes of PTCs have been proposed: BRAF-like (BRL) and RAS-like (RL) [28]. BRL tumours include not only carcinomas with $BRAF^{V600E}$ mutation but also tumours with *TERT* promoter mutations, *RET/PTC* and *BRAF* gene rearrangements, whereas RL tumours include all cases with *BRAF* point mutations other than V600E, *EIF1AX, RAS* point mutations, and *PAX8/PPARgamma* rearrangements. BRL tumours are most typically CV-PTC, while RL tumours have FV-PTC features. Moreover, the two molecular subtypes differ in terms of genomic, epigenetic, and proteomic profiles, however, the BRL tumours displayed a more heterogeneous nature. BRL PTCs activated mainly the MAPK signaling pathway, whereas

in the RL tumours activation of MAPK, as well as PI3K pathway was observed [28].

As far as the epigenetic modifications, increased expression of miR-146b, miR-221 and miR-222 has been documented in PTC [28, 63]. Moreover, *BRAF*^{V600E} mutation seems to be associated with hypermethylation of several tumour suppressor genes, including *TIMP3, DAPK1*, or *RARB* [64], and with hypomethylation of other genes [65].

Molecular Alterations in FTC

RAS point mutations and *PAX8/PPARgamma* (PAX8/PPARG fusion protein, PPFP) rearrangements are the most frequent molecular events in FTC. Noteworthy, these genetic lesions are also present in FTA. The frequency of *RAS* mutations in FTC is about 40%, mainly in codons 12, 13, and 61. The alterations in *NRAS* 61 codon are related to a more aggressive FTC course [44, 66]. FTCs harbouring *RAS* mutation may demonstrate a higher risk of dedifferentiation [67]. PPFP rearrangements are present in 26–56% of FTCs, in 0–13% of FTAs, 0–3% of HCCs, and 0–1% of PTCs [20]. They result from the translocation between chromosomes 2 and 3 [t(2;3)(q13;p25)] and, although the function of both genes is known, the role of PPFP fusion protein is still unclear. It leads to the downregulation of *PPARG* suppressor gene and the deregulation of *PAX8*, which may contribute to cancer development [68–71]. In addition to PPFP translocation, another *PPARG* rearrangement has been reported in FTC with *CREB3L2* gene as a fusion partner [72]. *TERT* promoter mutations are more frequent in FTC compared to PTC and concern, on average, 17% of FTC [49]. A small number of mutations recently detected in FTC involve *DICER1, IDH1, EZH1*, and genes related to the PI3K-PTEN-AKT pathway [73].

In contrary to PTC, copy number changes are common in FTC. Losses of chromosome 3p are common alterations, however, additional loss of heterozygosity (LOH) sites were detected in FTC, including 1p, 6p/q, 8p/q, 9p, 11q, 13q, 18q, and 22q [74]. Two amplifications, in turn, concerning 11p and 17q, were suggested as potential early molecular events involved in cancer initiation [75].

The knowledge about the miRNA profile in FTC is very limited. Increased expression of miR-181 and miR-200 family and downregulation of miR-199 was reported in FTC [76], and two miRNAs, miRNA-7-5p and miR-206, were proposed as potential markers able to distinguish FTA from FTC [77]. Moreover, miR-146b and miR-221, previously reported as upregulated in PTC, showed increased expression in FTC as well, which suggests these two miRNAs are common players in both DTC types [27].

Molecular Events in HCC

HCC, especially aggressive cases, display molecular signature typical for the activation of PI3K-PTEN-AKT and WNT-β-catenin pathways [78]. The presence of *TP53* mutations in 10–20% of cases speaks for its aggressive nature [79, 80]. A characteristic HCC feature is an increased mitochondrial mass [78]. This phenomenon results from mutations of genes encoding subunits of mitochondrial respiration chain complexes [81]. Mutations in mitochondrial as well as in nuclear DNA were described, however, alterations of the *GRIM-19* gene are the only nuclear gene mutations typical for oxyphilic tumours [53].

Molecular Basis of PDTC and ATC

PDTC and ATC may develop through dedifferentiation of DTC with *BRAF*^{V600E} noticed in 33% and 55% of cases, respectively, and *RAS* mutations present in 29% and 52%, respectively [44, 46, 82, 83]. *BRAF* and *RAS* mutations are frequently accompanied by *TERT* promoter mutations, which are described in 40% and 73% of PDTC and ATC, respectively [46, 48, 82, 84]. Importantly, *TERT* promoter mutations, which are subclonal in PTC, become clonal in PDTC and ATC [82]. *EIF1AX* mutations, in turn, are present in 11% of these malignancies, and in most cases, coexist with *RAS* mutations [46, 82]. *TP53* mutations, with the frequency of 59%, are characteristic for ATC, however, they are also present in PDTC but with a significantly lower frequency (10%) [46, 82, 83]. These alterations may reflect genomic instability and are considered as a late molecular event in TC progression. In contrast to DTC, ATC is associated with more frequently altered PI3K-PTEN-AKT pathway, WNT-β catenin pathway, SWI/SNF complex, and mutations of genes responsible for repair mechanisms [85].

Copy number changes are common alterations in PDTC and ATC. PDTC was described to harbour chromosomal deletions 1p, 13q, and 15q not associated with any known driver mutation and 22q deletion associated to *RAS* mutations [82]. Amplification of 20q and deletions 8p and 17p were found in ATC.

miRNAs profile is strongly deregulated in PDTC and ATC, with many studies reporting a number of miRNAs with altered expression in these tumours. miR-183-3p was reported to be upregulated in PDTC, whereas miR-150 and miR-23b were downregulated, with the latter two being associated with tumour relapse and mortality [27]. The downregulation of miR-200c was observed in ATC. This particular miRNA is related to *TP53* mutations as it is transcriptionally regulated by *TP53*. miR-30 family distinguishes undifferentiated from well-differentiated thyroid carcinomas with downregulation of miR-30 only in ATC. Noteworthy, the miR-17-92 cluster, which expression is stimulated by *BRAF* ^{V600E}, may play an important role in cancer initiation and progression to ATC [27].

REFERENCES

1. Jung CK, Little MP, Lubin JH, *et al.* The increase in thyroid cancer incidence during the last four decades is accompanied by a high frequency of BRAF mutations and a sharp increase in RAS mutations. *J Clin Endocrinol Metab*, 2014; **99**: E276–85.
2. Guan H, Ji M, Bao R, *et al.* Association of high iodine intake with the T1799A BRAF mutation in papillary thyroid cancer. *J Clin Endocrinol Metab*, 2009; **94**: 1612–17.
3. Marcello MA, Malandrino P, Almeida JFM, *et al.* The influence of the environment on the development of thyroid tumors: a new appraisal. *Endocr Relat Cancer*, 2014; **21**: T235–54.
4. Livolsi VA. Papillary thyroid carcinoma: an update. *Mod Pathol*, 2011; **24**: S1–9.
5. Nikiforov YE, Seethala RR, Tallini G, *et al.* Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol*, 2016; **2**: 1023–9.
6. Thompson LD. Ninety-four cases of encapsulated follicular variant of papillary thyroid carcinoma: a name change to noninvasive follicular thyroid neoplasm with papillary-like nuclear

features would help prevent overtreatment. *Mod Pathol*, 2016; **29**: 698–707.
7. Lloyd RO, Kloppel G, Rosai J. *WHO Classification of Tumours of Endocrine Organs/IARC Classification of Tumours*, 4th edn. Geneva: WHO, 2017.
8. Efanov AA, Brenner A V, Bogdanova TI, *et al*. Investigation of the relationship between radiation dose and gene mutations and fusions in post-Chernobyl thyroid cancer. *J Natl Cancer Inst*, 2018; **110**: 371–8.
9. Liu Y, Su L, Xiao H. Review of factors related to the thyroid cancer epidemic. *Int J Endocrinol*, 2017; **2017**: 5308635.
10. Leitzmann MF, Brenner A, Moore SC, *et al*. Prospective study of body mass index, physical activity and thyroid cancer. *Int J Cancer*, 2010; **126**: 2947–56.
11. Rinaldi S, Lise M, Clavel-Chapelon F, *et al*. Body size and risk of differentiated thyroid carcinomas: findings from the EPIC study. *Int J Cancer*, 2012; **131**: E1004–14.
12. Choi WJ, Kim J. Dietary factors and the risk of thyroid cancer: a review. *Clin Nutr Res*, 2014; **3**: 75.
13. Pellegriti G, Frasca F, Regalbuto C, *et al*. Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. *J Cancer Epidemiol*, 2013; **2013**: 965212.
14. Rajoria S, Suriano R, Shanmugam A, *et al*. Metastatic phenotype is regulated by estrogen in thyroid cells. *Thyroid*, 2010; **20**: 33–41.
15. Derwahl M, Nicula D. Estrogen and its role in thyroid cancer. *Endocr Relat Cancer*, 2014; **21**: T273–83.
16. Zahid M, Goldner W, Beseler CL, *et al*. Unbalanced estrogen metabolism in thyroid cancer. *Int J cancer*, 2013; **133**: 2642–9.
17. Luo J, Hendryx M, Manson JE, *et al*. Hysterectomy, oophorectomy, and risk of thyroid cancer. *J Clin Endocrinol Metab*, 2016; **101**: 3812–19.
18. Moleti M, Sturniolo G, Di Mauro M, *et al*. Female reproductive factors and differentiated thyroid cancer. *Front Endocrinol (Lausanne)*, 2017; **8**: 111.
19. Solarski M, Rotondo F, Foulkes WD, *et al*. DICER1 gene mutations in endocrine tumors. *Endocr Relat Cancer*, 2018; **25**: R197–208.
20. Parameswaran R, Brooks S, Sadler GP. Molecular pathogenesis of follicular cell derived thyroid cancers. *Int J Surg*, 2010; **8**: 186–93.
21. Takano T. Fetal cell carcinogenesis of the thyroid: a modified theory based on recent evidence. *Endocr J*, 2014; **61**: 311–20.
22. Namba H, Matsuo K, Fagin JA. Clonal composition of benign and malignant human thyroid tumors. *J Clin Invest*, 1990; **86**: 120–5.
23. Hicks DG, LiVolsi VA, Neidich JA, *et al*. Clonal analysis of solitary follicular nodules in the thyroid. *Am J Pathol*, 1990; **137**: 553–62.
24. Apel RL, Ezzat S, Bapat B V, *et al*. Clonality of thyroid nodules in sporadic goiter. *Diagn Mol Pathol*, 1995; **4**: 113–21.
25. Rivlin N, Brosh R, Oren M, *et al*. Mutations in the p53 tumor suppressor gene: important milestones at the various steps of tumorigenesis. *Genes Cancer*, 2011; **2**: 466–74.
26. Faam B, Ghaffari MA, Ghadiri A, *et al*. Epigenetic modifications in human thyroid cancer. *Biomed reports*, 2015; **3**: 3–8.
27. Boufraqech M, Klubo-Gwiezdzinska J, Kebebew E. MicroRNAs in the thyroid. *Best Pract Res Clin Endocrinol Metab*, 2016; **30**: 603–19.
28. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*, 2014; **159**: 676–90.
29. Tavares C, Melo M, Cameselle-Teijeiro JM, *et al*. Endocrine tumours: genetic predictors of thyroid cancer outcome. *Eur J Endocrinol*, 2016; **174**: R117–26.
30. Rusinek D, Swierniak M, Chmielik E, *et al*. BRAFV600E-associated gene expression profile: early changes in the transcriptome, based on a transgenic mouse model of papillary thyroid carcinoma. *PLoS One*, 2015; **10**: e0143688.
31. Knauf JA, Ma X, Smith EP, *et al*. Targeted expression of BRAFV600E in thyroid cells of transgenic mice results in papillary thyroid cancers that undergo dedifferentiation. *Cancer Res*, 2005; **65**: 4238–45.
32. de Biase D, Cesari V, Visani M, *et al*. High-sensitivity BRAF mutation analysis: BRAF V600E is acquired early during tumor development but is heterogeneously distributed in a subset of papillary thyroid carcinomas. *J Clin Endocrinol Metab*, 2014; **99**: E1530–8.
33. Oler G, Ebina KN, Michaluart P, *et al*. Investigation of BRAF mutation in a series of papillary thyroid carcinoma and matched-lymph node metastasis reveals a new mutation in metastasis. *Clin Endocrinol (Oxf)*, 2005; **62**: 509–11.
34. Vasko V, Hu S, Wu G, *et al*. High prevalence and possible de novo formation of BRAF mutation in metastasized papillary thyroid cancer in lymph nodes. *J Clin Endocrinol Metab*, 2005; **90**: 5265–9.
35. Walts AE, Pao A, Sacks W, *et al*. BRAF genetic heterogeneity in papillary thyroid carcinoma and its metastasis. *Hum Pathol*, 2014; **45**: 935–41.
36. Guerra A, Sapio MR, Marotta V, *et al*. The primary occurrence of BRAF(V600E) is a rare clonal event in papillary thyroid carcinoma. *J Clin Endocrinol Metab*, 2012; **97**: 517–24.
37. Ghossein RA, Katabi N, Fagin JA. Immunohistochemical detection of mutated BRAF V600E supports the clonal origin of BRAF-induced thyroid cancers along the spectrum of disease progression. *J Clin Endocrinol Metab*, 2013; **98**: E1414–21.
38. Xing M. BRAFV600E mutation and papillary thyroid cancer: chicken or egg? *J Clin Endocrinol Metab*, 2012; **97**: 2295–8.
39. Adeniran AJ, Zhu Z, Gandhi M, *et al*. Correlation between genetic alterations and microscopic features, clinical manifestations, and prognostic characteristics of thyroid papillary carcinomas. *Am J Surg Pathol*, 2006; **30**: 216–22.
40. Liu Z, Zeng W, Chen T, *et al*. A comparison of the clinicopathological features and prognoses of the classical and the tall cell variant of papillary thyroid cancer: a meta-analysis. *Oncotarget*, 2016; **8**: 6222–32.
41. Xing M, Westra WH, Tufano RP, *et al*. BRAF mutation predicts a poorer clinical prognosis for papillary thyroid cancer. *J Clin Endocrinol Metab*, 2005; **90**: 6373–9.
42. Torregrossa L, Viola D, Sensi E, *et al*. Papillary thyroid carcinoma with rare exon 15 braf mutation has indolent behavior: a single-institution experience. *J Clin Endocrinol Metab*, 2016; **101**: 4413–20.
43. Nikiforov YE, Nikiforova MN. Molecular genetics and diagnosis of thyroid cancer. *Nat Rev Endocrinol*, 2011; **7**: 569–80.
44. Howell GM, Hodak SP, Yip L. RAS mutations in thyroid cancer. *Oncologist*, 2013; **18**: 926–32.
45. Karunamurthy A, Panebianco F, J Hsiao S, *et al*. Prevalence and phenotypic correlations of EIF1AX mutations in thyroid nodules. *Endocr Relat Cancer*, 2016; **23**: 295–301.
46. Kunstman JW, Juhlin CC, Goh G, *et al*. Characterization of the mutational landscape of anaplastic thyroid cancer via whole-exome sequencing. *Hum Mol Genet*, 2015; **24**: 2318–29.

47. Huang FW, Hodis E, Xu MJ, *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science*, 2013; **339**: 957–9.
48. Liu X, Bishop J, Shan Y, *et al.* Highly prevalent TERT promoter mutations in aggressive thyroid cancers. *Endocr Relat Cancer*, 2013; **20**: 603–10.
49. Liu R, Xing M. TERT promoter mutations in thyroid cancer. *Endocr Relat Cancer*, 2016; **23**: R143–55.
50. Low KC, Tergaonkar V. Telomerase: central regulator of all of the hallmarks of cancer. *Trends Biochem Sci*, 2013; **38**: 426–34.
51. Shen X, Liu R, Xing M. A six-genotype genetic prognostic model for papillary thyroid cancer. *Endocr Relat Cancer*, 2017; **24**: 41–52.
52. Tallini G, Asa SL. RET oncogene activation in papillary thyroid carcinoma. *Adv Anat Pathol*, 2001; **8**: 345–54.
53. Acquaviva G, Visani M, Repaci A, *et al.* Molecular pathology of thyroid tumours of follicular cells: a review of genetic alterations and their clinicopathological relevance. *Histopathology*, 2018; **72**: 6–31.
54. Santoro M, Melillo RM, Fusco A. RET/PTC activation in papillary thyroid carcinoma: European Journal of Endocrinology prize lecture. *Eur J Endocrinol*, 2006; **155**: 645–53.
55. Ishizaka Y, Kobayashi S, Ushijima T, *et al.* Detection of retTPC/PTC transcripts in thyroid adenomas and adenomatous goiter by an RT-PCR method. *Oncogene*, 1991; **6**: 1667–72.
56. Rhoden KJ, Unger K, Salvatore G, *et al.* RET/papillary thyroid cancer rearrangement in nonneoplastic thyrocytes: follicular cells of Hashimoto's thyroiditis share low-level recombination events with a subset of papillary carcinoma. *J Clin Endocrinol Metab*, 2006; **91**: 2414–23.
57. Leeman-Neill RJ, Kelly LM, Liu P, *et al.* ETV6-NTRK3 is a common chromosomal rearrangement in radiation-associated thyroid cancer. *Cancer*, 2014; **120**: 799–807.
58. Ciampi R, Knauf JA, Kerler R, *et al.* Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest*, 2005; **115**: 94–101.
59. Ricarte-Filho JC, Li S, Garcia-Rendueles MER, *et al.* Identification of kinase fusion oncogenes in post-Chernobyl radiation-induced thyroid cancers. *J Clin Invest*, 2013; **123**: 4935–44.
60. Kelly LM, Barila G, Liu P, *et al.* Identification of the transforming STRN-ALK fusion as a potential therapeutic target in the aggressive forms of thyroid cancer. *Proc Natl Acad Sci U S A*, 2014; **111**: 4233–8.
61. Pérot G, Soubeyran I, Ribeiro A, *et al.* Identification of a recurrent STRN/ALK fusion in thyroid carcinomas. Toland AE (ed.). *PLoS One*, 2014; **9**: e87170.
62. Chou A, Fraser S, Toon CW, *et al.* A detailed clinicopathologic study of ALK-translocated papillary thyroid carcinoma. *Am J Surg Pathol*, 2015; **39**: 652–9.
63. He H, Jażdżewski K, Li W, *et al.* The role of microRNA genes in papillary thyroid carcinoma. *Proc Natl Acad Sci*, 2005; **102**: 19075–80.
64. Hu S, Liu D, Tufano RP, *et al.* Association of aberrant methylation of tumor suppressor genes with tumor aggressiveness and BRAF mutation in papillary thyroid cancer. *Int J Cancer*, 2006; **119**: 2322–9.
65. Hou P, Liu D, Xing M. Genome-wide alterations in gene methylation by the BRAF V600E mutation in papillary thyroid cancer cells. *Endocr Relat Cancer*, 2011; **18**: 687–97.
66. Motoi N, Sakamoto A, Yamochi T, *et al.* Role of ras mutation in the progression of thyroid carcinoma of follicular epithelial origin. *Pathol Res Pract*, 2000; **196**: 1–7.
67. Wang H-M, Huang Y-W, Huang J-S, *et al.* Anaplastic carcinoma of the thyroid arising more often from follicular carcinoma than papillary carcinoma. *Ann Surg Oncol*, 2007; **14**: 3011–18.
68. Kroll TG, Sarraf P, Pecciarini L, *et al.* PAX8-PPARgamma1 fusion oncogene in human thyroid carcinoma [corrected]. *Science*, 2000; **289**: 1357–60.
69. Gregory Powell J, Wang X, Allard BL, *et al.* The PAX8/PPARgamma fusion oncoprotein transforms immortalized human thyrocytes through a mechanism probably involving wild-type PPARgamma inhibition. *Oncogene*, 2004; **23**: 3634–41.
70. Reddi H V, McIver B, Grebe SKG, *et al.* The paired box-8/peroxisome proliferator-activated receptor-gamma oncogene in thyroid tumorigenesis. *Endocrinology*, 2007; **148**: 932–5.
71. Raman P, Koenig RJ. Pax-8-PPAR-gamma fusion protein in thyroid carcinoma. *Nat Rev Endocrinol*, 2014; **10**: 616–23.
72. Lui W-O, Zeng L, Rehrmann V, *et al.* CREB3L2-PPAR fusion mutation identifies a thyroid signaling pathway regulated by intramembrane proteolysis. *Cancer Res*, 2008; **68**: 7156–64.
73. Yoo S-K, Lee S, Kim S-J, *et al.* Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PLoS Genet*, 2016; **12**: e1006239.
74. Zitzelsberger H, Thomas G, Unger K. Chromosomal aberrations in thyroid follicular-cell neoplasia: in the search of novel oncogenes and tumour suppressor genes. *Mol Cell Endocrinol*, 2010; **321**: 57–66.
75. Da Silva L, James D, Simpson PT, *et al.* Tumor heterogeneity in a follicular carcinoma of thyroid: a study by comparative genomic hybridization. *Endocr Pathol*, 2011; **22**: 103–7.
76. Wójcicka A, Kolanowska M, Jażdżewski K. Mechanisms in endocrinology: microRNA in diagnostics and therapy of thyroid cancer. *Eur J Endocrinol*, 2016; **174**: R89–98.
77. Stokowy T, Wojtaś B, Fujarewicz K, *et al.* miRNAs with the potential to distinguish follicular thyroid carcinomas from benign follicular thyroid tumors: results of a meta-analysis. *Horm Metab Res*, 2014; **46**: 171–80.
78. Ganly I, Ricarte Filho J, Eng S, *et al.* Genomic dissection of Hürthle cell carcinoma reveals a unique class of thyroid malignancy. *J Clin Endocrinol Metab*, 2013; **98**: E962–72.
79. Evangelisti C, de Biase D, Kurelac I, *et al.* A mutation screening of oncogenes, tumor suppressor gene TP53 and nuclear encoded mitochondrial complex I genes in oncocytic thyroid tumors. *BMC Cancer*, 2015; **15**: 157.
80. Nikiforova MN, Wald AI, Roy S, *et al.* Targeted next-generation sequencing panel (thyroseq) for detection of mutations in thyroid cancer. *J Clin Endocrinol Metab*, 2013; **98**: E1852–60.
81. Maximo V, Lima J, Prazeres H, *et al.* The biology and the genetics of Hürthle cell tumors of the thyroid. *Endocr Relat Cancer*, 2012; **19**: R131–47.
82. Landa I, Ibrahimpasic T, Boucai L, *et al.* Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*, 2016; **126**: 1052–66.
83. Jeon MJ, Chun S-M, Kim D, *et al.* Genomic alterations of anaplastic thyroid carcinoma detected by targeted massive parallel sequencing in a BRAF(V600E) mutation-prevalent area. *Thyroid*, 2016; **26**: 683–90.
84. Landa I, Ganly I, Chan TA, *et al.* Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. *J Clin Endocrinol Metab*, 2013; **98**: E1562–6.
85. Xu B, Ghossein R. Genomic landscape of poorly differentiated and anaplastic thyroid carcinoma. *Endocr Pathol*, 2016; **27**: 205–12.

3.5.5 Pathology of Thyroid Cancer

Fulvio Basolo and Clara Ugolini

Introduction

Thyroid cancer is a rare neoplasia, accounting for about 1% of all malignancies and less than 0.5% of deaths for disease. However, thyroid cancer represents the most common type of endocrine malignancy, with incidence differently distributed in the various geographical areas, ranging from 0.5 to 10 cases per 100 000 population [1]. Women are more commonly affected, with rates that are two or three times higher than those of men [2].

Thyroid cancers can be classified into two broad types: primary or secondary. Primary cancers can have epithelial or non-epithelial origins. Primary epithelial cancers originate from follicular epithelial cells or from C cells. Only medullary thyroid carcinoma originates from C cells. Most primary epithelial thyroid tumours originate from follicular cells and show a variety of morphological features and clinical behaviours. They include papillary carcinoma with all its variants, follicular carcinomas, Hürthle carcinomas, poorly differentiated carcinomas, and anaplastic carcinomas. Tumours originating from follicular and from C cells (mixed medullary and follicular carcinomas) as well as primary thyroid squamous carcinomas are very rare. Non-epithelial tumours include malignant lymphomas and other mesenchymal tumours. Finally, secondary or metastatic tumours are infrequent. The most common tumours that can metastasize to thyroid are lung, kidney, breast carcinoma and also colon carcinoma [3].

The histological classification of thyroid cancer of the World Health Organization (WHO) has been recently re-elaborated and published [3].

As previously described, thyroid cancer is rare and usually indolent. For this reason, preoperative screening is essential to determine the most suitable clinical or surgical treatment.

Fine needle aspiration (FNA) biopsy is the most commonly used technique. It occurs under the supervision of an expert cytopathologist, and permits a high-accuracy diagnosis of nodules.; the use of this approach in the clinical practice can lead to a more conservative strategy, aimed at reducing surgical interventions [4].

Nowadays in the world there are different cytological classifications, very similar from several points of view, but the best known and more frequently used is the Bethesda System for reporting thyroid cytopathology [5].

According to the American Joint Committee, which develops the American Joint Committee on Cancer (AJCC) staging system describing the extent of cancer progression, last compiled and published thyroid cancer histological staging references in 2017 [6, 7]. The TNM (tumour, node, metastasis) classification considers 'T' the tumour size and its extension on the basis of the following scheme:

T1: tumour < or = 2 cm in greatest dimension limited to the thyroid
- T1a: tumour < or = 1 cm in greatest dimension limited to the thyroid
- T1b: tumour >1 but < or = 2 cm in greatest dimension limited to the thyroid

T2: tumour > 2 but < or = 4 cm in greatest dimension limited to the thyroid

T3: tumour >4 cm in greatest dimension limited to the thyroid or gross extrathyroidal extension, invading only strap muscles (sternohyoid, stenothyroid, or omohyoid muscles)
- T3a: tumour >4 cm in greatest dimension limited to the thyroid
- T3b: tumour of any size with gross extrathyroidal extension invading only strap muscles)

T4: tumour with gross extrathyroid extension
- T4a: gross extrathyroid extension invading subcutaneous soft tissues, larynx, trachea, oesophagus, recurrent laryngeal nerve
- T4b: gross extrathyroid extension invading prevertebral fascia or encasing the carotid artery or mediastinal vessels

With regard to 'N', lymph node staging:
- N0: no evidence of lymph node metatasis
- N1: evidence of loco-regional lymph node metastasis
- pN1a metastasis to VI or VII level lymph nodes
- pN1b metastasis to unilateral, bilateral, or contralateral lateral neck lymph nodes (I, II, III, IV or V levels), or retropharyngeal lymph nodes

With regard to 'M', distant metastasis staging:
- M0: no evidence of distant metastasis
- M0: no evidence of distance metastasis

Papillary Carcinoma

Papillary carcinoma is the most common malignant neoplasia of the thyroid gland, representing about 80% of thyroid cancer. It is more frequent in females in their III-IV decades of life and it is the most frequent thyroid cancer in the paediatric population. A familiar form of papillary carcinoma is frequent, but papillary carcinoma is rarely associated with familial syndromes such as familial adenomatous polyposis or Cowden syndrome [3].

Papillary carcinoma takes the name from its most frequent growth pattern: making papillae. The diagnosis of papillary carcinoma is however based on typical nuclear alterations. Papillary carcinoma generally has a good prognosis with long disease-free survival and high probability of healing. The mortality rate ranges from 5 to 17%. Negative prognostic factors are older age at time of diagnosis, large size, male sex, extension to extrathyroidal tissues (in particular to stripe muscles) and, finally, presence of distant

metastases. Some authors consider variants of papillary carcinoma (e.g. tall cells or hobnail variants) as negative prognostic factors. Papillary carcinoma can indeed show variable clinical behaviour: few variants of papillary carcinoma are very aggressive and may result in distant metastases and sometimes causing fatal outcomes. Papillary carcinoma usually metastasizes lymphatically., Vascular invasion can more rarely be observed, especially in follicular variants or tall cell variants of papillary carcinoma [8]. The most important etiological cause of papillary carcinoma is the exposure to ionizing radiation during childhood and adolescence. Dramatic cases were the nuclear accidents of Chernobyl in 1986 and the more recent one of Fukushima, in which the massive emissions of radioactive elements, in particular of radioiodine^{131} (I131), increased the incidence of papillary carcinoma by more than 100 times. Other evidence is the use of radiotherapy in the head and neck and in the mediastinum regions (in the past used for acne or tinea capitis or thymic hyperplasia in children and adolescents). The risk of cancer has a strong inverse correlation with age at exposure.

The molecular alterations of papillary carcinoma are numerous, of which one of the most important is the V600E mutation of the *BRAF* gene, also considered a diagnostic marker for papillary carcinoma.

Other mutations of the *BRAF* gene have been described, in particular the K601E mutation of *BRAF* has been reported in follicular variants of papillary carcinoma and sometimes in follicular adenomas.

Other important gene alterations are the rearrangements involving the *RET* gene (called *RET/PTC*) alongside a series of other genes, the most frequent being *RET/PTC1* and 3. Their frequency is extremely variable in the different study cases and could be due to a high heterogeneity of papillary carcinoma. Other important gene alterations of papillary carcinoma are the point mutations of *RAS* genes; these mutations are located at several specific sites, which are codons 12, 13, 61 of the N-, H-, K-RAS genes. Rearrangements of the *NTRK1* gene, also called TRK rearrangement, can be present in papillary carcinoma with a frequency of about 10%. It is important to underline that almost all these mutations are mutually exclusive, and are capable of activating the mitogen-activated protein kinase (MAPK) pathway regulating cell growth, differentiation, and survival [9].

Morphology

At gross examination, papillary carcinoma can show different aspects and sizes, and usually arises as a white-grey, irregular area with central scar formation; or as completely solid nodule; or as a solid and partially cystic nodule; or sometimes as a totally cystic lesion. The size of papillary carcinoma can change from only a few millimetres to numerous centimetres, forming voluminous masses with extension in the extrathyroidal tissues and sometimes in closer structures and organs. At microscopic examination papillary carcinoma can show different growth patterns: the most frequent growth pattern, the papillary pattern, shows exclusively papillary structures with a central fibro-vascular core lined by tumour cells (see Figure 3.5.5.1). The papillary carcinoma tumour cells are very peculiar. Accordingly with the WHO, the histological and cytological diagnosis of papillary carcinoma must be made only in presence of typical nuclear alterations.

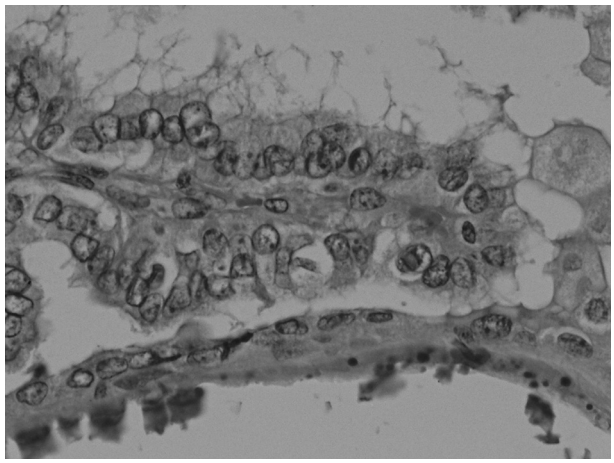


Figure 3.5.5.2 Nuclear features of papillary carcinoma: carcinoma cells are oval or elongated with irregular, large nuclei. The nuclei have a clear appearance with a marked nuclear membrane, show nuclear grooves, usually arranged in parallel to long axis of the nucleus, obtaining a typical coffee bean shape and they can show also nuclear pseudo-inclusions that are cytoplasmatic intranuclear inclusions.

The papillary carcinoma cells are oval or elongated with irregular, large nuclei. These have a clear appearance with a very marked and pronounced nuclear membrane on account of chromatin clearing and lateralization along the nuclear membrane. The nuclei of papillary carcinoma show nuclear grooves, generally arranged in parallel to the long axis of the nucleus, resulting in a typical coffee bean shape. The nuclei of papillary carcinoma can also show nuclear pseudoinclusions that are cytoplasmatic intranuclear inclusions similar to bubbles (see Figure 3.5.5.2). Papillary carcinoma can present psammoma bodies that are calcific, lamelled, and concentric structures. They are the result of cyclical depositions of calcium on tumoural death cells and they are particularly frequent in classical variants of papillary carcinoma [3, 8].

Other Variants of Papillary Carcinoma

Papillary carcinoma can present different histological variants. They are numerous and very different for their aspect and sometimes for their clinical behaviour. The histological variants are described on the basis of cell type, cell size and shape, architecture, and stroma. It is important to repeat that the diagnosis of papillary carcinoma must be made only in the presence of the typical nuclear alterations, independent of variants or growth patterns [3, 8].

Follicular Variant of Papillary Carcinoma

The follicular variant is the most common following the classical variant. As underlined by its name, it is characterized by the exclusive presence of follicular structures lined with cells with nuclear alterations of papillary carcinoma (see Figure 3.5.5.3). The follicular growth pattern can be micro, macrofollicular, or mixed [3, 8].

The follicular variant of papillary carcinoma can show an infiltrative attitude and lacks a fibrous capsule. Instead, its most frequent form presents a complete fibrous capsule, with no kind of infiltration or invasion. This form has an extremely low (quite absent) malignant potential.

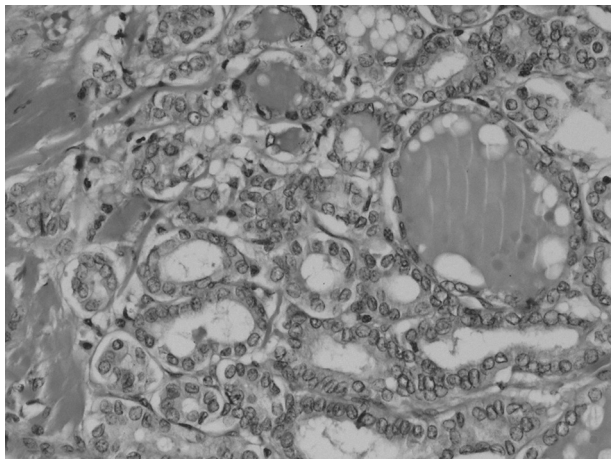


Figure 3.5.5.3 Follicular variant of papillary carcinoma: exclusive presence of follicular structures lined by cell with the typical nuclear alterations of papillary carcinoma.

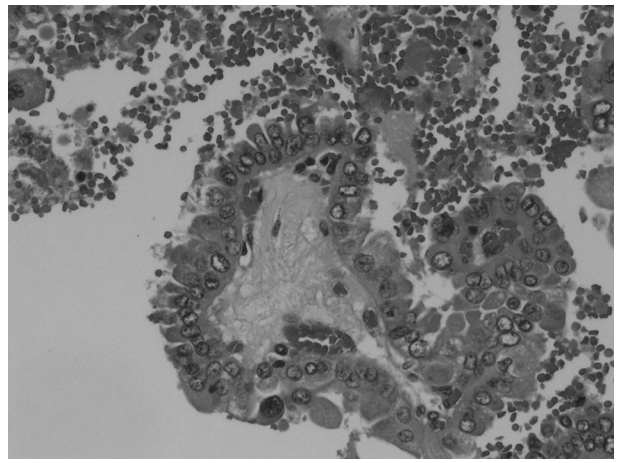


Figure 3.5.5.4 Hobnail variant of papillary carcinoma: papillary or micropapillary structures lined by cell with a wide eosinophilic cytoplasm and with apically located nuclei.

Recently, a task force of thyroid pathologists and molecular pathologists, together with clinicians, has declassified the completely encapsulated follicular variant of papillary carcinoma in Non-invasive follicular thyroid neoplasm with papillary-like nuclear features (named by its acronym NIFT-P) [3–10].

Tall Cell Variant of Papillary Carcinoma

The tall cell variant of papillary carcinoma is a typical variant of older age and male gender. It commonly shows papillary or solid growth patterns. The papillary or solid structures are lined with very particular elongated cells, the so-called tall cells, with a height two or three times the width and with the typical nuclear features of papillary carcinoma. The clinical behaviour of this variant is normally more aggressive than that of the classical variant, showing larger dimensions at the time of tumour diagnosis, with high likelihood of extrathyroidal infiltration and vascular invasion, and a high probability of recurrences and distant metastases. Disease-free survival is usually shorter than that of the classic variant and the mortality rate can reach 25% and more [3, 8].

Oncocytic Variant of Papillary Carcinoma

The oncocytic variant of papillary carcinoma is a rare variant. It is formed by large cells with wide eosinophilic cytoplasm and nuclear features of papillary carcinoma. The growth pattern is generally papillary, with a very low infiltrative behaviour. The clinical outcome is very similar to that of the classical variant [3, 8].

Hobnail Variant of Papillary Carcinoma

The hobnail variant of papillary carcinoma is a rare variant showing a papillary or micropapillary growth pattern, but rarely presents follicular or clustered structures. These structures are lined by cells with a wide eosinophilic cytoplasm and with apically located nuclei (see Figure 3.5.5.4). The nucleus cytoplasm ratio is decreased and the cells lose cellular cohesion. This variant has an aggressive behaviour and presents frequent necroses, mitoses, lymphatic invasion, extrathyroidal extension, distant metastases, and recurrences [3, 8].

Solid Variant of Papillary Carcinoma

The solid variant of papillary carcinoma is a rare variant, representing about 1% of adult papillary carcinoma. It is more common in children and adolescents or in patients exposed to ionizing radiations. The solid variant should be taken into consideration when all the other variants are not represented inside the tumour. It shows a solid, trabecular, or nested growth pattern. This variant has an aggressive behaviour showing distant metastases (in particular in the lung) and higher mortality rates than those of the classical variant [3, 8].

Other rarer variants are: **Warthin's-like variant,** morphologically very similar to **Warthin's-like tumour of salivary gland**; the columnar cell variant, that can be very aggressive and that is characterized by a papillary growth pattern with nuclear overlapping and stratification; the cribriform variant associated with familiar adenomatous polyposis; **papillary carcinoma** with nuclear fasciitis-like stroma in which papillary carcinoma is almost completely substituted by spindle cells arranged in irregular fascicles in a background of fibro-myxoid stroma (see Table 3.5.5.1 for complete list of variants of papillary carcinoma) [3, 8].

Table 3.5.5.1 List of papillary carcinoma variants

Variants of papillary carcinoma	
Papillary microcarcinoma	Papillary thyroid carcinoma with fibromatosis/fasciitis-like stroma
Encapsulated variant	Solid/trabecular variant
Follicular variant	Oncocytic variant
Diffuse sclerosing variant	Spindle cell variant
Tall cell variant	Clear cell variant
Columnar cell variant	Warthin-like variant
Cribriform-morular variant	Hobnail variant

Used with permission of the American College of Surgeons, Chicago, Illinois. The original source for this information is the *AJCC Cancer Staging Manual*, Eighth Edition (2017) published by Springer International Publishing. Tuttle RM, Morris LF, Haugen BR, Shah JP, Sosa JA, Subramaniam RM, *et al*. Thyroid (Differentiated and Anaplastic Carcinoma). In: Amin MB, Edge S, *et al*. (eds). *American Joint Committee on Cancer (AJCC) Cancer Staging Manual*, 8th ed. Chicago, IL: Springer; 2017:873–90.

Follicular Carcinoma

Follicular carcinoma represents 5–15% of thyroid cancer. It is more frequent in women and extremely rare in children and more common in areas with iodine deficiency [2, 3].

Most follicular carcinomas are sporadic. The most important factors implicated in the aetiology of follicular cancer are iodine deficiency and exposure to ionizing radiations. A proof is that the increment of iodine with diet substantially decrements the incidence of follicular carcinoma. This is probably due to the increment of TSH (thyroid stimulating hormone), stimulated by iodine deficiency and involved in thyroid cancer development.

Moreover, exposure to ionizing radiations results in the development of follicular cancer, although less than papillary carcinoma.

Another important risk factor is the pre-existence of a benign neoplasia, which may represent a definite risk factor in about 15% of follicular carcinomas. The presence of follicular adenoma could be the precursor of follicular cancer, possibly resulting from *RAS* gene family mutations. On the other hand, the development of follicular cancer in the multinodular goitre could be due to prolonged stimulation of TSH [9].

Follicular carcinoma is a well-differentiated cancer originating from follicular epithelium and characterized by an exclusive follicular growth pattern.

The cells of follicular carcinoma do not show any nuclear features of papillary carcinoma.

Clinically, the follicular carcinoma presents itself as a solitary cold nodule with very slow growth but with high metastatic power. It usually metastasizes by vascular invasion, while lymphatic invasion is extremely rare. Metastases to lung and bone are the most frequent and they sometimes represent the early onset of the disease [3, 8].

We can subdivide follicular carcinoma into a minimally invasive form and in a widely invasive form. The minimally invasive form is an encapsulated tumour, very similar to an adenoma, with histological evidence of tumour capsular and/or vascular infiltration (see Figure 3.5.5.5). The second form of follicular carcinoma is widely invasive, presenting wide infiltrative aspects of vascular, peritumoral, and sometimes perithyroidal tissues. These types differ for their clinical and biological behaviour. In the former type the 10-year survival rate is about 90%, while in the latter form it decreases to 35%.

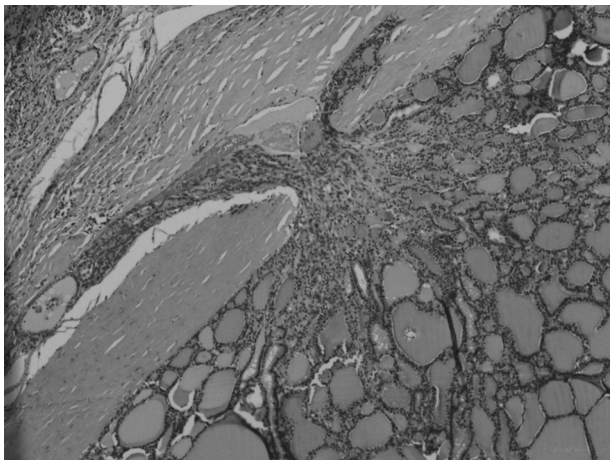


Figure 3.5.5.5 Minimally invasive follicular thyroid carcinoma: histological evidence of tumour capsular infiltration with the typical mushroom-like appearance.

The Hürthle (oncocytic or oxyphilic) tumour was considered the most important 'variant' of follicular carcinoma [3, 8].

As previously described, the most important genetic alterations involved in follicular cancer are the point mutations of the *RAS* gene family. These mutations affect codons 12, 13, and 61 of the H, K, and N-Ras genes. *RAS* gene mutations are present in more than 50% of follicular carcinomas, while the oncocytic or oxyphilic carcinomas present a lower percentage of *RAS* gene mutations. Another important genetic alteration is the PAX8/PPARgamma (peroxisome proliferator-activated receptor gamma) rearrangement, which involves about one-third of follicular carcinomas and about 5% of oncocytic or oxyphilic carcinomas [9].

Morphology

On gross examination follicular carcinoma, usually no larger than just a few centimetres in dimension, appears as an encapsulated, generally oval, solid nodule. A thick fibrous capsule generally surrounds the lesion and foci of invasion can rarely be detected in the tumour capsule; for this reason, extensive sampling of the tumour capsule is essential for a correct diagnosis of follicular carcinoma. In the widely invasive form of follicular carcinoma, thyroid tissue or sometimes extrathyroidal tissue invasion can be macroscopically evident. Vascular emboli are rarely observed, especially in the large vein of the neck.

On light microscopy follicular carcinoma presents a typical follicular architecture (microfollicular, macrofollicular, normofollicular, or mixed), sometimes associated with insular, or solid, or trabecular areas. In these cases, it is fundamental to distinguish the follicular carcinoma from the poorly differentiated thyroid carcinoma, which presents the same growth pattern but worse prognosis. Follicular carcinoma cells are cubic, small, and round, with an even chromatin pattern and small nuclei with small nucleoli. These nucleoli are more frequent and evident in oxyphilic follicular carcinoma, where they acquire a cherry colour. The cells of the follicular carcinoma obviously lack the nuclear features of papillary carcinoma [3–8].

The morphology of the cells and the presence of a thick fibrous capsule make it difficult to differentially diagnose between follicular adenoma and follicular carcinoma. The only fundamental diagnostic features are infiltration of the tumoural capsule and vascular invasion (Figure 3.5.5.5). To date, there is no real consensus about the invasion of the tumour capsule. Most authors require that neoplastic cells penetrate the entire thickness of the fibrous capsule for a diagnosis of follicular carcinoma. Capsule ruptures caused by surgical lacerations or changes due to precedent biopsy or FNAs are considered pseudo-invasions, insufficient for a diagnosis of follicular carcinoma. The typical appearance of infiltration of the tumour capsule is similar to that of a mushroom (Figure 3.5.5.5). Vascular invasion is referred to the invasion of tumour cells into the veins, and such cells must be adherent to the endothelium or take place in the context of a thrombus. Vascular invasion, whose extent has prognostic importance, should be localized in or beyond the capsule. Limited invasion (less than four vessels) has a good prognosis, while extensive invasion (more than four vessels) is associated with a worse outcome. Widely invasive follicular carcinomas show intrathyroidal or extrathyroidal tissue infiltration. They might also

Table 3.5.5.2 Classification of follicular thyroid carcinoma

Traditional	Armed Forces Institute of Pathology (AFIP) 2014		WHO 2017
Minimally invasive	Minimally invasive	With capsular invasion	Minimally invasive
		With limited vascular invasion (<4 vessels)	Encapsulated angioinvasive
		With extensive vascular invasion (> or = 4 vessels)	
Widely invasive	Widely invasive		Widely invasive

Reproduced with permission from LiVolsi V, Abdulkader Nallib I, Baloch ZW, Bartolazzi A, Chan JKC, DeLellis RA, *et al.* Follicular Thyroid Carcinoma. In: Lloyd RV, Osamura RY, *et al.* (eds). *WHO Classification of Tumors of Endocrine Organs.* Lyon: IARC Press. 2017: 66–143. Copyright © 2017 Virginia A. LiVolsi.

show wide vascular infiltration which, alone, is insufficient to be classified as widely invasive. Widely invasive follicular carcinomas are often large tumours with solid or insular or trabecular growth patterns, and they must be distinguished from poorly differentiated thyroid carcinomas or from solid or trabecular variants of papillary carcinomas [3–8] (see Table 3.5.5.2 for classification scheme modified from WHO 2017).

Hürthle (Oncocytic or Oxyphilic) Cell Carcinoma

The Hürthle (oncocytic or oxyphilic) cell carcinoma is a common tumour representing about 20% of follicular carcinomas. It is formed by Hürthle or oncocytic or oxyphilic cells, which are large cells with voluminous granular cytoplasm, and large centrally located nuclei with prominent nucleoli [2, 3, 8]. Most authors require that Hürthle cells should be most of the tumour (more than 75%) for a diagnosis of Hürthle cell carcinoma. Hürthle cell carcinomas present solid, trabecular, and follicular growth patterns with fibrous bands between clusters of cells. The tumours have little stroma and for this reason may show a pseudopapillary pattern. Hürthle cell carcinoma can metastasize through the bloodstream, invading in particular the liver, lungs, and bone. Hürthle cell carcinoma can also metastasize lymphatically and spread to cervical lymph nodes. Tumours with large dimension (more than 4 cm), necrosis, high mitotic rate, presence of atypical mitoses, and foci of small cells have been recently classified as poorly differentiated Hürthle cell carcinoma [3, 8, 9].

Medullary Carcinoma

At the end of the 50s medullary carcinoma was exhaustively described by Hazard and colleagues, who performed a detailed study of its morphology, presence of amyloid, frequency of lymph node metastasis, clinical behaviour and prognosis [7, 8]. The cells of origin, the so-called C cells, were described only some years later. Medullary carcinoma actually derives from C cells, which are usually present in normal thyroid glands localized in the interfollicular zone. Normal C cells do not stain by traditional histochemical methods such as haematoxylin-eosin staining, but they are very reactive to some types of immunohistochemical staining, the most common of which is currently calcitonin [3]. Medullary carcinoma, representing about 10% of malignant thyroid tumours, is particularly studied because of its clinical behaviour, high percentage of familiarity, and probability of being part of a syndrome. However, most medullary carcinomas are sporadic. Mean patient age is 50 years with a minimal prevalence for the female sex [3, 7, 8, 9]; up to 20% are familial cases with an autosomal dominant inheritance. These medullary cancers can appear as tumours that are not associated with the syndrome (familial non-MEN medullary carcinomas), or as a part of multiple endocrine syndromes (MEN), in particular MEN type 2A, MEN type 2B, MEN type 3. In familial cancer (familial non-MEN medullary carcinomas) the onset age is usually less than 50 years, while medullary cancer in MEN can already be present in the first years of life. The most frequent genetic alterations of medullary carcinoma are the RET proto-oncogene mutations present in more than 90% of the MEN2 family. Families with MEN 2A or FMTC present exon-10 and exon-11 mutations of the *RET* gene in 95 and 85% of cases, respectively. Families with MEN 2B present the mutation of codon 918 (in exon 16) in about 95% of cases. Genetic screening is recommended in first-degree relatives of patients with these types of alterations. If positive, they must be subject to prophylactic thyroidectomy. The risk of very early development of carcinomas and metastases makes prophylactic thyroidectomy fundamental in the very first years of life. Finally, also the sporadic form of medullary carcinoma presents RET gene mutations, the most frequent being found in codon 918 [9, 11, 12].

On gross examination, medullary cancer can be single or multifocal with dimensions that can range from few millimetres to some centimetres. The lesion usually presents a white-grey aspect and a clear edge.

On light microscopy, medullary carcinoma can show different growth patterns and may sometimes imitate other thyroid tumours. The cells are in most cases polygonal and present small-size nuclei, sometimes with spindle aspects. Nuclear chromatin is often finely granular and the nucleoli are absent. The cells rarely mimic the small lung cancer cell or can appear spindle as thymic or mesenchymal lesions. The tumour cells of medullary cancer are usually arranged in trabeculae and/or nests, which are separated by dense collagenous or hyalinized material containing amyloid that can be stained with Congo Red. Medullary carcinoma can be definitively diagnosed by immunostaining neoplastic cells show positive staining for chromogranin A and synaptophysin, CEA (carcinoembryonic antigen), NSE (neuro-specific enolase) and, in particular, calcitonin, which represents the most sensitive marker. The number of calcitonin-positive cells can change according to the different cases: of these, about 1.5% show negativity for calcitonin and they are considered medullary cancers only if they present also C-cell hyperplasia or familial genetic alterations. Medullary cancer also shows positive staining for cytokeratin and negative staining for follicular cell markers like thyroglobulin [3, 8, 9].

Sporadic medullary carcinoma appears as a single, painless nodule. More than 50% of cases present cervical or mediastinic

lymph node metastases at diagnosis, and more than 15% show distant metastases to bone, lung, liver, and adrenal gland.

Familial cancers are frequently multifocal and bilateral. Prognosis is associated with the rapidity of diagnosis and prophylaxis, although medullary cancer in FMTC or MEN2A shows a better prognosis than cancer in MEN2B [3, 8].

Poorly Differentiated Thyroid Carcinoma

Poorly differentiated thyroid carcinoma is still a controversial entity. Some authors have proposed to include, in poorly differentiated carcinoma category, tumours with clinical behaviour, morphology, and biological features intermediate between well-differentiated and undifferentiated carcinomas. A large number of carcinomas has been added over the years, making the poorly differentiated carcinoma category extremely heterogeneous. The tumours nowadays considered to be poorly differentiated include some variants of papillary carcinoma such as tall cell variant, hob nail variant, Hürthle cell carcinoma, and insular carcinoma [3, 8, 9].

The poorly differentiated carcinoma represents 5% of thyroid malignancies. The mean age of the patients is about 10 years higher than that of patients with differentiated carcinomas, but younger than that of patients with anaplastic cancer. Poorly differentiated carcinoma shows a prevalence for the female sex (male/female ratio:1/2) [1, 2].

The aetiology can develop in three different directions: poorly differentiated cancer can progress from papillary cancer, or from follicular cancer, or 'de novo'. From a molecular point of view, the most frequent alterations of poorly differentiated cancers are the point mutations. They include very early events that are important for tumourigenesis and that predispose for further mutations (like Ras mutations or BRAF mutations) and also for gene alterations driving de-differentiation (like TP-53 and β-catenin mutations) [9].

On gross examination, poorly differentiated cancer displays a large size (range 1–10 cm) with an aggressive and infiltrative behaviour. This neoplasia very often shows an extrathyroidal extension and very rarely shows a partial tumour capsule. Poorly differentiated carcinoma is usually hard to cut, its aspect is firm and yellowish-white-, and it may present haemorrhagic or necrotic areas.

On light microscopy the diagnosis of poorly differentiated carcinoma displays de-differentiated morphological features. On the basis of Torino's criteria, they result as following [13, 14]:

1. Solid/trabecular/insular growth pattern
2. Lack of nuclear features typical of papillary cancer
3. Presence of one of the following features: necrosis; three or more mitoses per high power fields; convoluted nuclei

Poorly differentiated carcinoma commonly shows a bad prognosis with 5-year survival lower than 50%. It generally presents compression symptoms associated with cough, dyspnoea, dysphagia. Infiltration of the recurrent nerve is frequent and an early event and distant metastases are present at the time of diagnosis in more than 20% of patients. Lymph node metastases are also frequent. Poorly differentiated carcinoma shows a partial response to routine radiometabolic therapy [3, 8].

Anaplastic (or Undifferentiated) Carcinoma

Anaplastic or undifferentiated carcinoma is a malignant neoplasm behaving in a very aggressive fashion, and is constituted by undifferentiated cells. It has an epithelial origin but, in the most cases, it lacks immunoreactivity for follicular markers such as thyroglobulin or TTF-1. It typically presents itself as a large mass (medium 6 cm) with a rapid growth (it may duplicate its volume in less than one week) and with a fatal outcome of a few months [3].

Anaplastic carcinoma is a rare neoplasm, representing about 5% of all malignant thyroid tumours, and it is a typical neoplasm of adult patients, the majority of whom are older than 60 years, whereas patients younger than 40 years are very rarely affected by the tumour. It is a little more common in females than in men (ratio F/M:1.5:1). The etiological factors of anaplastic cancer include the pre-existence of a malignant or of a benign thyroid neoplasia. The aetiology is described by the clinical histories of patients followed-up for goitre or for differentiated thyroid cancer and who have developed anaplastic carcinoma. Another proof of the possibility of transformation from differentiated to undifferentiated or anaplastic carcinoma is sometimes the presence of residual areas of differentiated carcinoma (i.e. papillary carcinoma areas) in the context of anaplastic cancer. Another certain etiological factor is the lack of iodine. The introduction of iodine in a diet (in areas with chronic deficiency) has demonstrated a high decrease of anaplastic cancer. The genetic alterations include somatic point mutations. As already explained for poorly differentiated carcinoma, they include very early event, important for tumorigenesis and predisposition for further mutations (such as Ras mutations or *BRAF* mutations), and also gene alterations that drive the de-differentiation (like TP-53 and β-catenin mutations) [9].

On gross examination, anaplastic carcinoma appears as a large mass infiltrating all the thyroid gland and often the adjacent structures. Size varies from 1 cm to 20 cm (medium about 6 cm). The lesions are friable with multiple necrotic or haemorrhagic foci. In the paucicellular variant of anaplastic carcinoma the lesions may be hard to cut and present a whitish aspect. On light microscopy, anaplastic carcinoma may show heterogeneous aspects, which can coexist in the same tumour. The cells display marked anaplasia and can show spindle, epithelial, giant, rhabdoid features. The nuclei are extremely pleomorphic with irregular edge and granular chromatin. The mitoses, especially atypical mitoses, are very frequent and coagulative necrosis is a recurring feature (see **Figure 3.5.5.6**). Regardless of its morphology, anaplastic carcinoma infiltrates the thyroid gland and its nearby structures as well as the veins and arteries.

The cells of anaplastic carcinoma may show immunohistochemical positivity for epithelial markers (i.e. cytokeratin), but usually lack expression for follicular markers such as thyroglobulin or TTF-1.

Anaplastic cancer is rare but, if present, it rapidly leads to fatal neoplasia. It presents itself as a large infiltrating mass that extends to the neck structures such as the oesophagus, trachea, or big vessels. Distant metastases are present in more than 45% of patients, at diagnosis time, and they are localized in the bone, lung, or brain. Anaplastic carcinoma does not respond to radiometabolic therapy [8].

Non-Epithelial Tumours of the Thyroid Gland

Mesenchymal Tumours

Primary mesenchymal tumours of the thyroid are extremely rare. We can cite smooth muscle tumours such as leiomyomas, leiomyosarcomas, and solitary fibrous tumours. Leiomyoma is a benign neoplasia, typical of women, which shows histological and immunophenotypical features of the smooth muscles and which can be cured, generally by complete surgical removal or by lobectomy. Leiomyosarcoma is a typical neoplasia of older patients, with no gender predilection. On gross examination, it usually appears as a large mass with infiltrative behaviour. On light microscopy pleomorphism, mitosis, atypical mitosis, coagulative necrosis can be found. Primary leiomyosarcoma of the thyroid is a very rare malignant neoplasia and its diagnosis can be made also after the exclusion of secondarism from other sites [3].

Primitive solitary fibrous tumour of the thyroid is a very rare mesenchymal neoplasia, for which only few cases have been reported. Solitary fibrous tumour shows prevalent fibroblastic differentiation. Some tumours can present focal or extended myofibroblastic differentiation that is generally benign [3].

Vascular Tumours

Vascular primitive thyroid tumours are rare and include haemangioma, lymphangioma, and angiosarcoma. In particular, the angiosarcoma is a rare and aggressive lesion representing less than 4% of thyroid lesions, with a geographical predilection of Alpine regions of Italy and of Europe. Its behaviour is similar to anaplastic carcinoma with rapid fatal outcome [3, 8].

Lymphomas

Primitive thyroid lymphomas are rare neoplasms, accounting for less than 5% of all malignant thyroid tumours. They include a heterogeneous group of lesions and originate from B lymphocytes. The most frequent histotypes are large B-cell lymphomas as well as the extranodal marginal zone B-cell lymphomas of mucosa-associated lymphoid tissue (MALT), with its variants. The aetiology of primary thyroid lymphomas is strictly linked to chronic lymphocytic thyroiditis. Clinically, primitive thyroid lymphomas occur in females with mean age 60–65 years, but they can potentially arise at any age (14–90 years). They can present different sizes, from very small to large masses with compression symptoms including pain, dysphagia, dyspnoea, coughing. Differential diagnosis must exclude secondarisms from diffuse lymphomas [8].

REFERENCES

1. Stewart BW, Kleihues P. *World Cancer Report*. Lyon: IARC Press 2003.
2. Ries LAG, Melbert D, Krapcho M, *et al. SEER Cancer Statistic Review 1975–2005*. Bethesda, MD: National Cancer Institute, 2007.
3. Lloyd RV, Osamura RY, Kloppel G, Rosai J. Tumours of the thyroid gland. In: *WHO Classification of Tumors of Endocrine Organs*. Lyon: IARC Press, 2017: 66–143.
4. Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, 2016; **26**: 1–133.
5. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid*, 2017; **27**: 1341–6.
6. Tuttle RM, Morris LF, Haugen BR, *et al.* Thyroid (differentiated and anaplastic carcinoma). In: Amin MB, Edge S, *et al.*, eds. *American Joint Committee on Cancer Staging Manual*, 8th edition. Chicago, IL: AJCC, Springer, 2017: 873–90.
7. Rosen JE, Lioyd RV, Brierley JD, *et al.* Thyroid (medullary carcinoma). In: Amin MB, Edge S, *et al.*, eds. *American Joint Committee on Cancer staging Manual*, 8th edition. Chicago, IL: AJCC, Springer, 2017: 891–901.
8. Baloch ZW, LiVolsi VA. Pathology of thyroid gland. In: LiVolsi VA, Asa SL, eds. *Endocrine Pathology*. Philadelphia, PA: Churchill Livingstone, 2002: 61–101.
9. Nikiforov YE. *Diagnostic Pathology and Molecular Genetics of the Thyroid.* Philadelphia, PA: Wolters Kluwer, 2010.
10. Nikiforov YE, Seethala RR, Tallini G, *et al.* Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol*, 2016; **1**: 1023–9.
11. Romei C, Tacito A, Molinaro E, *et al.* Twenty years of lesson learning: how does the RET genetic screening test impact the clinical management of medullary thyroid cancer? *Clin Endocrinol*, 2015; **82**: 892–9.
12. Romei C, Ciampi R, Casella F, *et al.* RET mutation heterogeneity in primary advanced medullary thyroid cancers and their metastases. *Oncotarget*, 2018; **4**: 9875–84.
13. Volante M, Collini P, Nikiforov YE, *et al.* Poorly differentiated thyroid carcinoma: the Turin proposal for the use of uniform diagnostic criteria and an algorithmic diagnostic approach. *Am J Surg Pathol*, 2007; **31**: 1256–64.
14. Volante M, Bussolati G, Papotti M. The story of poorly differentiated thyroid carcinoma: from Langhans' description to the Turin proposal via Juan Rosai. *Semin Diagn Pathol*, 2016; **33**: 277–83.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

3.5.6 Papillary, Follicular, and Anaplastic Thyroid Carcinoma and Thyroid Lymphoma

Ruxandra Dobrescu and Corin Badiu

Introduction

Thyroid cancer is the most frequent endocrine neoplasia. Most of these cancers (>90%) are derived from the follicular epithelium, that is, differentiated thyroid cancers (DTC)—papillary (PTC), and follicular thyroid cancer (FTC). Thyroid cancer incidence has increased significantly in the past 40 years [1, 2], probably mostly due to increased detection of infracentimetric PTC. On the other hand, this rise in incidence has not been paralleled by an increase in mortality [1], which, coupled with the high prevalence rates (up to 35%) of incidentally detected PTCs at autopsy studies, suggests that these newly diagnosed small PTCs might be clinically insignificant. Therefore, care should be taken to avoid overtreatment and treatment-associated morbidity in these patients. At the other end of the spectrum of disease are patients with invasive tumours, aggressive histology, and distant metastases, for whom early diagnosis and correct multimodal treatment is essential. Dynamic risk assessment during follow-up allows a better characterization of tumour behaviour and guides further management decisions.

Diagnosis

Thyroid cancer typically presents as a palpable cervical lump ± lymph nodes, particularly suspicious if consistency is hard, if it occurs in a child or as a single nodule in a man, if personal history reveals head and neck radiation exposure or if family history is positive for familial DTC syndromes (polyposis, Carney complex, Cowden, or Werner syndromes).

It is often discovered incidentally in an asymptomatic patient undergoing ultrasound, CT, or MRI of the neck and can be an isolated finding or part of multinodular goitre. A non-palpable nodule has the same risk of malignancy as a palpable nodule of the same size. Occasionally patients present with locally advanced, invasive disease, involving the adjacent neck structures resulting in hoarseness, dysphagia, dyspnoea, and signs of neurological compression (brachial plexus and recurrent laryngeal nerve compression caused by the primary tumour and spinal cord compression due to bone metastases).

Thyroid function tests are usually normal in these patients, although concurrent autoimmune thyroiditis, particularly frequent in women, can present with hypothyroidism. Hyperthyroidism can rarely occur in patients with FTC with distant metastases and it is often T_3 thyrotoxicosis due to increased activity of type 1 and 2 deiodinases in tumour tissue. Malignant *struma ovarii* is also occasionally associated with T_3 thyrotoxicosis.

Thyroid ultrasound evaluation is essential, as it provides the morphological characterization of the nodule and can identify other significant non-palpable thyroid nodules or pathological lymph nodes. The ultrasound pattern of a thyroid nodule has been widely used to predict the risk of malignancy and to select the nodules for which FNAB might be indicated. The ultrasound features with highest specificity for thyroid cancer are hypoechogenicity, the presence of microcalcifications, 'taller than wide' shape, irregular margins, and presence of suspicious lymph nodes [3].

Fine needle aspiration biopsy is the gold standard for the diagnosis of thyroid cancer. For cytologically indeterminate lesions (follicular neoplasm, follicular lesion of undetermined significance (FLUS) or atypia of undetermined significance (AUS)), with a risk of malignancy between 10% and 40% [4], molecular studies on fine needle aspirates using mutational analysis and mRNA expression profiles have shown promise in better risk-stratifying the patients to avoid un-necessary thyroidectomies [5, 6].

Staging

Thyroid cancer management is based on the concept of risk assessment and stratification. Treatment decisions are taken individually after balancing the risk of recurrence and disease-specific mortality with the risk of treatment-associated morbidity. The aggressiveness of the treatment and follow-up is thus tailored to each patient's tumour characteristics [7].

The preferred initial staging system is the tumour node metastasis (TNM) system [8] (Table 3.5.6.1). It uses age at diagnosis and tumour extension to stratify patients according to their estimated risk of dying of cancer. Age is the most important prognostic factor; mortality increases progressively after around 35 years of age. Large tumours, gross extension beyond the thyroid capsule and clinically apparent lymph nodes are negative prognostic factors.

Most patients with DTC will be stage I and II at diagnosis, with excellent prognosis in terms of survival, but although mortality is generally very low, recurrences can be very frequent. The American Thyroid Association (ATA) risk stratification system [7], updated in 2015, uses clinical and pathology data to stratify patients into low, intermediate and high-risk groups, in order to guide the best initial treatment decision based on estimated risk of recurrence (Figure 3.5.6.1). Precise histological characterization is important

Table 3.5.6.1 Differentiated thyroid carcinoma TNM staging, AJCC UICC 2017

Age	Stage group	TNM	10 yrs estimated survival
<55	I	Any T, Any N, M0	98–100%
<55	II	Any T, Any N, M1	85–95%
≥55	I	T1, N0/x, M0 T2, N0/x, M0	98–100%
≥55	II	T1, N1, M0 T2, N1, M0 T3, Any N, M0	85–95%
≥55	III	T4a, Any N, M0	60–70%
≥55	IV A	T4b, Any N, M0	<50%
≥55	IV B	Any T, Any N, M1	

TNM, tumour, node, metastasis; AJCC, American Joint Committee on Cancer; UICC, Union of International Cancer Control.

Primary tumour: T1: tumour ≤2 cm in greatest dimension limited to the thyroid; T2: tumour >2 but ≤4 cm limited to the thyroid; T3: tumour >4 cm limited to the thyroid or any size tumour with gross extrathyroidal extension invading only the strap muscles; T4: Gross extrathyroidal extension: T4a: if invading subcutaneous soft tissues, larynx, trachea, oesophagus, or recurrent laryngeal nerve; T4b: if invading prevertebral fascia or encasing the carotid artery or mediastinal vessels.

Regional lymph nodes: N0: no evidence of locoregional lymph node metastasis; N1: Metastasis to regional nodes.

Distant metastases: M0: no metastases, M1: distant metastases.

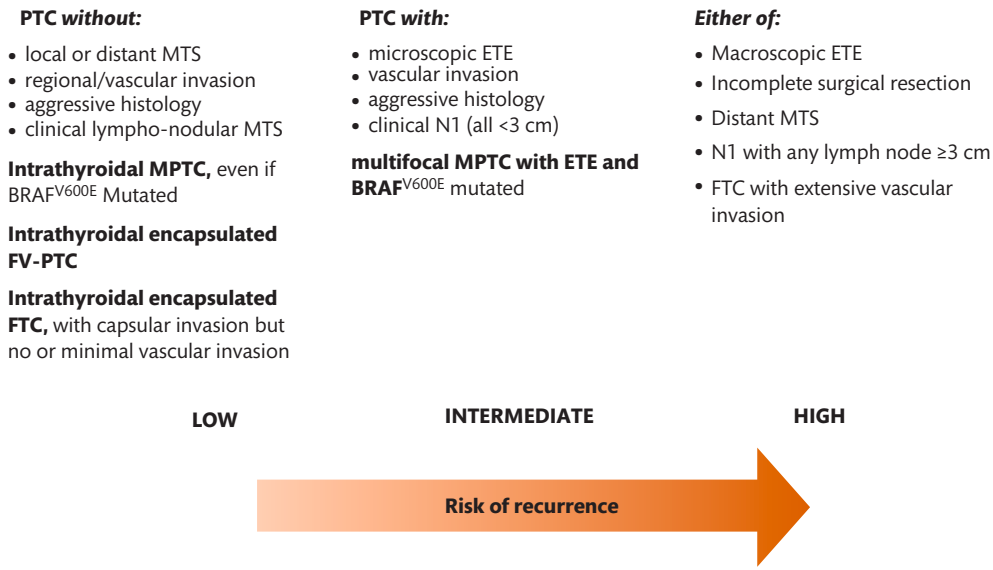


Figure 3.5.6.1 Risk of recurrence stratification system according to ATA 2015. ATA, American Thyroid Association; PTC, papillary thyroid carcinoma; FV-PTC, follicular variant papillary thyroid carcinoma; MPTC, micropapillary thyroid carcinoma; FTC, follicular thyroid carcinoma; MTS, metastases; ETE, extrathyroidal extension.

as certain histological subtypes (tall cell, hobnail variant, columnar cell carcinoma) tend to be more aggressive and can respond poorly to standard treatment. Nodal micrometastases can be present in up to 90% of patients with PTC but evidence suggests that adverse prognosis is limited to those patients with macroscopic, clinically evident node involvement [9]. The extent of extrathyroid extension (ETE) and the presence of vascular invasion are also important. The presence of $BRAF^{V600E}$ mutation is generally associated with aggressive histology, ETE, and lymph nodes metastases, and appears to convey a higher risk of recurrence [10], particularly if associated with another oncogenic mutation, such as the TERT promoter mutation [11].

During follow-up, new clinical, biological, and imaging data that reflect each patient's individual response to treatment are used to restratify that patient according to his/her current risk of recurrence, in a dynamic risk stratification system [7].

Treatment

Surgical resection is the essential first step in the management of thyroid cancer. Due to the ability of DTC to concentrate iodine and respond to thyroid-stimulating hormone (TSH) stimulation, adjuvant treatment is based on a combination of **radioiodine therapy** and **TSH suppression**, tailored on each patient's risk of recurrence. Treatment decisions should ideally be taken in a multidisciplinary tumour board (MDT) [12].

The optimal management is constantly re-evaluated in light of accumulating new evidence related to prognostic factors, risk of recurrence, and optimal treatment modalities, with guidelines published by several societies [7, 12–14]. Despite numerous biological differences, papillary and follicular thyroid carcinomas are generally treated in the same way, and have similar prognosis.

Surgery

Complete surgical tumour resection is essential, irrespective of stage and the possible presence of distant metastases, as this increases disease-related survival, decreases the risk of recurrence, and allows staging and decision-making for further management.

Presurgical imaging is important for proper planning of the procedure. Neck ultrasound will provide morphological details related to the tumour, extension, possible multifocality, and can identify suspicious lymph nodes. High resolution imaging with CT or MRI is needed in extensive local disease to assess the degree of infiltration of the trachea, oesophagus, and vascular structures and identify an intrathoracic extension for which thoracotomy might be necessary.

Extent of Surgery

Lobectomy is recommended for infracentimetric thyroid nodules with a positive FNA diagnosis in the absence of extrathyroidal extension and lymph node involvement, as the risk of recurrence is extremely low. Lobectomy can also be performed in selected patients with tumour size between 1 and 4 cm without any suspicious features, as the outcomes appear similar compared to total thyroidectomy, with fewer side effects. These patients do not require radioiodine ablation and can be followed-up with ultrasound and thyroglobulin measurements [7, 14].

Total thyroidectomy should be performed for multifocal tumours and in patients with positive family history of thyroid cancer or personal history of neck irradiation, as their risk of recurrence is higher. It is also preferred if ultrasound reveals contralateral lobe changes (e.g. thyroiditis) or non-specific adenopathy that can complicate follow-up. Total thyroidectomy is required in patients with large tumours bigger than 4 cm, with ETE, or in the presence of metastases [7, 12, 14].

Clinically evident lymph node metastases (on palpation, ultrasound or intraoperative examination) are present in 20–50% of patients with DTC and they convey a higher risk of recurrence [9] so **therapeutic node dissection** of clinically apparent node disease is recommended [7, 12, 14]. Up to 90% of patients have micrometastases evident at histology but their risk of recurrence is negligible [9] so prophylactic neck dissection is not recommended.

The most frequent complications after surgery are hypoparathyroidism and recurrent laryngeal nerve injury. The risk is higher with total thyroidectomy, node dissection, and with repeat interventions. Transient hypoparathyroidism is seen in up to 20% of thyroid cancer patients after surgery and becomes permanent in around 1%. Laryngeal nerve monitoring might help in redo surgery/recurrences, to decrease the recurrent damage.

Adjuvant Therapy Based on Risk Stratification

Radioiodine Treatment

Differentiated thyroid cancer arises from the follicular epithelium and similar to normal thyroid cells has the ability to take up iodine from the blood through the membrane sodium-iodide symporter. ^{131}I is taken up by normal thyroid and cancer cells and causes acute cell damage by emission of beta particles. Gamma radiation is also emitted by ^{131}I and can provide a map of thyroid cancer spread through the body after scanning with a gamma camera.

After total thyroidectomy, radioiodine therapy can be used in DTC patients for:

1. Remnant ablation—to destroy normal thyroid remnants, in order to minimize the risk of cancer recurrence in high-risk patients. After ablation, thyroglobulin can be used as a specific marker of disease persistence/recurrence and whole-body ^{131}I scanning can identify foci of disease activity.
2. Adjuvant therapy in case of subclinical, microscopic residual disease, or clinically significant residual disease that is not surgically accessible.

The necessity of radioiodine treatment is assessed individually based on the estimated risk of recurrence with the goal of administering the minimal amount of radioactivity needed in the best tolerated method [7, 14].

Low-risk patients: ^{131}I is not recommended for small tumours less than 1 cm without other risk factors, even with coexisting microscopic lymph node metastases, because no benefit was proven in terms of survival or risk of recurrence [15, 16]. ^{131}I is not required for tumours of 1–4 cm in diameter in the absence of other risk factors.

Intermediate risk patients: ^{131}I should be considered in selected patients with microscopic ETE, clinically documented lymph node metastases, vascular invasion, or in the presence of aggressive histologies, but the evidence so far concerning its benefit is conflicting [15].

High-risk patients: ^{131}I therapy should be administered to all high-risk patients as this improves survival and decreases the risk of recurrence [7, 12, 14].

Optimal uptake of ^{131}I by thyroid cancer cells is dependent on adequately increased TSH levels and a low-iodine diet in the weeks preceding radioiodine administration. Medications containing iodine and exposure to iodine-based contrast agents should be avoided, as a high load of iodine can take weeks to be eliminated (more than 4 weeks for CT with contrast/angiography; months/years after stopping amiodarone treatment). Twenty-four hour urinary iodine concentrations (UIC) can be helpful in monitoring iodine excretion.

The radioiodine uptake is increased by high TSH, achievable either by thyroid hormone withdrawal or by administrating recombinant human TSH.

Thyroid hormone withdrawal (THW) is essential in high-risk patients to treat gross residual disease for optimal ^{131}I uptake (7). Serum TSH should be above 30 mUI/L, which is usually achieved after 3–4 weeks of THW. To minimize symptoms of hypothyroidism, shorter-acting T_3 can be given (25 ug bds or tds) for 2 weeks after LT4 is withdrawn, but should be used with caution in older patients and in patients with ischaemic cardiac disease. T3 is stopped 2 weeks before radioiodine treatment.

Recombinant human TSH (rhTSH) has the advantage of avoiding prolonged hypothyroidism, thus improving quality of life, and reducing radiation exposure, as radioiodine is cleared more rapidly than in a severely hypothyroid patient. It can be used for remnant ablation and as adjuvant therapy of potential residual microscopic disease in low and intermediate risk patients [7]. It can also be used in high-risk patients in whom THW and subsequent hypothyroidism can seriously aggravate concurrent illnesses (i.e. congestive heart failure, depression, etc). rhTSH is essential in patients with central hypothyroidism of any cause in whom TSH stimulation is not possible with THW. rhTSH is administered as 0.9 mg intramuscular daily injections on two consecutive days and ^{131}I is administered on the third day, then after 2–7 days a whole-body scan (WBS) is performed.

^{131}I Dose

For remnant ablation in low-risk patients a low 30 mCi dose after rhTSH is as successful as a higher 100 mCi dose administered after THW, although long-term outcome data are lacking [17]. In intermediate risk patients with suspected microscopic residual disease, lower doses (30 mCi) appear to have similar ablation rates and long-term clinical outcomes to higher doses (100–150 mCi) [18, 19]; higher doses (75–100 mCi) might be appropriate for patients with higher risk features, such as vascular invasion. In high-risk patients with distant metastases higher doses (100–200 mCi) should be used after THW, as there is insufficient data concerning long-term outcomes with rhTSH administration [7]. Certain histological variants (Hürthle cell FTC, tall cell PTC or poorly DTC) have poor radioiodine sensitivity and they also require higher doses. Dosimetric studies might allow a more precise estimation of administered ^{131}I activity than empiric fixed doses.

Pretreatment radioiodine diagnostic scanning is rarely indicated. It can be useful when the extent of residual disease cannot be established with certainty after surgery using ultrasound and thyroglobulin measurement or if the results could potentially alter the need to administer ^{131}I or the required dose [7, 12, 14]. Diagnostic use of ^{131}I is hypothesized to cause 'stunning', i.e. lower uptake of the treatment dose of ^{131}I administered afterwards, due to sublethal radiation cell injury caused by the diagnostic dose. To minimize this

risk, diagnostic scanning should use low dose ^{131}I (1–3 mCi) or the lower energy isotope ^{123}I (1.5–3 mCi). The WBS should be obtained after 48–72 h for ^{131}I and 6–24 h for ^{123}I.

Side Effects and Complications

Acute and chronic side effects occur in up to 30% of patients after radioiodine treatment, especially if higher radiation doses are administered [20].

Most patients present with salivary and lacrimal gland dysfunction. Xerostomia, ageusia, and sialadenitis are frequent and lemon juice and candies can be helpful. Radiation thyroiditis is occasionally seen in the presence of large thyroid remnants. Similar to acute sialadenitis, it usually responds to non-steroidal anti-inflammatory drugs but steroids can sometimes be required.

In the acute setting, particularly after thyroxine withdrawal, high TSH values acts as a potent stimulus for cancer cell growth and together with the accompanying oedema can cause important tumour swelling, leading to airway obstruction, nerve compression, or severe bone pain, depending on the location. Steroids can be helpful (dexamethasone 8 mg bds or prednisone 60 mg/day), started 1 or 2 days before radioiodine administration then tapered during the following days.

Gonadal dysfunction is sometimes seen, ranging from transient oligospermia or ovarian dysfunction to infertility after high cumulative doses.

Patients are at increased risk of secondary malignancies, particularly leukaemia and solid tumours in sites that concentrate ^{131}I: breast tissue, urinary, and digestive tract [20]. This risk is higher in younger patients and with high cumulative doses (>500–600 mCi). Patients are encouraged to drink large volumes of fluids to avoid accumulation of the isotope in the bladder, and laxatives are prescribed to ensure regular bowel movements.

^{131}I administration is contra-indicated in pregnancy because it crosses the placenta and destroys the fetal thyroid leading to cretinism. Women are advised to avoid pregnancies for at least 6 months after radioiodine treatment. ^{131}I administration is forbidden during lactation as it crosses to the child through the milk; in addition, accumulation in the breast tissue leads to increased radiation exposure of the mother.

For radiation protection, most patients receive radioiodine therapy as inpatients in dedicated nuclear medicine units, with limited access. After treatment, patients must respect strict personal hygiene, avoid public gatherings and close contact, particularly with young children and pregnant women for a few days following ^{131}I administration, in order to reduce radiation exposure to the community.

Thyroid Hormone Suppression

After surgery, thyroxine treatment is tailored to counteract TSH stimulation of potential tumour remnants. The degree of TSH suppression in individual patients is decided based on the estimated risk of recurrence [7, 14]. A robust TSH suppression improves progression-free survival in high-risk patients but a moderate suppression of TSH is probably enough for intermediate risk patients and no benefit was proven for low-risk patients [21].

Low-risk patients with partial thyroidectomy or total thyroidectomy and non-detectable non-stimulated thyroglobulin should have a low-normal TSH (0.5–2 mUI/L). Certain patients have enough thyroid tissue left after a lobectomy to maintain normal thyroid function without thyroxine replacement.

Intermediate risk patients should have a TSH between 0.1 and 0.5 mUI/L. The same moderate TSH suppression is indicated for low-risk patients with low level detectable non-stimulated thyroglobulin.

High-risk patients require a TSH maintained below 0.1 mUI/L [7, 14].

Thyroxine dose can be reduced if required due to advanced age, the presence of atrial fibrillation, ischaemic heart disease, or severe osteoporosis.

Follow-up and Long-Term Management

The principal objectives of follow-up are to ensure adequate thyroid hormone replacement with the recommended level of TSH suppression and to monitor for possible disease recurrence. This active surveillance is mostly based on serial measurements of **thyroglobulin** and **neck ultrasound** [7, 14]. In patients with evidence of persistent disease further imaging will be required to define disease extension and to plan management. This will imply a combination of radioiodine WBS, CT/MRI, bone scan, ^{18}FDG-PET/CT.

Thyroglobulin

Synthesized exclusively in thyroid follicular cells, thyroglobulin (Tg) is a good marker for DTC providing information about persistent and recurrent disease.

Serum Tg is generally measured by immunometric assays based on antigen-antibody interactions. Up to 25% of patients with thyroid cancer have high antithyroglobulin antibody (TgAb) titres which can disrupt the thyroglobulin assay, making follow-up difficult. Additionally, there is significant interassay Tg variability, so in any given patient the same Tg assay should be used for serial measurements and TgAb should always be measured at the same time [7, 14]. High Tg values above the upper range of the assay should be re-measured after dilution until an accurate result is obtained, so precise data are documented allowing correct serial evaluation during follow-up.

Thyroglobulin reaches its nadir at 3–4 weeks after surgery; positive serum levels reflect presence of small thyroid remnants and/or disease remnants. There is no clear cut-off to discriminate between these scenarios, but levels above 30 ng/ml after lobectomy and above 5 ng/ml after total thyroidectomy should raise suspicion of residual/metastatic disease [7].

Radioiodine ablation destroys normal thyroid remnants thus increasing the specificity of thyroglobulin in detecting recurrences. TSH stimulation increases the sensitivity of thyroglobulin measurement. An undetectable TSH-stimulated thyroglobulin can safely identify disease-free patients. Newer super-sensitive thyroglobulin assays, with functional sensitivity of less than 0.05 ng/ml, might obviate the need for TSH stimulation, as non-stimulated thyroglobulin values correlate well with TSH-stimulated thyroglobulin measured by these assays [22].

For low-risk patients who have not received ^{131}I, the variation trend in serial Tg measurements can have an excellent negative (declining values) or positive (increasing values) predictive value for disease recurrence during long-term follow-up [15].

In a patient with positive TgAb, Tg measurement is unreliable due to assay interference, but rising TgAb levels suggest disease recurrence.

Ultrasound

The vast majority of DTC recurrences are localized in the neck and ultrasound is highly sensitive in detecting lymph node involvement, irrespective of ^{131}I ablation status and Tg levels [15]. Suspicious lymph nodes are defined by the presence of microcalcifications, cystic changes, peripheral vascularity, round shape and loss of hilar image and hyperechogenicity [23]. If repeat surgery is considered, FNAB should be previously performed with cytology and measurement of thyroglobulin in the aspirate fluid to confirm recurrent lymph node disease [24].

Radioiodine Scanning

The combination of neck ultrasound and serum thyroglobulin measurement is equal to or superior in detecting residual disease compared to radioiodine WBS, which should not be used routinely [15]. During follow-up, repeat WBS is indicated in patients with post-treatment scans showing areas of increased ^{131}I uptake suggestive for pathological lymph nodes/distant metastases. It is also indicated in patients with normal ultrasound but high TgAb in whom Tg measurement is unreliable to evaluate disease status [7]. SPECT/CT has the advantage of superimposing functional and anatomical images thus providing a more accurate localization of disease recurrence.

^{18}FDG-PET/CT

In patients with negative WBS despite high thyroglobulin levels (stimulated thyroglobulin >10 ng/ml), ^{18}FDG-PET can localize the residual disease [7]. The sensitivity ranges between 45% and 100% and is dependent on tumour burden, histological subtype, and degree of differentiation [25]. It is more sensitive in patients with more aggressive histological variants, such as poorly differentiated, tall cell, insular, Hürthle cell carcinomas, and identifies patients with poorer prognosis. The sensitivity is generally low in patients with stimulated thyroglobulin <10 ng/ml as this indicates low burden disease (with the exception of a poorly DTC where the low serum thyroglobulin is a sign of dedifferentiation). Integrating ^{18}FDG-PET and CT imaging allows colocalization of hypermetabolic and morphological data; CT has higher spatial resolution and can increase the specificity of PET scan by providing structural information about areas of physiological uptake or inflammatory lesions. PET is particularly useful in identifying mediastinal lymph nodes and in differentiating cervical residual disease from post-surgical fibrosis [25]. PET scan performed after rhTSH administration has slightly higher sensitivity than basal PET but this changes the planned management only in a minority of patients [26].

CT and MRI

High resolution axial CT of the neck and chest with iv contrast allows better definition of large nodal recurrences in the neck and mediastinum. It is also needed to describe extensive, invasive disease involving the trachea, oesophagus and adjacent blood vessels [7]. If dyspnoea or dysphagia are present, fibre optic laryngoscopy and endoscopy might be necessary to evaluate for possible intraluminal tumour invasion. CT of the chest is the most sensitive method to diagnose small lung metastases. CT and MRI of the brain or abdomen should be performed to evaluate for distant disease in patients with high thyroglobulin levels (Tg >10 ng/ml) or high and rising antithyroglobulin antibodies, if symptoms of organ-specific involvement are present or if radioiodine treatment is planned and tumour swelling after ^{131}I is a concern [7].

Follow-up Plan

During follow-up, data reflecting the patient's response to treatment is collected and the risk of recurrence is reassessed according to a dynamic risk stratification system [7] (Table 3.5.6.2).

Excellent response: no clinical, biochemical, structural signs of disease.

Table 3.5.6.2 Response to therapy definitions based on initial therapy (total thyroidectomy with/without radioiodine ablation), predicted outcomes, and recommended management plan

Category	Excellent response	Biochemical incomplete response	Structural incomplete response	Indeterminate response
Definition	• Negative imaging *and* • nsTg <0.2 ng/ml *or* • sTg <1 ng/ml after RAI sTg< 2 ng/ml without RAI • Negative TgAb	• Negative imaging *and* • nsTg ≥1 ng/ml after RAI nsTg ≥5 ng/ml without RAI *or* • sTg ≥10 ng/ml *or* • Rising TgAb	Structural or functional evidence of disease, regardless of Tg, TgAb levels	• Non-specific imaging findings • Faint uptake in thyroid bed on RAI scans • Detectable Tg (below cut-offs for biochemical incomplete response) *or* • TgAb stable/declining without structural/functional disease
Management plan	• nsTg every 1–2 yrs • sTg not necessary • Regular neck ultrasound • CT/MRI follow-up may be indicated in high-risk patients for 3–5 years after attaining excellent response	• nsTg every 6 months • If stable/declining Tg → observation; ultrasound yearly; sTg at 2 years to evaluate for excellent response • If rising Tg/TgAb → consider WBS, CT/MRI, ^{18}FDG-PET, and treatment of persistent disease	• nsTg every 6 months • Yearly ultrasound • Diagnostic WBS to evaluate iodine avidity of lesions • MRI, CT at 6–12 months depending on progression • 18FDG-PET—additional disease sites → Additional therapy of persistent disease	• nsTg 6–12 months • sTg after 2 yrs to evaluate for excellent response • Neck ultrasound every 6–12 months • If suspicious findings appear → consider WBS, CT/MRI, ^{18}FDG-PET, and treatment of persistent disease
TSH goal	0.5–2 mUI/L (0.1–0.5 mUI/L for 3–5 yrs in high-risk patients)	0.1–0.5 mUI/L	<0.1 mUI/L	0.1–0.5 mUI/L

nsTg, non-stimulated thyroglobulin; RAI, radioiodine ablation; sTg, stimulated thyroglobulin; TgAb, antithyroglobulin antibodies. Based on [7], [27], [28].

Box 3.5.6.1 Predicted outcomes in patients treated with total thyroidectomy and radioiodine ablation based on response to therapy

Excellent response:
<4% recurrence at 5–10 years (5–15% in high-risk patients)

Biochemical incomplete response:
30% spontaneously evolve to no evidence of disease (NED)
20% achieve NED after additional therapy
20% structural disease at end of follow-up

Structurally incomplete response:
50–85% persistent disease despite therapy
Deaths: 11% in regional disease
50% in distant disease

Indeterminate response:
15–20% with structural disease identified at follow-up
The rest—the changes are stable or resolve

Biochemical incomplete response: high thyroglobulin, increasing antithyroglobulin antibodies, but negative imaging.

Structurally incomplete response: persistent/recurrent structural disease, regardless of thyroglobulin or antibody levels.

Indeterminate response: unspecific biochemical/structural changes.

This system allows, for example, for a patient initially included in a high-risk group after surgery but with a good response after ^{131}I therapy, with stimulated thyroglobulin <1 ng/ml, negative antibodies, and imaging, to move to a lower risk group due to the excellent response to treatment, which would imply a decrease in the intensity of follow-up and a lesser degree of TSH suppression.

The definitions were originally developed and validated based on the outcome of patients with DTC treated with total thyroidectomy and radioiodine ablation (Box 3.5.6.1) [7], but the system was subsequently validated also for low and intermediate risk patients treated with lobectomy and total thyroidectomy without RIA [27, 28]. After lobectomy, a stable non-stimulated Tg of less than 30 ng/ml in the context of undetectable TgAb and negative imaging is considered an excellent response [28]. An increasing trend in non-stimulated thyroglobulin or TgAb in these patients represent an increased risk of structural disease and should prompt additional investigation [27].

Persistent/Recurrent Disease

In most patients with persistent disease the lesions are identified in the neck: lymph node recurrences or residual disease in the thyroid bed. Distant metastases can be present at diagnosis in up to 10% of patients or can develop over time. Persistent disease is suggested by high thyroglobulin or increasing TgAb titres and is certified by ultrasound, radioiodine scanning, CT/MRI, bone scan, or ^{18}FDG-PET which identify the site of disease.

Surgery is preferred if the location of the identified lesion allows it, followed by radioiodine if the lesion is ^{131}I avid [7, 14].

Clinically significant, biopsy proven, **cervical lymph node recurrences** should be resected. In high-risk surgical patients, percutaneous ethanol injection, laser, high intensity focused ultrasound or radiofrequency ablation can be considered as alternatives [7, 23, 29].

In **extensive, invasive disease**, surgical resection can be followed by external beam radiotherapy to lower the risk of recurrence [7].

Pulmonary metastases respond well to radioiodine therapy, particularly in younger patients. Complete remission is possible, particularly in case of micrometastases. In persistent disease therapy should be repeated every 6–12 months as long as it continues to concentrate ^{131}I, but high cumulative radioiodine doses can lead to pulmonary fibrosis.

Bone metastases usually require higher doses and ^{131}I therapy is usually not curative. Surgical resection and stabilization are sometimes needed to prevent fracture and neurological complications; bisphosphonate treatment with pamidronate or zoledronic acid can provide pain relief and even a small 2–10% decrease of bone tumour volume; external beam radiation therapy can help control bone pain.

Brain metastases usually occur in older patients and have a poorer prognosis. Surgical resection, if feasible, should be attempted before ^{131}I; otherwise external beam radiation therapy is an option. If the lesions are radioiodine-avid, ^{131}I therapy should be performed accompanied by glucocorticoid treatment to prevent tumour and neighbouring brain structures swelling [7].

Radioiodine Resistant Differentiated Thyroid Carcinoma

DTC usually caries a good prognosis due, in great measure, to its good response to radioiodine therapy. There are, however, patients with radioiodine refractory structurally evident disease in whom the lesions lose the ability to concentrate ^{131}I or progress despite continuing to exhibit ^{131}I uptake, probably due to dedifferentiation [26]. ^{18}FDG-PET/CT can identify distant metastases in these patients and should be used in conjunction with radioiodine scans as patients can present with both radioiodine-avid and resistant lesions. ^{18}FDG-PET uptake generally signals lesions with aggressive histologies and poor prognosis. For these patients, new therapeutic options are required. Tyrosine kinase inhibitors (sorafenib, lenvatinib, vandetanib) partially inhibit multiple kinases and affect multiple signalling pathways, including angiogenesis [30]. They are recommended in patients with radioiodine refractory, progressive, metastatic disease and have been shown to increase progression-free survival and induce sustained tumour regression [31–33]. The risk-benefit must be carefully weighted in each patient, as treatment is frequently accompanied by numerous side effects: hypertension, nausea, fatigue, weight loss, skin changes, risk of gastrointestinal perforation, hepatotoxicity, bleeding, etc. [7]. New research is focused on finding ways to increase radioiodine sensitivity. Two small studies suggested that treatment with selumetinib (an inhibitor of MAPK pathway) and dabrafenib (a selective *BRAF* inhibitor) can restore radioiodine sensitivity in patients with ^{131}I resistant DTC [34, 35]. Cytotoxic chemotherapy with doxorubicin or taxanes has limited efficacy in DTC patients but can be attempted in patients who cannot tolerate or progress despite tyrosine kinase therapy.

Anaplastic Thyroid Carcinoma

Anaplastic thyroid cancer (ATC) is an undifferentiated tumour of the follicular epithelium. It is rare but extremely aggressive, with a median survival of 5–7 months after diagnosis and a 20% 1-year survival rate [36]. It tends to develop in older individuals (mean age 65 years) and can be associated with DTC. Genetically, anaplastic cancer presents with activating *BRAF* and *RAS* mutations (similar to DTC), but also new mutations involving p53 tumour suppressor protein, TERT, or PIK3CA. It is thus believed that anaplastic cancer develops from more differentiated tumours after a dedifferentiation event, which is probably the acquisition of a new mutation.

ATC generally presents as a rapidly growing, fixed cervical tumour mass. It can extend into the perithyroidal fat to the strap muscles, trachea, oesophagus, blood vessels, and laryngeal recurrent nerves which leads to pain, dysphonia, dysphagia, dyspnoea, and cough. Distant metastases are present in up to 50% of patients at diagnosis (often in the lungs, bone, brain, skin, adrenal glands) and can manifest as systemic symptoms: fatigue, anorexia, weight loss or focal signs of compression: bone pain, neurological symptoms.

The initial evaluation should include a full blood count, biochemistry, thyroid function tests and neck ultrasound [37]. Marked leukocytosis is sometimes present due to tumour secretion of lymphokines. Some patients present with hypocalcaemia due to invasion of the parathyroid glands or humoral hypercalcaemia of malignancy. Thyroid function is usually normal, but hypothyroidism can occur in extensive disease destroying the normal parenchyma. Ultrasound reveals a hypoechoic, infiltrative mass, often with areas of extensive necrosis and lymph node metastasis. Diagnosis requires FNAB or core needle biopsy and immunohistochemistry is needed to distinguish ATC from poorly differentiated and medullary thyroid carcinomas, thyroid lymphoma, melanoma, and sarcoma [37].

Cross-sectional CT or MRI studies should be performed for staging purposes and to define local extension and the potential for surgical resection. ^{18}FDG-PET/CT is helpful to identify distant metastases. According to the TNM staging system, all ATCs are considered stage IV: intrathyroidal tumours are stage IVA, gross extrathyroidal extension means stage IVB and distant metastases—IVC [8].

Due to their aggressive nature, prompt management of ATC is essential. Surgery should be performed for all intrathyroidal tumours as complete resection increases survival. In invasive tumours surgical resection should be followed by radiotherapy and chemotherapy. Combination therapy decreases the risk or recurrence but doesn't influence survival [36]. In patients with distant metastases no therapy was proven effective and median survival is 4 months. External beam radiation therapy (EBRT) can provide pain relief in patients with bone metastases. Cytotoxic chemotherapy with doxorubicin, doxorubicin + cisplatin, or paclitaxel usually provides only short-term responses. If available, patients should be enrolled in clinical trials with new therapies based on the molecular profile of the tumour [37]. Radioiodine therapy and TSH suppression have no role in ATC, because the undifferentiated cells are not capable of ^{131}I uptake, but can be useful in patients with associated DTC. Palliative treatment should focus on securing the airway and maintaining access for nutritional support (tracheostomy and gastrostomy), as death usually occurs due to aerodigestive obstruction. Adequate management of pain should be provided as well as emotional and spiritual support.

Thyroid Lymphoma

Primary thyroid lymphoma (PTL) is a rare malignancy, representing less than 2% of all thyroid cancers. It typically presents in the seventh life decade, affecting predominantly women (female:male ratio—4:1). The major known risk factor for PTL is Hashimoto's thyroiditis, which was identified in up to 90% of patients [38].

Thyroid lymphomas are almost always of the non-Hodgkin type and of B-cell lineage; the most frequent are diffuse large B-cell lymphomas (DLBCL) and mucosa-associated lymphoid tissue (MALT) lymphomas [38, 39].

PTL typically presents as a rapidly expanding neck mass which is often hard, sometimes tender, fixed to the trachea, and often extending substernally. Because of invasion into the adjoining neck structures, patients can present with hoarseness, dyspnoea, and dysphagia. Lymph node enlargement can be identified in approximately 50% of cases. Systemic B-type symptoms can also be present: fever, night sweats, and significant weight loss. Occasionally, MALT lymphomas present as a single, asymptomatic, slow-growing thyroid nodule.

Patients are usually euthyroid; hypothyroidism sometimes occurs because of coexisting Hashimoto's thyroiditis/severe thyroid infiltration by the lymphoma. Rarely, transient thyrotoxicosis occurs due to rapid destruction of thyroid follicles.

Ultrasound reveals an intensely hypoechoic, pseudo-cystic structure, sometimes nodular, with 'broccoli-like borders', other times diffuse, infiltrative. Enhancement of posterior echoes is always present and can help to distinguish it from severe thyroiditis [40]. FNAB can suggest the diagnosis, but management is decided after large-bore or surgical biopsy with immunohistochemistry which is needed to evaluate the histological subtype.

^{18}FDG-PET, CT, or MRI are performed for staging purposes, to define local extension, to plan treatment, and monitor response to treatment.

Management depends on histological subtype and tumour stage. MALT lymphomas usually have an indolent clinical course and can be treated by surgery alone, radiotherapy alone or a combination of them [39]. The benefit of surgical resection in these patients is debatable, so it is not typically recommended, but MALT lymphomas are sometimes detected incidentally after thyroidectomy performed for other indications, and in this case no further treatment may be needed. DLBCL are generally aggressive and require combined modality treatment using the monoclonal antibody rituximab, combined chemotherapy (cyclophosphamide, doxorubicin, vincristine, prednisone) and radiotherapy [39]. Radiotherapy provides local control and combined chemotherapy is used to decrease the incidence of distant recurrences. If critical airway obstruction occurs, surgical resection may be an option; however, tumours typically respond in a few hours to chemotherapy and glucocorticoids and airway permeability is restored.

The prognosis of PTL depends largely on the histological subtype, patient age, disease stage, and treatment. The reported 5-year disease-free survival is 95% for MALT lymphomas and 75% for DLBCL [39].

REFERENCES

1. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg*, 2014; **140**: 317–22.
2. Wiltshire JJ, Drake TM, Uttley L, Balasubramanian SP. Systematic review of trends in the incidence rates of thyroid cancer. *Thyroid*, 2016; **26**: 1541–52.
3. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J*, 2017; **6**: 225–37.
4. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid*, 2017; **27**: 1341–6.
5. Nikiforova MN, Mercurio S, Wald AI, *et al*. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. *Cancer*, 2018; **124**: 1682–90.
6. Labourier E, Shifrin A, Busseniers AE, *et al*. Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J Clin Endocrinol Metab*, 2015; **100**: 2743–50.
7. Haugen BR, Alexander EK, Bible KC, *et al*. 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2016; **26**: 1–133.
8. Tuttle ML, Haugen B, Shah J, *et al*. Thyroid-differentiated and anaplastic carcinoma. In: Amin MB ES, Greene F, Byrd D, *et al*., eds. *AJCC Cancer Staging Manual*, 8th edn. New York, NY: Springer International Publishing, 2017: Ch. 73.
9. Randolph GW, Duh QY, Heller KS, *et al*. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. *Thyroid*, 2012; **22**: 1144–52.
10. Elisei R, Viola D, Torregrossa L, *et al*. The BRAF(V600E) mutation is an independent, poor prognostic factor for the outcome of patients with low-risk intrathyroid papillary thyroid carcinoma: single-institution results from a large cohort study. *J Clin Endocrinol Metab*, 2012; **97**: 4390–8.
11. Moon S, Song YS, Kim YA, *et al*. Effects of coexistent BRAF(V600E) and TERT promoter mutations on poor clinical outcomes in papillary thyroid cancer: a meta-analysis. *Thyroid*, 2017; **27**: 651–60.
12. Perros P, Boelaert K, Colley S, *et al*. Guidelines for the management of thyroid cancer. *Clin Endocrinol (Oxf)*, 2014; **81** Suppl 1: 1–122.
13. Pacini F, Schlumberger M, Dralle H, Elisei R, Smit JW, Wiersinga W. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur J Endocrinol*, 2006; **154**: 787–803.
14. National Comprehensive Cancer Network. Thyroid Carcinoma—NCCN Clinical Practice Guidelines in Oncology 2018 [updated 22 May 2018].
15. Lamartina L, Durante C, Filetti S, Cooper DS. Low-risk differentiated thyroid cancer and radioiodine remnant ablation: a systematic review of the literature. *J Clin Endocrinol Metab*, 2015; **100**: 1748–61.
16. Nixon IJ, Ganly I, Patel SG, *et al*. The results of selective use of radioactive iodine on survival and on recurrence in the management of papillary thyroid cancer, based on Memorial Sloan-Kettering Cancer Center risk group stratification. *Thyroid*, 2013; **23**: 683–94.
17. Schlumberger M, Catargi B, Borget I, *et al*. Strategies of radioiodine ablation in patients with low-risk thyroid cancer. *N Engl J Med*, 2012; **366**: 1663–73.
18. Han JM, Kim WG, Kim TY, *et al*. Effects of low-dose and high-dose postoperative radioiodine therapy on the clinical outcome in patients with small differentiated thyroid cancer having microscopic extrathyroidal extension. *Thyroid*, 2014; **24**: 820–5.
19. Castagna MG, Cevenini G, Theodoropoulou A, *et al*. Post-surgical thyroid ablation with low or high radioiodine activities results in similar outcomes in intermediate risk differentiated thyroid cancer patients. *Eur J Endocrinol*, 2013; **169**: 23–9.
20. Clement SC, Peeters RP, Ronckers CM, *et al*. Intermediate and long-term adverse effects of radioiodine therapy for differentiated thyroid carcinoma--a systematic review. *Cancer Treat Rev*, 2015; **41**: 925–34.
21. Jonklaas J, Sarlis NJ, Litofsky D, *et al*. Outcomes of patients with differentiated thyroid carcinoma following initial therapy. *Thyroid*, 2006; **16**: 1229–42.
22. Smallridge RC, Meek SE, Morgan MA, *et al*. Monitoring thyroglobulin in a sensitive immunoassay has comparable sensitivity to recombinant human TSH-stimulated thyroglobulin in follow-up of thyroid cancer patients. *J Clin Endocrinol Metab*, 2007; **92**: 82–7.
23. Leenhardt L, Erdogan MF, Hegedus L, *et al*. 2013 European thyroid association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer. *Eur Thyroid J*, 2013; **2**: 147–59.
24. Grani G, Fumarola A. Thyroglobulin in lymph node fine-needle aspiration washout: a systematic review and meta-analysis of diagnostic accuracy. *J Clin Endocrinol Metab*, 2014; **99**: 1970–82.
25. Leboulleux S, Schroeder PR, Schlumberger M, Ladenson PW. The role of PET in follow-up of patients treated for differentiated epithelial thyroid cancers. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 112–21.
26. Schlumberger M, Brose M, Elisei R, *et al*. Definition and management of radioactive iodine-refractory differentiated thyroid cancer. *Lancet Diabetes Endocrinol*, 2014; **2**: 356–8.
27. Momesso DP, Vaisman F, Yang SP, *et al*. Dynamic risk stratification in patients with differentiated thyroid cancer treated without radioactive iodine. *J Clin Endocrinol Metab*, 2016; **101**: 2692–700.
28. Momesso DP, Tuttle RM. Update on differentiated thyroid cancer staging. *Endocrinol Metab Clin N Am*, 2014; **43**: 401–21.
29. Jeong SY, Baek JH, Choi YJ, Lee JH. Ethanol and thermal ablation for malignant thyroid tumours. *Int J Hyperthermia*, 2017; **33**: 938–45.
30. Dobrescu R, Badiu C. An expanding class in the treatment of thyroid cancer: tyrosine kinase inhibitors. *Acta Endocrinologica-Bucharest*, 2015; **11**: 536–9.
31. Brose MS, Nutting CM, Jarzab B, *et al*. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet*, 2014; **384**: 319–28.
32. Leboulleux S, Bastholt L, Krause T, *et al*. Vandetanib in locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 2 trial. *Lancet Oncol*, 2012; **13**: 897–905.

33. Schlumberger M, Tahara M, Wirth LJ, *et al*. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *N Engl J Med*, 2015; **372**: 621–30.
34. Ho AL, Grewal RK, Leboeuf R, *et al*. Selumetinib-enhanced radioiodine uptake in advanced thyroid cancer. *N Engl J Med*, 2013; **368**: 623–32.
35. Rothenberg SM, McFadden DG, Palmer EL, Daniels GH, Wirth LJ. Redifferentiation of iodine-refractory BRAF V600E-mutant metastatic papillary thyroid cancer with dabrafenib. *Clin Cancer Res*, 2015; **21**: 1028–35.
36. Smallridge RC, Copland JA. Anaplastic thyroid carcinoma: pathogenesis and emerging therapies. *Clin Oncol (R Coll Radiol)*, 2010; **22**: 486–97.
37. Smallridge RC, Ain KB, Asa SL, *et al*. American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid*, 2012; **22**: 1104–39.
38. Watanabe N, Noh JY, Narimatsu H, *et al*. Clinicopathological features of 171 cases of primary thyroid lymphoma: a long-term study involving 24553 patients with Hashimoto's disease. *Br J Haematol*, 2011; **153**: 236–43.
39. Walsh S, Lowery AJ, Evoy D, McDermott EW, Prichard RS. Thyroid lymphoma: recent advances in diagnosis and optimal management strategies. *Oncologist*, 2013; **18**: 994–1003.
40. Ota H, Ito Y, Matsuzuka F, *et al*. Usefulness of ultrasonography for diagnosis of malignant lymphoma of the thyroid. *Thyroid*, 2006; **16**: 983–7.

3.5.7 Medullary Thyroid Carcinoma

Friedhelm Raue and Karin Frank-Raue

Classification and Epidemiology

Medullary thyroid carcinoma (MTC) is a rare tumour of the parafollicular or C cells of the thyroid, that produce calcitonin (Ctn) and not iodine-rich thyroid hormone, making it not susceptible to treatment with radioactive iodine. New genetic studies in mice indicating that the genuine progenitors to C cells arise in the endoderm germ layer and are not of a neural crest origin [1]; but MTC often have the clinical and histological features of neuroendocrine tumours. They account for 3–5% of all thyroid carcinomas and occur in both sporadic and hereditary forms [2]. The majority of patients have sporadic MTC (70%), while 30% have hereditary MTC. The sex ratio in sporadic MTC is 1:1.3 (male to female), while both sexes are nearly equally affected in the familial variety [3]. The highest incidence of sporadic disease occurs in the fifth decade of life, while hereditary disease can be diagnosed earlier, depending on the possibility of genetic and biochemical screening.

Table 3.5.7.1 Classification of medullary thyroid carcinoma

Variety of MTC	Incidence (%)	Age at onset	Associated endocrinopathies
Sporadic MTC	70	Fifth decade	None
Hereditary MTC	30		
- MEN 2A	~27	Third decade	Phaeochromocytoma, parathyroid adenoma/ hyperplasia, Hirschsprung disease, cutaneous lichen amyloidosis, only MTC (familial MTC)
- MEN 2B	~3	First decade	Phaeochromocytoma, mucosal neuromas, marfanoid habitus

MEN 2A, multiple endocrine neoplasia type 2A; MTC, medullary thyroid carcinoma.

The familial variety of MTC is inherited as an autosomal dominant trait with a high degree of penetrance and MTC is the dominant component of the multiple endocrine neoplasia type 2 syndrome (MEN2) [4]. It is caused by germline-activating mutations of the RET proto-oncogene. Two distinct hereditary varieties of MTC are known, MEN2A and MEN2B, they differ with respect to incidence, genetics, age of onset, association with other diseases, aggressiveness of MTC, and prognosis (Table 3.5.7.1.):

1. The MEN 2A syndrome (OMIM 171400), characterized by the association of MTC, pheochromocytoma, and tumours of the parathyroids, sometimes Hirschsprung's disease, and cutaneous lichen amyloids. In some cases MTC is isolated without other endocrinopathies (familial MTC, FMTC). The subtype FMTC has been diagnosed more frequently in recent years and is reported to account for 35–40% of all MEN 2 cases [5].
2. The MEN 2B syndrome (OMIM 162300), consisting of MTC, phaeochromocytoma, ganglioneuromatosis of the gastrointestinal tract, and marfanoid habitus; in this phenotype hyperparathyroidism is absent. It is the most aggressive form (5–10% of all cases) [6].

Many patients with MEN 2B have an earlier onset in the first year of life and more aggressive MTC with a higher morbidity and mortality than in patients with MEN 2A. They often do not have a family history of the disease. Their tumours and characteristic appearance are therefore due to de novo germline RET mutations that present as sporadic cases of potentially hereditary disease. In contrast, the clinical course of MTC in FMTC is more benign than MEN 2B with a late onset or no clinically manifest disease, and the prognosis is relatively good. Therefore, a family history is often inadequate in establishing familial disease and more thorough evaluation by genetic and biochemical screening often reveals a family history of MTC in a patient originally thought to have the sporadic form of the disease [7].

Detection of MTC in patients has changed in recent years with the introduction of specific strategies: Ctn screening in patients with thyroid nodules and screening with molecular methods for

RET proto-oncogene mutations in patients with apparently sporadic MTC and in family members at risk for MTC. By earlier identification of patients with MTC, the presentation has changed from clinical tumours to preclinical disease, resulting in a high cure rate of affected patients with much better prognosis [8].

Pathology and Biochemical Markers

The histological appearance of MTC is enormously variable with regard to cytoarchitecture (solid, trabecular, or insular) and cell shape (spindle, polyhedral, angular, or round). The presence of stromal amyloid is characteristic in about 50–80% of MTC patients. This feature had been an auxiliary diagnostic criterion for MTC before the use of calcitonin immunocytochemistry. MTC without stromal desmoplasia is associated with a very low potential for metastasis [9].

Hereditary MTC characteristically presents as a multifocal process with neoplastic C-cell hyperplasia in areas distinct from the primary tumour. Bilateral neoplastic C-cell hyperplasia is a precursor lesion to hereditary MTC, in contrast non-MEN2-associated C-cell hyperplasia which is found with a great variety of pathological conditions, including non-medullary thyroid tumours, nodular goitre, autoimmune thyroiditis, renal failure, and disorders of calcium metabolism, and is not related to the development of sporadic MTC [10]. Metastasis may be found first in central and lateral cervical and mediastinal lymph nodes of the neck in 10% of patients with a micro-MTC operated on after discovery by familial screening, and in up to 90% of patients operated on for clinical MTC. Metastases outside the neck and mediastinum may occur during the course of the disease in the liver, lung, and bone.

The primary secretory product of MTC is Ctn, which serves as a highly sensitive tumour marker. Measurement of monomeric Ctn with two-site assays remains the definitive test for prospective diagnosis of MTC [11]. The test is widely available, accurate, reproducible, and cost-effective. Reference intervals are gender-dependent with higher levels in men than in women. Basal Ctn concentrations usually correlate with tumour mass and are almost always high in patients with palpable tumours. Similarly, elevated plasma Ctn levels following surgery to remove the tumour are indicative of persistent or recurrent disease. Measurement of plasma Ctn has been part of the routine evaluation of patients with thyroid nodules; up to 3% of patients with thyroid nodules have pathological Ctn concentrations and about 0.6% have an MTC [12]. The prevalence of MTC was nearly 100% when basal Ctn levels were more than 100 pg/ml. For basal Ctn, the best cut-off points able to separate non-MTC (including normal and C-cell hyperplasia cases) from MTC patients were 26 pg/ml in females and 68 pg/ml in males [13]. The predictive value for detection of micro-MTC in the grey zone between upper reference value and cut-off value is low; small MTC could be diagnosed sufficiently early enough allowing curative thyroidectomy by increasing Ctn values during follow-up. All patients with basal Ctn below 100 pg/ml could be cured by surgery [14]. Ctn assays with improved functional sensitivity and gender specific reference values may avoid stimulation testing with pentagastrin or calcium. In rare cases MTC does not secrete Ctn, the prevalence of non-secretory MTC with no measurable Ctn is 0.83% [15].

After careful evaluation, Ctn measurement in nodular thyroid disease allows early diagnosis and early surgery of MTC, reducing the significant mortality associated with this malignant tumour. There are a number of other substances, including carcinoembryonic antigen (CEA), PDN-21 (katacalcin), procalcitonin, chromogranin A, neuron-specific enolase, somatostatin, and adrenocorticotrophic hormone (ACTH), that are produced by MTC and which may help to differentiate it from other tumours.

Genetic Abnormalities

The responsible gene for MEN 2 (OMIM 171400, 162300, 155240) was localized to centromeric chromosome 10 (10q11.2) by genetic linkage analysis in 1987. Activating germline point mutations of the RET proto-oncogene were identified in 1993 [16]. Analysis of RET in families with MEN 2 revealed that only affected family members had germline missense mutations in eight closely located exons (5, 8, 10, 11, and 13–16) [16, 17]. Genetic testing detects nearly 100% of mutation carriers and is considered the standard of care for all first-degree relatives of patients with newly diagnosed MTC [18].

The *RET* gene has 21 exons and encodes a receptor tyrosine kinase that appears to transduce growth and differentiation signals in several developing tissues including those derived from the neural crest. It is expressed in cells such as C cells, the precursors of MTC, and in adrenal medullary cells, the precursors of pheochromocytomas, and parasympathetic and sympathetic and colonic ganglia cells, precursors of the ganglioneuromas. Unregulated *RET* activation leads to various cancers, including hereditary MTC, while humans lacking *RET* expression due to inactivating germline mutations develop Hirschsprung's disease.

The *RET* protein consists of an extracellular segment with a ligand-binding domain, a cadherin (Ca^{2+}-dependent cell adhesion)-like domain, and a cysteine-rich domain that is positioned near the cell membrane. RET protein has a single transmembrane domain and an intracellular segment with two tyrosine kinase subdomains, TK1 and TK2. It is activated by ligand-induced dimerization [19].

Hereditary MTC is caused by autosomal dominant gain-of-function mutations in the RET proto-oncogene. Mutation of the extracellular cysteine at exon 11 codon 634 causes ligand-independent dimerization of receptor molecules, enhanced phosphorylation of intracellular substrates, and cell transformation. Mutation of the intracellular tyrosine kinase (codon 918) has no effect on receptor dimerization but causes constitutive activation of intracellular signalling pathways and also results in cellular transformation. There is a significant age-related progression from C-cell hyperplasia to MTC, which correlates with the transforming capacity of the respective RET mutations.

At present, mutation analysis has identified over 100 different missense mutations associated with the development of MEN 2 (**Figure 3.5.7.1.**). Although some overlap exists between RET mutations and the resulting clinical subtype of MEN 2, most of patients with MEN 2A have a mutation of codon 634 (exon 11), followed by mutations of codons 609, 611, 618, and 620 (exon 10). In FMTC, germline mutations are distributed throughout the RET gene with an accumulation in exon 13 (codons 768, 790, and 791), exon 14 (codons 804 and 844), and rarely exon 10 (codons 618 and 620); MEN 2B patients have mutations in codon 918 (exon 16), but mutations are rarely identified at codon 883 exon 15 [20]. Phaeochromocytomas are associated with codon 634 and 918 mutations in approximately

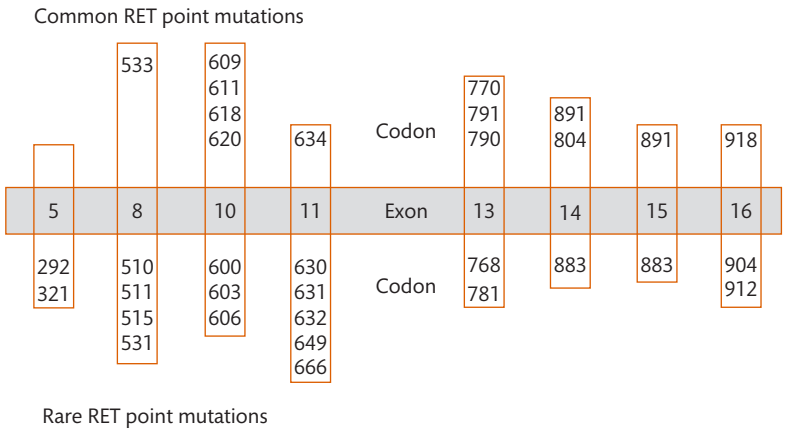


Figure 3.5.7.1 Germline mutations of the RET proto-oncogene associated with MEN 2 and FMTC. Numbers indicated mutated codons of the RET gene.

40–50% of patients, and are associated with mutations in exon 10 (codon 609, 611, 618, 620) in about 30% of patients and rarely in exon 15 (codon 791, 804) [21]. Hyperparathyroidism in MEN 2A is most commonly associated with codon 634 mutations, and in particular with the C634R mutation. The association between disease phenotype and RET mutation genotype has important implications for the clinical management of MEN 2 patients and their families. There is a correlation between the specific germline RET mutation and the age of onset and aggressiveness of MTC development and the presence of nodal metastases [22].

This information is used to stratify RET mutations into three risk levels namely moderate, high, and highest risk concerning the age of onset and aggressiveness of MTC: MEN2-patients with ATA (American Thyroid Association)-moderate risk have mutations at exon 10 (codons 609, 611, 618, 620), exon 13–15 (codon 768, 790, 791, 804, and 891), with ATA higher risk at exon 11 (codon 634) and at exon 15 (codons 883) and with ATA highest risk at exon 16 (codon 918). This risk classification is important in presymptomatic *RET* mutation carriers because prophylactic thyroidectomy must be performed before MTC development or while the tumour is confined to the thyroid gland. This strategy for preventing familial MTC should be tailored according to the specific risk of the mutation carried by each patient [2] (see next).

Approximately 23–60% of sporadic MTC have a somatic (present in tumour only) mutation at codon 918 identical to the germline mutation found in MEN 2B. Some reports suggest that patients with sporadic MTC with codon 918 somatic mutations have more aggressive tumour growth and a poorer prognosis [23].

Clinical Syndrome and Diagnostic Procedure

Sporadic Medullary Thyroid Carcinoma

The most common clinical presentation of sporadic MTC is a single nodule or thyroid mass found incidentally during routine examination [24]. The presentation does not differ from that observed in papillary or follicular thyroid carcinoma. A thyroid nodule identified by physical examination is generally evaluated by ultrasonography and radioisotopic scanning (Figure 3.5.7.2). MTC shows hypoechogenic regions, sometimes with calcifications, and a thyroid scan almost always shows no trapping of radioactive iodine or technetium. Cytological examination of the cold hypoechogenic nodule will lead to a strong suspicion, or a correct diagnosis in most cases, of sporadic MTC.

A plasma Ctn measurement can clarify the diagnosis, since preoperative Ctn levels correlate significantly with tumour size and, in the presence of a palpable MTC, the plasma Ctn concentration will usually be greater than 100 pg/ml. The CEA level will be elevated in most cases with clinically evident tumours. Therefore, measurement of plasma Ctn in patients with thyroid nodules has been advocated as a routine procedure by some European consensus groups [25].

Genetic testing for RET mutations in patients with elevated Ctn levels may also be helpful in apparently sporadic cases of MTC, since, if a mutation is found, it will imply that the disease is hereditary and that the family should be screened. The frequency of germline mutations, either inherited or de novo, in a larger series of apparently sporadic MTC patients varied between 1% and 7% [7].

Metastases to cervical and mediastinal lymph nodes are found in two-thirds of patients at the time of initial presentation. Distant metastases to lung, liver, and bone occur late in the course of the disease. Diarrhoea is the most prominent of the hormone-mediated clinical features of MTC and is often seen in patients with advanced disease. In addition, occasional tumours secrete ACTH causing Cushing's syndrome. Given the possibility that any patient with MTC may have MEN 2, preoperative testing must also include a 24-hour urinary excretion of catecholamines (to rule out phaeochromocytoma) and measurement of calcium (to rule out hyperparathyroidism).

Hereditary Medullary Thyroid Carcinoma

The clinical presentation and manifestation of familial MTC in index cases does not appear to differ from that in patients with sporadic MTC. MTC is often the initial manifestation of MEN 2 syndrome, as the other manifestations, phaeochromocytoma and hyperparathyroidism, develop later in the course of the disease. Less common presentations of MTC include recognition during search initiated after an associated disease such as bilateral pheochromocytoma or multiglandular hyperparathyroidism becomes apparent. The diagnosis of familial MTC in index cases is often made postoperatively

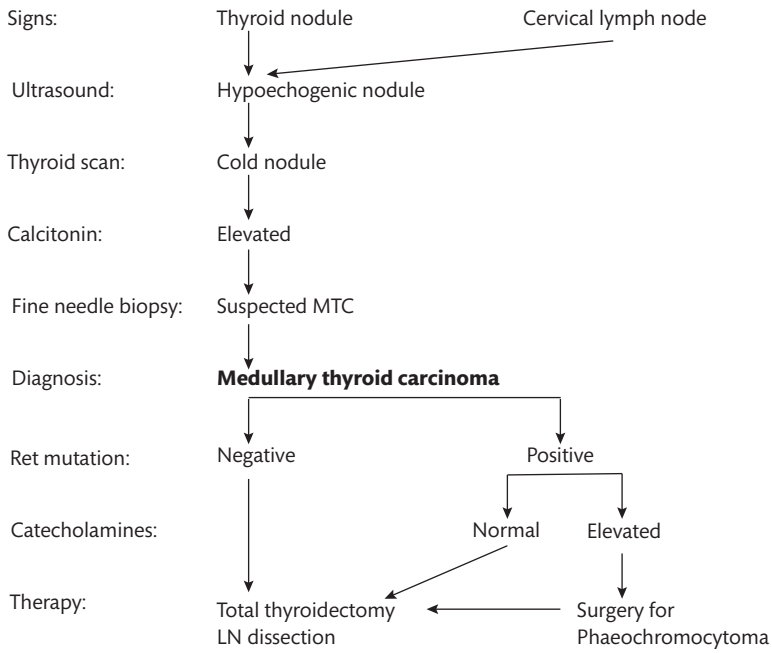


Figure 3.5.7.2 Clinical evaluation of patients at risk for medullary thyroid carcinoma.

when pathohistological examination may show multifocal bilateral MTC accompanied by diffuse C-cell hyperplasia. Rare variants of MEN 2A exist, including MEN 2A with cutaneous lichen amyloidosis and FMTC (or MEN 2A) with Hirschsprung's disease. The diagnosis of FMTC can only be considered when four or more family members across a wide range of ages have isolated MTC. In general, the clinical course of MTC in familial MTC is more benign and typically has a late onset or is not a clinically manifest disease [26].

MEN 2B has a typical phenotype with visible physical stigmata such as raised bumps on the lips and tongue (due to cutaneous neuromas), ganglioneuromas throughout the gastrointestinal tract, and a marfanoid habitus (long thin extremities, an altered upper–lower body ratio, slipped femoral epiphysis, pectus excavatum) with skeletal deformations and joint laxity. These patients have disease onset in the first year of life with the most aggressive form of MTC.

DNA testing becomes the optimal test for early detection of MEN 2 especially in 'at risk' families. At present, genetic testing is performed before the age of 5 years in all first-degree relatives of an index case (in MEN 2B patients directly after birth). Mutations in the RET proto-oncogene can be used to confirm the clinical diagnosis and identify asymptomatic family members with the syndrome (Figure 3.5.7.3). Those who have a negative test can be reassured and require no further biochemical screening.

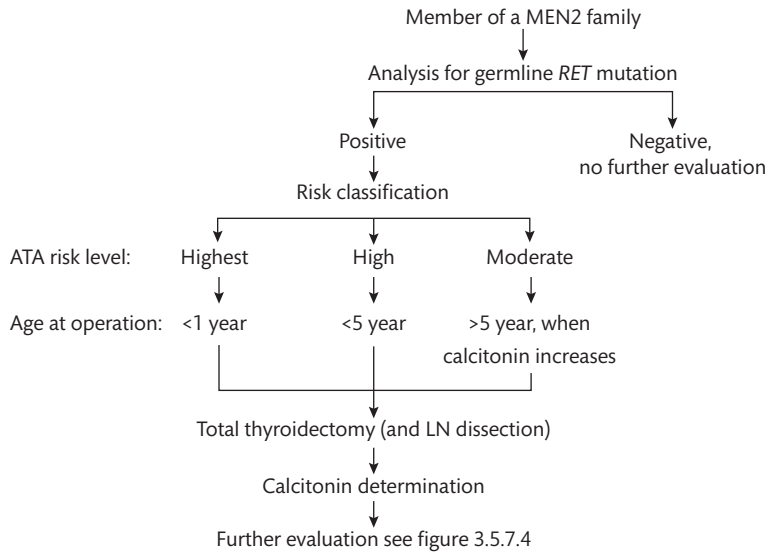


Figure 3.5.7.3 Work-up of family members at risk for medullary thyroid carcinoma/multiple endocrine neoplasia type 2.

The age of onset of MTC and tumour aggressiveness in MEN 2 depends on the codon mutated. This genotype–phenotype correlation is the basis for stratifying mutations into three risk levels concerning the risk for MTC development and growth. Decision making in the clinical management of MEN 2 patients depends on the risk level classification and calcitonin level, especially the timing of prophylactic thyroidectomy and the extent of surgical resection in presymptomatic RET mutation carriers.

Phaeochromocytoma

Phaeochromocytomas occur in approximately 20–50% of MEN 2A patients depending on the mutation. Phaeochromocytomas are associated with codon 634 and 918 mutations in approximately 50% of patients, and are associated with mutations in exon 10 (codons 609, 611, 618, and 620) in about 30% of patients and rarely in exon 15 (codons 791 and 804) [21]. As with MTC, the phaeochromocytomas of MEN 2 are also multicentric with diffuse adrenomedullary hyperplasia developing bilateral phaeochromocytomas in one-half of the cases, but often after an interval of several years. Almost all phaeochromocytomas are located in an adrenal gland, and malignant phaeochromocytomas are rare. In index cases, the clinical manifestation of phaeochromocytoma associated with MEN 2 is similar to that in sporadic cases with signs and symptoms such as headache, palpitations, nervousness, tachycardia, and hypertension. However, phaeochromocytomas are usually identified early as a result of regular biochemical screening in gene carriers, and clinical manifestations are thus subtle or absent. It is unusual for phaeochromocytoma to precede the development of MTC and be the initial manifestation of MEN 2. Annual biochemical screening by measuring plasma and/or 24-h urinary excretion of catecholamines and metanephrines should be performed. Once the biochemical diagnosis is made, imaging studies such as MRI or metaiodobenzylguanidine (MIBG) scanning are appropriate. The presence of phaeochromocytoma must be ruled out before any surgical procedure. Patients with MTC should be evaluated for possible phaeochromocytoma. A coexisting phaeochromocytoma should be removed before thyroidectomy.

Primary Hyperparathyroidism

Primary hyperparathyroidism, with hypercalcaemia and an elevated serum parathyroid hormone level occurs in 10–25% of MEN 2 gene carriers (especially codon 634). Hyperparathyroidism develops slowly, is usually mild, and clinical features do not differ from those seen in mild sporadic hyperparathyroidism. The diagnosis is established by finding high parathyroid hormone concentrations in the presence of hypercalcaemia. Pathological findings show chief cell hyperplasia involving multiple glands. Annual measurement of serum calcium concentration in gene carriers is probably adequate for screening purposes [27].

Treatment and Prognosis

Surgery

The definitive treatment for MTC is surgery no matter whether MTC is sporadic or familial. Several studies have shown that survival in patients with MTC is dependent upon the adequacy of the initial surgical procedure. The appropriate surgery for MTC is total thyroidectomy and careful lymph node dissection of the central and if necessary lateral compartment of the neck [28]. The presence of 10 and more cervical lymph node metastases is inconsistent with surgical cure [8]. Central lymph node dissection is also necessary for tumour staging and prevention of later midline complications related to local metastatic disease. If there is no evidence of local lymph node metastases during the primary surgical procedure, a surgical cure is likely and further neck dissection is probably unnecessary. The preoperative Ctn level can predict nodal metastasis and guide extent of neck dissection: Ctn levels of 20, 50, 200, and 500pg/ml predict nodal metastases in the ipsilateral central and lateral neck, contralateral central neck, contralateral lateral neck, and upper mediastinum, respectively [14]. Total thyroidectomy is absolutely necessary in hereditary cases because of the bilateral and multifocal nature of MTC. In contrast, unilateral lobectomy is sufficient in a patient with incidentally discovered sporadic MTC showing a single unilateral tumour focus and normal/ not measurable plasma Ctn levels after operation. All patients should receive adequate thyroxine replacement therapy after total thyroidectomy.

Patients with persistent elevation of plasma Ctn after total thyroidectomy should be thoroughly evaluated to define the extent of local and distant disease. Persistent elevation of plasma Ctn implies the presence of tumour. If there is no evidence of distant metastases, Ctn level below 1000pg/ml and if local disease is suspected/ found in the neck, and/or the initial surgical procedure was inadequate, then reoperation is advocated using meticulous dissection and microsurgical techniques. Unfortunately, many patients with MTC who have regional lymph node metastases also have systemic disease and are not cured biochemically despite aggressive surgery, including bilateral lateral neck dissection [14, 29]. If distant metastases are found, there is no indication for surgical reintervention unless the patient develops symptoms like local complications, for which tumour debulking may be beneficial.

Recommendations for the timing of prophylactic thyroidectomy in MEN 2 patients are based upon a model that utilizes genotype–phenotype correlations to stratify mutations into three risk levels. In the cases of highest risk mutations, patients with mutation of codons 918, a thyroidectomy is recommended as early as possible in the first year after birth, with higher risk mutation, patients with codon 634 and 883 mutation at the age of 5 years and with moderate risk mutations (codons 609, 611, 618, 620, 768, 790, 791, 804, and 891) thyroidectomy is recommended at age of 5, or surgery may be postponed until the Ctn value increases in the pathological range. By operating these patients before MTC has spread to neck lymph nodes or beyond they could be cured. Currently, the generally accepted practice is to use a combination of genetic testing and the basal serum Ctn level to decide the timing of thyroidectomy. The surgeon can avoid dissecting the central zone of the neck if calcitonin is in the normal range, since lymph node metastases rarely occur by this age.

Surgery for phaeochromocytoma in MEN 2 should precede surgery for MTC. Before adrenalectomy all patients should receive appropriate pharmacotherapy (α - with/or without β -adrenergic antagonist). Approximately one-third of patients who undergo a unilateral adrenalectomy will eventually require a second operation for contralateral phaeochromocytoma, but this may not occur for many years, during which time the patient will not be steroid dependent. Adrenal cortical-sparing adrenalectomy is an appropriate technique for preventing adrenal insufficiency [21, 30].

The parathyroid glands in MEN 2 patients are frequently found to be enlarged at thyroidectomy for MTC and should therefore be carefully evaluated. The goal in MEN 2 patients with primary hyperparathyroidism is to excise the enlarged glands and to leave at least one normal parathyroid gland intact. If they are all enlarged, a subtotal parathyroidectomy or total parathyroidectomy with autotransplantation should be performed [31].

Postsurgical Follow-up and Management

After surgery, patients with MTC should be assessed regarding the presence of residual disease, the localization of metastases, and the identification of progressive disease. Postoperative staging is used to separate low-risk from high-risk patients with MTC. A risk stratification system using TNM/AJCC staging system, nadir of CTN and CEA measured within the first year after initial treatment, and imaging studies identifying metastases allows to stratify MTC patients in three risk groups: excellent (biochemical cured), biochemical incomplete (detectable Ctn, and no evidence of disease initially), and structural incomplete (metastases at diagnosis) [32].

Normal/not measurable Ctn levels suggest an excellent tumour-free state and thus patients require no further treatment (biochemically cured) [33]. They can be followed-up at yearly intervals with physical examination and Ctn determination (Figure 3.5.7.4). Only in a small number of patients Ctn may rise after years.

In patients remaining Ctn-positive with evidence of occult disease (no local recurrence is found and adequate operation has been done: biochemical incomplete), or non-curable and non-operable disease (diffuse distant metastases: structural incomplete), close observation of changes in tumour markers and tumour growth is required. If the postoperative serum Ctn level exceeds 150 pg/ml patients should be evaluated by imaging procedures including ultrasonography of the neck and abdomen, CT (computed tomography) of neck, mediastinum, lung, and liver, or an MRI technique, and, bone scintigraphy. [18F]2-fluoro-2-deoxy-glucose (FDG), or the more sensitive 6-.^{8}F-fluoro-L-dihydroxyphenylalanine (.^{8}F-DOPA) positron emission tomography (PET) scanning in combination with CT (PET/CT) may also be helpful, especially in identifying yet unknown metastases [34]. In the absence of treatment, imaging should be repeated every 6–12 months or at more frequent intervals if progressive disease is present.

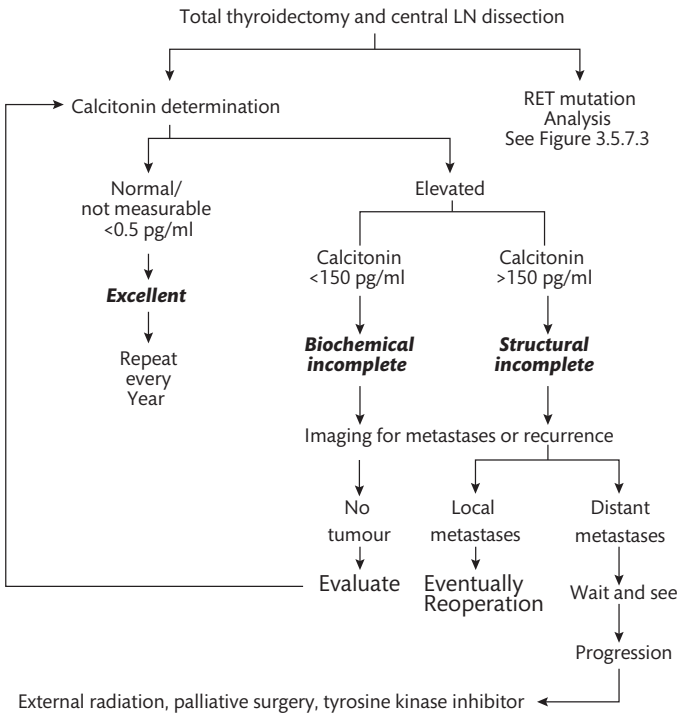


Figure 3.5.7.4 Recommended postoperative management of patients with medullary thyroid carcinoma. LN, lymph node.

One can estimate the growth rate of MTC metastases from sequential imaging studies using response evaluation criteria in solid tumours (RECIST) [35] that document increases in tumour size over time and by measuring serum levels of Ctn or CEA over multiple time points to determine the tumour marker doubling time [36–38]. Because serum Ctn concentration is closely associated with tumour volume, the doubling time is a strong predictor of tumour progression and survival.

Many patients may exhibit a remarkably stable indolent course over years and no further treatment is recommended; a 'watchful waiting' approach is advocated, as experience with non-surgical therapy in the management of slowly growing metastatic MTC has been disappointing [39].

Local treatment approaches are used in patients with local symptoms or those at high risk of local complications, preferably before initiation of systemic treatment. Palliative surgery and external radiation therapy to the neck or mediastinum might also be indicated in patients with tumours entering the aerodigestive structures. In patients with inoperable symptomatic tumour, radiotherapy can offer prolonged palliation and achieve local tumour control. External radiation therapy can procure rapid relief of bone pain and increase recalcification of bone lesions. Surgery could be indicated for bone metastases in patients with orthopaedic or neurological complications. Bisphosphonates (especially zoledronic acid) and the drug denosumab (directed at receptor activator of nuclear factor-κ B ligand) have been shown to improve local pain and to delay time to occurrence of skeletal-related events in other solid tumours with bone metastases. Symptomatic treatments of diarrhoea are done with loperamide, diphenoxylate, or atropine methonitrate, and codeine.

In a small subset of patients whose disease shows rapid and steady progress, e.g. more than 20% RECIST criterion and doubling of tumour marker in less than 1 year, intervention with tyrosine kinase inhibitors (TKI) like vandetanib (200–300 mg/d) [40] or cabozantinib (80–120 mg/d) [41] can be considered as a palliative therapeutic modality. With these substances, significant improvement of progression-free survival but not overall survival in patients with progressive MTC can be achieved. Life quality, toxic side effects, and survival have to be taken into account when TKI is recommended.

Once metastases appear the clinician must decide which patients require therapy, balancing the often slow rate of tumour progression associated with a good quality of life, against the limited efficacy and potential toxicities of local and systemic therapies. Considering that metastatic MTC is incurable the management goals are to provide loco-regional disease control, palliate symptoms of hormonal excess such as diarrhoea, palliate symptomatic metastases like pain or bone fracture, and control metastases that threaten life such as bronchial obstruction or spinal cord compression.

Prognostic Factors

The natural history of sporadic MTC is variable. The spectrum ranges from years of dormant residual disease after surgery to rapidly progressive disseminated disease and death related to either metastatic thyroid tumour or complications of phaeochromocytoma in MEN 2. The 10-year survival rates for all MTC patients ranges from approximately 61% to 76%. The overall prognosis is inferior to differentiated papillary and follicular carcinoma of the thyroid and much better than in the more aggressive anaplastic thyroid cancer. There is general agreement that tumour stage at diagnosis and surgical management have a favourable influence on the clinical course of the disease [8]. Early detection and surgical treatment of MTC is likely to be curative; more than 95% of patients detected at an early stage of disease remain disease-free (normal or undetectable Ctn values). The main factors that influence survival are the stage of disease at the time of diagnosis, size of the tumour, and lymph node involvement. The excellent prognosis associated with identification of MTC at its earliest stage underscores the importance of prospective screening (calcitonin screening) and early diagnosis (RET mutation analysis) which must be followed by adequate therapy.

Future

Early detection of MTC can enhance clinical outcome with a high surgical cure rate before it has spread to neck nodes or beyond. This goal is reached in most MEN2 patients by stratifying them depending on RET mutation into three risk groups concerning the age of onset and aggressiveness of MTC, enabling personalized approaches to appropriate treatment. This goal is not reached in patients with sporadic MTC and index cases of MEN2; increased clinical awareness especially in MEN2B patients, systematic Ctn screening in patients with thyroid nodules, primary operation of MTC by high volume surgeons, identifying high-risk patients in the follow-up, monitoring them appropriately, and treating advanced MTC with molecular targeted therapeutics is necessary. Advances in our understanding of the molecular pathways in MTC tumour cells may aid in the development of individualized risk stratification and therapeutic modalities based on target specific inhibition of tumour growth.

REFERENCES

1. Nilsson M, Williams ED. Medullary thyroid carcinoma—distinction towards neuroectodermal tumours: reply to the letter by Paschou and Vryonidou. *Eur Thyroid J*, 2016; **5**: 279–80.
2. Wells SA, Jr., Asa SL, Dralle H, *et al.* Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*, 2015; **25**: 567–610.
3. Raue F, Frank-Raue K. Update multiple endocrine neoplasia type 2. *Fam Cancer*, 2010; **9**: 449–57.
4. Wells SA, Jr., Pacini F, Robinson BG, Santoro M. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. *J Clin Endocrinol Metab*, 2013; **98**: 3149–64.
5. Elisei R, Romei C, Cosci B, *et al.* RET genetic screening in patients with medullary thyroid cancer and their relatives: experience with 807 individuals at one center. *J Clin Endocrinol Metab*, 2007; **92**: 4725–9.
6. Raue F, Dralle H, Machens A, Bruckner T, Frank-Raue K. Long-term survivorship in multiple endocrine neoplasia type 2B diagnosed before and in the new millennium. *J Clin Endocrinol Metab*, 2018; **103**: 235–43.
7. Romei C, Cosci B, Renzini G, *et al.* RET genetic screening of sporadic medullary thyroid cancer (MTC) allows the preclinical diagnosis of unsuspected gene carriers and the identification of a relevant percentage of hidden familial MTC (FMTC). *Clin Endocrinol (Oxf)*, 2011; **74**: 241–7.
8. Machens A, Dralle H. Surgical cure rates of sporadic medullary thyroid cancer in the era of calcitonin screening. *Eur J Endocrinol*, 2016; **175**: 219–28.
9. Scheuba C, Kaserer K, Kaczirek K, Asari R, Niederle B. Desmoplastic stromal reaction in medullary thyroid cancer-an intraoperative 'marker' for lymph node metastases. *World J Surg*, 2006; **30**: 853–9.
10. Schmid KW. Histopathology of C cells and medullary thyroid carcinoma. *Recent Results Cancer Res*, 2015; **204**: 41–60.
11. Kahaly GJ, Algeciras-Schimnich A, Davis TE, *et al.* United States and European multicenter prospective study for the analytical performance and clinical validation of a novel sensitive fully automated immunoassay for calcitonin. *Clin Chem*, 2017; **63**: 1489–96.
12. Costante G, Meringolo D, Durante C, *et al.* Predictive value of serum calcitonin levels for preoperative diagnosis of medullary thyroid carcinoma in a cohort of 5817 consecutive patients with thyroid nodules. *J Clin Endocrinol Metab*, 2007; **92**: 450–5.
13. Mian C, Perrino M, Colombo C, *et al.* Refining calcium test for the diagnosis of medullary thyroid cancer: cutoffs, procedures, and safety. *J Clin Endocrinol Metab*, 2014; **99**: 1656–64.
14. Machens A, Dralle H. Biomarker-based risk stratification for previously untreated medullary thyroid cancer. *J Clin Endocrinol Metab*, 2010; **95**: 2655–63.
15. Frank-Raue K, Machens A, Leidig-Bruckner G, *et al.* Prevalence and clinical spectrum of nonsecretory medullary thyroid carcinoma in a series of 839 patients with sporadic medullary thyroid carcinoma. *Thyroid*, 2013; **23**: 294–300.
16. Donis-Keller H, Dou S, Chi D, *et al.* Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum Mol Genet*, 1993; **2**: 851–6.
17. Mulligan LM, Kwok JB, Healey CS, *et al.* Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature*, 1993; **363**: 458–60.
18. Krampitz GW, Norton JA. RET gene mutations (genotype and phenotype) of multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma. *Cancer*, 2014; **120**: 1920–31.
19. Romei C, Ciampi R, Elisei R. A comprehensive overview of the role of the RET proto-oncogene in thyroid carcinoma. *Nat Rev Endocrinol*, 2016; **12**: 192–202.
20. Mathiesen JS, Habra MA, Bassett JHD, *et al.* Risk profile of the RET A883F germline mutation: an international collaborative study. *J Clin Endocrinol Metab*, 2017; **102**: 2069–74.
21. Mucha L, Leidig-Bruckner G, Frank-Raue K, *et al.* Phaeochromocytoma in multiple endocrine neoplasia type 2: RET codon-specific penetrance and changes in management during the last four decades. *Clin Endocrinol (Oxf)*, 2017; **87**: 320–6.
22. Voss RK, Feng L, Lee JE, *et al.* Medullary thyroid carcinoma in MEN2A: ATA moderate- or high-risk RET mutations do not predict disease aggressiveness. *J Clin Endocrinol Metab*, 2017; **102**: 2807–13.

23. Schilling T, Burck J, Sinn HP, *et al*. Prognostic value of codon 918 (ATG-->ACG) RET proto-oncogene mutations in sporadic medullary thyroid carcinoma. *Int J Cancer*, 2001; **95**: 62–6.
24. Ahmed SR, Ball DW. Clinical review: incidentally discovered medullary thyroid cancer: diagnostic strategies and treatment. *J Clin Endocrinol Metab*, 2011; **96**: 1237–45.
25. Chambon G, Alovisetti C, Idoux-Louche C, *et al*. The use of preoperative routine measurement of basal serum thyrocalcitonin in candidates for thyroidectomy due to nodular thyroid disorders: results from 2733 consecutive patients. *J Clin Endocrinol Metab*, 2011; **96**: 75–81.
26. Frank-Raue K, Rybicki LA, Erlic Z, *et al*. Risk profiles and penetrance estimations in multiple endocrine neoplasia type 2A caused by germline RET mutations located in exon 10. *Hum Mutat*, 2011; **32**: 51–8.
27. Scholten A, Schreinemakers JM, Pieterman CR, Valk GD, Vriens MR, Borel Rinkes IH. Evolution of surgical treatment of primary hyperparathyroidism in patients with multiple endocrine neoplasia type 2A. *Endocr Pract*, 2011; **17**: 7–15.
28. Jin LX, Moley JF. Surgery for lymph node metastases of medullary thyroid carcinoma: a review. *Cancer*, 2016; **122**: 358–66.
29. Machens A, Dralle H. Prognostic impact of N staging in 715 medullary thyroid cancer patients: proposal for a revised staging system. *Ann Surg*, 2013; **257**: 323–9.
30. Castinetti F, Qi XP, Walz MK, *et al*. Outcomes of adrenal-sparing surgery or total adrenalectomy in phaeochromocytoma associated with multiple endocrine neoplasia type 2: an international retrospective population-based study. *Lancet Oncol*, 2014; **15**: 648–55.
31. Moley JF, Skinner M, Gillanders WE, *et al*. Management of the parathyroid glands during preventive thyroidectomy in patients with multiple endocrine neoplasia type 2. *Ann Surg*, 2015; **262**: 641–6.
32. Lindsey SC, Ganly I, Palmer F, Tuttle RM. Response to initial therapy predicts clinical outcomes in medullary thyroid cancer. *Thyroid*, 2015; **25**: 242–9.
33. Jung KY, Kim SM, Yoo WS, *et al*. Postoperative biochemical remission of serum calcitonin is the best predictive factor for recurrence-free survival of medullary thyroid cancer: a large-scale retrospective analysis over 30 years. *Clin Endocrinol (Oxf)*, 2016; **84**: 587–97.
34. Santhanam P, Taieb D. Role of (18) F-FDOPA PET/CT imaging in endocrinology. *Clin Endocrinol (Oxf)*, 2014; **81**: 789–98.
35. Therasse P, Arbuck SG, Eisenhauer EA, *et al*. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*, 2000; **92**: 205–16.
36. Laure Giraudet A, Al Ghulzan A, Auperin A, *et al*. Progression of medullary thyroid carcinoma: assessment with calcitonin and carcinoembryonic antigen doubling times. *Eur J Endocrinol*, 2008; **158**: 239–46.
37. Barbet J, Campion L, Kraeber-Bodere F, Chatal JF, Group GTES. Prognostic impact of serum calcitonin and carcinoembryonic antigen doubling-times in patients with medullary thyroid carcinoma. *J Clin Endocrinol Metab*, 2005; **90**: 6077–84.
38. Meijer JA, le Cessie S, van den Hout WB, *et al*. Calcitonin and carcinoembryonic antigen doubling times as prognostic factors in medullary thyroid carcinoma: a structured meta-analysis. *Clin Endocrinol (Oxf)*, 2010; **72**: 534–42.
39. Hadoux J, Pacini F, Tuttle RM, Schlumberger M. Management of advanced medullary thyroid cancer. *Lancet Diabetes Endocrinol*, 2016; **4**: 64–71.
40. Wells SA, Jr., Robinson BG, Gagel RF, *et al*. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol*, 2012; **30**: 134–41.
41. Elisei R, Schlumberger MJ, Muller SP, *et al*. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol*, 2013; **31**: 3639–46.

SECTION 4

Parathyroid, Calcium and Bone Metabolism Disorders

4.1

Parathyroid Anatomy, Hormone Synthesis, Secretion, Action, and Receptors

David Goltzman and Geoffrey N. Hendy†

Parathyroid Embryology, Anatomy, and Morphology

Humans have two pairs of parathyroid glands lying in the anterior cervical region. The fetal parathyroid glands begin developing at 5 weeks from the third and fourth pharyngeal pouches. The third pharyngeal pouch that contains tissue that will become the thymus and parathyroid migrates downward and gives rise to the two inferior parathyroid glands normally located at the lower poles of the thyroid. The fourth pharyngeal pouch does not migrate and gives rise to the two upper parathyroid glands that normally are attached to the upper poles of the thyroid [1].

Eighty-five percent of normal adults have four parathyroid glands, but the number can vary markedly in some individuals. The location of the glands is also variable with the upper glands sometimes located behind the pharynx or the oesophagus. The lower glands may be found close to or within the thymus in the superior mediastinum. Because of the variability in location surgical exploration of the neck can be problematic especially in hyperparathyroidism of chronic kidney disease [2]. Use of preoperative localization techniques, such as technetium-99-sestamibi scanning, ultrasound, MRI, or CT scanning may be helpful in localizing enlarged parathyroid glands, however referral to an experienced parathyroid surgeon is essential to maximize localization of affected glands and minimize complications. Conversely, hypoparathyroidism most commonly occurs as a result of surgical excision of, or damage to, the parathyroid glands during non-parathyroid surgery, for example, total thyroidectomy for thyroid cancer and radical neck dissection for laryngeal or oesophageal carcinoma, as well as repeated surgery for hyperparathyroidism.

Most patients with primary hyperparathyroidism, about 80%, have a single benign adenoma [3]. Multiple (so-called) adenomas are rarely found and probably represent asynchronous parathyroid hyperplasia. Hyperplasia accounts for 15–20% of cases, and malignant parathyroid carcinoma is extremely rare, less than 1% of cases. In secondary hyperparathyroidism, all four glands are enlarged.

The chief cell is the predominant parathyroid cell type in humans, with some oxyphil cells, which have an acidophilic cytoplasm and mitochondria also present. Parathyroid cells have limited numbers of secretory granules containing parathyroid hormone (PTH), indicating that relatively little hormone is stored in the gland. Parathyroid cells normally divide at an extremely slow rate—mitoses are rarely observed.

Knowledge of the embryological formation of the parathyroids has been gained by study, on the one hand, of mouse models in which deletion of specific genes has led to lack of parathyroid gland development [4], and, on the other, of human inherited (autosomal dominant, autosomal recessive, or X-linked) familial hypoparathyroidism in which the formation of the parathyroid glands is defective [1].

The mouse has a single pair of parathyroid glands, and at day e10 both the precursor thymus and parathyroid cells in the third pharyngeal pouch endoderm, express the four transcription factors, Hoxa3, Pax1, Eya1, and Pax9. The conjoined thymus and parathyroid rudiment develop at day e11 and the primordium also expresses transcription factors Six1 and Pbx1. At day e12, separate pathways distinct for the different parts of the rudiment that will develop into

† It is with regret that we report the death of Geoffrey N. Hendy.

the thymus and parathyroid become apparent. Signalling molecules including sonic hedgehog, bone morphogenetic protein-4, noggin, and fibroblast growth factor-8 act in a complex fashion to affect the outgrowth of the parathyroid precursor. By day e13.5, the parathyroid cell mass and thymus cell mass are separate. The thymic cells express *Foxn1* that is not present in the parathyroid cells and that in turn specifically express glial cells missing-2 (Gcm2). *Gcm2* expression continues into adulthood and it transactivates the calcium-sensing receptor (*Casr*) gene and thereby influences the expression of the parathyroid calciostat [5].

In humans, hypoparathyroidism is part of several syndromes, for example, the DiGeorge type 1 syndrome that occurs because of a 22q11.2 microdeletion. Congenital defects arise as a result of the failure to develop the derivatives of the third and fourth pharyngeal pouches leading to agenesis or hypoplasia of the parathyroid glands and thymus [6]. Haploinsufficiency of the *TBX1* transcription factor gene appears to play an important role in the disorder. In DiGeorge syndrome type 2 patients harbour 10p13-14 deletions encoding the actin-binding protein nebulette (NEBL) although its mechanistic link to hypoparathyroidism is not clear. Some DiGeorge patients have features of the CHARGE syndrome which is caused by heterozygous mutations in the chromodomain helicase DNA-binding 7 gene (CHD7) on 8q12.2 that is expressed within pharyngeal ectoderm. Hypoparathyroidism is a part of the Barakat or HDR (Hypoparathyroidism nerve Deafness, and Renal dysplasia) syndrome that maps to 10p15. HDR is due to haploinsufficiency and loss-of-function mutations in the *GATA3* gene that encodes a zinc finger transcription factor [7]. *GATA3* is essential for normal embryonic development of the parathyroids, auditory system, and kidney in humans. Hypoparathyroidism together with growth and mental Retardation, and characteristic Dysmorphism (HRD) occur in autosomal recessive Kenny-Caffey type 1 and Sanjad-Sakati syndromes. The HRD syndrome is due to mutations in the tubulin chaperone E (*TBCE*) gene that maps to 1q42.3 [8]. Autosomal dominant forms of Kenny-Caffey type 2 and Gracile bone dysplasia with hypoparathyroidism are due to mutations in the family with sequence similarity 111 member A (*FAM111A*) gene on 11q12.1 involved in DNA replication and chromatin maturation.

Rare cases of primary hypoparathyroidism inherited in either an autosomal recessive or dominant manner due to mutations in the *GCM2* gene on chromosome 6p24 have been identified. In the latter case, the mutated GCM2 acts in a dominant-negative fashion [5]. Heterozygous and homozygous mutations in the *PTH* gene cause rare cases of hypoparathyroidism and more commonly heterozygous activating mutations in the *CASR* gene are implicated in hypoparathyroidism (see next). In an X-linked recessive form of hypoparathyroidism there is an interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1 near the *SOX3* gene that encodes a high mobility group box transcription factor. It is proposed that the hypoparathyroidism is caused by disruption of regulatory elements of the *SOX3* gene [1].

Parathyroid Hormone Synthesis

In mammals PTH is the 84-amino-acid product [9, 10] of a single-copy gene (**Figure 4.1.1**). The gene, which encodes a larger precursor molecule of 115 amino acids, preproparathyroid hormone (preproPTH), is organized into three exons. Exon I encodes the 5'-untranslated region of the messenger RNA, exon II encodes the NH_2-terminal pre- or signal peptide and a part of the short propeptide, and exon III encodes the Lys^{-2}–Arg^{-1} of the prohormone cleavage site, the 84 amino acids of the mature hormone, and the 3'-untranslated region of the mRNA (**Figure 4.1.2**). The importance of correct splicing of the primary *PTH* gene transcript, or premessenger RNA, was emphasized by the identification of a donor splice mutation in the *PTH* gene in affected members of a family with autosomal recessive isolated hypoparathyroidism, resulting in the loss of exon II which encodes the initiation codon and signal peptide [1].

A second member of the *PTH* gene family encodes the parathyroid hormone-related protein (PTHrP) which is the responsible causal factor in the majority of cases of hypercalcaemia associated with malignancies. PTHrP plays a critical role in fetal development, especially skeletogenesis [11, 12], but is not involved in normal calcium homeostatic control in the adult. In postnatal life, PTHrP regulates the epithelial mesenchymal interactions that are critical for development of the mammary gland, skin, and hair follicle. The *PTH* gene and *PTHLH* gene that encodes PTHrP map to chromosomes 11p15 and 12p12.1-11.2, respectively. These two human chromosomes are thought to have arisen by an ancient duplication of a single chromosome, and their respective gene clusters have been maintained as syntenic groups across the genomes of several species. Because of the similarity in NH_2-terminal sequence of their mature peptides, their gene organization, and chromosomal locations, it is likely that the *PTH* and *PTHLH* genes evolved from a common ancestral gene, with that for PTHrP being the more ancient gene.

The gene for tuberoinfundibular peptide of 39 residues (*TIP39*), a more distantly related member of the *PTH* gene family, resides on chromosome 19q13.33. TIP39 is a neuropeptide [13]. The *TIP39* gene shares organizational features with the *PTH* and *PTHrP* genes having one exon encoding the 5'UTR, one encoding the precursor leader sequence, and one encoding the prohormone cleavage site and the mature peptide (**Figure 4.1.2**).

Transcription of the *PTH* gene occurs almost exclusively in the endocrine cells of the parathyroid gland, and is subject to strong repressor activity in all other cells. Ectopic PTH synthesis (i.e. synthesis outside parathyroid tissue) has been documented in only a very few cases of malignancies associated with hypercalcaemia. Activation of genes in a particular tissue is often related to demethylation of cytosine residues, and the *PTH* gene in parathyroid cells is hypomethylated at CpG residues relative to other tissues. In one of the few cases of true ectopic PTH production, involving a pancreatic tumour, the upstream regions of the *PTH* gene were abnormally hypomethylated [14]. The human *PTH* gene has two functional TATA box-controlled transcription start sites, a cyclic AMP response element (CRE), and a negative vitamin D response element (VDRE) in its proximal promoter. While *PTH* gene transcription is negatively regulated by the hormonally active metabolite of vitamin D, $1,25(OH)_2D$, any regulation by extracellular calcium remains to be established. Also located distally are sequences that function to silence transcription in non-parathyroid cells. In a further case of ectopic PTH production, an ovarian carcinoma, this repressor regulatory region was replaced by a foreign

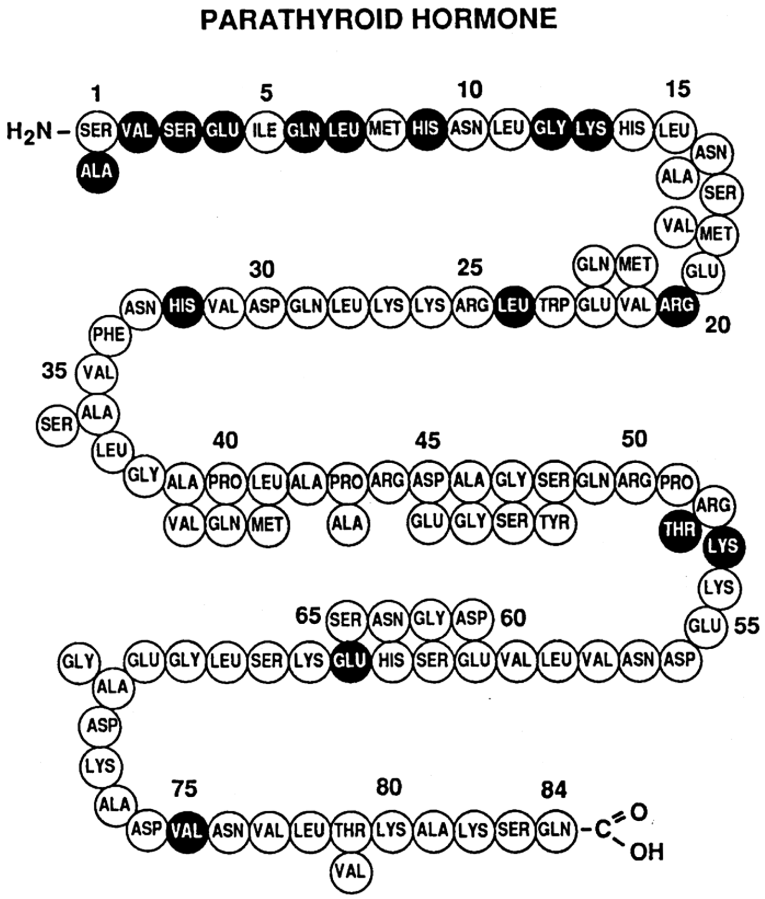


Figure 4.1.1 Amino acid sequence of mammalian PTH. The backbone sequence is that of the human with substitutions in the rat hormone shown at specific sites. Biological activity is a property of the amino-terminal one-third of the molecule [PTH(1-34)]. The solid circles show those amino acids that are identical in the human and rat PTH and PTH-related peptide (PTHrP) molecules.

sequence that allowed inappropriate transcription of the *PTH* gene to take place [3].

The human PTH produced by patients with hyperparathyroidism is structurally normal [9, 10]. In a small number of parathyroid tumours examined, the *PTH* gene sequence is rearranged, and the 5' flanking region of the *PTH* gene is placed upstream of the cyclin D1 (*CCND1*) gene located on the long arm of chromosome 11. This is thought to lead to deregulated expression of the *CCND1* gene that contributes to tumour development [3]. However, this type of gene arrangement occurs very infrequently in parathyroid tumours.

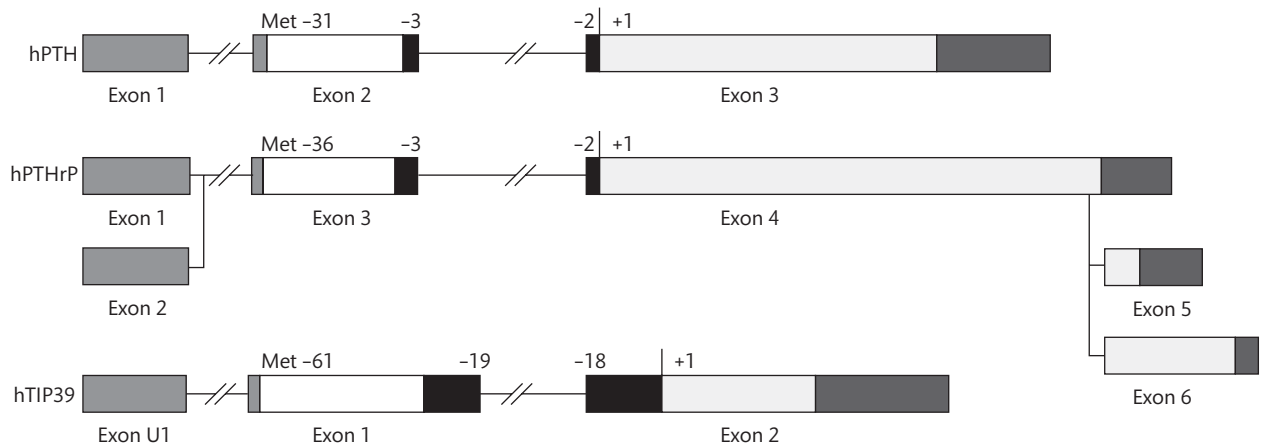


Figure 4.1.2 Comparison of structural organization of the human *PTH*, *PTHrP*, and *TIP39* genes. Exons are boxed: from left to right, dark grey boxes denote 5'UTRs, white boxes denote presequences, black boxes denote prosequences, light grey boxes denote mature polypeptide sequences, and dark grey boxes denote 3'UTRs.

A more common event involves the loss or inactivation of the multiple endocrine neoplasia type 1 (*MEN1*) gene, also on the long arm of chromosome 11. The protein encoded by the *MEN1* gene [15, 16] called menin, is a 610-amino acid nuclear protein [17]. Germline mutations in the *MEN1* gene cause familial and sporadic MEN1 and are found in 20% of non-MEN1 parathyroid adenomas. Loss of heterozygosity at 11q13 is found in MEN1 tumours and sporadic parathyroid adenomas consistent with *MEN1* being a tumour suppressor gene.

A target of the Wnt pathway, β-catenin, encoded by the *CTNNB1* gene, is a candidate for involvement in parathyroid neoplasia. Very few of the parathyroid adenomas examined so far have stabilizing missense *CTNNB1* mutations suggesting that mutation of the β-catenin gene itself is unlikely to be involved in the initiation or early progression of parathyroid adenomatosis. However, other components of the Wnt signalling pathway (e.g. a constitutively active LRP5 receptor derived from an alternatively spliced mRNA may be implicated in parathyroid tumorigenesis) [18].

Early onset recurrent parathyroid tumours occur as part of the uncommon autosomal dominant hyperparathyroidism and jaw tumour (HPT-JT) syndrome in which parathyroid carcinoma is frequent. The responsible gene, *HRPT2* (also known as *CDC73*), at 1q31.2, encodes a novel transcription factor, parafibromin, of 531 amino acids [19]. Sporadic parathyroid carcinomas very commonly contain somatic mutations of the *HRPT2* gene and some of these patients harbour germline mutations. In these cases, genetic testing in family members provides for early diagnosis [3]. Loss of heterozygosity at chromosome 1q occurs in carcinomas of the familial and sporadic disorder usually by intragenic mutations.

PTH follows a pattern of biosynthesis and of vectorial transport through organelles of the cell similar to that of many other peptide hormones. It is biosynthesized on the polyribosomes of the rough endoplasmic reticulum (ER) of the parathyroid endocrine cell. The *PTH* gene encodes a precursor, preproPTH, extended at the aminoterminus of PTH 1-84 by 31 residues. The NH_2-terminal 25-residue portion, characterized by its hydrophobicity, is called the signal, leader, or presequence, and it facilitates entry of the nascent hormone into the cisternae of the ER. One patient with autosomal dominant hypoparathyroidism was reported to have a mutation within the protein coding region of the *PTH* gene in which there was a single base substitution (T->C) in exon II, resulting in the replacement of arginine (CGT) for cysteine (TGT) in the signal peptide. This places a charged amino acid in the hydrophobic core of the signal peptide, leading to inefficient processing of the mutant preproPTH to PTH [3]. In cases like this it is suggested that the mutant polypeptide acts in a dominant-negative fashion by promoting ER stress leading to apoptosis and pharmacological chaperones may be beneficial in restoring proper processing of the PTH [20].

Normally, as the signal sequence of the synthesized hormone emerges from the ribosome, it binds to a signal recognition particle that stops further synthesis of the nascent protein. The signal recognition particle carrying the ribosome then binds to an integral membrane protein of the ER, called the docking protein or signal recognition particle receptor. This protein releases the block in protein synthesis, and the nascent peptide is transported across the membrane into the cisternae of the ER. The signal sequence is simultaneously removed at the inner surface of the ER, at a glycyl-lysyl bond, by a signalase enzyme. (Therefore, note that under normal circumstances the preproPTH molecule never exists as a complete entity.) The resultant precursor molecule, proparathyroid hormone (proPTH), is extended at the NH_2-terminus of PTH 1-84 by only six amino acids. The pro sequence is necessary for efficient translocation and cleavage of the signal peptide. Once formed, proPTH is transported to the Golgi apparatus.

The prohormone hexapeptide has several basic residues that serve as a recognition sequence to yield the mature hormone. Unlike many other prohormones, proPTH does not contain another sequence at the COOH-terminus and has not been detected within the circulation even in states of parathyroid gland hyperfunction. ProPTH has little biological activity until cleaved to create the hormonal form [21]. The conversion of proPTH to PTH takes place within the *trans*-Golgi network rather than the secretory granules as occurs with other prohormones like proinsulin. The enzymes involved include furin and PC7, mammalian proprotein convertases that are related to bacterial subtilisins [22]. Little proPTH is stored within the gland.

The resultant mature 84-amino acid form of the hormone is packaged in secretory granules and transported to the region of the plasma membrane. The hormone is released by exocytosis in response to the principal stimulus to secretion, hypocalcaemia. The calcium ion does not influence the enzymatic cleavages involved in the processing of preproPTH or proPTH.

Parathyroid Hormone Secretion

Relatively little PTH is stored in secretory granules within the parathyroid glands. In the absence of a stimulus for release, intraglandular metabolism occurs, causing complete degradation to its constituent amino acids or partial degradation to fragments (**Figure 4.1.3**). This has been postulated to occur through a specific calcium-regulated enzymatic mechanism. In the case of hypercalcaemia, the predominant hormonal entities released from the gland are fragments comprising midregion or COOH-terminal sequences. In response to hypocalcaemia, degradation of PTH within the parathyroid cell is minimized, and the major hormonal entity released is the bioactive PTH 1-84 molecule. Thus, in the presence of hypocalcaemia, increased amounts of bioactive PTH are secreted, even in the absence of additional synthesis of hormone. Hormone stores are insufficient, however, to maintain secretion for more than a few hours in the presence of a sustained severe hypocalcaemic stimulus, and other mechanisms—transcriptional and posttranscriptional—come into play to increase hormone production. For example, hypocalcaemia promotes stabilization of the preproPTH mRNA leading to increased PTH synthesis. In the presence of a sustained severe hypocalcaemic stimulus, additional PTH secretion depends on an increase in the number of parathyroid cells. Such an increase may also be stimulated by the reduction in circulating 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] that often accompanies hypocalcaemia. Normally, the sterol inhibits parathyroid cell proliferation by suppressing expression of immediate early response genes, such as the *MYC* proto-oncogene [23].

A circadian rhythm has been reported for PTH secretion, with increased blood levels occurring at night and small amplitude pulses of PTH secretion occurring at much shorter intervals. This suggests neural or central nervous system influences on PTH secretion, or reflects circadian alterations in the levels of extracellular calcium.

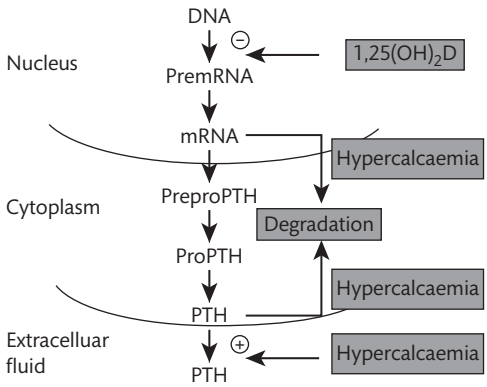


Figure 4.1.3 Schema of the sites of regulation of parathyroid hormone (PTH) biosynthesis, intraglandular degradation, and secretion. Both extracellular fluid calcium and 1,25-dihydroxyvitamin D levels negatively regulate transcription of the PreproPTH gene. Hypercalcaemia increases PreproPTH mRNA turnover and PTH degradation while hypocalcaemia stabilizes PreproPTH mRNA and promotes the production and synthesis of mature PTH.

Calcium

There is an inverse relationship between ambient calcium levels and PTH release that is curvilinear rather than proportional [24]. This relationship between PTH and extracellular calcium contrasts with the influence of the calcium ion as a secretagogue in most other secretory systems in which elevations in this ion enhance release of the secretory product. This distinction between the parathyroid cell and other secretory cells is maintained intracellularly, where elevations rather than decreases in cytosolic calcium correlate with decreased PTH release. Alterations in extracellular fluid calcium levels are transmitted through a parathyroid plasma membrane calcium-sensing receptor (CaSR) that couples through a Gq/11-protein complex to phospholipase C. Increases in extracellular calcium lead to increases in inositol 1,4,5-trisphosphate (IP_3) and mobilization of intracellular calcium stores. The CaSR also couples to a Gi-protein complex thereby inhibiting cyclic AMP production.

The human CaSR comprises 1078 amino acids with a large extracellular domain (ECD) (~600 amino acids), a seven-transmembrane-spanning domain and a cytoplasmic tail [25]. The CaSR is a member of group C of the G-protein-coupled receptor (GPCR) superfamily that includes the metabotropic glutamate, γ-aminobutyric acid-B, and taste and odorant receptors. These receptors function as dimers with the ECDs of each monomer having a so-called Venus flytrap domain consisting of two lobes that close upon the ligand leading to conformational changes in the transmembrane domain of the receptor allowing coupling of G proteins to the intracellular loops and the cytoplasmic tail. The CaSR has a low affinity for Ca^{2+} appropriate for it monitoring the relatively high levels of the mineral ion in the blood. Besides the parathyroid, the CaSR is also expressed in other cells having Ca^{2+}-sensing functions, such as those of the kidney tubule, the calcitonin secreting thyroid C-cells and in diverse other organs and tissues such as brain, bone, and cartilage, haematopoietic stem cells, keratinocytes, gastrointestinal tract, mammary gland, placenta, and vascular smooth muscle. Neomycin binds the receptor, which may account for the toxic renal effects of aminoglycoside antibiotics.

Inherited abnormalities of the *CASR* gene located on chromosome 3q13.3-21 can lead to either hypercalcaemia or hypocalcaemia depending upon whether they are inactivating or activating, respectively [25]. Heterozygous loss-of-function mutations give rise to familial (benign) hypocalciuric hypercalcaemia type 1 (FHH1) in which the lifelong hypercalcaemia is asymptomatic. The homozygous condition manifests itself as neonatal severe hyperparathyroidism (NSHPT), a rare disorder characterized by extreme hypercalcaemia and the bony changes of hyperparathyroidism. Cases of neonatal hyperparathyroidism (NHPT) may be caused by a paternal or de novo mutation in the *CASR* gene. FHH is heterogeneous and FHH2 and FHH3 are due to heterozygous inactivating mutations in the GNA11 gene (encoding the Gα11 protein) and *AP2S1* gene (encoding a subunit of the adaptor protein complex important for cell internalization of the CaSR), respectively. The disorder autosomal dominant hypocalcaemia type 1 (ADH1) is due to gain-of-function mutations in the *CASR* gene. ADH1 may be asymptomatic or present with neonatal or childhood seizures. Because of the overactive CaSR in the nephron, these patients are at a greater risk of developing renal complications during vitamin D therapy than patients with idiopathic hypoparathyroidism. ADH2 is due to heterozygous activating mutations in the GNA11 gene. A common polymorphism in the intracellular tail of the CaSR, Ala to Ser at position 986, has a modest effect on the serum calcium concentrations in healthy individuals [26]. *CASR* polymorphisms might also affect urinary calcium excretion and therefore the *CASR* is a candidate gene for involvement in disorders such as idiopathic hypercalciuria, and primary hyperparathyroidism.

The CaSR is a target for phenylalkylamine compounds—so-called calcimimetics—which are allosteric stimulators of the CaSR's affinity for cations. These orally active compounds have been approved for use in patients with chronic kidney disease and tertiary hyperparathyroidism, in parathyroid cancer, and in primary hyperparathyroidism when patients are not surgical candidates for parathyroidectomy. By their direct action on the parathyroid gland CaSR they provide an effective medical means of lowering PTH secretion and reducing hypercalcaemia [27]. Ongoing clinical trials in patients with mild primary hyperparathyroidism (PHPT) have shown that calcimimetics reduce serum calcium and PTH levels and increase serum phosphate levels but do not significantly affect bone turnover or bone mineral density (BMD).

The CaSR expressed in the developing parathyroid glands—and in the placenta—plays an important role in regulating fetal calcium concentrations. Normally, the fetal blood calcium level is elevated above the maternal level. This depends upon the action of PTHrP released from the placenta on placental calcium transport. Disruption of the CaSR, as shown by studies in CaSR-deficient mice, causes fetal hyperparathyroidism and hypercalcaemia due to fetal bone resorption. The transfer of calcium across the placenta is reduced and renal calcium excretion is increased.

Some patients with inactivating anti-CaSR autoantibodies, often associated with autoimmune disorders such as sprue or autoimmune thyroid disease, present as an FHH phenocopy termed acquired hypocalciuric hypercalcaemia (AHH). The anti-CaSR antibodies are directed against the ECD and interfere with elevated extracellular Ca^{2+}-mediated suppression of PTH release and perturb Ca^{2+}-sensing in the kidney, thereby closely mimicking FHH [24]. Activating autoantibodies that inhibit PTH secretion have

been identified in some patients with autoimmune hypoparathyroidism and the CaSR has also been identified as a self-antigen in patients with autoimmune polyendocrine syndrome type 1 (APS1) or acquired hypoparathyroidism associated with autoimmune hypothyroidism or idiopathic hypoparathyroidism. The activating autoantibodies are directed against epitopes in the ECD of the receptor.

In vivo, PTH mRNA levels are markedly stimulated by decreased circulating calcium concentrations. This occurs, in part, by a posttranscriptional mechanism whereby hypocalcaemia stabilizes and hypercalcaemia destabilizes the PTH mRNA. Prolonged hypocalcaemia *in vivo* may stimulate DNA replication, cell division, and the production of increased numbers of parathyroid cells or parathyroid hyperplasia. This would increase the synthesis of proteins, including PTH, within the hypercellular parathyroid gland and ultimately would increase PTH release. In primary parathyroid gland hyperfunction resulting in hyperparathyroidism, alterations in the calcium-sensing mechanism may manifest as a set-point error, producing a shift to the right of the curve relating PTH secretion to extracellular calcium levels. Consequently, elevated concentrations of extracellular fluid calcium may be required to reduce PTH secretion, resulting in an adenomatous or hyperplastic parathyroid gland that is incompletely suppressed by calcium. Such a mechanism may underlie the observation that an increase in the mass of parathyroid tissue like that produced by transplantation can be associated with hypercalcaemia. The parathyroid glands of patients with primary and severe uremic secondary hyperparathyroidism have reduced CaSR expression as assessed by immunostaining. Loss of a functional CaSR as in humans with NSHPT or in mice in which the *Casr* gene has been ablated leads to severe parathyroid hyperplasia. If basal secretion per cell produces a significant amount of bioactive PTH, the cumulative increase in this basal or non-calcium-suppressible secretion arising from an increase in parathyroid cells could also be responsible for the hypercalcaemia. The precise mechanistic relationship of extracellular calcium to parathyroid cell growth remains to be determined.

1,25-Dihydroxyvitamin D

Vitamin D metabolites modulate PTH release. There is a feedback loop between PTH-induced increase in $1,25(OH)_2D$ and vitamin D metabolite-induced decrease of PTH levels [28]. This latter effect is achieved by a direct action, via the vitamin D Receptor (VDR) on *PTH* gene transcription, thus altering the quantities of hormone available for immediate release by secretagogues.

Other Factors

In addition to calcium and vitamin D metabolites, several other factors influence the release of PTH from parathyroid glands. The cation magnesium affects PTH release like calcium, although with reduced efficacy. (The CaSR is also a magnesium sensor.) High concentrations of aluminium also suppress PTH release. Hyperphosphatemia is associated with increased levels of PTH, an effect that is most often indirect and a result of the hypocalcaemia and/or the decreased $1,25(OH)_2D$ production that accompanies the rise in serum phosphate. However, the anion can exert a more direct effect on PTH synthesis with hyperphosphatemia stabilizing and hypophosphatemia destabilizing PTH mRNA levels. Glucocorticoids (in some studies) increase PTH secretion. Agents such as biogenic amines, which increase parathyroid gland cAMP levels, induce PTH secretion, and those that lower cAMP levels within the parathyroid gland decrease PTH secretion. Lithium, often used to treat bipolar disease, has been associated with hypercalcaemia presumably due to hyperparathyroidism, however, the precise mechanism of lithium action, in this respect, remains to be determined.

PTH Measurement

Circulating PTH is heterogeneous. The major circulating bioactive moiety is similar or identical to intact PTH(1-84). This is metabolized by the liver, which releases midregion and COOH-terminal fragments into the circulation for subsequent clearance by the kidney. These biologically inert moieties generated by metabolism and secretion from the parathyroid gland are cleared more slowly than intact PTH. Circulating bioactive PTH is best measured by sensitive double antibody immunoassays that simultaneously recognize NH_2 and COOH epitopes on the PTH molecule, and detect intact PTH(1-84). This is the method of choice for the accurate diagnosis of patients with hypercalcaemia, especially in distinguishing patients with primary hyperparathyroidism from those with hypercalcaemia of malignancy and in assessing hyperparathyroidism in chronic kidney disease.

Actions of Parathyroid Hormone

€of extracellular fluid calcium [24] (Figure 4.1.4). The hormone exerts important effects on bone and kidney and indirectly influences the gastrointestinal tract. In response to a fall in the extracellular fluid ionized calcium concentration, PTH is released from the parathyroid cell and acts directly on the kidney to enhance renal calcium reabsorption and promote the conversion of 25-hydroxyvitamin D to $1,25(OH)_2D$. In the cortical thick ascending limb of the loop of Henle (CTAL) PTH binds to the PTH receptor (PTHR1) [29, 30], and enhances Ca reabsorption., by increasing the activity of the Na/K/2Cl cotransporter that drives NaCl reabsorption and stimulates paracellular Ca and Mg reabsorption [31]. In the distal convoluted tubule (DCT) PTH, after binding to its receptor, increases luminal Ca transfer into the renal tubule cell by augmenting the epithelial apical Ca channel of the transient receptor potential vanilloid family, TRPV5, followed by translocation of Ca across the cell from apical to basolateral surface via proteins such as calbindin-D28K, and then active extrusion of Ca from the cell into the blood via a Na+/Ca exchanger, NCX1, whose activity is stimulated by a cyclic AMP-mediated mechanism [32]. PTH can also stimulate the 1α(OH)ase, in the proximal tubule leading to increased synthesis of $1,25(OH)_2D$ from 25OHD [33]. The $1,25(OH)_2D$ increases gastrointestinal absorption of calcium and, with PTH, induces skeletal resorption, causing the restoration of extracellular fluid calcium and the neutralization of the signal initiating PTH release. In bone, PTHR1 is localized on cells of the osteoblast phenotype which are of mesenchymal origin [34] but not on osteoclasts which are of hematogenous origin. PTH then enhances release of the cytokine,

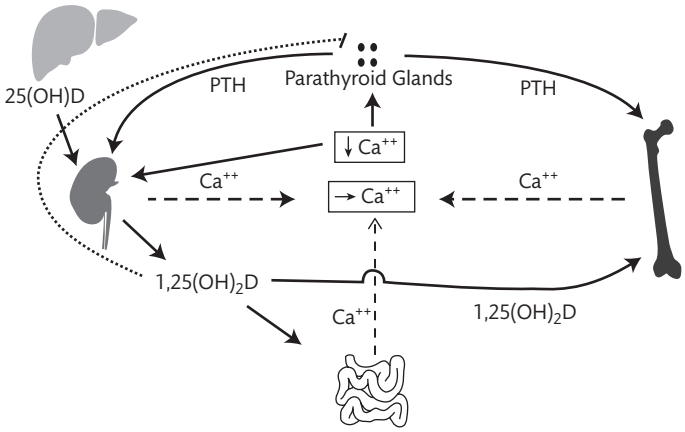


Figure 4.1.4 Parathyroid hormone (PTH) and vitamin D control calcium (as shown) and phosphate homeostasis. A fall in extracellular calcium concentration triggers PTH secretion. PTH directly acts on the kidney to promote renal calcium reabsorption and conversion of 25-hydroxyvitamin D [25(OH)D] to 1,25-dihydroxyvitamin D [1,25$(OH)_2$D]. 1,25$(OH)_2$D increases intestinal absorption of calcium (and phosphate) and, with PTH, mobilizes calcium (and phosphate) from bone. Thus, extracellular fluid (ECF) calcium is restored to normal neutralizing the signal initiating PTH release. PTH inhibits renal phosphate reabsorption promoting phosphaturia. Fibroblast growth factor 23 (FGF23),not shown, may be stimulated by 1,25$(OH)_2$D and also enhance phosphaturia, however FGF23 may then inhibit further 1,25$(OH)_2$D production.

receptor activator of NFκB ligand (RANKL) which binds to its receptor, RANK, on osteoclast precursors and osteoclasts, increasing the formation of mature osteoclasts from precursors and increasing the resorptive activity of existing osteoclasts especially in cortical bone. PTH may also reduce the osteoblastic protein, osteoprotegerin, which normally binds to RANKL, forms an inactive complex, and prevents it from binding to RANK, thus reducing osteoclastic activity [35, 36]. It has also been suggested that PTH can acutely release mineral at the bone surface in an osteoclast-independent manner by modifying its solubility [37]. The opposite series of homeostatic events occur in response to a rise in extracellular fluid calcium levels and a fall in PTH.

Although this scheme outlines the overall events that occur after a fall in calcium, aspects of the response may vary. Certain actions of PTH, such as renal calcium retention may predominate at relatively low circulating concentrations of PTH. Furthermore, PTH appears to be essential as a bone anabolic factor in the fetus [38] and neonate [39], particularly in trabecular bone but may be predominantly resorptive, especially in the cortical compartment of bone, in older animals [40] when the source of external calcium changes. PTH and PTHrP regulate osseous cellular differentiation, proliferation, and development, and via their actions on osteoblastic cells can function as anabolic skeletal agents when administered intermittently rather than continuously *in vivo*. Thus, intermittent doses of PTH(1-34)—and PTHrP(1-34) and related analogues—promote bone formation [41] directly by stimulating osteoblastic activity and indirectly, both by stimulating IGF-I production and suppressing sclerostin, thereby increasing Wnt signalling. Daily injections of human PTH(1-34), teriparatide, or a human PTHrP analogue, abaloparatide, increase hip and spine BMD, and prevent vertebral and non-vertebral fractures in osteoporosis, and each has been approved clinically for use as a bone anabolic agent to treat osteoporosis.

Besides regulating calcium homeostasis, PTH elicits various other responses. Among these responses are perturbations of other ions, the most marked of which are those involving phosphate. As a consequence of PTH-enhanced 1,25$(OH)_2$D production, the gastrointestinal absorption of phosphate is facilitated to some extent, and with PTH-induced skeletal lysis, phosphate and calcium are released. These effects increase the extracellular fluid phosphate levels. However the predominant effect of PTH on phosphate homeostasis is to inhibit renal phosphate reabsorption in the proximal renal tubule by enhancing rapid internalization and lysosomal degradation of the type II sodium phosphate cotransporters, NPT2a (SLC34A1) and Npt2c (SLC34A3) thus decreasing luminal transport of phosphate and causing phosphaturia [42, 43]. Consequently, a net decrease in extracellular fluid phosphate concentration occurs which is adjunctive to the role of PTH in raising calcium levels.

Parathyroid Hormone Receptors

Like other peptide hormones, PTH interacts through a receptor on the plasma membrane of target cells. This same receptor binds PTHrP [44, 45]. The type 1 PTH/PTHrP receptor (PTHR1) is a seven-transmembrane G-protein-linked receptor that has the 'signature' GPCR topology, a seven-membrane-spanning, 'serpentine' domain, as well as an extracellular ligand-binding domain and an intracellular COOH-terminal domain [46]. It is a member of group B of the GPCR superfamily. The receptor can couple to the stimulatory G protein, G_s, leading to increased adenylate cyclase activity, the generation of cAMP, and activation of the protein kinase A (PKA) pathway, and can couple to G_q, leading to an increase in the protein kinase C (PKC) pathway and to an increase in IP_3, diacylglycerol, and intracellular Ca^{2+} [46, 47]. As with other GPCRs, PTHR1 undergoes cyclical receptor activation, desensitization, and internalization [48]. After ligand binding and endocytosis, the PTHR1 is either recycled to the cell membrane or targeted for degradation. High circulating levels of PTH in hyperparathyroid states have been associated with hormonal desensitization in target tissues. Arrestins contribute to the desensitization of both G_s and G_q mediated PTHR1 signalling. PTHR1 activation and internalization can be selectively dissociated [49]. PTHR1 signalling can be modified by scaffolding proteins such as the Na^+/H^+ exchanger regulatory factor (NHERF) 1 and 2 through PDZ1 and PDZ2 domains [50, 51]. PTHR1 signalling via the cAMP pathway, leading to PKA activation, results in phosphorylation of the cyclic AMP response element binding protein (CREB). CREB binds to the CRE in the promoter region of many genes and transcriptionally modulates their expression. The difference in the extent of anabolic versus catabolic activity of PTH and PTHrP analogues may be explained, at least in part, by the way in which these two peptides interact at two PTHR1 conformations, R0 and RG. Higher affinity for R0 is associated with increased internalization, more cyclic AMP generation and longer-lived catabolic effects, whereas greater binding to RG is associated with shorter-duration anabolic responses.

The PTHR1 is highly expressed in kidney and bone, the primary target tissues of PTH, but is also expressed in a wide variety of embryonic and adult tissues, including cartilage, liver, brain, smooth

muscle, spleen, testis, and skin. In most of these tissues, the receptor appears to mediate the autocrine/paracrine actions of locally produced and secreted PTHrP. Nevertheless, PTHrP may also exert some of its bioactivity through domains of the molecule that do not interact with PTHR1 [52].

The human PTH/PTHrP receptor gene *(PTHR1)* localizes to chromosome 3p21.1-22 [53]. A second related receptor which is the product of a distinct gene (*PTHR2* on chromosome 2q33), and which binds PTH, TIP39, but not PTHrP, has been identified [54]. It is expressed in brain, pancreas, testis, and placenta, and its endogenous ligand is TIP39.

Direct evidence that the PTHR1 mediates the calcium homeostatic actions of PTH and the skeletal growth plate actions of PTHrP in humans has come from the study of rare genetic disorders. Jansen's metaphyseal chondrodysplasia (JMC) is inherited in an autosomal dominant fashion although most reported cases are sporadic [55]. The disorder comprises short-limbed dwarfism secondary to severe growth plate abnormalities, asymptomatic hypercalcaemia, and hypophosphatemia. There is increased bone resorption similar to that in primary hyperparathyroidism and urinary cAMP levels are elevated, but circulating PTH and PTHrP levels are low or undetectable. Although PTHR1 is found widely in fetal and adult tissues, it is most abundant in three major organs, the kidney, bone, and metaphyseal growth plate. The changes in mineral ion homeostasis and the growth plate in JMC are caused by heterozygous gain-of-function mutations (Figure 4.1.5) in the *PTHR1* giving rise to constitutively active receptors.

Inactivating or loss-of-function mutations in the *PTHR1* have been implicated in the molecular pathogenesis of Blomstrand

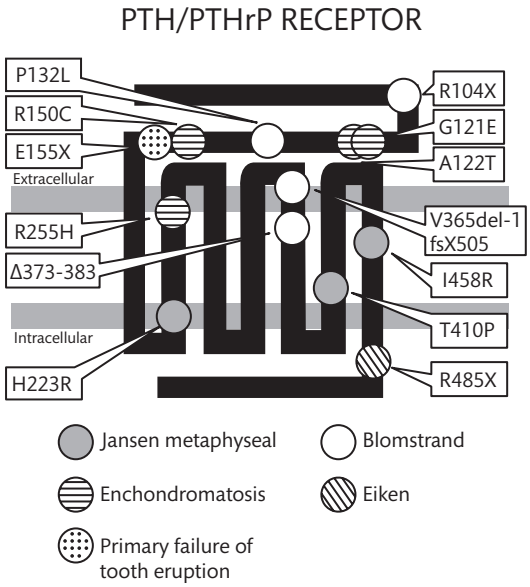


Figure 4.1.5 Schematic representation of the type 1 human PTH/PTHrP receptor. The locations of the H223R, T410P, and I458R activating mutations identified in patients with Jansen's metaphyseal chondrodysplasia, the R104X, P132L, V365del-1fsX505, and D373-383 inactivating mutations found in patients with Blomstrand chondrodysplasia, the R485X Eiken syndrome mutation, the G121E, A122T, R150C and R255H endochondromatosis mutations, and the E155X primary failure of tooth eruption (PFE) mutation, are indicated. Splice-site mutations that would result in predicted mutant C351fsX485 and E182fsX203 proteins have been identified in additional PFE cases.

lethal chondrodysplasia (BLC) [55]. This rare disease is characterized by advanced endochondral bone maturation, short-limbed dwarfism, abnormal breast and tooth morphogenesis, and fetal death, thus mimicking the phenotype of *Pthr1*-less mice [56]. The majority of BLC cases were born to phenotypically normal, consanguineous parents, suggesting an autosomal recessive mode of inheritance. Mutant PTHR1s (Figure 4.1.5) identified in BLC fetuses fail to bind ligand or stimulate cAMP or inositol phosphate production. A milder form of recessively inherited skeletal dysplasia, known as Eiken syndrome, has been linked to mutations of *PTHR1*, suggesting a wider range of skeletal phenotypes to this gene. Dominantly acting heterozygous *PTHR1* mutations have been identified in familial, non-syndromic primary failure of tooth eruption [57]. Heterozygous *PTHR1* mutations have been identified in endochondromas of patients with endochondromatosis (Ollier's disease), a familial disorder with evidence of autosomal dominance characterized by multiple benign cartilage tumours, and a predisposition to malignant chondrosarcomas [58]. As many patients with Ollier's disease do not apparently have *PTHR1* mutations the condition may be genetically heterogeneous.

Heterozygous inactivating mutations in the *GNAS1* gene encoding Gαs cause an ~50% reduction in amount/activity of the protein leading to resistance to PTH and other hormones in the disorder, pseudohypoparathyroidism (PHP) type 1a [59–61]. In contrast, patients with PHP type 1b have end-organ resistance to PTH without the typical physical stigmata—termed Albright's hereditary osteodystrophy—of PHP type 1a. Linkage to chromosome 20q13.3, which includes the *GNAS1* locus was established in kindreds with PHP type 1b [62]. In addition, the genetic defect is imprinted paternally and is inherited in the same fashion as the PTH resistance in kindreds with PHP type 1a, and in a mouse model heterozygous for ablation of the *GNAS* gene [63]. In PHP type 1b patients, mutations some distance upstream of the *GNAS1* coding regions affect the normal differential methylation of maternal and paternal alleles leading to silencing of the *GNAS* gene specifically in the renal proximal tubules [64].

PTH controls renal phosphate reabsorption. Mutations in the genes encoding the two renal sodium phosphate cotransporters, NPT2a and NPT2c, have been identified in a few patients with hyperphosphaturia. The sodium-hydrogen exchanger regulatory factor 1 (NHERF1) interacts with the PTHR1 and NPT2a. Study of hyperphosphaturic patients referred initially for nephrolithiasis or osteopaenia identified a few cases having NHERF1 mutations that could contribute to the renal phosphate loss [65].

Summary

PTH is responsible for the minute-to-minute maintenance of calcium homeostasis. PTH secretion is controlled via the parathyroid CaSR, and inactivating or activating mutations in this receptor lead to inherited hypercalcaemic and hypocalcaemic disorders, respectively. Both PTH (and the related gene family member, PTHrP) act through the PTHR1 that is widely expressed and signals through multiple second messenger pathways. Inactivating mutations in the PTHR1 cause Blomstrand's lethal chondrodysplasia, whereas activating mutations are found in JMC.

REFERENCES

1. Mannstadt M, Bilezikian JP, Thakker RV, *et al.* Hypoparathyroidism. *Nat Rev Dis Primers*, 2017; **3**: 17055.
2. Meakins JL, Milne CA, Hollomby DJ, Goltzman D. Total parathyroidectomy: parathyroid hormone levels and supernumerary glands in hemodialysis patients. *Clin Invest Med*, 1984; **7**: 21–5.
3. Hendy GN, Arnold A. Molecular basis of PTH overexpression. In: Bilezikian JP, Martin TJ, Raisz LG, eds. *Principles of Bone Biology*, 3rd edn. San Diego, CA: Academic Press, 2008: 1311–26.
4. Gunther T, Chen ZE, Kim J, *et al.* Genetic ablation of parathyroid glands reveals another source of parathyroid hormone. *Nature*, 2000; **40**: 199–203.
5. Canaff L, Zhou X, Mosesova I, Cole DEC, Hendy GN. Glial cells missing-2 transactivates the calcium-sensing receptor gene: effect of a dominant-negative GCM2 mutant associated with autosomal dominant hypoparathyroidism. *Hum Mutat*, 2009; **30**: 85–92.
6. Hendy GN, Cole DEC, Bastepe M. Hypoparathyroidism and Pseudohypoparathyroidism. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, eds. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. Last Update: 19 February 2017.
7. Ali A, Christie PT, Grigorieva IV, *et al.* Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into the mechanisms of DNA binding by the GATA3 transcription factor. *Hum Molec Genet*, 2007; **16**: 265–75.
8. Parvari R, Diaz GA, Hershkovitz E. Parathyroid development and the role of tubulin chaperone E. *Horm Res*, 2007; **67**: 12–21.
9. Keutmann HT, Sauer MM, Hendy GN, O'Riordan JLH, Potts JT Jr. Complete amino acid sequence of human parathyroid hormone. *Biochemistry*, 1978; **17**: 243–4.
10. Hendy GN, Kronenberg HM, Potts JT Jr, Rich A. Nucleotide sequence of cloned DNAs encoding human preproparathyroid hormone. *Proc Natl Acad Sci USA*, 1981; **78**: 7365–9.
11. Karaplis AC, Luz A, Glowacki J, *et al.* Lethal skeletal dysplasia from targeted disruption of the parathyroid hormone-related peptide gene. *Gene Develop*, 1994; **8**: 277–89.
12. Amizuka N, Warshawsky H, Henderson JE, Goltzman D, Karaplis AC. Parathyroid hormone-related peptide-depleted mice show abnormal epiphyseal cartilage development and altered endochondral bone formation. *J Cell Biol*, 1994; **126**: 1611–23.
13. Fegley DB, Holmes A, Riordan T, *et al.* Increased fear- and stress-related anxiety-like behavior in mice lacking tuberoinfundibular peptide of 39 residues. *Gene Brain Behavior*, 2008; **7**: 933–42.
14. VanHouten JN, Yu N, Rimm D, *et al.* Hypercalcemia of malignancy due to ectopic transactivation of the parathyroid hormone gene. *J Clin Endocrinol Metabol*, 2006; **91**: 580–3.
15. Chandrasekharappa SC, Guru SC, Manickam P, *et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science*, 1997; **276**: 404–7.
16. Lemmens I, Van der Ven WJ, Kas K, *et al.* Identification of the multiple endocrine neoplasia type 1 (MEN1) gene. European Consortium on MEN1. *Hum Mol Genet*, 1997; **6**: 1177–83.
17. Hendy GN, Kaji H, Canaff L. Cellular functions of menin. *Adv Exp Med Biol*, 2009; **668**: 37–50.
18. Bjorklund P, Akerstrom G, Westin G. An LRP5 receptor with internal deletion in hyperparathyroid tumors with implications for deregulated WNT/β-catenin signalling. *PloS Med*, 2007; **4**: e328.
19. Carpten JD, Robbins CM, Villablanca A, *et al. HRPT2*, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet*, 2002; **32**: 676–80.
20. Cinque L, Sparaneo A, Penta L, *et al.* Autosomal dominant PTH signal sequence mutation in a family with familial isolated hypoparathyroidism. *J Clin Endocrinol Metab*, 2017; **102**: 3961–9.
21. Rabbani SA, Kaiser SM, Henderson JE, *et al.* Synthesis and characterization of extended and deleted recombinant analogues of parathyroid hormone-(1-84): correlation of peptide structure with function. *Biochemistry*, 1990; **29**: 10080–9.
22. Hendy GN, Bennett HPJ, Gibbs BF, *et al.* Proparathyroid hormone (ProPTH) is preferentially cleaved to parathyroid hormone (PTH) by the prohormone convertase furin: a mass spectrometric analysis. *J Biol Chem*, 1995; **270**: 9517–25.
23. Kremer R, Bolivar I, Goltzman D, Hendy GN. Influence of calcium and 1,25-dihdroxycholecalciferol on proliferation and proto-oncogene expression in primary cultures of bovine parathyroid cells. *Endocrinology*, 1989; **125**: 935–41.
24. Brown EM. Role of the calcium-sensing receptor in extracellular calcium homeostasis. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 333–43.
25. Hendy GN. *Calcium-Sensing Receptor. ASBMR Primer.* Bilezikian J, ed. In press: 2019.
26. Cole DEC, Peltekova VD, Rubin LA, *et al.* A986S polymorphism of the calcium-sensing receptor and circulating calcium concentrations. *Lancet*, 1999; **353**: 112–15.
27. Nemeth EF, Goodman WG. Calcimimetic and calcilytic drugs: feats, flops, and futures. *Calcif Tissue Int*, 2016; **98**: 341–58.
28. Hendy GN. The calcium-sensing receptor and vitamin D. In: Feldman DJ, Pike W, Bouillon R, eds. *Vitamin D, Volume 1, Biochemistry, Physiology and Diagnostics*, 4th edn. Elsevier Inc, 2018: 477–95.
29. Jüppner H, Abou-Samra AB, Freeman M, *et al.* A G protein-linked receptor for parathyroid hormone and parathyroid-related peptide. *Science*, 1991; **254**: 1024–6.
30. Abou-Samra AB, Jüppner H, Force T, *et al.* Expression cloning of a common receptor for parathyroid hormone and parathyroid hormone-related peptide from rat osteoblast-like cells: a single receptor stimulates intracellular accumulation of both cAMP and inositol trisphosphates and increases intracellular free calcium. *Proc Natl Acad Sci USA*, 1992; **89**: 2732–6.
31. De Rouffignac C, Quamme G. Renal magnesium handling and its hormonal control. *Physiol Rev*, 1994; **74**: 305–22.
32. van Abel M, Hoenderop JG, van der Kemp AW, *et al.* Coordinated control of renal Ca(2+) transport proteins by parathyroid hormone. *Kidney Int*, 2005; **68**: 1708–21.
33. Brenza HL, Kimmel-Jehan C, Jehan F, *et al.* Parathyroid hormone activation of the 25-hydroxyvitamin D3-1alpha-hydroxylase gene promoter. *Proc Natl Acad Sci USA*, 1998; **95**: 1387–91.
34. Rouleau MF, Mitchell J, Goltzman D. *In vivo* distribution of parathyroid hormone receptors in bone: evidence that a predominant osseous target cell is not the mature osteoblast. *Endocrinology*, 1988; **123**: 187–91.
35. Huang JC, Sakata T, Pfleger LL *et al.* PTH differentially regulates expression of RANKL and OPG. *J Bone Miner Res*, 2004; **19**: 235–44.
36. Boyce BF, Xing L. Functions of RANKL/RANK/OPG in bone modeling and remodeling. *Arch Biochem Biophys*, 2008; **473**: 139–46.
37. Talmage DW, Talmage RV. Calcium homeostasis: how bone solubility relates to all aspects of bone physiology. *J Musculoskel Neuron Interact*, 2007; **7**: 108–12.

38. Miao D, He B, Karaplis AC, Goltzman D. Parathyroid hormone is essential for normal fetal bone formation. *J Clin Invest*, 2002; **109**: 1173–82.
39. Miao D, He B, Lanske B, *et al*. Skeletal abnormalities in Pth-null mice are influenced by dietary calcium. *Endocrinology*, 2004; **145**: 2046–53.
40. Xue Y, Karaplis AC, Hendy GN, Goltzman D, Miao D. Genetic models show that parathyroid hormone and 1,25-dihydroxyvitamin D3 play distinct and synergistic roles in postnatal mineral ion homeostasis and skeletal development. *Hum Mol Genet*, 2005; **14**: 1515–28.
41. Silva BC, Bilezikian JP. Parathyroid hormone: anabolic and catabolic actions on the skeleton. *Curr Opin Pharmacol*, 2015; **22**: 41–50.
42. Bacic D, Lehir M, Biber J, *et al*. The renal Na+/phosphate cotransporter NaPi-Iia is internalized via the receptor-mediated endocytic route in response to parathyroid hormone. *Kidney Int*, 2006; **69**: 495–503.
43. Segawa H, Yamanaka S, Onitsuka A, *et al*. Parathyroid hormone-dependent endocytosis of renal type Iic Na-Pi cotransporter. *Am J Physiol Renal Physiol*, 2007; **292**: F395–F403.
44. Goltzman D. Interactions of PTH and PTHrP with the PTH/PTHrP receptor and with downstream signaling pathways: exceptions that provide the rules. *J Bone Miner Res*, 1999; **14**: 173–7.
45. Vilardaga JP, Romero G, Friedman PA, Gardella TJ. Molecular basis of parathyroid hormone receptor signaling and trafficking: a family B GPCR paradigm. *Cell Mol Life Sci*, 2011; **68**: 1–13.
46. Cheloha RW, Gellman SH, Vilardaga JP, Gardella TJ. PTH receptor-1 signalling-mechanistic insights and therapeutic prospects. *Nat Rev Endocrinol*, 2015; **11**: 712–24.
47. Datta NS, Abou-Samra AB. PTH and PTHrP signaling in osteoblasts. *Cell Signal*, 2009; **21**: 1245–54.
48. Weinman EJ, Hall RA, Friedman PA, Liu-Chen LY, Shenolikar S. The association of NHERF adaptor proteins with G protein-coupled receptors and receptor tyrosine kinases. *Ann Rev Physiol*, 2006; **68**: 491–505.
49. Sneddon WB, Syme CA, Bisello A, *et al*. Activation independent parathyroid hormone receptor internalization is regulated by NHERF1 (EBP50). *J Biol Chem*, 2003; **278**: 43787–96.
50. Mahon MJ, Donowitz M, Yun CC, Segre GV. Na+/H+ exchanger regulatory factor 2 directs parathyroid hormone 1 receptor signalling. *Nature*, 2002; **417**: 858–61.
51. Wang B, Ardura JA, Romero G, *et al*. Na/H exchanger regulatory factors control parathyroid hormone receptor signaling by facilitating differential activation of Gα protein subunits. *J Biol Chem*, 2010; **285**: 26976–86.
52. Miao D, Su H, He B, *et al*. Severe growth retardation and early lethality in mice lacking the nuclear localization sequence and C-terminus of PTH-related protein. *Proc Natl Acad Sci USA*, 2008; **105**: 20309–14.
53. Pausova Z, Bourdon J, Clayton D, *et al*. Cloning of a parathyroid hormone/parathyroid-related peptide receptor (PTHR) cDNA from a rat osteosarcoma (UMR 106) cell line: chromosomal assignment of the gene in the human, mouse, and rat genomes. *Genomics*, 1994; **20**: 20–6.
54. Usdin TB, Gruber C, Bonner TI. Identification and functional expression of a receptor selectively recognizing parathyroid hormone, the PTH2 receptor. *J Biol Chem*, 1995; **270**: 15455–8.
55. Hendy GN, Bastepe M, Cole DEC. Parathyroid disorders. In: Rimoin DL, Connor JM, Pyeritz RE, Korf BE. *Emery and Rimoin's Principles and Practice of Medical Genetics*, 6th edn, Vol. 2. Churchill Livingstone, Edinburgh, 2013: Chapter 89.
56. Lanske B, Karaplis AC, Lee K, *et al*. PTH/PTHrP receptor in early development and Indian Hedgehog-regulated bone growth. *Science*, 1996; **273**: 663–6.
57. Decker E, Stellzig-Eisenhauer A, Fiebig BS, *et al*. PTHR1 loss-of-function mutations in familial, nonsyndromic primary failure of tooth eruption. *Amer J Hum Genet*, 2008; **83**: 781–6.
58. Couvineau A, Wouters V, Bertrand G, *et al*. PTHR1 mutations associated with Ollier diseases result in receptor loss of function. *Hum Molec Genet*, 2008; **17**: 2766–75.
59. Bastepe M. Genetics and epigenetics of parathyroid hormone resistance. *Endocr Dev*, 2013; **24**: 11–24.
60. Puzhko S, Goodyer CG, Kerachian MA, *et al*. Parathyroid signalling via Gαs is selectively inhibited by an NH(2)-terminally truncated Gαs: implications for pseudohypoparathyroidism. *J Bone Miner Res*, 2011; **26**: 102473–85.
61. Mantovani G, Bastepe M, Monk D, *et al*. Diagnosis and management of pseudohypoparathyroidism and related disorders: First International Consensus Statement. *Nat Rev Endocrinol*, 2018; **14**: 476–500.
62. Jüppner H, Schipani E, Bastepe M, *et al*. The gene responsible for pseudohypoparathyroidism type 1b is paternally imprinted and maps in four unrelated kindreds to chromosome 20q13.3. *Proc Natl Acad Sci USA*, 1998; **95**: 11798–803.
63. Yu S, Yu D, Lee E, *et al*. Variable and tissue-specific hormone resistance in heterotrimeric Gs protein α-subunit (Gsα) knockout mice is due to tissue-specific imprinting of the Gsα gene. *Proc Natl Acad Sci USA*, 1998; **95**: 8715–20.
64. Bastepe M, Frolich LF, Hendy GN, *et al*. Autosomal pseudohypoparathyroidism type 1b is associated with a heterozygous microdeletion that likely disrupts a putative imprinting control element of GNAS. *J Clin Invest*, 2003; **112**: 1255–63.
65. Karim Z, Gérard B, Bakouh N, *et al*. NHERF1 mutations and responsiveness of renal parathyroid hormone. *New Engl J Med*, 2008; **359**: 1128–11.

4.2

Hypercalcaemia

Claudio Marcocci, Federica Saponaro, and Filomena Cetani

Introduction

Hypercalcaemia is a common finding in general medical practice and in emergency departments. Hypercalcaemia is incidentally discovered in most patients evaluated for unrelated disorders, particularly after the widespread availability of multichannel biochemistry autoanalyser which also include the measurement of serum calcium. The principal challenge in the evaluation of hypercalcaemia is distinguishing parathyroid hormone (PTH)-related hypercalcaemia from other causes, that may need different diagnostic approach and treatment.

The total calcium content of the human body is about 1–2 kg and the large majority (98–99%) is stored in the skeleton, with two main functions: structural support for the human body and calcium deposit which can be exchanged with other compartments. The remaining 1–2% is present in the extracellular and intracellular space: the maintenance of calcium in a very narrow range in both compartments is very important for the adequate functioning of cellular calcium-dependent effects. Approximately 50% of the total amount of calcium present in blood is bound to proteins (mainly albumin) or complexed to some anions, namely phosphate and citrate [1]. The remaining part is in the ionized state (ionized calcium, Ca^{2+}), that represents the only biologically active form and is involved in the regulation of several functions, namely neuromuscular activity, cells membrane structure, bone homeostasis, coagulation, and endocrine functions.

Intestine, kidney, and bone are involved in the maintenance of calcium homeostasis. When dietary calcium intake is normal (1000 mg per day), approximately 200 mg of calcium is absorbed by the gastrointestinal apparatus and constitutes the only source to maintain calcium deposits. Duodenum, jejunum, and ileum display a high calcium absorptive capacity, depending from the length of each segment and the characteristics of the food bolus. Two mechanisms of calcium absorption are involved, depending from the calcium concentration in the intestinal lumen. A passive transport through a paracellular space between intestinal cells is operating when calcium level in the intestinal lumen is high. This pathway is only indirectly controlled by $1,25(OH)_2$vitamin D ($1,25(OH)_2D$), that activates a transduction signalling mediated by protein kinase C, leading to more permeable tight junctions between cells [2]. On the other hand, when calcium levels in the intestinal lumen are low, an active, $1,25(OH)_2D$ dependent, transport process mediates the entry of calcium into the enterocytes through a calcium channel [3].

When renal function is normal, 10 grams of calcium are filtrated daily, the large majority (98–99%) is reabsorbed by renal tubules, and about 100–200 mg are excreted. Reabsorption of filtered calcium mostly (90–95%) occurs in the proximal tubule and loop of Henle by a Na-Ca exchanger, the remaining being absorbed in the distal nephron by a finely-regulated PTH-dependent transcellular mechanism [2].

Calciotropic hormones, and particularly PTH and $1,25(OH)_2$vitamin D ($1,25(OH)_2D$), regulate serum calcium levels, acting on bone, kidney, and intestine

Increased bone resorption and gastrointestinal absorption and/or decreased renal excretion of calcium may lead to hypercalcaemia.

Hypercalcaemia is defined by a value of serum calcium levels two standard deviation above the normal mean. The total serum calcium concentration is commonly used in general practice, even though the measure of Ca^{2+} is more accurate, since the protein content may affect the total serum calcium levels. However, in many laboratories the measurement of $Ca^{2+}$is not available, and the following formula can be used to calculate the albumin-adjusted serum calcium (Alb-Ca):

$$\text{Alb-Ca (mg/dl)} = \text{measured serum calcium (mg/dl)} + [(4 - \text{plasma albumin g/dl}) \times 0.8].$$

The Alb-Ca can be used as an acceptable measure in most clinical circumstances.

Epidemiology

Epidemiological data regarding calcium disorders are challenging, because they are dependent from the specific settings where patients are evaluated, geographical differences, and clinical features of the underlying diseases.

The prevalence of persistent hypercalcaemia in the general population is about 1%, but it rises up to 3% in hospitalized subjects [4]. In a recent retrospective study on 12 334 hospitalized subjects, hypercalcaemia was found in 585 patients (4.74%) and its rate slightly, but not significantly, increased from 3.5% in the year 2011 to 6.9% in the year 2014 [5], with no gender differences.

Primary hyperparathyroidism (PHPT) is one of the most common causes of hypercalcaemia in free-living individuals. High variability in the incidence of PHPT has been reported in United Kingdom and United States studies, ranging from ~0.4 to 82 cases per 100 000, according to the different population analysed [6, 7]. Hypercalcaemia of malignancy (HCM) is the second cause of hypercalcaemia and the most common in hospitalized patients. The estimated prevalence based on retrospective studies ranges between 3% and 30%, depending upon the kind of primary tumour and age. In the paediatric population HCM occurs in 0.4–1.3% of patients with cancer mainly in those with cancer of the haematopoietic system. In 2016 Gastanaga *et al.* evaluated a large series (about 12 000) of cancer patients in the United States collected between 2009 and 2013 and found a prevalence of HCM of 2.8%, with a higher rate in patients with multiple myeloma (7.5%) and lower in those with prostate cancer (1.4%) and its prevalence was found to decrease over the study period [8]. Conversely, a lower (1%) prevalence of HCM was reported in another large series (37 442 over the 2003–2012 period) of cancer patients in the United Kingdom, but its rate increased during the follow-up, because serum calcium measurement was most commonly measured in the more recent years of the study [8].

Clinical Signs and Symptoms

Hypercalcaemia is commonly encountered in emergency department. A typical case could be 60-year-old women presenting with lethargy, confusion, profound weakness, progressive anorexia, and constipation since a few weeks. These clinical manifestations are usually present when serum calcium concentration is above 14 mg/dl (severe hypercalcaemia). These classical symptoms include neuropsychiatric symptoms (anxiety and depression), musculoskeletal involvement and bone pain, renal (polyuria, dehydration, and nephrolithiasis) and gastrointestinal manifestations (anorexia, nausea, constipation, and peptic ulcer). When serum calcium is above 13 mg/dl and hyperphosphatemia is also present, calcifications in the cardiovascular system, kidney, cornea, and joints usually occur.

Symptoms generally, but not always, correlate with the degree of hypercalcaemia and its duration. When hypercalcaemia is rapidly progressing reaching severe levels, nausea, vomiting, ventricular fibrillation, and coma may occur. Severe hypercalcaemia represents an emergency and treatment is urgently required.

The clinical manifestations of hypercalcaemia are reported in Table 4.2.1.

Pathophysiology

As mentioned before, calcium homeostasis is maintained by the equilibrium among renal excretion, bone resorption, and intestinal absorption of calcium. Parfitt *et al.* in 1979 proposed three different pathophysiologic mechanisms of hypercalcaemia: (i) the 'equilibrium' hypercalcaemia, which typically occurs in patients with mild-to-moderate PHPT, where the average serum calcium is elevated but constantly maintained at the same level, with few fluctuations. This is related to a perturbance of the so-called calcium homeostatic system on bone surface, where the increased calcium resorption is balanced by an increased renal tubular excretion; (ii) the 'disequilibrium' hypercalcaemia, which is characterized by a rapid increase in serum calcium, increased bone resorption, and negative calcium balance at the bone surface. The general mechanism involves osteoclast progenitor's differentiation, mature osteoclasts recruitment, and stimulation of bone resorption activity. 'Disequilibrium' hypercalcaemia can develop when 'equilibrium' hypercalcaemia is broken by precipitating conditions, or it can arise from the beginning in all situations in which a disturbance of the bone remodelling system occurs [9]; (iii) hyperabsorption hypercalcaemia, which is associated with increased intestinal calcium absorption and positive calcium

Table 4.2.1 Clinical manifestations of hypercalcaemia

Manifestations	Mild hypercalcaemia <12 mg/dl (3.0 mmol/L)	Moderate hypercalcaemia 12–14 mg/dl (3.0–3.5 mmol/L)	Severe hypercalcaemia >14 mg/dl (3.5 mmol/L)
Neuropsychiatric	Lack of concentration and memory, anxiety, depression, irritability	Confusion, cognitive and personality changes	Confusion, lethargy, hallucinations, psychosis, coma
Muscle-skeletal	Asymptomatic	Muscle weakness, osteoporosis	Muscle weakness, bone pain, arthritis, osteitis fibrosa-cystica
Renal	Polyuria and polydipsia	Dehydration, polyuria, polydipsia, nocturia	Nephrolithiasis and colic, renal failure
Gastrointestinal	Nausea, constipation, weight loss	Anorexia, nausea, vomiting	Abdominal pain, pancreatitis, peptic ulcer
Cardiovascular	Hypertension	ECG: shortened QT interval, prolonged PR interval and widened QRS complex, vascular calcification	Arrhythmias, ventricular tachycardia, and cardiovascular arrest, syncope
Others	Asymptomatic	Itching, keratitis, conjunctivitis	Cornea calcification, haematological disorders (anaemia)

balance, as in vitamin D intoxication [1]. All types of hypercalcaemia are associated with increased calcium absorption and hypercalciuria is the first biochemical abnormalities in patients developing severe hypercalcaemia, unless renal function is markedly deteriorate, followed by an increase of serum calcium concentration [10].

Severe hypercalcaemia develops when there is some limitation in the capacity of the kidney to excrete the increased calcium overload, most commonly due to a combination of decreased renal function and increased renal calcium reabsorption

Specific pathophysiology issues will be discussed in the dedicated sections.

Diagnosis

The diagnosis of hypercalcaemia is primarily made on the basis of serum calcium and preferably serum Alb-Ca or, whenever possible, Ca^{2+} levels (Figure 4.2.1). When serum calcium is above the normal range of the laboratory, the first step should be to confirm hypercalcaemia by a second measurement.

It is important to know that serum calcium results are affected by important preanalytical, analytical, and postanalytical factors. Fasting blood samples should be collected, avoiding venous stasis. Haemoconcentration and haemodilution are the most common preanalytical causes of misleading results. Haemoconcentration is usually due to dehydration and venous stasis and may increase serum calcium concentration. Other factors can be posture, abnormal proteins levels, temperature, anions, and pH [11]. Moreover, some analytical variables should be considered: the most common method for total serum calcium measurement is the spectrometry analysis, using metallochromic indicators or an ion-specific electrode (ISE). The within- and between subjects' variation of total serum calcium ranges between 1.9 and 2.8%, and the desirable value for inaccuracy, imprecision, and total assay error are 0.8%, 1%, and 2.4%, respectively. It is very important to validate the assay using appropriate quality control (QC) methods and have adequate laboratory coefficients of variation [11]. Finally, post-analytical interpretation is also very important. The large majority of laboratories that use high accuracy methods for total calcium measurement have a reference interval of approximately 8.8–10.2 mg/dl*. The reference interval is generally established at the central 95% of the Gaussian distribution, so at the extreme

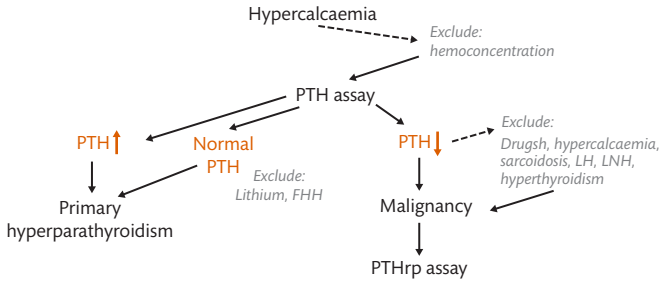


Figure 4.2.1 Differential diagnosis of hypercalcaemia. PTH, parathormone; FHH, familial hypocalciuric hypercalcaemia; LH, Hodgkin's lymphoma; NHL, non-Hodgkin's lymphoma.

* Use the formula serum calcium mg/dl: 2.5 = serum calcium mmol/L.

2.5% two sites of the curve values obtained in normal individuals will be interpreted as hyper- or hypocalcemia. In bordeline cases we recommend using the serum Alb-Ca obtained by the previously reported formula. However, values derived from this formula does not take into account the binding of calcium to proteins other than albumin and therefore do not accurately reflect the free serum calcium concentration [11].

Wherever available, serum Ca^{2+} should be measured, because it provides a more accurate assessment of the calcium status. It has been demonstrated that serum Ca^{2+} has a higher sensitivity for malignant hypercalcaemia and for the diagnosis of PHPT [12].

Moreover, in particular conditions, the measurement of serum Ca^{2+} is critical. In patients with kidney failure, comparison of results of total and Alb-Ca with Ca^{2+} demonstrated that the former measurements improperly classified as hypercalcaemic 20% of patients. The correct interpretation of calcium levels in those patients is important for the appropriate treatment [13]. In patients with critical diseases there is a redistribution of free, protein-bound, and complexed calcium and, therefore, the measurement of serum Ca^{2+} should be preferred [14]. The measurement of serum Ca^{2+} is also preferable in neonates and infants [15]. Exercise and hyperventilation may influence the concentration of serum Ca^{2+}: exercise may increase serum Ca^{2+} largely because exercise-induced acidosis decreases the binding of calcium to serum proteins, but reduction of plasma volume and influx form extracellular sources may also contribute; conversely hyperventilation may decrease Ca^{2+} because of hyperventilation-induced alkalosis increases the binding of calcium to serum proteins [11].

Once the diagnosis of hypercalcaemia is confirmed plasma PTH levels should be measured to distinguish PTH-dependent and PTH-independent hypercalcaemia. Second and third generation immunoradiometric PTH assays demonstrated a higher sensitivity (88–97%) in the diagnosis of PHPT compared to first generation assays and are the most commonly used [11, 16]. When PTH measurement is not easily available serum phosphate may be used to guide the differential diagnosis, being usually normal in PTH-independent hypercalcaemia and low or in the lower normal range in PTH-dependent hypercalcaemia. Serum creatinine and kidney function should also be evaluated.

A 24-hour urine collection for calcium and creatinine measurements should be obtained to calculate the calcium to creatinine clearance ratio and rule out familial hypocalciuric hypercalcaemia (FHH) (see next). Indeed, a calcium to creatinine ratio of less than 0.01 is suggestive of FHH. The measurement of serum 25(OH) vitamin D (25(OH)D) and, especially, $1,25(OH)_2D$ may help in the diagnosis of vitamin D-dependent forms of hypercalcaemia.

Aetiology

Hypercalcaemic disorders are highly variable and can be classified according to the underling diseases or conditions, as summarized in the Box 4.2.1.

Hypercalcaemia Mediated by Parathormone (PTH)

Primary hyperparathyroidism (PHPT) is the most common cause of PTH-dependent hypercalcaemia. PHPT is a common endocrine disease, characterized by increased serum calcium and high or

Box 4.2.1 Aetiological classification of hypercalcaemia

PTH-mediated hypercalcaemia
- Primary hyperparathyroidism
- Tertiary hyperparathyroidism
- Familial hypocalciuric hypercalcaemia (FHH)
- Authentic ectopic secretion of PTH

Hypercalcaemia of malignancy (HCM)
- Humoral hypercalcaemia of malignancy (HHM)
- Local osteolytic hypercalcaemia (LOH)

Vitamin D-mediated hypercalcaemia
- Hypercalcaemia associated with granulomatous diseases
- Malignant hypercalcaemia
- Hypercalcaemia associated with vitamin D intoxication
- Hypercalcaemia associated with 24α hydroxylase mutations

Drug-related hypercalcaemia
- Vitamin A
- Oestrogen and oestrogen modulators
- Lithium
- Thiazides

Endocrine conditions
- Hyperthyroidism
- Addison's disease/hypoadrenalism
- Others endocrinopathies

Others
- Pregnancy/lactation
- Milk-alkali syndrome
- Immobilization
- Post-acute kidney failure
- Miscellaneous

inappropriately normal serum levels of PTH [17]. PHPT is prevalent in the sixth decade of life, with a female to male ratio of 3–4:1 [18], whereas there is no gender difference below the age of 40 years [19]. PHPT generally occurs sporadically, but occasionally it may be part of familial syndromes: multiple endocrine neoplasia type 1, 2 A, and 4 (MEN1, MEN2A, and MEN4), hyperparathyroidism-jaw tumour syndrome (PHPT-JT), familial isolated PHPT (FIHP). In the majority of patients PHPT is due to a single parathyroid adenoma, rarely to a multiglandular disease and very rarely (less than 1%) to parathyroid carcinoma. In the latter case, hypercalcaemia is usually severe. During the last decades the clinical presentation of PHPT changed and in industrialized countries the disease commonly occurs as an asymptomatic disorder: patients rarely present hypercalcaemic symptoms, overt bone disease, and neuromuscular weakness; conversely silent nephrolithiasis is still rather common (up to 35%) [18, 20, 21]. Parathyroidectomy is the only definitive cure for PHPT and should be recommended in symptomatic patients. Moreover, parathyroidectomy should also be recommended in those with asymptomatic PHPT with target organs involvement, as suggested by international guidelines. Some important novelties in patient's evaluation and management were introduced in the lastly updated version in 2013, particularly regarding the evaluation of bone and kidney involvement and the impact of related abnormalities on patient's management [22].

Tertiary hyperparathyroidism In patients with severe vitamin D deficiency, resistance to 1,25 $(OH)_2$ vitamin D and endstage renal failure, parathyroid hyperplasia may be followed by autonomous overgrowth of a clone of parathyroid cells and development of adenomas and hypercalcaemia, a condition known as tertiary hyperparathyroidism [23].

Familial hypocalciuric hypercalcaemia FHH is a rare autosomal dominant disorder characterized by moderate hypercalcaemia and relative hypocalciuria, namely the urinary calcium excretion is low in relationship to the increased levels of serum calcium and renal calcium load. The reported prevalence of the disease is about 1 in 78 000 individuals, but it is probably underestimated, because of the subclinical picture and misdiagnosis [24]. FHH is due to a genetic disorder associated in the most cases with a loss-of-function mutation in the calcium-sensing receptor (CaSR) gene. This accounts for lower sensitivity to the inhibitory effect of serum Ca^{2+} on the parathyroid synthesis, secretion of PTH and increased renal tubular reabsorption of calcium. Patients are generally asymptomatic. The diagnosis should be considered in asymptomatic patients with mild-to-moderate hypercalcaemia and inappropriately normal or slightly increased PTH levels. The clue for the diagnosis of FHH is the finding of a calcium to creatinine clearance ratio, less than 0.01 (in patients with PHPT this ratio is usually above 0.02–0.03). Vitamin D deficiency, if present, should be corrected before urine collection. Laboratory testing (serum and urinary calcium) in relatives may be of value for supporting the diagnosis, which should be confirmed by genetic testing. It is important to remember that FHH is extremely rare as compared with PHPT and therefore in equivocal cases the latter diagnosis is the most likely one.

Authentic ectopic secretion of PTH. Since 1990 rare cases of patients with malignancies were reported, in whom hypercalcaemia was associated with authentic PTH secretion by the tumours. Streweler *et al.* for the first time described a patient with a neuroectodermal neck tumour shown to produce intact PTH [25]. Circulating PTH levels are usually particularly high in patients with a true intact PTH ectopic secretion. In the last 20 years 32 articles describing authentic ectopic secretion of PTH have been published. PTH secreting tumours can be found in the neck (medullary and papillary thyroid carcinoma, cervical paraganglioma, nasopharyngeal rhabdomyosarcoma, squamous tonsil carcinoma), in the thorax (lung carcinoma, thymoma), and in the gastrointestinal tract and pelvis (hepatocarcinoma, pancreatic adenocarcinoma, oesophageal carcinoma, pancreatic islet carcinoma, ovarian carcinoma, adenosquamous endometrium carcinoma) [26]. Ectopic PTH secretion should be considered when moderate-to-severe hypercalcaemia is found in a cancer patient, particularly if associated with high levels of PTH. When hypercalcaemia is mild, the diagnosis of PHPT, which is much more common than ectopic PTH secretion, should be considered, and therefore we cannot exclude that these two conditions may coexist in the same patient. Neck imaging studies and selective venous blood sampling for PTH measurement may be of help in selected cases.

Hypercalcaemia of Malignancy

Hypercalcaemia associated with malignancy (HCM) is the most common cause of PTH-independent hypercalcaemia and is responsible of the majority of cases of hypercalcaemia in hospitalized patients. HCM can be distinguished in humoral hypercalcaemia of malignancy (HHM), the most common form due to endocrine or paracrine effects on bone of tumour-produced molecules, and local osteolytic hypercalcaemia (LOH) that is mainly, but not only, due to the direct invasion of bone by tumour cells.

HCM can be associated at any type of cancer. It generally occurs at endstage of the disease, is associated with severe gastrointestinal, renal, neuropsychic, and cardiovascular symptoms and is predictive of a poor prognosis. In some patients serum calcium may increase over a short period of time and hypercalcaemia may be of difficult control leading to fatal hypercalcaemic crisis [27].

Lung and breast cancers and multiple myeloma account for more than half of all cases of HCM. Other cancers frequently associated with HCM are squamous cells carcinoma of the head and neck, and renal and kidney cancers [28].

Humoral hypercalcaemia of malignancy. HHM is due to the paracrine or endocrine actions of molecules secreted by the tumour cells, mainly the PTH-related peptide (PTHrP) (Figure 4.2.2). The hypothesis that a factor similar to PTH could be responsible for hypercalcaemia in cancer patients was firstly proposed by Albright in 1940, but it was only in 1980s that PTHrP was identified and shown to be similar to native PTH in its N-terminal end [29]. PTH and PTHrP are encoded by the same gene, but the two molecules have different structure and mechanism of control, even if they share the same PTH1 receptor [30]. PTHrP mimics most of the actions of PTH: it increases bone resorption and distal tubular calcium reabsorption, and inhibits proximal tubular phosphate transport. Conversely, PTHrP does not stimulate the production of $1,25(OH)_2$ vitamin D and has no effect on intestinal calcium absorption. Despite the use of the same cellular receptor, PTHrP and PTH act in a different way on bone, likely because of the activation of different intracellular signal pathways [31]. PTH induces a high, but still coupled bone turnover, whereas PTHrP produces an uncoupling of bone remodelling (non-equilibrium hypercalcaemia) and a negative bone balance. Moreover, However, PTHrP-induced hypercalcaemia is also and largely caused by its action of on the renal tubule with increased calcium reabsorption and phosphate excretion. PTHrP production is common in patients with adenocarcinomas of breast, prostate, and ovary. An epidemiological follow-up study demonstrated a strong association between ovarian cancer and hypercalcaemia, with a 63% increased risk of mortality when hypercalcaemia occurs [32]. HHM can also occur in patients with squamous cell carcinoma (lung, cervix, oesophagus), renal carcinomas, and lymphomas, although in the latter cases overproduction of $1,25(OH)_2$ vitamin D may also play an important role.

Local osteolytic hypercalcaemia. Since the 1940s, it has been recognized that some solid tumours (e.g. breast cancer) and haematological malignancies (e.g. myeloma, lymphoma, and leukaemia) may led to hypercalcaemia due to a widespread skeletal involvement (LOH). In these patients an extensive bone marrow invasion by the tumour was found at pathology and the initial understanding was that hypercalcaemia was due to the extensive bone marrow destruction by osteolytic tumours. Subsequently, it was clear that some tumours produced factors that could act in a paracrine way, inducing calcium mobilization from bone [33]. The current understanding is that tumour cells, after their interaction with the bone marrow microenvironment, produce different factors with activating (osteoclast activating factors, OAFs) or inhibitory (osteoblast inhibitory factors, OIFs) effects. OAFs, namely interleukin (IL)1, IL6, IL11, macrophage inflammatory protein 1 a (MIP1a) and, particularly PTHrP, directly increase maturation and activity of osteoclasts. This effect is mediated by an increased production by osteoblasts and stromal cells of RANK ligand (RANKL), associated with and inhibition of osteoprotegerin (OPG) production by the same cells. RANKL transduction signal induces osteoclast maturation by the mitogen activated protein kinase (MAPK) pathway, thus leading to bone resorption. The bone matrix destruction releases some growth factors such as TGFβ, FGF, PDGF, and IGF1 that stimulates tumour cells growth in a vicious cycle [34]. At the same time, the Wnt/dikkopf-1 pathway is inhibited, leading to a decreased osteoblast activation. The final effect is an uncoupling of bone remodelling, with prevalent bone destruction and calcium release [35]. Clinical features of LOH are hypercalcaemia, bone pain, bone loss, and fragility fractures.

Vitamin D-mediated Hypercalcaemia

$1,25(OH)_2D$, the active form of vitamin D produced in the kidney, acts as a classical hormone on tissues which express the vitamin D

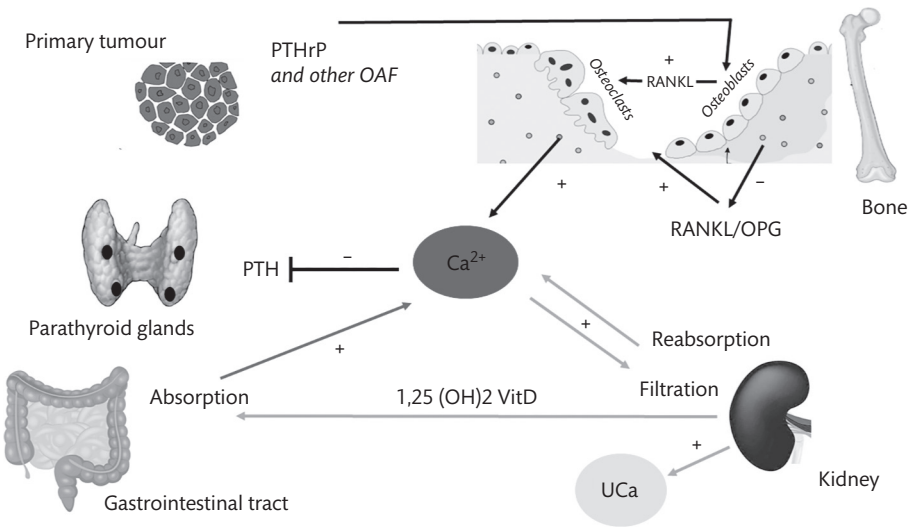


Figure 4.2.2 Mechanism involved in the humoral hypercalcaemia of malignancy (HHM). PTH: parathormone; PTHrP, PTH-related peptide; OAF, osteoclast activating factors; Ca2+, ionized serum calcium; UCa, urinary calcium; 1,25 (OH)2 VitD, 1,25 didroxy vitamin D.

receptor (VDR), like bone and intestine. It may also be locally produced by several non-renal cells expressing the 1-α-hydroxylase enzyme. This extrarenal production of 1,25 $(OH)_2D$ acts in a paracrine/autocrine fashion and mediates the so-called extraskeletal effects of vitamin D, which have received great attention in the recent years [36]. Vitamin D-mediated hypercalcaemia is caused by different mechanisms that are discussed next: (i) ectopic 1-α-hydroxylase overactivation, which may occur in granulomatous diseases and malignant tumours; (ii) intoxication of vitamin D and (iii) inactivating mutations of genes encoding key enzymes of $1,25(OH)_2$ catabolism and degradation (namely mutations of 24-hydroxylase gene) (see Box 4.2.1).

Hypercalcaemia associated with granulomatous diseases. Activation of 1-α-hydroxylase enzyme in macrophage is a well-known cause of hypercalcaemia associated with granulomatous diseases. It was originally described in sarcoidosis, tuberculosis, Crohn's, and Wegener's diseases.

Sarcoidosis. The first case of hypercalcaemia associated with sarcoidosis was reported in 1930 and since then several case reports were published, claiming that 10% of patients with sarcoidosis may experience hypercalcaemia [37]. It was subsequently demonstrated that serum levels of 1,25 $(OH)_2D$ are increased in patients with sarcoidosis and that the source of vitamin D is extrarenal. Indeed, elevated serum 1,25 $(OH)_2D$ levels were shown in anephric patients and in patients with endstage kidney failure [38]. It was shown that sarcoid-associated pulmonary alveolar macrophages produce their own 1-α-hydroxylase, which is stimulated by interferon γ, but not under the control of PTH [39, 40]. Hypercalcaemia is usually mild to moderate, and symptoms and laboratory findings are similar to those of other forms of PTH-independent hypercalcaemia. In addition, the serum level of angiotensin-converting enzyme is increased, and its concentration generally correlates with the severity of the disease. Serum levels of $1,25(OH)_2D$ are increased whereas those of 25(OH)D are either normal or decreased. The activity of sarcoid-associated pulmonary alveolar macrophages is efficiently suppressed by glucocorticoids, which are effective in lowering serum calcium. Ketoconazole has also been successfully used in the management of sarcoidosis-associated hypercalcaemia [40].

Tuberculosis. Patients with tuberculosis may experience hypercalcaemia with a highly variable (2.3–48%) prevalence in different studies. Hypercalcaemia is mainly due to the production of $1,25(OH)_2D$ by macrophage of the pleural cavity. $1,25(OH)_2D$ has been shown to have an antimicrobial action on mycobacterium tuberculosis, stimulating the secretion of cathelicidins and other antimicrobial peptides. Rifampicin and isoniazid, the most commonly used drugs in the management of patients with tuberculosis, are also helpful in controlling hypercalcaemia: rifampicin induces enzymes that degrade 25(OH)D, and isoniazid inhibits the synthesis of $1,25(OH)_2D$.

Other granulomatous diseases. Several other granulomatous diseases have been reported to be associated with hypercalcaemia, including infection (mycobacterium avium-complex disease, leprosy, fungal infections, pneumocystis pneumonia, and cat-scratch disease) [40] as well as non-infectious diseases. A recent meta-analysis reviewed all cases of hypercalcaemia associated to cosmetics injections [41]. The authors identified 23 patients, who mostly used silicon, polymethylmethacrylate, and paraffin oil injected in buttock and breast. Hypercalcaemia was severe with a mean serum Alb-Ca of 3.43 mmol/L and detected years after the initial injection. In these patient's serum $1,25(OH)_2D$ was elevated and PTH low or suppressed in the majority of cases [41]. Finally, hypercalcaemia has been reported in some cases of Wegener's granulomatosis, Chron's diseases, giant cell polymyositis, and berylliosis.

Malignant hypercalcaemia. Hypercalcaemia complicates up to 13% of non-Hodgkin and 5% of Hodgkin lymphoma, and when occurs is associated with a worse prognosis. The pathogenesis of lymphoma-related hypercalcaemia involves the increased production of ectopic $1,25(OH)_2D$, probably in lymph nodes, as suggested by *in vivo* studies [42]. Increased serum levels of $1,25(OH)_2D$ have been demonstrated in 20–40% of patients with non-Hodgkin and 100% of those with Hodgkin lymphoma [43, 44].

Hypercalcaemia associated with vitamin D intoxication. Vitamin D insufficiency is very common worldwide in the general population, particularly in older people and in subjects living in the Eastern countries, as well as in patients with several disease states affecting vitamin D absorption and metabolism (celiac diseases, malabsorption, etc.). Guidelines developed by the Institute of Medicine (IOM) in 2011 recommend different dietary vitamin D allowances in children, adults, elderly, and pregnant women and indicate that the daily intake of vitamin D should not be greater of 4000 UI in adults, 3000 UI in children 4–8 years, 2500 UI in children 1–3 years, and 100–1500 UI in infants [45]. Diseases unrelated to bone and calcium homeostasis may be associated with a low vitamin D status and decreased locally produced $1,25(OH)_2D$ by circulating 25OHD may have a pathogenic role in these conditions. This knowledge led to wide use vitamin D supplements not only to correct insufficiency in the general population but also to increase serum 25OHD in patients with several diseases, despite lack of evidence of benefit from vitamin D supplementation. Vitamin D intoxication is a rare event, but when occurring may lead to life-threatening hypercalcaemia [46, 47]. A recent meta-analysis reported 13 patients presenting with severe hypercalcaemia and related symptoms such as vomiting, dehydration, pain, and anorexia. In all cases, vitamin D intoxication was due to either excessive dosing by the patients or the prescribing physicians, and, rarely, by manufacturing errors. The daily dose of vitamin D ranged between 50 000 IU and 2 604 000 IU. Serum calcium ranged between 11.1 and 23.1 mg/dl and serum 25(OH)D concentration was greater than 150 ng/ml [48]. Vitamin D intoxication is different from vitamin D hypersensitivity, a condition in which a pre-existing hypercalcaemia is exacerbated by high doses of vitamin D.

Hypercalcaemia associated with CYP24A1 hydroxylase mutations. Idiopathic infantile hypercalcaemia (IIH, OMIM 143880) is a rare disease characterized by hypercalcaemia, hypercalciuria, nephrocalcinosis, usually diagnosed in infant between 3 and 7 months of age. The disease is rare and is related to an impairment of vitamin D catabolism. IIH is a genetic disorder due to loss-of-function mutations in the human cytochrome P450 24 subfamily A member 1 (CYP24A1) gene, encoding the 24-hydroxilase enzyme, which is involved in the catabolism of vitamin D [49]. The diseases can present with different phenotypes ranging from severe forms diagnosed early in the infancy and sometimes lethal [50] to milder forms, often diagnosed in the adulthood during workout for recurrent nephrolithiasis [51]. The first case described in 2011 lead to the identification of the genetic cause and thereafter around 200 cases have been reported leading to an increased insight into the diagnostic and therapeutic management of this disease [49].

Drug-Related Hypercalcaemia

Vitamin A. Vitamin A and its analogues are often used in the treatment of neoplastic and dermatologic diseases. Hypervitaminosis A can rarely be complicated by hypercalcaemia. The mechanisms proposed include either an increased bone resorption due to the fact that osteoclasts and osteoblasts express retinoic acid receptors or an action of vitamin A on calcium release following lysosomal degradation [52]. The susceptibility to develop hypervitaminosis A is highly variable, and the daily intake up to 50 000 UI is safe. Effective treatments, when needed later after discontinuation of vitamin A, include hydration, bisphosphonates, and glucocorticoids.

Oestrogen receptor modulators. Hypercalcaemia is rarely seen in women with breast cancer, in the early phase of treatment with tamoxifen, a selective oestrogen receptor modulator. The exact mechanism is not completely understood, and it is thought to be related to the presence of oestrogen receptor on cancer cells [53]. Hypercalcaemia is generally transient and, in some cases, may be moderate to severe.

Lithium. Lithium, effectively used worldwide in the management of patients with bipolar disorders, has been shown to interfere with thyroid and parathyroid functions. There are many studies that demonstrate that lithium can associated with a biochemical profile similar to that of PHPT. Lithium competes with calcium for the binding to the CaSR in the parathyroid glands, leading to a shift to the right of the calcium set-point and hypersecretion of PTH [54]. This effect is transient and resolves after lithium discontinuation, when used for limited time, but persists when lithium is chronically used [54]. In a few cases treatment with lithium has been associated with the development of parathyroid adenomas and, therefore, a 'true' PHPT. The optimal treatment is surgical as suggested in other forms of PHPT. Recently calcimimetics have been shown to be effective in lowering hypercalcaemia due to lithium treatment [55].

Thiazides. Treatment with thiazides has been associated with hypercalcaemia in 2% of cases and this side effect has been known for decades, even if the exact pathogenetic mechanism is poorly understood. Two possible triggers are the volume depletion induced by thiazides and the possible inhibition of the transient receptor potential cation channel subfamily V member 5 (TRVP5), a specific calcium selective channel [56]. Hypercalcaemia due to thiazides is generally mild and resolves after thiazides withdrawal. However, some cases of occult PHPT and persistent hypercalcaemia after drug discontinuation have been described. This can be due to the fact that thiazides may worsen hypercalcaemia and reveals the presence of a previously undiagnosed mild PHPT characterized by normal or intermittently mildly elevated serum calcium [57].

Hypercalcaemia Related to Endocrine Diseases

Hyperthyroidism. Thyroid hormones are known to have a role in the regulation of calcium metabolism and about 20% of patients with hyperthyroidism, especially in those with severe disease, may experience mild hypercalcaemia. Hypercalcaemia related to hyperthyroidism is generally mild to moderate, rarely above 3 mmol/L. Only few cases of serious hypercalcaemia have been described [58]. The pathophysiology of hyperthyroidism-related hypercalcaemia is still not completely clear. One hypothesis is that thyroid hormones induce an increase of the osteoclastic activity, as shown in fetal rat bones cultures exposed to an excess of thyroid hormones [59]. Other studies suggested that thyroid hormones excess could act on bone metabolism increasing the sensitivity of bone to PTH and also catecholamines [60]. Finally, serum IL6 is increased in thyrotoxicosis and correlated with serum thyroid hormone levels [61]. IL6 is known to activate osteoclastogenesis through the RANK-RANKL signalling pathway, leading to an increased bone resorption and calcium release [58]. As matter of fact, bone resorption markers are often increased in patients with hypercalcaemia associated with hyperthyroidism [62]. The primary treatment of patients with hyperthyroidism-induced hypercalcaemia is the adequate control of hyperthyroidism, followed by the other general measurement described next.

Addison's disease/hypoadrenalism. Hypercalcaemia has been reported in 6% of primary adrenal insufficiency, associated with PTH and $1,25(OH)_2$ D suppression and increased bone turnover [63]. Its pathophysiology is still unclear,

Others endocrinopathies. Some cases of hypercalcaemia have been reported in patients with VIPoma and pheochromocytoma, with uncertain pathophysiology [64].

Hypercalcaemia Related to Other Conditions

Hypercalcaemia during pregnancy. Hypercalcaemia rarely occurs in pregnant women (0.03%), but its recognition is challenging since the pregnancy itself induces significant changes in calcium homeostasis. During the whole pregnancy total serum calcium is decreased because of haemodilution-dependent hypoalbuminemia, Ca^{2+} is normal, and urinary calcium is increased. The renal 1α-hydroxylase activity is stimulated by oestradiol, prolactin, placental lactogen, calcitonin, and PTHrP of placenta and mammary gland origin [65, 66]. The increased 1α-hydroxylase activity may lead to hypercalcaemia. Symptoms of hypercalcaemia during pregnancy are the same as in the general population but can overlap with the symptoms related to the pregnancy: nausea, vomiting, constipation. Some cases of severe hypercalcaemia have been reported in pregnancy presenting with pancreatitis or kidney failure [67]. Fetal complication may occur, depending on the severity of hypercalcaemia: neonatal death (2%), neonatal hypocalcaemia and tetany (15%), and intrauterine growth decline (2%) [40]. PHPT is the most common cause of hypercalcaemia during pregnancy. Other rare causes of pregnancy-associated hypercalcaemia have been reported: FHH [68], milk-alkali syndrome, malignancy and benign tumours (dermoid cysts, pheochromocytomas, fibromas) [69]. The same pharmacological and surgical approach used for general population could be theoretically used in women with hypercalcaemia during pregnancy, even though there are few studies on safety and maternal-fetal outcomes. Hydration may help to control hypercalcaemia in mild cases. When this approach in insufficient parathyroidectomy could be considered in pregnant women with PHPT when serum calcium is above 2.85 mmol/L, preferably in the second trimester [70]. In a recent series of 28 pregnant women with PHPT followed between 2000 and 2015, 22 were treated with intravenous fluids, antihypertensive drugs, and cinacalcet (one case), and six patients underwent PTx. In the former group, 30% had pre-eclampsia and 50% a preterm delivery. On the other hand, in the latter group no complications were observed. Moreover, in a retrospective study of 109 pregnant women with PHPT, PTx was shown to reduce the rate of fetal complications [71]. Cinacalcet use has been described in few cases. Even though there is some concern about a potential negative effect of this drug in the placenta, where CaSR is expressed, animal

studies have shown no toxic effects in pregnant rats and rabbits [72]. Moreover, no adverse maternal and fetal effects were reported in pregnant women [73]. Bisphosphonates are not extensively studied during pregnancy, but in few series of newborns exposed at intrauterine bisphosphonates no serious adverse events were reported after delivery [74]. Finally, calcitonin does not seem to cross the placenta and is considered safe during pregnancy and lactation [75].

Milk-alkali syndrome. This is a rare syndrome in which hypercalcaemia and alkalosis occur in patients treated with calcium carbonate for peptic ulcer. Hypercalcaemia arises when the intestinal absorption of calcium is higher than its renal excretion, as it may occur in patients with reduced kidney function. Factors that contribute to the maintenance of hypercalcaemia are volume depletion, reduced glomerular filtration rate, and metabolic alkalosis that stimulates renal resorption. Hypercalcaemia may arise acutely or can be a chronic complication [76].

Immobilization. Hypercalcaemia is often seen after long period of immobilization in young patients with spinal injury and in elderly hospitalized people. In 1998 Zerwekh *et al.* studied a group of young healthy volunteers, in whom 12 weeks of skeletal unloading induced hypercalcaemia associated with reduced levels of PTH and 1,25(OH)2D, hypercalciuria and hypophosphaturia [77]. An increased bone resorption is thought to be the cause. Indeed, osteocytes seem to act as mechanoreceptors and are sensitive to unloading. Unloading results in an increased expression and synthesis of sclerostin, which leads to an inhibition of the Wnt signalling pathway. This results in a reduction of the production of osteoprotegerin, the decoy receptor for RANKL and regulator of bone resorption [78].

Post-acute kidney failure. Hypercalcaemia may occur during the recovery phase after an acute kidney failure because of secondary hyperparathyroidism caused by hyperphosphatemia and hypocalcaemia of the oliguric initial phase [79].

Miscellaneous. Several other conditions may be rarely associated with hypercalcaemia such as manganese intoxication, hypophosphatasia, Paget's disease, chondrodysplasia, William's syndrome.

Therapy

The management of hypercalcaemia should be addressed at the underlying cause/mechanisms and the prevention/management of target organ damages. The timing and the type of therapy are related to the cause and severity of hypercalcaemia, clinical symptoms, and renal function impairment. When hypercalcaemia is mild and the underlying cause has been identified, the treatment could be limited at the underlying cause. When hypercalcaemia in moderate to severe, the aim is also and most importantly directed to decrease serum calcium, by increasing renal calcium excretion and decreasing bone resorption and intestinal calcium absorption.

Increase Renal Calcium Excretion

Dehydration occurs in hypercalcaemic states because of a decreased fluid intake due to anorexia, nausea and vomiting, and polyuria due to a decreased renal water reabsorption. Therefore, signs and symptoms of vascular volume depletion are present in patients with hypercalcaemia. Hydration should be the first therapeutic approach to promote extracellular volume expansion, using oral fluids in patients with mild hypercalcaemia and intravenous saline in those with severe hypercalcaemia.

Patients with severe hypercalcaemia require an immediate treatment which includes intravenous 0.9% saline administration 3–4 litres per day or a bolus of 1–2 litres followed by 200–250 ml saline/hour [80]. Patients with heart or kidney failure should be carefully monitored for the risk of fluid overload. After volume expansion, loop diuretics, such as furosemide, may be considered to further increase renal calcium excretion, but caution is mandatory for the risk of dehydration and electrolytes imbalance. This view is not generally accepted nowadays. As a matter of fact, a metanalysis published in 2008 did not show benefit of adding loop diuretics to hydration in RCT [81]. When the aforementioned treatments fail, haemodialysis with low-calcium dialysate could be an option, which does not treat the cause of hypercalcaemia, but leads to a rapid solution in emergency situations. Some studies have shown that continuous veno-venus haemofiltration using citrate as calcium chelator could be more effective that haemodialysis. Haemodialysis may be complicated by infections, thromboembolism, and hypotension [82].

Decrease Bone Resorption

Increased osteoclastic activity and bone resorption is the most common mechanism responsible of hypercalcaemia. Bisphosphonates decrease bone resorption by several mechanisms: inhibition of monocyte/macrophages differentiation into mature osteoclasts, increased osteoclasts apoptosis, and binding to hydroxyapatite crystals thus decreasing their degradation. The use of bisphosphonates is limited to patients with moderate-to-severe hypercalcaemia. The intravenous route is preferable since bisphosphonates are poorly adsorbed when given orally and large doses are needed. Etidronate and clodronate were initially used but have been subsequently replaced by the novel and more potent aminobisphosphonates. Pamidronate and zoledronate have been approved by the Food and Drug Administration (FDA) in the United States and by the European Medicines Agency (EMEA) in Europe for the management of hypercalcaemia of malignancy. Both agents have proved their efficacy and safety in several randomized clinical trials [42, 83]. Either pamidronic acid or zoledronic acid should be started as soon as possible, because both require 48–72 hours for the initial response and the serum calcium nadir is obtained 4–7 days after the intravenous administration. Zoledronic acid compared with pamidronic acid has a more rapid onset of action and greater hypocalcaemic effect [84]. Zoledronic acid is usually administered at the dose of 4 mg in 100 ml of saline over 15 minutes and normocalcaemic state is restored in a few days in the large majority of patients. In patients with kidney failure and serum creatinine exceeding 3 mg/dl, the American Society of Oncology recommends a reduction in the dose of zoledronic acid, because of potential nephrotoxic effect. A worsening of renal failure has been reported after multiple infusions of zoledronic acid [85]. When there are concerns for renal toxicity, ibandronate, a less potent but less toxic bisphosphonate, could be used. Subcutaneous administration of clodronate is an option in those endstage cancer patients who need domiciliary therapy.

Calcitonin also is effective in rapidly reducing serum calcium and its effect begins within 2 hours. Salmon calcitonin is generally more

effective than human calcitonin and is preferably used at the dose of 4–8 units/kg of body weight, either intramuscularly or subcutaneously [86, 87]. However, tolerance to the drug (or tachyphylaxis) generally occurs and the hypocalcaemic effect wanes. In clinical practice calcitonin could be used when a rapid hypocalcaemic effect is aimed, while waiting for the more potent hypocalcaemic effect of bisphosphonates.

Denosumab, a fully human monoclonal antibody directed to the RANK ligand, decreases osteoclasts differentiation and survival. Denosumab is approved for the management of postmenopausal osteoporosis and in the prevention and management of osteoporosis due to adjuvant therapy in patients with breast and prostate cancers. Preclinical studies in a mice model of malignant hypercalcaemia due to PTHrP showed that denosumab rapidly decreases serum calcium [88]. Recently some case reports have shown a potent hypocalcaemic effect of denosumab in cancer patients [89, 90]. The value of denosumab in the management of hypercalcaemia of malignancy has been further proved in a single-arm, multicentre, international phase II study who enrolled 33 patients with malignant hypercalcaemia and no-response to bisphosphonates treatment. Patients were treated with 120 mg sc of denosumab every 4 weeks. The large majority (70%) of patients shoved a decrease of serum calcium from 13.7 mg/dl to value ≤11.5 mg/dl and 64% reached serum calcium levels ≤10.8 mg/dl; the median duration of response was 104 days. Serious adverse events were reported in two cases [91]. In 2014 the FDA approved denosumab for the treatment of malignant hypercalcaemia refractory to bisphosphonate therapy. More recently a meta-analysis of two randomized double-blind trials in cancer patients with bone metastases has shown that denosumab treatment significantly delayed, compared to zoledronic acid, the occurrence of malignant hypercalcaemia [92].

Preclinical studies, in a mice model of PTHrP-dependent malignant hypercalcaemia administration of monoclonal PTHrP antibodies was shown to have a hypocalcaemic effect, opening the potential development of this therapeutic approach also in selected cases in humans [93].

Decrease Intestinal Calcium Absorption

Increased intestinal calcium absorption contributes to the development of hypercalcaemia, particularly when due to vitamin D intoxication. A general measure in these patients consists of avoiding calcium and vitamin D intake. Glucocorticoids represent the first-choice option, because of their inhibitory effect on 1α hydroxylase enzyme, thus reducing intestinal calcium absorption. No guidelines are available to guide this treatment: a generally accepted schedule is the intravenous administration of hydrocortisone at a daily dose of 200–400 mg for 3–5 days, followed by oral daily prednisone at the dose of 10–20 mg for 7 days [86].

Calcimimetics

Cinacalcet, a calcimimetic agent, is approved for the management of hypercalcaemia in patients with PHPT, in whom parathyroidectomy would be indicated on the basis of serum calcium concentration but not performed because contraindicated or refused by the patient. The efficacy of cinacalcet in lowering and often normalizing serum calcium has been shown in several PHPT settings [94], including patients with intractable hypercalcaemia related to parathyroid cancer [95]. Interestingly, cinacalcet was effective in reducing hypercalcaemia in two cases of humoral malignant hypercalcaemia due to metastatic renal cell and breast carcinomas [96].

REFERENCES

1. Maier JD, Levine SN. Hypercalcemia in the intensive care unit: a review of pathophysiology, diagnosis, and modern therapy. *J Intensive Care Med*, 2015; **30**: 235–52.
2. Blaine J, Chonchol M, Levi M. Renal control of calcium, phosphate, and magnesium homeostasis. *Clin J Am Soc Nephrol*, 2015; **10**: 1257–72.
3. Keller J, Schinke T. The role of the gastrointestinal tract in calcium homeostasis and bone remodeling. *Osteoporos Int*, 2013; **24**: 2737–48.
4. French S, Subauste J, Geraci S. Calcium abnormalities in hospitalized patients. South Med J, 2012; **105**: 231–7.
5. Catalano A, Chilà D, Bellone F, *et al*. Incidence of hypocalcemia and hypercalcemia in hospitalized patients: is it changing? *J Clin Transl Endocrinol*, 2018; **13**: 9–13.
6. Yu N, Donnan PT, Murphy MJ, Leese GP. Epidemiology of primary hyperparathyroidism in Tayside, Scotland, UK. *Clin Endocrinol (Oxf)*, 2009; **71**: 485–93.
7. Wermers RA, Khosla S, Atkinson EJ, *et al*. Incidence of primary hyperparathyroidism in Rochester, Minnesota, 1993–2001: an update on the changing epidemiology of the disease. *J Bone Miner Res*, 2006; **21**: 171–7.
8. Jick S, Li L, Gastanaga VM, Liede A, Hernandez RK. Prevalence of hypercalcemia of malignancy among pediatric cancer patients in the UK clinical practice research datalink database. *Clin Epidemiol*, 2017; **9**: 339–43.
9. Parfitt AM. Quantum concept of bone remodeling and turnover: implications for the pathogenesis of osteoporosis. *Calcif Tissue Int*, 1979; **28**: 1–5.
10. Žofková I. Hypercalcemia. Pathophysiological aspects. *Physiol Res*, 2016; **65**: 1–10.
11. Endres DB. Investigation of hypercalcemia. *Clin Biochem*, 2012; **45**: 954–63.
12. Nordenström E, Katzman P, Bergenfelz A. Biochemical diagnosis of primary hyperparathyroidism: analysis of the sensitivity of total and ionized calcium in combination with PTH. *Clin Biochem*, 2011; **44**: 849–52.
13. Gøransson LG, Skadberg Ø, Bergrem H. Albumin-corrected or ionized calcium in renal failure? What to measure? *Nephrol Dial Transplant*, 2005; **20**: 2126–9.
14. Byrnes MC, Huynh K, Helmer SD, Stevens C, Dort JM, Smith RS. A comparison of corrected serum calcium levels to ionized calcium levels among critically ill surgical patients. *Am J Surg*, 2005; **189**: 310–14.
15. Wandrup J. Critical analytical and clinical aspects of ionized calcium in neonates. *Clin Chem*, 1989; **35**: 2027–33.
16. Silverberg SJ, Clarke BL, Peacock M, *et al*. Current issues in the presentation of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3580–94.
17. Marcocci C, Cetani F. Clinical practice. Primary hyperparathyroidism. *N Engl J Med*, 2011; **365**: 2389–97.
18. Khan AA, Hanley DA, Rizzoli R, *et al*. Primary hyperparathyroidism: review and recommendations on evaluation, diagnosis, and management. A Canadian and international consensus. *Osteoporos Int*, 2017; **28**: 1–19.

19. Saponaro F, Cetani F, Repaci A, *et al*. Clinical presentation and management of patients with primary hyperparathyroidism in Italy. *J Endocrinol Invest*, 2018; **41**: 1339–48.
20. Mihai R, Wass JA, Sadler GP. Asymptomatic hyperparathyroidism--need for multicentre studies. *Clin Endocrinol*, 2008; **68**: 155–64.
21. Cipriani C, Biamonte F, Costa AG, *et al*. Prevalence of kidney stones and vertebral fractures in primary hyperparathyroidism using imaging technology. *J Clin Endocrinol Metab*, 2015; **100**: 1309–15.
22. Bilezikian JP, Brandi ML, Eastell R, *et al*. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the fourth international workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3561–9.
23. Jamal SA, Miller PD. Secondary and Tertiary Hyperparathyroidism. *J Clin Densitom*, 2013; **16**: 64–8.
24. Varghese J, Rich T, Jimenez C. Benign familial hypocalciuric hypercalcemia. *Endocr Pract*, 2011; **17** Suppl 1: 13–17.
25. Mol JA, Visser TJ. Rapid and selective inner ring deiodination of thyroxine sulfate by rat liver deiodinase. *Endocrinology*, 1985; **117**: 8–12.
26. Kandil E, Noureldine S, Khalek MA, Daroca P, Friedlander P. Ectopic secretion of parathyroid hormone in a neuroendocrine tumor: a case report and review of the literature. *Int J Clin Exp Med*, 2011; **4**: 234–40.
27. Inzucchi SE. Understanding hypercalcemia: its metabolic basis, signs, and symptoms. *Postgrad Med*, 2004; **115**: 69–70, 73–6.
28. Basso U, Maruzzo M, Roma A, Camozzi V, Luisetto G, Lumachi F. Malignant hypercalcemia. *Curr Med Chem*, 2011; **18**: 3462–7.
29. Martin TJ, Suva LJ. 9 Parathyroid hormone-related protein: a novel gene product. *Baillieres Clin Endocrinol Metab*, 1988; **2**: 1003–29.
30. Mundy GR, Edwards JR. PTH-related peptide (PTHrP) in hypercalcemia. *J Am Soc Nephrol*, 2008; **19**: 672–5.
31. Gensure RC, Gardella TJ, Jüppner H. Parathyroid hormone and parathyroid hormone-related peptide, and their receptors. *Biochem Biophys Res Commun*, 2005; **328**: 666–78.
32. Schwartz GG, Skinner HG. Prospective studies of total and ionized serum calcium in relation to incident and fatal ovarian cancer. *Gynecol Oncol*, 2013; **129**; 169–72.
33. Silbermann R, Roodman GD. Myeloma bone disease: pathophysiology and management. *J Bone Oncol*, 2013; **2**: 59–69.
34. Wu MY, Li CJ, Yiang GT, *et al*. Molecular regulation of bone metastasis pathogenesis. *Cell Physiol Biochem*, 2018; **46**: 1423–38.
35. Walker RE, Lawson MA, Buckle CH, Snowden JA, Chantry AD. Myeloma bone disease: pathogenesis, current treatments and future targets. *Br Med Bull*, 2014; **111**: 117–38.
36. Wacker M, Holiack MF. Vitamin D-effects on skeletal and extraskeletal health and the need for supplementation. *Nutrients*, 2013; **5**: 111–48.
37. Rizzato G. Clinical impact of bone and calcium metabolism changes in sarcoidosis. *Thorax*, 1998; **53**: 425–9.
38. Maesaka JK, Batuman V, Pablo NC, Shakamuri S. Elevated 1,25-Dihydroxyvitamin D levels: occurrence with sarcoidosis with end-stage renal disease. *Arch Intern Med*, 1982; **142**: 1206–7.
39. Adams JS, Mercedes GA. Characterizaton of 1alpha-hydroxylation of vitain D3 sterols by cultured alveolar macrophages from patients with sarcoidosis. *J Exp Med*, 1985; **161**; 755–65.
40. Tebben PJ, Singh RJ, Kumar R. Vitamin D-mediated hypercalcemia: mechanisms, diagnosis, and treatment. *Endocr Rev*, 2016; **37**: 521–47.
41. Tachamo N, Donato A, Timilsina B, *et al*. Hypercalcemia associated with cosmetic injections: a systematic review. *Eur J Endocrinol*, 2018; **178**: 425–30.
42. Stewart AF. Hypercalcemia associated with cancer. *N Engl J Med*, 2005; **352**: 373–9.
43. Sargent JTS, Smith OP. Haematological emergencies managing hypercalcaemia in adults and children with haematological disorders: state of the art review. 2010; *Br J Haematol*, 2010;149: 465–477.
44. Seymour JF, Gagel RF, Hagemeister FB, Dimopoulos MA, Cabanillas F. Calcitriol production in hypercalcemic and normocalcemic patients with non-Hodgkin lymphoma. *Ann Intern Med*, 1994; **121**: 633–40.
45. Institute of Medicine (US). *Dietary Reference Intakes for Calcium and Vitamin D*. Ross CA, Taylor CL, Yaktine AL, Del Valle HB, eds. Washington, DC: National Academies Press, 2011.
46. Bell DA, Crooke MJ, Hay N, Glendenning P. Prolonged vitamin D intoxication: presentation, pathogenesis and progress. *Intern Med J*, 2013; **43**: 1148–50.
47. Taylor PN, Davies JS. A review of the growing risk of vitamin D toxicity from inappropriate practice. *Br J Clin Pharmacol*, 2018; **84**: 1121–7.
48. Galior K, Grebe S, Singh R. Development of vitamin D toxicity from overcorrection of vitamin D deficiency: a review of case reports. *Nutrients*, 2018; **10**: DOI: 10.3390/nu10080953.
49. Schlingmann KP, Kaufmann M, Weber S, *et al*. Mutations in *CYP24A1* and Idiopathic Infantile Hypercalcemia. *N Engl J Med*, 2011; **365**: 410–21.
50. Martin NDT, Snodgrass GJAI, Cohen RD. Idiopathic infantile hypercalcaemia-a continuing enigma. *Arch Dis Child*, 1984; **59**: 605.
51. Rubin MR, Bilezikian JP, McMahon DJ, *et al*. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*, 2008; **93**: 3462–70.
52. Penniston KL, Tanumihardjo SA. The acute and chronic toxic effects of vitamin A. *Am J Clin Nutr*, 2006; **83**: 191–201.
53. Carroll MF, Schade DS. A practical approach to hypercalcemia. *Am Fam Physician*, 2003; **67**: 1959–66.
54. Saunders BD, Saunders EFH, Gauger PG. Lithium therapy and hyperparathyroidism: an evidence-based assessment. *World J Surg*, 2009; **33**: 2314–23.
55. Szalat A, Mazeh H, Freund HR. Lithium-associated hyperparathyroidism: report of four cases and review of the literature. *Eur J Endocrinol*, 2009; **160**: 317–23.
56. Nijenhuis T, Vallon V, Van Der Kemp AWCM, Loffing J, Hoenderop JGJ, Bindels RJM. Enhanced passive Ca2+ reabsorption and reduced Mg2+ channel abundance explains thiazide-induced hypocalciuria and hypomagnesemia. *J Clin Invest*, 2005; **115**: 1651–8.
57. Griebeler ML, Kearns AE, Ryu E, *et al*. Thiazide-associated hypercalcemia: incidence and association with primary hyperparathyroidism over two decades. *J Clin Endocrinol Metab*, 2016; **101**: 1166–73.
58. Iqbal AA, Burgess EH, Gallina DL, Nanes MS, Cook CB. Hypercalcemia in hyperthyroidism: patterns of serum calcium, parathyroid hormone, and 1,25-dihydroxyvitamin D_3 levels during management of thyrotoxicosis. *Endocr Pract*, 2003; **9**: 517–21.

59. Cardoso LF, Maciel LMZ, de Paula FJA. The multiple effects of thyroid disorders on bone and mineral metabolism. *Arq Bras Endocrinol Metabol*, 2014; **58**: 452–63.
60. Dhanwal D. Thyroid disorders and bone mineral metabolism. *Indian J Endocrinol Metab*, 2011; **15**(Suppl 2): S107–12.
61. Salvi M, Pedrazzoni M, Girasole G, *et al*. Serum concentrations of proinflammatory cytokines in Graves' disease: effect of treatment, thyroid function, ophthalmopathy and cigarette smoking. *Eur J Endocrinol*, 2000; **143**: 197–202.
62. Fujikawa M, Kamihira K, Sato K, Okamura K, Kidota S, Lida M. Elevated bone resorption markers in a patient with hypercalcemia associated with post-partum thyrotoxicosis and hypoadrenocorticism due to pituitary failure. *J Endocrinol Invest*, 2004; **27**: 782–7.
63. Montoli A, Colussi G, Minetti L. Hypercalcemia in Addison's disease: calciotropic hormone profile and bone histology. *J Intern Med*, 1992; **232**: 535–40.
64. Edafe O, Webster J, Fernando M, Vinayagam R, Balasubramanian SP. Phaeochromocytoma with hypercortisolism and hypercalcaemia. *BMJ Case Rep*, 2015; **2015**: DOI: 10.1136/bcr-2014-208657.
65. Bertelloni S, Baroncelli GI, Pelletti A, Battini R, Saggese G. Parathyroid hormone-related protein in healthy pregnant women. *Calcif Tissue Int*, 1994; **54**: 195–7.
66. Rey E, Jacob C-E, Koolian M, Morin F. Hypercalcemia in pregnancy—a multifaceted challenge: case reports and literature review. *Clin Case Rep*, 2016; **4**: 1001–8.
67. Richa CG, Saad KJ, Chaaban AK, El Rawas MS. A rare case of hypercalcemia-induced pancreatitis in a first trimester pregnant woman. *Endocrinol Diabetes Metab Case Rep*, 2018; **2018**: DOI: 10.1530/EDM-17-0175.
68. Ghaznavi SA, Saad NMA, Donovan LE. The biochemical profile of familial hypocalciuric hypercalcemia and primary hyperparathyroidism during pregnancy and lactation: two case reports and review of the literature. *Case Rep Endocrinol*, 2016; **2016**: 1–6.
69. Kolnick L, Harris BD, Choma DP, Choma NN. Hypercalcemia in pregnancy: a case of milk-alkali syndrome. *J Gen Intern Med*, 2011; **26**: 939–42.
70. Walker A, Fraile JJ, Hubbard JG. 'Parathyroidectomy in pregnancy'—a single centre experience with review of evidence and proposal for treatment algorithim. *Gland Surg*, 2014; **3**: 158–64.
71. Malekar-Raikar S, Sinnott BP. Primary hyperparathyroidism in pregnancy—a rare cause of life-threatening hypercalcemia: case report and literature review. *Case Rep Endocrinol*, 2011; **2011**: 520516.
72. Colloton M, Shatzen E, Wiemann B, *et al*. Cinacalcet attenuates hypercalcemia observed in mice bearing either Rice H-500 Leydig cell or C26-DCT colon tumors. *Eur J Pharmacol*, 2013; **712**: 8–15.
73. Nadarasa K, Bailey M, Chahal H, *et al*. The use of cinacalcet in pregnancy to treat a complex case of parathyroid carcinoma. *Endocrinol Diabetes Metab Case Rep*, 2014; **2014**: 140056.
74. Ioannis SP, Chrysoula LG, Aikaterini K, *et al*. The use of bisphosphonates in women prior to or during pregnancy and lactation. *Hormones*, 2011; **10**: 280–91.
75. Rey E, Jacob C-E, Koolian M, Morin F. Hypercalcemia in pregnancy—a multifaceted challenge: case reports and literature review. *Clin Case Rep*, 2016; **4**: 1001–8.
76. Felsenfeld AJ, Levine BS. Milk alkali syndrome and the dynamics of calcium homeostasis. *Clin J Am Soc Nephrol*, 2006; **1**: 641–54.
77. Zerwekh JE, Ruml LA, Gottschalk F, Pak CYC. The effects of twelve weeks of bed rest on bone histology, biochemical markers of bone turnover, and calcium homeostasis in eleven normal subjects. *J Bone Miner Res*, 1998; **13**: 1594–601.
78. Gaudio A, Pennisi P, Bratengeier C, *et al*. Increased sclerostin serum levels associated with bone formation and resorption markers in patients with immobilization-induced bone loss. *J Clin Endocrinol Metab*, 2010; **95**: 2248–53.
79. Kelly KJ. Acute kidney injury. In: McQueen C, ed. *Comprehensive Toxicology*, 3rd edn, 2017: 98–127.
80. Koh LKH. The diagnosis and management of hypercalcaemia. Ann Acad Med Singapore, 2003: **32**: 129–39.
81. Practice C, Legrand SB, Leskuski D, Zama I. Annals of internal medicine review narrative review: furosemide for hypercalcemia: an unproven yet. *Ann Intern Med*, 2008; **149**: 259–63.
82. Brain MJ, Roodenburg OS, McNeil J. Comparison of pre-filter and post-filter ionised calcium monitoring in continuous veno-venous hemodiafiltration (CVVHD-F) with citrate anticoagulation. *PLoS One*, 2017; **12**: e0189745.
83. Kawada K, Minami H, Okabe K, *et al*. A multicenter and open label clinical trial of zoledronic acid 4 mg in patients with hypercalcemia of malignancy. *Jpn J Clin Oncol*, 2005; **35**: 28–33.
84. Major P, Lortholary A, Hon J, *et al*. Zoledronic acid is superior to pamidronate in the treatment of hypercalcemia of malignancy: a pooled analysis of two randomized, controlled clinical trials. *J Clin Oncol*, 2001; **19**: 558–67.
85. Markowitz GS, Fine PL, Stack JI, *et al*. Toxic acute tubular necrosis following treatment with zoledronate (Zometa). *Kidney Int*, 2003; **64**: 281–9.
86. Sternlicht H, Glezerman IG. Hypercalcemia of malignancy and new treatment options. *Ther Clin Risk Manag*, 2015; **11**: 1779–88.
87. Thosani S, Hu MI. Denosumab: a new agent in the management of hypercalcemia of malignancy. *Futur Oncol*, 2015; **11**: 2865–71.
88. Morony S, Warmington K, Adamu S, *et al*. The inhibition of RANKL causes greater suppression of bone resorption and hypercalcemia compared with bisphosphonates in two models of humoral hypercalcemia of malignancy. *Endocrinology*, 2005; **146**: 3235–43.
89. Fountas A, Andrikoula M, Giotaki Z, *et al*. The emerging role of denosumab in the long-term management of parathyroid carcinoma-related refractory hypercalcemia. *Endocr Pract*, 2015; **21**: 468–73.
90. Vellanki P, Lange K, Elaraj D, Kopp PA, Muayed M El. Denosumab for management of parathyroid carcinoma-mediated hypercalcemia. *J Clin Endocrinol Metab*, 2014; **99**: 387–90.
91. Hu MI, Glezerman IG, Leboulleux S, *et al*. Denosumab for treatment of hypercalcemia of malignancy. *J Clin Endocrinol Metab*, 2014; **99**: 3144–52.
92. Diel IJ, Body JJ, Stopeck AT, *et al*. The role of denosumab in the prevention of hypercalcaemia of malignancy in cancer patients with metastatic bone disease. *Eur J Cancer*, 2015; **51**: 1467–75.
93. Nissenson RA. Parathyroid hormone and parathyroid hormone-related protein. In: Marcus R, Feldman D, Cauley JA, *et al.*, eds. *Osteoporosis*, 4th edn. Waltham, MA: Elsevier/Academic Press, 2013: 259–81.

94. Marcocci C, Chanson P, Shoback D, *et al*. Cinacalcet reduces serum calcium concentrations in patients with intractable primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2009; **94**: 2766–72.
95. Silverberg SJ, Rubin MR, Faiman C, *et al*. Cinacalcet hydrochloride reduces the serum calcium concentration in inoperable parathyroid carcinoma. *J Clin Endocrinol Metab*, 2007; **92**: 3803–8.
96. Asonitis N, Kassi E, Kokkinos M, Giovanopoulos I, Petychaki F, Gogas H. Hypercalcemia of malignancy treated with cinacalcet. Endocrinol Diabetes Metab Case Rep, 2017; **2017**: 17–0118.

4.3

Primary Hyperparathyroidism

John P. Bilezikian

Introduction

Primary hyperparathyroidism (PHPT), the most common cause of hypercalcaemia, is characterized by excessive parathyroid hormone (PTH) secreted from one or more of the four parathyroid glands [1–3]. It is virtually always a benign disorder with 80% of patients harbouring a single parathyroid adenoma. Within the past 50 years, the disease has undergone a transformation in its clinical presentation from a disorder that was virtually always symptomatic to one that is discovered incidentally in the course or routine biochemical screening. Target organs such as the skeleton and the kidneys, while still a focus of the prudent evaluation, requires application of imaging and other technologies in order to discern what used to be clear-cut manifestations. When these technologies are applied even to patients who are discovered incidentally, it is evident the PHPT still has the potential to be a pervasive disease. This potential, however, is part of a variable natural history that ranges from an asymptomatic disorder to one with overt symptomatology. Even more recently, we recognize that PHPT, as defined by elevated PTH levels in the absence of any other aetiologies, can present with normal total and ionized serum calcium concentrations, giving rise to the phenotype known as normocalcaemic PHPT [4]. With these different clinical presentations, the challenge has been to apply guidelines with which decisions for surgical or medical management can be reasonably offered and based upon the best evidence. These guidelines have been regularly updated over the past 25 years [5–9].

From a historical perspective, it is interesting to note that in the 1930s, PHPT was so rare that when encountered, it often led to a case report [10]. Almost a century later, a casual literature search turns up over 6000 published papers in the 2000s alone. Another key ramification of ongoing interest in PHPT has been interest in the PTH molecule itself, with ongoing insights into its pleotropic cellular and molecular mechanisms of action [11, 12].

Epidemiology

PHPT is seen most often in postmenopausal women with a gender difference of 3–4:1. The prevalence of the disease varies depending upon country and representation by race. In the United States, prevalence figures centre around 0.86% [13]. African Americans are more prone that Caucasian women [14]. Prevalence figures have been reported to be lower (0.4%) and much higher (11%), the latter representing cohorts of patients with normocalcaemic PHPT [15–17]. Epidemiology is also influenced by medical practices, such as biochemical screening. When screening is a routine aspect of medical care, the prevalence of PHPT is higher, because the serum calcium is typically included in the multichannel screening test and patients are discovered in that context. When biochemical screening is commonly used in countries, as demonstrated in North America and Western Europe, the disease is common [3]. When biochemical screening is not a routine feature of medical care, such as in many locations in India, PHPT is uncommon. Screening, however, is not a complete explanation for the increasing prevalence of PHPT. In the United States, for example, after screening had become well established, the incidence of PHPT has continued to increase [14]. Perhaps, more proactive approaches to skeletal diseases, particularly in the context of mild elevations in the serum calcium, could account, at least in part, for this observation [14, 18, 19]. It is also noteworthy, but not surprising, that the incidence of PHPT increases in countries that adopt biochemical screening, as in China and in Latin America [20–23].

In most patients, there is no clear-cut aetiology to the disease although a history of childhood face and neck radiation can be obtained in a small percentage of individuals [24]. The genetics of PHPT are described next.

Diagnosis

PHPT is almost always a benign expression of excessive production and secretion of PTH. In 80% of adult patients, the cause is a single parathyroid adenoma while in the remaining 20%, all four parathyroid glands are involved with a hyperplastic histology [2]. Double parathyroid adenomas are seen infrequently [25]. Classically, PHPT is a hypercalcaemic disorder. In patients whose serum albumin is low, the total serum calcium measurement will read lower than it would if the level were 4.0 g/dl. The serum calcium is adjusted upwards by 0.8 mg/dl for every gram by which the albumin is below 4.0 g/dl. In centres where the ionized calcium is routinely measured, it will be elevated and, of course, no adjustment for albumin is needed. In PHPT, the PTH level is elevated but not infrequently, the PTH level will be within the normal reference range for the assay. The hypercalcaemic state defines that 'normal' PTH level as abnormal because normal parathyroid tissue would instantly reduce PTH secretion to undetectable levels if the cause of the hypercalcaemia were not parathyroid hormone dependent. Theoretically, any detectable level of PTH in the context of frank hypercalcaemia is compatible with a diagnosis of PHPT. Since second generation PTH assays, that are most commonly used, can detect inactive circulating fragments, it is helpful to set a lower limit of detectability below which one would look for other causes of hypercalcaemia. Well-documented PHPT has been described with levels of PTH as low as 20–25 pg/ml, assuming a normal reference range of 10–65 pg/ml. A PTH concentration of 20 pg/ml would seem to be a reasonable lower limit below which non PTH-dependent hypercalcaemia becomes more plausible. Bear in mind that PTH levels increase with age. While reference ranges have not been completely established by age, in those under the age of 45, an upper limit of 45 pg/ml, not 65 pg/ml would seem to be reasonable.

The recent controversy over normal reference ranges for 25-hydroxyvitamin D [26, 27] has stimulated discussion about normal reference ranges for PTH. The generally accepted reference ranges for PTH, as noted earlier, do not take into account the levels of 25-hydroxyvitamin D. Since there is a linear, inverse relationship between 25-hydroxyvitamin D levels and PTH [28], the reference range for PTH may well be lower in a vitamin D replete population [29].

In addition, the PTH assay can read low, in certain assays, by virtue of moieties that interfere with the assay, such as biotin [30]. If a patient who reports the use of biotin is rechecked after discontinuing it, the level PTH may become more compatible with the diagnosis of PHPT.

Thiazides and lithium can be associated with PTH-dependent hypercalcaemia, that is with elevated calcium and PTH levels [31, 32]. Most of these patients have PHPT and, if one were able to discontinue the thiazide or lithium, the serum calcium would remain elevated in the clear majority of patients. If there is doubt, thiazides can be substituted for other agents, but stopping or substituting lithium should only be done under careful supervision.

In the differential diagnosis of PTH-dependent hypercalcaemia, familial hypocalciuric hypercalcaemia (FHH) is often considered. This is clearly appropriate because one does not want to mistake PHPT for FHH, particularly if parathyroid surgery is being considered for the patient. The consideration of FHH is highly relevant in young patients under the age of 30 because the disease has high penetrance by then. If there is family history of hypercalcaemia, particularly if it includes members who have had unsuccessful parathyroid surgery one should consider FHH [33]. These patients represent a very small proportion of the hyperparathyroid universe. Since most patients with PHPT are postmenopausal women whose serum calcium was known to be normal years before, the development of hypercalcaemia in their adult years virtually rules out FHH. It should be emphasized that FHH is a rare disorder and, from a statistical point of view, a postmenopausal woman who develops hypercalcaemia and elevated levels of PTH does not have FHH. A key point of distinction between PHPT and FHH, as the name implies, is that urinary excretion of calcium is very low. When compared to the creatinine clearance, the ratio of the calcium clearance:creatinine clearance (CCCR) is less than 0.01. Patients with PHPT can also have a low urinary excretion of calcium because PTH is a calcium-conserving hormone. For any amount of calcium filtered at the glomerulus, under the influence of PTH, there will be greater tubular reabsorption and less calcium excretion. A CCCR less than 0.01 can, therefore, be compatible with PHPT particularly if the patient is on a low calcium diet. If there is doubt, genetic testing for an inactivating mutation in the calcium-sensing receptor can be done readily.

Patients with PHPT can demonstrate a normal albumin-corrected serum calcium level from time to time. As noted just now, if the albumin is low, the total serum calcium can read normal. If the patient has severe vitamin D deficiency, the serum calcium can read normal. If the patient has PHPT, correction of the vitamin D deficiency can unmask the hypercalcaemia of PHPT.

Variants of Classical Primary Hyperparathyroidism

Normocalcaemic primary hyperparathyroidism. The entity now known as normocalcaemic PHPT was discovered over 15 years ago [34]. During evaluation, but not necessarily long-term, these patients demonstrate normal corrected serum calcium levels [35]. Discover of this phenotype of PHPT is due, in large part, to the proactive approach to low bone mass with regard to testing in which the PTH is measured even when the serum calcium is normal. When these patients are confirmed to have normal serum calcium levels and elevated PTH, a consideration of secondary causes for an elevated PTH level becomes mandatory. One of the most important secondary causes to rule out is vitamin D deficiency. The Institute of Medicine has stated that vitamin D insufficiency can be defined when the 25-hydroxyvitamin D level is less than 20 ng/ml and that levels between 20 and 30 are acceptable. It stated, furthermore, that levels from 30 to 20 ng/ml are not reliably associated with elevated levels of PTH [27]. It is also true, though, that some patients will show an inflection upwards in their PTH levels when the 25-hydroxyvitamin D level is between 30 and 20 ng/ml [36]. In order to be sure that vitamin D insufficiency is not contributing to

the elevated PHT, a level of more than 30 ng/ml seems reasonable. In some situations, a goal might be to set the 25-hydroxyvitamin D level at 40 ng/ml, on the chance that this form of PHPT may be associated with an element of vitamin D resistance [37].

When considering the diagnosis of normocalcaemic PHPT, other secondary causes for an elevated PTH, besides vitamin D insufficiency, should be considered. Renal insufficiency with an eGFR of less than 60 cc/min can be associated with an elevated PTH level, but PTH levels do not tend to show levels in the abnormal range until renal function has declined to chronic kidney disease has reached stage 3B (i.e. <45 cc/min).

Medications, as noted earlier with regard to thiazide diuretics and lithium, can be associated with a high PTH. While not generally appreciated, bisphosphonates and denosumab can be associated with elevated levels of PTH [38, 39]. Patients with other metabolic bone disease such as Paget's disease can occasionally show elevated levels of PTH.

The diagnosis of normocalcaemic PHPT depends upon ruling out secondary causes of elevated PTH levels. Even with due diligence, one may miss a secondary cause such as a subclinical malabsorption syndrome.

The discussion of normocalcaemic PHPT should take into account also the possibility that the patient doesn't have a disease at all but that the PTH level is merely representing the fringe of the normal distribution curve. The normal distribution curve captures about 95% of the normal population with about 2.5% falling either above or below those defined limits. Such patients are not necessarily abnormal. Another consideration in the discussion of normocalcaemic PHPT is the 'normal' serum calcium level. Is the serum calcium normal for that patient? For an analyte like calcium, the control in a given individual is much tighter than it is for the entire population. If patients have spent most of their years with serum calcium levels averaging 9.0 mg/dl and develops PHPT, the serum calcium might increase to10.0 mg/dl. Without a history of those patients' earlier serum calcium levels, the measured calcium appears to be normal. If one had a history of those patients' earlier calcium levels, the levels measured as 'normal' would not be normal for that patient. The 1 mg/dl increase in serum calcium would still be within the normal population range but for that given patient, it is decidedly high. We describe the patient as one with normocalcaemic PHPT but in fact the patient is hypercalcaemic with regard to her own calcium homeostasis.

Finally, a form of abnormal parathyroid tissue has been described with a normal serum calcium and PTH concentration made only by pathological examination of abnormal parathyroid tissue has been proposed [40, 41].

Acute Primary Hyperparathyroidism

The onset of marked hypercalcaemia and elevated levels of PTH can occur in patients with a history of mild hypercalcaemia. Typically, mild hypercalcaemia due to PHPT is not likely to lead to acute, symptomatic hypercalcaemia, the incidence of which is less than 1% [42–45]. If there is an antecedent history of mild hypercalcaemia, the possibility of acute PHPT becomes likely. The inciting factors are not understood but an intercurrent illness in which the patient does not maintain good hydration and is bed ridden is a typical scenario [45]. Haemorrhage or infarction of the offending parathyroid gland is almost never seen at the time of parathyroidectomy. A previous history of mild hypercalcaemia is not always obtained, leading to the possibility that the patient has parathyroid cancer. While this is certainly a possibility, the rarity of parathyroid cancer makes acute primary hyperparathyroidism statistically more likely.

Parathyroid Carcinoma

The differential diagnosis of marked hypercalcaemia and very high levels of PTH usually comes down to either acute PHPT or parathyroid cancer. Because parathyroid cancer is so unlikely, the patient presenting with acute symptomatic hypercalcaemia is more likely to have a benign parathyroid adenoma. Among patients with PHPT, malignant disease constitutes less than 0.5%. In parathyroid cancer, the serum calcium tends to high (>14 mg/dl) and associated with hypercalcaemic symptoms, along with markedly elevated levels of PTH. It is not unusual for levels of PTH to be 20 times normal. It is exceedingly rare for a parathyroid cancer to be non-secretory. There are other differentiating features when parathyroid cancer becomes a consideration. For example, the carcinoma can present with a neck mass that is easily palpable. They are 40–50 years old, on average, about a decade younger than patients with benign disease [46, 47]. Different from its benign counterpart, there is no difference in incidence by gender. The natural history of parathyroid cancer can be cure, if surgery if successful at the outset. But many patients are not cured but follow, rather, a slow, indolent clinical course. Patients will typically live for several decades after the diagnosis is made. Recurrent parathyroid cancer can resurface 20 years later [48]. While parathyroid cancer is typically a sporadic disease, it has seen in association with the hyperparathyroid-jaw tumour syndrome [49]. That presentation and other aspects of parathyroid cancer genetics are covered elsewhere in this chapter.

Hyperparathyroidism due to true ectopic PTH production. While malignancies have been associated with hypercalcaemia, it is very rare for the tumour to be producing authentic PTH [50]. Much more common is the situation in which the patient has both a malignancy and PHPT. The production of parathyroid hormone related peptide (PTHrP) by the tumour is very well described but this peptide does not cross-react in the commercially available immunoassays for PTH. Thus, in this setting, the PTH level will be undetectable.

Hereditary Primary Hyperparathyroidism: Forms and Genetics

The Multiple Endocrine Neoplasia (MEN) syndromes, known classically as type 1 and type 2, are autosomal disorders. Of the glands that can be involved in MEN 1 or 2, PHPT is the most common and often the first to be involved. By the age of 50, penetrance is virtually complete with rare de novo cases seen thereafter [51]. In MEN 1, the other common glands that are involved include the pancreas and the anterior pituitary. In MEN 2, PHPT is seen with medullary thyroid cancer and pheochromocytoma. In the syndrome known as MEN 2b or MEN 3, PHPT is not seen. Along with medullary thyroid cancer and pheochromocytoma, MEN 2b (MEN 3), can feature mucosal neuromas, autonomic ganglia dysfunction of the gastrointestinal (GI) tract and a marfanoid habitus. In MEN 4, PHPT is seen in association with anterior pituitary, gonadal, adrenal, or renal tumours. PHPT, which can be a frank parathyroid malignancy or as part of the hyperparathyroidism-jaw tumour (HPT-JT) syndrome [52].

These genetic disorders are of great interest, in part, because they have helped to elucidate genetic controls of normal endocrine glands. But, they constitute a small percentage of patients with PHPT. One is advised to pursue a genetic aetiology only when setting is suggestive such as PHPT in those less than 30 years old or in those with a family history of PHPT or other hypersecretory endocrine syndromes. Familial isolated PHPT has also been well-described in the absence of any other endocrine disorder [53].

The tumour suppressor gene known as *MENIN* is inactivated in MEN 1 [51, 54]. Mutations in the RET proto-oncogene are responsible for MEN2 and MEN 2b (MEN3) [55]. The gene that controls the cyclin-dependent kinase inhibitor (p27kip1), CKNK1B, is affected in MEN 4. The *HRPT2* gene (CDC73) is abnormal in the HPT-JT syndrome. This gene directs the production of parafibromin which can be immunochemically detected in parathyroid tissue [56].

Other familial syndromes cannot be neatly classified. Familial isolated hyperparathyroidism (FIH) [55], for example, can be associated with multiglandular syndromes [57, 58]. It will be of great interest when these genes are identified and compared to regulatory aspects of the genes that have already been characterized in the familial syndromes of PHPT [59]. Thakker *et al.* presents a well-reasoned approach to questions related to the genetic basis for the hyperparathyroid states [54].

Familial hypocalciuric hypercalcaemia (FHH). The clinical aspects of FHH have already been discussed. The genetic basis of this disease is inactivation of the calcium-sensing receptor (CaSR).

Neonatal severe primary hyperparathyroidism (NSHPT). This rare but life-threatening presentation in newborns is identified by marked hypercalcaemia, very high levels of PTH, hypotonia, and respiratory distress. A homozygous mutation of the CaSR gene is responsible for this disease [60].

Molecular Pathogenesis

The search for genetic abnormalities has focused upon the familial syndromes described earlier.

The two genes that have been most clearly implicated in the sporadic (non-familial) forms of PHPT are cyclin D1 [61] and the gene associated with MEN1 [62]. Thus, both familial and sporadic forms of PHPT can be associated with the MEN 1 gene. With regard to sporadic PHPT, some studies have reported that as many as 25–40% of parathyroid adenomas are due to allelic loss of chromosome 11 markers [63, 64]. In the particular cases of parathyroid cancer and the hyperparathyroidism = jaw–tumour syndrome, the *HRPT2* gene (also known as CDC73), has been implicated [61].

Pathology

The single, benign parathyroid adenoma is the most common abnormality in PHPT. Multiple parathyroid adenomas are much less common, with an incidence less than 5% [25]. While double adenomas can be found, it is important to ascertain that the other glands are normal because some of these patients actually have four-gland disease [65]. Even if normal parathyroid tissue has been documented at the time of the parathyroid surgery for double adenomas, there is always the possibility that the patient will in time develop hyperfunction of the remaining parathyroid glands. In these situations, time is the key determinant and, thus, monitoring is highly recommended. Parathyroid adenomas are notoriously variable in location. While they are usually found at the inferior–anterior and superior–posterior poles of the thyroid gland, they can be found almost anywhere in the neck. Some typical 'hiding places' for adenomas in ectopic locations include the thyroid gland parenchyma, itself, and the superior mediastinum. In the mediastinum, the gland can be found within the thymus. More unusual locations for ectopic parathyroid glands include the lateral neck, the carotid sheath, behind the oesophagus and even in the oropharynx.

Four-gland parathyroid hyperplasia is present in approximately 15% of patients with primary hyperparathyroidism. While one immediately considers a familial syndrome, in this context, only about 30% of patients will be shown to have the familial form of PHPT [66]. Before localization procedures and surgery, itself, no clinical features raise suspicion for multiglandular disease. Invariably, the pathology of PHPT is benign, overactive tissue with the malignant parathyroid gland found in fewer than 0.5% of patients. In the case of parathyroid cancer, clinical clues are often apparent, such as marked hypercalcaemia and very high levels of PTH, along with a palpable neck mass. The pathological features raise suspicion for malignancy because the tissue may have invaded local tissues and show histological features of cancer. On the other hand, parathyroid cancers can present pathologically as rather benign tissue. Careful examination of the pathological specimen in parathyroid cancer will often show the cardinal features of malignancy, namely tissue invasion and local or distant metastases. Parathyroid cancers present as functional tumours, secreting PTH.

Biochemical Features

The biochemical features of PHPT in most publications depict the mild hypercalcaemic form of the disease as shown in Table 4.3.1. The serum calcium concentration is usually within 1 mg/dl of the upper limits of normal. The serum phosphorus is rarely frankly below normal but values usually are in the low normal range [67]. Circulating markers of bone turnover are often in the upper quartile of the normal range, but can be frankly elevated. The specific markers of bone turnover that are readily available and can be helpful in assessing the activity of the disease are the bone formation markers: bone-specific alkaline phosphatase activity and osteocalcin—and the bone resorption markers—serum CTx and urinary NTx [68].

One has to go back to almost 50 years ago to appreciate the prescient work of Lumb and Stanbury who proposed that the clinical manifestations of PHPT are worsened by vitamin D deficiency [69]. This association has since been substantiated by several studies [70, 71]. Even in mild, asymptomatic PHPT, indices of disease activity are more apparent when the 25-hydroxyvitamin D concentration is low [72]. This association may be more difficult to detect in view of the fact that vitamin D deficiency in PHPT [73, 74], is not as widespread as it used to be. This is because vitamin D supplements are now commonly used [75]. The hypothesis that parathyroid disease is worse when vitamin D deficiency is present describes a stimulus for further stimulation of PTH synthesis by the abnormal parathyroid gland(s). Stein *et al.* has correlated low 25-hydroxyvitamin D levels with higher PTH level [76]. The

Table 4.3.1 Changes in biochemical indices in primary hyperparathyroidism over the past several decades. The data represent cohorts from earlier and more recent series

Index	1984–1991 (n = 103)	2000–2014 (n = 100)
Calcium (mg/dl)	10.6 ± 0.06	10.7 ± 0.6
PTH (pg/ml)	127 ± 69	85 ± 48[1]
25-hydroxyvitamin D (ng/ml)	23 ± 10	29 ± 10[1]
1,25-dihydroxyvitamin D (pg/ml)	57 ± 20	69 ± 24[1]
Urinary calcium (mg/24 hrs)	229 ± 119	250 ± 144
P <0.005		

Adapted with permission from Silverberg SJ, Shane E, Dempster DW, Bilezikian JP. The effects of vitamin D insufficiency in patients with primary hyperparathyroidism. *Am J Med* 1999;107:561–56. Copyright © 1999 Excerpta Medica Inc. (ref 72) with data from Walker MD, Cong E, Lee JA, Kepley A, Zhang C, McMahon DJ, et al. Low vitamin D levels have become less common in primary hyperparathyroidism. *Osteoporosis Int* 2015;26:2837–2843. (ref 75).

microstructural and densitometric profile is not as consistent as the biochemical correlates [77–79].

Levels of 25-hydroxyvitamin D will vary according to the country's nutritional status. In the United States, a prototypical country in which the modern profile of PHPT the most common, 25-hydroxyvitamin D levels no longer tend to be below 30 ng/ml. Over the past 15 years, with the population routinely taking vitamin D supplements, the average 25-hydroxyvitamin D from the centre in New York has increased from 21 ng/ml to 36 ng/ml [75]. 1,25-dihydroxyvitamin D levels are no longer routinely measured in PHPT but as per previous experience, the level of this active form of vitamin D would be expected to be in the upper range of normal or frankly elevated. As previously shown, elevations in 1,25-dihydroxyvitamin D concentration can be seen in as many as 25% of the hyperparathyroid population [80]. This observation is due to the catalytic effect of PTH on renal 1-alpha hydroxylase that converts 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D. PTH conserves renal tubular handling of calcium and, thus, for any level of the serum calcium, patients with PHPT will show reduced concentrations. The reason why hypercalciuria can be seen in as many as 40% of individuals with PHPT is because the filtered calcium load overwhelms the conserving capacity of the kidneys.

Target Organ Features

The Skeleton

With the change in clinical profile of PHPT to a milder hypercalcaemic disorder, it is not surprising that the frequency of disease-specific radiological manifestations would decline. The historic Cope series reported a prevalence of 23% [81] while the series of Silverberg *et al.*, coming several decades later, reported a prevalence of less than 2% [82]. With the advent of technologies that permit evaluation of the skeleton in ways that are more sensitive than the routine skeletal X-ray, we now appreciate that skeletal involvement, even in mild 'asymptomatic' PHPT, is pervasive.

Bone densitometry. Bone mineral densitometry (BMD) by dual energy X-ray absorptiometry (DXA) came of age as a major clinical tool when PHPT was being recognized in the western world as primarily a biochemical disorder without classic radiographic findings. DXA made it possible to appreciate skeletal involvement even when radiographic signs were absent.

By DXA, one can measure three sites that each represents different proportions of trabecular and cortical bone. The lumbar spine is predominantly trabecular bone while the distal third radius is predominantly cortical bone. The hip is comprised of an even proportion of cortical and trabecular bone. In a disorder characterized by excessive PTH, the cortical compartment would be expected to be compromised preferentially. By DXA, this was shown by the distal third radius being preferentially reduced [83]. The lumbar spine, on the other hand, was the site least involved with the hip regions showing reductions mid-way between the lumbar spine and the distal third radius (Figure 4.3.1). This pattern of involvement necessitates three-site DXA in all patients with PHPT. If only the lumbar spine and hip are measured, an incomplete picture emerges that, in fact, could be misleading. It should be recognized, though, that other patterns of skeletal involvement have been described in PHPT, from preferential reduction of the lumbar spine to equivalent involvement of all three sites to completely normal bone density [84].

One of the tenets of DXA is that it predicts fracture risk. There is certainly true for osteoporosis and was logically assumed to be the case for PHPT. DXA would predict that fractures in PHPT would be more common at non-vertebral sites which are predominantly cortical bone. This expectation was brought into doubt first by Khosla *et al.* [85] and a decade later by Vignali *et al.* [86]. Both reports demonstrated that vertebral fractures were also seen in PHPT with a frequency that argued against trabecular bone being protected in

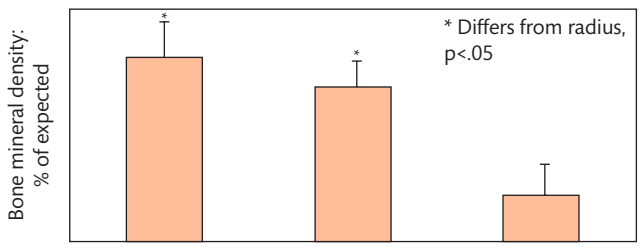


Figure 4.3.1 The pattern of bone loss in primary hyperparathyroidism. A typical pattern of bone loss is seen in asymptomatic patients with primary hyperparathyroidism. The lumbar spine is relatively well preserved while the distal radius (1/3 site) is preferentially affected.

Reproduced with permission from Silverberg SJ, Shane E, DeLaCruz L, Dempster DW, Feldman F, Seldin D, Jacobs TP, Siris ES, Mafferty M, Parisien MV, Lindsay R, Clemens TL, Bilezikian JP: Skeletal disease in primary hyperparathyroidism. *J Bone Mineral Res* 1989;4:283–91. Copyright © 1989 American Society for Bone and Mineral Research.

PHPT. Imaging technology advanced [87, 88] with high resolution peripheral computed tomography (HRpQCT), showing that both cortical and trabecular bone are in fact affected in PHPT [89–91]. When the HRpQCT image was assessed further by individual trabecular segmentation analysis, Stein, Silva *et al.* [89] showed that microstructural orientation of trabeculae was suboptimally configured for bone strength.

While HRpQCT gave insight into skeletal microstructure that not even histomorphometry was able to discern [92], it is a research tool. The practising physician does not have access to HRpQCT. To meet a practical need for assessment of microstructure in medical practice, yet another advance in imaging technology was developed, namely the trabecular bone score (TBS). TBS is a semiquantitative textural analysis of the lumbar spine image by DXA [93, 94]. Anyone with TBS software can apply it to the DXA image. The textual analysis distinguishes between homogeneity and heterogeneity of the image. The homogenous image is akin to normal bone while the markedly heterogeneous image is akin to osteoporotic bone. As is the case for DXA, TBS of the lumbar spine DXA image is measuring predominantly trabecular bone. Silva *et al.* and others have documented lower mean TBS scores in PHPT [95, 96].

With these advances in imaging technology applied to the form of PHPT known as 'asymptomatic', It is now apparent that both cortical and trabecular compartments of bone can be affected in the absence of radiographic manifestations or even reductions in lumbar spine BMD by DXA. The observations are consistent with the clinical documentation that fracture risk is increased, overall, in PHPT.

Renal Involvement

The kidneys, along with the skeleton, constitute the major target organ for complications of PHPT. Similar to the decline in overt skeletal disease, the incidence of kidney stones has declined from approximately 60%, when biochemical screening was not routine, to more recent series that document an incidence of kidney stones in fewer than 20% of patients [20]. When applying sensitive renal imaging technologies, such as computed tomography (CT) and ultrasound, some studies have confirmed the incidence of stones and/or nephrolithiasis at about 20% but work of Cristiani *et al.* observed a much higher incidence of 55% [97–99].

Clearly, the composition of urine is a likely predisposing factor in the development of kidney stones. This extends beyond elevated urinary calcium excretion. Peacock has summarized these other aspects of stone risk factors in PHPT [100]. Schillitani *et al.* has raised the possibility that the calcium receptor gene, in terms of specific polymorphisms, might be an important contributing factor [101].

The persistence of renal involvement in PHPT has led to a consensus that renal imaging should be an integral aspect of its evaluation. The imaging modalities can be ultrasonography, CT, or routine abdominal X-ray. A 24-hour urinary calcium, on a normal calcium intake, should also be part of every evaluation. Since, the urinary calcium excretion, alone, is not always indicative of stone risk in PHPT [102], a complete urinary stone risk profile should be undertaken when hypercalciuria is apparent (100).

In addition to concerns about nephrolithiasis and nephrocalcinosis, creatinine clearance may be reduced in PHPT. While the mechanisms by which the creatinine clearance may be reduced in PHPT are not clear, reductions below 60 ml/min/1.73 m^2 are seen, in the absence of any other explanation for a reduction in creatinine clearance. Thus, a reduction in renal function can be and often is ascribed to this disorder. A concern about reduced renal function in PHPT, even if it isn't related specifically to this disorder, raises concern because it is a potential stimulus for secondary increases in PTH, thus aiding and abetting the primary hyperparathyroid process. It would seem from the work of Walker *et al.*, that clearance values above 30 cc/min are unlikely to contribute independently to higher PTH levels [103]. This point does not obviate another one, namely that by dynamic histomorphometric assessment of bone biopsies, clearance values below, <60 cc/min are associated with evidence for more active parathyroid disease [104]. Thus, renal involvement in PHPT can take the forms of hypercalciuria, nephrolithiasis, nephrocalcinosis, and/or reduced renal function [105].

Neuromuscular Dysfunction

The apparent change in the clinical profile of PHPT has had an impact on the incidence of recognizable neuromuscular dysfunction. Once, it was a regular feature of the disease [106] but now it is seen only when the disorder presents in its overtly symptomatic fashion. With the emergence of the mild form of the disease, Turken *et al.* [107] found no evidence either clinically or electromyographically for neuromuscular dysfunction. Sural nerve conduction velocity was reduced in the study of Diniz *et al.* but as in the study of Turken *et al.* these patients had no clinical neurological manifestations [108].

Gastrointestinal Manifestations

The three major gastrointestinal aspects of PHPT are all uncertain from an etiological view. Glucose intolerance has been seen [109] but the association is far from established [110, 111]. Similarly, peptic ulcer disease occurs with a frequency that approximates the disease in the general population, about 10%. The only exception to this lack of association is in those whose PHPT occurs with MEN1. The frequency of peptic ulcer disease in MEN1 can approach 40% [112]. Some mechanistic credence can be attributed to the relationship between PHPT and pancreatitis because it is well known that hypercalcaemia can be causative. However, when hypercalcaemia is mild, as it often is, most series no longer report this relationship to be a significant one [113, 114].

Non-Classical Manifestations of Primary Hyperparathyroidism

Neurobehavioral and Neurocognitive Features

While viewed as non-classical manifestations of PHPT, these features such as easy fatigability, anxiety, poor concentration, cognitive decline, and reduced quality of life are among the most vexing [115, 116]. The clinician is often faced with these complaints in the context of many other chronic diseases. Thus, they lack specificity. Another problem is that they are very hard to quantitate with reproducible metrics that can tract changes after parathyroid surgery. The confounding variables among the reports that have attempted to link neurocognitive and neuropsychiatric features to PHPT include

study design (observational vs. prospective), selection bias, small numbers of study subjects, heterogeneity of bassline manifestations, short testing intervals after parathyroid surgery, and problems with control groups selected for comparison.

Even with newer testing instruments to test quality of life and cognitive function, some studies have shown improvements while others have not [43]. As a result of these inconclusive studies, it is difficult to make any recommendations regarding surgical intervention in these patients who otherwise do not meet the generally agreed upon guidelines for parathyroid surgery in this disease.

Cardiovascular Manifestations

The cardiovascular manifestations of primary hyperparathyroidism used to be heralded by hypertension, but now, with many patients demonstrating only modest hypercalcaemia, hypertension is no longer regarded as a feature of the disease [117]. The problem with a common disease such as hypertension is that to demonstrate a clear association, one would have to show unequivocally that after parathyroid surgery, the condition is ameliorated. The older published literature provided some support for this link [118–120], but it is much more likely now not to see a difference in blood pressure after parathyroid surgery [120, 121].

Coronary artery disease was linked to PHPT when the disease was associated with higher calcium levels than are seen today [122]. Most current studies do not show that patients with PHPT have a greater incidence of coronary calcification beyond those acknowledged risk factors for coronary calcification [123, 124]. In a similar manner, calcifications of the myocardium and cardiac valves are seen only when patients have marked hypercalcaemia [125, 126].

Left ventricular mass index (LVMI) is another focus for studies that have attempted to link PHPT to the cardiovascular system [127, 128]. While some studies have associated the PTH level itself, and not the blood pressure, with LVMI [126–130], other studies have not been confirmatory when other cardiovascular risk factors are taken into account [131, 132]. Again, conflicting data emerge with regard to regression of LVMI after parathyroidectomy [127–130, 133–135].

Calcium concentration and carotid plaque thickness may be related [136] but as shown by Walker *et al.* [137] after adjustment for cardiovascular risk factors, parathyroidectomy was not associated with improvement [138].

Vascular dysfunction has been observed both in severe PHPT and in patients whose serum calcium is only mildly elevated [139–143].

Thus, despite a plethora of data regarding PHPT and cardiovascular manifestations, the results are inconclusive with regard to all the indices that have been measured. This is not to say, however, that the symptomatic form of PHPT is not be associated with cardiovascular manifestations. In fact, it is probably only in the symptomatic cohorts of patients with PHPT that one is likely to observe these manifestations.

Malignancy

Reports associating cancers with PHPT [144, 145] are subject to qualification due to the likelihood of selection bias. Hypercalcaemia usually triggers a search for malignancy since in its differential diagnosis, cancer is high on the list. Another aspect of selection bias relates to the common discovery of thyroid malignancy in patients who undergo parathyroidectomy. Thyroid cancer probably would not have been discovered were it not for the parathyroid surgery. In contrast, Wermers *et al.* have reported death due to cancer is reduced in PHPT with a hazard ratio of 0.58 [146].

Mortality

Mortality rates in PHPT appear to follow the dictum that the more severe the disease is the more likely mortality will be increased. In cohorts of mild hypercalcaemia, mortality does not seem to be increased [147] while more symptomatic cohorts show an increase in mortality [148–149]. Supporting this expectation, mortality rates do correlate with the level of the serum calcium as well as parathyroid gland weight [150]. The experience from the Mayo Clinic also attributed higher mortality only in the quartile that had the highest serum calcium level [146]. However, an epidemiological study has reported increased mortality and morbidity in 'mild hyperparathyroidism' [144]. Nevertheless, with the common phenotype of mild hypercalcaemia seen in many parts of the world, it would appear that mortality rates are not increased in PHPT.

The Variable Clinical Expressions of Primary Hyperparathyroidism

The symptomatic disorder. Primary hyperparathyroidism was ushered into clinical recognition as a symptomatic disorder in the period from 1930 and 1970, It was truly a disease of bones and stones with serum calcium levels that were invariably over 12 mg/dl, and often much higher [151].

The asymptomatic disorder. In the 1970s, when biochemical screening became widespread in many countries, the form of PHPT that we call 'asymptomatic' began to dominate the clinical landscape. Identifying these patients with mild hypercalcaemia and no overt bone or stone disease as asymptomatic seemed reasonable. As described in this chapter, it is now evident that among patients who are discovered incidentally, more diligent studies using modern technologies that can detect skeletal and renal involvement, upon further inspection, will reveal that the disease often does involved those classic organ systems. With signs of skeletal and/or renal involvement, it might be time to reclassify these patients. Perhaps they are asymptomatic but they do demonstrate target organ involvement. Taking this concept one step further, if we reach a point where we can be confident of tools to measure neurocognitive and neuropsychiatric features, some of these patients might be classified as symptomatic.

The normocalcaemic disorder. After mild hypercalcaemia become the most common biochemical feature of PHPT, another four decades passed before the third variant of this disease, namely normocalcaemic PHPT became widely recognized [34]. These patients are not being discovered incidentally, as patients with mild hypercalcaemia are identified, but rather in the context of being evaluated for low bone density or frank osteoporosis [34–37, 152–155]. Paradoxically, these patients demonstrate at the time of presentation more signs of the disease than those with mild hypercalcaemia. Routine measurement of PTH in these settings explains how normocalcaemic PHPT became recognized. We now know that normocalcaemic PHPT can be discovered without low BMD if population screening is undertaken [156]. Thus, as is the case for the hypercalcaemic variant, normocalcaemic PHPT can be asymptomatic or symptomatic.

An International View of Clinical Variability in Primary Hyperparathyroidism

The historical perspective of PHPT implies a chronology of evolution. Symptomatic PHPT evolved to an asymptomatic disease which in turn evolved into a normocalcaemic disorder. This 90-year historical perspective can be misleading because it implies that the disease has actually changed. Given that all three forms of PHPT exist concurrently in the world today, another perspective argues that these three forms of PHPT have always coexisted. What sets the clinical stage for one or another variant may well be dependent upon country-specific factors. For example, in countries where biochemical screening is not part of the healthcare system, such as India [157], patients are unlikely to be discovered incidentally. In these countries, symptomatic PHPT is more likely and, as was the case throughout the world prior to the 1970s, the disease in those countries is considered to be uncommon. In our more modern approach to country-wide health and biochemical screening, asymptomatic PHPT has become the more common variant. It is not considered to be a rare disorder. Parenthetically, as noted already, the use of the term asymptomatic probably needs some adjustment, even if only semantically, because with greater imaging tools applied to this population, it is apparent that such patients can often be shown to have involvement of the skeleton and/or kidneys. Finally, in countries where PTH is routinely measured in the context of low BMD, normocalcaemic PHPT becomes a recognized clinical entity. The point to be made here is that PHPT has probably always existed in all three forms. The proportion of the hyperparathyroid population that will show them will depend upon these factors. Another consideration is the extent to which the population is vitamin D deficient. As noted already, vitamin D deficiency is associated with more signs of active disease. In a country that is markedly vitamin D deficient, on average, and does not employ widespread population screening, PHPT is likely to be viewed as an uncommon, symptomatic disorder. In a country that is not markedly vitamin D deficient, on average, and employs widespread population screening, PHPT is likely to be viewed as a common, asymptomatic disorder.

Management of Primary Hyperparathyroidism

Surgery for Signs and/or Symptoms

Patients who have clear signs of hyperparathyroidism whether they are skeletal (fractures) or renal (stones) or symptomatic hypercalcaemia should have parathyroid surgery, unless there are medical or other contraindications [8]. Areas of uncertainly relate to patients who have the asymptomatic and normocalcaemic forms of the disease.

Surgery in Asymptomatic Primary Hyperparathyroidism

Since surgery is curative when successful, patients and physicians can opt for surgery even if the patients do not meet guidelines that have been proposed for surgical management. Some patients find this to be an attractive option while others are inherently reluctant to consider surgery under almost any circumstances.

Over the past three decades, four international workshops on the management of asymptomatic primary hyperparathyroidism [5–8, 158] have periodically reviewed and updated the evidence that would argue either for a surgical or non-surgical management approach. The 4th International workshop [8] offered the following recommendations. They are listed here and shown in Table 4.3.2:

1. Serum calcium concentration greater than 1 mg/dl above the upper limit of normal;
2. a. BMD by DXA T score <–2.5 at lumbar spine, total hip, femoral neck, or distal third radius;
   b. vertebral fracture by X-ray, CT, MRI, or vertebral fracture assessment (VFA)
   c. evidence of a clinical fracture
3. a. estimated GFR (eGFR) <60 cc/min;
   b. 24-hour urine for calcium >400 mg/day (>10 mmol/day) and increased stone risk by biochemical stone risk analysis;
   c. nephrolithiasis or nephrocalcinosis by X-ray, ultrasound, or CT;
4. Age <50.

Any one of these criteria is sufficient to recommend surgery. These recommendations have not been updated recently, but we look forward to refinements and modifications based upon new data that have become available at the time of the next international conference.

It should be emphasized that these guidelines are not rules and, in fact, like many other guidelines, they are not based on strong evidence. Physicians and patients modify them according to the individual clinical situation. Some physicians feel that all patients with PHPT should undergo parathyroid surgery, as long as there are no contraindications. Some patients will resist surgery even though there are clear candidates and are strongly advised to have parathyroid surgery. Others will base their recommendations on other criteria.

Surgical Approach in Primary Hyperparathyroidism

A key dictum in this field is that parathyroid surgery should be undertaken by those who are experienced and highly skilled in the procedure [159–161]. While this dictum has not changed, over the past several decades, the approach to parathyroid surgery has. Rather than full exploration of the neck with visualization of all four parathyroid glands, many centres perform a limited procedure

Table 4.3.2 Guidelines for parathyroid surgery in asymptomatic primary hyperparathyroidism. Surgery is also indicated in patients for whom medical surveillance is neither desired nor possible. If any one of these criteria are met, the patient is considered to be a candidate for parathyroid surgery

Measurement	Surgery recommended
Serum calcium	>1.0 mg/dl (0.25 mmol/L) above normal
Renal system	Marked hypercalciuria (>400) with other stone risk factors; nephrolithiasis, nephrocalcinosis or creatinine clearance <60 cc/min/1.73 m²
Bone mineral density	T score <–2.5 at spine, hip (total or femoral neck), or radius (distal 1/3 site); or fragility fracture
Age	<50 years

Data derived from Bilezikian JP, Brandi ML, Eastell R, Silverberg SJ, Udelsman R, Marcocci C, Potts JT Jr: Consensus Statement: Guidelines for the Management of Asymptomatic Primary Hyperparathyroidism: Summary Statement from the Fourth International Workshop. *J Clin Endocrinol Metab* 2014;99:3561–3569.

when there is pre-intraoperative evidence for single gland disease and that at the time of surgery, there is intraoperative evidence for cure [162]. This is ascertained by intraoperative measurement of the PTH level at baseline and 2, 5, and 10 minutes following the removal of the parathyroid adenoma. If the PTH level falls by greater than 50% into the normal range, the resected parathyroid adenoma is considered to be the only gland involved and the operation is ended. With successful preoperative localization (see next), this operation can be performed by those skilled in the procedure under local anaesthesia and conscious sedation [163]. The minimally invasive parathyroidectomy (MIP), performed in an ambulatory care hospital setting, is a shorter operation without general anaesthesia, and with fewer complications [164–167]. If the intraoperative PTH level does not meet the criteria for success, namely it doesn't fall by greater than 50% and/or remains above normal, the operation is converted to a full neck exploration to seek other glands that are presumably overactive [168]. Success rates for this procedure that generally does not require an overnight hospital stay are greater than 90%, similar to the outcome of the full neck exploration [163, 169].

When the patient has multigland disease, a full exploration is necessary. Options are subtotal parathyroidectomy, consisting of removal of 3.5 glands, or total parathyroidectomy with transplantation of parathyroid tissue slices into the forearm. If the autotransplantation is successful, the patient will re-establish normal homeostatic control of the serum calcium by this ectopically placed parathyroid tissue. It is important to recognize that the tissue transplanted is abnormal and, thus, could become overactive in time. In that eventuality, the forearm site is convenient because the tissue could be accessed and debulked. Some centres have cryopreservation facilities which can be very helpful should the initial graft not take [170]. Lebastchi *et al.* have reported using the minimally invasive approach in multigland disease [171].

Surgery for Parathyroid Cancer

The surgical approach to parathyroid cancer is total removal of all malignant tissue and local neck dissection, when indicated. Sometimes, it is not known whether all malignant tissue has been removed because local and distant metastases can appear years after resection. Parathyroid cancer is one of the unusual malignancies in which metastases, if surgically approachable, can be removed with satisfactory results than can be measured sometimes in years [46]. Multiple therapeutic approaches along with resection of the metastases have also been associated with long-term survival [172].

Parathyroid Gland Imaging

While parathyroid imaging was for years considered to be a useful adjunct, prior to parathyroid surgery, advances in imaging technology have made this a mandatory aspect of the surgical protocol [9, 173, 174]. Preoperative parathyroid imaging is not done to make the diagnosis of PHPT. The decision to proceed with parathyroid surgery should have been made prior to the preoperative imaging test. Many approaches are available with endocrinologists, surgeons and radiologists all showing preferences for one or another procedure [175, 176].

Ultrasound is very convenient and with many endocrinologists and parathyroid surgeons having an ultrasound instrument in their offices, this approach is often utilized. It is unusual for the ultrasound image to be sufficient for confident localization without a confirmation by another approach. The sensitivity of neck ultrasound ranges from 42 to 82% but specificity is close to 90% [177].

Sestamibi imaging in many centres is a mainstay of the preoperative imaging repertoire. In some hands, sestamibi imaging can show sensitivity of about 90% and specificity of about 95% [178, 179]. When sestamibi is performed with single photon emission computed tomography (SPECT) alone [180] or in combination with CT, successful localization is even better [181]. A negative aspect of sestamibi is that its sensitivity for small adenomas or double adenomas or hyperplasia is poor. In a situation characterized by multigland disease, but one predominant parathyroid gland, the sestamibi scan typically will show only the dominant gland. To identify ectopic tissue, particularly in the mediastinum, sestamibi can be useful.

CT scanning of the neck and mediastinum is being used more frequently for parathyroid localization, particularly with contrast. The use of contrast identifies the procedure as 4D-CT with time being the fourth dimension. The contrast also gives functional identification along with anatomical localization of the parathyroid glands [182, 183]. It is also useful when the parathyroid tissue is ectopically present in the mediastinum [184]. The recent experience of Yeh *et al.*, compared 4DCT with sestamibi and found that 4DCT alone was more successful than sestamibi and that sestamibi did not provide additional value to the high success rate of 4DCT alone [185].

MRI has utility when the parathyroid gland is in the mediastinum and in the case of persistent disease after parathyroid surgery [186]. MRI has been used with positron emission tomography (PET) scanning with promising results [187].

Arteriography and Selective Venous Sampling. Rarely used these days, arteriography and selective venous sampling can be helpful in very special situations of persistent PHPT without successful localization by the aforementioned imaging modalities [188].

Clinical Course of Primary Hyperparathyroidism

Natural History Without Surgery

Prospective follow-up of asymptomatic PHPT for up to 15 years without surgery or specific medical therapy [189, 190] has shown that for the first 10 years, biochemistries are all stable. In years 10–15, however, the serum calcium shows a tendency to increase slightly, but renal function, as measured by the creatinine clearance does not decline. The serum calcium increased above 12 mg/dl in 4% while 15% became markedly hypercalciuric with urinary calcium excretion >400 mg/day. By BMD, the first decade of observation shows stability at all three sites. In years 11–15, however, the femoral neck and more evidently the distal 1/3 radius show progressive declines. Overall, 12% showed declines in BMD at any site, meeting guidelines for parathyroid surgery. Overall, 37% of subjects followed for this 15-year period met one or more indications for parathyroid surgery. No specific index could be shown to be predictive of disease progression with the one exception be age under 50. In those younger patients, the risk of progression was over 60%. In those over 50, risk of progression was 25%, over the same period of time.

Among patients who met one or more indications for surgery but who opted against surgery, progression was the rule. Recurrent nephrolithiasis for example was often experienced by these patients.

Monitoring Patients with PHPT Who Do Not Undergo Parathyroid Surgery

The natural history of PHPT followed without surgery has led to the following recommendations for monitoring. They are shown in Table 4.3.3 and summarized here.

1. Yearly or twice-yearly serum calcium measurement.
2. Yearly or every other year—three-site DXA (lumbar spine, hip regions, and distal 13 radius). While yearly DXA is unlikely to show significant changes, from year to year, having these data points allows one to spot downward trends more expeditiously. If clinically indicated, X-ray or VFA of the spine should be performed (i.e. height loss, back pain).
3. eGFR annually. If renal stones are suspected in the interim, a 24-hour urine for biochemical stone profile along with renal imaging by X-ray, ultrasound, or CT are advised.

Natural History of Normocalcaemic PHPT

In one study of normocalcaemic PHPT [35], 37 patients were followed for 3 years on average (range 1–9). Hypercalcaemia emerged in 7 (19%) individuals. A greater percentage, 40%, developed evidence of kidney stones, fractures, marked hypercalciuria or greater than a 10% decline in BMD. Seven patients had successful parathyroidectomy of whom three were hypercalcaemic and the rest met other criteria for surgery.

The Fourth International Workshop on the Management of Asymptomatic PHPT, suggested, for the first time, a management protocol for normocalcaemic disease [8]. In patients who develop hypercalcaemia, the recommendations for the hypercalcaemic form of PHPT should be followed as noted as described. If complications of PHPT develop, even if the serum continues to be normal, they should be regarded as candidates for parathyroidectomy [191].

Natural History After Parathyroidectomy

Successful parathyroid surgery is followed by correction of hypercalcaemia along with a return of the PTH level to normal. Urinary calcium levels, if elevated, will also fall. Bone turnover makers also decline [192, 193]. Bone resorption markers fall more rapidly after successful parathyroid surgery than bone formation markers [192]. The persistence of bone formation markers coupled with rapid declines in bone resorption markers is compatible with a pattern that should indicate accrual of bone mineral postoperatively.

In fact, one of the reliable aspects of successful parathyroid surgery is an increase in BMD. The lumbar spine shows increases within a 1 year (Figure 4.3.2), following by slower but eventual increases in the hip and distal one-third radius BMD [189, 190, 193–199]. In a meta-analysis of these postoperative findings, Sankaran *et al.* confirmed these findings [199]. Surgery also appears to be associated with a reduction in fracture risk [198, 199]. The marked increase in BMD following successful parathyroid surgery has led experts to suggest that pharmacological intervention for low bone mass can be safely withheld because bone mass accrues so quickly without intervention.

A dramatic example of this point was demonstrated by Kulak *et al.* [200] who reported two young patients with severe parathyroid bone disease (*osteitis fibrosa cystica*) who improved their BMD by 260 to 430%, within 4 years of after successful surgery.

Table 4.3.3 Management guidelines for patients with asymptomatic primary hyperparathyroidism who do not undergo parathyroid surgery

Measurement	Frequency
Serum calcium	Annually
Renal	Clcr annually; stone risk profile and/or abdominal imaging if clinically indicated
Bone mineral density	DXA every 1 or 2 years; vertebral imaging with CT or VFA if clinically indicated

Data derived from Bilezikian JP, Brandi ML, Eastell R, Silverberg SJ, Udelsman R, Marcocci C, Potts JT Jr: Consensus Statement: Guidelines for the Management of Asymptomatic Primary Hyperparathyroidism: Summary Statement from the Fourth International Workshop. *J Clin Endocrinol Metab* 2014;99:3561–3569.

Non-Surgical Approaches to Primary Hyperparathyroidism

Patients who are not going to have parathyroid surgery should maintain hydration at all times. Their intake of calcium should follow established nutritional guidelines. While it seems intuitive to want to restrict calcium intake in PHPT, there are reasons not to restrict dietary calcium. Calcium excretion is not substantially affected by calcium intake [201]. There is reason to be concerned about low calcium intakes as this could be a signal for the abnormal parathyroid gland(s) to become even more active. On the other hand, one does not want to encourage high calcium intake which can be a particular problem if the 1,25-dihydroxyvitamin D_3 is elevated [202]. As is the case for subjects without PHPT, dietary sources of calcium are to be recommended instead of supplemental forms.

Phosphate

Oral phosphate is virtually never used in PHPT anymore. Concerns about GI tolerance, further increases in PTH levels, and ectopic calcifications have essentially relegated oral phosphate to a historical footnote [201].

Oestrogen

Studies of oestrogen in PHPT have typically been high, but lower doses have also seemed to be beneficial [203]. Given the limited data, it is nevertheless, reasonable to consider oestrogen among postmenopausal women for whom surgery is not an option, are willing to take oestrogen, and there are no contraindications. When oestrogen is effective, the serum calcium will fall by 0.5 to 1.0 mg/d. PTH levels will remain stable. BMD will increase in the lumbar spine and femoral neck [203].

Selective Oestrogen Receptor Modulator (SERM)

There are very limited data on the use of SERMs in PHPT. In a small study of 18 postmenopausal women, Rubin *et al.* [204] reported a significant decline in the serum calcium by about 0.5 mg/dl after 8 weeks of raloxifene. The results were transient with return of the serum calcium and reduced bone turnover markers to baseline

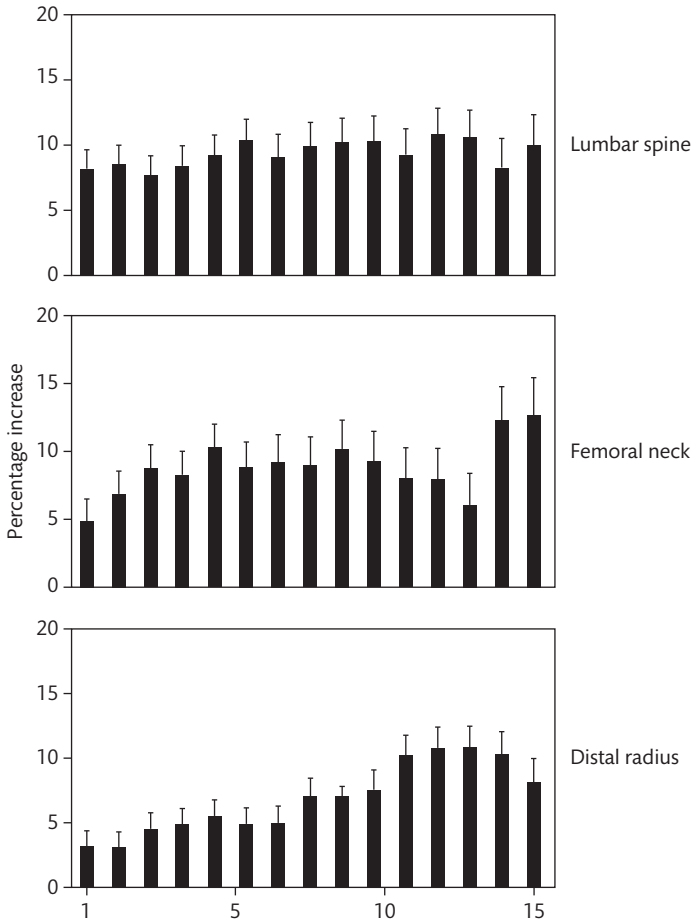


Figure 4.3.2 Increases in bone mineral density after successful parathyroid surgery in primary hyperparathyroidism. Data shown are the cumulative percentage changes from baseline over 15 years of follow-up.

Reproduced with permission from Rubin MR, Bilezikian JP, McMahon DJ, et al. The Natural History of Primary Hyperparathyroidism With or Without Parathyroid Surgery after 15 Years. *J Clin Endocrinol Metab* 2008;93:3462–3470. Copyright © 2008, Oxford University Press. (ref 190).

during the 4-week wash out period. There were no changes in PTH or urinary calcium excretion. Zanchetta and Bogado [205], with an even smaller number of subjects and without a blinded experimental design, showed a similar reduction in serum calcium with raloxifene.

Vitamin D

Referencing the discussion earlier in this chapter on vitamin D deficiency being related to activity of disease, the logic follows that vitamin D replacement should be associated with better control of the hyperparathyroid state. Grey *et al.* [206] administered vitamin D_3 to 21 patients with mild PHPT. The repletion regimen in these subjects whose 25-hydroxyvitamin levels averaged 20 ng/ml, was 50 000 international unit (IU)s weekly for 4 weeks followed by 50 000 IU monthly for the next 11 months. After 12 months, the average 25-hydroxyvitamin D levels increased to 31 ng/ml. The serum PTH levels fell by about 25% without any change in the serum calcium. Three subjects developed marked hypercalciuria (>400 mg/day). Although this report supports the idea that vitamin D insufficiency can worsen the hyperparathyroid state, it does not provide guidance as to how vitamin D should be replaced. A more reasonable approach employed 2800 IU/day for 6 months and showed, in a randomized trial, that mean PTH levels fell by 17% without any change in the serum calcium concentration [207]. In addition to the biochemical effects of vitamin D, muscle strength might be improved as shown by Amstrup *et al.* [208]. This suggestion could not be confirmed by Rolighed *et al.* in a short 6-month study [209].

In short, the 25-hydroxyvitamin D level should be normal in PHPT. While the definition of 'normal' is controversial, that controversy relates to the 'right number' in subjects without evidence for metabolic bone diseases. Given the upside safety of vitamin D and the work of Walker *et al.*, in which the inflection point below which the PTH level appears to rise further in PHPT, namely at around 25 ng/ml, it seems reasonable to aim for a level greater than 30 ng/ml (75 nmol/L).

Bisphosphonates

The bisphosphonates are attractive in PHPT, not because they affect PTH secretion directly but because they reduce bone turnover in a disease characterized by high bone turnover. Theoretically, they could reduce serum and urinary calcium levels and increase BMD. The amino substituted bisphosphonates, in common use for osteoporosis, have been of interest. A limited but positive experience with risedronate [210], has been followed by more detailed studies with alendronate. Several promising studies of alendronate

in PHPT [211–212] have been followed by three well-controlled studies [213–215]. The work of Khan *et al.* [215] utilizing a randomized, double-blinded, placebo-controlled design, in 44 patients with mild, asymptomatic PHPT, showed a significant 5.3% increase in BMD of the lumbar spine, which rose further to almost 7% % by year 2 (Figure 4.3.3). Total hip BMD increased by 4.01% in year 2. As expected, bone turnover markers fell by more than 50%. Calcium, phosphorus, and PTH levels did not change. Men with PHPT have also been studied with similar results [216, 217]. Chow *et al.* [214] have reported an experience similar to Khan *et al.*, except that the serum calcium fell by 0.34 mg/dl.

These promising results suggest that in patients with low BMD who are not candidates for parathyroid surgery, for whatever reason, might show improvements in BMD when alendronate is use. An unusual finding of increased fracture risk in those receiving alendronate compared to those who underwent successful parathyroidectomy deserves comment [218]. The selection bias of the two groups would appear to be the best explanation for these unexpected results. Those who received alendronate were at greater risk for fracture than those who underwent parathyroid surgery. Obviously, if a patient is a candidate for parathyroid surgery, that is preferred but the body of reports with alendronate do substantiate its utility in situations where parathyroid surgery is not a consideration.

Denosumab

The use of denosumab is theoretically attractive in PHPT because it interferes with RANK L, a cytokine intimately related to a catabolic pathway utilized by PTH. In the study of Eller-Vainicher *et al.* (Figure 4.3.4), denosumab was shown to improve BMD to a significantly greater extent than a comparable group of postmenopausal women with osteoporosis but without PHPT [219]. One looks forward to further controlled studies of denosumab in PHPT.

Calcimimetics

Different from the aforementioned pharmacological approaches to PHPT, an approach that targets the parathyroid gland secretory apparatus is most attractive. Interference with the CaSR would provide this more specific approach. Calcimimetics increase the calcium signal to the parathyroid cell and should reduce the secretion of PTH. Following promising early studies with an early calcimimetic, the phenylalkylamine (R)-N-(3-methoxy-alpha-phenylethyl)-3-(2-chlorophenyl)-1- propylamine [R-568], a more potent calcimimetic, cinacalcet hydrochloride was studied and approved by the US Food and Drug Administration (FDA). The indication is for control of the serum calcium in patients with PHPT who present a challenge to control [220–221]. The work of Shoback, Peacock, and their associates have documented the efficacy of cinacalcet in reducing the serum calcium in PHPT [222–224]. The studies have been carried out for up to 5 years, demonstrating continuous control of the serum calcium [224]. The drug has been shown to effective among a spectrum of clinical manifestations and in those with severe disease [225–227]. One negative aspect of this drug is that BMD does not improve.

Cinacalcet has also been shown to be effective in parathyroid cancer but high doses may be required [228].

ACKNOWLEDGEMENTS

The work in this chapter that refers specifically to the author's research was supported by NIH grant, DK32333 from NIDDK.

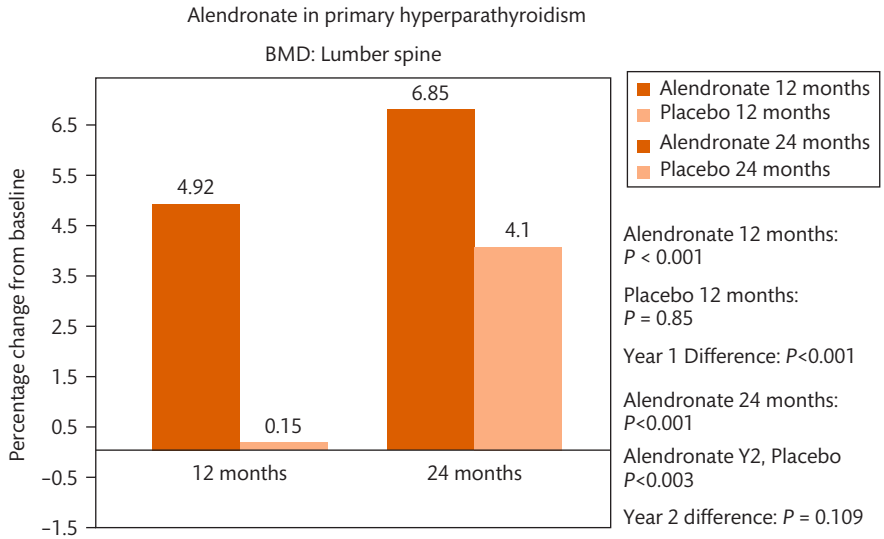


Figure 4.3.3 The effect of alendronate on bone mineral density in primary hyperparathyroidism. With alendronate, bone mineral density increases significantly after 1 year, while the placebo group shows no change until it is crossed over to alendronate in year 2.

Adapted with permission from Khan AA, Bilezikian JP, Kung AWC, Ahmed MM, Dubois SJ, Ho AYY, Schussheim D, Rubin MR, Shaikh AM, Silverberg SJ, Standish TI, Syed Z, Syed ZA: Alendronate in primary hyperparathyroidism: a double-blind, randomized, placebo-controlled trial. *J Clin Endocrinol Metab* 2004;89:3319–3325. Copyright © 2004, Oxford University Press. (ref 215).

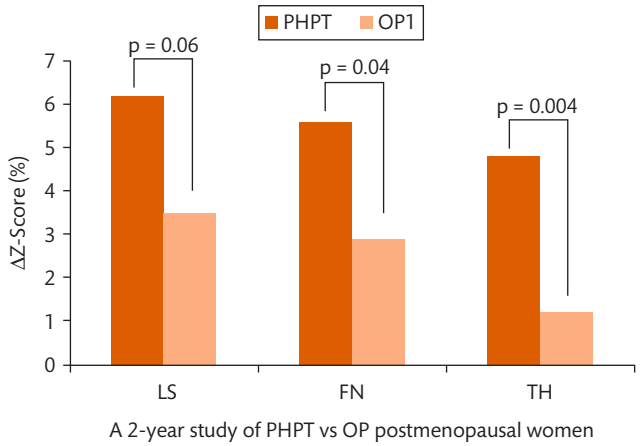


Figure 4.3.4 Denosumab in primary hyperparathyroidism. 1-year changes in BMD after denosumab in osteoporotic postmenopausal women with or without primary hyperparathyroidism. The patients with primary hyperparathyroidism demonstrated significant greater increases in BMD at all sites in comparison to those without primary hyperparathyroidism.

REFERENCES

1. Bilezikian JP, Cusano NE, Khan AA, Liu JM, Marcocci C, Bandeira F. Primary hyperparathyroidism. *Nat Rev Dis Primers*, 2016; **2**: 1–16.
2. Bilezikian JP. Primary hyperparathyroidism. In: Feingold KR, Anawalt B, Boyce A, *et al.*, eds. www.ENDOTEXT.org (Version of 2017). South Dartmouth (MA): MDText.com, Inc.; 2000–2017.
3. Bilezikian JP, Bandeira L, Khan A, Cusano NE. Hyperparathyroidism. *Lancet*, 2018; **391**: 168–78.
4. Cusano NE, Silverberg SJ, Bilezikian JP. Normocalcemic PHPT. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 331–40.
5. Consensus Development Conference Panel. Diagnosis and management of asymptomatic primary hyperparathyroidism. *Ann Intern Med*, 1991; **114**: 593–7.
6. Bilezikian JP, Potts JT, El-Hajj Fuleihan G, *et al.* Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century. *J Bone Min Res*, 2002; **17** (Suppl 2): N2–11.
7. Bilezikian JP, Khan A, Potts Jr JT. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the third international workshop. *J Clin Endocrinol Metab*, 2009; **94**: 335–9.
8. Bilezikian JP, Brandi ML, Eastell R, *et al.* Consensus statement: guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the fourth international workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3561–9.
9. Khan AA, Hanley DA, Rizzoli R, *et al.* Primary hyperparathyroidism: review and recommendations on evaluation, diagnosis and management. A Canadian and international perspective. *Osteoporosis Int*, 2017; **28**: 1–19.
10. Albright F, Aub J, Bauer W. Hyperparathyroidism: common and polymorphic condition as illustrated by seventeen proven cases in one clinic. *JAMA*, 1934; **102**: 1276.
11. Revollo L, Civitelli R. Molecular actions of parathyroid hormone. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 119–26.
12. Silva B and Kousteni S. Cellular actions of PTH: osteoblasts, osteoclasts, and osteocytes. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 127–38.
13. Press DM, Siperstein AE, Berber E, *et al.* The prevalence of undiagnosed and unrecognized primary hyperparathyroidism: a population-based analysis from the electronic medical record. *Surgery*, 2013; **154**: 1232–7.
14. Yeh MW, Ituarte PH, Zhour HC, *et al.* Incidence and prevalence of primary hyperparathyroidism in a racially mixed population. *J Clin Endocrinol Metab*, 2013; **98**: 1122–9.
15. Kontogeogos G, Trimpou P, Laine CM, Olerod G, Lindahl A, Landin-Wilhelmsen K. Normocalcemic, vitamin D-sufficient hyperparathyroidism—high prevalence and low morbidity in the general population: a long-term follow-up study, the WHO MONICA project, Gothenburg, Sweden. *Clin Endocrinol*, 2015; **83**: 277–84.
16. Vignali E, Cetani F, Chiavistelli S, *et al.* Normocalcemic primary hyperparathyroidism: a survey in a small village of Southern Italy. *Endocr Connect*, 2015; **4**: 172–8.
17. Cusano NE, Maalouf N, Wang PY, *et al.* Normocalcemic hyperparathyroidism and hypoparathyroidism in two community-based non-referral populations. *J Clin Endocrinol Metab*, 2013; **98**: 2734–41.
18. Abood A, Vestergaard P. Increasing incidence of primary hyperparathyroidism in Denmark. *Dan Med J*, 2013; **60**: A4567.
19. Collier A, Portelli M, Chosh S, Nowell S, Clark D. Primary hyperparathyroidism: increasing prevalence, social deprivation and surgery. *Endocr Res*, 2017; **42**: 31–5.
20. Bilezikian JP, Bandeira L, Khan A, Cusano NE. Hyperparathyroidism. *Lancet*, 2018; **391**: 168–78.
21. Liu J, Cusano NE, Silva BC, *et al.* Primary hyperparathyroidism: a tale of two cities revisited—New York and Shanghai. *Bone Res*, 2013; **2**: 162–9.
22. Bandeira F, Griz L, Caldas G, Bandeira C, Freese E. From mild to severe primary hyperparathyroidism: the Brazilian experience. *Arq Bras Endocrinol Metabol*, 2006; **50**: 657–63.
23. Eufrazino C, Veras A, Bandeira F. Epidemiology of primary hyperparathyroidism and its non-classical manifestations in the city of Recife, Brazil. *Clin Med Insights Endocrinol Diabetes*, 2013; **6**: 69–74.
24. Melton LJ. The epidemiology of primary hyperparathyroidism in North America. *J Bone Miner Res*, 2002; **17**(Suppl.2): N12–17.
25. DeGregorio L, Lubitz CC, Hodin RA, *et al.* The truth about double adenomas: incidence, localization, and intraoperative parathyroid hormone. *J Am Coll Surg*, 2016; **222**: 1044–52.
26. Holick MF, Binkley NC, Bischoff-Ferrari HA, *et al.* Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 1911–30.
27. Rosen CJ, Abrams SA, Aloia JF, *et al.* IOM committee members respond to Endocrine Society vitamin D guideline. *J Clin Endocrinol Metab*, 2012; **97**: 1146–52.
28. Bjorkman M, Sorva A, Tilvis R. Responses of parathyroid hormone to vitamin D supplementation: a systematic review of clinical trials. *Arch Gerontol Geriatr*, 2009; **48**: 160–6.
29. Fillee C, Keller T, Mourad M, Brinkmann T, Ketelslegers JM. Impact of vitamin D-related serum PTH reference values on the

diagnosis of mild primary hyperparathyroidism, using bivariate calcium/PTH reference regions. *Clin Endocrinol (Oxf)*, 2012; **76**: 785–9.
30. Piketty ML, Prie D, Sedel F, et al. High-dose biotin therapy leading to false biochemical endocrine profiles: validation of a simple method to overcome biotin interference. *Clin Chem Lab Med*, 2017; **55**: 817–25.
31. Griebeler ML, Kearns AE, Ryu E, *et al*. Thiazide-associated hypercalcemia: incidence and association with primary hyperparathyroidism over two decades. *J Clin Endocrinol Metab*, 2016; **101**: 1166–73.
32. Szalat A, Mazeh H, Freund HR. Lithium-associated hyperparathyroidism: report of four cases and review of the literature. *Eur J Endocrinol*, 2009; **160**: 317–23.
33. Marx SJ. Familial hypocalciuric hypercalcemia as an atypical form of primary hyperparathyroidism. *J Bone Miner Res*, 2018; **33**: 27–31.
34. Silverberg SJ, Bilezikian JP. 'Incipient' primary hyperparathyroidism: a 'forme fruste' of an old disease. *J Clin Endocrinol Metab*, 2003; **88**: 5348–52.
35. Lowe H, McMahon DJ, Rubin MR, Bilezikian JP, Silverberg SJ. Normocalcemic primaryhyperparathyroidism: further characterization of a new clinical phenotype. *J Clin Endocrinol Metab*, 2007; **92**: 3001–5.
36. Thomas MK, Lloyd-Jones DM, Thadhani RI, *et al*. Hypovitaminosis D in medical inpatients. *N Eng J Med*, 1998; **338**: 778–83
37. Maruani G, Hertig A, Paillard M, Houillier P. Normocalcemic primary hyperparathyroidism: evidence for a generalized target-tissue resistance to parathyroid hormone. *J Clin Endocrinol Metab*, 2003; **88**: 4641–8.
38. Greenspan SL, Holland S, Maitland-Ramsey L, *et al*. Alendroante stimulation of nocturnal parathyroid hormone secretion: a mechanism to explain the continued improvement in bone mineral density accompanying alendronate therapy. *Proc Assoc Am Physicians*, 1996; **108**: 230–8.
39. Mazokopakie EE. Denosumab-induced normocalcemic hyperparathyroidism in a woman with postmenopausal osteoporosis and normal renal function. *Curr Drug Saf*, 2018; **13**: 214–16.
40. Laryyk OA, Siperstein AE. Use of calcium and parathyroid hormone nomogram to distinguish between atypical primary hyperparathyroidism and normal patients. *World J Surg*, 2017; **41**: 122–8
41. Rosario RW. Primary hyperparathyroidism with normal calcium and PTH. *World J Surg*, 2017; **41**: 1649–50.
42. Maselly MJ, Lawrence AM, Brooks M, *et al*. Hyperparathyroid crisis. Successful treatment of ten comatose patients. *Surgery*, 1981; **90**: 741–60.
43. Nudelman IL, Deutsch AA, Reiss R. Primary hyperparathyroidism due to mediastinal parathyroid adenoma. *Int Surg*, 1987; **72**: 104–8.
44. Corlew DS, Bryda SL, Bradley EL, DiGirolamo M. Observations on the course of untreated primary hyperparathyroidism. *Surgery*, 1985; **98**: 1064–71.
45. Fitzpatrick LA. Acute primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 401–408.
46. Cetani F, Marcocci C. Parathyroid carcinoma. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 409–22.
47. De Lellis, RA, Arnold, A, Bilezikian JP, *et al.* Parathyroid carcinoma. In: Lloyd RV, Osamuna RY, Kloppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. Lyon, France: Int'l Agency for Research on Cancer, 2017: 147–52.
48. Marcocci C, Cetani F, Rubin MR, Silverberg SJ, Pinchera A, Bilezikian JP. Parathyroid carcinoma. *J Bone Min Res*, 2008; **23**: 1869–80.
49. Chen JD, Morrison C, Zhang C, Kahnoski K, Carpten JD, Teh BT. Hyperparathyroidism-jaw tumour syndrome. *J Intern Med*, 2003; **253**: 634–42.
50. Nakajima K, Tamai M, Okaniwa S, *et al*. Humoral hypercalcemia associated with gastric carcinoma secreting parathyroid hormone: a case report and review of the literature. *Endocr J*, 2013; **60**: 557–62
51. Thakker RV, Newey PJ, Walls GV, *et al*. Clinical practice guidelines for multiple endocirne neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*, 2012; **97**: 2990–3011.
52. Frederiksen A, Rossing M, Hermann P, Ejersted C, Thakker RV, Nielsen MF. Clinical features of multiple endocrine neoplasia type 4—novel pathogenic variant and review of published cases. *J Clin Endocrinol Metab*, 2019; **104**: 3637–46.
53. Marx SJ. New concepts about familial isolated hyperparathyroidism. *J Clin Endocrinol Metab*, 2019; **104**: 4058–66.
54. Thakker RV. Familial and hereditary forms of primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 341–63.
55. Kloos RT, Eng C, Evans DB, *et al*. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid*, 2009; **19**: 565–612.
56. Newey PJ, Bowl, MR, Cranston T, Thakker RV. Cell division cycle protein 73 homolog (CDC73) mutations in the hyperparathyroidism jaw-tumor syndrome (HPT-JT) and parathyroid tumors. *Hum Mutat*, 2010; **31**: 295–307.
57. Hannan FM, Nesbit MA, Christie PT, *et al*. Familial isolated primary hyperparathyroidism caused by mutations of the MEN1 gene. *Nat Clin Pract Endocrinol Metab*, 2008; **4**: 53–8.
58. Pannett AA, Kennedy AM, Turner JJ, *et al*. Multiple endocrine neoplasia type 1 (MEN1) germline mutations in familial isolated primary hyperparathyroidism. *Clin Endocrinol (Oxf)*, 2003; **58**: 639–46.
59. Warner JV, Nyholt DR, Busfield F, *et al*. Familial isolated hyperparathyroidism is linked to a 1.7Mb region on chromosome 2p13.3–14. *J Med Genet*, 2006; **43**: e12.
60. El-Hajj Fuleihan G, Brown E. Familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn San Diego, CA: Elsevier, 2015: 365–88.
61. Arnold A, Levine MA. Molecular basis of primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn San Diego, CA: Elsevier, 2015: 279–96.
62. European consortium on MEN1. Identification of the multiple endocrine neoplasia type 1 (*MEN1*) gene. *Hum Mol Genet*, 1997; **6**: 1177–83.
63. Tahara H, Smith AP, Gas RD, *et al*. Genomic localization of novel candidate tumor suppressor gene loci in human parathyroid adenomas. *Cancer Res*, 1996; **56**: 599–605.
64. Palanisamy N, Imanishi Y, Rao PH, *et al*. Novel chromosomal abnormalities identified by comparative genomic hybridization in parathyroid adenomas. *J Clin Endocrinol Metab*, 1998; **83**: 1766–70.
65. Baloch ZW, Livolsi VA. Parathyroids: morphology and pathology. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 23–36.

66. Villablanca A, Wassif WS, Smith T, *et al*. Involvement of the MEN1 gene locus in familial isolated hyperparathyroidism. *Eur J Endocrinol*, 2002; **147**: 313–22.
67. Silverberg SJ, Bilezikian JP. Asymptomatic primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 317–30.
68. Costa AG, Marcocci C, Bilezikian JP. Bone turnover markers in primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 423–8.
69. Lumb GA, Stanbury SW. Parathyroid function in vitamin D deficiency in primary hyperparathyroidism. *Am J Med*, 1974; **54**: 833–9.
70. Bilezikian JP, Meng X, Shi Y, Silverberg SJ. Primary hyperparathyroidism in women: New York and Beijing (a tale of two cities). *Int'l J Fertility and Women's Health*, 2000; **45**: 158–65.
71. Mithal A, Bandeira F, Meng X, *et al.* Primary hyperparathyroidism in India, Brazil and China. In: Bilezikian JP, Marcus R, Levine MA, eds. *The Parathyroids*, 2nd edn. San Diego, CA: Academic Press, 2001: 375–86.
72. Silverberg SJ, Shane E, Dempster DW, Bilezikian JP. Vitamin D deficiency in primary hyperparathyroidism. *Am J Med*, 1999; **107**: 561–56.
73. Fraser WD. Hyperparathyroidism. *Lancet*, 2009; **374**: 145–58.
74. Bouillon R, Bollerslev J, Silverberg SJ. Vitamin D and primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 481–4887.
75. Walker MD, Cong E, Lee JA, *et al*. Low vitamin D levels have become less common in primary hyperparathyroidism. *Osteoporosis Int*, 2015; **26**: 2837–43.
76. Stein EM, Dempster DW, Udesky J, *et al*. Vitamin D deficiency influences histomorphometric features of bone in primary hyperparathyroidism. *Bone*, 2011; **48**: 557–61.
77. Walker MD, Saeed I, Lee JA, *et al*. Effect of concomitant vitamin D deficiency or insufficiency on lumbar spine volumetric bone mineral density and trabecular bone score in primary hyperparathyroidism. *Osteoporosis Int*, 2016; **27**: 3063–71.
78. Walker MD, Nishiyama KK, Zhou B, *et al*. Effect of low vitamin D on volumetric bone mineral density, bone microarchitecture, and stiffness in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 905–13.
79. Walker MD, Bilezikian JP. Vitamin D and primary hyperparathyroidism: more insights into a complex relationship. *Endocrine*, 2017; **55**: 3–5.
80. Silverberg SJ, Bilezikian JP. Primary hyperparathyroidism. In: DeGroot LJ, Jameson JL, eds.; Potts JT, section ed. *Endocrinology*, 6th edn. Philadelphia, PA: Saunders-Elsevier, 2010: 1176–97.
81. Cope O. The story of hyperparathyroidism at the Massachusetts General Hospital. *N Engl J Med*, 1966; **21**: 174–82.
82. Silverberg SJ, Bilezikian JP. Primary hyperparathyroidism. In: Wass JAH, Stewart PM, eds. *Oxford Textbook of Endocrinology and Diabetes*, 2nd edn, Oxford: Oxford University Press, 2011: 653–64.
83. Silverberg SJ, Shane E, DeLaCruz L, *et al*. Skeletal disease in primary hyperparathyroidism. *J Bone Mineral Res*, 1989; **4**: 283–91.
84. Bilezikian JP, Silverberg SJ, Gartenberg F, *et al.* Clinical presentation of primary hyperparathyroidism. In: Bilezikian JP, Marcus R, Levine MA, eds. *The Parathyroids*. New York, NY: Raven Press, 1994: 457–70.
85. Khosla S, Melton LJ III, Wermers RA, Crowson CS, O'Fallon WM, Riggs BL. Primary hyperparathyroidism and the risk of fracture: a population-based study. *J Bone Min Res*, 1999; **14**: 1700–7.
86. Vignali E, Viccica G, Diacinti D, *et al*. Morphometic vertebral fractures in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2009; **94**: 2306–12.
87. Pawlowska M, Bilezikian JP. Beyond DXA: advances in clinical applications of new bone imaging technology. *Endocr Pract*, 2016; **22**: 990–8.
88. Silva BC, Cusano NE, Hans D, Lewiecki EM, Bilezikian JP. Skeletal Imaging in Primary Hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn San Diego, CA: Elsevier, 2015: 447–54.
89. Stein EM, Silva BC, Boutroy S, *et al*. Primary hyperparathyroidism is associated with abnormal cortical and trabecular microstructure and compromised bone stiffness in postmenopausal women. *J Bone Miner Res*, 2013; **28**: 1029–40.
90. Hansen S, Beck Jensen JE, Rasmussen L, Hauge EM, Brixen K. Effects on bone geometry, density, and microarchitecture in the distal radius but not the tibia in women with primary hyperparathyroidism: a case-control study using HR-pQCT. *J Bone Miner Res*, 2010; **25**: 1041–947.
91. Vu T, Wang XF, Wang Q, *et al*. New insights into the actions of PTH in primary hyperparathyroidism on cortical and trabecular compartments of bone. *Bone*, 2013; **55**: 57–63.
92. Dempster DW, Silverberg SJ, Shane E, Bilezikian JP. Bone histomorphometry and bone quality in primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 429–46.
93. Silva BC, Bilezikian JP. Trabecular bone score; perspectives of an imaging technology coming of age. *Arq Bras Endocrinol Metabol*, 2014; **58**: 493–502.
94. Silva BC, Broy SB, Boutroy S, Schousboe JT, Shepherd JA, Leslie WD. Fracture rick prediction by non-BMD DXA Measures: the 2015 ISCD official positions Part 2: trabecular bone score. *J Clin Densitom*, 2015; **18**: 309–30.
95. Silva BC, Leslie WD, Resch H, *et al*. Trabecular bone score: a non-invasive analytical method based upon the DXA image. *J Bone Miner Res*, 2014; **29**: 518–30.
96. Romagnoli E, Cipriani C, Nofroni I, *et al*. 'Trabecular Bone Score' (TBS): an indirect measure of bone micro-architecture in postmenopausal patients with primary hyperparathyroidism. *Bone*, 2013; **53**: 154–9.
97. Cipriani E, Biamonte F, Costa AG, *et al*. Prevalence of kidney stones and vertebral fractures in primary hyperparathyroidism using imaging technology. *J Clin Endocrinol Metab*, 2015; **100**: 1309–15.
98. Tay YD, Liu M, Bandeira L, Bucovsky M, Lee JA, Silverberg SJ, Walker MD. Occult urolithiasis in asymptomatic primary hyperparathyroidism. *Endocr Res*, 2018; **43**: 106–15.
99. Reid LJ, Muthukrishnan B, Patel D, Seckl Jr, Gibb FW. Predictors of nephrolithiasis, osteoporosis, and mortality in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2019: **104**: 3692–700.
100. Peacock M. Primary hyperparathyroidism and the kidney. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 455–67.
101. Scillitani A, Guarnieri V, Battista C, *et al*. Primary hyperparathyroidism and the presence of kidney stones are associated with different haplotypes of the calcium-sensing receptor. *J Clin Endocrinol Metab*, 2007; **92**: 277–83.
102. Odvina CV, Sakhaee K, Heller HJ, *et al*. Biochemical characterization of primary hyperparathyroidism with and without kidney stones. *Urol Res*, 2007; **35**: 123–8.

103. Walker MD, Nickolas T, Kepley A, *et al*. Predictors of renal function in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2014; **99**: 1885–92.
104. Walker MD, Dempster DW, McMahon DJ, *et al*. Effect of renal function on skeletal health in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2012; **97**: 1501–7.
105. Silverberg SJ, Clarke BL, Peacock M, *et al*. Consensus statement: current issues in the presentation of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3580–94
106. Patten BM, Bilezikian JP, Mallette LE, *et al*. The neuromuscular disease of hyperparathyroidism. *Ann Inter Med*, 1974; **80**: 182–94.
107. Turken SA, Cafferty M, Silverberg SJ, *et al*. Neuromuscular involvement in mild, asymptomatic primary hyperparathyroidism. *Am J Med*, 1989; **87**: 553–7.
108. Diniz ET, Bandeira F, Linds OG, *et al*. Primary hyperparathyroidism is associated with subclinical peripheral neural alterations. *Endocr Pract*, 2013; **19**: 219–25.
109. Procopio M, Borretta G. Derangement of glucose metabolism in hyperparathyroidism. *J Endocrinol Invest*, 2003; **11**: 17–35.
110. Kumar A, Singh S. Parathyroidectomy ameliorates glucose and blood pressure control in a patient with primary hyperparathyroidism, type 2 diabetes, and hypertension. *Clin Med Insights Endocrinol Diabetes*, 2015; **8**: 63–6.
111. Cvijovic G, Micic D, Kendereski A, *et al*. The effect of parathroidectomy on insulin sensitivity in patients with primary hyperparathyroidism—a never ending story? *Exp Clin Endocrinol Diabetes*, 2015; **123**: 336–41.
112. Thakker RV, Newey PJ, Walls GV, *et al*. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN 1). *J Clin Endocrinol Metab*, 2012; **97**: 2990–3011.
113. Khoo TK, Vege SS, Abu-Lebdeh HS, Ryu E, Nadeem S, Wermers RA. Acute pancreatitis in primary hyperparathyroidism: a population-based study. *J Clin Endocrinol Metab*, 2009; **94**: 2115–18.
114. Silverberg SJ, Walker MD, Bilezikian JP. Asymptomatic primary hyperparathyroidism. *J Clin Densitom*, 2013; **16**: 14–21.
115. Walker MD, Silverberg SJ. Non-traditional manifestations of primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 469–80.
116. Walker MD, Bilezikian JP. Primary hyperparathyroidism: recent advances. *Curr Opin Rheumatol*, 2018; **30**: 427–39.
117. Lumachi F, Camozzi V, Luisetto G, Zanella S, Basso SM. Arterial blood pressure, serum calcium and PTH in elderly men with parathyroid tumors and primary hyperparathyroidism. *Anticancer Res*, 2011; **31**: 3969–72.
118. Nainby-Luxmoore JC, Langford HG, *et al*. A case-comparison study of hypertension and hyperparathyroidism. *J Clin Endocrinol Metab*, 1982; **55**: 303–6.
119. Ringe JD Reversible hypertension in primary hyperparathyroidism—pre- and posteroperative blood pressure in 75 cases. *Klin Wochenschr*, 1984; **62**: 465–9.
120. Farahnak P, Ring M, Caldahl K, Farnebo LO, Eriksson MJ, Nilsson IL. Cardiac function in mild primary hyperparathyroidism and the outcome after parathyroidectomy. *Eur J Endocrinol*, 2010; **163**: 461–7.
121. Person A, Bollesleve J, Rosen T, *et al*. Effect of surgery on cardiac structure and function in mild primary hyperparathyroidism. *Clin Endocrinol (Oxf)*, 2011; **74**: 174–80.
122. Roberts WC, Waller BF. Effect of chronic hypercalcemia on the heart. An analysis of 18 necropsy patients. *Am J Med*, 1981; **71**: 371–84.
123. Streeten EA, Munir K, Hines S, *et al*. Coronary artery calcification in patients with primary hyperparathyroidism in comparison with control subjects from the multi-ethnic study of atherosclerosis. *Endocr Pract*, 2008; **14**: 155–61.
124. Kepez A, Harmanci A, Hazirolan T, *et al*. Evaluation of subclinical coronary atherosclerosis in mild asymptomatic primary hyperparathyroidism. *Int J Cardiovasc Imaging*, 2009; **25**: 187–93.
125. Iwata S, Walker MD, Di Tullio MR, *et al*. Aortic valve calcification in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2012; **97**: 132–7.
126. Walker MD and Silverberg SJ. Cardiovascular aspects of primary hyperparathyroidism. *J Endocrinol Investig*, 2008; **31**: 925–31.
127. Nuzzo V, Tauchmanova L, Fonderico F, *et al*. Increased intima-media thickness of the carotid artery wall, normal blood pressure profile and normal left ventricular mass in subjects with primary hyperparathyroidism. *Eur J Endocrinol*, 2002; **147**: 453–9.
128. Nilsson IL, Aberg J, Rastad J, Lind L. Left ventricular systolic and diastolic function and exercise testing in primary hyperparathyroidism-effects of parathyroidectomy. *Surgery*, 2000; **128**: 895–902.
129. Piovesan A, Molineri N, Casasso F, *et al*. Left ventricular hypertrophy in primary hyperparathyroidism. Effects of successful parathyroidectomy. *Clin Endocrinol (Oxf)*, 1999; **50**: 321–8.
130. Almqvist EG, Bondeson AG, Bondeson L, *et al*. Cardiac dysfunction in mild primary hyperparathyroidism assessed by radionuclide angiography and echocardiography before and after parathyroidectomy. *Surgery*, 2002; **132**: 1126–32.
131. Farahnak P, Ring M, Caldahl K, Farnebo LO, Eriksson MJ, Nilsson IL. Cardiac function in mild primary hyperparathyroidism and the outcome after parathyroidectomy. *Eur J Endocrinol*, 2010; **163**: 461–7.
132. Walker MD, Fleischer JB, DiTullio MR, *et al*. Cardiac structure and diastolic function in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2010; **95**: 2172–9.
133. Stefenelli T, Abela C, Frank H, *et al*. Cardiac abnormalities in patients with primary hyperparathyroidism: implications for follow-up. *J Clin Endocrinol Metab*, 1997; **82**: 106–112.
134. Persson A, Bollerslev J, Rosen T, *et al*. Effect of surgery on cardiac structure and function in mild primary hyperparathyroidism. *Clin Endocrinol (Oxf)*, 2011; **74**: 174–80.
135. Luigi L, Chiara FM, Laura Z, *et al*. Arterial hypertension, metabolic syndrome and subclinical cardiovascular organ dame in patients with asymptomatic primary hyperparathyroidism before and after parathyroidectomy: preliminary results. *Int J Endocrinol*, 2012; 2012: 408295.
136. Rubin MR, Rundek T, McMahon DJ, Lee HS, Sacco RL, Silverberg SJ. Carotid artery plaque thickness is associated with increased serum calcium levels. *Atherosclerosis*, 2007; **194**: 426–32.
137. Walker M, Fleischer J, DiTullio MR, *et al*. Cardiac structure and diastolic function in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2010; **95**: 2172–9.
138. Walker MD, Rundek T, Homma S, DiTullio M, Iwata S, Lee JA, *et al*. Effect of parathyroidectomy on subclinical cardiovascular disease in mild primary hyperparathyroidism. *Eur J Endocrinol*, 2012; **167**: 277–85.
139. Baykan M, Erem C, Erdogan T, *et al*. Assessment of left ventricular diastolic function and the Tei index by tissue Doppler

imaging in patients with primary hyperparathyroidism. *Clin Endocrinol (Oxf)*, 2007; **66**: 483–8.
140. Walker, MD, Carotid vascular abnormalities in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2009; **94**: 3849–56.
141. Baykan M, Erem C, Erdogan T, *et al*. Impairment of flow mediated vasodilatation of brachial artery in patients with primary hyperparathyroidism. *Int J Cardiovasc Imaging*, 2007; **23**: 323–8.
142. Smith JC, Page MD, John R, *et al*. Augmentation of central arterial pressure in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2000; **85**: 3515–19.
143. Rubin MR, Maurer MS, McMahon DJ, Bilezikian JP, Silverberg SJ. Arterial stiffness in mild primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2005; **90**: 3326–30.
144. Yu N, Donnan PT, Flynn RW, *et al*. Increased mortality and morbidity in mild primary hyperparathyroid patients. The Parathyroid Epidemiology and Audit Research Study (PEARS). *Clin Endocrinol (Oxf)*, 2010; **73**: 30–4.
145. Michels KB, Xue F, Brandt L, Ekbom A. Hyperparathyroidism and subsequent incidence of breast cancer. *Int J Cancer*, 2004; **110**: 149–51.
146. Wermers RA, Khosla S, Atkinson EJ, Grant CS, Hodgson SF, O'Fallon M, Melton LJ. Survival after the diagnosis of hyperparathyroidism. *Am J Med*, 1998; **104**: 115–22.
147. Pepe J, Cipriani C, Sonato C, Raimo O, Biamonte F, Minisola S. Cardiovascular manifestations of primary hyperparathyroidism: a narrative review. *Eur J Endocrinol*, 2017; **177**: R297–308.
148. Hedback G, Oden A. Increased risk of death from primary hyperparathyroidism—an update. *Eur J Clin Invest*, 1998; **28**: 271–6.
149. Nilsson IL, Yin L, Lundgren E, Rastad J, Ekbom A. Clinical presentation of PHPT in Europe-nationwide cohort analysis on mortality from nonmalignant causes. *J Bone Miner Res*, 2002; **17**(Suppl2): N68–74.
150. Hedback G, Oden A, Tisell L-E. Parathyroid adenoma weight and the risk of death after treatment for primary hyperparathyroidism. *Surgery*, 1995; **117**: 134–9.
151. Albright F. Reinfenstein E. *Parathyroid Glands and Betabolic Bone Disease* Baltimore, MD: Williams & Wilkins, 1948: 46–114.
152. Pawlowska M, Cusano NE. An overview of normocalcemic primary hyperparathyroidism. *Curr Opin Endocrinol Diabetes Obes*, 2015; **22**: 413–21.
153. Tordjman KM, Greenman Y, Osher E, Shenkerman G, Stern N. Characterization of normocalcemic primary hyperparathyroidism. *Am J Med*, 2004; **117**: 861–3.
154. Tuna MM, Caliskan M, Unal M, *et al*. Normocalcemic hyperparathyroidism is associated with complications similar to those of hypercalcemic hyperparathyroidism. *J Bone Miner Res*, 2016; **34**: 331–5.
155. Amaral LM, Queiroz DC, Marques TG, Mendes M, Bandeira F. Normocalcemic versus hypercalcemic hyperparathyroidism: more stone than bone? *J Osteoporos*, 2012; **2012**: 128352.
156. Corbetta S. Normocalcemic hyperparathyroidism. *Front Horm Res*, 2019; 51: 23–39.
157. Bilezikian JP. Primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2018; **103**: 3993–4004.
158. Bilezikian JP, Brandi ML, Eastell R, *et al.* Guidelines for the management of asymptomatic primary hyperparathyroidism. In: Bilezikian JP, ed. *The Parathyroids*, 3rd edn. San Diego, CA: Elsevier, 2015: 489–97.
159. Udelsman R, Akerstrom G, Biagini C, et al. The surgical management of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3595–606.
160. Silverberg SJ, Bilezikian JP. Primary hyperparathyroidism: physiology and surgical indications. In: Randolph GW, ed. *Surgery of the Thyroid and Parathyroid Glands*, 3rd edn. Philadelphia, PA: Elsevier-Saunders, 2019: 1219–37.
161. Stack BC Jr, Tolley NS, Bartel TB, *et al*. AHNS Series: do you know your guidelines? Optimizing outcomes in re-operative parathyroid surgery: definitive multidisciplinary joint consensus guidelines of the American Head and Neck Society and the British Association of Endocrine and Thyroid Surgeons. *Head Neck*, 2018; **40**: 1617–29.
162. Claflin J, Dhir A, Expinosa NM, *et al*. Intraoperative parathyroid hormone levels < 40 pg/mL are associated with the lowest persistence rates after parathyroidectomy for primary hyperparathyroidism. *Surgery*, 2019; **166**: 50–4.
163. Udelsman R, Lin Z, Donovan P. The superiority of minimally invasive parathyroidectomy based on 1650 consecutive patients with primary hyperparathyroidism. *Ann Surg*, 2011; **253**; 585–91.
164. Karahan O, Okus A, Sevinc B, *et al*. Minimally invasive parathyroidectomy under local anesthesia. *J Postgrad Med*, 2013; **59**: 21–4.
165. Morris LF, Lee S, Warneke CL, *et al*. Fewer adverse events associated with minimally invasive parathyroidectomy. *Am J Surg*, 2014; **208**: 850–5.
166. Chen H, Pruhs Z, Starling JR, Mack E. Intraoperative parathyroid hormone testing improves cure rates in patients undergoing minimally invasive parathyroidectomy. *Surgery*, 2005; **138**: 583–7.
167. Kunstman JW, Udelsman R. Superiority of minimally invasive parathyroidectomy. *Adv Surg*, 2012; **46**: 171–89.
168. Van Udelsman B, Udelsman R. Surgery in primary hyperparathyroidism: extensive personal experience. *J Clin Densitom*, 2013; **16**: 54–9.
169. Udelsman R, Donovan P, Shaw C. Cure predictability during parathyroidectomy. *World J Surg*, 2014; **38**: 525–33.
170. Cohen MS, Dilley WG, Wells SA, *et al*. Long-term functionality of cryopreserved parathyroid autografts: 13-year prospective analysis. *Surgery*, 2005; **138**: 1033–40.
171. Lebastchi AH, Donovan PI, Udelsman R. Paradigm shift in the surgical management of multigland parathyroid hyperplasia: an individualized approach. *JAMA Surg*, 2014; **149**: 1133–77.
172. Storvall S, Rhyhanen E, Bensch FV, *et al*. Recurrent metastasized parathyroid carcinoma-long-term remission after combined treatments with surgery, radiotherapy, cinacalcet, zoledronic acid, and temozolomide. *JBMR Plus,* 2018; **3**: e10114.
173. Kunstman JW, Kirsch JD, Mahajan A, Udelsman R. Clinical review: parathyroid localization and implications for clinical management. *J Clin Endocrinol Metab*, 2013; **98**: 902–12.
174. Wachtel H, Bartlett Ek, Kelz RR, Cerullo I, Karakousis GC, Fraker DL. Primary hyperparathyroidism with negative imaging: a significant clinical problem. *Ann Surg*, 2014; **260**: 474–80.
175. Kedarisetty S, Fundakowski C, Ramakrishnan K, Dadparvar S. Clinical value of Tc99m-MIBI SPECT/CT versus 4D-CT or US

in management of patients with primary hyperparathyroidism. *Ear Nose Throat*, 2019; **98**: 149–57.
176. Zafereo M, Yu J, Angelos P, *et al*. American Head and Neck Society Endocrine Surgery Section update on parathyroid imaging for surgical candidates with primary hyperparathyroidism. *Head Neck*, 2019; **41**: 2398–409.
177. Van Husen R, Kim LT. Accuracy of surgeon-performed ultrasound in parathyroid localization. *World J Surg*, 2005; **28**: 1122–6.
178. Saengsuda Y. The accuracy of 99m Tc-MIBI scintigraphy for preoperative parathyroid localization in primary and secondary-tertiary hyperparathyroidism. J Med Assoc Thai, 2012; **95**: Suppl 3: S81–91.
179. Singer MC, Pucar D, Mathew M, Terris DJ. Improved localization of sestamibi imaging at high volume centers. *Laryngoscope*, 2013; **123**: 298–301.
180. Ciappuccini R, Morera J, Pascal P, *et al*. Dual-phase 99m Tc sestamibi scintigraphy with neck and thorax SPECT/CT in primary hyperparathyroidism: a single institution experience. *Clin Nucl Med*, 2012; **37**: 223–8.
181. Pata G, Casella C, Besuzio S, Mitternpergher F, Salemi B. Clinical appraisal of 99m technetium-sestamibi SPECT/CT compared to conventional SPECT in patients with primary hyperparathyroidism and concomitant nodular goiter. *Thyroid*, 2010; **20**: 1121–7.
182. Hunter GJ, Schellinger D, Vu TH, Perrier ND, Hamberg LM. Accuracy of four-dimensional CT for the localization of abnormal parathyroid glands in patients with primary hyperparathyroidism. *Radiology*, 2012; **264**: 789–95.
183. Starker LF, Mahajan A, Bjorklund P, Sze G, Udelsman R, Carling T. 4D parathyroid CT as the initial localization study for patients with de novo primary hyperparathyroidism. *Ann Surg Oncol*, 2011; **18**: 1723–8.
184. Roach PJ, Schembri GP, Ho Shon JA, Bailey EA, Bailey DL. SPECT/CT scanner for anatomical localization; impact on diagnostic accuracy and reporter confidence in clinical practice. *Nucl Med Commun*, 2006; **27**: 977–87.
185. Yeh R, Tay YD, Tabacco G, *et al*. Diagnostic performance of 4D CT and sestamibi SPECT/CT in localizing parathyroid adenomas in primary hyperparathyroidism. *Radiology*, 2019; **29**: 469–76.
186. Grayev AM, Gentry LR, Hartman MJ, Chen H, Perlman SB, Reeder SB. Presurgical localization of parathyroid adenomas with magnetic resonance imaging at 3.0 T: an adjunct method to supplement traditional imaging. *Ann Surg Oncol*, 2012; **19**: 981–9.
187. Khafif A, Masalha M, Landsberg R, *et al*. The role of F18-fluorocholine positron emission tomography/magnetic resonance imaging in localizing parathyroid adenomas. *Eur Arch Otorhinolaryngol*, 2019; **276**: 1509–16.
188. Udelsman R, Donovan PI. Remedial parathyroid surgery: changing trends in 130 consecutive cases. *Ann Surg*, 2006; **244**: 471–9.
189. Silverberg SJ, Shane E, Jacobs TP, Siris E, Bilesikian JP. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. *N Engl J Med*, 1999; **341**: 1249–55.
190. Rubin MR, Bilezikian JP, McMahon DJ, *et al*. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. *J Clin Endocrinol Metab*, 2008; **93**: 3462–70.
191. Cusano NE, Cipriani C, Bilezikian JP. Management of normocalcemic primary hyperparathyroidism. *Best Pract Res Clin Endocrinol Metab*, 2018; **32**: 837–45.
192. Seibel MJ, Gartenberg F, Silverberg SJ, *et al*. Urinary hydroxypyridinium crosslinks of collagen in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 1992; **74**: 481–6.
193. Bollerslev J, Jansson S, Mollerup CL, *et al*. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. *J Clin Endocrinol Metab*, 2007; **92**: 1687–92.
194. Ambrogini E, Cetani F, Cianferotti L, *et al*. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. *J Clin Endocrinol Metab*, 2007; **92**: 3114–21.
195. Tamura Y, Araki A, Chiba Y, *et al*. Remarkable increase in lumbar spine bone mineral density and amelioration in biochemical markers of bone turnover after parathyroidectomy in elderly patients with primary hyperparathyroidism: a 5-year follow-up study. *J Bone Miner Metab*, 2007; **25**: 226–31.
196. Sitges-Serra A, Garcia L, Prieto R, Perla MJ, Nogues X, Sancho JJ. Effect of parathyroidectomy for primary hyperparathyroidism on bone mineral density in postmenopausal women. *Br J Surg*, 2010; **97**: 1012–19.
197. Fuleihan GEH, Moore F Jr, LeBoff MS, *et al*. Longitudinal changes in bone density in hyperparathyroidism. *J Clin Densitometry*, 1999; **2**: 153–62.
198. Sankaran S, Gamble G, Bolland M, Reid IR, Grey A. Skeletal effects of interventions in mild primary hyperparathyroidism: a meta-analysis. *J Clin Endocrinol Metab*, 2010; **95**: 1653–62.
199. Vestergaard P, Mosekilde L. Parathyroid surgery is associated with a decreased risk of hip and upper arm fractures in primary hyperparathyroidism: a controlled cohort study. *Ann Int Med*, 2004; **255**: 108–14.
200. Kulak CAM, Bandeira C, Voss D, *et al*. Marked improvement in bone mass after parathyroidectomy in osteitis fibros cystica. *J Clin Endocrinol*, 1998; **83**: 732–5.
201. Stock JL and Marcus R. Medical management of primary hyperparathyroidism. In: Bilezikian JP, Marcus R, Levine MA, eds. *The Parathyroids*, 2nd edn. San Diego, CA: Academic Press, 2001: 459–74.
202. Locker FG, Silverberg SJ, Bilezikian JP. Optimal dietary calcium intake in primary hyperparathyroidism. *Am J Med*, 1997; **102**: 543–50.
203. Marcus R. The role of estrogen and related compounds in the management of primary hyperparathyroidism. *J Bone Miner Res*, 2002; **17**(Suppl. 2): N146–9.
204. Rubin MR, Lee K, McMahon DJ, Silverberg SJ. Raloxifene lowers serum calcium and markers of bone turnover in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2003; **88**: 1174–8.
205. Zanchetta JR, Bogado CE. Raloxifene reverses bone loss in postmenopausal women with mild asymptomatic primary hyperparathyroidism. *J Bone Miner Res*, 2001; **16**: 189–90.
206. Grey A, Lucas J, Horne A, Gamble G, Davidson JS, Reid I. Vitamin D repletion in patients with primary hyperparathyroidism and coexistent vitamin D deficiency. *J Clin Endocrinol Metab*, 2005; **90**: 2122–6.
207. Rolighed L, Rajnmark L, Sikjaer T, *et al*. Vitamin D treatment in primary hyperparathyroidism: a randomized placebo controlled trial. *J Clin Endocrinol Metab*, 2014; **99**: 1072–80.
208. Amstrup AK, Rejnmark L, Vestergaard P, *et al*. Vitamin D status, physical performance and body mass in patients surgically cured for primary hyperparathyroidism compared with healthy controls—a cross-sectional study. *Clin Endocrinol (Oxf)*, 2011; **74**: 130–6.

209. Rolighed L, Rejnmark L, Sikjaer T, *et al*. No beneficial effects of vitamin D supplementation on muscle function or quality of life in primary hyperparathyroidism: results from a randomized controlled trial. *Eur J Endocrinol*, 2015; **172**: 609–17.
210. Reasner CA, Stone MD, Hosking DJ, Ballah A, Mundy GR. Acute changes in calcium homeostasis during treatment of primary hyperparathyroidism with risedronate. *J Clin Endocrinol Metab*, 1993; **77**: 1067–71.
211. Rossini M, Gatti D, Isaia G, Sartori L, Braga V, Adami S. Effects of oral alendronate in elderly patients with osteoporosis and mild primary hyperparathyroidism. *J Bone Miner Res*, 2001; **16**: 113–19.
212. Hassani S, Braunstein GD, Seibel MJ, *et al*. Alendronate therapy of primary hyperparathyroidism. *The Endocrinologist*, 2001; **11**: 459–64.
213. Parker CR, Blackwell PJ, Fairbairn KJ, Hosking DJ. Alendronate in the treatment of primary hyperparathyroid-related osteoporosis. *J Clin Endocrinol Metab*, 2002; **87**: 4482–9.
214. Chow CC, Chan WB, Li JKY, *et al*. Oral alendronate increases bone mineral density in postmenopausal women with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2003; **88**: 581–7.
215. Khan AA, Bilezikian JP, Kung AWC, *et al*. Alendronate in primary hyperparathyroidism: a double-blind, randomized, placebo-controlled trial. *J Clin Endocrinol Metab*, 2004; **89**: 3319–25.
216. Khan AA, Bilezikian JPO, Kung A, Syed ZA, Dubois SJ, Standish TI. Alendronate in men with primary hyperparathyroidism. *Endocr Pract*, 2009; **15**: 705–13.
217. Marcocci C, Cianferotti L, Silverberg SJ, Bilezikian JP. Primary hyperparathyroidism in men. In: Orwoll E, Bilezikian JP, Vanderscheuren D, eds. *Osteoporosis in Men*. San Diego, CA: Elsevier, 2010: 465–78.
218. Yeh MW, Zhou H, Adams AL, Ituarte PH, Li N, Liu IL, Haigh PI. The relationship of parathyroidectomy and bisphosphonates with fracture risk in primary hyperparathyroidism: an observational study. *Ann Intern Med*, 2016; **164**: 715–23.
219. Eller-Vainicher C, Palmieri S, Cairoli E, *et al*. Protective effect of denosumab on bone in older women with primary hyperparathyroidism. *J Am Geriatr Soc*, 2018; **66**: 518–24.
220. Messa P, Alfieri C, Brezzi B. Clinical utilization of cinacalcet in hypercalcemic conditions. *Expert Opin Drug Metab Toxicol*, 2011, **7**: 517–28.
221. Cetani F, Marcocci C. Cinacalcet in the management of primary hyperparathyroidism. *Expert Rev Endocrinol Metab*, 2012; **7**: 45–53.
222. Shoback DM, Bilezikian JP, Turner SA, McCary LC, Guo MD, Peacock M. The calcimimetic AMG 073 normalizes serum calcium in patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2003; **88**: 5644–9.
223. Peacock M, Bilezikian JP, Klassen PS, Guo MD, Turner SA, Shoback DM. Cinacalcet hydrochloride maintains long-term normocalcemia in patients with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2005; **90**: 135–41.
224. Peacock M, Bolognese MA, Borofsky M, *et al*. Cinacalcet treatment of primary hyperparathyroidism: biochemical and bone densitometric outcomes in a five-year study. *J Clin Endocrinol Metab*, 2009; **94**: 4860–7.
225. Peacock M, Bilezikian JP, Bolognese MA, *et al*. Cinacalcet HCl reduces hypercalcemia in primary hyperparathyroidism across a wide spectrum of disease severity. *J Clin Endocrinol Metab*, 2011; **96**: E9–18.
226. Cetani F, Saponaro F, Banti C, *et al*. Cinacalcet efficacy in patients with moderately severe primary hyperparathyroidism according to the European Medicine Agency prescription labeling. *J Endocrinol Invest*, 2012; **35**: 655–60.
227. Marcocci C, Chanson P, Shoback D, *et al*. Cinacalcet reduces serum calcium concentrations in patients with intractable primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2009; **94**: 2766–72.
228. Marcocci C, Cetani F, Rubin MR, Silverberg SJ, Pinchera A, Bilezikian JP. Parathyroid cancer. *J Bone Min Res*, 2008; **23**: 1869–80.

4.4

Familial Hypocalciuric Hypercalcaemia Types 1–3 and Neonatal Severe Primary Hyperparathyroidism

Muriel Babey and Dolores M. Shoback

Introduction

Familial hypocalciuric hypercalcaemia (FHH) is a benign autosomal dominant disorder of calcium homeostasis most often caused by a loss-of-function mutation in the *CASR* gene encoding the extracellular calcium-sensing receptor (CaSR) [1]. This form of FHH is now called FHH1. In FHH1, the mutant CaSR in the parathyroid glands responds less potently to changes in extracellular calcium concentration, leading to higher levels of parathyroid hormone (PTH), associated with an altered calcium set-point for the regulation of PTH secretion (**Figure 4.4.1**). CaSRs in the kidney directly inhibit renal calcium reabsorption, thereby counteracting the effect of PTH)[2]. An inactivating mutation in G alpha 11, one of the guanine nucleotide binding (G) proteins which links the CaSR to activation of phospholipase C, is responsible for FHH2 [3]. The phenotype of FHH type 1 and 2 is characterized biochemically by mild-to-moderate, PTH-dependent hypercalcaemia and relative to absolute hypocalciuria)[4]. FHH3 is due to a loss-of-function mutation of adaptor-related protein complex 2, sigma 1 subunit (AP2S1) which is involved in clathrin-mediated endocytosis of G-protein-coupled receptors)[5]. The FHH3 phenotype may be the more severe variant with symptomatic hypercalcaemia, low bone mineral density, and cognitive deficits in some affected individuals in childhood [6, 7].

Mutations in the *CASR, G11* or *AP2S1* are typically heterozygous and lead to loss of protein function with their predominant phenotypic features affecting parathyroid and kidney function. Individuals with FHH1 and FHH2 are usually asymptomatic. Neonates with homozygous mutations in the *CASR* tend to lack functioning CaSRs and present at the other end of the clinical spectrum with life-threatening hypercalcaemia—a disorder called neonatal severe hyperparathyroidism (NSHPT). In addition to the common presentation of FHH, occasionally patients with heterozygous inactivating mutations of *CASR* can present as NSHPT)[8], neonatal hyperparathyroidism (NHPT) [9], which is intermediate in severity between FHH and NSHPT, but is a more severe degree of hypercalcaemia in infancy and childhood than usually encountered in FHH [10], or mild hypercalcaemia characteristic of primary hyperparathyroidism (PHPT) [11].

This chapter will review key concepts in calcium homeostasis and the role of the CaSR and associated proteins Gα11 and AP2σ in CaSR function. The clinical and laboratory features of FHH types 1–3, NSHPT, NHPT, and severe hypercalcaemia in infancy will be discussed. We highlight the differential diagnosis of PTH-dependent forms of hypercalcaemia (especially PHPT), the molecular pathogenesis of CaSR signalling defects that produce hypercalcaemia, and the treatment and monitoring for these conditions.

Calcium Homeostasis and the Role of the CaSR

The extracellular calcium concentration is tightly regulated, given its involvement in many key biological processes such as muscle and heart contraction, neuronal excitability, bone mineralization, and hormone secretion. The CaSR plays a key role in calcium homeostasis by regulating the secretion of PTH from the parathyroid glands and the reabsorption of urinary calcium by the kidney. The transfer of calcium between the environment and extracellular fluid takes place in the intestine. Bone constitutes the main

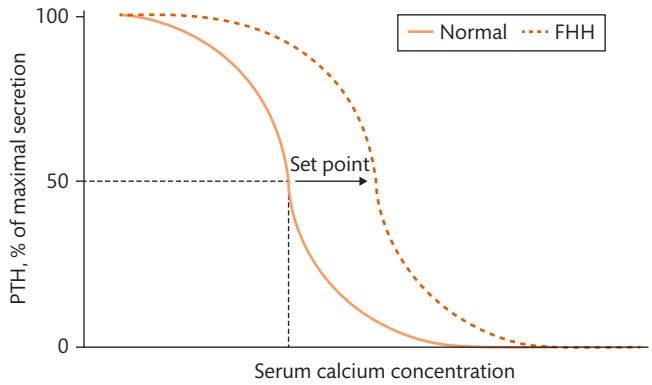


Figure 4.4.1 Relationship between levels of serum calcium and plasma PTH. Depicted is the inverse sigmoid relationship between levels of serum calcium and plasma PTH in patients with familial hypocalciuric hypercalcaemia (FHH; dotted line) and normal individuals (solid line). The calcium set-point is defined as the extracellular calcium level which produces half-maximal inhibition of PTH secretion. FHH results in a shift of the curve to the right. Increased levels of serum calcium are needed to suppress plasma PTH levels in FHH. The therapeutic effect of calcimimetic leads to a shifting of the set-point to the left.

calcium reservoir. Calciotropic hormones such as PTH and 1,25-hydroxyvitamin D_3 mediate changes in calcium homeostasis via actions in bone, kidney, intestine and parathyroid gland [12]. The relationship between extracellular calcium and PTH levels is represented by an inverse sigmoidal curve (Figure 4.4.1). The activity and/or the expression level of parathyroid CaSRs predicts the calcium set-point, defined at which the extracellular calcium level at which PTH secretion is half-maximal.

CaSRs serve as the calcium-sensing mechanism *in vivo*. CaSRs are predominantly expressed in the chief cells of the parathyroid gland and along the whole nephron and are also found in other tissues such as pancreatic islet, intestine, bone, breast and skin and in some areas of the brain [13–16]. The CaSR is a G-protein coupled receptor (GPCR) and belongs to the Family C of this receptor superfamily. The CaSR belongs to the group II subfamily, which contains the CaSR, the vomeronasal receptors (VRs) and odorant receptors, and has more than 20% amino acid identity over their transmembrane domains (TMD) comprising the 7-TM-spanning region.

The human *CASR* gene maps to 3q13.3-21 and has eight exons. Six exons, exon 2–6, encode the CaSR protein which is made up of 1078 amino acids. The large extracellular domain (ECD) of the CaSR has approximately 600 amino acids including the bi-lobed Venus flytrap-like domain (VFT). The VFT is responsible for ligand binding and connects via cysteine-rich region and peptide linker to the seven TMD. The intracellular COOH-terminal tail is 216 amino acids long [17]. The CaSR functions as a dimer on the cell surface, which is in part due to intermolecular disulphide bonds in the ECD [18]. Three to five calcium ions bind to the cleft of each VFT and cause closure of the VFT lobes with ensuing rotation of the VFT and a transfer of a conformational change to the cysteine-rich region of a CaSR dimer. Upon activation, the CaSR forms a novel dimer interface between subunits [19]. The conformational change reconfigures some TM-helices such that intracellular loops contact G proteins and initiate cell signalling. In addition to calcium, other polycations such as magnesium and barium, charged polyvalent molecules such as spermine, and β-amyloid peptides can activate the CaSR [20]. The sensitivity to calcium can also be positively modulated by pH, ionic strength, and aromatic L-α-amino acids such as phenylalanine, tryptophan, and histidine [19, 20]. According to the crystal structures of the CaSR, there are not only multiple binding sites for calcium, but unexpectedly also for phosphate ions. While calcium ions promote the active state, phosphate ions stabilize the inactive confirmation [19].

After ligand binding and likely dimerization of individual receptor molecules and conformational changes, the CaSR couples to multiple intracellular signalling pathways, depending on the target cell. CaSRs can couple with several G proteins (Gi, Gq/11, G 12/13). The CaSR stimulates activation of phospholipase A2, C, and D as well as various mitogen-activated protein kinases (MAPKs) and inhibits adenylate cyclase [9]. In particular, G-protein subunit α11 stimulates phospholipase C-β (PLC-β) activity (Figure 4.4.2) [21, 22]. PLC-β increases inositol 1,4,5-triphosphate, a mediator of rapid mobilization of calcium from intracellular stores [23], whereas the diacylglycerol (DAG) produced activates protein kinase C (PKC). PKC in turn initiates the MAPK pathway (Figure 4.4.2) [24].

Current thinking, based on work in model *in vitro* systems with high levels of receptor expression, is that cell surface CaSR expression is regulated by two main mechanisms. The first mechanism has been called agonist-driven insertional signalling or ADIS [25]. ADIS leads to increased anterograde trafficking of newly synthesized receptors to the plasma membrane after prolonged exposure to calcium. The second mechanism depends on an endocytic complex comprising clathrin, β-arrestin and the AP2 complex [5]. The endocytic complex transports CaSRs to the endosomal-lysosomal degradation pathway or recycles them to the cell surface through retrograde trafficking [26] (Figure 4.4.2).

Elevated serum calcium levels are sensed by CaSRs, and activation of downstream signalling leads to downregulation of PTH secretion [1, 2]. The expression of the CaSR on the cell surface seems to be upregulated by extracellular calcium and 1,25-dihyroxyvitamin D_3 [27, 28]. Decreased PTH secretion will reduce bone resorption, urinary calcium reabsorption [15–19], and renal synthesis of 1,25-hydroxyvitamin D_3, thereby decreasing intestinal calcium absorption. CaSRs respond to elevated calcium levels in the kidney, and urinary calcium reabsorption by PTH in the cortical thick ascending limb decreases by interfering directly with sodium/chloride reabsorption [29, 30]. The interference with sodium/chloride reabsorption impairs the generation of the medullary osmotic gradient which is critical for urinary concentration. In the proximal tubule, CaSRs reduce the inhibitory effect of PTH on renal phosphate reabsorption [31]. CaSRs also reduce the urinary concentrating ability in the inner medullary collecting duct by counteracting the action of vasopressin [32]. A decrease in serum or extracellular calcium reduces CaSR activation.

Familial Hypocalciuric Hypercalcaemia

Clinical and Laboratory Features of FHH

Familial hypocalciuric hypercalcaemia was first described in 1966 [33]. The disorder was initially named 'familial benign hypercalcaemia' in the case description of a 7-year-old boy and his extended 11-member family by Foley *et al.* in 1972 [34]. The distinctive clinical features were lack of clinical symptoms, mild hypercalcaemia, normal levels of PTH, normal or low serum phosphate for age, and

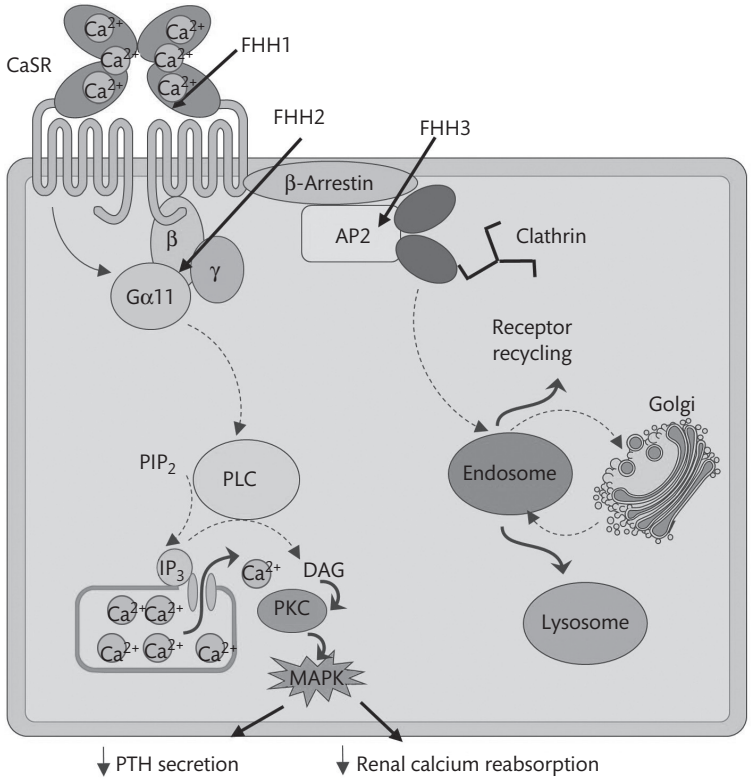


Figure 4.4.2 Role of the CaSR, Gα11 and AP2 complex in regulation of PTH secretion and renal calcium reabsorption. Multiple calcium ions bind to the extracellular bi-lobed Venus fly trap (VFT) domain of the CaSR. Upon activation, CaSR binds to the G11 and Gq (not shown) proteins which dissociate into the respective Gα-subunits and Gβγ heterodimers. Gα11 stimulates phospholipase C (PLC). PLC catalyses the formation of inositol 1,4,5-triphosphate (IP3) and diacylglycerol (DAG) from phosphatidylinositol 4,5-biphosphate (PIP2). IP3 accumulates and leads to a rapid release of calcium intro the cytosol from intracellular stores. DAG activates protein kinase C (PKC) which in turn initiates the mitogen-activated protein kinase (MAPK) pathway. Ultimately, these intracellular signalling pathways decrease PTH secretion from parathyroid chief cells and reduce renal tubular calcium absorption. CaSR cell surface expression is dependent on the endocytic complex, which consists of clathrin, β-arrestin and the AP2 complex, and on agonist-driven insertional signalling (ADIS) (not shown). The endocytic complex transports the CaSR to the endosomal-lysosomal degradation pathway or recycles it back to the cell surface. Loss-of-function mutations in CASR lead to FHH type 1, whereas loss-of-function mutations of GNA11 are associated with FHH type 2. FHH type 3 is caused by loss-of-function mutations of AP2S1.

low urinary calcium excretion. The pattern of inheritance in that family was compatible with an autosomal dominant mode of transmission. The authors postulated that there might be an abnormality in the calcium-sensing mechanism in the parathyroid glands [34]. Marx *et al.* examined kindreds with the same syndrome and named it FHH to highlight the absolute to relative hypocalciuria for the degree of hypercalcaemia being demonstrated [4].

The prevalence of FHH is estimated to be 1% or less to that of primary hyperparathyroidism (PHPT). Given its benign clinical features which may lead to cases escaping medical detection, FHH may be more common than the estimated prevalence of ~1/78 000 from data from developed countries [35]. FHH is inherited in an autosomal dominant manner and approaches a 100% penetrance of the trait.

Patients are usually asymptomatic despite chronic hypercalcaemia. Typical manifestations of hypercalcaemia (nausea, anorexia, constipation, nephrolithiasis, nephrocalcinosis) are usually not present [36]. Even patients with FHH with the higher serum calcium levels, are often asymptomatic. Chondrocalcinosis with advanced age and cases of pancreatitis associated with FHH1 have been described [37, 38]. Patients with three most frequent FHH2 mutations (p.I200del, p.L135Q, p.T54M) display a mild FHH phenotype with serum adjusted calcium concentrations of less than 2.8 mmol/L [3, 39]. FHH3 patients may experience symptomatic hypercalcaemia, reduced fractional excretion of calcium, low bone mineral density and cognitive dysfunction, and/or behavioural disturbances in children with the p.R15C or p.R15L AP2σ mutations [40, 41]. Adults with FHH3 with the heterozygous germline p.R15C mutation display elevated PTH levels and mild hypophosphatemia, and even osteomalacia after the age of 30 has been reported in at least one family [6, 7].

Biochemically, FHH1 is characterized by lifelong mild-to-moderate hypercalcaemia, inappropriately low urinary calcium excretion, a normal or mildly elevated circulating PTH level, and high-normal to elevated serum magnesium levels. Serum calcium levels for most affected families range from 2.6 to 2.9 mmol/L and average at about 10% above the upper limit of normal. In some kindreds, the serum calcium levels can range from the upper half of the normal range [42] to as high as 3.24–3.49 mmol/L [10]. Serum calcium levels within a kindred are usually clustered within a relatively narrow range [4]. Serum magnesium levels in FHH are measured at the upper half of normal or mildly elevated [4]. Serum magnesium levels correlate positively with the serum calcium concentration [4]. Serum phosphate levels in FHH are usually in the lower half of the normal range and rarely low [43]. Renal function remains intact in FHH despite hypercalcaemia, and renal complications such as nephrocalcinosis are not present [4, 43].

Table 4.4.1 Characteristics of patients with familial hypocalciuric hypercalcaemia type 1–3 (FHH) and primary hyperparathyroidism (PHPT)

Variable	FHH type 1–3	PHPT
Age of onset and gender distribution	At birth, equal gender distribution	Usually >50 years, women affected twice as often as men
Genetic pathogenesis	Loss-of-function germline mutation in *CASR* for FHH type 1, in *GNA11* for FHH type 2 and *AP2S1* for FHH type 3	Most often implicated in the development of sporadic PHPT: somatic mutations in *CCND1* and *MEN1* in parathyroid adenoma/hyperplasia
Symptoms and clinical findings	Usually asymptomatic for FHH type 1 and 2, although occasionally pancreatitis is seen There is possibly a more severe phenotype for FHH type 3, including symptomatic hypercalcaemia, low bone mineral density, and cognitive dysfunction	Asymptomatic in >80% Cortical bone loss, urolithiasis in <20%
Serum levels		
Calcium	Elevated	Elevated
Magnesium	Normal to elevated	Variable, usually normal
Phosphorus	Normal to slightly low	Normal to very low
Intact PTH	Normal, 80–85%; elevated, 15–20%	Elevated, >80%
$1,25(OH)_2D$	Normal	Normal to elevated
Urinary excretion		
Calcium	Normal to low	Normal to elevated
Ca/Cr clearance	Usually <0.01	Usually >0.02
Magnesium	Normal to low	Normal to elevated

Inappropriately low urinary calcium excretion is found in 95% of patients [44]. In a Danish study, 24-hour urinary calcium excretion was measured within the lower range of normal at approximately 2.7446 mmol/24 hours in FHH patients vs. a normal cohort in which it ranged from 1.996 to 7.984 mmol/24 hours [43]. Inappropriately low urinary calcium excretion demonstrates the enhanced renal tubular absorption of calcium. The calcium-to-creatinine clearance ratio (CCCR) which is also known as fractional excretion of calcium is defined by:

$$CCCR = \frac{urinary\ calcium}{plasma\ calcium} \times \frac{plasma\ creatinine}{urinary\ creatinine}.$$

Approximately 80% of individuals with FHH demonstrate a CCCR of <0.01 equivalent to less than 1%, whereas approximately 80% of patients with PHPT have a CCCR higher than 0.01 (Table 4.4.1). Patients with FHH are able to concentrate urine normally, while patients with PHPT show reduced urinary concentrating capacity if they are dehydrated [45].

Approximately 80% of patients with FHH have normal plasma PTH levels, with the other 20% having mildly elevated levels [46]. Elevated PTH levels may be mirroring the functional effect of a specific mutation [47]. 25-hydroxyvitamin levels and 1,25-dihydroxyvitamin levels are generally within normal range [43]. The elevated 1,25-dihyrdoxyvitamin levels of PHPT are usually not encountered and might be due to the generally normal PTH levels in FHH (Table 4.4.1). Bone mineral density is comparable to age-matched controls, even though markers of bone turnover can be mildly elevated [48]. In patients with FHH, bone mineral density does not need to be measured, unless there are other reasons to consider excessive bone loss. The parathyroid glands are usually of normal size and histology in patients with FHH. In a few cases of FHH, a single or multiple parathyroid adenomas or mild chief cell proliferation were shown [49, 50]. Parathyroid localization studies need not be done, unless the case is one of NSHPT or NHPT or the rare patient with FHH with a coexistent superimposed parathyroid adenoma. Most cases of FHH that have gone to surgery demonstrate four-gland parathyroid hyperplasia. After an erroneous subtotal parathyroidectomy for FHH, hypercalcaemia will recur, despite normal appearing parathyroid glands in most affected individuals with FHH.

Differential Diagnosis

The differential diagnosis in FHH is limited to the various aetiologies of PTH-dependent hypercalcaemia, since PTH is not appropriately suppressed for the degree of hypercalcaemia. The distinction between FHH and PHPT is important, given the different management approaches (Table 4.4.1). Parathyroid surgery should be avoided in FHH.

Ten to twenty per cent (10–20%) of patients with PHPT display a CCCR of <0.01 [51]. A similar percentage of FHH patients show a value higher than this level [4]. The value of CCCR providing the optimal diagnostic identification between FHH and PHPT was calculated as 0.0115 (Figure 4.4.3) [52] similar to the clinically used value of 0.01 which delivers 80% sensitivity and 88% specificity. There is an inverse correlation between the serum calcium and magnesium levels in patients with PHPT. Patients with PHPT and concomitant very low calcium intake, vitamin D deficiency, or mild-to-moderate renal insufficiency, or patients taking thiazide diuretic or lithium therapy can have low CCCR. In the evaluation of PTH-dependent (hypocalciuric) hypercalcaemia, these associated factors should be investigated by thorough history-taking and laboratory tests. Only a few patients have been described in the literature with acquired autoimmune hypocalciuric hypercalcaemia due to antibodies that inactivate or dampen CaSR signalling. Other autoimmune disorders in such patients (Hashimoto's thyroiditis or celiac sprue) might hint at that diagnosis.

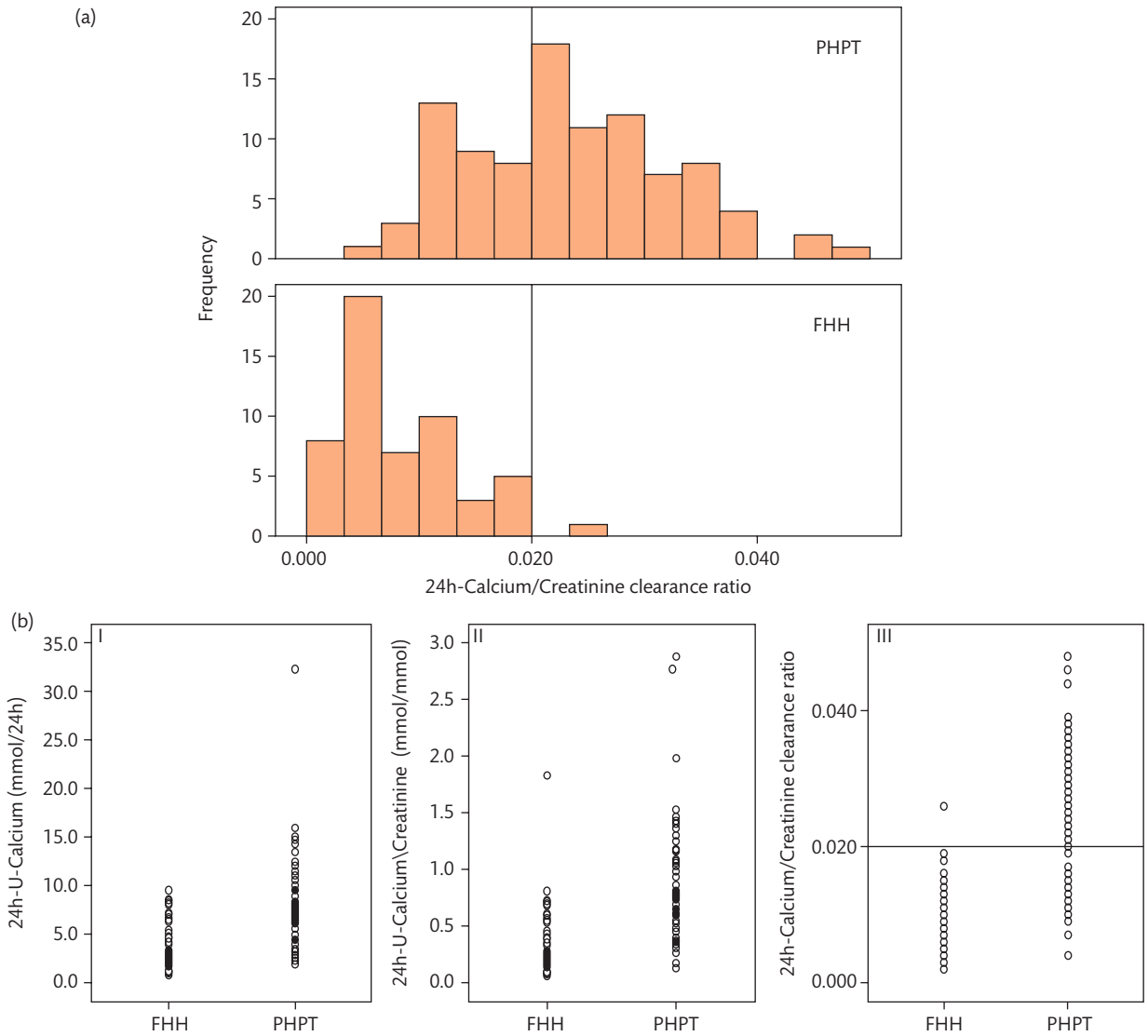


Figure 4.4.3 Ability of three indices of renal calcium excretion to differentiate between familial hypocalciuric hypercalcaemia and primary hyperparathyroidism. (a) The calcium/creatinine clearance ratio (CCCR) in patients with familial hypocalciuric hypercalcaemia (FHH, N = 54) and primary hyperparathyroidism (PHPT, N = 97) overlaps considerably, as shown in the two histograms. A CCCR cut-off value of < 0.02 included 98% of the FHH patients in this study, but a group so defined also included 35% of PHPT patients. (b) Scatterplots of 24-h calcium excretion (CE), 24-h calcium/creatinine excretion (CR) and 24-h CCCR in patients with FHH (N = 54) and PHPT (N = 97) are depicted. Based on overlap analyses, the CCCR misclassified fewer PHPT patients compared to the CE or CR index.

Reproduced with permission from Christensen SE, Nissen PH, Vestergaard P, Heickendorff L, Brixen K, Mosekilde L. Discriminative power of t+I294hree indices of renal calcium excretion for the distinction between familial hypocalciuric hypercalcaemia and primary hyperparathyroidism: a follow-up study on methods. *Clin Endocrinol (Oxf)*. 2008;69(5):713–20 (48).

In patients with PTH-dependent hypercalcaemia and CCCR <0.01 or in the CCCR overlap range of 0.01–0.02 between FHH and PHPT, the diagnosis usually can be made using the family history and biochemical screening of family members, if appropriate, to demonstrate an autosomal dominant pattern of hypercalcaemia in the patient and in two or more additional family members. Recently, a risk equation named Pro-FHH was developed and tested in two prospective cohorts to predict whether a patient has PHPT [53]. Pro-FHH incorporates the serum calcium, plasma PTH, and serum osteocalcin concentrations, and the calcium-to-creatinine clearance ratio calculated from 24-hour urine collection (24h-CCCR). This approach had a 100% specificity and 100% positive predictive value for the diagnosis of PHPT. This approach correctly categorized 60% of patients in one cohort, and the remainder of patients (40%) was recommended to undergo genetic testing. A prospective trial is necessary to assess its usefulness in a larger population and in patients with elevated PTH concentrations.

In patients with PTH-dependent hypercalcaemia in whom the diagnosis of FHH is suspected, screening for *CASR* mutations and mutations in *AP2S1* codon 15 is appropriate. If the screening is negative, then *GNA11* sequencing should be considered [54]. Patients with FHH may not have a family history, either because they may have a *de novo* mutation and/or because family members have not been previously studied. Genetic testing should be performed in children less than 10 years old with PTH-dependent hypocalciuric hypercalcaemia, including in children with NHPT and NSHPT and in patients and associated family members with hypocalciuric hypercalcaemia and recurrent pancreatitis [55]. Inactivating *CASR* mutations have been found in kindreds with familial isolated hyperparathyroidism (FIH)

as well [11, 56]. Their clinical presentation could not be differentiated from that of PHPT including at times the presence of hypercalciuria. Parathyroid surgery in such cases of FIH have led to persistent or recurrent hypercalcaemia in most patients [11].

Molecular Pathogenesis of FHH

FHH encompasses three genetically distinct disorders known as FHH types 1–3.

Familial Hypocalciuric Hypercalcaemia Type 1 (FHH1)

FHH1 (OMIM #145980) is caused by heterozygous loss-of-function mutations in the *CASR* gene on chromosome 3q21 as described here [1]. FHH1 is transmitted as an autosomal dominant disorder and is highly penetrant. FHH1 is the most common cause of the FHH phenotype accounting for approximately 65% of FHH cases in the database from McGill University summarizing many known *CASR* mutations (CaSRdb: calcium-sensing receptor database: https://www.casrdb.mcgill.ca) [57].

More than 130 different mutations of the *CASR* have been reported in patients with FHH1. Most commonly (≥85% of cases), a missense mutation reduces the function of the CaSR. Twenty-eight codons, equivalent to more than 20% of all mutated codons, are found to be mutational hot spots for missense mutations. They are clustered in three regions: the second loop of the ECD, the VFT cleft region with calcium binding sites, and the region encompassing TMD 6 and 7 [57].

The CaSR and its mutant forms have been functionally characterized *in vitro* in overexpression systems in which a mammalian expression vector with a cDNA insert is transfected into HEK293 (human embryonic kidney 293) or COS (CV-1 in Origin with SV40 gene) cells either transiently or stably. The cDNA encodes either a wild-type or mutant CaSR. More physiological renal or parathyroid cell model systems do not exist in which to study the function of the CaSR and its mutant forms. Altered CaSR signalling in renal tubular or parathyroid chief cells *in vivo* leads to altered CaSR signalling and excessive PTH secretion in the presence of hypercalcaemia and to increased renal calcium reabsorption and hypocalciuria, respectively. Mutated residues within the VFT cleft region likely affect binding of calcium or alter the conformational changes of the CaSR upon calcium binding [57, 58]. Specific VFT residues may act as an intramolecular switch to allow access to and binding of calcium within the VFT cleft region. When mutated, this can lead to opposing effects on CaSR signalling [57, 59]. Approximately 50% of mutations in *CASR* in FHH1 result in reduced cell surface CaSR levels due to impaired synthesis and/or increased degradation of the CaSR. Mutant CaSRs can be trapped intracellularly being unable to exit the endoplasmic reticulum (ER) or Golgi apparatus [60, 61], or there may be defective trafficking of the receptor from the ER to the plasma membrane [62]. Some mutant CaSRs are unable to couple to G proteins to activate intracellular cell signalling pathways. In addition, there may be dominant-negative effects of the mutant CaSR to reduce the function of the wild-type CaSR through the dimerization. As noted just here, mature cell surface CaSRs are thought to assemble as dimers [63]. However, coexpressed wild-type CaSRs might also increase transport of mutant CaSR receptors to the plasma membrane via ADIS [64]. Rarely, a nonsense, deletion, insertion, or splice-site mutation will lead to a truncated, inactive CaSR, but this occurs it seems in less than 15% of cases [36].

Familial Hypocalciuric Hypercalcaemia Type 2 (FHH2)

FHH2 (OMIM #145981) is caused by heterozygous loss-of-function mutations in the *GNA11* gene on chromosome 19p13.3, which encodes the $G\alpha_{11}$ protein [3, 39]. $G\alpha_{11}$ protein functions as a signalling partner for the CaSR (21) and is highly expressed in the parathyroid gland [65]. There are three most frequent FHH2 mutations (p.I200del, p.L135Q, p.T54M). $G\alpha_{11}$ mutations that cause FHH2 have given key information that has helped to elucidate the critical functions of the $G\alpha_{11}$ protein. All of the studies on the coupling between CaSR and $G\alpha_{11}$ were conducted in overexpression systems, described here, and not in a parathyroid or renal cellular context. The p.I200del mutation is localized by molecular modelling to the GTPase domain of $G\alpha_{11}$, which undergoes considerable conformational changes upon guanosine triphosphate (GTP) binding [3]. The p.L135Q mutation likely reduces CaSR signal transduction with downstream effectors. The p.T54M mutation likely diminishes coupling and/or dissociation of $G\alpha_{11}$ [39].

Familial Hypocalciuric Hypercalcaemia Type 3 (FHH3)

FHH3 (OMIM #600749) is caused by heterozygous loss-of-function mutations in the *AP2S1* gene on chromosome 19q13.3, which encodes the σ-subunit of the AP2 complex. The AP2 molecular complex is part of clathrin-coated vesicles and promotes endocytosis of GPCRs and other plasma membrane proteins [66]. Over 60 FHH3 patients and families have been reported with *AP2S1* mutations. At present, only three mutations (p.R15C, p.R15H, and p.R15L) have been identified [5, 40, 41, 67, 68]. According to *in vitro* overexpression cell models, these FHH3-causing mutations at the Arg 15 residue likely diminish the interaction between the AP2 complex and the intracellular carboxyl terminus of the CaSR and, therefore, impact the endocytosis of the CaSR [5]. In expression studies *in vitro*, AP2σ mutations change CaSR cell surface expression levels and affect signal transduction in a dominant-negative way [5, 40]. Other AP2σ mutants that are not observed in patients (p.R15G, p.R15P, and p.R15S) have been shown to disrupt cell growth *in vitro* and are likely embryonically lethal [40].

Treatment and Monitoring

The majority of patients with FHH do not need specific treatment due to its typically benign clinical course. Patients with FHH and their affected family members need to be screened and counselled against surgical intervention. Vitamin D and calcium can be supplemented in the same way as for persons of the same age and sex. There is no evidence that supplementation of vitamin D and calcium exacerbates hypercalcaemia or induces hypercalciuria in FHH.

There have been no systematic studies on how often patients with classically mild FHH should be monitored. Serum calcium, serum creatinine, and estimated glomerular filtration rate can be examined annually or biannually. Bone mineral density (BMD) at 3 skeletal sites (lumbar spine, hip, and distal third radius) can be determined according to country-specific guidelines initially, based on age, gender, and menopausal status. One can follow the suggested monitoring recommendations for patients with asymptomatic PHPT who do not meet current guidelines for parathyroid surgery [69]. Less frequent monitoring can also be employed given the typically benign clinical course of FHH. Dual-energy X-ray absorptiometry (DXA) may show reduced cortical bone mass if the

distal third radius site is measured [70]. However, ongoing cortical bone loss is not expected in patients with FHH [71]. A more extensive renal evaluation in FHH patients, as might be done in patients with PHPT, is not needed in patients with FHH, since the risk of renal stone disease and nephrocalcinosis is not significant. Tests of cardiovascular function need not be part of the evaluation or monitoring of patients with FHH for this indication. Monitoring aims to detect changes in the serum calcium level, significant reduction in BMD, or change in renal parameters. There is no increased risk of fracture at non-vertebral (cortical) or vertebral (trabecular) sites due to FHH as a sole aetiology.

Pregnancy in an FHH patient or in the spouse of an FHH carrier requires special consideration.

An unaffected newborn of a mother with hypercalcaemia due to FHH might experience transient hypocalcaemia due to suppression of the infant's PTH secretion *in utero* by the mother's hypercalcaemia. This transient hypoparathyroid state in the infant might last up to two months [72]. An affected newborn of a mother with FHH will be hypercalcaemic and could display features of NSHPT, depending on the specific CaSR mutation that the baby has. The affected newborn of a mother without FHH, but whose father has FHH and conferred the mutation to the offspring, that infant could show severe neonatal hypercalcaemia as well. Under those circumstances, the affected infant who has defective CaSR function is being exposed to a normocalcaemic milieu from its mother. This could produce a form of secondary hyperparathyroidism in the fetus *in utero* as well. The severe nature of the hypercalcaemia in such an infant may be transient. Paediatric endocrinologists and geneticists will need to distinguish this presentation from NSHPT due to homozygous CaSR mutations.

Efficacy of Calcimimetics in Patients with FHH

Calcimimetics mimic or enhance the effects of calcium on CaSRs and are categorized as two types. Polyvalent cations including the naturally occurring ligands of the CaSR (e.g. calcium, magnesium, barium, strontium, and others) are type I calcimimetics. Type II calcimimetics are positive allosteric modulators that increase the sensitivity of the CaSR to its main physiologic ligand—calcium [73]. Type II calcimimetics may have some intrinsic agonistic activity [74], but they generally require extracellular calcium to be present to be active. Calcimimetics likely stabilize the transmembrane domain of the CaSR in a conformation that enhances CaSR coupling to downstream signalling molecules. To enhance CaSR function, calcimimetics may only need to bind to one monomer of a putative dimeric form of the CaSR [75]. In addition, calcimimetics have been proposed to act via ADIS to increase CaSR synthesis and trafficking to the cell surface [64]. Cinacalcet is the first CaSR allosteric modulator approved by the US Food and Drug Administration (FDA) for the treatment of patients with secondary hyperparathyroidism due to end-stage renal failure, inoperable PHPT, and parathyroid carcinoma [76].

Recent case reports have described treating FHH with calcimimetics [77–81]. Successful treatment included subjective improvement of the patient and a decreased serum calcium level, although not necessarily into the normal range. The conclusions from these case reports are limited. A controlled clinical trial would be the ideal means to test such therapy, but this has not been done. The calcimimetic cinacalcet can be used in adults with FHH in whom the serum calcium level is elevated to a level sufficient, in the assessment of the treating clinician, to cause symptoms of hypercalcaemia [77–82]. Adverse effects were mild in the cases reported to date [77–81]. Information on the safety and efficacy of calcimimetics in FHH is limited to a few reports in children [83] and adolescents [84].

In Vitro and *in Vivo* Studies with Calcimimetic: FHH Type 1

In the systems *in vitro* that have been studied, prolonged exposure of CaSR-expressing cells to a calcimimetic enhances the signalling responses and/or expression of most loss-of-function mutant CaSRs [60, 61, 77, 85–87]. These agents likely function as pharmacochaperones to promote mutant receptor synthesis and trafficking by binding to the transmembrane domain of newly synthesized CaSR within the ER [60, 61, 87]. In adults with FHH1 who are symptomatic from hypercalcaemia (usually with serum calcium levels >0.25 mM above the upper limit of normal), cinacalcet can be initiated (as deemed necessary) and increased according to the product prescribing information for PHPT [88]. Recurrent pancreatitis, which is rare in FHH1, may suggest the need for considering drug treatment [81], but this has not been studied in any prospective or controlled manner. At the beginning of treatment, serum calcium should be monitored at short intervals and then after any dose adjustments until the serum calcium level has stabilized. If there is an amelioration of symptoms or improvement in serum calcium levels during a trial of 8 to 12 weeks, the decision to continue treatment should be discussed with the patient. Serum calcium levels do not need to be normalized. Serum PTH levels often decrease, but not in all cases [79, 89]. In patients with FHH1, cinacalcet lowered serum calcium and PTH levels without major adverse effects such as hypocalcaemia and improved hypercalcaemia-related symptoms in the majority of reported case reports [77–79, 90]. A periodic trial off the drug every 1 to 2 years should be considered for one to three weeks to gauge the need for long-term treatment.

In Vivo and *in Vitro* Studies with Calcimimetic: FHH Type 2

Cinacalcet improves the signalling responses of CaSRs *in vitro* when $G\alpha_{11}$ proteins bear the loss-of-function mutations associated with FHH2 [91, 92]. The response to cinacalcet is likely determined by the location of the $G\alpha_{11}$ mutation [91]. The p.I200del mutation within the GTPase domain has required higher cinacalcet dosing to correct the intracellular calcium responses that mutant G-protein subunit generates, compared to cells expressing the p.L135Q mutation that is located within the $G\alpha_{11}$ helical domain [91]. In a patient with FHH2, treatment with cinacalcet (60 mg daily) normalized the serum calcium concentration. The treatment, however, failed to improve the patient's symptoms of headaches, constipation and pruritus and was discontinued after 4 months [92].

In Vitro and *in Vivo* Studies with Calcimimetic: FHH Type 3

Cinacalcet tends to correct the intracellular calcium signalling responses to high extracellular calcium in HEK293 cells in which mutant CaSR proteins are overexpressed. All three FHH3-causing AP2σ mutations (p.R15C, p.R15H, and p.R15L), when expressed in

HEK293 cells, demonstrated increased intracellular calcium signalling and MAPK responses to extracellular calcium after cells were treated with cinacalcet. This treatment, therefore, caused the cells to have signalling responses to high extracellular calcium comparable to cells expressing wild-type CaSRs and AP2σ proteins [80].

Three patients with FHH3, each with a different heterozygous mutation (p.R15C, p.R15H, or p.R15L), were treated with 30–60 mg cinacalcet daily. These patients included a 34-year-old female with a 12-year history of hypercalcaemia and fatigue, headaches, and persistent generalized aches which did not resolve after parathyroidectomy; a 22-year-old male suffering from hypercalcaemia, fatigue, and generalized rib pain; and a 52-year-old woman with hypercalcaemia, headaches, abdominal pain, vomiting, fatigue, and musculoskeletal pain that did not resolve after pamidronate infusion or parathyroidectomy. Cinacalcet reduced the serum calcium by ≥20% and lowered serum PTH levels to within the normal range [80]. Cinacalcet improved hypercalcaemia-related symptoms (fatigue, musculoskeletal pain, and headaches) without causing adverse effects such as hypocalcaemia. Treatment was well tolerated over a period of over 30 months [80].

One child with a chromosome 22q11.2 deletion syndrome and a FHH3-causing AP2σ mutation was treated with cinacalcet (30–60 mg daily). He developed symptoms of hypocalcaemia with a serum calcium in the lower half of the normal range, and therapy was discontinued [84]. Long-term surveillance of patients with different forms of FHH taking cinacalcet is recommended to assess the safety of this approach especially with regard to hypocalcaemia [80].

Prognosis

The prognosis of typical FHH is usually excellent. Patients with FHH in multiple kindreds have not demonstrated increased mortality, compared to unaffected family members. Atypical forms of FHH including type 3, particularly those with higher serum calcium and/or PTH levels or recurrent pancreatitis, have not been systematically examined for overall prognosis and long-term complications.

Neonatal Severe Primary Hyperparathyroidism (NSHPT)

NSHPT (OMIM #239200) is caused most commonly by homozygous or compound heterozygous loss-of-function mutations in the *CASR* [93]. NSHPT is typically a life-threatening disorder in early infancy, if left untreated. Newborns with NSHPT usually suffer from severe hypercalcaemia with levels as high as 7.7 mmol/L reported (normal range for a full-term newborn with birth weight over 1500 g, 2.10–2.65 mmol/L) [42], marked (5-to-10-fold) elevations of serum PTH concentrations, enlarged parathyroid glands, hypotonia, respiratory distress, and hyperparathyroid bone disease leading to bone deformities and multiple fractures [94]. Without functioning CaSRs in the parathyroid glands, PTH is not suppressed even at these very high serum calcium levels. If the mutant CaSR retains some capacity to signal, PTH might be suppressed but to a very limited degree. Emergent parathyroidectomy is usually required and lifesaving. More than 25 different mutations are associated with NSHPT, of which more than 40% are either nonsense or frameshift mutations, that lead to a truncated CaSR [36]. Often neonates with NSHPT are born to parents who are first-degree relatives and who both have FHH and the same mutation in the *CASR*. This leads to the homozygous form of the disease (NSHPT) in an offspring with both alleles affected identically [95]. NSHPT has also been described in a neonate with compound heterozygous mutations—one in each *CASR* allele [96]. Since such neonates (with either homozygous or compound heterozygous mutations) have no or very low levels of normal CaSRs, these infants present with severe hypercalcaemia and typically (but not always) will require urgent parathyroidectomy. Not every patient with homozygous inactivating mutation of the *CASR* suffers from NSHPT. One patient was described who was homozygous for a p.P39A mutation in the CaSR and was only identified in adulthood, despite serum calcium levels of 3.74–4.24 mmol/L [97]. When the mutation was studied *in vitro*, it displayed only mild functional impairment. Another patient presented at age 2 with moderate hypercalcaemia and fractional excretion of calcium below 0.01. Genetic analysis of the *CASR* in that case demonstrated a homozygous p.Q459R mutation within the ECD of the receptor. The patient displayed an autosomal recessive inheritance pattern. All affected individuals in the family who were heterozygous for the mutation were normocalcaemic, although in the upper range of normal. Unaffected individuals had serum calcium levels which were significantly lower [42].

Some infants with only a mild impairment of CaSR function, due to the presence of compound heterozygous *CASR* mutations, develop instead neonatal hyperparathyroidism (NHPT). These infants present with symptomatic hypercalcaemia and often fail to thrive. This presentation seems to occur instead of classical FHH (mild asymptomatic hypercalcaemia often discovered incidentally) due to additional environmental and/or genetic factors that result in fetal calcium deficiency [98]. Infants with NHPT may show less elevated but more symptomatically transient hypercalcaemia than the dramatic presentations of NSHPT. They may further show moderate to severe elevations in PTH and hyperparathyroid bone disease. Following conservative treatment, some infants even transition to the more asymptomatic form of FHH and have not even required parathyroid surgery. Factors contributing to a more severe phenotype *in utero* include gestation in an unaffected mother or coexistence of maternal vitamin D deficiency, which can worsen the degree of hyperparathyroidism [99]. The neonate might also experience a dominant-negative effect of an abnormal *CASR* gene on the remaining normal allele, due to CaSR surface dimerization. When a mutant receptor is coexpressed with the wild-type CaSR *in vitro*, the affinity for calcium may be reduced, possibly due to the abnormally functioning heterodimers that can form under such conditions [10]. Sequence analysis of the *CASR* will help to distinguish NSHPT and NHPT.

Before parathyroid surgery, bisphosphonates and the calcimimetic cinacalcet have been used in newborns to treat severe hypercalcaemia and slow bone demineralization. The response to cinacalcet is dependent on the underlying *CASR* mutation. If cinacalcet can bind to the mutated CaSR as a positive allosteric modulator, it can often improve CaSR signalling associated with many loss-of-function mutations and improve hypercalcaemia and bone mineralization [100–102]. However, in infants with NSHPT due to biallelic truncating

CASR mutations, for example, cinacalcet has been shown to be ineffective, due to the inability of the drug to bind to the CaSR and/or couple to downstream signalling pathways [103, 104].

Hypercalcaemia After Infancy and Primary Hyperparathyroidism (PHPT) in Patients With FHH

Cases have been reported in which loss-of-function *CASR* mutations become apparent in childhood or early adulthood, often because of pronounced hypercalcaemia. Homozygous loss-of-function mutations at the N-terminal region of the CaSR have led to symptomatic hypercalcaemia after infancy and required parathyroidectomy [105, 106]. Patients with PHPT due to parathyroid adenomas or hyperplasia have even been diagnosed with heterozygous or homozygous loss-of-function *CASR* mutations [107]. Homozygous *CASR* mutations in these adult-onset PHPT patients seem to affect extracellular calcium-sensing and intracellular calcium signalling less compared to homozygous mutations in infants with NSHPT [107]. *CASR* mutations located in the ECD of the receptor in patients with PHPT or severe FHH are associated with more pronounced hypercalcaemia and multigland disease. Surgical treatment might therefore require a four-gland rather than a focused minimally invasive parathyroid exploration [36, 107].

Conclusions

Several decades ago, the clinical entities of FHH and NSHPT emerged from clinical reports, which described the disorders as separate ones initially and then as related familial syndromes. As individual patients were studied in more depth with biochemical and physiologic testing, abnormalities in the calcium-sensing mechanism in the parathyroid glands and the kidneys emerged as the means to understand the clinical presentation. Finally, the pathogenesis of FHH types 1–3 and NSHPT were delineated, once *CASR* sequencing became available, and a deeper understanding of the signalling, biosynthetic, and trafficking pathways for the CaSR in target cells began to emerge.

The CaSR is the key molecule responsible for sensing tiny changes in calcium levels that are physiologically relevant and ultimately sets the serum calcium level. $G\alpha_{11}$ protein mediates downstream CaSR signal transduction by coupling to phospholipase C and eventually the mobilization of intracellular calcium. The AP2σ protein likely participates in CaSR signalling and trafficking.

In patients with PTH-dependent hypercalcaemia and CCCR <0.01 or in the CCCR overlap range of 0.01–0.02 between FHH and PHPT, the diagnosis of FHH usually can be established using the family history and biochemical screening of family members, if appropriate, to demonstrate an autosomal dominant pattern of hypercalcaemia in the patient and two or more additional family members. Mutational screening for *CASR* and *AP2S1* codon 15 is appropriate in patients in whom the diagnosis of FHH is suspected. If the screening is negative, then *GNA11* sequencing should be considered. Cinacalcet has been effective in lowering serum calcium and PTH concentrations in selected patients with FHH 1–3 and in some infants with NSHPT. Other types of CaSR modulators with different modes of action and potency may be developed in the future that will benefit those patients with hypercalcaemia-related symptoms.

REFERENCES

1. Pollak MR, Brown EM, Chou YH, al. Mutations in the human Ca(2+)-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell*, 1993; **75**: 1297–303.
2. Brown EM, MacLeod RJ. Extracellular calcium sensing and extracellular calcium signaling. *Physiol Rev*, 2001; **81**: 239–97.
3. Nesbit MA, Hannan FM, Howles SA, *et al*. Mutations affecting G-protein subunit alpha11 in hypercalcemia and hypocalcemia. *N Engl J Med*, 2013; **368**: 2476–86.
4. Marx SJ, Attie MF, Levine MA, Spiegel AM, Downs RW, Jr., Lasker RD. The hypocalciuric or benign variant of familial hypercalcemia: clinical and biochemical features in fifteen kindreds. *Medicine (Baltimore)*, 1981; **60**: 397–412.
5. Nesbit MA, Hannan FM, Howles SA, *et al*. Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3. *Nat Genet*, 2013; **45**: 93–7.
6. Nesbit MA, Hannan FM, Graham U, *et al*. Identification of a second kindred with familial hypocalciuric hypercalcemia type 3 (FHH3) narrows localization to a <3.5 megabase pair region on chromosome 19q13.3. *J Clin Endocrinol Metab*, 2010; **95**: 1947–54.
7. McMurtry CT, Schranck FW, Walkenhorst DA, *et al*. Significant developmental elevation in serum parathyroid hormone levels in a large kindred with familial benign (hypocalciuric) hypercalcemia. *Am J Med*, 1992; **93**: 247–58.
8. Bai M, Pearce SH, Kifor O, *et al*. *In vivo* and *in vitro* characterization of neonatal hyperparathyroidism resulting from a de novo, heterozygous mutation in the Ca2+-sensing receptor gene: normal maternal calcium homeostasis as a cause of secondary hyperparathyroidism in familial benign hypocalciuric hypercalcemia. *J Clin Invest*, 1997; **99**: 88–96.
9. Pearce SH, Bai M, Quinn SJ, Kifor O, Brown EM, Thakker RV. Functional characterization of calcium-sensing receptor mutations expressed in human embryonic kidney cells. *J Clin Invest*, 1996; **98**: 1860–6.
10. Bai M, Quinn S, Trivedi S, *et al*. Expression and characterization of inactivating and activating mutations in the human Ca2+o-sensing receptor. *J Biol Chem*, 1996; **271**: 19537–45.
11. Simonds WF, James-Newton LA, Agarwal SK, Yang B, Skarulis MC, Hendy GN, *et al*. Familial isolated hyperparathyroidism: clinical and genetic characteristics of 36 kindreds. *Medicine (Baltimore)*, 2002; **81**: 1–26.
12. Brown EM. Extracellular Ca2+ sensing, regulation of parathyroid cell function, and role of Ca2+ and other ions as extracellular (first) messengers. *Physiol Rev*, 1991; **71**: 371–411.
13. Kameda T, Mano H, Yamada Y, *et al*. Calcium-sensing receptor in mature osteoclasts, which are bone resorbing cells. *Biochem Biophys Res Commun*, 1998; **245**: 419–22.
14. Hebert SC, Cheng S, Geibel J. Functions and roles of the extracellular Ca2+-sensing receptor in the gastrointestinal tract. *Cell Calcium*, 2004; **35**: 239–47.
15. Cheng I, Klingensmith ME, Chattopadhyay N, *et al*. Identification and localization of the extracellular calcium-sensing receptor in human breast. *J Clin Endocrinol Metab*, 1998; **83**: 703–7.
16. Brown EM, Hebert SC. Calcium-receptor-regulated parathyroid and renal function. *Bone*, 1997; **20**: 303–9.
17. Hu J, Spiegel AM. Structure and function of the human calcium-sensing receptor: insights from natural and engineered

mutations and allosteric modulators. *J Cell Mol Med*, 2007; **11**: 908–22.
18. Ward DT, Brown EM, Harris HW. Disulfide bonds in the extracellular calcium-polyvalent cation-sensing receptor correlate with dimer formation and its response to divalent cations *in vitro*. *J Biol Chem*, 1998; **273**: 14476–83.
19. Geng Y, Mosyak L, Kurinov I, *et al*. Structural mechanism of ligand activation in human calcium-sensing receptor. *Elife*, 2016; **5**.
20. Zhang C, Zhang T, Zou J, *et al*. Structural basis for regulation of human calcium-sensing receptor by magnesium ions and an unexpected tryptophan derivative co-agonist. *Sci Adv*, 2016; **2**: e1600241.
21. Hofer AM, Brown EM. Extracellular calcium sensing and signalling. *Nat Rev Mol Cell Biol*, 2003; **4**: 530–8.
22. Conigrave AD, Ward DT. Calcium-sensing receptor (CaSR): pharmacological properties and signaling pathways. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 315–31.
23. Breitwieser GE, Gama L. Calcium-sensing receptor activation induces intracellular calcium oscillations. *Am J Physiol Cell Physiol*, 2001; **280**: C1412–21.
24. Corbetta S, Lania A, Filopanti M, Vicentini L, Ballare E, Spada A. Mitogen-activated protein kinase cascade in human normal and tumoral parathyroid cells. *J Clin Endocrinol Metab*, 2002; **87**: 2201–5.
25. Grant MP, Stepanchick A, Cavanaugh A, Breitwieser GE. Agonist-driven maturation and plasma membrane insertion of calcium-sensing receptors dynamically control signal amplitude. *Sci Signal*, 2011; **4**: ra78.
26. Breitwieser GE. The calcium sensing receptor life cycle: trafficking, cell surface expression, and degradation. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 303–13.
27. Brown AJ, Zhong M, Finch J, *et al*. Rat calcium-sensing receptor is regulated by vitamin D but not by calcium. *Am J Physiol*, 1996; **270**(3 Pt 2): F454–60.
28. Yarden N, Lavelin I, Genina O, *et al*. Expression of calcium-sensing receptor gene by avian parathyroid gland *in vivo*: relationship to plasma calcium. *Gen Comp Endocrinol*, 2000; **117**: 173–81.
29. Loupy A, Ramakrishnan SK, Wootla B, *et al*. PTH-independent regulation of blood calcium concentration by the calcium-sensing receptor. *J Clin Invest*, 2012; **122**: 3355–67.
30. Riccardi D, Valenti G. Localization and function of the renal calcium-sensing receptor. *Nat Rev Nephrol*, 2016; **12**: 414–25.
31. Ba J, Brown D, Friedman PA. Calcium-sensing receptor regulation of PTH-inhibitable proximal tubule phosphate transport. *Am J Physiol Renal Physiol*, 2003; **285**: F1233–43.
32. Sands JM, Naruse M, Baum M, *et al*. Apical extracellular calcium/polyvalent cation-sensing receptor regulates vasopressin-elicited water permeability in rat kidney inner medullary collecting duct. *J Clin Invest*, 1997; **99**: 1399–405.
33. Jackson CE, Boonstra BC. Hereditary hypercalcemia and parathyroid hyperplasia without definite hyperparathyroidism. *J Lab Clin Med*, 1966; **68**: 883–90.
34. Foley TP, Jr., Harrison HC, Arnaud CD, Harrison HE. Familial benign hypercalcemia. *J Pediatr*, 1972; **81**: 1060–7.
35. Hinnie J, Bell E, McKillop E, Gallacher S. The prevalence of familial hypocalciuric hypercalcemia. *Calcif Tissue Int*, 2001; **68**: 216–18.
36. Hannan FM, Thakker RV. Calcium-sensing receptor (CaSR) mutations and disorders of calcium, electrolyte and water metabolism. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 359–71.
37. Volpe A, Guerriero A, Marchetta A, Caramaschi P, Furlani L. Familial hypocalciuric hypercalcemia revealed by chondrocalcinosis. *Joint Bone Spine*, 2009; **76**: 708–10.
38. Pearce SH, Wooding C, Davies M, Tollefsen SE, Whyte MP, Thakker RV. Calcium-sensing receptor mutations in familial hypocalciuric hypercalcaemia with recurrent pancreatitis. *Clin Endocrinol (Oxf)*, 1996; **45**: 675–80.
39. Gorvin CM, Cranston T, Hannan FM, *et al*. A G-protein subunit-alpha11 loss-of-function mutation, Thr54Met, causes familial hypocalciuric hypercalcemia type 2 (FHH2). *J Bone Miner Res*, 2016; **31**: 1200–6.
40. Hannan FM, Howles SA, Rogers A, *et al*. Adaptor protein-2 sigma subunit mutations causing familial hypocalciuric hypercalcaemia type 3 (FHH3) demonstrate genotype-phenotype correlations, codon bias and dominant-negative effects. *Hum Mol Genet*, 2015; **24**: 5079–92.
41. Vargas-Poussou R, Mansour-Hendili L, Baron S, *et al*. Familial hypocalciuric hypercalcemia types 1 and 3 and primary hyperparathyroidism: similarities and differences. *J Clin Endocrinol Metab*, 2016; **101**: 2185–95.
42. Lietman SA, Tenenbaum-Rakover Y, Jap TS, *et al*. A novel loss-of-function mutation, Gln459Arg, of the calcium-sensing receptor gene associated with apparent autosomal recessive inheritance of familial hypocalciuric hypercalcemia. *J Clin Endocrinol Metab*, 2009; **94**: 4372–9.
43. Christensen SE, Nissen PH, Vestergaard P, *et al*. Plasma 25-hydroxyvitamin D, 1,25-dihydroxyvitamin D, and parathyroid hormone in familial hypocalciuric hypercalcemia and primary hyperparathyroidism. *Eur J Endocrinol*, 2008; **159**: 719–27.
44. Marx SJ. Letter to the editor: distinguishing typical primary hyperparathyroidism from familial hypocalciuric hypercalcemia by using an index of urinary calcium. *J Clin Endocrinol Metab*, 2015; **100**: L29–30.
45. Marx SJ, Attie MF, Stock JL, Spiegel AM, Levine MA. Maximal urine-concentrating ability: familial hypocalciuric hypercalcemia versus typical primary hyperparathyroidism. *J Clin Endocrinol Metab*, 1981; **52**: 736–40.
46. Firek AF, Kao PC, Heath H, 3rd. Plasma intact parathyroid hormone (PTH) and PTH-related peptide in familial benign hypercalcemia: greater responsiveness to endogenous PTH than in primary hyperparathyroidism. *J Clin Endocrinol Metab*, 1991; **72**: 541–6.
47. Nissen PH, Christensen SE, Heickendorff L, Brixen K, Mosekilde L. Molecular genetic analysis of the calcium sensing receptor gene in patients clinically suspected to have familial hypocalciuric hypercalcemia: phenotypic variation and mutation spectrum in a Danish population. *J Clin Endocrinol Metab*, 2007; **92**: 4373–9.
48. Law WM, Jr., Wahner HW, Heath H, 3rd. Bone mineral density and skeletal fractures in familial benign hypercalcemia (hypocalciuric hypercalcemia). *Mayo Clin Proc*, 1984; **59**: 811–15.
49. Carling T, Szabo E, Bai M, *et al*. Familial hypercalcemia and hypercalciuria caused by a novel mutation in the cytoplasmic tail of the calcium receptor. *J Clin Endocrinol Metab*, 2000; **85**: 2042–7.
50. Burski K, Torjussen B, Paulsen AQ, Boman H, Bollerslev J. Parathyroid adenoma in a subject with familial hypocalciuric hypercalcemia: coincidence or causality? *J Clin Endocrinol Metab*, 2002; **87**: 1015–16.
51. Gunn IR, Wallace JR. Urine calcium and serum ionized calcium, total calcium and parathyroid hormone concentrations in the

diagnosis of primary hyperparathyroidism and familial benign hypercalcaemia. *Ann Clin Biochem*, 1992; **29**(Pt 1): 52–8.
52. Christensen SE, Nissen PH, Vestergaard P, Heickendorff L, Brixen K, Mosekilde L. Discriminative power of three indices of renal calcium excretion for the distinction between familial hypocalciuric hypercalcaemia and primary hyperparathyroidism: a follow-up study on methods. *Clin Endocrinol (Oxf)*, 2008; **69**: 713–20.
53. Bertocchio JP, Tafflet M, Koumakis E, *et al*. Pro-FHH: a risk equation to facilitate the diagnosis of parathyroid-related hypercalcemia. *J Clin Endocrinol Metab*, 2018; **103**: 2534–42.
54. Hovden S, Rejnmark L, Ladefoged SA, Nissen PH. AP2S1 and GNA11 mutations—not a common cause of familial hypocalciuric hypercalcemia. *Eur J Endocrinol*, 2017; **176**: 177–85.
55. Law WM, Jr., Heath H, 3rd. Familial benign hypercalcemia (hypocalciuric hypercalcemia). Clinical and pathogenetic studies in 21 families. *Ann Intern Med*, 1985; **102**: 511–19.
56. Warner J, Epstein M, Sweet A, *et al*. Genetic testing in familial isolated hyperparathyroidism: unexpected results and their implications. *J Med Genet*, 2004; **41**: 155–60.
57. Hannan FM, Nesbit MA, Zhang C, *et al*. Identification of 70 calcium-sensing receptor mutations in hyper- and hypocalcaemic patients: evidence for clustering of extracellular domain mutations at calcium-binding sites. *Hum Mol Genet*, 2012; **21**: 2768–78.
58. Huang Y, Zhou Y, Yang W, *et al*. Identification and dissection of Ca(2+)-binding sites in the extracellular domain of Ca(2+)-sensing receptor. *J Biol Chem*, 2007; **282**: 19000–10.
59. Zhang C, Mulpuri N, Hannan FM, *et al*. Role of Ca2+ and L-Phe in regulating functional cooperativity of disease-associated 'toggle' calcium-sensing receptor mutations. *PLoS One*, 2014; **9**: e113622.
60. Huang Y, Breitwieser GE. Rescue of calcium-sensing receptor mutants by allosteric modulators reveals a conformational checkpoint in receptor biogenesis. *J Biol Chem*, 2007; **282**: 9517–25.
61. White E, McKenna J, Cavanaugh A, Breitwieser GE. Pharmacochaperone-mediated rescue of calcium-sensing receptor loss-of-function mutants. *Mol Endocrinol*, 2009; **23**: 1115–23.
62. Pidasheva S, Grant M, Canaff L, Ercan O, Kumar U, Hendy GN. Calcium-sensing receptor dimerizes in the endoplasmic reticulum: biochemical and biophysical characterization of CASR mutants retained intracellularly. *Hum Mol Genet*, 2006; **15**: 2200–9.
63. Bai M, Trivedi S, Kifor O, Quinn SJ, Brown EM. Intermolecular interactions between dimeric calcium-sensing receptor monomers are important for its normal function. *Proc Natl Acad Sci U S A*, 1999; **96**: 2834–9.
64. Grant MP, Stepanchick A, Breitwieser GE. Calcium signaling regulates trafficking of familial hypocalciuric hypercalcemia (FHH) mutants of the calcium sensing receptor. *Mol Endocrinol*, 2012; **26**: 2081–91.
65. Varrault A, Pena MS, Goldsmith PK, Mithal A, Brown EM, Spiegel AM. Expression of G protein alpha-subunits in bovine parathyroid. *Endocrinology*, 1995; **136**: 4390–6.
66. Kim YM, Benovic JL. Differential roles of arrestin-2 interaction with clathrin and adaptor protein 2 in G protein-coupled receptor trafficking. *J Biol Chem*, 2002; **277**: 30760–8.
67. Fujisawa Y, Yamaguchi R, Satake E, *et al*. Identification of AP2S1 mutation and effects of low calcium formula in an infant with hypercalcemia and hypercalciuria. *J Clin Endocrinol Metab*, 2013; **98**: E2022–7.
68. Hendy GN, Canaff L, Newfield RS, *et al*. Codon Arg15 mutations of the AP2S1 gene: common occurrence in familial hypocalciuric hypercalcemia cases negative for calcium-sensing receptor (CASR) mutations. *J Clin Endocrinol Metab*, 2014; **99**: E1311–15.
69. Bilezikian JP, Khan AA, Potts JT, Jr., Third International workshop on the management of asymptomatic primary H. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the third international workshop. *J Clin Endocrinol Metab*, 2009; **94**: 335–9.
70. Castellano E, Attanasio R, Gianotti L, Cesario F, Tassone F, Borretta G. Forearm DXA increases the rate of patients with asymptomatic primary hyperparathyroidism meeting surgical criteria. *J Clin Endocrinol Metab*, 2016; **101**: 2728–32.
71. Isaksen T, Nielsen CS, Christensen SE, Nissen PH, Heickendorff L, Mosekilde L. Forearm bone mineral density in familial hypocalciuric hypercalcemia and primary hyperparathyroidism: a comparative study. *Calcif Tissue Int*, 2011; **89**: 285–94.
72. Powell BR, Buist NR. Late presenting, prolonged hypocalcemia in an infant of a woman with hypocalciuric hypercalcemia. *Clin Pediatr (Phila)*, 1990; **29**: 241–3.
73. Nemeth EF, Heaton WH, Miller M, *et al*. Pharmacodynamics of the type II calcimimetic compound cinacalcet HCl. *J Pharmacol Exp Ther*, 2004; **308**: 627–35.
74. Leach K, Gregory KJ, Kufareva I, *et al*. Towards a structural understanding of allosteric drugs at the human calcium-sensing receptor. *Cell Res*, 2016; **26**: 574–92.
75. Jacobsen SE, Gether U, Brauner-Osborne H. Investigating the molecular mechanism of positive and negative allosteric modulators in the calcium-sensing receptor dimer. *Sci Rep*, 2017; **7**: 46355.
76. Nemeth EF, Goodman WG. Calcimimetic and calcilytic drugs: feats, flops, and futures. *Calcif Tissue Int*, 2016; **98**: 341–58.
77. Mayr B, Schnabel D, Dorr HG, Schofl C. Genetics in endocrinology: gain and loss of function mutations of the calcium-sensing receptor and associated proteins: current treatment concepts. *Eur J Endocrinol*, 2016; **174**: R189–208.
78. Timmers HJ, Karperien M, Hamdy NA, de Boer H, Hermus AR. Normalization of serum calcium by cinacalcet in a patient with hypercalcaemia due to a de novo inactivating mutation of the calcium-sensing receptor. *J Intern Med*, 2006; **260**: 177–82.
79. Rasmussen AQ, Jorgensen NR, Schwarz P. Clinical and biochemical outcomes of cinacalcet treatment of familial hypocalciuric hypercalcemia: a case series. *J Med Case Rep*, 2011; **5**: 564.
80. Howles SA, Hannan FM, Babinsky VN, *et al*. Cinacalcet for symptomatic hypercalcemia caused by AP2S1 mutations. *N Engl J Med*, 2016; **374**: 1396–8.
81. Gunganah K, Grossman A, Druce M. Recurrent pancreatitis in a patient with familial hypocalciuric hypercalcaemia treated successfully with cinacalcet. *Endocrinol Diabetes Metab Case Rep*, 2014; **2014**: 140050.
82. Marx SJ. Calcimimetic use in familial hypocalciuric hypercalcemia-a perspective in endocrinology. *J Clin Endocrinol Metab*, 2017; **102**: 3933–6.
83. Alon US, VandeVoorde RG. Beneficial effect of cinacalcet in a child with familial hypocalciuric hypercalcemia. *Pediatr Nephrol*, 2010; **25**: 1747–50.
84. Tenhola S, Hendy GN, Valta H, *et al*. Cinacalcet treatment in an adolescent with concurrent 22q11.2 deletion syndrome and

familial hypocalciuric hypercalcemia type 3 caused by AP2S1 mutation. *J Clin Endocrinol Metab*, 2015; **100**: 2515–18.
85. Lu JY, Yang Y, Gnacadja G, Christopoulos A, Reagan JD. Effect of the calcimimetic R-568 [3-(2-chlorophenyl)-N-((1R)-1-(3-methoxyphenyl)ethyl)-1-propanamine] on correcting inactivating mutations in the human calcium-sensing receptor. *J Pharmacol Exp Ther*, 2009; **331**: 775–86.
86. Nakamura A, Hotsubo T, Kobayashi K, Mochizuki H, Ishizu K, Tajima T. Loss-of-function and gain-of-function mutations of calcium-sensing receptor: functional analysis and the effect of allosteric modulators NPS R-568 and NPS 2143. *J Clin Endocrinol Metab*, 2013; **98**: E1692–701.
87. Leach K, Wen A, Cook AE, Sexton PM, Conigrave AD, Christopoulos A. Impact of clinically relevant mutations on the pharmacoregulation and signaling bias of the calcium-sensing receptor by positive and negative allosteric modulators. *Endocrinology*, 2013; **154**: 1105–16.
88. Peacock M, Bilezikian JP, Bolognese MA, *et al*. Cinacalcet HCl reduces hypercalcemia in primary hyperparathyroidism across a wide spectrum of disease severity. *J Clin Endocrinol Metab*, 2011; **96**: E9–18.
89. Mastromatteo E, Lamacchia O, Campo MR, *et al*. A novel mutation in calcium-sensing receptor gene associated to hypercalcemia and hypercalciuria. *BMC Endocr Disord*, 2014; **14**: 81.
90. Festen-Spanjer B, Haring CM, Koster JB, Mudde AH. Correction of hypercalcaemia by cinacalcet in familial hypocalciuric hypercalcaemia. *Clin Endocrinol (Oxf)*, 2008; **68**: 324–5.
91. Babinsky VN, Hannan FM, Gorvin CM, *et al*. Allosteric modulation of the calcium-sensing receptor rectifies signaling abnormalities associated with G-protein alpha-11 mutations causing hypercalcemic and hypocalcemic disorders. *J Biol Chem*, 2016; **291**: 10876–85.
92. Gorvin CM, Hannan FM, Cranston T, *et al*. Cinacalcet rectifies hypercalcemia in a patient with familial hypocalciuric hypercalcemia Type 2 (FHH2) caused by a germline loss-of-function galpha11 mutation. *J Bone Miner Res*, 2018; **33**: 32–41.
93. Chattopadhyay N, Brown EM. Role of calcium-sensing receptor in mineral ion metabolism and inherited disorders of calcium-sensing. *Mol Genet Metab*, 2006; **89**: 189–202.
94. Marx SJ, Fraser D, Rapoport A. Familial hypocalciuric hypercalcemia. Mild expression of the gene in heterozygotes and severe expression in homozygotes. *Am J Med*, 1985; **78**: 15–22.
95. Cole DE, Janicic N, Salisbury SR, Hendy GN. Neonatal severe hyperparathyroidism, secondary hyperparathyroidism, and familial hypocalciuric hypercalcemia: multiple different phenotypes associated with an inactivating Alu insertion mutation of the calcium-sensing receptor gene. *Am J Med Genet*, 1997; **71**: 202–10.
96. Kobayashi M, Tanaka H, Tsuzuki K, *et al*. Two novel missense mutations in calcium-sensing receptor gene associated with neonatal severe hyperparathyroidism. *J Clin Endocrinol Metab*, 1997; **82**: 2716–19.
97. Aida K, Koishi S, Inoue M, Nakazato M, Tawata M, Onaya T. Familial hypocalciuric hypercalcemia associated with mutation in the human Ca(2+)-sensing receptor gene. *J Clin Endocrinol Metab*, 1995; **80**: 2594–8.
98. Glaudo M, Letz S, Quinkler M, *et al*. Heterozygous inactivating CaSR mutations causing neonatal hyperparathyroidism: function, inheritance and phenotype. *Eur J Endocrinol*, 2016; **175**: 421–31.
99. Zajickova K, Vrbikova J, Canaff L, Pawelek PD, Goltzman D, Hendy GN. Identification and functional characterization of a novel mutation in the calcium-sensing receptor gene in familial hypocalciuric hypercalcemia: modulation of clinical severity by vitamin D status. *J Clin Endocrinol Metab*, 2007; **92**: 2616–23.
100. Reh CM, Hendy GN, Cole DE, Jeandron DD. Neonatal hyperparathyroidism with a heterozygous calcium-sensing receptor (CASR) R185Q mutation: clinical benefit from cinacalcet. *J Clin Endocrinol Metab*, 2011; **96**: E707–12.
101. Gannon AW, Monk HM, Levine MA. Cinacalcet monotherapy in neonatal severe hyperparathyroidism: a case study and review. *J Clin Endocrinol Metab*, 2014; **99**: 7–11.
102. Fisher MM, Cabrera SM, Imel EA. Successful treatment of neonatal severe hyperparathyroidism with cinacalcet in two patients. *Endocrinol Diabetes Metab Case Rep*, 2015; **2015**: 150040.
103. Garcia Soblechero E, Ferrer Castillo MT, Jimenez Crespo B, Dominguez Quintero ML, Gonzalez Fuentes C. Neonatal hypercalcemia due to a homozygous mutation in the calcium-sensing receptor: failure of cinacalcet. *Neonatology*, 2013; **104**: 104–8.
104. Atay Z, Bereket A, Haliloglu B, *et al*. Novel homozygous inactivating mutation of the calcium-sensing receptor gene (CASR) in neonatal severe hyperparathyroidism-lack of effect of cinacalcet. *Bone*, 2014; **64**: 102–7.
105. Chikatsu N, Fukumoto S, Suzawa M, *et al*. An adult patient with severe hypercalcaemia and hypocalciuria due to a novel homozygous inactivating mutation of calcium-sensing receptor. *Clin Endocrinol (Oxf)*, 1999; **50**: 537–43.
106. Miyashiro K, Kunii I, Manna TD, *et al*. Severe hypercalcemia in a 9-year-old Brazilian girl due to a novel inactivating mutation of the calcium-sensing receptor. *J Clin Endocrinol Metab*, 2004; **89**: 5936–41.
107. Hannan FM, Nesbit MA, Christie PT, *et al*. A homozygous inactivating calcium-sensing receptor mutation, Pro339Thr, is associated with isolated primary hyperparathyroidism: correlation between location of mutations and severity of hypercalcaemia. *Clin Endocrinol (Oxf)*, 2010; **73**: 715–22.

4.5

Hypocalcaemic Disorders, Hypoparathyroidism, and Pseudohypoparathyroidism

Fadil M. Hannan, Bart L. Clarke, and Rajesh V. Thakker

Introduction

The concentration of ionized extracellular calcium is tightly regulated through the actions of the parathyroid glands (**Figure 4.5.1**). The parathyroids synthesize and secrete parathyroid hormone (PTH) in response to a reduction in the prevailing extracellular calcium concentration, which is sensed by the calcium-sensing receptor (CaSR) (**Figure 4.5.1**) [1]. The secreted PTH acts on the PTH1 receptor (**Figure 4.5.1**) to increase osteoclast-mediated bone resorption, promote urinary calcium reabsorption and the synthesis of 1,25-dihydroxyvitamin D, which in turn increases intestinal calcium absorption [1]. The causes of hypocalcaemia (**Box 4.5.1**) can be classified according to whether serum PTH concentrations are low (that is, hypoparathyroid disorders) or high (that is, disorders associated with secondary hyperparathyroidism or PTH resistance, as occurring in pseudohypoparathyroidism). Hypocalcaemia is most commonly caused by chronic renal failure, vitamin D deficiency, hypomagnesaemia, and hypoparathyroidism [2].

Hypocalcaemia

Epidemiology

Relatively few studies have described the epidemiology of hypocalcaemia, hypoparathyroidism, or pseudohypoparathyroidism. A recent large internal medicine hospital series including 30 813 patients showed that 3% had recognized hypocalcaemia at hospital admission [3]. Post-thyroid surgical clinical series describe transient hypocalcaemia in 25.4–83% of patients [4], with chronic hypocalcaemia more than 6 months after surgery occurring in 0.12–4.6% [5].

The prevalence of chronic hypoparathyroidism has been reported from various European countries and the United States to range between 10.1/100 000 in Norway [6], and 40/100 000 in Tayside, Scotland [7]. Most studies have shown the prevalence to be in the 20–30/100 000 range [8–12]. Only two studies have evaluated the incidence of hypoparathyroidism. One study in Denmark estimated the incidence to be 0.8/100 000/year [8], and the other study in India estimated the incidence to be 2.6/100 000/year [13].

The best prevalence estimate of chronic hypoparathyroidism in the United States is based on analysis of a large health plan claims database, which identified a total of 77 000 cases [9]. The longitudinal population-based Rochester Epidemiology Project was used to derive a prevalence estimate of 37/100 000 [10]. In the nationwide Danish historic cohort study, the prevalence of postsurgical hypoparathyroidism [8, 11, 12] was estimated to be 22/100 000, with the prevalence of non-surgical hypoparathyroidism 2.3/100 000. The study of hospitalized patients with hypoparathyroidism in Norway [6] showed the prevalence of postsurgical hypoparathyroidism to be 6.4/100 000, and non-surgical hypoparathyroidism 3.0/100 000.

The same Norwegian study estimated the prevalence of pseudohypoparathyroidism to be 0.8/100 000 [6]. The same study in India estimated the incidence of pseudohypoparathyroidism to be 0.16/100 000/year [13].

Clinical Features and Investigations

The clinical presentation of hypocalcaemia (serum ionized calcium <1.10 mmol/L or albumin-adjusted calcium <2.20 mmol/L) relates to its severity and rapidity of onset, and may range from an asymptomatic biochemical abnormality to a life-threatening disorder [2]. Severe hypocalcaemia (serum ionized calcium <0.95 mmol/L or serum albumin-adjusted calcium <1.90 mmol/L), particularly in the setting of acute-onset hypoparathyroidism (e.g. following surgical resection of the parathyroid glands), may lead to symptoms of neuromuscular irritability such as paraesthesiae of the circumoral region, fingers

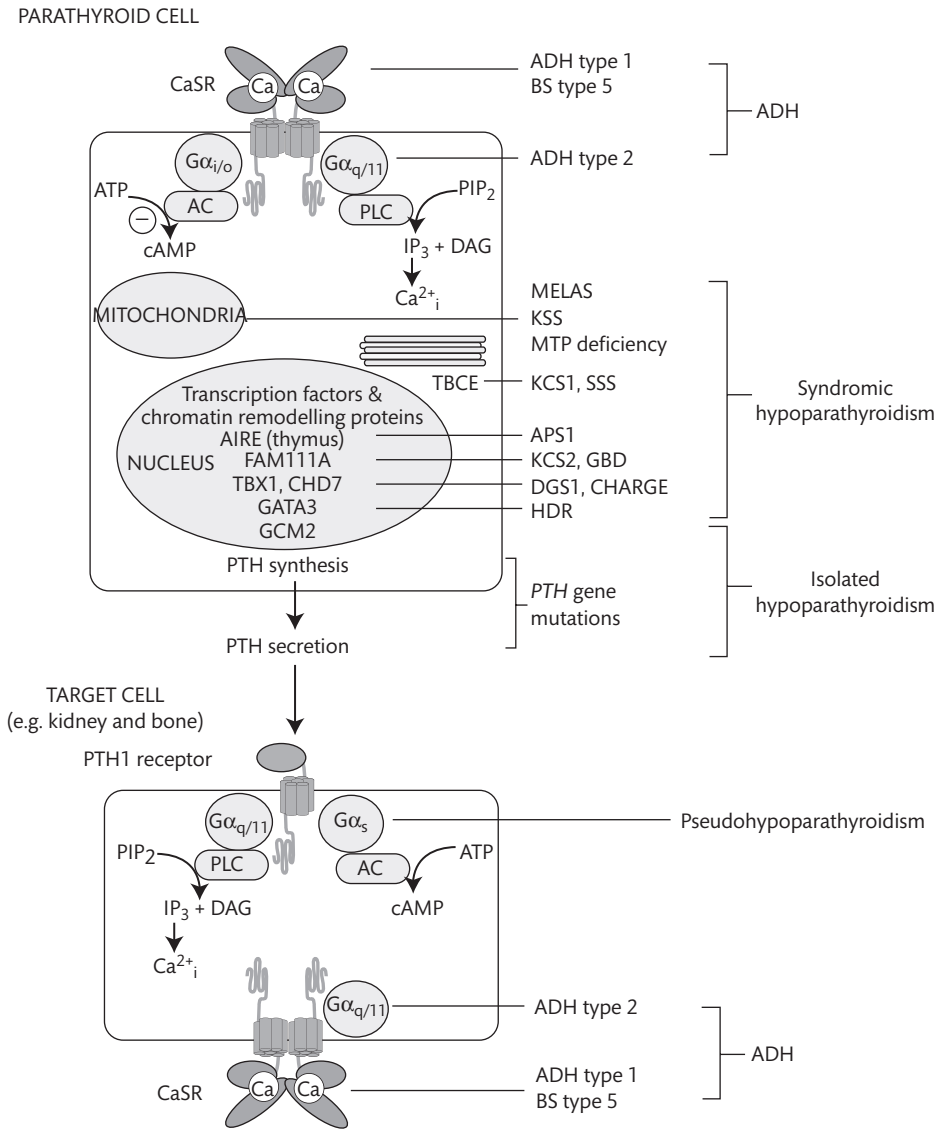


Figure 4.5.1 Components involved in the parathyroid regulation of calcium homeostasis. Alterations in extracellular calcium are detected by the calcium-sensing receptor (CaSR), which is expressed in parathyroid, kidney, and bone cells. The CaSR signals via the Gαq/11 proteins to stimulate phospholipase C (PLC), which catalyses the hydrolysis of phosphoinositide (PIP2) to inositol triphosphate (IP3), thereby increasing intracellular calcium (Ca2+i), and diacylglycerol (DAG). The CaSR also signals via the Gαi/o proteins, which inhibit adenylate cyclase (AC), thereby leading to a reduction in the formation of cAMP from adenosine triphosphate (ATP). In parathyroid cells, these proximal signals modulate downstream pathways, which lead to alterations in the synthesis and secretion of PTH. Secreted PTH acts on the PTH1 receptor in kidney and bone cells, which mediates signalling via the Gαs and Gαq/11 proteins to regulate calcium reabsorption, 1,25-dihydroxyvitamin D synthesis and bone resorption. Abnormalities in several genes and encoded proteins in these pathways—which regulate calcium-sensing and G-protein function; mitochondrial activity; tubulin formation; and gene transcription and chromatin remodelling—have been identified in patients with hypoparathyroid disorders (Table 4.5.2). ADH type 1, autosomal dominant hypocalcaemia type 1; ADH type 2, autosomal dominant hypocalcaemia type 2; BS type 5, Bartter syndrome type 5; MELAS, mitochondrial encephalopathy, stroke-like episodes, and lactic acidosis; KSS, Kearns-Sayre syndrome; MTP; mitochondrial trifunctional protein; KCS1, Kenny-Caffey syndrome type 1; KCS2, Kenny-Caffey syndrome type 2; SSS, Sanjad-Sakati syndrome; GBD, Gracile bone dysplasia; DGS1, DiGeorge syndrome type 1; HDR, hypoparathyroidism, deafness, and renal anomalies.

and toes; and also muscle cramps, carpopedal spasms, laryngospasm and bronchospasm, seizures of all types, and cardiac arrhythmias associated with prolongation of the QT interval on electrocardiogram (ECG) [1, 2]. Chronic hypocalcaemia (which has occurred over a period of months to years) may be associated with subcapsular cataracts, papilloedema and abnormal dentition [1, 2]. Moreover, patients with chronic hypocalcaemia caused by hypoparathyroidism frequently develop ectopic mineralization. Indeed, more than 70% of patients with idiopathic hypoparathyroidism have been reported to have calcifications within intracerebral regions such as in the basal ganglia [14].

Biochemical investigations are helpful in identifying the cause of hypocalcaemia and establishing the cause (Figure 4.5.2). For example, in hypoparathyroidism, serum calcium is low, phosphate is high, and PTH is low; renal function and concentrations of

Box 4.5.1 Causes of hypocalcaemia

Low parathyroid hormone levels (hypoparathyroidism)

- Parathyroid agenesis
  Isolated or part of complex developmental anomaly (e.g. DiGeorge syndrome)
- Parathyroid destruction
  Surgery*, irradiation
  Infiltration by metastases or systemic disease (e.g. amyloidosis or sarcoidosis) or by copper or iron overload (e.g. Wilson's disease, haemochromatosis, recurrent iron transfusions for thalassaemia)
  Autoimmune
  – Isolated
  – Pluriglandular Autoimmune Hypoparathyroidism
- Reduced parathyroid function
  Parathyroid hormone gene defects
  Hypomagnesaemia*
  Neonatal hypocalcaemia (may be caused by maternal hypercalcaemia)
  Autosomal dominant hypocalcaemia†

High parathyroid hormone levels (secondary hyperparathyroidism or parathyroid hormone resistance)

- Vitamin D deficiency*
  Reduced sunlight exposure*, nutritional lack*, malabsorption*, liver disease, chronic renal failure*
- Vitamin D resistance (rickets)
  Renal tubular dysfunction (Fanconi's syndrome), vitamin D receptor defects
- Drugs
  Calcium chelators (e.g. citrated blood transfusions, phosphate)
  Antiresorptives (e.g. bisphosphonates, denosumab)
  Calcimimetics (e.g. cinacalcet)
  Altered vitamin D metabolism (e.g. phenytoin, ketoconazole)
- Miscellaneous
  Acute pancreatitis
  Rhabdomyolysis
  Massive tumour lysis
  Osteoblastic metastases (e.g. from prostate)
  Toxic shock syndrome
  Hyperventilation
- Parathyroid hormone resistance
  Hypomagnesaemia, pseudohypoparathyroidism

* Most common causes
† Plasma parathyroid hormone concentrations may be normal

the 25-hydroxy and 1,25-dihydroxy metabolites of vitamin D are normal [2]. The features of pseudohypoparathyroidism (PHP) are similar to those of hypoparathyroidism except for PTH, which is markedly increased [1, 15]. In chronic renal failure, serum phosphate is high and alkaline phosphatase, creatinine and PTH are elevated; 25-hydroxyvitamin D is normal and 1,25-dihydroxyvitamin D is low. In vitamin D deficiency and osteomalacia, serum calcium, and phosphate are low, alkaline phosphatase and PTH are elevated, renal function is normal, and 25-hydroxyvitamin D is low [2].

Management of Acute Hypocalcaemia

Treatment should be given to symptomatic patients or asymptomatic patients with a serum calcium of less than 1.90 mmol/L who may be at high risk of developing complications. The preferred treatment for acute symptomatic hypocalcaemia is 10 ml 10% calcium gluconate (~2.25 mmol of calcium per 10 ml) i.v., diluted in 50 ml of 5% dextrose or 0.9% sodium chloride and given by slow injection (>5 min); this can be repeated to control symptoms. ECG monitoring and regular serum calcium assessment is required [16]. Continuing hypocalcaemia may be managed acutely by administration of a calcium gluconate infusion, e.g. 100 ml 10% gluconate (~22.5 mmol of calcium), in 1 litre of 5% dextrose or 0.9% sodium chloride. Generally, 0.30–0.40 mmol/kg of elemental calcium infused over 4–6 hours increases serum calcium by 0.5–0.75 mmol/L. If hypocalcaemia is likely to persist, oral vitamin D therapy should also be commenced. In hypocalcaemic patients who are also hypomagnesaemic, the hypomagnesaemia must be corrected before the hypocalcaemia will resolve. This may occur in the post-parathyroidectomy period or in those with severe intestinal malabsorption (e.g. as in coeliac disease) [17].

Management of Persistent Hypocalcaemia

The two major groups of drugs available for the treatment of hypocalcaemia are: supplemental calcium, about 10–20 mmol calcium 6–12 hourly and vitamin D preparations [18–20]. Patients with hypoparathyroidism seldom require calcium supplements after the early stages of stabilization on vitamin D. A variety of vitamin D preparations have been used (Table 4.5.1). These include: alfacalcidol (1α-hydroxycholecalciferol), 0.25–1.0 μg/day; and calcitriol (1,25-dihydroxycholecalciferol), 0.25–2.0 μg/day. In children, these preparations are prescribed in doses based on body weight. Calcitriol is probably the drug of choice because it is the active metabolite and, unlike alfacalcidol, does not require hepatic 25-hydroxylation. Close monitoring (at 1–2 week intervals) of the patient's serum and urine calcium are required initially, and at 3–6 monthly intervals once stabilization is achieved. The aim is to avoid hypercalcaemia, hypercalciuria, nephrolithiasis, and renal failure. In patients with hypoparathyroidism, which is inadequately controlled on vitamin D preparations, recombinant human PTH (rPTH) has been used to normalize serum calcium while reducing the risk of hypercalciuria [1]. A phase 3 randomized controlled clinical trial has demonstrated that once daily administration of rPTH(1-84) is effective at reducing the requirements for calcium and active vitamin D preparations in hypoparathyroid patients while maintaining stable serum calcium values [21]. In addition to its effects on mineral metabolism, rPTH(1-84) therapy has been shown to improve quality of life outcomes for hypoparathyroid patients [22]. This recombinant form of PTH is now licenced in the United States and Europe as an adjunct to calcium and vitamin D for the long-term management of hypoparathyroidism.

Hypocalcaemic Disorders

The application of molecular biology to the study of hypocalcaemic disorders has provided insights into the regulation of parathyroid gland development, PTH secretion, and action. Thus, studies of patients and mouse models with syndromic and non-syndromic forms of hypoparathyroidism have revealed that transcription factors (e.g. TBX1, GCM2, and GATA3) mediate patterning of the third pharyngeal pouch and parathyroid organogenesis, and are also involved in the elimination of auto-reactive T cells within the thymus (e.g. AIRE). Whereas characterization of the mutations causing autosomal dominant hypocalcaemia has demonstrated that the CaSR and G-protein

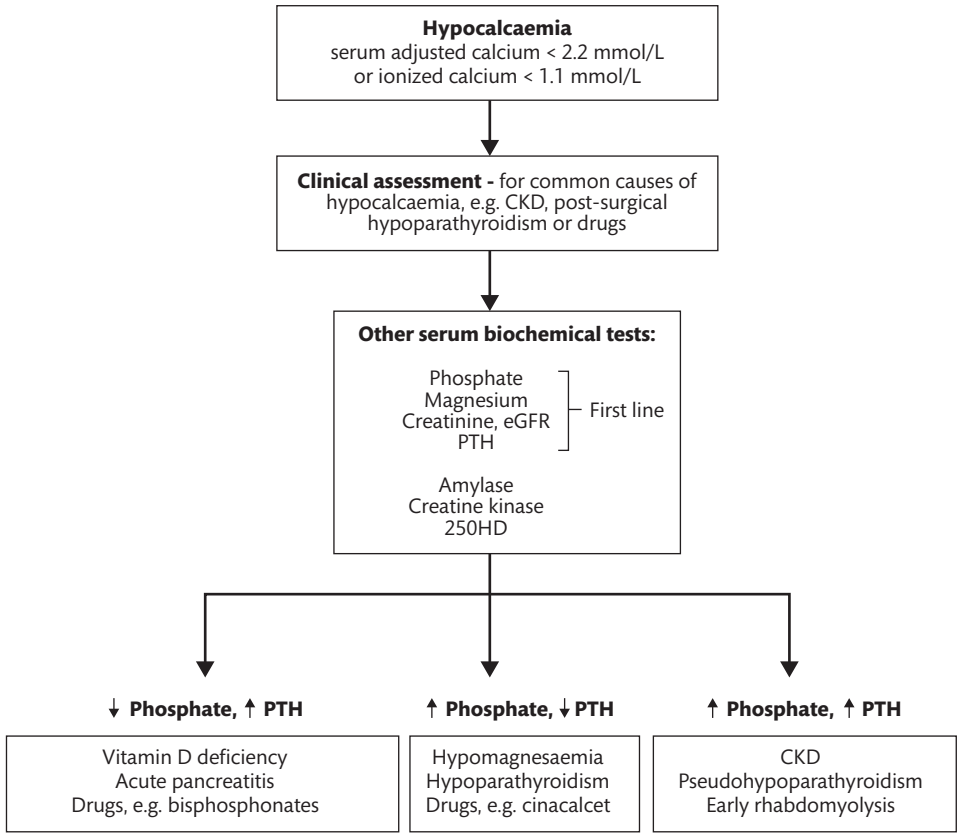


Figure 4.5.2 Clinical approach to the investigation of common causes of hypocalcaemia. CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; 25OHD, 25-hydroxyvitamin D.

Adapted with permission from Hannan FM, Thakker RV. Investigating hypocalcaemia. *BMJ*. 2013;346:f2213. Copyright © 2013, British Medical Journal Publishing Group. (ref 2).

subunit α-11 (Gα11) proteins regulate PTH secretion. Furthermore, the identification of *PTH* mutations as a cause of hypoparathyroidism has provided insights into the cellular biosynthesis of PTH and the mechanisms governing its binding to the PTH1 receptor. Moreover, the demonstration of G-protein subunit α-s ($G\alpha_s$) mutations as the cause of PHP type Ia, PHP type Ib, and pseudopseudohypoparathyroidism, has highlighted the importance of $G\alpha_s$ in mediating the target organ actions of PTH. These advances together with the clinical features of these disorders will be reviewed in this chapter.

Hypoparathyroidism

Hypoparathyroidism is characterized by hypocalcaemia and hyperphosphataemia, which arise from a deficiency in PTH secretion or action (Box 4.5.1) [1]. Hypoparathyroidism may result from agenesis (e.g. DiGeorge syndrome) or parathyroid gland destruction (e.g. following neck surgery, or in autoimmune diseases), from reduced secretion of PTH (e.g. hypomagnesaemia or autosomal dominant hypocalcaemia), or resistance to PTH (which may occur as a primary disorder, e.g. pseudohypoparathyroidism or secondary to

Table 4.5.1 Pharmaceutical preparations of vitamin D and active metabolites

	Calciferol[1]	Dihydrotachysterol	Calcifediol	Calcitriol	Alfacalcidol
Drug	Vitamin D_3 or D_2	DHT	25-OHD_3	1,25$(OH)_2D_3$	1α$(OH)D_3$
	• Capsules, 0.25 mg and 1.25 mg	• Liquid, 0.25 mg/ml	• Capsules, 20 and 50 μg	• Capsules, 0.25 and 0.5 μg	• Capsules, 0.25, 0.50 and 1 μg
				• Injection, 1 μg/ml	• Liquid, 2 μg/ml • Injection, 2 μg/ml in propylene glycol
Time to maximum effect	4–10 weeks	2–4 weeks	4–20 weeks	0.5–1 week	0.5–1 week
Persistence of effect after cessation	6–30 weeks	2–8 weeks	4–12 weeks	0.5–1 week	0.5–1 week

[1] Calciferol may contain cholecalciferol or ergocalciferol

Table 4.5.2 Inherited forms of hypoparathyroidism and their chromosomal locations

Disease	Inheritance	Gene	Chromosome
Isolated hypoparathyroidism			
Autosomal	Autosomal dominant	*GCM2, PTH*	6p24.2, 11p15
	Autosomal recessive	*GCM2, PTH, AIRE*	6p24.2, 11p15, 21q22.3
X-linked	X-linked recessive	*SOX3*†, *FHL1*†	Xq26-27, Xq26.3
Syndromic hypoparathyroidism			
DiGeorge syndromes (DGS)			
DGS type 1	Autosomal dominant	*TBX1*	22q11.2
DGS type 2	Autosomal dominant	*NEBL*†	10p13–14
CHARGE syndrome	Autosomal dominant	*CHD7*	8q12.1
Syndromes associated with deafness and/or renal anomalies			
HDR	Autosomal dominant	*GATA3*	10p15
Barakat syndrome	Autosomal recessive†	Unknown	?
Nephropathy, nerve deafness	Autosomal dominant†	Unknown	?
Nerve deafness without renal dysplasia	Autosomal dominant	Unknown	?
Lymphedema, hypoparathyroidism, nephropathy, mitral valve prolapse, and brachytelephalangy	Autosomal recessive†	Unknown	?
Pluriglandular autoimmune hypoparathyroidism	Autosomal recessive	*AIRE*	21q22.3
Mitochondrial disorders			
Kearns–Sayre syndrome	Maternal	Mitochondrial genome	–
MELAS	Maternal	Mitochondrial genome	–
MTP deficiency	Autosomal recessive	*HADHB*	2p23
Syndromes associated with growth failure/dwarfism			
Kenny-Caffey syndrome type 1	Autosomal recessive	*TBCE*	1q42.3
Sanjad-Sakati syndrome	Autosomal recessive	*TBCE*	1q42.3
Kenny-Caffey syndrome type 2	Autosomal dominant	*FAM111A*	11q12.1
Dubowitz syndrome	Autosomal recessive†	Unknown	?
Autosomal dominant hypocalcaemia (ADH)			
ADH type 1	Autosomal dominant	*CASR*	3q21.1
ADH type 2	Autosomal dominant	*GNA11*	19p13.3
Pseudohypoparathyroidism (PHP)			
PHP type 1a	Autosomal dominant*	*GNAS*	20q13.3
Pseudopseudohypoparathyroidism	Autosomal dominant*	*GNAS*	20q13.3
PHP type 1b	Autosomal dominant*	*GNAS, NESP55, STX16*	20q13.3

HDR, hypoparathyroidism, deafness, and renal anomalies; MELAS, mitochondrial encephalopathy, stroke-like episodes, and lactic acidosis; MTP, mitochondrial trifunctional protein.
† Most likely gene (or inheritance) shown; * Parentally imprinted.

hypomagnesaemia). In addition, hypoparathyroidism may occur as an inherited disorder (Table 4.5.2) that may either be part of a complex congenital defect (e.g. DiGeorge syndrome), or as part of a pluriglandular autoimmune disorder, or as a solitary endocrinopathy, which has been referred to as isolated hypoparathyroidism [1].

Isolated Hypoparathyroidism

Isolated hypoparathyroidism may either be inherited as an autosomal disorder caused by *PTH* or *GCM2* mutations, or as an X-linked disorder (Table 4.5.2), or it may be acquired as a consequence of surgical or autoimmune destruction of the parathyroid glands.

Autosomal Hypoparathyroidism Due to PTH Mutations

Autosomal dominant and recessive forms of isolated hypoparathyroidism are caused by mutations of the *PTH* gene, which consists of three exons and is located on chromosome 11p15 (Figure 4.5.3) [23]. Exon 1 of the *PTH* gene is untranslated, whereas exons 2 and

3 encode the 115-amino acid pre-pro-*PTH* peptide. Exon 2 encodes the initiation (ATG) codon, the prehormone sequence and part of the prohormone sequence, while exon 3 encodes the remainder of the prohormone sequence, the mature 84-amino acid PTH peptide, and the 3' untranslated region (Figure 4.5.3) [24].

Autosomal dominant hypoparathyroidism due to heterozygous *PTH* mutations has been reported in two unrelated families [25, 26]. In one family, a heterozygous single base substitution (T→C) in exon 2 of the *PTH* gene, which resulted in a p.Cys18Arg missense substitution within the hydrophobic core of the signal peptide was shown to impair processing of the mutant pre-pro-PTH to pro-PTH *in vitro* [25]. However, secretion of the preproPTH peptide harbouring the p.Cys18Arg mutation could be enhanced by the addition of 4-phenylbutyric acid, which is a pharmacological chaperone [27]. In another family, a heterozygous single base substitution (T→A) in exon 2 of the *PTH* gene, which resulted in a p.Met14Lys missense substitution in the signal peptide has been reported [26]. This mutation was predicted to impair cleavage of pre-pro-PTH, and *in vitro* studies showed intracellular retention of the mutant pre-pro-PTH peptide, which could be partially rescued by the addition of 4-phenylbutyric acid [26].

Autosomal recessive hypoparathyroidism due to homozygous *PTH* mutations that impaired processing of the PTH peptide have been identified in three unrelated families [24, 28, 29]. In one such family a donor splice-site mutation at the exon 2-intron 2 boundary has been identified [24]. This mutation involved a single base transition (g→c) at position 1 of intron 2 and an assessment of the effects of this alteration in the invariant gt dinucleotide of the 5' donor splice-site consensus on mRNA processing revealed that the mutation resulted in exon skipping, in which exon 2 of the *PTH* gene was lost and exon 1 was spliced to exon 3. The lack of exon 2 would lead to a loss of the initiation codon (ATG) and the signal peptide sequence, which are required respectively for the commencement of PTH mRNA translation and for the translocation of the PTH peptide [24]. In another family with autosomal recessive hypoparathyroidism a homozygous single base substitution (T→C) in exon 2, which resulted in a p.Ser23Pro missense substitution in the signal peptide was detected [29]. This mutation leads to the introduction of a mutant proline residue at the –3 position of the pre-pro-PTH protein cleavage site and is predicted to disrupt cleavage of the mutant pre-pro-PTH protein [29]. In the third unrelated family with autosomal recessive hypoparathyroidism, a homozygous single base substitution (C→A) in exon 2, which resulted in a stop codon (p.Ser23Stop) was detected [28]. This stop codon was predicted to be introduced at the pre-pro-PTH protein cleavage site, and would lead to a truncated and inactive PTH peptide [28]. A homozygous arginine-to-cysteine mutation has also been identified at codon 25, (p.Arg25Cys), of the mature PTH(1-84) peptide in a family with hypocalcaemia [30]. The plasma PTH levels of affected family members varied from low-normal to markedly elevated, depending on the type of PTH assay used [30]. In contrast to previously reported *PTH* gene mutations, which affect secretion of PTH, the p.Arg25Cys missense substitution was shown to diminish the binding of the mutant PTH peptide with the PTH/PTHrP receptor [30]. Moreover, the *PTH* p.Arg25Cys mutation interfered with PTH immunoassays that utilized antibodies affinity-purified using PTH 1-34 and 13-34 fragments, thereby explaining why some assays were unable to detect the mutant PTH peptide [30].

Autosomal Hypoparathyroidism Due to GCM2 Mutations

GCM2 (glial cells missing homolog 2) encodes the GCM2 transcription factor, which is expressed exclusively in the parathyroid glands, and promotes the differentiation and survival of parathyroid cells in the developing embryo [31], as well as upregulating CaSR and PTH expression (Figure 4.5.4) [32–34]. In support of this, mice homozygous for the deletion of *Gcm2* (*Gcm2*$^{-/-}$) lacked parathyroid glands and developed hypocalcaemia and hyperphosphatemia [35]. However, *Gcm2*$^{-/-}$ mice had PTH levels indistinguishable from

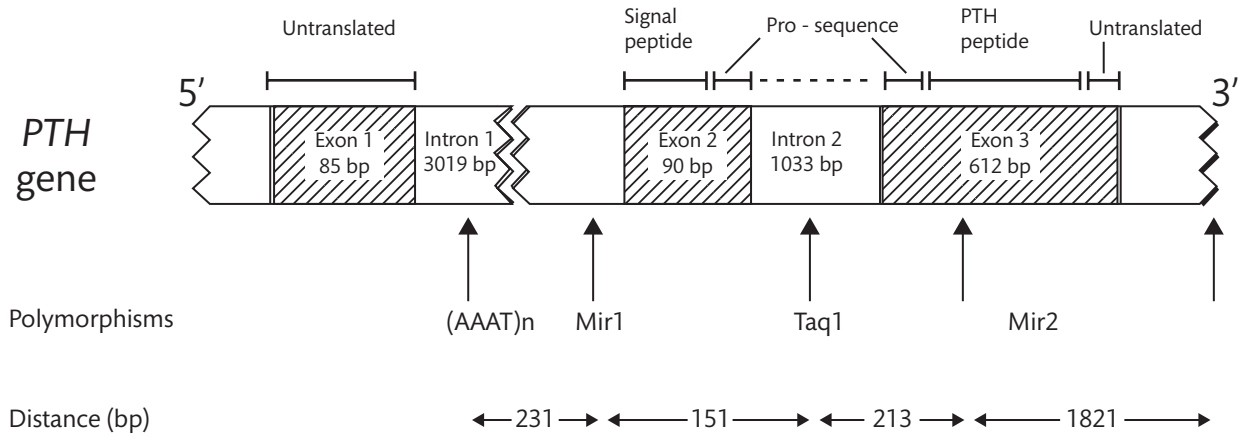


Figure 4.5.3 Schematic representation of the *PTH* gene. The *PTH* gene consists of three exons and two introns; the peptide is encoded by exons 2 and 3. The PTH peptide is synthesized as a precursor which contains a pre- and a pro-sequence. The mature PTH peptide, which contains 84 amino acids, and larger carboxy-terminal PTH fragments are secreted from the parathyroid cell. The polymorphic sites associated with the *PTH* gene are indicated. Two restriction fragment length polymorphisms (RFLPs) are associated with the PTH gene, and the TaqI polymorphic site is within intron 2 and the PstI polymorphic site is 1.7 kbp downstream in the 3′ direction of the gene. Two other polymorphisms of the PTH gene designated Mir1 and Mir2 are located in intron 1 and exon 3, respectively, and the tetranucleotide (AAAT)n polymorphism is in intron 1. Linkage disequilibrium between the (AAAT)n, TaqI, and PstI polymorphic sites has been established.

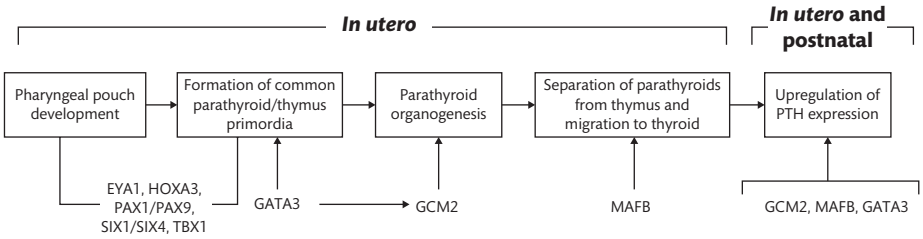


Figure 4.5.4 Transcription factors involved in parathyroid gland development and function. In humans, the parathyroid glands are derived from the endoderm of the third and fourth pharyngeal pouches. Whereas, in mice, the parathyroid glands develop with the thymus from the third pharyngeal pouch endoderm. A network of transcription factors such as the *EYA1*, *HOXA3*, *TBX1*, PAX1/PAX9, SIX1/SIX4, and GATA3 proteins mediate patterning of the third pharyngeal pouch and formation of the common parathyroid-thymus primordia. These transcription factors act in a spatiotemporal manner, for example, TBX1 is expressed in the pharyngeal endoderm and required for the development of the third pharyngeal pouch; whereas GATA3 is expressed in the common parathyroid-thymus primordia and mediates the differentiation of parathyroid and thymus progenitor cells. Moreover, GATA3 regulates the expression of GCM2, which is expressed in the parathyroid domain of the common primordia and mediates the initial stages of parathyroid organogenesis. MAFB is also expressed in the parathyroid domain and facilitates separation of the parathyroid glands from the thymus, and also the migration of the parathyroids towards the thyroid gland. GATA3, GCM2, and MAFB also act synergistically to upregulate PTH expression.

those of wild-type ($Gcm2^{+/+}$) mice, and the source of this auxiliary PTH in $Gcm2^{-/-}$ mice was identified to be a cluster of thymic PTH-expressing cells [35].

GCM2 mutations are associated with autosomal recessive and dominant forms of isolated hypoparathyroidism (Table 4.5.2). Homozygous *GCM2* mutations have been identified in patients with autosomal recessive hypoparathyroidism [36], while in another family a homozygous missense mutation (p.Arg47Leu) of the DNA binding domain was identified [37]. Functional analysis, using electrophoretic mobility shift assays (EMSAs), of this p.Arg47Leu *GCM2* mutation revealed that it resulted in a loss of DNA binding ability of the mutant *GCM2* protein [37]. Heterozygous *GCM2* mutations, which consist of single nucleotide deletions (c.1389delT and c.1399delC) that introduce frame-shifts and premature truncations, and a missense mutation, p.Asn502His, have also been identified in three unrelated families with autosomal dominant hypoparathyroidism [38, 39]. These mutations were shown, by using a *GCM2*-associated luciferase reporter, to inhibit the action of the wild-type transcription factor, thereby indicating that these *GCM2* mutants have dominant-negative properties [38, 39].

X-Linked Recessive Hypoparathyroidism

Hypoparathyroidism with an X-linked recessive transmission pattern was initially reported in two multigenerational kindreds [40, 41]. Only males were noted to be hypocalcaemic, which occurred in association with an absence of parathyroid tissue [42]. The relatedness of the two kindreds has been established by mitochondrial DNA sequence analysis [43]. Studies utilizing X-linked polymorphic markers in these families localized the mutant gene to chromosome Xq26-q27 [44], and a molecular deletion-insertion that involves chromosome 2p25 and Xq27 has been identified [45]. This deletion-insertion is located ~67 kb downstream of the SRY-box 3 (*SOX3)* gene (Table 4.5.2), which is related to *SRY*, the sex determining gene on the Y chromosome. The deletion-insertion is likely to result in altered *SOX3* expression, as *SOX3* expression has been reported to be sensitive to position-effects caused by X-chromosome abnormalities [46]. Indeed, reporter-construct studies of the mouse *Sox3* gene have demonstrated the presence of both 5' and 3' regulatory elements [47] and thus it is possible that the deletion-insertion in the X-linked recessive hypoparathyroid patients may have a position-effect on *SOX3* expression. More recently, another unrelated family with X-linked hypoparathyroidism was identified and whole genome sequence (WGS) analysis revealed an interstitial deletion-insertion that cosegregated with the hypoparathyroidism [48]. This deletion-insertion involves an ~50 kb region of chromosome 2p25.3 that has been duplicated and inserted into the X-chromosome, and which has resulted in a 1.4 kb deletion located 81.5 kb downstream of *SOX3* [48]. Thus, these findings support a role for *SOX3* in parathyroid gland development, and consistent with this, *SOX*3 has been shown to be expressed in the developing parathyroids of mouse embryos [45].

A recent report has also highlighted that the four and a half LIM domains 1 (*FHL1*) gene, which is located on chromosome Xq26.3, may also represent an additional gene for X-linked hypoparathyroidism [49], as a *FHL1* missense variant (p.Arg95Trp) was detected in a 4 year old male with isolated severe hypoparathyroidism [49]. Functional studies showed that knockdown of the FHL1 paralog in zebrafish increased plasma calcium concentrations, and that introduction of wild-type FHL1 into cells transfected with a luciferase reporter gene that is under the control of the human PTH promoter, resulted in increased PTH promoter activity [49]. However, FHL1 is predicted to have a low level of expression in human parathyroids [49], and thus the role of this protein in calcium homeostasis remains to be established. The possible deleterious effects on mammalian parathyroid gland activity of this p.Arg95Trp missense variant of FHL1 also remain to be established.

Acquired Forms of Isolated Hypoparathyroidism

Anterior neck surgery represents the major cause of acquired hypoparathyroidism (Box 4.5.1) and accounts for ~75% of cases [1]. Hypoparathyroidism may arise in patients undergoing total parathyroidectomy, or because of intraoperative trauma to the parathyroid glands, inadvertent gland removal, or gland devascularization; and occurs most commonly following total thyroidectomy or after radical neck dissection for head and neck malignancies [50]. Hypoparathyroidism may also occur following neck

irradiation, or because of parathyroid gland infiltration by metastases or systemic disease (e.g. amyloidosis and sarcoidosis), or from copper or iron overload (e.g. Wilson's disease, haemochromatosis, or recurrent blood transfusions for thalassaemia) [1]. Transient or reversible forms of acquired hypoparathyroidism include neonatal hypoparathyroidism caused by maternal hypercalcaemia, and hypoparathyroidism caused by hypomagnesaemia (serum magnesium <0.5 mmol/L) [17, 51]. Some cases of isolated hypoparathyroidism, in which no other cause can be identified, have been considered to be caused by autoimmune destruction of the parathyroid glands. However, formal criteria for diagnosing isolated autoimmune hypoparathyroidism have not been established and suitable autoantibody tests are not available for clinical use [1].

Complex Syndromes Associated with Hypoparathyroidism

Hypoparathyroidism may be associated with congenital developmental anomalies such as the DiGeorge and hypoparathyroidism, deafness, and renal anomalies (HDR) syndromes, and also syndromes associated with growth failure or dwarfism (Table 4.5.2). Syndromic forms of hypoparathyroidism may also be caused by mitochondrial disorders or the pluriglandular autoimmune syndrome.

DiGeorge Syndrome

Patients with the DiGeorge syndrome (DGS) typically suffer from hypoparathyroidism, immunodeficiency, cardiac outflow tract malformations; facial dysmorphia and palatal dysfunction [52]. DGS has been reported in up to 60% of children with familial or idiopathic forms of hypoparathyroidism [53], and has a wide spectrum of severity. Most commonly, hypoparathyroidism associated with DGS presents in the neonatal period with marked hypocalcaemia, which may cause laryngospasm or seizures [54]. However, the hypoparathyroidism may be transient and characterized by serum calcium concentrations that normalize during infancy and early childhood. Some individuals with DGS may only develop hypocalcaemic symptoms in adolescence or adulthood [54, 55]. Most cases of DGS are sporadic but an autosomal dominant inheritance of DGS has been observed in up to 28% of cases [52]. DGS is most commonly caused by a heterozygous 3 Mb microdeletion of chromosome 22q11.2 [52], which is referred to as DGS type 1 (DGS1). The deleted region encompasses the *TBX1* gene, which encodes a T-box transcription factor that regulates histone modification and is involved in parathyroids and thymic development (Figure 4.5.4) [52]. *TBX1* mutations have been identified in some DGS patients who did not have deletions of chromosome 22q11.2, and demonstrated to lead to all of the major phenotypes of DGS [56]. Furthermore, transgenic mice with deletion of *Tbx1* have been reported to have a phenotype that is similar to that of DGS1 patients [57].

Chromosome 10p deletions have been detected in some DGS patients [58] and this is referred to as DGS type 2 (DGS2). The nebulette (*NEBL*) gene, located on chromosome 10p, has been reported to be deleted in DGS2 patient cell lines, and thus may be the responsible gene for DGS2 (Table 4.5.2) [59]. Moreover, some DGS patients have features of the CHARGE syndrome, which is characterized by coloboma, heart abnormalities, choanal atresia, growth retardation, and genitourinary and/or ear anomalies [60]; and is caused by heterozygous mutations of the *CHD7* gene (Table 4.5.2). *CHD7* encodes the chromodomain helicase DNA binding protein 7 [60], which is expressed within the pharyngeal ectoderm [61], and may play a role in pharyngeal region development.

Hypoparathyroidism, Deafness, and Renal Anomalies (HDR) Syndrome

The combined inheritance of HDR anomalies as an autosomal dominant trait was reported in one family in 1992 [62]. Patients had asymptomatic hypocalcaemia with undetectable or inappropriately normal serum concentrations of PTH, and normal brisk increases in plasma cAMP in response to the infusion of PTH. The patients also had bilateral, symmetrical, sensorineural deafness involving all frequencies. The renal abnormalities consisted mainly of bilateral cysts that compressed the glomeruli and tubules and lead to renal impairment in some patients. Cytogenetic and deletion mapping studies involving additional HDR patients defined a critical 200-kb region on 10p14–10pter that contained the *GATA3* gene (Table 4.5.2) [63]. GATA3 belongs to a family of dual zinc-finger transcription factors that are involved in vertebrate embryonic development. DNA sequence analysis in HDR patients has identified mutations that resulted in a haploinsufficiency and loss of GATA3 function [63–65]. To date, more than 50 different heterozygous germline *GATA3* abnormalities, which comprise point-mutations and whole gene deletions have been reported [63–71]. The majority (>75%) of these HDR associated mutations are predicted to result in truncated forms of the GATA3 protein. Over 90% of patients with two or three of the major clinical features of the HDR syndrome (i.e. hypoparathyroidism, deafness, or renal abnormalities), have a *GATA3* mutation [64]. The HDR phenotype is consistent with the expression pattern of GATA3 during human and mouse embryogenesis in the developing kidney, otic vesicle, and parathyroids. However, GATA3 is also expressed in developing central nervous system (CNS) and the haematopoietic organs in man and mice, and this suggests that GATA3 may have a more complex role. Indeed, homozygous *Gata3* null mice ($Gata3^{-/-}$) have defects of the CNS and a lack of T-cell development, and die *in utero* [72]. Whereas, heterozygous *Gata3* null mice ($Gata3^{+/-}$) are viable and have hearing loss [73], in addition to having a smaller parathyroid-thymus primordia and an inadequate increase in plasma PTH in response to induced hypocalcaemia [74]. These studies of HDR patients and *Gata3* mutant mice indicate an important role for GATA3 in parathyroid development (Figure 4.5.4).

Additional familial syndromes characterized by hypoparathyroidism, deafness, and/or renal anomalies such as Barakat syndrome have been reported (Table 4.5.2) [75]. Molecular genetic studies have not been reported for these families.

Mitochondrial Disorders Associated with Hypoparathyroidism

Hypoparathyroidism has been reported to occur in three mitochondrial disorders: the Kearns–Sayre syndrome (KSS), the MELAS syndrome (mitochondrial encephalopathy, stroke-like episodes, and lactic acidosis) and the mitochondrial trifunctional protein (MTP) deficiency syndrome [76–78]. KSS is characterized by progressive external ophthalmoplegia and pigmentary retinopathy occurring at less than 20 years of age [77]. The MELAS syndrome consists of a childhood onset of mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes [76]. Mitochondrial DNA abnormalities such as deletions, duplications, and missense substitutions cause KSS and MELAS (Table 4.5.2, Figure 4.5.1) [53, 58, 76]. Whereas,

MTP deficiency syndrome, which is a disorder of fatty-acid oxidation associated with cardiomyopathy, peripheral neuropathy, pigmentary retinopathy and liver dysfunction, is caused by mutations of the *HADHB* gene on chromosome 2p23 (Table 4.5.2), which encodes the MTP β-subunit [78].

Hypoparathyroid Syndromes Associated with Growth Failure or Dwarfism

Hypoparathyroidism occurs in more than 50% of patients with Kenny-Caffey syndrome (KCS) [58], which is also characterized by short stature and osteosclerotic bone dysplasia. KCS may be inherited as an autosomal recessive (KCS type 1, KCS1) or dominant (KCS type 2, KCS2) disorder [58]. The Sanjad-Sakati syndrome (SSS), which is also known as the hypoparathyroidism-retardation-dysmorphism syndrome, affects Middle Eastern populations and has a similar phenotype to KCS [79]. SSS and KCS1 are caused by mutations of the *TBCE* gene (Table 4.5.2) [80], which encodes a chaperone protein required for the correct folding of α-tubulin subunits (Figure 4.5.1) [80], and is postulated to play a role in parathyroid gland migration [81]. KCS2 is caused by heterozygous missense mutations of the family with sequence similarity 111 member A (*FAM111A*) gene (Table 4.5.2), which modulates DNA replication and chromatin maturation (Figure 4.5.1), and may be involved in embryonic development [82, 83]. A truncating *FAM111A* mutation, p.Ser342del, has been shown to cause gracile bone dysplasia (GBD) (Figure 4.5.1), which occurs in association with hypoparathyroidism, and represents a perinatally lethal condition [83]. Hypoparathyroidism has also been reported in association with the Dubowitz syndrome (Table 4.5.2), which is characterized by multiple congenital anomalies that include microcephaly and growth retardation [84].

Pluriglandular Autoimmune Hypoparathyroidism

Hypoparathyroidism may occur in association with candidiasis and autoimmune Addison's disease, and the disorder has been referred to as either the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) syndrome or the autoimmune polyglandular syndrome type 1 (APS1) [85]. This disorder has a high incidence in Finland [86], and a genetic analysis of Finnish families indicated autosomal recessive inheritance of the disorder, and mapped the *APECED* gene to chromosome 21q22.3 (Table 4.5.2) [87]. Further positional cloning approaches led to the isolation of the autoimmune regulator (*AIRE)* gene from chromosome 21q22.3. This gene encodes a 545 amino acid protein that contains transcriptional factor motifs such as zinc-finger motifs, a proline-rich region and LXXLL motifs [88, 89]. Greater than 100 APECED-causing *AIRE* mutations have been reported to date, and four *AIRE* mutations have been shown to commonly occur in APECED families and these are: p.Arg257Stop in Finnish, German, Swiss, British and Northern Italian families; p.Arg139Stop in Sardinian families; p.Tyr85Cys in Iranian Jewish families; and a 13 bp deletion in exon 8 in British, Dutch, German, and Finnish families [89–91]. Occasionally, biallelic *AIRE* mutations have been reported in patients with isolated hypoparathyroidism [92, 93]. AIRE has been shown to regulate the elimination of organ-specific T cells in the thymus, and thus APECED is likely to be caused by a failure of this specialized mechanism for deleting forbidden T cells, and establishing immunologic tolerance [94].

Autosomal Dominant Hypocalcaemia

Autosomal dominant hypocalcaemia (ADH) is comprised of two genetically distinct disorders, designated as ADH types 1 and 2, which are caused by germline gain-of-function mutations of the CaSR and $G\alpha_{11}$ proteins, respectively (Table 4.5.2) [95].

CaSR and $G\alpha_{11}$

The human CaSR is a $G\alpha_{q/11}$ and $G\alpha_{i/o}$ -coupled G-protein coupled receptor (Figure 4.5.1), encoded by the *CASR* gene on chromosome 3q21.1 [95]. The CaSR is highly expressed in the parathyroid glands and kidneys, where it regulates the set-point for PTH secretion and decreases renal tubular calcium reabsorption, respectively [95]. CaSR cell-surface expression is regulated by agonist-driven insertional signalling (ADIS), which promotes anterograde trafficking of newly synthesized CaSRs to the plasma membrane [96], and also by retrograde trafficking of cell-surface CaSRs, which is mediated by the adaptor-related protein complex-2 (AP2) [97].

Autosomal Dominant Hypocalcaemia Type 1

Autosomal dominant hypocalcaemia type 1 (ADH1), which is due to gain-of-function CaSR mutations (Table 4.5.2), is characterized by hypocalcaemia, mild hypomagnesaemia, hyperphosphataemia, and serum PTH concentrations which are often detectable and within the lower half of the reference range [98]. Patients with ADH also have significantly increased urinary calcium excretion compared to hypoparathyroid patients [99]. Some patients with ADH have been reported to have ectopic calcifications affecting the basal ganglia and/or elevations in bone mineral density (BMD) [100], and some patients with severe gain-of-function CaSR mutations may also develop Bartter syndrome type 5, which is characterized by renal salt wasting leading to volume depletion, hyper-reninaemic hyperaldosteronism and hypokalaemic alkalosis [101]. Over 70 different CaSR mutations have been identified in ADH1 patients and families to date [95, 102].

Autosomal Dominant Hypocalcaemia Type 2

Mutations of the *CASR* gene are not detected in ~30% of ADH patients, and germline mutations of $G\alpha_{11}$, which is encoded by the *GNA11* gene on chromosome 19p13.3 (Table 4.5.2), have been reported in seven such ADH patients and families [98, 103–106]. *In vitro* studies of these $G\alpha_{11}$ mutations, which all comprise missense substitutions, have shown cells expressing the mutant $G\alpha_{11}$ proteins to have enhanced CaSR-mediated signalling responses, consistent with a gain-of-function [98, 103, 105]. These individuals and families with gain-of-function $G\alpha_{11}$ mutations were designated as having ADH type 2 (ADH2). ADH2 is associated with a milder urinary phenotype, with significantly reduced urinary calcium excretion compared to ADH1 [103]. Moreover, short stature caused by postnatal growth insufficiency has been reported in two ADH2 kindreds [103, 106].

Pseudohypoparathyroidism

Pseudohypoparathyroidism is caused by renal resistance to PTH, and is mainly due to abnormalities of Gαs, which mediates PTH1 receptor signalling (Figure 4.5.1) [15]. The Gαs protein is encoded by *GNAS*, which is a complex imprinted locus, and maternally

Table 4.5.3 Clinical, biochemical, and genetic features of hypoparathyroid and pseudohypoparathyroid disorders

	Hypoparathyroidism	Pseudohypoparathyroidism (PHP)				
		PHPIa	PPHP	PHPIb	PHPIc	PHPII
AHO manifestations	No	Yes	Yes	No	Yes	No
Serum calcium	↓	↓	N	↓	↓	↓
Serum PO_4	↑	↑	N	↑	↑	↑
Serum PTH	↓	↑	N	↑	↑	↑
Response to PTH:						
Urinary cAMP*a* (Chase–Aurbach test)	↑	↓	↑	↓	↓	↑
Urinary PO_4 (Ellsworth–Howard test)	↑	↓	↑	↓	↓	↓
Gsα activity	N	↓	↓	N	N	N
Inheritance	AD/AR/X	AD	AD	AD	AD	Sporadic
Molecular defect	PTH/CaSR/ GATA3/ Gcm2/ others	GNAS1	GNAS1	?GNAS1	?adenyl cyclase	?cAMP targets
Other hormonal resistance	No	Yes	No	No	Yes	No

↓ = decreased, ↑ = increased, N = normal, AD = autosomal dominant, AR = autosomal recessive, X = X-linked, AHO = Albright's hereditary osteodystrophy,? = presumed, but not proven.
*a*plasma cAMP responses are similar to those of urinary cAMP.

inherited inactivating coding-region mutations give rise to PHP type 1a (PHP1a) (Table 4.5.3), which is characterized by PTH resistance (hypocalcaemia, hyperphosphataemia, elevated serum PTH, and blunted cAMP response to exogenous PTH), together with features of Albright's hereditary osteodystrophy (AHO) such as short stature, obesity, round facies, subcutaneous calcifications, and brachydactyly [15]. Whereas, paternally inherited inactivating coding-region *GNAS* mutations cause pseudopseudohypoparathyroidism, which is characterized by AHO in the absence of PTH resistance (Table 4.5.3). Abnormalities of the upstream regions of the *GNAS* locus, or of the closely linked *STX16* and *NESP55* genes, which affect *GNAS* methylation, have been reported to cause PHP type 1b (Table 4.5.3) [107, 108], which is associated with PTH resistance in the absence of AHO [15]. Epigenetic or coding-region *GNAS* abnormalities may also cause PHP type 1c (PHP1c) (Table 4.5.3), which has a similar phenotype to PHP1a, except that altered Gαs activity is not observed in *in vitro* assays involving PHP1c patient-derived erythrocytes [109]. In contrast to PHP type 1, which has a blunted cAMP response; PHP type 2 is associated with a conserved cAMP response, and in some cases may represent an acquired condition caused by vitamin D deficiency [110].

REFERENCES

1. Mannstadt M, Bilezikian JP, Thakker RV, *et al*. Hypoparathyroidism. *Nat Rev Dis Primers*, 2017; **3**: 17055.
2. Hannan FM, Thakker RV. Investigating hypocalcaemia. *BMJ*, 2013; **346**: f2213.
3. Akirov A, Gorshtein A, Shraga-Slutzky I, Shimon I. Calcium levels on admission and before discharge are associated with mortality risk in hospitalized patients. *Endocrine*, 2017; **57**: 344–51.
4. Page C, Strunski V. Parathyroid risk in total thyroidectomy for bilateral, benign, multinodular goitre: report of 351 surgical cases. *J Laryngol Otol*, 2007; **121**: 237–41.
5. Asari R, Passler C, Kaczirek K, Scheuba C, Niederle B. Hypoparathyroidism after total thyroidectomy: a prospective study. *Arch Surg*, 2008; **143**: 132–7.
6. Astor MC, Lovas K, Debowska A, *et al*. Epidemiology and health-related quality of life in hypoparathyroidism in Norway. *J Clin Endocrinol Metab*, 2016; **101**: 3045–53.
7. Vadiveloo T, Donnan PT, Leese GP. A population-based study of the epidemiology of chronic hypoparathyroidism. *J Bone Miner Res*, 2018; **33**: 478–85.
8. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. Cardiovascular and renal complications to postsurgical hypoparathyroidism: a Danish nationwide controlled historic follow-up study. *J Bone Miner Res*, 2013; **28**: 2277–85.
9. Powers J, Joy K, Ruscio A, Lagast H. Prevalence and incidence of hypoparathyroidism in the United States using a large claims database. *J Bone Miner Res*, 2013; **28**: 2570–6.
10. Clarke BL, Leibson C, Emerson J, Ransom JE, Lagast H. Co-morbid-medical conditions associated with prevalent hypoparathyroidism: a population-based study. *J Bone Miner Res*, 2011; **26**: S182 (Abstract SA1070).
11. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. Postsurgical hypoparathyroidism—risk of fractures, psychiatric diseases, cancer, cataract, and infections. *J Bone Miner Res*, 2014; **29**: 2504–10.
12. Underbjerg L, Sikjaer T, Mosekilde L, Rejnmark L. The epidemiology of nonsurgical hypoparathyroidism in Denmark: a nationwide case finding study. *J Bone Miner Res*, 2015; **30**: 1738–44.
13. Hari Kumar KVS, Patnaik SK. Incidence of parathyroid disorders in Indian adult male population: a 25-year follow-up study. *Clin Endocrinol (Oxf)*, 2017; **87**: 605–8.
14. Goswami R, Sharma R, Sreenivas V, *et al*. Prevalence and progression of basal ganglia calcification and its pathogenic mechanism in patients with idiopathic hypoparathyroidism. *Clin Endocrinol (Oxf)*, 2012; **77**: 200–6.
15. Lemos MC, Thakker RV. GNAS mutations in Pseudohypoparathyroidism type 1a and related disorders. *Hum Mutat*, 2015; **36**: 11–19.

16. Shoback DM. Hypocalcemia Management. In: De Groot LJ, Chrousos G, Dungan K, *et al.*, eds. Endotext. South Dartmouth, MA, 2000.
17. Ayuk J, Gittoes NJ. How should hypomagnesaemia be investigated and treated? *Clin Endocrinol (Oxf)*, 2011; **75**: 743–6.
18. Bilezikian JP, Brandi ML, Cusano NE, *et al.* Management of hypoparathyroidism: present and future. *J Clin Endocrinol Metab*, 2016; **101**: 2313–24.
19. Bollerslev J, Rejnmark L, Marcocci C, *et al.* European Society of Endocrinology clinical guideline: treatment of chronic hypoparathyroidism in adults. *Eur J Endocrinol*, 2015; **173**: G1–20.
20. Brandi ML, Bilezikian JP, Shoback D, *et al.* Management of hypoparathyroidism: summary statement and guidelines. *J Clin Endocrinol Metab*, 2016; **101**: 2273–83.
21. Mannstadt M, Clarke BL, Vokes T, *et al.* Efficacy and safety of recombinant human parathyroid hormone (1-84) in hypoparathyroidism (REPLACE): a double-blind, placebo-controlled, randomised, phase 3 study. *Lancet Diabetes Endocrinol*, 2013; **1**: 275–83.
22. Cusano NE, Rubin MR, McMahon DJ, *et al.* The effect of PTH(1-84) on quality of life in hypoparathyroidism. *J Clin Endocrinol Metab*, 2013; **98**: 2356–61.
23. Naylor SL, Sakaguchi AY, Szoka P, *et al.* Human parathyroid hormone gene (PTH) is on short arm of chromosome 11. *Somatic Cell Genet*, 1983; **9**: 609–16.
24. Parkinson DB, Thakker RV. A donor splice site mutation in the parathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *Nat Genet*, 1992; **1**: 149–52.
25. Arnold A, Horst SA, Gardella TJ, *et al.* Mutation of the signal peptide-encoding region of the preproparathyroid hormone gene in familial isolated hypoparathyroidism. *J Clin Invest*, 1990; **86**: 1084–7.
26. Cinque L, Sparaneo A, Penta L, *et al.* Autosomal dominant PTH gene signal sequence mutation in a family with familial isolated hypoparathyroidism. *J Clin Endocrinol Metab*, 2017; **102**: 3961–9.
27. Datta R, Waheed A, Shah GN, Sly WS. Signal sequence mutation in autosomal dominant form of hypoparathyroidism induces apoptosis that is corrected by a chemical chaperone. *Proc Natl Acad Sci U S A*, 2007; **104**: 19989–94.
28. Ertl DA, Stary S, Streubel B, Raimann A, Haeusler G. A novel homozygous mutation in the parathyroid hormone gene (PTH) in a girl with isolated hypoparathyroidism. *Bone*, 2012; **51**: 629–32.
29. Sunthornthepvarakul T, Churesigaew S, Ngowngarmratana S. A novel mutation of the signal peptide of the preproparathyroid hormone gene associated with autosomal recessive familial isolated hypoparathyroidism. *J Clin Endocrinol Metab*, 1999; **84**: 3792–6.
30. Lee S, Mannstadt M, Guo J, *et al.* A homozygous [Cys25]PTH(1-84) mutation that impairs PTH/PTHrP receptor activation defines a novel form of hypoparathyroidism. *J Bone Miner Res*, 2015; **30**: 1803–13.
31. Liu Z, Yu S, Manley NR. Gcm2 is required for the differentiation and survival of parathyroid precursor cells in the parathyroid/thymus primordia. *Dev Biol*, 2007; **305**: 333–46.
32. Canaff L, Zhou X, Mosesova I, Cole DE, Hendy GN. Glial cells missing-2 (GCM2) transactivates the calcium-sensing receptor gene: effect of a dominant-negative GCM2 mutant associated with autosomal dominant hypoparathyroidism. *Hum Mutat*, 2009; **30**: 85–92.
33. Han SI, Tsunekage Y, Kataoka K. Gata3 cooperates with Gcm2 and MafB to activate parathyroid hormone gene expression by interacting with SP1. *Mol Cell Endocrinol*, 2015; **411**: 113–20.
34. Kamitani-Kawamoto A, Hamada M, *et al.* MafB interacts with Gcm2 and regulates parathyroid hormone expression and parathyroid development. *J Bone Miner Res*, 2011; **26**: 2463–72.
35. Gunther T, Chen ZF, Kim J, *et al.* Genetic ablation of parathyroid glands reveals another source of parathyroid hormone. *Nature*, 2000; **406**: 199–203.
36. Ding C, Buckingham B, Levine MA. Familial isolated hypoparathyroidism caused by a mutation in the gene for the transcription factor GCMB. *J Clin Invest*, 2001; **108**: 1215–20.
37. Baumber L, Tufarelli C, Patel S, *et al.* Identification of a novel mutation disrupting the DNA binding activity of GCM2 in autosomal recessive familial isolated hypoparathyroidism. *J Med Genet*, 2005; **42**: 443–8.
38. Mannstadt M, Bertrand G, Muresan M, *et al.* Dominant-negative GCMB mutations cause an autosomal dominant form of hypoparathyroidism. *J Clin Endocrinol Metab*, 2008; **93**: 3568–76.
39. Mirczuk SM, Bowl MR, Nesbit MA, *et al.* A missense glial cells missing homolog B (GCMB) mutation, Asn502His, causes autosomal dominant hypoparathyroidism. *J Clin Endocrinol Metab*, 2010; **95**: 3512–16.
40. Peden VH. True idiopathic hypoparathyroidism as a sex-linked recessive trait. *Am J Hum Genet*, 1960; **12**: 323–37.
41. Whyte MP, Weldon VV. Idiopathic hypoparathyroidism presenting with seizures during infancy: X-linked recessive inheritance in a large Missouri kindred. *J Pediatr*, 1981; **99**: 608–11.
42. Whyte MP, Kim GS, Kosanovich M. Absence of parathyroid tissue in sex-linked recessive hypoparathyroidism. *J Pediatr*, 1986; **109**: 915.
43. Mumm S, Whyte MP, Thakker RV, Buetow KH, Schlessinger D. mtDNA analysis shows common ancestry in two kindreds with X-linked recessive hypoparathyroidism and reveals a heteroplasmic silent mutation. *Am J Hum Genet*, 1997; **60**: 153–9.
44. Thakker RV, Davies KE, Whyte MP, Wooding C, O'Riordan JL. Mapping the gene causing X-linked recessive idiopathic hypoparathyroidism to Xq26-Xq27 by linkage studies. *J Clin Invest*, 1990; **86**: 40–5.
45. Bowl MR, Nesbit MA, Harding B, *et al.* An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. *J Clin Invest*, 2005; **115**: 2822–31.
46. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*, 2005; **76**: 8–32.
47. Brunelli S, Silva Casey E, Bell D, Harland R, Lovell-Badge R. Expression of Sox3 throughout the developing central nervous system is dependent on the combined action of discrete, evolutionarily conserved regulatory elements. *Genesis*, 2003; **36**: 12–24.
48. Taylor JC, Martin HC, Lise S, *et al.* Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*, 2015; **47**: 717–26.
49. Pillar N, Pleniceanu O, Fang M, *et al.* A rare variant in the FHL1 gene associated with X-linked recessive hypoparathyroidism. *Hum Genet*, 2017; **136**: 835–45.
50. Lorente-Poch L, Sancho JJ, Ruiz S, Sitges-Serra A. Importance of in situ preservation of parathyroid glands during total thyroidectomy. *Br J Surg*, 2015; **102**: 359–67.
51. Jain A, Agarwal R, Sankar MJ, Deorari A, Paul VK. Hypocalcemia in the newborn. *Indian J Pediatr*, 2010; **77**: 1123–8.

52. Kobrynski LJ, Sullivan KE. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet*, 2007; **370**: 1443–52.
53. Kim JH, Shin YL, Yang S, *et al*. Diverse genetic aetiologies and clinical outcomes of paediatric hypoparathyroidism. *Clin Endocrinol (Oxf)*, 2015; **83**: 790–6.
54. Hieronimus S, Bec-Roche M, Pedeutour F, *et al*. The spectrum of parathyroid gland dysfunction associated with the microdeletion 22q11. *Eur J Endocrinol*, 2006; **155**: 47–52.
55. Nakada Y, Terui K, Kageyama K, *et al*. An adult case of 22q11.2 deletion syndrome diagnosed in a 36-year-old woman with hypocalcemia caused by hypoparathyroidism and Hashimoto's thyroiditis. *Intern Med*, 2013; **52**: 1365–8.
56. Yagi H, Furutani Y, Hamada H, *et al*. Role of TBX1 in human del22q11.2 syndrome. *Lancet*, 2003; **362**: 1366–73.
57. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet*, 2001; **27**: 286–91.
58. Thakker RV, Bringhurst FR, Juppner H. *Regulation of Calcium Homeostasis and Genetic Disorders that Affect Calcium Metabolism*, 7th edn. Philadelphia, PA: Saunders/Elsevier, 2016.
59. Villanueva MP, Aiyer AR, Muller S, *et al*. Genetic and comparative mapping of genes dysregulated in mouse hearts lacking the Hand2 transcription factor gene. *Genomics*, 2002; **80**: 593–600.
60. Inoue H, Takada H, Kusuda T, *et al*. Successful cord blood transplantation for a CHARGE syndrome with CHD7 mutation showing DiGeorge sequence including hypoparathyroidism. *Eur J Pediatr*, 2010; **169**: 839–44.
61. Randall V, McCue K, Roberts C, *et al*. Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. *J Clin Invest*, 2009; **119**: 3301–10.
62. Bilous RW, Murty G, Parkinson DB, *et al*. Brief report: autosomal dominant familial hypoparathyroidism, sensorineural deafness, and renal dysplasia. *N Engl J Med*, 1992; **327**: 1069–74.
63. Van Esch H, Groenen P, Nesbit MA, *et al*. GATA3 haploinsufficiency causes human HDR syndrome. *Nature*, 2000; **406**: 419–22.
64. Ali A, Christie PT, Grigorieva IV, *et al*. Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet*, 2007; **16**: 265–75.
65. Nesbit MA, Bowl MR, Harding B, *et al*. Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem*, 2004; **279**: 22624–34.
66. Adachi M, Tachibana K, Asakura Y, Tsuchiya T. A novel mutation in the GATA3 gene in a family with HDR syndrome (hypoparathyroidism, sensorineural deafness and renal anomaly syndrome). *J Pediatr Endocrinol Metab*, 2006; **19**: 87–92.
67. Fukami M, Muroya K, Miyake T, *et al*. GATA3 abnormalities in six patients with HDR syndrome. *Endocr J*, 2011; **58**: 117–21.
68. Gaynor KU, Grigorieva IV, Nesbit MA, *et al*. A missense GATA3 mutation, Thr272Ile, causes the hypoparathyroidism, deafness, and renal dysplasia syndrome. *J Clin Endocrinol Metab*, 2009; **94**: 3897–904.
69. Mino Y, Kuwahara T, Mannami T, *et al*. Identification of a novel insertion mutation in GATA3 with HDR syndrome. *Clin Exp Nephrol*, 2005; **9**: 58–61.
70. Muroya K, Hasegawa T, Ito Y, *et al*. GATA3 abnormalities and the phenotypic spectrum of HDR syndrome. *J Med Genet*, 2001; **38**: 374–80.
71. Zahirieh A, Nesbit MA, Ali A, *et al*. Functional analysis of a novel GATA3 mutation in a family with the hypoparathyroidism, deafness, and renal dysplasia syndrome. *J Clin Endocrinol Metab*, 2005; **90**: 2445–50.
72. Pandolfi PP, Roth ME, Karis A, *et al*. Targeted disruption of the *GATA3* gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet*, 1995; **11**: 40–4.
73. van der Wees J, van Looij MA, de Ruiter MM, *et al*. Hearing loss following Gata3 haploinsufficiency is caused by cochlear disorder. *Neurobiol Dis*, 2004; **16**: 169–78.
74. Grigorieva IV, Mirczuk S, Gaynor KU, *et al*. Gata3-deficient mice develop parathyroid abnormalities due to dysregulation of the parathyroid-specific transcription factor Gcm2. *J Clin Invest*, 2010; **120**: 2144–55.
75. Barakat AY, D'Albora JB, Martin MM, Jose PA. Familial nephrosis, nerve deafness, and hypoparathyroidism. *J Pediatr*, 1977; **91**: 61–4.
76. El-Hattab AW, Adesina AM, Jones J, Scaglia F. MELAS syndrome: clinical manifestations, pathogenesis, and treatment options. *Mol Genet Metab*, 2015; **116**: 4–12.
77. Harvey JN, Barnett D. Endocrine dysfunction in Kearns-Sayre syndrome. *Clin Endocrinol (Oxf)*, 1992; **37**: 97–103.
78. Naiki M, Ochi N, Kato YS, *et al*. Mutations in HADHB, which encodes the beta-subunit of mitochondrial trifunctional protein, cause infantile onset hypoparathyroidism and peripheral polyneuropathy. *Am J Med Genet A*, 2014; **164A**: 1180–7.
79. Naguib KK, Gouda SA, Elshafey A, *et al*. Sanjad-Sakati syndrome/Kenny-Caffey syndrome type 1: a study of 21 cases in Kuwait. *East Mediterr Health J*, 2009; **15**: 345–52.
80. Parvari R, Hershkovitz E, Grossman N, *et al*. Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. *Nat Genet*, 2002; **32**: 448–52.
81. Parvari R, Diaz GA, Hershkovitz E. Parathyroid development and the role of tubulin chaperone E. *Horm Res*, 2007; **67**: 12–21.
82. Alabert C, Bukowski-Wills JC, Lee SB, *et al*. Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat Cell Biol*, 2014; **16**: 281–93.
83. Unger S, Gorna MW, Le Bechec A, *et al*. FAM111A mutations result in hypoparathyroidism and impaired skeletal development. *Am J Hum Genet*, 2013; **92**: 990–5.
84. Lerman-Sagie T, Merlob P, Shuper A, *et al*. New findings in a patient with Dubowitz syndrome: velopharyngeal insufficiency and hypoparathyroidism. *Am J Med Genet*, 1990; **37**: 241–3.
85. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med*, 1990; **322**: 1829–36.
86. Perheentupa J. APS-I/APECED: the clinical disease and therapy. *Endocrinol Metab Clin North Am*, 2002; **31**: 295–320, vi.
87. Aaltonen J, Bjorses P, Sandkuijl L, Perheentupa J, Peltonen L. An autosomal locus causing autoimmune disease: autoimmune polyglandular disease type I assigned to chromosome 21. *Nat Genet*, 1994; **8**: 83–7.
88. consortium TF-GA. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc finger domains. *Nat Genet*, 1997; **17**: 399–403.
89. Nagamine K, Peterson P, Scott HS, *et al*. Positional cloning of the *APECED* gene. *Nat Genet*, 1997; **17**: 393–8.

90. Heino M, Peterson P, Kudoh J, *et al*. *APECED* mutations in the autoimmune regulator (AIRE) gene. *Hum Mutat*, 2001; **18**: 205–11.
91. Pearce SH, Cheetham T, Imrie H, *et al*. A common and recurrent 13-bp deletion in the autoimmune regulator gene in British kindreds with autoimmune polyendocrinopathy type 1. *Am J Hum Genet*, 1998; **63**: 1675–84.
92. Li D, Streeten EA, Chan A, *et al*. Exome sequencing reveals mutations in AIRE as a cause of isolated hypoparathyroidism. *J Clin Endocrinol Metab*, 2017; **102**: 1726–33.
93. Sahoo SK, Zaidi G, Srivastava R, *et al*. Identification of autoimmune polyendocrine syndrome type 1 in patients with isolated hypoparathyroidism. *Clin Endocrinol (Oxf)*, 2016; **85**: 544–50.
94. Liston A, Lesage S, Wilson J, Peltonen L, Goodnow CC. Aire regulates negative selection of organ-specific T cells. *Nat Immunol*, 2003; **4**: 350–4.
95. Hannan FM, Babinsky VN, Thakker RV. Disorders of the calcium-sensing receptor and partner proteins: insights into the molecular basis of calcium homeostasis. *J Mol Endocrinol*, 2016; **57**: R127–42.
96. Grant MP, Stepanchick A, Cavanaugh A, Breitwieser GE. Agonist-driven maturation and plasma membrane insertion of calcium-sensing receptors dynamically control signal amplitude. *Sci Signal*, 2011; **4**: ra78.
97. Gorvin CM, Rogers A, Hastoy B, *et al*. AP2? mutations impair calcium-sensing receptor trafficking and signaling, and show an endosomal pathway to spatially direct G-protein selectivity. *Cell Rep*, 2018; **22**: 1054–66.
98. Nesbit MA, Hannan FM, Howles SA, *et al*. Mutations affecting G-protein subunit alpha11 in hypercalcemia and hypocalcemia. *N Engl J Med*, 2013; **368**: 2476–86.
99. Yamamoto M, Akatsu T, Nagase T, Ogata E. Comparison of hypocalcemic hypercalciuria between patients with idiopathic hypoparathyroidism and those with gain-of-function mutations in the calcium-sensing receptor: is it possible to differentiate the two disorders? *J Clin Endocrinol Metab*, 2000; **85**: 4583–91.
100. Pearce SH, Williamson C, Kifor O, *et al*. A familial syndrome of hypocalcemia with hypercalciuria due to mutations in the calcium-sensing receptor. *N Engl J Med*, 1996; **335**: 1115–22.
101. Watanabe S, Fukumoto S, Chang H, *et al*. Association between activating mutations of calcium-sensing receptor and Bartter's syndrome. *Lancet*, 2002; **360**: 692–4.
102. Hannan FM, Thakker RV. Calcium-sensing receptor (CaSR) mutations and disorders of calcium, electrolyte and water metabolism. *Best Pract Res Clin Endocrinol Metab*, 2013; **27**: 359–71.
103. Li D, Opas EE, Tuluc F, *et al*. Autosomal dominant hypoparathyroidism caused by germline mutation in GNA11: phenotypic and molecular characterization. *J Clin Endocrinol Metab*, 2014; **99**: E1774–83.
104. Mannstadt M, Harris M, Bravenboer B, *et al*. Germline mutations affecting Galpha11 in hypoparathyroidism. *N Engl J Med*, 2013; **368**: 2532–4.
105. Piret SE, Gorvin CM, Pagnamenta AT, *et al*. Identification of a G-protein subunit-alpha11 gain-of-function mutation, Val340Met, in a family with autosomal dominant hypocalcemia type 2 (ADH2). *J Bone Miner Res*, 2016; **31**: 1207–14.
106. Tenhola S, Voutilainen R, Reyes M, *et al*. Impaired growth and intracranial calcifications in autosomal dominant hypocalcemia caused by a GNA11 mutation. *Eur J Endocrinol*, 2016; **175**: 211–18.
107. Linglart A, Gensure RC, Olney RC, Juppner H, Bastepe M. A novel STX16 deletion in autosomal dominant pseudohypoparathyroidism type Ib redefines the boundaries of a cis-acting imprinting control element of GNAS. *Am J Hum Genet*, 2005; **76**: 804–14.
108. Richard N, Abeguile G, Coudray N, *et al*. A new deletion ablating NESP55 causes loss of maternal imprint of A/B GNAS and autosomal dominant pseudohypoparathyroidism type Ib. *J Clin Endocrinol Metab*, 2012; **97**: E863–7.
109. Brix B, Werner R, Staedt P, *et al*. Different pattern of epigenetic changes of the GNAS gene locus in patients with pseudohypoparathyroidism type Ic confirm the heterogeneity of underlying pathomechanisms in this subgroup of pseudohypoparathyroidism and the demand for a new classification of GNAS-related disorders. *J Clin Endocrinol Metab*, 2014; **99**: E1564–70.
110. Srivastava T, Alon US. Stage I vitamin D-deficiency rickets mimicking pseudohypoparathyroidism type II. *Clin Pediatr (Phila)*, 2002; **41**: 263–8.

4.6

Bones and the Kidney

The Practical Conundrum: Distinguishing Between Osteoporosis and the Bone Diseases that Accompany Chronic Renal Failure

Paul D. Miller and Michael Pazianas

Introduction

Osteoporosis may coexist with chronic kidney disease–mineral and bone disorder (CKD-MBD) and osteoporotic fractures occur in all stages of CKD (**Box 4.6.1**) [1]. Management of osteoporosis in CKD should consider the pathophysiology of both disorders. Diagnosis and management of osteoporosis in patients with stages 1–3 CKD and patients without CKD are similar, but diagnosis and management decisions differ greatly once patients have stages 4–5 CKD. Discriminating between osteoporosis and CKD-MBD is best accomplished with quantitative bone histomorphometry. Biochemical markers, especially intact parathyroid hormone and bone-specific alkaline phosphatase also may be helpful. When the diagnosis of osteoporosis is established, management in stages 4–5 CKD may include antiresorptive or anabolic agents, though evidence for efficacy is marginal in advanced CKD.

Mechanisms of Reduction in GFR: Age vs. Intrinsic Renal Disease

Glomerular filtration rate (GFR) decreases with increasing age especially after age 60 [2]. Bone strength declines with age, in part due to the decline in bone quality [3, 4], which contributes ~50% to the overall strength of bone. Bone quality is a broad term that defines the microarchitectural changes in bone (trabecular number and connectivity, cortical thinning, and increased cortical porosity) that makes an older bone more likely to break when compared with younger bone, even at the same bone density. Many of the biochemical derangements that accompany progressive CKD are contributing to the altered bone quality as well as increased fracture risk seen in CKD patients when compared to the lower risk in subjects of same age and bone mineral density (BMD) but without CKD [5].

The average rate of GFR reduction that accompanies increasing age is ~1–2 ml/min/year [6]. The mechanism for age-related reduction in renal function is multifactorial and is certainly related, in part, to vascular changes that accompany ageing. Age-related decrease in GFR is not associated with meaningful proteinuria or reduction in renal size as measured by ultrasound (US) or computerized tomography (CT) [7]. While GFR may decrease with advancing age, it is unusual that age alone is responsible for GFR reduction to levels lower than 30 ml/min. Patients with GFR values less than 30 ml/min usually have other reasons for their more severe, stage 4–5 CKD (diabetes mellitus and hypertension, for example). In all of the postmenopausal osteoporosis clinical registration trials, patients were excluded if they had a GFR under 30 ml/min.

Fractures in Patients with CKD: The Pathophysiological Origins of Skeletal Fragility

Fracture risk doubles in CKD beginning with stage 3A as compared to subjects with comparable BMD and age without CKD [8]. The

Box 4.6.1 The National Kidney Foundation (NKF) has classified chronic kidney disease (CKD) into five stages according to the estimated glomerular filtration rate (eGFR)

*Stage 3 CKD has been subdivided by the Kidney Disease Improving Global Outcome (KDIGO) into two separate stages

Stage 1 CKD: eGFR ≥ 90 ml/min with evidence of intrinsic renal damage (proteinuria or microscopic haematuria).

Stage 2 CKD: eGFR = 60–89 ml/min with evidence of intrinsic renal damage (proteinuria or microscopic haematuria).

Stage 3 CKD: eGFR = 30–59 ml/min without evidence of intrinsic renal damage (proteinuria or microscopic haematuria)*.

- **Stage 3A: eGFR 30–59 ml/min, and**
- **Stage 3B: eGFR 30–44 ml/min**

Stage 4 CKD: eGFR = 15–29 ml/min without evidence of intrinsic renal damage.

Stage 5 CKD: eGFR <15 ml/min and has also been divided into those not on dialysis and those on dialysis (stage 5D).

exact mechanism(s) of why there is greater skeletal fragility by stage 3A CKD than in persons without CKD or even in stage 1–2 CKD are not clearly delineated (Box 4.6.2) [9, 10].

Parathyroid hormone (PTH) levels rise progressively once the GFR falls to below 60 ml/min [11]. Some of the highest levels of PTH are seen in severe CKD (stage 5). PTH levels above 350 pg/ml are highly associated with quantitative bone histomorphometry features of osteitis fibrosa cystica [12]. On the opposite side of the bone histomorphometry spectrum that is seen in CKD, PTH levels below 150 pg/ml and especially below 100 pg/ml are highly correlated with very low bone turnover and often adynamic bone disease [13–15]. The latter diagnosis requires the complete absence of tetracycline labels, thin osteoid (matrix), and few or absent bone cells (osteoblasts and osteoclasts) [16].

Phosphate retention has a direct negative effect on bone mineralization [17]. It stimulates PTH secretion and reduces the production of 1,25(OH)2D and phosphate reabsorption [18] in the kidney. Phosphate retention is the stimulus for the increased secretion of osteocyte-derived fibroblast growth factor 23 (FGF-23), a phosphaturic hormone [19] which also reduces the renal production of $1,25(OH)_2D$ and rises in the serum before PTH—the other phosphaturic hormone [20]. Serum FGF-23 is inversely correlated with bone formation. Excessive production or absence of FGF-23 with normal renal function leads to defects in mineralization with accumulation of osteoid [21–23].

The active metabolite of vitamin D [$1,25(OH)D_2$] in concert with PTH and FGF-23, controls calcium and phosphorus homeostasis, and its serum concentration is a mirror image of serum phosphorus. The kidney is the main site of $1,25(OH)_2D$ production with PTH being the major stimulator. As part of the feedback loop, $1,25(OH)_2D$ suppresses PTH production. The production of $1,25(OH)_2D$ in the kidney however is decreased by FGF-23, an effect which is augmented by increased degradation as well. The reciprocal action of $1,25(OH)_2D$ is a lower PTH, increased FGF-23 production in bone, and upregulation of klotho, a cofactor for FGF-23 signalling [24].

In the intestine, $1,25(OH)_2D$ increases calcium and phosphate absorption. In the kidney, it enhances the PTH action on calcium transport and induces the synthesis of TRPV5 and the calbindins. A major indirect effect of 1,25(OH)2D on bone is calcium and phosphate absorption from the intestine. It could also have direct effects on bone formation and resorption.

Chronic metabolic acidosis has profound effects on bone [25, 26]. The retained hydrogen ions (H^+) are buffered by bone calcium carbonate ($CaCO_3$) and, as the negative ion carbonate (CO_3^-) binds to H^+ ion, calcium is released. In addition, chronic metabolic acidosis is associated with impaired bone mineralization and osteoid accumulation.

Box 4.6.2 The factors that are believed to contribute to greater bone fragility in CKD

It is very likely that the cause is multifactorial

1. Sustained secondary hyperparathyroidism
2. Phosphorus retention and impairment in bone mineralization
3. Elevated Fibroblast Growth Factor 23 (FGF 23)
4. Low renal production of 1,25 dihydroxy-vitamin D
5. Chronic metabolic acidosis (either an anion gap or non-anion gap, e.g. renal tubular acidosis)
6. Greater risk for falling (multifactorial)

Renal Bone Diseases and Osteoporosis: Partners Intertwined

The National Institutes of Health (NIH) defines osteoporosis as: 'A systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue with a consequent increase in bone fragility and susceptibility to fracture' [27]. Since all forms of renal bone disease may not always have a low BMD despite the altered microachitecture, one could make the case that the traditional renal bone diseases defined by quantitative histomorphometry (osteomalacia, hyperparathyroid bone disease, adynamic renal bone disease, and mixed renal bone disease) (Figures 4.6.1, 4.6.2, and 4.6.3) also have a form of osteoporosis by NIH criteria [28]. Recently, Professor Sharon Moe has coined the phrase 'Renal Osteodystrophy or Kidney-Induced Osteoporosis?' that recognizes the combination of traditional renal bone diseases accompanying the NIH definition of osteoporosis [29]. While scientifically correct, the clinician must decide what specific metabolic bone disease the patient has in order to implement proper treatment strategies. Are the fractures due to osteoporosis or renal osteodystrophy (defined by histomorphometry)?

The term 'Chronic Kidney Disease–Mineral and Bone Disorder (CKD-MBD)' was introduced in 2005 to describe the broader clinical syndrome encompassing mineral, bone, and calcific cardiovascular abnormalities that develop as a complication of CKD [30]. CKD-MBD defines the interactions between the kidney, bone, and parathyroid glands that lead to hyperphosphatemia and its accompanying vascular disease. Since then, the term renal osteodystrophy is used exclusively to define the bone pathology associated with CKD. Since CKD-MBD is a multisystem disease complex, it has no ICD 10 code nor any specific treatment intervention other than attempting to mitigate hyperphosphatemia with oral phosphate binders, and the magnitude of severity of secondary hyperparathyroidism. Administration of monoclonal antibodies to FGF-23, only used experimentally so far, lead to more severe hyperphosphatemia [31].

The spectrum of renal osteodystrophy consists of the predominant hyperparathyroid-mediated high-turnover bone disease

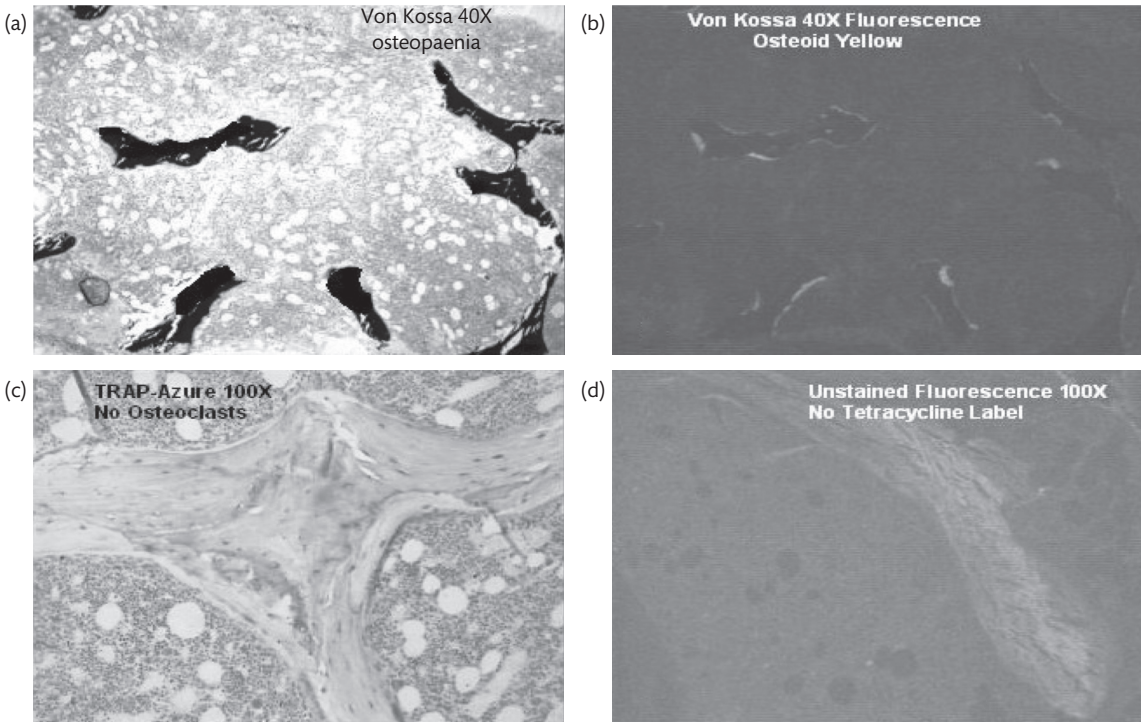


Figure 4.6.1 (a to d) Adynamic renal bone disease. *For a colour version of this figure, please see colour plate section.*

Reproduced with permission from Miller PD, Huffer W. Renal osteodystrophy. In: Schrier RW, Bennett W. *Essential Atlas of Nephrology and Hypertension*. Philadelphia: Current Medicine (2006);261–76.

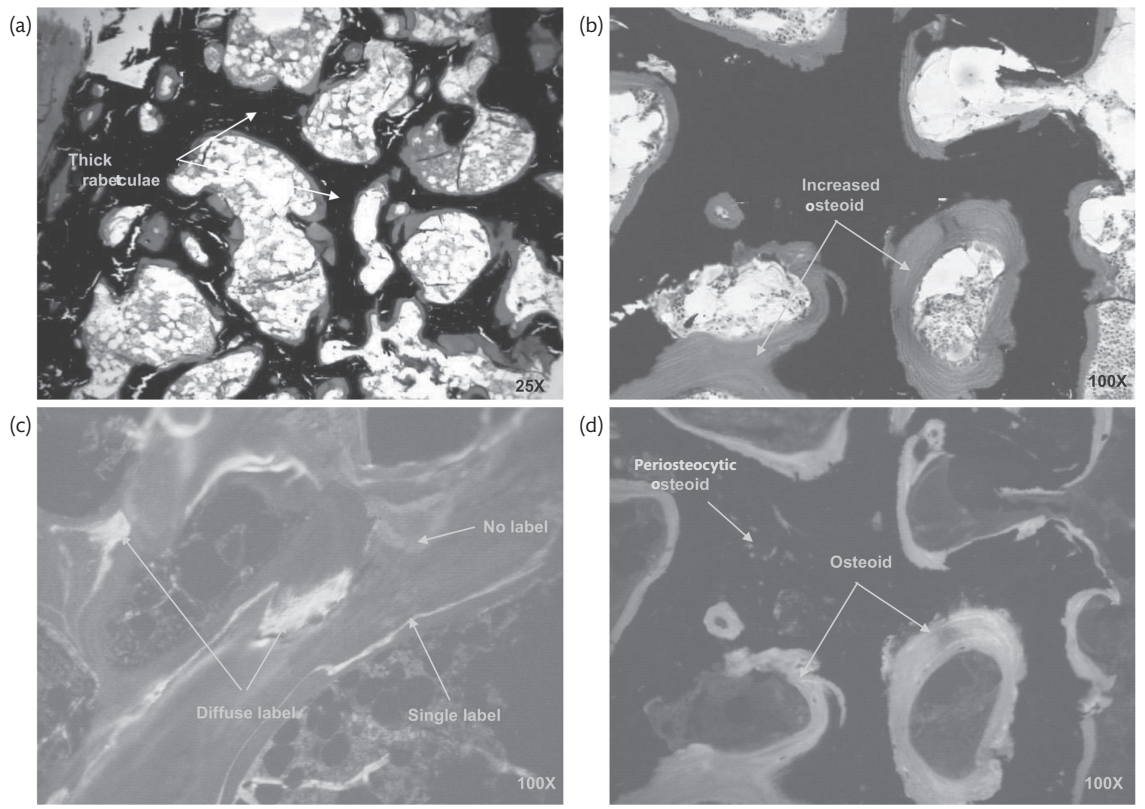


Figure 4.6.2 (a and b) Von Kossa, H&E stain for calcium and osteoid: osteomalacia. (c) Unstained, fluorescent for tetracycline. (d) Von Kossa, H&E stain, fluorescent for osteoid. *For a colour version of this figure, please see colour plate section.*

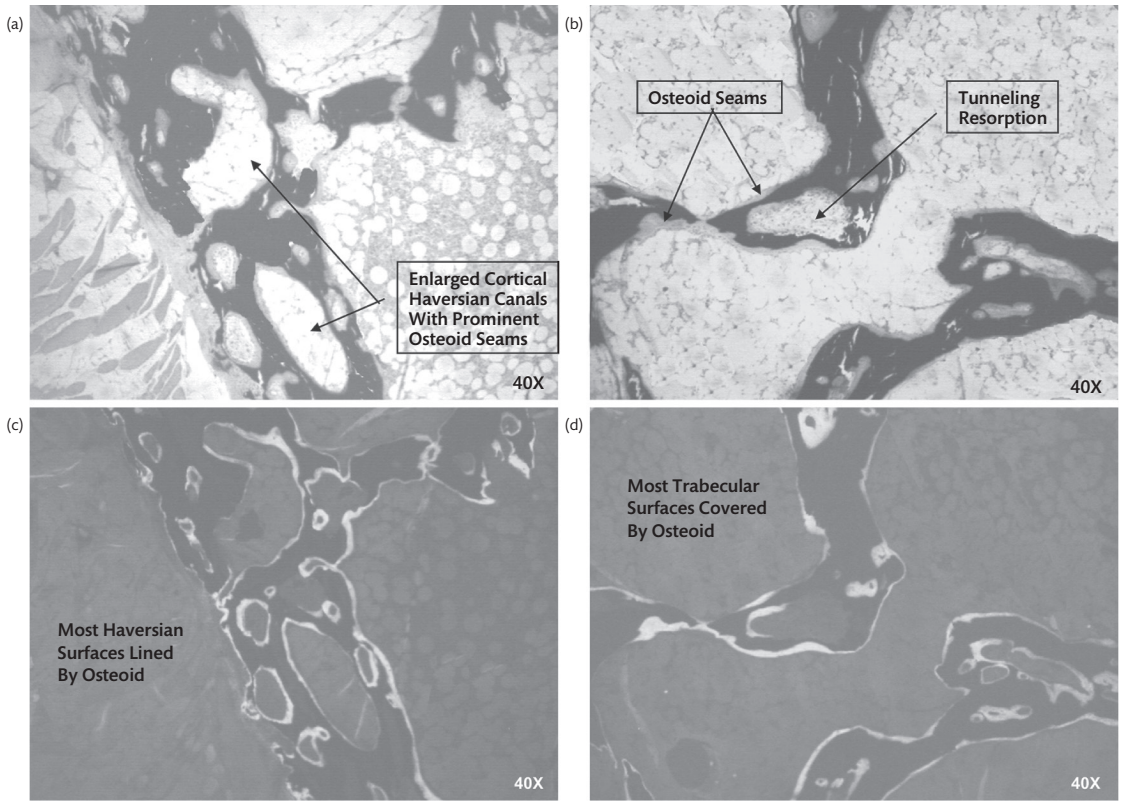


Figure 4.6.3 (a and b) Von Kossa, H&E stain for calcium: hyperparathyroidism. (c and d) Von Kossa, H&E stain, fluorescent for osteoid. *For a colour version of this figure, please see colour plate section.*

(osteitis fibrosa), osteomalacia (defined as a mineralization lag time greater than 100 days), low-turnover osteomalacia (defective mineralization in association with low osteoclast and osteoblast activities), mixed uremic osteodystrophy (hyperparathyroid bone disease with a superimposed mineralization defect), and the iatrogenic and idiopathic adynamic bone (diminished bone formation and resorption), but not osteoporosis [32, 33]. Two expert collaborative groups in the classification of renal bone diseases have suggested that renal osteodystrophy be divided into two models: low and high bone turnover [34, 35].

Idiopathic adynamic renal bone disease, more often seen in diabetic patients, could be due to an increased production of sclerostin, an inhibitor of osteoblast function [36]. Iatrogenic adynamic renal bone disease is due to oversuppression of parathyroid hormone production by therapies which inhibit parathyroid hormone synthesis (vitamin D analogues and/or Cinacalcet). Sustained lower levels of PTH below 150 pg/ml, and even more so below 100 pg/ml, have a high positive predictive value of predicting biopsy-proven adynamic bone disease [36, 37].

Discriminating Between Osteoporosis and Renal Bone Diseases

The clinician, confronted with the CKD patient who has low trauma fractures, and often a low T-score, has to decide what is the underlying metabolic bone disease in order to decide on the proper management strategy. It is also important to emphasize that both osteoporosis (defined by its highest order of criteria: fractures) as well as all forms of renal bone disease may have perfectly normal BMD (and T-score). This is because if one examines the equation for bone strength, 50% of bone strength is due to altered bone quality:

$$\text{Bone Strength} = \text{Bone Mineral Density} + \text{Bone Quality}.$$

Both osteoporosis as well as all forms of renal bone disease may have poor bone quality as a major contributor to bone fragility [38, 39]. There are no readily available office-based tools to measure bone quality. High resolution peripheral quantitative computerized tomography (HRpQCT) can provide resolution of the microstructure of bone down to levels of 70 microns. While predominantly a research tool, lower cost and smaller HRpQCT devices have recently been developed that may soon change the landscape of bone health into providing readily available office-based bone quality technologies [40].

Biochemical Markers of Bone Turnover and Quantitative Bone Histomorphometry: The Two Means of Making the Correct Diagnosis

How do we discriminate between osteoporosis and classical renal bone disease? Biochemical bone profile could be helpful but the 'gold standard' for making the correct diagnosis of the specific renal metabolic bone disease causing fragility fractures in CKD is double tetracycline labelled quantitative bone histomorphometry. The American Society for Bone and Mineral Research (ASBMR) has published a highly scientific white paper on defining both the static as well as dynamic parameters in quantitative histomorphometry [33]. Once a specific renal bone disease is excluded, then the firm

Box 4.6.3 The broad principles that may help in distinguishing between the renal bone diseases

PTH and BSAP combining the best of both worlds

- PTH "extremes" (< 100 pg/ml) or (> 600pg/ml) high specificity for adynamic/OFC.
- Bone specific alkaline phosphatase < 5 IU/L) has a high PPV (80%) for low bone turnover TINSAP <40 IU/L
- B .SAP correlate with PTH values in stage 5D CKD: both are increased on bone biopsy in established high bone turnover.
- Combining the lower quartile BSAP and PTH < 100-150 have a high PPV (90%) for adynamic bone disease.

Garrett G et al CJASN 2013
Couttenye C et al Kid internat 1999
Sprague SM et al Am J Kid Dis 2016

diagnosis for a CKD patient with either low BMD or fragility fractures, is osteoporosis.

Transiliac bone biopsies in experienced operator hands can be done very safely and painlessly. The procedure is well described in previous publications [41]. However, there are fewer and fewer skilled operators that are still active in the United States and even fewer places that perform the quantitative histomorphometry. Hence, great progress has been made in the utilization of specific biochemical markers of bone turnover (BTM), especially regarding two of them: bone-specific alkaline phosphatase and intact PTH, in helping to discriminate among the various forms of renal bone disease (Box 4.6.3).

The combination of intact PTH and either total tissue non-specific alkaline phosphatase or bone-specific alkaline phosphatase provide perhaps the best means of either predicting or excluding hyperparathyroid or adynamic renal bone diseases [42–44]. These two biochemical markers are only useful in patients not receiving any medications that alter PTH or alkaline phosphatase production (e.g. vitamin D analogues, cinacalcet, and antiresorptive or anabolic osteoporosis pharmacological therapies). The combination of a lower quartile alkaline phosphatase and intact PTH less than 150 pg/ml has a very high positive predictive value (PPV) for diagnosing adynamic bone disease. Two other 'pearls' that are helpful: an intact PTH 6× or higher than the upper limit of normal for the laboratory reference range (usually 65 pg/ml) is highly predictive of hyperparathyroid bone disease, while a high alkaline phosphatase excludes adynamic renal bone disease.

The question often asked by physicians trying to discriminate between osteoporosis and a renal bone disease is: what if the bone specific Alk Phos (BSAP) is in the mid-range of the normal reference range (neither low nor high) and the serum PTH is also neither low (<150 pg/ml) nor high (>350 pg/ml) in a stage 4–5 CKD patient that has fragility fractures, and a bone biopsy is not available? In these situations, the one renal bone disease we don't want to treat with an antiresorptive therapy is renal adynamic bone disease, since it seems illogical to reduce bone turnover in situations where bone turnover is low to begin with. Additionally, since the field of osteoporosis is moving towards using anabolic agents as first line therapy in high risk patients, based on data that initiating therapy with an anabolic agent followed by an antiresorptive agent is a better sequence than the opposite, we will use anabolic therapy first line [45, 46]. In a previous study of a single patient with biopsy-proven adynamic bone disease, the anabolic agent teriparatide was capable of reversing the adynamic bone disease by paired bone biopsy [47] (Figures 4.6.4 and 4.6.5).

Conclusions

Discriminating between osteoporosis and the various forms of renal bone disease can be done. The proper utilization of a menu of bone turnover markers can very often provide good diagnostic accuracy and discriminatory capabilities. While transiliac bone biopsy is the gold standard for diagnosis, it is often not readily available for many clinicians. The one renal bone disease where antiresorptive osteoporosis therapies would be potentially unsafe

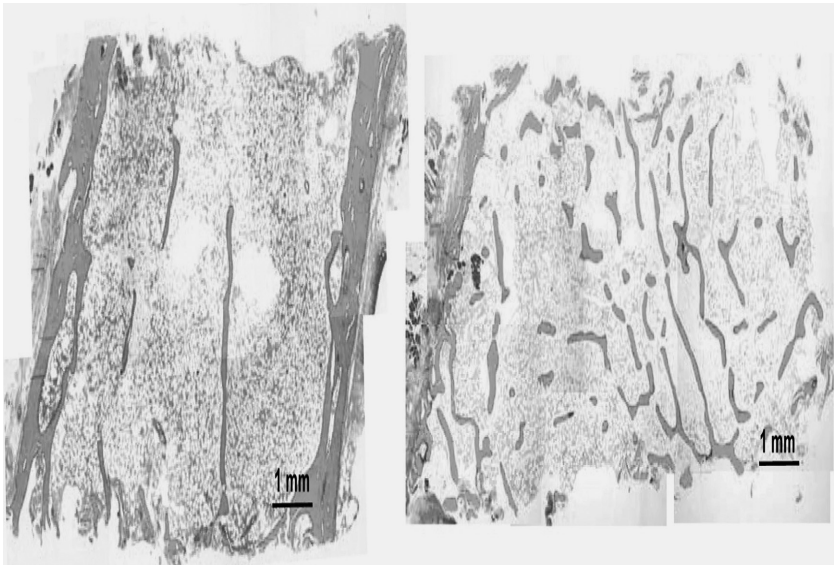


Figure 4.6.4 Effect of teriparatide on idiopathic renal (5 D) adynamic bone disease.

Reproduced with permission from Palcu P, Dion N, Ste-Marie LG, Goltzman D, Radziunas I, Miller P, *et al.* Teriparatide and Bone Turnover and Formation in a Hemodialysis Patient with Low-Turnover Bone Disease: A Case Report. *American Journal of Kidney Diseases* 65 (6): 933–6. 2015. Copyright © 2015 National Kidney Foundation, Inc.

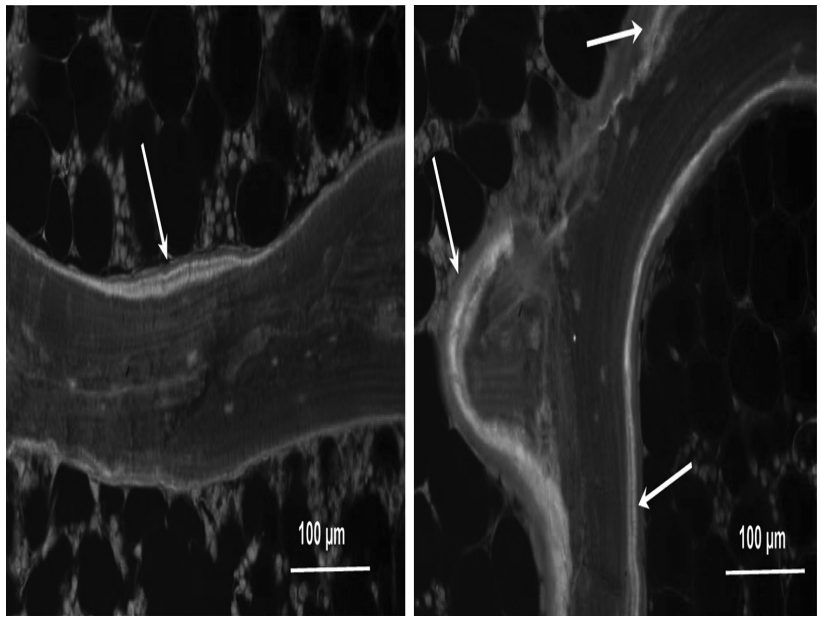


Figure 4.6.5 Effect of teriparatide to improve mineralization in stage 5D CKD.

Reproduced with permission from Palcu P, Dion N, Ste-Marie LG, Goltzman D, Radziunas I, Miller P, *et al.* Teriparatide and Bone Turnover and Formation in a Hemodialysis Patient with Low-Turnover Bone Disease: A Case Report. *American Journal of Kidney Diseases* 65 (6): 933–6. 2015. Copyright © 2015 National Kidney Foundation, Inc.

is idiopathic renal adynamic bone disease. In those patients who have fragility fractures, stage 4–5 CKD, and the biochemical profile is not as discriminatory as it can be, the initial use of an anabolic agent would be safe and certainly therapeutically beneficial in both osteoporosis as well as idiopathic renal adynamic bone disease. The two renal bone diseases where an osteoporosis pharmacological agent would not be advised are osteomalacia and hyperparathyroid bone disease. The latter two metabolic bone diseases can be excluded by defining the underlying cause of a high bone-specific alkaline phosphatase or defining the cause of an intact PTH that is also very high.

If a stage 4–5 CKD patient with fragility fractures is felt to have osteoporosis as the major underlying metabolic bone disease causing fractures, FDA approved pharmacological agents for the treatment of osteoporosis can be beneficial on or off label [48].

ACKNOWLEDGEMENTS

We would like to thank Dr I. Stanciu for reading the paper and helping with the administrative work.

REFERENCES

1. Ketteler M, Block GA, Evenepoel P, *et al.* Diagnosis, evaluation, prevention, and treatment of chronic kidney disease–mineral and bone disorder: synopsis of the kidney disease: improving Global Outcomes 2017 Clinical Practice Guideline Update. *Ann Intern Med*, 2018; **168**: 422–30.
2. Rowland J, Akbarov A, Maan A, Eales J, Dormer J, Tomaszewski M. Tick-tock chimes the kidney clock—from biology of renal ageing to clinical applications. *Kidney Blood Press Res*, 2018; **43**: 55–67.
3. Greenwood C, Clement J, Dicken A, *et al.* Age-related changes in femoral head trabecular microarchitecture. *Aging Dis*, 2018; **9**: 976–87.
4. Harvey NC, Odén A, Orwoll E, *et al.* Measures of physical performance and muscle strength as predictors of fracture risk independent of FRAX, falls, and aBMD: a meta-analysis of the Osteoporotic Fractures in Men (MrOS) study. *J Bone Miner Res*, 2018; **33**: 2150–7.
5. Damasiewicz MJ, Nickolas TL. Rethinking bone disease in kidney disease. *JBMR Plus*, 2018; **2**: 309–22.
6. Denic A, Glassock RJ, Rule AD. A structural and functional changes with the aging kidney. *Adv Chronic Kidney Dis*, 2016; **23**: 19–28.
7. HommOs MS, Glassock RJ, Rule AD. Structural and functional changes in human kidneys with healthy aging. *J Am Soc Nephrol*, 2017; **28**: 2838–44.
8. Whitlock RH, Leslie WD, Shaw J, *et al.* The Fracture Risk Assessment Tool (FRAX®) predicts fracture risk in patients with chronic kidney disease. *Kidney Int*, 2019; **95**: 447–54.
9. Kistler BM, Khubchandani J, Jakubowicz G, Wilund K, Sosnoff J. Falls and fall-related injuries among us adults aged 65 or older with chronic kidney disease. *Prev Chronic Dis*, 2018; **15**: 170518.
10. Papakonstantinopoulou K, Sofianos I. Risk of falls in chronic kidney disease. *JFSF*, 2017; **6**: 33–8.
11. Graciolli FG, Neves KR, Barreto F, *et al.* The complexity of chronic kidney disease–mineral and bone disorder across stages of chronic kidney disease. *Kidney Int*, 2017; **91**: 1436–46.
12. Ureña-Torres PA, Vervloet M, Mazzaferro S, *et al.*; ERA-EDTA CKD-MBD Working Group. Novel insights into parathyroid hormone: report of The Parathyroid Day in Chronic Kidney Disease. *Clin Kidney J*, 2018; **12**: 269–80.
13. Massy Z, Drüeke T. Adynamic bone disease is a predominant bone pattern in early stages of chronic kidney disease. *J Nephrol*, 2017; **30**: 629–34.

14. Bover J, Ureña P, Brandenburg V, *et al*. Adynamic bone disease: from bone to vessels in chronic kidney disease. *Semin Nephrol*, 2014; **34**: 626–40.
15. Garrett G, Sardiwal S, Lamb EJ, Goldsmith DJ. PTH—a particularly tricky hormone: why measure it at all in kidney patients? *Clin J Am Soc Nephrol*, 2013; **8**: 299–312.
16. Parfitt AM. Renal bone disease: a new conceptual framework for the interpretation of bone histomorphometry. *Curr Opin Nephrol Hypertens*, 2003; **12**: 387–403.
17. Endo A, Glimcher MJ. The effect of complexing phosphoproteins to decalcified collagen on *in vitro* calcification. *Connect Tissue Res*, 1989; **21**: 179–90.
18. Hernando N, Wagner CA. Mechanisms and regulation of intestinal phosphate absorption. *Compr Physiol*, 2018; **8**: 1065–90.
19. Dhillon-Jhattu S, Sprague SM. Should phosphate management be limited to the KDIGO/KDOQI guidelines? *Semin Dial*, 2018; **31**: 377–81.
20. Stubbs J, Liu L S, Quarles D. Role of fibroblast growth factor 23 in phosphate homeostasis and pathogenesis of disordered mineral metabolism in chronic kidney disease. *Semin Dialysis*, 2007; **20**: 302–8.
21. Shimada T, Urakawa I, Yamazaki Y, *et al*. FGF-23 transgenic mice demonstrate hypophosphatemic rickets with reduced expression of sodium phosphate cotransporter type IIa. *Biochem Biophys Res Commun*, 2004; **314**: 409–14.
22. Shimada T, Kakitani M, Yamazaki Y, *et al*. Targeted ablation of FGF 23 demonstrates an essential physiological role of FGF23 in phosphate and vitamin D metabolism. *J Clin Invest*, 2004; **113**: 561–8.
23. Erben RG. Physiological actions of fibroblast growth factor-23. *Front Endocrinol (Lausanne)*, 2018; **9**: 267–75.
24. Bikle DS, Adams J, Christakos S. Vitamin D: production, metabolism, mechanism of action, and clinical requirements. In: Rosen CJ, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, 8th edn. Danvers, MA: Wiley-Blackwell, 2013: 235–47.
25. Raphael KL. Metabolic acidosis and subclinical metabolic acidosis in CKD. *Am Soc Nephrol*, 2018; **29**: 376–82.
26. Raphael KL. Metabolic acidosis in CKD: core curriculum 2019. *Am J Kidney Dis*, 2019: **74**: 263–75.
27. NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy. *JAMA*, 2001; **285**: 785–95.
28. Miller PD. Chronic kidney disease and osteoporosis: evaluation and management. *Bonekey Rep*, 2014; **3**: 542–54.
29. Moe S. Renal osteodystrophy or kidney-induced osteoporosis? *Curr Osteoporos Rep*, 2017; **15**: 194–7.
30. Moe S, Drüeke T, Cunningham J, *et al*. Definition, evaluation, and classification of renal osteodystrophy: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int*, 2006; **69**: 1945–53.
31. Gutierrez O, Isakova T, Rhee E, *et al*. Fibroblast growth factor-23 mitigates hyperphosphatemia but accentuates calcitriol deficiency in chronic kidney disease. *J Am Soc Nephrol*, 2005; **16**: 2205–15.
32. Quarles LD. Bone biopsy and the diagnosis of renal osteodystrophy. Sterns HR, Section ed. UpToDate, 2017. Available at: https://www.uptodate.com/contents/bone-biopsy-and-the-diagnosis-of-renal-osteodystrophy (accessed: 24 February 2020).
33. Dempster DW, Compston JE, Drezner MK, *et al*. Standardized nomenclature, symbols, and units for bone histomorphometry: a 2012 update of the report of the ASBMR Histomorphometry Nomenclature Committee. *J Bone Miner Res*, 2013; **28**: 2–17.
34. Salam S, Gallagher O, Gossiel F, Paggiosi M, Khwaja A, Eastell R. Diagnostic accuracy of biomarkers and imaging for bone turnover in renal osteodystrophy. *J Am Soc Nephrol*, 2018; **29**: 1557–65.
35. Sprague SM, Bellorin-Font E, Jorgetti V, *et al*. Diagnostic accuracy of bone turnover markers and bone histology in patients with CKD treated by dialysis. *Am J Kidney Dis*, 2016; **67**: 559–66.
36. Massy Z, Drüeke T. Adynamic bone disease is a predominant bone pattern in early stages of chronic kidney disease. *J Nephrol*, 2017; **30**: 629–34.
37. Sista SK, Arum SM. Management of adynamic bone disease in chronic kidney disease: a brief review. *J Clin Transl Endocrinol*, 2016; **5**: 32–5.
38. Seeman E, Delmas PD. Bone quality—the material and structural basis of bone strength and fragility. *N Engl J Med*, 2006; **354**: 2250–61.
39. McNerny EMB, Nickolas TL. Bone quality in chronic kidney disease: definitions and diagnostics. *Curr Osteoporos Rep*, 2017; **15**: 207–13.
40. West SL, Lok CE, Langsetmo L, *et al*. Bone mineral density predicts fractures in chronic kidney disease. *J Bone Miner Res*, 2015; **30**: 913–19.
41. Miller PD. The role of bone biopsy in patients with chronic renal failure. *Clin J Amer Soc Nephrol*, 2008; **3**: S140–50.
42. Garrett G, Sardiwal S, Lamb EJ, Goldsmith DJ. PTH—a particularly tricky hormone: why measure it at all in kidney patients? *Clin J Am Soc Nephrol*, 2013; **8**: 299–12.
43. Couttenye MM, D'Haese PC, Verschoren WJ, Behets GJ, Schrooten I, De Broe ME. Low bone turnover in patients with renal failure. *Kidney Int Suppl*, 1999; **73**: S70–6.
44. Sprague SM, Bellorin-Font E, Jorgetti V, *et al*. Diagnostic accuracy of bone turnover markers and bone histology in patients with CKD treated by dialysis. *Am J Kid Dis*, 2016; **67**: 559–66.
45. Leder BZ. Optimizing sequential and combined anabolic and antiresorptive osteoporosis therapy. *JBMR Plus*, 2018; **2**: 62–8.
46. Cosman F, Nieves JW, Dempster DW. Treatment sequence matters: anabolic and antiresorptive therapy for osteoporosis. *J Bone Miner Res*, 2017; **32**: 198–202.
47. Palcu P, Dion N, Ste-Marie LG, *et al*. Teriparatide and bone turnover and formation in a hemodialysis patient with low-turnover bone disease: a case report. *Am J Kid Dis*, 2015; **65**: 933–6.
48. Lewiecki EM, Bilezikian JP, Giangregorio L, *et al*. Proceedings of the 2018 Santa Fe Bone Symposium: advances in the management of osteoporosis. *J Clin Densitom*, 2019; **22**: 1–19.

4.7

Hypercalcaemic and Hypocalcaemic Syndromes in Children

Laleh Ardeshirpour, Thomas O. Carpenter, and Cemre Robinson

Introduction

The calcium regulating system employs an intricate network of homeostatic signals and targets in order to meet the body's mineral demands. Mineral requirements vary considerably throughout progressive stages of development, in large part reflecting the mineral demands of *in utero* mineralization of the fetus and later skeletal growth. Features of the calcium homeostatic system during childhood years are in flux as relative rates of growth and development change throughout the life cycle. Furthermore, numerous compensatory mechanisms limit the severity of disease when an isolated insult occurs to the system. This chapter illustrates that many heritable disorders of mineral homeostasis become evident in infancy and childhood and are best recognized when viewed within the appropriate context of mineral requirements during the early stages of life. An understanding of the relevant physiology is central to formulating age-specific approaches to management of these various clinical problems, and so the disorders are reviewed in the context of physiology specific to childhood.

Features of Calcium Homeostasis Specific to Children

Perinatal Calcium Metabolism

Skeletal Development and Mineral Requirements of the Fetus

The growing fetus must be supplied with sufficient calcium for the formation and growth of a mineralizing skeleton. In addition, the physiologic milieu of the fetus must be maintained in an environment appropriate for normal cellular function. Thus, adequate extracellular calcium must be provided for normal function of the clotting factors, and avoidance of neuromuscular hyperexcitation. Yet at the same time, the supply must be appropriately limited to prevent soft tissue calcification or other toxicity to the developing fetus. A critical calcium dependent process in fetal life is skeletal accrual of mineral. Most of the skeleton is formed by the complex process referred to as endochondral ossification [1]. Cartilage templates are organized in concert with the transition of undifferentiated mesenchymal cells to differentiated chondrocytes. The cartilage templates serve as a nidus for eventual development into the skeleton. A system of chondrocyte maturation and proliferation occurs at what will become the ends of long bones, allowing for the continued linear growth of the skeleton. Regulation of this early formative process is dependent upon a variety of local and systemic factors, such as insulin-like growth factors (IGFs), fibroblast growth factors (FGFs), parathyroid hormone-related peptide (PTHrP), and Indian hedgehog protein [2]. Once mature cartilage forms, chondrocytes hypertrophy, and blood vessels penetrate the region, with the appearance of marrow stroma and osteoblasts soon to follow. Mineralization of the newly established skeleton begins, and growth results in an increasing mineral demand in order to effectively mineralize the newly formed tissue. Indeed, the fetus has substantial mineral demands: approximately 26–30 grams of calcium accumulate in the human through a term gestation, and accretion of more than three-fourths of this amount occurs in the third trimester. Calcium supply from the maternal circulation must be regulated by specific mechanisms in order to meet these demands throughout the later weeks of gestation.

The Fetal Calcium Regulating System

The maternal circulation is the source of calcium provided to the fetus. An abundance of calcium occurs in the mother primarily as a result of a pregnancy-induced doubling of maternal circulating 1,25-dihydroxyvitamin D [1,25 $(OH)_2D$] levels, which in turn increases fractional absorption of calcium at the intestine [3]. This occurs in the absence of significant increases in circulating levels of parathyroid hormone (PTH) in the mother.

The placenta is the site of nutrient transfer from the maternal circulation to the fetus. Calcium may be transported by several mechanisms across the placenta; the dominant direction of flow is from maternal to the fetal circulation, requiring active transport. A Ca^{++}-ATPase located in the fetus-directed basement membrane of the syncytial trophoblast cells appears to mediate this important function [4]. Although it is not clear exactly when in gestation active calcium transport begins, it is present by the beginning of the third trimester. The fetal circulating calcium level is maintained at a slightly higher concentration (1.2–2.0 mg/dl) than the maternal circulation as early as 15 weeks of gestation [5]. Active placental calcium transfer plays an important role in determining fetal circulating calcium level, but other factors may play a role, including PTH and PTHrP. The relative hypercalcaemia in the fetus appears to be necessary for complete skeletal growth and mineralization [6, 7].

Placental calcium transfer seems to be mainly regulated by PTHrP, and to a lesser extent by PTH [4, 8]. The mid- and carboxy-terminal portions of the PTHrP molecule are required, whereas the amino-terminus (most related to PTH in sequence) does not have activity in this regard [4, 9]. PTHrP plays an important role in embryonic growth and development of many tissues, and is produced by multiple tissues. Major sources of PTHrP production are the placenta, and to a lesser extent, the parathyroid glands. In a fetal mice model, disruption of PTHrP results in hypocalcaemia and severe chondrodysplasia [8, 10]. Fetal circulating PTH levels are low, probably due to Ca-sensing receptor (CaSR) mediated suppression of PTH secretion by fetal parathyroid glands [4]. Nevertheless, aparathyroid fetal mice develop hypocalcaemia and defective bone mineralization, pointing towards a role for PTH in maintaining normal serum calcium, and thereby perhaps supporting normal bone mineralization [11]. PTH may also exert its effect on bone formation, to some extent, via direct interaction with osteoblasts.

Fetal circulating 1,25 $(OH)_2D$ levels are low, and derived from placenta and fetal kidney, and may be related to the concurrent low PTH and high serum phosphorus levels. 1,25 $(OH)_2D$ does not play a major role in placental calcium transfer or maintenance of serum calcium level as evidenced by the fully mineralized skeleton and normal fetal serum calcium levels at term in vitamin D receptor-null (*vdr*-null) mice. In human cases of maternal vitamin D deficiency, skeletal mineralization seems to be unaffected but the newborn will be at risk of developing hypocalcaemia [4]. The presence of CaSR in both human and murine placenta suggests a possible role for this membrane receptor in fetal calcium homeostasis. Some insight has been provided by CaSR null mice: fetuses of this strain demonstrate increased PTH levels, reduced placental calcium transport, increased amniotic fluid calcium, and increased markers of bone resorption. This constellation of findings suggests that an increase in skeletal resorption can occur, when inadequate calcium levels are sensed by the fetus [11, 12]. The fetal kidneys may be involved in regulation of fetal calcium levels as well, but is of limited impact in that fetal urine is excreted into the amniotic fluid and recycled with swallowing.

Limited data regarding FGF23 levels in the human fetus suggest that intact FGF23 levels are lower than levels in adults [13] but that C-terminal levels are approximately twofold greater than in adults [14]. Circulating levels of alpha-klotho are reported as elevated in the fetus, likely reflecting a placental source [13]. Other relevant measures in the fetal circulation include relatively increased levels of phosphorus, magnesium, and relatively low concentrations of calcitonin compared to postnatal life [7].

Transition from Fetal Life to Infancy

At birth the generous supply of maternal calcium is abruptly withdrawn from the fetus, as well as the placental sources of PTHrP and 1,25 $(OH)_2D$, with a resultant acute decrease in serum calcium of approximately 1 mg/dl in term infants, and slightly more in preterm infants. One study indicates that the decrement in serum calcium and rise in PTH is greater in babies born by Caesarean section than in babies born spontaneously by the vaginal route. This decrease in calcium then stimulates secretion of PTH, suppressed during fetal life, which in turn, stimulates the kidney to generate adult normal levels of 1,25 $(OH)_2D$ within the next several days. Levels of PTHrP are reduced; this hormone likely plays a lesser role in postnatal calcium homeostasis than *in utero*. The serum calcium gradually increases to normal childhood levels within a few days of the acute postnatal decrement. The relative hypercalcaemia of the fetus likely protects the newborn infant from acute severe hypocalcaemia during the transition to the ex utero environment.

The intestine and kidney assume major roles in mineral homeostasis with this transition. The neonatal skeleton continues to accrue calcium at rates close to that attained in late gestation (averaging 100–150 mg/kg/day). Thus, the newborn infant becomes dependent upon exogenous nutritional sources of calcium. Renal excretion of calcium increases over the first few weeks of life, as glomerular filtration rate (GFR) increases. As the kidney matures, it begins to play a minor role in regulation of calcium. The newborn infant, however, becomes primarily dependent upon the *intestine* to maintain its calcium supply. In the first few days to weeks of neonatal life, passive or facilitated calcium transport (not vitamin D-mediated mechanisms) are the dominant means by which calcium is brought into the body. After several weeks, vitamin D appears to be useful in enhancing calcium absorption in term infants. Fractional calcium absorption can be relatively high in infancy particularly in very low birthweight children, who may develop hypercalcaemia during high calcium intake, as may occur with the administration of breast-milk fortifiers. This phenomenon may occur independently of vitamin D status (with normal circulating levels of 25-hydroxyvitamin D (25-OHD), and appropriately low circulating PTH and 1,25 $(OH)_2D$), implying that passive or facilitated, non-vitamin D-mediated calcium transport in the immature intestine can be remarkably efficient. The relatively low phosphate content of breast milk likely contributes to the greater bioavailability of calcium from breast milk as compared to most infant formulas. Formation of calcium-phosphate complexes in the higher phosphate containing formulas limit the availability of calcium for intestinal absorption.

Childhood Growth: A Period of Intensive Mineral Accretion

Growth and Accrual of Bone Density

Skeletal growth and mineralization continue at a very rapid pace throughout the first 2 years of life. The growth velocity on average during the first 4 months of life can be annualized to approximately 28 cm/yr. From that time on a child's growth rate asymptotically

decreases from a rate of one cm per month (approximately 12.5 cm/yr) to about 5–6 cm/yr at the time of the pubertal growth spurt, when a rate of about 10 cm/yr is transiently achieved prior to the cessation of growth. This linear growth represents the growth of the appendicular skeleton, which must be adequately mineralized; thus rapid growth in infancy places considerable mineral demands on the skeleton. Bone mineral content (BMC) and areal bone mineral density (BMD), as assessed by standard two-dimensional techniques, such as dual energy X-ray absorptiometry (DXA) proceeds at a steady pace until approximately age 11 in girls and slightly later in boys [15]. Specific guidelines for the use and interpretation of bone densitometry in children have been previously published and updated by the International Society of Clinical Densitometry [16, 17]. Bone size is a confounding variable in interpretation of DXA results, which is a particularly important issue in growing children. The BMD is an areal measurement, rather than a true volumetric measurement, since depth cannot be measured directly. Thus, DXA has been shown to underestimate the BMD in those children with smaller bone size and overestimate it in those with larger bone size. To resolve this problem, standard paediatric reference data for BMD have been established, taking into account the age, sex, ethnicity (African American or non-African American) and height of the patient, with the Bone Mineral Density in Childhood Study (BMDCS); a multicentre, longitudinal study of bone accrual in healthy children and adolescents performed at five US clinical centres [18].

Although the focus on bone activity during these years of rapid growth is primarily on formation and mineralization, there must also be extremely active turnover in general. The growing bone must be constantly modelled in order to maintain an appropriate structure. Weight-bearing forces begin to correct the physiologic bowing of childhood, as lower extremity alignment becomes more linear. As metaphyseal long bone segments accrue mineral at growth plate cartilage, extending the length of the long bone calls for an eventual narrowing of the former metaphyseal segment as it assumes a diaphyseal position. These processes require extensive bone resorptive activity. Thus, when investigating disorders of the bone and mineral system, one must recognize this relatively hyperdynamic state of bone turnover. The established normal ranges of bone activity for adults, which primarily represent bone remodelling, do not apply to children. In children and adolescents, serum markers of bone formation and resorption are released into the circulation during linear growth, bone modelling, and remodelling [19, 20]. The normal ranges of values for such markers as serum osteocalcin, alkaline phosphatase activity, or urinary excretion of deoxypyridinoline cross-links of collagen, or the N-telopeptide of type I collagen are quite wide (Table 4.7.1). Several investigators have compiled normative data on these and other biomarkers of bone turnover throughout childhood and/or adolescent age groups [3, 21] and reference curves for five bone markers in the serum; osteocalcin (OC), bone specific alkaline phosphatase (BALP), the carboxy-terminal telopeptide region of type I collagen (ICTP), the carboxy-terminal α1 chain telopeptide of type I collagen (CTX), and serum tartrate-resistant acid phosphatase 5b (TRAP5b) have been established in a large cohort of healthy children aged 2 months to 18 years, along with calculation of sex- and age-specific standard deviation scores [20, 21]. There is a consistent peak in the concentrations of most serum markers of bone formation and resorption in adolescence that decreases in adulthood, and this rise occurs approximately 2.5 years earlier in girls than in boys [21]. Circulating levels of P1NP (procollagen type I N-propeptide) has emerged as a frequently used marker of bone formation [21], and circulating CTX is often used as a marker of bone resorptive activity, and both of these markers are found to be greater in childhood than in adulthood (Table 4.7.1). Indeed, all bone markers tend to show significant variation with age whereas mainly BALP shows a significant variation with sex, with higher concentrations observed in boys [20]. Values in adolescence for the more widely used markers are shown in Table 4.7.1. It is recommended to use a set of bone turnover markers, rather than a single marker in evaluating those children with metabolic bone disorders, monitoring for disease progression or during antiresorptive therapy [20].

Puberty

In addition to the rapid growth spurt beginning in early puberty in girls, and later stages of puberty in boys, BMD accrues at an accelerated pace. The rate of increase in BMD in girls between the ages of 11 and 16 years is more rapid than at any time in late childhood or during adult life. The National Academy of Sciences (USA) has set 'adequate intake' levels for calcium by age ranges. These levels are 210 mg daily through the first 6 months of life, 270 mg/day for months 6 to 12, 500 mg/day from years 1 to 3, and 800 mg/day for ages 4 to 8. In keeping with the rapid rate of bone accretion in adolescence, calcium 'adequate intake' has been set at 1300 mg/day for the 9–18-year-old group [22]. Some have thought this number underestimates calcium requirements and have suggested that teenage girls consume 1500 mg of calcium daily.

Commensurate with the pubertal growth spurt are transient rises in the markers of bone formative activity, serum osteocalcin, and alkaline phosphatase activity. Markers of bone resorption, which decrease somewhat throughout later childhood, decrease substantially in late puberty (Tanner stages IV and V), reflecting more quiescent bone turnover than in earlier childhood, as described here and in Table 4.7.1. The postpubertal period of elevation in turnover markers persists in males longer than in females, suggesting a longer period of active mineral accrual in young men than in young women.

In addition to the changes in bone markers, geometric properties of long bones change during puberty, and appear to differ between boys and girls. These changes may be reflected in the differential changes in biomarker levels described earlier. However, the finding of wider long bones of males remains largely unexplained. Male long bones progressively grow in circumferential diameter beyond female growth in this regard, in part due to a prolonged period of generalized prepubertal growth [23]. Furthermore, recent data suggests that such sex differences in geometry are evident in prepubertal years, determined by complex genetic traits and environmental stimuli [23].

Age-Dependent Changes in Serum Minerals and Calciotropic Hormones

Appropriate diagnosis of disease or monitoring of therapy requires an understanding of changes in the biochemical parameters used to facilitate an evaluation of mineral metabolism in children.

Table 4.7.1 Normal values of biochemical markers of bone turnover in childhood

	Male	Female
Formative markers		
Serum osteocalcin (ng/ml)[1]		
<10 yrs:	6–35	6–40
10–18 yrs:	9–84	7–50
Serum alkaline phosphatase activity (IU/L)[2]		
0–14 days:	83–248	83–248
15 days <1 yr:	122–469	122–469
1–13 yrs:	135–376	135–376
13–15 yrs:	57–254	116–468
15–17 yrs:	50–117	82–331
17–19 yrs:	45–87	55–149
Serum PINP (N-terminal propeptide of type 1 procollagen) (ng/ml)[3]		
6–14 yrs:	250–1800	*6–12 yrs:* 250–1500
14–18 yrs:	80–1500	*12–16 yrs:* 80–1000
18–26 yrs:	40–300	*16–26 yrs:* 20–200
Resorptive markers		
Serum N-Tx (cross-linked N-terminal-telopeptide of type I collagen) (pmol/ml)[4]		
6–14 yrs:	20–200	*6–12 yrs:* 25–120
14–16 yrs:	14–180	*12–18 yrs:* 8–12
16–26 yrs:	7–50	*18–26 yrs:* 5–25
Urinary N-Tx (pmol equivalent of bone collagen/µmol creatinine)[5]		
<1 yr:	500–5000	870–5700
1 yr:	120–2800	475–2750
2–4 yr:	320–2100	155–2010
5–10 yr:	110–1275	115–1620
11–12 yr:	210–2600	235–2430
13–14 yr:	105–1900	45–1335
15–18 yr:	34–1146	45–400
Serum C-Tx (Cross-linked C-terminal-telopeptide of type I collagen) (pg/ml)[6, 7, 8]		
10–17 yr:	370–2560	240–1530
30–34 yr:	260–450	140–330

[1] Extrapolated from Figures 5 and 8, Calvo, M *et al.* [19]. Note that values will vary with respect to the assay employed and to the laboratory performing the test.
[2] Values rounded from table 3, Estey *et al.* [141].
[3] Extrapolated from Figure 2, van der Sluis *et al.* [21].
[4] Extrapolated from Figure 3, van der Sluis *et al.* [21].
[5] Values rounded from Bollen and Eyre [142].
[6] Extrapolated from Figure 3, Fares *et al.* [143].
[7] Reference [144].
[8] Adult reference from Table 4.7.2 [145].

Note the consistent peak in the concentrations of most serum markers of bone formation and resorption in adolescence that decreases in adulthood.

Figure 4.7.1 illustrates the changes in circulating minerals and related hormones during early infancy. The serum levels of calcium and magnesium do not change significantly after the first few days of life until adulthood. On the other hand, urinary excretion of calcium is much greater in infancy than in later childhood and adulthood. A convenient measure for urinary excretion of calcium is the ratio of calcium to creatinine (Ca/Cr) in a random urine sample. Urinary calcium excretion varies with type of feedings, vitamin D nutrition, and gestational age [24]. In the older child a fasting urine sample should have a Ca/Cr less than 0.21 mg/mg. For infants and younger children, age-specific normal values for spot urine samples have been reported and can range from 0.39 mg/mg for children between ages 1 and 5 years old to 0.81 mg/mg for those younger than 12 months of age [25]. A 24-hr urine collection should be confirmed

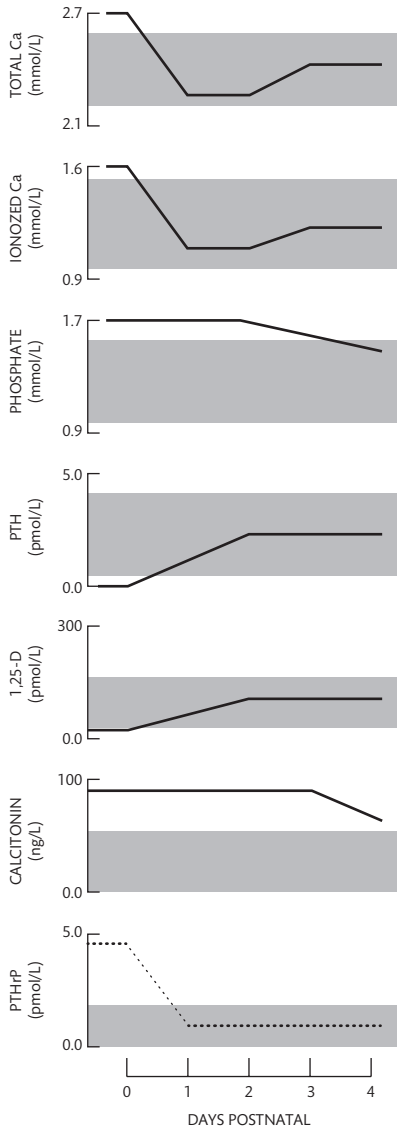


Figure 4.7.1 Longitudinal change in circulating concentrations of minerals, PTH, calcitonin, and 1,25 (OH)2 D during fetal and neonatal period in humans. Shaded areas represent the adult normal range for the parameter.

Reproduced with permission from Kovacs, C.S., Bone development and mineral homeostasis in the fetus and neonate: roles of the calciotropic and phosphotropic hormones. *Physiol Rev*, 2014. 94(4): p. 1143–218. Copyright © 2014 The American Geriatrics Society [ref 146].

by measurement of total creatinine (which should be 10–20 mg/kg/24 hr in most children), and the calcium should be less than 4 mg/kg/24hr. Circulating phosphate concentrations decrease considerably throughout the first year of life, and even further throughout later childhood. The normal ranges are substantially greater than that seen in older adults. This change is primarily due to increased reclamation of filtered phosphate in the proximal renal tubule early in life. The confusion in interpretation of age-related normal ranges has continued to result in missed diagnoses and inappropriate interpretation of mineral status. The assessment of urinary phosphate excretion should be performed on a 2-hour fasting urine specimen, with a blood sample obtained at the midpoint of the urine collection. The tubular reabsorption of phosphate (%TRP) is calculated as:

$$\%\mathbf{TRP} = \mathbf{1} - \frac{[U_P \times P_{Cr}]}{[P_P \times U_{Cr}]} \times 100,$$

where U_P refers to the urinary concentration of phosphate, P_P to plasma concentration of phosphate, U_{Cr} to urinary concentration of creatinine, and P_{Cr} to the plasma concentration of creatinine.

The %TRP can be plotted on the nomogram of Walton and Bijvoet [26] to obtain the TMP/GFR, or tubular maximum for phosphate expressed per GFR. This value reflects the value of serum P above which one will tend to stop reclaiming P in the tubule. The normal ranges vary with age and approximate the normal phosphate concentrations for age. Alternatively, a calculated method termed TP/GFR, can be used employing the formula:

$$\text{TP/GFR} = \text{Serum phosphate} - (\text{Urine phosphate} \times \text{Serum creatinine / Urine creatinine}). \quad [27]$$

Finally, values for circulating PTH do not change after early infancy throughout childhood. Circulating 25-OHD levels are relatively stable in early childhood (Figure 4.7.2), but are influenced by a variety of factors including ethnicity and season of sampling. Levels tend to be greater in children of European descent, with formula feeding, and during the summer and fall months [28]. Circulating 1,25 $(OH)_2D$ levels tend to be higher in childhood than in later life, and decline with time in early childhood, particularly during the first 3 years of life (Figure 4.7.2). Furthermore, there appears to be less stringent regulation of conversion of 25-OHD to 1,25 $(OH)_2D$ in early life. The authors have generally observed values for 1,25 $(OH)_2D$ to range up to 30% in excess of the adult normal range in normal children. Furthermore, there appears to be less stringent regulation of conversion of 25-OHD to 1,25 $(OH)_2D$ in early life.

Disorders of Hypocalcaemia

Serum Calcium

Total serum calcium is comprised of a free or ionized calcium component, a protein (primarily albumin) bound component, and a small component of filtrable calcium that is complexed to other ions such as sulphate, citrate, or phosphate. The ionized and protein bound components each represent approximately 45–50% of the total calcium. The ionized fraction is the biologically active component, and derangements in this fraction result in clinical symptoms. As discussed elsewhere, total serum calcium can be low with a simultaneously normal ionized fraction. This finding is typical of hypoalbuminaemia or acidosis. Various correction factors have been proposed, and are applicable to children as well as adults, however accurate measures of ionized calcium are preferable to calculated corrections.

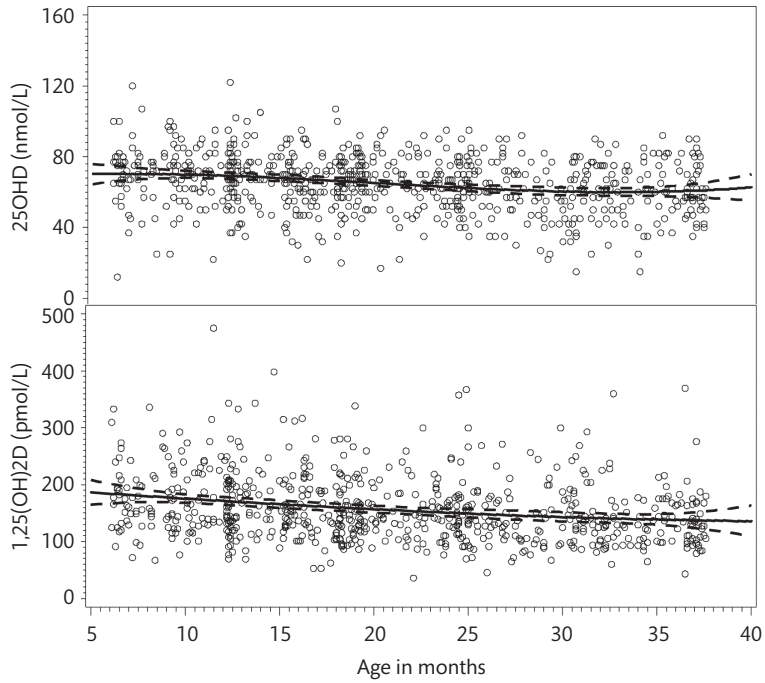


Figure 4.7.2 Scatterplot of 25(OH)D (top) or 1,25(OH)2D (bottom) related to age in months. The regression equation for the relation of 25(OH)D to age was as follows: 25(OH)D (nmol/L) = 74.9–0.44 (age in mo) (R = 0.18, *P* = 0.001). The regression equation for the relation of 1,25(OH)2D to age was as follows: 1,25(OH)2D (pmol/L) = 188.7–1.5 (age in mo) (R = 0.23, *P* = 0.001). Regression lines (solid lines) and their 95% CIs (dashed lines) are shown. 1,25(OH)2D, 1,25-dihydroxyvitamin D; 25(OH)D, 25-hydroxyvitamin D. Normal ranges of circulating 25(OH)D and 1,25(OH)2D levels in infancy. Circulating 25(OH)D and to a greater extent, circulating 1,25(OH)2D levels decrease with age over the first 3 years of life.

Hypocalcaemia in Childhood

This section discusses disorders of calcium homeostasis in childhood, with a primary focus on abnormalities in the maintenance of serum calcium. Several of these disorders, however, primarily affect the skeletal calcium compartment, and bone disease may be a more significant abnormality than perturbations in the serum concentrations. Thus, certain disorders are described in which serum calcium levels are often normal in the clinical setting, but at the expense of osteopaenic or rachitic abnormalities.

The presentation of hypocalcaemia in the newborn period typically includes facial twitching, limb jitteriness, or other features of neuromuscular irritability, occasionally progressing to focal or generalized convulsions. Poor feeding, hyperacusis, and laryngospasm mimicking croup have been described. On the other hand, non-specific findings such as apnoea, tachypnoea, tachycardia, cyanosis, or vomiting may be the only features evident. In older children tetany or perioral tingling, presentations more characteristically seen in adults, are more likely to be encountered. As with infants, focal seizures and generalized convulsions may occur. Carpo-pedal spasm in school age children has often been attributed to a writer's cramp. This phenomenon may be exacerbated by the hypomagnesaemia and alkalosis frequently encountered in states of parathyroid insufficiency. Lethargy, vomiting, and other non-specific signs have also been reported. The electrocardiogram may reveal a prolonged corrected QT interval, Q-T_c, which is determined by dividing the Q-T interval by the square root of the EKG cycle. The upper limit of normal in children is 0.44. The musculature may be affected by chronic hypocalcaemia, as witnessed by the cardinal sign of tetany. Serum creatine kinase activity (CK, or creatine phosphokinase, CPK) may be elevated in chronic hypocalcaemia; over time actual myopathic changes may occur. Severe clinical compromise with dilated cardiac myopathy has been described in infants with hypocalcaemia due to severe vitamin D deficiency [29].

Neonatal Hypocalcaemia

Neonatal hypocalcaemia seen transiently in the first few days of life is commonly referred to as **early neonatal hypocalcaemia**. This is often seen in preterm infants and has been explained as an exaggeration of the normal postnatal decrease in serum calcium levels. Early neonatal hypocalcaemia appears to occur with greater frequency in asphyxiated babies and in infants of diabetic mothers (IDM) than otherwise. The hypocalcaemia seen in IDM is probably multifactorial. Magnesium deficiency has been implicated, as well as alterations in maternal metabolism secondary to poor glucose control throughout gestation. Whether the normal postnatal increase in PTH secretion is blunted is not entirely clear.

Late neonatal hypocalcaemia occurs after 5–7 days of life and is more characteristic of the term infant. Late neonatal hypocalcaemia often presents with seizures and is less likely to be transient in nature. Hypoparathyroidism and magnesium deficiency often present in this time frame. Hypocalcaemia in babies with congenital heart disease of many types has been reported as a relatively common finding. Hypocalcaemia related to vitamin D deficiency may present at several weeks of age, however, radiographic evidence of rickets is usually not observed until over 2 months of age.

One classic situation in which prolonged neonatal hypocalcaemia occurs is in **the infant of the hyperparathyroid mother**. Presumably the maternal hypercalcaemia results in increased transport of calcium from the maternal to fetal circulation. The resultant excess calcium supply to the fetus is thought to suppress parathyroid responsivity, and prolonged hypoparathyroidism results. Symptomatic hypocalcaemia and hyperphosphataemia are typical biochemical features; hypomagnesaemia may occur as well. The disorder is usually transient, but some cases have been prolonged for months. Unrecognized maternal hyperparathyroidism should be carefully investigated in children that present with the characteristic features of the disorder. Maternal familial hypocalciuric hypercalcaemia (FHH) can result in this syndrome. We have encountered similar cases due to excessive maternal ingestion of calcium-containing antacids during pregnancy, suggesting that other causes of chronic maternal hypercalcaemia can result in a similar clinical picture.

Hypocalcaemia in the newborn setting may also occur during blood transfusions using citrated blood products. Citrate complexes with ionized calcium, reducing its circulating concentration to a level where neuromuscular hyperexcitability may occur. Total serum calcium is usually not decreased. Hypocalcaemia can occur in the **congenital nephrotic syndrome**. Persistent hypocalcaemia may present in this time frame as well. Congenital hypoparathyroidism may be present, as in the **DiGeorge syndrome**. The classic triad of this chromosome 22 deletion syndrome (hypoparathyroidism, athymia, and cono-truncal defects of the heart) typically results in long-standing hypoparathyroidism, although 'partial' hypoparathyroidism has been described. (See next for a detailed description of congenital disorders of the parathyroid glands.) Severe **osteopetrosis** may present with hypocalcaemia secondary to impaired mobilization of calcium from bone. Typically, PTH levels are elevated in order to compensate for the hypocalcaemia engendered by defective osteoclastic bone resorption in such cases. Severe **vitamin D deficiency** is generally an acquired condition manifest as hypocalcaemia as early as 2 to 3 months, but low maternal stores have rarely contributed to its development in an even younger age range.

Osteopaenia of prematurity is commonly encountered in premature infants. Poor bone mineralization is evident on radiographs or other measures of BMD. In general, the problem is more severe in children of lower birth weight. With increasing survival of children with birth weights less than 1000 grams the severity of this problem is increasing. Classical rachitic changes of flared and frayed epiphyses, craniotabes, and a rachitic rosary may develop over the first months of life. The histologic pattern of bone is thought to be a combined lesion with components of osteomalacia and osteoporosis. This disorder is a consequence of premature withdrawal of the maternal mineral supply. The enteral route, even with maximum feeding delivery, cannot provide for the mineral demands of the skeleton as it rapidly grows and mineralizes throughout the latter weeks of gestation. The problem often occurs in the setting of normocalcaemia. In preterm infants fed solely with breast milk, a phosphate deficiency syndrome may occur, as the phosphate content of breast milk is considerably less than that of commonly used in cow's milk formulas. Although breast milk phosphate is adequate for the growth of the term infant's skeleton, human breast milk fortifiers are routinely used to increase the mineral intake of the preterm infant. One caveat regarding the use of such fortifiers: calcium intake, if excessive, can result in hypercalcaemia, as its absorption is not tightly regulated in early infancy, and the fractional absorption of calcium can be very high in a low birth weight premature infant.

Treatment of Neonatal Hypocalcaemia

Symptomatic infants are treated with calcium, but there is controversy regarding treatment of hypocalcaemic infants who are asymptomatic. The emergency treatment of neonatal hypocalcaemia consists of intravenous administration of 1 ml/min of 10% calcium gluconate, which should not exceed 2.0 ml per kg body weight (which is equivalent to ~19 mg of elemental calcium per kilogram). This may be repeated three to four times in 24 hours. After acute symptoms have been managed, 5.0 ml/kg of 10% calcium gluconate may be given with intravenous fluids over 24 hours. It should be emphasized that repetitive and frequent boluses of high-dose intravenous calcium only temporarily raise the serum calcium level and an enormous renal calcium load result. Thus the favoured way of longer-term management when the parenteral route is necessary employs a constant intravenous infusion, which results in a lower calcium load per unit time to the kidneys, and maintains serum calcium at a relatively steady level. Calcium supplements may be introduced orally if tolerated. In persistent cases, the load of dietary phosphate should be lessened with a formula such as Similac PM 60/40. When hypomagnesaemia is identified, it can be treated with 0.1–0.2 ml/kg of a 50% solution of magnesium sulphate ($MgSO_4 \bullet 7H_2O$).

Hypocalcaemia of Later Onset

Disorders of Parathyroid Insufficiency

Congenital hypoparathyroidism. A wide variety of hypocalcaemic syndromes occur in children due to abnormalities in parathyroid synthesis or secretion. These syndromes provide classic examples of the critical role PTH plays in protecting the organism from acute decreases in the serum calcium level. The clinical manifestations of hypocalcaemia described earlier are the typical presenting features of hypoparathyroidism. Biochemical features usually include low blood total and ionized calcium levels, and an elevated blood phosphate level. The serum magnesium level may be low, and alkalosis may be present.

Hypoparathyroidism may result from a variety of causes. A number of genetic disorders may cause agenesis of parathyroid glands, disrupt PTH synthesis, processing, and secretion, and/or result in autoimmune destruction of parathyroid glands. As noted earlier, agenesis of the parathyroid glands occurs in the classic DiGeorge triad (OMIM #188400) of hypoparathyroidism, athymia, and cono-truncal heart defects. This syndrome is now known to be part of the larger spectrum of disease referred to as CATCH 22, a sequence of contiguous microdeletion syndromes localized to chromosome 22q11.2. These mutations are most notable in the *TBX-1* gene, a transcription factor that plays an important role in development of thymus and parathyroid glands [30]. Recently, some cases of DiGeorge syndrome have been found to harbour a deletion in 10p13-p12 (*NEBL* gene), which has been defined as DiGeorge type 2 (OMIM #601362) [31, 32]. Hypoparathyroidism usually occurs in those disorders most related to the classic DiGeorge syndrome, but has also been described in the velo-cardio-facial syndrome (OMIM #192430), another CATCH-22 pattern of anomalies.

Fluorescent *in situ* hybridization (FISH) using DNA probes which hybridize to the 22q11.2 locus is helpful in establishing the diagnosis. Not all hypocalcaemia in this microdeletion syndrome is due to parathyroid dysgenesis however; a recent description of pseudohypoparathyroidism in DiGeorge syndrome noted a deletion on chromosome 22 incorporating a G-protein locus ($G\alpha_s$)[33].

Mutations of the pre-proPTH gene, resulting in disruption of PTH secretion [34], or processing and translocation of PTH from the endoplasmic reticulum, [35] can cause **familial isolated hypoparathyroidism** (OMIM #146200) [36]. Likewise mutations of the transcription factor glial cells missing-2 (GCM2) [also referred to as glial cell missing-B (GCM-B)] on 6p24.2, resulting in loss of activity, will result in **familial isolated hypoparathyroidism** (OMIM #307700) [34, 37–39] A form of **X-linked hypoparathyroidism** has been reported in patients with mutations in the gene for the transcription factor Syr-box 3 (SOX3) [40].

A number of other patients with congenital hypoparathyroidism have concomitant involvement of other organ systems. The syndrome of hypoparathyroidism, sensorineural deafness, and renal anomalies (**HDR syndrome**; OMIM #146255) is an autosomal dominant disorder due to mutations or deletions of the gene for the transcription factor GATA3 on chromosome 10p14 [41–43]. Loss-of-function mutations in tubulin chaperone E (*TBCE*) gene on chromosome 1q42.3 cause the autosomal recessive syndrome of infantile onset hypoparathyroidism, mental retardation, intrauterine growth retardation, and dysmorphism (**Sanjad-Sakati syndrome**) (OMIM #241410) [44–46] and **Kenny-Caffey syndrome type 1** (OMIM #244460) (hypoparathyroidism, dwarfism, medullary stenosis of the long bones, and eye abnormalities) [47]. In autosomal dominant forms of Kenny-Caffey syndrome Type 2 patient has the clinical features, transient hypocalcaemia, but normal intelligence due to heterozygous mutations in *FAM111A* gene on chromosome 11q12 [48]. Mitochondrial gene defects, ranging from deletions to mutations and rearrangement, have been associated with hypoparathyroidism in patients with **Kearns-Sayre syndrome** (external ophthalmoplegia, pigmentary retinopathy, cardiomyopathy, diabetes, and hypoparathyroidism; OMIM #530000), **MELAS syndrome** (mitochondrial encephalomyopathy with lactic acidosis and stroke like syndrome, with diabetes and hypoparathyroidism; OMIM #540000), and **mitochondrial trifunctional protein deficiency (MTPDS) syndrome** manifest by peripheral neuropathy, retinopathy, and hypoparathyroidism due to mutations in *HADHA* at the 2p23.3 locus (OMIM #609015) [49, 50]. Other mitochondrial disorders associated with hypoparathyroidism include **long-chain 3-hydroxyacyl-coenzyme A dehydrogenase deficiency** (OMIM# 609016 LCHAD), and **propionic acidaemia** [51].

Gain-of-function mutations of the CaSR at chromosome 3q21.1 have been found to cause an autosomal dominant variety of hypocalcaemia. The CaSR protein, on parathyroid cell membranes, regulates PTH secretion. Mutations causing the protein to enhance its affinity for ionized calcium will allow relatively low concentrations of this ion to suppress PTH secretion, and a steady state serum calcium level in the hypocalcaemic range is established. This autosomal dominant form of hypocalcaemia has been termed ADH1 (OMIM #601198). It may be possible to distinguish **autosomal dominant hypocalcaemia** (ADH) from other forms of parathyroid insufficiency because of the relative hypercalciuria that occurs. Circulating PTH levels may not be undetectable in the untreated state [50, 52, 53]. This distinction from hypoparathyroidism has important consequences regarding long-term management. That is, the standard treatments for hypoparathyroidism, vitamin D, or its metabolites, and calcium, can further exaggerate this hypercalciuria such that nephrocalcinosis, nephrolithiasis, and renal impairment may result, particularly when serum calcium levels are kept in the usual normal range. ADH may present with features seen in **Bartter syndrome** (Bartter syndrome type 5), including hypokalaemia, metabolic alkalosis, hypomagnesaemia, hyperreninemia and hyperaldosteronaemia [54, 55]. A decrease in the distal tubular fractional chloride reabsorption rate with negative NaCl balance has been reported in some cases, with a resultant secondary hyperaldosteronism and hypokalaemia [54]. Hydrochlorothiazide and recombinant human PTH have been suggested as treatment modalities [56, 57].

Finally, heterozygous mutations in the alpha subunit of the G protein G11 (*GNA11*) on chromosome 19p13 have been identified in patients with ADH (ADH type 2; OMIM #615361) [58].

The CaSR and associated disorders are discussed in Chapter 4.4.

Acquired hypoparathyroidism. Causes of acquired hypoparathyroidism are mainly autoimmune, post-surgical, or due to infiltration of the parathyroid gland. A major cause of acquired hypoparathyroidism is due to the autoimmune polyglandular syndrome (or APS), type I. The disorder is also referred to by other names such as autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy (APECED) [59, 60]. The primary manifestations of this disorder include hypoparathyroidism, primary adrenal insufficiency, and mucocutaneous candidiasis. A variety of other autoimmune phenomena may occur, often grouped with the endocrine, gastrointestinal, or dermatologic systems (Box 4.7.1). Defects in cellular or humoral immunity may occur. The presentation typically begins in early childhood with candidiasis. In later childhood the onset of hypocalcaemic symptoms related to hypoparathyroidism occurs, and Addison's disease often presents during adolescence. There is considerable variability, however, and often patients do not develop the classic triad. The elevation in serum calcium that may occur during an acute adrenal crisis in an undiagnosed individual may result in an increase of the serum calcium from a hypocalcaemic to a normocalcaemic value, thus masking a coexistent presentation of hypoparathyroidism. Serum calcium should be determined during the initial presentation of suspected Addison's disease, but also shortly after recovery from the acute crisis.

APS-1 syndrome is due to mutations in the gene encoding the autoimmune regulator protein AIRE [61–63]. Mutations resulting in a single amino acid substitution at position 257 (arginine –> glutamic acid) were present in more than three-quarters of the affected Finnish cases. The same mutation has been identified in the majority of cases in another series from Italy and Switzerland. Mutations in AIRE have been identified in cases of isolated hypoparathyroidism as well [64]. Autoantibodies directed to parathyroid cells have been described in patients with APS type 1, but the frequency of these findings vary greatly dependent upon the series studied [65]. Antibodies to the parathyroid CaSR have been reported in acquired hypoparathyroidism both with and without coexisting autoimmune diseases [66–70]. Yet, CaSR antibodies are not found to be specific or sensitive markers for hypoparathyroidism in APECED [71]. In addition to the usual measures taken in the management of hypoparathyroidism, treatment of this disorder may require aggressive mineral replacement, due to the often-complicating issue of malabsorption. Acute illness can be associated with such severe impairment of gastrointestinal absorption of calcium and magnesium

Box 4.7.1 Clinical features of autoimmune polyendocrinopathy syndrome type I

Major manifestations
Chronic hypoparathyroidism
Chronic candidiasis
Autoimmune Addison's disease

Other manifestations
Autoimmune hypogonadotropic hypogonadism
Alopecia
Chronic hepatitis
Chronic atrophic gastritis
Pernicious anaemia
Vitiligo
Malabsorption
Nephritis
Sjögren's syndrome
Autoimmune thyroid disease
Keratoconjunctivitis
Hypophysitis
Insulin-dependent diabetes mellitus
Exocrine pancreatic insufficiency
Pneumonitis
Vasculitis
Enamel hypoplasia
Nail dystrophy
Haemolytic anaemia
Turner syndrome

In decreasing order of incidence, adapted with permission from Betterle, C., N.A. Greggio, and M. Volpato, Clinical review 93: Autoimmune polyglandular syndrome type 1. *J Clin Endocrinol Metab*, 1998. 83(4): p. 1049–55 [59] and from Husebye, E.S., M.S. Anderson, and O. Kampe, Autoimmune Polyendocrine Syndromes. *N Engl J Med*, 2018. 378(12): p. 1132–1141 [60].

so that parenteral replacement of these minerals may be required. Continuous nocturnal nasogastric calcium supplementation may be a useful temporary measure in the affected child unable to tolerate standard bolus feeding, and where prolonged parenteral infusions are not practical.

The long-term prognosis of this condition has improved greatly over the past generation. Early cases succumbed to such problems as unrecognized adrenal crises or diabetic keto-acidosis. Although numerous complications of this disorder are recorded, including overwhelming candida sepsis, oesophageal carcinoma, and chronic active hepatitis, these severe features are rare. One review records a 50-year survival of greater than 75%.

Surgical hypoparathyroidism is rarely encountered in childhood but can be a complication of thyroid surgery. Average rates of transient or permanent hypoparathyroidism are 5–15%, respectively, following thyroidectomy [72]. Standard guidelines for thyroid surgery include identification and preservation of parathyroid tissue. The use of radioactive iodine to ablate any thyroid remnants following surgery for cancer has resulted in a less aggressive approach to thyroid surgery. Transient hypocalcaemia in the 36 hours acutely following thyroid surgery is common, although no mechanism has been clearly established which accounts for this finding. Extent of surgery correlates with the risk of hypoparathyroidism as manipulation of the parathyroid glands can result in transient or permanent hypoparathyroidism [72]. Approaches to identifying those patients at an increased risk of developing hypocalcaemia after thyroidectomy include measurement of perioperative intact PTH (iPTH), early (1 hour) postoperative iPTH measurement [73, 74], and postoperative phosphorus monitoring. A level of postoperative iPTH <10–15pg/ml has been found to correlate with an increased risk of clinically significant hypocalcaemia in adults [75, 76]. In randomized controlled trials where patients were either given oral calcium and vitamin D supplements or no supplements, fewer patients on supplements developed hypocalcaemia [77, 78]. Thus, empiric calcium with or without calcitriol replacement therapy can be considered in those patients undergoing total thyroidectomy [72].

Hypoparathyroidism may occur as a complication in disorders related to metal toxicity and infiltration of the parathyroid glands. **Thalassaemia** has been shown to result in a variety of endocrinopathies related to iron deposition. Although hypoparathyroidism was recognized as an occasional complication in most clinics managing such patients, the incidence of this complication has decreased over the past 20 years with increasing use of chelating therapies such as deferoxamine [79]. Hypoparathyroidism has been reported to occur in **Wilson's disease** presumably related to copper deposition in the parathyroid glands [80].

Functional hypoparathyroidism occurs in severe **magnesium deficiency**. Both impairment of parathyroid secretion and resistance to PTH activity at the renal tubule have been described in the setting of chronic magnesium deficiency. In classic studies by Anast *et al.* [81], the dependence of the parathyroid glands on magnesium for secretion of PTH was described in a girl with a congenital magnesium wasting syndrome. It appears that more subtle defects in PTH secretion may occur with less severe magnesium depletion. The serum magnesium level may not reflect total body magnesium status, as magnesium is predominantly an intracellular ion. However, serum magnesium level of greater than 1.3 mg/dl is unlikely to result in clinically significant changes in parathyroid secretion. In children chronic magnesium deficiency occurs in **familial hypomagnesaemia** [82]. **Familial hypomagnesaemia with secondary hypocalcaemia** is an autosomal recessive disease due to mutations in the TRPM6 ion channel, resulting in electrolyte abnormalities in the newborn period [83]. This disorder may present with hypocalcaemic/hypomagnesaemic seizures in the first 2 months of life. If diagnosed early, severe neurologic impairment may be prevented. Mutations in paracellin (encoded by *CLDN16* (OMIM #248250)), a renal tubular paracellular transport protein of the claudin family, may also cause hypomagnesaemia, hypocalcaemia, and hypercalciuria [84]. Another member of this family, *CLDN19*, (OMIM #248190) may also cause a similar syndrome [85]. **Gitelman's syndrome** (OMIM #263800) is an autosomal recessive disorder of magnesium and potassium wasting with metabolic alkalosis and hypocalciuria, due to mutations in *SLC12A3* which encodes a thiazide-sensitive NaCl cotransporter [86].

Hypomagnesaemia in children occurs more frequently with the use of chemotherapeutic agents such as cisplatin, and with aminoglycoside antibiotics such as tobramycin. Hypomagnesaemia has been reported with ibuprofen overdosage in a 21-month-old child simultaneously treated with furosemide [87], and following ingestion of ammonium bifluoride-containing automobile wheel cleaner [88]. Recent data has suggested that moderate decreases in serum magnesium levels accompany the hypocalcaemia encountered during rehabilitation from burn injuries in children [89].

Treatment of chronic hypoparathyroidism. The mainstay of the management of chronic hypoparathyroidism is replacement

with oral calcium supplements and an active metabolite of vitamin D, such as 1,25 $(OH)_2D_3$ (calcitriol) or $1\alpha-(OH)D_3$. Various liquid forms of calcium carbonate (often sold as antacids for children) are useful for the young child unable to take tablets. With conventional therapy, larger doses are often utilized but may not always achieve a physiological normalization of calcium-phosphate homeostasis. The doses are titrated to maintain the serum calcium in the slightly low to low-normal range, with care not to render the child hypercalciuric. Hypercalciuria is especially a concern in patients with ADH due to activating mutations in the CaSR that can be managed by adjunctive therapy with thiazide diuretics. Novel approaches using teriparatide [parathyroid hormone(1-34); PTH(1-34)] and calcilytic agents (antagonists of the CaSR) have recently been reported as potential therapeutics for these disorders.

Limited studies of PTH in children have shown promising results [90–95] with respect to maintenance of normal serum calcium levels [93]. Twice-daily subcutaneous injections were superior to conventional therapy with calcium and calcitriol [93]. Twice-daily dosing resulted in reduced variation in serum calcium levels compared to single daily doses, and with lower total daily doses [94]. Renal function, bone density Z-scores, bone accrual and linear growth remained normal in children over the 3-year treatment period in the study [93]. In children with severe congenital hypoparathyroidism, pump delivery of PTH(1-34) was superior to twice-daily injection therapy, maintaining even more physiologic calcium homeostasis [92].

Although PTH(1-34) tended to lower urinary calcium excretion, results did not reach statistical significance in comparison to conventional therapy [93] or were unchanged from baseline [95]. Only one adult study of pump delivered of PTH(1-34) demonstrated a statistically significant reduction in urinary calcium excretion in comparison to subcutaneous injection [96].

Effects of PTH(1-34) on different skeletal compartments have been studied in both adults and adolescents [95], showing that daily injections of PTH(1-34) stimulated bone turnover, increased bone volume, with an apparent increase in trabecular bone and trabeculation of cortical bone [95].

PTH(1-34) was used safely in a child with hypoparathyroidism due to a sporadic calcium receptor mutation [97] resulting in improved serum mineral control, increased bone mass with normal growth, but this approach did not prevent nephrocalcinosis [97].

These findings lead us to advocate that the first line of chronic therapy for children with hypoparathyroidism should be oral calcium and calcitriol. Use of PTH may be considered in those cases resistant to conventional therapy. However, as long-term effects on developing skeleton are not known, future studies in this regard will be informative.

Calcilytics, which are allosteric antagonists of CaSR, have been considered for treatment of ADH. Studies of calcilytics in mutant *CASR* knock-in mice with the ADH phenotype showed dose-dependent effects on serum calcium and phosphate levels [98]. Urinary calcium levels were improved and renal calcification prevented [98]. Although both calcilytics and PTH(1-34) resulted in improvements in serum calcium and phosphate levels, hypercalciuria and renal calcification persisted with PTH(1-34) [98]. Further studies in humans are needed to better understand the efficacy and potential side effects of calcilytics.

PTH Resistance and Pseudohypoparathyroidism

Pseudohypoparathyroidism (PHP) and pseudopseudohypoparathyroidism (PPHP) are caused by mutations and/or epigenetic changes at the complex *GNAS* locus on chromosome 20q13.3 that undergoes parent-specific methylation changes at several sites [99].

GNAS encodes the alpha subunit of the stimulatory guanine nucleotide regulatory protein ($G\alpha_s$) and several splice variants. There is less paternal $G\alpha_s$ expression in certain tissues, such as the proximal renal tubules, thyroid, and pituitary gland, thus in the presence of maternal *GNAS* mutations, the absence of this protein will lead to hormonal resistance. The syndrome represents a fascinating aspect of hormone receptor biology, and is discussed in detail in Chapter 4.5. For our purposes here, we will describe only a few of the clinical features pertinent to children with the disorder.

Heterozygous inactivating mutations involving the maternal *GNAS* exons 1–13 cause PHP type Ia (PHP1A-OMIM #103580) with PTH-resistant hypocalcaemia and hyperphosphatemia. Additional characteristics of these patients include a short stocky build, round face, early onset obesity, ectopic intramembranous calcifications, brachydactyly, and neurodevelopmental abnormalities; these clinical findings are referred to as Albright hereditary osteodystrophy (AHO). $G\alpha_s$ can be found in most tissues, thus when defective could account for PTH resistance in PHP1A patients, resistance to multiple other hormones that act by stimulating adenylate cyclase [100] including thyroid-stimulating hormone (TSH), gonadotropins, GHRH (growth hormone releasing hormone), and central nervous system (CNS) neurotransmitters can be seen. The respective hormone resistance would be manifest by thyrotrope resistance to TSH, menstrual disorder/irregularities, cryptorchidism in males, and growth hormone deficiency with short stature. Presentation in infancy may include short stature and congenital elevations in TSH, but manifestations of hypocalcaemia usually happens later in childhood.

GNAS mutations inherited on the paternal allele, cause PPHP (OMIM #612463), evidenced by the skeletal phenotype of Albright hereditary osteodystrophy, but in the absence of biochemical abnormalities associated with hormone resistance, described here earlier, in PHP1A.

PHP1B (OMIM #603233) patients demonstrate PTH-resistant, hypocalcaemia, and hyperphosphatemia without the somatic abnormalities of absence of heterozygosity (AOH). These patients have normal $G\alpha_s$ bioactivity [101]. Autosomal dominant PHP type IB (AD-PHP1B) is caused by heterozygous maternal deletions within *GNAS* or *STX16* (Syntaxin-16). *STX16* encodes a protein that is a member of the Syntaxin or t-SNARE (target-SNAP receptor) family. These proteins are found on cell membranes and permit specific synaptic vesicle docking and fusion. Loss-of-methylation at maternally methylated *GNAS* result in reducing $G\alpha_s$ expression, leading to hormonal resistance [99, 101, 102].

PHP1C (OMIM #612462) is a subgroup of PHP1A, with clinical and laboratory findings typical for PHP1A (multihormonal resistance and AHO features), but normal $G\alpha_s$ activity [99, 103, 104].

Treatment of PHP is similar to that for primary hypoparathyroidism, although there is usually little risk of hypercalciuria in the classic (PHP1A) form of the disease. Subcutaneous ectopic ossification must be monitored, which if severe may require surgical removal. Although many treat the TSH resistance with thyroxine supplementation because of the elevated TSH levels, others do not,

in view of normal thyroxine levels. No data has confirmed efficacy of the approach.

Disorders Related to Vitamin D Metabolism

Hypocalcaemia may result from a deficiency of vitamin D, usually related to limited dietary content of vitamin D or limited exposure to ultraviolet light (UV light), critical in the early steps of endogenous vitamin D production in skin. Sufficient UV light is necessary for the production of pre-vitamin D in the stratum spinosum of the dermis. UVB light of wavelength 290–315 nm provides the energy to disrupt the 9–10 C-C bond in the B ring of the steroid nucleus of 7-dehydrocholesterol (**Figure 4.7.3**). Previtamin D is then rapidly isomerized in the skin to vitamin D. Thus, vitamin D deficiency is most frequent in parts of the world where sunlight exposure is limited or exposure to sunlight is prevented. In North America, there have been numerous reports of vitamin D deficiency rickets in breastfed children not supplemented with vitamin D. Breast milk contains little vitamin D, unless the mother is taking pharmacologic doses of the vitamin D. Pigmented individuals are at higher risk for this problem, as melanin absorbs UV light external to the layer of skin where vitamin D synthesis occurs. There is a greater incidence of the problem in the late winter as compared to other times of the year. In some children vitamin D deficiency is compounded by the coincident problem of calcium deficiency. This problem may occur in lactose intolerant children avoiding dairy products, or when the diet has a very high phytate content (as with certain grains and cereals) which can limit the bioavailability of ingested calcium.

Vitamin D deficiency may often present with rachitic bone disease. Symptomatic hypocalcaemia generally occurs late in the course of development of vitamin D deficiency. The initial decreases in intestinal calcium absorption which result from vitamin D deficiency are readily compensated for by the resultant secondary elevations in PTH. When frank hypocalcaemia with tetany or seizures occurs due to vitamin D deficiency, substantial chronicity of vitamin D deficiency has usually been present.

Vitamin D is further metabolized to its most abundant circulating metabolite, 25-OHD, and its best-known active metabolite, 1,25 $(OH)_2D$. The critical mechanism of action for this involves binding to its receptor, a DNA binding protein which is part of the large superfamily of steroid/thyroid/retinoid receptors. Pathophysiology similar to vitamin D deficiency may result from defects in the 1α-hydroxylase enzyme instrumental in the synthesis of 1,25 $(OH)_2D$, or in mutations in the vitamin D receptor.

Clinical features of rickets in children include bowing of the lower extremities, craniotabes, and rachitic rosary (hypertrophy of the costochondral junctions). These syndromes are discussed in detail in Chapter 4.11. The laboratory findings of vitamin D-related hypocalcaemic disorders are compared in **Table 4.7.2**.

Establishment of an appropriate biochemical threshold for the definition of vitamin D deficiency has received considerable attention as there is no consensus on the optimal level of total serum 25-OHD. The Institute of Medicine provides extensive data suggesting that a threshold value of 20 ng/ml provides adequate vitamin D for healthy bone and mineral status in children. In high risk individuals and in individuals prone to metabolic bone disease and/or vitamin D deficiency, keeping 25-OHD above 30 ng/mL can be considered as suggested by the Endocrine Society Task Force [105]. U.S. Preventive Services Task Force issued its own recommendation in adults and concluded that the evidence on screening for vitamin D deficiency in asymptomatic adults to improve health outcomes is insufficient. Thus, current recommendations are for those individuals who are at risk for developing vitamin D deficiency to be screened.

In addition to its well-known effects on calcium/phosphate homeostasis and bone metabolism, various health benefits have been associated with vitamin D status, mostly related to a variety of associations of 25-OHD level and prevalence of certain disorders, including diabetes, multiple sclerosis, cancers (particularly colorectal), and obesity [106]. These data are epidemiologic in nature, and the possibility that vitamin D status serves as a marker for other unidentified contributors to disease remains unknown though there is a suggestion that low vitamin D levels can be an independent predictor of all-cause mortality. Studies in adults have reported associations between dyslipidaemia and cardiovascular disease (CVD) and vitamin D deficiency [107, 108]. Data are limited in children, but similarly, vitamin D deficiency has been associated with markers of CVD in children with obesity, specifically with an increase in atherogenic lipids, suggesting increased cardiovascular risk [109]. A convincing example of 'non-classical' vitamin D effects (i.e. effects apart from those influencing systemic calcium homeostasis) has been demonstrated directly in human macrophages [110]. In the presence of activated toll-like receptors 1/2, macrophages are able to express their own 1α-hydroxylase and its own vitamin D receptor (VDR). Thus, the macrophage has the capacity to metabolize 25-OHD and to use the activated product, 1,25$(OH)_2D$, in autocrine fashion, as it can produce this molecule's receptor. The stimulation of the macrophage in this way leads to the production of a unique antimicrobial peptide, cathelicidin, which is inhibitory to the growth of *M. tuberculosis*. This mechanism has been posited as a means by which vitamin D may play a role in fighting infection. Elevations in circulating PTH occur as circulating 25-OHD decreases, suggesting a mechanism by which 25-OHD deficiency might be defined. However, the threshold values of 25-OHD at which an increase in PTH levels occurs is quite variable [111] suggesting caution in using this measure as a generalizable means of establishing vitamin D deficiency. Nevertheless, these findings, in sum, raise the issue of revising the threshold for 25-OHD level as a measure of optimal vitamin D status, and toxicity information in adults appear to indicate that modest increases in supplementation is safe. Data is not yet available to establish a clear benefit to this approach, and the application of such measures to infants and children needs to be carefully examined. Indeed, the administration of vitamin D to infants and children at the increased levels recently suggested for supplementation in adults could be risky. We have observed hypercalcaemia in a newborn infant given 1400 units of supplemental vitamin D daily, with concomitant circulating 25-OHD levels over 90 ng/ml (225 nmol/L) and thus support a conservative definition of vitamin D deficiency, using threshold values for 25(OH) D of 20 ng/ml (50 nmol/L). In the normal healthy term infant, we advise adherence to current recommendations of a daily vitamin D intake of 400 IU, and 600 IU in children 1 year of age and older.

Treatment: Vitamin D in dosages of 1000–2000 IU per day is a standard approach to the initial treatment of vitamin D deficiency. As rachitic lesions heal, the dosage is decreased to 600 IU per day, the generally recognized recommended daily allowance. An assessment after 8–12 weeks of therapeutic dosing is indicated, with extension of therapeutic doses if healing is insufficient. Consideration

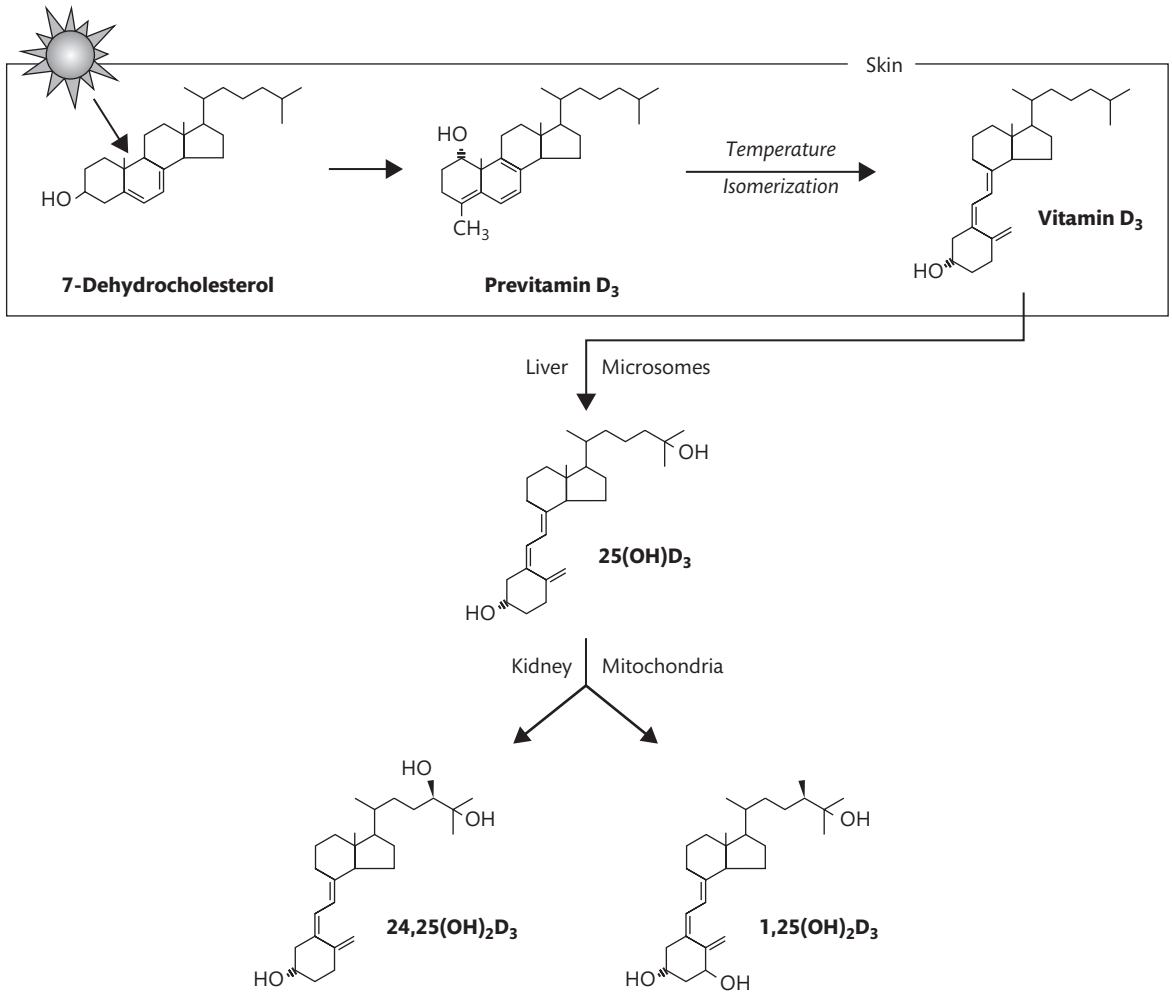


Figure 4.7.3 The vitamin D biosynthetic pathway. The steroid nucleus of 7-dehydrocholesterol is converted in skin to pre-vitamin D_3 with exposure to UVB light and rapidly isomerized to vitamin D_3, which is found in ng/ml concentrations in the circulation. This metabolite is converted to 25-OH D3 in hepatic microsomes, and is also found in ng/ml amounts in the circulation. Measurement of circulating 25-OH D_3 is a biomarker of total body vitamin D stores. 25-OH D_3 is converted in renal mitochondria to the best-known active metabolite, 1,25 (OH)2 D_3 which circulates in pg/ml concentrations in serum. Conversion of 25-OH D_3 to 24,25 (OH)2 D_3 also occurs in renal mitochondria.

of an alternative diagnosis may also be an important consideration at this stage. A single intramuscular dose of 600 000 units of vitamin D, or in two oral doses of 300 000 units each can be given in the outpatient clinic if the clinical situation would indicate that limited follow-up will occur.

The specialty clinician may happen to evaluate such a patient after a change in season, and sunlight exposure has concomitantly increased since the disease onset, or after vitamin supplementation has begun. This situation should be recognized clinically, as low-normal values of the 25(OH)D level may confuse the diagnosis.

Table 4.7.2 Laboratory findings in childhood syndromes presenting with hypocalcaemia

Serum biochemical measures[1]	Ca	P	Alk phos	PTH	25D	1,25D
Vitamin D deficiency	N,↓	N,↓	↑	↑	↓	↓,N,↑
Calcium deficiency	N,↓	N,↓	↑	↑	N	↑
Vitamin D 1-α hydroxylase defect	↓,N	↓,N	↑	↑	N	↓,N
Hereditary resistance to vitamin D	↓,N	↓,N	↑	↑	N	↑
Hypoparathyroidism	↓	↑	N	↓	N	↓N,↑
Pseudohypoparathyroidism	↓	↑	N	↑	N	↓,N

[1] Serum measures for calcium, phosphorus, alkaline phosphatase, parathyroid hormone, 25-hydroxyvitamin D, and 1,25 dihydroxyvitamin D are outlined. N, normal levels; ↓, decreased levels; ↑, increased levels.

Radiographs may demonstrate a thin dense line of opacity at the metaphyses of long bones, which indicates that recent rapid mineralization has occurred at the edge of the growth plate.

As mentioned earlier, children with vitamin D deficiency often require supplemental calcium. Some children may manifest hungry bone syndrome, in which mineralization is rapid and serum calcium levels may decrease as the bone mineralizes. Thus, supplemental calcium is given in many cases to provide a total daily intake of 30–50 mg/kg of elemental calcium. Vitamin D stores may be depleted rapidly during calcium insufficiency [112], suggesting that dietary calcium deficiency itself may be yet another risk factor for the development of vitamin D deficiency.

Deficiency of 1α-hydroxylase is best treated with physiologic doses of 1,25 $(OH)_2D_3$. Hereditary resistance to vitamin D may respond to high dosages of 1,25 $(OH)_2D_3$, but some patients require parenteral calcium infusions [113]. These heritable disorders are discussed in detail in Chapter 4.11.

Other Causes of Hypocalcaemia

Hypocalcaemia may also result from rapid loading of phosphate into the circulation. This phenomenon occurs in settings of tissue destruction, such as in **tumour lysis syndrome** observed during early phases of chemotherapy of large solid tumours. **Rhabdomyolysis** may decrease serum calcium levels for similar reasons. Several cases of hypocalcaemia and seizures have occurred following high-dose administration of phosphate either by enema or by the oral route. The use of phosphate-based cathartics in infants and small children is contraindicated. We are aware of a case in which severe hyperphosphataemia and hypocalcaemia occurred repetitively in a small child surreptitiously administered oral phosphate by her mother. Hypocalcaemia may also occur in the setting of pancreatitis due to precipitation of calcium-containing salts in the inflamed pancreatic tissue and it often correlates with the severity of the episode. Children with acute or chronic renal failure will also develop mild hypocalcaemia, which is due to a multitude of precipitating factors, such as hyperphosphataemia, and decreased 1-alpha hydroxylation of 25-hydroxyvitamin D. Hypocalcaemia has been associated with high volume (1.5 L or more per week) of soft drinks containing phosphoric acid [114]. Hypocalcaemia solely due to **low dietary calcium** intake has also been reported [115]. This syndrome has occurred in areas of South Africa in areas where food content is relatively low in calcium, and rickets is the usual presenting feature. The combination of dietary vitamin D and calcium deficiency is more commonly seen in North American children (see earlier).

Medications typically associated with the development of hypocalcaemia include bisphosphonates and aminoglycoside antibiotics. The use of intravenous bisphosphonates may precipitate hypocalcaemia in children although this is not a frequent occurrence. The effect of reduction in osteoclast activity with this class of compounds results in limitations of movement of calcium in skeletal tissue to the circulation. The practice of providing calcium supplementation for several days following an infusion is commonly used to abrogate the risk of hypocalcaemia. In small children we suggest that oral supplementation of 50 mg/kg of elemental calcium per day be provided in 2–3 dividend doses for 4–5 days following pamidronate or zoledronate infusions.

Aminoglycosides and cisplatin may affect renal tubular function and thereby result in hypocalcaemia due to renal tubular losses. Hypocalcaemia may occur in the setting of critical illness due to multiple mechanisms as described earlier. Transient hypocalcaemia often occurs post-thyroidectomy, due to manipulation of the parathyroid glands with disruption in their reverse function. Finally, the 'Hungry Bone syndrome' has been observed in the setting of treatment of severe nutritional rickets. This phenomenon is attributed to rapid uptake of calcium in the undermineralized skeleton following application of vitamin D therapy in the absence of providing concomitant calcium supplementation. Hungry Bone syndrome may also occur following parathyroidectomy when sudden removal of a parathyroid adenoma in a high bone turnover state reduces bone resorption in the setting where the remaining parathyroid glands have been chronically suppressed and unable to rapidly respond to the resultant hypocalcaemia. Continuous infusion of calcium may be necessary for several days in this clinical situation.

Disorders of Hypercalcaemia

Persistent hypercalcaemia is usually attributed to some combination of the following mechanisms: (1) excessive intestinal absorption of calcium; (2) excessive bone resorption of mineral; and (3) abnormal renal retention of calcium. Infants are usually asymptomatic with mild to moderate hypercalcaemia (11.0–13.0 mg/dl).

More severe hypercalcaemia may lead to failure to thrive, poor feeding, abdominal pain, constipation, hypotonia, vomiting, seizures, lethargy, polyuria, and hypertension. Hypercalcaemia is discussed in detail in Chapter 4.2. Congenital causes of hypercalcaemia are more frequent in children than acquired causes such as malignancy, which are more common in adults. Several syndromes with specific childhood features are described next.

Severe neonatal hyperparathyroidism is a rare condition presenting with hypercalcaemic symptoms in the first few days of life. Serum calcium levels may range as high as 15–30 mg/dl. The serum phosphate level is usually low, and serum PTH is elevated. The hypercalcaemia is predominantly due to increased bone resorption, but elevated intestinal absorption of calcium, as well as increased renal calcium retention probably occur. Radiographs of the clavicles typically reveal features of primary hyperparathyroidism. Nephrocalcinosis may be present on ultrasonographic examination. Severe neonatal hyperparathyroidism may occur in families with **FHH**. FHH is manifest by modest asymptomatic hypercalcaemia with relative hypocalciuria and normal or slightly increased serum PTH levels. FHH is an autosomal dominant, genetically heterogeneous disorder with three clinically indistinguishable variants acting in a negative manner on parathyroid cells (FHH1-FHH3) [116–118]. FHH1 (OMIM #145980) is a loss-of-function mutation in the *CASR* gene (encoding the CaSR, a G-protein-coupled receptor) on parathyroid cells. FHH2 is due to loss-of-function mutations of *GNA11* which encodes G protein family member $G\alpha_{11}$ and FHH3 (OMIM #600740) is due to loss of function of the adaptor-related protein complex 2, sigma 1 subunit (*AP2σ1*) [117, 118]. This gene product protein forms a heterotetramer with other subunits and mediates the endocytosis of the CaSR.

Severe neonatal hyperparathyroidism has been shown to result in homozygous or compound heterozygous of loss-of-function mutations in *CASR* [119]. This disorder usually requires emergency parathyroidectomy. Intravenous bisphosphonates (such as

pamidronate) has been used for treatment while waiting for surgery [120]. Another approach has employed cinacalcet, a calcimimetic, or positive allosteric modulator of CaSR, that partially restores the sensitivity of a mutant CaSR to calcium. However, this approach is not effective in the setting where homozygous null mutations of CaSR are present [116, 121]. Severe hypercalcaemia that requires surgery has also been described in infants in FHH families that have only one mutant copy of the *CaSR* gene.

In severe **Williams syndrome** (OMIM #194050) symptoms may be present from the neonatal period, but more frequently recognized later in the first few years of life. Infantile hypercalcaemia may be a presenting feature, in addition to pre- and postnatal growth failure. Characteristic unusual facies are often present, as well as cardiovascular abnormalities (usually supravalvular aortic stenosis or peripheral pulmonic stenosis), delayed psychomotor development, and selective mental deficiency. The genetic marker of the syndrome is a deletion at chromosome 7q11.23 involving elastin and LIM-Kinase genes [122]. The serum calcium levels may range as high as 12–19 mg/dl. The hypercalcaemia usually subsides spontaneously by the age of 4 years. The pathogenesis of hypercalcaemia is uncertain, although various metabolic disturbances have been described including abnormal $1,25(OH)_2D$ production, and decreased calcitonin production [123, 124]. Treatment has traditionally consisted of placing the child on a low calcium diet, free of vitamin D. Short-term therapy with corticosteroids has been applied, however a more recent approach employs intravenous bisphosphonate therapy, which has been quite effective in controlling hypercalcaemia in Williams syndrome patients in our hands. We usually use pamidronate at a dose of 0.25–0.5 mg/kg/dose, and have found that one to three doses have been sufficient to manage this problem permanently.

Jansen's metaphyseal chondrodysplasia (OMIM #156400) is an autosomal dominant disease with short limbs, severe hypercalcaemia, and hypercalciuria (despite normal or undetectable PTH and PTHrP levels) due to heterozygous mutations of the PTH/PTHrP receptor, causing ligand-independent activation of the receptor [125].

Hypophosphatemia increases circulating levels of $1,25(OH)_2D$ due to the decreased levels of the phosphate-regulating hormone, fibroblast growth factor 23 (FGF23). FGF23 regulates the genes that encode the vitamin D 1α- and 24-hydroxlyases (*CYP27B1* and *CYP24A1*, respectively) in reciprocal fashion, such that in the setting of low FGF23, 1α-hydroxylase activity increases and 24-hydroxylase decreases. This results in higher circulating levels of $1,25(OH)_2D$. This phenomenon is evident with loss-of-function mutations in *SLC34A1* (OMIM #616963), which encodes a major renal proximal tubal phosphate transporter (NaPi-IIa), as recently identified in some forms of idiopathic infantile hypercalcaemia (IIH).

Idiopathic infantile hypercalcaemia (IIH) may also result from loss-of-function mutations in the *CYP24A1* (OMIM #616963) resulting in reduced catabolism of both 25-OHD and $1,25(OH)_2D$. Most forms of this disorder result from biallelic mutations although milder phenotypes have been attributed to heterozygous mutations. The degree of hypercalcaemia can be quite variable among children with a classic phenotype [126].

Subcutaneous fat necrosis is a self-limited disorder which presents in infancy with symptoms of hypercalcaemia and erythematous or violaceous discoloration of the skin and firm indurated tissue palpable underneath. There is often a history of birth trauma or asphyxia. It is thought that the insult to the fat tissue causes cellular necrosis and infiltration of mononuclear cells and macrophages which may become granulomatous in nature. Small calcified areas may be present. Increased production of $1,25(OH)_2D$ by the inflammatory cells has been considered in some cases as the cause of hypercalcaemia [127]. Thus, a vitamin D-free, low calcium diet and glucocorticoids have traditionally been used to treat this disorder. However, like with Williams syndrome, dietary restriction of calcium intake alone may not be sufficient to correct the serum calcium level, indicating that a component of bone resorption must play a role in the development of hypercalcaemia. Therefore, we have used bisphosphonates (e.g. pamidronate) to successfully control persistent hypercalcaemia in this condition. Often a single dose is sufficient.

Intoxication with vitamin D or vitamin A should be excluded in the older infant with hypercalcaemia [128–131]. In vitamin D intoxication, it is important to measure the circulating level of 25-OHD, the most abundant circulating vitamin D metabolite. Levels of $1,25(OH)_2D$ are usually low. Toxicity may be mediated by the overwhelming large dosages of 25-OHD interacting with the VDR. Alternatively because 25-OHD has a far greater affinity than $1,25(OH)_2D$ for the circulating vitamin D binding protein (DBP), it has been proposed that the latter, more active metabolite is displaced from DBP, with toxicity resulting from the increase in free levels of $1,25(OH)_2D$. Excess intestinal absorption of calcium is present, and there may also be evidence for increased bone resorption. Elevated circulating $1,25(OH)_2D$ levels can cause hypercalcaemia but are not usually elevated in typical cases of vitamin D intoxication. Malignant and granulomatous diseases may cause extrarenal synthesis of $1,25(OH)_2D$ and resultant hypercalcaemia may ensue. Likewise, endogenous overproduction of $1,25(OH)_2D$ has been described in twins with **cat-scratch disease** induced granulomata [132]. Vitamin A intoxication results in bone pain, hypercalcaemia, headache, pseudotumor cerebri, and a characteristic erythematous skin rash with exfoliation. Alopecia and profuse ear discharge may be present. The hypercalcaemia is thought to be mediated by osteoclastic bone resorption [133]. Although toxicity is not thought to occur when less than 50 000 units of vitamin A or equivalent is ingested on a daily basis, reports of toxicity with less ingestion have been recorded in children [134]. Unrecognized liver disease may decrease the tolerance of vitamin A. In order to establish the diagnosis of vitamin A intoxication, serum retinyl ester levels should be determined in addition to the more common test for serum retinol.

Other conditions in which hypercalcaemia may be manifest in children include Down's syndrome, renal tubular acidosis, and osteogenesis imperfecta. Indeed, we have observed mild elevations in serum calcium during infancy in association with a variety of skeletal dysplasias. This appears to be a transient phenomenon. Endocrinopathies including pheochromocytoma, Addison's disease [135], thyrotoxicosis [136], and severe congenital hypothyroidism [137, 138] may be associated with development of hypercalcaemia.

Other major causes of hypercalcaemia include those commonly encountered in adults: immobilization, malignancy, and acquired hyperparathyroidism including parathyroid adenomas.

Treatment of Hypercalcaemia

Acute: The medical management of acute symptomatic hypercalcaemia consists of the administration of intravenous saline.

Intravenous infusion of bisphosphonates has also been useful in this setting. Specific long-term therapy depends on the specific hypercalcaemic disorder.

The use of bisphosphonate therapy in children has increased in recent years, both when associated with childhood cancers [139], and in the treatment of hypercalcaemia in Williams syndrome and subcutaneous fat necrosis, as noted earlier. Side effects are minimal and that the therapy appears to be safe. One should be aware of the potential complications of electrolyte disturbances, particularly hypocalcaemia, hypophosphataemia, and hypomagnesaemia.

One interesting approach in patients with IIH due to biallelic mutations of *CYP24A1* employs daily rifampin administration. The approach relies on the capacity of rifampin to induce overexpression of *CYP3A4*, which encodes an enzyme serving as an alternate mechanism for vitamin D inactivation in patients who lack *CYP24A1* function [140].

REFERENCES

1. Marks SC, Hermey DC. The structure and development of bone. In: Bilezikian JP, Raisz LG, Rodan GA, eds. *Principles of Bone Biology*. San Diego, CA: Academic Press, 1996: 3–14.
2. Vortkamp A, Lee K, Lanske B, Segre GV, Kronenberg HM, Tabinet CJ. Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science*, 1999; **273**: 613–22.
3. Gertner JM, Coustan DR, Kliger AS, Mallette LE, Ravin N, Broaduset AE. Pregnancy as state of physiologic absorptive hypercalciuria. *Am J Med*, 1986; **81**: 451–6.
4. Care AD. The placental transfer of calcium. *J Dev Physiol*, 1991; **15**: 253–7.
5. Moniz CF, Nicolaides KH, Tzannatos C, Rodeck CH. Calcium homeostasis in second trimester fetuses. *J Clin Pathol*, 1986; **39**: 838–41.
6. Simmonds CS, Karsenty G, Karaplis AC, Kovacs CS. Parathyroid hormone regulates fetal-placental mineral homeostasis. *J Bone Miner Res*, 2010; **25**: 594–605.
7. Kovacs CS. Calcium, phosphorus, and bone metabolism in the fetus and newborn. *Early Hum Dev*, 2015; **91**: 623–8.
8. Kovacs CS, Lanske B, Hunzelman JL, Guo J, Karaplis AC, Kronenberget HM. Parathyroid hormone-related peptide (PTHrP) regulates fetal-placental calcium transport through a receptor distinct from the PTH/PTHrP receptor. *Proc Natl Acad Sci U S A*, 1996; **93**: 15233–8.
9. Kovacs CS, Kronenberg HM. Maternal-fetal calcium and bone metabolism during pregnancy, puerperium, and lactation. *Endocr Rev*, 1997; **18**: 832–72.
10. Karaplis AC, Luz A, Glowacki J, *et al*. Lethal skeletal dysplasia from targeted disruption of the parathyroid hormone-related peptide gene. *Genes Dev*, 1994. **8**: 277–89.
11. Kovacs CS, Chafe LL, Fudge NJ, Friel JK, Manley NR. PTH regulates fetal blood calcium and skeletal mineralization independently of PTHrP. *Endocrinology*, 2001; **142**: 4983–93.
12. Kovacs CS, Ho-Pao CL, Hunzelman JL, *et al*. Regulation of murine fetal-placental calcium metabolism by the calcium-sensing receptor. *J Clin Invest*, 1998; **101**: 2812–20.
13. Ohata Y, Arahori H, Namba N, *et al*. Circulating levels of soluble alpha-Klotho are markedly elevated in human umbilical cord blood. *J Clin Endocrinol Metab*, 2011; **96**: E943–7.
14. Takaiwa M, Aya K, Miyai T, *et al*. Fibroblast growth factor 23 concentrations in healthy term infants during the early postpartum period. *Bone*, 2010; **47**: 256–62.
15. Zanchetta JR, Plotkin H, Alvarez Filgueira ML. Bone mass in children: normative values for the 2–20-year-old population. *Bone*, 1995; **16**(4 Suppl): 393s–9s.
16. Gordon CM, Bachrach LK, Carpenter TO, *et al*. Dual energy X-ray absorptiometry interpretation and reporting in children and adolescents: the 2007 ISCD Pediatric Official Positions. *J Clin Densitom*, 2008; **11**: 43–58.
17. Crabtree NJ, Arabi A, Bachrach LK, *et al*. Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: the revised 2013 ISCD Pediatric Official Positions. *J Clin Densitom*, 2014. **17**: 225–42.
18. Zemel BS, Kalkwarf HJ, Gilsanz V, *et al*. Revised reference curves for bone mineral content and areal bone mineral density according to age and sex for black and non-black children: results of the bone mineral density in childhood study. *J Clin Endocrinol Metab*, 2011; **96**: 3160–9.
19. Calvo MS, Eyre DR, Gundberg CM. Molecular basis and clinical application of biological markers of bone turnover. *Endocr Rev*, 1996; **17**: 333–68.
20. Rauchenzauner M, Schmid A, Heinz-Erian P, *et al*. Sex- and age-specific reference curves for serum markers of bone turnover in healthy children from 2 months to 18 years. *J Clin Endocrinol Metab*, 2007; **92**: 443–9.
21. van der Sluis IM, Hop WC, van Leeuwen J, *et al*. A cross-sectional study on biochemical parameters of bone turnover and vitamin d metabolites in healthy Dutch children and young adults. *Horm Res*, 2002; **57**: 170–9.
22. Institute of Medicine Standing Committee on the Scientific Evaluation of Dietary Reference Intakes. In: The National Academies Collection: Reports funded by National Institutes of Health. *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academies Press (US) National Academy of Sciences, 1997.
23. Iuliano-Burns S, Hopper J, Seeman E. The age of puberty determines sexual dimorphism in bone structure: a male/female co-twin control study. *J Clin Endocrinol Metab*, 2009; **94**: 1638–43.
24. Hillman LS, *et al*. Vitamin D metabolism, mineral homeostasis, and bone mineralization in term infants fed human milk, cow milk-based formula, or soy-based formula. *J Pediatr*, 1988; **112**: 864–74.
25. Hoppe B, Kemper MJ. Diagnostic examination of the child with urolithiasis or nephrocalcinosis. *Pediatr Nephrol*, 2010; **25**: 403–13.
26. Walton RJ, Bijvoet OL. Nomogram for derivation of renal threshold phosphate concentration. *Lancet*, 1975; **2**: 309–10.
27. Alon U, Hellerstein S. Assessment and interpretation of the tubular threshold for phosphate in infants and children. *Pediatr Nephrol*, 1994; **8**: 250–1.
28. Carpenter TO, Herreros F, Zhang JH, *et al*. Demographic, dietary, and biochemical determinants of vitamin D status in inner-city children. *Am J Clin Nutr*, 2012; **95**: 137–46.
29. Högler W. Complications of vitamin D deficiency from the foetus to the infant: one cause, one prevention, but who's responsibility? *Best Pract Res Clin Endocrinol Metab*, 2015; **29**: 385–98.
30. Zweier C, Sticht H, Aydin-Yaylagül I, Campbell CE, Rauch A. Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions. *Am J Hum Genet*, 2007; **80**: 510–17.

31. Daw SC, Taylor C, Kraman M, *et al.* A common region of 10p deleted in DiGeorge and velocardiofacial syndromes. *Nat Genet*, 1996; **13**: 458–60.
32. Schuffenhauer S, Lichtner P, Peykar-Derakhshandeh P, *et al.* Deletion mapping on chromosome 10p and definition of a critical region for the second DiGeorge syndrome locus (DGS2). *Eur J Hum Genet*, 1998; **6**: 213–25.
33. Craigen WJ, Lindsay EA, Bricker JT, Hawkins EP, Baldini A. Deletion of chromosome 22q11 and pseudohypoparathyroidism. *Am J Med Genet*, 1997; **72**: 63–5.
34. Ding C, Buckingham B, Levine MA. Familial isolated hypoparathyroidism caused by a mutation in the gene for the transcription factor GCMB. *J Clin Invest*, 2001; **108**: 1215–20.
35. Arnold A, Horst SA, Gardella TJ, Baba H, Levine MA, Kronenberg HM. Mutation of the signal peptide-encoding region of the preproparathyroid hormone gene in familial isolated hypoparathyroidism. *J Clin Invest*, 1990; **86**: 1084–7.
36. Thakker RV. Genetics of endocrine and metabolic disorders: parathyroid. *Rev Endocr Metab Disord*, 2004; **5**: 37–51.
37. Thomee C, Schubert SW, Parma J, *et al.* GCMB mutation in familial isolated hypoparathyroidism with residual secretion of parathyroid hormone. *J Clin Endocrinol Metab*, 2005; **90**: 2487–92.
38. Baumber L, Tufarelli C, Patel S, *et al.* Identification of a novel mutation disrupting the DNA binding activity of GCM2 in autosomal recessive familial isolated hypoparathyroidism. *J Med Genet*, 2005; **42**: 443–8.
39. Canaff L, Zhou X, Mosesova I, Cole DEC, Hendy GN. Glial cells missing-2 (GCM2) transactivates the calcium-sensing receptor gene: effect of a dominant-negative GCM2 mutant associated with autosomal dominant hypoparathyroidism. *Hum Mutat*, 2009; **30**: 85–92.
40. Bowl MR, Nesbit MA, Harding B, *et al.* An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. *J Clin Invest*, 2005; **115**: 2822–31.
41. Van Esch H, Groenen P, Nesbit MA, *et al.* GATA3 haploinsufficiency causes human HDR syndrome. *Nature*, 2000; **406**: 419–22.
42. Ali A, Christie PT, Grigorieva IV, *et al.* Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet*, 2007; **16**: 265–75.
43. Ferraris S, Del Monaco AG, Garelli E, *et al.* HDR syndrome: a novel 'de novo' mutation in GATA3 gene. *Am J Med Genet A*, 2009; **149a**: 770–5.
44. Parvari R, Diaz GA, Hershkovitz E. Parathyroid development and the role of tubulin chaperone E. *Horm Res*, 2007; **67**: 12–21.
45. Padidela R, Kelberman D, Press M, *et al.* Mutation in the TBCE gene is associated with hypoparathyroidism-retardation-dysmorphism syndrome featuring pituitary hormone deficiencies and hypoplasia of the anterior pituitary and the corpus callosum. *J Clin Endocrinol Metab*, 2009; **94**: 2686–91.
46. Ratbi I, Lyahyai J, Kabiri M, *et al.* The Bedouin mutation c.155-166del of the TBCE gene in a patient with Sanjad-Sakati syndrome of Moroccan origin. *Ann Saudi Med*, 2015; **35**: 170–2.
47. Parvari R, Hershkovitz E, Grossman N, *et al.* Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. *Nat Genet*, 2002; **32**: 448–52.
48. Isojima T, Doi K, Mitsui J, *et al.* A recurrent de novo FAM111A mutation causes Kenny-Caffey syndrome type 2. *J Bone Miner Res*, 2014; **29**: 992–8.
49. Labarthe F, Benoist JF, Brivet M, *et al.* Partial hypoparathyroidism associated with mitochondrial trifunctional protein deficiency. *Eur J Pediatr*, 2006; **165**: 389–91.
50. Shoback D. Clinical practice. Hypoparathyroidism. *N Engl J Med*, 2008; **359**: 391–403.
51. Clarke BL, Brown EM, Collins MT, *et al.* Epidemiology and diagnosis of hypoparathyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 2284–99.
52. Pollak MR, Brown EM, Estep HL, *et al.* Autosomal dominant hypocalcaemia caused by a Ca(2+)-sensing receptor gene mutation. *Nat Genet*, 1994. **8**: 303–7.
53. Baran N, ter Braak M, Saffrich R, *et al.* Novel activating mutation of human calcium-sensing receptor in a family with autosomal dominant hypocalcaemia. *Mol Cell Endocrinol*, 2015. **407**: 18–25.
54. Vargas-Poussou R, Huang C, Hulin P, *et al.* Functional characterization of a calcium-sensing receptor mutation in severe autosomal dominant hypocalcemia with a Bartter-like syndrome. *J Am Soc Nephrol*, 2002; **13**: 2259–66.
55. Kamiyoshi N, Nozu K, Urahama Y, *et al.* Pathogenesis of hypokalemia in autosomal dominant hypocalcemia type 1. *Clin Exp Nephrol*, 2016; **20**: 253–7.
56. Sato K, Hasegawa Y, Nakae J, *et al.* Hydrochlorothiazide effectively reduces urinary calcium excretion in two Japanese patients with gain-of-function mutations of the calcium-sensing receptor gene. *J Clin Endocrinol Metab*, 2002; **87**: 3068–73.
57. Mittelman SD, Hendy GN, Fefferman RA, *et al.* A hypocalcemic child with a novel activating mutation of the calcium-sensing receptor gene: successful treatment with recombinant human parathyroid hormone. *J Clin Endocrinol Metab*, 2006; **91**: 2474–9.
58. Roszko KL, Bi RD, Mannstadt M. Autosomal dominant hypocalcemia (hypoparathyroidism) types 1 and 2. *Front Physiol*, 2016; **7**: 458.
59. Betterle C, Greggio NA, Volpato M. Clinical review 93: autoimmune polyglandular syndrome type 1. *J Clin Endocrinol Metab*, 1998; **83**: 1049–55.
60. Husebye ES, Anderson MS, Kampe O. Autoimmune polyendocrine syndromes. *N Engl J Med*, 2018; **378**: 1132–41.
61. Alimohammadi M, Björklund P, Hallgren A, *et al.* Autoimmune polyendocrine syndrome type 1 and NALP5, a parathyroid autoantigen. *N Engl J Med*, 2008; **358**: 1018–28.
62. Perheentupa J. Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *J Clin Endocrinol Metab*, 2006; **91**: 2843–50.
63. Shikama N, Nusspaumer G, Hollander GA. Clearing the AIRE: on the pathophysiological basis of the autoimmune polyendocrinopathy syndrome type-1. *Endocrinol Metab Clin North Am*, 2009; **38**: 273–88, vii.
64. Li D, Streeten EA, Chan A, *et al.* Exome sequencing reveals mutations in aire as a cause of isolated hypoparathyroidism. *J Clin Endocrinol Metab*, 2017; **102**: 1726–33.
65. Brown EM. Anti-parathyroid and anti-calcium sensing receptor antibodies in autoimmune hypoparathyroidism. *Endocrinol Metab Clin North Am*, 2009; **38**: 437–45, x.
66. Mayer A, Ploix C, Orgiazzi J, *et al.* Calcium-sensing receptor autoantibodies are relevant markers of acquired hypoparathyroidism. *J Clin Endocrinol Metab*, 2004; **89**: 4484–8.

67. Gavalas NG, Kemp EH, Krohn KJE, *et al*. The calcium-sensing receptor is a target of autoantibodies in patients with autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*, 2007; **92**: 2107–14.
68. Kemp EH, Gavalas NG, Krohn KJE, *et al*. Activating autoantibodies against the calcium-sensing receptor detected in two patients with autoimmune polyendocrine syndrome type 1. *J Clin Endocrinol Metab*, 2009; **94**: 4749–56.
69. Harris HE, Kemp EH, Brown EM, *et al*. First report of anti-calcium-sensing receptor antibodies in a patient with Sjögren's syndrome and primary hypoparathyroidism. *Rheumatology (Oxford)*, 2011; **50**: 1173–5.
70. Tomar N, Gupta N, Goswami R. Calcium-sensing receptor autoantibodies and idiopathic hypoparathyroidism. *J Clin Endocrinol Metab*, 2013; **98**: 3884–91.
71. Kemp EH, Habibullah M, Kluger N, *et al*. Prevalence and clinical associations of calcium-sensing receptor and NALP5 autoantibodies in Finnish APECED patients. *J Clin Endocrinol Metab*, 2014; **99**: 1064–71.
72. Francis GL, Waguespack SG, Bauer AJ, *et al*. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2015; **25**: 716–59.
73. Gentileschi P, Gacek IA, Manzelli A, *et al*. Early (1 hour) postoperative parathyroid hormone (PTH) measurement predicts hypocalcaemia after thyroidectomy: a prospective case-control single-institution study. *Chir Ital*, 2008; **60**: 519–28.
74. Lim JP, Irvine R, Bugis S, Holmes D, Wiseman SM. Intact parathyroid hormone measurement 1 hour after thyroid surgery identifies individuals at high risk for the development of symptomatic hypocalcemia. *Am J Surg*, 2009; **197**: 648–53; discussion 653–4.
75. Grodski S, Lundgren CI, Sidhu S, Sywak M, Delbridge L. Postoperative PTH measurement facilitates day 1 discharge after total thyroidectomy. *Clin Endocrinol (Oxf)*, 2009; **70**: 322–5.
76. Grodski S, Serpell J. Evidence for the role of perioperative PTH measurement after total thyroidectomy as a predictor of hypocalcemia. *World J Surg*, 2008; **32**: 1367–73.
77. Bellantone R, Lombardi CP, Raffaelli M, *et al*. Is routine supplementation therapy (calcium and vitamin D) useful after total thyroidectomy? *Surgery*, 2002; **132**: 1109–12; discussion 1112–13.
78. Roh JL, Park CI. Routine oral calcium and vitamin D supplements for prevention of hypocalcemia after total thyroidectomy. *Am J Surg*, 2006; **192**: 675–8.
79. Gamberini MR, De Sanctis V, Gilli G. Hypogonadism, diabetes mellitus, hypothyroidism, hypoparathyroidism: incidence and prevalence related to iron overload and chelation therapy in patients with thalassaemia major followed from 1980 to 2007 in the Ferrara Centre. *Pediatr Endocrinol Rev*, 2008; **6 Suppl 1**: 158–69.
80. Carpenter TO, Carnes DL Jr, Anast CS. Hypoparathyroidism in Wilson's disease. *N Engl J Med*, 1983; **309**: 873–7.
81. Anast CS, Mohs JM, Kaplan SL, Burns TW. Evidence for parathyroid failure in magnesium deficiency. *Science*, 1972; **177**: 606–8.
82. Shalev H, Phillip M, Galil A, Carmi R, Landau D. Clinical presentation and outcome in primary familial hypomagnesaemia. *Arch Dis Child*, 1998; **78**: 127–30.
83. Schlingmann KP, Weber S, Peters M, *et al*. Hypomagnesemia with secondary hypocalcemia is caused by mutations in TRPM6, a new member of the TRPM gene family. *Nat Genet*, 2002; **31**: 166–70.
84. Simon DB, Lu Y, Choate KA, *et al*. Paracellin-1, a renal tight junction protein required for paracellular Mg2+ resorption. *Science*, 1999; **285**: 103–6.
85. Konrad M, Schaller A, Seelow D, *et al*. Mutations in the tight-junction gene claudin 19 (CLDN19) are associated with renal magnesium wasting, renal failure, and severe ocular involvement. *Am J Hum Genet*, 2006; **79**: 949–57.
86. Schlingmann KP, Konrad M, Seyberth HW. Genetics of hereditary disorders of magnesium homeostasis. *Pediatr Nephrol*, 2004; **19**: 13–25.
87. al-Harbi NN, Domrongkitchaiporn S, Lirenman DS. Hypocalcemia and hypomagnesemia after ibuprofen overdose. *Ann Pharmacother*, 1997; **31**: 432–4.
88. Klasaer AE, Scalzo AJ, Blume C, *et al*. Marked hypocalcemia and ventricular fibrillation in two pediatric patients exposed to a fluoride-containing wheel cleaner. *Ann Emerg Med*, 1996; **28**: 713–18.
89. Klein GL, Nicolai M, Langman CB, Cuneo BF, Sailer DE, Herndon DN. Dysregulation of calcium homeostasis after severe burn injury in children: possible role of magnesium depletion. *J Pediatr*, 1997; **131**: 246–51.
90. Saraff V, Rothenbuhler A, Högler W, Linglart A. Continuous subcutaneous recombinant parathyroid hormone (1-34) infusion in the management of childhood hypoparathyroidism associated with malabsorption. *Horm Res Paediatr*, 2018; **89**: 271–7.
91. Mishra PE, Schwartz BL, Sarafoglou K, *et al*. Short-term PTH(1-34) therapy in children to correct severe hypocalcemia and hyperphosphatemia due to hypoparathyroidism: two case studies. *Case Rep Endocrinol*, 2016; **2016**: 6838626.
92. Winer KK, Fulton KA, Albert PS, Cutler GB Jr. Effects of pump versus twice-daily injection delivery of synthetic parathyroid hormone 1-34 in children with severe congenital hypoparathyroidism. *J Pediatr*, 2014; **165**: 556–63.e1.
93. Winer KK, Sinaii N, Reynolds J, *et al*. Long-term treatment of 12 children with chronic hypoparathyroidism: a randomized trial comparing synthetic human parathyroid hormone 1-34 versus calcitriol and calcium. *J Clin Endocrinol Metab*, 2010; **95**: 2680–8.
94. Winer KK, Sinaii N, Peterson D, Sainz B Jr, Cutler GB Jr. Effects of once versus twice-daily parathyroid hormone 1-34 therapy in children with hypoparathyroidism. *J Clin Endocrinol Metab*, 2008; **93**: 3389–95.
95. Gafni RI, Brahim JS, Andreopoulou P, *et al*. Daily parathyroid hormone 1-34 replacement therapy for hypoparathyroidism induces marked changes in bone turnover and structure. *J Bone Miner Res*, 2012; **27**: 1811–20.
96. Winer KK, Zhang B, Shrader JA, *et al*. Synthetic human parathyroid hormone 1-34 replacement therapy: a randomized crossover trial comparing pump versus injections in the treatment of chronic hypoparathyroidism. *J Clin Endocrinol Metab*, 2012; **97**: 391–9.
97. Theman TA, Collins MT, Dempster DW, *et al*. PTH(1-34) replacement therapy in a child with hypoparathyroidism caused by a sporadic calcium receptor mutation. *J Bone Miner Res*, 200; **24**: 964–73.
98. Dong B, Endo I, Ohnishi Y, *et al*. Calcilytic ameliorates abnormalities of mutant calcium-sensing receptor (CaSR) knock-in mice mimicking autosomal dominant hypocalcemia (ADH). *J Bone Miner Res*, 2015; **30**: 1980–93.
99. Tafaj O, Juppner H. Pseudohypoparathyroidism: one gene, several syndromes. *J Endocrinol Invest*, 2017; **40**: 347–56.
100. Levine MA, Downs RW Jr, Moses AM, *et al*. Resistance to multiple hormones in patients with pseudohypoparathyroidism.

Association with deficient activity of guanine nucleotide regulatory protein. *Am J Med*, 1983; **74**: 545–6.
101. Weinstein LS, Yu S, Warner DR, Liu J. Endocrine manifestations of stimulatory G protein alpha-subunit mutations and the role of genomic imprinting. *Endocr Rev*, 2001; **22**: 675–705.
102. Liu J, Litman D, Rosenberg MJ, Yu S, Biesecker LG, Weinstein LS. A GNAS1 imprinting defect in pseudohypoparathyroidism type IB. *J Clin Invest*, 2000; **106**: 1167–74.
103. Haldeman-Englert CR, Hurst ACE, Levine MS. Disorders of GNAS inactivation. In: Pagon RA, Adam MP, Ardinger HH, *et al.*, eds. Seattle, WA: GeneReviews(R), 2018.
104. Giovanna M, Francesca Marta E. Multiple hormone resistance and alterations of G-protein-coupled receptors signaling. *Best Pract Res Clin Endocrinol Metab*, 2018; **32**: 141–54.
105. Holick MF, Binkley NC, Bischoff-Ferrari HA, *et al*. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 1911–30.
106. Maalouf NM. The noncalciotropic actions of vitamin D: recent clinical developments. *Curr Opin Nephrol Hypertens*, 2008; **17**: 408–15.
107. Jorde R, Grimnes G. Vitamin D and metabolic health with special reference to the effect of vitamin D on serum lipids. *Prog Lipid Res*, 2011; **50**: 303–12.
108. Al Mheid I, Quyyumi AA. Vitamin D and cardiovascular disease: controversy unresolved. *J Am Coll Cardiol*, 2017; **70**: 89–100.
109. Censani M, Hammad HT, Christos PJ, Schumaker T. Vitamin D deficiency associated with markers of cardiovascular disease in children with obesity. *Glob Pediatr Health*, 2018; **5**: 2333794X17751773.
110. Adams JS, Ren S, Liu PT, *et al*. Vitamin D-directed rheostatic regulation of monocyte antibacterial responses. *J Immunol*, 2009; **182**: 4289–95.
111. Lips P. Which circulating level of 25-hydroxyvitamin D is appropriate? *J Steroid Biochem Mol Biol*, 2004; 89–90: 611–14.
112. Clements MR, Johnson L, Fraser DR. A new mechanism for induced vitamin D deficiency in calcium deprivation. *Nature*, 1987; **325**: 62–5.
113. Balsan S, Garabédian M, Larchet M, *et al*. Long-term nocturnal calcium infusions can cure rickets and promote normal mineralization in hereditary resistance to 1,25-dihydroxyvitamin D. *J Clin Invest*, 1986; **77**: 1661–7.
114. Mazariegos-Ramos E, Guerrero-Romero F, Rodríguez-Morán M, Lazcano-Burciaga G, Paniagua R, Amato D. Consumption of soft drinks with phosphoric acid as a risk factor for the development of hypocalcemia in children: a case-control study. *J Pediatr*, 1995; **126**: 940–2.
115. Marie PJ, Pettifor JM, Ross FP, Glorieux FH. Histological osteomalacia due to dietary calcium deficiency in children. *N Engl J Med*, 1982; **307**: 584–8.
116. Stokes VJ, Nielsen MF, Hannan FM, Thakker RV. Hypercalcemic disorders in children. *J Bone Miner Res*, 2017; **32**: 2157–70.
117. Nesbit MA, Hannan FM, Howles SA, *et al*. Mutations affecting G-protein subunit alpha11 in hypercalcemia and hypocalcemia. *N Engl J Med*, 2013; **368**: 2476–86.
118. Nesbit MA, Hannan FM, Howles SA, *et al*. Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3. *Nat Genet*, 2013; **45**: 93–7.
119. Pollak MR, Brown EM, Chou YH, *et al*. Mutations in the human Ca(2+)-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell*, 1993; **75**: 1297–303.
120. Wilhelm-Bals A, Parvex P, Magdelaine C, Girardin E. Successful use of bisphosphonate and calcimimetic in neonatal severe primary hyperparathyroidism. *Pediatrics*, 2012; **129**: e812–16.
121. Garcia Soblechero E, Ferrer Castillo MT, Jiménez Crespo B, *et al*. Neonatal hypercalcemia due to a homozygous mutation in the calcium-sensing receptor: failure of cinacalcet. *Neonatology*, 2013; **104**: 104–8.
122. Frangiskakis JM, Ewart AK, Morris CA, *et al*. LIM-kinase1 hemizygosity implicated in impaired visuospatial constructive cognition. *Cell*, 1996; **86**: 59–69.
123. Palacios-Verdu MG, Segura-Puimedon M, Borralleras C, *et al*. Metabolic abnormalities in Williams-Beuren syndrome. *J Med Genet*, 2015; **52**: 248–55.
124. Kim YM, Hyang Cho J, Kang E, *et al*. Endocrine dysfunctions in children with Williams-Beuren syndrome. *Ann Pediatr Endocrinol Metab*, 2016; **21**: 15–20.
125. Schipani E, Langman C, Hunzelman J, *et al*. A novel parathyroid hormone (PTH)/PTH-related peptide receptor mutation in Jansen's metaphyseal chondrodysplasia. *J Clin Endocrinol Metab*, 1999; **84**: 3052–7.
126. Pronicka E, Ciara E, Halat P, *et al*. Biallelic mutations in CYP24A1 or SLC34A1 as a cause of infantile idiopathic hypercalcemia (IIH) with vitamin D hypersensitivity: molecular study of 11 historical IIH cases. *J Appl Genet*, 2017; **58**: 349–53.
127. Farooque A, Moss C, Zehnder D, Hewison M, Shaw NJ. Expression of 25-hydroxyvitamin D3-1alpha-hydroxylase in subcutaneous fat necrosis. *Br J Dermatol*, 2009; **160**: 423–5.
128. Vanstone MB, Oberfield SE, Shader L, Ardeshirpour L, Carpenter TO. Hypercalcemia in children receiving pharmacologic doses of vitamin D. *Pediatrics*, 2012; **129**: e1060–3.
129. Nimesh M, Singh P, Jhamb U, Dubey AP. An unsuspected pharmacological vitamin D toxicity in a child and its brief review of literature. *Toxicol Int*, 2015; **22**: 167–9.
130. Wani M, Wani I, Banday K, Ashraf M. The other side of vitamin D therapy: a case series of acute kidney injury due to malpractice-related vitamin D intoxication. *Clin Nephrol*, 2016; **86**: 236–41.
131. Gorris MA, Arora H, Lieb DC, Aloi JA, *et al*. A word of caution when prescribing high-dose vitamin D. *Am J Med*, 2017; **130**: e129–30.
132. Bosch X. Hypercalcemia due to endogenous overproduction of active vitamin D in identical twins with cat-scratch disease. *JAMA*, 1998; **279**: 532–4.
133. Valentic JP, Elias AN, Weinstein GD. Hypercalcemia associated with oral isotretinoin in the treatment of severe acne. *JAMA*, 1983; **250**: 1899–900.
134. Carpenter TO, Pettifor JM, Russell RM, *et al*. Severe hypervitaminosis A in siblings: evidence of variable tolerance to retinol intake. *J Pediatr*, 1987; **111**: 507–12.
135. Vasikaran SD, Tallis GA, Braund WJ. Secondary hypoadrenalism presenting with hypercalcaemia. *Clin Endocrinol (Oxf)*, 1994; **41**: 261–4.
136. Eriksen EF, Mosekilde L, Melsen F. Trabecular bone remodeling and bone balance in hyperthyroidism. *Bone*, 1985; **6**: 421–8.
137. Tau C, Garabédian M, Farriaux JP, *et al*. Hypercalcemia in infants with congenital hypothyroidism and its relation to vitamin D and thyroid hormones. *J Pediatr*, 1986; **109**: 808–14.

138. Laufer J, Noff D, Orda S, Sack J. Effect of short-term hyperthyroxinemia on vitamin D metabolism in congenital hypothyroidism. *Horm Metab Res*, 1993; **25**: 386–8.
139. Lteif AN, Zimmerman D. Bisphosphonates for treatment of childhood hypercalcemia. *Pediatrics*, 1998; **102**(4 Pt 1): 990–3.
140. Hawkes CP, Li D, Hakonarson H, *et al*. CYP3A4 induction by rifampin: an alternative pathway for vitamin D inactivation in patients with CYP24A1 mutations. *J Clin Endocrinol Metab*, 2017; **102**: 1440–6.
141. Estey MP, Cohen AH, Colantonio DA, *et al*. CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens Clinical Chemistry Assays: direct validation using reference samples from the CALIPER cohort. *Clin Biochem*, 2013; **46**: 1197–219.
142. Bollen AM, Eyre DR. Bone resorption rates in children monitored by the urinary assay of collagen type I cross-linked peptides. *Bone*, 1994; **15**: 31–4.
143. Fares JE, Choucair M, Nabulsi M, *et al*. Effect of gender, puberty, and vitamin D status on biochemical markers of bone remodeleding. *Bone*, 2003; **33**: 242–7.
144. Crofton PM, Evans N, Taylor MRH, Holland CV. Serum CrossLaps: pediatric reference intervals from birth to 19 years of age. *Clin Chem*, 2002; **48**: 671–3.
145. Michelsen J, Wallaschofski H, Friedrich N, *et al*. Reference intervals for serum concentrations of three bone turnover markers for men and women. *Bone*, 2013; **57**: 399–404.
146. Kovacs CS. Bone development and mineral homeostasis in the fetus and neonate: roles of the calciotropic and phosphotropic hormones. *Physiol Rev*, 2014; **94**: 1143–218.

4.8

Osteoporosis

Richard Eastell

Importance

Osteoporosis affects an estimated 75 million people in Europe, the United States, and Japan combined [1]. It is a preventable and a treatable condition, yet many patients with fractures remain unrecognized and untreated.

Definition

Definitions for osteoporosis have usually been conceptual. An example was produced by a Consensus Development Conference [1] as 'a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration with a consequent increase in bone fragility and susceptibility to fracture'. This definition is elegant but difficult to apply to an individual patient. An operational definition of osteoporosis has been proposed by a Working Group of the World Health Organization (WHO) [2]. This defines osteoporosis by the patient's bone mineral density (BMD) in relation to the mean value in normal, young subjects. Specifically, osteoporosis is a value for BMD level equal to or less than 2.5 standard deviation (SD) below the mean value in young subjects (T score ≤2.5) (**Figure 4.8.1**). This definition is useful as an entry criterion to a clinical trial or as a tool to study the epidemiology of osteoporosis, but it has limitations in clinical practice. It elevates a risk factor for fracture to the status of a diagnostic criterion, it ignores the importance of other determinants of bone strength [3], it ignores higher fracture risk associated with a certain level of BMD in older women, and it does not specify the technique or the site at which BMD should be measured. Bone density results can also be compared to the mean value in normal subjects of the same age (Z score). A Z score below –0.67 would indicate a value in the lowest 25% of the reference range, a level indicating a high lifetime risk of fracture. A Z score below –2 would indicate a value in the lowest 2.5% of the reference range, a level likely to be associated with a high risk factor for osteoporosis.

It is now possible to determine individuals' risk of osteoporosis using a combination of BMD and clinical risk factors as proposed by a WHO working group [4]. A prediction algorithm is available (FRAX, which allows estimation of 10-year risk) and treatment guidance may be based on this (http://www.shef.ac.uk/FRAX). In the United Kingdom, the threshold for treatment differs by age [4]. In other countries, such as the United States, treatment is recommended in a woman with osteopaenia (BMD T score –1.0 or less, and either a 10-year risk of hip fracture of 3% or 10-year risk of major osteoporotic fractures of 20% [5].

Epidemiology

The commonest osteoporosis-related fractures are the proximal femur (hip), vertebrae (spine), and distal forearm (wrist). Hip fracture rate increases exponentially with age and is three times more common in women than men. About 50% of hip fractures occur after age 80 years. There is an excess mortality of 18% following a hip fracture. Vertebral fractures are asymptomatic in over half of cases, and so their epidemiology is more difficult to define. Morphometric approaches have been applied to large population samples. Up to 25% of women over age 50 years have vertebral fractures and the prevalence in men is similar. Most vertebral fractures in women are a consequence of mild-to-moderate trauma, but in men almost half result from severe trauma. Wrist fracture incidence increases at the time of the menopause but reaches a plateau after age 70 years. There is no increase with age in men and 85% of wrist fractures occur in women. This plateau may be explained by a cessation of bone loss after age 70 years at the wrist or to a different way of falling in older people.

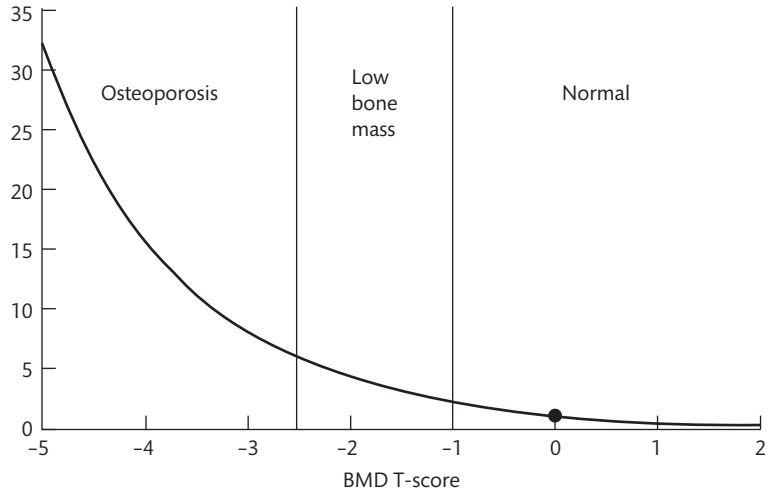


Figure 4.8.1 The risk of fracture increases by a factor of 2 for every 1–SD decrease in bone mineral density (BMD). The WHO definition of osteoporosis is based on the BMD in relation to SDs from the young normal mean (the T score). Osteoporosis is defined as a BMD that is equal to or less than 2.5 SD below the young normal mean.

These fractures carry a large cost to the Health Service, £1.7 billion in the United Kingdom. With the ageing of the population (and a secular increase in age-specific fracture incidence) the burden will increase in the future.

Pathophysiology of Postmenopausal Osteoporosis

Osteoporosis-related fractures result from a combination of decreased BMD and a deterioration in bone microarchitecture.

A BMD T score of –2.5 or less indicating osteoporosis is present in about 50% of patients presenting with wrist, spine, or hip fracture. A BMD below average for age can be considered a consequence of inadequate accumulation of bone in young adult life (low peak bone mass) or of excessive rates of bone loss. The microarchitectural changes occur in parallel with the bone loss but will be considered separately.

Determinants of Peak Bone Mass

The increase in bone mass that occurs during childhood and puberty results from a combination of growth of bone at the endplates (endochondral bone formation) and of change in bone shape (modelling) [6]. The rapid increase in bone mass at puberty is associated with an increase in sex hormone levels and the closure of the growth plates. Within 3 years of menarche, there is little further increase in bone mass. The small increase in BMD over the next 5–15 years is referred to as 'consolidation'. The resulting peak bone mass is achieved by age 20–30 years old [7].

Genetic factors are the main determinants of peak bone mass [8]. This has been shown by studies made on twins or on mother–daughter pairs. Hereditability appears to account for about 50–85% of the variance in bone mass, depending on the skeletal site. Variants in the genes encoding factors known to regulate bone mass (such as oestrogen receptor-α (ERα), vitamin D receptor, transforming growth factor-β (TGFβ), low-density lipoprotein receptor-related protein 5 (LRP5) and LRP6 (both involved in WNT signalling) and bone matrix components (e.g. collagen type I α1) were found to be associated with bone mass and fracture risk, but these variants could be used to explain only very little of the interindividual variation. The non-genetic factors include low calcium intake during childhood, low body weight at maturity and at 1 year of life, sedentary lifestyle, and delayed puberty. Each of these results in decreased bone mass.

Bone Loss

Mechanisms of Bone Loss

Bone loss occurs in the postmenopausal woman as a result of an increase in the rate of bone remodelling and an imbalance between the activity of osteoclasts and osteoblasts. Bone remodelling occurs at discrete sites within the skeleton and proceeds in an orderly fashion with bone resorption always being followed by bone formation, a phenomenon referred to as 'coupling'. In cortical and cancellous bone, the sequence of bone remodelling is similar [9]. The quiescent bone surface is converted to activity (origination) and the osteoclasts resorb bone (progression) forming a cutting cone (cortical bone) or a trench (cancellous bone). The osteoblasts synthesize bone matrix which subsequently mineralizes. The sequence takes up to 8 months. If the processes of bone resorption and bone formation are not matched then there is 'remodelling imbalance'. In postmenopausal women, this imbalance is magnified by the increase in the rate of initiation of new bone remodelling cycles (activation frequency).

Remodelling imbalance results in irreversible bone loss. There are two other causes of irreversible bone loss, referred to as 'remodelling errors'. First is excavation of overlarge haversian spaces in cortical bone [10]. Radial infilling is regulated by signals from the outermost osteocytes and is generally no more than 90 μm. Hence, large external diameters, which may simply occur randomly, lead to large central haversian canals, which then accumulate with age, leading to increased cortical porosity. In a similar way, osteoclast penetration of trabecular plates, or severing of trabecular beams, removes the scaffolding needed for osteoblastic replacement of resorbed bone. In both ways random remodelling errors tend to reduce both cancellous and cortical bone density and structural integrity.

Causes of Bone Loss

Oestrogen Deficiency

Bone loss in the postmenopausal woman occurs in two phases [11]. There is a phase of rapid bone loss that lasts for 5 years (about 3% per year in the spine). Subsequently, there is lower bone loss that is more generalized (about 0.5% per year at many sites). This slower phase of bone loss affects men, starting at about age 55 years. The rapid phase of bone loss in women is caused by oestrogen deficiency. The circulating level of oestradiol decreases by 90% at the time of the menopause. This bone loss can be prevented by the administration of oestrogen and progestins to the postmenopausal woman. It has been estimated that this rapid phase of bone loss contributes 50% to the spinal bone loss across life in women. The main effect of oestrogen deficiency is on bone, where it increases activation frequency, and may contribute to the remodelling imbalance. Oestrogen deficiency may increase bone resorption by stimulating the synthesis of *RANKL* by osteoblasts (or their precursors). *RANKL* binds to its receptor RANK on the osteoclast and promotes differentiation to osteoclasts, increases osteoclast activity, and inhibits osteoclast apoptosis. Oestrogen deficiency also increases the apoptosis of osteoblasts and osteocytes.

Oestrogen deficiency may be a determinant of bone loss in men [12]. Decreased BMD has been reported in men with an inactivating mutation of the genes for the oestrogen receptor or for aromatase (the enzyme that converts androgens to oestrogens). In older men, oestrogen levels correlate more closely with BMD than testosterone levels. In men with osteoporosis, oestradiol (but not testosterone) levels have been reported to be decreased.

Ageing

The slow phase of bone loss is attributed to age-related factors such as an increase in parathyroid hormone (PTH) levels (Figure 4.8.2) and to osteoblast senescence. An increase in PTH levels (and action) occurs in both men and women with ageing. PTH levels correlate with those of biochemical markers of bone turnover and both may be returned to those found in young adults by the intravenous infusion of calcium. The increase in PTH results from decreased renal calcium reabsorption and decreased intestinal calcium absorption. The latter may result from vitamin D deficiency (e.g. in the housebound elderly), decreased 1α-hydroxylase activity in the kidney resulting in decreased synthesis of 1,25-dihydroxyvitamin D, or resistance to vitamin D. Whatever the cause, a diet high in calcium returns both PTH and bone turnover markers to levels found in healthy young adults. It has been proposed that the age-related increase in PTH could result from indirect effects of oestrogen deficiency [11]. This proposal is based on the following evidence. In older women treated with oestrogen, [1] there is a decrease in bone turnover markers and PTH levels; [2] there is an increase in calcium absorption, possibly mediated by an increase in 1,25-dihydroxyvitamin D; [3] there is an increase in the PTH-independent calcium reabsorption in the kidney; and [4] there is a decrease in the parathyroid secretory reserve.

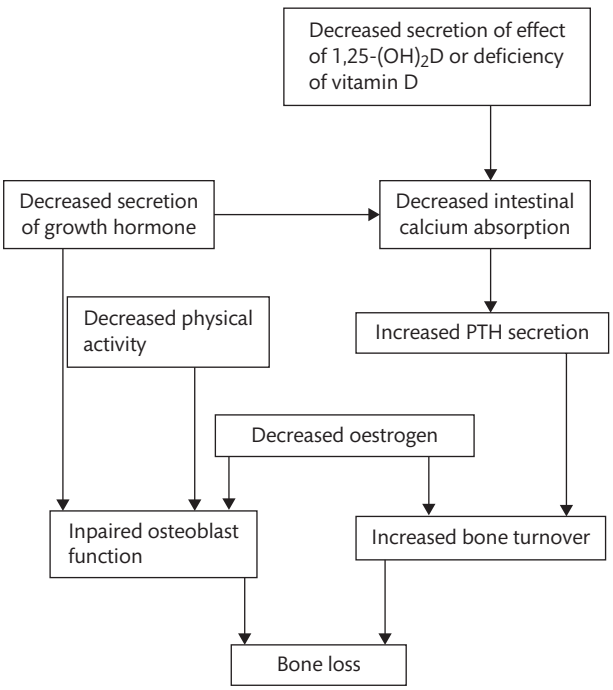


Figure 4.8.2 The causes of bone loss with ageing.

Accelerating Factors

A number of diseases and drugs are clearly related to accelerated bone loss (Box 4.8.1). Their effects are superimposed on those described just now. Thus, a patient starting on corticosteroid therapy is more likely to have an osteoporosis-related fracture if she has low BMD resulting from low peak bone mass and the accelerated bone loss of the menopause.

Identification of Mechanism of Bone Loss in an Individual

In a woman presenting with osteoporosis at age 70 years it is often possible to identify several reasons for the low BMD (Figure 4.8.3). Some of these may be identified from history taking (early menopause, drugs that accelerate bone loss), but some cannot be identified in retrospect (low peak bone mass and rapid losers).

Other Determinants of Bone Strength

Bone Geometry

Bone geometry has a major effect on fracture risk [3]. One example is hip axis length, the distance from the lateral surface of the trochanter to the inner surface of the acetabulum, along the axis of the femoral neck. Short hip axis length results in an architecturally stronger structure for any given bone density. This is probably the reason why Japanese and other Orientals have about half the hip fracture rate of Caucasians, despite similar bone density values.

Fatigue Damage

Fatigue damage consists of ultramicroscopic rents in the basic bony material, resulting from the inevitable bending that occurs when a structural member is loaded. Fatigue damage is the principal cause of failure in mechanical engineering structures; its prevention is the responsibility of the remodelling apparatus which detects and removes fatigue-damaged bone. Fractures related to fatigue damage occur whenever the damage occurs faster than remodelling can repair it or whenever the remodelling apparatus is defective. March fractures and the fractures of radiation necrosis are well-recognized examples of fractures due to these two mechanisms.

Box 4.8.1 Risk factors for osteoporosis in postmenopausal women

- Genetic factors
  - First-degree relative with low-trauma fracture (e.g. mother with hip fracture)
- Environmental factors
  - Cigarette smoking
  - Alcohol abuse
  - Physical inactivity or prolonged immobilization
  - Thin habitus (e.g. less than 57 kg)
  - Diet low in calcium (e.g. less than 500 mg/day)
  - Little exposure to sunlight (e.g. housebound elderly)
- Menstrual status
  - Early menopause, that is, before age 45 years
  - Previous amenorrhoea (e.g. anorexia nervosa, hyperprolactinaemia)
- Drug therapy
  - Glucocorticoids (e.g. 7.5 mg/day of prednisolone or more, for 6 months or more)
  - Antirejection therapy after organ transplantation (e.g. ciclosporin)
  - Antiepileptic drugs (e.g. phenytoin)
  - Excessive substitution therapy (e.g. thyroxine, hydrocortisone)
  - Anticoagulant therapy (e.g. heparin, warfarin)
  - Aromatase inhibitors and gonadotropin-releasing hormone agonist therapy for breast (and prostate) cancer
- Endocrine diseases
  - Primary hyperparathyroidism
  - Thyrotoxicosis
  - Cushing's syndrome
  - Addison's disease
- Haematological diseases
  - Multiple myeloma
  - Systemic mastocytosis
  - Lymphoma, leukaemia
  - Pernicious anaemia
- Rheumatological diseases
  - Rheumatoid arthritis
  - Ankylosing spondylitis
- Gastrointestinal diseases
  - Malabsorption states (e.g. coeliac disease, Crohn's disease, surgery for peptic ulcer)
  - Chronic liver disease (e.g. primary biliary cirrhosis)

Loss of Trabecular Connectivity

Bone structures loaded vertically, such as the vertebral bodies and femoral and tibial metaphyses, derive a substantial portion of their structural strength from a system of horizontal, cross-bracing trabeculae which support the vertical elements and limit lateral bowing and consequent snapping under vertical loading. Severance of such trabecular connections is known to occur preferentially in postmenopausal women and is considered to be an important reason for the large female/male preponderance of low-trauma vertebral fractures.

Risk Factors for Osteoporosis

A number of risk factors have been identified for osteoporosis (Box 4.8.1). Bone loss can be stopped, or reversed, if risk factors such as primary hyperparathyroidism (see Chapter 4.3) are identified and treated. The patient who presents with a vertebral fracture and low BMD is likely to have had one of four causes (Figure 4.8.3). It is impossible to know the importance of peak bone mass and the rate of bone loss in retrospect. Questions can be asked about early menopause and about the drugs and diseases known to accelerate bone loss. We usually assess risk factors by administering a questionnaire before first attendance at the clinic, by carrying out a limited biochemical work-up before the clinic visit (Box 4.8.2), and, after the clinical, evaluation exploring alternative diagnoses.

Genetic Factors

Family History

A maternal history of hip fracture increases the risk of hip fracture in an individual.

Osteogenesis Imperfecta

Late-onset forms (e.g. Sillence type I) may present with vertebral fracture. The clinical clues are the blue sclerae, hypermobile joints, lax skin, cardiac murmurs, and deafness (see Chapter 4.12).

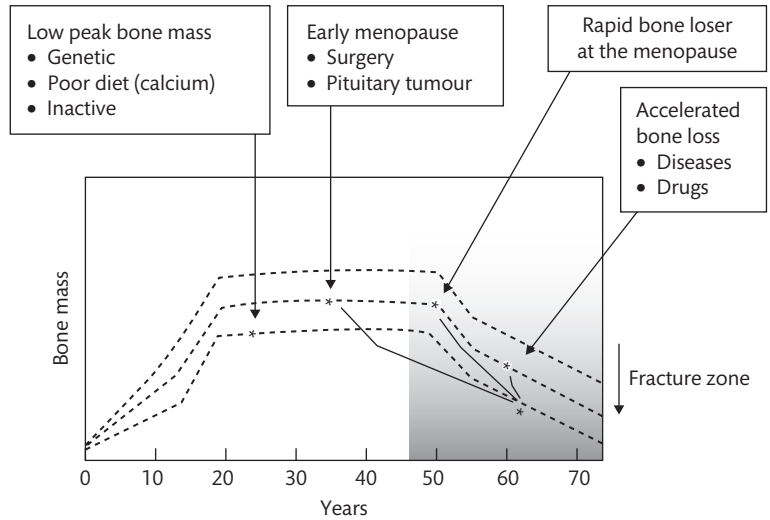


Figure 4.8.3 The possible causes of low bone mass in a 70-year-old woman. Note how peak bone mass is attained about the age of 30 years and the phase of accelerated bone loss begins at the menopause. The lower the bone density falls, the greater the risk of fracture.

Box 4.8.2 Diagnostic evaluation of osteoporosis

- Establish presence of low-trauma fracture (fall from standing height or less)
  - Spine radiographs
- Evaluate degree of bone loss
  - Bone densitometry (see Box 4.8.3)
- Laboratory tests to exclude secondary osteoporosis
  - Primary hyperparathyroidism (serum calcium)
  - Thyrotoxicosis (thyroid-stimulating hormone)
  - Multiple myeloma (erythrocyte sedimentation rate, protein electrophoresis, and urinary Bence Jones protein)
  - Osteomalacia (serum calcium, phosphate (fasting, morning), alkaline phosphatase, 24-h urinary calcium and creatinine)
  - Malabsorption syndrome (full blood count and, if necessary, red cell folate, serum vitamin B_{12}, antiendomysial antibodies, magnesium)
  - Hypogonadism in men (testosterone and, if necessary, sex hormone binding globulin, luteinizing hormone, follicle-stimulating hormone, prolactin)

Environmental Factors

Cigarette Smoking and Chronic Obstructive Pulmonary Disease

Smoking results in lower oestrogen levels and early menopause, and smokers often have a slender stature (see next). Chronic lung disease is associated with chronic respiratory acidosis and decreased physical activity.

Alcohol Abuse

The relationship between alcohol and bone loss is complex (and there may even be a protective effect at a low level of intake) [13]. Alcoholism results in low BMD because of poor nutrition and pseudo-Cushing's syndrome, and a direct suppressive effect of alcohol on osteoblasts. Fractures result from the increased propensity to fall.

Physical Inactivity and Immobilization (Neurological)

Athletes have high BMD. However, bone loss only results from complete immobilization (or space flight). The bone loss after paralysis (e.g. stroke) is regional.

Thin Habitus

This is a risk factor for fracture through decreased oestrogen production from adrenal androgens (in adipose tissue) and through decreased padding (to cushion a fall). Women with hip fracture weigh about 8 kg less than the average woman.

Diet—Low Dietary Calcium

Low dietary calcium and high dietary sodium are considered risk factors for osteoporosis. Calcium requirement increases during growth and in the postmenopausal period. A postmenopausal woman should take 1500 mg/day of calcium.

Little Exposure to Ultraviolet Light

Ultraviolet light (UVB) acts on the skin as the main source of vitamin D (see Chapter 4.11). The housebound are liable to vitamin D insufficiency. This does not result in clinical osteomalacia, but the decreased calcium absorption (see earlier) results in secondary hyperparathyroidism.

Menstrual Status

Early Menopause

A menopause before the age of 45 years is associated with increased risk of fracture. A menopause before the age of 40 years is often associated with some endocrine cause and should be investigated further.

Amenorrhoea

A late onset of the menarche and periods of amenorrhoea of any cause (e.g. exercise related, are associated with decreased bone mass later in life).

Anorexia Nervosa

This is associated with bone loss and increased risk of fracture. The bone loss is probably irreversible after 4 years of amenorrhoea. The mechanism of the bone loss is not just oestrogen deficiency. The diet is low in calcium and serum IGF-1 levels are low, and cortisol secretion may be increased.

Hyperprolactinaemia

This results in oestrogen deficiency. Not all studies have reported bone loss, and it may be that prolactin has some beneficial effects on calcium homeostasis, such as an increase in calcium absorption.

Drug Therapy

Corticosteroids

In the United Kingdom, over 250 000 patients take continuous oral glucocorticoids, yet no more than 14% receive any therapy to prevent bone loss, a serious complication of glucocorticoid treatment. Bone loss is rapid, particularly in the first year, and fracture risk may double [14]. The mechanism of the bone loss is mainly a suppression of osteoblast activity. This differs from oestrogen deficiency, in which the mechanism is mainly increased activation frequency. A treatment algorithm has been presented for adults receiving glucocorticoid doses for 6 months or more [15]. General measures (e.g. alternative glucocorticoids and routes of administration) and therapeutic interventions such as bisphosphonates, are recommended. Glucocorticoid-induced osteoporosis is discussed in greater detail in Chapter 4.12.

Antiepileptic Drugs

Phenobarbitone and phenytoin are known to affect vitamin D metabolism and result in osteomalacia. More commonly, they may cause secondary hyperparathyroidism and osteoporosis.

Excessive Substitution Therapy

Thyroxine doses sufficient to suppress thyroid-stimulating hormone, and hydrocortisone doses that result in 24-h urinary free cortisol above the reference range, have adverse effects on bone turnover and bone density (see Chapter 4.9).

Anticoagulant Drugs

Heparin stimulates bone resorption by a direct effect on osteoclasts. Its long-term use (e.g. in pregnancy) results in bone loss at the spine

and hip of 8–10% over 6 months. Warfarin may interfere with the γ-carboxylation of bone proteins, and its use is associated with an increased risk of fracture.

Endocrine Diseases

Primary Hyperparathyroidism

This is associated with an increase in bone turnover and a decrease in bone mass, particularly at sites rich in cortical bone. It is likely that there is an increase in fracture rates. These changes are reversible with surgical removal of the tumour (see Chapter 4.3).

Thyrotoxicosis

This topic is discussed in Chapter 4.9.

Cushing's Syndrome

Cushing's disease may present with vertebral fracture. The bone loss in the first few years after pituitary surgery is between 10% and 20% at the spine.

Addison's Disease

This is associated with decreased bone mass, resulting from excess substitution therapy and deficiency of adrenal androgens (precursors for oestrogen synthesis in men and postmenopausal women).

Haematological Diseases

Multiple Myeloma

This may present with vertebral fracture. It is usually identified with serum protein electrophoresis, and urinary Bence Jones testing, but occasionally the myeloma may be non-secretory and can usually be diagnosed by bone marrow examination.

Systemic Mastocytosis

This may cause decreased or increased bone density. It can be identified by urticaria pigmentosa, and mast cells are identified in the bone biopsy (see Chapter 6.15).

Pernicious Anaemia

This has been associated with low bone density and increased risk of fractures. The mechanism is unclear, as the absorption of calcium from food is normal despite the absence of gastric acid.

Rheumatological Diseases

Rheumatoid Arthritis and Ankylosing Spondylitis

The immobility may be an important cause, as may be the local (and circulating) cytokines, which promote bone resorption. The corticosteroid therapy for rheumatoid arthritis may also contribute.

Gastrointestinal Diseases

Malabsorption Syndrome

Diseases such as coeliac disease may present with osteoporosis. Other inflammatory bowel diseases, such as Crohn's disease, may require treatment with corticosteroids. Patients who have had peptic ulcer surgery have low bone density and increased risk of fracture. This may also be due to their habits—such patients are usually thin, commonly smoke, and may take excess alcohol.

Chronic Liver Disease

Chronic obstructive liver diseases, such as primary biliary cirrhosis, are associated with osteoporosis. Bilirubin has been associated with osteoblast suppression *in vitro*. Liver transplantation results in further bone loss and about one-third of patients suffer fractures [16]. This bone loss is likely to be related to the immunosuppression (corticosteroids and cyclosporine).

Clinical Presentation of Vertebral Fracture

Osteoporosis does not cause pain or deformity in the absence of fractures. Its importance lies in the fact that it greatly increases the risk of fracture, notably forearm (Colles') fracture, hip fracture, and vertebral fracture. The commonest presentation in clinical practice is vertebral fracture and so this will be described in more detail.

History of Back Pain

The back pain of vertebral fracture has some characteristic features. The pain often comes on within a day of some strain on the back, such as lifting a suitcase or a grandchild, a jolt on a bus or working in the garden. The pain soon becomes very severe and the patient may need to stay in bed for several days. The pain is usually localized to the back and it is uncommon for pain to radiate into the legs, and symptoms of cord compression such as bladder dysfunction are rare. The pain is present throughout the day and night. The pain gradually eases and goes by 4 to 6 months. If pain persists longer, or if there is a second peak of pain during the first 6 months, this usually indicates a second vertebral fracture. This is not an uncommon occurrence. Patients commonly do not complain of back pain. Indeed, it has been estimated that at least half of vertebral fractures are asymptomatic. These asymptomatic fractures appear to be particularly common in patients taking corticosteroids. Episodes of back pain may have been forgotten. The patient commonly recalls a painful episode when confronted with the appearance of a fracture on the spinal radiograph. This incontrovertible evidence prompts the recall of a painful event occurring many decades previously, often in relation to heavy manual work in a man or after pregnancy in a woman.

Loss of Height

Loss of height is an effect of ageing, resulting from the change of posture caused by degenerative changes in the intervertebral discs. Patients do not report this symptom often and it needs to be sought by asking the patient's height in early adult life. The patient may have noticed that it is more difficult to reach high shelves. It is unusual to have sudden height loss as the presenting complaint for vertebral fracture.

Kyphosis

This may have been noticed by a relative or the patient may report being 'round-shouldered'. Clothes may no longer fit. These symptoms are not specific to vertebral fracture and are more commonly caused by disc degeneration. In a young person, kyphosis may be caused by Scheuermann's disease (see next).

Other Symptoms of Vertebral Fracture

Vertebral fractures in the lumbar region result in decreased abdominal volume. This causes the abdomen to protrude. Patients with osteoporosis are commonly slender, so this appearance is new. They may also result in impingement of the costal margin on the iliac crest. This 'iliocostal friction syndrome' causes pain and a grating sensation. This pain is postural, occurring on sitting. Vertebral fractures in the thoracic spine result in reduced lung volume. This may result in respiratory symptoms, such as dyspnoea, or in delayed recovery from chest infections.

Clinical Examination

Two aspects of the clinical examination are useful in the patient with suspected vertebral fracture. The first relates to the location of the pain. It is often assumed that the deformity on the radiograph and the patient's back pain are associated. However, a careful palpation of the spinal processes counting down from vertebra prominens (seventh cervical vertebra) often reveals that the site of the pain does not correspond with the level of the deformity on the radiograph. It is helpful to evaluate the size of the gap between the costal margin and the iliac crest. This is normally three finger's breadths (as measured by the patient's fingers).

Diagnostic Evaluation

Spinal Radiographs

See Box 4.8.2. Plain radiographs are required of the thoracic and lumbar spine in the anteroposterior and lateral position. It is a common mistake to take the radiograph only of the painful area. This would miss asymptomatic fractures in other parts of the spine. It is common only to take lateral radiographs. The anteroposterior radiograph is useful to identify the vertebral level of the fracture and to exclude other causes of deformity, such a malignancy (associated with absent pedicles). The critical feature of an osteoporosis-related vertebral fractures is fracture of the vertebral endplate.

Types of Deformity

Vertebral deformities may be wedge, endplate ('biconcave', when both endplates are affected), and compression (also called 'crush') [17]. Wedge deformities are particularly common in the thoracic spine, because the normal kyphosis in this region results in the main force running anteriorly. Biconcavity deformities are particularly common in the lumbar spine, because the normal lordosis results in the main force running through the middle of the vertebra. There appears to be no association between the type of deformity and the severity of pain or with the level of BMD. The level of the deformities should be recorded to allow comparison at follow-up visits.

Deformities that Mimic Fractures

The most common deformity to mimic fracture is Scheuermann's disease. This is a form of epiphysitis that occurs during adolescence ('juvenile epiphysitis') and gives the appearance of wedging and elongation of the vertebral bodies. The characteristic feature is the wavy appearance of the superior and inferior borders.

Malignancy may cause vertebral deformity. The isotope bone scan is particularly useful in this situation as it is unusual to have a single bone lesion. Increased uptake in multiple sites in the skeleton is typical of malignancy. This may occur in prostate cancer, which affects the sacrum, lumbar spine, and ribs (via Batson's venous plexus). Malignancy may cause erosion of the pedicle, a typical appearance not found with osteoporotic vertebral fractures.

Paget's disease of bone commonly affects the spine (see Chapter 4.10). The bone may appear sclerotic, but it is weak and can fracture. The bone texture has a disorganized appearance and the vertebra may be enlarged.

Osteomalacia may result in vertebral deformities (see Chapter 4.11). Often adjacent vertebrae are affected and the endplates are deformed. This gives rise to a 'cod-fish' appearance. There may be other radiological clues, such as the ground-glass appearance of the vertebral body bone and the presence of pseudofractures (Looser's zones) in the pelvis, long bones, or ribs.

Use of Vertebral Morphometry in the Clinic

Dual-energy X-ray absorptiometry (DXA) can be used to image the spine as well as to obtain a measurement of bone density (see next). Vertebral fractures may be identified by careful focus on the appearance of the vertebral endplate [17]. Fracture is likely if the endplate is deformed and should be confirmed with a radiograph.

Isotope Bone Scan

This can be a useful diagnostic tool in certain cases. There is increased isotope uptake in a vertebra for at least 6 months after it has fractured and typically has a uniform distribution in the vertebral body. This can be useful if the radiological appearances are borderline and yet the symptoms are characteristic. In patients with a suspicion of malignancy (previous breast cancer and history of weight loss) the scan is helpful in that metastases often affect many bones. The scan is helpful in a patient with corticosteroid-induced osteoporosis with pelvic pain. These patients commonly develop insufficiency fractures and these show up as symmetrical appearance affecting the sacral alae and the pubic rami. Single photon emission computed tomography is a variant of the isotope bone scan and is particularly useful in identifying the cause of back pain. If the facet joints show increased uptake then the patient may benefit from an injection of local anaesthetic into the facet joint.

Magnetic Resonance Imaging

This approach can be useful in identifying a recent deformity and distinguishing a fracture from a malignant deposit. It is very useful in identifying cord compression. It is an essential test before kyphoplasty or vertebroplasty are considered (see next); only recent fractures clearly benefit from these procedures.

Bone Density Measurement

See Table 4.8.1. DXA is precise, accurate, involves exposure to only low doses of X-rays, and allows measurement of sites of clinical interest (that is, lumbar spine and proximal femur). In DXA, two energy peaks of X-rays are absorbed to different extents by bone and soft tissue, and the density of bone is calculated, in g/cm^2, using simultaneous equations. The measurement is compared with two reference ranges—one for young adults (age 30 years) to give T scores

Table 4.8.1 Techniques for the non-invasive measurement of bone mass

Technique	Site	Comments
Single (or dual) energy X-ray absorptiometry	Forearm and heel	Inexpensive, precise, uses low doses of radiation, measures sites unresponsive to therapy
Dual-energy X-ray absorptiometry	Lumbar spine	Fairly expensive, precise, uses low doses of radiation, measures site responsive to therapy, needs skilled operator, subject to artefacts (spondylosis)
	Proximal femur	Fairly expensive, less precise, uses low doses of radiation, measures site best for fracture prediction, needs skilled operator
	Total body	Expensive, precise, uses low doses of radiation, measures sites unresponsive to therapy, needs skilled operator, allows assessment of body composition
Quantitative computed tomography	Spine	Expensive, less precise, uses high doses of radiation, measures sites responsive to therapy, needs skilled operator, allows assessment of trabecular bone alone
	Forearm and ankle	Inexpensive, precise, uses low doses of radiation, measures sites unresponsive to therapy, does not need skilled operator
Ultrasonometry	Heel, fingers, etc.	Inexpensive, less precise, uses no radiation, measures sites unresponsive to therapy, does not need skilled operator, fairly portable

and one for age-matched adults to give Z scores. This has become the standard technique for bone density assessment and guidelines have been proposed (Box 4.8.3). The two sites most commonly used in practice are lumbar spine and total hip. Low BMD at the total hip is a strong predictor of hip fracture [18].

Single energy X-ray (or photon) absorptiometry has similar advantages to DXA, and the equipment is less expensive. However, the sites in which bone density can be measured (distal forearm and calcaneum) may not be of clinical interest.

Quantitative computed tomography (CT) allows three-dimensional measurements of the bone density of the lumbar spine. This technique also allows measurement of trabecular bone alone (i.e. the type of bone usually lost first in the development of osteoporosis). In the research setting, finite element modelling can be applied to the information provided by this technique to estimate bone strength. However, quantitative CT is more expensive, less precise, and involves a higher radiation dose then DXA. Reformatting of routine CT scans of chest and abdomen can help identify vertebral fractures.

Quantitative ultrasound measurements are usually made on the calcaneum. The ultrasound signal has a lower frequency (200–600 kHz) than that used in obstetrics (more than 1 MHz). The attenuation of the signal (broad-band ultrasound attenuation) may reflect both the density and the architecture of bone, and the velocity of the signal reflects the density and biomechanical properties (elasticity). Quantitative ultrasonometry is currently used only in research but, if recent studies of its predictive ability in osteoporosis are confirmed, it could become an established technique.

Investigating Secondary Osteoporosis

A secondary cause is present in approximately 40% of women and 60% of men with osteoporosis. The most commonly found abnormalities are those of low vitamin D and either high or low urinary calcium [19]. Investigations to identify a secondary cause are recommended if the BMD is more than 2 SD below the age-matched mean or if the patient has low-trauma vertebral fractures (Box 4.8.2).

Bone Biopsy

This may be useful in unusual forms of osteoporosis (e.g. idiopathic osteoporosis in young adults). It provides information about the rate of bone turnover and the presence of secondary forms of osteoporosis (e.g. systemic mastocytosis). Patients with high bone turnover usually respond better to antiresorptive drugs.

Biochemical Markers of Bone Turnover

These reflect the processes of bone resorption and bone formation (Box 4.8.4). Markers that are specific to bone (e.g. osteocalcin and deoxypyridinoline) may be useful for monitoring the effect of drugs used in the treatment of osteoporosis [20]. Biochemical markers of bone resorption (such as chest X-ray) may be particularly useful because they are maximally suppressed by 3 months' treatment with oestrogens or bisphosphonates (see next and Chapter 4.10). They could be more useful than bone density for monitoring treatment because changes in bone density may not be detected for 2 years and not all patients have access to bone densitometry.

Box 4.8.3 Clinical indications for bone densitometry

- Presence of strong risk factors (see Box 4.8.1)
- Radiological evidence of vertebral fracture or osteopaenia
- Previous fragility fracture of the spine, hip, or wrist (after age 40 years)
- Monitoring of therapy

Box 4.8.4 Biochemical markers of bone turnover

- Bone formation markers (products of the osteoblast)
  - Serum alkaline phosphatase (bone isoform)
  - Serum osteocalcin
  - Serum C- and N-propeptides of type I collagen
- Bone resorption markers (degradation products of type I collagen or enzymes)
  - Urinary excretion of pyridinium cross-links of collagen (e.g. deoxypyridinoline)
  - Serum or urinary excretion of C- and N-telopeptides of type I collagen
  - Serum or urinary excretion of galactosyl hydroxylysine
  - Urinary excretion of hydroxyproline
  - Serum tartrate-resistant acid phosphatase

Treatment

General Measures

The treatment of acute back pain due to a recent vertebral fracture includes:

- bed rest (as short as possible), back support
- analgesics/non-steroidal anti-inflammatory drugs (NSAIDs)
- heat and gentle massage
- insertion of cement into the vertebral body (balloon kyphoplasty or percutaneous vertebroplasty; https://www.nice.org.uk/guidance/ta279/chapter/1-Guidance)

The treatment of chronic back pain due to vertebral fractures is difficult but includes:

- analgesics/NSAIDs
- physiotherapy
- intermittent use of spinal support for some activities
- exercise programme to maintain muscle strength and flexibility of the spine
- injection of local anaesthetic into the facet joints of the spine

In all patients (with or without fractures), it is important to treat diseases that can increase bone loss and contribute to osteoporosis (Box 4.8.1). An important part of the management of patients with osteoporosis, especially those following hip fracture or other frail patients, consists of attention to their general health status, such as ensuring adequate dietary protein intake, and measures to decrease the risk of falls or the degree of trauma that results from falling. These included better lighting, provision of handrails, removal of obstacles, attention to drugs such as sedatives and antihypertensives that may predispose to falls, or carpeted surfaces rather than hard floors. Regular exercise may be of value in maintaining mobility and improving muscle mass, thus reducing the risk of falling. Heavy weight-bearing and vigorous exercise programmes should be avoided by patients with osteoporosis as they may trigger the occurrence of a new fracture.

Calcium and Vitamin D

Low calcium intake and vitamin D deficiency should be prevented or effectively treated in all patients. As many hip fractures occur in patients over age 80 years and this population is particularly prone to low calcium intake and vitamin D deficiency (see Chapter 4.11), it is particularly important to ensure that these patients receive adequate calcium and vitamin D as part of their management. Ambulatory patients who receive periodic sunlight exposure generally produce sufficient vitamin D through skin photoconversion, but others should receive a supplement containing at least 800 IU of vitamin D daily. It is best to avoid high dose, infrequent dosing as this may increase the risk of falls. Total intake of calcium, including supplements if necessary, should be at least 1000 mg. It is best to obtain this with diet, as there has been an association described of high dose calcium supplements and myocardial infarction. Despite the necessity of adequate calcium and vitamin D for bone health, it should be appreciated that treatment with calcium and vitamin D alone is insufficient to prevent postmenopausal bone loss or to markedly reduce fracture risk in patients with osteoporosis.

Pharmacological Intervention

Drugs to increase bone mass inhibit bone resorption or stimulate bone formation. Most drugs approved for use in osteoporosis inhibit bone resorption, but some of these (e.g. hormone replacement therapy (HRT), bisphosphonates) increase BMD by 5–10% over the first 2 years of treatment [21].

Antiresorptive drugs—bisphosphonates—are considered the treatment of choice (Table 4.8.2). Their strict dosing instructions may reduce compliance in older people. In the United Kingdom, five agents are currently approved for use in osteoporosis: etidronate, alendronate, risedronate, ibandronate, zoledronic acid, and raloxifene. The most effective alternative treatments are raloxifene and HRT.

- Calcium, 1000 mg/day, and vitamin D, 500 IU/day, have been shown to prevent hip fracture in housebound, elderly patients. This treatment is safe and inexpensive, and does not require

Table 4.8.2 Antifracture efficacy of approved treatments for postmenopausal women with osteoporosis when given with calcium and vitamin D

Intervention	Vertebral fracture	Non-vertebral fracture	Hip fracture
Alendronate	A	A	A
Risedronate	A	A	A
Ibandronate	A	A*	NAE
Zoledronic acid	A	A	A
Denosumab	A	A	A
Raloxifene	A	NAE	NAE
HRT	A	A	A
Calcitriol	A	NAE	NAE
Teriparatide	A	A	NAE
Abaloparatide	A	A	NAE

A, grade A recommendation; NAE, not adequately evaluated; HRT, hormone replacement therapy; A*, in subset of patients only (post hoc analysis).

Modified from: Compston J, Cooper A, Cooper C, Gittoes N, Gregson C, Harvey N, Hope S, Kanis JA, McCloskey EV, Poole KES, Reid DM, Selby P, Thompson F, Thurston A, Vine N; National Osteoporosis Guideline Group (NOGG). UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos.* 2017 Dec;12(1):43.

monitoring. It is commonly given with other treatments for osteoporosis.

- Alendronate is given in a dose of 10 mg/day continuously or 70 mg once weekly. Alendronate must be taken at least 30 minutes before breakfast (to help absorption) with a full glass of water, and the patient must not lie down after taking the tablet (to avoid oesophagitis). Alendronate is equally effective on the hip, forearm, and spine, and has been shown to prevent fracture at all of these sites.
- Risedronate is given in a dose of 5 mg/day continuously or 35 mg once weekly. Risedronate can be taken at least 30 min before breakfast or 2 h after a meal. It has been shown to prevent spine, hip, and other fractures.
- Ibandronate is given in a dose of 150 mg once monthly; calcium recommendations and instructions for use are as for risedronate. Ibandronate reduces the risk of vertebral fracture (and other fractures, if the BMD T score is ≤3). This treatment can also be given by intravenous injection (3 mg) given every 3 months.
- Zoledronic acid is given by intravenous infusion (5 mg over 15 min) given every 12 months. It reduces fractures at the spine and hip and all other fracture sites; it has been shown to reduce further fractures in patients presenting with hip fractures.
- Denosumab is given by subcutaneous injection (60 mg) every 6 months. It reduces fractures at the spine and hip and all other fracture sites for up to 10 years on treatment.
- Raloxifene is given in a dose of 60 mg/day continuously. Raloxifene has been shown to reduce the risk of spine (but not other) fracture, and may reduce the risk of breast cancer. It may increase the risk of deep vein thrombosis and does not prevent hot flushes.
- HRT is no longer recommended for the first-line prevention of osteoporosis because the risks outweigh the benefits. Risks with HRT include breast cancer (50% increase in risk after 10 years' treatment) and deep vein thrombosis (threefold increase in risk, particularly in patients with previous deep vein thrombosis); it is associated with increased risk of stroke and ischaemic heart disease. The benefits of HRT include relief of hot flushes and vaginal dryness. Tibolone has the advantage that it reduces fractures (spine and non-vertebral) and the risk of breast cancer, but it does increase the risk of stroke.
- Testosterone therapy is effective in men with hypogonadism. It is not currently used in eugonadal men, because of concerns about the increased risk of prostate cancer and ischaemic heart disease (via lowering of high-density lipoprotein).

One other agent can be useful in special circumstances:

- Calcitriol stimulates calcium absorption and may stimulate osteoblasts directly. It appears to be effective in corticosteroid-induced osteoporosis, in which it can be considered an alternative to HRT or bisphosphonates, particularly in younger patients and in patients with chronic kidney disease. Regular monitoring of serum calcium is required because hypercalcaemia is a common adverse effect.

Formation-stimulating drugs have been licensed for osteoporosis:

- Use of a recombinant fragment of PTH (teriparatide) may be advised by specialists in osteoporosis for patients who have failed to respond to antiresorptive therapy or are intolerant of it and have severe osteoporosis. Teriparatide treatment increases the thickness of cortical bone and the connectivity of trabecular bone. These improvements in bone quality (and in quantity—spine BMD is increased by about 9% on average at 1 year) are associated with reductions in fractures of the spine and elsewhere. The treatment is given by daily subcutaneous injection (20 μg/day), with calcium supplementation for a period of 2 years. Abaloparatide is an analogue of PTH-related protein and is licensed in some countries as an anabolic treatment.
- An antibody to sclerostin (romosozumab) has recently become available in some countries for the treatment of osteoporosis. It is more effective in preventing fractures than alendronate alone [22]. The treatment is given as a subcutaneous injection (210 mg) every month for a year and then is followed by an antiresorptive treatment such as alendronate or denosumab. It should not be used in patients at high risk of myocardial infarction or stroke.

Follow-up

Once patients have been identified and treatment initiated, it is important to ensure adequate follow-up to reinforce the importance of compliance to treatment and evaluate response. At a minimum, all treated patients should be seen initially after 3 to 6 months and thereafter at least annually. The importance of adherence to treatment should be stressed. After several years of therapy, the potent antiresorptive therapies (alendronate, risedronate, ibandronate, zoledronic acid, and denosumab) have been associated with an increased risk of atypical femur fracture and osteonecrosis of the jaw. Thus, if the risk of fracture is low after five years (e.g. femoral neck T score is above –2.5) then treatment is stopped for bisphosphonates (a drug holiday) or switched for denosumab (e.g. zoledronic acid); if the risk of fracture is high then treatment is continued for ten years and then stopped [23].

Monitoring of response to therapy can be achieved with the use of biochemical markers or repeated BMD measurements, and may be of value in assessing compliance and providing feedback to patients. Although not required, treatment response to oral antiresorptive treatments, such as bisphosphonates, can be evaluated after 3 to 6 months by assessing the change in biochemical markers of bone turnover, such as N-terminal or C-terminal cross-links of type I collagen, serum osteocalcin or serum procollagen I N-propeptide (Box 4.8.4). In most patients these markers decrease by more than 30% relative to pretreatment baseline measurements, and/or are reduced to within the premenopausal reference range, providing evidence that the treatment is having its desired effect to decrease bone turnover. Amino-terminal propeptide (PINP) has proven to be particularly useful in clinical practice as it has little diurnal rhythm and is stable. Changes in BMD occur over a longer time frame, and it is generally not useful to repeat the BMD measurement before the end of 1 to 2 years of therapy, and every 2 years thereafter (Figure 4.8.4). Most patients receiving efficacious therapy can be expected to have a measurable increase in BMD at the spine and total hip after 2 years of treatment. The increases in BMD are small at the peripheral sites of measurement, such as the heel or forearm, in relation to the precision of these measurements. Therefore, peripheral sites are unreliable for assessing response in individual patients.

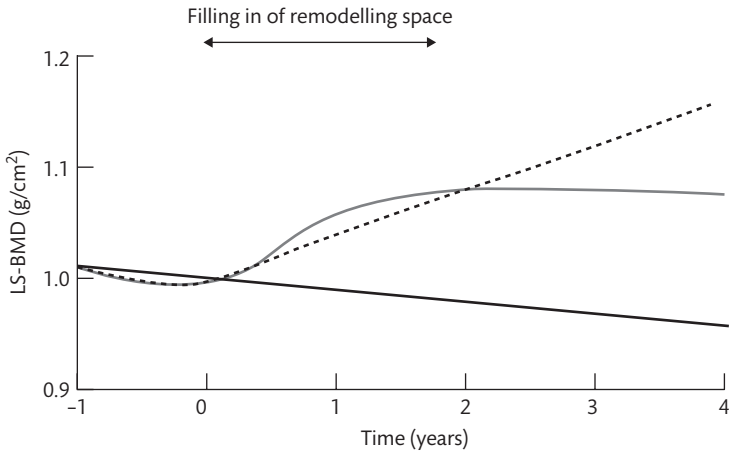


Figure 4.8.4 Changes in bone mineral density (BMD) with treatment. If no treatment is given (solid line) to someone with osteoporosis there is progressive bone loss. Antiresorptive treatments (stippled line) prevent this bone loss and result in bone gain because of filling in of the remodelling space (over a period of about 2 years). Formation-simulating treatments result in a year-on-year increase in BMD (broken line).

Other Forms of Primary Osteoporosis–Male Osteoporosis

Although osteoporosis is generally regarded as a disease of women, up to 30% of hip fractures and 20% of vertebral fractures occur in men [12]. The risk factors for osteoporotic fractures in men include low body mass index, smoking, high alcohol consumption, corticosteroid therapy, physical inactivity, diseases that predispose to low bone mass, and conditions increasing the risk of falls. The key drugs and diseases that definitely produce a decrease in BMD and/or an increase in fracture rate in men are long-term corticosteroid use, hypogonadism, alcoholism, and transplantation. Age-related bone loss may be a result of declining renal function, vitamin D deficiency, increased PTH levels, low serum testosterone levels, low calcium intake, and absorption. Osteoporosis can be diagnosed on the basis of radiological assessments of bone mass or clinically when it becomes symptomatic. Various biochemical markers have been related to bone loss in healthy and osteoporotic men. Their use as diagnostic tools, however, needs further investigation. A practical approach would be to consider a bone density more than 2.5 SD below the young normal mean value (T ≤–2.5) as an indication for therapy [18]. The treatment options for men with osteoporosis include agents to influence bone resorption or formation and specific therapy for any underlying pathological condition. Testosterone treatment increases BMD in hypogonadal men and is most effective in those whose epiphyses have not closed completely. Bisphosphonates (such as alendronate and risedronate) are the treatment of choice in idiopathic osteoporosis [24, 25], with teriparatide in more severe cases [26].

REFERENCES

1. NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy, March 7–29, 2000: highlights of the conference. *South Med J*, 2001; **94**: 569–73.
2. Kanis JA, Melton LJIII, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. *J Bone Miner Res*, 1994; **9**: 1137–41.
3. Fyhrie DP1, Christiansen BA. Bone material properties and skeletal fragility. *Calcif Tissue Int*, 2015; **97**: 213–28.
4. Kanis JA, McCloskey EV, Johansson H, Strom O, Borgstrom F, Oden A. Case finding for the management of osteoporosis with FRAX—assessment and intervention thresholds for the UK. *Osteoporos Int*, 2008; **19**: 1395–408.
5. Eastell R, Rosen CJ, Black DM, Cheung AM, Murad MH, Shoback D. Pharmacological management of osteoporosis in postmenopausal women: an Endocrine Society* clinical practice guideline. *J Clin Endocrinol Metab*, 2019; **104**: 1595–1622.
6. Eastell R. Role of oestrogen in the regulation of bone turnover at the menarche. *J Endocrinol*, 2005; **185**: 223–34.
7. Walsh JS, Henry YM, Fatayerji D, Eastell R. Lumbar spine peak bone mass and bone turnover in men and women: a longitudinal study. *Osteoporos Int*, 2008; **16**: 355–62.
8. Rivadeneira F, Mäkitie O. Osteoporosis and bone mass disorders: from gene pathways to treatments. *Trends Endocrinol Metab*, 2016; **27**: 262–81.
9. Eastell R. Treatment of postmenopausal osteoporosis. *N Engl J Med*, 1998; **338**: 736–46.
10. Seeman E, Delmas PD. Bone quality—the material and structural basis of bone strength and fragility. *N Engl J Med*, 2006; **354**: 2250–61.
11. Khosla S, Monroe DG. Regulation of bone metabolism by sex steroids. *Cold Spring Harb Perspect Med*, 2018; **8**: pii: a031211.
12. Ebeling PR. Clinical practice. Osteoporosis in men. *N Engl J Med*, 2008; **358**: 1474–82.
13. Kanis JA, Johansson H, Johnell O, *et al*. Alcohol intake as a risk factor for fracture. *Osteoporos Int*, 2005; **16**: 737–42.
14. van Staa TP. The pathogenesis, epidemiology and management of glucocorticoid-induced osteoporosis. *Calcif Tissue Int*, 2006; **79**: 129–37.
15. Compston J. Glucocorticoid-induced osteoporosis: an update. *Endocrine*, 2018; **61**: 7–16.
16. Early C, Stuckey L, Tischer S. Osteoporosis in the adult solid organ transplant population: underlying mechanisms and available treatment options. *Osteoporos Int*, 2016; **27**: 1425–40.
17. Jiang G, Eastell R, Barrington NA, Ferrar L. Comparison of methods for the visual identification of prevalent vertebral fracture in osteoporosis. *Osteoporos Int*, 2004; **15**: 887–96.

18. Cummings SR, Bates D, Black DM. Clinical use of bone densitometry: scientific review. *JAMA*, 2002; **288**: 1889–97.
19. Tannenbaum C, Clark J, Schwartzman K, *et al*. Yield of laboratory testing to identify secondary contributors to osteoporosis in otherwise healthy women. *J Clin Endocrinol Metab*, 2002; **87**: 4431–7.
20. Eastell R, Szulc P. Use of bone turnover markers in postmenopausal osteoporosis. *Lancet Diabetes Endocrinol*, 2017; **5**: 908–23.
21. Eastell R, O'Neill TW, Hofbauer LC, *et al*. Postmenopausal osteoporosis. *Nat Rev Dis Primers*, 2016; **2**: 16069.
22. Saag KG, Petersen J, Brandi ML, *et al*. Romosozumab or alendronate for fracture prevention in women with osteoporosis. *N Engl J Med*, 2017; **377**: 1417–27.
23. Adler RA, El-Hajj Fuleihan G, Bauer DC, *et al*. Managing osteoporosis in patients on long-term bisphosphonate treatment: report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res*, 2016; **31**: 1910.
24. Orwoll E, Ettinger M, Weiss S, *et al*. Alendronate for the treatment of osteoporosis in men. *N Engl J Med*, 2000; **343**: 604–10.
25. Boonen S, Orwoll ES, Wenderoth D, Stoner KJ, Eusebio R, Delmas PD. Once-weekly risedronate in men with osteoporosis: results of a 2-year, placebo-controlled, double-blind, multicenter study. *J Bone Miner Res*, 2009; **24**: 719–25.
26. Orwoll ES, Scheele WH, Paul S, *et al*. The effect of teriparatide [human parathyroid hormone (1–34)] therapy on bone density in men with osteoporosis. *J Bone Miner Res*, 2003; **18**: 9–17.

4.9

Thyroid Disorders and Bone Disease

Laura M. Watts, Bernard Freudenthal, J.H. Duncan Bassett, and Graham R. Williams

Introduction

Osteoporosis is defined as a bone mineral density (BMD) of 2.5 or more standard deviations below that of a young adult (T-score ≤ −2.5). It is characterized by reduced bone mass, low BMD, deterioration of bone microarchitecture, and an increased susceptibility to fragility fracture. The prevalence of postmenopausal osteoporosis increases with age from 6% at 50 years of age to over 50% at age 80 and the lifetime incidence of fracture for a 50-year-old in the United Kingdom is 50% for women and 20% for men [1]. Osteoporosis is a worldwide public health burden that costs an estimated $16 billion in the United States and €37 billion in Europe per annum [2, 3].

Bone strength and fracture susceptibility are determined by the acquisition of peak bone mass and the rate of bone loss in adulthood. Risk factors for osteoporosis and fracture include low BMD, a prior or parental history of fracture, low body mass index (BMI), use of glucocorticoids, smoking, and excessive alcohol consumption. Euthyroid status is essential for skeletal development, bone mineralization, acquisition of peak bone mass, and the regulation of bone maintenance in adults. Accordingly, large population studies have shown that hypothyroidism, thyrotoxicosis, and even subclinical hyperthyroidism are associated with an increased risk of fracture.

Thyroid disease is common and, similar to osteoporosis, is more prevalent in women and with increasing age. The prevalence of hypothyroidism is between 1 and 2% while 0.5–2% of women have thyrotoxicosis. As a result, thyroid replacement is very common for either primary hypothyroidism or the consequences of surgical or radioiodine treatment for thyrotoxicosis. Subclinical hypothyroidism (defined by the presence of an increased circulating thyroid-stimulating hormone (TSH) but normal thyroid hormone concentrations) may affect up to 10% of women over 55 and such patients may be overtreated [4, 5].

In children, hypothyroidism results in delayed bone age and growth arrest. Treatment with thyroxine (T_4) reverses these changes by inducing rapid 'catch-up' growth. Juvenile Graves' disease, the commonest cause of thyrotoxicosis in children, is characterized by advanced bone age and accelerated growth that results in short stature due to premature fusion of the growth plates [6]. In adults, histomorphometry studies reveal that hypothyroidism results in reduced bone turnover, however a prolonged remodelling cycle may cause a net gain in bone mass. Thyrotoxicosis increases bone resorption and formation but induces a net 10% loss of bone per remodelling cycle [7].

Thyroid Hormone Action

Circulating thyroid hormone levels are maintained in the euthyroid range by a classical negative feedback loop. Thyrotropin releasing hormone (TRH) is synthesized in the paraventricular nucleus of the hypothalamus and stimulates synthesis and secretion of TSH by the anterior pituitary gland. TSH, acting via the G-protein coupled TSH receptor, stimulates growth of thyroid follicular cells and the synthesis and release of thyroid hormones. Thyroid hormones act via thyroid hormone receptors in the hypothalamus and pituitary to inhibit TRH and TSH synthesis and secretion. This feedback loop maintains circulating thyroid hormones and TSH in a physiological inverse relationship that defines the hypothalamic-pituitary-thyroid (HPT) axis set point [6].

The thyroid gland secretes the pro-hormone T_4 and a small amount of the active hormone 3,5,3′-L-triiodothyronine (T_3). Circulating T_4 levels are maintained at approximately 3–4-fold higher concentrations than T_3. Intracellular availability of T_3 is determined by active uptake by specific cell membrane transporters including monocarboxylate transporter-8 (MCT8), MCT10 and organic acid transporter protein-1c1 (OATP1c1), and by the activities of the type 2 and 3 deiodinase enzymes (D2 and D3). D2 converts T_4 to T_3 by catalysing removal of a 5′-iodine atom. By contrast,

D3 prevents activation of T_4 and inactivates T_3 by removal of a 5-iodine atom to generate the metabolites 3,3′,5′-L-triiodothyronine (reverse T_3) and 3,3′-diiodothyronine (T_2), respectively. Thus, the relative levels of D2 and D3 determine the concentration of intracellular T_3 available to the nuclear receptors (TRs) [8].

TRs act as hormone inducible transcription factors that regulate expression of T_3-responsive target genes. The *THRA* and *THRB* genes encode three functional TRs, TRα1, TRβ1, and TRβ2. TRα1 and TRβ1 are widely expressed but with developmental and tissue specific regulation of their relative levels. Expression of TRβ2 is mainly restricted to the hypothalamus and pituitary where it mediates inhibition of TRH and TSH expression [7].

Skeletal Development and Bone Maintenance

Long bones form by endochondral ossification, in which a cartilage anlage develops from mesenchyme condensations to form a scaffold for subsequent bone formation. Mesenchyme progenitor cells differentiate into chondrocyte precursors, which undergo a sequence of clonal expansion, proliferation, hypertrophic differentiation, and apoptosis. Chondrocytes secrete a cartilage matrix that mineralizes and is subsequently remodelled by bone-resorbing osteoclasts and bone-forming osteoblasts to form the diaphysis. Linear growth occurs by a similar process within the epiphyseal growth plates, located at the proximal and distal ends of long bones. By contrast, the skull vertex forms by intramembranous ossification, in which mesenchymal cells differentiate directly into osteoblasts resulting in bone formation in the absence of a cartilage scaffold. Linear growth continues until fusion of the growth plates during puberty but bone mineralization and consolidation of bone mass accrual continues into early adulthood when peak bone mass is achieved [9].

Functional integrity of the skeleton is maintained by the process of bone remodelling, which is achieved by the integrated activities of osteocytes, osteoclasts, and osteoblasts. Osteocytes derive from differentiated osteoblasts that have become embedded in bone matrix and comprise 90–95% of all adult bone cells. The osteocyte network senses changes in mechanical load and regulates bone remodelling by the release of cytokines and chemotactic signals that stimulate differentiation and recruitment of osteoclasts to sites of altered mechanical load or microdamage. Osteoclasts excavate a resorption cavity until this process is terminated by apoptosis. Subsequently, osteoblasts mature, secrete, and mineralize osteoid to replace the resorbed bone. Signalling between osteoclasts and osteoblasts couples their activities to result in skeletal homeostasis with preservation of bone strength [10].

Thyroid Hormone Action in Bone

TRα1 and TRβ1 are expressed in growth plate chondrocytes, bone marrow stromal cells, and osteoblasts but it is uncertain whether they are present in osteoclasts [6, 11]. Expression of TRα1 is around ten times higher than TRβ1 in bone, suggesting the major T3 actions in bone are mediated via TRα1 [12]. Osteoblasts, osteoclasts, and chondrocytes also express thyroid hormone transporters to regulate local availability of thyroid hormone within bone [13].

In vivo and *in vitro* studies have shown that T_3 acts to inhibit growth plate chondrocyte proliferation and stimulate hypertrophic chondrocyte differentiation. In childhood hypothyroidism, growth arrest and delayed bone formation are consequences of gross disruption of growth plate architecture (epiphyseal dysgenesis) and a failure of hypertrophic chondrocyte differentiation. By contrast thyroid hormone excess accelerates chondrocyte differentiation resulting in advanced bone formation [6, 11].

In osteoblasts, T_3 increases proliferation and differentiation via complex cytokine and growth factor signalling pathways. For example, T_3 induces *FGFR1* in osteoblasts. Consistent with this, activating mutations of *FGFR1* cause Pfeiffer's craniosynostosis syndrome, and craniosynostosis is a recognized manifestation of severe juvenile thyrotoxicosis [14]. Thyroid hormone also stimulates osteoclast proliferation and activity, but it is uncertain whether this is a direct cellular response to T_3 [7].

The TSH receptor is expressed on osteoblasts, chondrocytes, and osteoclasts, however TSHα/β expression has not been detected within bone [15–17]. Direct actions of TSH on bone are uncertain. TSH may either inhibit or have no effect upon osteoclast differentiation, promote or have no effect upon osteoblast differentiation [17, 18] and promote chondrocyte differentiation [16, 19].

Studies in Genetically Modified Mice

Studies in mutant mice have demonstrated that TRα1 mediates T_3 action in bone. Mutation or deletion of TRα causes skeletal hypothyroidism due to disrupted T_3 action resulting in transient growth retardation, impaired ossification, and reduced bone mineralization during growth (Table 4.9.1). In adults there is a defect in bone remodelling, a marked increase in bone mass and increased bone mineralization [7, 20]. By contrast, mutation or deletion of TRβ disrupts the HPT axis leading to elevated levels of thyroid hormone which act upon TRα in bone causing skeletal hyperthyroidism. Mice display accelerated growth, advanced ossification with increased mineralization during growth but short stature from premature quiescence of the growth plates (Table 4.9.1). In adults increased bone remodelling results in osteoporosis and reduced bone mineralization [7, 12]. Together, these features indicate that thyroid hormones exert anabolic actions during skeletal growth but catabolic responses in adult bone [7, 21].

Control of T_3 availability is also required for normal postnatal skeletal development and strength. Mice lacking D2 have brittle bones with increased bone mineral content as reduced T_3 production within osteoblasts leads to decreased thyroid hormone action in bone [22]. Mice lacking the thyroid hormone transporter MCT8 have normal prenatal skeletal development but delayed endochondral ossification due to impaired T_3 transport into chondrocytes [23].

Mouse models, however, have been less clear regarding direct effects of TSH on bone. TSH receptor (TSHR) knockout mice have congenital hypothyroidism with undetectable T_3 and T_4 and elevated TSH. They show decreased BMD and increased bone turnover only partially normalized by thyroid supplementation from three weeks [15]. However, the extent to which this is due to TSH acting as a direct negative regulator of bone turnover or whether it results from thyroid hormone deficiency at a critical early period of

Table 4.9.1 Skeletal phenotypes of thyroid hormone receptor mutant mice

	TRα mutant mice	TRβ mutant mice
Systemic thyroid status	Euthyroid	Elevated T_4, T_3, and TSH
Skeletal thyroid status	Hypothyroid	Thyrotoxic
Juvenile skeleton	Transient growth delay Delayed endochondral and intramembranous ossification Impaired chondrocyte differentiation Reduced calcified bone	Persistent short stature Advanced endochondral and intramembranous ossification Enhanced chondrocyte differentiation Increased calcified bone
Adult skeleton	Osteosclerosis Increased bone volume Increased mineralization Reduced osteoclastic bone resorption	Osteoporosis Reduced bone volume Reduced mineralization Increased osteoclastic bone resorption

See review [7].

development followed by catch-up growth remains uncertain [7]. However, intermittent administration of TSH to ovariectomized animals has been shown to prevent bone loss associated with the hypogonadal state [18, 24]. Nevertheless, comparison of two mouse models of congenital hypothyroidism, suggests that the effects of thyroid hormone may predominate over those of TSHR signalling during skeletal development. Pax8 knockout mice lack a transcription factor that is essential for thyroid follicular cell development and have undetectable thyroid hormone levels, a 2000-fold elevation of TSH and a fully functional TSHR. By contrast, hyt/hyt mice have congenital hypothyroidism also accompanied by a 2000-fold increase in TSH but they harbour a loss-of-function point mutation in the *Tshr* gene leading to complete loss of TSHR protein function. Both mutants display a similar phenotype of growth retardation and delayed ossification typical of hypothyroidism despite the divergence in TSH signalling [7].

Skeletal Consequences of Altered Thyroid Status in Humans

Children

Childhood Hypothyroidism

Congenital hypothyroidism results in growth arrest, epiphyseal dysgenesis, delayed bone age, and short stature. Thyroxine replacement induces rapid catch-up growth and children that are treated early reach their predicted adult height and achieve normal BMD after 8.5 years follow-up [25, 26]. Nevertheless, a single study suggested that adult BMD may be reduced despite treatment from the neonatal period [27]. Children with juvenile acquired hypothyroidism also display growth arrest, delayed bone maturation and short stature [28]. T_4 replacement induces rapid catch-up growth, but these individuals may fail to achieve final predicted height and the resulting permanent height deficit is related to the duration of thyroid hormone deficiency prior to replacement [28]. A study of 25 children and adolescents with untreated subclinical hypothyroidism for over 3 years showed no significant effect of moderately raised TSH (4.58–8.88 IU/L) on BMD [29].

Childhood Hyperthyroidism

Juvenile thyrotoxicosis results in accelerated growth, advanced bone age, and short stature due to accelerated skeletal maturation. In severe cases in young children, early closure of the cranial sutures may result in craniosynostosis [30]. Thyrotoxicosis in children causes increased bone resorption, reduced BMD with major effects on cortical bone, and increased urinary N-terminal telopeptide of type I collagen (NTX) [31, 32]. After 1–2 years of antithyroid treatment, BMD and bone turnover markers return to normal [32–34].

Impaired Sensitivity to Thyroid Hormone

Impaired sensitivity to thyroid hormone can be caused by mutations affecting thyroid hormone action (thyroid hormone action defects (THAD), including RTHα and RTHβ), transporters (thyroid hormone cell membrane transporter defects, THCMTD) or generation of T_3 from T_4 (Thyroid hormone metabolism defects: THMD) [35]. Resistance to thyroid hormone (RTH) β is an autosomal dominant condition caused by a dominant negative mutation of *THRB* [36]. The mutant TRβ disrupts negative feedback in the HPT axis leading to increased circulating thyroid hormone concentrations in the presence of inappropriately normal or elevated TSH levels. The syndrome results in a complex mixed phenotype of hyperthyroidism and hypothyroidism depending on the target tissue studied and the specific mutation. Thus, an individual patient can have features of both thyroid hormone deficiency and excess with skeletal abnormalities similar to both hyper and hypothyroidism. These include craniofacial abnormalities, craniosynostosis, delayed or advanced bone age, short stature, increased bone turnover, osteoporosis, and fracture [7].

RTHα results from dominant negative mutations in *THRA*. Affected individuals have features of congenital hypothyroidism due to lack of response in TRα expressing tissues. Unlike RTHβ, there is no disruption to the negative feedback mechanisms of the HPT axis: serum TSH is normal, T_4 is normal/low and T_3 is normal/high resulting in a reduced T_4:T_3 ratio [7]. Skeletal features include disproportionate short stature due to short lower limbs, delayed tooth eruption, delayed closure of fontanelles, delayed bone age, epiphyseal dysgenesis and a thickened calvarium [7, 37, 38].

Sec insertion sequence binding protein 2 (SBP2) is a key factor mediating selenocysteine incorporation into thyroid hormone deiodinases, essential for their normal function. *SBP2* mutations result in defective conversion of T_4 to T_3, resulting in low T_3 despite high T_4 levels. Skeletal consequences include prepubertal growth retardation, transient delayed bone age and craniofacial dysmorphism [39, 40].

Mutations in the thyroid transporter *MCT8* result in low T_4, variable TSH, and elevated total T_3 levels, with a neurological phenotype of dystonia and hypotonia and severe developmental delay [41, 42]. The bone phenotype of these individuals has not been studied in detail.

Adults

Many studies have investigated the skeletal consequences of altered thyroid function in adults. Early studies frequently lacked statistical power because of small numbers of patients and the absence of long-term follow-up. Some were further confounded by inclusion of patients with a variety of thyroid diseases and by comparison of mixed cohorts of patients that included pre- and postmenopausal women or men. Other unadjusted confounders included prior or family history of fracture, BMI, physical activity, use of oestrogens, glucocorticoids, bisphosphonates, and vitamin D, smoking, or alcohol intake. Nevertheless, a number of meta-analyses have attempted to reconcile this heterogenous literature (Table 4.9.2).

Normal Individuals

Bone Turnover Markers

A study of 60 euthyroid postmenopausal women reported that high circulating TSH correlated with low urinary deoxypyridinoline concentrations but not with serum procollagen type I C propeptide (PICP) levels [43]. However, the number of subjects was small and included patients with treated hypothyroidism, subclinical hyperthyroidism, and secondary hyperparathyroidism. Two larger studies of euthyroid postmenopausal women [44] and euthyroid men over 70 years of age [45] showed no association between TSH and serum osteocalcin, procollagen type 1 N-terminal propeptide (P1NP), β-C-terminal telopeptide (CTX), and urinary NTX.

BMD and Fracture

A variety of mainly cross-sectional population studies have investigated the relationship between thyroid status and BMD (Table 4.9.3). With the exception of one prospective study of 4936 men and women over 65 years of age that showed no significant association with BMD [46], the others all reported an association between TSH or T_4 levels and BMD in either hip or lumbar spine.

Other studies have investigated the association of thyroid status with fracture risk, including longitudinal cohort studies (Table 4.9.4). The majority demonstrated an association between hyperthyroidism and fracture risk that was related to disease duration, although several large studies found no association between thyroid status and fracture risk.

Patients with Hypothyroidism

Bone Turnover Markers and BMD

Histomorphometry analyses have demonstrated that bone turnover is decreased in hypothyroidism [47] and the bone resorption markers urinary pyridinoline and deoxypyridinoline are lower in patients newly diagnosed with severe hypothyroidism [48]. Consistent with histomorphometric data showing normal bone volume in hypothyroidism, BMD was normal in newly diagnosed patients [49].

Fracture

Large population studies have mostly demonstrated an association between hypothyroidism and long-term fracture risk (Table 4.9.4). Patients with a prior history of hypothyroidism have been shown to have a 2–3-fold increased risk of fracture, which persists for 10 years following initial diagnosis [49–51]. In a recent study of 8414 hypothyroid adults, though elevated baseline TSH was not associated with hip fracture or major osteoporotic fracture, in post-hoc analysis, fracture risk was increased in postmenopausal women relative to the cumulative time they were over-replaced with levothyroxine, and was increased in all hypothyroid men below 75 years of age [52].

Patients with Thyrotoxicosis

The severe bone disease associated with uncontrolled thyrotoxicosis in now rare because of early diagnosis and treatment, although several studies have investigated the skeleton in thyrotoxic patients either prior to treatment, or by observing epidemiological data.

Table 4.9.2 Meta-analyses and literature reviews

Reference	Population	Studies (n)	Type	Conclusions
Vestergaard 2003 [56]	Thyrotoxicosis	20 BMD 5 Fracture	Meta-analysis	Spine and hip BMD reduced in untreated thyrotoxicosis. Fracture risk increases with age at diagnosis
Wirth 2014 [62]	Population cohort	7 Fracture	Meta-analysis	No significant association of thyroid status with fracture risk (NB heterogenous study design)
Blum 2015 [105]	Population cohort	13 Fracture	Individual participant data meta-analysis	Subclinical hyperthyroidism associated with increased risk of hip and other fractures
Yan 2016 [106]	Population cohort	5 Fracture	Meta-analysis	Subclinical hyperthyroidism associated with fracture risk over 60 years of age, but no association with subclinical hypothyroidism
Papaleontiou 2016 [107]	Suppressive T4	25 BMD	Literature review	Greater evidence of reduced BMD in postmenopausal than premenopausal women; no association with reduced BMD in men
Yang 2017 [108]	Population and subclinical thyroid dysfunction cohorts	19 Fracture and/or BMD	Meta-analysis	Subclinical hyperthyroidism associated with fracture risk and reduced BMD
Aubert 2017 [109]	Euthyroid individuals	13 Fracture	Individual participant data meta-analysis	Low-normal TSH and high-normal T4 associated with increased risk of hip fracture
Segna 2018 [110]	Population cohort	6 BMD	Individual participant data meta-analysis	Subclinical hyperthyroidism associated with reduced BMD

Table 4.9.3 Large studies of thyroid status and BMD

Reference	Study design	Subjects (n)	Patient group	Fracture risk
BMD associated with TSH				
Jamal 2005 [61]	Cross-sectional	15 316 women	Postmenopausal women	Low TSH associated with decreased hip BMD
Kim 2006 [111]	Cross-sectional	959 women	Postmenopausal women	Low-normal TSH associated with decreased spine and hip BMD
Morris 2007 [112]	Cross-sectional	581 women	Postmenopausal women	Low-normal TSH associated with decreased spine BMD
Grimnes 2008 [113]	Cross-sectional	968 men 993 women	Men and women >55 years of age	Low-normal TSH associated with decreased forearm BMD
Kim 2010 [114]	Cross-sectional	1478 men	Men of 22–85 years of age with no history of thyroid disease	Higher TSH associated with higher lumbar spine BMD
Noh 2015 [115]	Cross-sectional	756 women	Women >65 years of age	Lower TSH associated with lower BMD in lumbar spine
Lee 2016 [116]	Cross-sectional	343 men 674 women	Men and women >65 years of age	Low-normal TSH associated with reduced hip BMD in women
Acar 2016 [117]	Cross-sectional	1217 women	Postmenopausal women	Low TSH associated with decreased hip BMD
BMD associated with T_4				
Van der Deure 2008 [97]	Prospective cohort	479 men 672 women	Men and women >55 years of age	T_4 negatively associated with spine and hip BMD
Murphy 2010 [44]	Prospective cohort	593 women	Postmenopausal women	Higher T_3 and T_4 associated with lower BMD, but no association with TSH
Roef 2011 [118]	Cross-sectional	677 men	Healthy men at peak bone mass (25–45 years of age)	Higher T_4 and T_3 associated with lower BMD, but no association with TSH
Lin 2011 [119]	Cross-sectional	1343 men 1614 women	Men and women >45 years of age	No association with TSH, but T_4 correlated with BMD in older females
van Rijn 2014 [120]	Cross-sectional	2584 women	Perimenopausal women with no history of thyroid disease	High-normal T_4 levels associated with decreased BMD, but no association with TSH
Hwangbo 2016 [121]	Cross-sectional	648 women 728 men	Postmenopausal women and age-matched men	High-normal T_4 associated with higher trabecular bone score in women
No association with BMD				
Garin 2014 [46]	Prospective cohort	598 men and 719 women	Men and women >65 years of age	No association of TSH with BMD

Bone Turnover Markers and BMD

The effect of thyrotoxicosis on bone turnover markers is consistent with histomorphometric data that shows high turnover of trabecular bone with net increase in bone resorption [53]. Levels of resorption markers such as carboxy terminal telopeptide of type 1 collagen, urinary pyridinoline and deoxypyridinoline are increased, and formation markers including bone specific alkaline phosphatase and osteocalcin are also elevated [54]. Studies have shown that BMD at the time of diagnosis of thyrotoxicosis is reduced compared to age-matched controls [54–56].

Fracture

Various observational studies have identified an association between fracture and a prior history of thyrotoxicosis (Table 4.9.4). Most studies did not determine whether the increased fracture risk could be accounted for by reduced BMD. Nevertheless, other studies failed to demonstrate an association between thyrotoxicosis and fracture, though they may have lacked statistical power (Table 4.9.4).

Individuals with Subclinical Hyperthyroidism

Bone Turnover Markers and BMD

Elevated levels of urinary deoxypyridinoline, pyridinoline, osteocalcin and alkaline phosphatase have been reported in patients with subclinical hyperthyroidism [57–59]. Subclinical hyperthyroidism has also been associated with reduced BMD at the femoral neck and other sites (Table 4.9.3), although other studies have been negative [46, 60]. Meta-analyses have supported an association with reduced BMD that is strongest in postmenopausal women (Table 4.9.2).

Fracture

Prospective studies have suggested that suppressed TSH levels may be associated with an increased risk of hip fracture (Table 4.9.4), and this has been corroborated by meta-analyses (Table 4.9.2). A subanalysis of the Fracture Intervention Trial demonstrated that TSH suppressed below 0.5 mU/L was associated with increased risk of vertebral fracture [61], though some subjects may have had untreated thyrotoxicosis. A minority of smaller

Table 4.9.4 Large studies of thyroid status and fracture risk

Reference	Study design	Subjects (n)	Patient group	Fracture risk
Positive studies				
Bauer 2001 [122]	Prospective longitudinal	686 women, including 397 women with previous fracture	Women >65 years of age	Suppressed TSH associated with increased risk of hip and vertebral fracture
Vestergaard 2002 [49]	Cross-sectional	11 776 thyrotoxic 4473 hypothyroid 48 710 controls	National register	Increased risk of femur fracture at time of diagnosis of thyrotoxicosis and hypothyroidism
Sheppard 2002 [79]	Cross-sectional	23 183 patients and 92 732 matched controls (88% female)	Patients prescribed T4 replacement	Increased risk of femur fracture in males only, but no adjustment for TSH and T_4 levels
Vestergaard 2005 [50]	Cross-sectional fracture case control	124 655 men and women with current fractures 373 962 fracture-free controls	All fractures; 52% female	Fracture risk increased within 5 years after thyrotoxicosis and 10 years after hypothyroidism diagnosis
Jamal 2005 [61]	Cross-sectional	15 316 women	Postmenopausal women	Low TSH associated with increased risk of vertebral fracture
Flynn 2010 [78]	Retrospective cohort	17 684 patients on T_4 replacement	15 196 female cases and 2488 male cases	Both suppressed and high TSH associated with fracture risk
Murphy 2010 [44]	Prospective cohort	1278 euthyroid individuals	Postmenopausal women	Increased T_3 and T_4 associated with increased fracture risk; increased TSH associated with reduced fracture risk
Lee 2010 [123]	Prospective cohort	3567 subjects including 714 with subclinical thyroid dysfunction	1372 men and 2195 women over 65 years of age	Subclinical hypothyroidism and hyperthyroidism associated with hip fracture in men but no clear association in women
Waring 2013 [124]	Prospective cohort	397 men with fracture and 1420 men without fracture	Men >65 years of age	Low-normal TSH associated with increased risk of hip fracture but not with other non-spine fractures
Abrahamsen 2014 [125]	Prospective cohort	9217 hyperthyroid 222 138 controls	6629 female cases and 2588 male cases	Thyrotoxicosis associated with increased risk of hip fracture, proportionate to disease duration
Abrahamsen 2015 [52]	Prospective cohort	8414 hypothyroid 222 138 controls	2386 male cases, and 6027 female cases	Association of hypothyroidism with fracture only in men below 75 years of age; in postmenopausal women, fracture associated with cumulative time over-replaced
Negative studies				
Van den Eeden 2003 [77]	Cross-sectional case controlled	501 women hospitalized for hip fracture 533 controls	Women >65 years of age	No association of T_4 replacement with fracture risk
Van der Deure 2008 [97]	Prospective cohort	479 men 672 women	>55 years of age	No association of TSH or T_4 with fracture risk
Svare 2013 [126]	Prospective cohort	25 205 normal individuals	16 610 women and 8595 men aged 40 years or more	No association of TSH with fracture risk
Garin 2014 [46]	Cross-sectional	2765 women 2171 men	>65 years of age	No association of TSH with fracture risk
Siru 2018 [45]	Prospective cohort	4248 men	>70 years of age	No association of TSH or T_4 with fracture risk

studies failed to show a significant association with fracture risk, as did one meta-analysis that included only seven population-based studies [62].

Patients Treated with Suppressive Doses of Thyroxine

The long-term management of patients with differentiated thyroid cancer involves treatment with doses of thyroxine that suppress TSH and which may have detrimental effects on the skeleton [63].

Bone Turnover Markers and BMD

Small studies have investigated the effect of suppressive doses of T_4 on bone turnover, reporting increased levels of markers of bone resorption [64, 65] and bone formation [66–68]. Most studies of BMD show an association of TSH suppression therapy with reduced BMD at the lumbar spine, femur, or radius only in postmenopausal women [65, 69], and studies have shown no effect in men [70]. The reduction in BMD also correlates with duration of TSH suppression [71]. These findings are supported by

meta-analyses (Table 4.9.2). Of several cross-sectional studies that included male patients, only one reported a reduction in lumbar spine and femur BMD in subjects receiving suppressive doses of T_4 [72].

With the long-term effects of suppressive doses of T_4 on BMD most severe in postmenopausal women, monitoring of BMD in such patients has been recommended [73]. In addition, international thyroid cancer guidelines now recommend risk-stratifying patients and limiting the use and severity of TSH suppression in low-risk differentiated thyroid cancer [63].

Fracture

There are no studies with sufficient statistical power to determine the effect of treatment with suppressive dose of T_4 on fracture risk, except for one study showing markedly increased risk of radiological vertebral fracture [74].

Patients Treated for Hypothyroidism

Bone Turnover Markers and BMD

Histomorphometric studies have suggested an increase in bone turnover in response to T_4 replacement in hypothyroidism [47] and studies of bone turnover markers have shown that low levels of urinary pyridinoline and deoxypyridinoline in hypothyroidism normalize with treatment [48]. Once stabilized on treatment, bone turnover markers are comparable with normal controls [75], but plasma CTX correlates with the T_4 replacement dose [76]. Observational studies that include patients receiving long-term T_4 replacement for hypothyroidism have not identified any significant effect on BMD, particularly when adjusted for circulating thyroid hormone levels.

Fracture

Population studies have not identified an overall association between T_4 replacement therapy and fracture risk [77]. In one large prospective study investigating the effects of T_4 replacement, increased fracture risk was found to be associated with cumulative time of T_4 overreplacement in postmenopausal women [52]. In another observational cohort, fracture risk was increased in patients with suppressed (≤0.03 mU/L) or high TSH, but not low TSH (0.04–0.4 mU/L) [78]. One study showed an association of T_4 treatment with fracture only in males, but thyroid hormone levels were neither measured nor adjusted for [79]. Overall, fracture risk appears not to be affected in patients receiving physiological T_4 replacement, and this conclusion is consistent with studies of patients treated for subclinical hypothyroidism [80].

Patients Treated for Thyrotoxicosis

Studies of patients with thyrotoxicosis have shown that elevated resorption and formation markers return to normal two months after initiation of antithyroid medication [81–83]. A meta-analysis investigating the effect of treatment for thyrotoxicosis on BMD [56] demonstrated that the low BMD at diagnosis returned to normal within one to four years, and subsequent studies have shown treatment for thyrotoxicosis results in a 4% increase in BMD within one year [84–86]. Nevertheless, one large population study reported that an increased risk of fracture persisted for 5 years following diagnosis [50].

Human Genetics

BMD is a highly heritable trait, and genome-wide association studies have identified numerous genes associated with BMD. However, no thyroid related genes have yet been identified in these studies [87–93]. Thyroid status is also a complex genetic trait. In healthy individuals, T_3, T_4, and TSH levels fluctuate over a range that is less than 50% of the normal reference range and each individual has their own HPT axis set point [94]. Heritability of T_3 concentration is 23–64%, T_4 is 39–65%, and TSH is 64–65% [95, 96].

Candidate gene studies have shown variable but contradictory association of thyroid pathway genes with bone parameters. The D727E polymorphism in *TSHR* has been associated with increased femoral neck BMD [68, 97] but has been reported at increased frequency in those with osteoporosis [98]. A recent large study found no association between polymorphisms at the *TSHR* locus and femoral neck or lumbar spine BMD. Using Mendelian randomization to avoid confounding and reverse causation, this study also found no evidence for a causal relationship between circulating TSH concentrations and BMD [99].

Polymorphisms in *THRB* have been associated with trabecular but not cortical BMD at the hip [100], but no studies have associated *THRA* variants with BMD or fracture [101–103]. A polymorphism in *DIO2* has been linked to reduced femoral neck BMD and increased markers of bone turnover in a small study. However, the significance of this is unclear as no such association was found in a study of a larger group of individuals and no *DIO2* mutations were identified in 100 women with high BMD [103, 104]. The clinical significance of genetic variants in thyroid hormone metabolism and receptors is currently unclear and studies in very large populations will be required.

Conclusions

- A negative feedback loop maintains circulating thyroid hormones and TSH in an inverse relationship that defines the HPT axis set point.
- The skeleton is exquisitely sensitive to thyroid status. T_3 exerts anabolic responses during skeletal growth and has catabolic effects on adult bone.
- Bone strength and fracture susceptibility are determined by peak bone mass acquisition during growth and the rate of bone loss in adulthood.
- Population studies indicate that thyroid status in the upper normal range is associated with decreased BMD and fracture.
- Untreated hyperthyroidism results in increased bone turnover, reduced BMD, and an increased risk of fracture. A prior history of thyrotoxicosis may be associated with reduced BMD and a long-term increased risk of fracture.
- Subclinical hyperthyroidism is associated with increased bone loss, decreased BMD, and an increased risk of fracture.
- Treatment with suppressive dose of T_4 results in decreased BMD especially in postmenopausal women.
- Physiological T_4 replacement in hypothyroidism does not affect BMD or fracture risk.
- Overreplacement of patients with hypothyroidism is associated with an increased risk of fracture.

REFERENCES

1. van Staa TP, Dennison EM, Leufkens HG, *et al*. Epidemiology of fractures in England and Wales. *Bone*, 2001; **29**: 517–22.
2. Blume SW, Curtis JR. Medical costs of osteoporosis in the elderly Medicare population. *Osteoporos Int*, 2011; **22**: 1835–44.
3. Hernlund E, Svedbom A, Ivergard M, *et al*. Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). *Arch Osteoporos*, 2013; **8**: 136.
4. Vanderpump MP. The epidemiology of thyroid disease. *Br Med Bull*, 2011; **99**: 39–51.
5. Stott DJ, Rodondi N, Kearney PM, *et al*. Thyroid hormone therapy for older adults with subclinical hypothyroidism. *N Engl J Med*, 2017; **376**: 2534–44.
6. Bassett JH, Williams GR. Critical role of the hypothalamic-pituitary-thyroid axis in bone. *Bone*, 2008; **43**: 418–26.
7. Bassett JH, Williams GR. Role of thyroid hormones in skeletal development and bone maintenance. *Endocr Rev*, 2016; **37**: 135–87.
8. St Germain DL, Galton VA, Hernandez A. Minireview: defining the roles of the iodothyronine deiodinases: current concepts and challenges. *Endocrinology*, 2009; **150**: 1097–107.
9. Kronenberg HM. Developmental regulation of the growth plate. *Nature*, 2003; **423**: 332–6.
10. Raggatt LJ, Partridge NC. Cellular and molecular mechanisms of bone remodeling. *J Biol Chem*, 2010; **285**: 25103–8.
11. Bassett JH, Williams GR. The molecular actions of thyroid hormone in bone. *Trends Endocrinol Metab*, 2003; **14**: 356–64.
12. O'Shea PJ, Harvey CB, Suzuki H, *et al*. A thyrotoxic skeletal phenotype of advanced bone formation in mice with resistance to thyroid hormone. *Mol Endocrinol*, 2003; **17**: 1410–24.
13. Williams AJ, Robson H, Kester MHA, *et al*. Iodothyronine deiodinase enzyme activities in bone. *Bone*, 2008; **43**: 126–34.
14. Stevens DA, Harvey CB, Scott AJ, *et al*. Thyroid hormone activates fibroblast growth factor receptor-1 in bone. *Mol Endocrinol*, 2003; **17**: 1751–66.
15. Abe E, Marians RC, Yu W, *et al*. TSH is a negative regulator of skeletal remodeling. *Cell*, 2003; **115**: 151–62.
16. Endo T, Kobayashi T. Excess TSH causes abnormal skeletal development in young mice with hypothyroidism via suppressive effects on the growth plate. *Am J Physiol Endocrinol Metab*, 2013; **305**: E660–6.
17. Bassett JH, Williams AJ, Murphy E, *et al*. A lack of thyroid hormones rather than excess thyrotropin causes abnormal skeletal development in hypothyroidism. *Mol Endocrinol*, 2008; **22**: 501–12.
18. Sampath TK, Simic P, Sendak R, *et al*. Thyroid-stimulating hormone restores bone volume, microarchitecture, and strength in aged ovariectomized rats. *J Bone Miner Res*, 2007; **22**: 849–59.
19. Bagriacik EU, Yaman M, Haznedar R, *et al*. TSH-induced gene expression involves regulation of self-renewal and differentiation-related genes in human bone marrow-derived mesenchymal stem cells. *J Endocrinol*, 2012; **212**: 169–78.
20. Fraichard A, Chassande O, Plateroti M, *et al*. The T3R alpha gene encoding a thyroid hormone receptor is essential for postnatal development and thyroid hormone production. *EMBO J*, 1997; **16**: 4412–20.
21. Bassett JH, Williams GR. The skeletal phenotypes of TRalpha and TRbeta mutant mice. *J Mol Endocrinol*, 2009; **42**: 269–82.
22. Bassett JH, Boyde A, Howell PG, *et al*. Optimal bone strength and mineralization requires the type 2 iodothyronine deiodinase in osteoblasts. *Proc Natl Acad Sci U S A*, 2010; **107**: 7604–9.
23. Leitch VD, Di Cosmo C, Liao XH, *et al*. An essential physiological role for MCT8 in bone in male mice. *Endocrinology*, 2017; **158**: 3055–66.
24. Sun L, Vukicevic S, Baliram R, *et al*. Intermittent recombinant TSH injections prevent ovariectomy-induced bone loss. *Proc Natl Acad Sci U S A*, 2008; **105**: 4289–94.
25. Leger J, Ruiz JC, Guibourdenche J, *et al*. Bone mineral density and metabolism in children with congenital hypothyroidism after prolonged L-thyroxine therapy. *Acta Paediatr*, 1997; **86**: 704–10.
26. Kooh SW, Brnjac L, Ehrlich RM, *et al*. Bone mass in children with congenital hypothyroidism treated with thyroxine since birth. *J Pediatr Endocrinol Metab*, 1996; **9**: 59–62.
27. Salerno M, Lettiero T, Esposito-del Puente A, *et al*. Effect of long-term L-thyroxine treatment on bone mineral density in young adults with congenital hypothyroidism. *Eur J Endocrinol*, 2004; **151**: 689–94.
28. Rivkees SA, Bode HH, Crawford JD. Long-term growth in juvenile acquired hypothyroidism: the failure to achieve normal adult stature. *N Engl J Med*, 1988; **318**: 599–602.
29. Di Mase R, Cerbone M, Improda N, *et al*. Bone health in children with long-term idiopathic subclinical hypothyroidism. *Ital J Pediatr*, 2012; **38**: 56.
30. Segni M, Leonardi E, Mazzoncini B, *et al*. Special features of Graves' disease in early childhood. *Thyroid*, 1999; **9**: 871–7.
31. Numbenjapon N, Costin G, Gilsanz V, *et al*. Low cortical bone density measured by computed tomography in children and adolescents with untreated hyperthyroidism. *J Pediatr*, 2007; **150**: 527–30.
32. Lucidarme N, Ruiz JC, Czernichow P, *et al*. Reduced bone mineral density at diagnosis and bone mineral recovery during treatment in children with Graves' disease. *J Pediatr*, 2000; **137**: 56–62.
33. Mora S, Weber G, Marenzi K, *et al*. Longitudinal changes of bone density and bone resorption in hyperthyroid girls during treatment. *J Bone Miner Res*, 1999; **14**: 1971–7.
34. Numbenjapon N, Costin G, Pitukcheewanont P. Normalization of cortical bone density in children and adolescents with hyperthyroidism treated with antithyroid medication. *Osteoporos Int*, 2012; **23**: 2277–82.
35. Refetoff S, Bassett JH, Beck-Peccoz P, *et al*. Classification and proposed nomenclature for inherited defects of thyroid hormone action, cell transport, and metabolism. *Eur Thyroid J*, 2014; **3**: 7–9.
36. Weiss RE, Refetoff S. Resistance to thyroid hormone. *Rev Endocr Metab Disord*, 2000; **1**: 97–108.
37. Bochukova E, Schoenmakers N, Agostini M, *et al*. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*, 2012; **366**: 243–9.
38. Demir K, van Gucht AL, Buyukinan M, *et al*. Diverse genotypes and phenotypes of three novel thyroid hormone receptor-alpha mutations. *J Clin Endocrinol Metab*, 2016; **101**: 2945–54.
39. Dumitrescu AM, Liao XH, Abdullah MS, *et al*. Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. *Nat Genet*, 2005; **37**: 1247–52.
40. Azevedo MF, Barra GB, Naves LA, *et al*. Selenoprotein-related disease in a young girl caused by nonsense mutations in the *SBP2* gene. *J Clin Endocrinol Metab*, 2010; **95**: 4066–71.
41. Dumitrescu AM, Liao XH, Best TB, *et al*. A novel syndrome combining thyroid and neurological abnormalities is associated

with mutations in a monocarboxylate transporter gene. *Am J Hum Genet*, 2004; **74**: 168–75.
42. Friesema EC, Grueters A, Biebermann H, *et al*. Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet*, 2004; **364**: 1435–7.
43. Zofkova I, Hill M. Biochemical markers of bone remodeling correlate negatively with circulating TSH in postmenopausal women. *Endocr Regul*, 2008; **42**: 121–7.
44. Murphy E, Gluer CC, Reid DM, *et al*. Thyroid function within the upper normal range is associated with reduced bone mineral density and an increased risk of nonvertebral fractures in healthy euthyroid postmenopausal women. *J Clin Endocrinol Metab*, 2010; **95**: 3173–81.
45. Siru R, Alfonso H, Chubb SAP, *et al*. Subclinical thyroid dysfunction and circulating thyroid hormones are not associated with bone turnover markers or incident hip fracture in older men. *Clin Endocrinol (Oxf)*, 2018; **89**: 93–9.
46. Garin MC, Arnold AM, Lee JS, *et al*. Subclinical thyroid dysfunction and hip fracture and bone mineral density in older adults: the cardiovascular health study. *J Clin Endocrinol Metab*, 2014; **99**: 2657–64.
47. Eriksen EF, Mosekilde L, Melsen F. Kinetics of trabecular bone resorption and formation in hypothyroidism: evidence for a positive balance per remodeling cycle. *Bone*, 1986; **7**: 101–8.
48. Nakamura H, Mori T, Genma R, *et al*. Urinary excretion of pyridinoline and deoxypyridinoline measured by immunoassay in hypothyroidism. *Clin Endocrinol (Oxf)*, 1996; **44**: 447–51.
49. Vestergaard P, Mosekilde L. Fractures in patients with hyperthyroidism and hypothyroidism: a nationwide follow-up study in 16,249 patients. *Thyroid*, 2002; **12**: 411–19.
50. Vestergaard P, Rejnmark L, Mosekilde L. Influence of hyper- and hypothyroidism, and the effects of treatment with antithyroid drugs and levothyroxine on fracture risk. *Calcif Tissue Int*, 2005; **77**: 139–44.
51. Vestergaard P, Rejnmark L, Weeke J, *et al*. Fracture risk in patients treated for hyperthyroidism. *Thyroid*, 2000; **10**: 341–8.
52. Abrahamsen B, Jorgensen HL, Laulund AS, *et al*. The excess risk of major osteoporotic fractures in hypothyroidism is driven by cumulative hyperthyroid as opposed to hypothyroid time: an observational register-based time-resolved cohort analysis. *J Bone Miner Res*, 2015; **30**: 898–905.
53. Eriksen EF, Mosekilde L, Melsen F. Trabecular bone remodeling and bone balance in hyperthyroidism. *Bone*, 1985; **6**: 421–8.
54. El Hadidy el HM, Ghonaim M, El Gawad S, *et al*. Impact of severity, duration, and etiology of hyperthyroidism on bone turnover markers and bone mineral density in men. *BMC Endocr Disord*, 2011; **11**: 15.
55. Barbosa AP, Rui Mascarenhas M, Silva CF, *et al*. Prevalence of silent vertebral fractures detected by vertebral fracture assessment in young Portuguese men with hyperthyroidism. *Eur J Endocrinol*, 2015; **172**: 189–94.
56. Vestergaard P, Mosekilde L. Hyperthyroidism, bone mineral, and fracture risk—a meta-analysis. *Thyroid*, 2003; **13**: 585–93.
57. Kumeda Y, Inaba M, Tahara H, *et al*. Persistent increase in bone turnover in Graves' patients with subclinical hyperthyroidism. *J Clin Endocrinol Metab*, 2000; **85**: 4157–61.
58. Ohishi T, Takahashi M, Kushida K, *et al*. Quantitative analyses of urinary pyridinoline and deoxypyridinoline excretion in patients with hyperthyroidism. *Endocr Res*, 1992; **18**: 281–90.
59. Kisakol G, Kaya A, Gonen S, *et al*. Bone and calcium metabolism in subclinical autoimmune hyperthyroidism and hypothyroidism. *Endocr J*, 2003; **50**: 657–61.
60. Saler T, Ahbab S, Saglam ZA, *et al*. Endogenous subclinical hyperthyroidism may not lead to bone loss in premenopausal women. *Hippokratia*, 2014; **18**: 240–4.
61. Jamal SA, Leiter RE, Bayoumi AM, *et al*. Clinical utility of laboratory testing in women with osteoporosis. *Osteoporos Int*, 2005; **16**: 534–40.
62. Wirth CD, Blum MR, da Costa BR, *et al*. Subclinical thyroid dysfunction and the risk for fractures: a systematic review and meta-analysis. *Ann Intern Med*, 2014; **161**: 189–99.
63. Freudenthal B, Williams GR. Thyroid stimulating hormone suppression in the long-term follow-up of differentiated thyroid cancer. *Clin Oncol (R Coll Radiol)*, 2017; **29**: 325–8.
64. Heijckmann AC, Huijberts MS, Geusens P, *et al*. Hip bone mineral density, bone turnover and risk of fracture in patients on long-term suppressive L-thyroxine therapy for differentiated thyroid carcinoma. *Eur J Endocrinol*, 2005; **153**: 23–9.
65. Tournis S, Antoniou JD, Liakou CG, *et al*. Volumetric bone mineral density and bone geometry assessed by peripheral quantitative computed tomography in women with differentiated thyroid cancer under TSH suppression. *Clin Endocrinol (Oxf)*, 2015; **82**: 197–204.
66. Karner I, Hrgovic Z, Sijanovic S, *et al*. Bone mineral density changes and bone turnover in thyroid carcinoma patients treated with supraphysiologic doses of thyroxine. *Eur J Med Res*, 2005; **10**: 480–8.
67. Toivonen J, Tahtela R, Laitinen K, *et al*. Markers of bone turnover in patients with differentiated thyroid cancer with and following withdrawal of thyroxine suppressive therapy. *Eur J Endocrinol*, 1998; **138**: 667–73.
68. Heemstra KA, van der Deure WM, Peeters RP, *et al*. Thyroid hormone independent associations between serum TSH levels and indicators of bone turnover in cured patients with differentiated thyroid carcinoma. *Eur J Endocrinol*, 2008; **159**: 69–76.
69. Sugitani I, Fujimoto Y. Effect of postoperative thyrotropin suppressive therapy on bone mineral density in patients with papillary thyroid carcinoma: a prospective controlled study. *Surgery*, 2011; **150**: 1250–7.
70. Reverter JL, Colome E, Holgado S, *et al*. Bone mineral density and bone fracture in male patients receiving long-term suppressive levothyroxine treatment for differentiated thyroid carcinoma. *Endocrine*, 2010; **37**: 467–72.
71. Moon JH, Kim KM, Oh TJ, *et al*. The effect of TSH suppression on vertebral trabecular bone scores in patients with differentiated thyroid carcinoma. *J Clin Endocrinol Metab*, 2017; **102**: 78–85.
72. Jodar E, Begona Lopez M, Garcia L, *et al*. Bone changes in pre- and postmenopausal women with thyroid cancer on levothyroxine therapy: evolution of axial and appendicular bone mass. *Osteoporos Int*, 1998; **8**: 311–16.
73. Biondi B, Cooper DS. Benefits of thyrotropin suppression versus the risks of adverse effects in differentiated thyroid cancer. *Thyroid*, 2010; **20**: 135–46.
74. Mazziotti G, Formenti AM, Frara S, *et al*. High prevalence of radiological vertebral fractures in women on TSH-suppressive therapy for thyroid carcinoma. *J Clin Endocrinol Metab*, 2018; **103**: 956–64.
75. Langdahl BL, Loft AG, Eriksen EF, *et al*. Bone mass, bone turnover and body composition in former hypothyroid patients receiving replacement therapy. *Eur J Endocrinol*, 1996; **134**: 702–9.
76. Christy AL, D'Souza V, Babu RP, *et al*. Utility of C-terminal telopeptide in evaluating levothyroxine replacement therapy-induced bone loss. *Biomark Insights*, 2014; **9**: 1–6.

77. Van Den Eeden SK, Barzilay JI, Ettinger B, *et al*. Thyroid hormone use and the risk of hip fracture in women > or = 65 years: a case-control study. *J Womens Health (Larchmt)*, 2003; **12**: 27–31.
78. Flynn RW, Bonellie SR, Jung RT, *et al*. Serum thyroid-stimulating hormone concentration and morbidity from cardiovascular disease and fractures in patients on long-term thyroxine therapy. *J Clin Endocrinol Metab*, 2010; **95**: 186–93.
79. Sheppard MC, Holder R, Franklyn JA. Levothyroxine treatment and occurrence of fracture of the hip. *Arch Intern Med*, 2002; **162**: 338–43.
80. Pedrera-Zamorano JD, Roncero-Martin R, Calderon-Garcia JF, *et al*. Treatment of subclinical hypothyroidism does not affect bone mass as determined by dual-energy X-ray absorptiometry, peripheral quantitative computed tomography and quantitative bone ultrasound in Spanish women. *Arch Med Sci*, 2015; **11**: 1008–14.
81. Skowronska-Jozwiak E, Lewandowski KC, Adamczewski Z, *et al*. Mechanisms of normalisation of bone metabolism during recovery from hyperthyroidism: potential role for sclerostin and parathyroid hormone. *Int J Endocrinol*, 2015; **2015**: 948384.
82. Nagasaka S, Sugimoto H, Nakamura T, *et al*. Antithyroid therapy improves bony manifestations and bone metabolic markers in patients with Graves' thyrotoxicosis. *Clin Endocrinol (Oxf)*, 1997; **47**: 215–21.
83. Pantazi H, Papapetrou PD. Changes in parameters of bone and mineral metabolism during therapy for hyperthyroidism. *J Clin Endocrinol Metab*, 2000; **85**: 1099–106.
84. Udayakumar N, Chandrasekaran M, Rasheed MH, *et al*. Evaluation of bone mineral density in thyrotoxicosis. *Singapore Med J*, 2006; **47**: 947–50.
85. Dhanwal DK, Gupta N. Bone mineral density trends in Indian patients with hyperthyroidism—effect of antithyroid therapy. *J Assoc Physicians India*, 2011; **59**: 561–2, 7.
86. Karunakaran P, Maharajan C, Mohamed KN, *et al*. Rapid restoration of bone mass after surgical management of hyperthyroidism: a prospective case control study in Southern India. *Surgery*, 2016; **159**: 771–6.
87. Medina-Gomez C, Kemp JP, Trajanoska K, *et al*. Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. *Am J Hum Genet*, 2018; **102**: 88–102.
88. Kemp JP, Medina-Gomez C, Estrada K, *et al*. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet*, 2014; **10**: e1004423.
89. Kemp JP, Morris JA, Medina-Gomez C, *et al*. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet*, 2017; **49**: 1468–75.
90. Mullin BH, Zhao JH, Brown SJ, *et al*. Genome-wide association study meta-analysis for quantitative ultrasound parameters of bone identifies five novel loci for broadband ultrasound attenuation. *Hum Mol Genet*, 2017; **26**: 2791–802.
91. Moayyeri A, Hsu YH, Karasik D, *et al*. Genetic determinants of heel bone properties: genome-wide association meta-analysis and replication in the GEFOS/GENOMOS consortium. *Hum Mol Genet*, 2014; **23**: 3054–68.
92. Estrada K, Styrkarsdottir U, Evangelou E, *et al*. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet*, 2012; **44**: 491–501.
93. Rivadeneira F, Styrkarsdottir U, Estrada K, *et al*. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet*, 2009; **41**: 1199–206.
94. Andersen S, Bruun NH, Pedersen KM, *et al*. Biologic variation is important for interpretation of thyroid function tests. *Thyroid*, 2003; **13**: 1069–78.
95. Hansen PS, Brix TH, Sorensen TI, *et al*. Major genetic influence on the regulation of the pituitary-thyroid axis: a study of healthy Danish twins. *J Clin Endocrinol Metab*, 2004; **89**: 1181–7.
96. Panicker V, Wilson SG, Spector TD, *et al*. Heritability of serum TSH, free T4 and free T3 concentrations: a study of a large UK twin cohort. *Clin Endocrinol (Oxf)*, 2008; **68**: 652–9.
97. van der Deure WM, Uitterlinden AG, Hofman A, *et al*. Effects of serum TSH and FT4 levels and the TSHR-Asp727Glu polymorphism on bone: the Rotterdam Study. *Clin Endocrinol (Oxf)*, 2008; **68**: 175–81.
98. Liu RD, Chen RX, Li WR, *et al*. The Glu727 allele of thyroid stimulating hormone receptor gene is associated with osteoporosis. *N Am J Med Sci*, 2012; **4**: 300–4.
99. Van Vliet NA, Noordam R, Van Klinken JB, *et al*. Thyroid stimulating hormone and bone mineral density: evidence from a two-sample Mendelian randomization study and a candidate gene association study. *J Bone Miner Res*, 2018; **33**: 1318–25.
100. Yerges LM, Klei L, Cauley JA, *et al*. Candidate gene analysis of femoral neck trabecular and cortical volumetric bone mineral density in older men. *J Bone Miner Res*, 2010; **25**: 330–8.
101. Medici M, Rivadeneira F, van der Deure WM, *et al*. A large-scale population-based analysis of common genetic variation in the thyroid hormone receptor alpha locus and bone. *Thyroid*, 2012; **22**: 223–4.
102. Yerges LM, Klei L, Cauley JA, *et al*. High-density association study of 383 candidate genes for volumetric BMD at the femoral neck and lumbar spine among older men. *J Bone Miner Res*, 2009; **24**: 2039–49.
103. Gogakos A, Logan JG, Waung JA, *et al*. THRA and DIO2 mutations are unlikely to be a common cause of increased bone mineral density in euthyroid post-menopausal women. *Eur J Endocrinol*, 2014; **170**: 637–44.
104. Verloop H, Dekkers OM, Peeters RP, *et al*. Genetics in endocrinology: genetic variation in deiodinases: a systematic review of potential clinical effects in humans. *Eur J Endocrinol*, 2014; **171**: R123–35.
105. Blum MR, Bauer DC, Collet TH, *et al*. Subclinical thyroid dysfunction and fracture risk: a meta-analysis. *JAMA*, 2015; **313**: 2055–65.
106. Yan Z, Huang H, Li J, *et al*. Relationship between subclinical thyroid dysfunction and the risk of fracture: a meta-analysis of prospective cohort studies. *Osteoporos Int*, 2016; **27**: 115–25.
107. Papaleontiou M, Hawley ST, Haymart MR. Effect of thyrotropin suppression therapy on bone in thyroid cancer patients. *Oncologist*, 2016; **21**: 165–71.
108. Yang R, Yao L, Fang Y, *et al*. The relationship between subclinical thyroid dysfunction and the risk of fracture or low bone mineral density: a systematic review and meta-analysis of cohort studies. *J Bone Miner Metab*, 2018; **36**: 209–20.
109. Aubert CE, Floriani C, Bauer DC, *et al*. Thyroid function tests in the reference range and fracture: individual participant analysis of prospective cohorts. *J Clin Endocrinol Metab*, 2017; **102**: 2719–28.

110. Segna D, Bauer DC, Feller M, *et al*. Association between subclinical thyroid dysfunction and change in bone mineral density in prospective cohorts. *J Intern Med*, 2018; **283**: 56–72.
111. Kim DJ, Khang YH, Koh JM, *et al*. Low normal TSH levels are associated with low bone mineral density in healthy postmenopausal women. *Clin Endocrinol (Oxf)*, 2006; **64**: 86–90.
112. Morris MS. The association between serum thyroid-stimulating hormone in its reference range and bone status in postmenopausal American women. *Bone*, 2007; **40**: 1128–34.
113. Grimnes G, Emaus N, Joakimsen RM, *et al*. The relationship between serum TSH and bone mineral density in men and postmenopausal women: the Tromso study. *Thyroid*, 2008; **18**: 1147–55.
114. Kim BJ, Lee SH, Bae SJ, *et al*. The association between serum thyrotropin (TSH) levels and bone mineral density in healthy euthyroid men. *Clin Endocrinol (Oxf)*, 2010; **73**: 396–403.
115. Noh HM, Park YS, Lee J, *et al*. A cross-sectional study to examine the correlation between serum TSH levels and the osteoporosis of the lumbar spine in healthy women with normal thyroid function. *Osteoporos Int*, 2015; **26**: 997–1003.
116. Lee SJ, Kim KM, Lee EY, *et al*. Low Normal TSH levels are associated with impaired BMD and hip geometry in the elderly. *Aging Dis*, 2016; **7**: 734–43.
117. Acar B, Ozay AC, Ozay OE, *et al*. Evaluation of thyroid function status among postmenopausal women with and without osteoporosis. *Int J Gynaecol Obstet*, 2016; **134**: 53–7.
118. Roef G, Lapauw B, Goemaere S, *et al*. Thyroid hormone status within the physiological range affects bone mass and density in healthy men at the age of peak bone mass. *Eur J Endocrinol*, 2011; **164**: 1027–34.
119. Lin JD, Pei D, Hsia TL, *et al*. The relationship between thyroid function and bone mineral density in euthyroid healthy subjects in Taiwan. *Endocr Res*, 2011; **36**: 1–8.
120. van Rijn LE, Pop VJ, Williams GR. Low bone mineral density is related to high physiological levels of free thyroxine in peri-menopausal women. *Eur J Endocrinol*, 2014; **170**: 461–8.
121. Hwangbo Y, Kim JH, Kim SW, *et al*. High-normal free thyroxine levels are associated with low trabecular bone scores in euthyroid postmenopausal women. *Osteoporos Int*, 2016; **27**: 457–62.
122. Bauer DC, Ettinger B, Nevitt MC, *et al*. Risk for fracture in women with low serum levels of thyroid-stimulating hormone. *Ann Intern Med*, 2001; **134**: 561–8.
123. Lee JS, Buzkova P, Fink HA, *et al*. Subclinical thyroid dysfunction and incident hip fracture in older adults. *Arch Intern Med*, 2010; **170**: 1876–83.
124. Waring AC, Harrison S, Fink HA, *et al*. A prospective study of thyroid function, bone loss, and fractures in older men: the MrOS study. *J Bone Miner Res*, 2013; **28**: 472–9.
125. Abrahamsen B, Jorgensen HL, Laulund AS, *et al*. Low serum thyrotropin level and duration of suppression as a predictor of major osteoporotic fractures-the OPENTHYRO register cohort. *J Bone Miner Res*, 2014; **29**: 2040–50.
126. Svare A, Nilsen TI, Asvold BO, *et al*. Does thyroid function influence fracture risk? Prospective data from the HUNT2 study, Norway. *Eur J Endocrinol*, 2013; **169**: 845–52.

4.10

Paget's Disease of Bone

Socrates E. Papapoulos

Introduction

In 1876, Sir James Paget presented to the Royal Medical and Chirurgical Society of London an account of his experience with a previously unrecognized disease of the skeleton, which he termed osteitis deformans and has since born his name. Paget's disease of bone is a focal skeletal disorder which progresses slowly and leads to changes in the shape and size of affected bones and to skeletal, articular, and vascular complications. In some parts of the world it is the second most common bone disorder after osteoporosis. The disease is easily diagnosed and effectively treated but its pathogenesis is incompletely understood.

Epidemiology

Paget's disease affects typically older people, slightly more men than women, and seldom presents before the age of 35 years. Its prevalence increases with age and it affects 1 to 5% of those above 50 years of age. However, only a small proportion of individuals with Paget's disease comes to clinical attention, most commonly these with symptomatic or severe disease [1–3]. There is a distinct geographical distribution; the disease is common in central, western, and parts of southern Europe, the United States, Australia, New Zealand, and some countries of South America, while it is uncommon in Scandinavia, Asia, and Africa. There may also be variations in prevalence in regions within the same country [4]. For example, in northeast USA the prevalence is about fivefold higher than in south USA [5] and in parts of northwest England in 1974 the age- and gender-standardized prevalence rate was 8.3% compared to 4.6% in southern towns and cities [6]. Notably, a more recent radiographic survey in the same centres with identical methodology [7] reported a decline in the overall prevalence of the disease, as has also been observed in other, but not all, countries where comparative studies were performed. In addition, reports from New Zealand, United Kingdom, and Spain suggested that the clinical severity of the disease has attenuated in recent years [2, 8, 9]. These changes in prevalence and severity of the disease strongly suggest that environmental factors are involved in its pathogenesis.

Pathogenesis

Normal Bone Metabolism

The adult skeleton is continuously renewed throughout life by the process of bone remodelling. Old bone is removed by the osteoclasts whereas new bone is formed in the same location by the osteoblasts. This occurs in an orderly fashion through temporary anatomic structures called basic multicellular units (BMUs). A basic multicellular unit comprises a team of osteoclasts at the front and a team of osteoblasts at the back supported by blood vessels, nerves, and loose connective tissue. Osteoclasts and osteoblasts are derived from different precursors in the bone marrow. Osteoclasts originate from haematopoietic precursors of the monocyte/macrophage lineage while osteoblasts originate from multipotent mesenchymal stem cells, which give also rise to bone marrow stromal cells, chondrocytes, adipocytes, and muscle cells. The formation and lifespan of bone cells are controlled by mechanical, systemic, and local factors. Important regulators of osteoclast formation and activity belong to a ligand/receptor/soluble (decoy) receptor system involving proteins of the TNF receptor superfamily [10, 11]. These are RANK ligand, RANK, and OPG. RANKL is produced by osteocytes and osteoblastic/stromal cells, reacts with RANK, which is localized in haematopoietic osteoclast precursors, stimulates the formation and activity of osteoclasts, and prolongs their lifespan. OPG is a soluble receptor which counteracts the biological effects of RANKL preventing its binding to RANK and thereby suppressing bone resorption.

Pathology

Paget's disease of bone is a focal disorder of bone remodelling characterized by an increase in the number and size of osteoclasts in affected sites while the rest of the skeleton remains normal. The typically large osteoclasts, which may contain up to 100 nuclei per cell, induce excessive bone resorption associated with an increased recruitment of osteoblasts to the remodelling sites, resulting in increased bone formation and, hence, an overall increase in the rate of bone turnover. The increase in bone formation is thought to be secondary to the increased rate of bone resorption due to the coupling of the two processes. Some evidence, however, suggests that osteoblastic/ stromal cells might also be primarily affected in Paget's disease and contribute to the increased rate of bone formation [12, 13]. The accelerated rate of bone turnover is responsible for the deposition of bone with disorganized architecture and structural weakness. The bone packets lose their lamellar structure and are replaced by woven bone with a characteristic mosaic pattern while bone marrow is infiltrated by fibrous tissue and blood vessels.

Cell Biology

In clinical studies the likelihood of a bone being affected by Paget's disease was related to the amount of bone marrow present in that bone, leading to the postulation that the development of bone lesions may be related to specific properties of pagetic bone marrow [14]. In bone marrow cultures from patients with Paget's disease the rate of formation of osteoclasts and their number is markedly increased, suggesting that intrinsic abnormalities of the bone marrow microenvironment and/or of osteoclast precursors may contribute to the upregulation of osteoclastogenesis. A number of studies supported these notions and documented two major abnormalities. First, pagetic osteoclasts and their precursors express high levels of osteotropic factors (e.g. IL-6), a bone resorbing cytokine which has been proposed as a possible paracrine/autocrine factor contributing to the pathogenesis of the disease [13, 15, 16]. In addition, enhanced expression of RANKL was detected in bone marrow stromal cells from patients with Paget's disease and might contribute to the increased number of osteoclasts [17]. Second, compared to controls, bone marrow and peripheral cells from patients are hypersensitive to the action of RANKL and calcitriol [18, 19] and there is evidence suggesting that TAFII-17, a component of the transcription complex that binds vitamin D receptor, may be responsible for the hypersensitivity to calcitriol [20]. Thus, while the molecular characteristics of the cellular abnormalities of the disease are currently understood, the precise mechanism(s) that trigger these changes remain to be elucidated.

Aetiology

Several, not mutually exclusive, hypotheses have been proposed to explain the pathology of the disease, the most relevant being the viral and the genetic hypotheses. Studies of the distribution of bone lesions in patients with Paget's disease showed that the probability of a bone being affected is very similar to the probability of a bone being affected with haematogenous osteomyelitis, suggesting that the disease may be caused by a circulating infectious agent. An infection by a slow virus of the paramyxovirus family (measles virus, respiratory syncytial virus, canine distemper virus) was supported by the detection of nuclear and cytoplasmic inclusions resembling paramyxoviral nucleocapsids in osteoclasts and of measles virus nucleocapsid transcripts in bone marrow and peripheral blood monocytes from patients with the disease [21]. However, paramyxoviral-like structures have also been found in specimens from patients with other bone diseases, questioning the specificity of this finding. In addition, further search for viral presence in the osteoclasts provided conflicting results [22] and in a large cohort of patients with Paget's disease no evidence supporting an association between the disease and persistent infection with measles or other paramyxoviruses was found [23]. There is, however, good evidence that paramyxoviruses and viral proteins can promote the formation of osteoclasts with features similar to those of pagetic osteoclasts [24–26].

In familial aggregation studies the risk of first-degree relatives of patients with Paget's disease to develop the disorder was seven to ten times greater than the risk of individuals without such relatives. Furthermore, a positive family history has been reported in up to 40% of patients with Paget's disease [27–29]. Familial Paget's disease is inherited as an autosomal dominant trait and initial genetic analyses showed evidence of linkage to chromosome 18q21–22 in some families [30, 31]. This chromosome also contains the locus of the rare disease familial expansile osteolysis, which resembles Paget's disease and was found to be associated with activating mutations in the gene *TNFRSF11A*, which encodes RANK [32], while abnormalities of the same gene are responsible for another rare skeletal disease, expansile skeletal hyperphosphatasia [33]. Subsequent studies, however, failed to detect such mutations in patients with familial or sporadic Paget's disease. Other abnormal genes that have been identified in diseases with bone phenotypes similar to that of Paget's disease include *TNFRSF11B*, which encodes OPG in juvenile Paget's disease [34], and *VCP*, which encodes p97 in the syndrome of inclusion body myopathy associated with Paget's disease of bone and frontotemporal dementia [35]. All these genetic defects have in common the up-regulation of the NF-kB signal transduction, an essential process in the differentiation and activation of osteoclasts. These genes have also been investigated in patients with familial or sporadic Paget's disease but no mutations were identified. Analysis of families with Paget's disease identified further possible loci in other chromosomes indicating genetic heterogeneity. However, studies in different parts of the world identified mutations in the *SQSTM1* gene, located on chromosome 5q35, in up to 50% of patients with familial Paget's disease and up to 10% of those with sporadic disease [3, 22, 28, 36]. *SQSTM1* encodes p62, an adaptor protein that binds ubiquitin and is involved in various signalling pathways including the NF-κB pathway. The most common mutation of *SQSTM1* associated with Paget's disease (P329L) has been detected in patients from different European countries suggesting a founder gene defect [37]. In addition, animals overexpressing this mutation in cells of the osteoclast lineage formed more osteoclasts, which were hypersensitive to RANKL but did not develop bone lesions resembling those of Paget's disease in one study while in another they did [22]. In addition, genetic studies of patients with Paget's disease and no *SQSTM1* mutations identified seven loci that contribute substantially to the risk of developing the disease as well as affecting its severity in combination with mutations of *SQSTM1*. These variants include genes important in the differentiation and function of osteoclasts (e.g. *CSF1*) but also genes that have not been implicated in the regulation of bone metabolism [38–41].

The functional significance of these polymorphisms is unknown. Whether mutations of genes associated with Paget's disease are the cause of the disease or whether individuals with a mutation have an increased susceptibility to the disease when exposed to environmental factors, such as paramyxoviruses, is currently unclear.

Clinical Manifestations

The most commonly affected bones are the pelvis, the spine, the femora, and the skull but practically any bone of the skeleton may be affected and there is remarkable similarity in the frequency of affected bones in large series of patients from different countries [1, 42, 43]. About one-third of patients have only one lesion (Figure 4.10.1) but the frequency of single lesions varies among series, probably reflecting referral patterns, and is higher in asymptomatic patients. Notably, the anatomical spread of skeletal lesions in Paget's disease is log normally distributed and the commonly used distinction in clinical practice to monostotic and polyostotic disease is not pathophysiologically appropriate. The anatomical spread of the disease is not related to age or gender, shows no particular symmetry in the body, and remains largely unchanged throughout life. The disease progresses slowly within the affected bone but does not generally appear in other bones. Patients with limited bone involvement should, therefore, be reassured that the disease will not progress to other bones with time.

The majority of patients are asymptomatic and the disease may be diagnosed incidentally during investigation of an unrelated complaint by skeletal radiographs or by the finding of an unexplained elevation of serum alkaline phosphatase activity [44]. About 5–10% of affected patients have symptoms. Skeletal morbidity in Paget's disease is determined by the damage caused and the progression of the disease in affected sites as well as by the number and the localization of the lesions. Extensive disease, as originally described by Sir James Paget, occurs in about 5% of symptomatic patients. This is in agreement with the limited chance of an individual to develop extensive disease, as predicted by the distribution of lesions, but changing patterns of the disease to milder forms may also contribute to that.

The symptoms and complications of Paget's disease,(summarized in Table 4.10.1) [34], can have great impact on the quality of life of affected individuals [45, 46]. In the majority of patients, the presenting complaint is pain. This is related to the extent and site of the disease, it is usually persistent and present at rest, but is not specific. Pain due to secondary osteoarthritis is common and may hamper assessment of the relative contribution of bone and joint pains to the patient's disability. The origin of such pain can be assessed only retrospectively after treatment which reduces mainly the disease-related pain, having a rather limited effect on the arthritic pain. Deformities are present in about 15% of patients at the time of diagnosis and affect mainly weight-bearing bones, the most common deformity being bowing of the lower limbs. About 9% of patients present with fractures, which can be complete or fissure (incomplete) fractures. The latter occur more frequently, can be multiple, can cause pain, and may develop to complete fractures. Fractures heal generally well although in an older report of 182 fractures of the femur the incidence of non-union was 40% [47]. The skin overlying an affected bone may be warm as a result of increased

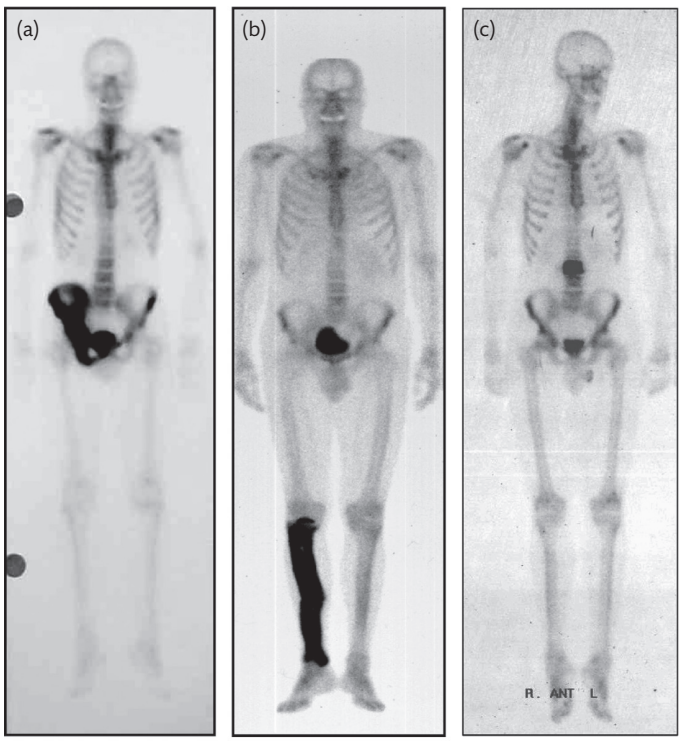


Figure 4.10.1 Monostotic Paget's disease illustrated by bone scintigraphy: (a) right pelvis; (b) right tibia (with deformity and fracture); (c) vertebra.

Table 4.10.1 Symptoms and complications of Paget's disease of bone

System	Complication
Musculoskeletal	Bone pain Bone deformity Osteoarthritis of adjacent joints Acetabular protrusion Fractures Spinal stenosis
Neurological	Hearing loss Cranial nerve deficits (rare) Basilar impression Increased cerebrospinal fluid pressure Spinal stenosis Vascular steal syndrome
Cardiovascular	Congestive heart failure and angina Increased cardiac output Aortic stenosis Generalized atherosclerosis Endocardial calcification
Metabolic	Immobilization hypercalcaemia Hypercalciuria Hyperuricaemia Nephrolithiasis
Neoplasia	Sarcoma (osteosarcoma, chondrosarcoma, and fibrosarcoma) Giant cell tumour

blood flow and bone turnover locally and hypervascularity of affected bones may cause ischaemia of adjacent structures (steal syndrome). Irreversible hearing loss is the most common neurological complication occurring in about one-third of patients with skull involvement. This is thought to be related to structural and/or density changes in the cochlear capsule bone [48]. Neoplastic transformation of pagetic bone is rare (less than 1%) and includes osteosarcoma and giant cell tumours associated more frequently with familial clustering of the disease [49]. Pathogenic mutations of the *ZNF687* gene were identified in a family with Paget's disease with increased prevalence of giant cell tumours of bone but also in patients with or without *SQSTM1* mutations [50, 51]. The *ZNF687* gene is highly expressed during osteoclast and osteoblast differentiation and patients with missense mutations of the gene had an earlier clinical expression and more severe bone disease.

Investigations

Radiographic changes are characteristic of the disease (Figure 4.10.2). Increased bone resorption may be detected as a decrease in the density of affected bones; sometimes a wedge- or flame-segment of bone resorption may be seen in long bones and extensive osteolytic areas in the skull (osteoporosis circumscripta). The osteolytic changes in long bones progress at a rate of about 1 cm/year. Older lesions usually have a mixed sclerotic and lytic appearance and in the last stage of the disease sclerotic lesions predominate. The involved parts of the skeleton are enlarged and deformed, and the cortex can be thickened and dense. The radiological changes can be considered pathognomonic but in some cases differential diagnosis may include fibrous dysplasia and bone metastases, particularly from prostate cancer. Bone scintigraphy is used to assess the extent of the disease. It is not specific but it is more sensitive than plain radiographs; up to 15% of lesions detected by bone scintigraphy may have normal radiographic appearance. Bone scintigraphy should always be included in the investigation of patients with Paget's disease and plain radiographs of the areas of increased radioisotope uptake should be subsequently made to confirm the diagnosis (Figure 4.10.3).

The pathology of Paget's disease is reflected in the proportional increase in biochemical indices of bone turnover [52]. In the past, urinary hydroxyproline excretion was used as an index of bone resorption and serum total alkaline phosphatase activity as an index of bone formation. These can be markedly increased in patients with extensive disease but can be also found within the reference range in patients with limited bone involvement. Patients with skull involvement tend to have the highest values of serum alkaline phosphatase activity. More specific and sensitive biochemical indices of bone formation include the bone-specific isoenzyme of alkaline phosphatase and the N-terminal extension peptide of collagen type I (procollagen I N-terminal peptide, P1NP). Serum osteocalcin concentrations are within the normal range in about half of the patients with elevated serum alkaline phosphatase values and should not be used in the management of patients with Paget's disease. Urinary hydroxyproline is neither specific nor sensitive enough and its determination depends on specific dietary advice. Peptides of the cross-linking domains of collagen type I, such as the N-telopeptide (NTX) or the C-telopeptide (CTX), measured in urine or serum are the most sensitive biochemical markers of bone resorption. Impaired isomerization of C-telopeptide has been reported in patients with Paget's disease but not in patients with increased bone turnover from other causes, leading to the postulation that this abnormality may reflect the defect in bone structure [53]. Degradation products of collagen type II are not increased in urines of patients [54].

In Paget's disease, despite the marked changes in the rate of bone turnover, extracellular calcium homeostasis is generally maintained but some disturbances may occur. Hypercalcaemia may develop in immobilized patients with active, extensive disease or may be due to concurrent primary hyperparathyroidism, the incidence of which

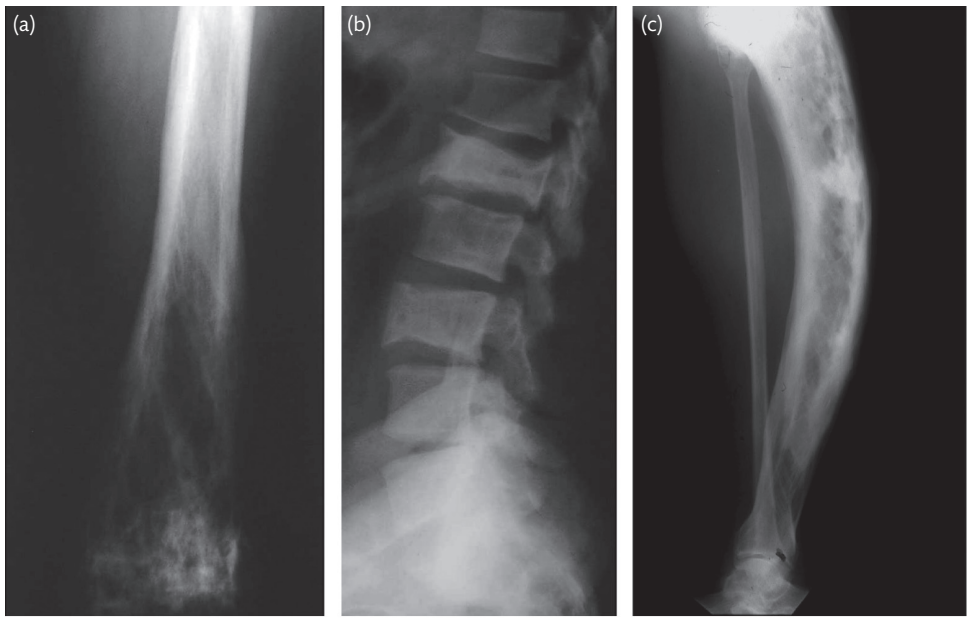


Figure 4.10.2 Radiographs of patients with Paget's disease: (a) distal femur showing extensive and flame-shaped osteolysis; (b) lumbar spine; (c) tibia with characteristic deformity.

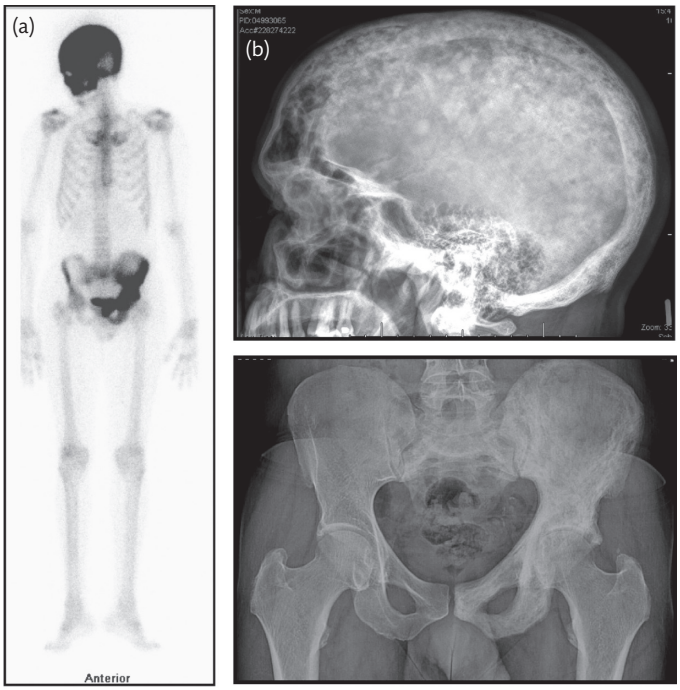


Figure 4.10.3 Bone scintigram of a patient with Paget's disease showing two areas of increased uptake of the isotope. Radiographs of these areas were diagnostic.

is thought to be higher in Paget's disease compared to the general population. Secondary hyperparathyroidism is present in about 20% of patients while serum concentrations of calcitriol are generally normal. Hypercalciuria and renal stone disease occur also more frequently in patients with Paget's disease.

Management

During the past 40 years, the management of patients with Paget's disease has changed dramatically due to the discovery of the therapeutic potential of the calcitonins and later of the bisphosphonates. Other, less frequently used treatments were plicamycin (mithramycin) and gallium nitrate. Bisphosphonates are currently the preferred treatment of Paget's disease.

Aims and Indications of Treatment

Classically treatment is given to patients with Paget's disease to relieve symptoms and improve their quality of life. The disease, however, is progressive and patients with symptoms were previously asymptomatic (Figure 4.10.4). It is currently impossible to identify patients who will develop symptoms and complications or to quantify the risk of complications in an individual. While this may change in the future with the use of genetic markers, anecdotal evidence indicates that patients with localizations close to large joints, in weight-bearing bones and the skull or with high levels of serum alkaline phosphatase activity are at higher risk for complications with time. Treatment with potent bisphosphonates does not only relieve symptoms due to the disease but restores bone quality and improves or even normalizes radiological appearances. Moreover, the bulk of evidence obtained with bisphosphonates strongly suggests that complications can be prevented if bone turnover is adequately suppressed, whereas there are indications that the contrary is true if bone turnover does not normalize [55]. Firm evidence, however, from prospective randomized controlled trials is lacking. A study of a large cohort of patients randomized to intensive bisphosphonate treatment (risedronate) or symptomatic management reported no differences in clinical outcomes between the two groups [56, 57]. Limitations of the study included the already advanced disease in most of the patients, use of bisphosphonates by the majority of patients before trial entry, and the fact that the disease was in biochemical remission in about half of the patients [58, 59].

These considerations led to the following treatment recommendations: (1) symptomatic disease, to reduce complaints and improve the quality of life; (2) preoperative treatment in preparation for an orthopaedic procedure on pagetic bone to reduce the increased blood flow and excessive bleeding; (3) treatment of asymptomatic patients with skeletal localizations at higher risk of future complications, such as those adjacent to large joints, in the skull, the spine, and the weight-bearing bones; and (4) young patients. Studies of treatment of asymptomatic patients with *SQSTM1* mutations are underway. The goal of treatment should be to normalize bone turnover, suppress serum alkaline phosphatase activity well within the normal range, and keep it adequately suppressed, if necessary with additional courses of treatment. Retreatment is generally advocated when a previously normal value of serum alkaline phosphatase activity exceeds the upper limit of normal or when it increases by 20–25% above its nadir value.

Bisphosphonates

The following properties render bisphosphonates as ideal agents for the treatment of Paget's disease: selective uptake at active skeletal sites; specific inhibition of bone resorption; short plasma half-life and lack of circulating metabolites; and persistence of the effect after stopping treatment. The general structure of the molecule of germinal bisphosphonates allows numerous substitutions, which has led to the synthesis of a variety of compounds with considerable differences in potency, activity to toxicity ratio, and mechanism of action [60]. Bisphosphonates are divided into two groups according to the presence or absence of a nitrogen atom in the molecule. The nitrogen increases the potency of the bisphosphonates and determines their mechanism of action. Compounds without a nitrogen atom in the side chain are etidronate, clodronate, and tiludronate. Nitrogen-containing bisphosphonates include alendronate, ibandronate, incadronate, neridronate, olpadronate, pamidronate, risedronate, and zoledronate. Practically all bisphosphonate, either approved or in clinical development, have been used in the treatment of Paget's disease, which in turn has served as a human model for investigating the pharmacological properties of these agents. Bisphosphonates approved around the world for the treatment of Paget's disease are listed in Table 4.10.2.

Pharmacodynamics

For the design of optimal therapeutic strategies of Paget's disease with bisphosphonates their pharmacodynamic properties need to be taken into consideration [61]. When a potent bisphosphonate is given to a patient with Paget's disease, the first measurable effect is the suppression of bone resorption. This occurs within a few days of starting treatment. During this initial period, bone

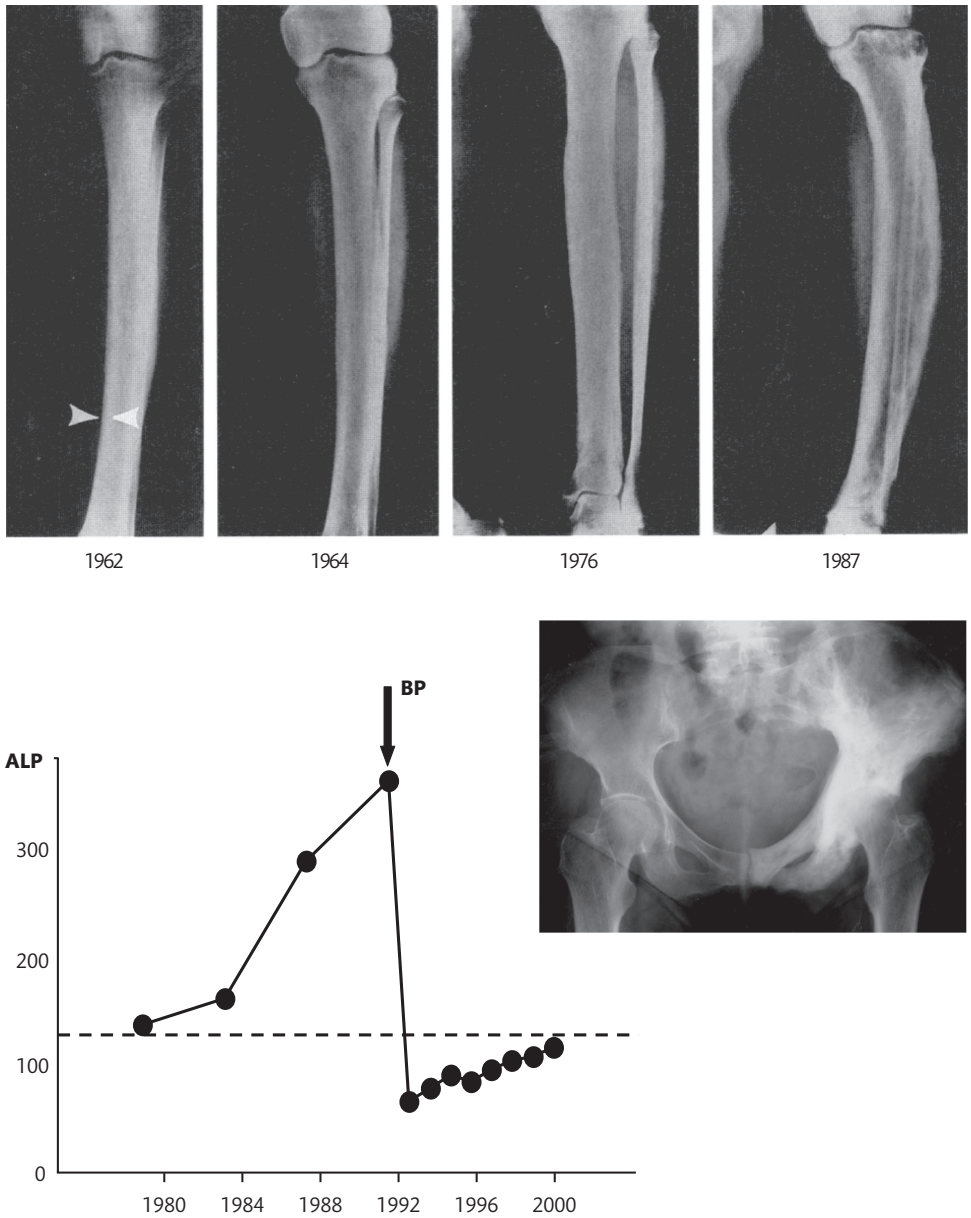


Figure 4.10.4 (upper panel) Serial radiographs (anteroposterior view) of the tibia of an untreated 68-year-old man with Paget's disease illustrating the progression of the disease. (lower panel) Sequential measurements of serum alkaline phosphatase (ALP) activity (U/L) over 20 years in a 51-year-old woman with Paget's disease of the pelvis. Note the progressive threefold increase in serum ALP activity on no treatment. Arrow indicates treatment with oral olpadronate 200 mg/day for 1 month inducing complete, long-lasting remission. Horizontal line represents the upper limit of the normal range. At the time of intervention, the patient had already developed osteoarthritis and required total hip arthroplasty despite successful treatment. BP, bisphosphonate.

formation does not change. This will decrease secondarily, at a slower rate, due to the coupling of bone resorption to bone formation, so that a new equilibrium will be reached after 3–6 months (Figure 4.10.5). Thus, adequate suppression of bone resorption will be predictably followed by an adequate suppression of bone formation. Suppression of biochemical indices of bone resorption early during the course of treatment provides, therefore, an indication of the pharmacological efficacy of the bisphosphonate and can subsequently determine the length of treatment [62]. Because of the predictable changes in bone remodelling that follow bisphosphonate therapy in Paget's disease, it is not necessary to prolong treatment until the lowest level of serum alkaline phosphatase is reached and short courses are usually sufficient to achieve remissions. Moreover, the retention of bisphosphonate in the skeleton is proportional to disease activity and inversely proportional to renal function [63]. Therefore, dose adjustments may be required in patients with impaired renal function, but no specific studies have addressed this issue. In addition, the wide variability of disease activity of affected patients strongly suggest that treatment needs to be individualized.

Table 4.10.2 Bisphosphonates approved for the treatment of Paget's disease

Generic name	Dose
Alendronate	Oral, 40 mg daily for 6 months
Clodronate	Oral, 1600 mg daily for 3 to 6 months
Etidronate	Oral, 400 mg daily for 6 months
Neridronate	Intravenous 100 mg daily for 2 days
Pamidronate	Intravenous, 30–60 mg daily for 3 days[a]
Risedronate	Oral, 30 mg daily for 2 months
Tiludronate	Oral, 400 mg daily for 3 months
Zoledronate	Intravenous, 5 mg (one 15 min infusion)[b]

[a] Lower dose recommended by the pharmaceutical industry, higher dose recommended by investigators; [b] Preferred treatment.

The long-term efficacy of treatment is best assessed by measuring biochemical indices of bone formation, serum alkaline phosphatase activity being still the most commonly used. In the past, the efficacy of treatment was evaluated by its ability to decrease serum alkaline phosphatase activity by more than 50% of its initial value. With the available potent bisphosphonates, this is no longer appropriate and treatment efficacy should be assessed only by its ability to decrease serum alkaline phosphatase values to the normal range (remission). In clinical practice there is no need to measure serum alkaline phosphatase activity earlier than 3 months after the start of treatment, 6 months being the optimal time.

During the initial phase of bisphosphonate treatment, when bone resorption and bone formation are still dissociated, the increased retention of calcium in the skeleton leads to changes in calcium metabolism. There is a fall in serum calcium concentration, which stimulates the secretion of parathyroid hormone secretion and consequently the renal production of calcitriol. These hormones, in turn, increase the renal tubular reabsorption of calcium (parathyroid hormone) and its intestinal absorption (calcitriol). The result is a marked, but transient, increase in calcium balance. The concomitant decrease in serum phosphate concentrations is due to the renal action of parathyroid hormone. Such responses are not observed during etidronate treatment, which has a weak action on bone metabolism. With the attainment of the new equilibrium of bone remodelling, calcium balance returns towards pretreatment levels and the values of the biochemical indices of calcium metabolism normalize. The adaptive changes of calcium metabolism to the marked alterations in bone remodelling prevent the development of symptomatic hypocalcaemia in calcium-replete patients. However, elderly patients frequently have calcium-deficient diets and some investigators advocate the use of calcium supplements during treatment of Paget's disease with potent bisphosphonates, especially if these are given intravenously or the disease is very active. Support for this logical assumption by clinical trials is, however, limited.

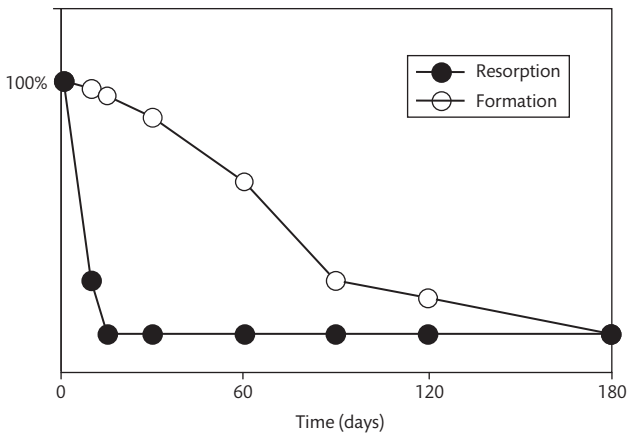


Figure 4.10.5 Schematic presentation of the changes in biochemical indices of bone resorption and bone formation following bisphosphonate treatment of Paget's disease.

Treatment Responses

Clinical responses to treatment include the disappearance or clear improvement of pain in more than 80% of treated patients, when this is due to the activity of the disease. A decrease of bone pain is generally observed 1 to 3 months after the start of treatment and the effect is maximal after 6 months and is maintained for as long as biochemical indices of bone turnover remain within the normal range. Soon after the start of therapy with a potent bisphosphonate, particularly if given intravenously, there may be a transient increase in pain at affected sites and patients should be reassured. Pain due to osteoarthritis is unresponsive to treatment in about 75% of patients; non-steroidal anti-inflammatory drugs can then be used. If the hip joint is affected, hip arthroplasty may be required to control the symptoms. Back pain resulting from involvement of lumbar vertebrae is frequently not relieved by treatment. About half of the patients with pain associated with deformity of the femur or the tibia will respond favourably to bisphosphonate therapy but pain may persist and a corrective osteotomy may be necessary. Deafness is usually not affected but its progression appears to be arrested. There have been also reports of improvement of spinal cord compression with bisphosphonate therapy and fracture frequency of pagetic bones appears to decrease with treatment.

Improvement in bone histology and formation of bone with normal lamellar structure and no evidence of a mineralization defect has been reported with currently used bisphosphonates. Radiologically, an arrest of the progression of the disease is usually seen. Radiological improvement can be dramatic, however, if lesions are lytic and are localized in long bones or in the skull. In other areas, improvement is slow and sometimes difficult to demonstrate by non-experienced radiologists. Treatment induces an exponential decrease in isotope uptake on bone scintigrams. However, even with normalization of disease activity, only about 10–30% of lesions normalize completely and residual uptake (up to 20% of the original) is detected [64]. Notably, more recent case studies showed normalization of radioisotope uptake in affected bones in most patients treated with zoledronate [65]. The possible relation of these scintigraphic changes to future recurrences has not been adequately studied but some investigators advocate normalization of bone scintigrams as one of the aims of treatment.

These clinical, histological, and radiological responses emphasize the need for an intervention with a bisphosphonate early in the course of the disease and before the development of complications.

All bisphosphonates given to patients with Paget's disease significantly decrease biochemical indices of bone turnover [66–73]. Considerable differences exist, however, in their ability to induce remissions. Generally, potent bisphosphonates induce better

responses. Head to head clinical trials have been performed with etidronate 400 mg daily for 6 months as comparator. In all these clinical trials, etidronate was less effective. The limited efficacy, relative to other bisphosphonates, together with the increased risk of osteomalacia, have made etidronate a treatment of the past. Normalization of serum alkaline phosphatase activity has been reported with tiludronate 400 mg daily for 3 months (35%), clodronate 1600 mg daily for 6 months (up to 70%), alendronate 40 mg daily for 6 months (63%), risedronate 30 mg daily for 2 months (up to 70%), and pamidronate, intravenously or orally in variable regimens (up to 90%). It should be noted that comparison of results obtained in different studies is not appropriate due to different selection criteria and disease activity of treated patients. The results of these studies showed, in addition, that despite the availability of effective and convenient treatment regimens with bisphosphonates, there was still need for further improvement. More recently, the efficacy and tolerability of a single 15-min intravenous infusion of zoledronate was compared to oral risedronate 30 mg per day for 2 months in patients with Paget's disease of moderate activity [74]. Results showed that zoledronate was significantly more efficacious than risedronate in inducing biochemical remission that lasts long [75] associated with improvements in some aspects of the quality of life of the patients. Zoledronate is currently considered the treatment of choice of Paget's disease.

Follow-up of patients in remission is indicated every 6–12 months. Remissions, estimated from the time of normalization of serum alkaline phosphatase activity, can be long and can last even longer than 10 years in some patients. We have observed, however, recurrences 12 or 13 years after induction of complete biochemical remission which illustrates the need for continuous follow-up. The duration of remission is determined by the degree of suppression of serum alkaline phosphatase activity and the number of affected bones but is not related to the length or to the mode of treatment (oral or intravenous) as long as a potent, efficacious bisphosphonate is given [70, 76–78]. The lower the serum alkaline phosphatase activity reached with treatment, the longer the period of remission. Suppression of serum alkaline phosphatase activity well within the normal range is a prerequisite for long-term remissions and should be part of treatment strategies. Notably, in a study of 152 (83% with pain complaints) patients with Paget's disease and median time of biochemical remission 78.9 months after short-term treatment with the bisphosphonate olpadronate, worsening of pain scores were significantly related to biochemical relapse probability [78]. These observations underscore the relationship between biochemical indices of bone turnover with pain and the need to retain them within the normal range.

Resistance to Bisphosphonate Treatment

Impaired response to repeated treatment courses with bisphosphonates is usually referred to as acquired resistance and should be distinguished from an intrinsic resistance to a particular compound. Acquired resistance has been reported for etidronate and pamidronate [79, 80] but the underlying mechanism is not known and it is important to differentiate between real and apparent resistance. Some patients may not respond to oral bisphosphonate but may show a prompt response to the same compound given intravenously. In such cases, factors interfering with the already low intestinal absorption of the drug are most likely responsible for the impaired response to oral treatment. Patients retreated with the same bisphosphonate during a recurrence of their disease may show a reduced fractional decrease in biochemical indices of bone turnover compared to earlier treatments. Some consider this response compatible with development of resistance to therapy. However, it has been shown in studies with clodronate and pamidronate that the actual level, rather than the fractional decrease of biochemical indices of bone turnover following every treatment, should be compared to those obtained after the initial therapy (Figure 4.10.6). This is because patients who are offered a new treatment course have generally a lower rate of bone turnover compared to that before the first treatment. Finally, in patients with Paget's disease and concurrent hyperparathyroidism, completeness of response is generally less and recurrences occur quicker which might be considered reduced responsiveness. For optimal responses of patients with Paget's disease and autonomous hyperparathyroidism to bisphosphonates,

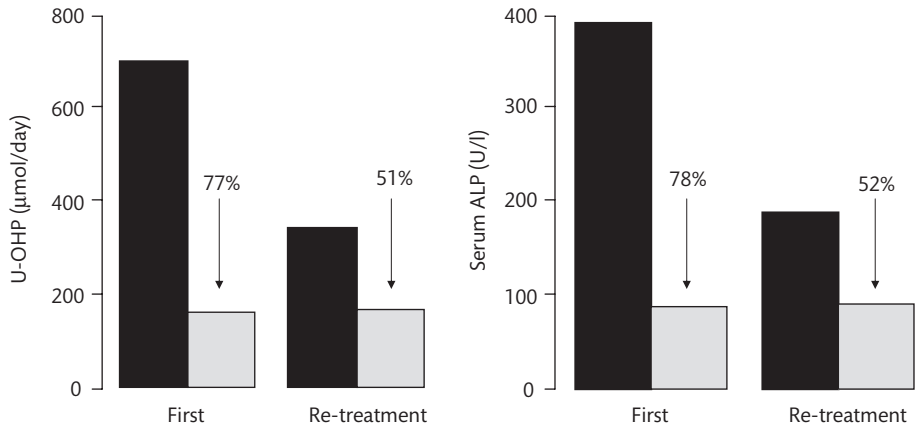


Figure 4.10.6 Apparent resistance to bisphosphonate therapy in Paget's disease. Absolute and percent changes of urinary hydroxyproline excretion (UOHP) and serum alkaline phosphatase (ALP) activity after first treatment with pamidronate or retreatment with the same bisphosphonate for a recurrence of the disease.

Modified from Harinck HI, Bijvoet OL, Blanksma HJ, Dahlinghaus-Nienhuys PJ. Efficacious management with aminobisphosphonate (APD) in Paget's disease of bone. *Clin Orthop Relat Res*, 1987; 217: 79–98. Copyright © 1987, © Lippincott-Raven Publishers.

parathyroidectomy should be considered. However, real resistance to pamidronate can develop. We showed, for example, progressive reduction in responsiveness to this bisphosphonate, which was mainly related to the extent of skeletal involvement but not to the dose of pamidronate or to the biochemical activity of the disease. Patients with three or more affected bones were most likely to develop resistance to pamidronate, a finding consistent with other reports. These patients respond readily to other bisphosphonates. There are scarce data of resistance to other nitrogen-containing bisphosphonates. Using the same approach as in the pamidronate studies, we found no resistance to consecutive treatments of patients with Paget's disease with olpadronate. Thus, within the limitations of existing studies, it appears that acquired resistance is specific for pamidronate and is limited to patients with extensive disease. Such resistance does not seem to occur with other nitrogen-containing bisphosphonates but the evidence for that is still weak. Finally, primary resistance to a specific bisphosphonate, if it exists, is rare.

Adverse Effects

All bisphosphonates given at very high doses can impair the mineralization of newly formed bone and induce osteomalacia. In clinical practice this is, however, relevant only for etidronate. Doses of potent nitrogen-containing bisphosphonates that induce osteomalacia exceed, by many orders of magnitude, those required for effective suppression of bone turnover. Consequently, in all reported controlled studies no adverse effects on bone mineralization have been observed. In only a few patients treated with intravenous pamidronate, at doses higher than those recommended, has impaired bone mineralization been reported. Histological osteomalacia induced by either etidronate or pamidronate is reversible.

In some patients treated for the first time with nitrogen-containing bisphosphonates there is a rise in body temperature and flu-like symptoms during the first 3 days of treatment. These symptoms are transient and subside with no specific measures even when treatment is continued [81, 82]. This response is dose-dependent and is associated more frequently with intravenous than oral treatment. Moreover, it does not generally recur upon retreatment, and, if it does, it is of lower intensity. Previous exposure to another nitrogen-containing bisphosphonate, but not to etidronate, precludes the development of this response. Laboratory findings are consistent with an acute phase reaction [82]. There is a transient decrease in blood lymphocytes and a transient increase in serum C-reactive protein, possibly due to increases in proinflammatory cytokines, such as IL-6 and TNF-α produced by γ,δ T lymphocytes in response to a metabolite of the mevalonate pathway, upstream farnesyl pyrophosphate synthase, which is inhibited by nitrogen-containing bisphosphonates [82, 83]. Rarely, high doses of nitrogen-containing bisphosphonates may induce ophthalmic reactions such as conjunctivitis, iritis, or uveitis. There are case reports of ototoxicity and central nervous toxicity after intravenous pamidronate. Allergic skin reactions have been occasionally observed with most of the bisphosphonates.

Mild gastrointestinal complaints occur with low frequency with the use of all bisphosphonates. Some nitrogen-containing bisphosphonates can induce more severe symptoms such as heartburn, nausea, and vomiting in a few patients, associated with oesophagitis or gastritis. The use of oral alendronate 40 mg daily was associated with higher frequency of epigastric complaints in an open, but not in a controlled, study and the latter was also the case with oral risedronate 30 mg daily. In a comparative study of alendronate 40 mg/day and risedronate 30 mg/day gastric ulcers and/or large numbers of gastric erosions were detected endoscopically in approximately 3% of patients, and their occurrence was comparable with both bisphosphonates. Nitrogen-containing bisphosphonates should be administered orally with one full glass of water and the patient should remain in an upright position for half an hour to allow quick passage through the oesophagus and to avoid oesophageal irritation. Rapid intravenous injection of bisphosphonates may chelate calcium in the circulation and form complexes, which can be nephrotoxic or can damage directly the renal tubule. Bisphosphonates should, therefore, be given by slow infusion. Zoledronate is administered by short intravenous infusion (15 min) because of the low effective dose. Aminobisphosphonates should not be injected intramuscularly because they can cause severe local irritation and necrosis but clodronate has been given intramuscularly. Osteonecrosis of the jaw is extremely rare in patients with Paget's disease treated either with oral or intravenous bisphosphonates.

Notably, impaired renal function is a contraindication for bisphosphonate treatment and zoledronate should not be administered to patients with creatinine clearance <35 ml/min. There is, therefore, an unmet therapeutic need of such patients. Anecdotal evidence suggests that treatment with the RANKL inhibitor, denosumab, which is not eliminated by the renal route, might be efficacious in patients with Paget's disease [84], an observation that needs, however, to be confirmed in appropriate patient groups.

REFERENCES

1. Kanis JA. *Pathophysiology and Treatment of Paget's Disease of Bone*, 2nd edn. London: Dunitz, 1998.
2. Siris ES, Roodman GD. Paget's disease of bone. In: Rosen CJ, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, 7th edn. Washington DC: American Society for Bone and Mineral Research, 2008: 335–43.
3. Cundy T, Bolland M. Paget disease of bone. *Trends Endocrinol Metab*, 2008; **19**: 246–53.
4. Corral-Gudino L, Borao-Cengotita-Bengoa M, Del Pino-Montes J, Ralston S. Epidemiology of Paget's disease of bone: a systematic review and meta-analysis of secular changes. *Bone*, 2013; **55**: 347–52.
5. Altman RD, Bloch DA, Hochberg MC, Murphy WA. Prevalence of pelvic Paget's disease of bone in the United States. *J Bone Miner Res*, 2000; **15**: 461–5.
6. Barker DJP, Clough PWL, Guyer PB, Gardner MJ. Paget's disease of bone in 14 British towns. *BMJ*, 1977; **1**: 1181–3.
7. Cooper C, Schafheutle K, Dennison E, Kellingray S, Guyer P, Barker D. The epidemiology of Paget's disease in Britain: is the prevalence decreasing? *J Bone Miner Res*, 1999; **14**: 192–7.
8. Corral-Gudino L, García-Aparicio J, Sánchez-González MD, *et al*. Secular changes in Paget's disease: contrasting changes in the number of new referrals and in disease severity in two neighboring regions of Spain. *Osteoporos Int*, 2013; **24**: 443–50.
9. Cundy T. Paget's disease of bone. *Metabolism*, 2018; **80**: 5–14
10. Suda T, Takahashi N, Udagawa N, Jimi E, Gillespie MT, Martin TJ. Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr Rev*, 1999; **20**: 345–57.

11. Kearns AE, Khosla S, Kostenuik PJ. Receptor activator of nuclear factor κB ligand and osteoprotegerin regulation of bone remodeling in health and disease. *Endocr Rev*, 2008; **29**: 155–92.
12. Gehron-Robey P, Bianco P. The role of osteogenic cells in the pathophysiology of Paget's disease. *J Bone Miner Res*, 1999; **14** (Suppl 2): 9–16.
13. Naot D, Bava U, Mattews B, *et al*. Differential gene expression in cultured osteoblasts and bone marrow stromal cells from patients with Paget's disease of bone. *J Bone Miner Res*, 2007; **22**: 298–309.
14. Reddy SV, Menaa S, Singer FR, Demulder A, Roodman GD. Cell biology of Paget's disease. *J Bone Miner Res*, 1999; **14**(Suppl 2): 3–8.
15. Hoyland JA, Freemond AJ, Sharpe PT. Interleukin-6, IL-6 receptor, and IL-6 nuclear factor gene expression in Paget's disease. *J Bone Miner Res*, 1994; **9**: 75–80.
16. Roodman GD, Kurihara N, Ohsaki Y, *et al*. Interleukin-6: a potential autocrine/paracrine factor in Paget's disease of bone. *J Clin Invest*, 1992; **89**: 46–52.
17. Menaa C, Reddy SV, Kurihara N, Anderson D, Roodman GD. Enhanced RANK ligand expression and responsivity of bone marrow cells in Paget's disease of bone. *J Clin Invest*, 2000; **105**: 1833–8.
18. Neale SD, Smith R, Wass JA, Athanasou NA. Osteoclast differentiation from circulating mononuclear precursors in Paget's disease is hypersensitive to 1,25-dihydroxyvitamin D_3 and RANKL. *Bone*, 2000; **27**: 409–16.
19. Menaa C, Barsony J, Reddy SV, *et al*. 1,25-dihydroxyvitamin D3 hypersensitivity of osteoclast precursors from patients with Paget's disease. *J Bone Miner Res*, 2000; **15**: 228–34.
20. Kurihara N, Reddy SV, Araki N, *et al*. Role of TAFII-17, a VDR binding protein, in the increased osteoclast formation in Paget's disease. *J Bone Miner Res*, 2004; **19**: 1154–64.
21. Singer FR. Update on the viral etiology of Paget's disease of bone. *J Bone Miner Res*, 1999; **14**(Suppl 2): 29–33.
22. Ralston SH. Pathogenesis of Paget's disease of bone. *Bone*, 2008; **43**: 819–25.
23. Visconti MR, Uategui-Martin R, Ralston SH. Antibody response to paramyxoviruses in Paget's disease of bone. *Calcif Tissue Int*, 2017; **101**: 141–7.
24. Reddy SV, Kurihara N, Menaa C, *et al*. Osteoclasts formed by measles virus-infected osteoclast precursors from hCD46 transgenic mice express characteristics of Pagetic osteoclasts. *Endocrinology*, 2001; **142**: 2898–905.
25. Wang FM, Sarmasik A, Hiruma Y, *et al*. Measles virus nucleocapsid protein, a key contributor to Paget's disease, increases IL-6 expression via down-regulation of FoxO3/Sirt1 signaling. *Bone*, 2013; **53**: 269–76.
26. Teramachi J, Nagata Y, Mohammad K, *et al*. Measles virus nucleocapsid protein increases osteoblast differentiation in Paget's disease. *J Clin Invest*, 2016; **126**: 1012–22.
27. Siris ES, Ottman R, Flaster E, Kelsey JL. Familial aggregation of Paget's disease of bone. *J Bone Miner Res*, 1991; **6**: 495–500.
28. Eekhoff EWM, Karperien M, Houtsma D, *et al*. Familial Paget's disease in the Netherlands. *Arthritis Rheum*, 2004; **50**: 1650–4.
29. Morales-Piga AA, Rey-Rey JS, Corres-Gonzales J, Garcia-Sagredo JM, Lopez-Abente G. Frequency and characteristics of familial aggregation of Paget's disease of bone. *J Bone Miner Res*, 1995; **10**: 663–70.
30. Haslam SI, van Hul W, Morales-Piga A, *et al*. Paget's disease of bone: evidence for a susceptibility locus on chromosome 18q and for genetic heterogeneity. *J Bone Miner Res*, 1998; **13**: 911–17.
31. Hocking L, Slee F, Haslam SI, *et al*. Familial Paget's disease of bone: patterns of inheritance and frequency of linkage to chromosome 18q. *Bone*, 2000; **26**: 1095–103.
32. Hughes AE, Ralston SH, Marken J, *et al*. Mutations in TNFRSF11A, affecting the signal peptide of RANK, cause familial expansile osteolysis. *Nat Genet*, 2000; **24**: 45–8.
33 Whyte MP, Hughes AE. Expansile skeletal hyperphosphatasia is caused by a 15-base pair tendem dublication in TNFRSF11A encoding RANK and is allelic to familial expansile osteolysis. *J Bone Miner Res*, 2002; **17**: 26–9.
34. Whyte MP, Obrecht SE, Finnegan PM, *et al*. Osteoprotegerin deficiency and juvenile Paget's disease. *N Engl J Med*, 2002; **347**: 175–84.
35. Watts GD, Wymer J, Kovach MJ, *et al*. Inclusion body myopathy associated with Paget's disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet*, 2004; **36**: 377–81.
36. Laurin N, Brown JP, Lemainque A, *et al*. Paget's disease of bone: mapping of two loci at 5q35-qter and 5q31. *Am J Hum Genet*, 2001; **69**: 528–43.
37. Chung PY, Beyens G, Guañabens N, *et al*. Founder effect in different European countries for the recurrent P392L SQSTM1 mutation in Paget's disease of bone. *Calcif Tissue Int*, 2008; **83**: 34–42.
38. Albagha OM. Genetics of Paget's disease of bone. *Bonekey Rep*, 2015; **4**: 756.
39. Albagha OM, Wani SE, Visconti MR, *et al*. Genome-wide association identifies three new susceptibility loci for Paget's disease of bone. *Nat Genet*, 2011; **43**: 685–9.
40. Albagha OM, Visconti MR, Alonso N, *et al*. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. *Nat Genet*, 2010; **42**: 520–4.
41. Chung PY, Beyens G, Boonen S, *et al*. The majority of the genetic risk for Paget's disease of bone is explained by genetic variants close to the CSF1, OPTN, TM7SF4, and TNFRSF11A genes. *Hum Genet*, 2010; **128**: 615–26.
42. Harinck HIJ, Bijvoet OLM, Vellenga CJLR, Blanksma HJ, Frijlink WB. Relation between signs and symptoms in Paget's disease of bone. *Q J Med*, 1986; **58**: 133–51.
43. Meunier PJ, Salson C, Mathieu L, *et al*. Skeletal distribution and biochemical parameters of Paget's disease. *Clin Orthop Rel Res*, 1987; **217**: 37–44.
44. Eekhoff EWM, van der Klift M, Kroon HM, *et al*. Paget's disease of bone in the Netherlands: a population-based radiological and biochemical survey; the Rotterdam study. *J Bone Miner Res*, 2004; **19**: 566–70.
45. Lyles KW, Siris ES, Singer FR, Meunier PJ. A clinical approach to diagnosis and management of Paget's disease of bone. *J Bone Miner Res*, 2001; **16**: 1379–87.
46. Gold DT, Boisture J, Shipp KM, Pieper CF, Lyles KW. Paget's disease of bone and quality of life. *J Bone Miner Res*, 1996; **11**: 1897–904.
47. Dove J. Complete fractures of the femur in Paget's disease of bone. *J Bone Joint Surg Br*, 1980; **62-B**: 12–17.
48. Monsell EM, Cody DD, Bone HG, Divine GW. Hearing loss as a complication of Paget's disease of bone. *J Bone Miner Res*, 1999; **14**(Suppl 2): 92–5.
49. Rendina D, De Filippo G, Ralston SH, *et al*. Clinical characteristics and evolution of giant cell tumor occurring in Paget's disease of bone. *J Bone Miner Res*, 2015; **30**: 257–63.
50. Divisato G, Formicola D, Esposito T, *et al*. ZNF687 mutations in severe Paget disease of bone associated with giant cell tumor. *Am J Hum Genet*, 2016; **98**: 275–86.

51. Divisato G, Scotto di Carlo F, Petrillo N, Esposito T, Gianfrancesco F. ZNF687 mutations are frequently found in pagetic patients from South Italy: implication in the pathogenesis of Paget's disease of bone. *Clin Genet*, 2018; **93**: 1240–4.
52. Alvarez L, Peris P, Pons F, *et al*. Relationship between biochemical markers of bone turnover and bone scintigraphy indices in assessment of Paget's disease activity. *Arthritis Rheum*, 1997; **40**: 461–8.
53. Garnero P, Fledelius C, Gineyts E, Serre CM, Vignot E, Delmas PD. Decreased ß-isomerization of C-telopeptides of α1 chain of type I collagen in Paget's disease of bone. *J Bone Miner Res*, 1997; **12**: 1407–15.
54. Christgau S, Garnero P, Fledelius C, *et al*. Collagen type II C-telopeptide fragments as an index of cartilage degradation. *Bone*, 2001; **29**: 209–15.
55. Meunier PJ, Vignot E. Therapeutic strategies in Paget's disease of bone. *Bone*, 1995; **17**(Suppl 5): 489S–91S.
56. Langston AL, Cambell MK, Fraser WD, MacLennan GS, Selby PL, Ralston SH. Randomised trial of intensive bisphosphonate treatment versus symptomatic management in Paget's disease of bone. *J Bone Miner Res*, 2010; **25**: 20–31.
57. Tan A, Goodman K, Walker A, *et al*. Long-term randomized trial of intensive versus symptomatic management in Paget's disease of bone: the PRISM-EZ study. *J Bone Miner Res*, 2017; **32**: 1165–73.
58. Singer FR, Bone HG 3rd, Hosking DJ, Lyles KW, Reid IR, Siris ES. Response to letter by Ralston et al. *J Clin Endocrinol Metab*, 2015; **100**: L47–8.
59. Cundy T. Treating Paget's disease—why and how much? *J Bone Miner Res*, 2017; **32**: 1163–4.
60. Papapoulos SE. Bisphosphonates: how do they work? *Best Pract Res Clin Endocrinol Metab*, 2008; **22**: 831–47.
61. Papapoulos SE. Pharmacodynamics of bisphosphonates in man; implications for treatment. In: Bijvoet OLM, Fleisch HA, Canfield RE, Russell RGG, eds. *Bisphosphonates on Bones*. Amsterdam: Elsevier, 1995: 231–63.
62. Papapoulos SE, Frölich M. Prediction of the outcome of treatment of Paget's disease of bone with bisphosphonates from short-term changes in the rate of bone resorption. *J Clin Endocrinol Metab*, 1996; **81**: 3993–7.
63. Cremers SCLM, Eekhoff MEMW, den Hartigh J, Vermeij P, Papapoulos SE. Relationships between pharmacokinetics and rate of bone turnover after intravenous bisphosphonate (olpadronate) in patients with Paget's disease of bone. *J Bone Miner Res*, 2003; **18**: 868–75.
64. Vellenga CJLR. Quantitative bone scintigraphy in the evaluation of Paget's disease of bone. In: Bijvoet OLM, Fleisch HA, Canfield RE, Russell RGG, eds. *Bisphosphonates on Bones*. Amsterdam: Elsevier, 1995: 279–91.
65. Reid IR, Maslowski K. Long-term bone scintigraphy results after intravenous zoledronate in Paget's disease of bone. *Calcif Tissue Int*, 2017; **101**: 43–9.
66. Delmas PD, Meunier PJ. The management of Paget's disease of bone. *N Engl J Med*, 1997; **336**: 558–66.
67. Miller PD, Brown JP, Siris ES, Hoseyni MS, Axelrod DW, Bekker PJ. A randomized, double-blind comparison of risedronate and etidronate in the management of Paget's disease of bone. *Am J Med*, 1999; **106**: 513–20.
68. Reid IR, Nicholson GC, Weinstein RS, *et al*. Biochemical and radiological improvement in Paget's disease of bone treated with alendronate: a randomized, placebo-controlled trial. *Am J Med*, 1996; **101**: 341–8.
69. Roux C, Gennari C, Farrerons J, *et al*. Comparative prospective, double-blind, multicenter study of the efficacy of tiludronate and etidronate in the treatment of Paget's disease of bone. *Arthritis Rheumat*, 1995; **38**: 851–8.
70. Schweitzer DH, Zwinderman AH, Vermeij P, Bijvoet OLM, Papapoulos SE. Improved treatment of Paget's disease with dimethylaminohydroxypropylidene bisphosphonate. *J Bone Miner Res*, 1993; **8**: 175–82.
71. Siris E, Weinstein RS, Altman R, *et al*. Comparative study of alendronate versus etidronate for the treatment of Paget's disease of bone. *J Clin Endocrinol Metab*, 1996; **81**: 961–7.
72. Siris ES, Chines AA, Altman RD, *et al*. Risedronate on the treatment of Paget's disease of bone; an open label, multicenter study. *J Bone Miner Res*, 1998; **13**: 1032–8.
73. Brown JP, Chines AA, Myers WR, Eusebio RA, Ritter-Hrncirik C, Hays CW. Improvement of pagetic bone lesions with risedronate treatment: a radiologic study. *Bone*, 2000; **26**: 263–7.
74. Reid IR, Miller P, Lyles K, *et al*. Comparison of a single infusion of zoledronic acid and risedronate for Paget's disease. *N Engl J Med*, 2005; **353**: 898–908.
75. Cundy T, Maslowski K, Grey A, Reid IR. Durability of response to zoledronate treatment and competing mortality in Paget's disease of bone. *J Bone Miner Res*, 2017; **34**: 753–6
76. Harinck HIJ, Papapoulos SE, Blanksma HJ, Moolenaar AJ, Vermeij P, Bijvoet OLM. Paget's disease of bone; early and late responses to three different modes of treatment with aminohydroxypropylidene bisphosphonate. *BMJ*, 1987; **295**: 1301–5.
77. Eekhoff EMW, Zwinderman AH, Haverkort DMAD, Cremers SCLM, Hamdy NAT, Papapoulos SE. Determinants of induction and duration of remission of Paget's disease of bone after bisphosphonate (olpadronate) therapy. *Bone*, 2003; **33**: 831–8.
78. Reid IR, Lyles K, Su G, *et al*. A single infusion of zoledronic acid produces sustained remissions in Paget disease: data to 6.5 years. *J Bone Miner Res*, 2011; **26**: 2261–70.
79. Cutteridge DH, Ward LC, Stewart GO, *et al*. Paget's disease: acquired resistance to one aminobisphosphonate with retained response to another. *J Bone Miner Res*, 1999; **14** (Suppl 2): 79–84.
80. Papapoulos SE, Eekhoff EMW, Zwinderman AH. Acquired resistance to bisphosphonates in Paget's disease of bone. *J Bone Miner Res*, 2006; **21** (Suppl 2): 88–91.
81. Adami S, Zamberlan N. Adverse effects of bisphosphonates. *Drug Safety*, 1996; **14**: 158–70.
82. Schweitzer DH, Oostendorp-van de Ruit M, van der Pluijm G, Löwik CWGM, Papapoulos S. Interleukin-6 and the acute phase reaction during treatment of patients with Paget's disease with the nitrogen-containing bisphosphonate dimethylaminohydroxypropylidene bisphosphonate. *J Bone Miner Res*, 1995; **10**: 956–62.
83. Coxon FP, Thompson K, Rogers MJ. Recent advances in understanding the mechanism of action of bisphosphonates. *Curr Opin Pharmacol*, 2006; **6**: 307–12.
84. Reid IR, Sharma S, Kalluru R, Eagleton C. Treatment of Paget's disease of bone with denosumab: case report and literature review. *Calcif Tissue Int*, 2016; **99**: 322–5.

4.11

Rickets and Osteomalacia (Acquired and Heritable Forms)

Michael P. Whyte

Introduction

Background

Mineralization of newly formed organic matrix within the skeleton is a complex and highly ordered process. Fundamentally, it requires both adequate extracellular concentrations of calcium (Ca^{2+}) and phosphorous as inorganic phosphate (Pi) as well as normal function of cartilage-forming chondrocytes and bone-forming osteoblasts. Compromise of either aspect can impair how well bones grow and harden. In fact, the causes and mechanisms for impaired skeletal mineralization are many and complex. The term 'rickets' describes the consequences of undermineralization of cartilage and bone matrix (primary spongiosa and osteoid, respectively) throughout a growing skeleton. Infants, children, and adolescents can be affected. The term 'osteomalacia' refers to what follows from impaired bone mineralization in adult life after the growth plates have fused and cartilage mineralization is no longer an issue. However, neither 'rickets' nor 'osteomalacia' denotes a specific aetiology. Both are generic labels for what follows from too little deposition of hydroxyapatite crystals into skeletal tissue. Nevertheless, there are several useful ways that rickets and osteomalacia can be classified. Nearly all individuals with rickets or osteomalacia have low circulating (extracellular) levels of Ca^{2+} and/or Pi. Often, inadequate vitamin D is involved. Sometimes there is instead impaired endogenous bioactivation of adequate vitamin D. Alternatively, kidney tubule dysfunction can lead to urinary Pi wasting and hypophosphataemia, often including impaired renal bioactivation of vitamin D. Skeletal cell-based explanations for undermineralization of cartilage and bone are rare and involve defective synthesis of constituents or conditioners of bone matrix. Hence, the number of disorders that cause rickets or osteomalacia is considerable; some are acquired, and others inherited (**Box 4.11.1**).

In contrast to osteomalacia in which predominantly bone remodelling (turnover) is impaired, rickets compromises all three components of skeletal formation:

- **Growth:** For long bones to lengthen and certain axial bones to enlarge, chondrocytes within their growth plates (i.e. physes, apophyses) must proliferate, hypertrophy, and then degenerate and the matrix they produced mineralize (i.e. endochondral bone formation).
- **Modelling:** Correct shaping of growing long bones requires deposition and removal of mineralized osseous tissue at their outer and inner cortical margins (periosteum and endosteum, respectively).
- **Remodelling:** On innumerable occasions over a lifetime, microscopic areas of mineralized cortical (compact) and trabecular (spongy) bone are resorbed and then reformed so that the skeleton's metabolic, structural, and repair roles can be fulfilled.

Hence, rickets often features short stature (physeal dysfunction), long bone bowing and metaphyseal widening (modelling defects), and frequent and poorly healing fractures (remodelling failure). These disturbances lead to diagnostic radiographic changes. In contrast to rickets, osteomalacia that begins in adult life is usually not deforming unless there are fractures. Because growth plates have fused and bone modelling has largely ceased, disturbed skeletal remodelling becomes paramount in osteomalacia. Osteomalacia is less apparent clinically and has few distinctive radiographic findings compared to rickets. In either disorder, however, defective bone turnover (remodelling) can manifest radiographically as generalized osteopaenia. Occasionally, if skeletal tissue accumulates, there can be hyperostosis (cortical bone widening) and osteosclerosis (trabecular bone thickening). The principal clinical, radiographic, and histological features of rickets or osteomalacia are largely shared among the numerous causes.

Alternatively for some years now, the term 'rickets' has been used by some histologists and clinicians strictly for the disturbance of

Box 4.11.1 Causes of rickets or osteomalacia

Vitamin D deficiency

- Deficient endogenous (cutaneous) synthesis
  - Inadequate sunshine
  - Other factors, e.g. ageing, pigmentation, sunscreens, clothing
- Dietary
  - Classic 'nutritional'
  - Fat-phobic

Malabsorption

- Gastric
  - Partial gastrectomy
- Intestinal
  - Small bowel disorders, e.g. coeliac disease (gluten-sensitive enteropathy)
- Hepatobiliary
  - Cirrhosis
  - Biliary fistula
  - Biliary atresia
- Pancreatic
  - Chronic pancreatic insufficiency

Disorders of Vitamin D bioactivation and action

- Hereditary
  - Vitamin D-hydroxylation deficient rickets, type 1A (OMIM #264700) (1α-hydroxylase deficiency)
  - Vitamin D-hydroxylation deficient rickets, type 1B (OMIM #600081) (25-hydroxylase deficiency)
  - Vitamin D-dependent rickets, type 2A (OMIM #277440) (vitamin D receptor defect)
  - Vitamin D-dependent rickets, type 2B (OMIM %600785) (intact vitamin D receptor)
- Acquired
  - Anticonvulsant therapy
  - Renal insufficiency
- Acidosis
- Distal renal tubular acidosis (classic, type I)
  - Primary (specific aetiology not determined)
    - Sporadic
    - Familial
  - Secondary
    - Galactosaemia
    - Hereditary fructose intolerance with nephrocalcinosis
    - Fabry disease
  - Hypergammaglobulinaemic states
  - Medullary sponge kidney
  - Postrenal transplantation
- Acquired
  - Ureterosigmoidostomy
  - Ileal conduit
  - Obstructive uropathies
  - Drug-induced
  - Acetazolamide
  - Ammonium chloride

Chronic renal failure

Phosphate depletion

- Dietary
  - Low phosphate intake
  - Antacid excess using aluminium hydroxide or other non-absorbable hydroxides
- Impaired renal tubular phosphate reabsorption ('phosphate diabetes')
  - Hereditary
    - X-linked hypophosphataemia (OMIM# 307800)
    - Autosomal dominant hypophosphatemic rickets
    - Autosomal recessive hypophosphatemic rickets, types I and II
    - Syndrome of lipoatrophic diabetes, vitamin D resistant rickets, persistent Müllerian ducts
    - Dent disease (OMIM #300009) (hypercalciuria, nephrolithiasis, X-linked)
  - Acquired
    - Oncogenic (tumour-induced)
    - Neurofibromatosis
    - McCune–Albright syndrome
    - Ifosfamide treatment
    - Cutaneous skeletal hypophosphatemic syndrome (epidermal nevus syndrome)

General renal tubular disorders (Fanconi's syndrome)

- Primary renal
  - Idiopathic
    - Sporadic
    - Familial
  - Associated with a systemic metabolic disease
    - Cystinosis
    - Glycogenosis
    - Lowe's syndrome
- Systemic disorder with associated renal disease
  - Hereditary
    - Inborn errors
    - Wilson's disease
    - Tyrosinaemia
  - Acquired
    - Multiple myeloma
    - Nephrotic syndrome
    - Transplanted kidney
  - Toxins
    - Cadmium
    - Lead
    - Outdated tetracycline

Primary mineralization defects

- Hereditary
  - Hypophosphatasia (OMIM #241500, #146300, #241510)
  - Osteogenesis imperfecta, type VI
  - Congenital sclerosing osteomalacia with cerebral calcification ('Raine syndrome')
- Acquired
  - Bisphosphonate (etidronate) toxicity
  - Skeletal fluorosis
  - Aluminium or strontium intoxication
  - Gallium intoxication

States of rapid bone formation

- Postoperative hypoparathyroidism with osteitis fibrosa cystica

Defective matrix synthesis

- Fibrogenesis imperfecta ossium

Miscellaneous

- Mg^{2+}-dependent
- Calciopaenic
- Steroid-sensitive
- Axial osteomalacia
- Osteopetrosis ('osteopetrorickets')

endochondral bone formation at physes, and 'osteomalacia' for the impaired remodelling typically observed using non-decalcified bone histology after tetracycline 'labelling'. In this context, rickets and osteomalacia both would be present in paediatric patients with impaired skeletal mineralization.

The fundamental pathogeneses of rickets or osteomalacia can be broadly but usefully considered: (1) primary Ca^{2+} deficiency (i.e. nutritional inadequacy for the mineral itself, or more commonly disrupted vitamin D homeostasis leading to inadequate absorption of dietary Ca^{2+}) and called hypocalcaemic rickets or osteomalacia; (2) primary phosphorous deficiency (i.e. binding and malabsorption of Pi within the gastrointestinal tract, or increased renal Pi clearance) and called hypophosphataemic rickets or osteomalacia; and (3) aberrant hydroxyapatite crystal deposition due to skeletal cell dysfunction or abnormal mineral composition.

Medical therapy for rickets or osteomalacia will have greatest success if the aetiology can be identified and thereby reveal a cure (e.g. environmental vitamin D deficiency) [1–3]. Failing a cure, precise understanding of the pathogenesis becomes key for safe and effective treatment. For some underlying disturbances of mineral homeostasis, directly circumventing the pathogenesis may be possible (e.g. providing the deficient form of vitamin D in certain inborn errors of vitamin D bioactivation, or alkaline phosphatase (ALP) in hypophosphatasia). Except for dosage, pharmacological approaches for the specific types of rickets or osteomalacia will generally be alike regardless of patient age. However, additional distinctive complications often trouble paediatric patients with soft bones, and the goals for treatment and necessary follow-up differ from adults. Management of rickets is especially complex and likely to require multidisciplinary skills including orthopaedic care and dentistry. Fortunately, all forms of rickets or osteomalacia now have some type of medical treatment and all patients can benefit.

Vitamin D and Mineral Metabolism

Vitamin D

Much is now understood concerning the biosynthesis, bioactivation, and peripheral actions of vitamin D [2, 3] and regulation of mineral homeostasis [4]. Sufficient antirachitic activity in people can come from cutaneous synthesis in sunshine of vitamin D_3 [5] derived from 7-dehydrocholesterol, but this is often not achieved so fortification of select foods with vitamin D_2 or vitamin D_3 (e.g. 400 IU per quart of milk or infant formula in the United States) acts as a 'safety net' in several countries. Ergocalciferol (vitamin D_2) is the product of UV irradiation of ergosterol extracted from animal or plant substances, and is used in supplements or as a pharmaceutical [2, 3]. For other individuals 290–310 nm ultraviolet (UV) light lamps are used for cutaneous synthesis of cholecalciferol (vitamin D_3). Age, skin pigmentation, and clothing as well as the angle, intensity, and duration of sunshine all condition how much vitamin D_3 is synthesized [5].

Vitamin D should be regarded as a fat-soluble steroid hormone rather than as a nutrient (i.e. 'vitamin') because cholecalciferol or ergocalciferol undergo two bioactivation (hydroxylation) steps to become the fully potent 1,25-dihydroxyvitamin D that circulates to target organs where it binds to the vitamin D receptor (VDR) [6–8]. Vitamin D_2 and D_3 are, therefore, both prohormones that are transported in the blood by their high-affinity binding protein to muscle or fat for storage, or to the liver and subsequently to the kidney for bioactivation or deactivation through targeted hydroxylation steps [2, 3]. The notion that this hormone is a vitamin came from cures of vitamin D-deficiency rickets by giving fish oils (e.g. cod liver oil) rich in this 'vital' substance. This bioactivation occurs first, with little regulation, when cholecalciferol or ergocalciferol are hydroxylated in hepatocyte mitochondria by the enzyme P450c25 to form 25-hydroxyvitamin D, called calcidiol. Then, with precise control reflecting circulating Ca^{2+}, Pi, and parathyroid hormone (PTH) levels, 25-hydroxyvitamin D is further hydroxylated by kidney proximal convoluted tubule cell mitochondrial P450c1α, more commonly called 25-hydroxyvitamin D,1α–hydroxylase (or 1α–hydroxylase) to 1,25-dihydroxyvitamin D, called calcitriol. Low circulating levels of Ca^{2+} and Pi and elevated circulating levels of PTH all activate renal 1α-hydroxylase in order to increase 1,25-dihydroxyvitamin D biosynthesis [6–8]. Other cells can contain 1α-hydroxylase: for example, keratinocytes, macrophages often in granulomas, some tumour cells, and placental decidual cells [7, 8]. However, under physiological conditions other than pregnancy, extrarenal production of 1,25-dihydroxyvitamin D does not significantly contribute to its circulating levels. Hydroxylation at carbon 24 to produce 24,25-dihydroxyvitamin D or 1,24,25-trihydroxyvitamin D (calcitroic acid) occurs in a wide range of healthy tissues and is important for the deactivation and elimination of vitamin D. All of these hydroxylases are mitochondrial mixed-function oxidases containing cytochrome P450 with ferredoxin and haem-binding domains. Vitamin D_2 and D_3 seem equally susceptible to these hydroxylation steps, and their bioactivated forms are essentially equipotent in conditioning mineral homeostasis [2, 3], although some evidence suggests vitamin D_3 is more effective than vitamin D_2 in humans.

In tissues targeted by vitamin D, genomic and non-genomic actions of 1,25-dihydroxyvitamin D can occur [9]. There, 1,25-dihydroxyvitamin D couples to the VDR encoded by its gene that is a member of the nuclear hormone receptor gene superfamily. The VDR has both 1,25-dihydroxyvitamin D-binding and DNA-binding domains. After the complex combines with a retinoid X receptor (RXR) heterodimeric partner, transcription of genes in bone, kidney, and enterocytes is activated to assure adequate extracellular concentrations of minerals [9]. 1,25-dihydroxyvitamin D is considered the active metabolite of vitamin D by its rapidity of action and potency to augment gut absorption of Ca^{2+}. Urinary Ca^{2+} reclamation by the kidneys and bone resorption to provide Ca^{2+} are also increased. However, 1,25-dihydroxyvitamin D suppresses PTH synthesis within the feedback loop [2, 3]. The nomenclature of vitamin D and its activated forms emphasizes the hormonal nature of these secosterols (Table 4.11.1).

Minerals

Extracellular levels of both Ca^{2+} and Pi are maintained principally by three organs: intestine (mainly by absorption), bone (by rapid fluxes as well as slower bone resorption and formation), and kidney (by ultrafiltration and then modulation of tubular reabsorption) [4]. Three hormones interact significantly to achieve mineral homeostasis: (1) PTH activates osteoblasts directly and osteoclasts indirectly, increases renal tubular reabsorption of Ca^{2+} while causing phosphaturia, and promotes 1,25-dihydroxyvitamin D production

Table 4.11.1 Nomenclature of vitamin D (calciferol) and its metabolites

Chemical	Vitamin	Abbreviation
Cholecalciferol	Vitamin D_3	D_3
Ergocalciferol	Vitamin D_2	D_2
25-Hydroxycholecalciferol	25-hydroxyvitamin D_3	25(OH) D_3
25-Hydroxyergocalciferol	25-hydroxyvitamin D_2	25(OH) D_2
1α-Hydroxyergocalciferol	1α-hydroxyvitamin D_2	1α(OH) D_2
1,25-Dihydroxycholecalciferol	1,25-dihydroxyvitamin D_3	$1,25(OH)_2D_3$
1,25-Dihydroxyergocalciferol	1,25-dihydroxyvitamin D_2	$1,25(OH)_2D_2$

Throughout the text, unless specified, the abbreviations D, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D refer to either the D_2 or D_3 compound, or a mixture of both.

in proximal renal tubule cells; (2) 1,25-dihydroxyvitamin D promotes Ca^{2+} (and Pi) absorption from the gut by increasing synthesis of the Ca^{2+} transport protein calbindin, probably activates osteoclasts, and regulates the biosynthesis of PTH and the phosphatonin fibroblast growth factor 23 (FGF23) [10]; (3) FGF23, produced by cells of the osteoblast lineage, circulates and then diminishes renal tubular reabsorption of Pi while inhibiting the biosynthesis of 1,25-dihydroxyvitamin D [7]. Thus, changes in circulating (i.e. extracellular) Ca^{2+} levels attributable to PTH come from its direct actions on the skeleton and kidneys and indirect action on the gut mediated instead by 1,25-dihydroxyvitamin D production [2, 3]. Extracellular Pi levels are regulated primarily by the kidney [11] where PTH and FGF23 cause phosphaturia, but to what degree other factors control Pi homeostasis is less well understood. Now, they include action by potentially a variety of phosphatonins [10], discovered from studies of a number of acquired or heritable forms of hypophosphataemic rickets or osteomalacia. Nevertheless, the precise interplay among these three hormones creates a means for controlling mineral homeostasis both acutely and long term, but with circulating levels of Ca^{2+} being more tightly controlled than Pi concentrations [4].

Box 4.11.2 Vitamin D deficiency: age-dependent signs and symptoms

- Metabolic
  - Hypocalcaemia
  - (See Box 4.11.3)
- Muscle
  - Proximal myopathy
  - Waddling gait
  - Asthenia
  - Pot belly with lumbar lordosis
- Dental
  - Delayed eruption
  - Enamel defects
  - Caries
- Skeletal
  - Frontal bossing
  - Widened cranial sutures
  - Craniotabes (skull asymmetry)
  - Harrison's groove
  - Rachitic rosary
  - Sternal indention or protrusion
  - Rib deformity → respiratory compromise
  - Pneumonia
  - Limb deformity
  - Flared wrists and ankles
  - 'String-of-pearls' deformity in hands
  - Bone tenderness
  - Fracture
  - Short stature
  - Kyphosis
  - Low back pain
  - Dystocia
  - Hypotonia
  - Lax ligaments
  - Listlessness

Diagnosis

Medical History

Disturbances in vitamin D and mineral homeostasis that cause rickets or osteomalacia can lead to a considerable number and variety of signs, symptoms, and complications—some of which depend importantly on the patient's age. These can be viewed as metabolic or skeletal in origin (**Box 4.11.2**), and become more likely as circulating Ca^{2+} or Pi levels become progressively lower (**Box 4.11.3**). Reduced levels or ineffective actions of vitamin D are particularly harmful for growing infants and young children (**Boxes 4.11.2** and **4.11.3**). Osteomalacia in adults is less apparent than rickets in children. Both disorders, however, can cause diffuse skeletal achiness and tenderness with focal worse pain due to fractures or pseudofractures. Some symptoms can instead be vague. The importance of a detailed medical history from, and then complete physical examination of, the patient to capture what is needed for diagnosing and treating all metabolic bone diseases was reviewed in 2018 [12], and cannot be overemphasized for rickets and osteomalacia.

Physical Examination

Rickets compromises especially the most rapidly growing bones [13, 14]. Thus, the nature of the skeletal changes depends on the patient's age when the metabolic perturbations began. Congenital forms of rickets may go unnoticed at birth because Ca^{2+} and Pi levels in the fetal circulation are controlled by placental transport from the mother. Such patients with healthy mothers usually develop physical features of rickets during the first 2 years of life. During infancy, the changes involve especially the cranium, ribs, and wrists causing widened cranial sutures, frontal bossing, posterior flattening of the skull (craniotabes), bulging of the costochondral junctions (rachitic rosary), indentation of the ribs at the insertion of the diaphragm

Box 4.11.3 Signs and symptoms of acute or chronic hypocalcaemia

- Nervous system
  - Increased irritability from latent or overt tetany
  - Seizures
  - Mental status change, retardation
  - Basal ganglia calcification
- Cardiovascular
  - Prolonged ST interval with arrhythmia
  - Cardiomyopathy with congestive heart failure
  - Hypotension
- Other
  - Papilloedema
  - Lenticular cataracts
  - Intestinal malabsorption
  - Dysplastic teeth
  - Rickets/osteomalacia
  - Integument changes
  - Joint contractures
  - Vertebral ligament calcification

Updated after reproduction with permission from Whyte MP. Hypocalcemia. In: *Metabolic Bone and Stone Disease*, 3rd edn. Nordin BEC, Need AG, Morris HA, eds. Edinburgh: Churchill Livingstone, 1993: 147–62.

(Harrison's groove), and distally widened forearm bones. The rib cage may become so deformed ('bell-shaped') that this predisposes to recurrent pneumonias and respiratory failure. Eruption of the dentition is delayed and defective mineralization ('shell teeth') can include enamel hypoplasia [13, 14]. After infancy, with standing and continued rapid linear growth, skeletal deformities can become severe in the lower limbs. Bowlegs (genu varum), or less often knock-knees (genu valgum), of variable severity develop together with widening of the ends of long bones due to physeal and metaphyseal expansion. Occasionally, the lower limbs bend in the same direction ('windswept' legs) [13]. However, extremity deformity may not occur if the rachitic child is too ill to bear weight or is not growing. When the bones first soften later in childhood or during the adolescent growth spurt, knock-knee deformity becomes particularly common. If unsuccessfully treated, rickets may compromise adult height, predispose to fractures, and leave permanent skeletal deformities that lead to osteoarthritis. Focal bone pain and tenderness can result from fracture. In osteomalacia, pressure on the sternum, percussion over the vertebrae, and squeezing of the ribs and long bones may elicit tenderness.

Skull shape and size can be distorted. Premature fusion of cranial sutures causing craniosynostosis is common in early-onset rickets. In X-linked hypophosphataemia, dolichocephaly often occurs due to closure of the sagittal suture, but is usually only a cosmetic finding unless additional sutures became prematurely fused. In severe instances of hypophosphatasia from ALP deficiency, 'functional' craniosynostosis from profound hypomineralization of the calvarium, or later due to true bony fusion of all cranial sutures, can lead to a scaphocephalic skull sometimes complicated by raised intracranial pressure [15].

In oncogenic (tumour-induced) rickets or osteomalacia, the causal neoplasm is usually benign and may be visible if not palpable, although some are smaller than pea-size and located essentially anywhere. Some have occurred intravaginally or in the nasopharynx, and not infrequently originate within the skeleton. Because extirpation of these tumours is curative, thorough physical examination to find them is essential. If the neoplasm remains elusive, including on radiological studies, the patients should periodically undergo physical examination for subcutaneous bumps because the tumours grow slowly and may appear.

Maternal dystocia due to a narrowed birth canal resulting from childhood vitamin D deficiency was a major cause of obstetrical mortality at the turn of the past century [14]. This deformity should be considered for women of childbearing age with a history of rickets.

Total alopecia represents an important finding concerning management and prognosis in patients with hereditary resistance to 1,25-dihydroxyvitamin D [6, 7].

Hypocalcaemia, if present, can enhance neuromuscular excitability [4]. Infants with compromised vitamin D homeostasis and hypocalcaemia are often 'floppy', hypotonic, listless, and irritable [13]. Symptoms and signs of latent or overt tetany may be reported during the medical history or elicited during the physical examination, respectively (**Box 4.11.3**). Such findings are particularly likely with severe and/or rapid reductions in circulating Ca^{2+} levels. There can be paresthesia of the lips and fingertips and spontaneous muscle contractions in the face, limbs, or elsewhere. Carpopedal spasm features thumb adduction, metacarpophalangeal joint flexion, and interphalangeal joint extension. Latent tetany can be unmasked if there is a positive Chvostek's sign or Trousseau's sign, yet both can be negative despite severe hypocalcaemia. Profound hypocalcaemia can also cause mental status changes, epileptic seizures, lethal stridor from laryngeal muscle spasm, and cardiomyopathy (**Box 4.11.3**) [4].

Additional problems from deranged vitamin D homeostasis and hypocalcaemia include a 'metabolic myopathy' with reduced muscle tone and strength as well as a waddling gait, but no (or nonspecific) changes on electromyography [13]. Weakness is a prominent feature of vitamin D deficiency and tumour-induced rickets or osteomalacia. Proximal muscle weakness is suspected when there is difficulty negotiating stairs, combing hair, or standing from sitting. Gower's sign detects muscle weakness if patients must push up with their hands on their thighs to stand. There can also be lax ligaments and *pectus excavatum* from diaphragmatic and intercostal muscle traction. Validated assessments of muscle strength by functional testing or dynamometry before treatment commences may be useful for follow-up if therapy will not be curative.

Radiological Studies

For rickets of any type, an anteroposterior radiograph of a knee and a posteroanterior radiograph of a wrist (perhaps non-dominant side) before treatment begins best document the characteristic physeal and metaphyseal abnormalities and will be an important basis to judge the patient's response to therapy [16]. Early on, rickets broadens growth plates uniformly (**Figures 4.11.1a** and **b**) and metaphyses are splayed and concave with their epiphysis appearing as though held within a ragged cup (**Figure 4.11.1b**). However, with chronic disease and long bone deformity mechanical forces acting on the lower limbs can cause asymmetrical widening of growth plates especially at the knees (**Figure 4.11.1c**). For a few years after the major appendicular growth plates fuse, rachitic changes may still be discernible in the apophyses of the ischium and ilium [16]. Long cassette radiographs of the lower limbs, taken while

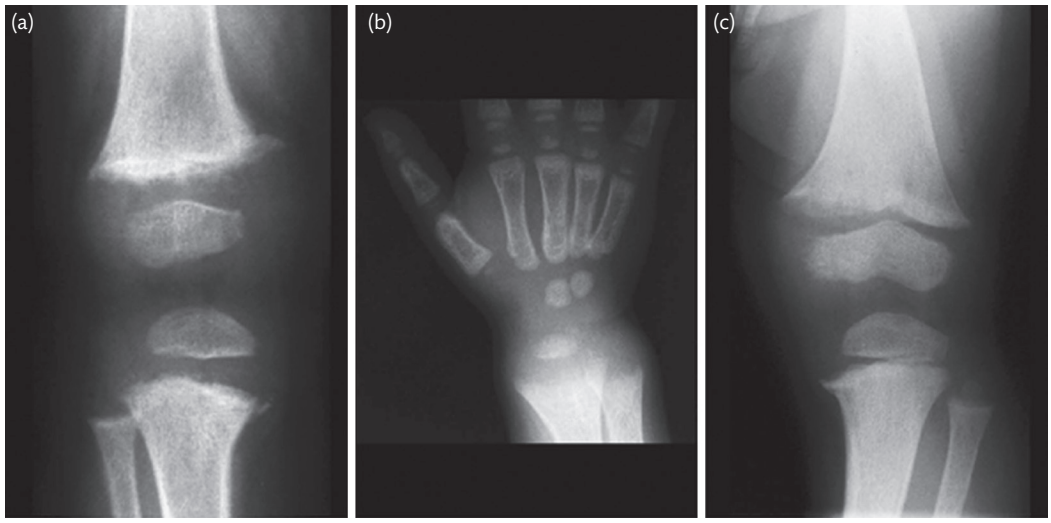


Figure 4.11.1 Rickets: (a) Before treatment, uniform physeal widening and metaphyseal irregularity and flaring are apparent in this knee of a 2-year-old child with vitamin D deficiency, poor dietary Ca^{2+} intake, and seizures treated with phenobarbital. (b) Physes are widened and metaphyses are irregular and flared in this wrist of a 2-year-old girl with untreated X-linked hypophosphataemia. A 'ball-in-cup' deformity is developing at her distal radius. (c) Physeal widening is less apparent in this knee of this 2-year-old girl beginning treatment for X-linked hypophosphataemia. The asymmetrical physeal widening and 'beaking' of the medial tibial metaphysis are due to the bowing deformity of the lower limbs.

the patient stands with reproducible positioning (ankles or knees together), help to explain, quantify, and follow bowing or knock-knee deformity. Now, low-dose X-ray imaging is available for this purpose.

Radiographs can also provide clues to the aetiology or pathogenesis of rickets [12, 16]. Disturbances in vitamin D homeostasis cause secondary hyperparathyroidism often featuring osteopaenia and sometimes evidence of subperiosteal bone resorption. Conversely, in X-linked hypophosphataemia changes of hyperparathyroidism are less usual, but increased skeletal radiodensity is not uncommon. Remarkable osteosclerosis and hyperostosis are characteristic of autosomal recessive hypophosphataemic rickets, type 1 due to dentin matrix protein 1 (DMP1) deficiency. In rickets from ALP deficiency in hypophosphatasia, peculiar 'tongues' of radiolucency commonly project from physes into metaphyses at the knees and elsewhere and there can be patchy osteosclerosis (Figure 4.11.2) [15]. However, not all disorders that cause radiographic growth plate distortion and limb deformity are forms of rickets [16]. Epiphyseal and metaphyseal dysplasias and Blount's disease may suggest rickets, but they do not compromise vitamin D or mineral homeostasis or cause osteomalacia [4].

In osteomalacia, pseudofractures (Looser's zones or Milkman fractures) can occur anywhere except in the skull, and most often involve the pubic and ischial rami, ribs, scapulae, and the medial cortex of the proximal femora (Figure 4.11.3). Intervertebral discs may compress softened vertebral endplates causing biconcave ('cod fish') vertebrae [16].

Radiographic signs of secondary hyperparathyroidism are seen best as subperiosteal erosions involving the radial border of the middle phalanx of the index finger, distal ends of the clavicles, and symphysis pubis [16].

The rapidity of radiographic response to therapy may be of diagnostic significance for rickets. In primary vitamin D deficiency, improvement occurs just several weeks after a single large oral dose of vitamin D [13]. Other forms of rickets often take months to improve radiographically, especially those due to renal Pi wasting treated with conventional medical therapy involving multiple daily doses of oral phosphate salts together with a bioactive form of vitamin D such as 1,25-dihydroxyvitamin D_3 [11].

Bone scintigraphy is useful for uncovering abnormalities in the skeleton, but does not provide a diagnosis. This procedure is usually unnecessary in children with rickets because radiographs are particularly useful. However, when physical examination fails to disclose the cause of tumour-induced rickets or osteomalacia, conventional bone scanning may help to detect the neoplasm. In general for this procedure, enhanced radioisotope uptake occurs where there is increased blood flow, osteoidosis, or rapid bone remodelling. In rickets or osteomalacia, a 'superscan' can result, featuring no apparent radioisotope within the kidneys and in osteomalacia will disclose complications such as fractures and pseudofractures. Octreotide scanning or (PET)/computed tomography (CT) are especially useful for detecting and locating the causal tumour [17]. Rarely, selective venous sampling with FGF23 assay is needed.

Bone densitometry using dual-energy X-ray absorptiometry (DXA) can be imprecise and difficult to interpret in rickets or osteomalacia because of short stature, skeletal deformities or fractures, and osseous tissue accumulation that mineralizes with effective treatment and then undergoes removal when remodelling resumes.

Biochemical Investigation

Laboratory investigation to differentiate among the three aforementioned principal types of rickets or osteomalacia is straightforward. Quantitation of serum Ca^{2+}, Pi, PTH, Mg^{2+}, and ALP is essential. If needed, assay of circulating 1,25-dihydroxyvitamin D, FGF23, and electrolyte levels and 24-hour urinary Ca^{2+} and Pi excretion will help to distinguish among the types of primary hypophosphataemia featuring defects in renal tubular Pi reabsorption.

Hypocalcaemia is usually more severe in vitamin D-deficiency rickets versus osteomalacia, and sometimes results paradoxically in hyperphosphataemia rather than in hypophosphataemia [13].

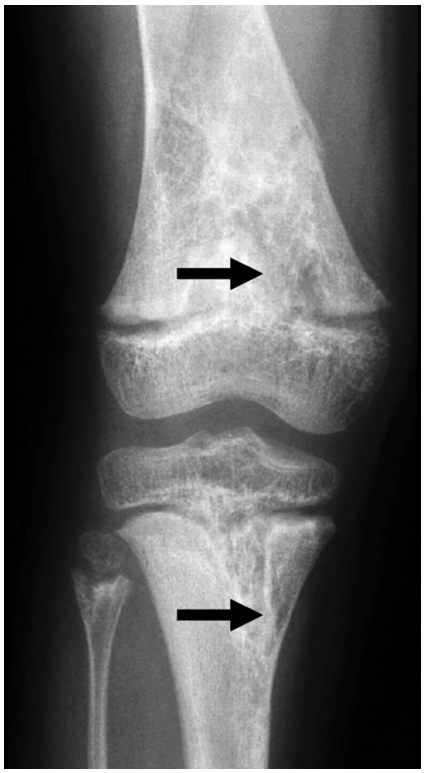


Figure 4.11.2 Hypophosphatasia: Characteristic tongues of radiolucency (arrows) project from the physes into the metaphyses of this 5-year-old girl who survived the infantile form of this heritable type of rickets.

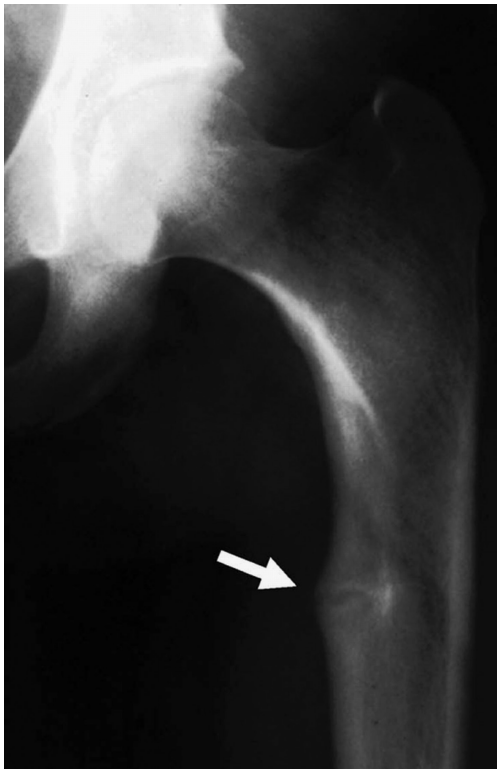


Figure 4.11.3 Pseudofracture: This 20-year-old woman with X-linked hypophosphataemia has a 'Looser zone' (arrow), characteristic of an osteomalacia, in her medial proximal femur.

Secondary hyperparathyroidism may cause a mild hyperchloraemic metabolic acidosis reflecting enhanced renal excretion of bicarbonate due to the elevated circulating PTH level. Significant metabolic acidosis, however, suggests Fanconi syndrome and therefore a different aetiology [11].

Although quantitation separately of circulating vitamin D_2 and D_3 might help to identify any predominance for an exogenous versus endogenous source, assays for these prohormones are not readily available [2, 3]. Fortunately, measuring serum 25-hydroxyvitamin D is a good surrogate for determining if there is vitamin D deficiency, sufficiency, or toxicity. Assay of serum 25-hydroxyvitamin D is offered by many commercial or hospital laboratories and typically detects 25-hydroxyvitamin D_2 and 25-hydroxyvitamin D_3 together. Assays for serum 1,25-dihydroxyvitamin D are also available commercially, but their utility for distinguishing among various types of rickets or osteomalacia is somewhat limited because the level can be low, normal, or high when there is vitamin D deficiency and depending upon the severity of secondary hyperparathyroidism and how much 25-hydroxyvitamin D substrate remains despite accelerated 1,25-dihydroxyvitamin D production [13, 14]. Quantitation of both 25-hydroxvitamin D and 1,25-dihydroxyvitamin D is, however, essential for differentiating among the various genetic disturbances in vitamin D bioactivation, including resistance to 1,25-dihydroxyvitamin D (Table 4.11.2).

Serum ALP activity (bone isoform) from chondrocytes and osteoblasts is elevated in most patients with rickets or osteomalacia. The striking exception is hypophosphatasia, which instead features hypophosphatas*aemia* [15] from deficiency of both the bone and liver isoforms of the 'tissue non-specific' isoenzyme of ALP. Blood and urine levels of other markers of skeletal turnover can be difficult to interpret in rickets or osteomalacia, and need not be measured routinely.

Histopathological Findings

Although the patient's medical history, physical findings, and results from routine radiographic and biochemical studies usually suffice to diagnose and to treat rickets or osteomalacia, if necessary histopathological assessment of bone tissue can provide definitive evidence of the expected defect in mineralization [1].

In clinical practice, the specimen used to detect impaired skeletal mineralization is routinely taken by trephine biopsy from the iliac crest. Beforehand, two 3-day courses of oxytetracycline or demeclocycline hydrochloride (20 mg/kg body weight per day in divided doses) are given orally (separated by a 2-week interval) for *in vivo* tetracycline labelling of mineralizing bone surfaces. The final dose is swallowed several days before the biopsy. The specimen must not be decalcified (see next). The histological picture is described as osteomalacia by histomorphometrists when excessive osteoid ('hyperosteoidosis') is present together with proof of defective bone matrix mineralization using time-spaced tetracycline labelling [18]. Histopathological confirmation of rickets is rarely necessary and would require risky biopsy of a growth plate.

Transiliac biopsy is performed at a standard anatomic site: 2 cm behind the anterosuperior iliac spine and just below the crest. The trephine has a 5.0 or 7.5 mm inner diameter for adults, and a 5 mm or narrower internal diameter for children depending on body size. Ideally, the specimen contains the outer and inner cortex connected by unbroken intervening trabeculae. It is fixed in 70% ethanol/30% water

Table 4.11.2 Biochemical parameters of mineral and skeletal homeostasis in rickets/osteomalacia, by aetiology

Aetiology	Biochemical properties Serum concentrations				
	Calcium	Phosphorous	PTH	Bone-specific ALP	24th urinary Ca^{2+} excretion
Hypocalcaemic, e.g. vitamin D deficiency	Low to low-normal	Low	Elevated	Elevated	Low
Hypophosphataemic, e.g. X-linked hypophosphataemia	Normal	Low	Normal to elevated	Elevated	Low to elevated
Tissue defects, e.g. hypophosphatasia	Normal or elevated	Normal or elevated	Normal to Low	Low	Normal to elevated

PTH, parathyroid hormone; ALP, alkaline phosphatase.

and then embedded in methylmethacrylate, sectioned undecalcified, and examined both unstained and with different stains. Histological assessment may include quantitation of various parameters compared to reference values (i.e. histomorphometry) (Figure 4.11.4).

Bone histology can be important in osteomalacia because radiographic studies are not characteristic as in rickets that is documented radiographically [16]. In osteomalacia or rickets, biopsy allows the increased width and extent of unmineralized osteoid covering mineralized bone surfaces to be quantitated (Figure 4.11.4a). Also, fluorescence microscopy will fail to show at osteoid/mineralized bone interfaces two discrete fluorescent tetracycline 'labels' produced if there is ongoing mineralization (Figure 4.11.4b). Instead, indistinct or absent fluorescence banding characterizes rickets or osteomalacia. Commonly used 'static' histomorphometric parameters [18] include trabecular bone volume, osteoid volume, osteoid surface, osteoid thickness, osteoblast surface, osteoclast surface, and osteoclast number whereas 'dynamic' parameters include double-labelled bone surface, mineral appositional rate, bone formation rate, and mineralization lag time.

Guidelines for Therapy and Follow-up

The last type of rickets or osteomalacia to acquire a medical therapy, hypophosphatasia, obtained one in 2015 [19]. Now, all forms of rickets or osteomalacia can be treated with at least some success. Ideally, the aetiology is corrected (e.g. vitamin D deficiency) (Box 4.11.1) and there is a cure. However, this is often not possible, and instead the pathophysiology must be understood and addressed for safe and effective treatment. Certain heritable forms of rickets or osteomalacia now have a pharmacological treatment aimed at a fundamental element of their pathogenesis; e.g. bone ALP replacement therapy for hypophosphatasia, monoclonal antibody treatment to decrease excessive circulating levels of the phosphatonin FGF23 in X-linked hypophosphatemia, and vitamin D metabolites to bypass errors in vitamin D bioactivation (see next). A correct diagnosis with appropriate follow-up is essential for a favourable clinical outcome while avoiding potential toxicities or complications. Pharmacological doses of cholecalciferol or ergocalciferol (or replacement doses of an active metabolite), sometimes with mineral supplementation, will be necessary when there is deranged bioactivation of vitamin D. Availability

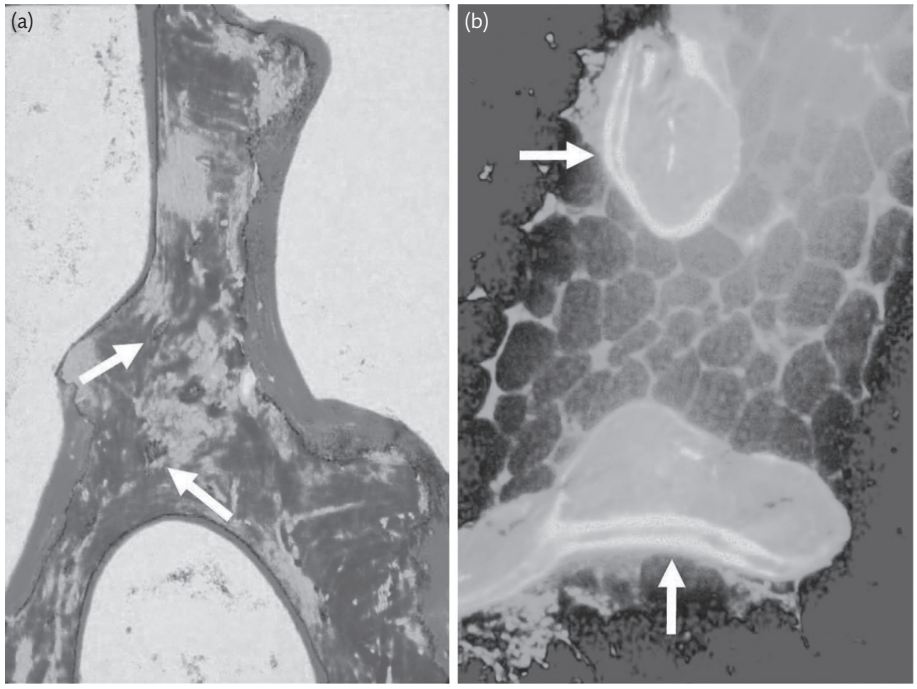


Figure 4.11.4 Pseudofracture: This 20-year-old woman with X-linked hypophosphataemia has a 'Looser zone' (arrow), characteristic of an osteomalacia, in her medial proximal femur.

Table 4.11.3 Pharmaceutical preparations of vitamin D and active metabolites

	Ergocalciferol	Dihydrotachysterol[a]	Calcifediol[b]	Calcitriol	Alfacalcidol
Abbreviation	**D_2**	**DHT**	**25(OH) D_3**	**1,25$(OH)_2D_3$**	**1α(OH)D**
Dosage form	Capsules: 1.25 mg Liquid: 200 μg/ml Intramuscular injection: 12.5 mg/ml in sesame oil (not in USA)	Tablets: 0.125, 0.200, and 0.400 mg Liquid: 0.250 mg/ml	Capsules: 20 and 50 μg	Capsules: 0.25 and 0.50 μg Liquid: 1.0 μg/ml Injection: 1.0 μg/ml	Capsules: 0.25 and 1.0 μg Liquid: 0.20 μg/ml Injection: 2.0 μg/ml in propylene glycol
Time to reach maximum biological effects	4–10 weeks	2–4 weeks	4–20 weeks	0.5–1 week	0.5–1 week
Persistence of biological effect after cessation	6–30 weeks	2–8 weeks	4–12 weeks	0.5–1 week	0.5–1 week

[a] This form of vitamin D is no longer generally available, but is listed for historical interest.
[b] Also called calcidiol.

of four vitamin D pharmaceuticals in the United States: vitamin D_2 and D_3, 25-hydroxyvitamin D_3, and 1,25-dihydroxyvitamin D_3 make it possible to replace, substitute for, or bypass various disturbances of vitamin D access, storage, or activation. They differ importantly in potency and biological half-life (Table 4.11.3). Alfacalcidol (1α-hydroxyvitamin D_3), rather than 1,25-dihyroxyvitamin D_3, is available in some countries recognizing it would undergo 25-hydroxylation in the liver. Dihydrotachysterol is no longer prescribed in the United States or the United Kingdom.

Rickets and osteomalacia have numerous and diverse explanations (Box 4.11.1). Successful treatment requires a correct diagnosis with an understanding of the aetiology and pathogenesis, including any deficiency or impaired bioactivation of vitamin D. Also, the nature of the deranged mineral metabolism must be appreciated. Wisely chosen, medical therapy for rickets or osteomalacia can greatly benefit the patient. For rickets, the three disturbances of skeletal development can often be improved, even corrected [1]. Given sufficient time before growth plate closure, reversal of deformity is a major goal of treatment. Relief from bone pain and fracture prevention also follow, and represent important goals of treating osteomalacia. Preventing short stature is also often a concern. Optimum therapy corrects or circumvents the underlying problem. However, a narrow range of safe and effective doses using appropriate pharmaceuticals and/or supplements should be anticipated for some patients. For hypophosphatasia and X-linked hypophosphatemia, weight-based dosing is prescribed precisely for their newer treatments [20–24]. Significant adverse consequences (failed treatment, hypercalcaemia, hyperphosphatemia, nephrocalcinosis, kidney stones, renal damage, and secondary or tertiary hyperparathyroidism) may result from an incorrect diagnosis or excessive and unmonitored therapy.

Physical examination provides important information. Supine length is measured in infants. Linear growth rates are monitored in infants, children, and adolescents. Height is determined best with a wall-mounted stadiometer, with several measurements averaged for distractible young children. Rapid changes in body size and weight during the pubertal growth spurt can increase dose requirements for chronic forms of rickets, and if not dealt with will worsen limb deformity and compromise final height. Furthermore, inordinate weight gain in girls can accelerate puberty and thereby close growth plates. When growth is complete, dosage requirements may diminish for various regimens.

Measurements of height and arm span as well as upper and lower body segment lengths will help to quantify skeletal distortion. With heritable forms of rickets, treatment will be necessary throughout growth, and months may pass before clinical improvement is achieved. It may be the individual who sees the patient perhaps every 6 months who best recognizes any changes in deformities. Accordingly, clinical photography, gait analysis, and videotaping can help assess for gradual changes. Radiographic alterations, such as widening of the growth plates and metaphyseal irregularity, are essential for diagnosing rickets and serve as important landmarks during follow-up to judge the efficacy of any medical treatment. Other radiographic findings are less helpful. Osteopaenia and coarsening of trabecular bone are observed in some patients [16]. Pseudofractures can heal with successful treatment.

The biochemical parameters used to diagnose rickets or osteomalacia are then followed to evaluate treatment. Typically, these are serum Ca^{2+}, Pi, 25-hydroxyvitamin D, ALP, and PTH. Depending on the aetiology and pathogenesis of this disorder, serum 1,25-dihydroxyvitamin D concentrations may or may not be helpful. Excretion of Ca^{2+} in 24-h urine collections (corrected for creatinine content) helps guide therapy as normal levels are aimed for, but also helps to monitor for excesses or toxicities. Eventual normalization of urinary levels of Ca^{2+} can indicate that treatment is now fully effective. If incomplete healing of rickets or osteomalacia is suspected, non-decalcified iliac crest histomorphometry will be definitive [1].

Treatment with vitamin D will be the appropriate way to overcome its deficiency, but pharmacologic doses can also help somewhat when there is defective bioactivation of vitamin D [7]. However, giving the correct active metabolite of vitamin D can circumvent such disturbances and therefore is useful for hereditary defects in vitamin D synthesis (see next) [7]. These pharmaceuticals are more potent than vitamin D_2 or D_3, have more rapid onset of action, and have shorter biological half-lives (Table 4.11.3). Thus, toxicity can appear sooner but will correct faster [7]. Nevertheless, they cannot replete deficient stores of vitamin D. Many different Ca^{2+} and Pi supplements are available [18]. The dose and type of Ca^{2+} may reflect the primary disturbance. $CaCO_3$ given orally is least expensive, but Ca^{2+} citrate is generally better absorbed. Ca^{2+} gluconate is especially costly. For the disorders that feature renal Pi wasting, conventional therapy has involved oral administration of Pi salts and bioactive vitamin D. Tablets rather than liquid preparations for oral Pi supplementation are more convenient, taste better,

and seem less prone to cause diarrhoea. Those with high amounts of sodium are best avoided. Now, since 2018, the anti-FGF23 monoclonal antibody, burosumab, has been available multinationally to treat X-linked hypophosphatemia [23, 24], and is effective for further types of FGF23-mediated rickets or osteomalacia.

Follow-up clinic visits are required for all types of rickets and osteomalacia, but the number and interval should reflect the specific diagnosis and therapeutic response. Because rickets and osteomalacia reflect deficient skeletal mineral content, decreases in treatment dose(s) might be necessary when healing is complete and maintenance becomes the principal issue. Because hypocalciuria characterizes most, but not all, forms of rickets or osteomalacia, urinary Ca^{2+} levels rising into the normal range will help to indicate effective therapy. Satiation of 'hungry bones' indicates the skeleton is no longer such an avid sump for mineral deposition. Normalization of previously aberrant biochemical findings could herald hypercalciuria as the definitive marker calling for dose reduction. Unless there is renal failure or fixed elevation in circulating PTH levels (reclaiming Ca^{2+} from the glomerular filtrate), hypercalciuria would precede hypercalcaemia. Cessation or lower maintenance doses of vitamin D and mineral supplements may then be needed. Thus, 24-h urine collections while consuming the usual diet, not random urine specimens, assayed for Ca^{2+} and creatinine can become important for patient follow-up and safety.

Consultation from the orthopaedic surgeon can be often helpful in managing rickets. As medical treatment is underway, he/she may consider leg bracing, physeal stapling (epiphysiodesis), or osteotomy to improve physical function and protect joint integrity. Straight lower limbs when growth ceases, with alignment of their physes parallel to the ground, helps prevent osteoarthritis. Intramedullary rodding may be necessary to heal pseudofractures and prevent fracturing for some adults with osteomalacia [25].

Pharmacological regimens for specific disorders are provided in the following sections that represent the three principal types of rickets or osteomalacia (i.e. hypocalcaemic, hypophosphataemic, and those with other pathogeneses).

Acquired Rickets and Osteomalacia

Primary Vitamin D Deficiency

Pathogenesis

Vitamin D deficiency reflects inadequate biosynthesis with insufficient acquisition from the diet, sometimes coupled with losses into the gastrointestinal tract. Primary ('nutritional') rickets or osteomalacia is a human-made disorder due to social, economic, and/or cultural factors that prevent sufficient exposure to sunlight for photosynthesis of vitamin D_3 in the skin. In modernity, hypovitaminosis D may also reflect too little intake from dietary (nutritional) sources or supplements containing vitamin D. Secondary deficiency could be from intestinal malabsorption of vitamin D_3 or D_2. Increased vitamin D clearance can result from accelerated catabolism, mainly in the liver, or increased loss via the intestine or the kidneys.

Vitamin D_3 synthesis in the skin requires UV light with a maximal effective wave length between 290 and 310 nm, and is affected by the intensity of the radiation, surface area of the exposed skin, and intrinsic properties of the epidermis [5]. At extreme latitudes during winter, almost no UV light reaches the ground. In the north of the United States, Canada, and north-western Europe, practically no vitamin D_3 is produced by exposed skin between October and March. Clothing, smog, window glass, plastic, and sunscreens [5] effectively block UV radiation and prevent cutaneous synthesis of vitamin D_3. Furthermore, vitamin D_3 production is less in people with dark skin (because melanin absorbs UV radiation), and in older people despite 'thin' skin [5]. Nevertheless, it has been estimated that 20-minute exposure thrice times weekly of the skin of the head and arms to summer sunshine prevents vitamin D deficiency in older people.

Vitamin D content of various unfortified food substances is very low, with the exception of cod liver oil and fatty fish such as herring and mackerel. During unfortified food consumption, it is estimated that less than 20% of total circulating 25-hydroxyvitamin D is from dietary sources. In some countries, therefore, dietary vitamin D is increased by supplementation of certain food products. In the United States, milk is fortified with 400 IU per quart. Improved vitamin D intake could result from multivitamins, which usually contain 400 IU or 800 IU per capsule, or some Ca^{2+} salt preparations that also contain vitamin D. These supplements will increase the relative contribution of dietary vitamin D to the total body pool, and be beneficial when cutaneous production is limited. Now in the United States, substantially higher doses of vitamin D can be obtained in various shops without a prescription.

Diagnosis

Hypovitaminosis D is diagnosed by assaying the serum concentration of 25-hydroxyvitamin D, which is a reliable measure of vitamin D status in almost all clinical situations. However, the necessary level of circulating 25-hydroxyvitamin D and the terms used to describe various low-end serum concentrations are not recognized universally. In fact, many of the clinical signs and symptoms and biochemical perturbations reflecting bone and mineral metabolism from hypovitaminosis D are shared among the disorders of vitamin D action and Ca^{2+} deficiency (Tables 4.11.2 and 4.11.4), including the subsequent low to low-normal serum Ca^{2+} levels, secondary hyperparathyroidism, hypocalciuria, hypophosphataemia, and the

Table 4.11.4 Serum levels of vitamin D metabolites in disorders of vitamin D action, by aetiology

Aetiology	Serum levels	
	25-hydroxyvitamin D	1,25-dihydroxyvitamin D
Vitamin D deficiency	Low	Low to elevated
25-hydroxyvitamin D, 1α-hydroxylase deficiency	Normal to elevated	Very low
Resistance to 1,25-dihydroxyvitamin D	Normal to elevated	Markedly elevated

hyperphosphatasaemia that seems to reflect enhanced willingness of the skeleton to ossify. Therefore, these biochemical parameters can support but not establish the diagnosis of rickets or osteomalacia, help assess the severity, and direct vitamin D dosing depending on the severity of the vitamin D requirement and the response to treatment. Importantly, circulating levels of 1,25-dihydroxyvitamin D can vary from low to elevated (Table 4.11.4), and thus are unhelpful for establishing this particular diagnosis.

In 1998, investigation of patients on a general medicine ward in Boston, Massachusetts, USA revealed that secondary hyperparathyroidism was common when serum 25-hydroxyvitamin D levels were at or below 15 ng/ml [26]. This led to studies to define vitamin D 'insufficiency' and 'deficiency'. Population-based reference values for serum 25-hydroxyvitamin D levels have been, for a considerable time, debatable. They could reflect age, geography, season, dress, eating habits, confinement to bed or home, local regulations concerning food fortification, and customs of vitamin supplementation. An alternative, however, was to define health-based concentrations (i.e. circulating 25-hydroxyvitamin D levels below which adverse health outcomes could occur). This would represent an intervention threshold to prevent detrimental effects on the skeleton, including relatively mild hypovitaminosis D not sufficient to compromise matrix mineralization but instead disrupt mineral homeostasis, cause secondary hyperparathyroidism, increase bone turnover, and lead to bone loss [27]. Multiple studies aimed to define serum 25-hydroxyvitamin D concentrations below which serum PTH levels increase, or for which vitamin D supplementation significantly decreased serum PTH concentrations [27]. Both approaches yielded similar functional thresholds. Accordingly, vitamin D 'adequacy' is accepted by most clinicians as serum 25-hydroxyvitamin D at or above 75 nmol/L (30 ng/ml), whereas some argue that the threshold level is instead between 50 and 75 nmol/L (20 and 30 ng/ml). However, serum 25-hydroxyvitamin D below 50 nmol/L (20 ng/ml) is considered 'inadequacy'. Vitamin D inadequacy has been subdivided into vitamin D 'insufficiency' when levels are between 25 nmol/L (10 ng/ml) and the threshold value, and vitamin D 'deficiency' when levels are below 25 nmol/L (10 ng/ml). Vitamin D deficiency represents a state in which Ca^{2+} homeostasis begins to fail (i.e. decreased serum Ca^{2+} levels despite increased serum PTH concentrations), with the high risk of developing rickets or osteomalacia. The same serum 25-hydroxyvitamin D threshold of ~75 nmol/L (30 ng/ml) was observed for additional physiological variables, such as intestinal Ca^{2+} absorption, changes in bone mineral density, and lower extremity physical performance [28]. Reduction of falls and fractures was positively correlated to the serum 25-hydroxyvitamin D level (up to a certain concentration) [29].

Prevalence of Vitamin D Insufficiency and Deficiency

Although the biosynthetic and bioactivation pathways for vitamin D are known, primary deficiency still seems common worldwide [14, 30]. In the United Kingdom, hypovitaminosis D resurged in the 1970s within the immigrant Asian community [13]. Those most vulnerable cannot move freely, and are at the beginning or end of life. In multiple studies worldwide, vitamin D deficiency was detected in 35 to 65% of older people; more so in institutionalized individuals [31]. In people hospitalized for an osteoporotic fracture, deficiency was recorded in 20–68%; only 1–3% had levels above 75 nmol/L (30 ng/ml). However, any age can be affected, especially when there are physical or mental handicaps. Additionally, the 'safety net' created in some countries by fortifying certain foods with vitamin D may not be available [13, 30]. Among adults, institutionalized or housebound individuals, the poor, older people, food faddists, and some religious groups (because of diet and dress) are at enhanced risk. Infants who are breastfed beyond 6 months of age or drink non-fortified milk or formula are also susceptible [5, 13]. In some populations, low dietary Ca^{2+} intake can be an important or exacerbating factor ('calciopaenic rickets') [30].

In postmenopausal women treated for osteoporosis in various regions of the world, vitamin D 'inadequacy'; that is, serum 25-hydroxyvitamin levels <75 nmol/L (30 ng/ml), was found in 52% of those in North America, 58% in Europe, 53% in Latin America, 71% in Asia, and 82% in the Middle East [32]. Pronounced differences in prevalences were observed among countries, ranging from 30% in Sweden in the summer to ~ 90% in Japan and South Korea in the winter and summer. Vitamin D 'deficiency' was, however, much less common; ~1% in North America, and ~6% and 8% in Latin America and the Middle East, respectively. Values below 25 nmol/L (10 ng/ml) were reported in ~40% of those tested in Sri Lanka and Beijing, and 18% in Hong Kong [33]. But, serum levels below 50 nmol/L (20 ng/ml) were observed in 78% of healthy hospital staff in India, and 90% of young women in Beijing and Hong Kong.

Of special concern is the high prevalence, in some geographic regions, of vitamin D deficiency in pregnant women, their children, and adolescent girls. Maternal serum 25-hydroxyvitamin D levels correlate positively with 25-hydroxyvitamin D levels in cord blood.

Fortunately, the prevalence of osteomalacia from hypovitaminosis D is lower than for the vitamin itself, but depends on the criteria for diagnosis (i.e. clinical, biochemical, bone histology, or bone histomorphometry). In a review of multiple publications describing histomorphometry of the femoral head or iliac crest in approximately 1400 patients with hip fracture, osteomalacia ranged from none to over 30% of patients. Perhaps this reflects different populations and magnitudes and durations of vitamin D deficiency, but also the histological criteria used to define osteomalacia.

Treatment

Although patient or parent education and correction of causal socioeconomic factors might seem best for preventing and treating primary vitamin D deficiency, this is often difficult to achieve. Fortunately, pharmacological or supplementation therapy should be inexpensive, effective, and work rapidly. Vitamin D deficiency should be treated using vitamin D_2 or D_3. Although 25-hydroxyvitamin D_3, 1α-hydroxyvitamin D_3, and 1,25-dihydroxyvitamin D_3 are more potent and act more rapidly than cholecalciferol or ergocalciferol, these active metabolites do not correct the depleted stores of vitamin D and their therapeutic windows are narrow and may not be sustained.

Adequate vitamin D intake has been recommended by the National Osteoporosis Foundation of the USA to be 800–1000 IU/day in adults age 50 years or older [34]. The recommendation for infants was increased to 400 IU/day of vitamin D_3. However, these goals are often unachievable unless certain foods are fortified with vitamin D.

Treatment should recognize those at greatest risk. Vitamin D supplementation for infants up to 1 year of age is required in many countries. However, supplementation may not be routine for older

people. Nursing home residents, institutionalized and hospitalized elderly, patients with hip fractures, and those with neurological disorders are among those in greatest jeopardy [31].

For vitamin D treatment to be fully effective, the recommended daily Ca^{2+} allowance should be achieved, including by Ca^{2+} salt supplementation if necessary. In fact, low dietary Ca^{2+} intake, especially common in some regions of Africa and North China, may itself cause rickets ('calciopaenic rickets') or exacerbate vitamin D-deficiency rickets.

The typical oral maintenance dose of vitamin D is 800 to 1000 IU daily. For patients with severe vitamin D deficiency causing symptomatic hypocalcaemia, it is helpful to administer a 'loading dose' of vitamin D_2 to replete body stores rapidly. Ca^{2+} intake must also be supplemented. Insufficient Ca^{2+} for a suddenly mineralizing skeleton could exacerbate the hypocalcaemia. A single oral dose of 5000 IU of vitamin D per kg body weight can be given. For a 70 kg adult, this is 350 000 IU of vitamin D. Although this quantity of vitamin D seems great, it illustrates the body's storage capacity for vitamin D [2, 3]. With symptomatic hypocalcaemia, Ca^{2+} can be given intravenously over 24 hours (as much as 20 mg of elemental Ca^{2+} per kg of body weight per day). Ca^{2+} infusions should be administered continuously, or slowly in portions, and always regulated by serum Ca^{2+} levels determined several times daily. Oral Ca^{2+} supplementation (1–2 g of elemental Ca^{2+} each day) can be initiated at this time. For patients who are not lactose intolerant and no longer hypocalcaemic, three to four glasses of milk each day will provide both Ca^{2+} and Pi to help remineralize the skeleton. In severe vitamin D deficiency, 4000–8000 IU could be given daily for the first 4–6 weeks. Alternatively, 50 000 IU of vitamin D could be administered two or three times a week for the first 2 weeks, followed by lower doses. Because vitamin D is stored in fat and released slowly, and the circulating half-life of 25-hydroxyvitamin D is 2–3 weeks, dosing can be once weekly, monthly, or every 3–6 months. An oral dose of 100 000 IU every 4 months for 5 years increased serum 25-hydroxyvitamin D to adequate levels. This approach can improve compliance by seeing the prescribed dose taken, both for independent elderly and for dependent institutionalized patients. However, experience with this high-dose regimen is relatively limited.

The response to vitamin D supplementation will depend on the severity of its deficiency and the consequent changes in mineral and skeletal homeostasis. Serum 25-hydroxyvitamin D should return to normal, but the timeframe will reflect the severity of the initial vitamin D deficiency. When the rickets or osteomalacia is severe, symptoms, signs, and laboratory parameters can change rapidly. Bone pain and muscle weakness will improve quickly, pseudofractures can heal, and serum Ca^{2+}, Pi, and PTH should return towards normal levels. In mild vitamin D deficiency or insufficiency, the response can be more subtle. Bone mineral density can increase, and the incidence of fractures should decrease and their rate of healing improve.

For infants and young children with vitamin D-deficiency rickets, liquid preparations of vitamin D_2 are available (Table 4.11.3) and can be dosed at 4000 IU (100 μg) orally each day for several months [13, 14]. If capsules can be chewed or swallowed directly, one 50 000 IU (1.25 mg) dose of vitamin D_2 orally each week for three or four doses is an inexpensive and straightforward regimen. It seems prudent that the physician observe that at least the first capsule is swallowed. Biochemical and radiographic improvement then typically occurs within just a few weeks [13, 14].

After healing their rickets, children who have inadequate sunlight exposure should receive 400 IU vitamin D_2 or D_3 per day, either by consuming fortified foods or using an over-the-counter multivitamin or vitamin D supplement [2, 3].

Failure to show biochemical and radiographic improvement with persistently low serum 25-hydroxyvitamin D levels could reflect failed compliance or malabsorption. Use of inexpensive vitamin D capsules may help to assure that the medication is taken. Alternatively, and if available, a single intramuscular injection of vitamin D in sesame oil will assure long-term access to antirachitic activity if patient compliance for oral treatment is poor, or if there is malabsorption (Table 4.11.3). Visits to a 'tanning salon' have also proved effective. If skeletal disease persists despite sustained correction of circulating 25-hydroxyvitamin D levels, calciopaenia, or one of the vitamin D-dependent or resistant syndromes (Box 4.11.1) must be considered (see next) [7, 9, 13, 30].

Prophylaxis against vitamin D-deficiency rickets could involve outdoor activity in sufficient sunshine but without sunburn, consumption of vitamin D-fortified foods, vitamin D supplements, or brief direct exposure to UV light wearing protective goggles [2, 3]. 'Stoss' therapy, used in Europe for children, consists of one depot intramuscular injection of 600 000 IU of vitamin D_2 during the autumn. However, this dose given once orally has caused hypercalcaemia and renal damage [35].

Secondary Vitamin D Deficiency

Vitamin D deficiency due to enhanced clearance is usually explained by systemic disorders (e.g. protein-losing nephropathy, intestinal malabsorption) or increased liver catabolism sometimes caused by certain drugs (e.g. barbiturates, other antiepileptics, etc.). Gastrointestinal malabsorption interferes with uptake of fat-soluble vitamin D from the gut, but sometimes disrupted enterohepatic recirculation of vitamin D metabolites contributes to the loss. Thus, intestinal malabsorption may decrease input and increase excretion of vitamin D.

Vitamin D deficiency from malabsorption may overcome normal amounts of sunlight exposure [5]. Gastrointestinal, pancreatic, or hepatobiliary disease may be at fault (Box 4.11.1) [2, 3], but the mechanism for the vitamin D deficiency and associated derangements in mineral metabolism is often complex. Malabsorption by the gut may also interfere with the uptake of dietary minerals.

Vitamin D is a fat-soluble secosterol, bile salts enable its absorption, and there is enterohepatic circulation of vitamin D and its derivatives [2, 3]. Hence, hepatobiliary or pancreatic disease or short bowel syndrome can cause deficiency of bile salts, steatorrhoea, and malabsorption leading to depletion of vitamin D stores. Furthermore, the small bowel mediates dietary Ca^{2+} uptake, and malabsorption of Ca^{2+} can exacerbate the consequences from vitamin D deficiency. With secondary hyperparathyroidism, conversion of 25-hydroxyvitamin D to both 1,25- and 24,25-dihydroxyvitamin D is enhanced, and 25-hydroxyvitamin D is depleted also by this mechanism. In fact, vitamin D deficiency and its clinical and biochemical consequences may be the first sign of occult malabsorption due, for example, to coeliac disease (non-tropical sprue). Nevertheless, in some conditions where osteomalacia might be anticipated (e.g. primary biliary cirrhosis), the

associated osteopathy is often osteoporosis. Iliac crest histology is especially useful for such patients.

Although the pathogenesis of secondary vitamin D-deficiency rickets or osteomalacia can be complicated, pharmacological therapy with appropriate follow-up should give gratifying results. As these patients reflect heterogeneous disturbances of wide-ranging severity, individualized therapy is key. Assay of serum 25-hydroxyvitamin D documents vitamin D status and is essential for monitoring progress [26]. Doses of vitamin D_2 or D_3 given orally should prove sufficient to correct any deficiency, and are relatively inexpensive. Repletion of vitamin D stores is a principal goal. Then, cholecalciferol or ergocalciferol can be converted to 25-hydroxyvitamin D, even if there is parenchymal liver disease (2, 3). Here too, a single oral 'loading' dose of about 125 µg (5000 IU) of vitamin D per kg of body weight can expedite treatment prior to maintenance dosing. Intravenous Ca^{2+} for symptomatic hypocalcaemia (as much as 20 mg of elemental Ca^{2+} per kg of body weight daily) over 24 hours by continuous infusion, or slowly in divided doses, is regulated by frequent measurements of serum Ca^{2+} and would be helpful for 'hungry bones'. Serum Mg^{2+} should be assayed for newly diagnosed hypocalcaemia, and treated if levels are low, as this may cause peripheral resistance to PTH or impair PTH biosynthesis [18].

After the loading dose(s) of vitamin D, patients with secondary vitamin D deficiency will require supplemental vitamin D unless the primary disorder can also be corrected. If not, it is impossible to predict the maintenance dose, and therefore clinical and biochemical follow-up is mandatory. Initially, outpatients should be evaluated every few weeks. Adjustments in dosing will be needed when the rickets or osteomalacia heals, or the gastrointestinal, hepatobiliary, or pancreatic disturbance evolves or responds to treatment, but considerable time may pass before the patient achieves a 'steady state' of vitamin D repletion.

For mild disease, a reasonable starting dose of vitamin D_2 or D_3 is 50 000 IU (1.25 mg) orally twice weekly. Assay of the circulating 25-hydroxyvitamin D level about 1 month later, and about every 4 months thereafter, will determine what dose of vitamin D proves useful. Serum 25-hydroxyvitamin D levels should be maintained somewhat above a threshold value of 75 nmol/L (30 ng/ml). If hypocalcaemia with secondary hyperparathyroidism is persisting, Ca^{2+} supplements can be added. Then, assay periodically of the Ca^{2+} and creatinine content of a 24-h urine collections, aiming for a normal calcium/creatinine ratio, can be used to optimize therapy. Attention to urinary Ca^{2+} levels will help to guard against vitamin D toxicity manifesting as hypercalciuria. This approach should show correction of hypocalciuria unless circulating levels of PTH (which reclaims urinary Ca^{2+}) are persistently elevated due to parathyroid gland hyperplasia and do not suppress with treatment. In this situation, assay of serum rather than urine Ca^{2+} levels becomes especially important.

Although oral vitamin D therapy at some dose is nearly always successful (unless there has been almost complete resection of the small intestine requiring total parenteral nutrition), intramuscular injection (if available) of depot vitamin D_2 in oil can be an alternative. Here, 12.5 mg of vitamin D_2 (500 000 IU) dissolved in 1 ml of sesame oil (Table 4.11.3) will provide sustained bioavailability of vitamin D [36]. An increment in the circulating 25-hydroxyvitamin D level may not appear for several weeks after the injection, but vitamin D_2 release will persist for months. Injections of 500 000 IU of vitamin D_2 every few months should provide effective and continuous supplementation for an adult, but biochemical monitoring is important.

Hypophosphataemia due to secondary hyperparathyroidism with renal Pi wasting contributes importantly to the pathogenesis of defective mineralization of skeletal matrix in most patients, but is not treated directly. Notably, some individuals with hypocalcaemia alone from hypoparathyroidism or pseudohypoparathyroidism develop rickets or osteomalacia despite elevated serum Pi levels.

Calciopaenic

Severe deficiency of dietary Ca^{2+} despite intact stores of vitamin D can also lead to defective skeletal mineralization [13, 30]. Calciopaenic rickets has been described in children fed a cereal-based diet [13] and in premature infants [14]. Several religious, ethnic, and other groups have vegetarian members who are at risk because they do not consume dairy products. Also, poor dietary Ca^{2+} intake can exacerbate vitamin D-deficiency rickets [30]. Altering the diet significantly, or using Ca^{2+} supplements, should readily reverse the disorder.

Drug-Induced

Anticonvulsant-Induced

Rickets and osteomalacia have been reported in institutionalized people receiving anticonvulsants, especially when multiple types [2, 3, 14]. Phenobarbital can alter hepatic vitamin D metabolism predisposing to vitamin D depletion [2, 3]. Also, primary deficiency of vitamin D affects many such individuals.

If their serum 25-hydroxyvitamin D level is low, the individual who takes phenobarbital or some other anticonvulsant can receive a 50 000 IU dose of vitamin D orally once weekly for several months, thereafter adjusted by following their serum 25-hydroxyvitamin D concentration.

Phosphate Binder-Induced

Osteomalacia, and rarely rickets, can result from excessive use of Pi-binders (e.g. magnesium and aluminium hydroxide antacids) [18]. Craniosynostosis together with rickets has been documented when such antacids were added to infant formula to treat colic [37]. Significant hypophosphataemia can occur. Assay of urinary phosphorous will reveal low levels, rather than seemingly normal or increased levels when defective skeletal mineralization is from renal phosphate wasting causing hypophosphatemia. Such patients may also hyperabsorb dietary Ca^{2+} and become hypercalciuric because their hypophosphataemia physiologically stimulates renal 25-hydroxyvitamin D, 1α-hydroxylase activity to increase biosynthesis of 1,25-dihydroxyvitamin D. Kidney stones can develop. Despite the increased dietary Ca^{2+} absorption, the hypophosphataemia alone impairs skeletal mineralization. Elimination of Pi-binder use will rapidly correct the disrupted mineral homeostasis and heal the skeletal defects. Pi supplementation or vitamin D therapy will not be necessary. It may take several months for serum ALP activity to correct.

Ifosfamide

This chemotherapeutic drug can cause transient or permanent kidney tubule damage leading to urinary Pi wasting and hypophosphataemic skeletal disease [38].

Intravenous Iron

Patients with iron deficiency anaemia given iron intravenously have developed rickets or osteomalacia. Circulating levels of biologically active FGF23 can rapidly increase and lead to renal phosphate wasting and suppressed 1,25-dihydroxyvitamin D synthesis causing secondary hyperparathyroidism [39].

Etidronate

This first-generation bisphosphonate used for Paget's bone disease and hypercalcaemia of malignancy can, with excessive or prolonged exposure, cause rickets or osteomalacia. Etidronate retains sufficient chemical similarity to the natural inhibitor of biomineralization, inorganic pyrophosphate, to inhibit hydroxyapatite crystal formation and growth [4].

Strontium-Induced

High skeletal levels of strontium have been associated with defective bone mineralization in haemodialysis patients [40]. Rickets seems more prevalent where soil contain high levels of strontium [41].

Toxin-induced

Rickets or osteomalacia can follow long-term exposure to several other inhibitors of skeletal mineralization.

Aluminium

Patients with uraemia exposed to aluminium-containing antacids, contaminated dialysate, or parenteral feedings containing aluminium have developed osteomalacia [18]. Treatment with deferoxamine has been helpful [18]. With use of newer agents for Pi binding and pure dialysate and feedings, this disorder is now rare.

Fluoride

Excessive fluoride from well water, industrial or 'huffing' inhalation, inordinate amounts of tea drinking, or sodium fluoride given for osteoporosis can cause rickets or osteomalacia [18, 42]. Defective bone mineralization will respond gradually to cessation of fluoride exposure, perhaps helped by oral Ca^{2+} supplementation [43].

Oncogenic

Oncogenic rickets or osteomalacia is a rare, sporadic, acquired disorder that is typically caused by a benign 'mixed mesenchymal' tumour in soft tissues (Figure 4.11.5) [11, 18]. However, a considerable variety of indolent neoplasms, non-ossifying fibroma, and (rarely) malignant bone tumours or other cancers can cause this condition [18]. These tumours secrete FGF23 and sometimes other phosphatonins that cause phosphaturia and inhibit renal 25-hydroxyvitamin D, 1α-hydroxylase activity (10). Low circulating 1,25-dihydroxyvitamin D levels can cause malabsorption of dietary Ca^{2+} and mild hypocalcaemia, secondary hyperparathyroidism, and hypocalciuria. Hypophosphataemia is, however, the principal biochemical abnormality, and lowers the blood $Ca^{2+} \times$ Pi product sufficiently to impair skeletal mineralization. Patients are often profoundly weak.

Definitive diagnosis of oncogenic rickets or osteomalacia is achieved by resection of the neoplasm and subsequent cure. Therefore, thorough diagnostic evaluation for this entity is especially important when there is sporadic acquired hypophosphataemic skeletal disease, particularly in an adult. If a soft tissue tumour is not apparent on physical examination, bone scintigraphy, whole body magnetic resonance imaging, octreotide scanning, or positron emission tomography CT using specific tracers with high sensitivity or high specificity have proven extremely useful [17, 18].

When removal of the causal neoplasm is not possible or unsuccessful, Pi supplementation together with 1,25-dihydroxyvitamin D_3 or 1α-hydroxyvitamin D_3 treatment can reverse patient weakness and heal the skeleton. In 2020, the anti-FGF23 monoclonal antibody, burosumab, gained approval for this disorder, and is promising for the similar condition cutaneous skeletal hypophosphatemia syndrome [44].

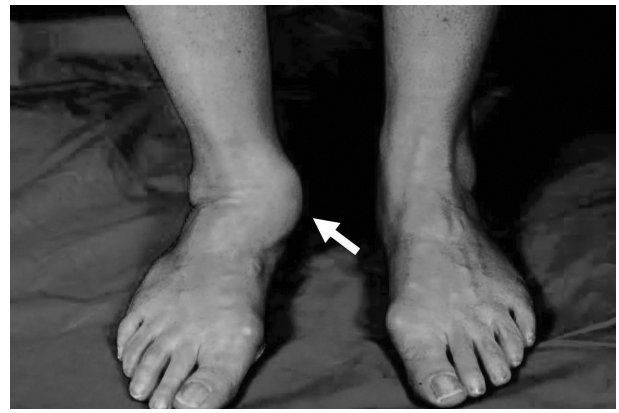


Figure 4.11.5 X-linked hypophosphataemia: Severe lower extremity bowing deforms an untreated mother and her daughters.

Metabolic Acidosis

Metabolic acidosis can cause rickets or osteomalacia (Box 4.11.1). The pathogenesis is not well understood but complex. Nevertheless, the skeletal disease responds well to vitamin D and alkali therapy. Ca^{2+} and potassium supplementation may be necessary at the onset of alkali therapy to prevent hypocalcaemia and hypokalaemia. Vitamin D (50 000 IU orally thrice weekly) can be used for adults, with careful follow-up until healing occurs. Alkali therapy should be continued after the mineralization defect is corrected. Urinary Ca^{2+} and creatinine levels must be monitored frequently because metabolic acidosis *per se* causes hypercalciuria.

Renal Failure

In uraemia, skeletal disease usually reflects the associated secondary or tertiary hyperparathyroidism leading to rapid bone remodelling (osteitis fibrosa cystica) [4]. However, some patients have instead manifested osteomalacia caused by aluminium or strontium exposure or excessive treatment that has suppressed calcitriol biosynthesis.

Aluminium Intoxication

Aluminium is toxic to osteoblasts and inhibits skeletal mineralization [4]. Contamination of dialysate with this metal has caused 'Newcastle bone disease'. Also, uraemic patients who used aluminium-containing antacids to bind dietary Pi deposited the metal into their skeletal tissue. Serum assays and bone histochemistry for aluminium support the diagnosis [4]. Therapy includes changing to be newer Pi-binders. Deferoxamine has been a useful chelating agent [18].

Hypophosphataemia From Phosphate Binders

Excessive use of Pi-binders in uraemic patients can cause hypophosphataemia leading to rickets or osteomalacia.

Parathyroid Insufficiency

Severe osteomalacia has occurred in chronic renal failure after excessive parathyroidectomy for secondary or tertiary hyperparathyroidism [4]. PTH is necessary for bone turnover in uraemia.

Epidermal Nevus Syndrome

Infants and children with epidermal nevus syndrome (cutaneous skeletal hypophosphataemia syndrome) can develop rickets due to renal Pi wasting [44]. 1,25-dihydroxyvitamin D_3 and Pi supplementation therapy is effective for the generalized impairment of skeletal mineralization, but not for the underlying focal pathogenetic bone disease. If circulating FGF23 levels are elevated, burosumab therapy may be useful [44].

Miscellaneous Disorders

Several rare sporadic conditions manifest with osteomalacia despite normal circulating concentrations of Ca^{2+}, Pi, PTH, and 25-hydroxyvitamin D [18]. There is no established therapy. A correct diagnosis is important in part because excessive doses of vitamin D or an active metabolite and Ca^{2+} supplementation could cause hypercalcaemia and hypercalciuria.

Fibrogenesis imperfecta ossium, reported in about a dozen patients, is an enigmatic acquired abnormality of the skeletal matrix. Growth hormone may be an effective treatment [45]. Axial osteomalacia is characterized radiographically by coarsening of trabecular bone in the axial skeleton, seemingly from a primary defect in osteoblasts, and may be heritable [46].

Heritable Rickets and Osteomalacia

Genetic disorders that cause rickets or osteomalacia are numerous (Box 4.11.1). Some feature renal Pi wasting, some reflect disturbances in the bioactivation or action of vitamin D, and some are inborn errors of metabolism due to enzyme deficiencies (throughout this section, heritable disorders are referred to by their McKusick symbol and OMIM number provided in Online Mendelian Inheritance In Man) [46].

Hypophosphataemic Bone Disease

Several heritable forms of rickets or osteomalacia reflect disturbances in vitamin D metabolism that lead to reduced levels of Ca^{2+} in the extracellular fluids [14]. These disorders can also diminish extracellular Pi concentrations partly from phosphaturia due to secondary hyperparathyroidism. Hypocalcaemia and hypophosphataemia together engender a decreased blood $Ca^{2+} \times$ Pi product that explains why there is impaired mineralization of newly synthesized osteoid [1].

The importance of Pi alone for skeletal mineralization is, however, best illustrated by the types of rickets or osteomalacia due to renal phosphate wasting without significant decreases in circulating Ca^{2+} levels [11]. 'Hypophosphataemic bone disease' is a generic term that emphasizes the skeletal consequences of this biochemical disturbance. Although pharmacological treatment for these disorders shares certain themes, optimal regimens may differ.

X-linked Hypophosphataemia

X-linked hypophosphataemia (XLH) is the most common heritable form of rickets or osteomalacia (OMIM #307800) [11]. Its prevalence in North America is approximately 1:20 000 live births. All races have affected individuals.

XLH was first identified in 1937 when vitamin D-deficiency rickets, a plague of Northern industrialized cities at the turn of the last century, became treatable with vitamin D [47]. Discovery of vitamin D in 1919, and then successful therapy and preventive measures for 'nutritional' rickets, represented a triumph of medical science [2, 3]. Nevertheless, some cases of refractory rickets were puzzling because they were not cured even by massive doses of vitamin D_2 [47]. Therefore, XLH was at first called 'vitamin D-resistant rickets' until 1958 when its X-linked dominant inheritance became appreciated [48]. Females and males are affected in a 2:1 ratio with no male-to-male transmission. Hypophosphataemia from renal Pi wasting became appreciated as a key pathogenetic factor in 1969 [49]. Inappropriately normal circulating levels of 1,25-dihydroxyvitamin D despite hypophosphataemia was recognized in 1982. In 1995, an international consortium identified the causal gene designated *PHEX* (phosphate-regulating gene with homology to endopeptidases on the X-chromosome) [50].

XLH causes short stature and bowing of the lower limbs beginning in affected toddlers when they bear weight to walk (Figure 4.11.6). Subsequently, they may seem clumsy, yet strong and are generally well in contrast to nearly all other forms of rickets. The skull is often dolichocephalic from early closure of the sagittal suture and Chiari

Figure 4.11.6 X-linked hypophosphataemia: Severe lower extremity bowing deforms an untreated mother and her daughters.

1 malformation from craniosynostosis can occur. The chest and upper extremities are not deformed. Fractures during childhood are uncommon. Occasionally, the skeletal disease presents later in childhood with knock-knees. Without treatment, height Z scores will be minus 2–3 standard deviations [51, 52].

Adults with XLH typically suffer five principal complications [51]. Arthralgias, primarily involving the lower limbs and especially the knees, are explained by osteoarthritis. The degree of residual lower extremity rachitic deformity predicts the likelihood of this knee joint deterioration [51]. Bone pain palpable in the thighs is often explained by femoral pseudofractures (**Figure 4.11.3**). Pseudofractures can occur elsewhere. Dental abscesses develop because brittle 'shell' teeth form early in life due to defective mineralization of dentine. Enthesopathy (calcification of tendons, ligaments, joint capsules, etc.) is common and spinal stenosis may be a result, but it is unclear how often symptoms develop. Sensorineural hearing loss seems to have increased prevalence. Obstetrical histories seem benign, but as for all types of rickets dystocia should be considered a potential problem [51]. The impact of XLH in older people is being studied. Life expectancy is probably not compromised.

Radiographs of children with XLH show physeal widening, that in the knees becomes especially pronounced medially when there is bowing deformity of the lower limbs (**Figure 4.11.1c**). Osteopaenia and evidence of secondary hyperparathyroidism are generally absent unless dietary Ca^{2+} intake is poor. In fact, the skeleton often appears radiographically dense, contrasting with other forms of rickets that characteristically increase circulating PTH levels. In adults with XLH, axial skeletal mass is typically normal, although sclerotic bones are not uncommon [53].

The biochemical hallmark of XLH is hypophosphataemia from selective renal Pi wasting [11]. Serum Ca^{2+} levels are low-normal, but usually not distinctly reduced [51, 52]. Hypophosphataemia must be documented recognizing important age-related effects on reference ranges for serum Pi. Healthy children have considerably higher serum Pi concentrations (and ALP activity) compared with healthy adults. Because circulating Pi levels may increase or decrease depending upon what is eaten, fasting blood specimens are best for assay [52]. In XLH, quantitation of renal Pi reclamation by calculating the transport maximum for phosphorous per glomerular filtration rate (TmP/GFR) shows that the hypophosphataemia reflects decreased renal tubular reabsorption ('phosphate diabetes') [51, 52]. Occasionally, trace glucosuria is detected, however, other parameters of proximal renal tubule function (e.g. serum potassium, bicarbonate, uric acid levels) are normal; that is, Fanconi's syndrome (see later) is absent. Serum 1,25-dihydroxyvitamin D levels in XLH are generally normal or low-normal despite the hypophosphataemia, which physiologically increases renal 25-hydroxyvitamin D, 1α-hydroxylase activity [11]. During treatment, unless Pi supplements are insufficiently matched by doses of 1,25-dihydroxyvitamin D_3 to prevent malabsorption of dietary Ca^{++} (see next), circulating PTH levels are usually normal [51, 52]. Without treatment, serum ALP is increased in children, but not always in adults. Serum FGF23 levels are elevated in nearly all untreated XLH patients, in keeping with this key factor in the disorder's pathogenesis [10].

Histopathological examination of the skeleton shows rickets or osteomalacia in untreated patients with XLH (**Figure 4.11.4a**) [51]. Elevated circulating PTH levels predict features of hyperparathyroidism, including abundant osteoclasts and peritrabecular fibrosis. Additionally, in appropriately stained non-decalcified sections [1], there are halos of hypomineralized bone surrounding osteocytes (**Figure 4.11.4a**) [51]. This peculiarity has been considered characteristic of XLH, reflecting an osteoblast defect that persists despite successful 1,25-dihydroxyvitamin D_3 and Pi therapy when these cells become osteocytes. Accumulation of osteopontin seems to be the explanation [54].

The aetiology of XLH is loss-of-function mutation of *PHEX*, but the pathogenesis is incompletely understood [10, 11]. Kidney transport of Pi is diminished across proximal tubule cells where normally PTH and FGF23 decrease urinary Pi reclamation [11]. Here, Pi movement across the brush border membrane is the rate-limiting step. In XLH, there is diminished Na^+-dependent Pi transport [11]. However, blunted response to the activators of 1,25-dihydroxyvitamin D biosynthesis is also present in kidney mitochondria [11] where Pi deprivation and supplementation physiologically accelerate and suppress, respectively, 1,25-dihydroxyvitamin D formation. Nevertheless, the precise intracellular disturbances that diminish renal Pi transport and alter vitamin D bioactivation are not known [10, 11]. Parabiosis and renal transplantation studies using the *Hyp* mouse implicated a phosphaturic factor(s) [55], now appreciated to be an excess of circulating FGF23 [10]. Malabsorption of dietary Ca^{2+} in XLH is poorly understood, but considered a feature of the vitamin D resistance.

In XLH, no gene dosage effect emerged from a study of prepubertal heterozygous girls and hemizygous boys (**Figure 4.11.6**) [56]. Nevertheless, the complications of pseudofracture and enthesopathy seem more severe in men compared to women [51]. Accordingly, gender (sex steroids and/or physical labour, etc.) may condition long-term outcomes [51].

XLH maps to chromosome Xp22.31–21.3 [46]. Many different mutations involving the coding sequence or splice sites of *PHEX* have been discovered in XLH worldwide [46]. They are expected to diminish PHEX protein function. An early preliminary study indicated that defects compromising the exons encoding the PHEX protein *per se* cause more severe XLH [57]. Indeed, a missense mutation within the *PHEX* 3'UTR has been found to cause a distinctly mild variant of XLH in American patients [58]. Nevertheless, the putative substrate for PHEX remains uncertain [10, 11]. PHEX could act at cell surfaces to inactivate a phosphaturic factor, or activate a suppressor of phosphatonins such as FGF23 [10, 11, 59].

Renal Pi wasting is the key pathogenetic abnormality that causes defective skeletal mineralization in XLH. TmP/GFR correlates positively with height Z score in paediatric patients [56], and decreases reflect the degree of bowing deformity in affected adults [51]. Accordingly, in conventional therapy, this disturbance is targeted together with decreases in 1,25-dihydroxyvitamin D biosynthesis.

Bioactivated forms of vitamin D have been used for decades to treat XLH because high doses of vitamin D (e.g. 100 000 IU daily) can somewhat improve, but will not heal, the rickets [60]. Large doses of vitamin D_2 are readily converted to 25-hydroxyvitamin D, however, the affinity of the VDR for 25-hydroxyvitamin D is two to three orders of magnitude lower than for 1,25-dihydroxyvitamin D [2, 3]. Furthermore, excessive vitamin D_2 given to patients with XLH sometimes caused prolonged hypercalcaemia, hypercalciuria, nephrocalcinosis, and renal failure [61] reflecting the long biological half-life of vitamin D (**Table 4.11.3**). Hypercalcaemia can persist for

months, requiring dietary Ca^{2+} restriction and glucocorticoid treatment. Additionally, high-dose vitamin D_2 therapy has required cessation months in advance of osteotomy to avoid hypercalciuria or hypercalcaemia if postoperative immobilization is prolonged.

When the pathogenetic renal Pi wasting of XLH became addressed, improved clinical, biochemical, and radiographic responses were noted [49, 52]. Transient augmentation of circulating Pi levels was achieved using frequent oral Pi dosing to supply, depending on body size, about 1–2 g of phosphorous (as Pi) each day. Combined use of 1,25-dihydroxyvitamin D_3 or 1α-hydroxyvitamin D_3 and Pi supplementation became the best regimen for XLH [52]. The active metabolite of vitamin D would augment both Ca^{2+} and Pi uptake from the gut, while better Ca^{2+} absorption would prevent secondary or tertiary hyperparathyroidism provoked by Pi lowering blood Ca^{2+} levels directly or binding Ca^{2+} in the intestine. Nevertheless, this regimen only transiently corrected hypophosphatemia and rarely would safely fully correct the radiographic rickets [23, 24]. This conventional therapy required a medical/orthopaedic approach for XLH best provided by experienced centres where 1,25-dihydroxyvitamin D_3 and Pi supplementation could improve the defects in skeletal growth, modelling, and remodelling in compliant patients [52]. However, since 2018, the human anti-FGF23 monoclonal antibody burosumab (Crysvita®), that neutralizes circulating FGF23, has been commercially available multinationally to better treat XLH in patients aged 6 months or older [23, 24].

There are two principal goals of therapy for XLH: (1) correction of limb deformity by the time growth plates fuse to prevent osteoarthritis; and (2) satisfactory height. The principal complications of active vitamin D and Pi treatment are: (1) secondary or tertiary hyperparathyroidism that compromise clinical outcomes and perhaps necessitate partial parathyroidectomy; and (2) nephrocalcinosis that may cause renal compromise.

This treatment may begin for toddlers to help promote their growth and to avoid lower extremity distortions, but oral dosing and monitoring is understandably difficult especially at first. Control, but not complete correction, of the rickets seemed a reasonable objective. Both 0.25 and 0.50 μg capsules of 1,25-dihydroxyvitamin D_3 are commercially available. The contents can be put into applesauce, etc., but a liquid preparation is also now marketed (Table 4.11.3). Approximately 40 ng (i.e. 0.040 μg) per kg of body weight of 1,25-dihydroxyvitamin D_3 daily (divided doses are ideal) may be achieved safely over 2 to 3 months by gradually increasing the dose and monitoring its biochemical effects. In the United Kingdom and Europe, a solution of 1α-hydroxyvitamin D is available. Pi supplementation, given three to four times daily, is introduced simultaneously and also gradually increased. Tablets of neutral sodium/ potassium phosphate (e.g. K-Phos Neutral; Beach Pharmaceuticals, Tampa, FL) are most convenient and generally well tolerated. Occasionally, Pi causes diarrhoea. In some ways, 1,25-dihydroxyvitamin D_3 and Pi produce opposite effects on Ca^{2+} homeostasis [52]. Accordingly, if either Pi or 1,25-dihydroxyvitamin D_3 is stopped, both should stop. Sudden decreases (or especially cessation) in Pi supplementation alone should be avoided, because 1,25-dihydroxyvitamin D effects can persist and urinary and then blood Ca^{2+} levels may rapidly rise. Accordingly, patients should be cautioned not to run out of these medications.

Careful biochemical surveillance is essential because 1,25-dihydroxyvitamin D_3 is especially potent in enhancing gastrointestinal Ca^{2+} absorption. Ca^{2+} and creatinine should be assayed in 24-hour urine collections (not random specimens) [52]. Initially, monitoring should occur monthly, but then every 3 months. Urinary Ca^{2+} to creatinine ratios of about 150–180 mg/g reflect adequate gut effects of 1,25-dihydroxyvitamin D_3 helping to suppress circulating PTH levels. If hypocalciuria is a persisting problem, increased dairy consumption or Ca^{2+} supplementation may be helpful. Unless PTH levels are elevated and non-suppressible (predicting hypercalcaemia before hypercalciuria), hypercalciuria will herald excessive 1,25-dihydroxyvitamin D_3 or 1α-hydroxyvitamin D_3 dosing. Fortunately, they both have short biological half-lives, permitting rapid corrections of excessive effect (Table 4.11.3). Urine levels of approximately 3.0 g phosphorous per g creatinine have been a good target in our experience, and seem less likely than greater values to cause nephrocalcinosis. Renal ultrasonography, creatinine clearance, and serum PTH levels should be monitored at least yearly. Dosage increases will be necessary as the child grows. Nephrocalcinosis in XLH seems to represent Ca^{2+}–Pi deposits, and this emphasizes the importance of Pi levels in this complication. However, it seems that subradiographic nephrocalcinosis does not compromise renal function [60]. Partial parathyroidectomy may become necessary if elevated serum PTH levels become persistent and are associated with hypercalcaemia. Some have used calcimimetics to control hyperparathyroidism. Hypercalciuria (>4 mg Ca^{2+} per kg of body weight, or >220 mg Ca^{2+} per g creatinine) can occur when skeletal mineralization is fully restored or when growth plates fuse. Lower doses may provide maintenance therapy until skeletal 'consolidation' is complete and cessation of medical treatment can be considered.

Orthopaedic evaluation should occur at least yearly during childhood, and twice yearly during the adolescent growth spurt, because limb bracing or epiphysiodesis may be helpful. Osteotomies are sometimes postponed until growth ceases to minimize the possibility of postoperative deformity. Unless patients are weight-bearing within 2 days of surgery (or fracture, etc.), 1,25-dihydroxyvitamin D_3 and then Pi therapy should be held until standing resumes to avoid immobilization hypercalciuria and hypercalcaemia.

Closure of physes after puberty does not mean that XLH is cured [51]. The metabolic derangements persist lifelong. Accordingly, affected adults should be followed perhaps yearly. Some may benefit from continuing 1,25-dihydroxyvitamin D_3 and Pi therapy to prevent fractures or worsening deformity [51].

In 2018, the medical treatment possibilities for paediatric and adult XLH changed substantially with the multinational approval of burosumab (Crysvita®) [23, 24]. Subcutaneous injection every two or four weeks, respectively, of this fully human monoclonal antibody against FGF23 rapidly and safely increases TmP/GFR and corrects hypophosphatemia with only a small increase within the normal range of circulating 1,25-dihydroxyvitamin D levels. Radiographic rickets can improve compared to continuing conventional therapy with oral 1,25-dihydroxyvitamin D and Pi supplementation [23, 24]. Growth and functionality also benefit [23, 24]. In adults with XLH given burosumab, pseudofractures may heal together with osteomalacia seen on iliac crest biopsy.

Dent's Disease

X-linked recessive hypophosphataemia (Dent disease 1) (OMIM #300009) maps to chromosome Xp11.22 and is due to

deactivation of the *CLCN5* gene involved in chloride transport [62]. Hypercalciuria nephrocalcinosis, β_2-microglobinuria, and progressive glomerular disease affect hemizygous males. Renal Pi wasting sometimes causes mild rickets. Treatment consists of Pi supplementation with caution not to cause hyperparathyroidism or to exacerbate nephrocalcinosis. Dent disease 2 (OMIM #300009) is caused by mutation of the *OCRL* gene.

Autosomal Dominant Hypophosphataemic Rickets

This rare form of renal Pi wasting (OMIM #193100) causes relatively mild rickets appearing during adolescence. The disorder maps to chromosome 12p13 and involves activating mutations in the gene encoding FGF23 (59). Treatment is similar to XLH, but lower doses of 1,25-dihydroxyvitamin D_3 and Pi are required. Treatment of any iron deficiency can be important.

Autosomal Recessive Hypophosphataemic Rickets

Deactivating mutations in the gene that encodes dentine matrix protein 1 (*DMP1*) causes a very rare autosomal recessive form of hypophosphataemic rickets featuring elevated circulating FGF23 and a diffusely sclerotic skeleton (OMIM #241520) [63]. A second type (OMIM #613312) involves deactivation of *ENPP1* gene.

Fanconi's Syndrome

Fanconi's syndrome features renal Pi wasting causing hypophosphatemia together with other manifestations of proximal renal tubule dysfunction including low serum levels of potassium, bicarbonate, and uric acid as well as aminoaciduria. There are many aetiologies including cystinosis, tyrosinaemia, and Lowe's syndrome (Table 4.11.1). Therapy with 1,25-dihydroxyvitamin D_3 and Pi supplementation (see XLH) seems helpful, but urinary Ca^{2+} levels must be monitored especially carefully because hypercalciuria can be present.

McCune–Albright Syndrome

McCune–Albright syndrome (OMIM #174800) often causes acquired hypophosphataemic rickets [18]. Treatment with 1,25-dihydroxyvitamin D_3 and Pi helps control only the rachitic skeletal disease superimposed on polyostotic fibrous dysplasia, but therapy in affected adults may be especially difficult to assess because of premature closure of growth plates and the underlying widespread fibrodysplastic disease. In fact, bone biopsy looking for osteomalacia may be challenging because of the fibrous dysplasia precluding a suitable biopsy site. Renal Pi wasting may diminish as the fibroblastic disease 'burns out' in middle age.

Vitamin D-Dependent Rickets

Vitamin D-dependent rickets (VDDR) types I and II (VDDR I and VDDR II) are rare, autosomal recessive disorders (OMIM #264700, #600081, #277440, %600785) that mimic vitamin D-deficiency rickets [5–8, 64]. However, there is no deficiency in cutaneous synthesis or accelerated loss of vitamin D. Patients are typically replete with vitamin D as shown by normal circulating levels of 25-hydroxyvitamin D [65].

VDDR I and II feature diminished biosynthesis of, and target tissue resistance to, 1,25-dihydroxyvitamin D, respectively. Because there is either disturbed conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D (VDDR I) or peripheral resistance to 1,25-dihydroxyvitamin D (VDDR II), serum levels of 1,25-dihydroxyvitamin D are low and high, respectively (Table 4.11.4) [5–8, 64]. Nevertheless, both types of VDDR alter mineral homeostasis in a similar way. Dietary Ca^{2+} is malabsorbed, leading to hypocalcaemia, secondary hyperparathyroidism, and hypophosphataemia. Decreased extracellular fluid levels of Ca^{2+} and Pi together impair mineralization of skeletal matrix. Because the pathogenesis of VDDR I involves defective production of 1,25-dihydroxyvitamin D by the kidney, physiological doses of 1,25-dihydroxyvitamin D_3 control the disorder [64]. However, in VDDR II even enormous doses of 1,25-dihydroxyvitamin D_3 may prove ineffective and Ca^{2+} may need to be administered intravenously [5–8]. Both VDDR I and II are understood at the gene level and as indicated next, now have more informative names [18, 46].

25-Hydroxyvitamin D, 1α-Hydroxylase Deficiency (Vitamin D-Dependent Rickets, Type I)

1,25-dihydroxyvitamin D deficiency can be defined as low circulating levels of this hormone despite normal or elevated (depending on preceding vitamin D therapy) concentrations of 25-hydroxyvitamin D. In theory, this situation could result from decreased production or increased clearance of 1,25-dihydroxyvitamin D. Indeed, decreased production of 1,25-dihydroxyvitamin D can be hereditary, but in other circumstances acquired. Acquired deficiency is usually explained by systemic disease, such as chronic renal failure or acquired Fanconi's syndrome, etc., which affect bone and mineral metabolism in multiple and complex ways. In contrast, increased clearance is uncommon and typically accompanies loss of other vitamin D metabolites, such as 25-hydroxyvitamin D, and would therefore fit within the definition of secondary vitamin D deficiency. The genetic entity discussed here (OMIM #264700) is now also called hereditary 1,25-dihydroxyvitamin D deficiency, and is an inborn error of metabolism featuring defective enzymatic biosynthesis of 1,25-dihydroxyvitamin D.

Prader and colleagues first characterized this disorder when they described two young children who showed all of the usual clinical features of vitamin D deficiency despite adequate intake of the vitamin. Complete remission depended upon continuous therapy with high doses of vitamin D—thus, the term 'vitamin D-dependent rickets'. They had instead coined the term 'pseudovitamin D deficiency'. Remission could, however, be achieved by physiological (microgram) doses of 1α-hydroxylated vitamin D metabolites [66, 67]. VDDR I is now understood at the molecular level and is, therefore, best described as 25-hydroxyvitamin D, 1α-hydroxylase deficiency [64].

Patients with 1α-hydroxylase deficiency appear healthy at birth. Subsequently, features consistent with nutritional rickets are usually noticed before 2 years of age, and often during the first 6 months of life. There is growth retardation and poor gross motor development. Muscle weakness, irritability, pneumonia, seizures, and failure to thrive are prominent findings.

Serum 1,25-dihydroxyvitamin D levels are low or undetectable despite normal levels of 25-hydroxyvitamin D. Malabsorption of dietary Ca^{2+} leads to hypocalcaemia, secondary hyperparathyroidism, and hypophosphataemia. Serum ALP activity is elevated.

The radiographic changes are in keeping with nutritional rickets. In addition to growth plate abnormalities and rachitic deformities, osteopaenia, and other features of secondary hyperparathyroidism

are present. Undecalcified bone documents defective matrix mineralization and secondary hyperparathyroidism including osteoclastosis and peritrabecular fibrosis [64].

Early reports of affected siblings in inbred kindreds indicated that VDDR I is an autosomal recessive condition, especially prevalent in French-Canadians [46]. A founder effect seems to have occurred in this population and, in 1990, linkage studies mapped the disorder to chromosome l2q14 [64]. The molecular defect compromises the kidney's mitochondrial cytochrome P450clα enzyme responsible for rate-limiting, hormonally regulated, 25-hydroxyvitamin D bioactivation to 1,25-dihydroxyvitamin D (i.e. 25-hydroxyvitamin D, lα-hydroxylase). Actually, this enzyme has several components, cytochrome P-450D10t, ferredoxin, and ferredoxin reductase [64]. A variety of mutations have been found in the P450clα gene (*CYP27B1*: OMIM 609506) [66]. French-Canadian patients are commonly homozygous for a 958ΔG defect in this single copy gene. None of these mutations engenders an enzyme with decreased (rather than absent) activity [66].

Serum concentrations of 25-hydroxyvitamin D are normal in VDDR I (elevated if pharmacological doses of vitamin D or 25-hydroxyvitamin D are given), yet 1,25-dihydroxyvitamin D levels are low, or remain only partially corrected by vitamin D or 25-hydroxyvitamin D therapy [63]. Because pharmacological doses of vitamin D_2 or D_3 or 25-hydroxyvitamin D_3 produce therapeutic responses in VDDR I similar to physiological (replacement) doses of 1,25-dihydroxyvitamin D_3, it is apparent that 25-hydroxyvitamin D (or some metabolite) at sufficient levels can activate the VDR. Alternatively, perhaps enhanced local 1,25-dihydroxyvitamin D biosynthesis occurs with pharmacological doses of the prohormones.

The 1α-hydroxylase gene from more than 25 families with this disorder has been studied by site-directed mutagenesis and cDNA expression in transfected cells. All patients had homozygous mutations. Most French-Canadian patients have the same mutation causing a frame shift and a premature stop codon in the putative haem-binding domain. However, the same mutation was then observed in additional families of diverse origin. All other patients had either a base-pair deletion causing a premature termination codon upstream from the putative ferredoxin and haem-binding domains, or missense mutations. No 1α-hydroxylase activity was detected when the mutant enzyme was expressed in various cells. The 1α-hydroxylase gene sequence in keratinocytes and peripheral blood mononuclear cells is identical with the renal gene.

The differential diagnosis for VDDR I includes especially defects in the VDR-effector system, where serum concentrations of 1,25-dihydroxyvitamin D and the response to treatment with 1α-hydroxylated vitamin D metabolites are greatly different (**Table 4.11.2**).

Clinical remission has followed daily, high-dose therapy with 1–3 mg of vitamin D_2, or with 0.2–0.9 mg of 25-hydroxyvitamin D. Because there is no defect in hepatic conversion of vitamin D to 25-hydroxyvitamin D, vitamin D rather than 25-hydroxyvitamin D is cheap yet effective. However, a physiological ('replacement') dose of 1,25-dihydroxyvitamin D, 0.25–1.0 μg daily, bypasses the lα-hydroxylase defect and provides an effective and direct treatment [49]. Although 25-hydroxyvitamin D_3 or 1,25-dihydroxyvitamin D_3 therapy is expensive, it has advantages. The physiological half-lives of these metabolites are much shorter than vitamin D, and excessive dosing will respond more rapidly to temporary cessation of therapy. Most patients, however, can be managed with vitamin D, but follow-up is essential for any regimen.

Hereditary Resistance to 1,25-Dihydroxyvitamin D (Vitamin D-Dependent Rickets, Type II)

This disorder was characterized in 1978 when a patient with features of 'pseudovitamin D deficiency' (see earlier) was found instead to have high serum levels of 1,25-dihydroxyvitamin D [66]. Thus, 'hereditary resistance to 1,25-dihydroxyvitamin D' or VDDR II refers to this condition (OMIM #277440) [5, 6]. Autosomal recessive inheritance is well established, and parental consanguinity has been reported in approximately 50% of cases [46].

There is a striking clustering of patients around the Mediterranean, including patients reported form Europe and America who originated from the same area [7, 8]. A notable exception is a cluster of kindreds from Japan [46]. Obligate heterozygotes do not have clinical or biochemical manifestations. Patients appear normal at birth, but typically then develop features resembling vitamin D deficiency during the first year of life [5–8]. Although several sporadic cases first manifested skeletal disease as late as their teenage years or middle age, they represent the mildest form of the disease and had complete remission when treated with vitamin D or its active metabolites. It was unclear if the adult-onset patients belong to this entity. In general, the earlier the presentation, the more severe the clinical and biochemical features [5–8].

Hypocalcaemia causes secondary hyperparathyroidism, hypophosphataemia, and elevated serum ALP activity. However, 1,25-dihydroxyvitamin D levels are elevated, sometimes as much as 10-fold [5–8]. This abnormality reflects peripheral resistance to 1,25-dihydroxyvitamin D causing malabsorption of dietary Ca^{2+} and the subsequent combined effects of three activators of renal 25-hydroxyvitamin D, lα-hydroxylase activity: hypocalcaemia, increased circulating PTH, and hypophosphataemia together with diminished feedback inhibition by 1,25-dihydroxyvitamin D on the kidney 1α-hydroxylase.

The radiographic and histological findings of VDDR II resemble those of nutritional rickets, as described before, including growth plate disturbances, rachitic deformities, osteopaenia, fractures, and evidence of secondary hyperparathyroidism.

A peculiar feature, appearing in more than half of the affected individuals, is total alopecia or sparse hair. Alopecia usually manifests during the first year of life and there may be additional ectodermal anomalies including oligodontia, epidermal cysts, and cutaneous milia [5–8]. In a patient with total alopecia, hair follicles were present.

Alopecia seems to be a marker for a more severe form of the disease, as judged by earlier onset, severity of the clinical features, proportion of patients who do not respond to treatment with high doses of vitamin D or its active metabolites, and the extremely elevated serum levels of 1,25-dihydroxyvitamin D during therapy. Although some patients with alopecia achieve clinical and biochemical remission of their bone disease, none have shown hair growth. The notion that total alopecia reflects a defective VDR-effector system is supported by the fact that alopecia has only been associated with hereditary defects in the VDR system (i.e. with end-organ resistance to the action of the hormone. Indeed, hair follicles normally contain the VDR).

Patients with VDDR II with normal hair can respond fully to high doses of bioactive vitamin D metabolites. However, only some with total alopecia do so. Remarkably, some patients with VDDR II may no longer need 1,25-dihydroxyvitamin D_3 therapy, or require lower doses, later in life [5–8]. A VDR-positive mild variant has been reported in Columbia, South America (OMIM % 600785) [46, 68].

The nature of the resistance to 1,25-dihydroxyvitamin D and aberrations in the VDR/effector system have been elucidated [5–8, 9]. A variety of VDR, or post-VDR, defects block the peripheral action of 1,25-dihydroxyvitamin D. There can be absence of the VDR, diminished binding capacity or binding affinity of the VDR for 1,25-dihydroxyvitamin D, or failure of the 1,25-dihydroxyvitamin D–VDR complex to localize to the nucleus or bind to DNA [8]. Patients without hormone or DNA binding by the VDR are the most difficult to treat [5, 6]. A mouse model has been developed by targeted ablation of the *VDR* gene.

If untreated, most patients with VDDR II die in early childhood [5–8]. However, good control of the disorder is possible with therapy, especially in individuals without alopecia. Depending upon severity, VDDR II may require calciferols that enhance endogenous production of 1,25-dihydroxyvitamin D, administration of high doses of both calciferols and Ca^{2+} to compensate for the target tissue resistance to 1,25-dihydroxyvitamin D, or high doses of Ca^{2+} alone (given orally or intravenously) to circumvent the target cell 1,25-dihydroxyvitamin D resistance [8, 60]. Whereas most patients may respond to very high oral doses of 1,25-dihydroxyvitamin D_3 (10–40 μg daily), some can have clinical, radiographic, and biochemical corrections with high doses of vitamin D_2 or 25-hydroxyvitamin D_3 [5–8]. Some patients have unexplained disease fluctuation.

Before therapy, serum 1,25-dihydroxyvitamin D concentrations range from the upper normal limit to markedly elevated. With vitamin D treatment, they may reach the highest levels found in any living system (≥100 times the upper normal limit).

The near ubiquity of a similar if not identical VDR-effector system among various cell types helped clarify the nature of the intracellular and molecular defects in these patients [7, 8]. A defect that compromises RXR heterodimerization with the VDR (essential for nuclear localization and probably for recognition of the vitamin D responsive element in the DNA as well) was characterized in several kindreds with and without alopecia [7, 8]. In kindreds with defects in VDR binding to DNA, different single nucleotide mutations in the DNA-binding region were found [7, 8]. All point mutations affected the region of the two zinc fingers of the VDR essential for functional interaction of the hormone receptor complex with DNA. Interestingly, all altered amino acids are highly conserved in the steroid receptor superfamily. In all such patients, no response followed very high doses of vitamin D or its active 1α-hydroxylated metabolites.

Normal hair is usually associated with milder and usually complete clinical and biochemical remission on high doses of vitamin D or its metabolites [7, 8]. In contrast, only ~50% of patients with alopecia have satisfactory clinical and biochemical remission to high doses of vitamin D or its active 1α-hydroxylated metabolites, and the dose requirement is about 10-fold higher than those with normal hair.

It seems that defects characterized as deficient hormone binding affinity and deficient heterodimerization with RXR achieve remission on high doses of vitamin D or its active 1α-hydroxylated metabolites. Most with other defects could not be cured.

Typical clinical and biochemical features (Table 4.11.4) support the diagnosis. The issue becomes more complicated when the clinical features are atypical, i.e. late onset, sporadic cases, and normal hair. Failure of a therapeutic trial with Ca^{2+} and/or physiological replacement doses of vitamin D or its active metabolites may support the diagnosis but now mutation analysis should reveal the aetiology and pathogenesis.

Based on the clinical and biochemical features, the following additional disease states should be considered: (1) extreme Ca^{2+} deficiency, and (2) severe vitamin D deficiency, because during the initial stages of vitamin D therapy in children with severe vitamin D-deficient rickets, the biochemical picture may resemble 1,25-dihydroxyvitamin D resistance, but this is a transient condition differentiated by the history of vitamin D deficiency and the final therapeutic response to vitamin D.

An adequate therapeutic trial must include vitamin D at sufficient doses to maintain high serum concentrations of 1,25-dihydroxyvitamin D because patients can produce high serum 1,25-dihydroxyvitamin D levels if supplied with substrate. If high serum levels are not achieved, 1α-hydroxylated vitamin D metabolites should be given in daily doses up to 6 μg/kg weight or a total of 30–60 μg and up to 3 g of elemental Ca^{2+} orally daily; therapy must continue for a period sufficient to mineralize the abundant osteoid (usually 3–5 months). Therapy may be considered a failure if no change in the clinical, radiological, or biochemical parameters occurs while serum 1,25-dihydroxyvitamin D concentrations are maintained at approximately 100 times average normal values. Several patients have shown unexplained fluctuations in response to therapy or at presentation of the disease [7, 8].

In some patients unresponsive to vitamin D or its metabolites, clinical and biochemical remission, including catch-up growth, accompanied administration of large amounts of Ca^{2+} achieved by long-term (months) intracaval infusions of up to 1000 mg of Ca^{2+} daily. Alternatively, increasing oral Ca^{2+} intake was used successfully in only very few patients and this approach was limited by dose and patient tolerability [69].

Hypophosphatasia

In 1948, 'hypophosphatasia' (OMIM #241500, #146300, #241510) was coined to distinguish a rare heritable form of rickets characterized biochemically by hypophosphatasaemia due to deficient activity of the tissue non-specific (liver/bone/kidney) isoenzyme of ALP (TNSALP) [15]. More than 400 different mutations in the *ALPL* gene (OMIM *171760) have been discovered in patients with hypophosphatasia (HPP) worldwide [70]. HPP is the instructive inborn error of metabolism that confirmed the theory promulgated by Robert Robison, PhD beginning in 1923 that ALP promotes the mineralization of cartilage and bone.

More than 800 cases of HPP have been reported. The severity is remarkably broad-ranging and spans intrauterine death from profound skeletal hypomineralization to loss of teeth or calcific arthritis in adults [15]. Now, six clinical forms are reported depending primarily on patient age when skeletal disease is documented. Perinatal, infantile, mild, or severe childhood, and adult HPP feature skeletal complications [15]. Children and adults who manifest tooth loss without skeletal disease (radiographically or on bone biopsy) have odonto HPP. Although artificial, this clinical classification provides information concerning recurrence risks and prognosis.

Perinatal HPP is diagnosed at birth and is almost always rapidly lethal [15, 70]. Stillbirth is common. Profound skeletal hypomineralization with caput membranaceum and short and deformed limbs is obvious. Severe osteogenesis imperfecta or cleidocranial dysplasia may be suspected, but can be distinguished radiographically and by gene testing [16]. Occasionally, bony spurs protrude from the shafts of major long bones. Failure to gain weight, irritability with a high-pitched cry, unexplained fever, anaemia, periodic apnoea with bradycardia, and intracranial haemorrhage can occur. Respiratory compromise proves fatal from chest weakness and deformity and perhaps pulmonary hypoplasia.

Infantile HPP becomes clinically apparent postnatally before 6 months of age with failure to thrive, poor feeding, rickets, widened fontanelles, hypotonia, and sometimes vitamin B_6-dependent seizures [15]. Hypercalcaemia and hypercalciuria may explain bouts of vomiting and nephrocalcinosis, sometimes with significant renal impairment. Rachitic deformity of the chest and rib fractures predispose to recurrent pneumonias. Seizures and spells of apnoea may occur. Despite the impression from palpation or radiographs that cranial sutures are widely open, functional craniosynostosis from skull hypomineralization is common. Infantile hypophosphatasia often features progressive clinical and radiographic deterioration, and untreated ~50% of patients die within the first year of life. However, the prognosis seems better if there is survival past infancy spontaneously, although persisting skeletal disease seems likely [15].

Childhood HPP is especially variable [14]. Premature loss of deciduous teeth (age <5 years) from hypoplasia and hypomineralization of cementum may be the most remarkable clinical manifestation. Cementum anchors dentition to the periodontal ligament, therefore, deciduous teeth in HPP are shed without root resorption. Mandibular incisors are usually lost first, but the entire dentition can be exfoliated early. Enlarged pulp chambers and root canals result in 'shell' teeth characteristic of disorders that cause rickets. Skeletal deformity can include scaphocephaly with frontal bossing, a rachitic rosary, bowed legs or knock-knees, short stature, and wrist, knee, or ankle enlargement. When radiographs disclose rickets, delayed walking and a characteristic waddling gait are common. Childhood HPP may improve clinically when growth plates fuse after puberty, but recurrence of symptoms seems likely sometime during adult life [15, 70].

Adult HPP presents during middle age [15, 70]. Approximately 50% of these patients recall being told of rickets and/or premature loss of teeth during childhood. Often, there are recurrent, poorly healing, metatarsal stress fractures. Subtrochanteric femoral pseudofractures may be found proximally in the lateral cortices [25]. Chondrocalcinosis is common and inorganic pyrophosphate deposition may also cause arthritis including pseudogout.

Radiographic findings in HPP are helpful for diagnosis, especially in paediatric patients. Perinatal HPP features pathognomonic changes as the skeleton can be so hypomineralized that only the skull base is apparent. Individual vertebrae can be 'missing', and bony spurs may protrude from major long bones. Alternatively, severe rachitic changes are seen. Calvarial bones can be mineralized only centrally, giving the illusion that sutures are widely patent. Fractures are common. In infants, abrupt transition from well mineralized diaphyses to hypomineralized metaphyses suggests sudden metabolic deterioration. Relentless skeletal demineralization with worsening rachitic disease and progressive bony deformity and fracturing as well as vitamin B_6-dependent seizures predict a lethal outcome. Bone scintigraphy showing little tracer uptake in widely separated cranial 'sutures' suggests functional suture closure. Severely affected patients who survive infancy can develop true premature cranial sutures fusion causing a 'beaten copper' radiographic appearance and raised intracranial pressure (**Figure 4.11.7**). In affected children, characteristic 'tongues' of radiolucency extend from physes into metaphyses of major long bones where there can also be patchy osteosclerosis (**Figure 4.11.2**). Adult HPP causes recurrent, poorly healing, metatarsal stress fractures and femoral pseudofractures that appear laterally (rather than medially as in other forms of osteomalacia). There can also be osteopaenia and chondrocalcinosis with changes of pyrophosphate arthropathy.

Subnormal serum ALP activity for age and sex (hypophosphatasaemia) is the biochemical hallmark of HPP. Generally, the levels reflect disease severity [15, 70]. Patients with odonto HPP have mild but discernible hypophosphatasaemia. In fact, this finding is especially impressive because rickets or osteomalacia typically cause hyperphosphatasaemia [14]. Several other conditions, some with skeletal manifestations, lower blood ALP levels [15], but are distinguished from HPP partly because the patients do not accumulate TNSALP substrates (see next). Serum levels of Ca^{2+} and Pi are not diminished in HPP. In fact, hypercalciuria and hypercalcaemia often complicate the infantile form where the pathogenesis seems to involve a 'dyssynergy' between gut absorption of dietary Ca^{2+} and defective skeletal mineralization; however, skeletal demineralization may also become a factor. Serum levels of PTH, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D are usually unremarkable in HPP unless there is hypercalcaemia or renal compromise. Serum Pi concentrations are above control mean levels, and mild hyperphosphataemia occurs in more than half of

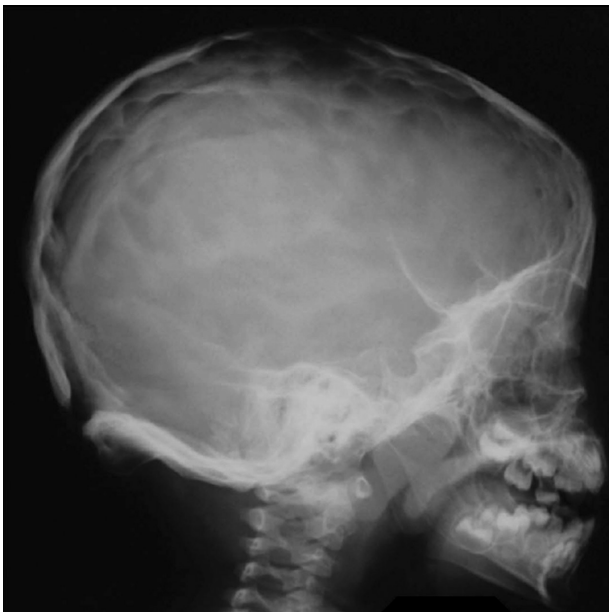


Figure 4.11.7 Hypophosphatasia: The 'beaten copper' skull of this 2-year-old boy with the childhood form of hypophosphatasia results from premature closure of cranial sutures. Previously, he underwent craniotomy.

affected children and perhaps adults. The pathogenesis involves enhanced renal reclamation of Pi [70].

Three phosphocompounds, natural substrates for TNSALP, accumulate endogenously in HPP: phosphoethanolamine, inorganic pyrophosphate (PPi), and pyridoxal 5′-phosphate (PLP) [15, 70]. Assays are commercially available for urinary and blood phosphoethanolamine and plasma PLP. Mild phosphoethanolaminuria occurs in several metabolic bone diseases [15]. Fortunately, increased plasma PLP concentration is a particularly sensitive and specific marker for HPP. However, patients must not be taking vitamin B_6 for at least one week before tested. Endogenous accumulation of PPi seems to be the key pathogenetic factor accounting for the defective hard tissue mineralization (see next), yet quantitation of PPi remains a research technique.

Defective skeletal mineralization occurs in all clinical forms of HPP except odonto HPP [2]. Unless evaluation of the ALP activity in the skeleton is undertaken, the histopathological findings are those of other types of rickets or osteomalacia lacking secondary hyperparathyroidism.

HPP occurs in all races, and is especially prevalent among Mennonites and Hutterites in Canada. In North America the incidence of severe disease is ~ 1/100 000 live births [15]. Perinatal and nearly all cases of infantile HPP are transmitted as an autosomal recessive trait. Carriers can have diminished or low-normal levels of serum ALP activity, and sometimes modest elevations in plasma PLP levels [15, 70]. The inheritance pattern(s) for childhood, adult, and odonto HPP is autosomal recessive or autosomal dominant. Generation-to-generation transmission is typically associated with relatively mild clinical expression.

The *ALPL* gene has 12 exons and to exists as a single copy in the haploid genome on the tip of the short arm of chromosome 1 (lp36.1–lp34). In 1988, a missense mutation was identified in a severely affected infant from an inbred Canadian kindred [70]. Subsequent studies have disclosed > 400 different *ALPL* mutations [70]. Most are missense defects. Perinatal and infantile HPP reflect homozygosity or compound heterozygosity for these mutations. The childhood and adult forms of HPP can be the 'same' disease [15]. Mouse models first recapitulated the infantile form of HPP [71], but now some match the mild forms.

ALP (orthophosphoric monoester phosphohydrolase (alkaline optimum), EC 3.1.3.1), found in nearly all organisms, is a glycosylated, plasma membrane-bound enzyme [70]. Discovery of the endogenous accumulation of phosphoethanolamine, PPi, and PLP in HPP revealed how TNSALP functions [15, 70]. Accumulation of PLP, the principal cofactor form of vitamin B, indicated that TNSALP acts as a cell-surface enzyme. Patients with HPP typically do not have symptoms or signs of vitamin B_6 deficiency or toxicity despite their markedly increased plasma PLP levels.

In 1965, discovery of elevated urinary and then blood levels of PPi in HPP disclosed the pathogenesis of the rickets and osteomalacia. PPi is a potent inhibitor of biomineralization. Matrix vesicles are devoid of ALP activity in HPP, but do contain hydroxyapatite crystals [4]. However, only a few isolated crystals are observed outside these extracellular structures. The excess PPi in HPP blocks hydroxyapatite crystal formation in the extracellular matrix of affected bone.

Unless documented, conventional treatments for rickets or osteomalacia are generally best avoided in HPP because patients are usually vitamin D replete and serum levels of Ca^{2+} and Pi are not low [15]. Indeed, such treatment could exacerbate or provoke hypercalcaemia and hypercalciuria. Hypercalcaemia in infantile HPP generally responds to reduction in dietary Ca^{2+} intake, but may require glucocorticoid therapy. Supportive therapy is important for HPP. Fractures do mend, but delayed healing after casting or osteotomy has been observed. In affected adults, placement of intramedullary rods, rather than load-sparing devices (e.g. plates), seems to be preferable for the acute or prophylactic treatment of fractures and pseudofractures [25]. Expert dental care is especially important for children, because their nutrition can be impaired by premature tooth loss. Craniotomy may be crucial for children with craniosynostosis. Fetuses that are severely affected (perinatal form) can be detected *in utero* early on by ultrasonography, but a relatively mild 'benign prenatal' form of non-lethal HPP must be considered. *ALPL* gene mutation testing has improved diagnosis [15].

Enzyme replacement by intravenous infusion of various natural soluble forms of ALP has generally been disappointing for HPP [70]. Two infants who seemed destined to die from infantile HPP showed clinical and radiographic improvement following transplantation of marrow or bone-derived cells [72]. This suggested that TNSALP must be increased within the skeleton itself. Hydroxyapatite-targeted enzyme replacement therapy (Strensig®) using subcutaneous injections of a recombinant TNSALP fusion protein (asfotase alfa) was approved multinationally in 2015 typically for paediatric-onset HPP [73]. Results of clinical trials of HPP patients of all ages have documented clinical and skeletal improvements [20–22].

ACKNOWLEDGEMENTS

Supported in part by Shriners Hospitals for Children, The Clark and Mildred Cox Inherited Metabolic Bone Disease Research Fund at The Barnes-Jewish Hospital Foundation.

REFERENCES

1. Parfitt AM. Vitamin D and the pathogenesis of rickets and osteomalacia. In: Feldman D, Pike JW, Glorieux FH, eds. *Vitamin D*, 2nd edn. Amsterdam: Elsevier/Academic Press, 2005: 1029–48.
2. Holick MF, ed. *Vitamin D: Physiology, Molecular Biology, and Clinical Applications*, 2nd edn. Totawa, NJ: Humana Press, 2010.
3. Feldman D, Pike JW, Bouillon R, eds. *Vitamin D*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2018.
4. Bilezikian JP, Raisz LG, Martin, TJ, eds. *Principles of Bone Biology*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2019.
5. Holick MF. Photobiology of vitamin D. Chapter 4. In: Feldman D ed. *Vitamin D*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2018: 45–55.
6. Malloy PJ, Tiosano D, Feldman D. Hereditary 1,25-dihydroxyvitamin D resistant rickets. Chapter 72. In: Feldman D, ed. *Vitamin D*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2018: 264–94.
7. Liberman UA, Marx SJ. Vitamin D and other calciferols. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn. New York, NY: McGraw-Hill, 2001: 4223–40.

8. Liberman UA. Hereditary deficiencies in vitamin D action. In: Bilizikian JP, Raisz LG, Rodan GA, eds. *Principles in Bone Biology*, 3rd edn. San Diego, CA: Elsevier/Academic Press, 2008: 1195–208.
9. Haussler MR, Haussler CA, Jurutka PW, *et al*. The vitamin D hormone and its nuclear receptor: molecular actions and disease states. *J Endocrinol*, 1997; **154**: S57–73.
10. White KE, Larsson TE, Econs MJ. The roles of specific genes implicated as circulating factors involved in normal and disordered phosphate homeostasis: frizzled related protein-4, matrix extracellular phosphoglycoprotein, and fibroblast growth factor 23. *Endocr Rev*, 2006; **27**: 221–41.
11. Tenenhouse HS, Econs MJ. Mendelian hypophosphatemias. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn. New York, NY: McGraw-Hill, 2001: 5039–67.
12. Whyte MP. Approach to the patient with metabolic bone disease. Chapter 48. In: Feldman D, ed. *Vitamin D*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2018: 887–902.
13. Pettifor JM Thandrayen K, Thacher TD. Vitamin D deficiency and nutritional rickets in children. Chapter 67. In: Feldman D, ed. *Vitamin D*, 4th edn. San Diego, CA: Elsevier Academic Press, 2018: 179–94.
14. Glorieux FH, ed. *Rickets*. New York, NY: Raven Press, 1991.
15. Whyte MP. Hypophosphatasia and how alkaline phosphatase promotes mineralization. Chapter 28. In: Thakker RV, Whyte MP, Eisman J, Igarashi T, eds. *Genetics of Bone Biology and Skeletal Disease*. San Diego, CA: Elsevier/Academic Press, 2018: 481–504.
16. Adams JE. Radiology of rickets and osteomalacia. Chapter 54. In: Feldman D, ed. *Vitamin D*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2018: 975–1001.
17. Mumford E, Marks J, Wagner T, *et al*. Oncogenic osteomalacia: diagnosis, localisation, and cure. *Lancet Oncol*, 2018; **19**: e365.
18. Bilezikian JP, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, 9th edn. Hoboken, NJ: Wiley Blackwell, 2019.
19. Whyte MP. Hypophosphatasia: aetiology, nosology, pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*, 2016; **12**: 233–46.
20. Whyte MP, Simmons JH, Moseley S, *et al*. Asfotase alfa for infants and young children with hypophosphatasia: 7 year outcomes of a single-arm, open-label, phase 2 extension trial. *Lancet Diabetes Endocrinol*, 2019; **7**: 93–105.
21. Whyte MP, Madson KL, Phillips D, *et al*. Asfotase alfa therapy for children with hypophosphatasia. *JCI Insight*, 2016; **1**: e85971.
22. Kishnani PS, Rockman-Greenberg C, Rauch F, *et al*. Efficacy and safety of asfotase alfa for adolescents and adults with hypophosphatasia treated for 5 years. *Bone*, 2019; **121**: 149–62.
23. Carpenter T, Whyte MP, Imel E, *et al*. Burosumab in children with X-linked hypophosphatemia. *N Engl J Med*, 2018; **378**: 1987–98.
24. Whyte MP, Carpenter T, Gottesman GS, *et al*. Efficacy and safety of burosumab in children aged 1–4 years with x-linked hypophosphatemia: a multicenter, open-label, phase 2 trial. *Lancet Diabetes Endocrinol*, 2019; **7**: 189–99.
25. Coe JD, Murphy WA, Whyte MP. Management of femoral fractures and pseudofractures in adult hypophosphatasia. *J Bone Joint Surg Am*, 1986; **68**: 981–90.
26. Thomas MK, Lloyd-Jones DM, Thadhani RI, *et al*. Hypovitaminosis D in medical inpatients. *N Engl J Med*, 1998; **338**: 777–83.
27. Lips P, Duong D, Oleksik A, *et al*. A global study of vitamin D status and parathyroid function in postmenopausal with osteoporosis: baseline data from the Multiple Outcomes of Raloxifene Evaluation Clinical Trial. *J Clin Endocrinol Metab*, 2001; **86**: 1212–21.
28. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. *Am J Clin Nutr*, 2006; **84**: 18–28.
29. Bischoff-Ferrari HA, Willett WC, Wong JB, Giovanucci E, Dietrich T, Dawson-Hughes B. Fracture prevention with vitamin D supplementation: a meta-analysis of randomized control trials. *JAMA*, 2005; **293**: 2257–64.
30. Thacher TD, Fischer PR, Pettifor JM, *et al*. A comparison of calcium, vitamin D, or both for nutritional rickets in Nigerian children. *N Engl J Med*, 1999; **341**: 563–8.
31. Bischoff-Ferrari HA. How to select the dose of vitamin D in the management of osteoporosis. *Osteoporos Int*, 2007; **18**: 401–7.
32. Lips P, Hosking D, Lippunes K, *et al*. The prevalence of vitamin D inadequacy amongst women with osteoporosis: an international epidemiological investigation. *Journal of Internal Medicine*, 2006; **260**: 245–54.
33. Mithal A, Wahl DA, Bonjour JP, *et al*. Global vitamin D status and determinants of hypovitaminosis D. *Osteoporos Int*, 2009; **20**: 1807–20.
34. National Osteoporosis Foundation. Physician Guide to Prevention and Treatment of Osteoporosis. Updated recommendation for calcium and vitamin D intake. Available at: https://www.nof.org/patients/treatment/calciumvitamin-d/ (accessed: 3 March 2020).
35. Hoppe B, Gnehm HE, Wopmann M, Neuhaus T, Willi U, Leumann E. Vitamin D poisoning in infants: a preventable cause of hypercalciuria and nephrocalcinosis. *Schweiz Med Wochenschr*, 1992; **122**: 257–62.
36. Whyte MP, Haddad JG Jr., Walters D, Stamp TCB. Vitamin D bioavailability: serum 25-hydroxyvitamin D levels in man following oral, subcutaneous, intramuscular, and intravenous Vitamin D administration. *J Clin Endocrinol Metab*, 1979; **48**: 906–11.
37. Pivnick EK, Kerr NC, Kaufman RA, Jones DP, Chesney RW. Rickets secondary to phosphate depletion. A sequela of antacid use in infancy. *Clin Pediatr*, 1995; **34**: 73–8.
38. Skinner R, Pearson AD, English MW, *et al*. Risk factors for ifosfamide nephrotoxicity in children. *Lancet*, 1996; **348**: 578–80.
39. Wolf M, Chertow GM, Macdougall IC, Kaper R, Krop J, Strauss W. Randomized trial of intravenous iron-induced hypophosphatemia. *JCI Insight*, 2018; **3**: e124486.
40. D'Haese PC, Schrooten I, Goodman WG, *et al*. Increased bone strontium levels in hemodialysis patients with osteomalacia. *Kidney Int*, 2000; **57**: 1107–14.
41. Ozgür S, Sümer H, Koçoğlu G. Rickets and soil strontium. *Arch Dis Child*, 1996; **75**: 524–6.
42. Tucci RJ, Whitford GM, Novack D, *et al*. Skeletal fluorosis due to inhalation abuse of a difluoroethane-containing computer cleaner. *J Bone Miner Res*, 2017; **32**: 188–95.
43. Kurland ES, Schulman RC, Zerwekh JE, Dempster DW, Reinus WR, Whyte MP. Recovery from skeletal fluorosis (an enigmatic, American case). *J Bone Miner Res*, 2007; **22**: 163–70.

44. Smith PS, Bayliss SJ, Shinawi MS, Gottesman GS, McAlister WH, Sugarman J, Whyte MP, Arbelaez AM. First report of burosumab (anti-FGF23 monoclonal antibody) for rickets complicating HRAS-associated cutaneous skeletal hypophosphatemia syndrome. Proceedings of the 2019 Annual Meeting of the Endocrine Society.
45. Bhadada SK, Dhiman V, Mukherjee S, *et al*. Fibrogenesis imperfecta ossium and response to human growth hormone: a potential therapy. *J Clin Endocrinol Metab*, 2017, **102**: 1750–6
46. Online Mendelian Inheritance in Man, OMIM (TM). *McKusick-Nathans Institute of Genetic Medicine*, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 12–21–2009. Available at: http://www.ncbi.nlm.nih.gov/omim/ (accessed: 5 March 2019).
47. Albright F, Butler AM, Bloomberg E. Rickets resistant to vitamin D therapy. *Am J Dis Child*, 1937; **54**: 529–47.
48. Winters RW, Graham JB, Williams TF, McFalls VW, Burnett CH. A genetic study of familial hypophosphatemia and vitamin D resistant rickets with a review of the literature. *Medicine (Baltimore)*, 1958; **37**: 97–142.
49. Menking M, Sotos JF. Effect of administration of oral neutral phosphate in hypophosphatemic rickets. *J Pediatr*, 1969; **75**: 1001–7.
50. The HYP Consortium. A gene (PEX) with homologies to endopeptidases is mutated in patients with X-linked hypophosphatemic rickets. *Nat Genet*, 1995; **11**: 130–6.
51. Reid IR, Hardy DC, Murphy WA, Teitelbaum SL, Bergfeld MA, Whyte MP. X-linked hypophosphatemia: a clinical, biochemical, and histopathologic assessment of morbidity in adults. *Medicine (Baltimore)*, 1989; **68**: 336–52.
52. Petersen DJ, Boniface AM, Schranck FW, Rupich RC, Whyte MP. X-linked hypophosphatemic rickets: a study (with literature review) of linear growth response to calcitriol and phosphate therapy. *J Bone Miner Res*, 1992; **7**: 583–97.
53. Reid IR, Hardy DC, Murphy WA, Teitelbaum SL, Bergfeld MA, Whyte MP. X-linked hypophosphatemia: skeletal mass in adults assessed by histomorphometry, computed tomography, and absorptiometry. *Am J Med*, 1991; **90**: 63–9.
54. Boukpessi T, Hoac B, Coyac BR, *et al*. Osteopontin and the dento-osseous pathobiology of X-linked hypophosphatemia. *Bone*, 2017; **95**: 151–61.
55. Nesbitt T, Coffman TM, Griffiths R, Drezner MK. Cross-transplantation of kidneys in normal and Hyp mice: evidence that the Hyp mouse phenotype is unrelated to an intrinsic renal defect. *J Clin Invest*, 1992; **89**: 1453–9.
56. Whyte MP, Schranck FW, Armamento-Villareal R. X-linked hypophosphatemia: a search for gender, race, anticipation, or parent-of-origin effects on disease expression in children. *J Clin Endocrinol Metab*, 1996; **81**: 4075–80.
57. Whyte MP, Christie PT, Podgornik MN, *et al*. X-linked hypophosphatemia (XLH): mutations compromising PHEX structure reflect a severe phenotype (abstract). *Am J Hum Genet*, 1999; **65**: A114.
58. Smith PS, Gottesman GS, Zhang F, *et al*. X-Linked hypophosphatemia: uniquely mild disease associated with PHEX 3'-UTRc.*231A>G mutation (a retrospective, case-control study). *J Bone Miner Res*. 2020; **35**: 920–31.
59. White KE, Evans WE, O'Riordan JLH, *et al*. Autosomal dominant hypophosphataemic rickets is associated with mutations in FGF-23. *Nat Genet*, 2000; **26**: 345–8.
60. Glorieux FH, Marie PJ, Pettifor JM, Delvin EE. Bone response to phosphate salts, ergocalciferol, and calcitriol in hypophosphatemic vitamin D-resistant rickets. *N Engl J Med*, 1980; **303**: 1023–31.
61. Eddy MC, McAlister WH, Whyte MP. X-linked hypophosphatemia: normal renal function despite medullary nephrocalcinosis 25 years after vitamin D_2-induced azotemia. *Bone*, 1997; **21**: 515–20.
62. Scheinman SJ, Guay-Woodford LM, Thakker RJ, Warnock DG. Genetic disorders of renal electrolyte transport. *Mechanisms of Disease*, 1999; **340**: 1177–87.
63. Whyte MP, Amalnath SD, McAlister WH, *et al*. Osteosclerotic hypophosphatemic osteomalacia associated with novel homozygous mutations of *DMP1* encoding dentin matrix protein 1 and *SPP1* encoding osteopontin: the first digenic sibling protein osteopathy. *Bone*, 2020; **132**: 115190.
64. Glorieux FH, St-Arnaud R. Vitamin D hydroxylation-deficient rickets, Type 1A: CYP27B1. In: Feldman D, Pike W, Adams J, Glorieux FH, eds. *Vitamin D*, 4th edn. San Deigo: Elsevier/Academic Press, 2018: 249–60.
65. Casella SJ, Reiner BJ, Chen TC, Holick MF, Harrison HE. A possible genetic defect in 25-hydroxylation as a cause of rickets. *J Pediatr*, 1994; **124**: 929–32.
66. Wang JT, Lin CJ, Burridge SM, *et al*. Genetics of vitamin D 1alpha-hydroxylase deficiency in 17 families. *Am J Hum Genet*, 1998; **63**: 1694–702.
67. Miller WL, Portale AA. Genetic causes of rickets. *Curr Opin Pediatr*, 1999; **11**: 333–9.
68. Giraldo A, Pino W, Garcia-Ramirez LF, Pineda M, Iglesias A. Vitamin D dependent rickets type II and normal vitamin D receptor cDNA sequence. A cluster in a rural area of Cauca, Colombia, with more than 200 affected children. *Clin Genet*, 1995; **48**: 57–65.
69. Wong GW, Leung SS, Law WY, Cheung NK, Oppenheimer SJ. Oral calcium treatment in vitamin D-dependent rickets type II. *J Paediatr Child Health*, 1994; **30**: 444–6.
70. Whyte MP. Hypophosphatasia: nature's window on alkaline phosphatase function in humans. Chapter 66. In: Bilezikian JP, Raisz LG, Martin TJ, eds. *Principles of Bone Biology*, 4th edn. San Diego, CA: Elsevier/Academic Press, 2019.
71. Millán JL, Narisawa S, Lemire I, *et al*. Enzyme replacement therapy for murine hypophosphatasia. *J Bone Miner Res*, 2008; **23**: 876–86.
72. Cahill RA, Wenkert D, Perlman SA, *et al*. Infantile hypophosphatasia: trial of transplantation therapy using bone fragments and cultured osteoblasts. *J Clin Endocrinol Metab*, 2007; **92**: 2923–30.
73. Whyte MP: Hypophosphatasia: enzyme replacement therapy bring new opportunities and new challenges. *J Bone Miner Res*, 2017; **32**: 667–75.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

4.12

Glucocorticoid-Induced Osteoporosis

Gherardo Mazziotti, Ernesto Canalis, and John P. Bilezikian

Introduction

Over the past two decades, significant advances in our understanding of glucocorticoid-induced osteoporosis (GIO), the most common form of secondary osteoporosis in women and men, have been made [1]. While it can be caused by endogenous glucocorticoid excess, as seen in Cushing's syndrome [2], GIO is almost always due to exogenous glucocorticoids which are used in a wide variety of disorders including autoimmune, pulmonary, and gastrointestinal diseases, as well as in patients following organ transplantation and in individuals with malignancies. Despite the widespread therapeutic use of glucocorticoids, and its acknowledged major role in causing osteoporosis, many patients receiving long-term glucocorticoid therapy neither undergo an evaluation for skeletal health nor do they often receive specific preventive or therapeutic agents when indicated [2].

This chapter deals with the pathophysiological, clinical, and therapeutic aspects of exogenous and endogenous GIO.

Epidemiology

Despite advances in targeted therapies for many autoimmune disorders, long-term use of glucocorticoids occurs in an estimated 1% of the population [3]. In a multinational population-based prospective, observational study of 60 393 postmenopausal women who had visited their primary care medical provider within the last 2 years, the Global Longitudinal Study of Osteoporosis in Women, up to 4.6% were currently taking oral glucocorticoids [4]. Vertebral fractures are the hallmark of GIO and they occur in about one-third of patients exposed to glucocorticoid excess [5]. After starting glucocorticoid therapy, fracture risk, a function also of dose and duration of use, increases rapidly. Prolonged exposure to doses of prednisolone—or the equivalent—as low as 2.5–5 mg daily are associated with an increased risk of hip and vertebral fractures [6]. In Amiche *et al.*'s Bayesian meta-analysis of fracture risk associated with oral glucocorticoid use, the annual incidences of vertebral fracture were 5.1% and 3.2% in individuals with exposure to glucocorticoid therapy shorter or longer than 6 months, respectively [7]. As expected, the greatest increase in fracture incidence was seen in postmenopausal women and elderly men. Even for those whose glucocorticoid exposure is cyclic, the risk of osteoporotic fractures remains increased [8]. It is noteworthy that fracture risk decreases after discontinuation of oral corticosteroids, although the time it takes to reduce risk is variable [1].

Inhaled glucocorticoids have effects on skeletal markers of bone metabolism and on bone mineral density (BMD), but the effects are variable and their significance as it relates to fracture risk is uncertain [9, 10]. Dose, duration of treatment, and degree of absorption, as well as the underlying disease, are all confounding factors.

Endogenous hypercortisolism is less frequently a cause of GIO because the disease itself is uncommon [11]. However, up to 10% of subjects attending an outpatient clinic of osteoporosis were suspected to have subclinical Cushing's syndrome [12]. Either clinical or subclinical fragility fractures can be the presenting manifestation of Cushing's syndrome [13]. In cross-sectional studies, the prevalence of fractures has been reported to be 15–50%, with the most frequent site of fracture involving the ribs and vertebrae [13]. These rates are likely to be an underestimation of real fracture risk in endogenous hypercortisolism, since not all patients underwent a complete morphometric and radiological evaluation to detect asymptomatic vertebral and rib fractures [13]. In fact, when searched by a systematic radiological approach, to detect vertebral fractures, the prevalence in those with Cushing's disease was a remarkable 78%, with only half clinically evident with pain, functional limitation, and loss of height [14]. Even among young patients with endogenous hypercortisolism, fracture prevalence is high [15]. Vestergaard *et al.* reported a sixfold increase in the risk of clinical fractures in the 2 years prior to the diagnosis of endogenous hypercortisolism, without a significant difference between Cushing's syndrome due to pituitary or adrenal disease [16]. High risk of fractures was reported

even in patients with subclinical endogenous hypercortisolism [17]. When Cushing's syndrome is effectively treated, recovery of bone loss occurs but it is gradual and often incomplete [13].

Pathophysiology

The effects of glucocorticoids on bone are biphasic. Under physiological conditions, endogenous glucocorticoids can be positive regulators of mesenchymal cell differentiation into mature osteoblasts, as suggested by some *in vitro* studies [13]. Cultures overexpressing 11β-hydroxysteroid dehydrogenate type 2 (11β-hsd2), are resistant to glucocorticoids and exhibit reduced osteoblastogenesis and predominant adipogenesis when compared with wild-type cultures [18]. Consistent with these observations, transgenic mice overexpressing 11β-hsd2 in mature osteoblasts exhibit vertebral osteopaenia [19].

In human subjects, the pathophysiology of bone loss in the context of glucocorticoid excess involves direct and indirect deleterious effects on the skeleton [20].

The central pathophysiological mechanism of GIO is reduced bone formation, due to actions that affect osteoblast differentiation and function. The early effect, however, is a significant increase in bone resorption, this mechanism perhaps accounting for the early increase of risk of fractures.

Direct Effects of Glucocorticoids on Bone Cells

Glucocorticoids decrease the number and the function of osteoblasts. These effects lead to a suppression of bone formation, a central feature of the pathogenesis of GIO. Glucocorticoids decrease the replication of cells of the osteoblastic lineage, reducing the pool of cells that may differentiate into mature osteoblasts. The mechanism by which this occurs may relate to redirecting multipotential bone marrow stromal cells into the adipocyte line and away from osteoblast lineage precursors. Thus, glucocorticoids impair osteoblastic differentiation and maturation at the early stages of osteoblast development.

Mechanisms involved in this redirection of stromal cells include induction of nuclear factors of the CAAT enhancer binding protein family and the induction of peroxisome proliferator-activated receptor γ 2 (PPARγ 2), both of which play essential roles in adipogenesis [21].

A fundamental mechanism by which glucocorticoids inhibit osteoblast cell differentiation is by opposing Wnt/β-catenin signalling. Wnt signalling has emerged as a key regulator of osteoblastogenesis [21] and osteoclastogenesis [22]. In skeletal cells, the canonical Wnt/β-catenin signalling pathway operates. When Wnt is absent (or inhibited), β-catenin is phosphorylated by glycogen-synthase kinase-3β (GSK-3β), and then degraded by ubiquitination in the cytoplasm, thus preventing translocation into the nucleus. The binding of Wnt to specific receptors, called Frizzled, and to coreceptors, such as low-density lipoprotein receptor related proteins (LRP)-5 and -6, leads to an inhibition of GSK-3β activity. When GSK-3β is not active, stabilized β-catenin is able to translocate to the nucleus, where it associates with transcription factors to regulate gene expression. Inactivation of either Wnt or *Ctnnb* (encoding for β-catenin) results in impairment of osteoblastogenesis, and increased osteoclastogenesis, since canonical Wnt signalling results in a direct inhibitory effect on osteoclast progenitors and induction of osteoprotegerin (OPG) production [22, 23]. The Wnt pathway can be opposed by Wnt antagonists, such as Dickkopf-1 and sclerostin that prevent Wnt from binding to its receptor complex. Glucocorticoids enhance Dickkopf-1 and sclerostin expression, and maintain GSK 3-β in an active state, leading ultimately to the inactivation of β-catenin [21]. The importance of sclerostin in mediating the effects of glucocorticoids on bone formation is emphasized by the demonstration, in mice with sclerostin deficiency, that bone integrity is maintained in the presence of glucocorticoid excess [24].

Glucocorticoids also suppress bone morphogenetic protein (BMP) signalling in osteoblasts either directly [25] or indirectly by microRNAs which are small endogenous single-stranded non-coding ribonucleic acids (RNAs) that repress gene expression and bind to mRNA [26]. BMPs are members of the transforming growth factor β (TGF-β) superfamily of polypeptides, which includes TGF-βs, activins, and inhibins [27]. BMPs signal through the small molecules against decapentaplegic (SMADs) and the mitogen activated protein kinase signalling pathways [28].

Notch is a family of transmembrane receptors regulating cell fate decisions. In immature cells of the osteoblastic lineage, Notch suppresses cell differentiation by inhibiting Wnt signalling and by interacting with Runx2 to prevent osteoblast maturation [29]. Although glucocorticoids stimulate Notch1 and Notch2 expression in osteoblasts, Notch signalling is not induced by glucocorticoids [30].

In addition to inhibiting the differentiation of osteoblasts, glucocorticoids inhibit the function of the differentiated mature cells directly, and indirectly by suppressing the expression of insulin-like growth factor 1 (IGF1) by skeletal cells [31]. Glucocorticoids inhibit osteoblast-driven synthesis of type I collagen, the major component of the bone extracellular matrix, with a consequent decrease in bone matrix available for mineralization. The decrease in type I collagen synthesis occurs by transcriptional and post-transcriptional mechanisms [20].

Glucocorticoids have pro-apoptotic effects on osteoblasts and osteocytes due to activation of caspase 3, a common downstream effector of several apoptotic signalling pathways. Caspases are synthesized as proenzymes and are activated through autocatalysis or a caspase cascade. Active caspases contribute to apoptosis by cleaving target cellular proteins. Caspase 3 is a key mediator of apoptosis and is a common downstream effector of multiple apoptotic signalling pathways. The inhibitory effects of glucocorticoids on osteoblastic cell replication and differentiation and the increased apoptosis of mature osteoblasts, all contribute to the depletion of the osteoblastic cellular pool and decreased bone formation [20].

Osteocytes serve as mechanosensors, and play a role in the repair of bone microdamage [32]. Loss of osteocytes disrupts the osteocyte-canalicular network resulting in failure to detect signals that normally stimulate processes associated with the replacement of damaged bone. Disruption of the osteocyte-canalicular network can disrupt fluid flow within the network adversely affecting the material properties of the surrounding bone, independently of changes in bone remodelling or architecture. Glucocorticoids affect the function of osteocytes, by modifying the elastic modulus surrounding osteocytic lacunae. Glucocorticoids induce the apoptosis of osteocytes. As a result, the normal maintenance of bone through

this mechanism is impaired and the biomechanical properties of bone are compromised [20].

The initial bone loss occurring in patients exposed to glucocorticoids may be secondary to increased bone resorption. In fact, glucocorticoids increase the expression of receptor activator of NF-k B ligand (RANKL) and decrease the expression of its soluble decoy receptor, OPG in stromal and osteoblastic cells [20]. The combination of an increase in RANKL, a necessary signal for osteoclastogenesis, and a reduction in OPG, an inhibitor of RANKL action, leads to the initial phase of rapid bone loss. Glucocorticoids also enhance the expression of macrophage colony-stimulating factor (M-CSF), which in the presence of RANKL induces osteoclastogenesis. Moreover, glucocorticoids upregulate receptor subunits for osteoclastogenic cytokines of the gp130 family. Furthermore, glucocorticoids may decrease apoptosis of mature osteoclasts. Consequently, there is increased formation of osteoclasts with a prolonged life span explaining, at the cellular level, the enhanced bone resorption observed in the initial phases of GIO.

Indirect Effects of Glucocorticoids on Bone Metabolism

Calcium-phosphate homeostasis is usually impaired in patients exposed to glucocorticoid excess. Glucocorticoids impair the complex process of transcellular calcium absorption in the duodenum which is regulated by the epithelial Receptor Potential–Vanilloid 6 (TRPV6) and TRPV5 (D ECaC1) and then by calcium binding calbindin-D9K [33]. Moreover, glucocorticoids inhibit renal tubular calcium reabsorption favouring calcium renal loss, leading to negative calcium balance. In addition to these effects on calcium metabolism, glucocorticoids elicit phosphaturia by reducing tubular reabsorption of phosphate through antagonism of the tubular expression of the sodium gradient-dependent phosphate transporter [34]. The abnormalities of calcium-phosphate homeostasis in GIO may be favoured by the effects of glucocorticoids on vitamin D metabolism and activity. Glucocorticoids do not directly compete with 1,25-dihydroxyvitamin D binding to its receptor, but rather regulate vitamin D receptor expression [35]. Prolonged exposure to glucocorticoids was associated with diminished expression of duodenal vitamin D receptor which in turn causes a` decrease in transcription of duodenal calbindin-D9K, TRPV5, or TRPV6 activity [36]. Glucocorticoids also modulate vitamin D metabolism with a decrease of 1α-hydroxylase and an increase in 24-hydroxylase activity [37]. Moreover, glucocorticoids inhibit hydroxylation of vitamin D3 in the liver which can lead to reduced plasma 25-hydroxyvitamin D levels [38].

Parathyroid hormone (PTH) and the PTH receptor play a major role in regulating calcium homeostasis. PTH regulates the conversion of 25-hydroxyvitamin D to the active metabolite 1,25-dihydroxyvitamin D. PTH activates dihydropyridine-sensitive channels that mediate calcium entry into the cell. 1,25-dihydroxyvitamin D inhibits PTH secretion and parathyroid cell proliferation via the vitamin D receptor. PTH secretion is regulated by serum ionized calcium levels. While negative calcium balance is expected to cause a secondary hyperparathyroidism in GIO, glucocorticoids do not induce hypertrophy of parathyroid glands in experimental animals [36]. Moreover, fasting PTH values are not consistently reported to be increased in patients exposed to glucocorticoid excess [39]. Perhaps more revealing than static measurements of PTH after glucocorticoid exposure are studies related to the pulsatility of PTH [39]. Chronic glucocorticoid exposure is associated with a redistribution of spontaneous PTH secretory dynamics by reducing tonic and amplifying pulsatile secretion of PTH [40]. This finding may be clinically relevant since it can represent either a compensatory or a pathophysiological mechanism [39].

Glucocorticoids modulate growth hormone secretion (GH) by various and competing effects on the hypothalamus and pituitary gland, all dependent on hormone concentrations and time of exposure [41]. Specifically, under physiological conditions, glucocorticoids stimulate GH secretion either by direct effects on pituitary cells or by increasing their GH secretory response to GH-releasing hormone [42]. However, when glucocorticoid levels exceed the physiological range, an increase in hypothalamic somatostatin tone may occur with a consequent impairment of GH secretion [41]. This inhibitory effect on pituitary GH secretion was observed even when glucocorticoid excess was mild, as in patients treated with inhaled corticosteroids and in those with subclinical endogenous hypercortisolism [2]. Moreover, glucocorticoid excess may suppress the peripheral expression of GH receptors impairing the GH-mediated synthesis of IGF-1 and thus amplifying the impact of a functional GH deficiency (GHD) on target tissues [2]. On the other hand, there is a cross-talk between glucocorticoids and the GH-IGF-I axis, since peripheral metabolism of glucocorticoids by 11-βHSD is modulated by GH and activation of cortisone to cortisol in target tissues is amplified by GHD [13]. The real impact of GHD on fracture risk in GIO is largely unknown. However, since GH is an osteoanabolic hormone, one could argue that GHD may contribute to the impairment in bone quality in patients exposed to glucocorticoid excess [41]. As a matter of fact, both exposure to glucocorticoid excess and GHD are associated with 'low-turnover osteoporosis' and high risk of vertebral fractures [13].

Glucocorticoids inhibit the release of gonadotropins, and as a result oestrogen and testosterone production [20]. Central hypogonadism may contribute to bone loss and fractures in GIO not only via negative effects on bone mass but also on muscle function. In particular, hypogonadism may enhance the glucocorticoid-induced sarcopenia, thereby increasing the likelihood of falls and consequent fractures.

Effects on Muscle

In addition to the direct effects of glucocorticoids on bone cells, the catabolic effects of glucocorticoids on muscle may contribute to fracture risk since these steroids cause muscle weakness, which can increase the incidence of falls [20]. Glucocorticoid-induced myopathy may occur following early or chronic exposure to glucocorticoids. Chronic glucocorticoid-induced myopathy is generally manifested by weakness particularly of the pelvic girdle muscles and may affect up to 60% of patients treated with glucocorticoids. The myopathy involves muscle loss due to glucocorticoid-induced proteolysis of myofibrils. This is mediated by activation of negative regulators of muscle mass, such as myostatin [43] and NF-κB-inducing kinase [44].

Effects of the Underlying Chronic Disease

An important point that should be considered is that many disorders for which the glucocorticoids are prescribed are themselves a cause of osteoporosis. One has to take into account, therefore, the underlying disease itself along with the use of glucocorticoids when

considering the management of GIO. Inflammatory bowel disease, rheumatoid arthritis, and chronic obstructive pulmonary disease (COPD), for example, all are associated with bone loss, independent of glucocorticoid treatment [20]. The systemic release of inflammatory cytokines, which affect bone formation and bone resorption seems to underlie the pathophysiology of the bone loss in these settings [20]. However, there are additional factors that may play a role in the bone loss. In inflammatory bowel disease, bone loss may be due, in part, to malabsorption of vitamin D, calcium and other nutrients [20]. In COPD, hypoxia, acidosis, reduced physical activity, and smoking may all contribute to bone loss, independent of the use of glucocorticoids [20].

Differential Skeletal Susceptibility to Glucocorticoids

Individual susceptibility to glucocorticoids varies considerably, possibly because of differences in the absorption, distribution, or metabolism of the steroid, or because of differences in the number and affinity of glucocorticoid receptors. Polymorphisms of the glucocorticoid receptor gene are associated with differences in BMD and body composition [45]. An attractive explanation to account for interindividual variability among those exposed to glucocorticoids is related to peripheral enzymes that interconvert active and inactive glucocorticoid molecules [46]. 11-βHSD regulate the interconversion of the inactive hormone cortisone and hormonally active cortisol. This enzyme plays a critical role in the regulation of glucocorticoid activity [46]. Two distinct 11-βHSD enzymes have been described in humans. 11-βHSD type 2 is expressed in tissues with high levels of glucocorticoid receptors, such as liver and adipose tissue, and acts as an inactivating enzyme by converting cortisol to cortisone. This enzyme was identified also in rat and human osteosarcoma cells where glucocorticoid inactivation by this mechanism was demonstrated. In contrast, 11-βHSD type 1 is primarily a glucocorticoid activator, converting cortisone to cortisol. This enzyme is widely expressed in target tissues of glucocorticoid action, including bone. The activity of 11-βHSD type-1 and its potential to generate cortisol from cortisone in human osteoblasts is increased by pro-inflammatory cytokines and by glucocorticoids [46]. These effects of glucocorticoids appear to be mediated by the C/EBP family of transcription factors. An inverse relationship between 11-βHSD type 1 activity and osteoblast differentiation appears to exist, although mice with targeted deletions of 11-βHSD type-1 do not develop a skeletal phenotype [45]. An increase of 11-βHSD type 1 activity occurs with ageing, possibly providing an explanation for the enhanced glucocorticoid effects in the skeleton of elderly subjects [45].

Clinical Manifestations

Despite the recognition that glucocorticoids can cause bone loss and fractures, many patients receiving, or being considered for, long-term glucocorticoid therapy are not evaluated for their skeletal health [2].

Fractures occur more frequently at sites enriched in cancellous bone, such as the vertebrae and femoral neck. As with vertebral fractures occurring in postmenopausal osteoporosis, vertebral fractures associated with glucocorticoid therapy or endogenous hypercortisolism often are asymptomatic, in which case a radiological approach with morphometric analysis is often necessary [5], as is the case with many causes of secondary osteoporosis [13]. Vertebral fractures occur early after exposure to glucocorticoids, at a time when BMD declines rapidly. However, a direct relationship between BMD and fracture risk in GIO has not been established. It is likely to be different from that established in postmenopausal osteoporosis because fractures in GIO occur at higher BMD values [47]. This point has to be considered when making treatment decisions in GIO, since therapeutic intervention should be considered at T-scores that may be in the osteopaenic range [48].

Trabecular bone score (TBS) provides an indirect index of trabecular bone architecture that can be obtained from bone densitometry (DXA) images of the lumbar spine, and has predictive value for fracture independent of BMD [49]. In postmenopausal women exposed to glucocorticoid excess, TBS was significantly lower as compared to control subjects, notwithstanding comparable BMD T-scores [50]. Interestingly, decreases in TBS are substantial in individuals with fractures [51], supporting the concept that TBS values may be useful for predicting fractures in this clinical setting [1]. A more direct measurement of bone architecture and strength can be provided by high-resolution peripheral computed tomography, but data in subjects with exogenous and endogenous GIO using this approach are sparse [52, 53].

Fracture Risk Assessment (FRAX), an algorithm that calculates the 10-year probability of major osteoporotic and hip fractures, can be used to estimate fracture risk in GIO. However, its value was initially limited by the fact that the use of glucocorticoids is entered as a dichotomous risk factor without considering the dose or length of exposure to glucocorticoids. Ways to correct for this limitation have been proposed [54], but this modification has not been rigorously tested. Another confounder to the utility of FRAX in the context of glucocorticoid exposure is the use of the femoral neck BMD which may result in underestimation of fracture risk in patients with differentially low spine BMD. A correction factor for this discordance when there is a greater than a 1 T-score difference between lumbar spine and femoral neck BMD has been proposed [55]. Assessment of fracture risk using FRAX is recommended in several guidelines for the management of GIO, including those from the National Osteoporosis Guideline Group (NOGG) [56], the updated recommendations by the American College of Rheumatology (ACR) [57] and those published by the International Osteoporosis Foundation (IOF) and the European Calcified Tissue Society (ECTS) [58]. However, FRAX can be used only in individuals at age 40 and older. In children and young adults, fracture risk assessment should be performed using BMD measurement, together with consideration of other risk factors, particularly previous fractures [1]. It should also be recognized that in children and young adults, their young age is a powerful protection against what otherwise in older subjects would be substantial fracture risk. Based on radiological/historical assessment of prevalent fractures, glucocorticoid dose, age of patients, BMD values, and FRAX calculations, the guidelines proposed a stratification of fracture risk in GIO for identifying patients for whom treatment with bone-active drugs is cost-effective and indicated (Table 4.12.1) [56–58]. It is unclear whether these algorithms may simplify the approach to the prevention and treatment of GIO in real-life clinical practice.

The guidelines recommend evaluation of calcium metabolism in all subjects in order to select those for whom vitamin D and/or

Table 4.12.1 Criteria proposed for starting treatment with bone-active drugs in patients exposed to glucocorticoid treatment

Criteria*	ACR [57]	IOF/ ECTS [58]	NOGG [56]
Clinical fractures/morphometric vertebral fractures	YES At any age	YES At any age	YES At any age
Age of patients	NS	≥70 yrs	≥70 yrs
Glucocorticoid dose**	≥30 mg/day (≥30 yrs)	≥7.5 mg/day (≥50 yrs)	≥7.5 mg/day (≥50 yrs)
BMD	T-score ≤–2.5 SD (≥50 yrs) Z score ≤–3.0 SD (≤40 yrs) ≥10%/year loss (≤40 yrs)	T-score ≤–1.5 SD (≥50 yrs)	NS*
FRAX	10-year risk of major fracture ≥10% (≥40 yrs) 10-year risk of hip fracture ≥1% (≥40 yrs)	10-year risk of major fracture ≥ threshold*** (≥50 yrs)	10-year risk of major fracture ≥ threshold**** (≥50 yrs)

ACR, American College of Rheumatology; BMD, bone mineral density; ECTS, European Calcified Tissue Society; FRAX, Fracture Risk Assessment; IOF, International Osteoporosis Foundation; NOGG, National Osteoporosis Guideline Group; NS, not specified; SD, standard deviation; yrs, years. *, one criterion is sufficient for starting treatment; **, expressed as daily dose of prednisolone; ***, variable according to the country and age of patients; ****, variable according to age of patients and dose of glucocorticoids.

calcium supplementation are indicated [58]. The role of biochemical markers of bone turnover in the diagnostic work-up in GIO has not been established, and their levels vary and are dependent on the different stages of the disease. Following the initial exposure to glucocorticoids, there is an increase in biochemical markers of bone resorption, which is followed by a prolonged suppression of markers of bone formation and bone resorption.

The assessment of gonadal function may be useful for the subsequent treatment of GIO. In men receiving glucocorticoids, low total and free-testosterone concentrations are frequently found. Combined with low or normal serum gonadotropin levels, such low testosterone levels are likely to be as manifestation of secondary hypogonadism [45].

Therapy

Various guidelines advocate the following measures for the prevention and treatment of GIO: general health awareness, administration of sufficient calcium and vitamin D, use of the minimal effective dose of corticosteroids and, when indicated, therapeutic intervention with bisphosphonates and other agents [57, 58]. A cost-effectiveness analysis demonstrated that treatment of GIO is cost-effective in patients with a prior fracture, in individuals ≥ 75 years of age or in younger subjects with T-scores ≤–2.0 [59]. Recent guidelines emphasize the importance of starting bone protective therapy early in high-risk individuals, because of the rapidity with which bone loss and increased fracture risk occur [56–58]. It is unknown whether such recommendations may be adapted to patients with endogenous GIO, since it is difficult to translate corticosteroid dosages to different degrees of endogenous hypercortisolism and there are no data on validation of the FRAX stratification method in patients with endogenous hypercortisolism [60].

Despite several guidelines and consensus recommendations stating that patients are at increased fracture risk, little attention is being paid to this risk and less than 50% of patients meeting guideline criteria for treatment are indeed being treated [61, 62].

Various pharmacological agents have been assessed for the prevention and treatment of GIO. In most studies, the primary endpoint was a change in BMD; fracture outcomes were measured in selected studies as secondary endpoints.

Calcium and Vitamin D

The negative calcium balance induced by glucocorticoids provides a rationale for maintaining a daily calcium intake of about 1200–1500 mg in patients with GIO [58]; the upper limit of daily calcium intake set by the Institute of Medicine is 2500 mg [63]. However, calcium alone does not prevent glucocorticoid-induced bone loss, even when administered at high doses [64]. Vitamin D is a key adjunct to the therapeutic use of calcium in GIO, particularly with regard to monitoring by BMD [65]. Vitamin D increases the intestinal absorption of calcium and reabsorption of calcium in the distal renal tubules leading to increased bone mineralization. Because glucocorticoids inhibit the 25 hydroxylation of vitamin D, the use of calcitriol or alfacalcidol has been proposed in place of vitamin D in individuals with GIO but the results have been variable [66, 67]. Since the potential advantages of using active analogues of vitamin D are unclear, native vitamin D_3 is considered the gold standard for treatment of hypovitaminosis D in patients exposed to glucocorticoid excess [68]. In patients with higher body mass index and in those receiving higher doses of glucocorticoids, calcidiol was shown to be more effective than cholecalciferol to reach adequate 25-hydroxyvitamin D levels most probably bypassing the inhibition of 25 hydroxylase of chronic corticosteroid therapy [69]. Moreover, some trials have demonstrated an advantage of calcidiol in improving BMD in transplant patients [70].

It is important to note that subjects receiving glucocorticoids may display vitamin D resistance. Consequently, patients often require up to 1000–2000 IU of vitamin D_3 daily, to maintain supraoptimal 25 hydroxyvitamin D_3 levels of ≥40 ng/ml or 110 nmol/L [71].

Hormonal Replacement Therapy

GIO can be associated with suppressed gonadal function in men and women since glucocorticoids inhibit gonadotropin release and as a consequence oestrogen and androgen synthesis. In men with GIO, testosterone administered intramuscularly induced a significant increase in lumbar spine BMD, without significant effects on BMD at the femoral neck. The studies had BMD as primary endpoint, whereas no information on bone fractures is available. Testosterone also was shown to improve muscular performance and quality of life in men with GIO [72]. Therefore, substitution treatment of hypogonadism in men and in women may prove beneficial in the management of GIO. However, potential benefits should be weighed against the risks of side effects.

Bisphosphonates

Bisphosphonates are currently the major class of drugs used for the treatment of osteoporosis, including GIO [73] and current guidelines recommend these drugs as first-line therapy in patients receiving glucocorticoids [56–58]. There are no specific guidelines for the management of skeletal fragility in individuals with endogenous hypercortisolism [13] since evidence of efficacy and safety of bone-active drugs in this specific clinical setting is insufficient based on results of few clinical studies evaluating the effects of bisphosphonates in patients with Cushing's syndrome [13, 60]. Obviously, correction of the endogenous glucocorticoid excess is a key therapeutic approach, if possible. Bisphosphonates are stable analogues of naturally occurring inorganic pyrophosphate and have been shown to be effective in postmenopausal osteoporosis and in osteoporotic men.

The evidence of benefit for bisphosphonate therapy is weaker than for postmenopausal osteoporosis, since fracture reduction has not been a primary end point of any study. This reflects the acceptance by regulatory authorities of bridging studies, using BMD, for agents proposed for GIO that reduce fractures in postmenopausal osteoporosis [74]. Oral alendronate (5 or 10 mg daily or 70 mg once weekly) and risedronate (5 mg daily or 35 mg once weekly), and intravenous zoledronic acid (5 mg once yearly by intravenous infusion) are approved for treatment of GIO. A meta-analysis performed on nine studies including 1134 patients revealed that alendronate increased BMD at the lumbar spine (mean difference = 3.66%), total hip (mean difference = 2.08%) and trochanter (mean difference = 1.68%), however alendronate had no effect on vertebral and non-vertebral fractures due to lack of statistical power [75]. Indeed, in some studies, bisphosphonates induced a significant decrease of vertebral fracture risk [76–78]. This effect was confirmed by an observational cohort study of 11 007 women aged >65 years treated with glucocorticoids and receiving alendronate or risedronate [79]. The incidence of fractures was significantly decreased in both alendronate (33% lower at non-vertebral sites and 59% at vertebral sites) and risedronate (28% lower at non-vertebral sites and 54% at vertebral sites) cohorts. Moreover, in three matched cohorts derived from healthcare administrative data from Ontario, Canada, Amiche *et al.*, reported that in individuals initiating long-term glucocorticoid therapy treated during the first six months with alendronate or risedronate there was a decrease in incident hip fractures (hazard ratios 0.49 and 0.58, for alendronate and risedronate, respectively) [80].

Zoledronic acid was shown to have a greater effect on BMD than risedronate in both prevention and treatment trials [81] but the fracture rate was too limited to assess antifracture efficacy.

The use of bisphosphonates in eugonadal premenopausal women has to be considered carefully, since bisphosphonates may cross the placenta and affect embryonic skeletal development. The guidelines recommend to use bisphosphonates in eugonadal premenopausal women with a history of fragility fractures, or in those continuing glucocorticoid treatment for more than 6 months at a dose of higher than 7.5 mg/day who have either a hip or spine BMD Z score <–3 SD or bone loss of at least 10% yearly at the hip or spine as assessed by dual X-ray absorptiometry [57].

Denosumab

Denosumab is a human monoclonal antibody that binds and neutralizes the activity of human RANKL similar to the action of endogenous OPG. In postmenopausal osteoporosis, denosumab increases BMD and reduces vertebral, non-vertebral, and hip fracture risk [82]. In a recent phase 3, randomized, double-blind, double-dummy, active-controlled study, denosumab was both non-inferior and superior to risedronate in increasing BMD at lumbar spine (+2.2% vs. risedronate after 12 months) and total hip (+1.5% vs. risedronate after 12 months) with a comparable safety profile, in 505 glucocorticoid-continuing and 290 glucocorticoid-initiating patients at moderate to high risk of fractures [83]. The increase in lumbar spine BMD induced by denosumab was of similar magnitude as the effect observed with alendronate or zoledronic acid in former studies on GIO [81, 84]. Denosumab is approved by the Food and Drug Administration and the European Medicines Agency for use in GIO.

Osteoanabolic Therapy

PTH is an attractive candidate for the treatment of GIO because it protects against osteoblast apoptosis and increases osteoblast cell number and activity [85]. PTH rapidly stimulates osteoblastic function, inducing an upregulation of osteoblast-derived RANKL, which eventually leads to osteoclast differentiation and bone resorption [21].

The efficacy of teriparatide was demonstrated in a randomized, double-blind, controlled study directly comparing teriparatide at 20 μg/day with alendronate at 10 mg/day for 18 months [84]. The study enrolled men and women with GIO whose average T-score was –2.5 SD, almost half of whom had evidence for non-vertebral fractures. Teriparatide induced a greater increase in BMD than alendronate at the lumbar spine, hip, and femoral neck. While not a prespecified endpoint, there was a significantly greater reduction in morphometric vertebral fractures in subjects on teriparatide than in those treated with alendronate. This statistically significant difference was present after 18 and 36 months of intervention [86]. Clinically evident vertebral fractures were statistically lower after 36 months in subjects receiving teriparatide than those receiving alendronate. The results were independent of the amount of glucocorticoid exposure [87]. Subgroup analyses showed BMD changes favouring teriparatide among men and premenopausal women in the same manner as postmenopausal women [88]. Similarly, teriparatide caused larger increments in spinal BMD as compared to risedronate along with greater effects on mechanical parameters of bone strength, as assessed by high-resolution peripheral quantitative computed tomography [89]. In addition to the favourable effects on bone structure and strength, an improvement of glucose homeostasis was demonstrated in patients with diabetes and GIO undergoing treatment with teriparatide [90].

Unresolved issues with the use of teriparatide and other forms of PTH in GIO are its use in the prevention of GIO, its use in patients who are resistant to bisphosphonate therapy, its use in younger populations with open epiphyses (e.g. theoretical risk of osteosarcoma), and whether teriparatide should be followed by antiresorptives in GIO, as it is recommended in postmenopausal osteoporosis. Moreover, there are no clinical studies evaluating the effects and safety of teriparatide in endogenous hypercortisolism,

while only anecdotal data showed favourable effects of this drug on BMD and TBS after cure of Cushing's syndrome [13].

GH or IGF-I administration could revert some of the negative effects of chronic glucocorticoids on the skeleton [31]. However, glucocorticoids decrease the activity of GH on skeletal cells and there are no controlled trials to determine the effectiveness of either GH or IGF-I as treatments for GIO. Increases in serum osteocalcin, carboxy-terminal propeptide of type I procollagen, and carboxy-terminal telopeptide of type I collagen are observed following short-term use of recombinant human GH treatment in a selected population of patients receiving chronic corticosteroid treatment for non-endocrine diseases [91]. Combined therapy of GH and IGF-I counteracts selected negative effects of glucocorticoids on bone in healthy volunteers, receiving short-term glucocorticoid therapy [41]. Observational and controlled studies in children receiving glucocorticoid therapy for juvenile idiopathic arthritis demonstrated that GH restored normal height velocity with a concomitant enhancement of bone mineralization [41]. However, the efficacy and safety of GH and IGF-I treatment in GIO is unknown and well-designed prospective controlled studies are necessary before their use can be recommended.

REFERENCES

1. Compston J. Glucocorticoid-induced osteoporosis: an update. *Endocrine*, 2018; **61**: 7–16.
2. Mazziotti G, Formenti AM, Adler RA, *et al*. Glucocorticoid-induced osteoporosis: pathophysiological role of GH/IGF-I and PTH/VITAMIN D axes, treatment options and guidelines. *Endocrine*, 2016; **54**: 603–11.
3. Overman RA, Yeh JY, Deal CL. Prevalence of oral glucocorticoid usage in the United States: a general population perspective. *Arthritis Care Res*, 2013; **65**: 294–8.
4. Diez-Perez A, Hooven FH, Adachi JD, *et al*. Regional differences in treatment for osteoporosis. The global longitudinal study of osteoporosis in women (GLOW). *Bone*, 2011; **49**: 493–8.
5. Angeli A, Guglielmi G, Dovio A, *et al*. High prevalence of asymptomatic vertebral fractures in post-menopausal women receiving chronic glucocorticoid therapy: a cross-sectional outpatient study. *Bone*, 2006; **39**: 253–9.
6. van Staa TP, Leufkens HG, Cooper C. The epidemiology of corticosteroid-induced osteoporosis: a meta-analysis. *Osteoporosis Int*, 2002; **13**: 777–87.
7. Amiche MA, Albaum JM, Tadrous M, *et al*. Fracture risk in oral glucocorticoid users: a Bayesian meta-regression leveraging control arms of osteoporosis clinical trials. *Osteoporosis Int*, 2016; **27**: 1709–18.
8. De Vries F, Bracke M, Leufkens HG, Lammers JW, Cooper C, Van Staa TP. Fracture risk with intermittent high-dose oral glucocorticoid therapy. *Arthritis Rheum*, 2007; **56**: 208–14.
9. Vestergaard P, Rejnmark L, Mosekilde L. Fracture risk associated with systemic and topical corticosteroids. *J Int Med*, 2005; **257**: 374–84.
10. Gonnelli S, Caffarelli C, Maggi S, *et al*. Effect of inhaled glucocorticoids and beta(2) agonists on vertebral fracture risk in COPD patients: the EOLO study. *Calcified Tissue Int*, 2010; **87**: 137–43.
11. Mazziotti G, Delgado A, Maffezzoni F, Formenti A, Giustina A. Skeletal fragility in endogenous hypercortisolism. *Front Hormone Res*, 2016; **46**: 66–73.
12. Chiodini I, Mascia ML, Muscarella S, *et al*. Subclinical hypercortisolism among outpatients referred for osteoporosis. *Ann Int Med*, 2007; **147**: 541–8.
13. Mazziotti G, Frara S, Giustina A. Pituitary diseases and bone. *Endocr Rev*, 2018; **39**: 440–88.
14. Tauchmanova L, Pivonello R, Di Somma C, *et al*. Bone demineralization and vertebral fractures in endogenous cortisol excess: role of disease etiology and gonadal status. *J Clin Endocrinol Metab*, 2006; **91**: 1779–84.
15. Belaya ZE, Hans D, Rozhinskaya LY, *et al*. The risk factors for fractures and trabecular bone-score value in patients with endogenous Cushing's syndrome. *Arch Osteoporos*, 2015; **10**: 44.
16. Vestergaard P, Lindholm J, Jorgensen JO, *et al*. Increased risk of osteoporotic fractures in patients with Cushing's syndrome. *Eur J Endocrinol*, 2002; **146**: 51–6.
17. Morelli V, Eller-Vainicher C, Salcuni AS, *et al*. Risk of new vertebral fractures in patients with adrenal incidentaloma with and without subclinical hypercortisolism: a multicenter longitudinal study. *J Bone Miner Res*, **26**: 1816–21.
18. Zhou H, Mak W, Zheng Y, Dunstan CR, Seibel MJ. Osteoblasts directly control lineage commitment of mesenchymal progenitor cells through Wnt signaling. *J Biol Chem*, 2008; **283**: 1936–45.
19. Sher LB, Woitge HW, Adams DJ, *et al*. Transgenic expression of 11beta-hydroxysteroid dehydrogenase type 2 in osteoblasts reveals an anabolic role for endogenous glucocorticoids in bone. *Endocrinology*, 2004; **145**: 922–9.
20. Canalis E, Mazziotti G, Giustina A, Bilezikian JP. Glucocorticoid-induced osteoporosis: pathophysiology and therapy. *Osteoporos Int*, 2007; **18**: 1319–28.
21. Canalis E, Giustina A, Bilezikian JP. Mechanisms of anabolic therapies for osteoporosis. *N Engl J Med*, 2007; **357**: 905–16.
22. Moverare-Skrtic S, Henning P, Liu X, *et al*. Osteoblast-derived WNT16 represses osteoclastogenesis and prevents cortical bone fragility fractures. *Nat Med*, 2014; **20**: 1279–88.
23. Wei W, Zeve D, Suh JM, *et al*. Biphasic and dosage-dependent regulation of osteoclastogenesis by beta-catenin. *Mol Cell Biol*, 2011; **31**: 4706–19.
24. Sato AY, Cregor M, Delgado-Calle J, *et al*. Protection from glucocorticoid-induced osteoporosis by anti-catabolic signaling in the absence of Sost/sclerostin. *J Bone Miner Res*, 2016; **31**: 1791–802.
25. Liu J, Lu C, Wu X, *et al*. Targeting osteoblastic casein kinase-2 interacting protein-1 to enhance Smad-dependent BMP signaling and reverse bone formation reduction in glucocorticoid-induced osteoporosis. *Sci Rep*, 2017; **7**: 41295.
26. Liu K, Jing Y, Zhang W, *et al*. Silencing miR-106b accelerates osteogenesis of mesenchymal stem cells and rescues against glucocorticoid-induced osteoporosis by targeting BMP2. *Bone*, 2017; **97**: 130–8.
27. Canalis E, Economides AN, Gazzerro E. Bone morphogenetic proteins, their antagonists, and the skeleton. *Endocr Rev*, 2003; **24**: 218–35.
28. Nohe A, Hassel S, Ehrlich M, *et al*. The mode of bone morphogenetic protein (BMP) receptor oligomerization determines different BMP-2 signaling pathways. *J Biol Chem*, 2002; **277**: 5330–8.
29. Zanotti S, Canalis E. Notch signaling and the skeleton. *Endocr Rev*, 2016; **37**: 223–53.
30. Zanotti S, Yu J, Adhikari S, Canalis E. Glucocorticoids inhibit notch target gene expression in osteoblasts. *J Cell Biochem*, 2018; **119**: 6016–23.

31. Giustina A, Mazziotti G, Canalis E. Growth hormone, insulin-like growth factors, and the skeleton. *Endocr Rev*, 2008; **29**: 535–59.
32. Verborgt O, Gibson GJ, Schaffler MB. Loss of osteocyte integrity in association with microdamage and bone remodeling after fatigue *in vivo*. *J Bone Miner Res*, 2000; **15**: 60–7.
33. Huybers S, Naber TH, Bindels RJ, Hoenderop JG. Prednisolone-induced Ca2+ malabsorption is caused by diminished expression of the epithelial Ca2+ channel TRPV6. *Am J Physiol Gastro Liver Physiol*, 2007; **292**: G92–7.
34. Levi M, Shayman JA, Abe A, *et al*. Dexamethasone modulates rat renal brush border membrane phosphate transporter mRNA and protein abundance and glycosphingolipid composition. *J Clin Invest*, 1995; **96**: 207–16.
35. Kim MH, Lee GS, Jung EM, Choi KC, Oh GT, Jeung EB. Dexamethasone differentially regulates renal and duodenal calcium-processing genes in calbindin-D9k and -D28k knockout mice. *Exp Physiol*, 2009; **94**: 138–51.
36. Kim MH, Lee GS, Jung EM, Choi KC, Jeung EB. The negative effect of dexamethasone on calcium-processing gene expressions is associated with a glucocorticoid-induced calcium-absorbing disorder. *Life Sci*, 2009; **85**: 146–52.
37. Akeno N, Matsunuma A, Maeda T, Kawane T, Horiuchi N. Regulation of vitamin D-1alpha-hydroxylase and -24-hydroxylase expression by dexamethasone in mouse kidney. *J Endocrinol*, 2000; **164**: 339–48.
38. Toloza SM, Cole DE, Gladman DD, Ibanez D, Urowitz MB. Vitamin D insufficiency in a large female SLE cohort. *Lupus*, 2010; **19**: 13–19.
39. Chiavistelli S, Giustina A, Mazziotti G. Parathyroid hormone pulsatility: physiological and clinical aspects. *Bone Res*, 2015; **3**: 14049.
40. Bonadonna S, Burattin A, Nuzzo M, *et al*. Chronic glucocorticoid treatment alters spontaneous pulsatile parathyroid hormone secretory dynamics in human subjects. *Eur J Endocrinol*, 2005; **152**: 199–205.
41. Mazziotti G, Giustina A. Glucocorticoids and the regulation of growth hormone secretion. *Nat Rev Endocrinol*, 2013; **9**: 265–76.
42. Giustina A, Veldhuis JD. Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. *Endocr Rev*, 1998; **19**: 717–97.
43. Wang R, Jiao H, Zhao J, Wang X, Lin H. Glucocorticoids enhance muscle proteolysis through a myostatin-dependent pathway at the early stage. *PloS One*, 2016; **11**: e0156225.
44. Fry CS, Nayeem SZ, Dillon EL, *et al*. Glucocorticoids increase skeletal muscle NF-kappaB inducing kinase (NIK): links to muscle atrophy. *Physiol Rep*, 2016; 4: e13014.
45. Mazziotti G, Angeli A, Bilezikian JP, Canalis E, Giustina A. Glucocorticoid-induced osteoporosis: an update. *Trends Endocrinol Metab*, 2006; **17**: 144–9.
46. Tomlinson JW, Walker EA, Bujalska IJ, *et al*. 11beta-hydroxysteroid dehydrogenase type 1: a tissue-specific regulator of glucocorticoid response. *Endocr Rev*, 2004; **25**: 831–66.
47. Van Staa TP, Laan RF, Barton IP, Cohen S, Reid DM, Cooper C. Bone density threshold and other predictors of vertebral fracture in patients receiving oral glucocorticoid therapy. *Arthritis Rheum*, 2003; **48**: 3224–9.
48. Mirza F, Canalis E. Management of endocrine disease: secondary osteoporosis: pathophysiology and management. *Eur J Endocrinol*, 2015; **173**: R131–51.
49. Ulivieri FM, Silva BC, Sardanelli F, Hans D, Bilezikian JP, Caudarella R. Utility of the trabecular bone score (TBS) in secondary osteoporosis. *Endocrine*, 2014; **47**: 435–48.
50. Paggiosi MA, Peel NF, Eastell R. The impact of glucocorticoid therapy on trabecular bone score in older women. *Osteoporos Int*, 2015; **26**: 1773–80.
51. Saag KG, Agnusdei D, Hans D, *et al*. Trabecular bone score in patients with chronic glucocorticoid therapy-induced osteoporosis treated with alendronate or teriparatide. *Arthritis Rheumatol (Hoboken, NJ)*, 2016; **68**: 2122–8.
52. Sutter S, Nishiyama KK, Kepley A, *et al*. Abnormalities in cortical bone, trabecular plates, and stiffness in postmenopausal women treated with glucocorticoids. *J Clin Endocrinol Metab*, 2014; **99**: 4231–40.
53. dos Santos CV, Vieira Neto L, Madeira M, *et al*. Bone density and microarchitecture in endogenous hypercortisolism. *Clin Endocrinol*, 2015; **83**: 468–74.
54. Kanis JA, Johansson H, Oden A, McCloskey EV. Guidance for the adjustment of FRAX according to the dose of glucocorticoids. *Osteoporos Int*, 2011; **22**: 809–16.
55. Johansson H, Kanis JA, Oden A, *et al*. Impact of femoral neck and lumbar spine BMD discordances on FRAX probabilities in women: a meta-analysis of international cohorts. *Calcified Tissue Int*, 2014; **95**: 428–35.
56. Compston J, Cooper A, Cooper C, *et al*. UK clinical guideline for the prevention and treatment of osteoporosis. *Arch Osteoporos*, 2017; **12**: 43.
57. Buckley L, Guyatt G, Fink HA, *et al*. 2017 American College of Rheumatology Guideline for the Prevention and Treatment of Glucocorticoid-Induced Osteoporosis. *Arthritis Care Res*, 2017; **69**: 1095–110.
58. Lekamwasam S, Adachi JD, Agnusdei D, *et al*. A framework for the development of guidelines for the management of glucocorticoid-induced osteoporosis. *Osteoporos Int*, 2012; **23**: 2257–76.
59. Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M. Glucocorticoid-induced osteoporosis: a systematic review and cost-utility analysis. *Health Technol Assess*, 2007; **11**: iii–iv, ix–xi, 1–231.
60. Scillitani A, Mazziotti G, Di Somma C, *et al*. Treatment of skeletal impairment in patients with endogenous hypercortisolism: when and how? *Osteoporos Int*, 2014; **25**: 441–6.
61. Albaum JM, Levesque LE, Gershon AS, Liu G, Cadarette SM. Glucocorticoid-induced osteoporosis management among seniors, by year, sex, and indication, 1996–2012. *Osteoporos Int*, 2015; **26**: 2845–52.
62. Trijau S, de Lamotte G, Pradel V, *et al*. Osteoporosis prevention among chronic glucocorticoid users: results from a public health insurance database. *RMD Open*, 2016; **2**: e000249.
63. Ross AC, Manson JE, Abrams SA, *et al*. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab*, 2011; **96**: 53–8.
64. Sambrook P, Birmingham J, Kelly P, *et al*. Prevention of corticosteroid osteoporosis. A comparison of calcium, calcitriol, and calcitonin. *N Engl J Med*, 1993; **328**: 1747–52.
65. Homik J, Suarez-Almazor ME, Shea B, Cranney A, Wells G, Tugwell P. Calcium and vitamin D for corticosteroid-induced osteoporosis. *Cochrane Database Syst Rev*, 2000; **2**: CD000952.
66. Ringe JD, Dorst A, Faber H, Schacht E, Rahlfs VW. Superiority of alfacalcidol over plain vitamin D in the treatment of glucocorticoid-induced osteoporosis. *Rheumatol Int*, 2004; **24**: 63–70.
67. de Nijs RN, Jacobs JW, Algra A, Lems WF, Bijlsma JW. Prevention and treatment of glucocorticoid-induced

osteoporosis with active vitamin D3 analogues: a review with meta-analysis of randomized controlled trials including organ transplantation studies. *Osteoporos Int*, 2004; **15**: 589–602.
68. Holick MF, Binkley NC, Bischoff-Ferrari HA, *et al*. Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 1911–30.
69. Cianferotti L, Cricelli C, Kanis JA, *et al*. The clinical use of vitamin D metabolites and their potential developments: a position statement from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO) and the International Osteoporosis Foundation (IOF). *Endocrine*, 2015; **50**: 12–26.
70. Garcia-Delgado I, Prieto S, Gil-Fraguas L, Robles E, Rufilanchas JJ, Hawkins F. Calcitonin, etidronate, and calcidiol treatment in bone loss after cardiac transplantation. *Calcified Tissue Int*, 1997; **60**: 155–9.
71. Heaney RP. The vitamin D requirement in health and disease. *J Steroid Biochem Mol Biol*, 2005; **97**: 13–19.
72. Tracz MJ, Sideras K, Bolona ER, *et al*. Testosterone use in men and its effects on bone health. A systematic review and meta-analysis of randomized placebo-controlled trials. *J Clin Endocrinol Metab*, 2006; **91**: 2011–16.
73. Doga M, Mazziotti G, Bonadonna S, *et al*. Prevention and treatment of glucocorticoid-induced osteoporosis. *J Endocrinol Invest*, 2008; **31**: 53–8.
74. Compston J, Reid DM, Boisdron J, *et al*. Recommendations for the registration of agents for prevention and treatment of glucocorticoid-induced osteoporosis: an update from the Group for the Respect of Ethics and Excellence in Science. *Osteoporos Int*, 2008; **19**: 1247–50.
75. Kan SL, Yuan ZF, Li Y, Ai J, Xu H, Sun JC, Feng SQ. Alendronate prevents glucocorticoid-induced osteoporosis in patients with rheumatic diseases: a meta-analysis. *Medicine*, 2016; **95**: e3990.
76. Adachi JD, Saag KG, Delmas PD, *et al*. Two-year effects of alendronate on bone mineral density and vertebral fracture in patients receiving glucocorticoids: a randomized, double-blind, placebo-controlled extension trial. *Arthritis Rheum*, 2001; **44**: 202–11.
77. Saag KG, Emkey R, Schnitzer TJ, *et al*. Alendronate for the prevention and treatment of glucocorticoid-induced osteoporosis. Glucocorticoid-Induced Osteoporosis Intervention Study Group. *N Engl J Med*, 1998; **339**: 292–9.
78. Wallach S, Cohen S, Reid DM, *et al*. Effects of risedronate treatment on bone density and vertebral fracture in patients on corticosteroid therapy. *Calcif Tissue Int*, 2000; **67**: 277–85.
79. Thomas T, Horlait S, Ringe JD, *et al*. Oral bisphosphonates reduce the risk of clinical fractures in glucocorticoid-induced osteoporosis in clinical practice. *Osteoporos Int*, 2013; **24**: 263–9.
80. Amiche MA, Levesque LE, Gomes T, Adachi JD, Cadarette SM. Effectiveness of oral bisphosphonates in reducing fracture risk among oral glucocorticoid users: three matched cohort analyses. *J Bone Miner Res*, 2018; **33**: 419–29.
81. Reid DM, Devogelaer JP, Saag K, *et al*. Zoledronic acid and risedronate in the prevention and treatment of glucocorticoid-induced osteoporosis (HORIZON): a multicentre, double-blind, double-dummy, randomised controlled trial. *Lancet*, 2009; **373**: 1253–63.
82. Cummings SR, San Martin J, McClung MR, *et al*. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*, 2009; **361**: 756–65.
83. Saag KG, Wagman RB, Geusens P, *et al*. Denosumab versus risedronate in glucocorticoid-induced osteoporosis: a multicentre, randomised, double-blind, active-controlled, double-dummy, non-inferiority study. *Lancet Diabetes Endocrinol*, 2018; **6**: 445–54.
84. Saag KG, Shane E, Boonen S, et al. Teriparatide or alendronate in glucocorticoid-induced osteoporosis. *N Engl J Med*, 2007; **357**: 2028–39.
85. Carpinteri R, Porcelli T, Mejia C, *et al*. Glucocorticoid-induced osteoporosis and parathyroid hormone. *J Endocrinol Invest*, 2010; **33**: 16–21.
86. Saag KG, Zanchetta JR, Devogelaer JP, *et al*. Effects of teriparatide versus alendronate for treating glucocorticoid-induced osteoporosis: thirty-six-month results of a randomized, double-blind, controlled trial. *Arthritis Rheum*, 2009; **60**: 3346–55.
87. Devogelaer JP, Adler RA, Recknor C, *et al*. Baseline glucocorticoid dose and bone mineral density response with teriparatide or alendronate therapy in patients with glucocorticoid-induced osteoporosis. *J Rheumatol*, 2010; **37**: 141–8.
88. Langdahl BL, Marin F, Shane E, *et al*. Teriparatide versus alendronate for treating glucocorticoid-induced osteoporosis: an analysis by gender and menopausal status. *Osteoporos Int*, 2009; **20**: 2095–104.
89. Gluer CC, Marin F, Ringe JD, *et al*. Comparative effects of teriparatide and risedronate in glucocorticoid-induced osteoporosis in men: 18-month results of the EuroGIOPs trial. *J Bone Miner Res*, 2013; **28**: 1355–68.
90. Mazziotti G, Maffezzoni F, Doga M, Hofbauer LC, Adler RA, Giustina A. Outcome of glucose homeostasis in patients with glucocorticoid-induced osteoporosis undergoing treatment with bone active drugs. *Bone*, 2014; **67**: 175–80.
91. Giustina A, Bussi AR, Jacobello C, Wehrenberg WB. Effects of recombinant human growth hormone (GH) on bone and intermediary metabolism in patients receiving chronic glucocorticoid treatment with suppressed endogenous GH response to GH-releasing hormone. *J Clin Endocrinol Metab*, 1995; **80**: 122–9.

SECTION 5

Adrenal Diseases

5.1

Adrenal Imaging

Peter Guest

Introduction

Evaluating the adrenal gland with imaging can be challenging. The adrenal glands may be morphologically within normal limits even in the presence of clear hyperfunction. Hyperplasia and small nodules may coexist. Non-functioning nodules are frequent and need to be differentiated from hyperfunctioning adenomas or malignancy. However, the high-resolution anatomical imaging provided by computed tomography (CT) and magnetic resonance imaging (MRI), together with the functional characterization afforded by radionuclide imaging, allows correlation with clinical and endocrine parameters.

Anatomy

Embryologically, the adrenal cortex derives from coelomic mesoderm and the medulla from neural crest cells. Development is independent of the kidney and adrenal glands will normally be present even in the absence of a kidney. In the newborn, the adrenal glands are large, being one-third of the size of the kidneys. The fetal adrenal zones involute rapidly, however, and in the adult the adrenal glands are small. They are situated in the retroperitoneum immediately above and anteromedial to the upper pole of the kidneys. The right adrenal lies immediately behind the inferior vena cava, alongside the right diaphragmatic crus. The left adrenal lies behind the splenic vein, lateral to the left crus.

The normal adrenal has a characteristic inverted Y- or V-shape, with the two limbs fusing anteromedially. The most cranial section has a triangular appearance. Cross-sectional appearance varies according to the exact level. Each limb measures 2.5–4 cm in length and 3–6 mm in thickness. Accessory adrenal tissue (rests) may be found in the kidney, testis, or ovary, and elsewhere in the retroperitoneum, arising from the fetal urogenital ridge during its separation into adrenal and gonads.

Arterial supply is from three sources: superior–multiple arteries from the inferior phrenic artery; middle—direct from the aorta; and inferior arising from the renal artery. A single vein drains each adrenal. The left is a tributary of the left renal vein, the right leads directly to the cava, although rarely may join a hepatic vein first. The right adrenal vein is short.

Imaging Modalities

The diagnosis of hyperfunction is made clinically and endocrinologically, not radiologically. Imaging is reserved for localization and characterization of adrenal lesions [1, 2].

Radiography

The adrenals may be calcified as a result of previous haemorrhage, infarction, or granulomatous infection (e.g. tuberculosis). Adrenal cysts may be large and show calcification of the wall. Of adrenal tumours, 10–14% are calcified on CT, but this is rarely demonstrable on plain films. Large adrenal masses may be inferred from displacement or distortion of bowel gas or adjacent organs such as the kidney. Rarely, malignant adrenal lesions will invade the kidney and masquerade as a renal tumour on intravenous urography.

Ultrasonography

Ultrasonography is widely available, and does not involve exposure to ionizing radiation. It is, however, very poor at visualizing the normal adrenal or small masses, and a normal

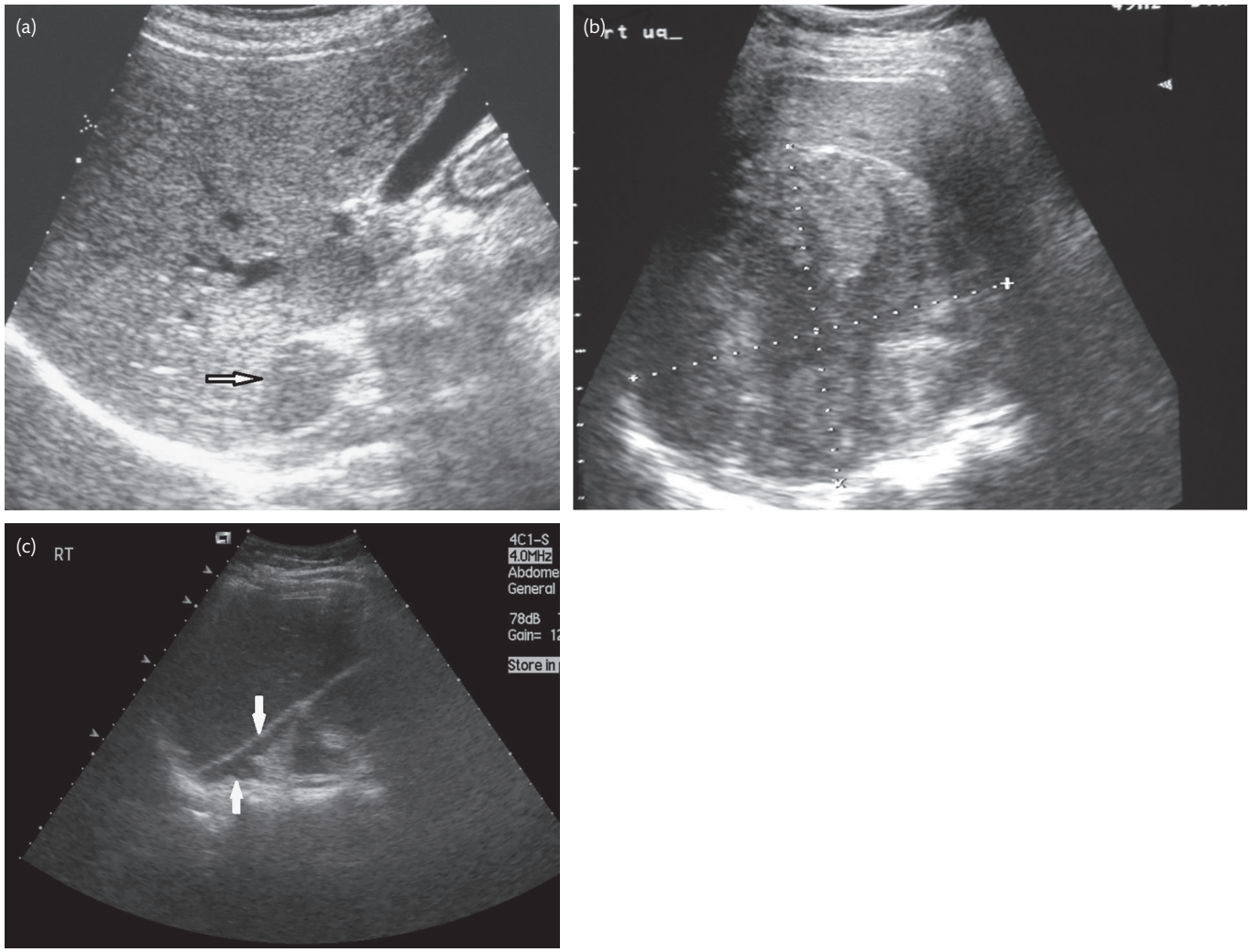


Figure 5.1.1 (a) 2 cm right adrenal Conn's adenoma (arrowed) adjacent to liver. (b) Large adrenocortical carcinoma (measured), with a heterogeneous appearance due to central necrosis (c) A v-shaped hyperplastic adrenal (arrowed) in congenital adrenal hyperplasia.

examination would therefore not exclude an adenoma, adrenal hyperplasia, or small malignant tumours. It could be expected to demonstrate tumours of 2 cm or more in size if the examination is technically complete (Figure 5.1.1) It is more helpful in children where body fat is less of a problem, and when it is particularly desirable to avoid the radiation exposure of a CT examination.

Computed Tomography

CT is the mainstay of adrenal imaging. The normal adrenal gland can almost always be visualized. The right may be more difficult to identify, being in close apposition to the back wall of the cava and the overlying liver. Modern multidetector scanners routinely use thin (1 mm) slices and are capable of resolving tumours as small as 5 mm. Multiplanar reconstructions can clarify anatomy or adjacent organ involvement. The use of contrast media is not necessary for detection of adrenal masses, the anatomical demonstration is sufficient. In some instances, however, the pattern of enhancement may help characterize the lesion (e.g. phaeochromocytomas) may be vascular and enhance strongly.

Staging of malignant adrenal tumours requires scanning of the chest, abdomen, and pelvis for local organ invasion, lymphadenopathy, and metastases. Intravenous contrast medium is necessary for maximum sensitivity for hepatic metastases. Hypervascular metastases (e.g. phaeochromocytoma) may be more conspicuous with scans obtained in the arterial phase rather than the portal venous phase of enhancement.

Magnetic Resonance Imaging

As with CT, MRI allows visualization of the normal and abnormal adrenal gland in the majority of cases. It does not involve exposure to ionizing radiation and hence is preferred in children, the young adult, and the pregnant patient. It has a valuable role particularly in characterization of the indeterminate adrenal lesion using chemical shift imaging. The ability to image in multiple planes allows improved recognition of adjacent organ involvement, and possibly determination of the suprarenal origin of a mass (Figure 5.1.2). However, when staging malignancy, it is usually difficult to evaluate the whole body as the examination time would be prolonged, and it is poorly sensitive to small metastases in the lung.

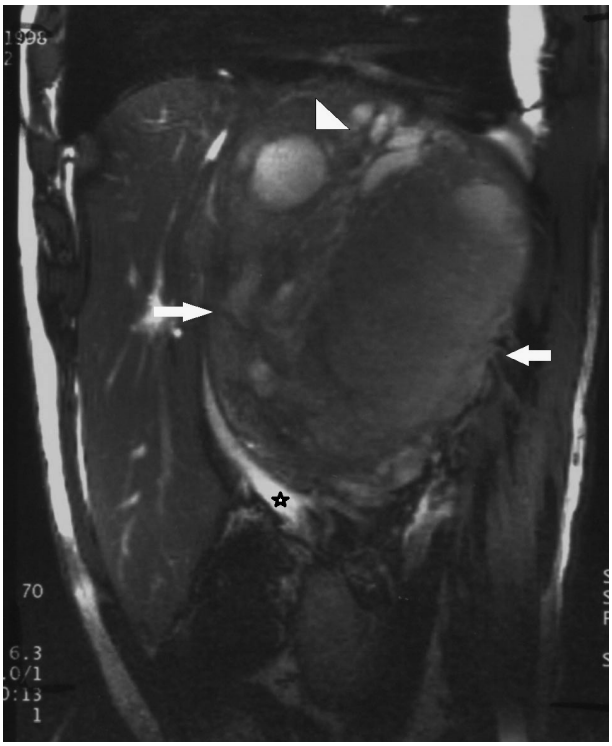


Figure 5.1.2 Large phaeochromocytoma (arrows). Sagittal T_2-weighted magnetic resonance image with flowing blood and areas of fluid necrosis (arrowhead) as white. Note anterior displacement of inferior vena cava (star) and heterogeneity of tumour.

Radionuclide Imaging

The functional information afforded by imaging with radioisotopes is complementary to the anatomical demonstrations of CT and MRI. Various radiopharmaceuticals are available. Standard isotope imaging of the adrenal medulla uses the noradrenaline analogue meta-iodobenzylguanidine (MIBG) [3]. The tracer is actively taken up in postsynaptic nerve terminals where it is resistant to degradation and can hence be used to demonstrate accumulation in such tissue as in phaeochromocytomas, paragangliomas, and neuroblastomas. Other neuroendocrine tumours such as carcinoid and medullary thyroid tumours also take up the radiopharmaceutical. MIBG is labelled with ^{123}I- or ^{131}I-iodine; ^{123}I gives a lower radiation dose and better quality images but is less readily available and more expensive. A number of drugs inhibit MIBG uptake: opioids, tricyclic antidepressants, sympathomimetics, antipsychotics, cocaine, and, importantly, antihypertensive agents including labetalol and calcium channel blockers, and such drugs need to be withdrawn, if possible, before this examination [3].

MIBG scans are usually performed at 24 h and 48 h. Planar whole-body images are best supplemented by SPECT-CT (single photon emission computed tomography-computed tomography), an image fusion technique of isotope and CT data, giving increased sensitivity, specificity and anatomical localization (**Figure 5.1.3**) [4]. Physiological uptake is seen in liver, spleen, myocardium, and salivary glands. Occasionally, bowel uptake is seen which may hinder interpretation, although SPECT-CT usually resolves any difficulties. Normal adrenal glands are usually not well visualized although there may be faint uptake.

An alternative radiopharmaceutical used for neuroendocrine tumours is the somatostatin analogue octreotide acetate labelled with ^{123}I or ^{111}In (indium), which localizes to somatostatin receptor-bearing tumours including phaeochromocytomas and neuroblastomas [3]. However new positron emission tomography (PET) pharmaceuticals are supplanting octreotide and MIBG [3].

Historically isotope imaging of the adrenal cortex has used labelled cholesterol analogues: 6-β-iodomethyl-19-norcholesterol labelled with ^{131}I (NP-59). Adrenocortical scintigraphy is not now widely available or used [5]. This may be due to the number of patient visits required, the lag time to final result, a relatively high radiation dose, limited availability, but perhaps most importantly the use of endocrine assessments in conjunction with high-resolution anatomical imaging with CT, and metabolic information from PET. Recent work has used radionuclides containing metomidate, derived from the specific CYP11B1/2 inhibitor etomidate, thereby targeting uptake to adrenocortical tissues. Both [^{123}I]-metomidate SPECT-CT [6] and [^{11}C]-metomidate PET-CT [7] have been used,

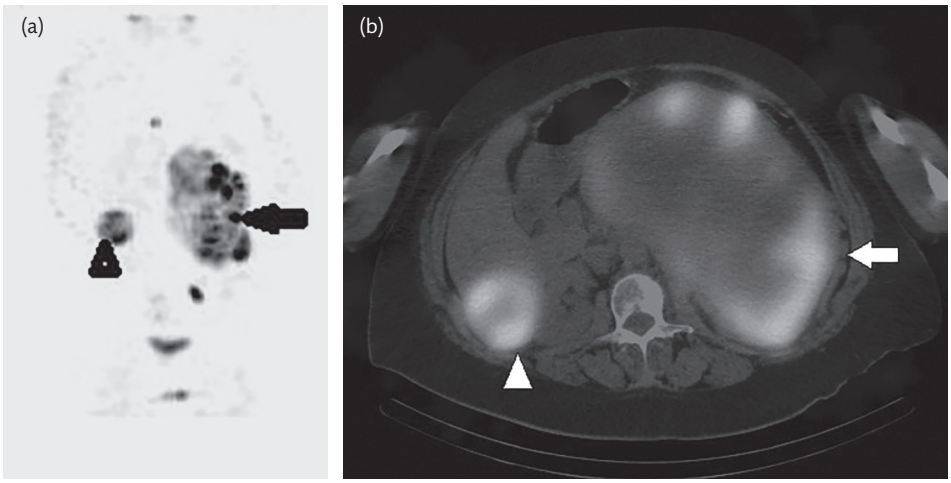


Figure 5.1.3 MIBG scan (a) showing a large MIBG avid left upper quadrant phaeochromocytoma (arrowed) also shown on the CT-SPECT fused image (b) with an FDG avid liver metastasis (arrowhead).

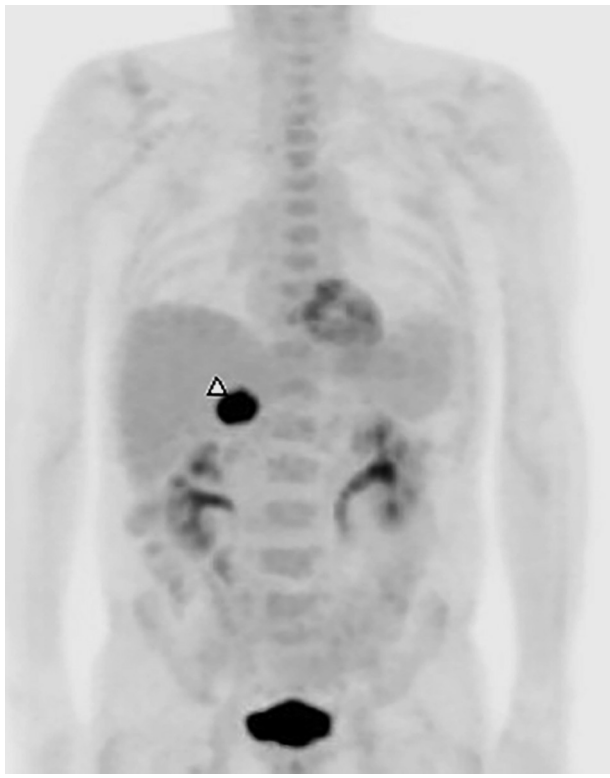


Figure 5.1.4 FDG-PET showing hypermetabolic mass above the right kidney (arrowhead).

but are not yet widely available, pending full assessment of their diagnostic potential in the assessment of adrenocortical tumours.

PET has an increasingly recognized role in the assessment of adrenal masses [8–10]. The most commonly used radiopharmaceutical is [^{18}F]2-fluoro-2-deoxy-D-glucose (FDG), a glucose analogue that is actively taken up and trapped in hypermetabolic cells, usually reflecting malignancy. FDG-PET can thus be used to detect metastases to the adrenals, to indicate that the likelihood that a mass is malignant, and whole-body staging. (Figure 5.1.4) [10]. Although not widely available, other fluorinated pharmaceuticals such as ^{18}F-fluorodopamine, ^{18}F-fluoroDOPA have been used for evaluation of adrenal and other neuroendocrine tumours [10, 11]. Gallium 68 (^{68}Ga) 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)–octreotate (DOTATATE) uses a different positron emitter (^{68}Ga) conjugated with a somatostatin receptor imaging agent [12, 13]. This is used for neuroendocrine tumour imaging—detection of primary tumours and staging. The physiological affinity of DOTATATE is greater than octreotide. The resolution, sensitivity, and specificity afforded by PET are much greater than that of octreotide CT-SPECT. Other similar analogues are available (DOTATOC (GaToc), and DOTANOC (GaNoc)). Therapeutic octreotide administration should be discontinued prior to imaging to avoid competitive inhibition of uptake. Physiological uptake is seen in the spleen, liver, urinary tract, salivary glands, thyroid, and the pituitary. There is high uptake in phaeochromocytoma and paragangliomas, and it is now the functional imaging technique of choice for neuroendocrine tumours (Figure 5.1.5). FDG-PET is complementary as it will highlight more aggressive tumours,

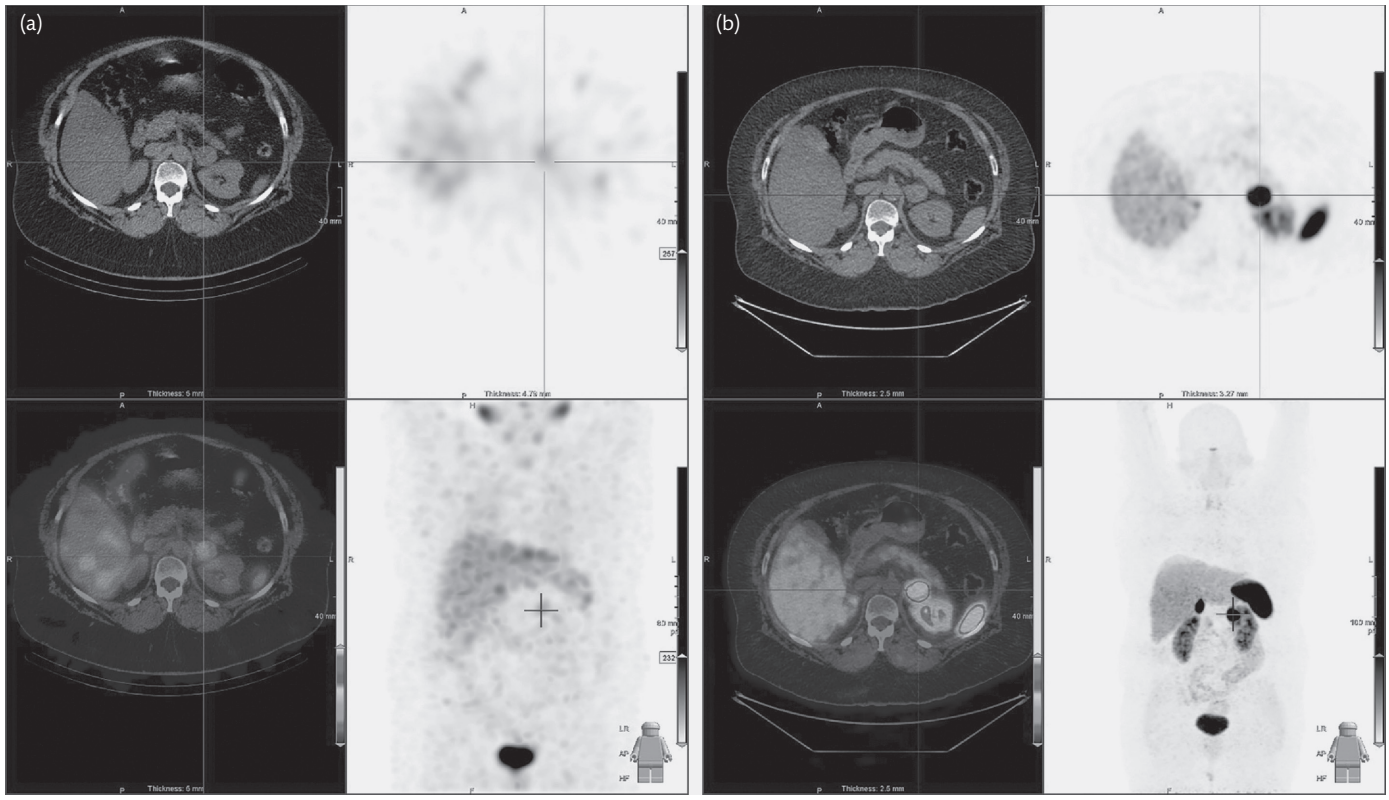


Figure 5.1.5 MIBG (a) and ^{68}Ga DOTATATE (b) scans showing a left sided phaeochromocytoma with only mild uptake on the MIBG scan, but marked avidity on DOTATATE (cross hairs). A second tumour is seen above the right kidney on the DOTATATE scan coronal image.

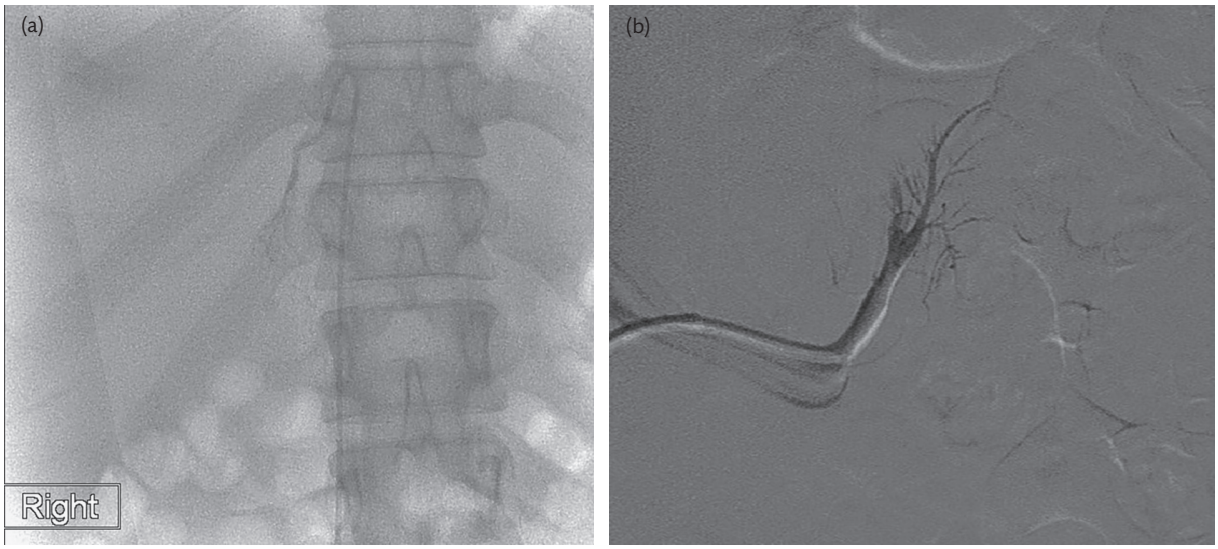


Figure 5.1.6 Adrenal vein sampling (a) right direct from the cava (b) left via the left renal vein.

whereas somatostatin receptor imaging shows better differentiated functioning tumours [14].

Arteriography

The vascular supply and drainage have been described in the anatomy section. Arteriography is rarely indicated for diagnostic purposes. Occasionally it may help indicate an adrenal origin of an uncertain abdominal mass, or be used as a prelude to embolization of vascular tumours.

Venous Sampling

Sampling of the adrenal effluent is used to determine whether aldosterone production originates from one or both adrenals, and to determine if nodules detected on CT are functional or not in patients with primary hyperaldosteronism (Conn's syndrome). The left adrenal vein is relatively easy to cannulate via the renal vein but the anatomy of the right is less favourable (direct drainage to the inferior vena cava between the T11 and L1 vertebrae.) and success rates between 74% and nearly 100% have been reported [15–17]. The venogram images are not used for diagnosis but to confirm correct catheter placement (Figure 5.1.6). Optimal results arise from close collaboration between clinicians, radiologists and laboratory staff; intraprocedural rapid cortisol measurement can improve the success rate. Extravasation and venous infarction may complicate injection of the adrenal veins.

Biopsy

The indeterminate lesion that is discovered when imaging for other purposes may need histological evaluation. However, endocrine assessment and the newer CT and MR characterization techniques have significantly reduced the requirement for biopsy [18]. Biopsy accuracy rates range between 80% and 100% [18, 19]. However, it is important to note that biopsy cannot differentiate between an adrenocortical adenoma and an adrenocortical carcinoma and thus should be avoided if an adrenocortical carcinoma is suspected. Also there is a risk of tumour seeding of the track and a higher likelihood of recurrence and metastasis if the capsule has been violated [20]. However, biopsy can be helpful in other scenarios (e.g. to differentiate between a tumour of adrenal origin and an adrenal metastasis of a solid organ tumour distinct from the adrenal). Haemorrhage and pneumothorax are the most common complications of adrenal biopsy.

CT guidance is usual except where the tumour is relatively large when ultrasonography is a good alternative. A posterior approach with the patient prone is the least hazardous but the posterior costophrenic angles may be deep, and transgression of the lungs may be unavoidable with consequent risk of pneumothorax. A transhepatic approach is an alternative on the right, or on rare occasions a safe anterior approach may be identified. Endoscopic ultrasound guided transgastric approach gives access to the left adrenal (Figure 5.1.7).

Most operators remain reluctant to biopsy adrenal masses that may be phaeochromocytomas due to the risk of precipitating a hypertensive crisis [16]. Prior blood or urine biochemistry for exclusion of catecholamine excess is therefore mandatory before any adrenal biopsy.

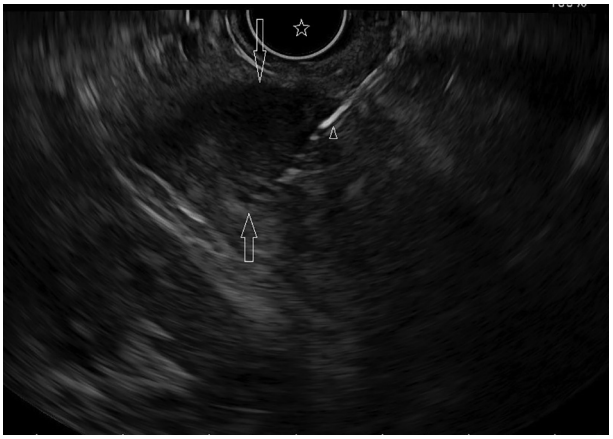


Figure 5.1.7 Left adrenal mass (between arrows) undergoing endoscopic biopsy (needle indicated by arrowhead, endoscopic probe by star).

Cushing's Syndrome

Cushing's syndrome is the result of overproduction of cortisol by the adrenal cortex. The distinction between pituitary or ectopic adrenocorticotrophic hormone (ACTH)-driven cortisol production and primary adrenal disorders is made on a clinical and biochemical basis and imaging directed appropriately.

MRI is the best imaging modality for pituitary tumours, which account for about 80% of cases of Cushing's (Figure 5.1.8a). The multiplanar presentation is ideal, and intravenous contrast medium administration mandatory, for the detection of small tumours. CT is a reasonable alternative. Thin-section coronal examinations with intravenous contrast are most sensitive for small tumours. There is no justification for imaging the adrenals in pituitary-driven or ectopic Cushing's, although bilateral adrenal hyperplasia can be expected (Figure 5.1.8b). Adrenal hyperplasia is manifest as thickening and elongation of the limbs of the adrenal gland. The hyperplasia may be smooth or multinodular, but the glands may look normal.

Adrenal tumours are the cause of Cushing's syndrome in 15–25% of cases and are well shown with CT and MRI [21]. Adenomas are usually between 2 cm and 4 cm in size. They are uniform in attenuation, rounded, and well demarcated (Figure 5.1.8c), identical to non-functioning tumours. They may be large enough to be seen on ultrasonography although increased body fat may hinder the examination. An active adrenal adenoma will be accompanied by atrophy of the rest of the gland and the contralateral adrenal. Carcinomas are usually larger (more than 4 cm) and may show necrosis, haemorrhage, or calcification [21]. Histology cannot be definitive of malignancy but large tumours are predictive of subsequent malignant behaviour, that is, metastasis or recurrence.

Rarely Cushing's syndrome is the result of primary pigmented nodular hyperplasia when multiple small nodules are shown arising from an otherwise atrophic gland; or ACTH-independent macronodular hyperplasia when the glands are markedly enlarged and nodular but maintain their shape [21].

There are other radiological features of Cushing's syndrome. Increased body fat may be evident, especially on CT and MRI.

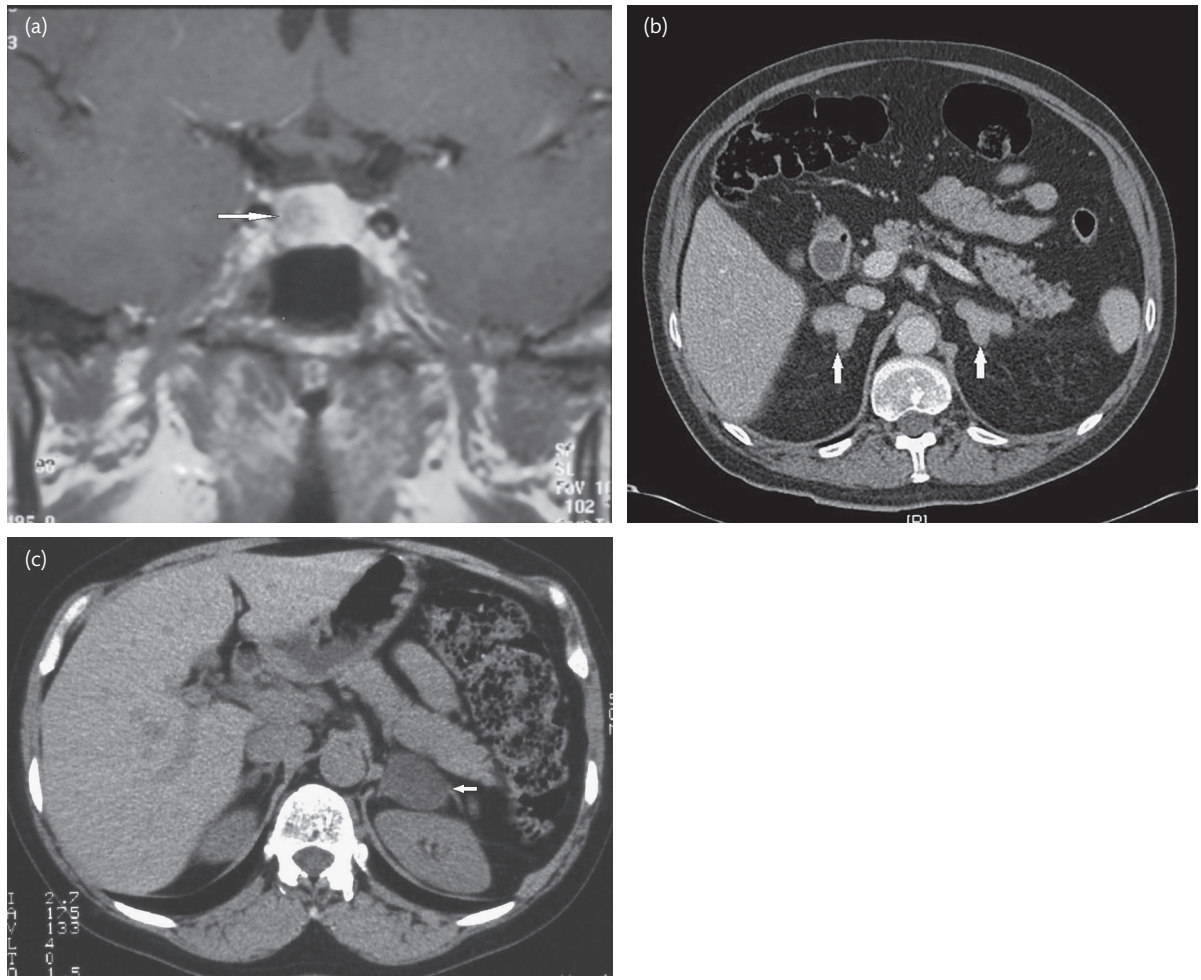


Figure 5.1.8 (a) Cushing's disease: pituitary adenoma (arrowhead) on coronal gadolinium-enhanced MR image. (b) Cushing's syndrome due to ectopic ACTH production. Marked bilateral adrenal hyperplasia (arrows) secondary to an unidentified source of ectopic ACTH. The patient was subjected to bilateral adrenalectomy for treatment. (c) Adrenal Cushing's syndrome: unenhanced CT. Low attenuation, 3 cm left adrenal mass (arrow).

Chronic glucocorticoid overproduction results in skeletal osteoporosis. Diagnosis of osteoporosis is best done by bone mineral densitometry usually with DEXA (dual energy X-ray absorptiometry).

Primary Hyperaldosteronism (Conn's Syndrome)

Conn's syndrome results from overproduction of aldosterone either from an adrenal adenoma or bilateral hyperplasia [16]. The distinction is crucial as it directly affects surgical management. A unilateral adrenalectomy is often curative for unilateral secretion while bilateral hyperaldosteronism is usually treated medically. The pitfall is to remove an incidental non-functional nodule from the innocent side. Adrenal venous sampling is therefore required. Conn's is very rarely (less than 1%) the result of an adrenocortical carcinoma, usually presenting as a larger tumour (>4 cm).

Aldosterone-producing adenomas are often small (less than 2 cm), compared to cortisol-producing adenomas requiring thin-section CT for detection (**Figure 5.1.9a**). The average diameter is between 12 mm and 18 mm and 20% are less than 10 mm. Tumours less than 1 cm can be difficult to identify. Conn's tumours are characteristically lipid rich and usually have the lowest attenuation values of all hyperfunctioning adenomas (**Figure 5.1.9b**) [22].

Hyperplastic glands may appear normal or show obvious symmetric enlargement. MRI has no real advantage over CT although high signal returned from a lesion on T_2-weighted imaging may aid detection.

The definitive test is adrenal venous sampling with an accuracy rate of close to 100% (**Figure 5.1.6**) [15–17]. The ratio between aldosterone and cortisol in the venous blood from each adrenal is compared with peripheral samples. Samples following ACTH stimulation may be taken in some centres though no consensus exists whether this improves diagnostic accuracy. A lateralizing result for a functioning adenoma is achieved if the aldosterone:cortisol ratio is four times the opposite side [16, 17].

^{11}C metomidate PET is not widely available but appears to be non-invasive alternative to adrenal venous sampling (AVS) for lateralization [23], pending more detailed, prospective evaluation of accuracy.

Androgen Excess and Oestrogen Excess

Androgen excess may be of ovarian (rarely testicular) or adrenal origin. Tumorous adrenal causes are mostly malignant. Imaging choices are similar to the investigation of Conn's syndrome, that is, CT or MRI with venous sampling (with the addition of ovarian venous aspirates).

Feminization, for example gynaecomastia, due to adrenal oestrogen excess is rare. As with androgen excess of adrenal origin, the cause is most often an adrenocortical carcinoma of such a size that CT and ultrasonography are invariably helpful.

Adrenocortical Carcinoma

This may be responsible for any of the aforementioned syndromes or be clinically non-functional. They are readily seen on ultrasonography, CT, or MRI, as tumours are usually large at presentation although hyperfunctioning tumours present at a smaller size, as a result of the endocrine effects [24]. Heterogeneity of some degree is usual. Calcification occurs in 30%. An adrenal origin may be difficult to determine on standard axial CT imaging if the tumours are large and invasive, but CT reconstructions or MRI are more informative, using multiple planes and different sequences. Significant growth over time, local invasion of adjacent organs or metastatic disease are features of malignancy (**Figure 5.1.10a, b**). As with renal cell carcinoma, there may be invasion of the adrenal vein and extension into the cava (**Figure 5.1.10c, d**). They rarely contain significant amounts of intracellular lipid, which can be exploited diagnostically as malignant tumours do not therefore show low attenuation density values (<10 Hounsfield units (HU)) on CT and rarely lose signal on opposed-phase MRI.

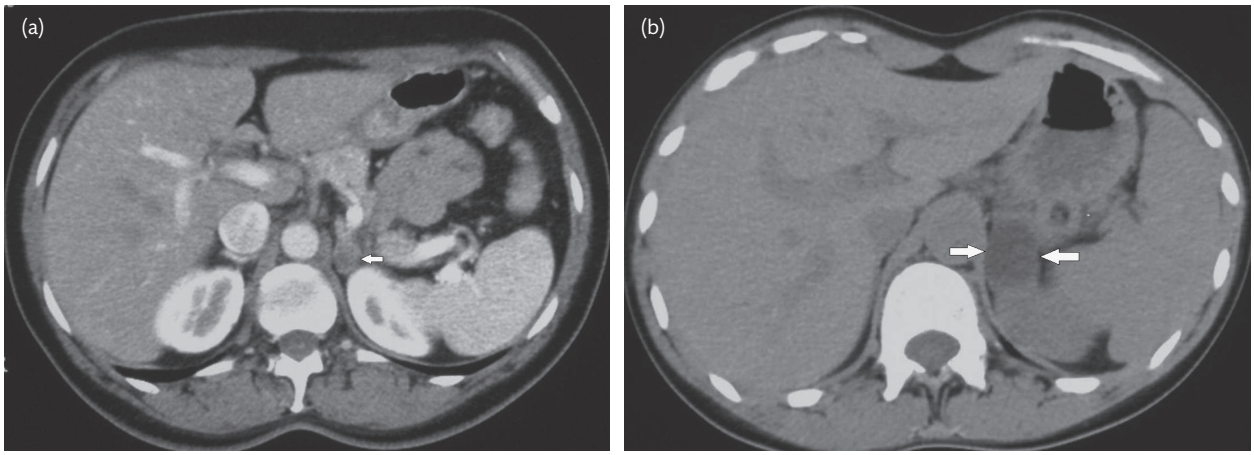


Figure 5.1.9 (a) Small Conn's tumour: enhanced CT showing a small mass (arrow) arising from the medial limb of the left adrenal. (b) Large Conn's tumour: unenhanced CT showing a typically low attenuation left adrenal mass (arrowhead).

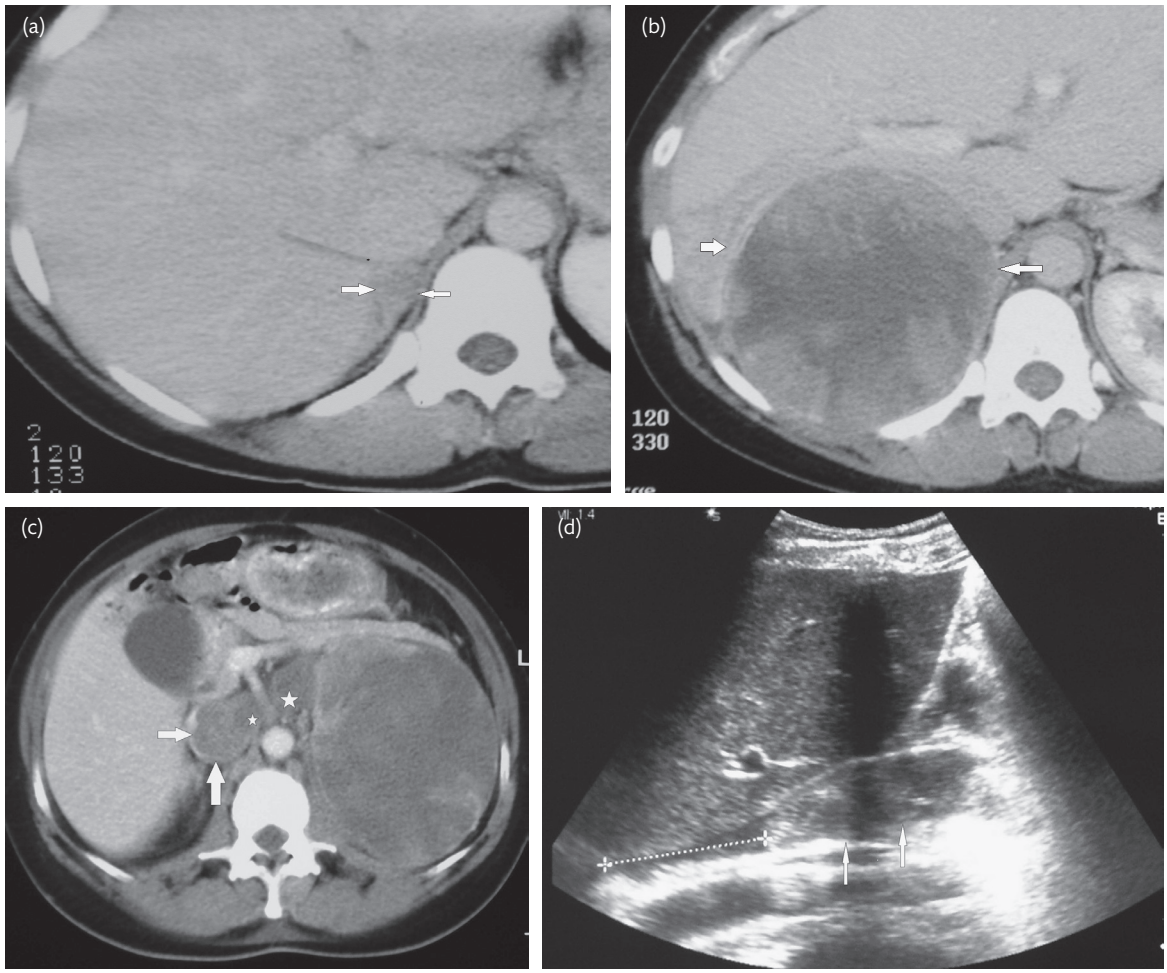


Figure 5.1.10 (a) CT shows an adrenocortical carcinoma, initially small (arrows). demonstrating growth over time (b, arrows). (c) Adrenocortical carcinoma: CT showing a large mass in the left suprarenal region, invading left renal vein (stars) and inferior vena cava on CT and ultrasound (d) (arrows). Measurement marks show length of uninvolved hepatic cava.

Addison's Disease

This term refers to adrenal insufficiency. Autoimmune mechanisms are the commonest cause now that the incidence of tuberculosis has reduced. CT may show atrophy or calcification. Tuberculous infection in the subacute stage produces enlarged adrenals (Figure 5.1.11a), which may show peripheral enhancement around central necrosis. In the long term the glands calcify (Figure 5.1.11b). Histoplasmosis produces similar appearances and half of patients with disseminated disease develop Addison's disease [25].

Bilateral adrenal metastases, even when large, can result in adrenal insufficiency, however usually in less than 20% of cases. This carries the risk that symptoms of Addison's disease may be confused for those of the underlying malignancy (Figure 5.1.12) [26].

Acute adrenal insufficiency as the result of bilateral adrenal haemorrhage or severe hypotension may complicate shock, sepsis, or bleeding disorders. High-attenuation swelling of the adrenals is the finding on CT performed acutely.

Rare causes of hypoadrenalism include haemochromatosis, when CT may demonstrate increased attenuation of liver and pancreas as a result of iron deposition and Wolman's disease (lipid storage abnormality due to a deficiency of liposomal acid lipase) when the adrenals are enlarged and show diffuse punctate calcification.

Phaeochromocytomas

These tumours arise from the chromaffin cells of the sympathetic nervous system. Thus, they most commonly arise in the adrenal medulla (=phaeochromocytoma), but can also be extra-adrenal (=paraganglioma) in the neck, the mediastinum (including intrapericardiac), in a para-aortic position, in an accumulation of sympathetic ganglia at the base of the inferior mesenteric artery known as the organ of Zuckerkandl, and in the pelvis and bladder (Figure 5.1.13) [27].

At least about 25% of apparently sporadic tumours are associated with familial conditions such as neurofibromatosis, von Hippel–Lindau and multiple endocrine neoplasia (MEN) syndromes. These patients are more likely to have bilateral or multiple lesions (Figure 5.1.14) [28].

Phaeochromocytomas are usually, but not always, benign. As with other adrenal tumours, benign and malignant phaeochromocytomas

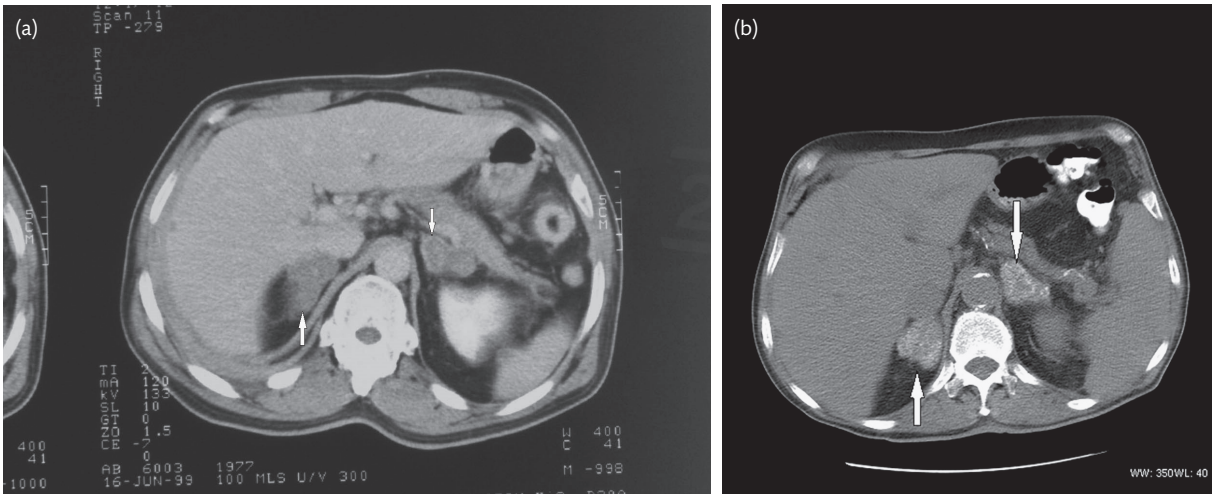


Figure 5.1.11 (a) CT—Active tuberculosis of adrenals—swollen peripherally enhancing adrenals (arrows). (b) Unenhanced CT in a different patient presenting with adrenal insufficiency shows calcified enlarged adrenals (arrows) following previous tuberculous infection.

are distinguished by behaviour (i.e. metastasis or local invasion) rather than histology or radiology [29]. However, the results of genetic analysis are usually predictive of malignancy risk and determine screening and follow-up strategies.

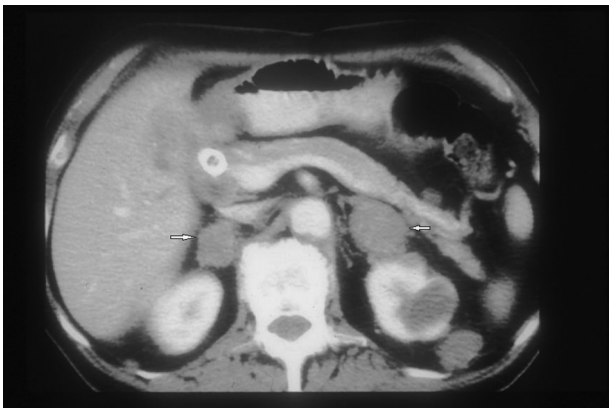


Figure 5.1.12 Bilateral adrenal metastases (arrows) that resulted in adrenal failure.

Adrenal medullary tumours are usually sizeable (greater than 5 cm) except when associated with the MEN syndromes. Therefore they can often be shown with ultrasonography and appear either homogeneous or heterogeneous with cystic or necrotic elements. CT is preferred, however, and will demonstrate the majority of adrenal phaeochromocytomas [28]. About 10% may show calcification. These lesions usually enhance strongly on CT and heterogeneity corresponding to haemorrhage or necrosis is better appreciated after contrast medium enhancement. They are characteristically of high signal on T_2-weighted MR sequences [24]. The older intravenous iodinated ionic contrast agents could precipitate hypertensive crisis in the absence of pharmacological alpha- and β-blockade [30]. The almost ubiquitous non-ionic contrast media used now do not carry the same risk and blockade is not now regarded as necessary anymore [31–33].

CT is generally used to evaluate the adrenals and to search for ectopic sources. MRI does not usually have any additional benefit. However, surgical planning for large or locally invasive tumours is helped by the multiplanar facility of MRI (Figure 5.1.2) (CT also now offers multiplanar reconstructions). The tumour can be

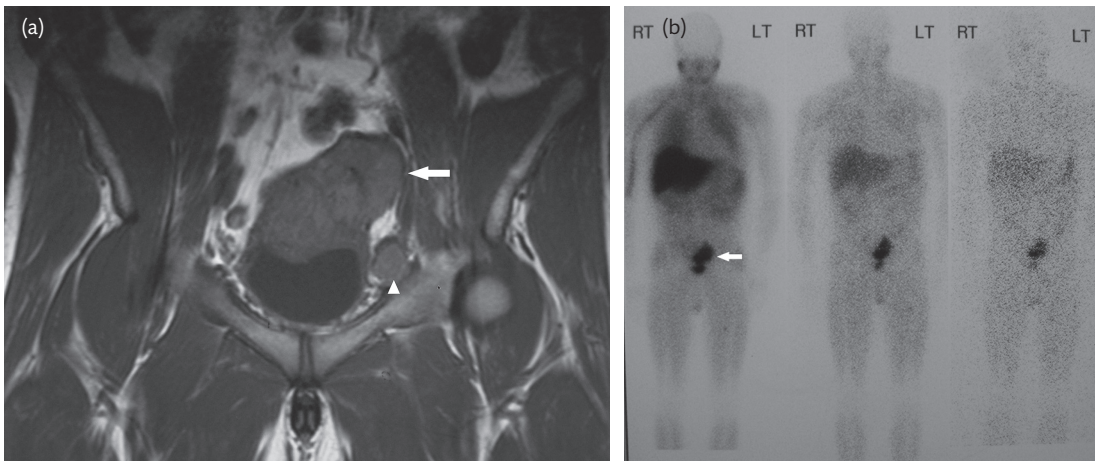


Figure 5.1.13 Bladder phaeochromocytoma. (a) Coronal MR of bladder showing a polypoid tumour (arrow) arising from the left side of the bladder dome (and an enlarged metastatic left pelvic node—arrowhead). (b) Corresponding anterior whole-body MIBG images at 1, 2 and 3 days show an MIBG avid tumour (arrow) above excreted activity in the bladder.

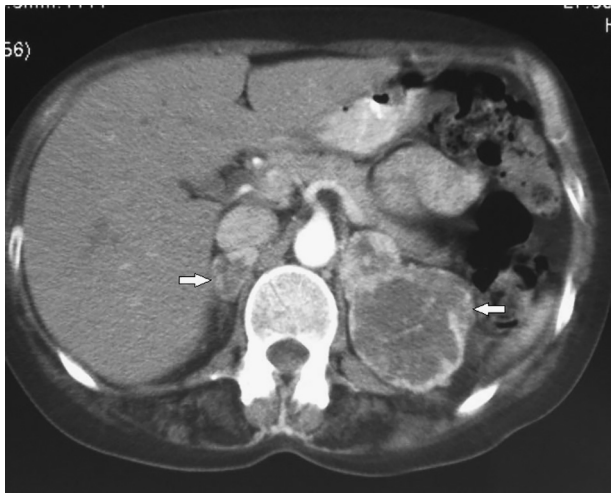


Figure 5.1.14 MEN type 2 associated bilateral phaeochromocytoma. Enhanced CT showing a large left adrenal and smaller right adrenal masses (arrowed) both showing typical marked peripheral enhancement, with central necrosis.

markedly hyperintense on T_2-weighted imaging ('lightbulb' sign) [29, 34, 35]. These lesions are often heterogeneous and vascular. Haemorrhage may occur leading to fluid–fluid levels and areas of high signal on T_1-weighted images. Extension of abdominal tumours into the inferior vena cava is shown with flow-sensitive sequences, or with intravenous contrast enhancement [27, 35]. Mediastinal tumours, and especially intrapericardiac lesions, are well shown with electrocardiographically gated MRI.

The functional technique of MIBG scanning is of value in the imaging of phaeochromocytomas. It is especially useful for the detection of extra-abdominal tumours and for the staging of malignant lesions, as the metastases are active [2, 3, 25]. Whole-body imaging is supplemented by SPECT-CT (Figure 5.1.3) and may be used for the initial imaging test, or to search for an ectopic tumour following a negative adrenal CT (Figure 5.1.15) [4]. MIBG is not specific for phaeochromocytoma; other tumours of neural crest origin such as neuroblastoma, carcinoid tumours, medullary carcinoma of the thyroid, Merkel-cell skin tumours, and non-functioning paragangliomas also show uptake, but this is not usually a clinical dilemma.

As discussed earlier, other radionuclide methods used to demonstrate phaeochromocytomas include ^{111}In-octreotide scanning and PET with FDG (Figure 5.1.16). ^{111}In octreotide is more useful for demonstration of metastatic disease than benign tumours [1]. FDG and MIBG scanning are complementary depending on the degree of differentiation and malignant potential [8]. Other radiopharmaceuticals such as ^{18}F-fluoroDOPA, ^{18}F-fluorodopamine, and particularly Ga-68 DOTATATE are being utilized [6, 10, 11, 36, 37]. Ga-68 DOTATATE is now the functional imaging agent of choice for neuroendocrine tumours), offering greater anatomical resolution and sensitivity (Figure 5.1.5) [12, 14].

Neuroblastoma

This tumour of infancy and childhood can arise in the adrenals (50%), in the abdominal sympathetic chain, or in the mediastinum. MIBG uptake is a feature, and can therefore be used for assessment

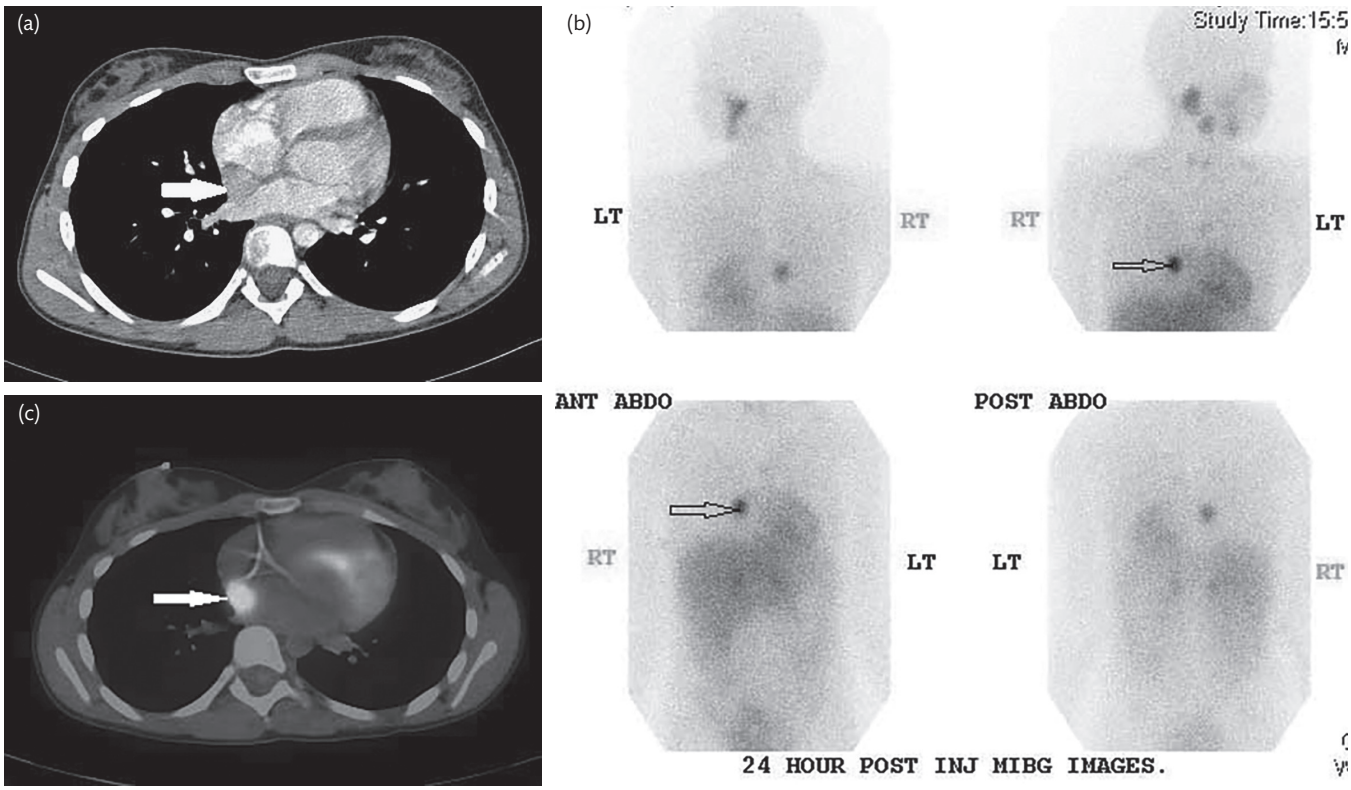


Figure 5.1.15 Paracardiac paraganglioma (arrow) demonstrated on (a) CT (not prospectively identified), (b) MIBG, and (c) 18-FDG-PET scans.

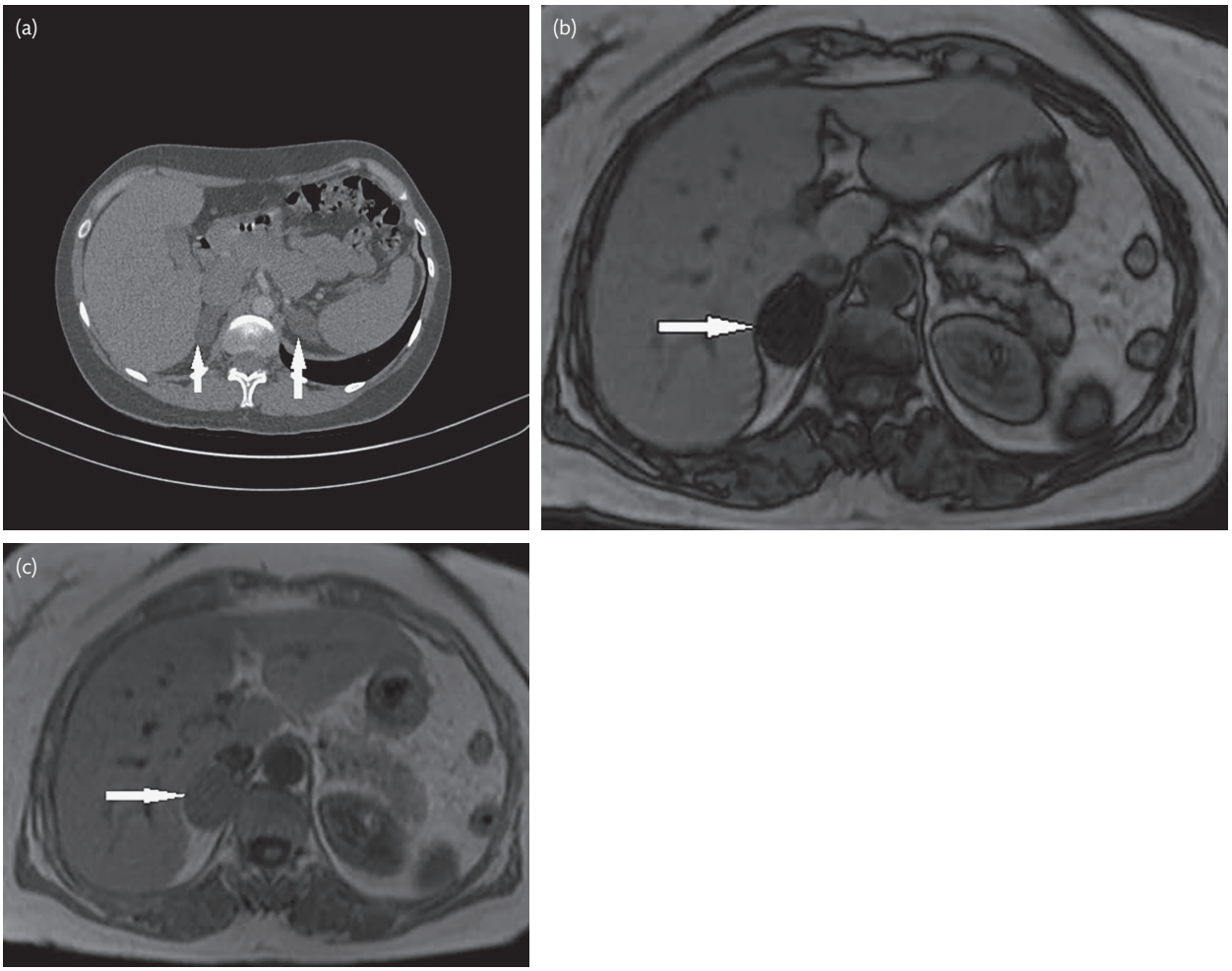


Figure 5.1.16 (a) Unenhanced CT showing bilateral low attenuation adrenal tumours (subjectively less than liver and spleen and less than 10 HU on measurement indicative of bilateral benign lipid-rich adenomas). MR showing marked signal loss (arrowed) on opposed-phase T_1-weighted MR (b); compared to the in-phase study (c); and therefore a lipid-rich right adrenal adenoma.

of metastatic disease, although CT or MRI is more appropriate for the assessment of the local disease. Tumours are often non-homogeneous on CT and MRI, and calcification is characteristic (which helps differentiate it from Wilm's tumour of the kidney). Local invasion into the spine, and skeletal metastatic disease, can occur. The multiplanar capability of MRI allows demonstration of vascular and liver involvement, intraspinal spread, and marrow disease [37].

The Incidental Adrenal Mass

With the exponential increase in the use of modern cross-sectional imaging techniques, adrenal masses are increasingly incidental findings (4–6%). They may be functional in terms of hormone synthesis and this is determined endocrinologically. The vast majority are, however, endocrinologically irrelevant, and are likely, statistically, to be benign if small. They are much more likely to be malignant if there is known primary, extra-adrenal malignancy, although even in this situation less than 50% are metastatic. The differentiation of metastasis or malignancy from an incidental adrenal adenoma is therefore very important [38].

Comparison with old scans or follow-up examination is important—a lesion that significantly changes in size over 3–6 months is highly likely to be malignant. Biopsies are helpful to clarify whether a new adrenal lesion is a metastasis of a known extra-adrenal malignancy; however, a biopsy is not useful for differentiating benign from malignant adrenocortical lesions [39] and should only ever be performed if the results are certain to affect management and after biochemical exclusion of phaeochromocytoma.

The CT appearance may give some indication of the nature of an incidentally discovered adrenal mass. It may be evidently a cyst (i.e. thin-walled, well-defined, and of fluid density). If solid, benign lesions are usually homogeneous, although there may rarely be calcification. Frank areas of fat is virtually diagnostic of a benign myelolipoma, particularly if calcification is also present. Malignant lesions may be irregular in outline, heterogeneous, perhaps with necrotic areas. The morphological appearance following contrast may be of value. Adjacent organ invasion or demonstration of metastasis is diagnostic of malignancy.

The presence of a high proportion of intracellular fat in 70% of adenomas is an important finding allowing the use of CT density measurements and chemical shift MR imaging to diagnose a benign

lesion. Adenomas are often readily apparent as hypoattenuating compared to kidney or liver on unenhanced CT (Figure 5.1.8c, 9b, and 16a). Mean attenuation values of adenomas are 2 HU, and of non-adenomas 30 HU [40]. Phaeochromocytomas invariably present with attenuation <10 HU [41]. On thin-section unenhanced scans a density of 10 HU or less is indicative of a high proportion of intracellular fat, which suggests an adenoma with high specificity (98%) though poorer sensitivity (71%) [42, 43]. Lipid content cannot be assessed on single phase contrast-enhanced scans where density measurement is not of any value (unfortunately, as these lesions are often incidental findings on enhanced scans).

Quantitative enhancement and washout characteristics have been used by many and relates to the fact that benign lesions lose enhancement more rapidly than malignant. The most commonly used threshold is >40% relative washout on a 15 min delayed scan, or >60% absolute washout [38, 44, 45]. Phaeochromocytomas, however, may also show early washout [45]. Refinements using dual energy subtraction techniques to generate a virtual unenhanced CT from an enhanced study has been promising, but dual energy scanners are not widely available [45–47]. However, in contrast to unenhanced CT attenuation, the evidence base for the use of washout CT in differentiating benign from malignant adrenal lesions is very limited [48].

Early users of MRI suggested that high signal intensity on T_2-weighted images indicated a malignant lesion, benign lesions generally having signal intensity similar to normal adrenal; however, it became apparent that there was too much overlap (20–30%) for this to be a useful feature. Gadolinium contrast medium enhancement was likewise unreliable, even with the use of dynamic acquisitions. The most robust technique appears to be the use of chemical shift imaging [38, 45, 49]. This utilizes the fact that the presence of fat within benign adenomatous cells alters the local magnetic environment, and hence the resonant frequency of the processing protons. This results in a reduction of signal intensity of benign lesions (whose cells contain both lipid and water) on out-of-phase imaging (Figures 5.1.16b and c). It appears to be useful in cases where CT attenuation is less than 30HU [34, 45]. Normal adrenals also display this phenomenon but non-adenomatous lesions do not. The principle is entirely different to saturation sequences used for fat suppression, which apply to macroscopic fat. Visual assessment can be supplemented by quantification (e.g. signal intensity index of adrenal/spleen intensity ration but results are similar).

There is work indicating that MR spectroscopy can characterize adrenal masses as adenomas, carcinomas, phaeochromocytomas, or metastases but at present this can only be used for masses larger than 2 cm [50, 51].

FDG-PET relies on the altered metabolism of cancer cells trapping this radiopharmaceutical which enters the glycolytic pathway in the place of glucose. It has been shown to accurately differentiate benign and malignant adrenal masses in the cancer patient with specificities of 90–96%, and sensitivities of 93–100% [8, 9]; however, this is best on series of limited numbers. It is the most accurate test therefore to confirm that an adrenal mass is metastatic in the patient with an underlying extra-adrenal malignancy.

A meta-analysis found little evidence for clear superiority of any imaging test for characterization of incidental adrenal masses (although unenhanced CT was of value) [48]. A reasonable approach to the incidentally discovered adrenal lesion, whether in the setting of a known malignancy or not, is to perform density measurement on thin-section unenhanced CT. On unenhanced CT if the density is less than 10 HU, malignancy is almost certainly excluded. In the range 10–20 HU in- and out-of-phase MRI might be helpful. If greater than 20 HU, MRI is less likely to resolve the issue due to the lack of intracellular lipid, and FDG-PET may be helpful [52].

Non-Adenomatous Adrenal Abnormalities

Myelolipomas are benign and contain elements of fat and bone marrow [38]. The fat is diagnostic and can be demonstrated on ultrasonography (hyperechoic), CT (low attenuation), or MRI (high signal on T_1-weighting, focal loss of signal on fat-suppressed images, but not on opposed-phase sequences) (Figure 5.1.17). However, fat is not always a dominant feature and the appearance of the lesion is then non-specific. Calcific elements may be present. Haemorrhage may complicate the imaging features.

Adrenal cysts are endothelial (lymphangiomas and haemangiomas), epithelial (retention cysts, embryonal, or cystic adenomas), pseudocysts (resulting from previous haemorrhage), or echinococcal (hydatid). Of adrenal cysts, 15% show mural calcification, particularly in hydatid disease [53, 54].

The CT appearance of acute adrenal haemorrhage is of high-attenuation material expanding the adrenal gland or periadrenal haemorrhage leading to stranding, and an indistinct adrenal contour (Figure 5.1.18a). The high attenuation may not be appreciated if only contrast-enhanced scans are available. Subacute haematomas are isodense with normal adrenal tissue and indistinguishable from adenomas. Old adrenal haematomas may lead to calcification. Because of the paramagnetic effects of blood such as haemosiderin and methaemoglobin, MRI can be diagnostic of adrenal haematoma, although there will be a complex variation depending on the age of the lesion (Figure 5.1.18b, c, and d).

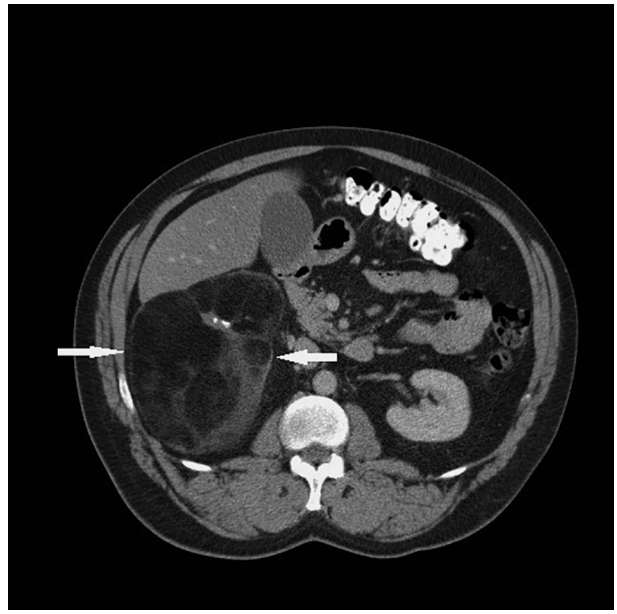


Figure 5.1.17 Adrenal myelolipoma. CT showing a left adrenal mass with macroscopic fat (dark grey elements) and small foci of calcium.

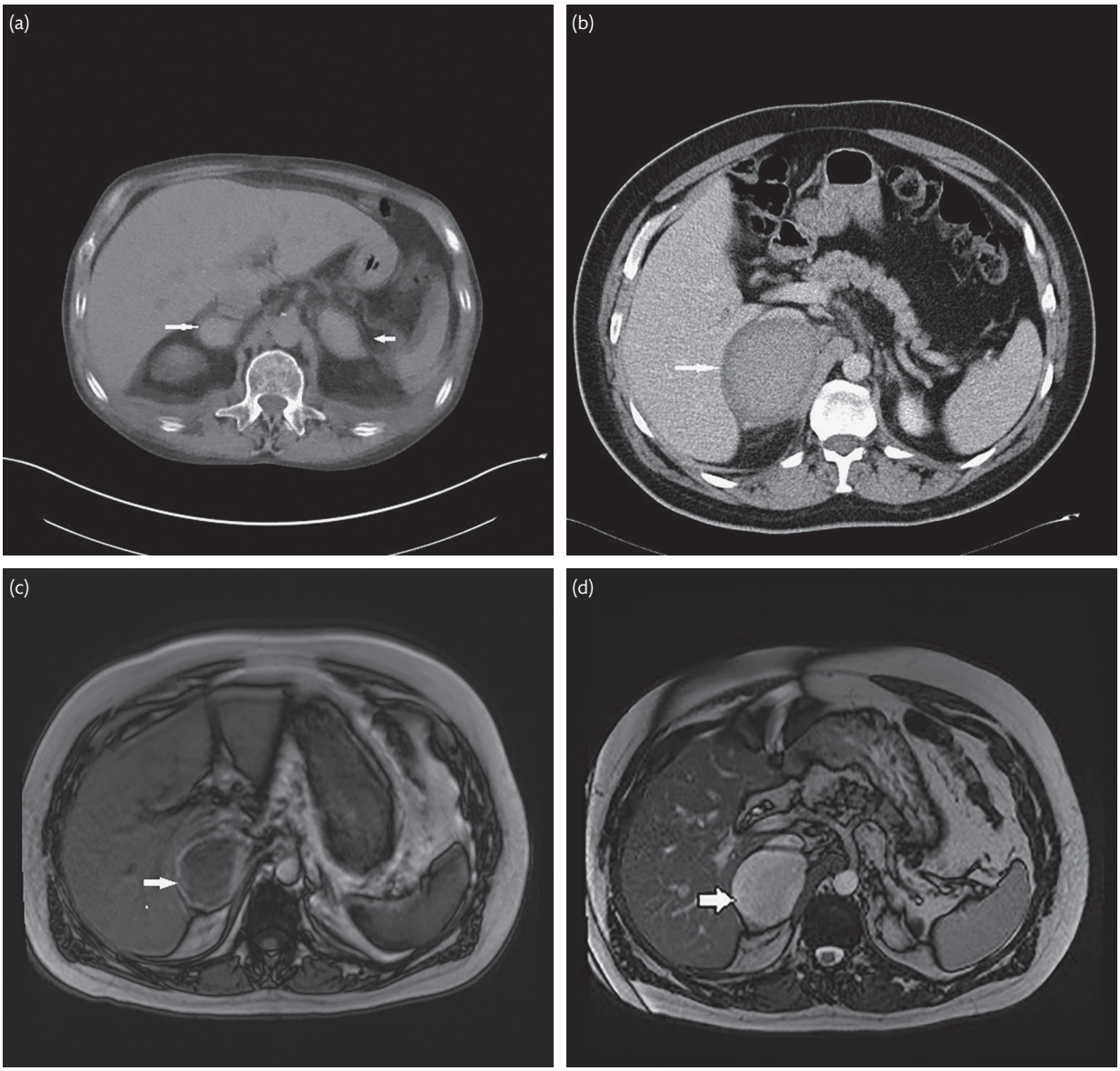


Figure 5.1.18 (a) Acute haemorrhage into both adrenals caused by excess anticoagulation. The non-enhanced CT scan shows increased density and expansion of both adrenals, diagnostic of haemorrhage (arrows). (b–d) Acute haemorrhage into the right adrenal caused by excess anticoagulation (arrows). Although an enhanced CT scan there is increased density and expansion of the right adrenal on indicating fresh blood. The T_1-weighted MRI (c) shows peripheral high signal consistent with blood products and high signal centrally on T_2-weighting (d) indicative of fluid.

Adrenal Rests

Adrenal tissue can be found in ectopic sites such as the coeliac plexus region, the broad ligaments, the testes, and the ovaries. These rests arise during fetal development when the urogenital ridge separate into adrenals and gonads, therefore, often have properties of both. Hence, adrenal rest tissue may enlarge due to ACTH stimulation (e.g. in pathological conditions such as congenital adrenal hyperplasia; Figure 5.1.19) [55], or Addison's or ectopic Cushing's disease. This phenomenon may lead to the misdiagnosis of testicular or other tumours [56], in particular in congenital adrenal hyperplasia.

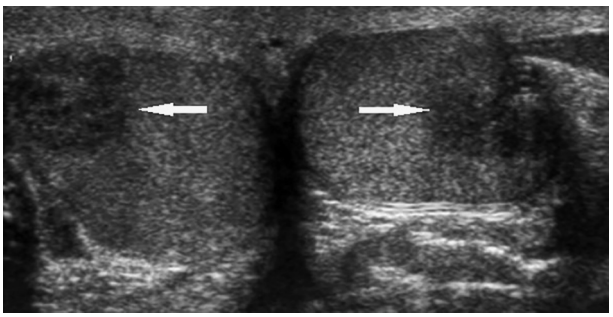


Figure 5.1.19 Bilateral adrenal rest 'tumours' in the testicles of a patient with congenital adrenal hyperplasia (arrows) (same patient as Figure 5.1.1b).

REFERENCES

1. Ilias I, Sahdev A, Reznek R, Grossman A, Pacak K. The optimum imaging of adrenal tumours: a comparison of different methods. *Endocr Relat Cancer*, 2007; **14**: 587–99.
2. Mayo-Smith WW, Boland GW, Noto RB, Lee M. State-of–the-art adrenal imaging. *Radiographics*, 2001; **21**: 995–1012.
3. Taieb D, Timmers H, Hindie E, *et al.* EANM 2012 guidelines for radionuclide imaging of phaeochromocytoma and paraganglioma, *Eur J Nuc Med Mol Imaging*, 2012; **39**: 1977–95.
4. Rozovsky K, Koplewitz B, Krausz Y, *et al*. Added value of SPECT/CT for correlation of MIBG scintigraphy and diagnostic CT in neuroblastoma and pheochromocytoma *Am J Roentgenol*, 2008; **190**: 1085–90.
5. Ryo U, Johnston S, Kim I, Pinsky S. Adrenal scanning and uptake with ^{131}I-6-beta-iodomethyl-nor-cholesterol. *Radiology*, 1978; **128**: 157–61.
6. Hahner S, Kreissl MC, Fassnacht M, *et al*. Functional characterization of adrenal lesions using [123I]IMTO-SPECT/CT. *J Clin Endocrinol Metab*, 2013; **98**: 1508–18.
7. Hennings J, Lindhe O, Bergström M, Långström B, Sundin A, Hellman P. [11C]metomidate positron emission tomography of adrenocortical tumors in correlation with histopathological findings. *J Clin Endocrinol Metab*, 2006; **91**: 1410–14.
8. Elaini AB, Shetty SK, Chapman VM, *et al*. Improved detection and characterization of adrenal disease with PET-CT. *Radiographics*, 2007; **27**: 755–67.
9. Chong S, Lee KS, Kim HY, *et al*. Integrated PET-CT for the characterization of adrenal gland lesions in cancer patients: diagnostic efficacy and interpretation pitfalls. *Radiographics*, 2006; **26**: 1811–26.
10. Dong A, Cui Y, Wang Y, Zuo C, Bai Y. ^{18}F-FDG PET/CT of adrenal lesions. *Am J Roentgenol*, 2014; **203**: 245–52.
11. Kaji P, Carrasquillo J, Linehan W, *et al.* The role of 6-(^{18}F) fluorodopamine postiron emission tomography in the localization of adrenal pheochromocytoma associated with von Hippel–Lindau syndrome. *Eur J Endocrinol*, 2007; **156**: 483–7.
12. Hofman M, Lau W, Hicks R. Somatostatin receptor imaging with ^{68}Ga DOTATATE PET/CT: clinical utility, normal patterns, pearls and pitfalls in interpretation. *Radiographics*, 2015; **35**: 500–16.
13. Sanli Y, Garg I, Kandathil A, *et al*. Neuroendocrine tumour and management: ^{68}Ga-DOTATATE PET/CT. *Am J Roentgenol*, 2018; **211**: 1–11.
14. Hope T, Bergsland, Bozcurt M, *et al.* Appropriate use criteria for somatostatin receptor imaging in neuroendocrine tumors. *J Nuc Med*, 2018; **59**: 66–74.
15. Daunt N. Adrenal vein sampling: how to make it quick, easy and successful. *Radiographics*, 2005: **25** (Suppl 1); S143–58.
16. Patel SM, Lingam R, Beaconsfield TI, Tran TL, Brown B. Role of radiology in the management of primary aldosteronism. *Radiographics*, 2007; **27**: 1145–57.
17. Young W, Stanson A. What are the keys to successful adrenal venous sampling (AVS) in patients with primary hyperaldosteronism? *Clin Endocrinol*, 2009; **70**: 14–17.
18. Paulsen S, Hghiem H, Korobkin M, Caoili E, Higgins E. Changing role of imaging-guided biopsy of adrenal masses: evaluation of 50 adrenal biopsies. *Am J Roentgenol*, 2004; **182**: 1033–7.
19. Mody MK, Kazeroooni EA, Korobkin M. Percutaneous CT-guided biopsy of adrenal masses: immediate and delayed complications. *J Comput Assist Tomogr*, 1985; **144**: 67–9.
20. Allolio B, Fassnacht M, Clinical review: adrenocortical carcinoma: clinical update. *J Clin Endocrinol Metab*, 2006; **91**: 2027–37.
21. Rockall AG, Babar SA, Sohaib SA, *et al.* CT and MR imaging of the adrenal glands in ACTH-independent Cushing syndrome. *Radiographics*, 2004; **24**: 435–52.
22. Kawashima A, Sandler CM, Fishman EK, *et al.* Spectrum of CT findings in nonmalignant disease of the adrenal gland. *Radiographics*, 1998; **18**: 393–412.
23. Burton T, Mackenzie I, Balan K, *et al.* Evaluation of the sensitivity and specificity of ^{11}C metomidate positron emission tomography (PET)-CT for lateralizing aldosterone production by Conn's adenomas. *J Clin Endocrinol Metab*, 2012; **97**: 100–9.
24. Elsayes KM, Mukundan G, Narra VR, *et al.* Adrenal masses: MR imaging features with pathologic correlation. *Radiographics*, 2004; **24**: S73–86.
25. Francis IR, Gross MD, Shapiro B, Korobkin M, Quint LE. Integrated imaging of adrenal disease. *Radiology*, 1992; **184**: 1–13.
26. Seidenwurm DJ, Elmer EB, Kaplan LM, *et al.* Metastases to the adrenal gland and development of Addison's disease. *Cancer*, 1984; **54**: 552–7.
27. Francis I, Korobkin M. Phaeochromocytoma. *Radiol Clin N Am*, 1996; **34**: 1101–12.
28. Chrisoulidou A, Kaltsas G, Ilias I, Grossman A. The diagnosis and management of malignant phaeochromocytoma and paraganglioma. *Endocr Relat Cancer*, 2007; **14**: 569–85.
29. Leung K, Stamm M, Raja A, Low G. Pheochromocytoma: the range of appearances on ultrasound, CT, MRI and functional imaging. *Am J Roentgenol*, 2013; **200**: 370–8.
30. Raisanen J, Shapiro B, Glazer GM, Desai S, Sisson JC. Plasma catecholamines in phaeochromocytoma: effect of urographic contrast media. *Am J Roentgenol*, 1984; **143**: 43–6.
31. Mukherjee JJ, Peppercorn PD, Reznek RH, *et al.* Pheochromocytoma: effect of non-ionic contrast medium in CT on circulating catecholamine levels. *Radiology*, 1997; **202**: 227–31.
32. Bessel-Browne R, Malley ME. CT of pheochromocytoma and paraganglioma: risk of adverse events with IV administration of nonionic contrast material. *Am J Roentgenol*, 2007; **188**: 970–4.
33. Baid S, Lai E, Wesley, *et al*. Brief communication: radiographic contrast infusion and catecholamine release in patients with pheochromocytoma. *Ann Intern Med*, 2009; **150**: 27–32.
34. Shin Y, Kim K. Imaging features of various adrenal neoplastic lesions on radiologic and nuclear medicine imaging. *Am J Roentgenol*, 2015; **205**: 554–63.
35. Gilfeather M, Woodward P. MR imaging of the adrenal glands and kidneys. *Semin Ultrasound CT MRI*, 1998; **19**: 53–66.
36. Kaji P, Carrasquillo J, Linehan W, *et al.* The role of 6-[^{18}F] florodopamine positron emission tomography in the localization of adrenal pheochromocytoma associated with von Hippel–Lindau syndrome. *Eur J Endocrinol*, 2007; **156**: 483–7.
37. Bilal MM, Brown JJ. MR imaging of renal and adrenal masses in children. *MRI Clin N Am*, 1997; **5**: 179–97.
38. Boland G; Blake M, Hahn P, Mayo-Smith W. Incidental adrenal lesions: principles, techniques, and algorithms for imaging characterization. *Radiology*, 2008; **249**: 756–75.
39. Bancos I, Tamhane S, Shah M, Delivanis DA, Alahdab F, Arlt W, Fassnacht M, Murad MH. Diagnosis of endocrine disease: the diagnostic performance of adrenal biopsy: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R65–80.

40. Boland GW, Lee ML, Gazelle GS, Halpern EF, McNicholas MJM, Mueller PR. Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. *Am J Roentgenol*, 1998; **171**: 201–4.
41. Canu L, Van Hemert JAW, Kerstens MN, *et al.* CT characteristics of pheochromocytoma: relevance for the evaluation of adrenal incidentaloma. *J Clin Endocrinol Metab*, 2019; **104**: 312–18.
42. Korobkin M, TJ Giordano, FJ Brodeur, *et al.* Adrenal adenomas: relationship between histological lipid and CT and MR findings. *Radiology*, 1996; **200**: 743–7.
43. Francis IR, Korobkin M. Incidentally discovered adrenal masses. *MRI Clin N Am*, 1997; **5**: 147–64.
44. Korobkin M, Brodeur FJ, Francis IR, Quint LE, Dunnick NR, Londy F. CT time-attenuation washout curves of adrenal adenomas and nonadenomas. *Am J Roentgenol*, 1998; **170**: 747–52.
45. Schieda N, Siegelman E. Update on CT and MRI of adrenal nodules. *Am J Roentgenol*, 2017; **208**: 1206–17.
46. Gnannt R, Fischer M, Goetti R, Karlo C, Leshka S, Alkadhi H. Dual-energy CT for characterization of the incidental adrenal mass: preliminary observations. *Am J Roentgenol*, 2011 **198**: 138–44.
47. Kim Y, Park B, Kim C, Park S. Adenoma characterization: adrenal protocol with dual-energy CT. *Radiology*, 2013; **267**: 155–63.
48. Dinnes J, Bancos I, di Duffano F, *et al. Eur J Endocrinol*, 2016; **175**: R51–64.
49. Israel G, Korobkin M, Wand C, Hecht E, Krinsky G. Comparison of unenhanced CT and chemical shift MRI in evaluating lipid-rich adrenal adenomas. *Am J Roentgenol*, 2004; **183**: 215–19.
50. Faria J, Goldman S, Szejnfeld J, *et al.* Adrenal masses: characterization with *in vivo* proton MR spectroscopy–initial experience. *Radiology*, 2007; **245**: 788–97.
51. Kim S, Salibi N, Hardie A, *et al.* Characterization of adrenal pheochromocytoma using respiratory-triggered proton MR spectrosopy: initial experience. *Am J Roentgen*, 2009; **192**: 450–4.
52. Garrett R, Nepute J, Hayek M, Albert S. Adrenal incidentalomas: clinical controversies and modified recommendations. *Am J Roentgenol*, 2016; **206**: 1170–8.
53. Pender SM, Boland GW, Lee MJ. The incidental nonhyperfunctioning adrenal mass: an imaging algorithm for characterization. *Clin Radiol*, 1998; **53**: 796–804.
54. Rozenblit A, Morehouse HT, Amis ES. Cystic adrenal lesions: CT features. *Radiology*, 1996; **201**: 541–8.
55. Martinez-Aguayo A, Rocha A, Rojas N, *et al.* Testicular adrenal rest tumors and Leydig and Sertoli cell function in boys with classical congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2007; **92**: 4583–9.
56. Avila NA, Premkumar A, Shawker TS, Jones JV, Laue L, Cutler GB. Testicular adrenal rest tissue in congenital adrenal hyperplasia: findings at gray-scale and color Doppler US. *Radiology*, 1996; **198**: 99–104.

FURTHER READING

Blake MA, Boland GWL, eds. *Adrenal Imaging*. Totowa, NJ: Springer, 2009.

Low G, Dhliwayo H, Lomas D. Adrenal neoplasms. *Clin Radiology*, 2012; **67**: 988–1000.

5.2

Adrenal Surgery

Fausto Palazzo and Radu Mihai

Introduction

The last two decades have witnessed a significant increase in the diagnosis of both functioning and non-functioning adrenal tumours. The widespread use of cross-sectional imaging has unveiled a previously undiagnosed reservoir of as many as 10% of adults with enlarged adrenal glands. Increasingly sensitive oncological staging has led to the identification of patients with previously undiagnosed solitary adrenal metastases. The (slow) acceptance of screening of hypertensive populations for hyperaldosteronism has been encouraged by the recognition of the potential benefits of adrenal surgery in hyperaldosteronism. All of these changes have increased the number of patients considered for adrenal surgery.

The operative approach to the adrenal gland, in particular for benign lesions, has now been consolidated towards one of two main minimal access approaches proven to be associated with a faster recovery and reduced morbidity compared to traditional open surgery. However, studies in both the United Kingdom and United States have demonstrated that the results of adrenal surgery tightly correlate with surgical volume and that most surgeons undertaking an adrenalectomy have unacceptably small volume practices [1, 2]. The centralization of adrenal surgery remains one of the challenges for the future.

Despite the changes in adrenal practice over the years, the surgical principles in adrenal disease remain unchanged and the perioperative management of adrenal patients mandates a multidisciplinary collaboration with colleagues in endocrinology, biochemistry, and radiology. The key principles in adrenal surgery are that:

- Functional studies are used to confirm or exclude a syndromic diagnosis;
- The patient should be made safe with the appropriate pharmacological treatment;
- Radiological studies are required to confirm the anatomical characteristics of the abnormality;
- The need for surgical intervention should be ratified in the context of a multidisciplinary team meeting.

This chapter will focus primarily on the background history, functional embryology, and anatomy of the adrenal glands and the surgical aspects of treatment. The pathology of adrenal disease, details of biochemical and radiological investigations, and the non-surgical modalities of treatment are covered elsewhere.

Historical Perspective

Landmark events in our knowledge about the existence and function of the adrenal glands are summarized in Table 5.2.1. While

Table 5.2.1 Timeline of understanding the role of adrenal glands

1552	The adrenal glands were first described by Bartholomaeus Eustachius in his *Opuscula Anatomica* as 'glandulae renis incumbentes' (glands lying on the kidney)
1629	Jean Riolan of Paris introduced the term *capsulae suprarenales*
1716	the Academie des Sciences de Bordeaux offered a prize for the answer to the question 'What is the purpose of the suprarenal glands?' but no progress was achieved
1805	Cuvier defined the anatomic division into a cortex and a medulla
1855	Thomas Addison of Guy's Hospital published his clinical observations of 11 patients with destruction of both adrenal glands and described the eponymous clinical syndrome
1856	Brown-Séquard performed adrenalectomy in animals and provided the first experimental confirmation of Addison's theory that the adrenal glands were essential to life
1901	Epinephrine was purified from the adrenal gland
1904	Epinephrine and norepinephrine were first synthesized by Frank Stolz in Germany
1950	Edward Kendall, Tadeus Reichstein, and Philip Hench receive the Nobel Prize in Physiology or Medicine for the discovery of the hormones of the adrenal cortex, their structure, and biological effects

the Italian anatomist Bartholomaeus Eustachi first described the adrenals in the sixteenth century, another 300 years passed until the role of the adrenal gland was clarified and the clinical impact of some adrenal tumours understood. In 1865, DeCrecchio first reported congenital adrenal hyperplasia in a female pseudohermaphrodite. In 1886, Frankel described a type of tumour for which the pathologist Pick in 1912 proposed the name phaeochromocytoma, from the Greek *phaios* (dark or dusky) and *chroma* (colour). In 1912, Harvey Cushing described the classic features of his eponymous syndrome and, in 1955, Conn reported the first patient with primary hyperaldosteronism.

The first reported adrenalectomy was performed by Knowsley-Thornton, in London in 1889, for what is presumed to have been an adrenal malignancy. The first successful adrenalectomy for a phaeochromocytoma is attributed to Cesar Roux, in 1926 in Switzerland, and phaeochromocytoma surgery was immediately recognized to be associated with a high morbidity and mortality. This remained the case at least until the arrival of adrenergic receptor blockers in the 1960s, which, along with improved surgical technique, probably made the largest impact on the safety of this type of surgery [3]. Starting from the 1990s adrenal surgery has benefited greatly from technological advances of fibreoptic scopes and laparoscopic instruments and devices that allow laparoscopic and retroperitoneoscopic surgery to be performed safely in all but the largest adrenal tumours.

Surgical Embryology

The adrenal gland has a dual origin: the cortex arises from the mesoderm and the medulla is neuroectodermal.

The cortex starts to develop in the fifth week of gestation as a proliferation of coelomic mesothelium close to the urogenital ridge (the mesonephros and the developing gonad). In the initial phase the *adrenogonadal primordium* is distinguished and after the eighth week of gestation it separates into two distinct structures, the adrenal and the gonadal primordial. The cortex is divided into three layers: the outer zona glomerulosa, the middle zona fasciculata, and the inner zona reticularis, predominantly secreting aldosterone, cortisol, and sex steroids, respectively.

The medulla develops after the fourth week of embryonic life, when a portion of the neuroectoderm adjacent to the neural tube separates and remains between the neural tube and the definitive ectoderm as the *neural crest*. Cells derived from the neural crest (sympathogonia, primitive spinal ganglia) migrate ventrally from the apex of the neural tube to the dorsal aorta (where they aggregate and differentiate into neuroblasts to form sympathetic neurons) or to the adrenal primordia (where they differentiate into pheochromoblasts to form chromaffin cells). Some primitive adrenal medulla cells remain closely associated with the developing sympathetic nervous system and give rise to the extra-adrenal chromaffin cells and chromaffin bodies. The adrenal medulla functions as postganglionic neural tissue, secreting catecholamines and their metabolites.

Surgical Anatomy

The adrenal glands are golden-brown in colour (Figure 5.2.1) and located in the retroperitoneum in relation to the upper poles of the kidneys at the level of the T11–T12 vertebrae (Figure 5.2.1). The right adrenal gland is a distorted pyramidal structure that lies anterior to the diaphragm and the right kidney and partially posterior to the inferior vena cava and the right lobe of the liver (Figures 5.2.2 and 5.2.3). Superiorly, the gland is in contact with the bare area of the liver. The left adrenal gland is semilunar in shape and extends further inferiorly on the medial margin of the kidney. It is related anteriorly to the stomach and pancreas and posteriorly to the diaphragm and posterior abdominal wall. Surgeons regularly undertaking adrenal surgery will become aware of not infrequent ectopic adrenal tissue or adrenal rests which may exist along the path of testicular descent in males and near the broad ligament and uterus in females but which may only become apparent as a source of continued cortisol production after bilateral adrenalectomy or in rare adrenal malignancies found separate from the adrenal gland (Figure 5.2.2).

Both adrenal glands are enclosed in a layer of fat within Gerota's fascia that envelops the kidney and perinephric fat. The adrenals are supplied by three main groups of arteries: the inferior phrenic artery, the aorta, and the renal artery. In addition to these three regular sources of blood supply, pathological glands may have additional vessels due to neovascularization and on occasion also branches originating from the intercostal, the left ovarian, or the left internal spermatic arteries. Arterial branches ramify over the capsule before entering the gland and divide within to form a *subcapsular plexus* from which short cortical arteries give rise to an extensive network of sinusoidal capillaries in the interstices among clusters of cells. The venous drainage of the gland is usually into a single large suprarenal (central) vein on the right, which tends to be short and drains directly into the inferior vena cava on its posterior aspect. Control of this vessel can be hazardous. On the left side, the suprarenal vein is longer and is usually joined by the inferior phrenic vein, often described surgically as the *accessory adrenal vein*, before draining together into the left renal vein. The lymphatic drainage of the adrenal is irregular and profuse and ends in the lateral aortic lymph nodes and in the para-aortic nodes near the crus of the diaphragm and the origin of the renal artery. Given the absence of a defined lymphatic pathway radical surgical lymphadenectomy is of unproven value in adrenal malignancy. Some lymphatic vessels pierce the diaphragm and drain towards the thoracic duct or the posterior mediastinum, which explains the development of distant and early metastases of cortical malignant tumours.

Indications for Adrenal Surgery

Adrenalectomy is indicated for clinically significant functional tumours of the adrenal medulla or cortex and/or the presence of suspected or established primary or solitary secondary malignant tumours.

Of the functional tumours, **phaeochromocytoma** represents the commonest indication for adrenalectomy among surgeons submitting data to the British Association of Endocrine and Thyroid Surgery national registry, but with the growing diagnosis of primary hyperaldosteronism this may change in time. However, department of health Hospital Episode Statistics (HES) data for England suggests that surgery for apparently non-functioning adrenal tumours is particularly common among lower volume

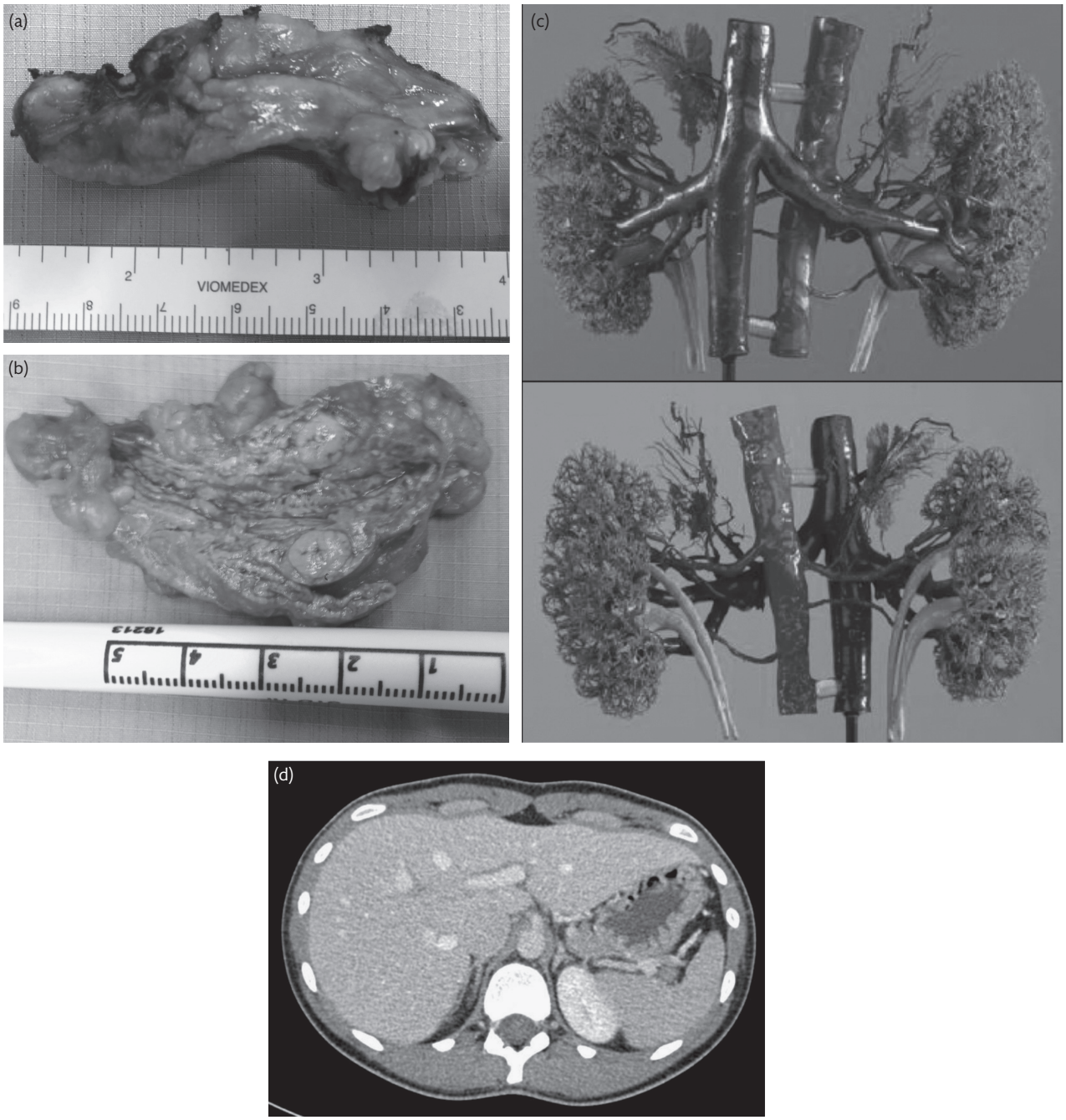


Figure 5.2.1 Macroscopic appearance of the adrenal gland.

surgeons. Bilateral adrenalectomy for familial cases with bilateral phaeochromocytomas tumours is uncommon (22 of 446 cases recorded on the BAETS database) and the *cortex sparing* or subtotal adrenalectomy is an option in such cases. Extra-adrenal chromaffin tumours (paragangliomas) represent approximately 10% of cases and may be treated in a similar fashion although their risk of malignancy is considerably higher than in adrenal phaeochromocytomas.

Cushing's syndrome or ACTH-independent hypercortisolism accounts for approximately 1 in 6 patients with endogenous hypercortisolism. Most patients with adrenal Cushing's syndrome have an underlying benign adrenocortical adenoma but the presence of mixed steroid secretion and combination with androgenization or feminization increase the index of suspicion for a functional adrenocortical carcinoma. Bilateral adrenalectomy is sometimes also considered in ACTH-independent hypercortisolism in patients with bilateral nodular hyperplasia or primary pigmented nodular adrenal disease (Figure 5.2.4). The role of surgery in patients with subclinical Cushing's syndrome is unclear and the current subject of a multicentre pan-European randomized controlled trial. The only randomized controlled trial (RCT) published to date was from a single centre and included only 45 patients recruited over a decade [4]. These results suggested that adrenalectomy resulted in significant improvement, not only of biochemical parameters, but also associated medical conditions such as diabetes, hypertension,

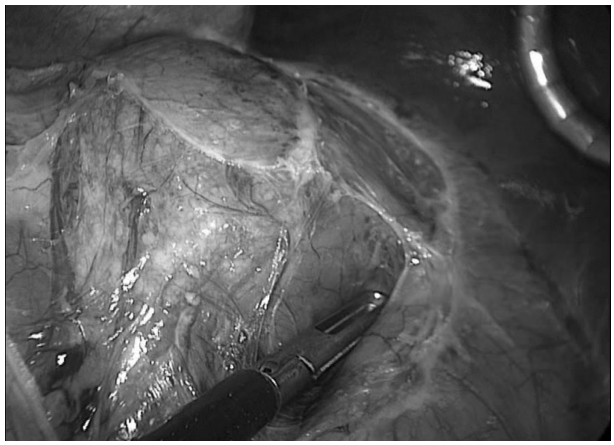


Figure 5.2.2 Anatomical view during a right laparoscopic adrenalectomy.

hyperlipidaemia, and obesity, which was supported by a systematic review [5] and a recent meta-analysis [6], albeit based on low-to-moderate quality evidence.

Cushing disease or **ACTH-dependent hypercortisolism secondary to unsuccessfully treated** pituitary tumours or paraneoplastic ectopic ACTH secretion may also require a bilateral adrenalectomy, especially when severe and when alternative treatment options have been attempted. The results of pituitary surgery have improved with centralization but medium-term recurrence remains a problem after trans-sphenoidal pituitary surgery even in the best hands [7, 8]. In both Cushing's syndrome and bilateral adrenalectomy for Cushing's disease, the hypercortisolaemia needs to be controlled for a prolonged preoperative period of at least one month to minimize the deleterious effects of excess glucocorticoids on perioperative morbidity. This can be achieved with metyrapone or ketoconazole. These patients also require steroid hormone replacement postoperatively since a cortisol secreting adrenal adenoma causes profound suppression of contralateral adrenal cortical activity as a result of inhibition of ACTH secretion, which may take up to 1 year to recover. The arrival of minimal access adrenal surgery has reduced the need for high perioperative steroid doses with physiological hydrocortisone replacement at 10/5/5 mg daily usually achieved within days. Patients that have undergone bilateral adrenal surgery require hydrocortisone as well as fludrocortisone for life.

Primary hyperaldosteronism is an increasingly frequent indication for adrenal surgery because of the realization that up to 10% of hypertensive patients may have primary hyperaldosteronism. Of these, up to 60% of have unilateral or dominant unilateral aldosterone hypersecretion that may benefit from adrenal surgery. Confirmation of lateralization still requires selective adrenal vein sampling (AVS) in all patients except younger patients with a clear adrenal adenoma. However, AVS requires interventional radiological skills that are not universally available; therefore, a reliable noninvasive lateralization imaging method would be welcome. It is hoped that the advent of ^{11}C-metomidate PET-CT or other modalities may fill this gap in the coming years, thus allowing more patients to benefit from intervention [9]. If patients are appropriately diagnosed with unilateral disease, 90% will benefit from the removal of the disease-causing adrenal, with as many as 50% coming off all hypertensive medication. The patients most likely to benefit are young and with a short history of hypertension and/or hypokalaemia. Another 40% achieve a reduction in their antihypertensive burden and lose the need for potassium supplements if present preoperatively [10].

Adrenal metastasectomy currently accounts for approximately one-tenth of the almost 800 adrenalectomies performed in England each year. Previously the presence of metastasis was treated as a terminal event but adrenal surgery is now considered a part of multimodality treatment in several malignancies. Lung cancer, colon cancer, renal cancer, melanoma are the most common causes of solitary adrenal metastases considered for surgery. Adrenal metastases should also be considered in the radiologically identified asymptomatic and non-functioning adrenal lesions in patients with a previous history of malignancy. Indeed, patients with a known history of cancer have a 50% chance of an incidentally discovered adrenal lesion being a metastasis, especially when the malignancy was recently diagnosed. The benefits of a solitary adrenal metastasectomy are contested and, therefore, decisions should be individualized in the context of risk-benefit analysis, especially if an R0 resection is unlikely [11].

A special category is represented by the unfortunately named adrenal 'incidentaloma'. Few diseases have been defined and named by their modality of discovery. **Adrenal incidentaloma** describes a clinically non-functioning adrenal lesion detected incidentally on imaging performed for reasons other than suspected adrenal disease. These lesions are usually solid and less frequently cystic and

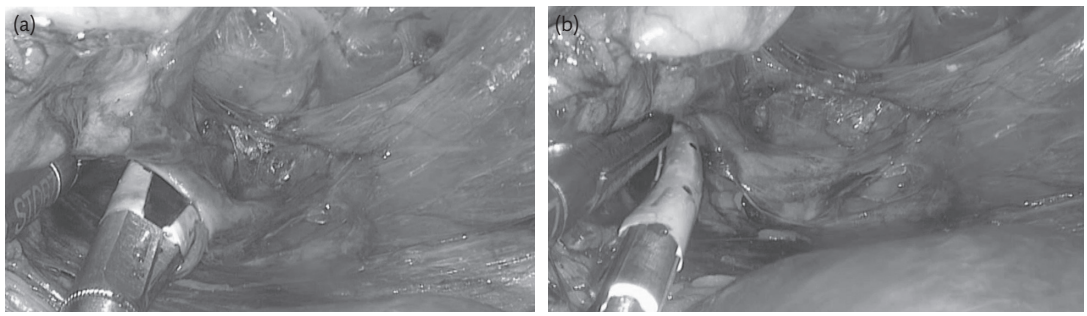


Figure 5.2.3 Anatomical view during a left retroperitoneoscopic adrenalectomy.

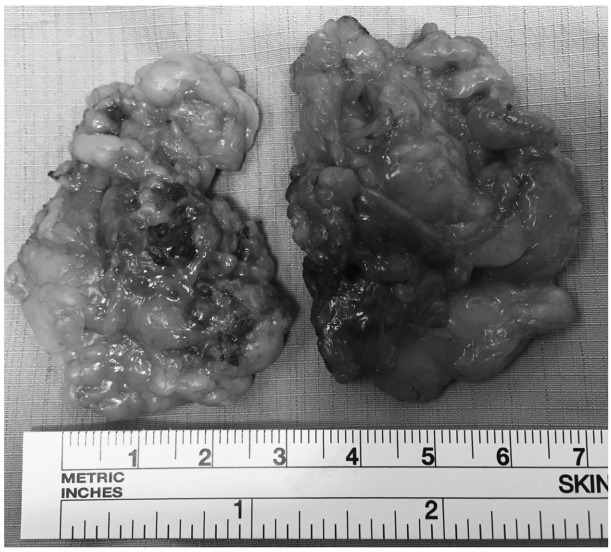


Figure 5.2.4 Bilateral pigmented micronodular hyperplasia.

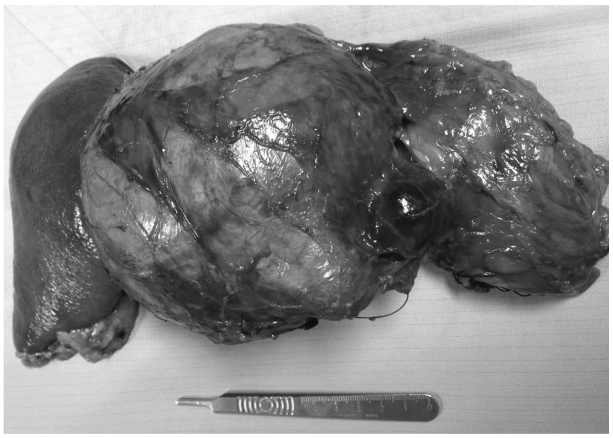


Figure 5.2.5 Large ACC resected with spleen, kidney, tail of pancreas.

the vast majority are benign. Computed tomography can usually aid in the definition of these lesions as benign, especially when of low density, indicating a high fat content, or the less frequent but more challenging radiological diagnosis of ganglioneuroma. Once identified, adrenal masses require pragmatic management with a functional endocrinological assessment. Functional adrenal adenomas are treated like any other functional adenoma. Most adenomas discovered incidentally are non-functioning and their subsequent management depends on their malignancy risk, with surgery considered only when the balance of risk and benefit shifts towards the risk of missing a malignancy. Data based on removed tumours demonstrates that the size of an adrenal growth appears to correlate with the risk of malignancy—less than 2% in lesions smaller than 4 cm in size but a significant climb in risk as the diameter increases [12]. Surgery is indicated for indeterminate lesions larger than 4 cm in size, but improvements in imaging may desirably reduce the amount of diagnostic adrenal surgery. Low density (less than 10 Hounsfield units on unenhanced CT) and well-defined MRI lesions are likely to be benign. A negative FDG PET scan may also permit a conservative approach, especially when used in combination with conventional imaging, and can allow patient discharge. The role of a multidisciplinary team (MDT) discussion to avoid overtreatment in such patients cannot be overemphasized.

Adrenal tumours presenting at a young age, those with rapid enlargement at any age and/or a picture of mixed hormonal secretion with virilizing or feminizing features represent those with the highest risk of being malignant. Surgery is the mainstay of treatment and the only potential cure for **adrenocortical cancer (ACC)**. Even with optimal care, ACC has a dismal prognosis and an overall 5-year survival of less than 30%. Locally advanced tumours require an en-bloc compartmental resection that mandates an open adrenalectomy approach (Figure 5.2.5). The complex perioperative decisions necessary for such patients have recently been reviewed in guidelines written by a collaboration between the European Society of Endocrine Surgeons (ESES) and European Network for the Study of Adrenal Tumours (ENSAT) [13]. Surgery may be considered in patients with locally advanced or recurrent tumours as resection may improve control of hormonal symptoms; however, recurrence within 12 months of primary surgery usually indicates a rapidly growing tumour that might suggest consideration of medical rather than surgical options. Non-surgical options such as radiofrequency ablation for local recurrence have been used with success [14] and may also be considered.

Minimal access adrenalectomy, whether laparoscopic or retroperitoneoscopic, is now the gold standard for benign tumours so long as tumour size does not affect the technical feasibility for a safe resection. There is no universally accepted maximum size since this will depend on the patient body size and habitus, the perceived malignancy risk, and the experience of the operating surgeon. The gold standard of open adrenalectomy for malignant adrenal tumours has in recent years been challenged in smaller tumours without local invasion where malignancy is suspected but not certain. The same applies to adrenal metastases from primaries elsewhere.

Surgical Preparation

In addition to the specific preparation described earlier, all patients undergoing adrenalectomy require a thorough preoperative clinical assessment to determine comorbidities that may need optimization prior to surgery. Hypertension is common in all adrenal hyperfunction syndromes and diabetes and ischaemic heart disease are frequent sequelae of long-standing cortisol excess. A full blood count, renal and liver function tests, clotting screen, and blood grouping are performed. For patients with large tumours/phaeochromocytoma, critical care facilities, while not always required, should be available. Patients with cortisol excess tend to have visceral obesity and have skin and musculoskeletal fragility such that special care is required to avoid soft tissue injury pre-, intra- and postoperatively. Single dose antibiotic prophylaxis and prophylaxis against venous thromboembolism with low molecular weight heparin (especially in patients with hypercortisolism) is given unless specifically contraindicated.

Particular emphasis is placed on the perioperative care of phaeochromocytoma patients that require pharmacological controlled vasodilatation with restoration of blood volume with oral fluid intake. In the United Kingdom, this is most typically achieved with a non-competitive α-blocker such as phenoxybenzamine, with a β-blocker such as propranolol to control the tachycardia that may follow. Alternatives include a selective α1-blocker such as doxazosin to reduce the side effects such as nasal stuffiness, postural hypotension, and somnolence that may occur with phenoxybenzamine. Calcium channel blockade is another alternative that is widely used in mainland Europe. Whichever pharmacological approach is adopted, preoperative hospital admission for at least 24 hours prior to surgery to optimize blockade is preferable. Ideally patients can achieve a systolic blood pressure of ≤120 mmHg with a postural drop without a postural tachycardia. Despite optimal blockade some intraoperative cardiovascular instability is commonplace and requires expert anaesthesia. Postoperatively, close observation is also required since postoperative hypotension requiring treatment beyond i.v. fluids may be required. This should not continue beyond several hours and ongoing hypotension in an ideally prepared patient necessitates the investigation for other causes of the hypotension. Postoperative hypertension is rarer still, and may suggest undetected bilateral or extra-adrenal disease. Hypoglycaemia may occasionally occur in diabetics, as a result of the withdrawal of the anti-insulin effects of catecholamines.

The process of informed consent of patients undergoing adrenal surgery requires careful explanation so that the patient is fully aware of indications, benefits, consequences, and risks of surgery quantified based on the surgeon's personal practice. A clear explanation of the alternatives to surgery where present is mandatory.

Surgical Techniques

Open Anterior Approach

Minimal access approaches—laparoscopic and retroperitoneoscopic—to the adrenal are now the gold standard approach so a direct adoption of an open approach is reserved for those cases where minimal access surgery is inappropriate. Currently open adrenalectomy is reserved for tumours that are technically too large to remove safely via minimal access approaches and those with characteristics suggestive of ACC. There is no absolute size criterion for the technical limitation of a laparoscopic approach since this will depend also on the size of the patient, the consistence of the tumour and the experience of the surgeon.

The open transperitoneal approach invariably involves a long incision that may be a curved transverse incision ('roof top') or unilateral subcostal incision extended to the flank. Other options include a vertical usually midline incision.

Left adrenalectomy for large left adrenal tumours requires mobilization of parietal peritoneum lateral to the left colon, continued upward, dividing the splenorenal ligament up to the diaphragm to allow the medial rotation of the spleen, tail of the pancreas and colonic flexure (Figure 5.2.5). Alternative potential routes to consider are through the gastrocolic omentum, opening the lesser sac (without injury of the gastroepiploic arcade) or through the left mesocolon but these are options in smaller tumours only.

Right adrenalectomy requires the mobilization of the hepatic flexure of the colon and the right lobe of the liver is carefully retracted upward and rotated medially. This is achieved via division of the falciform ligament and the right triangular ligaments. The duodenum is mobilized in its second portion (Kocher's manoeuvre) by incision on its lateral avascular peritoneal reflection, allowing exposure of the inferior vena cava, the right adrenal gland, and the upper pole of the right kidney. The critical step is the clamping of the right adrenal vein because it is short, leaves the gland on its anterior aspect, and enters the vena cava on its posterior surface.

Complex multidisciplinary operations. Malignant tumours invading the IVC (Figure 5.2.6) create operative challenges that are beyond the scope of this chapter. There is concern that such patients could be labelled as inoperable in centres where there has been no previous experience with such operations and where a collaboration with cardiac or liver surgeons is not feasible.

Laparoscopic Transperitoneal Adrenalectomy

Right laparoscopic adrenalectomy is achieved in the same way as an open adrenalectomy but via four ports of 10 mm and 5 mm under the right rib cage. The hepatic mobilization is performed in the same way, but without the need for the division of the falciform ligament. The peritoneum is divided under the liver edge—the right coronary ligament—and along the inferior vena cava (IVC). A 'groove' is created between the IVC and the adrenal gland, which is mobilized laterally (Figure 5.2.2). In this groove, the right adrenal vein can be identified, clipped, and divided. The lower border of the adrenal gland is dissected off the renal pole.

Left laparoscopic adrenalectomy is performed usually with three ports inserted below the left rib cage. The descending colon is mobilized by dividing the peritoneal reflection and identifying the plane in front of the Gerota's fascia. This plane facilitates mobilization of the splenic flexure and the division of the spleno-colic ligament. The spleen is mobilized by dividing the lateral peritoneal reflection with care shown to avoid injury to the diaphragm or the short gastric vessels, which can be encountered at the superior pole of the spleen. Gerota's fascia is incised vertically and the dissection aims to create a 'groove' with the spleen and the tail of the pancreas medially and with the adrenal and the perinephric fat laterally. A small left adrenal gland might not become apparent until the posterior aspect of the kidney is demonstrated.

Retroperitoneoscopic Adrenalectomy

This technique allows direct access to the adrenal gland via the retroperitoneum. The patient is placed in a prone position with hip flexion. Three ports are inserted using the twelfth rib as a landmark. A space is created using high-pressure insufflation of 20 mmHg (or more). The upper pole of the kidney is identified, and the anterior aspect of the kidney is mobilized. On the left the dissection continues towards the left renal vein to allow the identification of the left adrenal vein and from this landmark the dissection continues upwards to mobilize the entire adrenal gland (Figure 5.2.3). On the right the aim is to demonstrate the IVC and to identify the short right adrenal vein. Retroperitoneoscopic adrenalectomy unquestionably presents a challenging learning curve due to the need to familiarize with

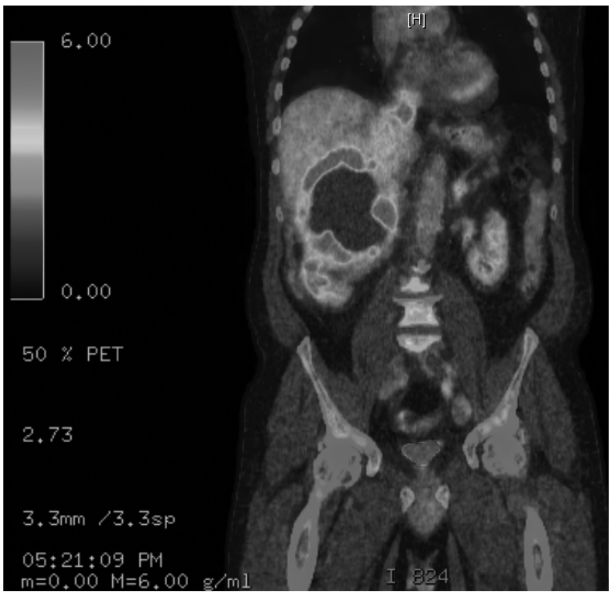


Figure 5.2.6 PET scan of patient with adrenocortical cancer extending to the IVC and the right atrium.

anatomical relationships and landmarks seen from another perspective due to the approach from the posterior abdominal wall.

Summary

Adrenal surgery is required for the treatment for syndromes of hormonal excess arising from the adrenal gland and adrenal tumours or neoplasms with malignant potential. A multidisciplinary approach with protocolled perioperative planning, and an experienced surgeon are essential for good outcomes. Most of the controversies in adrenal surgery have been resolved but some areas remain controversial. These include:

- the appropriateness of laparoscopic surgery when malignancy is likely (but without adrenal malignancy)
- the role of surgery in subclinical Cushing's syndrome
- the use of cortical sparing/subtotal adrenalectomy
- the role of adrenalectomy in patients with adrenal metastasis
- the centralization of services based on the volume-outcome data

The rarity of adrenal disease, the heterogeneity of clinical conditions, and the multifactorial influences on outcome makes it unlikely that surgical decisions can be based on evidence from randomized trials. The best available quality data from observational series and the insight and judgement of experienced endocrine surgeons and multidisciplinary working represents the best acceptable alternative.

REFERENCES

1. Palazzo F, Dickinson A, Phillips B, *et al.* Adrenal surgery in England: better outcomes in high-volume practices. *Clin Endocrinol (Oxf)*, 2016; **85**: 17–20.
2. Anderson KL Jr, Thomas SM, Adam MA, *et al.* Each procedure matters: threshold for surgeon volume to minimize complications and decrease cost associated with adrenalectomy. *Surgery*, 2018; **163**: 157–64.
3. Welbourn RB. *The History of Endocrine Surgery*. New York: Praeger, 1990.
4. Toniato A, Merante-Boschin I, Opocher G, Pelizzo MR, Schiavi F, Ballotta E. Surgical versus conservative management for subclinical Cushing syndrome in adrenalincidentalomas: a prospective randomized study. *Ann Surg*, 2009; **249**: 388–91.
5. Iacobone M, Citton M, Scarpa M, Viel G, Boscaro M, Nitti D. Systematic review of surgical treatment of subclinical Cushing's syndrome. *Br J Surg*, 2015; **102**: 318–30.
6. Bancos I, Alahdab F, Crowley RK, *et al.* Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and subclinical Cushing's syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R283–95.
7. Porterfield JR, Thompson GB, Young WF, Jr., *et al.* Surgery for Cushing's syndrome: an historical review and recent ten-year experience. *World J Surg*, 2008; **32**: 659–77.7
8. Morris LF, Harris RS, Milton DR, *et al.* Impact and timing of bilateral adrenalectomy for refractory adrenocorticotropic hormone-dependent Cushing's syndrome. *Surgery*, 2013; **154**: 1174–83; discussion 1183–4.
9. Stowasser M. Update in primary aldosteronism. *J Clin Endocrinol Metab*, 2015; **100**: 1–10
10. Funder JW, Carey RM, Mantero F, *et al.* The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 1889–916.
11. Huang SH, Kong QL, Chen XX, He JY, Qin J, Chen ZG. Adrenalectomy does not improve survival rates of patients with solitary adrenal metastasis from non-small cell lung cancer. *Ther Clin Risk Manag*, 2017; **13**: 355–60.
12. Bilimoria KY, Shen WT, Elaraj D, *et al.* Adrenocortical carcinoma in the United States: treatment utilization and prognostic factors. *Cancer*, 2008; **113**: 3130–6.
13. Gaujoux S, Mihai R; joint working group of ESES and ENSAT. European Society of Endocrine Surgeons (ESES) and European Network for the Study of Adrenal Tumours (ENSAT) recommendations for the surgical management of adrenocortical carcinoma. *Br J Surg*, 2017; **104**:358–76.
14. McDuffie LA, Aufforth RD. Adrenocortical carcinoma: modern management and evolving treatment strategies. *Int J Endocr Oncol*, 2016; **3**: 161–74.

5.3

Adrenal Incidentaloma

Irina Bancos, Massimo Terzolo, and Wiebke Arlt

Introduction

Adrenal incidentalomas are adrenal tumours discovered incidentally on around 5% of cross-sectional imaging studies performed for indications unrelated to adrenal disease [1, 2]. The incidental discovery of an adrenal mass has evolved into a common problem, because of the increasingly widespread use of cross-sectional imaging in current clinical practice. Over the last two decades, the number of abdominal computed tomography (CT) and magnetic resonance imaging (MRI) imaging studies performed for various reasons quadrupled, and increasing resolution of these imaging studies allows for detection of ever smaller tumours, explaining the growing numbers of patients with adrenal incidentaloma [3]. Any patient with a newly detected adrenal mass requires further evaluation to assess: (1) whether the adrenal mass is functioning (i.e. overproducing adrenal hormones), and (2) whether the adrenal mass is malignant or benign.

Epidemiology

In autopsy series, the mean prevalence of clinically inapparent adrenal masses is around 2.0%, ranging from 1.0% to 8.7% [4]. This variability reflects different definitions and also the difficulty in distinguishing larger nodules within adrenal hyperplasia from distinct adrenocortical adenomas. Large radiological series reported a 4.4–5.0% prevalence of adrenal tumours [1, 2]. In patients with a history of extra-adrenal malignancy the prevalence of adrenal tumours can be as high as 50–75%, as the adrenal gland is frequently targeted by metastatic spread [5, 6]. Adrenal metastases originate most frequently from lung cancer, kidney cancer, gastrointestinal cancer, melanoma, and lymphoma [7–10].

Adrenal tumours demonstrate different pathology distribution in the population dependent on the patient's age and sex. In clinical reports, adrenal incidentalomas show a peak incidence between 50 and 60 years of age [4]. This pattern could merely reflect a higher number of diagnostic procedures in these age decades or be the consequence of the ageing process of the adrenal glands, which may lead to increased formation of cortical nodules secondary to vascular changes [4]. The frequency of adrenal incidentalomas is very low in childhood and adolescence (0.3–0.4% of all neoplasms in children) [11].

While previous studies reported a higher proportion of women with adrenal tumours, possibly explained by a referral bias (i.e. more imaging studies are done in women due to higher prevalence of biliary disease), more recent studies describe a roughly equal sex distribution in patients with adrenal tumours overall, though higher in certain adrenal tumour subtypes (such as female predominance in patients with adrenocortical carcinoma and male predominance in patients with other adrenal malignancies) [1, 4, 12, 13]. Adrenocortical adenomas are more frequent on the left side and can be bilateral in 14–18% of cases, while no significant lateralization is observed in pheochromocytomas, adrenocortical carcinomas and other benign and malignant adrenal tumours [12, 14]. Adrenal metastases may affect both adrenal glands in 22–42% of patients [8, 10, 12], however bilateral adrenal involvement is less frequent in patients with pheochromocytoma (8–10%), other benign tumours (10%), and almost never observed in patients with adrenocortical carcinomas [12, 15].

Aetiology

Aetiology of adrenal incidentalomas includes both benign and malignant lesions. However, an adrenal incidentaloma is most likely benign, representing an adrenal adenoma in approximately 80–90% of cases (of which cortisol-secreting adenomas are seen in 1.0–2.9%

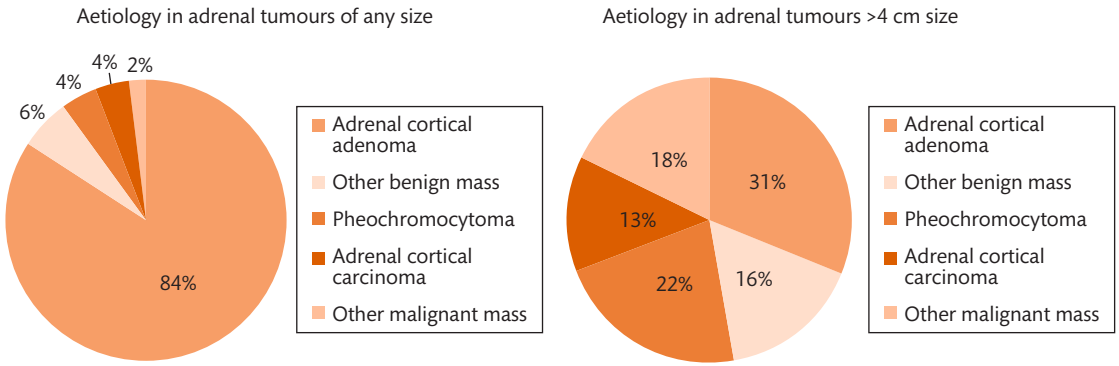


Figure 5.3.1 Distribution of diagnoses among patients with adrenal tumours. (Adapted from [2, 12, 16, 17, 21, 22]).

and aldosterone-secreting adenomas in 1.6–2.3%) [16, 17]. The frequency of pheochromocytoma is estimated to be 1.5–23%, that of adrenocortical carcinoma varies from 1.2% to 11% [16, 17]. It is important to note that despite the fact that the majority of adrenal incidentalomas are adrenocortical adenomas, 40–60% of adrenocortical carcinomas, pheochromocytomas, and adenomas with biochemically overt hormonal excess are diagnosed incidentally [12, 15]. The risk of an adrenal incidentaloma being an adrenal cancer is proportional to tumour size, especially in patients with adrenocortical lesions, as most adrenocortical carcinomas are larger than 4–6 cm (median size 10 cm), while most adrenocortical adenomas are smaller than 4 cm (median size of 1–3 cm) [2, 12, 14, 17, 18]. Other benign causes of adrenal incidentaloma are adrenal cysts, ganglioneuromas, myelolipomas, haematomas, while malignant causes include metastases of other malignancies, sarcomas, and primary adrenal lymphomas. Certain genetic syndromes predispose to an increased likelihood of developing an adrenal mass. Pheochromocytomas are common in patients with multiple endocrine neoplasia type 2, Von Hippel Lindau syndrome, Neurofibromatosis type 1, and Succinate dehydrogenase mutations [15, 19, 20] (see Chapter 5.5.1). Moreover, adrenal lesions are found in inherited endocrine cancer syndromes (McCune–Albright syndrome, multiple endocrine neoplasia type 1), and in suboptimally controlled congenital adrenal hyperplasia.

In a multi-institutional, retrospective survey performed in Italy including 1004 patients, of whom 380 underwent surgery, the most frequent pathological diagnoses were adrenocortical adenoma (52%), adrenocortical carcinoma (12%), phaeochromocytoma (11%), and myelolipoma (8%) (6). However, in more recent cohorts, lower prevalence of malignancy and pheochromocytoma was noted [12, 21] (Figure 5.3.1). Distribution of adrenal pathologies differs based on tumour size. For example, in a large retrospective single centre consecutive series of 705 patients with large adrenal tumours (>4 cm), 31% were malignant tumours (ACC—18%, other malignant tumours—13%), 22% pheochromocytomas, and 47% benign adrenal tumours [12].

Differential Diagnosis

The differential diagnosis of adrenal masses should take several key factors in consideration, including demographics, mode of discovery, clinical, and biochemical presentation, presence of any genetic predisposition syndrome, as well as imaging characteristics (Table 5.3.2). While prevalence of adrenal tumours in children is <0.5% [11, 17], risk of malignancy is high, with up to 60–80% represented by adrenocortical carcinomas [11]. Overall, adrenocortical carcinoma represents 1.3% of all malignancies in patients younger than 20 years of age [4]. Adrenal tumours discovered incidentally are less likely to be malignant or hormonally active in comparison to patients presenting with symptoms suggestive of hormone excess or undergoing imaging for the purposes of cancer staging [2, 8, 12]. However, malignant tumours and pheochromocytomas are not infrequently discovered incidentally, as was reported in 42–54% of large tumours, reflecting the non-specific of symptoms and/or subclinical presentation of adrenal hormonal excess [12]. See Table 5.3.1 and Figure 5.3.2.

Risk of malignancy increases with the size of the adrenal mass [12, 22–24]. In a multicentre study of 1096 patient with incidentally discovered adrenal tumours, sensitivity of tumour size cut-off of >4 cm to diagnose adrenocortical carcinoma was 93%, though specificity was only 42%. A higher cut-off of 6 cm improved specificity to 71–73%, but at a cost of lower sensitivity of 61–74% [12, 24]. In a large retrospective cohort of patient with adrenal tumours >4 cm, prevalence of malignancy was reported to be 31% (18% adrenocortical carcinoma and 13% other malignant tumours). Similarly, 20% of patients with adrenal tumours >4 cm were found to have adrenocortical carcinomas in a prospective cohort of patients [21].

Radiological Assessment

In addition to tumour size, imaging characteristics are helpful in diagnosis of adrenal malignancy. Unenhanced CT is the most commonly used in evaluation of adrenal tumours. Unenhanced attenuation values of less than 10 Hounsfield Units (HU) indicates a lipid-rich benign adrenocortical adenoma, while HU values >10HU, and especially >20HU are suspicious of malignancy. HU values can only be determined in homogeneous tumours, thus, heterogeneous tumour appearance on CT also has to be considered suspicious. CT washout studies do not have a sufficient evidence base for reliable use in clinical practice, as demonstrated in a recent meta-analysis [13]. MRI also evaluates tumour fat content, similarly to CT, where the loss of signal intensity between in-phase

Table 5.3.1 Key points on the differential diagnosis of adrenal incidentalomas: consider history, demographics, clinical presentation, adrenal mass imaging phenotype, and any presence of genetic predisposition syndromes

• History of extra-adrenal malignancy/active extra-adrenal malignancy	→ higher likelihood of adrenal metastasis
• Genetic syndromes	→ higher likelihood of pheochromocytoma (e.g. VHL, NF-1, MEN, SDH) or adrenocortical carcinoma (e.g. TP53, Lynch syndrome)
• Young age	→ higher likelihood of adrenocortical carcinoma
• Signs and symptoms of adrenal hormonal excess	→ adrenocortical adenoma, carcinoma, or pheochromocytoma, based on type of hormonal excess
• Bilateral adrenal masses	→ higher likelihood of bilateral metastases, bilateral adrenocortical adenomas, macronodular hyperplasia, or bilateral pheochromocytomas → extremely unlikely to be adrenocortical carcinoma
• Large tumour size (>4–6 cm)	→ higher likelihood of adrenocortical carcinoma, sarcoma, lymphoma, pheochromocytoma, as well as other benign tumours, such as myelolipomas
• Imaging characteristics: unenhanced CT attenuation <10 HU, or present chemical shift on MRI	→ extremely unlikely to be pheochromocytoma, adrenocortical carcinoma or any kind of malignancy, very likely to be lipid-rich adenoma or myelolipoma,
• Imaging characteristics: unenhanced CT attenuation 10–20 HU	→ low likelihood of malignancy and pheochromocytoma, likely a cortical adenoma
• Imaging characteristics: unenhanced CT attenuation >20 HU or absent chemical shift on MRI	→ higher likelihood of pheochromocytoma, adrenocortical carcinoma or other malignancy

and out-of-phase T2-weighted sequences (chemical shift analysis; also see Chapter 5.1) indicates a lipid-rich adrenal tumour. If an adrenal mass is considered indeterminate on CT or MRI, Fluorodeoxyglucose positron emission tomography (FDG-PET) can be considered in patients with and without a history of extra-adrenal malignancy. FDG-PET/CT offers superior specificity (84% for an adrenal-liver-ratio (ALR) >1.8 of the maximum standardized uptake value, SUVmax) for diagnosis of malignant adrenal mass [25–27]; however, at present, this is based on limited data. Additional imaging features such as irregular borders, heterogeneity of the adrenal mass, signs of local invasion, or evidence of rapid growth indicate underlying adrenal malignancy. See Table 5.3.3.

Adrenal Biopsy

Guidelines on the management of adrenal incidentalomas developed by the European Society of Endocrinology (ESE) and the

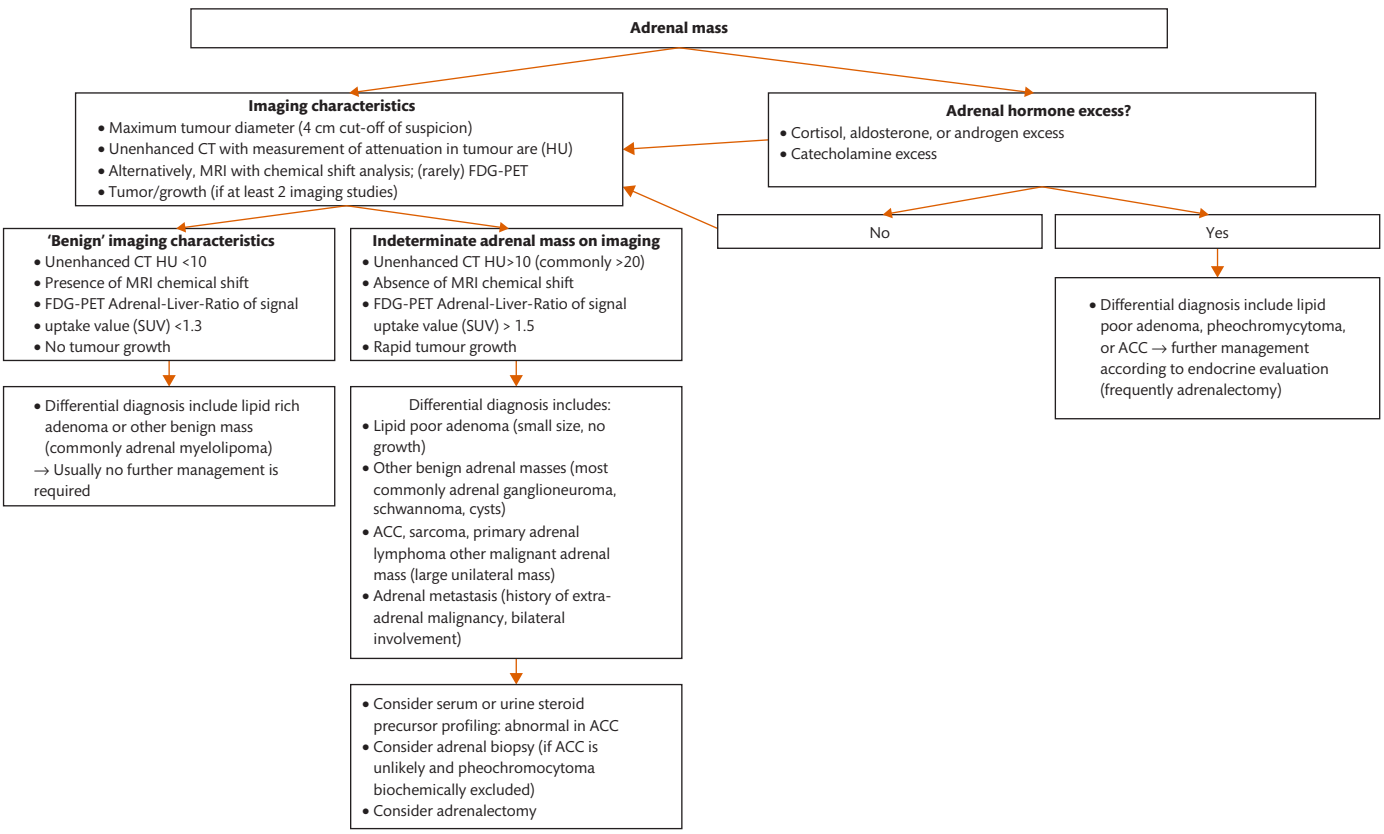


Figure 5.3.2 CT images of (a) adrenocortical carcinoma, (b) phaeochromocytoma, and (c) metastases from extra-adrenal cancer.

Table 5.3.2 Differential diagnosis of adrenal tumours

	Adrenocortical adenoma	Other benign tumour	Pheochromocytoma	Adrenocortical carcinoma	Other malignant tumours
Age, median, years	50–60	Wide spectrum	Wide spectrum	40–50	60–70
Women, %	50–60%	50%	50%	65%	35%
Genetic predisposition syndrome	Very rare	Very rare	Up to 40%	Rare	Very rare
Incidental mode of discovery	85%	90%	50–60%	40%	20–50%
Tumour size	Usually <4 cm	Wide spectrum	Usually >4 cm	Usually >6 cm	Wide spectrum
Unenhanced attenuation	Most <20 HU	Most <20 HU	>20 HU	>20 HU	>20 HU
Bilateral	15–20%	10%	10%	<1%	20–42%
Hormone excess	10–50%	0%	>95%	50%	0%

(Adapted from [7, 8, 10, 11, 12, 15–17, 23–25]).

European Network for the Study of Adrenal Tumors (ENSAT) in 2016 recommend adrenal biopsy only for patients with an indeterminate adrenal mass and a high suspicion for metastasis, infection, or another infiltrative process [16, 28]. The adrenal glands are common sites of metastases, mainly from lung, kidney, gastrointestinal, breast cancers, but also lymphomas and melanomas [10, 29]. While only 2–13% of adrenal incidentalomas represent metastases, the risk is 28–75% when patients have a history of and/or active extra-adrenal malignancy [5, 6, 12, 30–32]. In situations where an adrenal mass remains indeterminate after initial imaging, adrenal biopsy can be rarely considered for diagnostic purposes, especially in patients at high risk for metastatic disease [13, 33] (i.e. patients who underwent imaging as part of cancer staging or follow-up monitoring).

Adrenal biopsy is an invasive procedure with a 5% non-diagnostic rate. In addition, it is important to consider the risk of adrenal biopsy complications that was reported in 0 to 13%, likely dependent on location, technique, and experience [7–9, 34].

It is mandatory to exclude pheochromocytoma before proceeding with adrenal biopsy, even in the presence of extra-adrenal malignancy. Pheochromocytomas present with indeterminate imaging characteristics, similar to malignant adrenal tumours [23, 35, 36] and can present incidentally, without symptoms suggestive of catecholamine excess [15, 37]. Biopsy of pheochromocytoma may lead to hypertensive crisis, haemorrhage, and even death [37–40].

Guidelines do not recommend adrenal biopsy for suspected adrenocortical carcinoma in a patient with a unilateral, large adrenal mass, especially when evidence of adrenal hyperfunction is present.

Table 5.3.3 Diagnostic accuracy of commonly employed adrenal imaging studies

Diagnostic test	Sensitivity	Specificity
Tumour size >4 cm	80–93%	34–61%
Tumour size >6 cm	61–74%	71–73%
Unenhanced CT (HU >10)	93–100%	33–72%
Unenhanced CT (HU >20)	96–98%	64–70%
MRI, loss of signal intensity	86–99%	85–93%
PET ALR (adrenal to liver ratio) SUV max	87–100%	84–96%
PET SUV max	87–93%	69–73%

(Adapted from [7, 12, 13, 16, 26]).

There are several reasons for this recommendation, including suboptimal diagnostic performance in adrenocortical tumours (with pathologists often struggling to make a diagnosis of ACC on the whole tumour specimen) and risk of the needle track tumour dissemination [15, 33, 41–44]. Biopsy performs best in confirming metastasis of an extra-adrenal tumour, in particular if histology of the primary is available.

Hormonal Assessment

Around 15% of adrenal tumours present with adrenal hormone excess including clinically pheochromocytoma, overt Cushing syndrome, primary hyperaldosteronism and, uncommonly, adrenal androgen excess [4, 24]. Every patient with adrenal mass should undergo clinical and biochemical assessment for adrenal hormone excess (Table 5.3.4). Independent of clinical presentation, work-up should include diagnostic tests for Cushing syndrome and pheochromocytoma, while work-up for primary hyperaldosteronism should be performed in all patients with hypertension, whether hypokalaemia is present or not. In addition, presentation with hirsutism or virilization in a woman with adrenal mass is suspicious for adrenocortical carcinoma, and should prompt work-up for androgen excess. In very rare instances, isolated adrenal androgen excess can be seen in the context of a benign adrenal adenoma, but in most cases an adrenocortical carcinoma is the root cause, which can be taken as proven if androgen excess is combined with overproduction of other steroid hormones, such as cortisol, or also aldosterone or oestradiol.

Table 5.3.4 Initial work-up for adrenal hormone excess

Type of hormone excess	Initial diagnostic work-up
Cortisol excess (Cushing syndrome or MACS)	1 mg dexamethasone suppression test, consider 24 h urine cortisol, plasma ACTH, midnight serum, or salivary cortisol
Primary aldosteronism (in patients with hypertension)	Paired plasma renin and plasma aldosterone, serum potassium
Catecholamine excess	Plasma or 24 h urine metanephrines
Androgen excess (in patients with hirsutism or virilization)	Serum DHEAS, androstenedione, testosterone

Steroid profiling has recently emerged as a promising tool for differential diagnosis of adrenocortical carcinoma [21, 45]. In a proof-of-concept study published in 2011, authors reported on urinary steroid profiling by gas chromatography-mass spectrometry (GC-MS) combined with computational data analysis utilizing a machine learning-based approach [46]. This study revealed a distinct steroid pattern that diagnosed an adrenocortical carcinoma with 90% sensitivity and specificity. The three most informative steroid metabolites included the glucocorticoid precursor metabolite tetrahydro-11-deoxycortisol (THS) derived from 11-deoxycortisol, as well as pregnanediol (5-PD) and pregnanetriol (5-PT), derived from pregnenolone and 17-hydroxypregnenolone, respectively. Recently, a large prospective international multicentre test validation study (EURINE-ACT) carried out within the European Network for the Study of Adrenal Tumours (ENSAT) has been successfully completed and demonstrated that urine steroid metabolomics has superior accuracy to commonly used imaging tests in the detection of adrenocortical carcinoma [45]. Once introduced into clinical practice, steroid profiling will allow a noninvasive diagnosis of adrenocortical carcinoma in patients with indeterminate adrenal masses.

Mild Autonomous Cortisol Secretion (Previously Known as Subclinical Cushing's Syndrome)

Mild autonomous cortisol secretion (MACS, or also mild autonomous cortisol excess, MACE) is a syndrome characterized by a spectrum of hypothalamic-pituitary-adrenal axis abnormalities in a patient with adrenal mass presenting without clinically overt features of Cushing's syndrome [16]. The diagnosis of MACS represents a challenge due to limitations of the currently used diagnostic tests, differences in the definitions, and unclear clinical implications of MACS. Multiple definitions have been proposed for diagnosis of MACS [47]. These include dexamethasone suppression tests, 24 h urine cortisol measurements, midnight salivary and serum cortisol measurements, morning and afternoon cortisol measurements to assess cortisol secretory circadian rhythm, measurements of adrenocorticotrophic hormone (ACTH) (corticotropin), and dehydroepiandrosterone sulphate (DHEAS). The recent ESE-ENSAT guidelines on the management of adrenal incidentalomas suggested using the 1 mg overnight dexamethasone suppression test as a first step, with 'possible' MACS diagnosed when post-dexamethasone serum cortisol is >50 nmol/L and 'confirmed' MACS if cortisol is >140 nmol/L [16], in the absence of clinically overt signs of Cushing's syndrome. Guidelines further suggest that low or suppressed plasma ACTH and DHEAS concentrations can help confirm MACS [16, 48].

The reported prevalence of MACS among patients with adrenal incidentaloma ranges from 5–50% [16, 17, 49, 50]. This heterogeneity is explained, at least in part, by the different work-up protocols and variable criteria used to define MACS as well as in different inclusion criteria and probably referral bias of the reported series [7].

MACS very rarely progresses towards overt Cushing syndrome, however, patients with MACS present with increased prevalence of cardiovascular comorbidities, cardiovascular events, osteopaenia, osteoporosis, and fractures, as well as increased mortality [16, 25, 47, 51–57]. Rates of hypertension, prediabetes, or diabetes mellitus type 2, dyslipidaemia, obesity, and fractures have been reported to be higher than in controls or patients with non-functioning adrenal tumours [24, 47, 58]. Limited data on the effectiveness of adrenalectomy suggests that this leads to improvement of cardiovascular risk factors and decrease in fractures in patients with MACS, especially when the degree of cortisol abnormality is higher [25, 48, 57]. However, identifying patients most likely to benefit is still challenging and many patients with MACS are age-advanced and have multiple comorbidities, which increases the potential risk of surgery. At present, approach to patients with MACS remains both a diagnostic and management challenge. Diagnostic criteria for MACS are summarized in **Box 5.3.1**, and the major areas of uncertainty are summarized in **Box 5.3.2**.

Box 5.3.1 Diagnostic criteria for MACS

1. Incidentally discovered adrenocortical adenoma
2. Lack of features suggestive of overt Cushing syndrome (moon facies, proximal myopathy, striae, abdominal fat redistribution)
3. Abnormal cortisol after 1 mg overnight dexamethasone suppression test (>50 nmol/L), low ACTH, and DHEAS

Natural History and Management

Management of adrenal incidentaloma is a complex decision-making process, which involves considering a range of possible diagnoses as opposed to respective natural history, and weighing the risks and benefits of interventions in light of the patient's age and the tumour size and imaging characteristics. Surgery is the appropriate therapeutic measure for ACC, pheochromocytoma, and other functional adrenal tumours causing overt glucocorticoid, mineralocorticoid, or adrenal sex hormone excess; treatment of metastasis depends on individual clinical circumstances and the nature of the primary extra-adrenal cancer.

The available follow-up data of patients with clinically inapparent adrenal mass suggests that the large majority of adrenal lesions classified as benign at diagnosis remain stable over time [16, 24, 25]. The risk of malignant transformation at long-term follow-up is very low, estimated at 0 to <1:1000 incidentalomas [16, 24, 25]. Tumour growth or decrease in size over time may occur in benign tumours (around 5%) [4, 16, 24, 25]. Guidelines for management of adrenal incidentaloma suggest against imaging follow-up for adrenal tumours <4 cm and clear benign imaging characteristics, while in patient with an indeterminate adrenal mass, options include (1) adrenalectomy or (2) either another imaging by a different modality or repeat imaging after 6–12 months [16]. Similarly, guidelines recommend against repeated hormonal work-up in patients with normal initial evaluation unless new signs suggestive

Box 5.3.2 Unsolved issues with subclinical Cushing's syndrome

- Individualized diagnostic criteria to diagnose clinically relevant MACS?
- Identification of patients at risk of developing comorbidities related to MACS?
- Identification of patients with MACS to benefit from adrenalectomy prior to development of MACS-related comorbidities

of adrenal hormonal excess develop [16]. Management of patients with MACS requires an individual approach based on the degree of hypercortisolism and presence of comorbidities, with options including adrenalectomy or annual clinical reassessment for cortisol excess [16], with medical strategies for reducing cortisol excess currently undergoing evaluation by research studies.

REFERENCES

1. Bovio S, Cataldi A, Reimondo G, *et al.* Prevalence of adrenal incidentaloma in a contemporary computerized tomography series. *J Endocrinol Invest*, 2006; **29**: 298–302.
2. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *AJR Am J Roentgenol*, 2008; **190**: 1163–8.
3. OECD. Computed tomography (CT) exams (indicator), 2017. Available at: https://data.oecd.org/healthcare/computed-tomography-ct-exams.htm (accessed: 20 March 2010).
4. Barzon L, Sonino N, Fallo F, Palu G, Boscaro M. Prevalence and natural history of adrenal incidentalomas. *Eur J Endocrinol*, 2003; **149**: 273–85.
5. Frilling A, Tecklenborg K, Weber F, *et al.* Importance of adrenal incidentaloma in patients with a history of malignancy. *Surgery*, 2004; **136**: 1289–96.
6. Lenert JT, Barnett CC Jr, Kudelka AP, *et al.* Evaluation and surgical resection of adrenal masses in patients with a history of extra-adrenal malignancy. *Surgery*, 2001; **130**: 1060–7.
7. Bancos I, Tamhane S, Shah M, *et al.* Diagnosis of endocrine disease: the diagnostic performance of adrenal biopsy: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R65–80.
8. Delivanis DA, Erickson D, Atwell TD, *et al.* Procedural and clinical outcomes of percutaneous adrenal biopsy in a high-risk population for adrenal malignancy. *Clin Endocrinol (Oxf)*, 2016; **85**: 710–16.
9. Zhang CD, Delivanis DA, Eiken PW, Atwell TD, Bancos I. Adrenal biopsy: performance and use. *Minerva Endocrinol*, 2019; **44**: 288–300.
10. Mao JJ, Dages KN, Suresh M, Bancos I. Presentation, Disease Progression, and Outcomes of Adrenal Gland Metastases. *Clin Endocrinol (Oxf)*, 2020; Jun 22. doi: 10.1111/cen.14268. Online ahead of print. PMID: 32569405.
11. Ciftci AO, Senocak ME, Tanyel FC, Büyükpamukçu N. Adrenocortical tumors in children. *J Pediatr Surg*, 2001; **36**: 549–54.
12. Iñiguez-Ariza NM, Kohlenberg JD, Delivanis DA, *et al.* Clinical, biochemical, and radiological characteristics of a single-center retrospective cohort of 705 large adrenal tumors. *Mayo Clin Proc Innov Qual Outcomes*, 2018; **2**: 30–9.
13. Dinnes J, Bancos I, Ferrante di Ruffano L, *et al.* Management of endocrine disease: imaging for the diagnosis of malignancy in incidentally discovered adrenal masses: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R51–64.
14. Hao M, Lopez D, Luque-Fernandez MA, *et al.* The lateralizing asymmetry of adrenal adenomas. *J Endocr Soc*, 2018; **2**: 374–85.
15. Gruber LM, Hartman RP, Thompson GB, *et al.* Pheochromocytoma characteristics and behavior differ depending on method of discovery. *J Clin Endocrinol Metab*, 2018; **104**: 1386–93.
16. Fassnacht M, Arlt W, Bancos I, *et al.* Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2016; **175**: G1–G34.
17. Terzolo M, Stigliano A, Chiodini I, *et al.* AME position statement on adrenal incidentaloma. *Eur J Endocrinol*, 2011; **164**: 851–70.
18. Sturgeon C, Shen WT, Clark OH, Duh QY, Kebebew E. Risk assessment in 457 adrenocortical carcinomas: how much does tumor size predict the likelihood of malignancy? *J Am Coll Surg*, 2006; **202**: 423–30.
19. Yan Q, Bancos I, Gruber LM, *et al.* When biochemical phenotype predicts genotype: pheochromocytoma and paraganglioma. *Am J Med*, 2018; **131**: 506–9.
20. Gruber LM, Erickson D, Babovic-Vuksanovic D, *et al.* Pheochromocytoma and paraganglioma in patients with neurofibromatosis type 1. *Clin Endocrinol (Oxf)*, 2017; **86**: 141–9.
21. Bancos I, Chortis V, Lang K, *et al.* The natural history of adrenal incidentaloma—results from the international prospective multi-centre EURINE-ACT study. *Endocr Abstracts*, 2017; **49**: GP122.
22. Bancos I, Arlt W. Diagnosis of a malignant adrenal mass: the role of urinary steroid metabolite profiling. *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 200–7.
23. Canu L, Van Hemert JAW, Kerstens MN, *et al.* CT characteristics of pheochromocytoma–Relevance for the evaluation of adrenal incidentaloma. *J Clin Endocrinol Metab*, 2018; **104**: 312–18.
24. Mantero F, Terzolo M, Arnaldi G, *et al.* A survey on adrenal incidentaloma in Italy. Study Group on Adrenal Tumors of the Italian Society of Endocrinology. *J Clin Endocrinol Metab*, 2000; **85**: 637–44.
25. Elhassan YS, Alahdab F, Prete A, *et al.* Natural history of adrenal incidentalomas with and without mild autonomous cortisol excess: a systematic review and meta-analysis. *Ann Intern Med*, 2019; **171**: 107–16.
26. Delivanis DA, Bancos I, Atwell TD, *et al.* Diagnostic performance of unenhanced computed tomography and ^{18}F-fluorodeoxyglucose positron emission tomography in indeterminate adrenal tumours. *Clin Endocrinol*, 2018; **88**: 30–6.
27. Guerin C, Pattou F, Brunaud L, *et al.* Performance of 18F-FDG PET/CT in the characterization of adrenal masses in noncancer patients: a prospective study. *J Clin Endocrinol Metab*, 2017; **102**: 2465–72.
28. Herndon J, Nadeau AM, Davidge-Pitts CJ, Young WF, Bancos I. Primary adrenal insufficiency due to bilateral infiltrative disease. *Endocrine*, 2018; **62**: 721–8.
29. Delivanis DA, Erickson D, Atwell TD, *et al.* Procedural and clinical outcomes of percutaneous adrenal biopsy in a high-risk population for adrenal malignancy. *Clin Endocrinol*, 2016; **85**: 710–16.
30. Eloubeidi MA, Black KR, Tamhane A, Eltoum IA, Bryant A, Cerfolio RJ. A large single-center experience of EUS-guided FNA of the left and right adrenal glands: diagnostic utility and impact on patient management. *Gastrointest Endosc*, 2010; **71**: 745–53.
31. Schuurbiers OC, Tournoy KG, Schoppers HJ, *et al.* EUS-FNA for the detection of left adrenal metastasis in patients with lung cancer. *Lung Cancer*, 2011; **73**: 310–15.
32. Kuruba R, Gallagher SF. Current management of adrenal tumors. *Curr Opin Oncol*, 2008; **20**: 34–46.
33. Bancos I, Tamhane S, Shah M, *et al.* Diagnosis of endocrine disease: the diagnostic performance of adrenal biopsy: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R65–80.
34. Zhang CD, Erickson D, Levy MJ, *et al.* Endoscopic ultrasound-guided fine-needle aspiration in the diagnosis of adrenal

metastasis in a high-risk population. *Endocr Pract*, 2017; **23**: 1402–7.
35. Leung K, Stamm M, Raja A, Low G, *et al.* Pheochromocytoma: the range of appearances on ultrasound, CT, MRI, and functional imaging. *AJR Am J Roentgenol*, 2013; **200**: 370–8.
36. Iñiguez-Ariza NM, Kohlenberg JD, Delivanis DA, *et al.* Clinical, biochemical, and radiological characteristics of a single-center retrospective cohort of 705 large adrenal tumors. *Mayo Clin Proc Innov Qual Outcomes*, 2017; **2**: 30–9.
37. Sood SK, Balasubramanian SP, Harrison BJ. Percutaneous biopsy of adrenal and extra-adrenal retroperitoneal lesions: beware of catecholamine secreting tumours! *Surgeon*, 2007; **5**: 279–81.
38. Lumachi F, Borsato S, Brandes AA, *et al.* Fine-needle aspiration cytology of adrenal masses in noncancer patients: clinicoradiologic and histologic correlations in functioning and nonfunctioning tumors. *Cancer*, 2001; **93**: 323–9.
39. Paulsen SD, Nghiem HV, Korobkin M, Caoili EM, Higgins EJ. Changing role of imaging-guided percutaneous biopsy of adrenal masses: evaluation of 50 adrenal biopsies. *AJR Am J Roentgenol*, 2004; **182**: 1033–7.
40. Vanderveen KA, Thompson SM, Callstrom MR, *et al.* Biopsy of pheochromocytomas and paragangliomas: potential for disaster. *Surgery*, 2009; **146**: 1158–66.
41. Williams AR, Hammer GD, Else T. Transcutaneous biopsy of adrenocortical carcinoma is rarely helpful in diagnosis, potentially harmful, but does not affect patient outcome. *Eur J Endocrinol*, 2014; **170**: 829–35.
42. Mazzaglia PJ, Monchik JM. Limited value of adrenal biopsy in the evaluation of adrenal neoplasm: a decade of experience. *Arch Surg*, 2009; **144**: 465–70.
43. Mody MK, Kazerooni EA, Korobkin M. Percutaneous CT-guided biopsy of adrenal masses: immediate and delayed complications. *J Comp Assisted Tomography*, 1995; **19**: 434–9.
44. Habscheid W, Pfeiffer M, Demmrich J, Müller HA. [Puncture track metastasis after ultrasound-guided fine-needle puncture biopsy. A rare complication?]. *Dtsch Med Wochenschr*, 1990; **115**: 212–15.
45. Bancos I, Taylor AE, Chortis V, et al. Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. ENSAT EURINE-ACT Investigators. *Lancet Diabetes Endocrinol*, 2020 Sep;**8**(9):773–81.
46. Arlt W, Biehl M, Taylor AE, *et al.* Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*, 2011; **96**: 3775–84.
47. Bancos I, Alahdab F, Crowley RK, *et al.* Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and subclinical Cushing's syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R283–95.
48. Dennedy MC, Annamalai AK, Prankerd-Smith O, *et al.* Low DHEAS: a sensitive and specific test for the detection of subclinical hypercortisolism in adrenal incidentalomas. *J Clin Endocrinol Metab*, 2017; **102**: 786–92.
49. Giuseppe R, Elena C, Maurizio G, et al. Adrenal Incidentalomas Are Tied to Increased Risk of Diabetes: Findings From a Prospective Study. *J Clin Endocrinol Metab*, 2020 Apr 1;**105**(4):dgz284. doi: 10.1210/clinem/dgz284.
50. Prete A, Taylor AE, Sitch AJ, *et al.* Mild autonomous cortisol excess in adrenal incidentalomas—metabolic disease burden and urinary steroid metabolome in 1201 prospectively recruited patients. *Endocrine Abstracts*, 2017; **50** OC2.1, 2017.
51. Tauchmanovà L, Rossi R, Biondi B, *et al.* Patients with subclinical Cushing's syndrome due to adrenal adenoma have increased cardiovascular risk. *J Clin Endocrinol Metab*, 2002; **87**: 4872–8.
52. Terzolo M, Pia A, Alì A, *et al.* Adrenal incidentaloma: a new cause of the metabolic syndrome? *J Clin Endocrinol Metab*, 2002; **87**: 998–1003.
53. Chiodini I, Tauchmanovà L, Torlontano M, *et al.* Bone involvement in eugonadal male patients with adrenal incidentaloma and subclinical hypercortisolism. *J Clin Endocrinol Metab,* 2002; **87**: 5491–4.
54. Debono M, Bradburn M, Bull M, Harrison B, Ross RJ, Newell-Price J. Cortisol as a marker for increased mortality in patients with incidental adrenocortical adenomas. *J Clin Endocrinol Metab*, 2014; **99**: 4462–70.
55. Di Dalmazi G, Vicennati V, Garelli S, *et al.* Cardiovascular events and mortality in patients with adrenal incidentalomas that are either non-secreting or associated with intermediate phenotype or subclinical Cushing's syndrome: a 15-year retrospective study. *Lancet Diabetes Endocrinol*, 2014; **2**: 396–405.
56. Morelli V, Reimondo G, Giordano R, *et al.* Long-term follow-up in adrenal incidentalomas: an Italian multicenter study. *J Clin Endocrinol Metab*, 2014; **99**: 827–34.
57. Morelli V, Eller-Vainicher C, Palmieri S, *et al.* Prediction of vertebral fractures in patients with monolateral adrenal incidentalomas. *J Clin Endocrinol Metab*, 2016; **101**: 2768–75.
58. Bancos I, Alahdab F, Crowley RK, *et al.* Therapy of endocrine disease: improvement of cardiovascular risk factors after adrenalectomy in patients with adrenal tumors and Subclinical Cushing Syndrome: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R283–95.

5.4

Adrenocortical Cancer

Anne Jouinot, Rossella Libè, and Jérôme Bertherat

Essential Information

Adrenocortical cancer (ACC) is a rare tumour with an overall poor prognosis, with a 5-year survival rate lower than 35% in most series. By contrast, adrenocortical tumours (especially non-hypersecreting benign adenomas) are frequent, emphasizing the importance of a correct diagnosis of malignancy in incidentally discovered adrenal masses. Recent advances in the molecular genetics of adrenal tumours demonstrate the importance of genomic alterations in ACC tumorigenesis. Integrated genomic studies identify different subgroups of ACC, characterized by distinct genomic alterations, including specific patterns of gene expression, chromosome alterations, and DNA methylation profiles, as well as somatic mutations in *p53* (TP53, CDKN2A) or *ß-catenin* (CTNNB1, ZNRF3) pathways. These genomic subgroups are remarkably associated with patient outcomes. Correct management of ACC requires a multidisciplinary specialist team. The diagnosis and extensive work-up of a malignant adrenocortical tumour is an important step, and relies on careful investigations of clinical, hormonal, and imaging features before surgery, and precise pathological examination after tumour removal. Most tumours appear steroid-secreting after careful hormonal investigations. Mixed steroid excess, most frequently the cosecretion of androgens and cortisol, is highly suggestive of a malignant tumour. ACC are most often large, heterogeneous tumours with a low lipid content on computed tomography (CT) or magnetic resonance imaging (MRI). Imaging investigations are also important for preoperative tumour staging and follow-up. ^{18}F-fluorodeoxyglucose positron emission tomography ((18)-FDG-PET) is increasingly used for that purpose. Stage 1 and 2 tumours are limited to the adrenal and present a better prognosis. Histopathological analysis with the use of the Weiss score is important for ACC diagnosis, especially in such localized tumours. Ki-67 immunostaining is useful for ACC prognostication. Stage 3 and 4 tumours extend locally or present with distant metastases, mainly liver, lung, lymph nodes, and bone. These advanced tumours are most often associated with a poor survival. Surgery by an expert surgeon aiming at complete tumour removal and avoiding tumour spillage is the best initial treatment, when achievable. In case of tumour recurrence or in metastasized ACC, medical therapy with the adrenolytic drug mitotane is used, alone or in combination with etoposide-doxorubicin-platinum cytotoxic chemotherapy. Mitotane should be considered as adjuvant therapy after complete removal of a tumour with a significant risk of recurrence. It is expected that a better understanding of the pathophysiology of ACC and the use of molecular tools for tumour classification will beneficially impact patient management in the near future. Similarly, international efforts in therapeutic trials will help to better define the respective roles of the various medical therapies currently available and facilitate the development of new targeted therapies. For this goal; as this is often the case for aggressive rare diseases, patient management in coordination with expert centres and networks of clinical research are important drivers of progress.

Introduction

Adrenocortical cancer (ACC) is among the most aggressive endocrine tumours with an overall poor prognosis. Morbidity and mortality can be secondary to tumour-related steroid hormone excess and/or tumour growth and metastases. This potentially poor outcome explains why the early detection of adrenocortical malignancy is paramount for the investigation of adrenal masses, alongside exclusion of hormone excess. The diagnosis of adrenocortical carcinoma relies on careful investigations of clinical, endocrine, and imaging features before surgery, and histopathological examination after tumour removal. Appropriate management and follow-up by an expert multidisciplinary team is critical to improve prognosis and drive progress for this rare cancer.

Pathogenesis and Genetics

Germline Alterations: ACC and Hereditary Tumour Syndromes

ACC can occur in the context of several tumour susceptibility syndromes (Table 5.4.1).

Historically, ACC has been first described as part of Li–Fraumeni syndrome in 1969 [1], which was later linked to germline *TP53* mutations [2]. This syndrome displays dominant inheritance and confers susceptibility to breast cancer, soft-tissue sarcoma, brain tumours, osteosarcoma, leukaemia, and ACC. Germline mutations in *TP53* have been observed in 50–80% of children with apparently sporadic ACC in North America and Europe. In Southern Brazil, a founder germline mutation, R337H in exon 10, is observed in almost all paediatric cases [3]. Germline *TP53* mutations are found much more rarely in adult ACC (<5%).

Another historic germline alteration is observed in the imprinted 11p15 region, carrying *IGF2*. This locus is altered most often through uniparental disomy, with loss of the maternal allele, responsible for *IGF2* overexpression, implicated in Beckwith–Wiedemann syndrome [4, 5]. This overgrowth disorder is characterized by macrosomia, macroglossia, organomegaly and developmental abnormalities (in particular abdominal wall defects with exomphalos), nephroblastoma, ACC, neuroblastoma, and hepatoblastoma.

More recently, ACC was described as part of other tumour susceptibility syndromes, such as Lynch syndrome and multiple endocrine neoplasia type 1 (MEN1). The *MEN1* gene, located at the 11q13 locus, is a tumour suppressor gene. A heterozygous inactivating germline mutation of *MEN1* is found in about 90% of families affected by multiple endocrine neoplasia type 1. The principal clinical features of this autosomal dominant syndrome include parathyroid (95%), endocrine pancreas (45%) and pituitary (45%) tumours and thymic carcinoids. Benign adrenocortical tumours and/or hyperplasia are common (25–40%), whereas ACC is rarely observed in MEN1 patients (<2%) [6]. Lynch syndrome is dominantly inherited and due to germline mutations in DNA mismatch repair genes *MLH1, MSH2, MSH6*, and *PMS2*. This syndrome typically confers an increased risk for colorectal, endometrial, small bowel, and upper tract urothelial cancers (Amsterdam Criteria), but also for sebaceous tumours, ovarian and pancreatic cancers, and ACC. Lynch syndrome accounts for up to 5% of ACC cases [7].

ACC has been described in several other hereditary tumour syndromes, including *APC, NF1, PRKAR1A, FLCN*, or *BRCA2* germline mutations [8].

Overall, germline predisposition affects 10% of adult ACC patients. Therefore, patient with newly diagnosed ACC should be considered for a genetic evaluation. In the absence of obvious inherited genetic syndrome, a next generation sequencing panel including *MSH2, MSH6, PMS2, MLH1, EPCAM, MEN1, APC*, and *TP53* could be proposed [8].

Somatic Alterations: IGF2, TP53, and Wnt/β-Catenin Pathways

In various cancers, the study of chromosomal rearrangement led to the identification of the oncogenes or tumour suppressor genes

Table 5.4.1 Main genetic predispositions to adrenocortical tumours and the molecular genetics of sporadic ACC. The table describes the main hereditary syndromes associated with adrenocortical tumours for which the locus and/or genes have been identified at the germline level. The alterations of these genes and chromosomal regions as somatic defect observed on tumour DNA of sporadic tumours are listed.

Genetic hereditary syndrome and MIM reference number	Genes, chromosomal localization, and type of defect	Tumours and non-tumoural manifestations observed in the hereditary syndrome	Somatic genetic defect observed in sporadic adrenocortical tumours
Li–Fraumeni syndrome (MIM 151623)	*TP53* (17p13) Inactivative heterozygous mutations of the tumour suppressor gene *TP53*	Soft-tissue sarcoma, breast cancers, brain tumours, leukaemia, ACC	*TP53* somatic mutations in sporadic ACC (30%) 17p13 LOH in sporadic ACC (>80%)
Multiple endocrine neoplasia type 1 (MIM 131100)	*MEN1* (11q13) Inactivative heterozygous mutations of the tumour suppressor gene *Menin*	Parathyroid, pituitary, pancreas tumours, adrenal cortex (25–40%), among which are adrenocortical adenomas, adrenocortical hyperplasia, and rare ACC	Very rare somatic *menin* gene mutations in sporadic adrenocortical tumours Frequent 11q13 LOH in ACC (90%)
Beckwith–Wiedemann syndrome (MIM 130650)	11p15 locus alterations *IGF2* overexpression *p57kip2* (*CDKN1C*) (genetic defect) *KCNQ1OT* (epigenetic defect) *H19* (epigenetic defect)	Exomphalos, macroglossia, macrosomia, hemihypertrophy, nephroblastoma, ACC	Frequent 11p15 LOH in ACC (>80%) Frequent *IGF2* overexpression in ACC (>85%)
Familial adenomatous polyposis (MIM 175100)	*APC* (5q12-22) Inactivating heterozygous mutations of the tumour suppressor gene *APC*	Multiple adenomatous polyps and cancer of the colon and rectum. Possible extracolonic manifestations include periampullary cancer, thyroid tumours, hepatoblastoma. Adrenocortical tumours can be diagnosed as adrenocortical adenomas, possibly multiples and/or bilateral, and ACC	Transcriptome analysis shows overexpression of targets of the Wnt-signalling pathway in ACC Immunohistochemistry shows abnormal localization of β-catenin in ACC, suggesting activation of the Wnt/ β-catenin pathway. β-catenin activating somatic mutations in ACC (20–30%)

LOH, loss of heterozygosity. ACC, adrenocortical cancer; MIM, Mendelian inheritance in man.

involved in their development. However, so far in ACC, such genes have been mostly identified by candidate gene approaches through the study of familial diseases responsible for adrenocortical tumours. Nevertheless, the loci of these genes are frequently altered in sporadic ACC, suggesting the importance of these loci and genes in the development of these tumours (Table 5.4.1). Studies using microsatellite markers in ACC have demonstrated a high percentage of loss of heterozygosity (LOH) or allelic imbalance at 11p15 (≥ 90%) and 17p13 (≥ 85%), corresponding respectively to *IGF2* and *TP53* loci. Indeed, *IGF2* is strongly overexpressed in 90% of sporadic ACC [9] and somatic mutations of *TP53* are found in 25% of adult ACC [10].

Genetic alterations of the Wnt-signalling pathway were initially identified in familial adenomatous polyposis coli and have been extended to a variety of cancers. Furthermore, familial adenomatous polyposis coli patients with germline mutations of the APC (adenomatous polyposis coli) gene that lead to an activation of the Wnt-signalling pathway, may develop adrenocortical tumours. The Wnt-signalling pathway is normally activated during embryonic development. ß-catenin is a key component of this signalling pathway. In ACC, ß-catenin delocalization can be observed, consistent with an abnormal activation of the Wnt-signalling pathway. This is explained in a subset of ACC (25%) by somatic activating mutations of *CTNNB1*, the ß-catenin gene [11]. Recently, exome sequencing and single-nucleotide polymorphism (SNP) array studies revealed alterations—mostly homozygous deletions—of *ZNRF3* in 20% of sporadic ACC [12, 13]. *ZNRF3* encodes an E3-ubiquitin ligase and is a known inhibitor of the Wnt/ß-Catenin pathway, by regulating the Wnt receptor turnover [14].

Genomics: Recent Advances Towards a New Classification of ACC

The development of genomics in the last decade has been a source of new progress both for tumour classification and pathogenesis understanding.

Gene expression profile (i.e. transcriptome), of benign tumours differs markedly from that of ACC[15, 16]. *IGF2* appeared as one of the most highly expressed genes in ACC. Adenomas display a steroidogenic signature, whereas ACC are characterized by a proliferative signature, suggesting that a dedifferentiation process might occur during malignant transformation [17]. The differences in gene expression profiles offer new diagnostic tools to discriminate benign from malignant adrenocortical tumours [16].

Moreover, studies analysing genomic (exome and chromosome alterations), epigenomic (micro-RNA expression and methylation) and transcriptomic (gene expression) profiles provided a complete high-throughput molecular characterization of adrenal tumours, in particular two large scale, integrated genomic studies, one from the European Network for the Study of Adrenal Tumours (ENSAT) network [12], and the other from The Cancer Genome Atlas (TCGA) programme [13]. These studies converge in an unsupervised classification of ACC into three groups, characterized by distinct molecular alterations and associated with very different outcomes (Figure 5.4.1). The first group is characterized by the combination a 'C1A' transcriptome profile (with upregulation of proliferative genes), a CpG island hypermethylation, a 'noisy' chromosome alteration profile (i.e. numerous and anarchic alterations), and an accumulation of mutations in cell-cycle and Wnt/ß-catenin related genes. This subgroup is associated with very poor outcome. Conversely, a second group of ACCs is characterized by a 'C1B' transcriptome profile (enriched in immune-related genes), no hypermethylation, a 'chromosomal' genome profile (i.e. extended patterns of LOH), and low mutation rate. This subgroup is associated with a better outcome. Finally, a third group of ACCs is characterized by 'C1A' transcriptome, and either hypermethylation or 'noisy' profile but not both, and shows an intermediate prognosis. These findings led to the development of new molecular markers for the prognostication of ACC [16, 18].

Epidemiology

ACC is a rare tumour with an estimated incidence between 1 and 2 per million and per year in adults in North America and Europe. The prevalence has been estimated between 4 and 12 per million [19]. As for many rare tumours, the incidence is difficult to determine, and the true numbers might be higher than the current

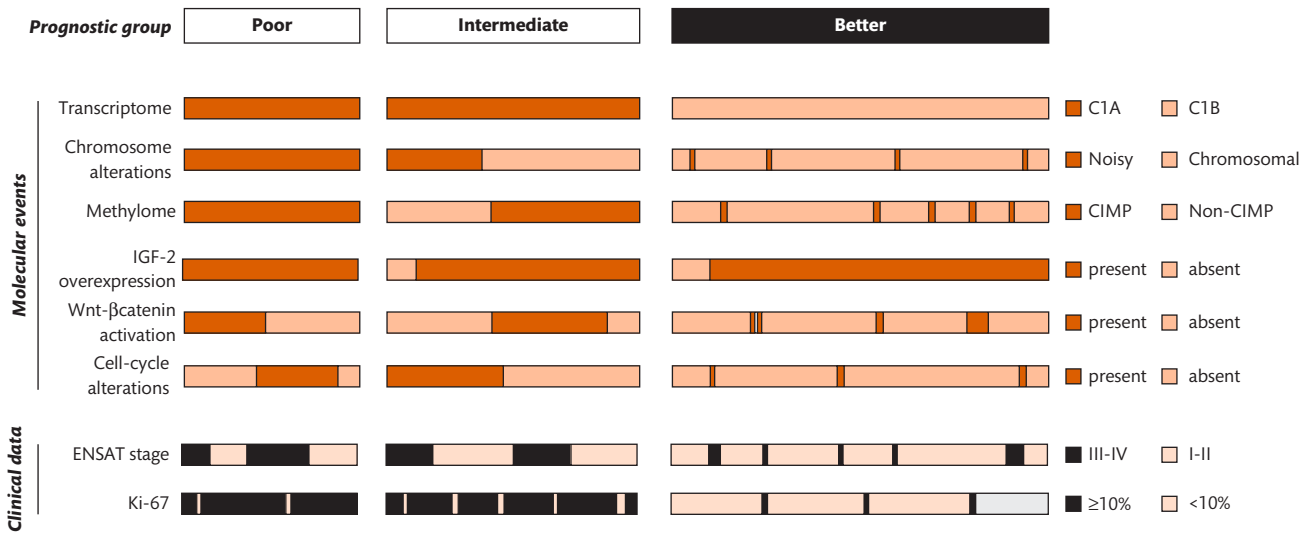


Figure 5.4.1 Genomic classification of adrenocortical cancer. This figure presents a schematic view of the genomic classification of ACC. Tumours are divided into three groups based on their transcriptome 'C1A' or 'C1B', chromosome alterations 'Noisy' or 'Chromosomal', and methylome—CIMP (CpG islands methylator phenotype) or non-CIMP profiles. These three groups are associated with very different prognosis. *IGF2* is overexpressed in 90% of ACC in all molecular groups. Activation of the Wnt-βcatenin pathway (including *CTNNB1* and *ZNRF3* alterations) and alterations in cell-cycle genes (including *TP53* and *CDKN2A*) are found mostly in aggressive ACC subgroups.

estimations. For instance, the prevalence of adrenal incidentaloma range in the general population from 1% in subjects younger than 30 years to 7% in subjects older than 70 years. Among the group of adrenal incidentaloma selected for surgery the frequency of ACC range between 2 to 12%.

In children, the incidence of ACC is considered as ten time lower as in adults, except in South Brazil where there is a higher incidence of paediatric ACC. This high incidence is explained by a specific germline *TP53* mutations as just discussed.

In some series, there is a slightly increased female to male ratio [20], although not always reported. Among female patients with Cushing's syndrome diagnosed during pregnancy, the frequency of ACC is higher than in non-pregnant female patients with Cushing's syndrome [21].

Clinical Features and Hormonal Investigations

Circumstances Leading to the Initial Diagnosis

Symptoms leading to the diagnosis of ACC can be due to steroid excess and/or tumour mass and metastases [22]. Although ACC is not the most frequent diagnosis of incidentalomas, nowadays the diagnosis of ACC is more often made during the diagnostic work-up of an adrenal incidentaloma. This circumstance is important since it might be a way to diagnose an ACC at an earlier stage and to improve the prognosis by an early complete surgical removal. This underlines the need for careful investigations of adrenal incidentalomas in order to decide on timely surgery if malignancy is suspected. Other specific features may be associated with rare genetic diseases such as the Li–Fraumeni and Beckwith–Wiedemann syndromes where ACC is part of a more complex syndrome as just discussed.

Less than a third of ACC are really 'non-hypersecreting' after careful hormonal investigations [22, 23]. In these cases, one should be cautious not to overdiagnose a tumour of the adrenal area as an ACC. These non-hypersecreting ACC can be diagnosed after investigation of adrenal incidentalomas or discovered by the manifestations of the tumour growth or extension: local symptoms (pain, palpation of a tumour, venous thrombosis), or distant metastases (liver, lung, bones). Fever may occur in some after tumour necrosis. However, in non-functional ACC, the general wellbeing of the patient is usually preserved until at a very late stage when the tumour causes signs and symptoms due to tumour expansion, which likely explains that non-hypersecreting ACC are typically diagnosed at a late stage.

Steroid Excess

Most patients will present with signs of steroid excess. Cushing's syndrome associated with signs of androgen excess with rapid onset and progression of clinical symptoms is the most characteristic presentation (**Figure 5.4.2**). Signs of mineralocorticoids or oestrogen excess are less frequent but highly suggestive of the diagnosis of ACC in a patient with an adrenal mass above 3 cm.

More than three-quarters of ACC are secreting tumours when assessed by careful hormonal investigations with current plasma and urine assays. In the absence of steroid excess, one should be cautious before diagnosing a mass of the adrenal area with suspicious imaging findings as an ACC. Hormonal investigations also provide important information for patient management and can serve as a tumour marker for follow-up. In 2018, the ACC guideline of the European Society of Endocrinology (ESE) and ENSAT has recommended a minimal hormonal work-up in patient with ACC [24] (**Box 5.4.1**).

In contrast to benign adrenocortical tumours, which usually oversecrete a single steroid, usually either cortisol or aldosterone, ACC can secrete various types of steroids. Cosecretion of androgens and cortisol is the most frequent and is highly suggestive of a malignant adrenocortical tumour. Cortisol oversecretion will induce centripetal obesity, protein wasting with skin thinning and striae, muscle atrophy (myopathy), and osteoporosis. Cortisol excess can also cause increased infections, diabetes, hypertension, psychiatric disturbances, and gonadal dysfunction in men and women. Androgen oversecretion may induce various manifestations in women: hirsutism, menstrual abnormalities, infertility, and eventually frank virilization (alopecia, deepening of the voice, clitoris hypertrophy). ACC can also secrete mineralocorticoids and steroid precursors. Oversecretion of oestrogens can be observed in rare cases and is very suggestive of ACC in a male patient with an

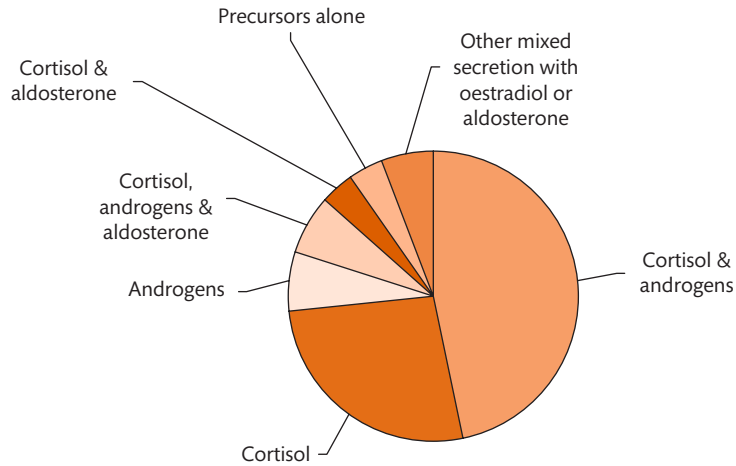


Figure 5.4.2 Steroid secretion. The figure shows the frequency of each steroid secretion profile (expressed as percentage) according to hormonal investigations in secreting ACC (Cochin endocrinology department series, investigated as reported in Abiven *et al*. 2006 [22]). Note that almost half of the secreting tumours are associated with mixed secretion of cortisol and androgens.

Box 5.4.1 Hormonal investigations in patients with ACC

These assays are adapted from the recommendation of the guideline of the European Society of Endocrinology (ESE) and the European Network for the Study of Adrenal Tumors (ENSAT). The steroids in italics are not part of the minimal work-up. This imply prior exclusion of a pheochromocytoma by the assay of urinary or plasma metanephrine and normetanephrine.

1 Glucocorticoid excess (a minimum of two tests):
- 24 hour urinary free cortisol and urinary creatinine
- Dexamethasone overnight suppression test (1 mg)
- Basal ACTH (plasma)
- *Basal cortisol*

2 Sex steroids:
- Testosterone (in women)
- Oestradiol (in men and postmenopausal women)
- Androstenedione
- DHEA-S (*or DHEA*)

3 Precursors:
- 17-hydroxyprogesterone
- 11-deoxycortisol
- *11-desoxycorticosterone*

4 Mineralocorticoids excess:
- Potassium
- Plasma aldosterone and renin (patients with hypertension or hypokalaemia)

adrenal tumour. Oestrogen excess is responsible for gynaecomastia in male.

Adrenocorticotrophic hormone (ACTH)-independent cortisol oversecretion is easily demonstrable: increased urinary cortisol excretion that is not suppressible with high doses of dexamethasone, and associated with undetectable ACTH plasma levels. Plasma 17-hydroxyprogesterone is often elevated at baseline and/or after ACTH stimulation, as well as the adrenal androgen precursor dehydroepiandrosterone sulphate (DHEA-S), which can lead to testosterone excess in females. Secretion of aldosterone by ACC is not frequent and can be detected by plasma aldosterone and renin assays.

Recently, mass spectrometry approaches have allowed detailed steroid profiling including precursor steroids usually not captured with routine assays, revealing a pattern of predominantly immature, early-stage steroidogenesis in ACC versus ACA and healthy controls [25]. Interestingly, urinary steroid metabolite profiling detected abnormal steroid precursor excretion even in ACC that were considered as non-hypersecreting using routine biochemical work-up. Thus, urine steroid metabolomics can be used as a biomarker tool for detecting malignancy in adrenal tumours [25].

Imaging Investigations

Computed Tomography and Magnetic Resonance Imaging

Imaging is an essential diagnostic step for ACC, especially in cases of adrenal incidentaloma. It is important both for the diagnosis of malignancy of an adrenal mass but also for the staging work-up. Adrenal CT is a very informative imaging procedure for adrenocortical tumours [26]. In ACC, it usually shows a unilateral mass, that is most often large (above 5–6 cm, and typically 10 cm and above), lowering the kidney. Apart from the size of the tumour, the features suggestive of malignancy are the lack of homogeneity with foci of necrosis and irregular margins or, in relatively homogenous tumours examined by unenhanced CT, an increased attenuation above 10 HU, indicating a low fat content (Figure 5.4.3). CT also informs about presence of local invasion and/or distant metastases, most frequently in liver and lung and all patients with suspected ACC should undergo preoperate staging of thorax, abdomen, and pelvis. Locoregional vessel invasion through the renal veins and the inferior vena cava can proceed up to the right atrium and result in metastatic pulmonary embolism [27]. Magnetic resonance imaging (MRI) is particularly useful for the detection of liver metastasis and venous invasion.

Nuclear Imaging

More recently, studies have demonstrated that many if not most ACC demonstrate a high uptake of ^{18}F-fluorodeoxyglucose ((18)-FDG). Thus (18)-FDG Positron Emission Tomography ((18)-FDG-PET) can help to distinguish between benign and malignant adrenal tumours [28] and provide useful information in pre- and postoperative staging [29] (Figure 5.4.4). It is especially informative in combination with CT.

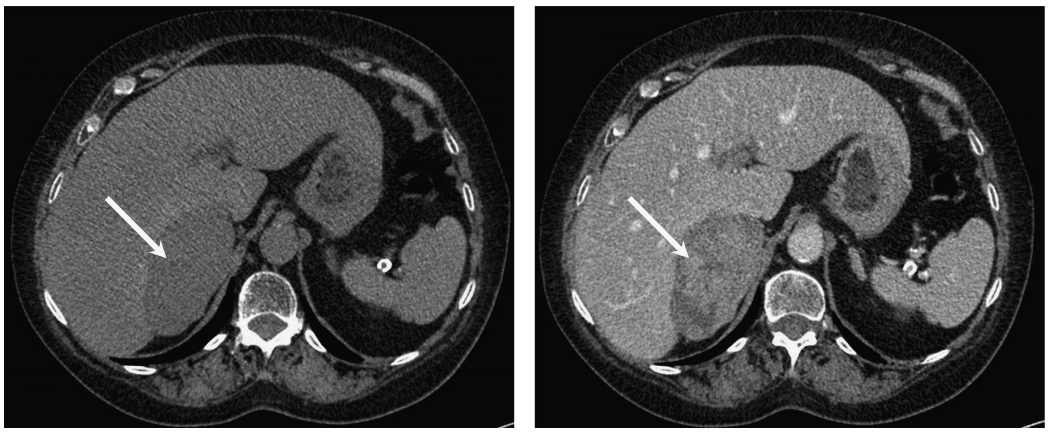


Figure 5.4.3 Computed tomography scan of an adrenocortical cancer. The white arrow points to an ACC arising from the left adrenal gland. The maximum diameter of the tumour is 10 cm. Unenhanced CT attenuation (left) is 27 HU (above the 10 HU cut-off for benign adenoma). After injection (right), the mass appears heterogeneous and the wash-out of contrast agent is below 50%.

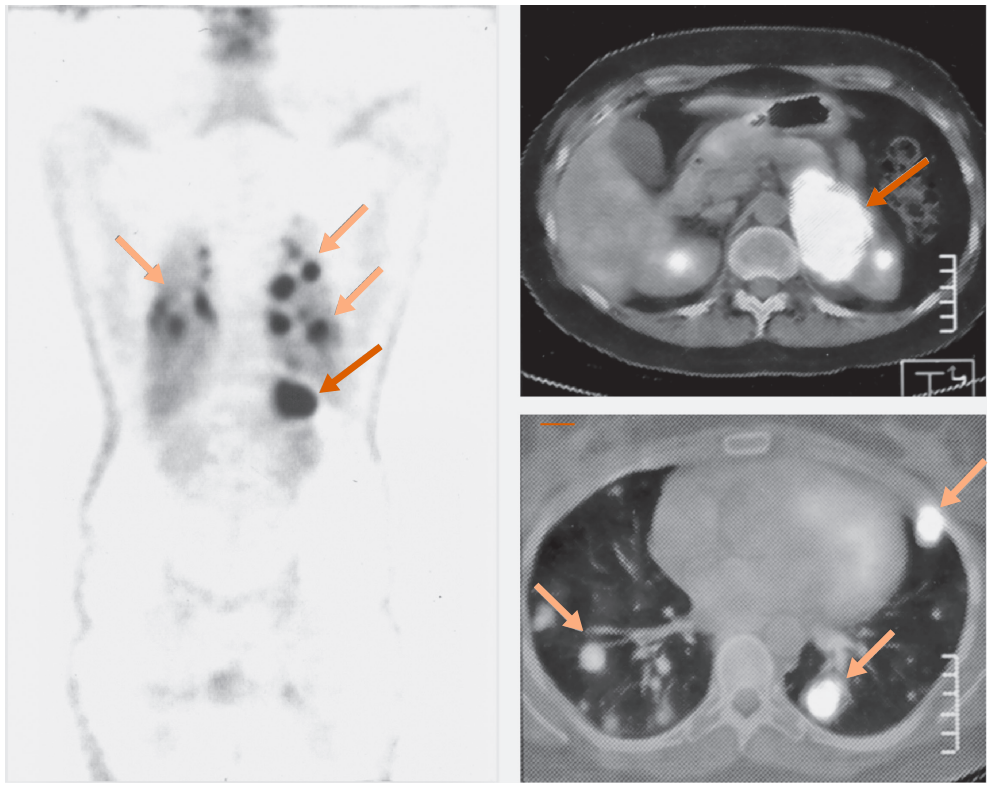


Figure 5.4.4 ^{18}F-fluorodeoxyglucose positron emission tomography ((18)-FDG-PET) of a metastatic adrenocortical cancer. Left panel: the left adrenal tumour shows a high uptake on the (18)FDG-PET scan (light orange arrow) and pulmonary metastases are detected = (dark orange arrows) at diagnosis of this stage 4 tumour. Right panel: combination of the PET imaging with a CT-scan shows the adrenal primary tumour (top) and the pulmonary metastases (bottom). *For a colour version of this figure, please see colour plate section.*

Bone scintigraphy may help evaluate bone metastases. However, in patients with Cushing's syndrome, bone remodelling and/or fracture can induce false positive results and the wider use of (18)-FDG-PET may replace it in the future. New adrenal cortex-specific radiolabelled tracers, such as [^{123}I]-metomidate, are currently investigated and may be promising options in the near future [30].

Diagnostic Criteria and Prognosis

Pathological Diagnosis and the Weiss Score

As just discussed, clinical, hormonal, and imaging investigations can be very suggestive of an ACC. Large adrenocortical tumours (>6 cm) are more likely to be malignant (Figure 5.4.5), but tumour size is clearly not a valid criterion to diagnose or exclude malignancy. On the other hand, a presentation with a large adrenal mass with ACTH-independent steroid excess is almost pathognomonic. However, pathological diagnosis is always a very important step. In the case of non-hypersecreting and/or localized tumours, pathology is key to diagnose both the adrenocortical origin and the malignant nature. The adrenocortical origin of the tumour is based on the histological analysis but also immunohistochemistry. Immunohistochemical markers are also employed to exclude other type of tumour, for instance a pheochromocytoma will stain positive for chromogranin A antibody while an ACC will not. Positive immunostaining in an adrenocortical tumour is either non-specific (e.g. Melan A) or not used on a routine basis (e.g. SF-1 or steroidogenic enzymes).

As often the case with endocrine tumours, the diagnosis of the malignant nature of an adrenocortical lesion can be difficult for the pathologist. There is not a single pathological feature which

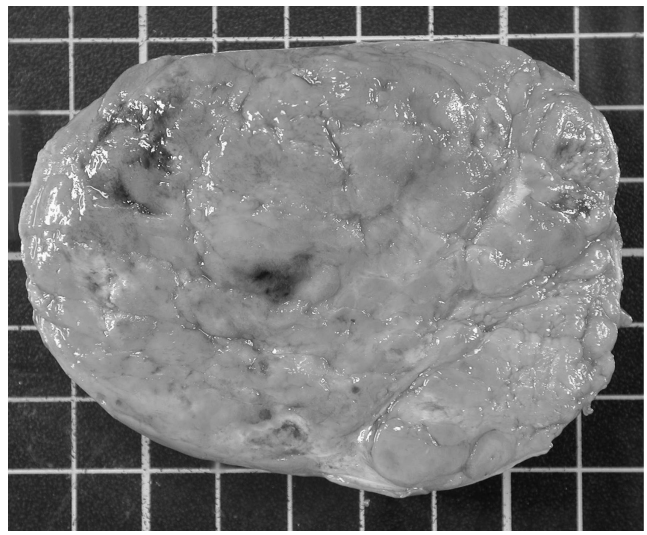


Figure 5.4.5 Macroscopic view of an adrenocortical cancer. The length of each scale is 1 cm. Note the characteristic large and heterogeneous aspect of the tumour.

Box 5.4.2 The nine criteria of the Weiss score for the diagnosis of malignancy of an adrenocortical tumour

The presence of three or more criteria classifies the tumour as a malignant one.

- High nuclear grade
- Mitotic rate above 5 per 50 high-power fields
- Atypical mitosis
- Less than 25 % of clear cells
- Diffuse architecture
- Necrosis
- Venous invasion
- Sinusoidal structures invasion
- Tumour capsule invasion

allows to conclusively diagnose a malignant adrenal cortical tumour. Combinations of various histological parameters that allow to establish a 'score' for a given tumour have been developed. The most widely used is the Weiss score comprising nine different items [31] (Box 5.4.2). Each item is given a value of one when it is present, zero when it is absent. The score represents the sum of the value for each individual item. It is assumed that a score equal or above 3 is most likely associated with a malignant tumour. Other approaches based on microscopic feature analysis have been developed, but are much less widely used and, therefore, less validated than the Weiss score. However, as all these approaches suffer limitations and are highly dependant on the experience of the pathologist, there is an effort to develop molecular markers of malignancy. As just described, *IGF2* overexpression and allelic losses at 17p13 have been suggested as potential molecular ACC markers [9]. More recently, large scale transcriptome analysis using DNA chips was used to develop molecular markers based on the expression level of two genes (*DLGAP5-PINK1*) for the diagnosis of malignancy [16, 32]. It is expected that in the near future such translational research programmes will have important impact for the classification and management of these tumours.

Tumour Staging

Tumour staging is the most important prognostic factor in ACC. Different staging classifications have been proposed over the years. The 2008 ENSAT staging provides superior survival stratification [33, 34] and is now the most commonly used classification. Four different stages are differentiated with this score. Stage 1 and stage 2 tumours are localized to the adrenal cortex and present with a maximum diameter below or above 5 cm, respectively. Stage 3 tumours present with local infiltration reaching the surrounding tissue, invasion into adjacent organs or lymph nodes. Stage 4 tumours are associated with distant metastases. The prognosis of stage 1 and 2 tumours is better than that of stage 3 or 4 tumours [33] (Figure 5.4.6).

Outcome and Long-Term Survival

The metastatic spread of ACC mostly involves liver and lung, observed in about 35 to 50% of the patients. Bone metastases are diagnosed in 10 to 15% of the cases. Metastatic spread to other organs is rare [22, 35].

The overall survival of patients with ACC is poor, with a 5-year survival rate below 35% in most series [36]. However, survival depends on tumour stage at diagnosis. It is likely that progress in the medical management of incidentalomas and better investigations of subtle degrees of steroid excess will increase the detection of localized ACC (stages 1 or 2). This should improve the overall survival rate. A better survival has been reported in younger patients, but this is not a constant finding [22]. Cortisol-secreting tumour appear to be associated with a poorer prognosis [22]. This could be due to the morbidity associated with Cushing's syndrome and/or to different tumour behaviour with more rapid progression. Some pathological features, such as a high mitotic rate or atypical mitotic figures as well as a high Ki-67 labelling, have been shown to be associated with a poor prognosis [37–40]. Recently, a Ki-67 proliferation index ≥10–20% was identified as a major indicator of poor prognosis in stage 1–3 ACC after apparently complete surgery [39]. The routine evaluation of Ki-67 expression in the pathology report is now recommended for the standard prognostic assessment of every resected ACC [24]. However, one limitation of Ki-67 is the low interobserver and intraobserver concordance for intermediate indices (10–30%) [41], which correspond to the cut-offs used for patient management.

Following the recent progress in genomics, targeted molecular markers have been developed for ACC prognostication [36] and may complete the standard pathology assessment. In particular, the prognostic markers derived from the analysis of transcriptome—differential expression of two genes (*BUB1B-PINK1*) [16, 32]—and

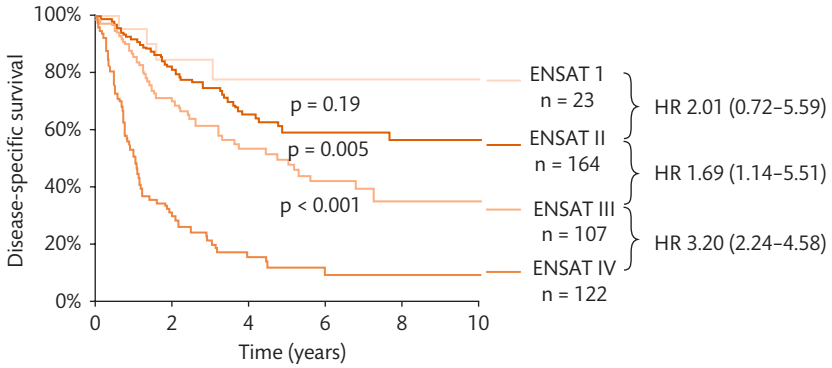


Figure 5.4.6 Survival of patients with adrenocortical cancer according to initial staging. The figure shows the disease-specific survival time (expressed in years) according to the ENSAT stage at presentation

Adapted with permission from Fassnacht M, Johanssen S, Quinkler M *et al.* Limited prognostic value of the 2004 International Union Against Cancer staging classification for adrenocortical carcinoma: proposal for a Revised TNM Classification. *Cancer* 2009;115: 243–50 [33].

from methylome—methylation level of four genes (*PAX5, PAX6, GSTP1*, and *PYCARD*) [18]— have demonstrated independent prognostic value from ENSAT stage and Ki-67, now also reported by large validation studies on the use of molecular markers for ACC prognostication [42, 43].

Treatment

Treatment aims to correct both steroid oversecretion and its clinical consequences in steroid-producing ACC, and, importantly, to eradicate the malignant tumour. The best way to achieve both goals is the complete removal of the tumour if possible, depending on tumour stage and the patient's condition (Figure 5.4.7).

Steroid oversecretion, when clinically significant and not curable by tumour removal, requires therapeutic intervention, in particular in the case of cortisol excess. Mitotane is the drug most often used in this context, because it also has a cytotoxic effect on the adrenocortical cells as discussed next. In the situation of clinically severe Cushing's syndrome requiring rapid control of steroid excess, other drugs can be also used, eventually in combination with mitotane, including ketoconazole, metyrapone, and etomidate [44].

Surgery

Surgery of the adrenal tumour is the major treatment of stage 1 to 3 ACC. It can also be discussed in stage 4 patients. The initial surgery is a crucial therapeutic step in the management of ACC. Therefore, it should be performed by highly trained specialist

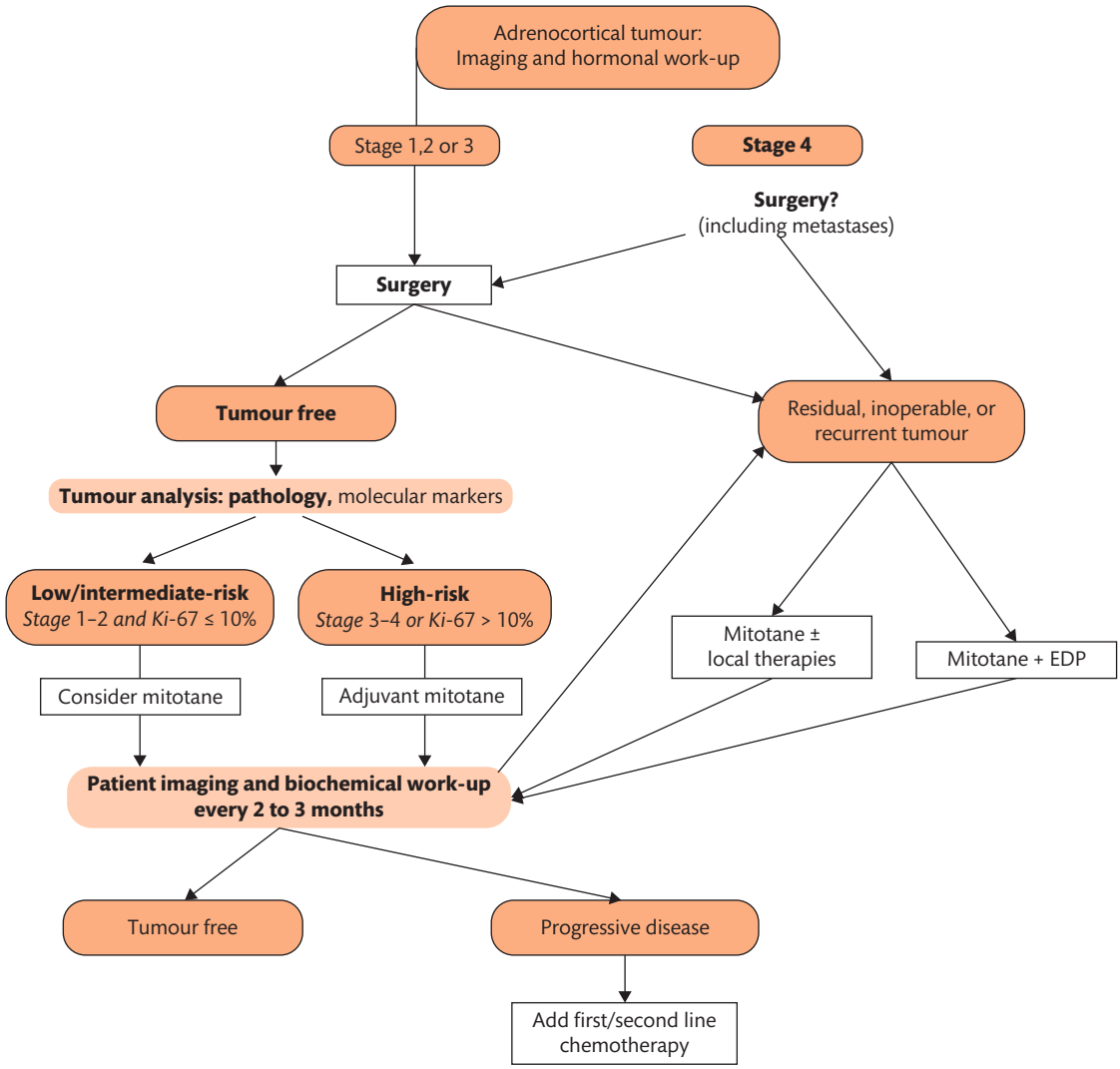


Figure 5.4.7 Schematic view of the management of patient with adrenocortical cancer. Surgery is the primary approach in patients with localized tumours (ENSAT stage 1, 2, and 3). Adjuvant mitotane is recommended for tumours with high risk of recurrence. In patients with distant metastases (stage 4), surgery of the primary tumour should be considered to reduce tumour mass, in particular if steroid excess is present, and also removal of metastases, if feasible, i.e. if no disseminated metastasis is present. Local recurrence without distant metastases usually requires surgery. Local treatments include: radiotherapy (especially for bone metastasis), chemoembolization (mostly for liver metastasis), radiofrequency thermal ablation of lung or liver metastasis, as well as surgical removal of limited metastasis. The delay between imaging and biochemical work-up can be extended to 3–4 months in patients presenting with complete remission and good prognostic factors and might be extended up to 6-month intervals after 2–3 years if there is still no recurrence. Mitotane and cytotoxic chemotherapy with etoposide, doxorubicin, platinum (EDP) is proposed for patients with rapidly progressive metastatic disease.

surgeons, experienced in the management of adrenal tumours and embedded into a multidisciplinary team, to achieve complete tumour removal without tumour spillage. Complete tumour removal is very important to increase the probability of long-term remission [24, 27, 45]. Open adrenalectomy is currently the standard of surgical care, with many advocating locoregional lymphadenectomy [46, 47]. Laparoscopic removal of malignant adrenocortical tumours could be associated with a high risk of peritoneal dissemination and should be restricted to suspected malignant adrenal masses with a diameter of less than 6 cm without evidence of local or lymph nodes invasion and surgical management by a high-volume centre [45]. Glucocorticoid replacement therapy should be started immediately after the removal of cortisol-secreting tumours to prevent acute adrenal deficiency as the contralateral adrenal cortex will be dormant consequent to the longstanding suppression of corticotroph ACTH secretion by the chronic cortisol oversecretion related to ACC.

In stage 4 patients with distant metastases, tumour debulking with removal of the adrenal tumour can be discussed, in order to both improve prognosis and reduce steroid excess. Surgery should be discussed considering tumour bulk and spread, as well as the velocity of observed tumour progression. However, it is important to consider the postoperative recovery period and the expected residual tumour mass as cytotoxic chemotherapy cannot be introduced in freshly operated patients. When the number of metastasis is limited, their surgical removal can also be discussed and downstaging by chemotherapy can also represent a feasible option in some cases.

Surgery is the primary approach in case of local recurrences, if operable. In general, recurrences occurring more than 12 months after the primary ACC surgery appear to be associated with a better prognosis, while early recurrence indicates a rapidly growing tumour.

Local Therapies and Radiotherapy

Radiofrequency thermal ablation of liver and lung metastasis with a maximum diameter less than 5 cm can be an alternative to surgical removal [48]. Bone metastasis can be operated to reduce fracture risk or neurological symptoms due to spinal metastasis. Radiation therapy has often been considered as not very effective to control tumour growth. However, it has been suggested that it could help to prevent local recurrence and potentially prolong survival [35]. Since all studies investigating radiotherapy in ACC are observational with potential confounding effects, adjuvant radiotherapy is not recommended for stage 1–2 tumours after complete tumour resection, but should be considered in addition to mitotane therapy in patients at high risk of recurrence, i.e. with uncomplete resection or stage 3 tumours [24].

Medical Therapy of ACC

When complete tumour removal is not possible or in case of recurrence, medical treatment with o,p'DDD (ortho, para', dichloro-, diphenyl-, dichloroethane, or mitotane) is recommended [24].

Mitotane has a specific cytotoxic effect on adrenal cortical cells [49]; it also inhibits steroid synthesis and, therefore, is usually effective for the control of steroid excess in ACC. Most studies reporting on the efficacy of mitotane in ACC are retrospective analyses, with variable results on tumour progression. An objective tumour regression could be observed in 25% of the cases [50]. Patients with a cortisol-secreting ACC might have a better survival when treated with mitotane in the 3 months following the surgery of the adrenal tumour [22]. A mitotane blood level of at least 14 mg/L seems to improve the tumour response rate [51]. However, the side effects of mitotane (mainly gastrointestinal and neurologic) can limit the ability to reach this suggested optimal level. The daily mitotane dose required to achieve this 14 mg/L level varies from patient to patient. Therefore, close monitoring of mitotane plasma levels is very helpful to remain in the narrow range between 14 and 20 mg/L, considered by most authors as the therapeutic range of mitotane in ACC, with lower levels rarely preventing tumour growth and higher levels associated with neurologic toxicity. Since mitotane invariably induces adrenal insufficiency, glucocorticoid replacement has to be initiated concurrently with mitotane and should be administered at increased doses (e.g. 40–80 mg per day) as mitotane strongly induces CYP3A4, which results in rapid inactivation of cortisol [52]. It is generally accepted that adjuvant therapy with mitotane should be proposed in patients considered to be at high risk of tumour recurrence. However, the benefit of mitotane treatment as an adjuvant medical treatment in low-intermediate risk ACC, i.e. stage 1–3, after 'complete' surgical removal, with Ki-67 index below 10%, remains to be demonstrated. Retrospective analysis have reported conflicting results [53]. This important point is expected to be clarified by an international randomized trial (ADIUVO, NCT00777244).

Several cytotoxic chemotherapy regimens have been used in ACC. Recently, the randomized phase III FIRM-ACT trial has established the association of mitotane with etoposide, doxorubicin, and platinum (EDP) chemotherapy as the best option for patients with unresectable disease [54]. The choice of first-line therapy, either mitotane alone or in combination with cytotoxic chemotherapy, depends on prognostic parameters. Mitotane monotherapy is usually recommended in case of slow progressive disease whereas cytotoxic chemotherapy is proposed as first-line for patients with rapidly progressive tumours, and as second-line therapy for patients with tumour progression despite therapeutic mitotane plasma levels or presenting with severe side effects limiting the use of mitotane [24]. The role of adjuvant Etoposide-Cisplatin chemotherapy will be investigated in the ADIUVO2 phase III trial (NCT03583710).

Several targeted therapies have been investigated in advanced ACC, with disappointing results. Antiangiogenic tyrosine kinase inhibitors such as sunitinib failed to demonstrate efficacy in advanced ACC, although drug efficacy might could have been hampered by pharmacologic interaction with mitotane [55]. Although inhibitors of the insulin-like growth factor (IGF) receptor seemed very attractive, given the major role of *IGF2* overexpression in the pathogenesis of ACC, the phase III GALACCTIC trial showed that linsitinib, an *IGF1R* inhibitor, did not improve progression-free or overall survival versus placebo [56].

Conclusion

Although ACC is a rare condition, significant advances have been made in the last decade to understand its pathophysiology. These have been important both for improving diagnosis of ACC and for a better assessment of disease prognosis. However, much more

progress needs to be achieved with regard to therapeutic options. Due to the rarity of ACC, collaborative work performed in national and international networks dedicated to adrenocortical tumours will be important, with ENSAT working very successfully in this field in Europe.

Areas of Uncertainty or Controversy

The pathological diagnosis of malignancy of an adrenocortical tumour can be difficult in some cases. Although careful analysis by an expert pathologist solves most cases, there are still some suspicious tumours with a borderline Weiss score [31] that are difficult to classify.

The prognosis of a tumour diagnosed as a malignant one, especially after complete surgical resection of a stage 1 or 2 ACC, is heterogeneous and still difficult to predict.

The surgical procedure has not been defined in a homogeneous way. The benefit of large 'en bloc' aggressive surgical resection that could lead to kidney ablation and the strategy for lymph node removal require further studies. The possibility to use laparoscopic resection of small ACC restricted to the adrenal gland by expert surgeons without increasing the risk of local recurrence or peritoneal metastasis need to be determined.

The benefit of radiotherapy remains uncertain. The place for radiotherapy as adjuvant or curative therapy will have to be established.

The benefit of mitotane as adjuvant therapy is most often accepted but needs to be demonstrated in prospective trials, in particular for ACC with lower risk of recurrence.

Likely Developments Over the Next 5–10 Years

The development of new immunohistochemical markers should improve the pathological diagnosis and prognostication of ACC.

Genomic studies have revealed a new classification of adrenocortical tumours leading to the development of molecular markers for the classification and prognostication of these tumours. The use of these markers in clinical practice, including its application to formalin-fixed and paraffin-embedded tissue is expected to improve patient management. Genomics studies are also giving new insights on the pathophysiology of ACC and this should help to define new targeted therapies.

The use of mass spectrometry for urinary steroid metabolite profiling has shown great promises in the diagnosis and follow-up of ACC, in particular when combined with machine learning-based algorithm-driven data analysis. Further studies will analyse the value of non-targeted metabolome and proteome analysis in ACC.

The development of new specific scintigraphies (as ^{123}I-iodometomidate ou ^{11}C-metomidate) is in progress and should improve tumour diagnosis and follow-up. Radiolabelled tracers could also be used for metabolic radiotherapy.

The results of ADIUVO study should clarify the role of adjuvant mitotane in low/intermediate risk patients after complete surgery.

The ADIUVO2 study will determine the role of adjuvant cytotoxic chemotherapy in patients at high risk of relapse.

New targeted therapies are currently in preclinical and clinical studies. Among them, new antiangiogenics inhibitors such as cabozantinib (NCT03612232) will be evaluated after mitotane discontinuation to prevent drug interactions.

ACKNOWLEDGEMENTS

We thank Dr Frédérique Tissier (service d'anatomopathologie, Hôpital Cochin) for help with the figures included in this chapter.

REFERENCES

1. Li FP, Fraumeni JF, Mulvihill JJ, *et al.* A cancer family syndrome in twenty-four kindreds. *Cancer Res*, 1988; **48**: 5358–62.
2. Malkin D, Li FP, Strong LC, *et al.* Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science*, 1990; **250**: 1233–8.
3. Ribeiro RC, Sandrini F, Figueiredo B, *et al.* An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. *Proc Natl Acad Sci U S A*, 2001; **98**: 9330–5.
4. Henry I, Jeanpierre M, Couillin P, *et al.* Molecular definition of the 11p15.5 region involved in Beckwith-Wiedemann syndrome and probably in predisposition to adrenocortical carcinoma. *Hum Genet*, 1989; **81**: 273–7.
5. Wilkin F, Gagné N, Paquette J, *et al.* Pediatric adrenocortical tumors: molecular events leading to insulin-like growth factor II gene overexpression. *J Clin Endocrinol Metab*, 2000; **85**: 2048–56.
6. Gatta-Cherifi B, Chabre O, Murat A, *et al.* Adrenal involvement in MEN1. Analysis of 715 cases from the Groupe d'etude des Tumeurs Endocrines database. *Eur J Endocrinol*, 2012; **166**: 269–79.
7. Raymond VM, Everett JN, Furtado LV, *et al.* Adrenocortical carcinoma is a lynch syndrome-associated cancer. *J Clin Oncol Off J Am Soc Clin Oncol*, 2013; **31**: 3012–18.
8. Petr EJ, Else T. Adrenocortical carcinoma (ACC): when and why should we consider germline testing? *Presse Medicale Paris Fr, 1983*, 2018; **47**: e119–25.
9. Gicquel C, Bertagna X, Gaston V, *et al.* Molecular markers and long-term recurrences in a large cohort of patients with sporadic adrenocortical tumors. *Cancer Res*, 2001; **61**: 6762–7.
10. Ohgaki H, Kleihues P, Heitz PU. p53 mutations in sporadic adrenocortical tumors. *Int J Cancer*, 1993; **54**: 408–10.
11. Tissier F, Cavard C, Groussin L, *et al.* Mutations of beta-catenin in adrenocortical tumors: activation of the Wnt signaling pathway is a frequent event in both benign and malignant adrenocortical tumors. *Cancer Res*, 2005; **65**: 7622–7.
12. Assié G, Letouzé E, Fassnacht M, *et al.* Integrated genomic characterization of adrenocortical carcinoma. *Nat Genet*, 2014; **46**: 607–12.
13. Zheng S, Cherniack AD, Dewal N, *et al.* Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*, 2016; **29**: 723–36.
14. Hao H-X, Xie Y, Zhang Y, *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature*, 2012; **485**: 195–200.
15. Giordano TJ, Thomas DG, Kuick R, *et al.* Distinct transcriptional profiles of adrenocortical tumors uncovered by DNA microarray analysis. *Am J Pathol*, 2003; **162**: 521–31.
16. de Reyniès A, Assié G, Rickman DS, *et al.* Gene expression profiling reveals a new classification of adrenocortical tumors and identifies molecular predictors of malignancy and survival. *J Clin Oncol*, 2009; **27**: 1108–15.
17. Bertherat J, Bertagna X. Pathogenesis of adrenocortical cancer. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 261–71.

18. Jouinot A, Assié G, Libe R, *et al.* DNA methylation is an independent prognostic marker of survival in adrenocortical cancer. *J Clin Endocrinol Metab*, 2017; **102**: 923–32.
19. Grumbach MM, Biller BMK, Braunstein GD, *et al.* Management of the clinically inapparent adrenal mass ('incidentaloma'). *Ann Intern Med*, 2003; **138**: 424–9.
20. Luton JP, Cerdas S, Billaud L, *et al.* Clinical features of adrenocortical carcinoma, prognostic factors, and the effect of mitotane therapy. *N Engl J Med*, 1990; **322**: 1195–201.
21. Lindsay JR, Nieman LK. The hypothalamic-pituitary-adrenal axis in pregnancy: challenges in disease detection and treatment. *Endocr Rev*, 2005; **26**: 775–99.
22. Abiven G, Coste J, Groussin L, *et al.* Clinical and biological features in the prognosis of adrenocortical cancer: poor outcome of cortisol-secreting tumors in a series of 202 consecutive patients. *J Clin Endocrinol Metab*, 2006; **91**: 2650–5.
23. Fassnacht M, Arlt W, Bancos I, *et al.* Management of adrenal incidentalomas: European Society of Endocrinology clinical practice guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2016; **175**: G1–34.
24. Fassnacht M, Dekkers O, Else T, *et al.* European Society of Endocrinology clinical practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2018: pii: EJE-18-0608.
25. Arlt W, Biehl M, Taylor AE, *et al.* Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*, 2011; **96**: 3775–84.
26. Hamrahian AH, Ioachimescu AG, Remer EM, *et al.* Clinical utility of noncontrast computed tomography attenuation value (Hounsfield units) to differentiate adrenal adenomas/hyperplasias from nonadenomas: Cleveland Clinic experience. *J Clin Endocrinol Metab*, 2005; **90**: 871–7.
27. Icard P, Goudet P, Charpenay C, *et al.* Adrenocortical carcinomas: surgical trends and results of a 253-patient series from the French Association of Endocrine Surgeons study group. *World J Surg*, 2001; **25**: 891–7.
28. Groussin L, Bonardel G, Silvéra S, *et al.* 18F-Fluorodeoxyglucose positron emission tomography for the diagnosis of adrenocortical tumors: a prospective study in 77 operated patients. *J Clin Endocrinol Metab*, 2009; **94**: 1713–22.
29. Leboulleux S, Dromain C, Bonniaud G, *et al.* Diagnostic and prognostic value of 18-fluorodeoxyglucose positron emission tomography in adrenocortical carcinoma: a prospective comparison with computed tomography. *J Clin Endocrinol Metab*, 2006; **91**: 920–5.
30. Hahner S, Stuermer A, Kreissl M, *et al.* [123 I]Iodometomidate for molecular imaging of adrenocortical cytochrome P450 family 11B enzymes. *J Clin Endocrinol Metab*, 2008; **93**: 2358–65.
31. Lau SK, Weiss LM. The Weiss system for evaluating adrenocortical neoplasms: 25 years later. *Hum Pathol*, 2009; **40**: 757–68.
32. Fragoso MCBV, Almeida MQ, Mazzuco TL, *et al.* Combined expression of BUB1B, DLGAP5, and PINK1 as predictors of poor outcome in adrenocortical tumors: validation in a Brazilian cohort of adult and pediatric patients. *Eur J Endocrinol*, 2012; **166**: 61–7.
33. Fassnacht M, Johanssen S, Quinkler M, *et al.* Limited prognostic value of the 2004 International Union Against Cancer staging classification for adrenocortical carcinoma: proposal for a Revised TNM Classification. *Cancer*, 2009; **115**: 243–50.
34. Lughezzani G, Sun M, Perrotte P, *et al.* The European Network for the Study of Adrenal Tumors staging system is prognostically superior to the international union against cancer-staging system: a North American validation. *Eur J Cancer Oxf Engl, 1990*, 2010; **46**: 713–19.
35. Fassnacht M, Allolio B. Clinical management of adrenocortical carcinoma. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 273–89.
36. Jouinot A, Bertherat J. Management of endocrine disease: adrenocortical carcinoma: differentiating the good from the poor prognosis tumors. *Eur J Endocrinol*, 2018, DOI: 10.1530/EJE-18-0027.
37. Miller BS, Gauger PG, Hammer GD, *et al.* Proposal for modification of the ENSAT staging system for adrenocortical carcinoma using tumor grade. *Langenbecks Arch Surg*, 2010; **395**: 955–61.
38. Assié G, Antoni G, Tissier F, *et al.* Prognostic parameters of metastatic adrenocortical carcinoma. *J Clin Endocrinol Metab*, 2007; **92**: 148–54.
39. Beuschlein F, Weigel J, Saeger W, *et al.* Major prognostic role of Ki67 in localized adrenocortical carcinoma after complete resection. *J Clin Endocrinol Metab*, 2015; **100**: 841–9.
40. Libè R, Borget I, Ronchi CL, *et al.* Prognostic factors in stage III-IV adrenocortical carcinomas (ACC): an European Network for the Study of Adrenal Tumor (ENSAT) study. *Ann Oncol Off J Eur Soc Med Oncol*, 2015; **26**: 2119–25.
41. Papathomas TG, Pucci E, Giordano TJ, *et al.* An international Ki67 Reproducibility study in adrenal cortical carcinoma. *Am J Surg Pathol*, 2016; **40**: 569–76.
42. Assié G, Jouinot A, Fassnacht M, *et al.* Value of molecular classification for prognostic assessment of adrenocortical carcinoma. *JAMA Oncol*, 2019. doi: 10.1001/jamaoncol.2019.1558.
43. Lippert J, Appenzeller S, Liang R, *et al.* Targeted molecular analysis in adrenocortical carcinomas: a strategy toward improved personalized prognostication. *J Clin Endocrinol Metab*, 2018; **103**: 4511–23.
44. Preda VA, Sen J, Karavitaki N, Grossman AB. Etomidate in the management of hypercortisolaemia in Cushing's syndrome: a review. *Eur J Endocrinol*, 2012; **167**: 137–43. Review. Erratum in: *Eur J Endocrinol*, 2013; 168: X1.
45. Gaujoux S, Mihai R, joint working group of ESES and ENSAT. European Society of Endocrine Surgeons (ESES) and European Network for the Study of Adrenal Tumours (ENSAT) recommendations for the surgical management of adrenocortical carcinoma. *Br J Surg*, 2017; **104**: 358–76.
46. Reibetanz J, Jurowich C, Erdogan I, *et al.*; German ACC study group. Impact of lymphadenectomy on the oncologic outcome of patients with adrenocortical carcinoma. *Ann Surg*, 2012; **255**: 363–9.
47. Reibetanz J, Rinn B, Kunz AS, *et al.* Patterns of lymph node recurrence in adrenocortical carcinoma: possible implications for primary surgical treatment. *Ann Surg Oncol*, 2019; **26**: 531–8.
48. Wood BJ, Abraham J, Hvizda JL, *et al.* Radiofrequency ablation of adrenal tumors and adrenocortical carcinoma metastases. *Cancer*, 2003; **97**: 554–60.
49. Sbiera S, Leich E, Liebisch G, *et al.* Mitotane inhibits sterol-O-acyl transferase 1 triggering lipid-mediated endoplasmic reticulum stress and apoptosis in adrenocortical carcinoma cells. *Endocrinology*, 2015; **156**: 3895–908.

50. Haak H, Hermans J, van de Velde C, *et al.* Optimal treatment of adrenocortical carcinoma with mitotane: results in a consecutive series of 96 patients. *Br J Cancer*, 1994; **69**: 947–51.
51. Kerkhofs TM, Baudin E, Terzolo M, *et al.* Comparison of Two Mitotane Starting Dose Regimens in Patients With Advanced Adrenocortical Carcinoma. *J Clin Endocrinol Metab*, 2013; **98**: 4759–67
50. Chortis V, Taylor AE, Schneider P, *et al.* Mitotane therapy in adrenocortical cancer induces CYP3A4 and inhibits 5α-reductase, explaining the need for personalized glucocorticoid and androgen replacement. *J Clin Endocrinol Metab*, 2013; **98**: 161–71.
53. Terzolo M, Angeli A, Fassnacht M, *et al.* Adjuvant mitotane treatment for adrenocortical carcinoma. *N Engl J Med*, 2007; **356**: 2372–80.
54. Fassnacht M, Terzolo M, Allolio B, *et al.* Combination chemotherapy in advanced adrenocortical carcinoma. *N Engl J Med*, 2012; **366**: 2189–97.
55. Kroiss M, Quinkler M, Johanssen S, *et al.* Sunitinib in refractory adrenocortical carcinoma: a phase II, single-arm, open-label trial. *J Clin Endocrinol Metab*, 2012; **97**: 3495–503.
56. Fassnacht M, Berruti A, Baudin E, *et al.* Linsitinib (OSI-906) versus placebo for patients with locally advanced or metastatic adrenocortical carcinoma: a double-blind, randomised, phase 3 study. *Lancet Oncol*, 2015; **16**: 426–35.

5.5

Phaeochromocytoma and Paraganglioma

5.5.1 Genetics of Phaeochromocytomas, Paragangliomas, and Neuroblastoma

Eamonn R. Maher and Ruth T. Casey

Introduction

Phaeochromocytomas, paragangliomas, and neuroblastomas are the main primary tumours that arise from the autonomic nervous system. The autonomic nervous system is subdivided into the sympathetic and parasympathetic systems. Phaeochromocytomas arise from sympathetic nervous system (chromaffin) cells in the adrenal medulla. Paragangliomas may arise from the sympathetic or parasympathetic system. Parasympathetic ganglia-derived paragangliomas (herein referred to as head and neck paraganglioma, HNPGL) develop along branches of the vagal and glossopharyngeal nerves (e.g. carotid body tumours, glomus jugulare) and are only rarely secretory. Phaeochromocytoma, paraganglioma, and HNPGL are rare in childhood but neuroblastomas, which are derived from neuroblasts in the developing sympathetic nervous system and are most common in children under the age of 5 years. Familial forms of neuroblastoma are rare but a major feature of phaeochromocytoma and paraganglioma (PPGL) and HNPGL is the high frequency of inherited cases and the major inherited syndromic and non-syndromic disorders that predispose to these tumours are described in Chapter 6.12.

Phaeochromocytoma and Paraganglioma

Clinical and Pathological Features

The estimated incidence of PPGL is 2 to 8/million of the population per year, giving an estimated annual incidence of 130–520 new patients a year in the United Kingdom [1]. Presenting symptoms of these rare tumours are most commonly related to catecholamine excess, and include headache, palpitations, paroxysmal hypertension, anxiety, abdominal pain, and excessive sweating. PPGL are often diagnosed incidentally following cross-sectional imaging studies [2].

Paragangliomas in the abdomen and pelvis commonly arise from the aortic bifurcation or the so-called organ of Zuckerkandl, but can also occur in the urinary bladder. Mediastinal paraganglioma are rare but can develop from para-aortic (middle mediastinum) and para-vertebral (posterior mediastinum) sympathetic chain ganglia [3]. PPGL and HNPGL may be familial and/or multifocal (in which case a genetic cause should be suspected) and/or there may be a previous history of wild-type gastrointestinal stromal tumour (GIST) or renal cell carcinoma (which are each associated with germline mutations in succinate dehydrogenase (SDH) subunit genes (see next)). Although most PPGL are benign (~85–90%), current predictors of malignant or aggressive tumours are imprecise and therefore in 2017, the World Health Organization (WHO) classification of adrenal tumours abolished the term 'benign' for PPGL, stating that all PPGL should be considered to have metastatic potential [4].

Histologically, PPGL are identified microscopically by a characteristic 'zellballen' pattern, referring to nests of polygonal/spindle shaped cells on a rich vascular network and immunohistochemical stains such as; synaptophysin, chromogranin A and S100 are used to confirm a diagnosis of PPGL [5]. In 2017, the first staging system for PPGL (but not HNPGL) was proposed by the American Joint Committee on Cancer (AJCC) and includes tumour size, lymph node metastases, and distant metastases [6]. Tumour size (>5 cm), local invasion and extra-adrenal abdominal location are predictors of malignancy [7] as is the detection of a germline pathogenic *SDHB* variant [8].

Surgery remains the first line treatment option for PPGL and, when possible, a minimally invasive approach is favoured. In the genomics era, surgical management can also be tailored based on the genetic aetiology of PPGL. Patients with certain hereditary syndromes (e.g. multiple endocrine neoplasia type 2 [MEN2], von Hippel Lindau disease [VHL], neurofibromatosis type 1 [NF1]) are at risk of bilateral phaeochromocytomas (Chapter 6.12) and cortical sparing surgery reduces the morbidity associated with bilateral adrenalectomy and requirement for lifelong glucocorticoid and mineralocorticoid replacement. Although patients with SDH subunit gene (*SDHX*) mutations and *MAX* gene mutations may develop bilateral phaeochromocytoma, the higher risk of malignancy negates the benefit of cortical sparing surgery and total adrenalectomy should be considered for these patients [9].

Long-term surveillance strategies are influenced by the tumour location and the detection of a germline predisposition. Current guidelines recommend lifelong surveillance for young patients, patients with a large or an extra-adrenal tumour and those with a known hereditary predisposition. Patients with a small (<5 cm) presumed sporadic phaeochromocytoma undergo a minimum of 10 years clinical surveillance for disease recurrence [10]. Postoperative surveillance includes annual biochemical testing for patients with secretory PPGL or cross-sectional imaging with CT/MRI every 1–2 years for patients with non-secretory PPGL [10].

Genetics

Despite clinical studies suggesting that PPGL is familial in ~10% of patients, the application of molecular genetic studies to PPGL has revealed that around 40% of individuals with PPGL will harbour a germline mutation in a PPGL predisposition gene and that most of these will not have a family history. The most common causes of inherited predisposition to PPGL are germline mutations in the various succinate dehydrogenase subunit genes (*SDHX*), *VHL, RET*, and *NF1* genes, but mutations in multiple other genes (including *MAX, TMEM127, HIF2A/EPAS1, FH, MDH2, SLC25A11*) have also been described to predispose to PPGL. The clinical characteristics of germline mutations in different inherited PPGL genes differ and are described in detail in Chapter 6.12.

The somatic (acquired) genetic and epigenetic alterations associated with PPGL tumourigenesis have been described in both inherited and sporadic forms of PPGL. Most inherited PPGL genes are classed as tumour suppressor genes and the germline mutations are loss of function variants. Initiation of tumourigenesis requires inactivation of the wild-type allele and this can occur through a variety of mechanisms including large chromosomal deletions (most often), somatic mutations, or promoter methylation with transcriptional silencing. For example, PPGL from patients with von Hippel–Lindau disease frequently demonstrate loss of the short arm of chromosome 3p (so the wild-type (non-mutated) VHL allele at 3p25 is lost) and in PPGL from patients with NF1 chromosome 17q loss is found (the *NF1* tumour suppressor gene is located at 17q11.2). As germline *RET* mutations activate the oncogenic function of RET, chromosome 10 deletions are not a feature of phaeochromocytomas from patients with MEN2. Secondary epigenetic alterations (e.g. genome-wide hypermethylation) are a characteristic feature of PPGL associated with germline mutations in *SDHX, FH, MDH2,* and *SLC25A11*. The gene products of these genes encode key components of cellular metabolic pathways and their inactivation leads to accumulation of metabolites (e.g. succinate with *SDHX* and fumarate with *FH* mutations) that inhibit alpha-ketoglutarate-dependent enzymes including the TET (ten-eleven translocation) proteins that are require for active DNA demethylation and lead to silencing of tumour suppressor genes through aberrant promoter methylation [11, 12].

Various genome-wide copy number abnormality studies have demonstrated recurrent areas of loss and gain in PPGL including chromosome 1p, 3p, 3q 11p, 17, 21q, and 22q loss. Interestingly 11p loss in PPGL preferentially affects the maternally inherited chromosome suggesting that imprinted genes (e.g. the paternally expressed *IGF2* and maternally expressed *CDKN1C*) are implicated in the pathogenesis of PPGL and the parent-of-origin effects on tumour risks seen with mutations in the chromosome 11q *SDHD* and *SDHAF2* genes (see Chapter 6.12).

Large-scale next-generation sequencing studies of somatic mutations in sporadic PPGL have identified mutations in a wide variety of genes including genes known to be associated with inherited PPGL (*VHL, NF1, RET*) and also genes in which only somatic mutations occur [13]. Though some important inherited PPGL tumour suppressor genes map to regions of recurrent copy number loss in sporadic PPGL (including *SDHB* to 1p36, *VHL* to 3p25, and *NF1* to 17q11.2) the frequency of somatic mutations in inherited PPGL genes in sporadic PPGL varies markedly. For example, whereas somatic mutations in *NF1* and *VHL* are relatively common, somatic mutations in *SDHX* subunit genes are infrequent. Gain of function mutations in *RET* and *HIF2A*/*EPAS1* can occur as both germline and somatic mutations. The most strongly activating mutations (e.g. *RET* M918T which is associated with MEN2B) are more often found as somatic than germline variants. Polycythaemia and PPGL can be features of somatic (rarely germline) *HIF2A* mutations [14]. Importantly, *HIF2A* variants have been associated with mosaicism and, therefore, if an *HIF2A* variant is identified in an apparently sporadic tumour, it is recommended that efforts are made to further clarify the presence of the variants in other tissue types, to inform the extent of disease and the appropriate surveillance approach for the patient and relatives [15]. Somatic mutations in *HIF2A* (which encodes the alpha subunit of the heterodimeric HIF-2 hypoxia responsive transcription factor) usually occur at or adjacent to the proline residue (P531) that is critical for binding to the VHL gene product (pVHL). The pVHL protein has a key role in targeting the HIF2A protein for proteosomal degradation and somatic inactivating mutations in *VHL* and activating mutations in *HIF2A* will both result in stabilization of HIF2A and activation of hypoxic gene response pathway [16]. Furthermore, *SDHX* and *FH* mutations are also associated with stabilization of HIF2A as both are associated with an inhibition of Pro531 hydroxylation which is required for pVHL binding and there are rare reports of PPGL patients with germline mutations in EGLN1/PHD2 which encodes one of the proline hydroxylase enzymes that act on HIF2A.

A wide variety of genes have been reported to be somatically mutated in PPGL but not implicated in familial PPGL [13]. These include activating mutations in the *HRAS* and *BRAF* proto-oncogenes and *CSDE1* (a *wnt*-pathway regulator). In a comprehensive genomic study, a further *wnt*-pathway related abnormality, *MAML3* fusion genes was identified in PPGL [13]. Other reported targets for somatic mutation include *SETD2, FGFR1, TP53, ATRX, ARNT, IDH1, H3F3A, MET* [13, 17] (see **Figure 5.5.1.1**). Though a large number of genes have been reported to harbour germline and/or somatic mutations in PPGL, the number of driver mutations per tumour is low and in many cases only a single event is detected [13]. Somatic *HRAS* mutations can be detected in about 10% of sporadic PPGL and, interestingly, it seems that somatic *HRAS* mutations and germline inherited PPGL gene mutations are usually mutually exclusive [18]. It has therefore been suggested that somatic mutation profiling of PPGL to identify driver gene mutations can facilitate management of patients with PPGL. If a mutation in an inherited PPGL gene is identified, then germline sequencing can be undertaken to determine if the mutation is somatic or germline. However, if a *HRAS* mutation is detected in the tumour then the risk of inherited disease will be low.

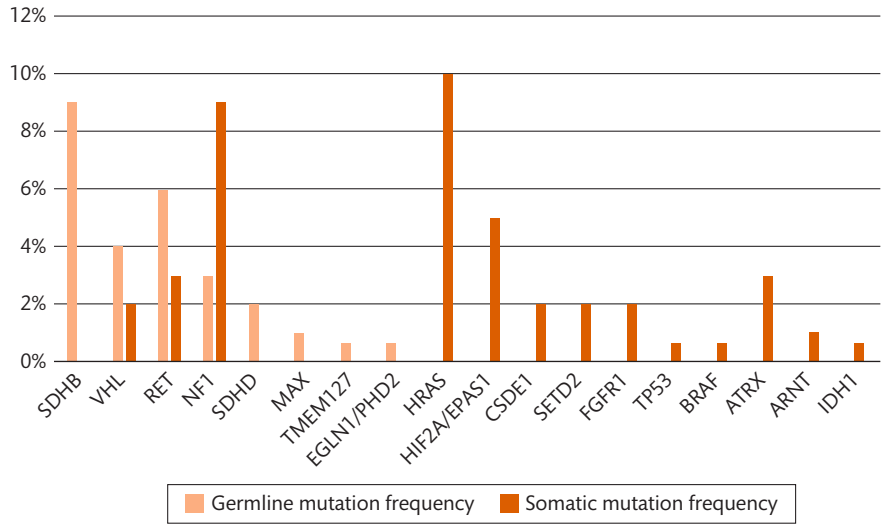


Figure 5.5.1.1 Summary of findings from a comprehensive genomic analysis of phaeochromocytoma and paraganglioma by Fishbein *et al.* [13].

Genotype-Function Correlations

Genomic profiling of PPGL has facilitated the diagnosis of hereditary cases, enabled a better understanding of the molecular pathways implicated in tumour development and provided an opportunity to develop new therapeutic approaches. Transcriptomic analysis of PPGL suggested two distinct categories: Cluster 1 (the 'pseudohypoxic cluster'), characterized by upregulation of hypoxia signalling pathways [19] and 'cluster 2', characterized by an upregulation of kinase signalling pathways [20]. Tumours caused by pathogenic variants in the *VHL, FH, SDHX, MDH, SLC25A11*, and *EPAS1* genes fall into the pseudohypoxic cluster [21], which is characterized by transcriptional upregulation of genes implicated in angiogenesis, and cellular proliferation secondary to hypoxia inducible factor (HIF) complex stabilization [19]. Cluster 2 tumours are associated with germline and somatic mutations in *RET, NF1, TMEM127, MAX*, and *HRAS*, which activate kinase pathways including the (i) PI3K/AKT, (ii) RAS/RAF/ERK, and (iii) mTORC1 pathways. More recently a Cancer Genome Atlas (TCGA) study has extended the molecular classification of PPGL to also include a *Wnt*-pathway altered cluster (associated with somatic *CSDE1* mutations and *MAML3* fusion genes and a small fourth category of cortical admixture PPGL) [13]. Included in this category were a small number (2) of *MAX* mutated tumours, but the relevance of this category as a distinct biological category remains under review. Looking at the somatic variant landscape of PPGL in isolation, the most common molecular pathway implicated is the RAS/RAF/ERK pathway [22]. This pathway is commonly implicated in tumourigenesis and consists of a series of serine-threonine kinases functional in a sequential pathway to regulate cell cycle factors. See Figure 5.5.1.2.

Reduced activity of PHD then stabilizes hypoxia inducible factor subunit alpha (HIFα) and upregulates HIF and downstream HIF targets.

Figure Figure 5.5.1.2B shows the interaction between PI3K/AKT, RAS/RAF/ERK, mTORC1 pathways, and the regulation of these pathways by known PPGL predisposition genes.

Genotype-Phenotype Correlations

Histopathology and Immunohistochemistry (IHC)

Biallelic inactivation of any of the *SDHX* genes usually results in destabilization of the SDH enzyme complex (Chapter 6.12), which can be detected by loss of staining for the SDHB protein by IHC [23]. Thus, SDHB IHC can be used to identify PPGL and HNPGL harbouring a *SDHX* mutation, provides a guide to likely of pathogenicity of an uncharacterized *SDHX* variant and is an indicator of malignancy risk [24]. IHC for *SDHA* expression can predict the presence of pathogenic variants in the *SDHA* gene and is also utilized in clinical practice. IHC may also be utilized for other hereditary causes of PPGL including assessment of MAX expression and the diagnosis of a pathogenic variant in *FH* is facilitated by IHC to detect protein succinylation (a post-translational modification resulting from the reaction of excess fumarate with cysteine residues to generate S-(2-succinyl)-cysteine (2SC)) [25].

Biochemistry

Biochemical testing is an integral component of the diagnostic pathway for PPGL and guidelines recommend urinary or plasma metanephrines and plasma 3-methoxytyramine (3MT) as the first line biochemical tests in the diagnosis of PPGL [1]. The pattern of catecholamine secretion can be determined by paraganglial cell differentiation and therefore biochemistry can be used to predict genotype and or malignant potential. Pseudohypoxic PPGL, including those tumours caused by *TCA* gene mutations and VHL, are characterized by poor differentiation of paraganglia cells and reduced expression of a key catecholamine conversion enzyme; Phenylethanolamine N-methyltransferase (PNMT) which results in reduced conversion of noradrenaline to adrenaline and a predominant noradrenergic secretory pattern [26]. In contrast, 'cluster 2' tumours are predominantly located within the adrenal gland and have a more mature phenotype associated with normal or elevated PNMT expression and a mixed or predominately adrenergic secretory pattern [26]. In addition

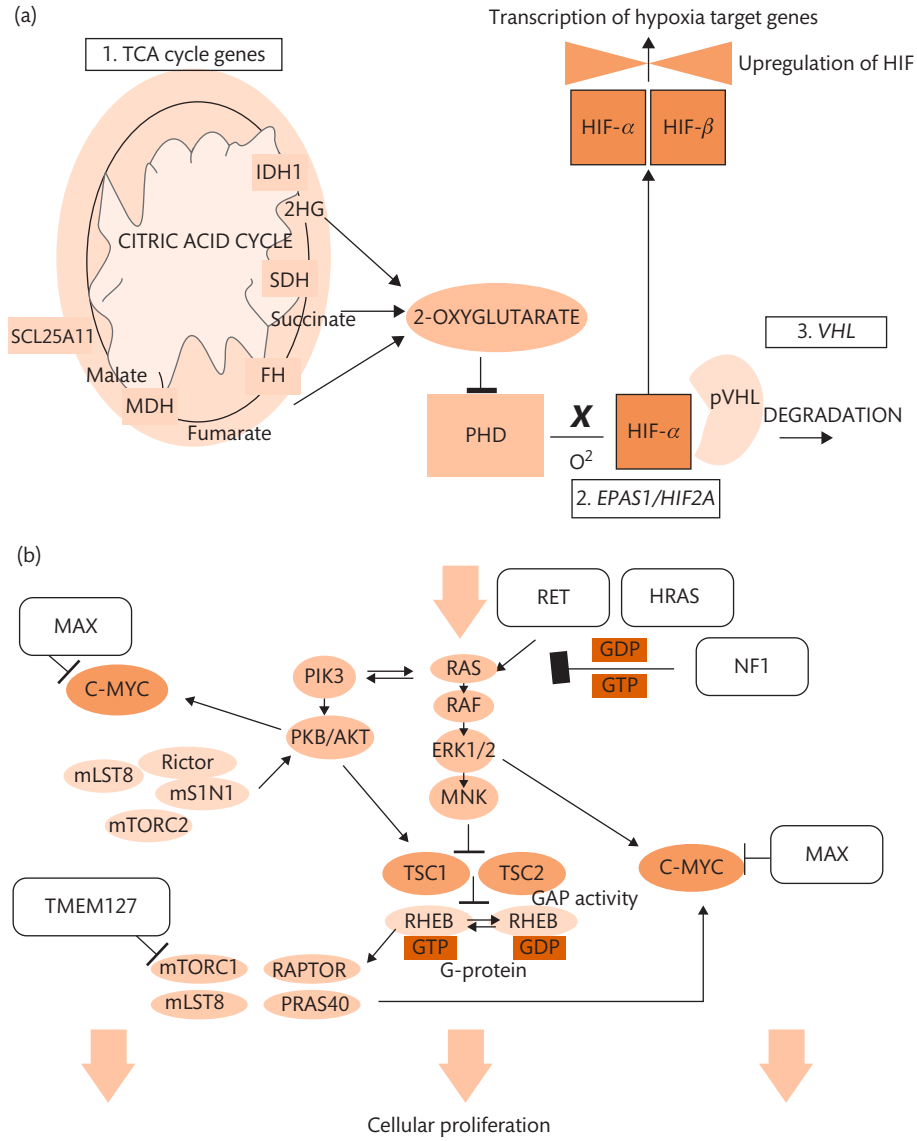


Figure 5.5.1.2 (a) Cluster 1/ 'Pseudohypoxic' genes. (b) Cluster 2 genes.

Figure A demonstrates how oncometabolite accumulation (e.g. succinate, fumarate, etc)

Modified with permission from Casey, R. (2019). A study of succinate dehydrogenase deficient tumourigenesis: From functional assessment of variant pathogenicity to the discovery of new disease biomarkers (Doctoral thesis). https://doi.org/10.17863/CAM.39029.

to reduced expression of PNMT, *SDHX* mutated tumours also have reduced activity of the enzyme dopamine-β-hydroxylase, responsible for the conversion of dopamine to norepinephrine in the catecholamine synthesis pathway. Therefore, elevated levels of dopamine or its metabolite 3-methoxytyramine is characteristic of SDH-deficient tumours [27]. Elevated dopamine is not typically associated with clinical symptoms but is a significant indicator of poor paraganglia cell differentiation and elevated 3-methoxytyramine levels have been validated as an independent biomarker of malignant disease [26]. Finally, non-secretory PPGL are also more commonly associated with *SDHX* mutations and these tumours frequently have reduced or absent activity of tyrosine hydroxylase, the rate limiting enzyme in catecholamine synthesis [27].

Radionuclide Imaging

Nuclear imaging techniques have been utilized as adjuncts to morphological cross-sectional imaging studies, in the diagnosis and management of PPGL for the past four decades and the tracers can be subclassified based on their target ligand into three groups; (i) catecholamine storage and synthesis (^{123}I-metaiodobenzylguanidine, ^{18}F-fluorodopamine (18F-FDA), and 18F-fluorodihydroxyphenylalanine (^{18}F-FDOPA)), (ii) somatostatin receptor (^{111}Indium-pentetreotide and 68Ga-labelled somatostatin analogue peptides) and iii) glucose metabolism (^{18}F-FDG) (Figure 5.4.1.3). It is widely accepted that the selection of the most appropriate tracer for surveillance or diagnosis can be guided by the mutational status of the PPGL [28]. The tracer ^{123}I-metaiodobenzylguanadine (MIBG) is taken up by the noradrenaline

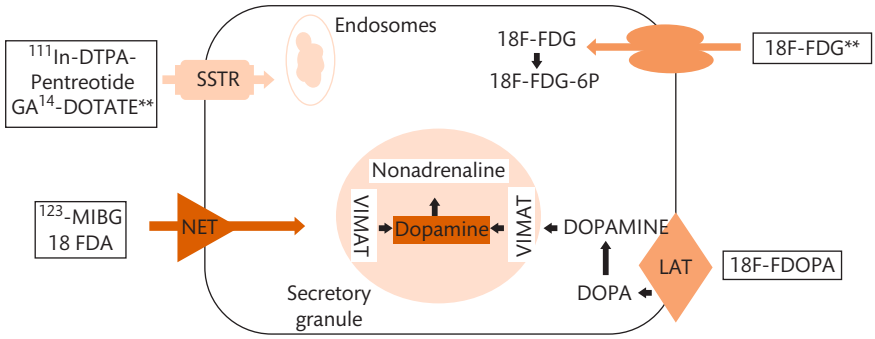


Figure 5.5.1.3 Functional Imaging ligands in PPGL.
** = Improved sensitivity of imaging tracers targeting this ligand in 'pesudohypoxic' PPGL.

transporter (NET) however, the sensitivity of ^{123}I-MIBG scintigraphy is affected by tumour de-differentiation associated with metastatic or aggressive disease, as this is associated with loss of NET expression, therefore increasing the risk of false-negative results using ^{123}I-MIBG scintigraphy. Furthermore, mutations in *SDHX* are associated with a higher risk of false-negative results because of an associated downregulation of the NET transporter in SDH-deficient tumours [29]. Therefore, current recommendations advise that ^{123}I-MIBG scintigraphy should be reserved for those cases being investigated for suitability of treatment with ^{123}I-MIBG radionuclide therapy [10]. Similar issues with sensitivity are seen with the tracer ^{18}F-FDA, also taken up by the NET transporter.

I^{8}F-FDG PET-CT is currently recommended as the molecular imaging modality of choice for the localization of primary and metastatic PPGL [1]. Intriguingly, the sensitivity of ^{18}F-FDG PET-CT also differs depending on the molecular pathways implicated in the development of PPGL, with tumours in the 'pseudohypoxic' cluster demonstrating increased standard uptake values (SUV) of ^{18}F-FDG due to attenuated glycolysis. Recently, evidence has emerged that pseudohypoxic TCA cycle–related PPGL have a high expression of somatostatin receptor type 2 (SSTR2) which improves the sensitivity of somatostatin receptor imaging studies such as 68Ga-DOTATATE PET/CT, which recently demonstrated a sensitivity of 98.6–100% and superior sensitivity to ^{18}F-FDG in two studies [17]. An added benefit of using ^{68}Ga-DOTATATE PET/CT is its predictive value in determining the efficacy of peptide receptor radionuclide therapy (PRRT) with ^{177}Lu-DOTATATE and in the future this may provide an additional therapeutic option for patients with malignant disease.

Chemotherapeutics

Some 10–20% of patients with PPGL will develop malignant disease with the highest risk in patients with germline pathogenic *SDHB* variants [8]. The 5-year survival if malignant disease develops is less than 50% [8]. After surgical resection, effective treatment options are limited and are aimed at controlling tumour proliferation but also tumour secretion of catecholamines to reduce the morbidity associated with excess catecholamines. Systemic treatments currently available include: cytotoxic chemotherapy consisting of a regime of cyclophosphamide, vincristine, and dacarbazine (CVD) and PRRT with radiolabelled (^{131})I-metaiodobenzylguanidine (MIBG) therapy. Recent advances in genomics have suggested novel targeted approaches to molecular therapies for patients with metastatic PPGL. Thus in patients with cluster 1 (pseudohypoxic) PPGL the tumours are characterized by stabilization of HIF and upregulation of glycolytic genes and vascular epidermal growth factor. Receptor tyrosine kinase inhibitors such as sunitinib and other antiangiogenic agents would be predicted to be suitable for PPGL and other tumours harbouring *HIF2A* mutations as well as germline and somatic variants in other cluster 1 genes. Hypermethylation is another unique biological feature of PPGL driven by mutations in the TCA cycle genes and phase 2 clinical trials for demethylating agents are in progress for patients with *SDHX* or *FH* mutated PPGL. In addition, it has been suggested that TCA cycle gene mutation-associated PPGL might be vulnerable to targeting with poly(ADP)-ribose polymerase (PARP) inhibitors [30]. In contrast, for cluster 2 PPGL, though malignant disease is rare in this category, targeted therapies might be directed towards specific kinase pathways (e.g. tyrosine kinase inhibitors (*RET*), ERK1/2 inhibitors (*RAS*), or MEK inhibitors (*NF1*)).

Head and Neck Paragangliomas

Clinical

HNPGL (parasympathetic-derived paragangliomas) are most common in the fourth and fifth decades and are more frequent in females than males. More than 50% of HNPGL arise in the carotid body (known as carotid body paraganglioma or chemodectoma) and, in contrast to sympathetic paragangliomas, metastatic disease is infrequent (<5%). Carotid body HNPGL usually present with a slowly enlarging neck mass and only rarely secrete catecholamines. Glomus jugulare or glomus tympanicum tumours may be associated with partial or complete hearing loss and pulsatile tinnitus. HNPGL may be familial or bilateral (in which case a genetic cause should be suspected) and/or there may be a previous history of PPGL, wild-type gastrointestinal stromal tumour (wtGIST) or renal cell carcinoma (which are each associated with germline mutations in *SDHX* genes. In patients with sporadic HNPGL, about 20% of patients may have a genetic cause [31].

Genetics

Familial and/or multicentric HNPGL most often results from germline mutations in the SDH complex genes (*SDHA, SDHB,*

SDHC, SDHD, and *SDHAF2*). Which *SDHX* gene harbours a mutation is critical for determining the risk to other family members. The most commonly mutated gene in HNPGL, *SDHD*, displays a parent-of-origin effect in penetrance such that individuals who inherit a *SDHD* mutation from their father are at high risk of HNPGL (and also PPGL) whereas if the mutation is inherited from their mother the risk of disease is very low (see Chapter 6.12). Inherited *SDHAF2* mutations show similar parent-of-origin effects on tumour risk. Mutations in *SDHB, SDHC*, and *SDHA* do not demonstrate any parent-of-origin effects and are inherited in a conventional autosomal dominant manner. Though each of these genes encodes an SDH complex protein the tumour-specific risks vary (e.g. for *SDHD* and *SDHC* the risks of HNPGL are higher than those for PPGL whereas the reverse is true for *SDHB* mutations). While *SDHA* mutations are infrequent in patients with HNPGL and PPGL it is the most commonly mutated gene in wtGIST [32]. Individuals presenting with a HNPGL who are found to harbour a *SDHX* gene mutation are at risk of further HNPGL and other *SDHX*-related tumours and require lifelong surveillance and their relatives should be offered genetic investigations/surveillance also (see Chapter 6.12). Occasionally HNPGL may be associated with other inherited causes of PPGL (e.g. *VHL, TMEM127, MAX*) [33].

Management

Symptomatic HNPGL are usually treated by surgery but in some locations (e.g. carotid body tumours, vagal, and jugular paragangliomas) resection of more advanced tumours can be associated with a significant risk of vascular/cranial nerve damage and external beam radiotherapy may be preferable [34]. Small asymptomatic HNPGL may be detected in hereditary mutation carriers and, if multiple, may be kept under surveillance though the risk of malignancy in *SDHB* mutation carriers should be considered.

Neuroblastoma

Clinical

This tumour of developing postganglionic sympathetic neurons occurs in the adrenal medulla or paraspinal ganglia and is the most common solid extracranial tumour in children. Most tumours arise in the abdomen. Mean age at diagnosis is 1.5 years and 90% of cases occur before the age of 10 years [35]. The prognosis is variable with survival ranging from 95% to 50% in low- and high-risk groups, respectively. A variety of clinical, pathological, and molecular characteristics (e.g. age at diagnosis, histopathology, tumour *MYCN* amplification, and chromosomal deletions) are used to categorize affected children into several groups (low to ultra-high) and guide management. A striking feature of neuroblastoma in some patients (particularly young infants) is spontaneous regression, even if metastatic disease is present [35]. Most cases are sporadic and familial cases are rare (up to 2% of all cases) but have provided important insights into the molecular basis of this disorder.

Genetics of Familial Neuroblastoma

Inherited susceptibility to neuroblastoma may be occur as part of well-recognized syndromes such as Beckwith–Wiedemann syndrome and hemihypertrophy, NF1 and RASopathy disorders, congenital central hypoventilation syndrome (Ondine's course) and Hirschsprung disease [36]. In addition, familial (non-syndromic) neuroblastoma may be inherited as an autosomal dominant trait with incomplete penetrance.

Genetic linkage studies in kindreds with familial non-syndromic neuroblastoma localized a gene to chromosome 2p23–p24 and sequencing of the *ALK* (anaplastic lymphoma kinase) proto-oncogene identified recurrent missense variants that act as activating mutations in the ALK tyrosine kinase domain [37]. Germline pathogenic variants in *ALK* are found in about 80% of familial neuroblastoma kindreds [38]. In addition, about 10% of sporadic neuroblastomas harbour somatic *ALK* mutations [39]. Functional studies of mutant *ALK* have identified genotype-phenotype-functional correlations such that the familial mutations with stronger activating effects (e.g. R1275Q) are associated with more complete penetrance than those with weaker activating effects (G1128A) and somatic mutations with the strongest activating effects (e.g. F1174* and F1245*) have only very rarely been detected in the germline but were associated with severe neurodevelopmental anomalies [36].

Idiopathic congenital central hypoventilation syndrome (CCHS) is a rare autosomal dominantly inherited disorder characterized by abnormal autonomic control of ventilation resulting in cyanosis and hypercapnia during sleep. Additional features may include Hirschsprung disease and other features of autonomic nervous system dysfunction (excessive sweating, pupillary abnormalities, disordered body temperature regulation, etc.) [40]. Though the disorder may be fatal in the neonatal period, long-term survivors have been reported. CCHS is the most frequent syndromic cause of neuroblastoma and, in 2003, idiopathic CCHS was shown to be associated with germline mutations in the paired-like homeobox gene *PHOX2B* [37].

RASopathy disorders comprise a group of rare congenital syndromes that result from mutations in genes that encode components of the RAS signalling pathway. These disorders include Costello syndrome (caused by activating mutations in the *HRAS* proto-oncogene), Noonan syndrome (*PTNP11, KRAS, NRAS, SOS1, BRAF, RAF1, MEK1, RIT1*), Cardiofaciocutaneous syndrome (*BRAF, MAP2K1, MAP2K2, KRAS*) and NF1 [41]. Of these disorders, the cancer risk is highest in Costello syndrome, which is characterized by developmental delay, short stature (though birth weight may be increased), facial dysmorphisms, and cardiac anomalies.

Beckwith–Wiedemann syndrome (BWS) is a congenital disorder characterized by variable features including macroglossia, anterior abdominal wall defects, pre-and/or postnatal overgrowth, neonatal hypoglycaemia, lateralized overgrowth (hemihypertrophy and embryonal tumour predisposition in 5–10% of cases. BWS is a human imprinting disorder that is characterized by altered expression/function of the IGF2 growth suppressor and/or the CDKN1C growth suppressor. BWS can be caused by several molecular mechanisms and the embryonal tumour risk varies between molecular subgroups [42]. Overall, the most common embryonal tumour associated with BWS is Wilms tumour, but the neuroblastoma is the most frequent tumour in children with germline *CDKN1C* mutations (~4% of cases) [42].

Neuroblastoma has been associated with germline *TP53* mutations in a few cases (though neuroblastoma is not considered a characteristic Li-Fraumeni syndrome tumour) and rarely has occurred with a germline *SDHB* mutation [41]. In addition, investigations of familial neuroblastoma to identify rare variants in inherited neuroblastoma genes, investigators have undertaken genome-wide association studies (GWAS) in sporadic cases to identify common variants predisposing to neuroblastoma variants. More than 15 susceptibility loci have been identified including variants at (or close to) *BARD1, MLF1, CPZ, CASC15, LIN28B, LMO1, HSD17B12*, and *TP53* [36]. Several of the genes linked to neuroblastoma susceptibility by GWAS studies have also been implicated in tumour behaviour.

Somatic Genetics of Neuroblastoma

More than 30 years ago, amplification of the *MYCN* proto-oncogene was identified as a key driver of neuroblastoma oncogenesis [35]. *MYCN* amplification is associated with advanced tumour stage and poorer patient survival and evaluation of tumour NMYC status is a key component of pretreatment evaluation protocols that use clinical and biological data to categorize patients into the risk groups that determine individual clinical management. A number of recurrent copy number abnormalities have been described in neuroblastoma including loss of 1p, 4p, 6q, 11q, and 14q and gain at 1q, 2p, and 17q. Loss of 1p and 11q are features of high risk neuroblastoma but whereas 1p loss and 17q gain are correlated with *MYCN* amplification, 11q loss is inversely correlated with *MYCN* amplification [35]. Other somatic findings in neuroblastoma include *ALK* amplification (*ALK* and *MYCN* are both map to 2p24) and mutations, *ATRX, ARID1A,* and *ARID1B* mutations, and *TERT* promoter rearrangements. Though investigation for tumour *MYCN* status and copy number abnormalities (e.g. 11q loss) is routinely performed to establish risk group and guide treatment, other genomic profiling investigations (e.g. *ALK* mutation analysis) are not universal though will likely be adopted as targeted therapies become available. A well-recognized feature of neuroblastoma is the occurrence of spontaneous regression in a subset of cases. Several mechanistic aetiologies have been suggested but no genomic biomarker has been identified to date to predict which cases will regress.

Genetic Testing and Surveillance

In non-syndromic patients, testing for germline mutations associated with inherited neuroblastoma is indicated if there is a family history of neuroblastoma or if there are multifocal or bilateral tumours. In addition, the presence of clinical features of a syndrome associated with neuroblastoma predisposition is an indication for specific genetic testing to be initiated.

The neuroblastoma risk in individuals with a germline *ALK* or *PHOX2B* mutation has been estimated at up to 50% and such individuals should be offered regular surveillance as recommended by AACR Childhood Cancer Predisposition Group and including abdominal ultrasound, measurement of urinary catecholamine metabolites (VMA and HVA) and chest radiography until age 10 years [41]. An International Consensus Group recommended three monthly abdominal ultrasound scans until age 7 years in children with *BWS* and a *CDKN1C* mutation [41]. Routine screening for neuroblastoma is generally not indicated in RASopathy disorders but should be considered in Costello syndrome (though elevated urinary VMA and HVA can occur in the absence of a tumour) [41]. About 15% of cases of familial neuroblastoma do not have an identifiable germline mutation and surveillance should also be considered for at risk children with a strong family history.

Conclusions

PPGL, HNPGL, and neuroblastomas are all tumours of the autonomic nervous system but display differing clinical and molecular characteristics. In the past two decades there has been important progress in elucidating the molecular basis of familial and sporadic forms of these disorders and this information has increasingly been used to inform the management of affected individuals and, in inherited cases, their families. For example, the identification of germline mutations in patients with PPGL and HNPGL is now recognized as an important aspect of clinical care. The identification of individuals with inherited neuroblastoma is also important but for this tumour, the characterization of somatic genetic changes (e.g. *MYCN* and 11q deletion status) is an established part of routine clinical management and the role of molecular characterization for prognostic and treatment stratification will undoubtedly continue to expand.

REFERENCES

1. Lenders JWM, Duh Q-Y, Eisenhofer G, *et al.* Pheochromocytoma and paraganglioma: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 1915–42.
2. Kopetschke R, Slisko M, Kilisli A, *et al.* Frequent incidental discovery of phaeochromocytoma: data from a German cohort of 201 phaeochromocytoma. *Eur J Endocrinol*, 2009; **161**: 355–61.
3. Young WF. Paragangliomas: clinical overview. *Ann N Y Acad Sci*, 2006; **1073**: 21–9.
4. Lloyd RV, Osamura RY, Kloppel G, Rosai J. *WHO Classification of Tumours: Pathology and Genetics of Tumours of Endocrine Organs*, 4th edn. Lyon: WHO, 2017.
5. McNichol AM. Differential diagnosis of pheochromocytomas and paragangliomas. *Endocr Pathol*, 2001; **12**: 407–15.
6. Jimenez C, Libutti S, Landry C, *et al.* Adrenal-neuroendocrine tumors. *AJCC Cancer Staging Manual*, 8th edn. New York: Springer, 2017: 919–27.
7. Ayala-Ramirez M, Feng L, Johnson MM, *et al.* Clinical risk factors for malignancy and overall survival in patients with pheochromocytomas and sympathetic paragangliomas: primary tumor size and primary tumor location as prognostic indicators. *J Clin Endocrinol Metab*, 2011; **96**: 717–25.
8. Gimenez-Roqueplo A-P, Favier J, Rustin P, *et al.* Mutations in the SDHB gene are associated with extra-adrenal and/or malignant phaeochromocytomas. *Cancer Res*, 2003; **63**: 5615–21.
9. Castinetti F, Taieb D, Henry JF, *et al.* Management of endocrine disease: outcome of adrenal sparing surgery in heritable pheochromocytoma. *Eur J Endocrinol*, 2016; **174**: R9–18.
10. Plouin PF, Amar L, Dekkers OM, *et al.* European Society of Endocrinology Clinical Practice Guideline for long-term follow-up of patients operated on for a phaeochromocytoma or a paraganglioma: Table 1. *Eur J Endocrinol*, 2016; **174**: G1–10.

11. Letouzé E, Martinelli C, Loriot C, *et al*. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 2013; **23**: 739–52.
12. Morin A, Letouzé E, Gimenez-Roqueplo A-P, Favier J. Oncometabolites-driven tumorigenesis: from genetics to targeted therapy. *Int J Cancer*, 2014; **135**: 2237–48.
13. Fishbein L, Leshchiner I, Walter V, *et al*. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell*, 2017; **31**: 181–93.
14. Därr R, Nambuba J, Del Rivero J, *et al*. Novel insights into the polycythemia-paraganglioma-somatostatinoma syndrome. *Endocr Relat Cancer*, 2016; **23**: 899–908.
15. Buffet A, Smati S, Mansuy L, *et al*. Mosaicism in HIF2A-related polycythemia-paraganglioma syndrome. *J Clin Endocrinol Metab*, 2014; **99**: E369–73.
16. Gossage L, Eisen T, Maher ER. VHL, the story of a tumour suppressor gene. *Nat Rev Cancer*, 2015; **15**: 55–64.
17. Toledo RA, Qin Y, Cheng Z-M, *et al*. Recurrent mutations of chromatin-remodeling genes and kinase receptors in pheochromocytomas and paragangliomas. *Clin Cancer Res*, 2016; **22**: 2301–10.
18. Luchetti A, Walsh D, Rodger F, *et al*. Profiling of somatic mutations in phaeochromocytoma and paraganglioma by targeted next generation sequencing analysis. *Int J Endocrinol*, 2015; **2015**: 1–8.
19. López-Jiménez E, Gómez-López G, Leandro-García LJ, *et al*. Research resource: transcriptional profiling reveals different pseudohypoxic signatures in SDHB and VHL-related pheochromocytomas. *Mol Endocrinol*, 2010; **24**: 2382–91.
20. Brouwers FM, Elkahloun AG, Munson PJ, *et al*. Gene expression profiling of benign and malignant pheochromocytoma. *Ann N Y Acad Sci*, 2006; **1073**: 541–56.
21. Crona J, Taïeb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification. *Endocr Rev*, 2017; **38**: 489–515.
22. Dahia PLM. Pheochromocytoma and paraganglioma pathogenesis: learning from genetic heterogeneity. *Nat Rev Cancer*, 2014; **14**: 108–19.
23. van Nederveen FH, Gaal J, Favier J, *et al*. An immunohistochemical procedure to detect patients with paraganglioma and phaeochromocytoma with germline SDHB, SDHC, or SDHD gene mutations: a retrospective and prospective analysis. *Lancet Oncol*, 2009; **10**: 764–71.
24. Kimura N, Takayanagi R, Takizawa N, *et al*. Pathological grading for predicting metastasis in phaeochromocytoma and paraganglioma. *Endocr Relat Cancer*, 2014; **21**: 405–14.
25. Bardella C, El-Bahrawy M, Frizzell N, *et al*. Aberrant succination of proteins in fumarate hydratase-deficient mice and HLRCC patients is a robust biomarker of mutation status. *J Pathol*, 2011; **225**: 4–11.
26. Eisenhofer G, Klink B, Richter S, Lenders JW, Robledo M. Metabologenomics of phaeochromocytoma and paraganglioma: an integrated approach for personalised biochemical and genetic testing. *Clin Biochem Rev*, 2017; **38**: 69–100.
27. Eisenhofer G, Pacak K, Huynh T-T, *et al*. Catecholamine metabolomic and secretory phenotypes in phaeochromocytoma. *Endocr Relat Cancer*, 2011; **18**: 97–111.
28. Timmers HJLM, Chen CC, Carrasquillo JA, *et al*. Staging and functional characterization of pheochromocytoma and paraganglioma by 18f-fluorodeoxyglucose (18F-FDG) positron emission tomography. *JNCI J Natl Cancer Inst*, 2012; **104**: 700–8.
29. Fonte JS, Robles JF, Chen CC, *et al*. False-negative ^{123}I-MIBG SPECT is most commonly found in SDHB-related pheochromocytoma or paraganglioma with high frequency to develop metastatic disease. *Endocr Relat Cancer*, 2012; **19**: 83–93.
30. Sulkowski PL, Sundaram RK, Oeck S, *et al*. Krebs-cycle-deficient hereditary cancer syndromes are defined by defects in homologous-recombination DNA repair. *Nat Genet*, 2018; **50**: 1086–92.
31. Piccini V, Rapizzi E, Bacca A, *et al*. Head and neck paragangliomas: genetic spectrum and clinical variability in 79 consecutive patients. *Endocr Relat Cancer*, 2012; **19**: 149–55.
32. Evenepoel L, Papathomas TG, Krol N, *et al*. Toward an improved definition of the genetic and tumor spectrum associated with SDH germ-line mutations. *Genet Med*, 2015; **17**: 610–20.
33. Boedeker CC, Hensen EF, Neumann HPH, *et al*. Genetics of hereditary head and neck paragangliomas. *Head Neck*, 2014; **36**: 907–16.
34. Suárez C, Rodrigo JP, Mendenhall WM, *et al*. Carotid body paragangliomas: a systematic study on management with surgery and radiotherapy. *Eur Arch Otorhinolaryngol*, 2014; **271**: 23–34.
35. Matthay KK, Maris JM, Schleiermacher G, *et al*. Neuroblastoma. *Nat Rev Dis Prim*, 2016; **2**: 16078.
36. Ritenour LE, Randall MP, Bosse KR, Diskin SJ. Genetic susceptibility to neuroblastoma: current knowledge and future directions. *Cell Tissue Res*, 2018; **372**: 287–307.
37. Mosse YP, Laudenslager M, Khazi D, *et al*. Germline PHOX2B mutation in hereditary neuroblastoma. *Am J Hum Genet*, 2004; **75**: 727–30.
38. Tolbert VP, Coggins GE, Maris JM. Genetic susceptibility to neuroblastoma. *Curr Opin Genet Dev*, 2017; **42**: 81–90.
39. Bresler SC, Weiser DA, Huwe PJ, *et al*. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. *Cancer Cell*, 2014; **26**: 682–94.
40. Weese-Mayer DE, Rand CM, Berry-Kravis EM, *et al*. Congenital central hypoventilation syndrome from past to future: model for translational and transitional autonomic medicine. *Pediatr Pulmonol*, 2009; **44**: 521–35.
41. Kamihara J, Bourdeaut F, Foulkes WD, *et al*. Retinoblastoma and neuroblastoma predisposition and surveillance. *Clin Cancer Res*, 2017; **23**: e98–106.
42. Brioude F, Kalish JM, Mussa A, *et al*. Expert consensus document: clinical and molecular diagnosis, screening and management of Beckwith–Wiedemann syndrome: an international consensus statement. *Nat Rev Endocrinol*, 2018; **14**: 229–49.

5.5.2 Management of Phaeochromocytoma and Paraganglioma

Henri Timmers

Introduction

Definition

A phaeochromocytoma is a rare neuroendocrine tumour arising from adrenomedullary chromaffin cells that produces one or more catecholamines: adrenaline, noradrenaline, and dopamine [1]. The name is derived from the Greek terms *phaios* (dark, dusky) and *chroma* (colour), referring to the chromium staining characteristics of tumour tissue when treated with chromium salts. Paragangliomas are histologically similar to phaeochromocytomas and also derived from the neural crest, but they do not arise from the adrenal medulla but develop from extra-adrenal sympathetic-derived chromaffin tissues, in particular sympathetic paravertebral ganglia of the thorax, abdomen, and pelvis. Phaeochromocytoma and paraganglioma together will be referred to here as PPGL. The majority of PPGLs, around 80–85%, are located in the adrenal gland. Typical locations for extra-adrenal PPGLs are: (1) the Zuckerkandl body, a sympathetic ganglion located at the root of the inferior mesenteric artery; (2) the sympathetic plexus of the urinary bladder, the kidneys and the heart; (3) the sympathetic ganglia in the mediastinum. Head and neck paragangliomas, located along the glossopharyngeal and vagal nerves in the neck and at the base of the skull, represent the parasympathetic counterparts of PPGL. They are also known as glomus tumours and they mainly derive from the glomus caroticum, glomus (jugulo-)tympanicum, and glomus vagale. In general they do not produce clinically significant amounts of catecholamines.

Epidemiology

The prevalence of PPGL in patients with hypertension in general outpatient clinics varies between 0.2% and 0.6%. In children with hypertension, the prevalence is estimated at 1.7%. Autopsy studies demonstrate undiagnosed PPGLs in 0.05–0.1% of cases and up to 50% of these unrecognized tumours may have contributed to mortality. Therefore, the true prevalence is probably higher. PPGLs are increasingly diagnosed following the incidental discovery of an adrenal mass on abdominal imaging studies for unrelated disorders, nowadays in approximately 30% of cases. Around 7% of adrenal incidentalomas prove to be a phaeochromocytoma [2]. PPGLs can occur at any age, but they mostly present in adults between the ages of 20 and 50 years. Both genders are affected equally. In at least one-third of all patients, PPGL is part of a hereditary syndrome caused by germline mutations. The genetic aspects of PPGL and associated syndromes are discussed in Chapter 5.5.1. Patients with hereditary PPGLs more often present with multifocal disease and at a younger age than those with sporadic neoplasms.

Clinical Significance

The secretion of catecholamines by PPGLs can be devastating due to potentially lethal cardiovascular complications. Therefore, prompt diagnosis is essential for effective treatment, usually by surgical resection, which is curative in most cases. A subset of PPGLs have malignant potential. Metastases develop in around 10–17% of patients, but the malignancy rate is highly variable among different hereditary subforms. Considering the high prevalence of underlying genetic disorders, the diagnosis of PPGL may trigger the discovery of neoplasms in family members. For these reasons, it is crucially important to recognize symptoms and signs and apply an optimal strategy for biochemical testing, tumour localization, and management. In 2014, an Endocrine Society USA clinical practice guideline on the diagnosis and management of PPGL has become available [3].

Pathogenesis

The pathophysiology of PPGL is largely driven by genomic alterations [4]. At least 35% of PPGLs occur as part of a hereditary disorder. In contrast to sporadic PPGLs, which usually involve a unilateral adrenal lesion, familial PPGLs are more often bilateral, extra-adrenal, multifocal, and occur at a younger age. In the context of hereditary disease, PPGL may concurrently present with other syndromal features, including tumours of other neuroendocrine or non-neuroendocrine tissues [5]. The genetic and syndromal aspects of PPGL are discussed in detail in Chapter 5.5.1.

The principal genetic drivers in hereditary PPGL are germline mutations in tumour susceptibility genes, in particular succinate dehydrogenase (SDH) subunits A, B, C, and D and assembly factor 2, Von Hippel–Lindau (VHL), *RET* (causing multiple endocrine neoplasia type 2), neurofibromatosis type 1 *(NF1), MAX, TMEM127*, and more recently discovered fumarate hydratase and malate dehydrogenase 2. Besides these germline mutations, tumourigenesis in 30–40% of PPGLs is driven by somatic mutations, mainly in *HRAS, NF1, EPAS1/HIF2A, RET, VHL*, and *CSDE1*, or by fusions of mastermind-like transcriptional coactivator 3 (*MAML3*).

Based on transcriptional profiling studies, PGLs can be classified into different clusters [6]. Tumours in the first cluster include those associated with *SDH, VHL, fumarate hydratase, malate dehydrogenase*, and *EPAS1/HIF2A* and are characterized by increased expression of genes involved in (pseudo)hypoxia, cell proliferation, angiogenesis, electron transport chain, and the Krebs cycle and abnormal function of oxidoreductases. Tumours in the second cluster include those associated with *RET, NF1, TMEM127* and *MAX* show an increased expression of genes involved in kinase signalling, protein synthesis, endocytosis, and maintenance of a differentiated

chromaffin cell catecholamine biosynthetic and secretory phenotype. The third cluster involves alterations in the Wnt signalling pathway of tumourigenesis and is driven by *MAML3* and *CSDE1*.

SDH subunit B mutations are of particular interest, because they have been associated with large sized aggressive tumours and increased risk of malignancy. SDH is part of the tricarboxylic acid (TCA) cycle where it converts succinate to fumarate and it also serves as complex II of the mitochondrial electron transport chain. *SDH* mutations result in the accumulation of succinate which acts as an oncometabolite and triggers the stabilization of hypoxia-inducible transcription factors which induce the expression of genes with hypoxia response elements that support tumour progression via different signalling pathways [7]. In addition, DNA methylation profiling revealed that *SDH subunit B* mutated PPGLs are characterized by a pattern of global histone and DNA hypermethylation [8]. These epigenetic changes are also induced by succinate through inactivation of histone demethylases and account for several of the characteristics of this subset of tumours, including cell migration/metastatic capacity and catecholamine secretory profile.

Pathology

In general, PPGLs are well-circumscribed, encapsulated tumours [9, 10]. Most PPGL range in size from 3 to 5 cm, but diameters up to 20 cm have been reported. Incidentally discovered PPGLs and those diagnosed during tumour screening in mutation carriers are usually smaller. PPGLs may show a variety of architectural patterns. Typical patterns include nests of cells ('Zellballen'), anastomosing trabecular cords of cells or, most commonly, a mixture of both. See Figure 5.5.2.1.

Occasionally, composite tumours, usually consisting of PPGL combined with ganglioneuroblastoma or ganglioneuroma, are seen. The tumour cells are usually polygonal with an intermediate amount of lightly coloured eosinophilic granular cytoplasm. Cells may vary in size from small to large. Nuclei are well demarcated and generally eccentric in location. Nuclear pleomorphism with enlargement and hyperchromatism may be seen.

According to the revised 2017 WHO classification of endocrine tumours, the qualifiers 'benign' or 'malignant' have no meaning in the context of PPGLs since these tumours all have some metastatic potential, regardless of location or histology [9]. Lesions that qualify as metastases occur in non-chromaffin tissues such as bone, lung, liver, and lymph nodes. Features that are associated with the development of metastases include extra-adrenal tumour location, tumour size larger than 5 cm and **succinate dehydrogenase subunit B** mutations. In addition, various histological features of the primary tumour may help to stratify the risk of subsequent development of metastatic disease. These include the presence of invasion of soft tissue, vascular penetration, hypercellularity, comedo-type necrosis, and a high mitotic count or Ki-67 proliferation index. Coarse nodularity, decreased numbers of sustentacular cells, or absence of hyaline globules may also be associated. On an individual basis, it is difficult to quantify the risk, however. Several histological grading systems have been proposed, such as the most well-known Phaeochromocytoma of the Adrenal Scaled (PASS) score. These scores are hampered by considerable interobserver variability among pathologist and require validation in independent cohorts.

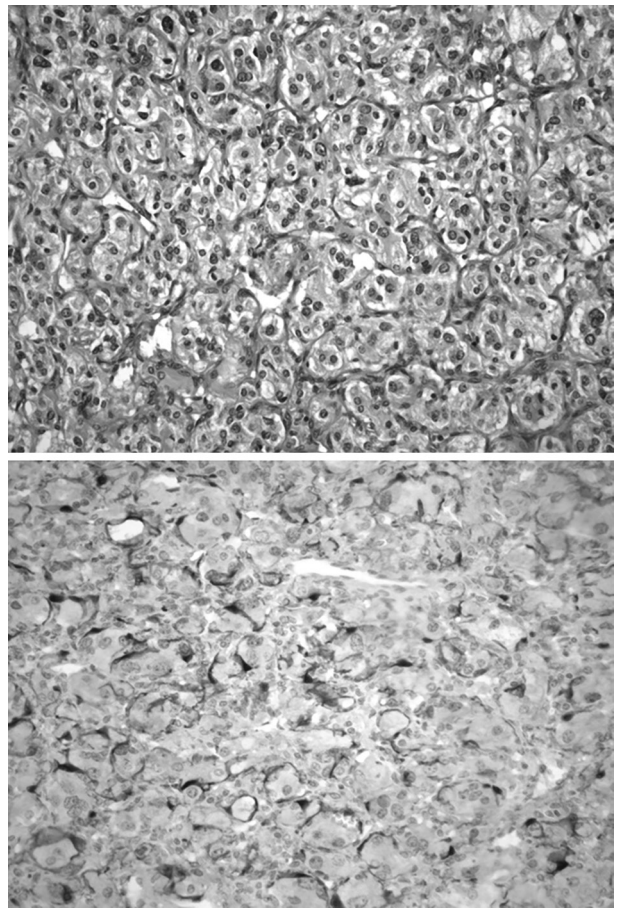


Figure 5.5.2.1 Histology of PPGL. High power view of left adrenal phaeochromocytoma showing typical nested growth pattern ('Zellballen'). Upper panel: haematoxylin & eosin stain. Lower panel: S-100 stain.

Currently there are also no reliable molecular markers for malignancy of PPGL.

The pathological differential diagnosis of PPGL includes oncocytic adrenocortical tumours and (metastases of) renal cell carcinoma and, in case of extra-adrenal paraganglioma, low-grade neuroendocrine tumours. Most, but not all PPGLs show immunohistochemical expression of chromogranin in tumour cells and S100 in sustentacular cells. Presence of steroidogenic factor-1 expression reliably distinguishes adrenocortical tumours from phaeochromocytoma. Staining patterns of keratins may be helpful to distinguish paragangliomas from neuroendocrine tumours.

Clinical Features

Symptoms and Signs

The manifestations of PPGL are diverse, and the tumour can mimic a variety of conditions, often resulting in erroneous diagnoses. Most but not all the clinical signs and symptoms of PPGL are due to the direct actions of secreted catecholamines and the associated short-term activation of the sympathetic nervous system [1]. Hypertension, tachycardia, pallor, headache, and feelings of panic

Table 5.5.2.1 Clinical features of PPGL

Symptoms	Signs
Headaches	Sustained hypertension
Palpitations	Paroxysmal hypertension
Sweating ('diaphoresis')	Tachycardia or reflex bradycardia
Anxiety/panic	Hyperglycaemia
Tremulousness	Postural hypotension
Nausea/vomiting	Weight loss
Abdominal pain	Pallor
Chest pain	Tremor
Tiredness/weakness	Increased respiratory rate
Dizziness	Decreased gastrointestinal motility
Heat intolerance	Psychosis (rare)
Paraesthesiae	Flushing (rare)
Constipation	
Dyspnoea	
Visual disturbances	
Seizures, grand mal (rare)	

Symptoms and signs are listed in order of decreasing frequency

or anxiety, usually dominate the clinical presentation. See Table 5.5.2.1.

Hypertension in patients with PPGL can be either persistent or paroxysmal, but is absent in a minority of patients. Metabolic effects include hyperglycaemia, lactic acidosis, and weight loss. Less common signs and symptoms are nausea, fever, and (pale) flushing. Hypertension is often paroxysmal in nature, in some patients occurring on a background of sustained hypertension, whereas others can have normal blood pressure between paroxysms. Hypertensive episodes can be severe and result in hypertensive emergencies. Blood pressure can also be consistently normal, especially in case of early discovery of small tumours as well as in (very) rare cases of dopamine secretion only or biochemically silent PPGL. Occasionally, patients with predominantly epinephrine-secreting tumours and/or cosecretion of other vasodilatory neuropeptides present with hypotension or even shock. Other substances besides catecholamines that can be secreted by PPGL include vasoactive intestinal peptide, serotonin, histamine, calcitonin, insulin-like growth factor-2, and adrenocorticotrophic hormone, the latter causing ectopic Cushing's syndrome in rare cases. Catecholamine excess may also trigger cardiovascular emergencies, such as myocardial infarction, cardiac arrhythmias, dissecting aortic aneurysm, toxic cardiomyopathy, hypertensive encephalopathy, cerebrovascular accident, neurogenic pulmonary oedema, and even sudden death. See Table 5.5.2.2.

Anaesthesia and tumour manipulation are the most well-known stimuli to elicit a catecholaminergic crisis. Tyramine containing foods, micturition (urinary bladder PPGL), and various drugs such as metoclopramide, tricyclic antidepressants, and glucocorticoids might also induce paroxysms. See Table 5.5.2.3.

Usually, however, spells are unpredictable. For most patients they last between several minutes and one hour. Intervals between attacks may vary from hours to months.

Table 5.5.2.2 Catecholamine-related complications of PPGL

Organ system	Complication
Cardiovascular	Hypertensive crisis
	Shock
	Acute heart failure
	Myocardial infarction
	Arrhythmias
	(Takutsubo) cardiomyopathy
	Dissecting aortic aneurysm
	Limb ischaemia
Pulmonary	Pulmonary oedema
Abdominal	Abdominal haemorrhage
	Paralytic ileus
	Acute intestinal obstruction
	Bowel ischaemia and perforation
	Mesenteric vascular occlusion
	Acute pancreatitis
	Cholecystitis
	Megacolon
Neurologic	Stroke
	Encephalopathy
Renal	Acute renal failure
Metabolic	Diabetic ketoacidosis
	Lactic acidosis
Multisystem	Multiple organ failure, fever

As in adults, PPGLs in children usually present with hypertension and symptoms of catecholamine excess, but the clinical picture may be dominated by central nervous system disturbances such as syncope, cyanotic episodes, cerebral encephalopathy, and blurred vision.

Differential Diagnosis

The individual symptoms are quite non-specific, in particular headaches, palpitations, and sweating, which are the most frequent. This

Table 5.5.2.3 Medications that are contraindicated in PPGL because of potential adverse effects

Dopamine D_2 receptor antagonists	(antiemetics and antipsychotics)
Sympathomimetics	(e.g. ephedrine)
Non-selective β-adrenoreceptor blockers	
Opioid analgesics	
Noradrenaline reuptake inhibitors	(including tricyclic antidepressants)
Monoamino oxidase inhibitors	
Corticosteroids	(particularly if high doses are given)
Peptides	(e.g. ACTH, glucagon)
Neuromuscular blocking agents	(used for anaesthesia)
Serotonin reuptake inhibitors	(rare)

Box 5.5.2.1 Differential diagnosis of PPGL

Cardiovascular
Arrhythmias
Angina pectoris
Heart failure
Baroreflex failure
Postural orthostatic tachycardia syndrome
(Pre)eclampsia

Endocrine
Thyrotoxicosis
Carcinoid
Menopausal syndrome
Hypoglycaemia
Mastocytosis
Medullary thyroid carcinoma

Neurological
Migraine
Stroke
Focal epilepsy
Intracranial lesions
Cerebral vasculitis

Miscellaneous
Anxiety disorder
Related to medication
Related to illegal drugs (e.g. cocaine)
Factitia
Porphyria

can cause a considerable delay in the diagnosis in many patients. Nevertheless, if all three symptoms present together, the specificity of this combination is reported to be more than 90%. Differential diagnostic considerations are provided in Box 5.5.2.1.

These include anxiety or panic disorder, drug use, baroreflex failure, hyperthyroidism, hypoglycaemia, menopausal flushes, and migraine must also be considered. Many of these conditions can be excluded readily on the basis of a good history and physical examination. When paroxysms involve chest pain, differentiation from an acute myocardial infarction may be difficult because angina pectoris and myocardial damage may occur in the absence of coronary artery disease in a patient with PPGL. Many non-specific electrocardiographic changes have been reported, as well as various supraventricular and ventricular tachycardias. Hypertension from a PPGL during pregnancy can mimic pre-eclampsia.

Biochemical Diagnosis

Clinical Settings for Testing

Testing for PPGL should be considered in several clinical settings [3]. The diagnosis should be prompted by symptoms of PPGL, in particular when they are paroxysmal and when they are precipitated by the use of medications that affect catecholamine release. See Table 5.5.2.3. Screening for PPGL should also be performed in all (non-symptomatic) patients with an accidentally discovered adrenal mass, regardless of the presence of hypertension. In addition, periodic testing is performed in patients with a relevant hereditary predispositions (see Chapter 5.5.1) or a history of previous PPGL [11]. Not all patients who present with new onset hypertension need to be tested but only those with additional clues for catecholamine excess. Initial testing is also required in patients with unexplainable variability of blood pressure or a paradoxical blood pressure response to anaesthesia, surgery, or medications.

Catecholamine Synthesis and Metabolization

The diagnosis of PPGL primarily relies on the measurement of catecholamine metabolites in plasma or urine. Biosynthesis of catecholamines starts with conversion of tyrosine to dihydroxyphenylalanine (DOPA) by the enzyme tyrosine hydroxylase. See Figure 5.5.2.2.

DOPA is converted to dopamine which is translocated from the cytoplasm into catecholamine storage vesicles of chromaffin cells. The presence of the enzyme dopamine-ß hydroxylase within these vesicles is responsible for conversion of dopamine to noradrenaline. In adrenal medullary chromaffin cells, noradrenaline is further converted to adrenaline by phenylethanolamine-*N*-methyltransferase. Since this enzyme is only present in these cells and requires the presence of cortisol as a cofactor, adrenaline is nearly exclusively produced within the adrenal medulla. Metabolism—and thereby inactivation—of catecholamines occurs through different pathways resulting in numerous metabolites. The majority of circulating noradrenaline is derived from noradrenergic neurons of the central and sympathetic nervous system. Deamination of neuronal noradrenaline occurs by monoamine oxidase after neuronal reuptake or after leakage of the transmitter from storage vesicles into the neuronal cytosol. Noradrenaline is also partially metabolized in extraneuronal tissues and adrenal chromaffin cells, where it is converted to normetanephrine by catechol-*O*-methyltransferase. Adrenaline is mainly metabolized within adrenal chromaffin cells into the O-methylated metabolite metanephrine. Metabolism of dopamine follows other pathways resulting in production of the *O*-methylated metabolite methoxytyramine. Plasma-free metanephrines and

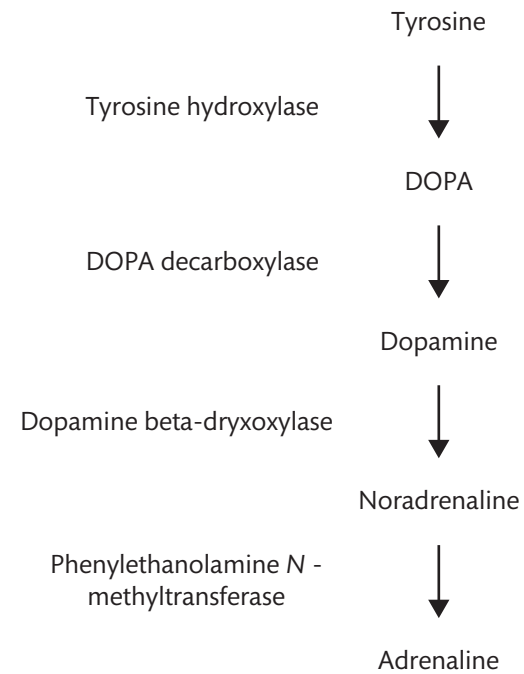


Figure 5.5.2.2 Catecholamine synthesis pathway. Tyrosine is converted to dihydroxyphenylalanine (DOPA) then to dopamine, noradrenaline, and adrenaline. The conversion of noradrenaline to adrenaline is catalysed by the enzyme phenoxyethanolamine N-methyltransferase in the adrenal medulla, which is dependent on cortisol as a cofactor.

methoxytyramine are subsequently conjugated to sulphates by gut wall enzymes and then excreted by the kidneys.

Choice of Biochemical Test

Measurements of plasma-free metanephrines or 24 h urinary fractionated metanephrines are the tests of first choice since they have a higher diagnostic accuracy than the measurement of catecholamines or other metabolites [3]. It is well established that measurements of urinary and plasma catecholamines are insufficiently reliable since catecholamine secretion in PPGLs is often episodic or even negligible in asymptomatic patients. This higher diagnostic accuracy of metanephrines can be attributed to continuous intratumoural production and secretion of metanephrines into the circulatory compartment. This secretion is independent of highly variable catecholamine release caused by the tumour or by sympathoadrenal excitation. Both plasma-free and 24-h urinary fractionated metanephrines offer high sensitivities between 96% and 99% and a specificity between 69% and 93% [12]. In children, the diagnostic performance of plasma or urine metanephrines is similarly high. Plasma concentrations of metanephrine on average are higher in children than in adults, but also higher in boys than in girls. It is advised to apply sex and age adjusted reference intervals [13].

The measurement of metanephrines can be achieved by several analytical techniques. Earlier procedures utilizing radioenzymatic assays have generally been superseded by highly sensitive and specific high-performance liquid chromatography methods utilizing electrochemical detection or, more efficient, in combination with tandem mass spectrometry.

Interpretation of Biochemical Test Results

The interpretation of biochemical test results for the diagnosis of PPGL can be challenging and several pitfalls need to be considered. Elevated plasma levels of catecholamines or metanephrines are not specific for PPGL and do not always prove the presence of PPGL but could also reflect increased sympathetic activity. Several preanalytical factors may affect test results, such as exercise, posture, food, stress, caffeine, smoking, hypoglycaemia, and medications; these factors may alter production or disposition of catecholamines and their metabolites. A blood sample for measurement of metanephrines should ideally be taken after supine rest in a quiet room for at least 20–30 minutes, since samples obtained in sitting position without preceding rest provide a lower diagnostic accuracy. Numerous food products, including banana, pineapple, nuts, and cereals, contain substantial quantities of biogenic amines that may produce false-positive test results. Dietary restrictions prior to testing, however, appear warranted for the measurement of the dopamine metabolite methoxytyramine only.

Medications and intoxications can interfere analytically, pharmacodynamically, and pharmacokinetically, resulting in usually false-positive test results. See Box 5.5.2.2.

Analytical interference can occur with some, but not all, liquid chromatography with electrochemical detection assays. Tandem mass spectrometry is less susceptible to analytical interference. Pharmacodynamic and pharmacokinetic interference involves the effects of drugs on secretion, metabolism, and excretion of catecholamines or metabolites. Numerous drugs are known to increase catecholamine and metabolite concentrations, resulting in falsely elevated test results [14]. Examples are sympathomimetic agents such as ephedrine, amphetamine, cocaine, caffeine, and nicotine, agents that inhibit reuptake of noradrenaline such as serotonin/noradrenaline reuptake inhibitors and tricyclic antidepressants, agents that inhibit catecholamine metabolism such as monoamino oxidase inhibitors and vasodilators such as calcium-channel blockers and α- and β-adrenergic receptor blockers. False-positive results can also occur with Levodopa, used in Parkinson's disease.

Box 5.5.2.2 Medications that may cause false elevated metanephrines

Pharmacodynamic or pharmacokinetic interference
Tricyclic antidepressants
Phenoxybenzamine
Monoamine oxidase inhibitors
Levodopa
Alpha-methyldopa
Sympathomimetics
Calcium-channel blockers
Caffeine
Nicotine
Sympathomimetics

Analytical interference (with some LC-ECD assays)
Coffee (including decaffeinated coffee)
Labetalol
Paracetamol
Sotalol
Alpha-methyldopa
Buspirone
Sulphasalazine
Levodopa
Alpha-methyldopa
Sympathomimetics

Clonidine Suppression Test

Initial test results of plasma-free metanephrines are not always immediately conclusive. When metanephrines are mildly elevated (<3–4 times the upper reference limit) the question arises whether the test result is true or false-positive. First, measurements of plasma-free metanephrines can be repeated after optimizing sampling conditions and elimination of potential interfering drugs and substances which may be responsible for false-positive results. If the tests are still inconclusive, a clonidine suppression test can be performed [15]. Clonidine suppresses release of neuronal noradrenaline (and thereby of normetanephrine) by stimulating brain and prejunctional neuronal α2-adrenergic receptors. For the test, clonidine is administered orally at a dose of 300 μg/70 kg body weight. An abnormal test result indicating a PPGL includes an elevation of plasma normetanephrine at 3 hours after clonidine administration and a less than 40% decrease in levels compared with baseline. If these criteria are not met, sympathetic activation rather than PPGL is the cause of elevation. Both sensitivity and specificity of the test are high. Since clonidine may induce hypotension, blood pressure should be intermittently monitored throughout the procedure. The glucagon stimulation test, previously used to unmask the presence of a PPGL, is obsolete due to substantial risk of hypertensive complications and insufficient sensitivity.

Biochemical Phenotype

Once biochemical diagnosis of PPGL is established, tumours can be divided into three groups according to their biochemical phenotype.

Noradrenergic tumours secrete mainly noradrenaline while adrenergic tumours secrete mainly adrenaline in addition to a varying amount of noradrenaline. The third group is formed by tumours which (co)secrete dopamine. The secretion of metanephrines is relatively higher in noradrenergic tumours than adrenergic tumours and is positively correlated with tumour size [16]. The biochemical phenotype can also be an indicator of tumour location since adrenaline and metanephrine secretion is usually confined to adrenal PPGLs, whereas extra-adrenal tumours secrete predominantly or exclusively noradrenaline and normetanephrine. These differences are attributed to the paracrine action of cortisol, which induces the expression of adrenal phenylethanolamine-*N*-methyltransferase, the enzyme that converts noradrenaline to adrenaline. The biochemical phenotype also strongly depends on the underlying genotype. PPGLs in the context of multiple endocrine neoplasia type 2 and neurofibromatosis type 1 are adrenergic, whereas those related to Von Hippel–Lindau syndrome and **succinate dehydrogenase** mutations are almost always noradrenergic and in the latter case regularly also dopaminergic [17]. Dopamine excess typically does not cause endocrine manifestations.

Tumour Localization

Anatomical Imaging

Localization of PPGL should be initiated only if a biochemical diagnosis has been established. In patients with a hereditary predisposition, less compelling biochemical evidence might justify imaging studies. Particularly, familial PPGLs related to mutations in succinate dehydrogenase subunit B and D can be completely negative on biochemical testing and therefore imaging represents the principal means for diagnosis. The same applies to most head and neck paragangliomas. The goals of imaging studies are localization of the primary tumour, detection of multiple primary tumours and of metastases. First line anatomical imaging modalities for imaging are computed tomography (CT) and magnetic resonance imaging (MRI), the latter being preferred in children and young adults to limit radiation exposure. Both techniques provide a high sensitivity (88–100%) and allow precise tumour delineation, which is critical for presurgical evaluation. Use of ultrasound is usually not recommended due to its suboptimal sensitivity. The administration of both CT and MRI contrast agents is safe and does not require adrenergic blockade [18]. The radiological appearance of PPGL is variable. The tumours can be homogeneous or heterogeneous, including features of necrosis, haemorrhage, cystic changes, and calcifications. In representative areas of the tumour, however, unenhanced attenuation on CT is almost invariably >10 Hounsfield units as opposed to most adrenal adenomas [19, 20]. Contrast washout is unreliable to distinguish between PPGL and adenoma. Disadvantages of CT besides radiation burden are contrast related allergies and nephropathy. On MRI, a high intensity (bright) T_2-weighted signal is typical of PPGL, but this occurs in only one-third of the tumours. Artefacts due to surgical clips of previous surgery may hamper the localization of recurrent disease by MRI. In addition, the specificity of anatomical imaging is limited and therefore functional imaging can be of additional value.

Functional Imaging

Lesions detected by anatomical imaging can be specifically identified as PPGL by functional imaging agents that target the catecholamine synthesis, storage, and secretion pathways of chromaffin tumour cells [21]. In this respect, the most widely used imaging modality is [^{123}I]-metaiodobenzylguanidine (MIBG) single-photon emission computed tomography (SPECT). See Figure 5.5.2.3.

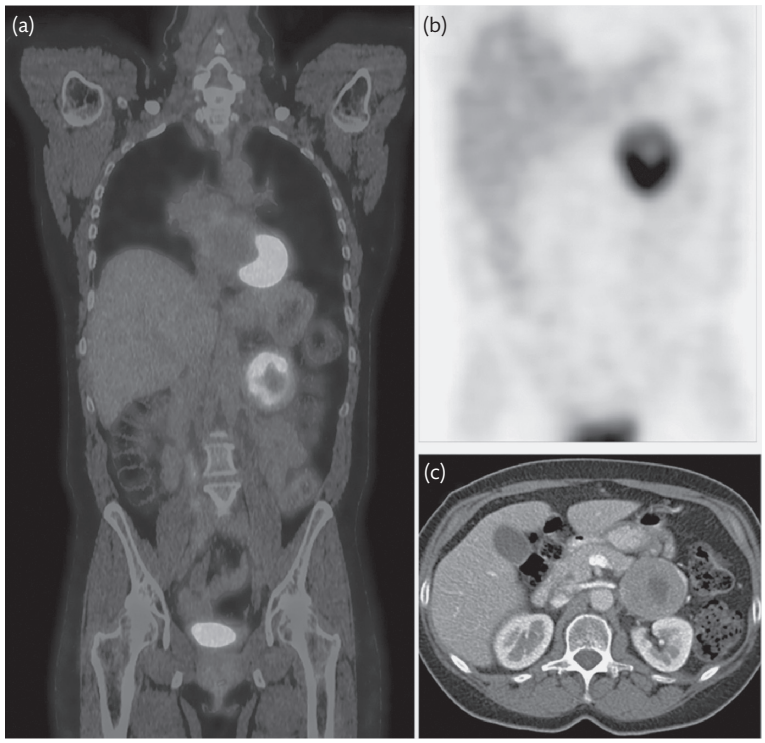


Figure 5.5.2.3 Imaging results in patient with left adrenal phaeochromocytoma. (a) [^{18}F]-FDG PET; (b) [^{123}I]-MIBG single-photon emission computed tomography; (c) computed tomography (CT).

Table 5.5.2.4 Radiopharmaceuticals for the functional imaging of PPGL

Molecular targets	Tracers
Norepinephrine transporter	[^{123}I]-MIBG, 6-[^{18}F]-fluorodopamine, [^{11}C]-epinephrine, [^{11}C]-hydroxyephedrine
Neutral amino acid transporter System L	[^{18}F]-fluoro-L-3,4-dihydroxyphenylalanine
Somatostatin receptors	[^{68}Ga]-DOTATATE, DOTANOC, DOTATOC
Glucose transporters	[^{18}F]-fluorodeoxyglucose

The norepinephrine analogue MIBG targets the norepinephrine transporter of the PPGL cell membrane and the vesicular monoamine transporters in the membrane of intracellular vesicles. See Table 5.5.2.4.

[^{123}I]-MIBG is typically preferred to [^{131}I]-MIBG for imaging because [^{123}I] lacks β emissions and has a shorter half-life (13.2 hours compared with 8 days), resulting in overall lower radiation exposure for the patient. Alternative tracers that accumulate in the tumour cells through the same mechanisms, but are detected by positron emission tomography (PET), are 6-[^{18}F]-fluorodopamine, [^{11}C]-epinephrine and [^{11}C]-hydroxyephedrine. However, the availability of these latter radiotracers for clinical practice is limited. MIBG scanning can be used in patients with an increased risk for metastatic disease due to large size of the primary tumour or to extra-adrenal, multifocal or recurrent disease [22]. Its use is essential in patients with metastatic PPGL when radiotherapy using [^{131}I]-MIBG is planned. Because up to 50% of normal adrenal glands demonstrate physiological uptake of MIBG, false-positive results can occur when the uptake is asymmetric. The sensitivity of MIBG scans to detect PPGL is 85–88% for adrenal, 56–75% for extra-adrenal and 56–83% for metastatic tumours. False negative results are most common (~50%) in PPGLs related to mutations in succinate dehydrogenase. Practical disadvantages of MIBG scintigraphy are a delay in scanning until 24 h post-injection, potential drug interference (e.g. false negative results due to labetalol, tricyclic antidepressants, sympathomimetics, and calcium-channel blockers) and the need for thyroid protection by pretreatment with iodine.

A clearly better sensitivity for detection of metastases than that achieved by MIBG SPECT can be provided by PET. 2-[^{18}F]-Fluoro-2-deoxy-D-glucose (FDG) PET provides an index of intracellular glucose metabolism and is taken up by the tumour cell through the membranous glucose transporters. Its routine application is recommended for the staging and monitoring of metastatic PPGL, in particular when caused by SDH subunit B mutations [23]. Similar or even higher sensitivities can be accomplished by using [^{18}F]-fluoro-L-3,4-dihydroxyphenylalanine (DOPA), which targets the amino acid transporters of neuroendocrine cells, or by using the [^{68}Ga] or [^{111}In] labelled somatostatin analogues DOTATATE, DOTANOC, or DOTATOC [24]. Moreover, for comprehensive localization of metastases of the bone, a frequent site of involvement in malignant PPGL, PET is superior to anatomical imaging.

Treatment and Prognosis

Medical and Surgical Management

Optimal management of PPGL warrants a multidisciplinary team of (paediatric) endocrinologists, surgeons, anaesthesiologists, genetic counsellors, radiologists, and cardiologists with expertise in the field. This calls for centralization of care for these patients. Non-metastatic PPGL can be cured by surgical resection. However, before any surgery is undertaken, the deleterious effects of catecholamine secretion should be under control. The main goals of preoperative management are to relieve symptoms and signs of catecholamine excess, including normalization of blood pressure and heart rate, volume repletion and prevention of anaesthesia and surgery-induced catecholamine storm and its consequences on the cardiovascular system [25]. Potentially lethal complications, including hypertensive crisis, myocardial infarction, congestive heart failure (e.g. due to (Takutsubo) cardiomyopathy) and cerebrovascular events, should be minimized. Retrospective studies suggested that risk factors for haemodynamic instability during surgery include large tumour size, high noradrenaline concentration, and high preoperative blood pressure. There are no evidence-based protocols for the medical management. Local practices are wide ranging and there are international differences in available and approved medical therapies. An example approach to the preoperative management is provided in Box 5.5.2.3.

There is extensive clinical experience with the irreversible and non-selective α-adrenergic receptor blocker phenoxybenzamine. Treatment is primarily aimed at blocking catecholamine-mediated vasoconstriction. Adverse reactions include postural hypotension, syncope, nasal congestion, reflex tachycardia, drowsiness, and fatigue. Over at least 2–4 weeks prior to surgery, doses are gradually increased until blood pressure is steadily controlled. This preoperative medical preparation takes place either in the outpatient or inpatient clinic, depending on patient-related factors (e.g. disease severity, geographical considerations) and local experience. In general, suggested targets are a supine systolic/diastolic blood pressure below 130/80 mmHg and a systolic blood pressure in the upright position between 90 and 110 mmHg. Doses of phenoxybenzamine up to 70 mg twice daily maybe necessary. Alternatively, doxazosin,

Box 5.5.2.3 Preparation for PPGL surgery (to be customized for the individual patient)

Start α-adrenoreceptor blocker at least 14 days before surgery
(e.g. phenoxybenzamine 10 mg twice daily or doxazosin 4 mg once daily)
Monitor side effects, supine and upright blood pressure, and heart rate (at least twice daily) and body weight (daily)
Oral salt and fluid loading
(e.g. NaCl 250 mmol = 15 grams daily)
Gradually increase α-adrenoreceptor blocker until haemodynamic goals* are met
(e.g. until maximum dose of phenoxybenzamine 70 mg twice daily or doxazosin 24 mg twice daily)
Add β-adrenoceptor blocker after at least 3–4 days of β-adrenoreceptor blocker, when significant tachycardia or a catecholamine-induced arrhythmia occurs
(e.g. metoprolol, atenolol, or propranolol)
Gradually increase β-adrenoceptor blocker until haemodynamic goals* are met
Add calcium-channel antagonist when haemodynamic goals* are not met despite maximum α-adrenoreceptor-blockade
(e.g. nipedipine or amlodipine)
Administer a saline infusion (e.g. 2L NaCL 0.9% in 24 hours) 24 h before surgery

* Haemodynamic goals: supine systolic/diastolic blood pressure <130/80 mmHg, systolic blood pressure in the upright position between 90 and 110 mmHg, heart

prazosin, or terazosin may be used. These are all competitive and selective α1-adrenergic receptor blockers, with no effect on presynaptic α2-adrenergic receptors and therefore the negative feedback on catecholamine release is preserved, reducing the risk of reflex tachycardia. Doxazosin may need to be titrated up to 24 mg twice daily. Retrospective studies demonstrated that α1-selective adrenergic receptor blockers were associated with lower preoperative diastolic blood pressure, a lower intraoperative heart rate, better postoperative haemodynamic recovery and fewer adverse effects than non-selective α-adrenergic blockers [26]. Another study did not show any difference between the two classes of drugs [27]. Prospective trials investigating the effectiveness and safety of different α-adrenergic blocking therapies are under way.

B-adrenergic blocking agents such as propranolol, atenolol, and metoprolol are needed only when significant tachycardia occurs after α-blockade or for treatment of catecholamine-induced arrhythmia. Suggested targets for heart rate are below 80 beats per minute in the supine position and <100 beats per minute while standing. A (non-selective) β-blocker should never be used in the absence of sufficient α-adrenergic blockade because the former will exacerbate epinephrine-induced vasoconstriction by inhibiting β2-mediated vasodilation. This will make hypertensive episodes worse in subjects on a β-blocker alone. The use of labetalol is not recommended because of a fixed ratio of α- to β-antagonistic activity (1:5).

A-methyl-paratyrosine (metyrosine) is a competitive inhibitor of tyrosine hydroxylase, the rate-limiting step in catecholamine biosynthesis. It can also be used in the treatment of patients with metastatic PPGL. The drug is not generally available in all countries and institutions. Side effects such as sedation, depression, anxiety, and extrapyramidal manifestations relate to inhibition of catecholamine synthesis in the brain and can be severe. Alternatively, calcium-channel blockers such as nifedipine, amlodipine, and verapamil can be used or added, if maximum doses of α-adrenergic blockers are insufficient.

Usually, oral salt and fluid loading or intravenous fluids are given to expand blood volume to prevent preoperative orthostatic hypotension and severe postoperative hypotension. The rationale behind this recommendation is based on the notion that PPGL is associated with a decreased intravascular volume, which is restored under influence of treatment with α-adrenoceptor antagonists. It is common practice to administer a saline infusion (e.g. 2L NaCL 0.9% in 24 hours) during the 24 h before surgery [27]. Medications that can elicit a catecholamine surge should be avoided. See Table 5.5.2.3.

There is no consensus on the optimal anaesthetic management, as randomized controlled trials on this subject are not available. As a minimum requirement, anaesthesiologic management should be in experienced hands. Despite preoperative blockade, intraoperative blood pressure can vary widely, necessitating the use of short acting drugs such as phentolamine, nicardipine, and sodium nitroprusside. Perioperative glucocorticoid stress dosing to avoid Addison crisis should be applied in case of bilateral adrenalectomy, (partial) rest-adrenalectomy, and the rare event of preoperative Cushing syndrome due to mixed pheochromocytoma/adrenocortical adenoma or ectopic adrenocorticotrophic hormone (ACTH) production [28].

With regard to the surgical resection of both adrenal and extra-adrenal retroperitoneal PPGL, open surgery has been largely replaced by minimally invasive laparoscopic surgery via either the retroperitoneal or transperitoneal route. Open resection is recommended for large (e.g. >6 cm) or invasive PPGLs to ensure complete tumour resection, prevent tumour rupture and avoid local recurrence [3]. The optimal surgical approach for extra-adrenal paraganglioma depends on the tumour location, invasion of surrounding structures and size. In one of the largest surgical series reported so far, perioperative mortality and morbidity were 2.4% and 23.6%, respectively [29]. In the setting of bilateral adrenal PPGL or a high chance of developing contralateral disease and a low chance of metastases, for example in case of multiple endocrine neoplasia type 2, cortex-sparing partial adrenalectomy can be attempted. The latter approach particularly pertains to situations where contralateral adrenalectomy has already been performed previously and loss of adrenocortical function and subsequent hormone replacement could be prevented. However, this is at the cost of an increased risk of local recurrence.

Postoperatively, the patient is monitored in the ICU for at least 24 hours to manage potential complications such as rebound hypotension and hypoglycaemia. Hypotension can be overcome by volume loading and vasopressor therapy with noradrenaline under strict monitoring of haemodynamic parameters. Post-surgical hypertension on the other hand can occur as a consequence of non-radical resection, pain, or volume overload and should be treated accordingly.

Follow-up After Surgical Excision

In 2016, a European Society of Endocrinology clinical practice guideline on the follow-up of patients operated on for a PPGL was published [30]. These recommendations were based on a systematic review of the literature regarding the incidence of recurrence (local or metastatic) or new tumours in patients operated on for PPGL with apparently complete tumour resection. The 5-year incidence of recurrence was estimated at 4.7% (95% CI: 4.0, 6.1), distributed as follows: new tumours 22%, local recurrences 23% and metastatic recurrences 55%. The risk of new events persists in the long-term and is higher for patients with genetic or syndromic diseases. Plasma or urinary metanephrines should be determined between 2 and 6 weeks after recovery from surgery to establish biochemical cure. Subsequently, a minimum follow-up of 10 years by yearly clinical and biochemical investigations is recommended. In high-risk patients, i.e. young patients and those with a hereditary disease, a large tumour, and/or an extra-adrenal PPGL, lifelong annual follow-up should be offered. For patients with known genetic diseases, follow-up should be tailored to the specific condition. For patients with a familial paraganglioma syndrome due to succinate dehydrogenase mutation, the follow-up should also include periodic imaging, regardless of the results of metanephrines, to screen for biochemically silent PPGL and head and neck paraganglioma, as well as other associated tumours such as renal cell carcinoma, gastrointestinal stromal tumours, pituitary adenoma, carcinoid tumours and, rarely, neuroblastoma.

Metastatic PPGL

A minority of patients with PPGL (10–17%) develop metastatic disease. In around half of these patients metastases are already present at the initial diagnosis ('synchronous'), whereas metachronous metastases may develop in the course of years or decades after surgery. Predictors of a malignant course include large tumour size,

Box 5.5.2.4 TNM classification of malignant PPGL

Primary tumour size

T_x Primary tumour cannot be assessed
T_1 Tumour <5 cm in greatest dimension, no extra-adrenal invasion
T_2 Tumour ≥5 cm or extra-adrenal paraganglioma of any size, no extra-adrenal invasion
T_3 Tumour of any size with invasion into surrounding tissues (e.g. liver, pancreas, spleen, kidney)

Regional lymph nodes

N_x Regional lymph nodes cannot be assessed
N_0 No lymph node metastasis
N_1 Regional lymph node metastasis

Distant metastasis

M_0 No distal metastasis
M_1 Distant metastasis
M1a: to only bone
M1b: to only distant lymph nodes/liver or lung
M1c: to bone plus multiple other sites

extra-adrenal tumour location, and presence of a germline mutation in SDH subunit B. The TNM staging for PPGL has adopted a cut-off of 5 cm as well as extra-adrenal location as important determinants of the risk of metastasis [31]. See Box 5.5.2.4.

Reliable histological and molecular markers to predict a malignant course in individual patients are lacking. Distant metastases of mainly occur in the lymph nodes (80%), skeleton (72%), liver (50%), and lungs (50%) [32]. See Figure 5.5.2.4.

Local infiltration of the liver, pancreas, kidneys, gastrointestinal tract, and adipose tissue may also occur. Metastases to the brain, breasts, skin, and ovaries have rarely been described.

The natural course of the disease is highly heterogeneous. The overall 5-year survival rate varies between 34% and 60% [33]. Overall survival, progression-free survival, and clinical outcome are difficult to predict on an individual basis. Some patients have indolent disease, despite the presence of extensive distant metastases, especially when confined to the skeleton. These patients may have an excellent quality of life and may need little or no therapeutic intervention for a prolonged period of time. In a French series, half of therapy-naive patients with metastatic PPGL achieved stable disease at 1 year without treatment [34]. On the other hand, some patients exhibit very aggressive disease and lack of response to systemic therapy. Most patients exhibit intermediate outcomes with progressive disease that will require intervention at some point. Several factors have been associated with survival. This includes the presence of a germline mutation in *SDH subunit B* [35], with a prevalence of 30% in patients with metastatic PPGL and up to 90% if the tumour originates from an extra-adrenal location. Other negative predictors of survival are overall tumour burden, timing, and location of metastases, and the degree of catecholamine excess.

No official guidelines on the treatment of metastatic PPGL exist. In any case, successful management again requires a multidisciplinary approach, involving endocrinology, oncology, surgery, radiotherapy, interventional radiology, and pathology. The goals of treatment are to reduce tumour size, treat catecholamine-related adverse events, palliate symptoms related to tumour burden and prevent tumour progression. In case of limited metastatic disease

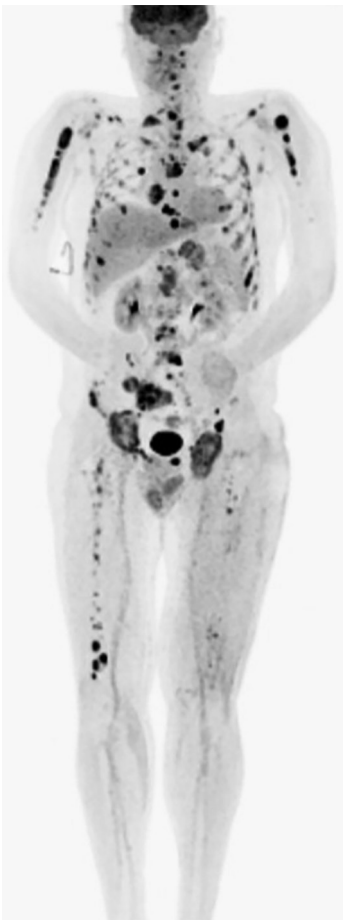


Figure 5.5.2.4 [^{18}F]-FDG PET in patient with widespread bone metastasis of extra-adrenal PPGL due to succinate dehydrogenase subunit B mutation.

such as local malignant lymphadenopathy, surgical metastasectomy can be undertaken with curative intent. Even in the presence of incurable disease, resection (or debulking) of the primary tumour has been shown to be of benefit for survival. Other local therapies include external beam irradiation, radiofrequency ablation, tumour embolization, cryotherapy, and percutaneous microwave coagulation, which can be applied for local symptom relief. Patients with epidural cord compression may benefit from spine surgery as well as external beam irradiation, or a combination of both. Stable compression fractures can be safely treated with percutaneous cement injection. Spinal instability can be treated only with surgery. Prior to these interventions, patients with hormonally active PPGL must be treated with α- and β-adrenergic blockers. Very importantly, these medications are also given to control symptoms of catecholamine excess. In some cases α-methyl-paratyrosine (metyrosine) might be added. Besides hypertension and other cardiovascular manifestations, patients with metastatic PPGL may suffer from severe constipation caused by the inhibitory effects of catecholamines on bowel movement, in particular noradrenaline. This requires vigorous treatment with laxatives.

In case of MIBG positive metastatic lesions on diagnostic scanning, targeted radiotherapy with [^{131}I]-MIBG can be applied. Both high-dose regimens and fractionated low-dose regimens have been shown to be effective in achieving a therapeutic response.

A meta-analysis on the effectiveness of different regimens of [^{131}I]-MIBG radiotherapy, yielded the following response data: a complete response in 3% of patients, a partial response in 27% of patients, and stable disease in 52% of patients [36]. Biochemical responses were as follows: 11% complete response, 40% partial response, and 21% stable disease. [^{131}I]-MIBG treatment is generally well-tolerated, although high-dose regimens may cause bone marrow toxicity. Recently, positive effects have also been reported for ultratrace iobenguane, a more efficiently radiolabelled form of [^{131}I]-MIBG. The use of radiolabelled somatostatin analogues for peptide receptor radionuclide therapy for patients with PPGL metastases that are positive on [^{68}Ga] DOTATATE, DOTANOC, or DOTATOC has not been systematically evaluated but may be effective in some cases.

Regarding chemotherapy, different agents have been used for metastatic PPGL, including mainly cyclophosphamide, doxorubicin, vincristine, dacarbazine, temozolomide, 5-fluorouracil, methotrexate, ifosfamide, streptozotocin, and platinum compounds. The best examined protocol, albeit in retrospective studies only, is a combination of cyclophosphamide, vincristine, and dacarbazine, usually administered in 21-day cycles. Complete and partial radiological response rates were 11% and 44%, respectively [37]. It was shown that besides yielding a potential survival benefit, this combination chemotherapy can facilitate control of blood pressure as well as symptoms [38]. On the other hand, the side effects may negatively impact the patient's quality of life. Patients likely to benefit from chemotherapy are those with rapidly progressive disease and tumour related symptoms that cannot be easily controlled by supportive therapies. The role of chemotherapy in a (neo)adjuvant setting of PPGL remains to be determined. As an oral alternative, temozolomide has been successfully applied to treat *SDH subunit B* related PPGL [39].

In a subset of patients, antiangiogenetic treatment with tyrosine kinase inhibitors such as sunitinib has been associated with partial response, disease stabilization and improved blood pressure control [40]. Prospective clinical trials are now ongoing (http://clinicaltrials.gov/). For future treatment, in the context of a tumour that is highly variable in its genotype, growth-rate, and prognosis, a multiomics approach is being investigated to better stratify PPGLs and guide tailor-made strategies [6]. A proposed algorithm for currently available treatments of metastatic PPGL is provided in Figure 5.5.2.5.

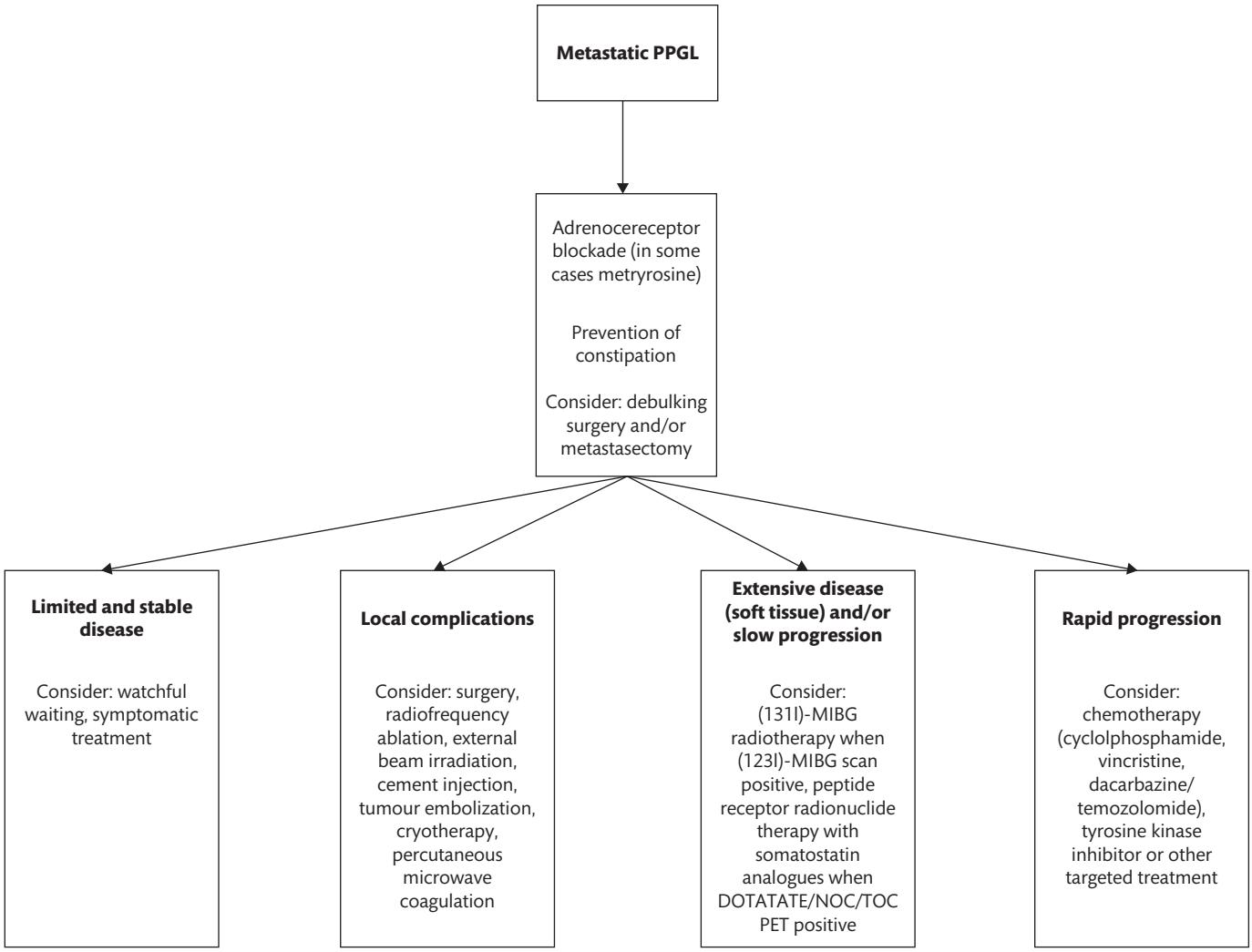


Figure 5.5.2.5 Proposed algorithm for the management of metastatic PPGL.

PPGL in Children

Considering the scarcity of PPGL in children, often in the context of hereditary disease, paediatricians face an even stronger challenge in diagnosing and treating this condition in an evidence-based way.

Although PPGLs are the most common endocrine tumours in children, they account for no more than 10% of all PPGLs, with an incidence of less than 2 per million per year [41]. As compared to adults, PPGLs in children are more commonly familial (80%), extra-adrenal (66%), multifocal (32%), and metastatic (49%) [42]. These features are probably interrelated. Childhood PPGLs peak at 10 to 13 years, with a male predominance (male vs. female, 2:1) before puberty. Regarding management, drug doses and haemodynamic treatment goals need to be adjusted according to age- and weight specific standards.

PPGL in Pregnancy

PPGL in pregnancy is extremely rare, occurring in around 0.002% of all pregnancies [43]. If undiagnosed, maternal and fetal mortality is around 50%. Confusion with the much more prevalent forms of pregnancy-related hypertension such as (pre)eclampsia is the main cause of overlooking the diagnosis. With early detection and proper treatment during pregnancy, mortality can be considerably decreased. For the biochemical diagnosis, plasma or urinary metanephrines can be used since they have a high negative predictive value. For localization, only MRI is suitable. When the diagnosis is made in the first 24 weeks of gestation, it should be (laparoscopically) removed after medical preparation. If the tumour is diagnosed in the third trimester, the patient should be managed using adrenoreceptor blockade until the fetus is viable and Caesarean section with tumour removal in the same session or at a later stage is then preferred.

REFERENCES

1. Lenders JW, Eisenhofer G, Mannelli M, Pacak K. Phaeochromocytoma. *Lancet*, 2005; **366**: 665–75.
2. Fassnacht M, Arlt W, Bancos I, *et al*. Management of adrenal incidentalomas: European Society of Endocrinology clinical practice guideline in collaboration with the European network for the study of adrenal tumors. *Eur J Endocrinol*, 2016; **175**: G1–G34.
3. Lenders JW, Duh QY, Eisenhofer G, *et al*. Pheochromocytoma and paraganglioma: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 1915–42.
4. Fishbein L, Leshchiner I, Walter V, *et al*. Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell*, 2017; **31**: 181–93.
5. Favier J, Amar L, Gimenez-Roqueplo A. Paraganglioma and phaeochromocytoma: from genetics to personalized medicine. *Nat Rev Endocrinol*, 2015; **11**: 101–11.
6. Crona J, Taieb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification. *Endocr Rev*, 2017; **38**: 489–515.
7. Lussey-Lepoutre C, Hollinshead KE, Ludwig C, *et al*. Loss of succinate dehydrogenase activity results in dependency on pyruvate carboxylation for cellular anabolism. *Nat Commun*, 2015; **6**: 8784.
8. Letouze E, Martinelli C, Loriot C, *et al*. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 2013; **23**: 739–52.
9. Lam AK. Update on adrenal tumours in 2017 World Health Organization (WHO) of endocrine tumours. *Endocr Pathol*, 2017; **28**: 213–27.
10. Turchini J, Cheung VKY, Tischler AS, De Krijger RR, Gill AJ. Pathology and genetics of phaeochromocytoma and paraganglioma. *Histopathology*, 2018; **72**: 97–105.
11. van Berkel A, Lenders JW, Timmers HJ. Diagnosis of endocrine disease: biochemical diagnosis of phaeochromocytoma and paraganglioma. *Eur J Endocrinol*, 2014; **170**: R109–19.
12. Lenders JW, Pacak K, Walther MM, *et al*. Biochemical diagnosis of pheochromocytoma: which test is best? *JAMA*, 2002; **287**: 1427–34.
13. Eisenhofer G, Lattke P, Herberg M, *et al*. Reference intervals for plasma free metanephrines with an age adjustment for normetanephrine for optimized laboratory testing of phaeochromocytoma. *Ann Clin Biochem*, 2013; **50**(Pt 1): 62–9.
14. Neary NM, King KS, Pacak K. Drugs and pheochromocytoma--don't be fooled by every elevated metanephrine. *N Engl J Med*, 2011; **364**: 2268–70.
15. Eisenhofer G, Goldstein DS, Walther MM, *et al*. Biochemical diagnosis of pheochromocytoma: how to distinguish true- from false-positive test results. *J Clin Endocrinol Metab*, 2003; **88**: 2656–66.
16. Eisenhofer G, Lenders JW, Goldstein DS, *et al*. Pheochromocytoma catecholamine phenotypes and prediction of tumor size and location by use of plasma free metanephrines. *Clin Chem*, 2005; **51**: 735–44.
17. Eisenhofer G, Lenders JW, Timmers H, *et al*. Measurements of plasma methoxytyramine, normetanephrine, and metanephrine as discriminators of different hereditary forms of pheochromocytoma. *Clin Chem*, 2011; **57**: 411–20.
18. Baid SK, Lai EW, Wesley RA, *et al*. Brief communication: radiographic contrast infusion and catecholamine release in patients with pheochromocytoma. *Ann Int Med*, 2009; **150**: 27–32.
19. Buitenwerf E, Korteweg T, Visser A, *et al*. Unenhanced CT imaging is highly sensitive to exclude pheochromocytoma: a multicenter study. *Eur J Endocrinol*, 2018; **178**: 431–7.
20. Canu L, Van Hemert JAW, Kerstens MN, *et al*. CT characteristics of pheochromocytoma: relevance for the evaluation of adrenal incidentaloma. *J Clin Endocrinol Metab*, 2019; **104**: 312–18.
21. Ilias I, Shulkin B, Pacak K. New functional imaging modalities for chromaffin tumors, neuroblastomas and ganglioneuromas. *Trends Endocrinol Metab*, 2005; **16**: 66–72.
22. van Berkel A, Pacak K, Lenders JW. Should every patient diagnosed with a phaeochromocytoma have a (1)(2)(3) I-MIBG scintigraphy? *Clin Endocrinol*, 2014; **81**: 329–33.
23. Timmers HJ, Chen CC, Carrasquillo JA, *et al*. Staging and functional characterization of pheochromocytoma and paraganglioma by 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography. *J Natl Cancer Inst*, 2012; **104**: 700–8.
24. Janssen I, Blanchet EM, Adams K, *et al*. Superiority of [68Ga]-DOTATATE PET/CT to other functional imaging modalities in the localization of SDHB-associated metastatic pheochromocytoma and paraganglioma. *Clin Cancer Res*, 2015; **21**: 3888–95.
25. Pacak K. Preoperative management of the pheochromocytoma patient. *J Clin Endocrinol Metab*, 2007; **92**: 4069–79.
26. Prys-Roberts C, Farndon JR. Efficacy and safety of doxazosin for perioperative management of patients with pheochromocytoma. *World J Surg*, 2002; **26**: 1037–42.

27. Bruynzeel H, Feelders RA, Groenland TH, *et al*. Risk factors for hemodynamic instability during surgery for pheochromocytoma. *J Clin Endocrinol Metab*, 2010; **95**: 678–85.
28. Pacak K, Timmers H, Eisenhofer G. Pheochromocytoma. In: De Groot LJ, Jameson JL, eds. *Endocrinology, Adult and Pediatric*. 2, 6th edn. Philadelphia, PA: Elsevier Saunders, 2010: 1990–2018.
29. Plouin PF, Duclos JM, Soppelsa F, Boublil G, Chatellier G. Factors associated with perioperative morbidity and mortality in patients with pheochromocytoma: analysis of 165 operations at a single center. *J Clin Endocrinol Metab*, 2001; **86**: 1480–6.
30. Plouin PF, Amar L, Dekkers OM, *et al*. European Society of Endocrinology Clinical Practice Guideline for long-term follow-up of patients operated on for a phaeochromocytoma or a paraganglioma. *Eur J Endocrinol*, 2016; **174**: G1–G10.
31. Jimenez C, Libutti SK, Landry CS, *et al*. The first TNM staging system for patients with pheochromocytomas and paragangliomas. In: Amin MB, ed. AJCC Cancer Staging Manual, 8th edn. New York, NY: Springer, 2017: 919–27.
32. Jimenez P, Tatsui C, Jessop A, Thosani S, Jimenez C. Treatment for malignant pheochromocytomas and paragangliomas: 5 years of progress. *Curr Oncol Rep*, 2017; **19**: 83.
33. Plouin PF, Fitzgerald P, Rich T, *et al*. Metastatic pheochromocytoma and paraganglioma: focus on therapeutics. *Horm Metab Res*, 2012; **44**: 390–9.
34. Hescot S, Leboulleux S, Amar L, *et al*. One-year progression-free survival of therapy-naive patients with malignant pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*, 2013; **98**: 4006–12.
35. Amar L, Baudin E, Burnichon N, *et al*. Succinate dehydrogenase B gene mutations predict survival in patients with malignant pheochromocytomas or paragangliomas. *J Clin Endocrinol Metab*, 2007; **92**: 3822–8.
36. van Hulsteijn LT, Niemeijer ND, Dekkers OM, Corssmit EP. (131)I-MIBG therapy for malignant paraganglioma and phaeochromocytoma: systematic review and meta-analysis. *Clin Endocrinol*, 2014; **80**: 487–501.
37. Huang H, Abraham J, Hung E, *et al*. Treatment of malignant pheochromocytoma/paraganglioma with cyclophosphamide, vincristine, and dacarbazine: recommendation from a 22-year follow-up of 18 patients. *Cancer*, 2008; **113**: 2020–8.
38. Ayala-Ramirez M, Feng L, *et al*. Clinical benefits of systemic chemotherapy for patients with metastatic pheochromocytomas or sympathetic extra-adrenal paragangliomas: insights from the largest single-institutional experience. *Cancer*, 2012; **118**: 2804–12.
39. Hadoux J, Favier J, Scoazec JY, *et al*. SDHB mutations are associated with response to temozolomide in patients with metastatic pheochromocytoma or paraganglioma. *Int J Cancer*, 2014; **135**: 2711–20.
40. Ayala-Ramirez M, Chougnet CN, Habra MA, *et al*. Treatment with sunitinib for patients with progressive metastatic pheochromocytomas and sympathetic paragangliomas. *J Clin Endocrinol Metab*, 2012; **97**: 4040–50.
41. Waguespack SG, Rich T, Grubbs E, *et al*. A current review of the etiology, diagnosis, and treatment of pediatric pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*, 2010; **95**: 2023–37.
42. Pamporaki C, Hamplova B, Peitzsch M, *et al*. Characteristics of pediatric vs adult pheochromocytomas and paragangliomas. *J Clin Endocrinol Metab*, 2017; **102**: 1122–32.
43. Lenders JW. Pheochromocytoma and pregnancy: a deceptive connection. *Eur J Endocrinol*, 2012; **166**: 143–50.

5.6

Primary Aldosteronism

5.6.1 Genetics of Primary Aldosteronism and Other Steroid-Related Causes of Endocrine Hypertension

Maria Christina Zennaro, Fabio Fernandes-Rosa, and Sheerazed Boulkroun

Introduction

Arterial hypertension is a major cardiovascular risk factor. Identification of secondary forms of hypertension is key for targeted management and prevention of cardiovascular complications. Aldosterone plays a major role in regulating blood pressure and electrolyte homeostasis, and primary aldosteronism (PA) has been recognized as the most frequent form of secondary arterial hypertension. The prevalence of PA has been estimated to be around 4 to 6% of hypertensives in primary care [1, 2] and 10% in reference centres [1]. PA is associated with a suppressed renin-angiotensin system and often hypokalaemia; it is due mainly to unilateral aldosterone-producing adenoma (APA) or to bilateral adrenal hyperplasia. In addition to aldosterone excess, some rare disorders affecting corticosteroid production in the adrenal gland or their actions in the kidney also lead to steroid dependent hypertension. These include 11β-hydroxylase (*CYP11B1*) and 17α-hydroxylase (*CYP17A1*) deficiencies, as well as activating mineralocorticoid receptor mutations and the syndrome of apparent mineralocorticoid excess.

Aldosterone regulates blood pressure and electrolyte homeostasis by activating the mineralocorticoid receptor (MR) in the so-called aldosterone sensitive distal nephron (late distal convoluted tubule, connecting tubule and collecting duct), which plays an important role in fine-tuning renal sodium excretion by reabsorbing around 5–10% of the filtered sodium load [3]. The MR is a ligand dependent transcription factor, member of the nuclear receptor superfamily [4]. Following hormone binding, the MR activates transcription of a number of genes coding for proteins involved in regulating transepithelial sodium transport, including the epithelial sodium channel ENaC and the Na^+,K^+-ATPase [5]. Importantly, the MR binds aldosterone, 11-deoxycorticosterone (DOC), corticosterone and cortisol with very similar and high affinity [6], with certain differences in their efficiency of activating gene transcription [7]. Cortisol circulates in plasma at ≈1000-fold higher levels than aldosterone [8]. In epithelial aldosterone target tissues like the distal parts of the kidney, aldosterone selectivity of MR binding is made possible by the action of 11β-hydroxysteroid dehydrogenase type 2 (*HSD11B2*), an enzyme which converts cortisol to its inactive metabolite cortisone [9, 10]. In tissues not expressing 11β-HSD2, or in cases where the enzyme is inactive, the MR is mainly bound by cortisol and acts as a bona fide high affinity glucocorticoid receptor.

Aldosterone is synthetized from cholesterol through sequential enzymatic steps in the adrenal zona glomerulosa (ZG). The enzyme aldosterone synthase (*CYP11B2*), is the rate-limiting enzyme for aldosterone biosynthesis and is expressed exclusively in the ZG. It catalyses the three terminal reactions involving 11β-hydroxylation of DOC to corticosterone and its subsequent 18-hydroxylation and 18-oxidation to 18-hydroxy-corticosterone and aldosterone. Cortisol, the main glucocorticoid in human, is produced in the zona fasciculata (ZF) of the adrenal, where expression of 17α-hydroxylase mediates conversion of 17-deoxy-21-carbon steroids into 17-hydroxylated precursors of cortisol. In the ZF, 11β-hydroxylase (*CYP11B1*) catalyses 11-hydroxylation of 11-deoxycortisol to cortisol.

Aldosterone biosynthesis is tightly regulated to maintain sodium, potassium, and fluid homeostasis by the kidney. The principal regulators are the renin-angiotensin system (RAS), extracellular potassium concentration, and, to a lesser extent adrenocorticotrophic hormone (ACTH) [11]. The RAS responds to changes in renal perfusion and sodium concentrations; increased levels of angiotensin II (AngII) bind to AT1 receptors expressed in the ZG and activate aldosterone production via a Gαq-phospholipase C mediated pathway, increasing inositol 1,4,5-triphosphhate (IP3) and 1,2-diacylglycerol concentrations [12]. IP3 increases intracellular calcium concentrations by calcium release from intracellular stores. Extracellular potassium is a major regulator of aldosterone biosynthesis. Indeed,

the main ionic conductance of ZG cells is that of K^+ and their cell membrane potential closely follows the equilibrium potential of K^+ over a wide range of extracellular K^+ concentrations [13]. Increased extracellular K^+ leads to cell membrane depolarization and opening of voltage-dependent calcium channels, followed by increase of intracellular calcium concentrations. AngII also inhibits potassium channels and the Na^+,K^+-ATPase, inducing cell membrane depolarization and activation of voltage-gated calcium channels [12] (see online Figure 5.6.1.1, left panel). Increase in intracellular calcium induced by either AngII or potassium activates calcium signalling, the main trigger for aldosterone biosynthesis [12]. This mediates, in particular, activation of protein kinases regulating the phosphorylation and activity of different transcription factors, which increase *CYP11B2* transcription and aldosterone synthase expression [14].

Inherited Forms of Primary Aldosteronism

Although the majority of cases of PA is sporadic, in 1–5% of cases familial inheritance can be observed, which is due to different Mendelian forms of the disease (see online Table 5.6.1.1). Those diseases are essentially classified according to the underlying genetic defect as familial hyperaldosteronism (FH) type I to x, with additional numbers being used each time a new gene is identified. In particular, the most frequent group classified as FH-II most likely represents a genetically heterogeneous population, with patients sharing a similar phenotype but different genetic defects, many of which still remain to be identified. Genetic testing of inherited forms of PA is routinely performed in some genetic laboratories and family screening is useful for early diagnosis and targeted management.

Familial Hyperaldosteronism Type I

Familial hyperaldosteronism type I (FH-I), also called glucocorticoid-remediable aldosteronism (GRA), is an autosomal dominant disease characterized by severe hypertension typically occurring during childhood [15, 16]. Patients present with PA of variable severity, bilateral adrenal hyperplasia, with associated adrenal nodules in some cases, and high production of the hybrid steroids 18-hydroxycortisol and 18-oxocortisol [15–17]. The genetic abnormality underlying FH-I is a chimeric gene generated by an unequal crossing-over event, fusing the regulatory regions of *CYP11B1* to the coding sequence of *CYP11B2* [18, 19]. This leads to ectopic expression of *CYP11B2* throughout the adrenal cortex, with inappropriate regulation of aldosterone biosynthesis by ACTH, following the circadian rhythmicity of cortisol production [17, 20]. Patients respond well to treatment with exogenous glucocorticoids, which inhibit ACTH production [21]. The prevalence of FH-I has been estimated at around 0.6% of PA patients [22, 23]. However, the frequency seems to be higher in the hypertensive paediatric population, where a prevalence of the chimeric *CYP11B1*/*CYP11B2* gene of 3% has been reported [24]. Among the peculiar features of FH-I is the high morbidity and mortality at early age, due to the occurrence of early-onset haemorrhagic stroke and ruptured intracranial aneurysms [25].

Familial Hyperaldosteronism Type II

FH-II is the most common form of familiar hyperaldosteronism. It is clinically and biochemically indistinguishable from sporadic PA, not remediable by glucocorticoids, and diagnosis is based on the presence of two or more affected family members and autosomal dominant transmission [26, 27]. Patients show varying aldosterone responses on postural test and to AngII within a given family, and different PA subtypes (APA or bilateral adrenal hyperplasia) are found [23, 27]. Its prevalence is estimated to be 1.2% to 6% in adults with PA [23, 27–29]. The phenotypic variability of FH-II has suggested that FH-II may be associated with a variety of genetic defects [27]. Indeed, following the first description of *KCNJ5* mutations in FH-III [30], germline *KCNJ5* gene mutations were reported in patients with a moderate phenotype resembling FH-II [31]. In particular, affected members of three families carrying the *KCNJ5* p.Gly151Glu mutation and affected members from one family carrying the *KCNJ5* p.Tyr152Cys mutation exhibited a mild phenotype compatible with FH-II [32–34]. Somatic mutations of *KCNJ5* were also reported in APA in patients with familial hyperaldosteronism initially diagnosed as FH-II [33], suggesting that in some cases FH-II may represent familial aggregation of sporadic PA, given the high frequency of PA in patients with hypertension.

The large original family with FH-II described by Stowasser *et al.* [26] has been linked to a locus on chromosome 7p22 [35], although genetic heterogeneity had been shown subsequently [36]. It was only very recently that germline mutations in *CLCN2*, coding for the voltage-gated chloride channel ClC-2, have been identified in young onset PA and in the very same family with FH-II [37, 38]. *CLCN2* is expressed in the human and mouse adrenal cortex and is the main chloride conductor of resting glomerulosa cells [37]. The ClC-2 mutations affect different amino acids located in well conserved domains of the protein. In particular, the ClC-2 p.Gly24Asp mutation, found in a 9-year-old girl with PA, is located in a domain whose inactivation leads to 'open' ClC-2 channels. Functional analysis of the mutant channel indicated that the mutation abolished the voltage- and time-dependent gating of ClC-2, strongly increasing chloride currents at resting potential. In adrenocortical cells, this induced increased aldosterone production, which was mediated by plasma membrane depolarization, opening of voltage-gated calcium channels, increased intracellular Ca^{2+} concentrations and increased expression of *CYP11B2* [37] (see online Figure 5.6.1.1, right panel). The family described by Stowasser carried a p.Arg172Gln mutation, which was identified in four other cases; four additional mutations were identified in unrelated subjects with early-onset PA [38]. Again, the mechanism of autonomous aldosterone production involved gain-of-function mutations, promoting cell membrane depolarization, and opening of voltage-gated calcium channels [38]. These data identify a new genetic defect in patients with early-onset PA and a subset of FH-II and highlight the important role for chloride currents in regulating aldosterone biosynthesis.

Familial Hyperaldosteronism Type III

Patients with FH-III show severe early-onset hypertension, usually before age 20, resistant to medical treatment, and associated with severe hypokalaemia. Urinary concentrations of hybrid steroids 18-oxocortisol and 18-hydroxycortisol are elevated and aldosterone production is not suppressed by dexamethasone [39]. Hyperaldosteronism is typically due to massive bilateral adrenal hyperplasia and requires bilateral adrenalectomy to control blood pressure [39]. The genetic cause of FH-III are mutations in the *KCNJ5* gene, coding for the G protein-activated inward

rectifier potassium channel GIRK4 [30]. In particular, mutations p.Gly151Arg, p.Thr158Ala, p.Ile157Ser and p.Glu145Gln all are associated with a severe form of the disease [31]; other mutations are found in patients with moderate hyperaldosteronism resembling FH-II [32–34]. More recently, aldosterone and cortisol cosecretion has been observed in a patient carrying a germline p.Glu145Gln *KCNJ5* mutation [40]. The patient had shown severe hypertension and hypokalaemia at the age of 2 years and subsequently developed Cushing's syndrome at the age of 20 years, due to massive bilateral adrenal hyperplasia.

KCNJ5 mutations are located within or near the selectivity filter of the channel, which confers the K^+ selectivity to GIRK4. Functional studies have shown that mutations change its ion selectivity, leading to sodium influx into the cell and cell membrane depolarization. This leads to activation of voltage-dependent Ca^{2+} channels, increased intracellular Ca^{2+} concentrations, activation of calcium signalling promoting increased expression of steroidogenic enzymes and aldosterone production [30, 41] (see online Figure 5.6.1.1, right panel). Abnormal coexpression of steroidogenic enzymes, particularly aldosterone synthase (CYP11B2) and 17α-hydroxylase (CYP17A1), in the hyperplastic adrenal cortex of patients with FH-III may explain the abnormally high secretion rate of hybrid steroids in FH-III [42].

Familial Hyperaldosteronism Type IV

Whole exome sequencing in subjects with early-onset PA has identified a recurrent germline mutation p.Met1549Val in *CACNA1H* (coding for the pore-forming α1 subunit of the T-type voltage-dependent calcium channel Cav3.2) in five children with PA manifesting before the age of 10 years [43]. Analysis of their relatives showed that the mutation was inherited in an autosomal dominant manner in four cases, with an incomplete penetrance of the phenotype. In two *CACNA1H* mutation carriers, developmental delay or attention deficit disorder was subsequently observed. Affected patients had normal adrenal imaging findings, but in one case histological examination of the adrenal gland following adrenalectomy showed micronodular adrenal hyperplasia [43]. Functional studies showed that the mutation shifts the activation of the calcium channel towards more negative potentials with loss of normal inactivation, producing increased intracellular calcium entry into the cell due to prolonged opening of the channel [43] (see online Figure 5.6.1.1, right panel). Adrenocortical cells expressing the mutated channel had higher aldosterone production and increased *CYP11B2* mRNA expression, which were reduced by treatment with the T-type calcium channel blocker mibefradil, suggesting that T-type calcium channel inhibitors may be useful for treating patients with *CACNA1H* gain-of-function mutations [44].

Subsequently, four different germline *CACNA1H* mutations were identified in patients with different phenotypic presentations of PA [45]. They included a patient with early-onset hypertension and hyperaldosteronism associated with multiplex development disorder, who carried a *de novo* germline p.Met1549Ile *CACNA1H* mutation. A germline p.Ser196Leu and a p.Pro2083Leu mutation were identified in two kindreds diagnosed as FH-II, and a germline *CACNA1H* missense variant (p.Val1951Glu) was identified in a patient with APA [45]. In all cases, the electrophysiological properties of the channel were affected, with gain-of-function leading to increased calcium entry into the cell. Again, this was followed by increased expression of steroidogenic enzymes and aldosterone production in adrenocortical cells [45] (see online Figure 5.6.1.1, right panel). Identification of *CACNA1H* mutations in patients with different phenotypic presentations of PA, including FH-II, suggests that *CACNA1H* may be a susceptibility gene for different forms of the disease.

CACNA1D Mutations in a Rare Disease of PA, Seizures, and Neurological Abnormalities (PASNA)

De novo germline *CACNA1D* mutations were described in two children with a new form of hyperaldosteronism associated with a complex neurologic disorder (Primary Aldosteronism, Seizures and Neurologic Abnormalities, PASNA) [46]. The syndrome is characterized by early-onset PA, severe hypertension, and hypokalaemia in children showing a complex neurologic syndrome including cerebral palsy and epileptic seizures, with no adrenal hyperplasia visible on imaging. In one patient, treatment with a calcium channel blocker normalized blood pressure, with cure of hypertension and PA on follow up [47]. The *CACNA1D* gene codes for Cav1.3, the voltage-dependent L-type calcium channel subunit alpha-1D, which is commonly mutated in APA at the somatic level [46]. Mutations affect highly conserved residues located in domains responsible for channel opening. They modify channel voltage sensitivity, promoting L-type calcium channel opening at lower voltages. Activation at less depolarizing potentials increases intracellular calcium influx and activates calcium signalling, thereby enhancing aldosterone production (see online Figure 5.6.1.1, right panel).

Somatic Mutations in Aldosterone-Producing Adenoma

Sporadic forms of PA represent the majority of cases and are due either to bilateral adrenal hyperplasia or APA. APA are found in approximately one-third of patients, although frequency may vary depending on diagnostic criteria and thresholds used for subtype identification [2, 48, 49]. By applying whole exome sequencing to paired tumour and germline DNA samples, it has been possible to identify recurrent somatic mutations in different genes coding for ion channels (*KCNJ5* [30] and *CACNA1D* [46, 50]) and ATPases (*ATP1A1* and *ATP2B3*, [50, 51]) regulating intracellular ionic homeostasis and cell membrane potential (see online Figure 5.6.1.1, right panel). Mutations of *KCNJ5* and *CACNA1D* are similar to those identified in FH-III and PASNA and lead to increased intracellular calcium concentrations, which activate calcium signalling and aldosterone biosynthesis. Mutations in *ATP1A1*, coding for the α1 subunit of the Na^+,K^+-ATPase, lead to a loss of pump activity and inward proton or sodium leak, which increase aldosterone production through cell membrane depolarization and increased calcium influx [50, 51] or increased intracellular acidification [52]. *ATP2B3* codes for the plasma membrane calcium-transporting ATPase 3 (PMCA3). Mutations lead to loss of the physiological pump function and reduced Ca^{2+} export; they also lead to severely depolarized membrane potential [51], which is explained by a pathological sodium leak associated with or conducted by the mutant PMCA3 [53]. Finally, somatic mutations in

CTNNB1, coding for β-catenin which plays an important role in adrenal cortex development and tumorigenesis, are found in 2–5% of APA [46, 54, 55], and mutations in *PRKACA* (coding for the cAMP-dependent protein kinase catalytic subunit alpha) in rare cases [56]. Mutations in those genes are also found in cortisol-producing adenoma and adrenocortical cancer; the mechanisms whereby the same mutations lead to different hormonal phenotypes remain yet to be established [57].

KCNJ5 mutations are the most prevalent mutations found in ≈40% of APA, with a higher prevalence in Asian populations [58, 59]. Carriers of *KCNJ5* mutations are younger and more often women [58, 59]. Some studies also report larger tumour size, a ZF-like cellular phenotype, and higher plasma aldosterone levels when compared to non-carriers [59, 60], but these results, as well as other genotype-phenotype correlations, are not replicated in other studies [58, 61–64]. ATPase mutations were reported to be more frequent in men [51] and *CACNA1D* mutations are associated mainly with smaller tumours [46, 50, 51, 54, 58, 65, 66]. *CTNNB1* mutations seem to be more prevalent in women [46, 54, 55, 67].

Recent data indicate that the frequency of somatic mutations in APA might be much higher than previously described, due to tumour heterogeneity and the presence of multiple nodules, which do not all express aldosterone synthase. Indeed, aldosterone synthase immunohistochemistry-guided targeted sequencing on DNA extracted from formalin-fixed paraffin embedded adrenal samples revealed an overall frequency of somatic mutations of 88%, with a similar mutational spectrum then in previous studies [68, 69].

Genetic testing in patients with APA prior to surgery is challenged by the difficulty of detecting somatic mutations from peripheral blood. This could be circumvented by identifying surrogate biomarkers of the mutation status. Genotype-specific steroid profiles have been recently associated with APA, in particular a 7-steroid fingerprint able to correctly classify 92% of the APA according to genotype [70]. Future use of this type of screening could allow to select patients for adrenal vein sampling and to apply targeted treatments, once these approaches have been prospectively validated. Indeed, mutated KCNJ5 channels have peculiar pharmacological characteristics and are blocked by calcium-channel blockers, such as verapamil, and Na-channel blockers, such as amiloride at high therapeutic doses [71]. Macrolide antibiotics, including roxithromycin, potently inhibit mutant channels as well as *CYP11B2* expression and aldosterone production in adrenocortical cells [72] and in primary cell lines from APA [73]. Current studies are ongoing to test whether these compounds can be used to identify and treat patients with PA due to APA with *KCNJ5* mutations [74].

Hypertension Due to Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH) is a group of autosomal recessive diseases of adrenal steroidogenesis caused by mutations in genes encoding enzymes or cofactors involved in cortisol biosynthesis. For a detailed description of CAH variants see Chapter 5.9, here we briefly describe those variants that are associated with increased mineralocorticoid receptor activation. While the most frequent form of CAH, 21-hydroxylase deficiency (21OHD), is characterized by salt wasting and hypotension, 11β-hydroxylase deficiency (11OHD) and 17α-hydroxylase deficiency (17OHD), are associated with hypertension (see online Table 5.6.1.1).

11β-Hydroxylase Deficiency

11OHD is the second most frequent cause of CAH, accounting for 2–5% of cases (overall incidence 1 in 100 000–200 000 live births) [75, 76]. The enzyme 11β-hydroxylase (CYP11B1) catalyses the conversion of 11-deoxycortisol to cortisol and DOC to corticosterone in the ZF. Defective cortisol biosynthesis results in ACTH secretion leading to accumulation of 11-deoxycortisol and DOC, which binds the MR with high affinity. Although a slightly less potent mineralocorticoid than aldosterone [7], the excessive production of DOC in 11OHD causes sodium retention, volume expansion and hypertension in 30–60% of cases [75, 77, 78]. Renin production is suppressed secondary to mineralocorticoid-induced sodium retention and volume expansion and aldosterone levels are decreased due to low serum potassium and low plasma renin (see online Table 5.6.1.1). Because newborns are relatively resistant to mineralocorticoids, hypertension might not become apparent during the neonatal period [79]. Increased ACTH secretion leads also to adrenal androgen excess resulting in fetal female virilization and precocious pseudopuberty in boys. 11OHD is caused by inactivating mutations in the *CYP11B1* gene located on chromosome 8q21 and is transmitted as an autosomal recessive trait [77]. *CYP11B1* is located in tandem with the highly homologous *CYP11B2* gene. Opposite to FH-I, unequal crossing-over between those genes, where the *CYP11B1* coding sequence is under the control of the *CYP11B2* promoter responding to angiotensin II and not ACTH, also results in a phenotype of classic 11OHD [80]. 11OHD treatment is based on cortisol replacement to normalize ACTH, which in turn removes the drive for overproduction of DOC, leading to remission of hypertension.

17α-hydroxylase Deficiency

17OHD is a rare form of CAH accounting for less than 1% of cases [81]. The enzyme 17α-hydroxylase (CYP17A1) is expressed in adrenal glands and gonads and catalyses 17α-hydroxylation of pregnenolone and progesterone as well as the 17,20 lyase reaction catalysing conversion of 17-hydroxypregnenolone to dehydroepiandrosterone (DHEA) and, with less efficiency, that of 17-hydroxyprogesterone to androstenedione. 17OHD shifts steroidogenesis towards the mineralocorticoid pathway, with accumulation of corticosterone and DOC. Patients exhibit hypokalaemic hypertension, male undervirilization at birth in affected 46,XY individuals, primary amenorrhea in 46,XX patients, and hypergonadotropic hypogonadism in both sexes [82]. 17OHD is caused by inactivating mutations in *CYP17A1*, located on the long arm of chromosome 10 [83]. To date, more than 100 mutations have been reported, 75% of which are missense mutations [84]. 17OHD treatment is based on partial glucocorticoid replacement to reduce DOC production, antihypertensive management, mainly with MR antagonists, and sex steroid replacement according to age and sex of rearing.

Apparent Mineralocorticoid Excess

Apparent mineralocorticoid excess (AME) is an autosomal recessive disease first described in a 3-years old boy with hypertension,

hypokalaemia, low plasma renin activity, and aldosterone concentration, and responsiveness to spironolactone [85], with a decreased rate of conversion of active cortisol to cortisone suggesting a defect in the oxidation of cortisol [86]. AME is caused by inactivating mutations in the *HSD11B2* gene encoding 11β-hydroxysteroid dehydrogenase type 2 (HSD11B2) [87]. Given that plasma free cortisol levels are ~1000-fold those of aldosterone and cortisol binds with high affinity to MR, lack of HSD11B2 is responsible for cortisol binding to MR in the distal nephron of the kidney and a phenotype of pseudohyperaldosteronism [88, 89].

In early life, AME is characterized by severe hypertension, suppressed renin and plasma aldosterone, and an elevated ratio of urinary cortisol to cortisone metabolites. Affected children may also exhibit lower birth weight, polydipsia, polyuria, failure to thrive, poor growth, hypokalaemia, and persistent metabolic alkalosis [90, 91]. AME is associated with variable organ damage, potentially affecting kidneys, retina, heart, and central nervous system [92]. A milder phenotype has been described in adults with hypertension and milder biochemical abnormalities [93]. The phenotypic variability is associated with different extents of impairment of HSD11B2 activity with distinct mutations [94, 95]. Differential diagnosis includes excessive ingestion of liquorice, which contains glycyrrhetinic and glycyrrhizic acid, both being potent competitive inhibitors of HSD11B2 activity [96]. Inadvertent overexposure to potent synthetic mineralocorticoids also mimics AME [97]. AME treatment is based on MR antagonism with spironolactone or eplerenone. Other antihypertensive drugs such as amiloride, calcium channel blockers, and thiazides, which can help to normalize blood pressure and lower hypercalciuria, may be associated [92]. The importance of early diagnosis and treatment was reinforced by the description of significant improvement in end-organ damage in patients who had undergone two to 13 years of treatment [90].

Activating Mutations of the Mineralocorticoid Receptor

A Mendelian form of human hypertension was described by Geller *et al.* in 2000, caused by an activating mutation of the *NR3C2* gene, coding for the MR. The heterozygous MR p.Ser810Leu mutation was identified in a 15-year-old boy with severe hypertension associated with suppressed plasma renin and low aldosterone levels [98]. Screening of his family showed that the p.Ser810Leu mutation was transmitted as an autosomal dominant trait and cosegregated with early-onset hypertension (before age 20) in 11 subjects. Two female carriers of the mutation had their pregnancies complicated by marked exacerbation of hypertension, accompanied by low serum potassium levels and undetectable aldosterone levels, but without signs of preeclampsia.

The MR p.Ser810Leu mutation is located in the MR ligand binding pocket, inducing a major change in its conformation and ligand specificity. While there is no difference in mutant or wild-type MR activity in response to aldosterone, both progesterone and spironolactone, two MR antagonists, are able to activate transcription via the p.Ser810Leu MR mutant. These findings suggest that progesterone may contribute to the exacerbation of hypertension during pregnancy [98]. The endogenous steroids responsible for hypertension in men and non-pregnant women carrying the p.Ser810Leu mutation were identified to be cortisone and 11-dehydrocorticosterone, the main cortisol and corticosterone metabolites produced by the action of HSD11B2 in the distal nephron, which bind with high affinity to the mutant MR and are potent activators of mutant MR-dependent transcription [98]. Given the much higher concentration of plasma cortisol than corticosterone, the authors suggested that cortisone is the endogenous steroid binding and activating the mutant MR [98]. While one study reported a high frequency of MR p.Ser810Leu in hypertensive pregnant women from Mexico [100], no further family carrying this mutation was described. Furthermore, the mutation is not listed in the large database of genetic variants ExAC (The Exome Aggregation Consortium, http://exac.broadinstitute.org) and its frequency is not described on the 1000 genomes project (http://www.internationalgenome.org). The use of MR antagonists such as spironolactone or eplerenone should be avoided in patients carrying the p.Ser810Leu mutation as they act as agonists on the mutant MR. In contrast, the third-generation MR antagonists BR-4628 and finerenone, both derived from dihydropyridines, are also antagonists on the mutant receptor [101, 102]. Finrenone, currently approved for clinical use, might be an option to treat carriers of the MR p.Ser810Leu mutation.

REFERENCES

1. Hannemann A, Wallaschofski H. Prevalence of primary aldosteronism in patient's cohorts and in population-based studies—a review of the current literature. *Horm Metab Res*, 2012; **44**: 157–62.
2. Monticone S, Burrello J, Tizzani D, *et al.* Prevalence and clinical manifestations of primary aldosteronism encountered in primary care practice. *J Am Coll Cardiol*, 2017; **69**: 1811–20.
3. Loffing J, Loffing-Cueni D, Valderrabano V, *et al.* Distribution of transcellular calcium and sodium transport pathways along mouse distal nephron. *Am J Physiol Renal Physiol*, 2001; **281**: F1021–7.
4. Nuclear Receptors Nomenclature, C. A unified nomenclature system for the nuclear receptor superfamily. *Cell*, 1999; **97**: 161–3.
5. Stockand JD. New ideas about aldosterone signaling in epithelia. *Am J Physiol Renal Physiol*, 2002; **282**: F559–76.
6. Arriza JL, Weinberger C, Cerelli G, *et al.* Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. *Science*, 1987; **237**: 268–75.
7. Hellal-Levy C, Couette B, Fagart J, *et al.* Specific hydroxylations determine selective corticosteroid recognition by human glucocorticoid and mineralocorticoid receptors. *FEBS Lett*, 1999; **464**: 9–13.
8. Eisenhofer G, Peitzsch M, Kaden D, *et al.* Reference intervals for plasma concentrations of adrenal steroids measured by LC-MS/MS: impact of gender, age, oral contraceptives, body mass index and blood pressure status. *Clin Chim Acta*, 2017; **470**: 115–24.
9. Edwards CRW, Stewart PM, Burt D, *et al.* Localization of 11b-hydroxysteroid dehydrogenase—tissue specific protector of the mineralocorticoid receptor. *Lancet*, 1988; **2**: 986–9.
10. Funder JW, Pearce PT, Smith R, Smith AI. Mineralocorticoid action: target tissue specificity is enzyme, not receptor, mediated. *Science*, 1988; **242**: 583–5.
11. Connell JM, Davies E. The new biology of aldosterone. *J Endocrinol*, 2005; **186**: 1–20.

12. Spat A, Hunyady L. Control of aldosterone secretion: a model for convergence in cellular signaling pathways. *Physiol Rev*, 2004; **84**: 489–539.
13. Bandulik S, Penton D, Barhanin J, Warth R. TASK1 and TASK3 potassium channels: determinants of aldosterone secretion and adrenocortical zonation. *Horm Metab Res*, 2010. **42**: 450–7.
14. Hattangady NG, Olala LO, Bollag WB, Rainey WE. Acute and chronic regulation of aldosterone production. *Mol Cell Endocrinol*, 2012; **350**: 151–62.
15. Sutherland DJ, Ruse JL, Laidlaw JC. Hypertension, increased aldosterone secretion and low plasma renin activity relieved by dexamethasone. *Can Med Assoc J*, 1966; **95**: 1109–19.
16. Halperin F, Dluhy RG. Glucocorticoid-remediable aldosteronism. *Endocrinol Metab Clin North Am*, 2011; **40**: 333–41, viii.
17. Pascoe L, Jeunemaitre X, Lebrethon MC, *et al*. Glucocorticoid-suppressible hyperaldosteronism and adrenal tumors occurring in a single French pedigree. *J Clin Invest*, 1995; **96**: 2236–46.
18. Lifton R, Dluhy RG, Powers M, *et al*. A chimaeric 11beta-hydroxylase aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*, 1992; **355**: 262–5.
19. Pascoe L, Curnow KM, Slutsker L, *et al*. Glucocorticoid-suppressible hyperaldosteronism results from hybrid genes created by unequal crossovers between CYP11B1 and CYP11B2. *Proc Natl Acad Sci U S A*, 1992; **89**: 8327–31.
20. Stowasser M, Gordon RD. Primary aldosteronism: learning from the study of familial varieties. *J Hypertens*, 2000; **18**: 1165–76.
21. Stowasser M, Bachmann AW, Huggard PR, Rossetti TR, Gordon RD. Treatment of familial hyperaldosteronism type I: only partial suppression of adrenocorticotropin required to correct hypertension. *J Clin Endocrinol Metab*, 2000; **85**: 3313–18.
22. Fardella CE, Mosso L, Gómez-Sánchez C, *et al*. Primary hyperaldosteronism in essential hypertensives: prevalence, biochemical profile, and molecular biology. *J Clin Endocrinol Metab*, 2000; **85**: 1863–7.
23. Mulatero P, Tizzani D, Viola A, *et al*. Prevalence and characteristics of familial hyperaldosteronism: the PATOGEN study (primary aldosteronism in TOrino-GENetic forms). *Hypertension*, 2011; **58**: 797–803.
24. Aglony M, Martínez-Aguayo A, Carvajal CA, *et al*. Frequency of familial hyperaldosteronism type 1 in a hypertensive pediatric population: clinical and biochemical presentation. *Hypertension*, 2011; **57**: 1117–21.
25. Litchfield WR, Anderson BF, Weiss RJ, Lifton RP, Dluhy RG. Intracranial aneurysm and hemorrhagic stroke in glucocorticoid-remediable aldosteronism. *Hypertension*, 1998; **31**(1 Pt 2): 445–50.
26. Stowasser M, Gunasekera TG, Gordon RD, *et al*. Familial hyperaldosteronism type II: five families with a new variety of primary aldosteronism. *Clin Exp Pharmacol Physiol*, 1992; **19**: 319–22.
27. Stowasser M, Gordon RD. Familial hyperaldosteronism. *J Steroid Biochem Mol Biol*, 2001; **78**: 215–29.
28. Medeau V, Assié G, Zennaro MC, Clauser E, Plouin PF, Jeunemaitre X. [Familial aspect of primary hyperaldosteronism: analysis of families compatible with primary hyperaldosteronism type 2]. *Ann Endocrinol (Paris)*, 2005; **66**: 240–6.
29. Pallauf A, Schirpenbach C, Zwermann O, *et al*. The prevalence of familial hyperaldosteronism in apparently sporadic primary aldosteronism in Germany: a single center experience. *Horm Metab Res*, 2012; **44**: 215–20.
30. Choi M, Scholl UI, Yue P, *et al*. K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science*, 2011; **331**: 768–72.
31. Mulatero P, Monticone S, Rainey WE, Veglio F, Williams TA. Role of KCNJ5 in familial and sporadic primary aldosteronism. *Nat Rev Endocrinol*, 2013; **9**: 104–12.
32. Scholl UI, Nelson-Williams C, Yue P, *et al*. Hypertension with or without adrenal hyperplasia due to different inherited mutations in the potassium channel KCNJ5. *Proc Natl Acad Sci U S A*, 2012; **109**: 2533–8.
33. Mulatero P, Tauber P, Zennaro MC, *et al*. KCNJ5 mutations in European families with nonglucocorticoid remediable familial hyperaldosteronism. *Hypertension*, 2012; **59**: 235–40.
34. Monticone S, Hattangady NG, Penton D, *et al*. A novel Y152C KCNJ5 mutation responsible for familial hyperaldosteronism type III. *J Clin Endocrinol Metab*, 2013; **98**: E1861–5.
35. Lafferty AR, Torpy DJ, Stowasser M, *et al*. A novel genetic locus for low renin hypertension: familial hyperaldosteronism type II maps to chromosome 7 (7p22). *J Med Genet*, 2000; **37**: 831–5.
36. So A, Duffy DL, Gordon RD, *et al*. Familial hyperaldosteronism type II is linked to the chromosome 7p22 region but also shows predicted heterogeneity. *J Hypertens*, 2005; **23**: 1477–84.
37. Fernandes-Rosa FL, Daniil G, Orozco IJ, *et al*. A gain-of-function mutation in the CLCN2 chloride channel gene causes primary aldosteronism. *Nat Genet*, 2018; **50**: 355–61.
38. Scholl UI, Stölting G, Schewe J, *et al*. CLCN2 chloride channel mutations in familial hyperaldosteronism type II. *Nat Genet*, 2018; **50**: 349–54.
39. Geller DS, Zhang J, Wisgerhof MV, *et al*. A novel form of human mendelian hypertension featuring nonglucocorticoid-remediable aldosteronism. *J Clin Endocrinol Metab*, 2008; **93**: 3117–23.
40. Tong A, Liu G, Wang F, *et al*. A novel phenotype of familial hyperaldosteronism type III: concurrence of aldosteronism and Cushing's syndrome. *J Clin Endocrinol Metab*, 2016; **101**: 4290–7.
41. Oki K, Plonczynski MW, Lam ML, Gómez-Sánchez EP, Gómez-Sánchez CE. The potassium channel, Kir3.4 participates in angiotensin II-stimulated aldosterone production by a human adrenocortical cell line. *Endocrinology*, 2012; **153**: 4328–35.
42. Gómez-Sánchez CE, Qi X, Gómez-Sánchez EP, *et al*. Disordered zonal and cellular CYP11B2 enzyme expression in familial hyperaldosteronism type 3. *Mol Cell Endocrinol*, 2017; **439**: 74–80.
43. Scholl UI, Stölting G, Nelson-Williams C, *et al*. Recurrent gain of function mutation in calcium channel CACNA1H causes early-onset hypertension with primary aldosteronism. *Elife*, 2015; **4**: e06315.
44. Reimer EN, Walenda G, Seidel E, Scholl UI. CACNA1HM1549V mutant calcium channel causes autonomous aldosterone production in HAC15 cells and is inhibited by Mibefradil. *Endocrinology*, 2016: **157**: 3016–22.
45. Daniil G, Fernandes-Rosa FL, Chemin J, *et al*. CACNA1H mutations are associated with different forms of primary aldosteronism. *EBioMedicine*, 2016; **13**: 225–36.
46. Scholl UI, Goh G, Stölting G, *et al*. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nat Genet*, 2013; **45**: 1050–4.
47. Korah HE, Scholl UI. An update on familial hyperaldosteronism. *Horm Metab Res*, 2015; **47**: 941–6.
48. Rossi GP, Bernini G, Caliumi C, *et al*. A prospective study of the prevalence of primary aldosteronism in 1,125 hypertensive patients. *J Am Coll Cardiol*, 2006; **48**: 2293–300.

49. Funder JW, Carey RM, Mantero F, *et al.* The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 1889–916.
50. Azizan EA, Poulsen H, Tuluc P, *et al.* Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. *Nat Genet*, 2013; **45**: 1055–60.
51. Beuschlein F, Boulkroun S, Osswald A, *et al.* Somatic mutations in ATP1A1 and ATP2B3 lead to aldosterone-producing adenomas and secondary hypertension. *Nat Genet*, 2013; **45**: 440–4, 444e1–2.
52. Stindl J, Tauber P, Sterner C, Tegtmeier I, Warth R, Bandulik S. Pathogenesis of adrenal aldosterone-producing adenomas carrying mutations of the Na(+)/K(+)-ATPase. *Endocrinology*, 2015; **156**: 4582–91.
53. Tauber P, Aichinger B, Christ C, *et al.* Cellular pathophysiology of an adrenal adenoma-associated mutant of the plasma membrane Ca(2+)-ATPase ATP2B3. *Endocrinology*, 2016; **157**: 2489–99.
54. Scholl UI, Healy JM, Thiel A, *et al.* Novel somatic mutations in primary hyperaldosteronism are related to the clinical, radiological and pathological phenotype. *Clin Endocrinol (Oxf)*, 2015.
55. Akerstrom T, Maharjan R, Willenberg HS, *et al.* Activating mutations in CTNNB1 in aldosterone producing adenomas. *Sci Rep*, 2016; **6**: 19546.
56. Rhayem Y, Perez-Rivas LG, Dietz A, *et al.* PRKACA Somatic mutations are rare findings in aldosterone-producing adenomas. *J Clin Endocrinol Metab*, 2016; **101**: 3010–17.
57. Zennaro MC, Boulkroun S, Fernandes-Rosa F. Genetic causes of functional adrenocortical adenomas. *Endocr Rev*, 2017; **38**: 516–37.
58. Fernandes-Rosa FL, Williams TA, Riester A, *et al.* Genetic spectrum and clinical correlates of somatic mutations in aldosterone-producing adenoma. *Hypertension*, 2014; **64**: 354–61.
59. Lenzini L, Rossitto G, Maiolino G, Letizia C, Funder JW, Rossi GP. A meta-analysis of somatic KCNJ5 K(+) channel mutations in 1636 patients with an aldosterone-producing adenoma. *J Clin Endocrinol Metab*, 2015; **100**: E1089–95.
60. Azizan EA, Lam BY, Newhouse SJ, *et al.* Microarray, qPCR, and KCNJ5 sequencing of aldosterone-producing adenomas reveal differences in genotype and phenotype between zona glomerulosa- and zona fasciculata-like tumors. *J Clin Endocrinol Metab*, 2012; **97**: E819–29.
61. Boulkroun S, Beuschlein F, Rossi GP, *et al.* Prevalence, clinical, and molecular correlates of KCNJ5 mutations in primary aldosteronism. *Hypertension*, 2012; **59**: 592–8.
62. Seccia TM, Mantero F, Letizia C, *et al.* Somatic mutations in the KCNJ5 gene raise the lateralization index: implications for the diagnosis of primary aldosteronism by adrenal vein sampling. *J Clin Endocrinol Metab*, 2012; **97**: E2307–13.
63. Zheng FF, Zhu LM, Nie AF, *et al.* Clinical characteristics of somatic mutations in Chinese patients with aldosterone-producing adenoma. *Hypertension*, 2015; **65**: 622–8.
64. Osswald A, Fischer E, Degenhart C, *et al.* Lack of influence of somatic mutations on steroid gradients during adrenal vein sampling in aldosterone-producing adenoma patients. *Eur J Endocrinol*, 2013; **169**: 657–3.
65. Akerstrom T, Willenberg HS, Cupisti K, *et al.* Novel somatic mutations and distinct molecular signature in aldosterone-producing adenomas. *Endocr Relat Cancer*, 2015; **22**: 735–44.
66. Kitamoto T, Suematsu S, Yamazaki Y, *et al.* Clinical and steroidogenic characteristics of aldosterone-producing adenomas with ATPase or CACNA1D gene mutations. *J Clin Endocrinol Metab*, 2016; **101**: 494–503.
67. Teo AE, Garg S, Shaikh LH, *et al.* Pregnancy, primary aldosteronism, and adrenal CTNNB1 mutations. *N Engl J Med*, 2015; **373**: 1429–36.
68. Nanba K, Omata K, Else T, *et al.* Targeted molecular characterization of aldosterone-producing adenomas in white Americans. *J Clin Endocrinol Metab*, 2018; **103**: 3869–76.
69. De Sousa K, Boulkroun S, Baron S, *et al.* Genetic, cellular, and molecular heterogeneity in adrenals with aldosterone-producing adenoma. *Hypertension*, 2020 Apr; **75(4)**: 1034–44.
70. Williams TA, Peitzsch M, Dietz AS, *et al.* Genotype-specific steroid profiles associated with aldosterone-producing adenomas. *Hypertension*, 2016; **67**: 139–45.
71. Tauber P, Penton D, Stindl J, *et al.* Pharmacology and pathophysiology of mutated KCNJ5 found in adrenal aldosterone-producing adenomas. *Endocrinology*, 2014; **155**: 1353–62.
72. Scholl UI, Abriola L, Zhang C, *et al.* Macrolides selectively inhibit mutant KCNJ5 potassium channels that cause aldosterone-producing adenoma. *J Clin Invest*, 2017.
73. Caroccia B, Prisco S, Seccia TM, Piazza M, Maiolino G, Rossi GP. Macrolides blunt aldosterone biosynthesis: a proof-of-concept study in KCNJ5 mutated adenoma cells ex vivo. *Hypertension*, 2017; **70**: 1238–42.
74. Maiolino G, Ceolotto G, Battistel M, *et al.* Macrolides for KCNJ5-mutated aldosterone-producing adenoma (MAPA): design of a study for personalized diagnosis of primary aldosteronism. *Blood Press*, 2018; **27**: 200–5.
75. Nimkarn S, New MI. Steroid 11beta-hydroxylase deficiency congenital adrenal hyperplasia. *Trends Endocrinol Metab*, 2008; **19**: 96–9.
76. Rosler A, Leiberman E, Cohen T. High frequency of congenital adrenal hyperplasia (classic 11 beta-hydroxylase deficiency) among Jews from Morocco. *Am J Med Genet*, 1992; **42**: 827–34.
77. White PC, Curnow KM, Pascoe L. Disorders of steroid 11 beta-hydroxylase isozymes. *Endocr Rev*, 1994; **15**: 421–38.
78. Zachmann M, Tassinari D, Prader A. Clinical and biochemical variability of congenital adrenal hyperplasia due to 11 beta-hydroxylase deficiency. A study of 25 patients. *J Clin Endocrinol Metab*, 1983; **56**: 222–9.
79. Holcombe JH, Keenan BS, Nichols BL, Kirkland RT, Clayton GW. Neonatal salt loss in the hypertensive form of congenital adrenal hyperplasia. *Pediatrics*, 1980; **65**: 777–81.
80. Hampf M, Keenan BS, Nichols BL, Kirkland RT, Clayton GW. Unequal crossing-over between aldosterone synthase and 11beta-hydroxylase genes causes congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2001;**s 86**: 4445–52.
81. Costa-Santos M, Kater CE, Auchus RJ. Two prevalent CYP17 mutations and genotype-phenotype correlations in 24 Brazilian patients with 17-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2004; **89**: 49–60.
82. Auchus RJ. The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol Metab Clin North Am*, 2001; **30**: 101–19, vii.
83. Matteson KJ, Picado-Leonard J, Chung BC, Mohandas TK, Miller WL. Assignment of the gene for adrenal P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase) to human chromosome 10. *J Clin Endocrinol Metab*, 1986; **63**: 789–91.

84. Baranowski ES, Arlt W, Idkowiak J. Monogenic disorders of adrenal steroidogenesis. *Horm Res Paediatr*, 2018; **89**: 292–310.
85. New MI, Levine LS, Biglieri EG, Pareira J, Ulick S. Evidence for an unidentified steroid in a child with apparent mineralocorticoid hypertension. *J Clin Endocrinol Metab*, 1977; **44**: 924–33.
86. Ulick S, Levine LS, Gunczler P, *et al.* A syndrome of apparent mineralocorticoid excess associated with defects in the peripheral metabolism of cortisol. *J Clin Endocrinol Metab*, 1979; **49**: 757–64.
87. Wilson RC, Krozowski ZS, Li K, *et al.* A mutation in the HSD11B2 gene in a family with apparent mineralocorticoid excess. *J Clin Endocrinol Metab*, 1995; **80**: 2263–6.
88. Farman N, Rafestin-Oblin ME. Multiple aspects of mineralocorticoid selectivity. *Am J Physiol Renal Physiol*, 2001; **280**: F181–92.
89. Rusvai E, Naray-Fejes-Toth A. A new isoform of 11 beta-hydroxysteroid dehydrogenase in aldosterone target cells. *J Biol Chem*, 1993; **268**: 10717–20.
90. New MI, Geller DS, Fallo F, Wilson RC. Monogenic low renin hypertension. *Trends Endocrinol Metab*, 2005; **16**: 92–7.
91. Funder JW. Apparent mineralocorticoid excess. *J Steroid Biochem Mol Biol*, 2017; **165**(Pt A): 151–3.
92. Dave-Sharma S, Wilson RC, Harbison MD, *et al.* Examination of genotype and phenotype relationships in 14 patients with apparent mineralocorticoid excess. *J Clin Endocrinol Metab*, 1998; **83**: 2244–54.
93. Wilson RC, Wei J-Q, Dave Sharma S, *et al.* A genetic defect resulting in mild low-renin hypertension. *Proc Natl Acad Sci U S A*, 1998; **95**: 10200–5.
94. Nunez BS, Rogerson FM, Mune T, *et al.* Mutants of 11beta-hydroxysteroid dehydrogenase (11-HSD2) with partial activity: improved correlations between genotype and biochemical phenotype in apparent mineralocorticoid excess. *Hypertension*, 1999; **34**(4 Pt 1): 638–42.
95. Yau M, Haider S, Khattab A, *et al.* Clinical, genetic, and structural basis of apparent mineralocorticoid excess due to 11beta-hydroxysteroid dehydrogenase type 2 deficiency. *Proc Natl Acad Sci U S A*, 2017; **114**: E11248–56.
96. Stewart PM, Wallace AM, Valentino R, Burt D, Shackleton CH, Edwards CR. Mineralocorticoid activity of liquorice: 11-beta-hydroxysteroid dehydrogenase deficiency comes of age. *Lancet*, 1987; **2**: 821–4.
97. Funder JW, Adam WR, Mantero F, Kraft N, Ulick S. The etiology of a syndrome of factitious mineralocorticoid excess: a steroid-containing nasal spray. *J Clin Endocrinol Metab*, 1979; **49**: 842–6.
98. Geller DS, Farhi A, Pinkerton N, *et al.* Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science*, 2000; **289**: 119–23.
99. Rafestin-Oblin ME, Souque A, Bocchi B, Pinon G, Fagart J, Vandewalle A. The severe form of hypertension caused by the activating S810L mutation in the mineralocorticoid receptor is cortisone related. *Endocrinology*, 2003; **144**: 528–33.
100. Ramirez-Salazar M, Romero-Gutiérrez G, Zaina S, Malacara JM, Kornhauser C, Pérez-Luque E. Relationship of aldosterone synthase gene (C-344T) and mineralocorticoid receptor (S810L) polymorphisms with gestational hypertension. *J Hum Hypertens*, 2011; **25**: 320–6.
101. Fagart J, Hillisch A, Huyet J, *et al.* A new mode of mineralocorticoid receptor antagonism by a potent and selective nonsteroidal molecule. *J Biol Chem*, 2010; **285**: 29932–40.
102. Amazit L, Le Billan F, Kolkhof P, *et al.* Finerenone impedes aldosterone-dependent nuclear import of the mineralocorticoid receptor and prevents genomic recruitment of steroid receptor coactivator-1. *J Biol Chem*, 2015; **290**: 21876–89.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

5.6.2 Management of Primary Aldosteronism

William M. Drake and Morris J. Brown

Introduction

The successful management of PA involves important contributions from primary, secondary, and tertiary physicians, clinical biochemists, cross-sectional and interventional radiologists, endocrine surgeons, specialist nurses experienced in conducting complex dynamic

testing and, increasingly, endocrine histopathologists who provide retrospective, molecular information as to the accuracy of the preceding clinical evaluation. Such collaborative, multidisciplinary working, involving mutual trust and shared learning over time, lends itself well to high-quality clinical care but complicates comparison of results between centres. Variations in assays for aldosterone and renin, subjective interpretation of adrenal imaging, and differences in the protocols for adrenal vein sampling (AVS), contrast with, for instance, evaluations of osteoporosis therapy. In PA, there is no equivalent to the strict, standardized bone mineral density scores used in osteoporosis as both entry criteria and outcome measures. The authors take the view that there is no 'perfect' way to manage PA, but we attempt to provide a balanced view of a fast-moving field. Where personal opinion is presented, it is done with appropriate caveats.

Evaluation of Published Evidence in PA

A decade after his original description [1], Conn published 145 cases of PA and suggested that the syndrome may account for up to 20% of patients with hypertension. He later revised this estimate down to 10% [2], which approximates to recent evidence suggesting a prevalence of 5–8% in unselected hypertensive populations. Allowing for inevitable variability between studies, the accepted prevalence of PA now stands on a par with type 1 diabetes and hepatitis B. In both of these examples, robust, prospective, outcome studies are plentiful and, indeed, mandated before any new intervention can receive approval for funding by most healthcare providers. Strikingly, the published long-term outcome data on PA are sparse and their quality substantially inferior to the two comparator conditions quoted. The vast majority of outcome data available for analysis are retrospective. To date, there exists only one published prospective intervention study in PA, discussed below [3], and it was only recently that international consensus data were produced defining what constitutes surgical 'cure' [4]. In contrast, the recent explosion of data on the molecular pathogenesis of PA [5–8] is incontrovertibly of high quality and provides an example of how such insights can re-energize an area of clinical research that had advanced only modestly in the preceding decades; and of how such findings can lead to the re-challenging of old 'dogmas' and 'rules'.

Suspicion of PA: Case Finding

Symptoms and physical signs (other than hypertension) are few in PA. Contrary to historical teaching, hypokalaemia occurs in only a minority of patients but, if present, may cause muscle weakness and cramps, palpitation, and, consequent upon an induced renal concentrating defect, symptoms of frequency and nocturia. Given that PA is most commonly diagnosed in middle age, these symptoms are often dismissed as consequent upon prostatic enlargement or childbirth. Transient hypokalaemia precipitated by the use of thiazide or other potassium-wasting diuretics is a common scenario, but this is often attributed to the drug alone, rather than raising suspicion that the diuretic is unmasking PA by increasing sodium delivery to the site of aldosterone action downstream in the nephron. When groups of patients with PA are compared with those with essential hypertension or other forms of endocrine-mediated hypertension, greater degrees of left ventricular hypertrophy (LVH), higher rates of atrial fibrillation and greater urinary albumin excretion are demonstrable, although none is diagnostically sensitive or specific. This bland clinical picture does not lend itself easily to diagnosis in the primary care setting, in which most cases of hypertension are seen, and mandates a more 'algorithmic' approach to case finding. Current consensus guidelines [9] recommend screening for PA in certain scenarios, shown in **Box 5.6.2.1**.

Box 5.6.2.1 Indications for testing for PA

Consensus guidance on when to test for primary aldosteronism (PA)

- Hypertension and spontaneous or diuretic-provoked hypokalaemia
- Hypertension resistant to three conventional antihypertensive drugs (including a diuretic)
- Hypertension controlled by the use of four conventional antihypertensive drugs (including a diuretic)
- Hypertension and an adrenal incidentaloma
- Hypertension and sleep apnoea
- Hypertension and a family history of early onset hypertension or stroke <40 years
- All hypertensive first-degree relatives of patients with diagnosed PA

Adapted with permission from Funder JW, Carey RM, Mantero F, *et al*. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2016; 101(5): 1889–916; to which the reader is referred for a more detailed discussion.

These guidelines for screening are not underpinned by a published body of evidence indicating that they favourably impact on morbidity, mortality, or quality of life. To date, no prospective study exists to demonstrate that widespread testing for PA is cost effective. The 'targeted strategy' recommended by those experienced in the field, towards groups of patients 'enriched' with PA, represents a compromise between the cost and logistical challenge of screening huge numbers of patients and the consequences of a missed/delayed diagnosis on the development of hypertension that is more severe and/or resistant to medical or surgical treatment.

There is little dispute about the biochemical investigations with which to undertake case finding for PA. Like other areas of endocrinology, paired measurement of both limbs of the feedback system (in this case aldosterone and renin, with which to calculate an aldosterone:renin ratio, ARR) is needed; isolated measurements of potassium, aldosterone, and renin are demonstrably less accurate. Individual centres use different assays for both plasma aldosterone concentration (PAC) and renin (either plasma renin activity, PRA, or direct renin concentration, DRC) with slightly different reference ranges. Hence (like other endocrine conditions), implicit in the cut-off values for ARR given in **Table 5.6.2.1** [9] is advice to clinicians to be acquainted with their own institution's assay performance. The clinical use of the ARR grew out of recognition that many patients with PA do not have plasma aldosterone levels above earlier thresholds for diagnosis, and that a reduction in these would lead to false positives unless an elevated ARR was also required; and it has an intrinsic flaw due to the logarithmic distribution of plasma renin, but arithmetic distribution of aldosterone. The limitations of the use of a ratio as a screening tool are self-evident; when denominator (renin) levels are low the ARR may be high even if the PAC is also low and conceptually incompatible

Table 5.6.2.1 ARR cut-off values, depending on assay and based on whether PAC, PRA, and DRC are measured in conventional or SI units

	PRA (ng/ml/h)	PRA (pmol/L/min)	DRC (mU/L)	DRC (ng/L)
PAC (ng/dL)	20	1.6	2.4	3.8
	30	2.5	3.7	5.7
	40	3.1	4.9	7.7
PAC (pmol/L)	750	60	91	144
	1000	80	122	192

Adapted with permission from Funder JW, Carey RM, Mantero F, *et al*. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2016; 101(5): 1889–916; to which the reader is referred for a more detailed discussion.

with a diagnosis of PA. To account for this issue, the concept of a 'minimum' PAC is advocated by some, although the precise value below which PA is excluded is not well defined.

The ARR test performs best when patients have been ambulant for at least 2 hours in the morning, with liberal salt intake in the preceding days and with hypokalaemia corrected if originally present. Blood is usually drawn with the patient seated for 10–15 minutes. The test should not be performed if a patient is taking spironolactone, epleronone, or high dose amiloride; discontinuation for a minimum of 4 weeks is mandatory for these drugs. However, in patients with difficult hypertension, safety is the first concern. Spironolactone has long-acting metabolites that have some activity even 6 weeks after discontinuation, whereas eplerenone and amiloride have shorter half-lives and no active metabolites. As a screening test, it is acceptable to measure the ARR on diuretics, angiotensin converting enzyme inhibitors (ACEIs) and angiotensin receptor blockers (ARBs), although the potential effect of these drugs to de-repress renin levels dictates that many patients will subsequently require further, confirmatory testing if PA is strongly suspected. Conversely, this generates a useful clinical practice point; the finding of a suppressed value during therapy with one of this class of drug is very suspicious of PA. β-adrenergic antagonists reduce renin levels, but aldosterone values also fall so the overall performance of the ARR as a screening tool is acceptably preserved.

It should be emphasized that the ARR is a 'case finding'/screening tool that is particularly effective at excluding PA in drug-naïve patients. It is accepted, in a substantial proportion of patients, that, because of indeterminate results and/or the effect of interfering medications, one of a choice of confirmatory tests will subsequently be required.

Confirmation of PA: Diagnostic Testing

The cut-off values for the ARR place a high priority on not missing a diagnosis of PA, so many patients will require a subsequent confirmatory test if the screening investigation does not confidently exclude autonomous aldosterone secretion. Conversely, for some patients the results will be sufficiently convincing not to require further testing. Published guidance on this is, perhaps, surprisingly strict (documented hypokalaemia, PAC>550 pmol/L and plasma renin below detection levels) [9] and many experienced clinicians adopt a more pragmatic approach.

Four confirmatory tests are in common use; each, in a slightly different way, seeks dynamically to interrogate the suspected abnormal prevailing physiology. None has been convincingly demonstrated to be superior to any other and individual choice is usually on the basis of pragmatic considerations such as local expertise and cost.

The oral sodium load and saline infusion tests are similar in concept; the provision of a surplus of sodium should, with a normally functioning renin–angiotensin–aldosterone feedback system, result in marked suppression of aldosterone production. In the former case, urinary aldosterone excretion is measured on day 3 of a diet that includes at least 200 mmol/day of sodium and during which correction of induced hypokalaemia is crucial. In the latter test, 2 L of 0.9% saline are infused over 4 hours into a fasted patient, with a single measurement of PAC at the conclusion of the infusion. Both tests have high reported sensitivities and specificities, and both carry the potential hazard, in patients with marked hypertension or cardiac failure, of precipitating clinical deterioration. The use of antihypertensive medications that do not interfere with the renin–angiotensin axis (e.g. doxazosin, verapamil, hydralazine) is therefore recommended to achieve blood pressure control prior to either of these confirmatory tests.

In the fludrocortisone suppression test, 0.1 mg fludrocortisone acetate is administered 6 hourly for 4 days, together with high doses of sodium and potassium supplements. In hypertensive patients without PA, PAC falls to below a generally agreed cut-off of 170 pmol/L; above this level, PA is highly likely. For precision and safety, many centres advocate conducting this test as an inpatient, which increases its cost considerably.

Where there is concern about tests that may exacerbate hypertension and/or precipitate cardiac failure, the captopril challenge test (CCT) is a potentially attractive alternative. Samples for PAC and renin are drawn at time zero and 1–2 hours after an oral dose of 25–50 mg captopril. PAC should fall by >30% during the test; failure to achieve this, in combination with a persistently suppressed renin value, is highly suggestive of PA. Head-to-head comparison of CCT with saline suppression testing suggests that failure to reduce plasma aldosterone to <230 pmol/L is a more accurate threshold. There are suspicions that the CCT may yield false positive results and, in clinical practice, the principle of the investigation is more commonly exploited by performing a screening ARR in a patient already treated with a longer-acting ACEI.

The issue arises as to what constitutes the 'gold standard' against which the performances of the ARR and confirmatory tests at detecting/diagnosing PA are compared. Many studies have compared the 'screening' ARR with the results obtained during

subsequent confirmatory testing, but, in turn, the thresholds applied during those confirmatory tests are often based on the lowest values of PAC recorded in patients subsequently shown to have had normal renin–angiotensin–aldosterone physiology restored after a unilateral adrenalectomy. Two problems arise from this (enforced) approach. First, if a patient with suspected PA has equivocal/indeterminate results from one of the confirmatory tests but does not undergo surgery, he/she can never contribute to the data set that gives rise to a diagnostic threshold. Second, in analysing data for a diagnostic threshold, the importance placed on data derived from patients who eventually undergo surgery (with histology and postoperative biochemical re-evaluation reasonably regarded as 'gold standard') assumes that the physiology of 'surgical' and 'non-surgical' PA is identical. This assumption may or may not be correct.

Subtyping of PA

Once the diagnosis of PA is securely established, the next stage of its management is accurate subtyping. Although rare causes of PA exist (adrenocortical carcinoma and various familial forms), the vast majority is accounted for by a benign, unilateral aldosterone-producing adenoma (APA) or bilateral adrenal hyperplasia (BAH). For this reason, the following discussion deals with the investigations and their interpretation that assist with the distinction between these two entities. This distinction is crucial because most patients with APA are managed by unilateral adrenalectomy, whereas patients with BAH should be treated medically. Unilateral adrenal hyperplasia (UAH) also exists and is grouped in with this discussion because, although much less common, the results from AVS and response to unilateral adrenalectomy are similar to APA.

Cross-Sectional Imaging

The role of adrenal imaging in the management of PA is to exclude a large, potentially malignant mass and provide anatomical information to the surgeon in the event of an adrenalectomy. By facilitating reconstruction of adrenal vein anatomy, it may also improve the success rates of AVS, although individual radiology opinion varies on this. No formal, prospective studies exist comparing computed tomography (CT) with magnetic resonance imaging (MRI), but most clinicians favour the former on account of its superior spatial resolution. MRI may be preferable in women of childbearing age; this may become more important once functional imaging (see below) becomes widely used in place of AVS. Thin slice, high-quality contrast-enhanced CT (and, crucially, interpretation) provides important information about the presence of nodules (macro ≥ 10 mm; micro <10 mm), adrenal limb thickening, or normal anatomy; all of which contribute to decision-making. A detailed description of radiological features and interpretation is beyond the scope of this chapter, but most patients will fall into one of the categories shown in Table 5.6.2.2, together with some important practice points. From the table, it can be seen that (other than concern about a malignant mass) the only scenario in which AVS can be avoided for accurate subtyping is the presence of a clear-cut, typical macronodule with unequivocally normal anatomy in a young person. Anecdotally, it is the authors' experience that it is the judgement of 'normality' in adrenal imaging that is often the most challenging issue. The detection of macronodules should be straightforward, but the molecular discoveries discussed in Chapter 5.6.1 have led to increasing recognition of 'scanning occult' microadenomas as important causes of PA (especially among African Caribbean patients), and of an inverse correlation between size of nodule and density of CYP11B2 (aldosterone synthase).

Table 5.6.2.2 Adrenal appearances on CT imaging and decision-making in PA

Imaging appearances	Interpretation/clinical management
Normal appearances	Does not imply BAH. Many APAs are small. AVS needed
Unilateral, typical macroadenoma and unequivocally normal contralateral gland	In young patients (<35 years), this is generally sufficient to avoid AVS and proceeding direct to unilateral adrenalectomy is reasonable. In patients ≥40 years, the presence of a macronodule cannot assumed to be an APA. AVS needed*
Bilateral micro and/or macronodules	Does not imply BAH. Some patients will have unilateral, surgically treatable, disease. AVS needed
Unilateral adrenal thickening	Implies neither UAH nor BAH. AVS needed.
Unilateral micronodule	May represent area of hyperplasia or be functionless. AVS needed
Unilateral nodule with ipsi- or contralateral thickening	AVS needed
Unilateral large (usually >4 cm), often heterogenous mass	Concern about malignancy highly likely to mandate surgery without recourse to AVS. Tests for other hormonal secretion (cortisol, androgens) also needed

* Opinion varies as to whether 35 or 40 represents the most appropriate 'threshold' for the avoidance of AVS with this anatomical scenario

It is important to appreciate the hazards of relying on imaging alone to guide clinical decision-making in PA. Errors occur in both directions: concluding that a nodule is the source of aldosterone excess may lead to erroneous adrenalectomy; whereas by relying on radiological appearances for a diagnosis of BAH patients may be denied potentially curative surgery. Rates of 'misdiagnosis by imaging alone' vary between studies but are consistently substantial [10].

AVS

AVS is widely accepted as the criterion standard to distinguish between unilateral and bilateral aldosterone production in patients with PA. The test involves the catheterization, via a femoral vein puncture, of both adrenal veins, guided by fluoroscopy and with confirmation of correct catheter placement by injection of contrast medium. Blood is obtained from both adrenal veins and the inferior vena cava (IVC) well below the renal veins (some advocate using an iliac vein), with all samples assayed for cortisol and aldosterone. The left adrenal vein is a tributary of the inferior phrenic vein and this sample is usually taken from close to the junction of the two. The right adrenal vein is the more difficult to cannulate on account of its

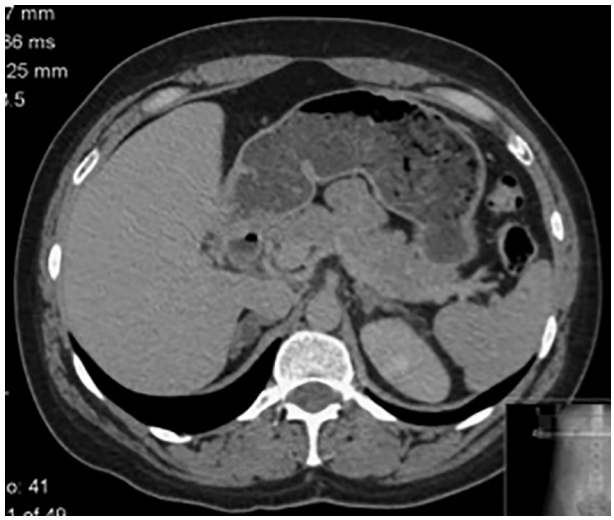


Figure 5.6.2.1 PA was diagnosed in a 38-year-old woman with severe hypertension who wished to conceive: PAC 780 pmol/L; PRA <0.01 pmol/ml/h. She was normokalaemic. CT confirmed a focal, hypodense nodule in the right adrenal, but concern about equivocal contralateral minor abnormal adrenal morphology mandated AVS, with results as shown in Table 5.4.4.3. The right adrenal vein was inadequately cannulated (SI 1.76), but the left adrenal vein: IVC cortisol-corrected aldosterone value was suppressed compared to that in the IVC (see Table 5.4.4.3). Together with the presence of the focal abnormality it was considered that this represented sufficient evidence, in a young patient, to offer a right adrenalectomy. She conceived shortly after surgery and remains normotensive several years later on no medication. In selected cases, careful interpretation of incomplete AVS may facilitate clinical decision-making.

length (approximately 2 mm), diameter (approximately 1 mm) and the fact that it enters the IVC posteriorly at an acute angle. It can sometimes be confused with adjacent small hepatic branches. With experience, repetition, persistence, and careful liaison with endocrine colleagues for feedback and audit, successful cannulation rates in the best centres rise above 90%. This process can be supported by intraprocedural rapid cortisol measurements, which provide the radiologists with additional information regarding correct catheter tip placement [11, 12]. Complication rates of AVS such as adrenal haemorrhage are low in high volume centres and are inversely correlated with radiological experience [13]. For clinical endocrinologists, the limited number of radiologists able to work to a high standard is a major contributor to the small proportion of patients with PA that are fully investigated.

Ideally, patients should be taking either no antihypertensive therapy at all, but in the majority of cases clinical safety dictates that non-renin/aldosterone interfering agents such as alpha$_1$ adrenoceptor antagonists (doxazosin, terazosin) and/or non-dihydropyridine calcium antagonists (verapamil) and/or hydralazine will be required for blood pressure control. Of note, a recent study showed good performance of AVS in differentiating APA from BAH irrespective of intake of mineralocorticoid receptor antagonists [14]. Potassium concentrations should be normal, as hypokalaemia restrains aldosterone secretion. Discontinuation of antiplatelet therapy is not necessary. Discontinuation of anticoagulants is at the discretion of the radiologist, although most experienced operators do not mandate it. In most centres, AVS is a day case procedure performed in the morning in order to allow sufficient time to elapse before safe discharge given the femoral vein puncture.

Variations and Refinements in AVS Protocols

The aforementioned is a 'generic' description of the principle of AVS, but different, very experienced, centres advocate various, subtle refinements for which controlled, prospective, comparative data are currently not available. Many units perform AVS under conditions of adrenocorticotrophic hormone (ACTH) stimulation, either continuously infused or administered as a bolus. Perceived advantages of this refinement include minimizing stress-associated fluctuations in cortisol and aldosterone; maximizing the gradient in cortisol from adrenal to peripheral (IVC or iliac) vein; and maximizing aldosterone secretion from an APA. The first of these is particularly relevant to the practice of sequential (as opposed to simultaneous) AVS, as a stress-induced surge of adrenal secretion has the potential to alter the results. Proponents of the unstimulated, bilateral simultaneous technique (a more challenging and prolonged protocol that slightly increases the risk of adrenal vein thrombosis if the catheters remain in position for long periods) point to the pulsatile nature of aldosterone secretion and the possibility of generating erroneous gradients between adrenals with sequential sampling. Bilateral, sequential, unstimulated sampling is probably the least practised technique. The authors (at whose centre sequential, ACTH-stimulated sampling is performed) take the pragmatic view that any subtle advantages of one variation over another are comfortably outstripped by the benefits of a single, experienced radiologist achieving bilateral cannulation rates in excess of 90% using a protocol with which he/she and the accompanying radiographer/nursing staff are comfortable and familiar.

Interpretation of AVS Data

The success of cannulation is determined using the selectivity index (SI), defined as the ratio of the cortisol concentrations in the adrenal veins to that in IVC/iliac vein. The 'cut-offs' used to determine successful AV catheterization are, inevitably, somewhat arbitrary. It is clearly appropriate to use a lower SI with unstimulated as opposed to stimulated AVS and, with both protocols, the more stringent the value used, the lower will be the proportion of patients judged to have had adequate AVS. A recent consensus statement recommended values of 3 and 2: 1 for stimulated and unstimulated protocols respectively [13]. In the authors' institution, a SI of 5:1 is used, but this is applied flexibly on a case-by-case basis.

The PAC at each of the sample sites (adrenal veins and IVC/iliac vein) is divided by the cortisol to generate a value termed the 'cortisol-corrected ratio', although a few centres refer to a 'normalized aldosterone'. 'Corrected' and 'normalized' denote the concept of correcting for the dilutional effect of non-adrenal blood (notably the inferior phrenic and accessory hepatic veins on the left and right, respectively). The cortisol-corrected ratios in each adrenal are then compared to produce a lateralization index (LI) and this is used to determine whether aldosterone excess is likely to be unilateral or bilateral. No prospective, randomized trials are available for determining the best LI cut-off to use for clinical decision-making; all published data come from retrospective, observational studies. Consensus criteria settle around LIs of 4.0 and 2.0 with and without ACTH stimulation, respectively, above which PA is highly likely to be unilateral and

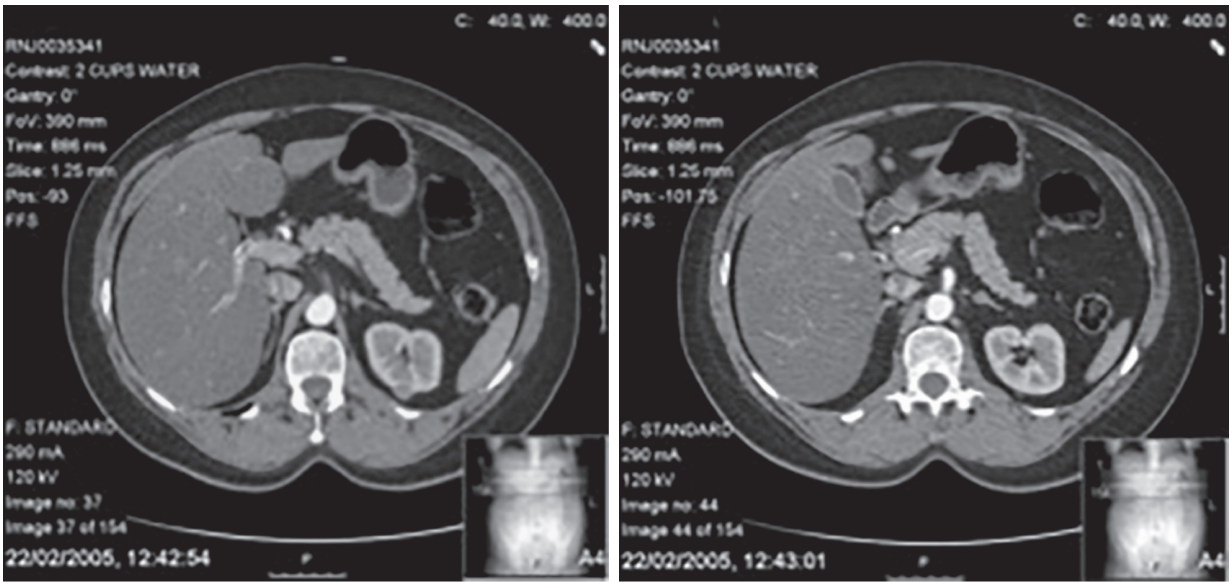


Figure 5.6.2.2 A 54-year-old woman with longstanding hypertension resistant to three agents was investigated for secondary causes. After withdrawal of interfering medication, PA was confirmed: serum potassium 3.1 mmol/L (normal 3.5–5.1), serum bicarbonate 31 mmol/L (normal 22–28), PRA <0.001 pmol/ml/hr, PAC 782 pmol/L. CT scanning showed normal appearances of the right adrenal with a small, hypodense nodule within the left adrenal. She underwent AVS, with results as shown. Lateralization was not confirmed (LI<4) and she was treated medically with spironolactone. This case highlights that, in PA patients over the age of 35–40, the presence of solitary, focal adrenal nodules cannot be assumed to imply a unilateral source of aldosterone excess.

Table 5.6.2.3 Data for Figure 5.6.2.1

Vein	Aldosterone (A), pmol/L	Cortisol (C), nmol/L	A/C	LI
Right adrenal vein	6620	1198	N/A	N/A
Left adrenal vein	9230	13470	0.7	N/A
Inferior vena cava	4148	682	6.2	

Table 5.6.2.4 Data for Figure 5.6.2.2

Vein	Aldosterone (A), pmol/L	Cortisol (C), nmol/L	A/C	Lateralization index
Right adrenal vein	76700	23830	3.7	2.2
Left adrenal vein	50500	29450	1.7	
Inferior vena cava	1368	1178	1.2	

Table 5.6.2.5 A 45-year-old man was referred to the authors' institution for a second opinion after a right adrenalectomy for PA was quickly followed by recurrent, hypokalaemic hypertension. He had presented 2 years earlier with hypertension resistant to three agents and thiazide-provoked mild hypokalaemia (serum potassium 3.4 mmol/L). PA was confirmed (PAC 1620 pmol/L; PRA undetectable) and adrenal imaging was reported as normal. The results from AVS are shown below. On the basis of the LI, the patient was advised to undergo a right adrenalectomy. Two months after the operation, the blood pressure was unchanged from its presenting value, identical biochemistry was recorded and he was referred to the authors' centre. Although the LI from right to left is >4, the contralateral (in this case left) adrenal vein A/C value is not suppressed compared to the peripheral (IVC) and the absolute aldosterone value is substantially greater than background. The authors' interpretation of this AVS is that there was bilateral aldosterone production, with predominance of the right over the left adrenal and that an adrenalectomy was unlikely to be curative. This case underscores the need to examine the raw data, in addition to calculating a LI. The patient continues MR antagonist therapy

Vein	Aldosterone (A), pmol/L	Cortisol (C), nmol/L	A/C	LI
Right adrenal vein	1445800	14559	99.3	4.8
Left adrenal vein	88 300	4317	20.5	
Inferior vena cava	3820	754	5.1	

patients should be considered for surgical treatment. With ACTH stimulation, a LI <3.0 is strongly indicative of bilateral disease; values between 3.0 and 4.0 represent an area of uncertainty. The authors, in general, work to the aforementioned consensus guidelines, but are mindful of the potential effects of cortisol cosecretion on AVS data and, in patients whose LI values fall close to the accepted cut-offs, also take into account, for any individual patient, the anatomy and other factors (age, PAC, severity of hypertension, marked hypokalaemia) that affect the *a priori* probability of PA being unilateral.

Controversies and Uncertainties in AVS

Common to all centres, at present, is the use of cortisol as the denominator for calculating a corrected ratio with which, subsequently, to determine a LI. This practice is increasingly being questioned, partly because of the growing recognition that cortisol cosecretion in PA affects a significant proportion of patients [15] and that this may influence AVS results. Low-grade autonomous cortisol secretion from one adrenal, sufficient to lower ACTH such that cortisol production by the contralateral gland is partially or completely suppressed, may result in the erroneous conclusion of failed cannulation in that adrenal vein. This issue applies to both stimulated and unstimulated AVS techniques. Even if, with ACTH stimulation, the SI is satisfied, it may still be the case that cortisol output is relatively restrained. By reducing the value of the denominator on the 'cortisol suppressed' side, the A/C value may be augmented, potentially leading to the erroneous conclusion that aldosterone production is bilateral. This has led to interest in the use of androstenedione and metanephrines in place of cortisol [16, 17], although at present this is not widely practised.

The requirement to discontinue mineralocorticoid receptor (MR) antagonists for 4–6 weeks prior to AVS poses problems for those patients whose severe hypertension is poorly responsive to other medications. This has led some groups to explore whether meaningful data may be derived from patients who, for safety reasons, cannot discontinue spironolactone or epleronone in preparation for AVS. In one small series, four patients underwent a unilateral adrenalectomy on the basis of AVS results obtained while taking spironolactone [18]. Importantly, the renin values were not derepressed by therapy and in all four patients the AVS correctly predicted the clinical and biochemical outcome after surgery. More, and larger, series will be needed before this practice can routinely be recommended and it represents a potential advantage of functional adrenal imaging, discussed next.

Given the technical difficulties of AVS, the question arises as to whether useful data for decision-making can be derived from studies in which only one vein (usually the left) is successfully cannulated. It has been suggested that suppression of the left adrenal vein A/C value ≤0.5 in comparison to the IVC value may imply that the source of aldosterone excess is the right adrenal and direct clinicians to recommend surgery [19]. Conversely, a left adrenal A/C value sufficiently greater than the IVC (≥5.5) may imply that a contribution from the right adrenal to aldosterone excess is improbable. It is hard to know if these thresholds can safely be applied to other centres using different protocols and their application does not address the underlying issue, for many patients, of lack of access to AVS.

Alternatives that Reduce the Requirement for AVS

Although the details and interpretation of AVS are debated, and some question its status as the 'criterion standard', most clinicians accept the need to divide PA patients into one of three groups: those with a high probability of unilateral disease, most of whom will be referred for surgery; those with a high probability of bilateral disease, treated medically; and those in whom there is sufficient uncertainty to require AVS. The discrepancy between the size of group 3 and the capacity for high-quality AVS has led to attempts to develop 'prediction scores' that can help move some extra patients into group 1. One scoring tool (a clinical prediction score, CPS) combined a clear-cut macroadenoma on CT with hypokalaemia and a threshold GFR, with a relative 'importance' weighting of radiology>severe hypokalaemia>mild hypokalaemia>GFR [20]. However, the reference standard against which the CPS was developed was AVS rather than surgical outcome. Given that glomerular filtration rate (GFR), for many patients, represents a surrogate for duration of hypertension (and, in turn, age of the patient), the CPS adds little to the *a priori* suspicion that a patient with a clear, unilateral nodule and normal contralateral adrenal morphology, marked hypokalaemia, and younger age has a high likelihood of unilateral PA. Other groups have found the CPS less accurate [21] and it is not widely used.

Nuclear Medicine Imaging in PA

The technical demands and lack of availability of AVS have, for many years, fuelled interest in the development of functional imaging in PA. The hazards of relying on CT imaging alone are discussed earlier, but combining anatomical detail with evidence of functional activity is an attractive concept and one that has the potential to allow more PA patients to be fully investigated. ^{131}I-iodomethyl norcholesterol and ^{75}Se-selenomethyl-19-norcholesterol have been used in past, but their overall accuracy and the requirement for prolonged use of high dose dexamethasone to suppress non-autonomous adrenal activity renders them unsatisfactory as an alternative to AVS.

More promising is the development of positron emission tomography (PET)-CT scanning, as it provides better spatial resolution than conventional planar scintigraphy. The most advanced tracer agent in use is metomidate, a methyl analogue of the anaesthetic agent etomidate. Metomidate is a potent inhibitor of CYP11B1 (11β-hydroxylase) and CYP11B2 (aldosterone synthase) and can be $C^{11}H_3$-labelled as a PET radiotracer (^{11}C-metomidate). The use of preparatory dexamethasone, to suppress non-adenomatous adrenal activity, appears to improve the diagnostic accuracy. In a non-randomized proof-of-concept study ^{11}C-metomidate PET-CT performed similar to AVS in distinguishing unilateral from bilateral PA [22]. Larger-scale, outcomes-based studies comparing metomidate PET-CT with AVS are awaited, but accumulated experience and published case reports suggest that a particular strength of the technique is its ability to identify small APAs, not easily visible on CT, whose histological architecture has a zona glomerulosa appearance and whose cells show high expression of CYP11B2 (aldosterone synthase). The requirement for an on-site cyclotron imposes limitations on the availability of this technique,

but newer tracers are in development and may be easier to deploy if published evidence supports their use. Alternatively, [^{123}I]-metomidate-SPECT-CT could be used, which has been evaluated for use in adrenal masses [23] but not yet systematically in primary aldosteronism; in this format local cyclotron facilities would not be required.

Other potential ligands for use in PET-CT scanning for PA include ^{68}Ga-pentixafor (which targets the CXC chemokine receptor-4) [24] and a compound with some structural homology to metomidate ^{8}F-CDP2230 [25]). To date, published experience with these ligands is confined to nine patients and rodents, respectively.

Steroid Metabolomics

Historically, pathological analysis of resected APAs was limited to size, colour (typical canary yellow), presence or absence of spironolactone inclusion bodies, and confirmation of benign histopathology, sometimes with a comment about surrounding hyperplasia. Recently, it has become apparent that APAs fall into one of two, broad, groups: those with cells that, paradoxically, resemble zona fasciculata (ZF) cells (often above the median for size) and those that resemble zona glomerulosa (ZG) cells. The latter type is often small and stains densely for CYP11B2 (aldosterone synthase). As described in Chapter 5.6.1, several somatic mutations have been described in APAs and there is an emerging consensus that KCNJ5 mutations are frequently associated with ZF-like histopathology, whereas mutations in certain calcium channels and ATPases tend to produce smaller tumours with ZG-like morphology. Interest has grown in the possibility that distinct steroid 'fingerprints' are associated with particular APA genotypes [26–29]. Increased concentrations of the hybrid steroids 18-hydroxycortisol and 18-oxocortisol have been documented in the adrenal and peripheral vein samples of patients carrying KCNJ5 mutated APAs [28, 30]. This has an attractive, plausible explanation: such tumours have increased expression of CYP11B1 (11β-hydroxylase) and CYP17A1, whereas these are absent from the ZG cells where aldosterone is physiologically produced. Increased availability of 11-deoxycortisol could result in its conversion by CYP11B2 to hybrid steroids, resulting in higher detected levels compared to other APAs. Evidence is beginning to point to KCNJ5-mutant adenomas as the group most likely to be solitary and unilateral, but these data will need to be robust before the finding of hybrid steroids reduces the requirement to perform AVS.

The use of liquid chromatography-tandem mass spectrometry (LC-MS/MS) to measure adrenal steroid levels including hybrid steroids in peripheral blood samples for the purpose of distinguishing APAs from BAH has been also tested. Statistically significant differences are demonstrable, although the degree of overlap is likely to limit the widespread use of this investigation. Again, such measurements, at best, can only provide information to suggest that BAH is likely (and therefore AVS is unnecessary) and it seems likely that advances in functional imaging will impact more on reducing AVS requirements and increasing the numbers of PA patients proceeding to surgery.

Surgical Treatment of PA

Laparoscopic adrenalectomy (see Chapter 5.2 for details) is now well-established as the treatment of choice for the majority of patients in whom it is possible to demonstrate a unilateral source of aldosterone excess. Morbidity is low in experienced hands and its development and subsequent refinement through the 1980s and 90s radically altered the threshold for surgical referral. (Prior to laparoscopic surgery, open adrenalectomy involved a large flank incision and sometimes resection of the twelfth rib; morbidity and mortality rates were substantial.) Various techniques are described: anterior or lateral transperitoneal, lateral or posterior retroperitoneal and (in selected centres) a single port retroperitoneal approach. The lateral transperitoneal route is the commonly used technique and involves insertion of three or four ports with the patient in a left or right lateral decubitus position for right- and left-sided operations, respectively. In both procedures the patient is flexed above the hip to open up the space between the iliac crest and rib cage. For a left adrenalectomy, ports are usually inserted below the rib cage along the anterior, posterior axillary, and mid-clavicular lines; for a right-sided procedure the ports are similarly sited, but with the addition of an extra subcostal midline port to facilitate liver retraction. Although cross-sectional imaging portrays the adrenal gland as being clearly distinct from the surrounding retroperitoneal fat, the reality is that, following mobilization of the surrounding structures (spleen, spleno-colic ligament, and tail of pancreas on the left; liver on the right) the adrenal gland may be challenging to identify, particularly if there is excess visceral fat. It is important to be aware of the surrounding structures and the planes which define the position of the adrenal gland, especially in patients with PA as the tumour may be too small to visualize and the adrenal gland is of normal dimensions. It is largely this 'judgement' aspect of adrenal surgery that dictates better results among more experienced surgeons, although no consensus exists as to what constitutes an appropriate minimum number of procedures per year to achieve good results. Unlike adrenal surgery for Cushing's syndrome or phaeochromocytoma, few, if any, anaesthetic issues arise over and above other laparoscopic (e.g. gall bladder) procedures, but close liaison regarding the possibility of cortisol cosecretion and postoperative contralateral adrenal suppression is important. Overt haemodynamic collapse due to postoperative adrenal insufficiency is rare, but described. Transient postoperative hyperkalaemia is well recognized and treated either with augmented sodium intake or temporary support with fludrocortisone.

Medical Therapy of PA

Medical treatment of PA is indicated for patients whose imaging and/or AVS results indicate that a unilateral adrenalectomy would be inappropriate or who, after careful discussion, are unfit/unwilling to undergo surgery. Given that aldosterone has deleterious effects on the heart and vasculature independent of blood pressure, the goals of pharmacological therapy of PA are not just control of hypertension, but also effective blockade of the MR.

General Measures

Although often overlooked and not formally studied, it is the authors' view that generic advice applicable to all patients with hypertension should be reinforced to patients with PA. This includes restriction of dietary sodium, maintenance of a healthy body weight, abstinence from tobacco and excessive alcohol intake, and regular aerobic exercise. All are likely to make useful contributions to making pharmacological therapy more effective.

Spironolactone

This is the drug for which most clinical experience exists in PA. Although no placebo-controlled trials of its use exist, there is sufficient consensus about its efficacy that most physicians would now consider such a study to be unethical. Its mechanism of action is competitive inhibition of cytosolic MRs, thereby modifying both transcription and translation of proteins that act to mediate sodium retention by the epithelial cells of the distal nephron. The drug is well absorbed and undergoes rapid hepatic metabolism to several metabolites including 7α-thiomethylspironolactone and canrenone. Both have a half-life substantially greater than the parent drug itself (1.4 hours) such that spironolactone is usually administered once daily. Oral canrenone is available in some countries but its use in PA is not widespread.

Individual physicians vary in their dosing strategies; the authors' practice is to start at modest doses (typically 25 mg daily), increasing slowly as tolerated and with frequent monitoring of blood potassium levels. An important clinical point is that the full hypotensive effects of spironolactone (typically 2 months or more) lag by several weeks behind the effect on potassium levels. An interesting, although not well studied, phenomenon is that, once blood pressure control is established, with time it is often possible to reduce the dose without loss of efficacy.

Various observational studies exist, some containing exclusively BAH patients and others a mixture of APA and BAH, to suggest excellent clinical efficacy either as monotherapy or by facilitating a reduction in other non-PA directed antihypertensive therapy. Dosing regimens vary markedly in the published series, from 25 to 400 mg daily, so rigid dosing protocols are hard to construct. Pragmatically, it is the blood potassium levels that govern dosing; once values approach the upper limit of the reference range further increments in dose are unlikely to be clinically beneficial without provoking problematic hyperkalaemia.

Spironolactone is not a 'pure' MR antagonist; its interaction with progesterone and androgen receptors explain its common, frequently dose-limiting, side effects of menstrual irregularity in premenopausal women and erectile dysfunction and painful gynaecomastia in men. The last of these is clearly dose-related, but such side effects tend to be under-reported in observational studies, so reports of an incidence of 50% at doses of 150 mg daily are, in the authors' view, improbably low. Other practical points about spironolactone usage include the need for regular monitoring of serum creatinine and potassium in patients with abnormal baseline renal function and/or diabetes mellitus, a reduction in efficacy with concomitant salicylate use and prolongation of its half-life with advanced liver disease or heart failure.

Inevitably, the relatively recent appreciation of adverse cardiometabolic consequences of aldosterone excess independent of blood pressure render historical observational reports of the efficacy of spironolactone somewhat incomplete. Although a systematic review suggested that regression of many of the adverse features of PA (LVH, proteinuria and renal function) were equivalent between medically and surgically treated patients [31], a recent, large, retrospective study suggested that current MR antagonist regimens do not restore the morbidity of PA to that of essential hypertension except (on subanalysis) in patients whose plasma renin was derepressed by an adequate dose of spironolactone [32]. This remains an area of uncertainty in the field.

Eplerenone

Developed and approved for the treatment of hypertension, this second-generation MR antagonist has fewer antiandrogen and antiprogesterone side effects than spironolactone consequent upon its 9.11-epoxide group which ensures more selective receptor affinity. It has few, if any, active metabolites and its half-life mandates twice daily dosing. Like spironolactone, no placebo-controlled studies for eplerenone exist in PA, but several trials have compared the two drugs against each other in this clinical setting [33]. Spironolactone appears consistently more effective at lowering blood pressure in PA, albeit at the expense of the side effects described earlier. One head-to-head study suggested a dose equivalence of 1:2 (spironolactone 50 mg = eplerenone 100 mg), but there was substantial crossover from spironolactone to eplerenone during the study [34]. Together with its greater cost, eplerenone should be considered second line medical therapy for PA and a good alternative when adverse symptoms limit the use of spironolactone. The maximum licensed dose is 100 mg daily, but this is often insufficient to achieve blood pressure control. The requirement to monitor blood potassium and creatinine levels is similar between the two MR antagonist drugs.

Non-MR Antagonist Medical Therapies

If adverse symptoms and/or cost and/or availability dictate that neither spironolactone nor epleronone can be used, high dose amiloride is an alternative. Although many of the mechanisms of action of aldosterone outside the kidney remain to be fully elucidated, at the distal tubule it is the amiloride-sensitive epithelial sodium channel (ENaC) that is the rate-limiting step on sodium reabsorption. Its perceived modest effect in PA may, in part, be due to historical underdosing and the authors consider the drug to be a useful adjunctive to a maximum tolerated dose of spironolactone in a patient with troublesome side effects and/or cost constraints on the use of MR antagonists. Sodium competes with amiloride for intraluminal binding to ENaC, such that the efficacy of amiloride can be improved by a reduction in sodium intake. The use of triamterene is uncommon, but, conceptually, is similar to amiloride.

Possible Future Medical Therapies

Finerenone is a non-steroidal MR antagonist that has shown promising results in patients with heart failure, but has yet to undergo clinical studies in PA. Aldosterone synthase (CYP11B2) inhibitors

represent an attractive concept for the treatment of PA. Older drugs in this class were compromised by their poor selectivity between the CYP11B2 and CYP11B1 enzymes involved in aldosterone and cortisol synthesis, respectively, whose structures are 93% identical. A few highly selective newer drugs are at the early stages of evaluation.

Decision-Making in PA: Medical or Surgical Management

An inevitable consequence of the way in which clinical trial and/or departmental audit/outcome data are presented is that decision-making can seem clear-cut to the point of 'rigidity'. The reality, for many patients, is that imaging and biochemical data can be equivocal and choices about medical or surgical treatment difficult. It has become commonplace in clinical teaching to state that 'decision-making should be within a multidisciplinary team (MDT)'. The authors subscribe to this practice, with some clarifying views. First, MDTs function best when constructive, robust audit programmes are embedded within the group culture, so that experience can be shared and augmented. Second, such audit and outcome data should be generously shared around the PA community, lowering the threshold for collaborative studies to improve practice. Last, although discussion among colleagues about imaging and AVS results is always helpful, individual knowledge of the case and personal clinician responsibility are paramount to good decision-making and clinical care.

It should be emphasized that the SI and LI values quoted earlier are guides rather than 'absolutes'. As with all endocrine testing (inferior petrosal sinus sampling and MRI scanning of the pituitary for ACTH-dependent Cushing's syndrome is another good example), results are always interpreted against a background of the *a priori* likelihood of a particular diagnosis. Careful review of the anatomy, results of AVS if performed, the presence or absence of cortisol cosecretion, patient age, severity of hypokalaemia (if present) and the diagnostic values of renin and aldosterone all contribute to an integrated decision-making process that improves with personal and shared experience.

Clinical Outcomes

The importance attached to distinguishing between unilateral and bilateral disease assumes that the former may be cured by adrenalectomy while the latter will not. However, the difference in outcomes is not as clear-cut as might be expected, and advances in molecular understanding of the pathogenesis also point to a blurring between the unilateral and bilateral subtypes. Immunohistochemistry with specific antisera against CYP11B2 shows most 'hyperplasia' to consist of discrete 'aldosterone-producing cell clusters', of which 30–60% have the same mutations as reported in the smaller, ZG-type APAs [35]. The frequency of these clusters, reported in at least 30% of post-mortem adrenals, and of the gene mutations they harbour, raises suspicion that fewer patients have unilateral disease than originally thought; and that apparently successful surgery may be due to 'debulking' rather than 'curing' disease. It is well known that only 30–60% of patients with rigorously proven unilateral PA are cured of hypertension, even though PA is biochemically cured in 90–95% of such patients. The likely explanation is the long duration of PA (usually decades) before diagnosis, resulting in irreversible vascular changes that sustain hypertension, even after the original cause is removed. This probably explains why younger patients, with less severe hypertension, have a greater likelihood of complete cure from hypertension.

The commonly agreed practice of integrating anatomical information and AVS results into clinical decision-making was recently challenged by the results of what is, to date, the only published prospective, randomized study in the field of PA and therefore worthy of particularly detailed discussion and analysis here. In SPARTACUS (Subtyping Primary Aldosteronism: A Randomized Trial Comparing Adrenal Vein Sampling and Computed Tomography Scanning), 200 patients were enrolled and randomized to have their clinical decision regarding medical or surgical management determined by AVS (preceded by CT, to exclude ACC and provide anatomical information for surgery, where relevant) or CT alone [3]. For those randomized to CT-based decision-making, surgery was undertaken only in the event of a unilateral radiological abnormality with unequivocally normal contralateral appearances. Patients with any other radiological findings (normal, bilaterally enlarged adrenal glands or a focal nodule with subtle contralateral thickening) were treated medically. The data from 184 patients who completed the protocol separated into four groups: CT-determined medical and surgical management; and AVS-determined medical and surgical management, with, neatly, 46 in each group. The primary endpoint of SPARTACUS was the burden of antihypertensive treatment required to achieve a predefined target blood pressure at 1 year.

There was no difference between the groups with regard to the intensity of antihypertensive drug treatment, mean blood pressure, proportion of patients reaching target blood pressure or quality of life at 1 year. The authors concluded that neither AVS nor CT scans should be considered as gold standard tests for identifying an APA and that the considerable expense of AVS could be jettisoned without compromising clinical outcomes.

The controversial nature of the reported findings has been associated with some criticism of the study design. First, the choice of primary endpoint is strikingly different from other reported outcome studies in PA and not in line with a recent international consensus statement on the definition of cure following adrenalectomy [4]. Second, the patients recruited into the study were at the more severe end of the PA spectrum; a group known to have more marked target organ damage, including LVH, renal damage, and vascular remodelling, each of which may contribute to persistence of hypertension after biochemically curative adrenalectomy. This, in turn, dictates that such patients would be less likely to be achieve satisfactory control of hypertension with either surgical or medical treatment, reducing the chances (with the size of study population) of detecting a difference between the randomized groups. Third, the perceived 'burden' of antihypertensive therapy depends, in part, on the choice of therapy for any residual, fixed hypertension as a consequence of longstanding PA or background essential hypertension, even if surgery leads to biochemical cure. In SPARTACUS, the antihypertensive medicines deployed in the study patients at 1 year was unusual. Fourth, the comparison of CT vs. AVS-determined medical therapy reduced the overall power

of the study and, given that APA patients respond perfectly satisfactorily to MR antagonists, did not provide significant clinical information.

With all those caveats, SPARTACUS is, to date, the only prospective, randomized study that has attempted to address/challenge the issue of whether AVS can justify its status as the 'criterion standard' for dictating clinical management. The investigators accept that CT scanning is associated with significant misclassification rates, but reasonably point out that, in their hands, AVS is flawed too and have re-energized the debate as to whether it is appropriate that a restricted, expensive investigation should continue severely to limit the access of PA patients to potentially curative surgery.

Cortisol Cosecretion in Primary Aldosteronism

Abnormal (non-ACTH mediated) cortisol secretion in PA has been known about for some time [35] but the availability of specific antisera to CYP11B1 has provided further insights. Total 24-h urinary glucocorticoid excretion in primary aldosteronism significantly correlates with immunohistochemical expression of CYP11B1 in the corresponding APA tissue [15]. Further, clinical surrogate parameters of metabolic risk, such as waist circumference and diastolic blood pressure, correlate significantly with cortisol, but not with aldosterone, production. This is potentially clinically relevant as studies in large cohorts of PA patients have demonstrated an increased risk of type 2 diabetes and osteoporosis compared to population-based controls [37, 38]; adverse outcomes usually associated with glucocorticoid rather than mineralocorticoid excess. It is conceivable, but not proven, that cortisol cosecretion in PA drives these adverse metabolic outcomes [39]. Of note, a recent study suggested that both LVH at diagnosis of PA and its reversibility correlated with cortisol co-secretion as assessed as total 24-h urinary glucocorticoid excretion [40].

In a separate report, 45 of 175 patients who underwent a postoperative short synacthen test after surgery for PA, 29% were shown to have a suboptimal response [15]; a figure supported by a separate study [42]. The clinical significance of this is uncertain and reports of overt post-operative hypadrenalism are very rare [43], but subtle abnormalities of cortisol requiring hydrocortisone supplementation may be more common and clinical awareness is paramount. This rate of adrenal insufficiency was confirmed in a recent prospective series carried out in 100 consecutive patients with primary aldosteronism due to unilateral APA; 27% of patients failed the postoperative synacthen test and required hydrocortisone replacement therapy for several months, with significant adrenal crisis risk [42]. The increasing awareness of the possible clinical implications for cortisol co-secretion are likely to feature more prominently in future iterations of consensus guidelines for the management of PA.

There is no doubt that these findings will be considered when the next American and European consensus guideline revisions will be undertaken and given the observed prevalence of postoperative adrenal insufficiency, the integration of an early postoperative short synacthen test in surgically treated primary aldosteronism patients may be prudent.

Rare Forms of PA

The previous discussion of APA and BAH applies to around 98% of cases of PA. Rare forms are recognized and briefly described here, but are dealt with in more detail elsewhere.

Aldosterone-Secreting Adrenocortical Carcinoma (ACC)

This diagnosis will be immediately apparent on cross-sectional imaging. Tumours are typically large (>4 cm) and often irregular, heterogeneous, and invasive. The only chance of cure is surgical resection. Other therapeutic options for non-curable ACC are covered in Chapter 5.4 and should be combined with MR antagonists and/or high dose amiloride therapy to control hypertension and hypokalaemia. Pure aldosterone-secreting ACCs are rare; more common is for an array of secreted steroids to include deoxycorticosterone (DOC), but the principles of management remain the same.

Familial Forms of PA

Various forms of inherited PA exist, distinguished by their clinical features and genetic causes; a detailed description of these conditions is provided in Chapter 5.6.1. The term is restricted to a small number of very rare conditions, in which germline mutation of a specific gene is identified. The current nomenclature, in which a different number is used for each genetic abnormality, is not practised in other areas of endocrinology and may not survive the discovery of further genes.

Special Situations in PA

As with many endocrine conditions, particular clinical situations arise for which few or no data exist to guide decision-making. Under these circumstances, physicians have to rely on judgement, backed up by collaborative discussion with experienced colleagues. Two such situations are the combination of PA and pregnancy and PA with advanced renal impairment.

PA in Pregnancy

The frequency with which PA occurs in pregnancy is not known. If the prevalence of hypertension in pregnancy (6–8%) is combined with the prevalence of PA in unselected hypertensive populations, then approximately 0.5–0.8% of pregnant women will have PA. One could argue that the majority of women with PA are diagnosed after childbearing years, making the aforementioned figure an overestimate, but most PA is diagnosed opportunistically and the chances of detecting asymptomatic hypertension are higher in pregnancy as blood pressure monitoring accompanies most obstetric/midwifery encounters. As discussed earlier, PA has few clinical features, so suspicion of PA arises in the context of hypokalaemia and/or severe hypertension before 20 weeks of gestation (after which the likelihood of pre-eclampsia as a cause rises). Supportive biochemical evidence for a diagnosis of PA is largely restricted to a suppressed

renin, as PAC rises in normal pregnancy. Standard confirmatory tests (SIT, CCT) should wait until after delivery. If PA is strongly suspected, then medical therapy with MR antagonists should be commenced. Classical teaching is that the antiandrogen properties of spironolactone dictate this drug is contraindicated in pregnancy, but in fact only one case report exists of ambiguous genitalia in a male infant associated with use of the drug in the first few weeks of gestation [44]. Eplerenone has no antiandrogenic action but published data on pregnancy outcomes are sparse. The authors recommend epleronone for use in pregnancy and breast feeding if PA is suspected or confirmed, but if this drug is unavailable or not tolerated, then given that sexual differentiation is largely complete by the end of the first trimester, and considering the context of the hazards of uncontrolled hypertension in pregnancy, the benefits of spironolactone on blood pressure and potassium control after this time probably outweigh any theoretical issue with male fetal development. If hypertension and/or hypokalaemia cannot be controlled medically and the diagnosis of PA is secured prepregnancy or strongly suspected during pregnancy on account of a suppressed renin measurement, then a mid-trimester adrenalectomy may be considered. Adrenal imaging should be with (non-ionizing) MRI; given the age of most pregnant women, incidental adrenal nodules are improbable, so the finding of a solitary nodule is usually sufficient evidence with which to proceed. AVS should not be performed in pregnancy.

PA and Renal Failure

The combination of PA and advanced renal failure presents particular diagnostic and therapeutic problems. Clinical suspicion is difficult, given that most patients with renal failure will have hypertension requiring therapy with more than one agent; the characteristic hypokalaemic alkalosis will be obscured by the renal disease; and secondary hyperaldosteronism is common as kidney function declines. Imaging quality may be compromised by the need to avoid iodinated contrast agents for CT scanning and some gadolinium containing agents are contraindicated in advanced renal disease because of the risk (albeit with repeated administration) of a characteristic dermopathy. Whereas some people may instinctively feel that, once renal failure is established, investigation and treatment of PA is too late and futile, the authors take the reverse view. It is the combination of hypertension and excess aldosterone exposure, in many cases, that has led to accelerated renal disease and correction of the latter almost invariably facilitates better treatment of the former, even in advanced cases. Further, in those patients who have progressed to end-stage renal disease (ESRD) requiring dialysis, it might be thought that the physician has a degree of control over salt and water balance and that the renal actions of aldosterone on sodium retention are irrelevant in an anuric patient. The fact that spironolactone improves blood pressure in patients with PA on dialysis provides compelling evidence for the extrarenal effects of aldosterone and that removing or pharmacologically blocking that excess will be clinically beneficial. In patients with ESRD contemplating transplantation, removal of an APA will prevent the scenario of a resumption of sodium and water retention and hypertensive damage to the grafted kidney. AVS requires the use of a modest volume of contrast agent and close nephrological liaison is advised.

Newer Techniques for Treating Unilateral Aldosterone Excess

Annual UK audit returns indicate that 250–300 adrenalectomies per year are performed for PA. Using even the most conservative estimates of the prevalence of PA (approximately 300 000) and assuming that 50% of cases are due to a unilateral APA, it is immediately apparent that, if case detection and access to accurate lateralization investigations were dramatically to improve, then endocrine surgical services would quickly become overwhelmed. This, together with a desire to avoid removing a whole gland as treatment for a small, benign tumour has led to a search for less invasive methods of removing the source of aldosterone excess. Radiofrequency ablation (RFA) is an established technique for the treatment of liver deposits from large bowel, neuroendocrine, and other malignant tumours and some centres have developed protocols to destroy functioning adrenal nodules. Slight variations are described, but, in general, patients are positioned prone in the lateral decubitus position either heavily sedated or under full general anaesthesia. Under CT guidance a monopolar radiofrequency electrode is introduced and, using a pulsed technique and generating a power of 50–150 W, RFA is applied for several minutes. Periadrenal fat stranding is observed and technical success judged by lack of enhancement on a follow-up CT scan. Interestingly, although the nodule being treated is steroid-producing, intraprocedural, urgent hypertensive episodes are not infrequent, necessitating continuous intra-arterial monitoring and appropriate pharmacological intervention.

No randomized studies are available that compare RFA with conventional adrenalectomy, but reports exist of extensive, accumulated single centre experience [45]; and a comparison of clinical outcomes between two neighbouring tertiary United States centres, one of which provides both surgery and RFA on the basis of patient choice [46]. The data are, at present, too preliminary to draw firm conclusions, but there is sufficient 'momentum' to justify a randomized study. Attractions of the technique include the possibility of shorter hospital stays and the potential to treat bilateral disease. With experience and refinement, the availability of RFA may, in time, lower the threshold for intervening in unilateral disease; at present, the scenario of a 'failed adrenalectomy' (one in which hopes for complete cure of hypertension are not met) often leads to caution in recommending abdominal surgery for a benign condition. If an intervention considered to be less invasive were available, it is possible that the threshold for that procedure would be lower than at present. Areas of uncertainty and controversy regarding RFA include the lack of histological proof and complete molecular diagnosis provided by surgery; and the long-term consequences of leaving untreated areas of hyperplasia that often surround the easily visible APA.

Conclusions

Conn's prediction of the number of hypertensive patients who would have the syndrome he described was uncannily accurate. The excess cardiometabolic morbidity associated with PA over and above essential hypertension is now widely accepted; as is the current shortfall in the rate of diagnosis, availability of subtyping investigations and access to potentially curative interventions. The

recent explosion in understanding of the molecular pathogenesis of the condition has fuelled a new wave of investigational protocols that promise to remove many of the barriers that currently block the management of patients with this condition. The authors predict that the next version of this chapter will look strikingly different.

REFERENCES

1. Conn JW. Presidential address. I. Painting background. II. Primary aldosteronism, a new clinical syndrome. *J Lab Clin Med*, 1955; **45**: 3–17.
2. Conn JW, Cohen EL, Rovner DR, Nesbit RM. Normokalaemic primary aldosteronism. A detectable casue of curable 'essential' hypertension. *JAMA*, 1965; **193**: 200–6.
3. Dekkers T, Prejbisz A, Kool LJS, *et al.* Adrenal vein sampling versus CT scan to determine treatment in primary aldosteronism: an outcome-based randomised diagnostic trial. *Lancet Diab Endocrinol*, 2016; **4**: 739–46.
4. Williams TA, Lenders JWM, Mulatero P, *et al.* Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diab Endocrinol*, 2017; **5**: 689–99.
5. Choi M, Scholl UI, Yue P, *et al.* K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science*, 2011; **331**: 768–72.
6. Scholl UI, Goh G, Stolting G, *et al.* Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nature Gen*, 2013; **45**: 1050–4.
7. Teo AE, Garg S, Shaikh LH, *et al.* Pregnancy, primary aldosteronism, and adrenal CTNNB1 mutations. *N Engl J Med*, 2015; **373**: 1429–36.
8. Azizan EA, Poulsen H, Tuluc P, *et al.* Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. *Nature Gen*, 2013; **45**: 1055–60.
9. Funder JW, Carey RM, Mantero F, *et al.* The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 1889–916.
10. Kempers MJ, Lenders JW, van Outheusden L, *et al.* Systematic review: diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. *Ann Int Med*, 2009; **151**: 329–37.
11. Rossi E, Regolisti G, Perazzoli F, *et al.* Intraprocedural cortisol measurement increases adrenal vein sampling success rate in primary aldosteronism. *Am J Hypertens*, 2011; **24**: 1280–5.
12. Yoneda T, Karashima S, Kometani M, *et al.* Impact of new quick gold nanoparticle-based cortisol assay during adrenal vein sampling for primary aldosteronism. *J Clin Endocrinol Metab*, 2016; **101**: 2554–61.
13. Rossi GP, Auchus RJ, Brown M, *et al.* An expert consensus statement on use of adrenal vein sampling for the subtyping of primary aldosteronism. *Hypertension*, 2014; **63**: 151–60.
14. Nanba AT, Wannachalee T, Shields JJ, *et al.* Adrenal vein sampling lateralization despite mineralocorticoid receptor antagonists exposure in primary aldosteronism. *J Clin Endocrinol Metab*, 2019; **104**: 487–92.
15. Arlt W, Lang K, Sitch AJ, *et al.* Steroid metabolome analysis reveals prevalent glucocorticoid excess in primary aldosteronism. *JCI Insight*, 2017; **2**: pii: 93136.
16. Freel EM, Stanson AW, Thompson GB, *et al.* Adrenal venous sampling for catecholamines: a normal value study. *J Clin Endocrinol Metab*, 2010; **95**: 1328–32.
17. Ueland GÅ, Methlie P, Jøssang DE, *et al.* Adrenal venous sampling for assessment of autonomous cortisol secretion. *J Clin Endocrinol Metab*, 2018; **103**: 4553–60.
18. Haase M, Riester A, Kropil P, *et al.* Outcome of adrenal vein sampling performed during concurrent mineralocorticoid receptor antagonist therapy. *J Clin Endocrinol Metab*, 2014; **99**: 4397–402.
19. Strajina V, Al-Hilli Z, Andrews JC, *et al.* Primary aldosteronism: making sense of partial data sets from failed adrenal venous sampling-suppression of adrenal aldosterone production can be used in clinical decision making. *Surgery*, 2018; **163**: 801–6.
20. Kupers EM, Amar L, Raynaud A, Plouin PF, Steichen O. A clinical prediction score to diagnose unilateral primary aldosteronism. *J Clin Endocrinol Metab*, 2012; **97**: 3530–7.
21. Sze WC, Soh LM, Lau JH, *et al.* Diagnosing unilateral primary aldosteronism—comparison of a clinical prediction score, computed tomography and adrenal venous sampling. *Clin Endocrinol*, 2014; **81**: 25–30.
22. Burton TJ, Mackenzie IS, Balan K, *et al.* Evaluation of the sensitivity and specificity of (11)C-metomidate positron emission tomography (PET)-CT for lateralizing aldosterone secretion by Conn's adenomas. *J Clin Endocrinol Metab*, 2012; **97**: 100–9.
23. Kreissl MC, Schirbel A, Fassnacht M, *et al.* [^{123}I]Iodometomidate imaging in adrenocortical carcinoma. *J Clin Endocrinol Metab*, 2013; **98**: 2755–64.
24. Heinze B, Fuss CT, Mulatero P, *et al.* Targeting CXCR4 (CXC chemokine receptor type 4) for molecular imaging of aldosterone-producing adenoma. *Hypertension*, 2018; **71**: 317–25.
25. Abe T, Naruse M, Young WF, Jr., *et al.* A novel CYP11B2-specific imaging agent for detection of unilateral subtypes of primary aldosteronism. *J Clin Endocrinol Metab*, 2016; **101**: 1008–15.
26. Fernandes-Rosa FL, Daniil G, Orozco IJ, *et al.* A gain-of-function mutation in the CLCN2 chloride channel gene causes primary aldosteronism. *Nature Gen*, 2018; **50**: 355–61.
27. Scholl UI, Stolting G, Schewe J, *et al.* CLCN2 chloride channel mutations in familial hyperaldosteronism type II. *Nature Gen*, 2018; **50**: 349–54.
28. Williams TA, Peitzsch M, Dietz AS, *et al.* Genotype-specific steroid profiles associated with aldosterone-producing adenomas. *Hypertension*, 2016; **67**: 139–45.
29. Yang Y, Burrello J, Burrello A, *et al.* Classification of microadenomas in patients with primary aldosteronism by steroid profiling. *J Steroid Biochem Mol Biol*, 2019; **189**: 274–82.
30. Tezuka Y, Yamazaki Y, Kitada M, *et al.* 18-oxocortisol synthesis in aldosterone-producing adrenocortical adenoma and significance of KCNJ5 mutation status. *Hypertension*, 2019; **73**: 1283–90.
31. Muth A, Ragnarsson O, Johannsson G, Wangberg B. Systematic review of surgery and outcomes in patients with primary aldosteronism. *Br J Surg*, 2015; **102**: 307–17.
32. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Cardiometabolic outcomes and mortality in medically treated primary aldosteronism: a retrospective cohort study. *Lancet Diab Endocrinol*, 2018; **6**: 51–9.
33. Deinum J, Riksen NP, Lenders JW. Pharmacological treatment of aldosterone excess. *Pharmacol Ther*, 2015; **154**: 120–33.

34. Karashima S, Yoneda T, Kometani M, *et al.* Comparison of eplerenone and spironolactone for the treatment of primary aldosteronism. *Hypertension Res*, 2016; **39**: 133–7.
35. Lau JH, Sze WC, Reznek RH, *et al.* A Prospective Evaluation of Postural Stimulation Testing, Computed Tomography and Adrenal Vein Sampling in the Differential Diagnosis of Primary Aldosteronism. *Clin Endocrinol,* 2012 Feb; **76**(2): 182–8.
36. Omata K Satoh F, Morimoto R, *et al.* Cellular and genetic causes of idiopathic hyperaldosteronism. *Hypertension*, 2018; **72**: 874–80.
37. Wu VC, Chueh SJ, Chen L, *et al.*; TAIPAI Study Group. Risk of new-onset diabetes mellitus in primary aldosteronism: a population study over 5 years. *J Hypertens*, 2017; **35**: 1698–708.
38. Wu VC, Chang CH, Wang CY, *et al.* Risk of fracture in primary aldosteronism: a population-based cohort study. *J Bone Miner Res*, 2017; **32**: 743–52.
39. Beuschlein F, Reincke M, Arlt W. The impact of Connshing's syndrome—mild cortisol excess in primary aldosteronism drives diabetes risk. *J Hypertens*, 2017; **35**: 2548.
40. Adolf C, Köhler A, Franke A, *et al.* Cortisol excess in patients with primary aldosteronism impacts left ventricular hypertrophy. *J Clin Endocrinol Metab*, 2018; **103**: 4543–52.
41. Gerards J, Heinrich DA, Adolf C, *et al.* Impaired glucose metabolism in primary aldosteronism is associated with cortisol cosecretion. *J Clin Endocrinol Metab*, 2019; 104: 3192–202.
42. Heinrich DA, Adolf C, Holler F, *et al.* Adrenal insufficiency after unilateral adrenalectomy in primary aldosteronism—long-term outcome and clinical impact. *J Clin Endocrinol Metab*, 2019. pii: jc.2019-00996.
43. Allan CA, Kaltsas G, Perry L *et al.* Concurrent Secretion of Aldosterone and Cortisol From an Adrenal Adenoma - Value of MRI in Diagnosis. *Clin Endocrinol,* 2000 Dec; **53**(6): 749–53.
44. Riester A, Reincke M. Progress in primary aldosteronism: mineralocorticoid receptor antagonists and management of primary aldosteronism in pregnancy. *Eur J Endocrinol*, 2015; **172**: R23–30.
45. Liu SY, Chu CC, Tsui TK, *et al.* Aldosterone-producing adenoma in primary aldosteronism: CT-guided radiofrequency ablation-long-term results and recurrence rate. *Radiology*, 2016; **281**: 625–34.
46. Sarwar A, Brook OR, Vaidya A, *et al.* Clinical outcomes following percutaneous radiofrequency ablation of unilateral aldosterone-producing adenoma: comparison with adrenalectomy. *J Vasc Interv Radiol*, 2016; **27**: 961–7.

5.7

Cushing's Syndrome

John Newell-Price

Introduction and Historical Perspective

> *I would like to see the day when somebody would be appointed surgeon somewhere who had no hands, for the operative part is the least part of the work.*
>
> Harvey Cushing: Letter to Dr Henry Christian, 20 November 1911

Harvey Cushing described the first case of Cushing's syndrome with a severe phenotype in 1912. Since that time, investigation and management of Cushing's syndrome has remained a significant clinical challenge [1–3] and patients suspected of this diagnosis warrant referral to major centres.

Endogenous Cushing's syndrome is due the chronic, excessive, and inappropriate secretion of cortisol. When presentation is florid, diagnosis is usually straightforward, but in modern practice Cushing's syndrome is frequently and increasingly considered in mild cases in the absence of the classical signs in the context of osteoporosis, diabetes, hypertension as well as in the context of gynaecological and psychiatric clinics. Achieving a diagnosis can be difficult. Appropriate management of Cushing's syndrome is dependent on correctly identifying the cause of excess cortisol. Separating ACTH-independent causes (adrenal tumours) from ACTH-dependent causes (pituitary or ectopic secretion of ACTH) is usually simple. However, many ectopic sources are occult and the identification of the source of ACTH secretion may require meticulous and repeated investigations.

In most circumstances the mainstay of therapy remains surgery to either an ACTH-secreting tumour or directly to the adrenal glands, but additional treatment with cortisol-lowering or opposing drugs and tumour-directed therapy is often needed.

Aetiology, Genetics, Pathogenesis, and Pathology

Endogenous Cushing's syndrome is usually sporadic and divided into ACTH-dependent, and ACTH-independent causes (Table 5.7.1). Overall, ACTH-dependent causes account for approximately 80% of cases, and of these 80% are due to corticotroph pituitary adenomas (Cushing's disease) with an excess female predominance, and the remaining 20% due to the ectopic ACTH syndrome with male predominance [1]. Cushing's disease, the ectopic ACTH syndrome, and adrenal adenomas may also be found in the context of multiple endocrine neoplasia 1 (MEN 1).

Cushing's Disease

Most cases of Cushing's disease are due to corticotroph microadenomas, a few millimetres in diameter, only being larger than 1 cm (macroadenoma) in 6% of cases [1, 3, 4]. These tumours express the proopiomelanocortin gene (*POMC* 176830), the peptide product of which is subsequently cleaved to ACTH. POMC-processing is usually efficient in corticotroph microadenomas,

Table 5.7.1 Aetiology of Cushing's syndrome

Cause of Cushing's syndrome	F:M	%
ACTH-dependent[a]		
Cushing's disease	3.5:1[b]	70%
Ectopic ACTH syndrome	1:1	10%
Unknown source of ACTH[c]	5:1	5%
ACTH-independent		
Adrenal adenoma	4:1	10%
Adrenal carcinoma	1:1%	5%
Other causes (PBMAH; PPNAD; McCune–Albright)		<2%

[a] In women 9:1 ratio of Cushing's disease to ectopic ACTH.
[b] Male preponderance in children.
[c] Patients may ultimately prove to have Cushing's disease.
PPNAD, primary pigmented nodular adrenal disease; AIMAH, ACTH-independent massive adrenal hyperplasia.

but less so in macroadenomas, which may secrete relatively large amounts of unprocessed POMC. Some pituitary macroadenomas are 'silent corticotroph adenomas', and may present with tumour mass effects (e.g. optic chiasm compression) alone; on follow-up, initial absence of cushingoid features may progress to overt clinical Cushing's syndrome. Approximately 90% of tumours express the corticotropin-releasing hormone (CRH)-1 receptor, as evidenced by the release of ACTH in response to exogenously administered CRH. Tumours also express the vasopressin-3 receptor, and respond to vasopressin and desmopressin.

Tumours causing Cushing's disease are relatively resistant to the effects of glucocorticoids, but *POMC* expression and ACTH secretion are reduced by higher doses of dexamethasone in 80% of cases [2, 5]. This may be caused by 'miss-expression' of the 'bridging protein' Brg1 (which is important for glucocorticoid inhibitory feedback on *POMC* expression) found in corticotroph tumours, and may be one event determining tumourigenesis [6]. Corticotroph tumours also show overexpression of cyclin E, low expression of the cyclin-dependent inhibitor, p27, and a high Ki-67 expression, all indicative of a relatively high proliferative activity [7]. The excess number of reproductive-aged women with Cushing's disease, and the fact that there is a male preponderance in prepubertal cases [8] suggest a potential aetiological role for oestrogens. Recent data has emphasized the role of epidermal growth factor (EGF) and its receptor (EGFR), which increases *POMC* mRNA transcription, cell proliferation and ACTH secretion [9–11]. Usually, EGFR is degraded by ubiquitination, but somatic missense mutations in ubiquitin-specific protease 8 (USP8), present in approximately 40% of patients with Cushing's disease, inhibits deubiquitinating activity and so EGFR is re-cycled and signalling is enhanced [9–11]. Antagonists of EGFR may have therapeutic potential in Cushing's disease.

Ectopic ACTH Secretion

Well differentiated neuroendocrine tumours from any source may cause the ectopic ACTH syndrome, and show a molecular phenotype close to that of pituitary corticotroph tumours. In contrast, data in small cell lung cancer cells have shown that *POMC* is activated by transcription factors distinct from those in the pituitary, including E2F factors [12], which are able to bind the promoter when it is in an unmethylated state [13], suggesting a different pathogenesis.

Adrenal Disease

Approximately 50% of adrenal adenomas causing overt Cushing syndrome exhibit hotspot mutations at L205R of *PRKACA*, which encodes the catalytic subunit of cAMP-dependent protein kinase A (PKA), and drives cortisol secretion [14–17].

In primary bilateral macronodular adrenal hyperplasia (PBMAH), previously also termed AIMAH (for ACTH-independent macronodular hyperplasia), excess cortisol secretion may be associated with either ectopically-expressed receptors or increased eutopic receptor expression [18], and activation by ligands not usually associated with adrenal steroidogenesis: gastric inhibitory peptide (food-dependent Cushing's); vasopressin; interleukin-1; luteinizing hormone; and serotonin. Activation of receptors increasing intracellular cAMP is thought to cause hyperplasia over many years, and hence Cushing's syndrome. Recently, it has been demonstrated that steroidogenic cell clusters in PBMAH adrenals produce corticotropin, which stimulates cortisol production by the adrenal in an intracrine/paracrine fashion. Therefore, in PBMAH, cortisol excess seems to mediated by ACTH, albeit of adrenal rather than pituitary or extra-adrenal ectopic origin; therefore, PBMAH is not truly ACTH-independent. For this reason, the term PBMAH is now preferred over AIMAH. The nodules can vary in size and may be massive. Germline inactivating mutations of Armadillo repeat 5 (*ARMC5*) [19] have now been identified as a common cause of PBMAH. ARMC5 is a cytosolic protein and little is known about its function; in rodent models a more complex phenotype ensued including compromised T-cell proliferation and differentiation; in humans, ARMC5 has also been reported to be associated with an increased incidence of meningioma alongside Cushing's syndrome. Characteristically, ARMC5 causes a mild Cushing's phenotype with onset over years and decades, often resulting in misdiagnosis as metabolic syndrome; family screening can reveal cases much earlier.

Primary pigmented nodular adrenal disease (PPNAD) causes small ACTH-secreting nodules on the adrenal, often not visualized on imaging. PPNAD can be sporadic or part of the Carney's complex and most cases occur in late childhood or in young adults, often with a mild or cyclical presentation [20, 21]. Germ line mutations of the regulatory subunit R1A of PKA (*PRKAR1A*) are present in approximately 45% of patients with Carney's complex [22, 23] and as well as in sporadic PPNAD. Another causative gene locus maps to chromosome 2p16. Interestingly, these patients show a paradoxical increase in cortisol secretion in response to dexamethasone.

McCune–Albright syndrome is due to a postzygotic activating mutation in the *GNAS1* gene. The resulting tissue mosaicism results in a varied phenotype, and the disease may present in the first few weeks of life. These mutations lead to constitutive steroidogenesis in the affected adrenal nodules [24]. Mutations of *GNAS1* have also been found in cases of PBMAH.

Epidemiology

The true prevalence of Cushing's syndrome is difficult to quantify. Earlier data suggest an incidence from 0.7 to 2.4/million population per year depending on the population studied. Biochemical Cushing's syndrome with no clear clinical features has, however, been shown to be common. Incidental adrenal lesions found on computed tomography (CT) scans are now a very common clinical problem and approximately 1–3% of the population aged 70 or more will have evidence of low-grade hypercortisolaemia from such a lesion. In addition, Cushing's syndrome has been reported in some series to be found in 1–5% of obese patients with type 2 diabetes, and up to 10.8% of older patients with osteoporosis and vertebral fracture [25]. The difficulty here, however, is whether detection of mild Cushing's syndrome in these populations is of clinical value as the outcomes of small and uncontrolled intervention studies are mixed. These data indicate that formal intervention trials are needed before widespread screening in these populations can be recommended, and there is a need for clinical decision-making tools to allow stratification for intervention on an individualized basis.

Clinical Features of Cushing's Syndrome

Glucocorticoid receptors are present in virtually all cells, reflecting the diverse actions of cortisol, and hence the symptoms and signs

of hypercortisolaemia encompass all organ systems. Many of the symptoms associated with hypercortisolaemia are common and of little specificity, such as weight gain, lethargy, weakness, menstrual irregularities, loss of libido, hirsutism, acne, depression, and psychosis (Table 5.7.2). While each symptom itself may be mild, the presence of a greater number of features in any given patient increases the likelihood of Cushing's syndrome. The signs most useful in differentiating Cushing's syndrome include the presence of proximal myopathy, and easy bruising, purplish striae, thinness, and fragility of the skin [1, 2, 25]. The sign of proximal weakness is most easily demonstrated by asking the patient to stand from sitting position without the use hands; an initial backwards movement of the buttocks is present in early myopathy, while in more severe cases rising from a chair may not be possible.

Presentation differs between genders, with purple striae, muscle atrophy, osteoporosis, and kidney stones being more common in men [26]. Gonadal dysfunction is common in both sexes. The adverse effects of glucocorticoids on bone metabolism are evidenced by decreased bone mineral density, especially of the axial skeleton. Over 70% of patients with Cushing's syndrome may present with psychiatric symptoms ranging from anxiety to frank psychosis; if present, depression is often agitated in nature, and some degree of psychiatric disturbance often persists following remission of Cushing's syndrome [3]. Impairment in short-term memory and cognition is common and can persist for at least a year following treatment. Cortisol excess predisposes to hypertension and glucose intolerance.

Classically, the ectopic ACTH syndrome due to small cell lung cancer may have a rapid onset with severe features: profound weakness, myopathy, hyperpigmentation, diabetes mellitus, and hypokalaemic alkalosis, while there is often neither weight gain nor the classical cushingoid appearance. In contrast, the clinical phenotype and biochemical features of neuroendocrine tumours (of any tissue origin) may be indistinguishable from that of Cushing's disease, causing diagnostic difficulty.

Clinical and biochemical features may commonly vary in a 'cyclical fashion', causing diagnostic difficulty. Signs and symptoms fluctuate with circulating cortisol, such as facial plethora, myopathy, mood, blood pressure, and blood glucose, and all investigations may be normal when hypercortisolaemia is absent. Great care is needed to seek for evidence of 'cyclicity' in the clinical history.

Table 5.7.2 Clinical features of Cushing's syndrome

Feature	%
Obesity or weight gain	95
Facial plethora	90
Rounded face	90
Decreased libido	90
Thin skin	85
Decrease linear growth in children	70–80
Menstrual irregularity	80
Hypertension	75
Hirsutism	75
Depression/emotional lability	70
Easy bruising	65
Glucose intolerance	60
Weakness	60
Acne	50
Osteopenia or fracture	50
Nephrolithiasis	50

Clinical Investigation and Diagnostic Criteria

Who to Test

Most patients initially suspected of possibly having Cushing's syndrome will not have this condition. The complete assessment of a patient known to have some form of Cushing's syndrome is complex, expensive, and often stressful for the patient, who is usually already significantly ill emotionally, psychologically, and physically. Thus efficient screening procedures are needed to identify the minority who will need intensive and expensive investigation leading to an accurate and precise differential diagnosis [1, 25].

It is recommended that clinical judgement is used to select patients for testing, which should be considered in: (1) patients with features that are unusual for age, such as hypertension and osteoporosis; (2) those with multiple and progressive features, especially if these include the signs that most reliably distinguish Cushing's syndrome: the presence of thin skin in the young, easy bruising, proximal myopathy, and purple striae; (3) in children with increasing weight percentile and decreased linear growth; and (4) patients with adrenocortical lesions consistent with an adenoma found on CT scans performed for other reasons, so-called adrenal 'incidentaloma' [25] (for details see Chapter 5.3).

It is essential that a careful drug history is taken prior to any biochemical testing seeking to exclude exogenous sources of glucocorticoids that may be present in prescribed oral, rectal, inhaled, topical, or parenteral medication as well as in many 'over the counter' preparations, including skin creams, 'skin-whitening' agents, and various 'tonics' and herbal preparations.

Biochemical Assessment

The biochemical hallmark of Cushing's syndrome is inappropriate cortisol secretion not subject to the normal negative feedback effects of circulating glucocorticoids on the regulatory centres, hypothalamus, and pituitary. The tests are based on demonstration of excessive cortisol secretion, loss of its circadian rhythm, and the abnormal feedback regulation of the hypothalamic–pituitary–adrenal (HPA) axis (Figure 5.7.1).

In florid cases of Cushing's syndrome the diagnosis may be obvious, but biochemical confirmation is still needed. Investigation of Cushing's syndrome is a two-step process. Hypercortisolaemia *must* be confirmed before *then* the cause is identified. Failure to follow this approach will result in inappropriate treatment and management.

Step 1: Diagnosis of Hypercortisolaemia

Several tests are usually needed. Investigation should be performed when there is no acute concurrent illness, such as infection or heart failure, as these may cause false-positive results. The three main tests in use are: 24-h urinary free cortisol; 'low-dose'

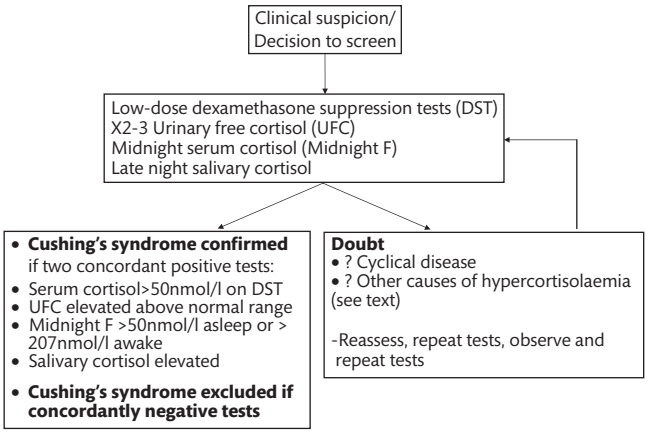


Figure 5.7.1 Biochemical diagnosis of Cushing's syndrome.

dexamethasone-suppression tests; and assessment of midnight plasma or late-night salivary cortisol. The best approach is to perform at least two different tests; if concordantly positive or negative, Cushing's syndrome is either likely or unlikely, respectively [1, 3, 25]. When there are discrepancies between tests, further evaluation and repeated testing is often required. Hypercortisolaemia is also found in some patients with depression, alcohol dependence, anorexia nervosa, and late pregnancy. However, in contrast to true endogenous Cushing's syndrome, the biochemistry of these pseudo-Cushing states improves when the underlying condition has resolved.

Urinary Free Cortisol

Urinary cortisol is a direct assessment of circulating free (biologically active) cortisol. Excess circulating cortisol saturates the binding proteins (cortisol-binding globulin, CBG) and is excreted in urine as free cortisol, and when collected for 24 h gives an integrated estimation of the level of hypercortisolaemia. A single measurement has low sensitivity, and three 24-h collections should be performed [2, 17]. There is a 50% day-to-day variability in urine-free cortisol (UFC) values in patients with Cushing's syndrome, reinforcing the need for multiple collections [27]. Values fourfold greater than the upper limit of normal are rare except in Cushing's syndrome. In contrast, if values are normal on repeated occasions, Cushing's syndrome is unlikely. Specificity is a common problem with antibody-based assays, but [1, 2, 25] high performance liquid chromatography (HPLC) and tandem mass spectrometry improves diagnostic accuracy, although substances such as digoxin and carbamazepine may produce peaks in the HPLC assay that give falsely high values [25]. Moreover, if there is renal impairment with a glomerular filtration rate (GFR) of less than 30.0 ml/min, or an incomplete collection, the urinary free cortisol may be falsely low [25]. Review of the collection volume and correction for creatinine concentration may be helpful in assessing whether the collection is complete. Use of urinary free cortisol is advised in the very rare situation of Cushing's syndrome being considered during pregnancy.

Low-Dose Dexamethasone-Suppression Tests

Two tests are in common use. In the overnight dexamethasone-suppression test, 1 mg of dexamethasone is administered at 23.00 hours and serum cortisol measured the next day at 08.00–09.00 hours. In the 48-h dexamethasone-suppression test, dexamethasone is administered at the dose of 0.5 mg every 6 hours for 2 days at 09.00, 15.00, 21.00, and 03.00 hours with measurements of serum cortisol at 09.00 hours at the start and end of the test. To exclude Cushing's syndrome the serum cortisol value should be less than 50 nmol/L following either test [1–3, 25]. The 48-h test, though more cumbersome, is more specific and with adequate regular instructions can easily be performed by outpatients. In both tests, caution needs to be exercised if there is potential malabsorption of dexamethasone or if patients are on drugs that increase hepatic clearance of dexamethasone, including carbamazepine, phenytoin, phenobarbital, or rifampicin. Simultaneous measurement of plasma dexamethasone may be useful to confirm compliance and adequate circulation levels, and a value of >3.3 nmol/L as measured by LC-MS/MS improves the accuracy of the test [28]. Patients taking oestrogen therapy, or who are pregnant, may have an increase in the cortisol-binding globulin. As commercial cortisol assays measure total cortisol, this may give a false-positive result on dexamethasone-suppression testing. Oral oestrogens need to be stopped for a period of 4–6 weeks so that cortisol-binding globulin may return to basal values. Even transdermal oestrogens may cause false-positive results, and tests should be repeated off transdermal oestrogens if positive results are obtained. In renal failure, suppression on dexamethasone testing is likely to exclude Cushing's syndrome. It is important to note that 3–8% of patients with proven Cushing's disease show suppression of serum cortisol to less than 50 nmol/L on either test. Thus, if clinical suspicion remains high, repeated tests and other investigations are indicated.

Late-Night Salivary Cortisol or Midnight Plasma Cortisol

The normal circadian rhythm of cortisol secretion is lost in patients with Cushing's syndrome. A single sleeping midnight plasma cortisol of less than 50 nmol/L effectively excludes Cushing's syndrome at the time of the test. This is one of the harder tests to perform as it requires hospitalization for at least 48 hours, and lack of intercurrent illness, but it can be of great utility to exclude Cushing's syndrome, especially when the patient is on drugs known to enhance metabolism of dexamethasone causing a false-positive on dexamethasone testing. Values above 50 nmol/L when asleep or above 207 nmol/L when awake are found in Cushing's syndrome, even in those who suppress on dexamethasone [29, 30]. An elevated midnight plasma cortisol does not provide additional information if clinical signs are florid and there is clear lack of suppression on dexamethasone testing.

Late-Night Salivary Cortisol

Salivary cortisol reflects free circulating cortisol and its ease of collection and stability at room temperature make it a highly suitable screening tool for outpatient assessment. The diagnostic ranges vary between reports due to the different assays and the comparison groups used to set cut-off points. The test has a sensitivity of between 95% and 98% [31, 32]. Comorbidities, such as poorly controlled diabetes and hypertension, as well as advancing age may cause false-positive results [33]. As the values of salivary cortisol are an order of magnitude lower than serum cortisol, it is essential that the performance of the local assay be known and that the appropriate cut-off point is utilized. The test is of particular use in the assessment of cyclical Cushing's syndrome, and in children.

Diagnostic Doubt

In cases of doubt the best option is to repeat the tests at a later date, or seek further opinion. The dexamethasone-suppressed CRH test, and the desmopressin test have been proposed as useful diagnostic tools but the dexamethasone-suppressed CRH test does not appear to be more accurate than the 48-h low-dose dexamethasone-suppression test [34–36].

Step 2: Establishing the Aetiology of Cushing's Syndrome: Differential Diagnosis

Once a diagnosis of Cushing's syndrome is established the next step is establish the cause. Investigation will vary depending upon the availability of the biochemical tests and imaging and expertise detailed next.

The first key procedure is to measure plasma ACTH. The plasma should be separated rapidly and stored at –40C to avoid degradation and a falsely low result. Levels consistently below 5 ng/L indicate ACTH-independent Cushing's syndrome and attention can be turned to imaging the adrenal with CT. Levels of ACTH persistently above 15 ng/L almost always reflect ACTH-dependent pathologies and require investigation, as detailed next. The values between these two need cautious interpretation as patients with Cushing's disease and adrenal pathologies may have intermediate values [1, 2]. A positive CRH test (see next) can identify an occasional patient with Cushing's disease with low baseline ACTH plasma levels.

Non-ACTH-Dependent Cushing's Syndrome

In established Cushing's syndrome, when plasma ACTH has been sampled and handled carefully, and levels are persistently undetectable, the cause is of adrenal origin. The next diagnostic procedure is to proceed to imaging with CT or MRI, which will most likely show an adrenocortical adenoma or carcinoma. If imaging is negative, the diagnosis may either be PPNAD, or due to surreptitious hydrocortisone absorption.

ACTH-Dependent Cushing's Syndrome

Localization of the source of ACTH secretion in ACTH-dependent Cushing's syndrome can constitute one of the most formidable challenges of clinical endocrinology. Neuroendocrine tumours may be clinically indistinguishable from Cushing's disease, and are frequently difficult to identify with imaging, especially if radiological (pituitary, thoracic, pancreatic) 'incidentalomas' complicate interpretation. As a result, biochemical evaluation rather than imaging is used to differentiate between pituitary and non-pituitary causes. In women with ACTH-dependent Cushing's syndrome, 9 out of 10 cases will be due to Cushing's disease. It is against this pre-test likelihood that the performance of any test needs to be judged. On occasion, despite all investigation, in some patients it may not be possible to locate the source of ACTH with confidence, and management of hypercortisolaemia may be needed without a precise diagnosis being reached.

Basal Testing: Plasma ACTH and Potassium

While very high levels of plasma ACTH may be seen in ectopic ACTH, the values frequently overlap those seen in Cushing's disease. High levels of cortisol of any aetiology may overwhelm the 11β-hydroxysteroid dehydrogenase type II enzyme in the kidney, allowing cortisol to act as a mineralocorticoid; approximately 70% of patients with ectopic ACTH syndrome due to carcinoid tumours have hypokalaemia, but it is also present in approximately 10% of patients with Cushing's disease with extremely high cortisol production [1, 2].

Dynamic Testing

The relative merits of each investigation will be discussed, but ultimately local experience of a given investigation, dependent on assays and radiological skill, will be an important determinant of the overall diagnostic success.

Dynamic Non-Invasive Tests

High-dose dexamethasone-suppression test: The high-dose dexamethasone-suppression test (2 mg given every 6 h for 48 h and serum cortisol measure at 09.00 h at the beginning and end, or a single 8 mg dose given at 23.00 h and serum cortisol measured the next day at 09.00 h) has been in widespread use for many years. The test relies upon the relative sensitivity of pituitary corticotroph adenomas to the effects of glucocorticoids, compared to the resistance exhibited by non-pituitary tumours. Approximately 80% of patients with Cushing's disease will demonstrate suppression of the serum cortisol to a value of less than 50% of the basal level [1, 2]. This is less than the pre-test likelihood of Cushing's disease and, thus, by itself the high-dose dexamethasone-suppression test has little diagnostic utility. Moreover, when utilizing the 48-h low-dose dexamethasone-suppression test, if there has already been the demonstration of suppression of serum cortisol by more than 30%, there is no further advantage to utilizing the high-dose dexamethasone-suppression test. Therefore, continued routine use of the high-dose dexamethasone-suppression test can no longer be recommended except when bilateral inferior petrosal sinus sampling (BIPSS) is not available. The positive predictive value for Cushing's disease is, however, high if there is a positive response (suppression of serum cortisol <50%) **and** a positive response on CRH testing (see next and Figure 5.7.2), but the negative predictive value for exclusion of Cushing's disease when both tests are negative, is low.

The corticotropin-releasing hormone (CRH) test: CRH was identified and sequenced in 1981, and is available for clinical practice as either the ovine (oCRH) or human sequence (hCRH) which differ by seven amino acid residues. oCRH has a longer duration of action and is the form available in North America, while the experience of hCRH dominates in Europe. In practice, the value of the test is the same [37, 38]. CRH is well tolerated, with side effects from systemic administration consisting of mild, short-lived facial flushing, a sensation of a metallic taste, and a transient sinus tachycardia. A single intravenous bolus of CRH (100 μg or 1 μg/kg) administered at 09.00 hours stimulates pituitary ACTH and cortisol release in healthy individuals, excessively in patients with pituitary-dependent Cushing's syndrome, but generally not in patients with ectopic ACTH secretion or adrenal tumours. A >35% rise above in mean plasma ACTH at 15 and 30 minutes when using oCRH [38], or a >14% rise in mean serum cortisol measured at these time points when using hCRH [37] give the best discrimination between pituitary and ectopic disease.

Desmopressin testing: Since the vasopressin-3 receptor is expressed in pituitary and many ectopic tumours secreting ACTH, the

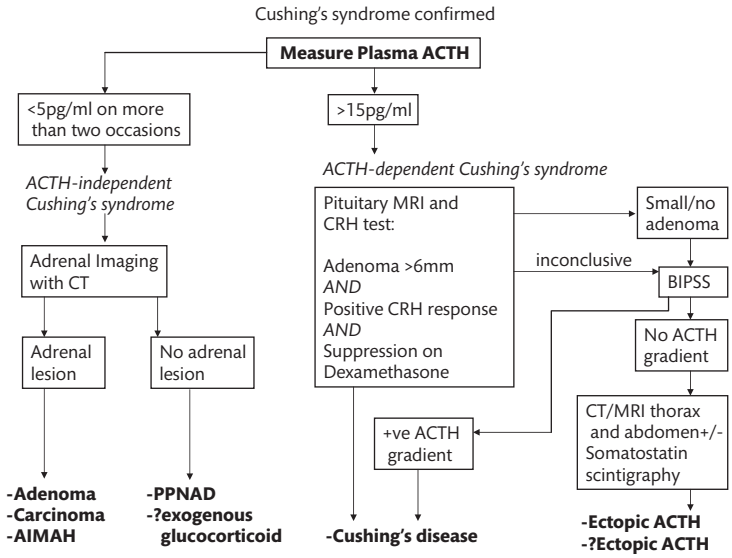


Figure 5.7.2 Diagnosis of cause of Cushing's syndrome. BIPSS, bilateral inferior petrosal sinus sampling; PPNAD, primary pigmented nodular adrenal disease; PBMAH, ACTH-independent macronodular hyperplasia.

desmopressin test is of limited utility in the differential diagnosis of ACTH-dependent Cushing's syndrome.

Dynamic Invasive Testing

Bilateral inferior petrosal sinus sampling (BIPSS): If a patient has ACTH-dependent Cushing's syndrome, with responses **both** on dexamethasone-suppression *and* CRH testing suggesting pituitary disease, and the pituitary MRI scan shows an isolated lesion of 6 mm or more, most will regard the diagnosis of Cushing's disease to have been made. A major problem is that up to 40% of patients with proven Cushing's disease have normal pituitary MRI scans [3]. In these cases, sampling of the gradient of ACTH from the pituitary to the periphery is the most reliable means for discriminating between pituitary and non-pituitary sources of ACTH, and is strongly recommended for most cases of ACTH-dependent Cushing's syndrome. Since the pituitary effluent drains via the cavernous sinuses to the petrosal sinuses and then jugular bulb, there is a gradient of the value of plasma ACTH compared to the simultaneous peripheral sample when there is a central source of ACTH. BIPSS is a highly skilled and invasive technique, requiring placement of catheters in both inferior petrosal sinuses. Plasma ACTH levels in peripheral blood fluctuate spontaneously by up to a factor of two, and hence a central to peripheral ratio greater than 2 is required to have confidence that ACTH secretion is pituitary and not the result of random variation from either a pituitary or ectopic source. Via a needle in a femoral vein, two catheters are passed up the inferior and superior vena cava into the neck. One each is then placed in a jugular vein and advanced into the inferior petrosal sinus. Catheter position and venous anatomy require confirmation by venography, as non-uniform drainage is not uncommon. The diagnostic accuracy of the test is improved with the administration of CRH. A basal central: peripheral ratio of more than 2:1 or a CRH-stimulated ratio of more than 3:1 is consistent with Cushing's disease. The combined data for many series indicate a sensitivity and a specificity of 94% [39]. Where CRH is unobtainable or too costly, desmopressin offers a reasonable alternative, but few patients with ectopic ACTH secretion have been studied in this way.

False-positive results may be caused by inadequate suppression of the normal corticotrophs; the duration and amount of hypercortisolism should be assessed prior to the test. For this reason pretreatment with cortisol-lowering agents prior to BIPSS is to be strongly discouraged as this increases the likelihood of a false-positive response in a patient with ectopic disease. ACTH secretion will always be localized to the pituitary in normal individuals and hence it is crucial to establish that all patients truly have ACTH-dependent Cushing's syndrome before undertaking this procedure. A false-negative result may be found in patients with cyclical Cushing's disease if the procedure is undertaken when the disease is inactive, and thus it is imperative to measure serum cortisol in the 24 hours prior to sampling to establish activity.

In adults, BIPSS is only 70% accurate for lateralization of the source of ACTH within the pituitary gland [1, 2], but in children it may have greater accuracy for this purpose than MRI [40]. False negatives may also occur if there is atrophic or plexiform venous drainage of the petrosal sinuses and this possibility should be checked for by venography at the time of BIPSS.

Imaging

Adrenal

CT gives the best resolution of adrenal anatomy (see Figure 5.7.3). Cortisol-secreting adenomas are typically less than 4 cm in diameter and associated with atrophy of the unaffected adrenal tissue, and the contralateral adrenal gland. Malignancy is more common with increasing tumour size, attenuation values of >10 on unenhanced CT and radiological evidence of vascular invasion, or cosecretion of sex steroids. In ACTH-dependent Cushing's syndrome, sustained stimulation of the adrenals may lead to the occurrence of adrenal nodules and adrenal hyperplasia that is not always symmetrical, causing diagnostic confusion with a unilateral primary adrenal cause if the biochemistry is not strictly assessed; in 30% of Cushing's disease the adrenal glands appear normal, while in ectopic ACTH the adrenals are virtually always homogeneously enlarged [41]. A recent report showed evidence for somatic mutations

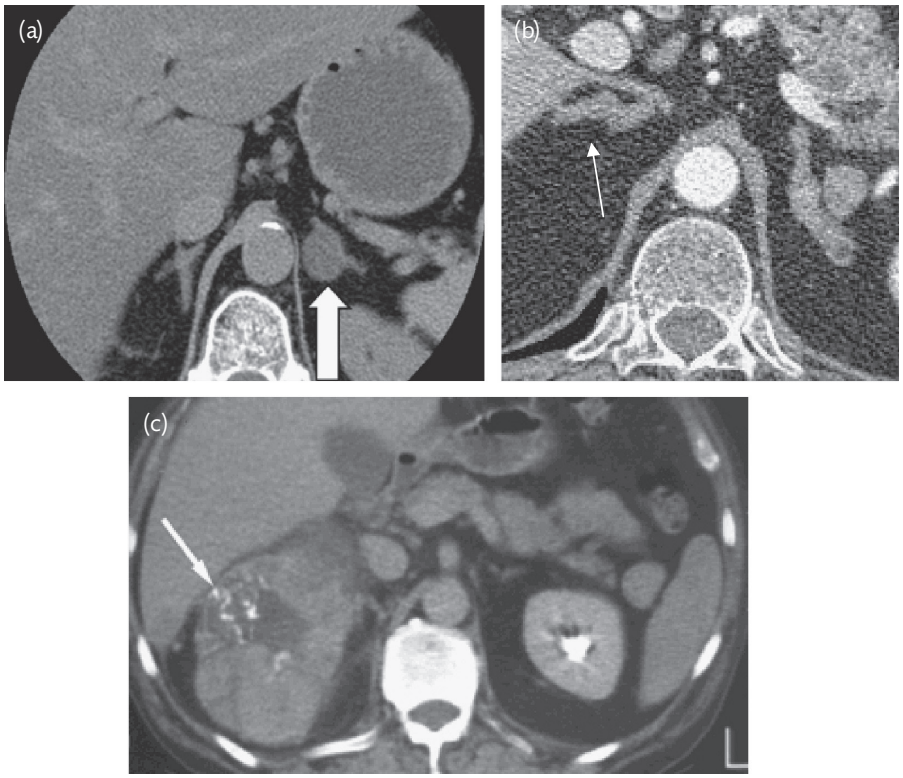


Figure 5.7.3 Adrenal imaging in Cushing's syndrome. (a) Non-contrast (= unenhanced) CT scan of left adrenal adenoma with low attenuation in the tumour area; (b) bilateral adrenal hyperplasia in ACTH-dependent Cushing's syndrome; (c) right-sided adrenocortical carcinoma.

in the PRKACA gene in nodular adrenal hyperplasia arising in two patients with long-standing adrenal stimulation due to long-standing Cushing's disease, demonstrating transition from pituitary to adrenal Cushing's.

Pituitary

Up to 40% of corticotroph adenomas causing Cushing's disease in adults are not visible on MRI scanning [3]. Those that are visible usually fail to enhance following gadolinium on T1-weighted imaging. The use of dynamic MRI, with the administration of intravenous contrast media and rapid sequence acquisition following this, does not improve the overall diagnostic rate. There is also a 10% rate of pituitary incidentalomas in the normal population [42], emphasizing the need for careful biochemical discrimination of pituitary from non-pituitary sources of ACTH. In the absence of a pituitary macroadenoma, an abnormal MRI alone is **not** conclusive evidence in favour of Cushing's disease.

Imaging in the Ectopic ACTH Syndrome

Small cell lung cancer may be obvious, but in most cases thoracic and abdominal imaging by high resolution CT with 1 mm 'cuts' with imaging early after intravenous contrast administration is needed to identify small neuroendocrine tumours, which may be extremely hard to localize, as a source of ACTH [43, 44]. Other than small cell lung cancer, bronchial neuroendocrine tumours are the most common sources of ectopic ACTH secretion, and are usually less than 1 cm in diameter. However, small, typically enhancing neuroendocrine tumours may be confused with pulmonary vascular shadows, but these tumours usually have high signal intensity on T_2-weighted and short-inversion-time inversion recovery on MRI. ACTH-secreting thymic neuroendocrine tumours are generally larger than 2 cm and readily visualized by CT. Although ectopic ACTH-secreting tumours often express somatostatin receptors and can be seen on radiolabelled octreotide scintigraphy, they are also almost always identified by CT. [^{68}Ga]-DOTATATE PET-CT is more sensitive and may detect tumours only a few millimetres in diameter, although even then the lesion is usually seen on CT The excess circulating cortisol in Cushing's syndrome may downregulate the sst2 receptor on the causative tumour and thus a period of medical therapy to lower cortisol, or offset the actions of cortisol, may result in a positive scan that had previously shown no uptake [45].

Treatment and Prognosis

The Endocrine Society has produced evidence-based clinical practice guidelines for the treatment of Cushing syndrome [46]. To deliver high-quality treatment to patients with Cushing's syndrome requires a team that includes specialized surgeons and physicians, radiologists, cytologists, histopathologists, and radiotherapists. The sustained hypercortisolaemia of Cushing's syndrome, of any aetiology, suppresses ACTH secretion from healthy corticotrophs and, hence, hypoadrenalism will be the consequence of complete excision of any tumour causing Cushing's syndrome, be it adrenal, pituitary, or an ectopic source of ACTH secretion, and this may be prolonged.

Management is aimed at lowering cortisol levels, removing tumour tissue, and, in the case of Cushing's disease, causing the least

harm to remaining pituitary function. Some centres use medical therapy to control hypercortisolaemia prior to surgery, and this makes intuitive sense, but there are no published data that this affects overall outcome. Hypertension and diabetes require treatment on their own merits, but both tend to improve, often dramatically, with control of hypercortisolaemia. Severe hypokalaemia secondary to Cushing's syndrome is extremely difficult to treat unless hypercortisolaemia is corrected. If it persists, or while cortisol control is being affected, triamterene or high-dose amiloride is helpful.

Medical Therapies

A potential side effect of all drugs used to control cortisol secretion is hypoadrenalism, particularly in patients with cyclical Cushing's syndrome, and hence all patients require close monitoring [47]. Although urinary free cortisol is easy to use for monitoring therapy, it has major limitations in that hypoadrenalism may be difficult to establish accurately, and failure to ensure a complete 24-h collection will result in spuriously low results. Calculation of the mean of five serum cortisol measurements across a single day, though cumbersome offers both the ability to identify transient hypoadrenalism and allows accurate monitoring, since a mean serum cortisol between 150 and 300 nmol/L has been demonstrated to equate to a normal cortisol production rate [48]. Measurement of salivary cortisol may offer and alternative at greater convenience [49], but this has not been formally assessed in clinical practice.

Adrenal Steroidogenesis Inhibitors

Hypercortisolaemia of any cause may be controlled by drugs that act on the adrenal glands to reversibly inhibit cortisol secretion [50]. Metyrapone and ketoconazole are currently licensed for this purpose. These drugs are not curative, and cortisol oversecretion will recur when they are discontinued. Steroidogenesis inhibitors are usually used to regulate cortisol secretion in very specific circumstances, namely: in preparation for surgery, in patients not cured by surgery, while waiting for radiotherapy to be effective, after chemotherapy, and to correct acute severe physical or psychiatric consequences of hypercortisolaemia. Cortisol-induced psychosis usually responds rapidly to lowering of circulating cortisol levels.

Metyrapone inhibits cortisol secretion by blocking the final step in cortisol synthesis, namely conversion from 11-deoxycortisol by the cytochrome P450 enzyme 11-hydroxylase (CYP11B1). Serum cortisol levels fall within 2 h of instigating therapy, but the effect is short-lived and metyrapone requires to be taken three times daily. Because 11-deoxycortisol levels are increased, it is best if LC-MS/MS assays are used for monitoring due the potential cross-reactivity of the accumulating 11-deoxycortisol with cortisol in immunoassays [46], where even a small amount of cross-reactivity of 11-deoxycortisol will produce artificially elevated apparent serum cortisol, potentially masking hypoadrenalism. The drug is usually given in doses ranging from 250 mg twice daily to 1.5 g every 6 hours, with lower doses for adrenal adenoma and higher for ectopic ACTH and overall, in clinical practice it is effective in around 50% of cases [51]. Nausea is a side effect that can be helped (if it is not caused by adrenal insufficiency) by giving the drug with milk.

Hypoadrenalism is the major unwanted effect of metyrapone and can occur for several reasons: overtreatment, inability to mount a cortisol response to intercurrent infection, and cyclical Cushing's syndrome, but this issue is circumvented by using an assay based on mass spectrometry [52]. Hirsutism and acne, if present in women patients before treatment, may worsen due to the accumulation of adrenal androgenic precursor steroids secondary to the blockade of cortisol synthesis.

Ketoconazole is an imidazole that was used as an antifungal agent, but because it blocks a variety of steroidogenic cytochrome P450–dependent enzymes it thus lowers plasma cortisol levels. In contrast to metyrapone, adrenal androgen concentrations fall with treatment. Treatment is usually initiated with 200 mg three times per day and adjusted depending on serum cortisol concentrations; with between 400 and 1200 mg/day required to normalize cortisol secretion rates in patients with Cushing's disease, and an acidic stomach is needed for absorption; in clinical practice the drug is effective in around 50% of patients [53]. An additional desirable characteristic is that ketoconazole lowers serum cholesterol concentrations, which are characteristically raised in Cushing's syndrome.

Ketoconazole is of slower onset of action than metyrapone, and dose adjustments should only be made every 2 to 3 weeks, although in patients with adrenal adenomas responsiveness is more rapid and hypoadrenalism has occurred within 24 h. Ketoconazole consistently induces a reversible rise in liver transaminase and γ-glutamyltransferase levels and increases by 2–3-fold are expected and should not preclude continuation of treatment, but rarely fulminant hepatic failure has been seen. Liver function must be monitored on initiation of treatment and closely thereafter [53]. Ketoconazole is teratogenic to male fetuses and is contraindicated in pregnancy.

Ketoconazole and metyrapone may be given in combination, allowing doses to be used that are lower than required as monotherapy, with ketoconazole lowering androgen levels and thereby greatly increases the acceptability of metyrapone in women. Combination therapy is also useful in cases of severe Cushing's [54].

Etomidate is an imidazole, and its principal clinical use is as an anaesthetic agent, however, already at low, subhypnotic doses intravenous etomidate is a potent inhibitor of cortisol secretion. The use of intravenous etomidate at subhypnotic doses is an option when oral adrenolytic therapy is not possible, also in the context of palliation (e.g. for metastatic cortisol-producing adrenocortical carcinoma). However, this usually requires ICU monitoring, though in some palliative situations cases of etomidate treatment via subcutaneous pump have been reported. Doses between 1.2 and 2.5 mg/h lower serum cortisol, sometimes to undetectable levels, when the patient needs to be maintained on a 'block and replace' regimen with the concomitant use of intravenous hydrocortisone (1–2 mg/h) [55, 56].

High-dose *o,p*'DDD (*ortho,para*' dichlorodiphenyl dichloroethane, mitotane) has been used widely for the treatment of cortisol excess due to inoperable adrenocortical carcinoma, but when given at a lower dose is effective in controlling cortisol secretion in Cushing's syndrome due to benign causes. It has a direct adrenolytic action, destroying adrenocortical cells, but also blocks cortisol synthesis by inhibiting 11β-hydroxylation and cholesterol side-chain cleavage. It is of slow onset of action; first effects are usually observed within 2 weeks but changes in dose usually require 6 weeks or more to be fully effective. Use of mitotane in adrenocortical cancer is addressed in Chapter 5.4. Low-dose treatment with mitotane for benign Cushing's syndrome, with a starting dose of 0.5 to 1 g/day, with gradual dose titration is well

tolerated with rare gastrointestinal upset and few neurological side effects, and is used more frequently for this purpose in mainland Europe [57]. Currently, it is mainly used for Cushing's disease only when metyrapone and ketoconazole cannot be used effectively. The major limitation of treatment is that it consistently causes hypercholesterolaemia, but if mitotane therapy is necessary, then the hypercholesterolaemia can be reversed by the use of a statin or ketoconazole.

Osilodrostat is a newer inhibitor of 11β-hydroxylase (CYP11B1) with greater potency and longer duration of action compared to metyrapone, that shows promise for the treatment of Cushing's disease. Phase 2 data indicates control, as assessed by UFC, in 80% of patients [58]. A large multinational phase 3 study is complete, but is yet to be formally reported.

Glucocorticoid Receptor Blockade

RU 486 (mifepristone) is a potent glucocorticoid receptor antagonist (GRA) that blocks cortisol action and reverses the consequences of hypercortisolaemia. It is licensed in the US for the treatment of hyperglycaemia associated with active endogenous Cushing's syndrome [59]. Efficacy of action is determined by clinical assessment because levels of cortisol remain high, and this may result in hypokalaemia through the action of cortisol binding to the mineralocorticoid receptor [59]. Antagonism of the progesterone receptor can cause endometrial hyperplasia. Newer GRAs are in development that lack progesterone activity.

Therapies to Lower ACTH and Thus Cortisol in Cushing's Disease

Somatostatin analogues such as octreotide and lanreotide that bind predominantly somatostatin receptor (sstr) subtype 2, are generally ineffective in Cushing's disease. However, the multireceptor somatostatin analogue, pasireotide, which demonstrates high-affinity binding to sstr subtypes 1, 2, 3, and 5, normalizes urinary free cortisol in 17–40% of patients with Cushing's disease depending on the severity of disease, with hyperglycaemia being a common side effect [60, 61].

Cabergoline has been assessed in Cushing's disease, and lowers cortisol and hence ACTH and in the recent largest series 20–25% of patients achieved good control of UFC using modest doses (1 mg per week) [62].

Surgery

Transsphenoidal Surgery

Transsphenoidal selective microadenectomy, by an experienced pituitary surgeon, is the treatment of choice for Cushing's disease, as it offers the prospect of a dramatic, rapid, and long-lasting cure without other hormonal deficiency (Figure 5.7.4a, b) [46, 63].

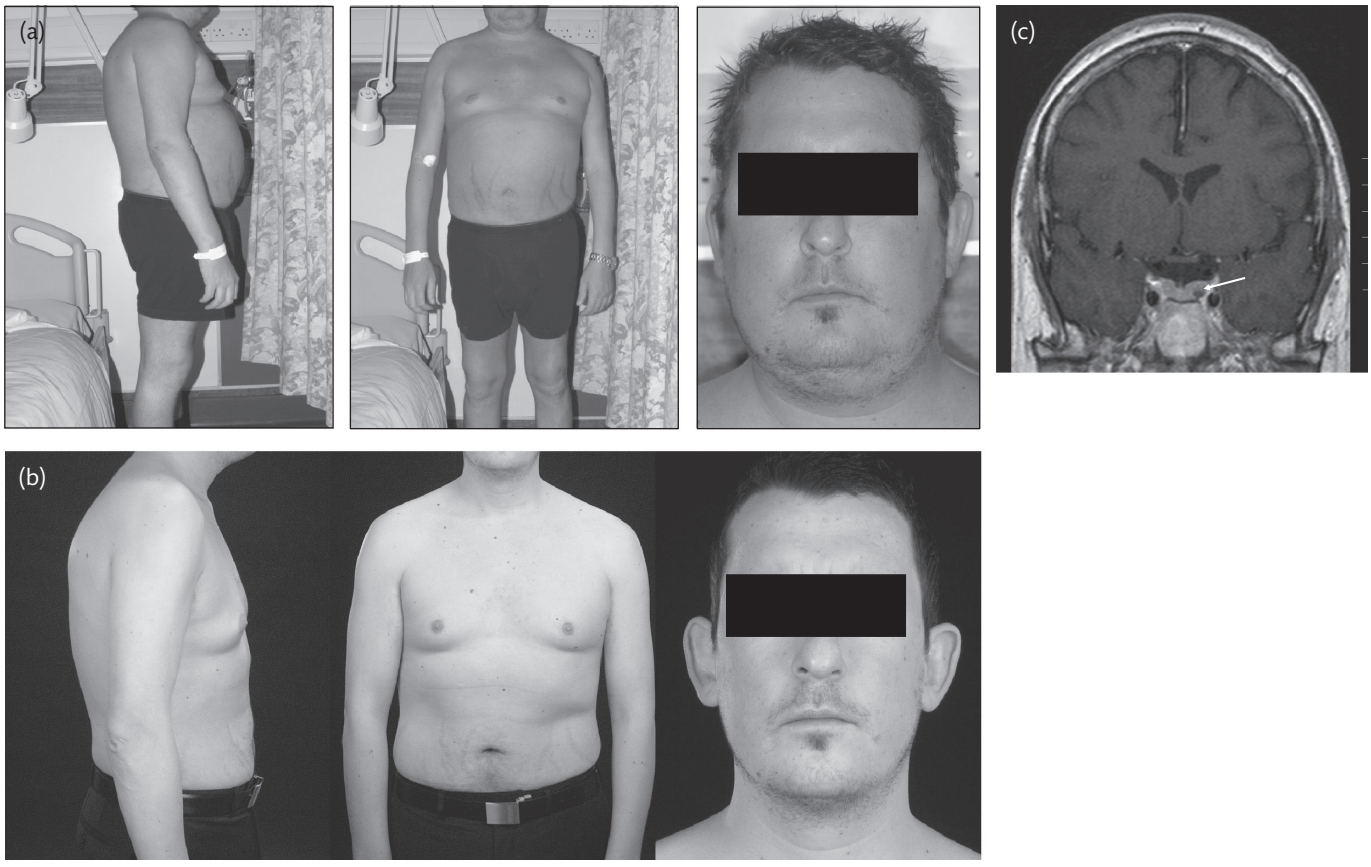


Figure 5.7.4 Resolution of clinical features following selective trans-sphenoidal microadenomectomy. (a) A 33-year-old man with florid Cushing's syndrome; note truncal obesity, striae, proximal muscle wasting, and facial plethora. (b) Dramatic resolution of clinical features 4 months after selective removal of ACTH-secreting microadenoma. (c) T_1-weighted gadolinium-enhanced MRI scan from the same patient showing 2.3 mm pituitary microadenoma causing Cushing's disease. Note non-enhancement (arrow). Patient underwent bilateral inferior petrosal sinus sampling to confirm pituitary source of ACTH. (See also Plate 27.)

In most cases, control of tumour volume is not a priority as the majority have either microadenomas (Figure 5.7.4c) or no visible tumour on MRI. Numerous series have reported the results and long-term follow-up following transsphenoidal surgery for Cushing's disease. Taking all series in the world literature together, the initial remission rate is 70% to 90% for microadenomas and 50% for macroadenomas but with a relapse rate of up to 20% when followed for many years, emphasizing the need for lifelong follow-up (Figure 5.7.5) [46, 63]. Rates for postoperative hypopituitarism and permanent diabetes insipidus depend on how aggressive the surgeon was in removing pituitary tissue.

It is likely that the variations in outcomes reflect surgical skill as well as the controversy regarding the characterization of remission or continuing disease in the postoperative period. Overall, with careful and prolonged follow-up (10 years) the long-term remission rate is approximately 60%; series suggesting rates higher than this usually either have shorter follow-up or less stringent criteria for remission. Patients who are hypocortisolaemic (low 09.00 h serum cortisol) in the immediate postoperative period require glucocorticoid therapy until the HPA axis recovers, usually 6–18 months postoperatively. While long-term remission is most likely when postoperative serum cortisol is low (<50 nmol/L), there is no threshold value that fully excludes possible recurrence [46, 63]. Care needs to be taken in the interpretation of postoperative serum cortisol in those patients who have received high-dose perioperative glucocorticoids, as these may suppress the level of cortisol in any remaining corticotroph tumour cells, with the patient appearing to be in remission, but then for the tumour cells to grow slowly and relapse appear years later. Similarly, suppression of normal serum cortisol levels on dexamethasone testing in the postoperative period is a poor indicator of long-term remission. Levels of basal 09.00 h postoperative serum cortisol of 100–200 nmol/L do not necessarily indicate failure of surgery, as some patients may remain in long-term remission [46]. On the other hand levels above 200 nmol/L will almost always indicate failure of surgery, although in some patients levels fall to very low levels 4–6 weeks after surgery Prompt postoperative assessment of the HPA axis is important, as in patients in whom hypercortisolaemia persists after an initial operation, repeat surgery, within 10 days, will allow remission in a further 50% of patients having a second operation. Care is needed in assessment of patients who have been taking cortisol-lowering therapy in the preoperative period as this may 'derepress' the suppressed normal corticotrophs and confound interpretation [64]. There is no agreement as to whether the presence or absence of a microadenoma on MRI makes remission more likely.

Complications of pituitary surgery include cerebrospinal fluid leakage (less than 5%) or meningitis (under 2%), but are unusual in experienced hands. Hypopituitarism may occur, but successful microadenectomy will leave pituitary function intact in more than 50%. Pituitary function needs to be tested in full, pre-, and postoperatively. The importance of preserving pituitary function has to be balanced against fitness for surgery and the consequences of deficiency, such as future fertility plans. It is important to note that functional deficiencies of growth hormone secondary to hypercortisolaemia may remain for 2 years after achieving remission by surgery. Thus the frail elderly patient, in whom a second operation would not be possible, might need an attempted total hypophysectomy at first operation, whereas in a fit young patient the tumour may be treated by a more limited procedure to attempt selective removal of the apparent local microadenoma, on the understanding that a second operation may be necessary if cure does not follow the first attempt, with almost inevitable hypopituitarism afterwards.

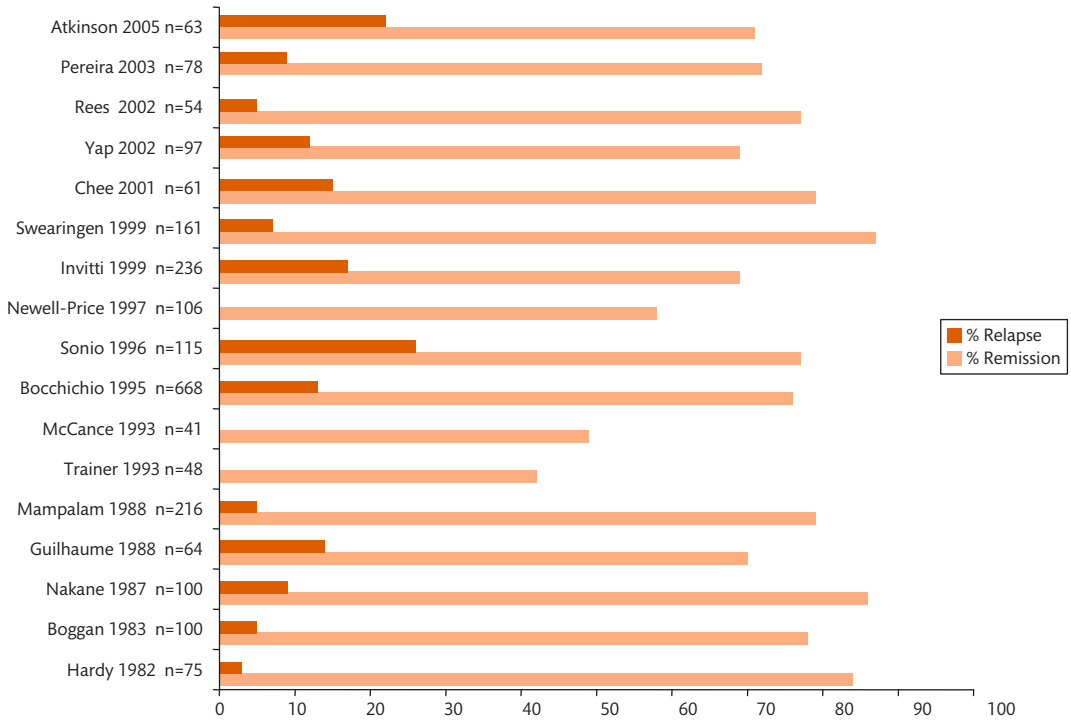


Figure 5.7.5 Modified from long-term outcome of transsphenoidal surgery for Cushing's disease. Initial remission rates in grey, relapse in black. Note that the lower initial remission rates are frequently associated with less relapse on follow-up.

Adrenal Surgery

For patients with an adrenocortical adenoma causing Cushing's syndrome the treatment of choice is a laparoscopic adrenalectomy by an experienced surgeon, as this is a safe and well-tolerated procedure [46]. Higher volume surgeons have better outcomes with fewer complications [65] (for details on adrenal surgery see Chapter 5.2).

The adrenal contralateral to a cortisol-secreting adrenal tumour will be atrophic, and glucocorticoid, but not mineralocorticoid, replacement therapy may be required for months or sometimes years, in rare cases permanently. In any cause of ACTH-dependent Cushing's syndrome, total bilateral adrenalectomy induces a rapid resolution of the clinical features. Following bilateral surgery, patients require lifelong treatment with glucocorticoid and mineralocorticoid replacement. With the low morbidity associated with laparoscopic adrenal surgery, this approach is being considered more frequently, and possibly even as primary therapy in some patients with Cushing's disease, especially when disease is severe or because of patient preference, and also in ectopic Cushing's due to an occult tumour; long-term outcomes are good [66, 67]. A major concern following bilateral adrenalectomy in patients with Cushing's disease is the development of Nelson's syndrome, a locally aggressive pituitary tumour that secretes high levels of ACTH, resulting in pigmentation. It remains controversial as to whether the tumour progression is a result of the lack of cortisol feedback following adrenalectomy, or whether the progression reflects corticotroph tumours that were always programmed to behave in an aggressive manner [68, 69]. If no tumour is visible on pituitary MRI at the time of adrenalectomy the likelihood of Nelson's syndrome is much less. Monitoring is by MRI and measurement of plasma ACTH. The tumour itself may be treated with further surgery or radiotherapy. Some advocate pituitary radiotherapy at the time of adrenalectomy to reduce the risk of this syndrome, but others have not confirmed this [69]. In contrast to adrenal adenomas most adrenal carcinomas have a very poor prognosis.

Fractionated External Pituitary Radiotherapy

Conventional external-beam radiotherapy has been available for 40 years, with large amounts of data demonstrating it to be safe and effective at controlling tumour growth and hormone secretion [46]. In Cushing's disease, its use is reserved for patients not cured by surgery, those in whom surgery is deemed inappropriate, and in the treatment of Nelson's syndrome. While waiting for the effect of radiotherapy to happen, which appears to be quicker in children, patients will usually require continued treatment with cortisol-lowering drugs, and regular biochemical assessment. Long-term hypopituitarism is likely in most cases, and fewer patients undergo radiotherapy than in previous decades [46].

Stereotactic Radiosurgery

Stereotactic radiotherapy, the most widely used variety being the 'Gamma Knife', is a means of delivering a high dose of radiation to a small volume in a single session without surrounding tissues being exposed to significant radiation. The main advantages of stereotactic over conventional radiotherapy are the convenience of a single session, more rapid correction of hypersecretion, and preservation of the function of surrounding healthy pituitary tissue. Rare, larger tumours are less easily treated, and the dose to the optic chiasm limited to less than 6–8 Gy. Gamma knife radiosurgery is effective [70, 71] but there is a relapse rate that is not apparent after remission following fractionated radiotherapy. It is likely that this outcome reflects case selection. In some circumstances Gamma Knife radiotherapy can be extremely effective, even as primary therapy, and may be more rapid in onset and in efficacy. This depends on absolute confidence in diagnosis, and an anatomically favourable lesion, especially if not approachable by surgery (Figure 5.7.6).

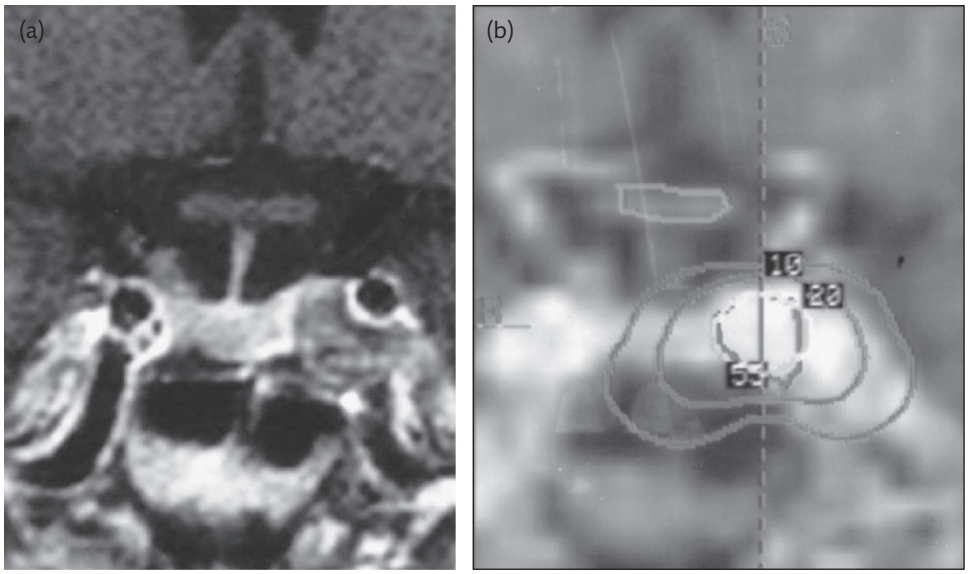


Figure 5.7.6 Gamma knife stereotactic radiosurgery for Cushing's disease. Figure shows diagnostic and planning MRI images of a patient with severe Cushing's disease treated by Gamma Knife radiosurgery as the primary and only definitive therapy at the author's institution, and who remains in remission with no pituitary deficit 10 years later. (a) Pretreatment the laterally placed tumour (arrow) was inaccessible to surgical approaches. (b) Tumour targeting with Gamma Knife 50% isodose to the tumour margin is shown; note the margin of safety from the 10% isodose to optic chiasm (outlined). (See also Plate 28.)

Prognosis

Uncontrolled and severe Cushing's syndrome has a 5-year survival of just 50%, with death due mainly to vascular and infective complications. Despite modern management, cardiovascular risk persists for many years after an apparent remission, but if remission is achieved at the first pituitary surgery the standardized mortality rate approaches the normal population, emphasizing the need for expert surgeons [72, 73].

Patients should be warned that they will often feel worse on correction of the hypercortisolism, with skin desquamation, profound lethargy, steroid-withdrawal arthropathy, and mood changes that may take several weeks or months to resolve [46]. These features can be ameliorated by a short-term increase in glucocorticoid replacement therapy. Patients are invariably growth hormone (GH) deficient, and GH replacement therapy may produce clinical benefit. Following pituitary surgery for a microadenoma the HPA axis is often suppressed for 9–12 months, with gradual recovery as evidenced by measuring 09.00 h serum cortisol every 3 months; once this is greater than 200 nmol/L it is reasonable to perform a synacthen test to assess the HPA axis. Once glucocorticoids are no longer needed patients need clinical and biochemical follow-up and surveillance for potential relapse. Even in remission, quality of life remains impaired and patients require careful counselling about this [74, 75].

Childhood Cushing's Syndrome

Cushing's syndrome should be considered in any child with obesity in combination with short stature, as children with simple obesity tend to be tall whereas hypercortisolaemia stunts growth. The elevated circulating androgens seen in Cushing's syndrome, particularly in adrenal adenomas, can result in apparent puberty and virilization without gonadal enlargement (pseudoprecocious puberty). Contrary to earlier reports, bone age is normal in 80% of children with Cushing's syndrome as although androgens accelerate bone maturation, with a consequent loss of linear growth potential, hypercortisolaemia appears to delay maturation.

The specific features that may be present on examination include inappropriate axillary or pubic hair, penile enlargement, scrotal pigmentation, and clitoral hypertrophy. Acne vulgaris can develop in children of any age. These features of hyperandrogenism are likely to be more pronounced with androgen-secreting adrenal tumours, but also occur with ACTH-dependent Cushing's syndrome. Primary or secondary amenorrhoea may be a feature in girls. Blood pressure is often mildly or moderately elevated, but diabetes mellitus is very unusual. School performance can suffer and psychiatric and emotional symptoms may occur.

Aetiology

The distribution of causes in childhood is different from that encountered in adults; there is a male predominance in Cushing's disease [8]. Fetal- and neonatal-onset Cushing's syndrome have been described, but the diagnosis remains exceptionally rare until approximately 8 years of age. Under 2 years of age, adrenal carcinoma accounts for 80% of cases of Cushing's syndrome, of which 80% occur in females [76].

Investigation in Childhood

Cushing's syndrome presenting in childhood is rare. Any child with weight gain and growth failure should be investigated but, as described earlier, the presenting symptoms may vary. The investigative algorithm is as described for adults, and experience shows that it is not necessary to alter the dose of dexamethasone or CRH used in

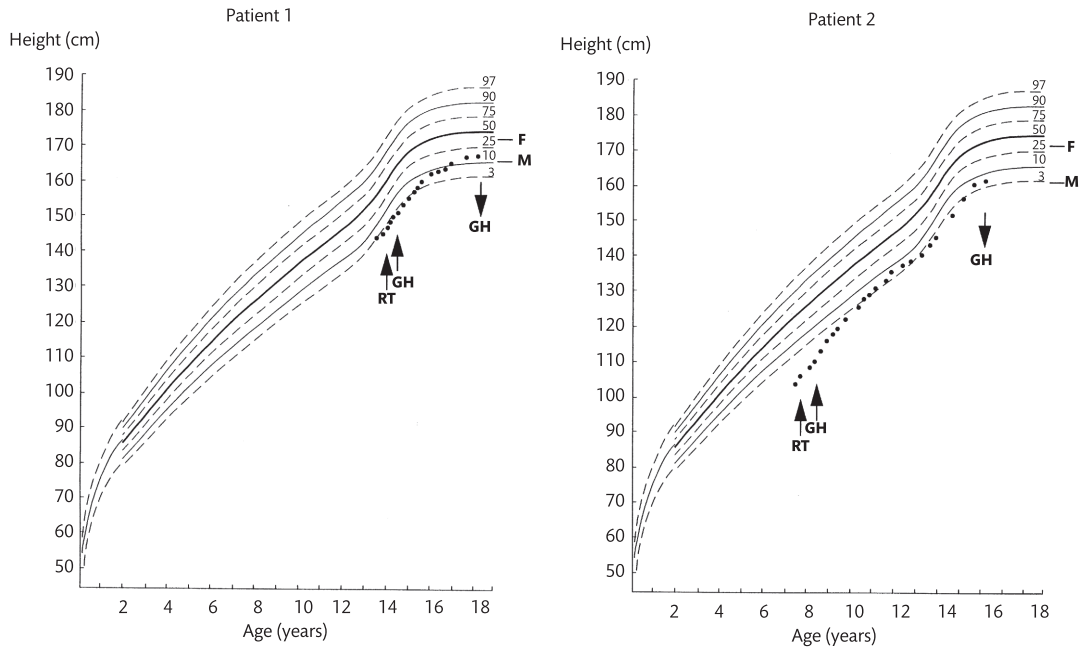


Figure 5.7.7 Growth charts of two children with Cushing's disease treated with pituitary irradiation. GHT, growth hormone therapy; RT, radiotherapy; M, mother's height; F, father's height.

Reproduced with permission from Johnston L. Grossmann AB, Plowman PN, Besser GM, Savage MO. *Clinical Endocrinology*, 1998; 48: 663–7.

adults. Although more technically challenging in children, inferior petrosal sinus sampling is, as in adults, vital for localizing the source of ACTH secretion [40].

Treatment

The principles of treatment of hypercortisolaemia are the same in children as in adults. Ketoconazole is preferable to metyrapone in children, as the former lowers, rather than increases, circulating androgen levels, but both are safe. Transsphenoidal surgery achieves remission in the majority of children with Cushing's disease. Pituitary radiotherapy is reserved for surgical failures, or those whose surgery is impossible because of the small size of the pituitary fossa; however, when required it controls ACTH secretion more promptly in children than in adults with Cushing's disease, but residual pituitary function requires close monitoring to ensure normal pubertal development and growth. Growth hormone deficiency occurs early after radiotherapy in children, but may recover in some [77].

Once hypercortisolaemia has been controlled, the management of growth and puberty is a major challenge. Glucocorticoids both inhibit growth hormone secretion and induce epiphyseal insensitivity to growth hormone action, and correction of hypercortisolaemia is a prerequisite to re-establishing linear growth. Adrenal androgen-induced pseudoprecocious puberty causes premature true gonadotropin-dependent puberty, and hence, even once adrenal androgen secretion has been controlled, bone age will continue to advance and potential for linear growth to diminish. These factors can be regulated by the combined use of gonadotropin-releasing hormone analogues to inhibit gonadotropin secretion and control puberty, and growth hormone treatment to induce linear growth. With effective treatment of hypercortisolaemia and careful management of puberty and growth, children with Cushing's syndrome will achieve a normal final height [78] (**Figure 5.7.7**).

Likely Developments Over the Next 5–10 Years

There remains a clear need for novel approaches to medical therapy. Antisense approaches to the GR that inhibit in the periphery, but do not cross the blood-brain barrier and so still allow GR-mediated feedback, and antibody approaches to lower ACTH are in development. Improved radiotracer ligands are likely to enhance the ability to discover occult sources of ACTH secretion.

REFERENCES

1. Newell-Price J, Bertagna X, Grossman AB, Nieman LK. Cushing's syndrome. *Lancet*, 2006; **367**: 1605–17.
2. Newell-Price J, Trainer P, Besser M, Grossman A. The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. *Endocr Rev*, 1998; **19**: 647–72.
3. Lacroix A, Feelders RA, Stratakis CA, Nieman LK. Cushing's syndrome. *Lancet*, 2015; **386**: 913–27.
4. Woo YS, Isidori AM, Wat WZ, *et al*. Clinical and biochemical characteristics of adrenocorticotropin-secreting macroadenomas. *J Clin Endocrinol Metab*, 2005; **90**: 4963–9.
5. Daniel E, Newell-Price JD. Diagnosis of Cushing's disease. *Pituitary*, 2015; **18**: 206–10.
6. Bilodeau S, Vallette-Kasic S, Gauthier Y, *et al*. Role of Brg1 and HDAC2 in GR trans-repression of the pituitary POMC gene and misexpression in Cushing disease. *Genes Dev*, 2006; **20**: 2871–86.
7. Liu NA, Araki T, Cuevas-Ramos D, *et al*. Cyclin E-mediated human proopiomelanocortin regulation as a therapeutic target for Cushing disease. *J Clin Endocrinol Metab*, 2015; **100**: 2557–64.
8. Storr HL, Isidori AM, Monson JP, Besser GM, Grossman AB, Savage MO. Prepubertal Cushing's disease is more common in males, but there is no increase in severity at diagnosis. *J Clin Endocrinol Metab*, 2004; **89**: 3818–20.
9. Reincke M, Sbiera S, Hayakawa A, *et al*. Mutations in the deubiquitinase gene USP8 cause Cushing's disease. *Nat Genet*, 2015; **47**: 31–8.
10. Ma ZY, Song ZJ, Chen JH, *et al*. Recurrent gain-of-function USP8 mutations in Cushing's disease. *Cell Res*, 2015; **25**: 306–17.
11. Perez-Rivas LG, Theodoropoulou M, Ferrau F, *et al*. The gene of the ubiquitin-specific protease 8 is frequently mutated in adenomas causing Cushing's disease. *J Clin Endocrinol Metab*, 2015; **100**: E997–1004.
12. Picon A, Bertagna X, de Keyzer Y. Analysis of the human proopiomelanocortin gene promoter in a small cell lung carcinoma cell line reveals an unusual role for E2F transcription factors. *Oncogene*, 1999; **18**: 2627–33.
13. Newell-Price J, King P, Clark AJ. The CpG island promoter of the human proopiomelanocortin gene is methylated in nonexpressing normal tissue and tumors and represses expression. *Mol Endocrinol*, 2001; **15**: 338–48.
14. Beuschlein F, Fassnacht M, Assie G, *et al*. Constitutive activation of PKA catalytic subunit in adrenal Cushing's syndrome. *N Engl J Med*, 2014; **370**: 1019–28.
15. Cao Y, He M, Gao Z, *et al*. Activating hotspot L205R mutation in PRKACA and adrenal Cushing's syndrome. *Science*, 2014; **344**: 913–17.
16. Goh G, Scholl UI, Healy JM, *et al*. Recurrent activating mutation in PRKACA in cortisol-producing adrenal tumors. *Nat Genet*, 2014; **46**: 613–17.
17. Sato Y, Maekawa S, Ishii R, *et al*. Recurrent somatic mutations underlie corticotropin-independent Cushing's syndrome. *Science*, 2014; **344**: 917–20.
18. Lacroix A, Ndiaye N, Tremblay J, Hamet P. Ectopic and abnormal hormone receptors in adrenal Cushing's syndrome. *Endocr Rev*, 2001; **22**: 75–110.
19. Assie G, Libe R, Espiard S, *et al*. ARMC5 mutations in macronodular adrenal hyperplasia with Cushing's syndrome. *N Engl J Med*, 2013; **369**: 2105–14.
20. Storr HL, Mitchell H, Swords FM, *et al*. Clinical features, diagnosis, treatment and molecular studies in paediatric Cushing's syndrome due to primary nodular adrenocortical hyperplasia. *Clin Endocrinol (Oxf)*, 2004; **61**: 553–9.
21. Stratakis CA, Kirschner LS, Carney JA. Clinical and molecular features of the Carney complex: diagnostic criteria and recommendations for patient evaluation. *J Clin Endocrinol Metab*, 2001; **86**: 4041–6.
22. Kirschner LS, Carney JA, Pack SD, *et al*. Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. *Nat Genet*, 2000; **26**: 89–92.
23. Casey M, Vaughan CJ, He J, *et al*. Mutations in the protein kinase A R1alpha regulatory subunit cause familial cardiac myxomas and Carney complex. *J Clin Invest*, 2000; **106**: R31–8.
24. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating mutations of the stimulatory G

protein in the McCune–Albright syndrome. *N Engl J Med*, 1991; **325**: 1688–95.
25. Nieman LK, Biller BM, Findling JW, *et al*. The diagnosis of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2008; **93**: 1526–40.
26. Pecori Giraldi F, Moro M, Cavagnini F. Gender-related differences in the presentation and course of Cushing's disease. *J Clin Endocrinol Metab*, 2003; **88**: 1554–8.
27. Petersenn S, Newell-Price J, Findling JW, *et al*. High variability in baseline urinary free cortisol values in patients with Cushing's disease. *Clin Endocrinol (Oxf)*, 2014; **80**: 261–9.
28. Hawley JM, Owen LJ, Debono M, Newell-Price J, Keevil BG. Development of a rapid liquid chromatography tandem mass spectrometry method for the quantitation of serum dexamethasone and its clinical verification. *Ann Clin Biochem*, 2018; **55**: 665–72.
29. Newell-Price J, Trainer P, Perry L, Wass J, Grossman A, Besser M. A single sleeping midnight cortisol has 100% sensitivity for the diagnosis of Cushing's syndrome. *Clin Endocrinol (Oxf)*, 1995; **43**: 545–50.
30. Papanicolaou DA, Yanovski JA, Cutler GB, Jr., Chrousos GP, Nieman LK. A single midnight serum cortisol measurement distinguishes Cushing's syndrome from pseudo-Cushing states. *J Clin Endocrinol Metab*, 1998; **83**: 1163–7.
31. Carroll T, Raff H, Findling JW. Late-night salivary cortisol for the diagnosis of Cushing syndrome: a meta-analysis. *Endocr Pract*, 2009; **15**: 335–42.
32. Carroll T, Raff H, Findling JW. Late-night salivary cortisol measurement in the diagnosis of Cushing's syndrome. *Nat Clin Pract Endocrinol Metab*, 2008; **4**: 344–50.
33. Liu H, Bravata DM, Cabaccan J, Raff H, Ryzen E. Elevated late-night salivary cortisol levels in elderly male type 2 diabetic veterans. *Clin Endocrinol (Oxf)*, 2005; **63**: 642–9.
34. Findling JW, Raff H. Diagnosis of endocrine disease: differentiation of pathologic/neoplastic hypercortisolism (Cushing's syndrome) from physiologic/non-neoplastic hypercortisolism (formerly known as pseudo-Cushing's syndrome). *Eur J Endocrinol*, 2017; **176**: R205–16.
35. Martin NM, Dhillo WS, Banerjee A, *et al*. Comparison of the dexamethasone-suppressed corticotropin-releasing hormone test and low-dose dexamethasone suppression test in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab*, 2006; **91**: 2582–6.
36. Martin NM, Dhillo WS, Meeran K. The dexamethasone-suppressed corticotropin-releasing hormone stimulation test and the desmopressin test to distinguish Cushing's syndrome from pseudo-Cushing's states. *Clin Endocrinol (Oxf)*, 2007; **67**: 476.
37. Newell-Price J, Morris DG, Drake WM, *et al*. Optimal response criteria for the human CRH test in the differential diagnosis of ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*, 2002; **87**: 1640–5.
38. Nieman LK, Oldfield EH, Wesley R, Chrousos GP, Loriaux DL, Cutler GB, Jr. A simplified morning ovine corticotropin-releasing hormone stimulation test for the differential diagnosis of adrenocorticotropin-dependent Cushing's syndrome. *J Clin Endocrinol Metab*, 1993; **77**: 1308–12.
39. Lindsay JR, Nieman LK. Differential diagnosis and imaging in Cushing's syndrome. *Endocrinol Metab Clin North Am*, 2005; **34**: 403–21.
40. Storr HL, Drake WM, Evanson J, *et al*. Endonasal endoscopic transsphenoidal pituitary surgery: early experience and outcome in paediatric Cushing's disease. *Clin Endocrinol (Oxf)*, 2014; **80**: 270–6.
41. Sohaib SA, Hanson JA, Newell-Price JD, *et al*. CT appearance of the adrenal glands in adrenocorticotrophic hormone-dependent Cushing's syndrome. *AJR Am J Roentgenol*, 1999; **172**: 997–1002.
42. Hall WA, Luciano MG, Doppman JL, Patronas NJ, Oldfield EH. Pituitary magnetic resonance imaging in normal human volunteers: occult adenomas in the general population. *Ann Intern Med*, 1994; **120**: 817–20.
43. Ilias I, Torpy DJ, Pacak K, Mullen N, Wesley RA, Nieman LK. Cushing's syndrome due to ectopic corticotropin secretion: twenty years' experience at the National Institutes of Health. *J Clin Endocrinol Metab*, 2005; **90**: 4955–62.
44. Isidori AM, Kaltsas GA, Pozza C, *et al*. The ectopic adrenocorticotropin syndrome: clinical features, diagnosis, management, and long-term follow-up. *J Clin Endocrinol Metab*, 2006; **91**: 371–7.
45. de Bruin C, Hofland LJ, Nieman LK, *et al*. Mifepristone effects on tumor somatostatin receptor expression in two patients with Cushing's syndrome due to ectopic adrenocorticotropin secretion. *J Clin Endocrinol Metab*, 2012; **97**: 455–62.
46. Nieman LK, Biller BM, Findling JW, *et al*. Treatment of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2015; **100**: 2807–31.
47. Feelders RA, Newell-Price J, Pivonello R, Nieman LK, Hofland LJ, Lacroix A. Advances in the medical treatment of Cushing's syndrome. *Lancet Diabetes Endocrinol*, 2019; **7**: 300–12.
48. Trainer PJ, Eastment C, Grossman AB, Wheeler MJ, Perry L, Besser GM. The relationship between cortisol production rate and serial serum cortisol estimation in patients on medical therapy for Cushing's syndrome. *Clin Endocrinol (Oxf)*, 1993; **39**: 441–3.
49. Findling JW, Fleseriu M, Newell-Price J, *et al*. Late-night salivary cortisol may be valuable for assessing treatment response in patients with Cushing's disease: 12-month, Phase III pasireotide study. *Endocrine*, 2016; **54**: 516–23.
50. Daniel E, Newell-Price JD. Therapy of endocrine disease: steroidogenesis enzyme inhibitors in Cushing's syndrome. *Eur J Endocrinol*, 2015; **172**: R263–80.
51. Daniel E, Aylwin S, Mustafa O, *et al*. Effectiveness of metyrapone in treating cushing's syndrome: a retrospective multicenter study in 195 patients. *J Clin Endocrinol Metab*, 2015; **100**: 4146–54.
52. Monaghan PJ, Owen LJ, Trainer PJ, Brabant G, Keevil BG, Darby D. Comparison of serum cortisol measurement by immunoassay and liquid chromatography-tandem mass spectrometry in patients receiving the 11beta-hydroxylase inhibitor metyrapone. *Ann Clin Biochem*, 2011; **48**(Pt 5): 441–6.
53. Castinetti F, Guignat L, Giraud P, *et al*. Ketoconazole in Cushing's disease: is it worth a try? *J Clin Endocrinol Metab*, 2014; **99**: 1623–30.
54. Kamenicky P, Droumaguet C, Salenave S, *et al*. Mitotane, metyrapone, and ketoconazole combination therapy as an alternative to rescue adrenalectomy for severe ACTH-dependent Cushing's syndrome. *J Clin Endocrinol Metab*, 2011; **96**: 2796–804.
55. Chan LF, Vaidya M, Westphal B, *et al*. Use of intravenous etomidate to control acute psychosis induced by the hypercortisolaemia in severe paediatric Cushing's disease. *Horm Res Paediatr*, 2011; **75**: 441–6.
56. Drake WM, Perry LA, Hinds CJ, Lowe DG, Reznek RH, Besser GM. Emergency and prolonged use of intravenous etomidate to

control hypercortisolemia in a patient with Cushing's syndrome and peritonitis. *J Clin Endocrinol Metab*, 1998; **83**: 3542–4.
57. Baudry C, Coste J, Bou Khalil R, *et al*. Efficiency and tolerance of mitotane in Cushing's disease in 76 patients from a single center. *Eur J Endocrinol*, 2012; **167**: 473–81.
58. Bertagna X, Pivonello R, Fleseriu M, *et al*. LCI699, a potent 11beta-hydroxylase inhibitor, normalizes urinary cortisol in patients with Cushing's disease: results from a multicenter, proof-of-concept study. *J Clin Endocrinol Metab*, 2014; **99**: 1375–83.
59. Fleseriu M, Biller BM, Findling JW, Molitch ME, Schteingart DE, Gross C. Mifepristone, a glucocorticoid receptor antagonist, produces clinical and metabolic benefits in patients with Cushing's syndrome. *J Clin Endocrinol Metab*, 2012; **97**: 2039–49.
60. Colao A, Petersenn S, Newell-Price J, *et al*. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med*, 2012; **366**: 914–24.
61. Lacroix A, Gu F, Gallardo W, *et al*. Efficacy and safety of once-monthly pasireotide in Cushing's disease: a 12 month clinical trial. *Lancet Diabetes Endocrinol*, 2018; **6**: 17–26.
62. Ferriere A, Cortet C, Chanson P, *et al*. Cabergoline for Cushing's disease: a large retrospective multicenter study. *Eur J Endocrinol*, 2017; **176**: 305–14.
63. Petersenn S, Beckers A, Ferone D, *et al*. Therapy of endocrine disease: outcomes in patients with Cushing's disease undergoing transsphenoidal surgery: systematic review assessing criteria used to define remission and recurrence. *Eur J Endocrinol*, 2015; **172**: R227–39.
64. Valassi E, Franz H, Brue T, *et al*. Preoperative medical treatment in Cushing's syndrome: frequency of use and its impact on postoperative assessment: data from ERCUSYN. *Eur J Endocrinol*, 2018; **178**: 399–409.
65. Palazzo F, Dickinson A, Phillips B, *et al*. Adrenal surgery in England: better outcomes in high-volume practices. *Clin Endocrinol (Oxf)*, 2016; **85**: 17–20.
66. Reincke M, Ritzel K, Osswald A, *et al*. A critical reappraisal of bilateral adrenalectomy for ACTH-dependent Cushing's syndrome. *Eur J Endocrinol*, 2015; **173**: M23–32.
67. Ritzel K, Beuschlein F, Mickisch A, *et al*. Clinical review: outcome of bilateral adrenalectomy in Cushing's syndrome: a systematic review. *J Clin Endocrinol Metab*, 2013; **98**: 3939–48.
68. Assie G, Bahurel H, Coste J, *et al*. Corticotroph tumor progression after adrenalectomy in Cushing's disease: a reappraisal of Nelson's syndrome. *J Clin Endocrinol Metab*, 2007; **92**: 172–9.
69. Barber TM, Adams E, Ansorge O, Byrne JV, Karavitaki N, Wass JA. Nelson's syndrome. *Eur J Endocrinol*, 2010; **163**: 495–507.
70. Cohen-Inbar O, Ramesh A, Xu Z, Vance ML, Schlesinger D, Sheehan JP. Gamma knife radiosurgery in patients with persistent acromegaly or Cushing's disease: long-term risk of hypopituitarism. *Clin Endocrinol (Oxf)*, 2016; **84**: 524–31.
71. Sheehan JP, Xu Z, Salvetti DJ, Schmitt PJ, Vance ML. Results of gamma knife surgery for Cushing's disease. *J Neurosurg*, 2013; **119**: 1486–92.
72. Clayton RN, Jones PW, Reulen RC, *et al*. Mortality in patients with Cushing's disease more than 10 years after remission: a multicentre, multinational, retrospective cohort study. *Lancet Diabetes Endocrinol*, 2016; **4**: 569–76.
73. Clayton RN, Raskauskiene D, Reulen RC, Jones PW. Mortality and morbidity in Cushing's disease over 50 years in Stoke-on-Trent, UK: audit and meta-analysis of literature. *J Clin Endocrinol Metab*, 2011; **96**: 632–42.
74. Lindsay JR, Nansel T, Baid S, Gumowski J, Nieman LK. Long-term impaired quality of life in Cushing's syndrome despite initial improvement after surgical remission. *J Clin Endocrinol Metab*, 2006; **91**: 447–53.
75. Valassi E, Feelders R, Maiter D, *et al*. Worse health-related quality of life at long-term follow-up in patients with Cushing's disease than patients with cortisol producing adenoma. Data from the ERCUSYN. *Clin Endocrinol (Oxf)*, 2018; **88**: 787–98.
76. Storr HL, Chan LF, Grossman AB, Savage MO. Paediatric Cushing's syndrome: epidemiology, investigation and therapeutic advances. *Trends Endocrinol Metab*, 2007; **18**: 167–74.
77. Storr HL, Plowman PN, Carroll PV, *et al*. Clinical and endocrine responses to pituitary radiotherapy in pediatric Cushing's disease: an effective second-line treatment. *J Clin Endocrinol Metab*, 2003; **88**: 34–7.
78. Savage MO, Storr HL, Grossman AB, Krassas GE. Growth and growth hormone secretion in paediatric Cushing's disease. *Hormones (Athens)*, 2003; **2**: 93–7.

5.8

Adrenal Insufficiency

5.8.1 Genetics of Adrenal Insufficiency

Li F. Chan and Shwetha Ramachandrappa

Introduction

Adrenal insufficiency refers to the inadequate secretion of glucocorticoids from the adrenal cortex, with or without mineralocorticoid deficiency. The production of glucocorticoid, cortisol in humans, and corticosterone in rodents, is regulated by circulating levels of adrenocorticotropic hormone (ACTH) which is secreted by the anterior pituitary gland in response to hypothalamic signals. The hypothalamic–pituitary–adrenal (HPA) axis is under negative feedback control with cortisol acting to inhibit the secretion of corticotropin-releasing hormone (CRH) and ACTH from the hypothalamus and pituitary, respectively.

Primary adrenal insufficiency is used to describe situations where the causal defect in glucocorticoid production lies within the adrenal gland, while in secondary adrenal insufficiency defects in the pituitary gland or hypothalamus lead to inadequate ACTH levels, with lack of adrenocortical stimulation resulting in secondary loss of cortisol production.

Causes of adrenal insufficiency can be genetic or non-genetic. Non-genetic causes are the most common (listed in Table 5.8.1.1) and include, autoimmune, infective, neoplastic, traumatic, and iatrogenic causes presenting at any age. Genetic causes, which are inherited or sporadic, as part of a syndrome or restricted to isolated adrenal insufficiency, generally present early in life although presentation in adolescence and adulthood has been described.

The incidence of primary adrenal insufficiency is reported to be 4 per million in Caucasian populations with a prevalence of 100–140 cases per million [1]. The most common cause of primary adrenal insufficiency in Western society is autoimmune adrenalitis while infective causes, especially tuberculous adenalitis, remain a leading cause of adrenal insufficiency worldwide. In comparison secondary adrenal insufficiency has an estimated prevalence of 150 to 280 per million [2, 3]. The most common cause of secondary adrenal insufficiency is suggested to be iatrogenic due to glucocorticoid therapy that leads to suppression of ACTH secretion and adrenal atrophy.

In this chapter we will focus on the causes of monogenic adrenal insufficiency. As expected, alterations in genes that play a direct role in the development and function of the HPA axis give rise to monogenic causes of adrenal insufficiency. More recently, defects in a number of ubiquitously expressed genes involved in a wide range of cellular processes have also been shown to cause adrenal insufficiency, providing novel insights into adrenal steroid regulation.

Genetics of Secondary Adrenal Insufficiency

There are many genetic causes of secondary adrenal insufficiency broadly categorized into those involved in pituitary development and genes involved in pro-opiomelanocortin (POMC) cell maturation and POMC synthesis and processing (Table 5.8.1.2). Defects in genes involved in pituitary development give rise to combined pituitary hormone deficiency. Often there is relative sparing of corticotroph function in these disorders, which are described in more detail in Chapter 2.3.1 ('Development of the Pituitary and Genetic Forms of Hypopituitarism'). Here we will focus on the few genes, in which alterations give rise to isolated ACTH deficiency.

Disorders of POMC Synthesis and Processing

The paraventricular nucleus of the hypothalamus secretes vasopressin and CRH, which regulate the synthesis of the precursor protein pro-opiomelanocortin (POMC) in the corticotropic cells of the anterior pituitary gland. POMC is then cleaved to produce ACTH, which binds to its unique receptor the ACTH receptor (also known as the melanocortin 2 receptor, MC2R) on the surface of adrenocortical cells to stimulate glucocorticoid secretion. POMC is also expressed in the arcuate nucleus of the hypothalamus where it is processed to generate melanocortin peptides, most importantly α-MSH. Central melanocortin signalling through the melanocortin 3 and 4 receptors (MC3R and MC4R) regulates energy balance and satiety [4]. Peripheral melanocortin signalling through the melanocortin 1 receptor (MC1R) is involved in pigment production in melanocytes [5]. Biallelic loss-of-function mutations in POMC in humans have been

Table 5.8.1.1 Non-genetic causes of adrenal insufficiency

Secondary adrenal insufficiency	
Iatrogenic	Adrenal suppression following exogenous steroid treatment
	Adrenal suppression following treatment of endogenous hypercortisolaemia due to Cushing's syndrome
	Following transsphenoidal surgery for ACTH-producing adenoma
	Pituitary irradiation and radiotherapy for brain tumours and craniospinal irradiation for other tumours
Tumours	Pituitary tumours (mostly adenomas, rarely carcinomas)
	Tumours affecting the hypothalamic–pituitary region (craniopharyngioma, meningioma, ependymoma, intra/suprasellar metastasis)
Infiltration/deposition	Tuberculosis, actinomycosis, sarcoidosis, histiocytosis X, Wegener's granulomatosis, Hemochromatosis
Autoimmune	Autoimmune hypophysitis; most frequently in relation to pregnancy; commonly associated with panhypopituitarism, but also presenting with isolated ACTH deficiency
Infarction/haemorrhage	Pituitary apoplexy/necrosis
	Postpartum pituitary infarction (Sheehan's syndrome)
Trauma	Pituitary stalk lesions
	Traumatic brain injury
Primary adrenal insufficiency	
Infections leading to adrenalitis	Tuberculosis
	HIV, CMV
	Cryptococcosis, histoplasmosis, coccidiomycosis, candidiasis, African trypanomiasis
	Treponema pallidum
Bilateral adrenal haemorrhage	Septic shock, specifically meningococcal sepsis (Waterhouse–Friderichsen syndrome)
	Anticoagulants
	Trauma
Adrenal infarction	Primary antiphospholipid syndrome
	Anticardiolipin syndrome
	Lupus anticoagulant syndrome
Adrenal infiltration	Adrenal metastases
	Primary adrenal lymphoma
	Sarcoidosis
	Amyloidosis
	Hemochromatosis
Bilateral adrenalectomy	Management of intractable Cushing's disease or ectopic ACTH secretion
	Following nephrectomy for tumour
	Treatment for bilateral pheochromocytoma
Drug-induced AI	Mitotane (steroid synthesis inhibitor/adrenolytic agent)
	Mifepristone (Glucocorticoid Receptor Antagonist antagonist)
	Abiraterone acetate (steroid synthesis inhibitor)
	Trilostane (steroid synthesis inhibitor)
	Etomidate (inhibits cortisol synthesis)
	Ketaconazole/fluconazole (inhibits cortisol synthesis)
	Suramin (steroid synthesis inhibitor/adrenolytic agent)
	Aminoglutethiamide (inhibits cortisol synthesis)
	Phenytoin (increased cortisol metabolism)
	Phenobarbital (increased cortisol metabolism)
	Rifampicin (enhanced autoimmunity)
	St John's wort (increased cortisol metabolism)
	CTLA-4 inhibitors (enhanced autoimmunity)
	Thyroxine (accelerated cortisol metabolism and stimulation of 11-beta-hydroxysteroid dehydrogenase)

Genetic SAI

Table 5.8.1.2 Genetic causes of secondary adrenal insufficiency

Disorder	Gene	OMIM	Mechanism of action	Inheritance	Additional phenotypic features
Pituitary development					
Combined pituitary hormone deficiency	Prophet of Pit-1 (PROP1)	601538	Loss of function	AR	Progressive development of combined pituitary hormone deficiency (CPHD) in the order GH, PRL, TSH, LH/FSH, (ACTH—late onset); anterior pituitary may be hypoplastic, normal, or enlarged
	Homeobox gene 1 (HESX1)	601802	Loss of function	AR/AD	CPHD + optic nerve hypoplasia and midline brain defects/ agenesis of corpus callosum (septo-optic dysplasia); anterior pituitary hypoplastic or ectopic;
	LIM homeobox 3 (LHX3)	600577	Loss of function	AR	CPHD with involvement of GH, TSH, gonadotrophins, PRL; ACTH may be deficient; limited neck rotation, short cervical spine, sensorineural deafness; anterior pituitary hypoplastic, normal, or enlarged
	LIM homeobox 4 (LHX4)	602146	Loss of function	AD	CPHD with involvement of GH, thyrotropin, and ACTH secretion, cerebellar abnormalities; anterior pituitary hypoplastic or ectopic
	SRY-box 3 (SOX3)	313430	Loss of function	XLR	Infundibular hypoplasia, CPHD, variable mental retardation
POMC synthesis and processing					
POMC deficiency	Pro-opiomelanocortin (POMC)	176830	Loss of function	AR	Hyperphagia, early onset obesity, cholestasis, pigmentary defect, central hypothyroidism, GH deficiency, hypogonadotropic hypogonadism
Proprotein convertase 1/3 deficiency	Proprotein convertase Subtilisin/Kexin type 1(PCSK1)	162150	Loss of function	AR	Obesity, gastrointestinal disturbance, other pituitary deficiencies possible, elevated proinsulin, low insulin
POMC lineage differentiation and transcription defects					
Congenital isolated ACTH deficiency	T-box transcription factor 19 (TBX19)	604614	Loss of function	AR	Neonatal presentation
Other syndromes associated with secondary adrenal insufficiency					
Prader–Willi syndrome	Imprinted gene cluster 15q11.2 (Including genes SNRPN, NDN)	-	Loss of function	IC	Syndrome of hypotonia, short stature, hyperphagia, obesity, hypogonadism, psychomotor delay, and sleep-related breathing disorders
Pallister-Hall syndrome	GLI3	165240	Loss of function	AD	Hypothalamic hamartomas, hypopituitarism, imperforate anus, and postaxial polydactyly

Genetic PAI

demonstrated to lead to adrenal deficiency in the first few months of life. Hyperphagia, early onset obesity and hypopigmentation of skin and hair are invariably also present [6–8]. Mutations in proprotein convertase 1 (PC1), a subtilisin-like enzyme responsible for the cleavage of POMC, gives rise to a related phenotype of secondary adrenal insufficiency associated with obesity [9].

POMC Lineage Differentiation and Transcription Defects

Autosomal recessive genetic variants in the gene T-box 19 (TBX19), which encodes a transcription factor, which is involved in the development of the POMC expressing cells of the anterior pituitary and the transcription of POMC, have been identified in patients with isolated ACTH deficiency, typically presenting with hypoglycaemia, seizures and prolonged jaundice in the neonatal period [10, 11].

Genetics of Primary Adrenal Insufficiency

The number of genes known to be involved in primary adrenal insufficiency has been steadily growing in light of recent breakthroughs in genetic techniques. These genes include those that affect adrenocortical cell function from ACTH signalling to defects in steroidogenesis. Other pathways such as redox and antioxidant mechanisms, and DNA replication and repair have more recently been shown to be important in the development of primary adrenal insufficiency. We have summarized the genes involved in primary adrenal insufficiency in Table 5.8.1.3 and selected those that affect specific pathways to discuss further here.

Defects in ACTH Signalling

Familial glucocorticoid deficiency (FGD), also known as isolated glucocorticoid deficiency or hereditary unresponsiveness to ACTH, is a subtype of primary adrenal insufficiency characterized by adrenal resistance to ACTH. The hallmark of FGD is low to absent serum cortisol in the presence of preserved mineralocorticoid production and extremely high levels of plasma ACTH. FGD classically presents in the neonatal period or infancy with symptoms of hypocortisolaemia leading to hypoglycaemia, failure to thrive, coma, susceptibility to infections, and death if untreated. Intense hyperpigmentation is a feature of FGD due to the action of ACTH on MC1R in skin.

Table 5.8.1.3 Genetic causes of primary adrenal insufficiency

Disorder	Gene	OMIM	Mechanism of action	Inheritance	Additional phenotypic features
Defects in ACTH signalling					
Familial glucocorticoid deficiency 1	MC2R	607397	Loss of function	AR	Isolated GC deficiency, occasionally mild MC deficiency, absent adrenarche, and tall stature
Familial glucocorticoid deficiency 2	MRAP	609196	Loss of function	AR	Isolated GC deficiency occasionally mild derangement of RAA
Defects affecting the steroidogenic pathways					
Congenital lipoid adrenal hyperplasia	StAR	600617	Loss of function	AR	46,XY sex reversal, partial or complete, gonadal insufficiency
Congenital lipoid adrenal hyperplasia due to cytochrome P450 SCC deficiency	CYP11A1	118485	Loss of function	AR	46,XY sex reversal, partial or complete, gonadal insufficiency
Congenital adrenal hyperplasia due to 3beta-HSD deficiency	HSD3B2	613890	Loss of function	AR	46,XY DSD, 46,XX DSD,
Congenital adrenal hyperplasia due to 21-hydroxylase deficiency	CYP21A2	613815	Loss of function	AR	46,XX DSD, androgen excess
Congenital adrenal hyperplasia due to 11beta hydroxylase deficiency	CYP11B1	610613	Loss of function	AR	46,XX DSD, androgen excess, hypertension
Congenital adrenal hyperplasia due to 17α-hydroxylase deficiency	CYP17A1	609300	Loss of function	AR	46,XY DSD, gonadal insufficiency, hypertension
Congenital adrenal hyperplasia due to cytochrome P450 oxidoreductase deficiency	POR	124015	Loss of function	AR	46,XY DSD, 46,XX DSD, gonadal insufficiency, dysmorphisms skeletal malformation, disorder of drug metabolism, developmental delay
Alterations in antioxidant mechanisms					
Familial glucocorticoid deficiency 4, with or without mineralocorticoid deficiency	NNT	607878	Loss of function	AR	Variable presentation with or without mineralocorticoid deficiency
Familial glucocorticoid deficiency	GPX1	138320	Loss of function	AR or digenic	
Familial glucocorticoid deficiency	PRDX3	604769	Loss of function	Possible Digenic	
Familial glucocorticoid deficiency, type 5	TXNRD2	606448	Loss of function	AR	Possible structural cardiac abnormalities
Defects in metabolic pathways					
Wolman disease and cholesteryl ester storage disease	LIPA	613497	Loss of function	AR	Malabsorption, gastrointestinal symptoms, liver dysfunction, atherosclerosis, serum lipid abnormalities, hypersplenism, xanthelasma, hypercholesterolemia, steatorrhea, enlarged adrenal glands with punctate calcifications
Smith–Lemi–Opitz disease	DHCR7	602858	Loss of function	AR	Mental retardation, craniofacial malformations, growth failure, polydactyly, 46,XY DSD, renal hypoplasia, genital abnormalities
Abetalipoproteinemia	MTP	157147	Loss of function	AR	Hypocholesterolaemia and malabsorption of lipid-soluble vitamins leading to retinal degeneration, neuropathy, and coagulopathy. Hepatic steatosis
Familial hypercholesterolaemia	LDLR	606945	Loss of function	AD	Tendinous xanthomas, corneal arcus, and coronary artery disease
Sitosterolemia (phytosterolemia)	ABCG5, ABCG8	605459, 605460	Loss of function	AR	Very high levels of plant sterols in the plasma and develop tendon and tuberous xanthomas, accelerated atherosclerosis, and premature coronary artery disease, short stature, gonadal failure
Adrenoleukodystrophy	ABCD1	300371	Loss of function	XLR	Cognitive and behavioural problems, motor problems, visual impairment, headache, seizures, bowel and bladder disturbance, sexual dysfunction, elevated VLCFA's

(*continued*)

Table 5.8.1.3 *Continued*

Disorder	Gene	OMIM	Mechanism of action	Inheritance	Additional phenotypic features
Peroxisome biogenesis disorder 1A (Zellweger)	PEX1	602136	Loss of function	AR	Congenital malformations, developmental delay, corneal clouding, retinal dystrophy, sensorineural hearing loss, neurological deficits, liver dysfunction, renal oxalate stones, failure to thrive, skeletal abnormalities, urogenital anomalies, liver dysfunction,
Sphingosine-1-phosphate lyase deficiency	SGPL1	603729	Loss of function	AR	Steroid-resistant nephrotic syndrome, Ichthyosis, primary hypothyroidism, hypogonadism, cryptorchidism, and immune and neurological abnormalities
Kearns–Sayre syndrome	mtDNA deletion	530000	Loss of function	Maternal	Ophthalmoplegia, pigmented retinopathy, cardiomyopathy, cerebellar ataxia, endocrinopathies, hypoparathyroidism, type 1 diabetes, rarely AI
Refsum disease	PHYH	602026	Loss of function	AR	Retinitis pigmentosa, peripheral neuropathy, cerebellar ataxia, ichthyosis, and elevated protein levels in the cerebrospinal fluid (CSF) without an increase in the number of cells
DNA replication and repair disorders					
Immunodeficiency 54,	MCM4	602638	Loss of function	AR	Severe intra and extrauterine growth retardation, microcephaly, decreased numbers of natural killer cells, and recurrent viral infections, hepatosplenomegaly, hyperpigmentation, increased susceptibility to lymphoproliferative disorders, chromosomal breakage
Achalasia–Addisonian–alacrimia syndrome	AAAS	605378	Loss of function	AR	Microcephaly, short stature, neurological symptoms, alacrimia, autonomic dysfunction, developmental delay, hyperkeratosis, achalasia, optic atrophy
Monogenic causes of autoimmune adrenalitis					
Autoimmune polyendocrinopathy syndrome, type I	AIRE	607358	Loss of function/ Dominant negative	AR/AD	Candidiasis, hypoparathyroidism, insulin dependent diabetes, alopecia, vitiligo, gastrointestinal complaints, eye abnormalities
Conditions associated with adrenal dysgenesis					
Congenital adrenal hypoplasia	NR0B1 (DAX1)	300473	Loss of function	XLR	Genital abnormalities, hypogonadotropic hypogonadism, occasionally gonadotrophin independent precocious puberty, Duchenne muscular dystrophy + glycerol kinase deficiency (psychomotor retardation) for Xp21.2 contiguous gene syndrome
IMAGe syndrome	CDKN1C	600856	Gain of function	AD—maternal transmission only	Intrauterine growth restriction, metaphyseal dysplasia, adrenal hypoplasia congenita, hypogonadotropic hypogonadism, and genitourinary abnormalities in males, facial dysmorphism
MIRAGE syndrome	SAMD9	610456	Gain of function	AD	Myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy, cognitive impairment, hypoplastic thymus, hyperpigmentation, skeletal anomalies, gastrointestinal problems,
Steroidogenic factor 1 deficiency	NR5A1	184757	Loss of function	AD	46,XY DSD, gonadal insufficiency, 46,XX Sex Reversal, Premature ovarian failure, Spermatogenic failure

FGD is a rare autosomal recessive disorder, first described by Shepherd *et al.* in 1959 [12], who reported siblings with 'familial Addison's disease'. This prompted subsequent reports of individuals with isolated glucocorticoid deficiency and normal aldosterone production [13–15]. The first inactivating mutation in the ACTH receptor was described some 30 years later [16, 17], made possible by the cloning of the receptor, the second melanocortin receptor to be cloned—hence the alternative name MC2R [18]. Since 1993 over 40 variants have been described, predominantly missense mutations spread throughout the length of the receptor [19]. Nonsense MC2R variants are less frequent and have been associated mild disturbance of the renin-angiotensin-aldosterone system [20, 21]. Mutations in MC2R account for approximately 25% of cases of FGD (FGD type 1). Tall stature has been associated with FGD type 1; it is unclear whether this is due to ACTH action on bone or to hypocortisolaemia, it is not seen in all FGD type 1 individuals. Absent adrenarche, i.e. loss of accelerating adrenal androgen production at age 6–9 years, has also been described, demonstrating the importance of ACTH in this process [22].

In 2005, homozygosity mapping using SNP arrays in consanguineous FGD families without causative mutations in MC2R enabled the identification of MRAP, melanocortin 2 receptor accessory protein [23]. It had long been suggested that the MC2R required an adrenal specific factor to enable expression at the cell surface. MRAP proved to be this factor. A small single transmembrane domain protein expressed in adrenal cells, MRAP was shown to be essential for the functional expression of the MC2R at the cell surface. Intriguingly, it has now been shown that this protein adopts an antiparallel dimer conformation—the only eukaryotic protein, except for its paralogue MRAP2, to do this [19]. Approximately 20% of cases of FGD are now known to be due to mutations in MRAP (FGD type 2). Unlike mutations in MC2R, MRAP mutations are usually nonsense or splice-site variations resulting in truncation or complete absence of the protein. As such, individuals with mutations in MRAP present earlier than those with defects in MC2R [24]. Rare MRAP missense mutations have been described and are associated with milder, late onset disease [25].

Studying FGD type 3 and beyond, i.e. those individuals without mutations MC2R and MRAP, has identified novel genes and pathways in adrenal gland physiology (described next). Such genetic discoveries have highlighted great phenotypic heterogeneity, prompting increased awareness of overlap in classical categorizations of primary adrenal insufficiency disorders.

Defects Affecting the Steroidogenic Pathways

The synthesis of steroid hormones in the adrenal cortex is dependent on the coordinated action of numerous enzymes [26]. Congenital adrenal hyperplasia (CAH) is a group of recessive conditions in which genetic alterations in the enzymes involved in steroid synthesis result in impaired glucocorticoid production. Depending on the enzyme involved and the point at which steroidogenesis is disrupted this is associated with excess or inadequate androgen production, accompanied by varying degrees of mineralocorticoid deficiency.

CAH is the most common genetic cause of primary adrenal insufficiency in childhood with mutations in *CYP21A2* accounting for 95% of cases of CAH [27]. The carrier rate of 21-hydroxylase deficiency (21OHD) is estimated to be around 1 in 50, but is higher in certain ethnicities. *CYP21A2* is located in a tandemly duplicated sequence within the human leukocyte antigen (HLA) cluster on chromosome 6 [28]. The region contains the C4A and C4B genes, which encode complement factor 4 as well as *CYP21A2* and its pseudogene *CYP21A1P* [29]. The tandem arrangement of these genes with a high degree of sequence similarity makes this region susceptible to misalignment during meiosis which can result in deletions and duplications as well as gene conversion events. The tenascin X gene, *TNXB*, which encodes an important constituent of the connective tissue, is located next to *CYP21A2*. Deletions of *CYP21A2* together with part or all of TNXB lead to connective tissue disease [30]. The high sequence similarly between *CYP21A2* and its pseudogene precludes the use of direct sequencing approaches in cell-free DNA analysis as part of non-invasive prenatal diagnosis for CAH. Relative haplotype dosage analysis has now been successfully developed for this purpose.

The genetics of CAH is covered in detail in Chapter 5.9.1, here we will focus on those genes that have been identified in studies of FGD.

The first step in steroid synthesis in the adrenal gland involves the rate-limiting step of cholesterol transport from the outer mitochondrial membrane to the inner mitochondrial membrane by the steroidogenic acute regulatory protein (StAR). Defects in the *STAR* gene have long been known to cause lipoid CAH, whereby both glucocorticoid and mineralocorticoid deficiency are present together with defective gonadal steroidogenesis, the latter presenting as 46,XY DSD or gonadal insufficiency. In 46,XX females, lipiod CAH can present as progressive hypogonadism. However, it is now clear that 5–10% of FGD patients harbour partial loss-of-function mutations in StAR which result in non-classical lipoid CAH and isolated glucocorticoid deficiency [31].

Once transported, cholesterol is converted to pregnenolone by the P450 side-chain cleavage enzyme (CYP11A1). Recessive genetic defects in the *CYP11A1* gene are known to cause a similar phenotype to lipoid CAH. Partial loss-of-function mutations, with retention of up to 20% of wild type function can lead to delayed onset adrenal insufficiency with minor malformations of the external genitalia such as hypospadias [32–34]. A relatively common variant in *CYP11A1* among European populations, traditionally characterized as a benign variant, has recently been shown to affect splicing, thus demonstrating the difficulties in assigning pathogenicity to variants and the limited utility in using population frequencies to inform pathogenicity in recessive disease.

Alterations in Antioxidant Mechanisms

The redox milieu of a cell is maintained by a tissue specific range of antioxidant mechanisms. There is growing evidence that adrenocortical cells are particularly susceptible to oxidative stress. Steroid production not only generates substantial levels of reactive oxygen species (ROS) but is also itself exquisitely sensitive to the intracellular redox environment. ROS have been shown to supress the synthesis of StAR protein and the steroidogenic enzymes *CYP11A1* and *CYP11B2* utilize the reducing power of NADPH [35]. Under normal physiological conditions, this negative feedback mechanism may have a role in the regulation of glucocorticoid secretion, however, a number of disease states illustrate that glucocorticoid secretion can be compromised when the redox state of the adrenal tissue is disturbed.

Examples of this include the discovery that mutations in the gene encoding Nicotinamide nucleotide transhydrogenase (NNT) account for 10% of FGD [36]. NNT is ubiquitously expressed in human tissues and is responsible for generation of NADPH, a reducing agent for the glutathione and thioredoxin systems as well as the ferredoxin system to remove ROS. The discovery of a homozygous truncating mutation in thioredoxin reductase 2 (TXNRD2) as a cause of FGD in a consanguineous family of Kashimiri origin suggests that individual components of these networks provide a rich source of putative candidate genes for FGD [37]. Alterations in glutathione peroxidase 1 (GPX1) and peroxiredoxin 3 (PRDX3) [38] have been described in a family with FGD, where they are hypothesized to have a synergistic effect. Furthermore, several observations suggest that oxidative stress may play a role in the pathogenesis of Triple A syndrome [39]. Seen in the context of the ubiquitous nature of these antioxidant pathways the narrow phenotype of FGD suggests that these antioxidant pathways are redundant/less important in most other tissue types.

Genes Affecting Metabolic Pathways That Are Essential for Steroid Synthesis

Esterified cholesterol and triglycerides are degraded by lysosomal acid lipase in lysosomes. The resulting non-esterified cholesterol can be used as a substrate for steroidogenesis. Lysosomal acid lipase (LAL) deficiency caused by biallelic genetic alterations in the gene *LIPA* leads to the accumulation of esterified cholesterol and triglycerides in the lysosomes [40, 41]. The deficiency of intracellular cholesterol can lead to impaired steroidogenesis affecting glucocorticoid and mineralocorticoid production. A wide range of phenotypes are seen in LAL deficiency. The infantile onset form known as Wolman disease is characterized by malabsorption, hepatomegaly, and adrenal insufficiency [42]. The later onset forms collectively known as cholesteryl ester storage disease (CESD) may present in childhood in a similar manner to Wolman disease or in later life with lipid abnormalities, hepatosplenomegaly, and elevated liver enzymes. Adrenal insufficiency can occasionally be seen in CESD [43].

Peroxisomes are found in virtually all eukaryotic cells. Their principal function is the β-oxidation of very long chain fatty acids (VLCFAs), amino acids, and uric acid. They are also important in producing cholesterol and plasmalogens, a distinct class of glycerophospholipids. Disruption of peroxisome function results in adrenal insufficiency, ostensibly as a result of the accumulation of VLCFAs initiating apoptotic mechanisms and resulting in the generation of reactive oxidative species.

X-linked adrenoleukodystrophy (ALD), caused by alterations in the gene *ABCD1* which mediates transport of VLCFA across the peroxisome membrane, manifests as a spectrum of different phenotypes affecting the adrenal gland and the nervous system [44]. There are three main patterns of involvement seen in affected males. In the childhood cerebral form, inflammatory cerebral demyelination often presents with mild learning disability or behavioural problems that are progressive. In the adrenomyeloneuropathy form loss of axons and demyelination in the gracile and corticospinal tracts leads to progressive spasticity and leg weakness presenting in adulthood. In both of these forms of X-linked ALD, affected individuals generally have adrenocortical dysfunction by the time their neurological problems are apparent. The third common presentation of X-linked ALD is isolated adrenal insufficiency, which can present in childhood or adulthood. A significant proportion of heterozygote females manifest adrenomyeloneuropathy symptoms, but adrenal function appears to be preserved. There is no correlation between the molecular insult in the *ABCD1* gene and the age of onset or severity of the resulting phenotype. The most common disease-causing alteration in *ABCD1* is c.1415_1416delAG, which leads to a frameshift and premature truncation of the resultant protein.

Zellweger spectrum disorder (ZSD) caused by mutations in the *PEX* genes leads to a range of phenotypes [45]. The *PEX* genes encode peroxin proteins required for the assembly and maintenance of peroxisome membranes and the import of peroxisome matrix proteins. In severe disease infants are born with congenital malformations, make poor developmental progress, and typically die in the first year of life. In milder disease progressive peroxisome dysfunction leads to retinal dystrophy, sensorineural hearing loss, neurological deficits, liver dysfunction, and renal oxalate stones. There is a high prevalence of adrenal insufficiency in ZSD [46]. The majority of cases of ZSD are related to alterations in PEX1. Two common variants in PEX1 have been identified p.Ile700TyrfsTer42 and p.Gly843Asp [47, 48]. These variants are found within specific haplotypes implying that they have arisen as a result of the dissemination of specific founder mutations [49].

DNA Replication and Repair Disorders Associated with Primary Adrenal Insufficiency

An autosomal recessive disorder characterized by adrenal insufficiency, severe intra- and extrauterine growth retardation, microcephaly, decreased numbers of natural killer cells, and recurrent viral infections has been described in a genetically isolated Irish population. This disorder is caused by genetic alterations in the gene *MCM4* (minichromosome maintenance complex component 4), which forms part of a helicase complex that is involved in DNA replication in the S phase of the cell cycle [50]. Affected individuals show increased chromosome breakage suggesting that DNA repair is compromised. To date, only one variant in MCM4 (c.71-1insG, p.Pro24ArgfsTer4) has been shown to cause FGD, identified in the Irish traveller community. In some individuals, adrenal insufficiency has been shown to develop over time.

Monogenic Causes of Autoimmune Adrenalitis

Autoimmune primary adrenal insufficiency also known as Addison's disease accounts for almost 90% of cases of primary adrenal insufficiency. This can be isolated or can be found together with other autoimmune conditions associated with genetic factors such as the HLA-DR3 haplotype or common polymorphisms in CTLA-4 (Chapter 1.7).

Monogenic causes of autoimmune adrenalitis are seen in polyglandular syndromes. Classical cases of autoimmune polyglandular syndrome type 1 (APS-1) characterized by hypoparathyroidism, adrenal insufficiency and chronic mucocutaneous candidiasis, are caused by genetic alterations in the autoimmune regulator gene *AIRE* [51, 52]. *AIRE* encodes a transcription factor, which is expressed in a subset of medullary T cells, where it mediates the expression of antigens leading to centrally mediated T cell tolerance. *AIRE* also appears to play a role in regulating the central and peripheral tolerance of B cells. Absence of functional *AIRE* resulting in APS-1 was first described as a recessive condition. However, dominant forms of the condition have now emerged, which are thought to be due to dominant negative mutations in the SAND and PHD domains.

Autoimmune polyglandular syndrome type 2 is characterized by adrenal insufficiency, hypothyroidism, and diabetes. Other endocrine glands may also be involved. APS-2 seems to be inherited within families in a dominant fashion with incomplete penetrance. The genetic basis of APS-2 is not yet known.

Disorders Associated with Adrenal Dysgenesis

Alterations in several genes can cause adrenal dysgenesis. Among these conditions, alterations in *CDKN1C* are notable for their unusual pattern of inheritance which derives from the location of *CDKN1C* within an imprinted region of the genome.

There are two imprinted clusters of genes at chromosome 11p15 which are known to regulate growth and development. Imprinting centre 1 regulates the expression of *IGF2* which is a fetal growth factor, and *H19*, a non-translated RNA. Imprinting centre 2 regulates the expression of *CDKN1C, KCNQ1OT1* a non-coding RNA, and *KCNQ1*. Differential methylation of the imprinting centres on the maternal and paternal chromosomes results in expression of *H19, CDKN1C*, and *KCNQ1* from the maternal chromosome and *IGF2* and *KCNQ1OT1* from the paternal chromosome. A number of growth disorders including Russell–Silver syndrome, Beckwith–Wiedemann syndrome (BWS) and hemihypertrophy are caused by perturbations of gene expression with this region.

CDKN1C is a cyclin dependent kinase inhibitor which plays a role in cell–cycle regulation, DNA replication and repair, and activation of apoptosis and is expressed in the developing adrenal gland. The CDKN1C protein has three functional domains, the cyclin-dependent kinase (CDK) inhibitor domain, the PAPA domain, and the PCNA domain. Approximately 5% of cases of BWS are caused by loss-of-function genetic alterations in the maternal copy of *CDKN1C* which are distributed throughout the gene. A broad spectrum of adrenal phenotypes are seen in BWS including cystic changes, adenomas, hyperplasia, and neuroendocrine tumours. In direct contrast, gain-of-function genetic alterations in the maternal copy of *CDKN1C* give rise to IMAGE syndrome characterized by intrauterine growth retardation, skeletal abnormalities, adrenal hypoplasia, and genital abnormalities in males [53, 54]. Adrenal insufficiency can be of variable severity, presenting soon after birth with adrenal crisis or in infancy with mild, non-specific symptoms. Gain-of-function alterations are clustered tightly in the PCNA domain where they have been shown in *in-vitro* expression models to disrupt PCNA binding and its subsequent ubiquitination.

The gene *NR0B1*, also known as *DAX1*, is situated at position Xp21.2. The gene encodes a nuclear receptor protein which plays a role in the development and physiological function of several endocrine tissues including the adrenal gland, the pituitary gland, the hypothalamus, and the gonads. Alterations in *NR0B1* are associated with X-linked adrenal hypoplasia congenita (AHC) a form of primary adrenal insufficiency resulting from the agenesis or underdevelopment of the adrenal gland [55, 56]. The condition classically presents in boys in early infancy or childhood with adrenal failure affecting both glucocorticoid and mineralocorticoid production. Absent or delayed pubertal development becomes apparent at adolescence due to combined hypothalamic and pituitary dysfunction. A subgroup of patients with AHC with precocious puberty have been described in whom a degree of sexual maturation may precede signs of adrenal insufficiency. Variants in *NR0B1* which are located in the ligand like binding region around the hydrophobic core and around the repression helix domain can result in late onset forms of AHC [57, 58]. It has also been observed that nonsense variants in the amino terminal region of the protein can be bypassed by translational initiation from codon 83, leading to a protein with some residual function resulting in late onset disease [59]. Deletions and duplications of Xp21.2 have been described in human disease. Deletions involving *NR0B1* lead to X-linked adrenal hypoplasia (AHC). When the neighbouring genes dystrophin and glycerol kinase are also involved contiguous gene deletion syndromes where features of AHC are compounded by the presence of muscular dystrophy, neurobehavioral features and the metabolic sequelae of glycerol kinase deficiency are seen [60]. Duplications encompassing *NR0B1* lead to 46,XY disorders of sexual development, indicating that *NR0B1* plays a dosage sensitive role in sex determination [61].

The gene *NR5A1* encodes the transcription factor SF1, which is involved in the development of the gonads and the adrenal glands [62]. Genetic alterations in SF1 cause a spectrum of reproductive phenotypes ranging from 46,XY disorders of sexual development and hypospadias to male factor infertility or primary ovarian insufficiency. Disruption of SF1 causing adrenal insufficiency has been described [63], but this is very rare.

Advances in Gene Discovery in Adrenal Insufficiency

The traditional approach to gene discovery in adrenal insufficiency centred on the sequencing of candidate genes implicated in adrenal development and/or function. Subsequent use of gene arrays and linkage analysis in informative families helped narrow down a region of interest in which selected candidate genes could then be studied. However, more recently, the increased accessibility of whole exome sequencing has led to the identification of several genes in novel pathways which may not have been implicated in adrenal insufficiency using candidate gene approaches.

An example of recent success includes the study of syndromic forms of adrenal hypoplasia, defined as MIRAGE syndrome (myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy), which has been shown to be caused by heterozygous variants in the gene *SAMD9* which is involved in endosome fusion [64, 65]. Although the majority of reported individuals with variants in *SAMD9* have primary adrenal insufficiency, the adrenal phenotype of MIRAGE syndrome is emerging as being more variable than early reports would suggest [66].

Another example is the discovery of sphingosine-1-phosphate (SGPL1), an intracellular and extracellular signalling molecule, which is involved in several fundamental cellular processes including differentiation, migration, and survival as well as in angiogenesis. SGPL1 is responsible for the breakdown of sphingosine-1-phosphate [67]. Recessive alterations in the *SGPL1* gene result in steroid-resistant nephrotic syndrome. Adrenal development and steroidogenesis are also commonly affected leading to adrenal insufficiency [68, 69]. Ichthyosis, primary hypothyroidism, hypogonadism, cryptorchidism, and immune and neurological abnormalities have also been described in these patients. Recessive alterations in *SGPL1* have recently been described in individuals with isolated adrenal insufficiency, and it remains to be seen whether these individuals will develop renal manifestations later in life [70].

Despite the list of genes associated with adrenal insufficiency continuing to grow, a significant proportion of cases of adrenal

insufficiency remain unsolved. This was illustrated by a recent study in which a molecular diagnosis was reached in approximately 80% of a large cohort of children with PAI recruited across multiple endocrine centres in Turkey [71]. As the challenge of novel gene discovery moves from the generation of sequence data to its accurate interpretation, it is important to appreciate that large scale databases of benign genetic variation may be subject to poor phenotyping for mild adrenal phenotypes. Furthermore, our understanding of how to use frequency data in the clinical interpretation of genetic variants is predominantly derived from models of dominant conditions in outbred populations, neither of which may be appropriate assumptions in primary adrenal insufficiency.

Conclusion

Our understanding of the genetic causes of adrenal insufficiency has greatly increased in recent years due to advances in sequencing techniques. Despite this, there are still a significant proportion of patients where causative mutations have not been found. Novel mechanisms of inheritance, genetic variation in non-coding portions of the genome as well as oligogenic and synergistic mechanisms are likely to contribute. Ultimately, future advances in our understanding of the genetics of adrenal insufficiency, and rare disease as a whole, are likely to depend heavily on the careful, precise and purposeful integration of patient cohorts/data and the coordinated mobilization of international resources and expertise to deal with the vast amount of data generated from advances in genetic technologies.

REFERENCES

1. Bornstein SR, Allolio B, Arlt W, *et al.* Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2016; **101**: 364–89.
2. Arlt W, Allolio B. Adrenal insufficiency. *Lancet*, 2003; **361**: 1881–93.
3. Bornstein SR. Predisposing factors for adrenal insufficiency. *N Engl J Med*, 2009; **360**: 2328–39.
4. Huszar D, Lynch CA, Fairchild-Huntress V, *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*, 1997; **88**: 131–41.
5. Raffin-Sanson ML, de Keyzer Y, Bertagna X. Proopiomelanocortin, a polypeptide precursor with multiple functions: from physiology to pathological conditions. *Eur J Endocrinol*, 2003; **149**: 79–90.
6. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Gruters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet*, 1998; **19**: 155–7.
7. Farooqi IS, Drop S, Clements A, *et al.* Heterozygosity for a POMC-null mutation and increased obesity risk in humans. *Diabetes*, 2006; **55**: 2549–53.
8. Clement K, Dubern B, Mencarelli M, *et al.* Unexpected endocrine features and normal pigmentation in a young adult patient carrying a novel homozygous mutation in the POMC gene. *J Clin Endocrinol Metab*, 2008; **93**: 4955–62.
9. Jackson RS, Creemers JW, Ohagi S, *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*, 1997; **16**: 303–6.
10. Couture C, Saveanu A, Barlier A, *et al.* Phenotypic homogeneity and genotypic variability in a large series of congenital isolated ACTH-deficiency patients with TPIT gene mutations. *J Clin Endocrinol Metab*, 2012; **97**: E486–95.
11. Lamolet B, Pulichino AM, Lamonerie T, *et al.* A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell*, 2001; **104**: 849–59.
12. Shepard TH, Landing BH, Mason DG. Familial Addison's disease; case reports of two sisters with corticoid deficiency unassociated with hypoaldosteronism. AMA *J Dis Child*, 1959; **97**: 154–62.
13. Thistlethwaite D, Darling JA, Fraser R, Mason PA, Rees LH, Harkness RA. Familial glucocorticoid deficiency. Studies of diagnosis and pathogenesis. *Arch Dis Child*, 1975; **50**: 291–7.
14. Kelch RP, Kaplan SL, Biglieri EG, Daniels GH, Epstein CJ, Grumbach MM. Hereditary adrenocortical unresponsiveness to adrenocorticotropic hormone. *J Pediatr*, 1972; **81**: 726–36.
15. Migeon CJ, Kenny EM, Kowarski A, *et al.* The syndrome of congenital adrenocortical unresponsiveness to ACTH. Report of six cases. *Pediatr Res*, 1968; **2**: 501–13.
16. Clark AJ, McLoughlin L, Grossman A. Familial glucocorticoid deficiency associated with point mutation in the adrenocorticotropin receptor. *Lancet*, 1993; **341**: 461–2.
17. Tsigos C, Arai K, Hung W, Chrousos GP. Hereditary isolated glucocorticoid deficiency is associated with abnormalities of the adrenocorticotropin receptor gene. *J Clin Invest*, 1993; **92**: 2458–61.
18. Mountjoy KG, Robbins LS, Mortrud MT, Cone RD. The cloning of a family of genes that encode the melanocortin receptors. *Science*, 1992; **257**: 1248–51.
19. Novoselova TV, Chan LF, Clark AJL. Pathophysiology of melanocortin receptors and their accessory proteins. *Best Pract Res Clin Endocrinol Metab*, 2018; **32**: 93–106.
20. Lin L, Hindmarsh PC, Metherell LA, *et al.* Severe loss-of-function mutations in the adrenocorticotropin receptor (ACTHR, MC2R) can be found in patients diagnosed with salt-losing adrenal hypoplasia. *Clin Endocrinol (Oxf)*, 2007; **66**: 205–10.
21. Chan LF, Metherell LA, Krude H, *et al.* Homozygous nonsense and frameshift mutations of the ACTH receptor in children with familial glucocorticoid deficiency (FGD) are not associated with long-term mineralocorticoid deficiency. *Clin Endocrinol (Oxf)*, 2009; **71**: 171–5.
22. Clark AJ, Weber A. Adrenocorticotropin insensitivity syndromes. *Endocr Rev*, 1998; **19**: 828–43.
23. Metherell LA, Chapple JP, Cooray S, *et al.* Mutations in MRAP, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2. *Nat Genet*, 2005; **37**: 166–70.
24. Chung TT, Chan LF, Metherell LA, Clark AJ. Phenotypic characteristics of familial glucocorticoid deficiency (FGD) type 1 and 2. *Clin Endocrinol (Oxf)*, 2010; **72**: 589–94.
25. Hughes CR, Chung TT, Habeb AM, Kelestimur F, Clark AJ, Metherell LA. Missense mutations in the melanocortin 2 receptor accessory protein that lead to late onset familial glucocorticoid deficiency type 2. *J Clin Endocrinol Metab*, 2010; **95**: 3497–501.
26. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*, 2011; **32**: 81–151.
27. New MI, Abraham M, Gonzalez B, *et al.* Genotype-phenotype correlation in 1,507 families with congenital adrenal hyperplasia

owing to 21-hydroxylase deficiency. *Proc Natl Acad Sci U S A*, 2013; **110**: 2611–16.
28. Carroll MC, Campbell RD, Porter RR. Mapping of steroid 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. *Proc Natl Acad Sci U S A*, 1985; **82**: 521–5.
29. White PC, Grossberger D, Onufer BJ, *et al*. Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. *Proc Natl Acad Sci U S A*, 1985; **82**: 1089–93.
30. Burch GH, Gong Y, Liu W, *et al*. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet*, 1997; **17**: 104–8.
31. Metherell LA, Naville D, Halaby G, *et al*. Nonclassic lipoid congenital adrenal hyperplasia masquerading as familial glucocorticoid deficiency. *J Clin Endocrinol Metab*, 2009; **94**: 3865–71.
32. Baker BY, Lin L, Kim CJ, *et al*. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*, 2006; **91**: 4781–5.
33. Fluck CE, Pandey AV, Dick B, *et al*. Characterization of novel StAR (steroidogenic acute regulatory protein) mutations causing non-classic lipoid adrenal hyperplasia. *PLoS One*, 2011; **6**: e20178.
34. Sahakitrungruang T, Tee MK, Blackett PR, Miller WL. Partial defect in the cholesterol side-chain cleavage enzyme P450scc (CYP11A1) resembling nonclassic congenital lipoid adrenal hyperplasia. *J Clin Endocrinol Metab*, 2011; **96**: 792–8.
35. Kil IS, Lee SK, Ryu KW, *et al*. Feedback control of adrenal steroidogenesis via H_2O_2-dependent, reversible inactivation of peroxiredoxin III in mitochondria. *Mol Cell*, 2012; **46**: 584–94.
36. Meimaridou E, Kowalczyk J, Guasti L, *et al*. Mutations in NNT encoding nicotinamide nucleotide transhydrogenase cause familial glucocorticoid deficiency. *Nat Genet*, 2012; **44**: 740–2.
37. Prasad R, Chan LF, Hughes CR, *et al*. Thioredoxin reductase 2 (TXNRD2) mutation associated with familial glucocorticoid deficiency (FGD). *J Clin Endocrinol Metab*, 2014; **99**: E1556–63.
38. Chan LF, Campbell DC, Novoselova TV, Clark AJ, Metherell LA. Whole-exome sequencing in the differential diagnosis of primary adrenal insufficiency in children. *Front Endocrinol (Lausanne)*, 2015; **6**: 113.
39. Prasad R, Metherell LA, Clark AJ, Storr HL. Deficiency of ALADIN impairs redox homeostasis in human adrenal cells and inhibits steroidogenesis. *Endocrinology*, 2013; **154**: 3209–18.
40. Ameis D, Brockmann G, Knoblich R, *et al*. A 5' splice-region mutation and a dinucleotide deletion in the lysosomal acid lipase gene in two patients with cholesteryl ester storage disease. *J Lipid Res*, 1995; **36**: 241–50.
41. Zschenker O, Jung N, Rethmeier J, *et al*. Characterization of lysosomal acid lipase mutations in the signal peptide and mature polypeptide region causing Wolman disease. *J Lipid Res*, 2001; **42**: 1033–40.
42. Browne M, Somers G, Savoia H, Kukuruzovic R. Wolman's disease in an infant. *Br J Haematol*, 2003; **122**: 522.
43. Stitziel NO, Fouchier SW, Sjouke B, *et al*. Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. *Arterioscler Thromb Vasc Biol*, 2013; **33**: 2909–14.
44. Cartier N, Sarde CO, Douar AM, Mosser J, Mandel JL, Aubourg P. Abnormal messenger RNA expression and a missense mutation in patients with X-linked adrenoleukodystrophy. *Hum Mol Genet*, 1993; **2**: 1949–51.
45. Braverman NE, Raymond GV, Rizzo WB, *et al*. Peroxisome biogenesis disorders in the Zellweger spectrum: an overview of current diagnosis, clinical manifestations, and treatment guidelines. *Mol Genet Metab*, 2016; **117**: 313–21.
46. Berendse K, Engelen M, Linthorst GE, van Trotsenburg AS, Poll-The BT. High prevalence of primary adrenal insufficiency in Zellweger spectrum disorders. *Orphanet J Rare Dis*, 2014; **9**: 133.
47. Poll-The BT, Gootjes J, Duran M, *et al*. Peroxisome biogenesis disorders with prolonged survival: phenotypic expression in a cohort of 31 patients. *Am J Med Genet A*, 2004; **126A**: 333–8.
48. Steinberg S, Chen L, Wei L, *et al*. The PEX gene screen: molecular diagnosis of peroxisome biogenesis disorders in the Zellweger syndrome spectrum. *Mol Genet Metab*, 2004; **83**: 252–63.
49. Collins CS, Gould SJ. Identification of a common PEX1 mutation in Zellweger syndrome. *Hum Mutat*, 1999; **14**: 45–53.
50. Hughes CR, Guasti L, Meimaridou E, *et al*. MCM4 mutation causes adrenal failure, short stature, and natural killer cell deficiency in humans. *J Clin Invest*, 2012; 122: 814–20.
51. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med*, 1990; **322**: 1829–36.
52. Finnish-German AC. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*, 1997; **17**: 399–403.
53. Arboleda VA, Lee H, Parnaik R, *et al*. Mutations in the PCNA-binding domain of CDKN1C cause IMAGe syndrome. *Nat Genet*, 2012; **44**: 788–92.
54. Vilain E, Le Merrer M, Lecointre C, *et al*. IMAGe, a new clinical association of intrauterine growth retardation, metaphyseal dysplasia, adrenal hypoplasia congenita, and genital anomalies. *J Clin Endocrinol Metab*, 1999; **84**: 4335–40.
55. Muscatelli F, Strom TM, Walker AP, *et al*. Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. *Nature*, 1994; **372**: 672–6.
56. Zanaria E, Muscatelli F, Bardoni B, *et al*. An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature*, 1994; **372**: 635–41.
57. Kyriakakis N, Shonibare T, Kyaw-Tun J, *et al*. Late-onset X-linked adrenal hypoplasia (DAX-1, NR0B1): two new adult-onset cases from a single center. *Pituitary*, 2017; **20**: 585–93.
58. Tabarin A, Achermann JC, Recan D, *et al*. A novel mutation in DAX1 causes delayed-onset adrenal insufficiency and incomplete hypogonadotropic hypogonadism. *J Clin Invest*, 2000; **105**: 321–8.
59. Ozisik G, Mantovani G, Achermann JC, *et al*. An alternate translation initiation site circumvents an amino-terminal DAX1 nonsense mutation leading to a mild form of X-linked adrenal hypoplasia congenita. *J Clin Endocrinol Metab*, 2003; **88**: 417–23.
60. Davies KE, Patterson MN, Kenwrick SJ, *et al*. Fine mapping of glycerol kinase deficiency and congenital adrenal hypoplasia within Xp21 on the short arm of the human X chromosome. *Am J Med Genet*, 1988; **29**: 557–64.
61. Barbaro M, Oscarson M, Schoumans J, Staaf J, Ivarsson SA, Wedell A. Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. *J Clin Endocrinol Metab*, 2007; **92**: 3305–13.

62. Achermann JC, Ito M, Ito M, Hindmarsh PC, Jameson JL. A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat Genet*, 1999; **22**: 125–6.
63. El-Khairi R, Achermann JC. Steroidogenic factor-1 and human disease. *Semin Reprod Med*, 2012; **30**: 374–81.
64. Buonocore F, Kuhnen P, Suntharalingham JP, *et al*. Somatic mutations and progressive monosomy modify SAMD9-related phenotypes in humans. *J Clin Invest*, 2017; **127**: 1700–13.
65. Narumi S, Amano N, Ishii T, *et al*. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet*, 2016; **48**: 792–7.
66. Shima H, Hayashi M, Tachibana T, *et al*. MIRAGE syndrome is a rare cause of 46,XY DSD born SGA without adrenal insufficiency. *PLoS One*, 2018; **13**: e0206184.
67. Reiss U, Oskouian B, Zhou J, *et al*. Sphingosine-phosphate lyase enhances stress-induced ceramide generation and apoptosis. *J Biol Chem*, 2004; **279**: 1281–90.
68. Prasad R, Hadjidemetriou I, Maharaj A, *et al*. Sphingosine-1-phosphate lyase mutations cause primary adrenal insufficiency and steroid-resistant nephrotic syndrome. *J Clin Invest*, 2017; **127**: 942–53.
69. Lovric S, Goncalves S, Gee HY, *et al*. Mutations in sphingosine-1-phosphate lyase cause nephrosis with ichthyosis and adrenal insufficiency. *J Clin Invest*, 2017; **127**: 912–28.
70. Settas N, Persky R, Faucz FR, *et al*. SGPL1 deficiency: a rare cause of primary adrenal insufficiency. *J Clin Endocrinol Metab*, 2018; **104**: 1484–90.
71. Guran T, Buonocore F, Saka N, *et al*. Rare causes of primary adrenal insufficiency: genetic and clinical characterization of a large nationwide cohort. *J Clin Endocrinol Metab*, 2016; **101**: 284–92.

5.8.2 Management of Adrenal Insufficiency

Wiebke Arlt

Introduction

In 1855, Thomas Addison identified a clinical syndrome characterized by wasting and hyperpigmentation as the result of adrenal gland destruction [1]. This landmark observation paved the way for progress in understanding and treating adrenal insufficiency, with the introduction of adrenal extracts for treatment of Addison's disease by the groups of Hartman and Pfiffner in 1929. However, long-term survival of patients with adrenal insufficiency only became possible after the seminal work of Edward Kendall, Philip Hench, and Tadeus Reichstein on the characterization and therapeutic use of cortisone. In 1946, Lewis Sarrett, a Merck scientist, achieved a partial synthesis of cortisone, which marked the beginning of industrial-scale production of cortisone. In 1948, in a fundamental clinical experiment at the Mayo Clinic, the first patient with Addison's received intravenous injections of Kendall's Compound E, cortisone, resulting in 'notable improvement of his condition'. This was followed by the ground-breaking trials on the use of cortisone in rheumatoid arthritis, yielding unanticipated clinical improvements, which quickly led to the labelling of cortisone as 'the wonder drug'. In November 1950, cortisone was made available to all physicians in the United States, a rapid translational development process, which culminated in the award of the 1950 Nobel Prize in Medicine to Kendall, Hench, and Reichstein. This progress reached other countries with variable delay and widespread availability of cortisone in the United Kingdom was achieved by joint efforts of Glaxo and the Medical Research Council. Though almost 150 years have passed since Addison's landmark observations and 60 years since the introduction of life-saving cortisone, there are still advances and challenges in the management of adrenal insufficiency, summarized in this chapter.

Physiology of Adrenal Steroid Synthesis

Adrenal Steroids and Steroidogenesis

When extracting steroids from the adrenal Kendall and Reichstein identified 28 separate steroids and today we classify the steroids produced by the adrenal glands, the corticosteroids, in three major classes—glucocorticoids (cortisol, corticosterone), mineralocorticoids (aldosterone, deoxycorticosterone), and adrenal sex steroid precursors (dehydroepiandrosterone (DHEA), androstenedione).

Cholesterol is the precursor for all adrenal steroidogenesis. The principal source of cholesterol is provided from the circulation in the form of low-density lipoprotein (LDL) cholesterol. Uptake is by specific cell-surface LDL receptors present on adrenal tissue; LDL is then internalized via receptor-mediated endocytosis, the resulting vesicles fuse with lysozymes, and free cholesterol is produced following hydrolysis. However, it is clear that this cannot be the sole source of adrenal cholesterol as patients with abetalipoproteinaemia, who have undetectable circulating LDL, and patients with defective LDL receptors in the setting of familial hypercholesterolaemia still have normal basal adrenal steroidogenesis. Cholesterol can be generated *de novo* within the adrenal cortex from acetyl coenzyme A. In addition, there is evidence that the adrenal can utilize high-density lipoprotein (HDL) cholesterol following uptake through the HDL receptor, scavenger receptor.

The biochemical pathways involved in adrenal steroidogenesis start with the rate-limiting step of the transport of intracellular cholesterol from the outer to the inner mitochondrial membrane. Within the mitochondrion cholesterol is then converted to pregnenolone by the cholesterol side chain cleavage enzyme, cytochrome P450scc (CYP11A1). The rapid transport of cholesterol into the mitochondria is importantly facilitated by steroidogenic acute regulatory protein, which is induced by an increase in intracellular cAMP following binding of adrenocorticotrophic hormone (ACTH) to its receptor.

Steroidogenesis involves the concerted action of several enzymes, including a series of cytochrome P450 (CYP) enzymes (for schematic overview, see Chapter 5.9). CYP11A1 and the CYP11B1 and CYP11B2 enzymes are localized to the mitochondria and require an electron shuttle system—provided through adrenodoxin/adrenodoxin reductase—for functional activity. Other CYP enzymes involved in steroidogenesis, namely 17α-hydroxylase (CYP17A1) and 21-hydroxylase (CYP21A2), are localized to the microsomal/endoplasmic reticulum fraction and depend on electron transfer from nicotinamide adenine dinucleotide phosphate (NADPH) via the electron donor enzyme P450 oxidoreductase (POR).

After the uptake of cholesterol to the mitochondrion, side chain cleavage of cholesterol by CYP11A1 forms pregnenolone, which is converted in the cytoplasm to progesterone by the type 2 isoenzyme of 3β-hydroxysteroid dehydrogenase. Progesterone is hydroxylated to 17-hydroxyprogesterone (17OHP) through the activity of 17α-hydroxylase. 17-hydroxylation is an essential prerequisite for glucocorticoid synthesis. CYP17A1 also possesses 17,20 lyase activity, which crucially facilitates the synthesis of the sex steroid precursor DHEA, a reaction that also requires allosteric interaction of the flavoprotein cytochrome b5 with both CYP17A1 and POR. In humans, 17OHP is not an efficient substrate for CYP17A1, and there is negligible conversion of 17-OH progesterone to androstenedione. Adrenal androstenedione secretion is dependent upon the conversion of dehydroepiandrosterone to androstenedione by 3β-hydroxysteroid dehydrogenase (3β-HSD). 21-hydroxylation of either progesterone (zona glomerulosa) or 17OHP (zona fasciculata) is carried out by 21-hydroxylase (CYP21A2) to yield deoxycorticosterone or 11-deoxycortisol, respectively. The final step in cortisol biosynthesis takes place in the mitochondria and involves the conversion of 11-deoxycortisol to cortisol by 11β-hydroxylase activity of the enzyme CYP11B1. In the zona glomerulosa, 11β-hydroxylase may also convert deoxycorticosterone to corticosterone. However, the enzyme CYP11B2, or aldosterone synthase, may also carry out this reaction and, in addition, is required for the conversion of corticosterone to aldosterone via the intermediate 18-OH corticosterone. Thus, CYP11B2 can carry out 11β-hydroxylation, 18-hydroxylation, and 18-methyl oxidation to yield the characteristic C11–18 hemiacetal structure of aldosterone.

Cortisol is inactivated to cortisone by action of the enzyme 11β-hydroxysteroid dehydrogenase type 2 (HSD11B2) mainly in the kidney, while the opposite reaction, activation of cortisone to cortisol, is carried out by the type 1 isoenzyme, HSD3B1, mainly in the liver (**Figure 5.8.2.1**). Recent years have highlighted the important role of this system in the tissue-specific activation and inactivation of glucocorticoids. Without the action of hepatic HSD11B1, Kendall would have observed no activity of his 'Compound E', as cortisone does not bind the glucocorticoid receptor and conversion to cortisol by HSD11B1 is a mandatory requirement for its biological activity.

Glucocorticoids are secreted in relatively high amounts (cortisol 10–20 mg/day) from the zona fasciculata, while mineralocorticoids are secreted in low amounts (aldosterone 100–150 μg/day) from the zona glomerulosa. The adrenal androgen precursors DHEA, its sulphate ester dehydroepiandrosterone sulphate (DHEAS), and androstenedione are produced in the adrenal zona reticularis and represent the most abundant steroids secreted by the adult adrenal gland (>20 mg/day). Recent work has shown that androstenedione is effectively converted to 11-hydroxyandrostenedione (11OHA4) by CYP11B1 and 11OHA4 is a major product of adrenal steroidogenesis, as measured in adrenal vein blood [2]. Adrenal zonal-specific steroidogenesis of the three different corticosteroid

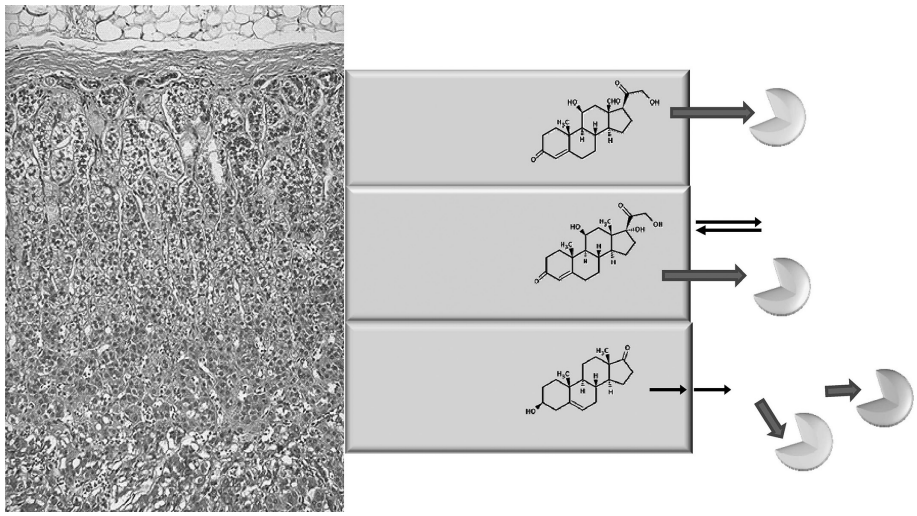


Figure 5.8.2.1 Schematic representation of adrenal zonation and steroidogenesis, depicting histology of the three adrenocortical and the major corticosteroids and the receptors mediating their action. While cortisol and aldosterone can bind and activate the glucocorticoid and mineralocorticoid receptor, respectively, DHEA requires conversion to active androgens and further aromatization to oestrogens prior to exerting sex steroid action. (See also Plate 29.)

classes is facilitated by tissue-specific expression of the required steroidogenic enzymes. The zona glomerulosa generally does not synthesize cortisol because it does not express 17α-hydroxylase. In contrast, aldosterone secretion is largely confined to the outer zona glomerulosa through the restricted expression of CYP11B2. Although CYP11B1 and CYP11B2 share 95% homology, the 5′ promoter sequences differ and permit regulation of the final steps in glucocorticoid and mineralocorticoid biosynthesis by ACTH and angiotensin II, respectively. DHEA is sulphated in the zona reticularis by the DHEA sulphotransferase (SULT2A1) to form DHEAS.

Regulation of Adrenal Corticosteroid Synthesis

Classical endocrine feedback loops are in place to control the secretion of both cortisol and aldosterone. Cortisol inhibits the secretion of both hypothalamic corticotrophin-releasing factor and pituitary ACTH, and the aldosterone-mediated increase in sodium retention inhibits renin secretion by the juxta-glomerular cells of the kidney (Figure 5.8.2.2).

Glucocorticoid synthesis is under negative feedback control of the hypothalamic–pituitary–adrenal (HPA) axis (Figure 5.8.2.2a). Adrenocorticotropic hormone (ACTH) secretion from the anterior pituitary is stimulated by hypothalamic corticotrophin-releasing hormone (CRH) following a circadian rhythm, with a peak around 03.00h to 04.00h. Other major effectors on CRH secretions are various forms of stress, including hypoglycaemia, hypotension, fever, trauma, and surgery. ACTH binds to its receptor (melanocortin receptor 2, MC2R) on the adrenocortical cell surface and stimulates import of cholesterol into the mitochondrion by steroidogenic acute regulatory protein. In parallel, transcription of genes encoding steroidogenic enzymes and proteins of the electron transfer shuttle is increased.

Mineralocorticoid synthesis is mainly under the control of the renin–angiotensin–aldosterone system (RAAS) and a potassium feedback loop (Figure 5.8.2.2b). A variety of factors stimulate renin secretion from renal juxtaglomerular cells, with renal perfusion being the most important regulator. Several other stimulators (β-adrenergic stimulation, prostaglandins) and inhibitors (α-adrenergic stimulation, dopamine, atrial natriuretic peptides, angiotensin II) are known. Angiotensinogen is an α_2-globulin synthesized within the liver which is cleaved by renin to form angiotensin I. Angiotensin I is converted to angiotensin II by angiotensin-converting enzyme in the lung and many other peripheral tissues. Angiotensin I has no apparent biological activity but angiotensin II is a potent stimulator of aldosterone secretion. In addition, angiotensin II acts as a potent vasoconstrictor. The rate-limiting step in the RAAS is the secretion of renin, which is also controlled through a negative feedback loop. Renin is secreted from juxtaglomerular epithelial cells within the macula densa of the renal tubule in response to underlying renal arteriolar pressure, oncotic pressure, and sympathetic drive. Thus, low perfusion pressure and/or low tubular fluid sodium content, as seen in haemorrhage, renal artery stenosis, dehydration, or salt loss, increase renin secretion. Conversely, secretion is suppressed following a high salt diet and by factors that increase blood pressure. Hypokalaemia increases and hyperkalaemia decreases renin secretion; in addition, potassium exerts a direct effect upon the adrenal cortex to increase aldosterone secretion. Angiotensin II and potassium stimulate aldosterone secretion principally by increasing the

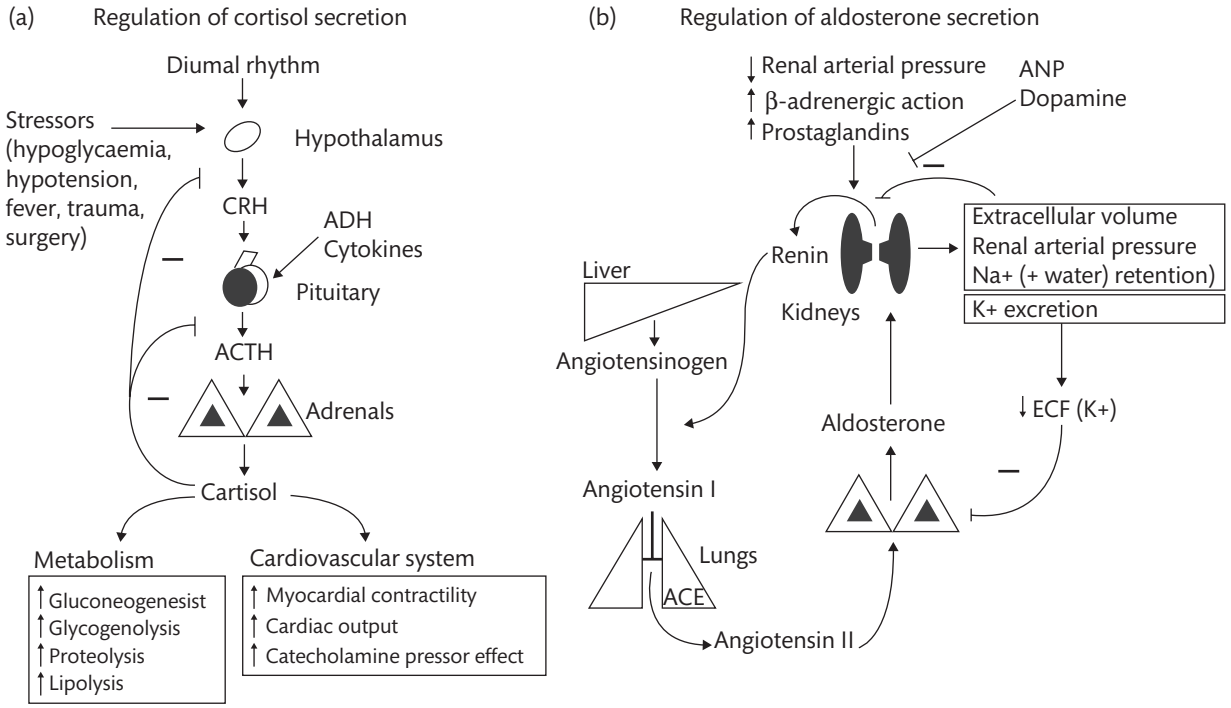


Figure 5.8.2.2 Negative feedback regulation of cortisol and aldosterone secretion. (a) Glucocorticoid feedback regulation by the hypothalamic–pituitary–adrenal (HPA) axis. CRH, corticotropin-releasing hormone; ACTH, adrenocorticotropic hormone; ADH, antidiuretic hormone.
(b) Mineralocorticoid regulation by the renin–angiotensin–aldosterone system (RAAS). The extracellular fraction (ECF) of potassium has an important direct influence on aldosterone secretion. ACE, angiotensin-converting enzyme; ANP, atrial natriuretic peptide.
Schematic graph: Dr Nils Krone, Birmingham.

transcription of *CYP11B2* through common intracellular signalling pathways. The potassium effect is mediated through membrane depolarization and opening of calcium channels, and the angiotensin II effect following binding of angiotensin II to the surface AT_1 receptor and activation of phospholipase C.

The separate control of glucocorticoid biosynthesis through the HPA axis and mineralocorticoid synthesis via the renin–angiotensin system has important clinical consequences. The overwhelming majority of patients with primary adrenal failure have both cortisol and aldosterone deficiency, whereas patients with ACTH deficiency due to pituitary disease have glucocorticoid deficiency only, as the intact renin–angiotensin system maintains normal aldosterone levels.

Corticosteroid Hormone Action

Both cortisol and aldosterone exert their effects following uptake of free hormone from the circulation and binding to intracellular receptors, termed the glucocorticoid and mineralocorticoid receptors (GR, MR). These are both members of the thyroid/steroid hormone–receptor superfamily of transcription factors, comprising a C-terminal ligand binding domain, a central DNA binding domain, interacting with specific DNA sequences on target genes, and an N-terminal hypervariable region. In both cases, although there is only a single gene encoding the GR and MR, splice variants have been described resulting in α and β variants.

The binding of glucocorticoid to the GR-α in the cytosol results in activation of the steroid–receptor complex through a process which involves the dissociation of heat-shock proteins HSP 90 and HSP 70. Following translocation to the nucleus, gene transcription is stimulated or repressed following binding of dimerized GR–ligand complexes to specific DNA sequences (glucocorticoid-response element) in the promoter regions of target genes. The GR-β variant may act as a dominant negative regulator of GR-α transactivation.

In contrast to the diverse actions of glucocorticoids, mineralocorticoids have a more restricted role, principally to stimulate epithelial sodium transport in the distal nephron, distal colon, and salivary glands. This is mediated through the induction of the apical sodium channel (comprising three subunits α, β, and γ) and the α_1 and β_1 subunits of the basolateral Na^+K^+-ATPase through transcriptional regulation of a specific aldosterone-induced gene that encodes serum and glucocorticoid-induced kinase. Aldosterone binds to the MR, principally in the cytosol (though there is evidence for expression of the unoccupied MR in the nucleus) followed by translocation of the hormone–receptor complex to the nucleus.

The MR and GR share considerable homology—57% in the steroid binding domain and 94% in the DNA binding domain. Therefore, it is perhaps not surprising therefore that there is promiscuity of ligand binding with aldosterone binding to the GR and cortisol binding to the MR. For the MR this is particularly impressive—*in vitro* the MR has the same inherent affinity for aldosterone, corticosterone, and cortisol. Specificity upon the MR is conferred through the 'prereceptor' metabolism of cortisol via the enzyme HSD11B2 in the kidney, which inactivates cortisol and corticosterone to 11-keto metabolites that do not activate the MR, thereby enabling aldosterone to bind to the MR.

For both glucocorticoids and mineralocorticoids, there is accumulating evidence for so-called non-genomic effects involving hormone response obviating the genomic GR or MR effects. A series of responses have been reported within seconds/minutes of exposure to corticosteroids and are thought to be mediated by, as yet uncharacterized, membrane-coupled receptors.

Cortisol-Binding Globulin and Corticosteroid Hormone Metabolism

Over 90% of circulating cortisol is bound, predominantly to the α_2-globulin cortisol-binding globulin (CBG). This 383-amino acid protein is synthesized in the liver and binds cortisol with high affinity. Affinity for synthetic corticosteroids (except prednisolone, which has an affinity for CBG of approximately 50% of that of cortisol) is negligible. Circulating CBG concentrations are approximately 700 nmol/L; levels are increased by oestrogens and in some patients with chronic active hepatitis but generally reduced in patients with cirrhosis, nephrosis, and hyperthyroidism. The oestrogen effect can be marked, with levels increasing two- to threefold across pregnancy, and this should also be taken into account when measuring plasma 'total' cortisol in pregnancy and in women taking oestrogens. Inherited abnormalities in CBG synthesis are much rarer than those described for thyroid-binding globulin but include patients with elevated CBG, partial and complete deficiency of CBG, or CBG variants with reduced affinity for cortisol. In each case, alterations in CBG concentrations change total circulating cortisol concentrations accordingly but 'free' cortisol concentrations are normal. Only this free circulating fraction is available for transport into tissues for biological activity. The excretion of 'free' cortisol through the kidneys is termed urinary-free cortisol and represents only 1% of the total cortisol secretion rate. Approximately 50% of secreted cortisol appears in the urine as tetrahydrocortisol (THF), 5alpha-tetrahydrocortisol (allo-THF), and tetrahydrocortisone (THE), 25% as cortols/cortolones, 10% as C19 steroids, and 10% as cortolic/cortolonic acids.

Aldosterone is also metabolized in the liver and kidneys. In the liver it undergoes tetrahydro reduction, followed by its excretion in the urine as a 3-glucuronide tetrahydroaldosterone derivative. However, glucuronide conjugation at the 18 position occurs directly in the kidney, as does 3α and 5α/5β metabolism of the free steroid. Because of the aldehyde group at the C18 position, aldosterone is not metabolized by HSD11B2. Hepatic aldosterone clearance is reduced in patients with cirrhosis, ascites, and severe congestive heart failure.

Epidemiology of Adrenal Insufficiency

The prevalence of Addison's disease, mostly due to primary adrenal failure due to autoimmune adrenalitis, is 93–140 per million while secondary insufficiency, mostly due to hypothalamic–pituitary tumours, has a prevalence of 125–280 per million [3]. The overall prevalence of adrenal insufficiency is 5 in 10 000 population; on average, three of those suffer from secondary adrenal insufficiency, one from primary adrenal insufficiency due to autoimmune adrenalitis, and one from congenital adrenal hyperplasia (see Chapter 5.9).

Primary Adrenal Insufficiency

According to recent studies, chronic primary adrenal insufficiency has a prevalence of 93 to 140 per million and an incidence of 4.7

to 6.2 per million in Caucasian populations [3, 4]. These numbers are considerably higher than reported earlier, despite a continuous decline in tuberculous adrenalitis in the developed world, and suggest an increasing incidence of autoimmune adrenalitis. The age at diagnosis peaks in the fourth decade of life, with women more frequently affected. For details on the rare inborn causes of primary adrenal insufficiency, please see Chapters 5.8.1 and 5.9.1.

Secondary Adrenal Insufficiency

Secondary adrenal insufficiency has an estimated prevalence of 150 to 280 per million [3, 4]. Again, women are more frequently affected and age at diagnosis peaks in the sixth decade.

It has been suggested that therapeutic glucocorticoid administration is the most common cause of adrenal insufficiency, as exogenous glucocorticoids induce atrophy of both pituitary corticotroph and adrenocortical cells. However, iatrogenic adrenal insufficiency only becomes potentially relevant during or after glucocorticoid withdrawal. As iatrogenic adrenal insufficiency is transient in the majority of cases it can be suspected that the prevalence of permanent iatrogenic adrenal insufficiency is clearly lower than that of endogenous adrenal insufficiency.

Causes of Adrenal Insufficiency

A large number of frequent and rare causes of adrenal insufficiency are summarized in Tables 5.8.2.1 and 5.8.2.2, and in the following sections more detailed information on some of the more frequent causes is provided. Two reviews have given an excellent overview of causes of adrenal insufficiency including citation of all original literature which cannot be provided here because of space constraints [3, 4].

Causes of Primary Adrenal Insufficiency

During the times of Thomas Addison, tuberculous adrenalitis was by far the most prevalent cause of adrenal insufficiency. In the developing world, tuberculosis still remains a major cause of adrenal insufficiency. In active tuberculosis, the incidence of adrenal involvement is 5%.

In North American and European countries, autoimmune adrenalitis accounts for more than 90% of cases with primary adrenal insufficiency; in 40% adrenal insufficiency is isolated while in 60% it arises as part of an autoimmune polyendocrine syndrome (APS) [3, 5]. APS type 1, also termed autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy, accounts for 15% of cases and is characterized by adrenal insufficiency, hypoparathyroidism, and chronic mucocutaneous candidiasis, the latter being the primary manifestation in most cases and already apparent in childhood [6]. APS 1 is caused by mutations in the autoimmune regulator gene (*AIRE*) [7–9] while APS 2 is thought to be inherited as a complex trait, associated with loci within the major histocompatibility complex [5] and distinct susceptibility genes [10–12]. APS 2 is much more common than APS 1 and in addition to adrenal insufficiency most frequently comprises

Table 5.8.2.1 Causes of primary adrenal insufficiency

Diagnosis	Clinical features	Pathogenesis/genetics
Autoimmune adrenalitis (AA)		
Isolated AA	AI	Associations with HLA-DR3, CTLA-4
AA as part of autoimmune polyendocrine syndromes (APS)		
APS 1 (=APECED) APS 2	AI + hypoparathyroidism + chronic mucocutaneous candidiasis ± other autoimmune disorders AI + thyroid disease (= Schmidt's syndrome) + type 1 diabetes mellitus (= Carpenter's syndrome) ± other autoimmune diseases	*AIRE* gene mutations (21q22.3) Associations with HLA-DR3, CTLA-4
Infectious adrenalitis		
Tuberculous adrenalitis	AI + other organ manifestations of tuberculosis	Tuberculosis
AIDS	AI +other AIDS-associated diseases	HIV, CMV
Fungal adrenalitis	AI + mostly immunosuppressed patients	Cryptococcosis, histoplasmosis, coccidioidomycosis
Genetic disorders leading to AI		
Adrenoleukodystrophy (ALD) Adrenomyeloneuropathy (AMN)	AI + demyelination of CNS (cerebral ALD) or spinal cord/peripheral nerves (AMN)	Mutation of the *X-ALD* gene encoding for the peroxisomal adrenoleukodystrophy protein (ALDP)
Congenital adrenal hyperplasia (CAH)		
21-hydroxylase deficiency 11β-hydroxylase deficiency 3β-HSD type 2 deficiency 17α-hydroxylase deficiency P450 oxidoreductase deficiency	AI + ambiguous genitalia in females AI + ambiguous genitalia in females + hypertension AI + ambiguous genitalia in males + postnatal virilization in females AI + ambiguous genitalia in males + lack of puberty in both sexes + hypertension AI + ambiguous genitalia in both sexes + skeletal malformations	*CYP21A2* mutation *CYP11B1* mutation *HSD3B2* mutation *CYP17A1* mutation *POR* mutation
Congenital lipoid adrenal hypoplasia (lipoid CAH)	AI + XY sex reversal	Mutations in the steroidogenic acute regulatory protein (*STAR*) gene Mutations in *CYP11A1* (encoding P450scc)
Smith–Lemli–Opitz syndrome (SLOS)	AI, mental retardation, craniofacial malformations, growth failure	Sterol delta-7-reductase gene (*DHCR7*) mutations

(*continued*)

Table 5.8.2.1 *Continued*

Diagnosis	Clinical features	Pathogenesis/genetics
Adrenal hypoplasia congenita (AHC)		
X-linked AHC Xp21 contiguous gene syndrome SF-1 linked AHC	AI + hypogonadotropic hypogonadism AI + Duchenne muscular dystrophy + glycerol kinase deficiency (psychomotor retardation) AI + XY sex reversal	Mutation in *NROB1* (encoding DAX1) Deletion of the Duchenne muscular dystrophy, glycerol kinase, and *DAX1* genes Mutation in *NR5A1* (encoding SF-1)
IMAGe syndrome	Intrauterine growth retardation + metaphyseal dysplasia + AI + genital anomalies	?
Kearns–Sayre syndrome	Progressive external ophthalmoplegia, pigmentary retinal degeneration and cardiac conduction defects; endocrinopathies include gonadal failure, hypoparathyroidism, type 1 diabetes, only rarely AI	Mitochondrial DNA deletions
ACTH insensitivity syndromes = familial glucocorticoid deficiency (FGD)	Glucocorticoid deficiency, excess plasma ACTH; no (or only very mild) impairment of mineralocorticoid synthesis; lack of adrenarche	
FGD 1 FGD 2 FGD 3 Triple A syndrome (=Allgrove's syndrome)	AI, tall stature AI AI AI + alacrimia + achalasia; additional symptoms (neurological impairment, deafness, mental retardation, hyperkeratosis)	Mutations in melanocortin-2-receptor (*MC2R*) encoding the ACTH receptor Mutations in MC2R accessory protein (*MRAP*) ? Mutations in the triple A gene (*AAAS*) encoding a WD repeat protein
Bilateral adrenal haemorrhage	AI + symptoms of underlying disease	Septic shock, specifically meningococcal sepsis (Waterhouse–Friderichsen syndrome) Primary antiphospholipid syndrome
Diagnosis	Clinical features	Pathogenesis/genetics
Adrenal infiltration	AI + symptoms of underlying disease	Adrenal metastases primary adrenal lymphoma sarcoidosis, amyloidosis, haemochromatosis
Bilateral adrenalectomy	AI + symptoms of underlying disease	e.g. in the management of Cushing's due to ectopic ACTH secretion of unknown source or following tumour nephrectomy
Drug-induced AI	AI	Treatment with mitotane, aminoglutethimide, arbiraterone, trilostane, etomidate, ketoconazole, suramin, RU486

AI, adrenal insufficiency; APECED, autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy.

autoimmune thyroid disease, albeit more often autoimmune hypothyroidism than Graves' disease.

X-linked adrenoleukodystrophy (ALD) is caused by a mutation in the *X-ALD* gene, which encodes a peroxisomal membrane protein (adrenoleukodystrophy protein), leading to accumulation of very long chain fatty acids (>24 carbon atoms). The clinical picture comprises adrenal insufficiency and neurological impairment due to white matter demyelination. The two major forms are cerebral ALD (50% of cases; early childhood manifestation, rapid progression) and adrenomyeloneuropathy (35% of cases; onset in early adulthood, slow progression) with restriction of demyelination to spinal cord and peripheral nerves. Adrenal insufficiency may precede the onset of neurological symptoms and is the sole manifestation of disease in 15% of cases.

Other causes of primary adrenal insufficiency (Table 5.8.2.1), e.g. adrenal infiltration or haemorrhage, are rare. Congenital or neonatal primary adrenal insufficiency accounts for only 1% of all cases. However, the recent elucidation of the genetic basis of underlying diseases has highlighted the importance of specific genes for adrenal development and steroidogenesis (see Chapter 5.8.1).

Causes of Secondary Adrenal Insufficiency

The most common cause of secondary adrenal insufficiency is a tumour of the hypothalamic–pituitary region, usually associated with panhypopituitarism as a result of tumour growth or treatment with surgery and/or irradiation (Table 5.8.2.2). Autoimmune lymphocytic hypophysitis is less frequent, mostly affecting women during or shortly after pregnancy. Isolated ACTH deficiency may also be of autoimmune origin as some patients concurrently suffer from other autoimmune disorders, most frequently thyroid disease. The differential diagnosis of postpartal autoimmune hypophysitis includes Sheehan's syndrome, which results from pituitary apoplexy, mostly due to pronounced blood loss during delivery. Very rarely mutations of genes important for pituitary development or for synthesis and processing of the corticotropin precursor proopiomelanocortin cause secondary adrenal insufficiency (Table 5.8.2.2).

Table 5.5.2.2 Causes of secondary adrenal insufficiency

Diagnosis	Comment
AI as the consequence of growth or therapeutic management of hypothalamic–pituitary mass lesions	
Pituitary tumours	Generally adenomas, carcinomas very rare Additional signs and symptoms consequent to impairment of other pituitary axes (thyroid, gonads, PRL, GH), visual field impairment due to compression of the optic chiasm
Other tumours of the hypothalamic–pituitary region	Craniopharyngioma, meningioma, ependymoma, intra-/suprasellar metastases
Pituitary irradiation	Radiation therapy for pituitary tumours, brain tumours outside the HPA axis and craniospinal irradiation in leukaemia and other cancers
Non-tumoural causes	
Lymphocytic hypophysitis isolated	Autoimmune hypophysitis; most frequently in relation to pregnancy; commonly associated with panhypopituitarism, but also presenting with isolated ACTH deficiency only
as part of autoimmune polyendocrine syndromes (APS)	associated with autoimmune thyroid disease, less frequently also with vitiligo, primary gonadal failure, type 1 diabetes, and pernicious anaemia
Genetic disorders leading to secondary AI	
Congenital isolated ACTH deficiency	Tpit or T-box 19 (*TBX19*) mutations; neonatal presentation; autosomal recessive
Combined pituitary hormone deficiency (CPHD)	Prophet of Pit-1 (*PROP1*) mutations: progressive development of CPHD in the order GH, PRL, TSH, LH/FSH, (ACTH– late onset); anterior pituitary may be hypoplastic, normal, or enlarged; autosomal recessive Homeobox gene 1 (*HESX1*) mutations: CPHD + optic nerve hypoplasia and midline brain defects/ agenesis of corpus callosum (=septo-optic dysplasia); anterior pituitary hypoplastic or ectopic; autosomal recessive, autosomal dominant Lim homeobox 3 (*LHX3*) mutations: CPHD with involvement of GH, TSH, gonadotrophins, PRLs; ACTH may be deficient; limited neck rotation, short cervical spine, sensorineural deafness; anterior pituitary hypoplastic, normal, or enlarged; autosomal recessive LIM homeobox 4 (*LHX4*) mutations: CPHD with involvement of GH, thyrotropin, and ACTH secretion, cerebellar abnormalities; anterior pituitary hypoplastic or ectopic; autosomal dominant SRY-box 3 (*SOX3*) mutations: infundibular hypoplasia, CPHD, variable: mental retardation
Proopiomelanocortin (POMC) deficiency syndrome	*POMC* gene mutations; clinical triad AI + early-onset obesity + red hair pigmentation
Prader–Willi syndrome	Imprinting disorder, manifests with AI, obesity, hypogonadism, variable learning difficulties, and hypotonia
Pituitary apoplexy—Sheehan's syndrome	Onset mainly with abrupt severe headache, visual disturbance, nausea/vomiting Pituitary apoplexy/necrosis with peripartal onset (e.g. due to high blood loss and/or hypotension)
Pituitary infiltration/ granuloma	Tuberculosis, actinomycosis, sarcoidosis, histiocytosis X, Wegener's granulomatosis
Trauma	Pituitary stalk lesions, traumatic brain injury
Drugs	Chronic glucocorticoid excess: exogenous glucocorticoid administration for more than 4 weeks endogenous glucocorticoid hypersecretion due to Cushing's syndrome

GH, growth hormone; LH/FSH, luteinizing hormone/follicle-stimulating hormone; PRL, prolactin; TSH, thyroid-stimulating hormone.

Clinical Presentation of Adrenal Insufficiency

The clinical signs and symptoms of both acute and chronic adrenal insufficiency are a logical consequence of the underlying pathology, i.e. mostly the deficiency of adrenal corticosteroid production arising from primary or secondary adrenal failure (Table 5.8.2.3).

Acute adrenal insufficiency (i.e. life-threatening adrenal crisis) typically presents with severe hypotension or hypovolaemic shock, acute abdominal pain, vomiting, and often with fever, and, therefore, is sometimes mistaken for acute abdomen. In a series of 91 patients with Addison's disease, adrenal crisis led to the initial diagnosis of adrenal insufficiency in half of the patients. In children, acute adrenal insufficiency often presents as hypoglycaemic seizures. Deterioration of glycaemic control with recurrent hypoglycaemia may be the presenting sign of adrenal insufficiency in patients with pre-existing type 1 diabetes. In APS 2, onset of autoimmune hyperthyroidism (or thyroxine replacement for newly diagnosed hypothyroidism) may precipitate adrenal crisis due to enhanced cortisol clearance.

The leading symptom of chronic adrenal insufficiency is fatigue, accompanied by lack of stamina, loss of energy, reduced muscle strength, and increased irritability. In addition, chronic glucocorticoid deficiency leads to weight loss, nausea, and anorexia (in children, failure to thrive) and may account for muscle and joint pain. Unfortunately, most of these symptoms are non-specific. Thus, every second patient suffers from signs and symptoms of Addison's disease for more than 1 year before diagnosis is established. In secondary adrenal insufficiency, diagnosis is mostly prompted by a history of pituitary disease, but may also be delayed (e.g. in isolated ACTH deficiency). A more specific sign for primary adrenal failure is hyperpigmentation (Figure 5.8.2.3), which is most pronounced in areas of the skin exposed to increased friction (e.g. hand lines, knuckles, scars, oral mucosa). Hyperpigmentation is due to enhanced stimulation of skin MC1-receptor by ACTH and other pro-opiomelanocortin-related peptides. Accordingly, patients with secondary adrenal insufficiency often present with pale, alabaster-coloured skin. Laboratory

Table 5.8.2.3 Clinical manifestations of adrenal insufficiency

Manifestations	Explained by deficiency of
Symptoms	
Fatigue, lack of energy/stamina, reduced strength	Glucocorticoids (adrenal androgens)
Anorexia, weight loss (in children: failure to thrive)	Glucocorticoids
Abdominal pain, nausea, vomiting (more frequent in primary AI)	Mineralocorticoids, glucocorticoids
Myalgia, joint pain	Glucocorticoids
Dizziness, postural hypotension	Mineralocorticoids
Salt craving (primary AI only)	Mineralocorticoids
Dry and itchy skin (in women)	Adrenal androgens
Loss/impairment of libido (in women)	Adrenal androgens
Signs	
Skin hyperpigmentation (primary AI only)	*Excess* of pro-opiomelanocortin (POMC) derived peptides (primary AI)
Alabaster-coloured pale skin (secondary AI only)	Deficiency of POMC derived peptides (secondary AI)
Loss of axillary/pubic hair (in women)	Adrenal androgens
Fever	Glucocorticoids
Low blood pressure (systolic RR <100 mm Hg), postural hypotension (pronounced in primary AI)	Mineralocorticoids, glucocorticoids
Anaemia, lymphocytosis, eosinophilia	Glucocorticoids
Serum creatinine ↑ (primary AI only)	Mineralocorticoids
Hyponatraemia	Mineralocorticoids, (glucocorticoids = SIADH)
hyperkalaemia (primary AI only)	mineralocorticoids
TSH ↑ (primary AI only)	Glucocorticoids (or autoimmune hypothyroidism)
Hypercalcaemia (primary AI only)	Glucocorticoids (rare, mostly observed if concurrent hyperthyroidism)
Hypoglycaemia	Glucocorticoids, (epinephrine deficiency?) (more frequent in children)

SIADH, syndrome of inappropriate antidiuretic hormone secretion; TSH, thyroid-stimulating hormone.

findings in glucocorticoid deficiency may include mild anaemia, lymphocytosis, and eosinophilia. Cortisol physiologically inhibits thyrotropin (TSH) release. Thus, TSH is often increased at initial diagnosis of primary adrenal insufficiency, but returns to normal during glucocorticoid replacement, unless there is coincident autoimmune thyroid failure. In rare cases, glucocorticoid deficiency may result in hypercalcaemia, which is due to increased intestinal absorption and decreased renal excretion of calcium and usually coincides with autoimmune hyperthyroidism, facilitating calcium release from bone.

Mineralocorticoid deficiency, which is only present in primary adrenal insufficiency, leads to dehydration and hypovolaemia, resulting in low blood pressure, postural hypotension, and sometimes even in prerenal failure. Deterioration may be sudden and is often due to exogenous stress such as infection or trauma. Combined mineralocorticoid and glucocorticoid replacement in primary adrenal insufficiency reconstitutes the diurnal rhythm of blood pressure and reverses cardiac dysfunction. Glucocorticoids contribute to this amelioration not only by mineralocorticoid receptor binding, but also by permissive effects on catecholamine action. The latter may account for the relative unresponsiveness to catecholamines in patients with unrecognized adrenal crisis. Mineralocorticoid deficiency accounts for hyponatraemia (90%), hyperkalaemia (65%), and salt craving (15%). Low serum sodium may also be present in secondary adrenal insufficiency due to the syndrome of inappropriate antidiuretic hormone secretion, which results from the loss of physiological inhibition of pituitary vasopressin release by glucocorticoids.

Adrenal insufficiency inevitably leads to DHEA deficiency. DHEA is the major precursor of sex steroid synthesis and loss of its synthesis results in pronounced androgen deficiency in women. As a consequence, women with adrenal insufficiency frequently show loss of axillary and pubic hair (absence of pubarche in children), dry skin, and reduced libido. DHEA also exerts direct action as a neurosteroid with potential antidepressant properties. Thus, DHEA deficiency may contribute to the impairment of well-being that is observed in patients with adrenal insufficiency despite adequate glucocorticoid and mineralocorticoid replacement.

Diagnostic Laboratory Evaluation of Adrenal Insufficiency

Presentation with acute adrenal insufficiency (i.e. life-threatening adrenal crisis), requires an immediate, combined diagnostic and therapeutic approach (Figure 5.8.2.4). Haemodynamically stable patients may undergo a cosyntropin stimulation test (also termed short synacthen test, as synacthen, i.e. $ACTH_{1-24}$, is used, rather than the full-length $ACTH_{1-39}$ peptide. In a severely unwell patient, baseline bloods for serum cortisol and plasma ACTH will suffice and if cortisol is less than 100 nmol/L while ACTH is considerably elevated, there is no doubt about the diagnosis of primary adrenal insufficiency. Formal confirmation of diagnosis can be performed via the short synacthen test following clinical improvement. Diagnostic measures must never delay treatment of a suspected adrenal crisis; therefore, treatment should be initiated upon strong clinical suspicion of adrenal insufficiency without awaiting biochemical results. It is of negligible risk to start hydrocortisone and stop it after adrenal insufficiency has been safely excluded; withholding potentially life-saving treatment, however, could have fatal consequences.

Adrenal insufficiency is readily diagnosed by the short synacthen test, a safe and reliable diagnostic tool with excellent long-term predictive value [13, 14]; it is important to be aware of the considerable variability between results of different cortisol assays [15] and when defining the cut-off for failure, commonly set at 500 nmol/L but with the use of newer, mass spectrometry-based assays now often at 400–450 nmol/L, one should ideally refer to results from a local reference cohort obtained with the same assay. The diagnostic value of the short synacthen test is compromised within the first 4 weeks following a pituitary insult [14, 16], as during this period the adrenals

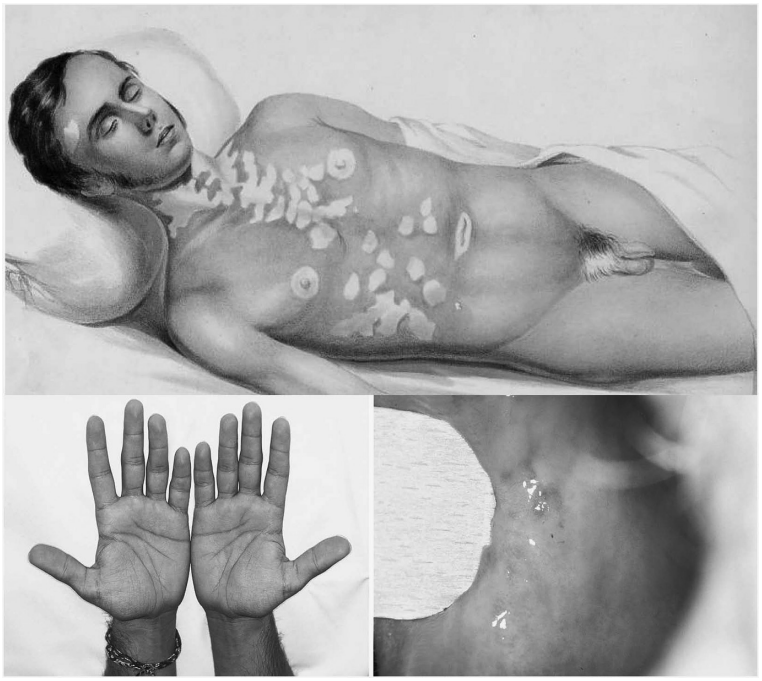


Figure 5.8.2.3 Skin changes observed in primary adrenal insufficiency (Addison's disease). (a) Panel drawn by Thomas Addison (1855) of a patient with Addison's disease, depicting generalized hyperpigmentation, in particular in areas of increased friction, and patchy vitiligo, indicative of autoimmune polyendocrine syndrome. (b) Hyperpigmentation of the palmar creases in a patient with acute primary adrenal insufficiency. (c) Patchy hyperpigmentation of the oral mucosa in a patient with acute primary adrenal insufficiency. (See also Plate 30.)

will still respond to exogenous ACTH stimulation despite the loss of endogenous ACTH drive. When suspecting secondary adrenal insufficiency, the insulin tolerance test is an alternative choice for diagnostic confirmation, considered by many as the gold standard, however it is associated with side effects and requires exclusion of cardiovascular disease and history of seizures. Again, hydrocortisone treatment could be initiated, with formal confirmation of the diagnosis of secondary adrenal insufficiency once the patient is clinically better and time since a suspected pituitary insult is more than 4 weeks. Formal confirmation of diagnosis by the short synacthen stimulation test should include blood samples for plasma ACTH, which will guide the way for further diagnostic assessment, by reliably differentiating primary from secondary adrenal insufficiency, i.e. adrenal from hypothalamic–pituitary disease (Figure 5.8.2.4).

Possible glucocorticoid deficiency is also indicated by normocytic anaemia as sufficient levels of cortisol are required for maturation of blood progenitor cells; other blood count changes may include lymphocytosis and eosinophilia. Sometimes mild metabolic acidosis or hypercalcaemia can be observed in affected patients, the latter mostly in the context of coincident hyperthyroidism. Serum glucose may be low; however, significant hypoglycaemia as a presenting sign plays a more important role in childhood adrenal insufficiency where it can result in significant brain damage. However, in a patient with pre-existing type 1 diabetes onset of recurrent hypoglycaemic episodes despite unchanged insulin regimen should raise the suspicion of adrenal insufficiency.

Mineralocorticoid deficiency is present in primary adrenal insufficiency only; the renin–angiotensin–aldosterone system in patients with hypothalamic–pituitary disease and intact adrenals is usually preserved. Mineralocorticoid deficiency is not only reflected by the arterial hypotension and deranged potassium and sodium but intravascular volume depletion is also indicated by the slightly raised creatinine, a common finding in Addison patients. Hyponatraemia is observed in about 80% of acute cases while less than half present with hyperkalaemia. In the first instance, baseline bloods for serum aldosterone and plasma renin should be taken.

Diagnosis of Primary Adrenal Insufficiency

The combined measurement of early morning serum cortisol and plasma ACTH separates patients with primary adrenal insufficiency from normal subjects and patients with secondary adrenal insufficiency. Plasma ACTH is usually grossly elevated and invariably higher than 22 pmol/L, with baseline serum cortisol usually below the lower end of the reference range (<150 nmol/L), sometimes also in the lower reference range. Establishment of the diagnosis of primary adrenal insufficiency always depends on the **combined** measurement of ACTH and cortisol at the same time. Serum aldosterone concentrations are subnormal or within the lower normal range with plasma renin activity concurrently increased above the normal range. In patients with adrenal insufficiency, serum DHEAS is invariably low, in women often below the limit of detection.

The impaired ability of the adrenal cortex to respond to ACTH is readily demonstrated by the short synacthen test (SST), employing serum cortisol measurements before and 30 min after IV injection of 250 µg $ACTH_{1-24}$ (= synacthen). In some areas, IM injections are used and others use a 60 min timepoint for analysing the adequacy of the cortisol response. However, there are no studies showing superiority of either administration or timepoint; hence, IV injection and the 30 min interval are generally preferred, as more convenient for the patients. In normal subjects, this leads to a physiological increase in serum cortisol to peak concentrations above 500 nmol/L. In primary adrenal insufficiency, the adrenal cortex is already

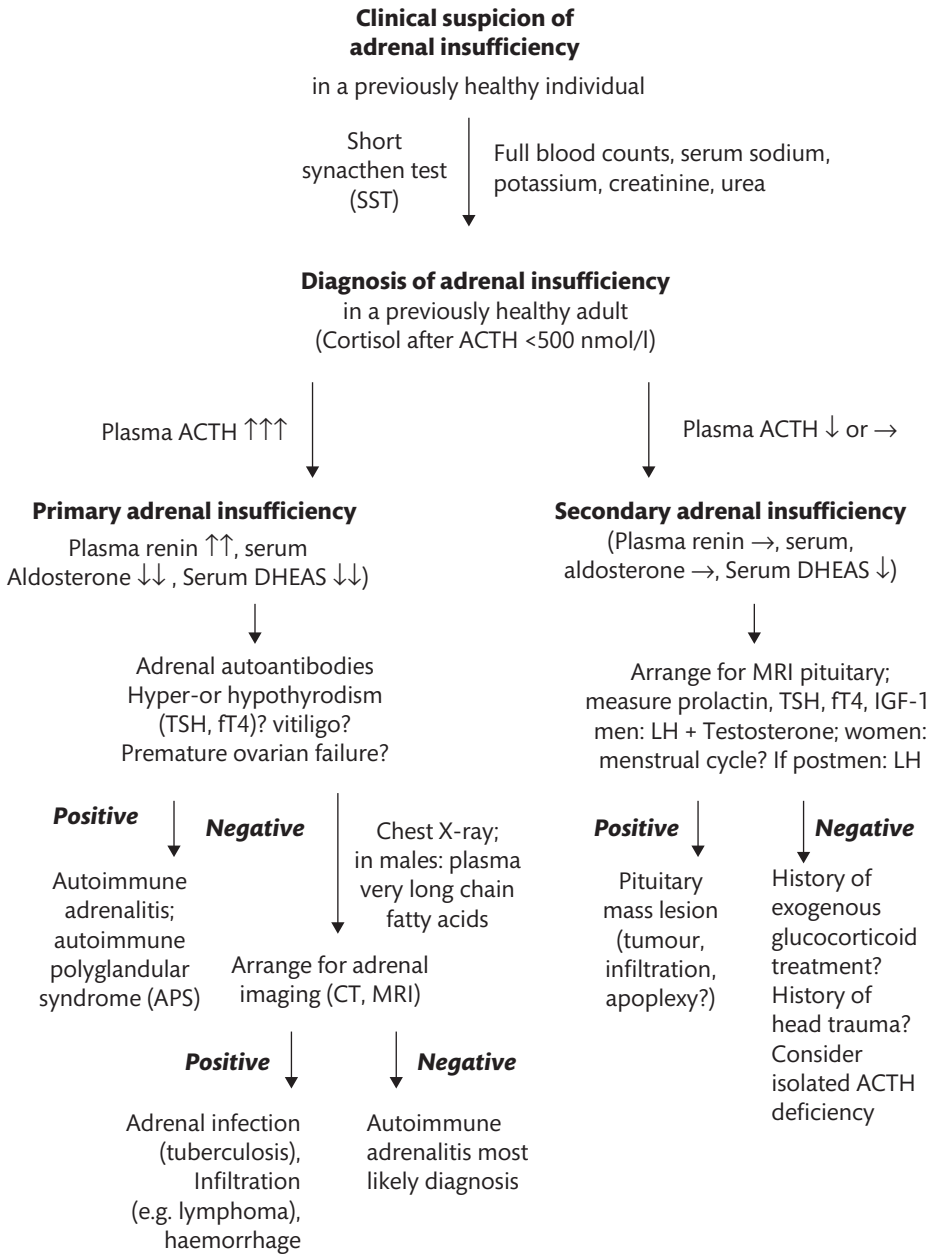


Figure 5.8.2.4 Flowchart outlining the steps to be taken for the diagnostic management of adults with newly diagnosed adrenal insufficiency. DHEAS, dehydroepiandrosterone sulphate ester; IGF-1, insulin-like growth factor 1; LH, luteinizing hormone; TSH, thyroid-stimulating hormone.

maximally stimulated by endogenous ACTH, exogenous ACTH administration therefore usually does not evoke any further increase in serum cortisol.

Adrenal cortex autoantibodies and/or antibodies against 21-hydroxylase are found in more than 80% of patients with recent-onset autoimmune adrenalitis. While 21-hydroxylase has been identified as the major autoantigen in autoimmune adrenalitis, autoantibodies against other steroidogenic enzymes (CYP11A1, CYP17A1) and steroid-producing cell antibodies are present in a lower percentage of patients. Measurement of autoantibodies is particularly helpful in patients with isolated primary adrenal insufficiency and no family history of autoimmune disease. In APS 2, autoimmune adrenalitis may be associated with autoimmune thyroid disease or type 1 diabetes and screening for concomitant disease should involve measurements of thyrotropin and fasting glucose but not of other organ-related antibodies. (For a state-of-the-art review of adrenal insufficiency in the context of APS, please see reference [17]).

In male patients with isolated primary adrenal insufficiency without unequivocal evidence of autoimmune adrenalitis, serum very long chain fatty acids (chain length of 24 carbons and more; C26, C26/C22, and C24/C22 ratios) should be measured to exclude adrenoleukodystrophy/adrenomyeloneuropathy.

Diagnosis of Secondary Adrenal Insufficiency

Baseline hormone measurements only poorly separate patients with secondary adrenal insufficiency from normal subjects. However, a morning cortisol below 100 nmol/L indicates adrenal insufficiency,

whereas a serum cortisol greater than 500 mmol/L is consistent with an intact HPA axis, unless the patient is clinically severely unwell and treated in the ICU. Thus, in most cases, dynamic tests of the HPA axis are required to establish the diagnosis of secondary adrenal insufficiency.

In addition to the widely used SST, the insulin tolerance test (ITT) is still regarded as a 'gold standard' in the evaluation of suspected secondary adrenal insufficiency, as hypoglycaemia (blood glucose <2.2 mmol/L) is a powerful stressor resulting in rapid activation of the HPA axis. An intact HPA axis is demonstrated by a peak cortisol above 500 nmol/L at any time during the test. The occasional patient will pass the ITT while exhibiting clinical evidence for adrenal insufficiency responding to hydrocortisone substitution and a higher cut-off value (550 nmol/L) may help to reduce misclassification. During ITT close supervision is mandatory and cardiovascular disease and history of seizures represent contraindications.

Another test for the diagnosis of secondary adrenal insufficiency, which is no longer widely used, is the overnight metyrapone test (30 mg metyrapone/kg (maximum 3 g) with a snack at midnight). Metyrapone inhibits the adrenal 11β-hydroxylase enzyme CYP11B1, thereby blocking the conversion of 11-deoxycortisol to cortisol. In normal subjects, HPA feedback activation will increase serum 11-deoxycortisol, while serum cortisol remains less than 230 nmol/L. In patients with secondary adrenal insufficiency, 11-deoxycortisol does not exceed 200 nmol/L at 8.00 hours after metyrapone. Shortcomings of the test are limited availability of reliable 11-deoxycortisol assays and of the drug itself, which cannot be obtained in all countries though it is readily available in the United Kingdom. As metyrapone may precipitate adrenal crisis in severe cortisol deficiency, a morning cortisol above 200 nmol/L should be documented prior to performing the test on an outpatient basis.

As both the ITT and the metyrapone test pose a significant burden to patients and physicians, there have been continuing efforts to replace these tests by more convenient tools. Sustained secondary adrenal insufficiency leads to adrenal atrophy and also to reduced adrenal ACTH receptor expression, as ACTH up-regulates its own receptor. Thus adrenal responsiveness to an acute exogenous ACTH challenge is impaired also in secondary adrenal insufficiency facilitating the use of the SST for the evaluation of HPA axis integrity (Figure 5.5.2.4). Several studies have reported excellent agreement between peak cortisol values in the short synacthen test and the ITT [14] and a definitive study has convincingly demonstrated the long-term predictive accuracy of the SST [13]. However, there is some evidence that some patients with secondary adrenal insufficiency will pass the SST while failing the ITT. The use of a higher cut-off value for passing the SST may minimize the risk of overlooking secondary adrenal insufficiency but this is largely assay dependent as different radioimmunoassays and now more widely mass spectrometry-based assays have different cut-offs [18] and it is now more important than ever to ascertain locally derived reference ranges with the cortisol assay used.

There are other tests that have been used in the diagnostic assessment of adrenal insufficiency but their use is not recommended as a routine procedure. As the administration of 250 μg 1–24 ACTH represents a massive supraphysiological challenge, a low-dose corticotropin test (LDT) employing only 1 μg ACTH has been proposed as a more sensitive test for the diagnosis of secondary adrenal insufficiency. The LDT has been successfully used to monitor recovery of adrenal function after withdrawal of oral glucocorticoids and to detect subtle impairment of adrenal reserve during inhalative steroid therapy. However, the administration of 1 μg ACTH IV still results in ACTH levels above those required for maximum cortisol release. Accordingly, in normal subjects serum cortisol concentrations measured 30 min after the ACTH challenge do not differ between the 250 μg SST and LDT. Thus it is currently a matter of debate whether employing the LDT represents any advantage [19, 20], which would be further offset by handling problems due to the necessity of dilution from the commercially available 250 μg ACTH 1–24 ampoule and due to the potential binding of ACTH to the surface of injection devices.

CRH has been used to differentiate hypothalamic from pituitary disease in secondary adrenal insufficiency. However, CRH stimulation is not of great help in actually diagnosing secondary adrenal insufficiency as individual responses to exogenous CRH are highly variable and cut-off values or even normal ranges are still not well defined.

Finally, a word of caution: none of the tests, including the ITT, will classify all patients correctly. Mild secondary adrenal insufficiency may pass as intact HPA axis and some healthy subjects may fail any single test by a small margin. Thus clinical judgement remains important. Persisting symptoms such as fatigue, myalgia, or reduced vitality should lead to reassessment.

Special Diagnostic Situations

Adrenal Insufficiency After Pituitary Surgery

Screening for adrenal insufficiency by SST should not be performed immediately after pituitary surgery, but only 4 to 6 weeks later, as adrenal atrophy may develop only gradually after onset of ACTH deficiency. Until then, patients with a morning cortisol not excluding secondary adrenal insufficiency (<450 nmol/L at 3 days and <350 nmol/L at 7 days after surgery, with cut-offs varying according to locally used cortisol assays) should receive hydrocortisone replacement paused 24 h prior to scheduled adrenal function testing. The impairment of other hormonal axes after pituitary surgery increases the likelihood of ACTH deficiency, whereas isolated corticotropin deficiency is uncommon.

Adrenal Insufficiency in Critically Ill Patients

In critically ill patients, the corticotropic axis is markedly activated [21, 22]. Moreover, patients in intensive care units are less sensitive to dexamethasone suppression and achieve higher peak ACTH and cortisol concentrations after CRH. In addition, patients with critical illness show relatively low serum aldosterone levels with concurrently elevated plasma renin activity. Cortisol concentrations correlate with illness-severity scores and are highest in patients with the highest mortality. On the other hand, cytokine activation may induce relative secondary adrenal insufficiency in some patients with severe illness, thus putting them at risk of dying from adrenal crisis. Chronic inhibition of cortisol production by etomidate has been associated with increased mortality in intensive care unit patients. Unfortunately, no consensus exists how to diagnose adrenal insufficiency in critically ill patients. In patients with primary adrenal insufficiency or severe secondary adrenal insufficiency the SST will establish the diagnosis by demonstrating

a low baseline cortisol (<150 nmol/L) not responding to corticotropin (peak cortisol <500 nmol/L). However, it has been suggested that 'relative' adrenal insufficiency may be present in a number of critically ill patients, characterized by a poor cortisol response (increment <248 nmol/L) to ACTH despite normal baseline cortisol. These patients often present with catecholamine-dependent hypodynamic shock responding to hydrocortisone administration. One study has reported decreased mortality in patients with septic shock and abnormal cortisol response in the SST (increment <248 nmol/L) after treatment with hydrocortisone [23] but a prospective study did not support this finding [24]. At present it seems prudent to collect a random sample of serum cortisol and plasma ACTH in critically ill patients with suspected adrenal insufficiency followed by immediate hydrocortisone administration. Depending on the results of these hormone determinations (it is agreed that serum cortisol >700 nmol/L rules out adrenal insufficiency) hydrocortisone therapy is terminated or a more detailed evaluation employing the SST is performed. For more details, please review the recently published, updated consensus statement of the American and European critical care societies [25].

Imaging Requirements in Adrenal Insufficiency

If there is no coexisting autoimmune disease and adrenal and steroid autoantibodies are negative, imaging of the adrenals, preferably by computed tomography (CT), is warranted (Figure 5.8.2.4). Tuberculosis should be considered, which is frequent in developing countries and, therefore, also in migrant populations. Chest radiography is helpful and imaging of the adrenals typically shows hyperplastic organs in the early phase and spotty calcifications in the late phase of tuberculous adrenalitis. Much rarer causes are bilateral infiltration by bilateral primary adrenal lymphoma, (predominantly lung cancer) metastases [26, 27], sarcoidosis, haemochromatosis, or amyloidosis. Bilateral adrenal haemorrhage is usually only seen during septic shock or in very rare instances in primary antiphospholipid syndrome [28]. In male patients with isolated Addison's and negative autoantibodies, imaging should be preceded by measurement of plasma very long chain fatty acids to safely exclude X-linked adrenoleukodystrophy which affects 1 in 20 000 males [29]. *ABCD1* gene mutations encoding for the peroxisomal ALD protein involved in cross-membrane transport manifest in 50% of cases in early childhood and primarily with central nervous system (CNS) symptoms. However, the adrenomyeloneuropathy variant, accounting for 35% of cases, can manifest with adrenal insufficiency prior to the development of spinal paraparesis during early adulthood [29].

If ACTH is inappropriately low in the presence of cortisol deficiency, imaging of the hypothalamic–pituitary region by MRI is the first diagnostic measure that should be arranged for, alongside an endocrine pituitary baseline profile (Figure 5.8.2.4). Pituitary adenomas are most common, craniopharyngiomas are much rarer and may present at any age; very rare causes include meningioma, metastases, and infiltration by sarcoidosis, Langerhans' cell histiocytosis, or other granulomatous disease. Careful history-taking should ask for previous head trauma [30, 31], surgery, radiotherapy, and for clinical indicators of pituitary apoplexy [32], i.e. the sudden onset of high-impact headache [33]. The latter may occur spontaneously in larger pituitary adenomas or may result from sudden hypocirculation during surgery or as a consequence of complicated deliveries with significant blood loss, the classical cause of Sheehan's syndrome. Lymphocytic hypophysitis of autoimmune origin [34] commonly presents with panhypopituitarism including diabetes insipidus and a pituitary mass effect. However, it may present with isolated ACTH deficiency, in some cases coinciding with autoimmune thyroid disease [35, 36].

Importantly, the most obvious should not be forgotten— suppression of the hypothalamic–pituitary axis by exogenous glucocorticoid treatment. This should always be excluded, considering not only oral steroid intake but also glucocorticoid inhalers, creams, or intra-articular injections.

Treatment of Adrenal Insufficiency

Glucocorticoid Replacement

A patient with suspected acute adrenal insufficiency needs immediate therapeutic attention. Symptoms such as fatigue and weight are non-specific; more specific signs and symptoms, and therefore more suggestive of adrenal insufficiency include generalized hyperpigmentation including patchy hyperpigmentation of the oral mucosa (sometimes the only visible hyperpigmentation in dark-skinned individuals) and the presence of severe hypovolaemic hypotension. If peripheral veins are collapsed, a central line for IV fluid resuscitation may be required, administered at an initial rate of 1 L/h and with continuous cardiac monitoring. In addition, hydrocortisone replacement should be immediately commenced upon suspicion of acute adrenal insufficiency, by intravenous bolus injection of 100 mg hydrocortisone followed by continuous infusion of 200 mg hydrocortisone per 24 h [37]. Mineralocorticoid replacement does not need to be added in the acute setting as long as the total daily hydrocortisone dose is greater than 50 mg, as such a dose will ensure sufficient mineralocorticoid receptor activation by cortisol.

Chronic glucocorticoid replacement requires additional considerations. Physiological daily cortisol production rates vary between 5 and 10 mg/m^2 [38], which is equivalent to the oral administration of 15 to 25 mg hydrocortisone (i.e. cortisol). After oral ingestion cortisol produces highly variable peak concentrations within the supraphysiological range followed by a rapid decline to below 100 nmol/L 5 to 7 h after ingestion (Figure 5.8.2.5). It is recommended to administer hydrocortisone in two to three divided doses (e.g. 15 mg in the morning upon awakening followed by 5 mg 6 h later, or 10 mg upon awakening followed by 5 mg 4 h and 8 h later). It is important to let the patient experiment with different timings to find the most suitable regimen for their individual needs. Importantly, patients who work shifts have to adjust the timing of the glucocorticoid doses to their working times and consequent sleep–wake cycle. Whether a thrice daily glucocorticoid regimen should be preferred over twice daily administration is not clear as well-designed and appropriately powered studies are lacking. As Figure 5.8.2.5 illustrates, neither of the two regimens will be able to achieve cortisol availability similar to that of physiological diurnal secretion. Some groups advocate weight-related dosing [39]

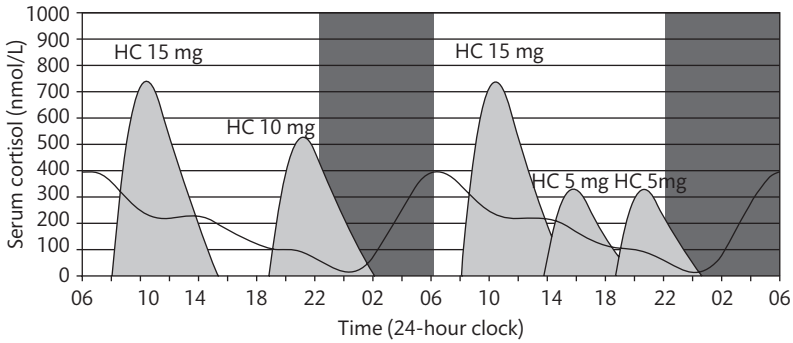


Figure 5.8.2.5 Schematic graph depicting the physiological diurnal rhythm of cortisol secretion and typical mean serum cortisol concentrations observed after different doses of oral hydrocortisone (HC) in patients with adrenal insufficiency.

and this appears to generate a smoother pharmacokinetic profile but data demonstrating superiority of such a regimen are lacking. However, body surface area adjusted glucocorticoid dosing is commonly used for guiding glucocorticoid replacement in children.

The oral administration of currently available cortisol preparations is not able to mimic the physiological pattern of cortisol secretion, which follows a distinct circadian rhythm. Physiologically, cortisol secretion begins to rise between 02.00 and 04.00 hours, peaks within an hour of waking and then declines gradually to low levels during the evening and nadir levels at and after midnight [40]. There is evidence for a diurnal variability in glucocorticoid sensitivity. Plat *et al.* [41] have demonstrated that a more unfavourable metabolic response occurs to evening administration of glucocorticoids. Also, high levels of glucocorticoids may disrupt sleep, thus late evening hydrocortisone administration should be avoided; sleep disturbances contributing to increased fatigue are a common feature in chronic adrenal insufficiency [42, 43]. The delivery of cortisol by intravenous infusion [44] or subcutaneous pump [45] can closely mirror diurnal secretion, but these administration modes are obviously not suited for routine delivery. Recently developed modified and delayed release hydrocortisone preparations mimicking physiological cortisol secretion represent a very promising therapeutic approach [46, 47].

Cortisone acetate requires intrahepatic activation to cortisol by HSD11B1, which contributes to a higher pharmacokinetic variability compared to hydrocortisone; 25 mg cortisone acetate are equivalent to 15 mg hydrocortisone [48, 49]. Long-acting glucocorticoids are also used for replacement (e.g. in 20% of respondents to the 2002 survey of the North American Addison Disease Foundation). Some countries do not have access to hydrocortisone or cortisone acetate and therefore have to resort to long-acting synthetic glucocorticoids. However, prednisolone and dexamethasone have considerably longer biological half-lives, likely to result in unfavourably high night-time glucocorticoid activity with potentially detrimental effects on insulin sensitivity and bone mineral density [50]. In addition, available preparations offer limited options for dose titration. Therefore, synthetic glucocorticoids should generally be avoided in the replacement therapy for adrenal insufficiency; the only exception are patients with concurrent insulin-dependent diabetes in whom the use of longer-acting prednisolone may help to avoid the peaks and troughs of hydrocortisone pharmacokinetics and thus also subsequent rapid changes in glucose control. For clinical purposes, one can assume equipotency to 1 mg hydrocortisone for 1.6 mg cortisone acetate, 0.2 mg prednisolone, 0.25 mg prednisone, and 0.025 mg dexamethasone, respectively. While equipotency doses of hydrocortisone and cortisone acetate are based on pharmacokinetic studies [48, 49], suggested doses for synthetic steroids are based on estimates from very old studies comparing the relative anti-inflammatory properties of various glucocorticoids based on rodent *in vivo* experiments.

Recent work has highlighted the potential of modified and delayed release hydrocortisone in delivering cortisol to mimic its physiological, diurnal secretion pattern in patients with congenital adrenal hyperplasia [50], with the outcomes of a phase 3 study currently awaited. Another hydrocortisone preparation, which is already approved by regulators, offers delayed release, which might be a good option in particular for elderly patients with secondary and partial adrenal insufficiency, who could then rely on the intake of a single glucocorticoid morning dose; however, current studies have been undertaken exclusively in primary rather than secondary adrenal insufficiency patients [51, 52]. Another option for modified hydrocortisone delivery to yield physiological, diurnal cortisol delivery is subcutaneous hydrocortisone administration via pump, for which encouraging results have been described [53, 54]. This option is particularly suitable for patients with impaired gastrointestinal absorption of hydrocortisone, but obviously require much more intense monitoring than oral hydrocortisone replacement.

Monitoring of glucocorticoid replacement is mainly based on clinical grounds as a reliable biomarker for glucocorticoid activity has yet to be identified (Table 5.8.2.4). Plasma ACTH cannot be used as a criterion for glucocorticoid dose adjustment. In primary adrenal insufficiency, ACTH is invariably high before the morning dose and rapidly declines with increasing cortisol levels after glucocorticoid ingestion [55, 56] (Figure 5.8.2.6a). Therefore, aiming at ACTH levels within the normal range would result in overreplacement. In secondary adrenal insufficiency, plasma ACTH is anyway low and thus not informative. Urinary 24-h free cortisol excretion has been advocated for monitoring of replacement quality by some [57]. However, after exogenous glucocorticoid administration, urinary cortisol excretion shows considerable interindividual variability. Also, following glucocorticoid absorption by the gut, CBG is rapidly saturated, resulting in transient but pronounced increases in renal cortisol excretion. Thus, one cannot refer to normal ranges for healthy subjects when judging urinary cortisol excretion during replacement therapy for adrenal insufficiency. Some authors have suggested regular measurements of

Table 5.8.2.4 Treatment and monitoring in chronic adrenal insufficiency

Chronic adrenal insufficiency	
Glucocorticoid replacement	Primary adrenal insufficiency: 20–25 mg hydrocortisone per 24 h
	Secondary adrenal insufficiency: 15–20 mg hydrocortisone per 24 h; if borderline fail in cosyntropin test consider 10 mg or stress dose cover only
	Administer in 2–3 divided doses with two-thirds and half of the dose, respectively, administered immediately after awakening
	Monitoring: • Check body weight, calculate body mass index • Check for signs of under-replacement (weight loss, fatigue, nausea, myalgia, lack of energy) • Check for signs of overreplacement (weight gain, central obesity, stretch marks, osteopenia/osteoporosis, impaired glucose tolerance, hypertension) • Take a detailed account of stress-related glucocorticoid dose self-adjustments since last visit, potential adverse events including emergency treatment and/or hospitalization
Mineralocorticoid replacement	Only required in primary adrenal insufficiency
	Not required as long as hydrocortisone dose >50 mg per 24 h
	Start on 100 µg fludrocortisone (doses vary between 50 and 250 µg per 24 h) administered as a single dose in the morning immediately after waking up
	Monitoring: • Blood pressure sitting and erect (postural drop ≥15 mm Hg indicative of under-replacement, high blood pressure may indicate overreplacement) • Check for peripheral oedema (indicative of overreplacement) • Check serum sodium and potassium • Check plasma renin activity (annually, upon clinical suspicion of over- and under-replacement and after significant changes in the hydrocortisone dose (40 mg hydrocortisone = 100 µg fludrocortisone)
Adrenal androgen replacement	Consider in patients with impaired well-being and mood despite apparently optimized glucocorticoid and mineralocorticoid replacement and in women with symptoms and signs of androgen deficiency (dry, itchy skin; reduced libido)
	DHEA 25–50 mg as a single morning dose
	If no perceived benefit after 6 months, consider stopping
	Monitoring: • In women, serum testosterone and SHBG (to calculate free androgen index) • In people on DHEA replacement, serum DHEAS and androstenedione levels • Blood should be sampled at steady state, i.e. ideally 24 h after the preceding DHEA dose
Additional monitoring requirements	Regular follow-up in specialist centre every 6–12 months
	In primary adrenal insufficiency of autoimmune origin (isolated Addison or autoimmune polyendocrine syndrome) serum TSH every 12 months
	In female patients: check regularity of menstrual cycle, consider measurement of ovarian autoantibodies if family planning not finalized
	Check emergency bracelet/steroid card, update as required
	Check knowledge of 'sick day rules' and reinforce emergency guidelines involving partner/family members
	Consider prescription of a hydrocortisone emergency self-injection kit, in particular if delayed access to acute medical care is likely (rural areas, travel)
	Check if other medication includes drugs known to induce (e.g. rifampicin, mitotane, anticonvulsants such as phenytoin, carbamazepine, oxcarbazepine, phenobarbital, topiramate) or inhibit (e.g. antiretroviral agents) hepatic cortisol inactivation by CYP3A4, which may require glucocorticoid dose adjustment

DHEA, dehydroepiandrosterone; SHBG, sex hormone-binding hormone; TSH, thyroid-stimulating hormone.

serum cortisol day curves to monitor replacement therapy [57, 58]. However, the efficacy of this approach is not supported by data that indicate a poor correlation between clinical assessment and cortisol levels [59]. Timed serum cortisol measurements can be of some value in selected patients (e.g. in case of suspected non-compliance or gastrointestinal malabsorption); however, random serum cortisol measurements without information on the timing of the hydrocortisone dose intake and without referring to a locally generated time course reference range are not informative.

Thus, in the absence of objective parameters, the physician has to rely primarily on clinical judgement, carefully taking into account signs and symptoms potentially suggestive of glucocorticoid over- or under-replacement, recognizing their relative lack of specificity. Glucocorticoid under-replacement bears the risk of incipient crisis and significant impairment of well-being. Conversely, chronic overreplacement may lead to substantial morbidity including impaired glucose tolerance, obesity, and osteoporosis. An increased incidence of osteoporosis has only been reported in patients receiving daily replacement doses of 30 mg hydrocortisone or higher [60–62] or 7.5 mg prednisone [60] whereas appropriate replacement doses of 15–25 mg hydrocortisone do not affect bone mineral density [59, 63]. Therefore, bone mineral density measurements are not routinely required in patients with adrenal insufficiency receiving recommended glucocorticoid replacement doses.

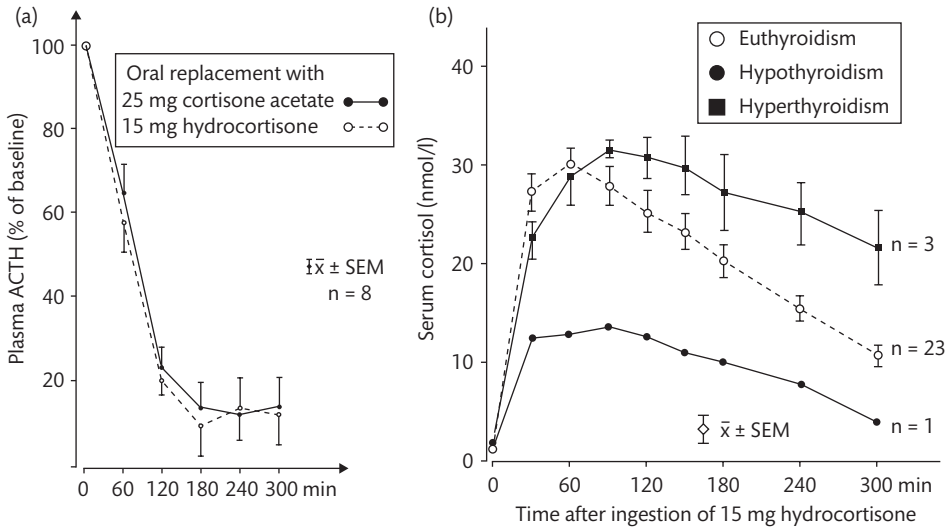


Figure 5.8.2.6 (a) Plasma ACTH concentrations before and after administration of the hydrocortisone morning dose in patients with primary adrenal insufficiency (n = 8). (b) Serum cortisol and thyroid function. Serum cortisol concentrations after administration of 15 mg hydrocortisone orally in 27 patients with primary adrenal insufficiency. Patients with concurrent overt hypothyroidism (n = 3) or hyperthyroidism (n = 1) differ from euthyroid patients (n = 23), which has to be considered when choosing appropriate glucocorticoid replacement doses.

Modified with permission from Allolio B, Kaulen D, Deuss U, Hipp FX, Winkelmann W. Comparison between hydrocortisone and cortisone acetate as replacement therapy in adrenocortical insufficiency. *Akt Endokr Stoffw*, 1985: 6: 35–9.

Mineralocorticoid Replacement

Patients with primary adrenal insufficiency require mineralocorticoid replacement, which usually consists of the oral administration of 9α-fludrocortisone; fluorination at the 9α position ensures selective binding to the MR and thus exclusive mineralocorticoid action. By contrast, cortisol binds with equal affinity to both the GR and MR. However, excessive MR binding of cortisol in the kidney is prevented by 11β-hydroxysteroid dehydrogenase type 2, which inactivates cortisol to cortisone. Oelkers has coined the term 'mineralocorticoid unit' (MCU), determining that 100 MCU are equivalent to 100 µg fludrocortisone and 40 mg hydrocortisone, respectively [64]. By contrast, prednisolone exerts only reduced mineralocorticoid activity and dexamethasone none at all; therefore, patients treated with synthetic glucocorticoids need particularly careful monitoring of their mineralocorticoid replacement requirements.

In the newly diagnosed patient, mineralocorticoid replacement should be initiated at 100 µg once daily; optimized doses may vary between 50 and 250 µg. Children, in particular neonates and infants, have considerably higher mineralocorticoid dose requirements and often need additional salt supplementation. However, also among adults there is a good degree of interindividual variability. A high dietary salt intake may slightly reduce mineralocorticoid requirements. An important additional factor is temperature and humidity (e.g. individuals living in Mediterranean summer or tropical climates will require a 50% increase in fludrocortisone dose due to increased salt loss through perspiration). Monitoring (Table 5.8.2.4) includes supine and erect blood pressure and serum sodium and potassium; plasma renin activity should be checked regularly, aiming at the upper normal range [64]. If essential hypertension develops, mineralocorticoid dose may be slightly reduced, accompanied by monitoring of serum sodium and potassium, but complete cessation of mineralocorticoid replacement should be avoided. It is important to recognize that plasma renin physiologically increases during pregnancy; therefore, monitoring in pregnancy should not rely on renin measurement, but comprise blood pressure, serum sodium and potassium, and, if required, urinary sodium excretion. During the last term of pregnancy fludrocortisone dose may require adjustment, also due to increased progesterone levels exerting antimineralocorticoid activity [65].

Prevention of Adrenal Crisis

Risk of adrenal crisis is higher in primary adrenal insufficiency and several factors such as coincident APS or age have been suggested as additional modifiers [2, 66]. Many crises are due to glucocorticoid dose reduction or lack of stress-related glucocorticoid dose adjustment by patients or general practitioners [3]. A survey in 526 patients found that 42% of patients (47% in primary adrenal insufficiency, 35% in secondary adrenal insufficiency) had experienced at least one adrenal crisis during the course of their disease. Precipitating causes were mainly gastrointestinal infections and fever but also several other causes, including major pain, surgery, psychological distress, heat, and pregnancy. This was corroborated by data from a large patient survey (n = 841) [67] that also highlighted gastrointestinal infections as the single most important cause of crisis. Importantly, a prospective study in 423 patients with primary and secondary adrenal insufficiency, respectively, showed a high incidence of adrenal crises, occurring at a rate of 8.3 crises and 0.5 deaths per 100 patient-years [68]. Thus, adrenal crisis is a predictable and frequent, but still undermanaged event that regularly leads to loss of life [69]. Crisis prevention and better education of patients and healthcare professionals is a key strategy that needs to be pursued to avoid morbidity and mortality. Currently, prompt initiation of glucocorticoid therapy is delayed in nearly half of adrenal insufficiency experiencing an adrenal crisis [70].

All patients and their partners should receive regular crisis prevention training, including verification of steroid emergency card/

bracelet and instruction on stress-related glucocorticoid dose adjustment (Table 5.8.2.4). Generally, hydrocortisone should be doubled during intercurrent illness, such as a respiratory infection with fever, until clinical recovery. Gastrointestinal infections, a frequent cause of crisis, may require parenteral hydrocortisone administration,. Preferably all patients, but at least patients travelling or living in areas with limited access to acute medical care should receive a hydrocortisone emergency self-injection kit (e.g. 100 mg for IM injection). In case of serious illness and trauma, major surgery, or if intensive care unit treatment is required, patients should receive major stress dose cover with an immediate bolus injection of 100 mg hydrocortisone, followed by 200 mg hydrocortisone per 24 h via perfuser; if IV access is lost, patients should receive 50 mg hydrocortisone per intramuscular injection immediately and every 6 hours until IV access is re-established [37, 71]. This recommendation is based on recently published experimental evidence [72] that has clarified the exact dose requirements for the prevention of adrenal crisis for the first time since the seminal observation that glucocorticoid replacement needs to be increased during periods of major stress to avoid adrenal crisis and death [73]. Recent work has also highlighted the feasibility of subcutaneous hydrocortisone emergency injections [74], albeit pharmacokinetic data were obtained in unstressed adrenal insufficiency patients, while in impending shock subcutaneous might be hypoperfused, possibly preventing good uptake of hydrocortisone; however, subcutaneous self-administration would certainly be better than no self-administration of hydrocortisone at all. The development of a 'Corti-Pen' (as an Epi-Pen equivalent for adrenal insufficiency) or similar easy-to-use devices or delivery modes for hydrocortisone self-administration would certainly be game changing, but are still awaited.

DHEA Replacement

The introduction of DHEA, the third major steroid produced by the adrenal gland, into the replacement regimen for adrenal insufficiency [75] represents a major advance, in particular for women who are invariably androgen deficient [75, 76]. DHEA has been shown to significantly enhance well-being, mood, and subjective health status in women with primary and secondary adrenal insufficiency [75, 77–80] and also recently in children and adolescents with adrenal failure [81]. Similar effects have been described for testosterone replacement in hypopituitarism [82], however, no study has yet directly compared DHEA to testosterone. In addition to acting as an androgen precursor, DHEA has neurosteroidal properties, exerting a primarily antidepressive effect, and also shows immunomodulatory properties [83]. Of note, DHEA has been shown to exert beneficial effects on subjective health status and energy levels not only in women but also in men with primary adrenal insufficiency [78, 79] including significant beneficial effects on bone mineral density and truncal lean mass [78].

Currently, DHEA replacement is hampered by the lack of pharmaceutically controlled preparations, with questionable quality and content of several over-the-counter preparations [84]. At present, DHEA should be reserved for patients with adrenal insufficiency suffering from significant impairment in well-being despite otherwise optimized replacement, in particular women with signs of androgen deficiency such as dry and itchy skin and loss of libido. DHEA should be taken as a single dose (25–50 mg) in the morning. Treatment monitoring (Table 5.8.2.4) should include blood sampling 24 h after the last preceding morning dose for measurement of serum DHEAS (in women also androstenedione, testosterone, sex hormone-binding hormone) aiming at the middle normal range for healthy young subjects. I usually start patients on 25 mg and increase to 50 mg after 2 to 4 weeks, advising them to halve the dose if androgenic skin side effects (greasy skin, spots) persist for more than a week. Transdermal testosterone represents an alternative androgen replacement tool in women with adrenal failure.

Special Therapeutic Situations Impacting on Corticosteroid Replacement

Thyroid Dysfunction

Hyperthyroidism results in increased cortisol metabolism and clearance and hypothyroidism the converse, principally due to an effect of thyroid hormone upon hepatic 11β-HSD1 and 5α/5β-reductases. Insulin-like growth factor 1 (IGF-1) increases cortisol clearance by inhibiting hepatic HSD11B1 (conversion of cortisone to cortisol). In patients with adrenal insufficiency and unresolved hyperthyroidism, glucocorticoid replacement should be doubled to tripled. To avoid adrenal crisis, thyroxine replacement for hypothyroidism should only be initiated after concomitant glucocorticoid deficiency has either been excluded or treated. Obviously overt endogenous hyperthyroidism will also increase hydrocortisone metabolism (Figure 5.8.2.6b). Therefore, the initiation of glucocorticoid replacement in patients with newly diagnosed hypopituitarism should always precede the initiation of thyroxine replacement as the reverse might precipitate adrenal crisis.

Pregnancy

Pregnancy is physiologically associated with a gradual increase in CBG and during the last term of pregnancy also with an increase in free cortisol. In addition, serum progesterone increases, exerting antimineralocorticoid action. Therefore, during the third trimester, hydrocortisone replacement should be increased by 50%. Plasma renin activity cannot serve as a monitoring tool because it physiologically increases during pregnancy. Peripartal hydrocortisone replacement should follow the requirements for major surgery, with 100 mg hydrocortisone per bolus injection at the onset of active labour, followed by administration of 200 mg hydrocortisone per 24 h [37]. After delivery, the dose can be immediately tapered back to normal replacement doses, unless there are clinical complications.

Concomitant Drug Therapy and Interactions

When deciding on the glucocorticoid dose, it is important to consider concurrent medication, in particular drugs known to increase hepatic glucocorticoid metabolism by CYP3A4 induction, which results in increased 6β-hydroxylation and hence cortisol inactivation [3, 4]. 6β-hydroxylation by CYP3A4 is normally a minor pathway but cortisol itself induces CYP3A4 so that 6β-hydroxycortisol excretion is markedly increased in patients with Cushing's syndrome. A multitude of drugs are known to induce CYP3A4 (Table 5.8.2.1), which require a two- to threefold increase in glucocorticoid dose. Conversely, the intake of drugs inhibiting CYP3A4 would require reduction of glucocorticoid replacement dose.

Of note, treatment of tuberculosis with rifampicin increases cortisol clearance but does not influence aldosterone clearance. Thus, glucocorticoid replacement should be doubled during rifampicin treatment. Mitotane (o,p′DDD, *ortho, para′*, dichlorodiphenyldichloroethane) decreases bioavailable glucocorticoid levels due to an increase in CBG and concurrently enhanced glucocorticoid metabolism following induction of CYP3A4. During chronic mitotane treatment, e.g. in adrenal carcinoma [85], usual glucocorticoid replacement doses should therefore be at least tripled.

Quality of Life, Disablement, and Prognosis in Adrenal Insufficiency

Studies demonstrate that current standard corticosteroid replacement fails to restore quality of life, which is significantly impaired in both patients with primary and secondary adrenal insufficiency [42, 86], with no apparent difference between prednisolone and hydrocortisone-treated patients [87]. Predominant complaints are fatigue, lack of energy, depression, anxiety, and reduced ability to cope with daily demands; the degree of impairment is comparable to that observed in congestive heart failure and chronic haemodialysis patients [42, 86]. Subjective health status is most reduced in younger patients but all age groups are significantly impaired [86], a persistent finding even if only analysing patients without any comorbidity [86]. This also has a socioeconomic perspective as patients with Addison's disease have a two- to threefold higher likelihood of receiving disablement pensions [42, 86].

In addition, large cohort studies have demonstrated an increased mortality not only in patients with secondary adrenal insufficiency due to hypopituitarism [88] but also in primary adrenal insufficiency, i.e. Addison's disease [89, 90], a finding still valid when the influence of comorbidities is excluded. The causes underlying this increased mortality remain unclear. However, a recent study demonstrated that patients on conventional, immediate release hydrocortisone replacement have considerably impaired natural killer cell function [91], an immune cell critically involved in protection from respiratory infections. Another study [92, 93] indicated that natural killer cell numbers increase after switching adrenal insufficiency patients from conventional to modified-release hydrocortisone. We certainly need to consider the possible impact of current replacement regimens on the observed increase in mortality from cardiovascular and cerebrovascular disease and respiratory infections.

Conclusions

More than 150 years after Thomas Addison first described a disease characterized by salt wasting and hyperpigmentation as the result of adrenal gland destruction [1], adrenal insufficiency is no longer an invariably fatal condition. The landmark achievement of the synthesis of cortisone in the late 1940s and its introduction into therapy in the early 1950s quickly lead to widespread availability of life-saving glucocorticoid replacement therapy. However, while initial survival is routinely achieved nowadays, current replacement regimens may not be able to achieve normal quality of life and the reliable prevention of adrenal crisis as a cause for the observed increased mortality is an important issue. Future research has to uncover the causes underlying the increased mortality in adrenal insufficiency and should further establish the role of novel glucocorticoid replacement modalities to more reliably mimic physiological diurnal cortisol delivery.

REFERENCES

1. Addison T. *On the Constitutional and Local Effects of Diseases of the Supra-Renal Capsules*. London: Warren and Son, 1855.
2. Rege J, Nakamura Y, Satoh F, *et al*. Liquid chromatography-tandem mass spectrometry analysis of human adrenal vein 19-carbon steroids before and after ACTH stimulation. *J Clin Endocrinol Metab*, 2013; **98**: 1182–8.
3. Bancos I, Hahner S, Tomlinson J, Arlt W. Diagnosis and management of adrenal insufficiency. *Lancet Diabetes Endocrinol*, 2015; **3**: 216–26.
4. Bornstein SR. Predisposing factors for adrenal insufficiency. *N Engl J Med*, 2009; **360**: 2328–39.
5. Betterle C, Dal PC, Mantero F, Zanchetta R. Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocr Rev*, 2002; **23**: 327–64.
6. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med*, 1990; **322**: 1829–36.
7. Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*, 1997; **17**: 399–403.
8. Nagamine K, Peterson P, Scott HS, *et al*. Positional cloning of the APECED gene. *Nat Genet*, 1997; **17**: 393–8.
9. Mathis D, Benoist C. A decade of AIRE. *Nat Rev Immunol*, 2007; **7**: 645–50.
10. Kemp EH, Ajjan RA, Husebye ES, *et al*. A cytotoxic T lymphocyte antigen-4 (CTLA-4) gene polymorphism is associated with autoimmune Addison's disease in English patients. *Clin Endocrinol (Oxf)*, 1998; **49**: 609–13.
11. Skinningsrud B, Husebye ES, Gervin K, *et al*. Mutation screening of PTPN22: association of the 1858 T-allele with Addison's disease. *Eur J Hum Genet*, 2008; **16**: 977–82.
12. Skinningsrud B, Husebye ES, Pearce SH, *et al*. Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. *J Clin Endocrinol Metab*, 2008; **93**: 3310–17.
13. Agha A, Tomlinson JW, Clark PM, Holder G, Stewart PM. The long-term predictive accuracy of the short synacthen (corticotropin) stimulation test for assessment of the hypothalamic-pituitary-adrenal axis. *J Clin Endocrinol Metab*, 2006; **91**: 43–7.
14. Stewart PM, Corrie J, Seckl JR, Edwards CR, Padfield PL. A rational approach for assessing the hypothalamo-pituitary-adrenal axis. *Lancet*, 1988; **1**: 1208–10.
15. Clark PM, Neylon I, Raggatt PR, Sheppard MC, Stewart PM. Defining the normal cortisol response to the short synacthen test: implications for the investigation of hypothalamic-pituitary disorders. *Clin Endocrinol (Oxf)*, 1998; **49**: 287–92.
16. Inder WJ, Hunt PJ. Glucocorticoid replacement in pituitary surgery: guidelines for perioperative assessment and management. *J Clin Endocrinol Metab*, 2002; **87**: 2745–50.

17. Husebye ES, Anderson MS, Kämpe O. Autoimmune polyendocrine syndromes. *N Engl J Med*, 2018; **378**: 1132–41.
18. Ueland GÅ, Methlie P, Øksnes M, *et al*. The short cosyntropin test revisited: new normal reference range using LC-MS/MS. *J Clin Endocrinol Metab*, 2018; **103**: 1696–703.
19. Kazlauskaite R, Evans AT, Villabona CV, *et al*. Corticotropin tests for hypothalamic-pituitary-adrenal insufficiency: a metaanalysis. *J Clin Endocrinol Metab*, 2008; **93**: 4245–53.
20. Stewart PM, Clark PM. The low-dose corticotropin-stimulation test revisited: the less, the better?. *Nat Clin Pract Endocrinol Metab*, 2009; **5**: 68–9.
21. Cooper MS, Stewart PM. Corticosteroid insufficiency in acutely ill patients. *N Engl J Med*, 2003; **348**: 727–34.
22. Marik PE, Pastores SM, Annane D, *et al*. Recommendations for the diagnosis and management of corticosteroid insufficiency in critically ill adult patients: consensus statements from an international task force by the American College of Critical Care Medicine. *Crit Care Med*, 2008; **36**: 1937–49.
23. Annane D, Sebille V, Charpentier C, *et al*. Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA*, 2002; **288**: 862–71.
24. Sprung CL, Annane D, Keh D, *et al*. Hydrocortisone therapy for patients with septic shock. *N Engl J Med*, 2008; **358**: 111–24.
25. Annane D, Pastores SM, Rochwerg B, *et al*. Guidelines for the diagnosis and management of critical illness-related corticosteroid insufficiency (CIRCI) in critically ill patients (Part I): Society of Critical Care Medicine (SCCM) and European Society of Intensive Care Medicine (ESICM) 2017. *Intensive Care Med*, 2017; **43**: 2078–88.
26. Lutz A, Stojkovic M, Schmidt M, Arlt W, Allolio B, Reincke M. Adrenocortical function in patients with ehaviourtases of the adrenal gland. *Eur J Endocrinol*, 2000; **143**: 91–7.
27. Lam KY, Lo CY. Metastatic tumours of the adrenal glands: a 30-year experience in a teaching hospital. *Clin Endocrinol (Oxf)*, 2002; **56**: 95–101.
28. Presotto F, Fornasini F, Betterle C, Federspil G, Rossato M. Acute adrenal failure as the heralding symptom of primary antiphospholipid syndrome: report of a case and review of the literature. *Eur J Endocrinol*, 2005; **153**: 507–14.
29. Moser HW, Mahmood A, Raymond GV. X-linked adrenoleukodystrophy. *Nat Clin Pract Neurol*, 2007; **3**: 140–51.
30. Agha A, Rogers B, Sherlock M, *et al*. Anterior pituitary dysfunction in survivors of traumatic brain injury. *J Clin Endocrinol Metab*, 2004; **89**: 4929–36.
31. Giordano G, Aimaretti G, Ghigo E. Variations of pituitary function over time after brain injuries: the lesson from a prospective study. *Pituitary*, 2005; **8**: 227–31.
32. Chanson P, Lepeintre JF, Ducreux D. Management of pituitary apoplexy. *Expert Opin Pharmacother*, 2004; **5**: 1287–98.
33. Schwedt TJ, Matharu MS, Dodick DW. Thunderclap headache. *Lancet Neurol*, 2006; **5**: 621–31.
34. Caturegli P, Lupi I, Landek-Salgado M, Kimura H, Rose NR. Pituitary autoimmunity: 30 years later. *Autoimmun Rev*, 2008; **7**: 631–7.
35. Kasperlik-Zaluska AA, Czarnocka B, Czech W. Autoimmunity as the most frequent cause of idiopathic secondary adrenal insufficiency: report of 111 cases. *Autoimmunity*, 2003; **36**: 155–9.
36. Manetti L, Lupi I, Morselli LL, *et al*. Prevalence and functional significance of antipituitary antibodies in patients with autoimmune and non-autoimmune thyroid diseases. *J Clin Endocrinol Metab*, 2007; **92**: 2176–81.
37. Bornstein SR, Allolio B, Arlt W, *et al*. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 364–89. Review.
38. Esteban NV, Loughlin T, Yergey AL, *et al*. Daily cortisol production rate in man determined by stable isotope dilution/mass spectrometry. *J Clin Endocrinol Metab*, 1991; **72**: 39–45.
39. Mah PM, Jenkins RC, Rostami-Hodjegan A, *et al*. Weight-related dosing, timing and monitoring hydrocortisone replacement therapy in patients with adrenal insufficiency. *Clin Endocrinol (Oxf)*, 2004; **61**: 367–75.
40. Krieger DT, Allen W, Rizzo F, Krieger HP. Characterization of the normal temporal pattern of plasma corticosteroid levels. *J Clin Endocrinol Metab*, 1971; **32**: 266–84.
41. Plat L, Leproult R, L'Hermite-Baleriaux M, *et al*. Metabolic effects of short-term elevations of plasma cortisol are more pronounced in the evening than in the morning. *J Clin Endocrinol Metab,* 1999; **84**: 3082–92.
42. Lovas K, Loge JH, Husebye ES. Subjective health status in Norwegian patients with Addison's disease. *Clin Endocrinol (Oxf)*, 2002; **56**: 581–8.
43. Lovas K, Husebye ES, Holsten F, Bjorvatn B. Sleep disturbances in patients with Addison's disease. *Eur J Endocrinol*, 2003; **148**: 449–56.
44. Merza Z, Rostami-Hodjegan A, Memmott A, *et al*. Circadian hydrocortisone infusions in patients with adrenal insufficiency and congenital adrenal hyperplasia. *Clin Endocrinol (Oxf)*, 2006; **65**: 45–50.
45. Lovas K, Husebye ES. Continuous subcutaneous hydrocortisone infusion in Addison's disease. *Eur J Endocrinol*, 2007; **157**: 109–12.
46. Newell-Price J, Whiteman M, Rostami-Hodjegan A, *et al*. Modified-release hydrocortisone for circadian therapy: a proof-of-principle study in dexamethasone-suppressed normal volunteers. *Clin Endocrinol (Oxf)*, 2008; **68**: 130–5.
47. Debono M, Ross R, Newell-Price J. Inadequacies of glucocorticoid replacement and improvements by physiological circadian therapy. *Eur J Endocrinol*, 2009; **160**: 719–29.
48. Allolio B, Kaulen D, Deuss U, Hipp FX, Winkelmann W. Comparison between hydrocortisone and cortisone acetate as replacement therapy in adrenocortical insufficiency. *Akt Endokr Stoffw*, 1985; **6**: 35–9.
49. Kehlet H, Binder C, Blichert-Toft M. Glucocorticoid maintenance therapy following adrenalectomy: assessment of dosage and preparation. *Clin Endocrinol (Oxf)*, 1976; **5**: 37–41.
50. Mallappa A, Sinaii N, Kumar P, *et al*. A phase 2 study of Chronocort, a modified-release formulation of hydrocortisone, in the treatment of adults with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2015; **100**: 1137–45.
51. Johannsson G, Nilsson AG, Bergthorsdottir R, *et al*. Improved cortisol exposure-time profile and outcome in patients with adrenal insufficiency: a prospective randomized trial of a novel hydrocortisone dual-release formulation. *J Clin Endocrinol Metab*, 2012; **97**: 473–81.
52. Nilsson AG, Bergthorsdottir R, Burman P, *et al*. Long-term safety of once-daily, dual-release hydrocortisone in patients with adrenal insufficiency: a phase 3b, open-label, extension study. *Eur J Endocrinol*, 2017; **176**: 715–25.
53. Øksnes M, Björnsdottir S, Isaksson M, *et al*. Continuous subcutaneous hydrocortisone infusion versus oral hydrocortisone replacement for treatment of Addison's disease: a randomized clinical trial. *J Clin Endocrinol Metab*, 2014; **99**: 1665–74.

54. Björnsdottir S, Øksnes M, Isaksson M, *et al*. Circadian hormone profiles and insulin sensitivity in patients with Addison's disease: a comparison of continuous subcutaneous hydrocortisone infusion with conventional glucocorticoid replacement therapy. *Clin Endocrinol (Oxf)*, 2015; **83**: 28–35.
55. Feek CM, Ratcliffe JG, Seth J, Gray CE, Toft AD, Irvine WJ. Patterns of plasma cortisol and ACTH concentrations in patients with Addison's disease treated with conventional corticosteroid replacement. *Clin Endocrinol (Oxf)*, 1981; **14**: 451–8.
56. Scott RS, Donald RA, Espiner EA. Plasma ACTH and cortisol profiles in Addisonian patients receiving conventional substitution therapy. *Clin Endocrinol (Oxf)*, 1978; **9**: 571–6.
57. Howlett TA. An assessment of optimal hydrocortisone replacement therapy. *Clin Endocrinol (Oxf)*, 1997; **46**: 263–8.
58. Peacey SR, Guo CY, Robinson AM, *et al*. Glucocorticoid replacement therapy: are patients over treated and does it matter?. *Clin Endocrinol (Oxf)*, 1997; **46**: 255–61.
59. Arlt W, Rosenthal C, Hahner S, Allolio B. Quality of glucocorticoid replacement in adrenal insufficiency: clinical assessment vs. timed serum cortisol measurements. *Clin Endocrinol (Oxf)*, 2006; **64**: 384–9.
60. Jodar E, Valdepenas MP, Martinez G, Jara A, Hawkins F. Long-term follow-up of bone mineral density in Addison's disease. *Clin Endocrinol (Oxf)*, 2003; **58**: 617–20.
61. Zelissen PM, Croughs RJ, van Rijk PP, Raymakers JA. Effect of glucocorticoid replacement therapy on bone mineral density in patients with Addison disease. *Ann Intern Med*, 1994; **120**: 207–10.
62. Florkowski CM, Holmes SJ, Elliot JR, Donald RA, Espiner EA. Bone mineral density is reduced in female but not male subjects with Addison's disease. *N Z Med J*, 1994; **107**: 52–3.
63. Braatvedt GD, Joyce M, Evans M, Clearwater J, Reid IR. Bone mineral density in patients with treated Addison's disease. *Osteoporos Int*, 1999; **10**: 435–40.
64. Oelkers W, Diederich S, Bahr V. Diagnosis and therapy surveillance in Addison's disease: rapid adrenocorticotropin (ACTH) test and measurement of plasma ACTH, renin activity, and aldosterone. *J Clin Endocrinol Metab*, 1992; **75**: 259–64.
65. Ehrlich EN, Lindheimer MD. Effect of administered mineralocorticoids or ACTH in pregnant women. Attenuation of kaliuretic influence of mineralocorticoids during pregnancy. *J Clin Invest*, 1972; **51**: 1301–9.
66. Erichsen MM, Lovas K, Fougner KJ, *et al*. Normal overall mortality rate in Addison's disease, but young patients are at risk of premature death. *Eur J Endocrinol*, 2009; **160**: 233–7.
67. White K, Arlt W. Adrenal crisis in treated Addison's disease: a predictable but under-managed event. *Eur J Endocrinol*, 2010; **162**: 115–20.
68. Hahner S, Spinnler C, Fassnacht M, *et al*. High incidence of adrenal crisis in educated patients with chronic adrenal insufficiency: a prospective study. *J Clin Endocrinol Metab*, 2015; **100**: 407–16.
69. Allolio B. Extensive expertise in endocrinology. Adrenal crisis. *Eur J Endocrinol*, 2015; **172**: R115–24. Review.
70. Hahner S, Hemmelmann N, Quinkler M, Beuschlein F, Spinnler C, Allolio B. Timelines in the management of adrenal crisis—targets, limits and reality. *Clin Endocrinol (Oxf)*, 2015; **82**: 497–502.
71. Glowniak JV, Loriaux DL. A double-blind study of perioperative steroid requirements in secondary adrenal insufficiency. *Surgery*, 1997; **121**: 123–9.
72. Prete A, Taylor AE, Bancos I, *et al*. Prevention of adrenal crisis in stress: cortisol responses in major stress in comparison to stress dose hydrocortisone deliver in adrenal insufficiency. *J Clin Endocrinol Metab*, 2020 Mar 14. pii: dgaa133.
73. Nicholas JA, Burstein CL, Umberger CJ, Wilson PD. Management of adrenocortical insufficiency during surgery. *AMA Arch Surg*, 1955; **71**: 737–42.
74. Hahner S, Burger-Stritt S, Allolio B. Subcutaneous hydrocortisone administration for emergency use in adrenal insufficiency. *Eur J Endocrinol*, 2013; **169**: 147–54.
75. Arlt W, Callies F, van Vlijmen JC, *et al*. Dehydroepiandrosterone replacement in women with adrenal insufficiency. *N Engl J Med*, 1999; **341**: 1013–20.
76. Miller KK, Sesmilo G, Schiller A, Schoenfeld D, Burton S, Klibanski A. Androgen deficiency in women with hypopituitarism. *J Clin Endocrinol Metab*, 2001; **86**: 561–7.
77. Brooke AM, Kalingag LA, Miraki-Moud F, *et al*. Dehydroepiandrosterone improves psychological well-being in male and female hypopituitary patients on maintenance growth hormone replacement. *J Clin Endocrinol Metab*, 2006; **91**: 3773–9.
78. Gurnell EM, Hunt PJ, Curran SE, *et al*. Long-term DHEA replacement in primary adrenal insufficiency: a randomized, controlled trial. *J Clin Endocrinol Metab*, 2008; **93**: 400–9.
79. Hunt PJ, Gurnell EM, Huppert FA, *et al*. Improvement in mood and fatigue after dehydroepiandrosterone replacement in Addison's disease in a randomized, double blind trial. *J Clin Endocrinol Metab*, 2000; **85**: 4650–6.
80. Johannsson G, Burman P, Wiren L, *et al*. Low dose dehydroepiandrosterone affects ehaviour in hypopituitary androgen-deficient women: a placebo-controlled trial. *J Clin Endocrinol Metab*, 2002; **87**: 2046–52.
81. Binder G, Weber S, Ehrismann M, *et al*. Effects of dehydroepiandrosterone therapy on pubic hair growth and psychological well-being in adolescent girls and young women with central adrenal insufficiency: a double-blind, randomised, placebo-controlled phase III trial. *J Clin Endocrinol Metab*, 2009; **94**: 1182–90.
82. Miller KK, Biller BM, Beauregard C, *et al*. Effects of testosterone replacement in androgen-deficient women with hypopituitarism: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2006; **91**: 1683–90.
83. Arlt W. Androgen therapy in women. *Eur J Endocrinol*, 2006; **154**: 1–11.
84. Parasrampuria J, Schwartz K, Petesch R. Quality control of dehydroepiandrosterone dietary supplement products. *JAMA*, 1998; **280**: 1565.
85. Hahner S, Fassnacht M. Mitotane for adrenocortical carcinoma treatment. *Curr Opin Investig Drugs*, 2005; **6**: 386–94.
86. Hahner S, Loeffler M, Fassnacht M, *et al*. Impaired subjective health status in 256 patients with adrenal insufficiency on standard therapy based on cross-sectional analysis. *J Clin Endocrinol Metab*, 2007; **92**: 3912–22.
87. Bleicken B, Hahner S, Loeffler M, Ventz M, Allolio B, Quinkler M. Impaired subjective health status in chronic adrenal insufficiency: impact of different glucocorticoid replacement regimens. *Eur J Endocrinol*, 2008; **159**: 811–17.
88. Tomlinson JW, Holden N, Hills RK, *et al*. Association between premature mortality and hypopituitarism. West Midlands Prospective Hypopituitary Study Group. *Lancet*, 2001; **357**: 425–31.

89. Bergthorsdottir R, Leonsson-Zachrisson M, Oden A, Johannsson G. Premature mortality in patients with Addison's disease: a population-based study. *J Clin Endocrinol Metab*, 2006; **91**: 4849–53.
90. Bensing S, Brandt L, Tabaroj F, *et al*. Increased death risk and altered cancer incidence pattern in patients with isolated or combined autoimmune primary adrenocortical insufficiency. *Clin Endocrinol (Oxf)*, 2008; **69**: 697–704.
91. Bancos I, Hazeldine J, Chortis V, *et al*. Primary adrenal insufficiency is associated with impaired natural killer cell function: a potential link to increased mortality. *Eur J Endocrinol*, 2017; **176**: 471–80.
92. Isidori AM, Venneri MA, Graziadio C, *et al*. Effect of once-daily, modified-release hydrocortisone versus standard glucocorticoid therapy on metabolism and innate immunity in patients with adrenal insufficiency (DREAM): a single-blind, randomised controlled trial. *Lancet Diabetes Endocrinol*, 2018; **6**: 173–85.
93. Venneri MA, Hasenmajer V, Fiore D, *et al*. Circadian rhythm of glucocorticoid administration entrains clock genes in immune cells: a DREAM trial ancillary study. *J Clin Endocrinol Metab*, 2018; **103**: 2998–3009.

5.9

Congenital Adrenal Hyperplasia

5.9.1 Genetics of Congenital Adrenal Hyperplasia

Nils P. Krone

Introduction

Congenital Adrenal Hyperplasia (CAH) represents a group of autosomal recessive conditions caused by the deficiency of one of the steroidogenic enzymes involved in cortisol biosynthesis or in the electron donor enzyme P450 oxidoreductase (POR) [1]. Steroidogenic enzymes belong to two major groups: cytochrome P450 (CYP) enzymes and hydroxysteroid dehydrogenases (HSD). CYP enzymes are subdivided in type I and type II enzymes. Steroidogenic CYP type I enzymes include P450 side-chain cleavage (CYP11A1) and 11β-hydroxylase (CYP11B1) enzymes. They are localized to the mitochondrion where they receive electrons from reduced nicotinamide dinucleotide phosphate (NADPH) via ferredoxin reductase and ferredoxin. Microsomal CYP type II enzymes include 21-hydroxylase (CYP21A2) and 17α-hydroxylase (CYP17A1) enzymes. These enzymes are localized to the endoplasmic reticulum and receive electrons from POR [1, 2] (**Figure 5.9.1.1**).

Commonly, three additional conditions are categorized as CAH. Inactivating mutations in the mitochondrial cholesterol transporter steroidogenic acute regulatory protein (StAR) cause congenital lipoid adrenal hyperplasia (CLAH) compromising all three steroidogenic pathways. CLAH is clinically indistinguishable from deficiency of CYP11A1 that catalyses the first step in steroidogenesis, the generation of pregnenolone from cortisol [1, 2]. Aldosterone synthase (CYP11B2) is a mitochondrial CYP enzyme facilitating aldosterone biosynthesis. CYP11B2 deficiency is associated with isolated aldosterone deficiency [3].

All genes causing CAH are well characterized and mutation analysis is widely available (**Table 5.9.1.1**). An overall strong genotype-phenotype correlation exists in these conditions. Molecular genetic diagnosis confirms the clinical and biochemical diagnosis and provides important information for clinical and genetic counselling.

21-Hydroxylase (CYP21A2) Deficiency

The most common cause of CAH is 21-hydroxylase deficiency (21OHD, MIM +201910), accounting for 90–95% of cases. 21OHD is caused by inactivating mutations in the 21-hydroxylase gene (*CYP21A2*, GeneID 1589) leading to impaired conversion of progesterone to 11-deoxycorticosterone and 17-hydroxyprogesterone (17OHP) to 11-deoxycortisol. Commonly, 21OHD is classified into severe classic CAH and the milder non-classic (NC) form. Classic CAH is clinically subdivided into simple virilizing (SV) and salt-wasting (SW) 21OHD, which both lead to glucocorticoid deficiency. SVCAH is characterized by hyperandrogenism resulting in 46,XX DSD and precocious pseudopuberty in both sexes. This is also observed in SWCAH, but affected patients present with additional severe renal salt loss due to impaired aldosterone synthesis, which leads to life-threatening salt-wasting crisis during the first weeks of life. NCCAH manifests later in life with precocious pseudopuberty due to androgen excess, or with clinical features resembling polycystic ovary syndrome comprising hirsutism, oligomenorrhea and anovulation [4]. In some cases, individuals with NCCAH are asymptomatic [2, 5]. Classic 21OHD affects about 1 in 9800 to 18 000 live births in most Caucasian populations [6]. However, the incidence varies depending on ethnicity and geographical area, with the highest frequency detected in isolated populations such as Yupik Eskimos (1 in 282) and on the island of La Reunion (1 in 2100). Significant lower incidences are reported in the Japanese (1 in 21 000) and Taiwanese (1 in 28 000) population

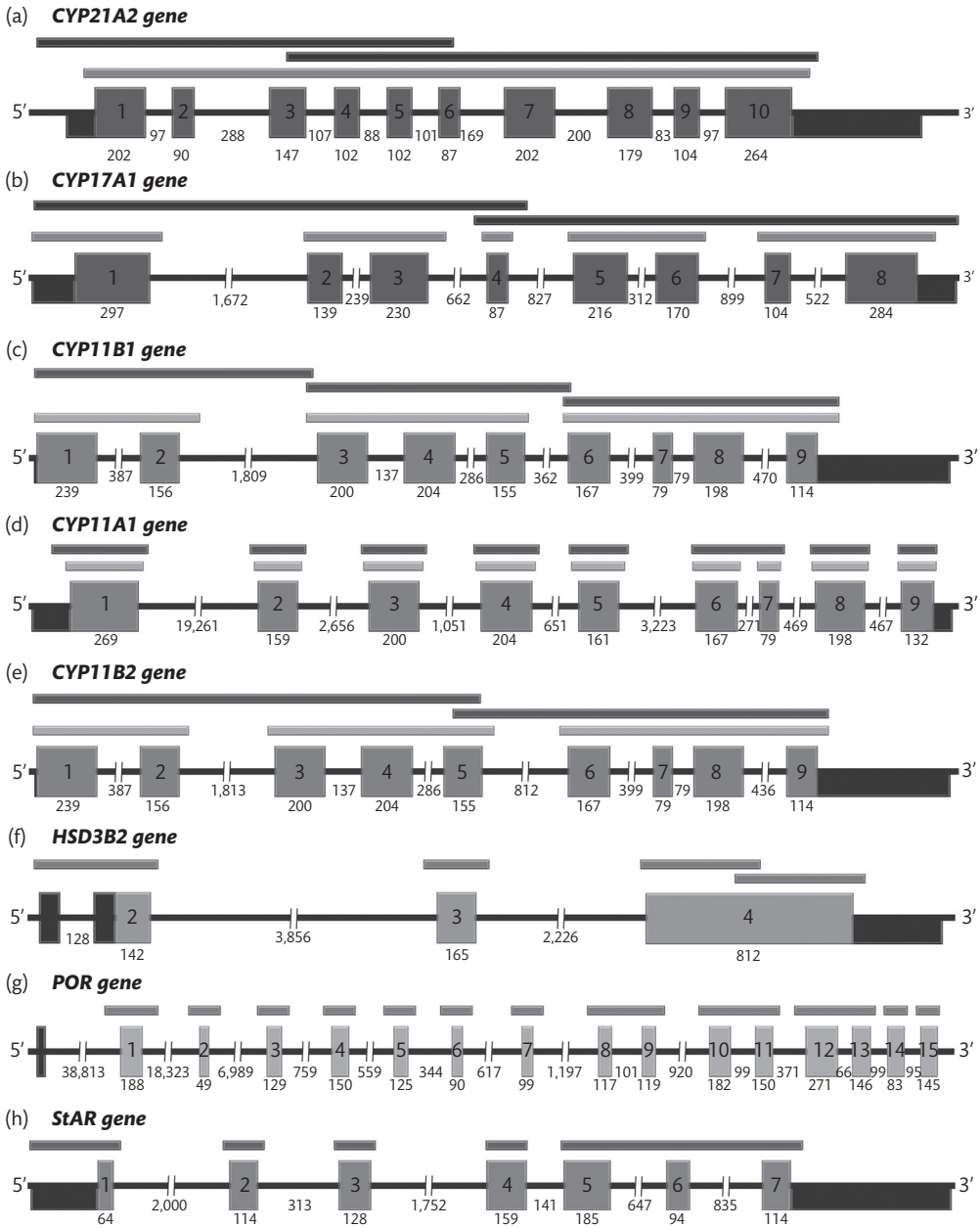


Figure 5.9.1.1 Genomic organization of genes causing different forms of congenital adrenal hyperplasia. (1) Genes encoding steroidogenic Cytochrome P450 type II enzymes: (a) The 21-hydroxylase (*CYP21A2*) gene consists of 10 exons and it is typically amplified in two overlapping fragments. (b) The 17-hydroxylase (*CYP17A1*) gene consists of 8 exons and different strategies have been employed either amplifying the gene in five or in two fragments. (2) Genes encoding steroidogenic Cytochrome P450 type I enzymes: (c) The 11-hydroxylase (*CYP11B1*) gene consists of 9 exons and is usually amplified in three overlapping fragments, although non-overlapping strategies have been described. (d) The P450 side chain cleavage (*CYP11A1*) gene consists of 9 exons and it is usually amplified in small non-overlapping fragments, although different PCR strategies have been described. (e) The aldosterone synthase (*CYP11B2*) gene consists of 9 exons normally amplified in either two overlapping fragments or three non-overlapping fragments. (3) Genes encoding hydroxysteroid dehydrogenases: (f) The hydroxysteroid dehydrogenase type 2 (*HSD3B2*) gene has 4 exons; exon 1 and the 5-prime part of exon 2 are not translated. (4) Gene encoding the electron donor of steroidogenic cytochrome P450 type II: (g) The P450 oxidoreductase (*POR*) gene has 15 translated exons and an untranslated exon (1U). PCR amplification is performed in several small fragments. (5) Gene encoding for a cholesterol transporter: (h) The steroid acute regulatory protein (*StAR*) gene consists of 7 exons commonly amplified in five fragments.

and in African Americans (1 in 42 000) [6]. NCCAH occurs with an estimated prevalence of about 1 in 200 to 1 in 1000 in Caucasian populations [7–10]. However, the incidence of NCCAH varies depending on ethnicity and geographical area ranging from 1 in 27 of Ashkenazi Jews and 1 in 50 to 1 in 100 of Hispanics and Yugoslavs (Table 5.9.1.2).

21-Hydroxylase Gene and Gene Locus

The *CYP21A2* gene encodes for the microsomal enzyme 21-hydroxylase of 495 amino acids. *CYP21A2* is localized in the HLA III region on the short arm of chromosome 6 (6p21.3), approximately 30-kilobases apart from its non-functional *CYP21A1P* pseudogene. The *CYP21A2* and *CYP21A1P* genes share a high nucleotide

Table 5.9.1.1 Differential diagnosis of congenital adrenal hyperplasia—clinical, biochemical, and genetic characteristics

Deficiency	21-hydroxylase		11β-hydroxylase		17α-hydroxylase	3β-HSD type 2	P450 oxidoreductase	Congenital lipoid adrenal hyperplasia	P450 side chain cleavage	Aldosterone synthase
OMIM No.	+201910		#202010		#202110	+201810	#201750	*600617	*118485	*124080
Gene	*CYP21A2*		*CYP11B1*		*CYP17A1*	*HSD3B2*	*POR*	*StAR*	*CYP11A1*	*CYP11B2*
GeneID	1589		1584		1586	3284	5447	6770	1583	1585
Chromosome	6p21.3		8q21		10q24.3	1p13.1	7q11.2	8q11.23	15.q24.1	18q24.3
Alias	P450c21		P450c11		P450c17	3β-HSD	CPR, CYPOR		P450scc	P450aldo
Subtype	Classic	Non-classic	Classic	Non-classic						
Incidence	1:10 000 to 15 000	1:500 to 1:1000	1:100 000 to 1:200 000	Unknown	Rare	Rare	Unknown	Rare	Rare	Rare
DSD	46,XX	No	46,XX	No	46,XY	46,XY[#]	46,XX + 46,XY*	46,XY	46,XY	No
Primary affected organ	Adrenal	Adrenal	Adrenal	Adrenal	Adrenal, gonads	Adrenal, gonads	Adrenal, gonads, liver, all CYP type II expressing tissues	Adrenal, gonads	Adrenal, gonads	Adrenal
Glucocorticoids	Reduced	Normal	Reduced	Normal	Reduced	Reduced	Reduced to normal, impaired stress response	Reduced	Reduced	Normal
Mineralocorticoids	Reduced in SW	Normal	Increased, mainly precursors	Normal	Increased	Reduced often	Reduced to increased	Reduced	Reduced	Reduced
Sex hormones	Increased	Increased	Increased	Increased	Reduced	Reduced in males Increased in females[%]	Reduced	Reduced	Reduced	Normal
Increased marker metabolite										
Plasma	17OHP 21-deoxycortisol		DOC, S		Pregnenolone, Progesterone DOC, S	17OH-Pregnenolone, DHEA	Pregnenolone, progesterone, 17OHP			DOC, B 18OH-B
Urine	Pregnanetriol, 17Preg, pregnanetriolone		THDOC, THS		THDOC, THB, Pregnanediol, pregnanediol	Pregnanetriol	Pregnenediol, pregnanediol pregnanetriol, 17Preg			
PRA	Increased	Normal –mildly increased	Reduced	Reduced	Reduced	Increased		Increased	Increased	Increased
Hypertension	No	No	Yes	Yes	Yes	No	No or mild	No	No	No
Plasma Sodium	Reduced in SW	Normal	Increased	Increased	Increased	Reduced in SW	Normal	Reduced	Reduced	Reduced
Plasma Potassium	Increased in SW	Normal	Reduced	Reduced	Reduced	Increased in SW	Normal	Increased	Increased	Increased
Urinary salt loss	Yes	No	No	No	No	Yes	No	Yes	Yes	Yes
Skeletal malformation	No	No	No	No	No	No	Yes[&]	No	No	No

[%] Steroid hormone conversion by 3β-HSD type 1 in peripheral tissues.
[#] Masculinization of the external genitalia in females at birth is rare and if present in most cases mild, signs of increased androgens usually present later.
* DSD observed in both sexes as well as normal sex specific sexual development reported.
[&] In majority of cases published thus far, but absence of skeletal malformations does not rule out P450 oxidoreductase deficiency.
17Preg, 17-hydroxypregenenolone, S,11-deoxycortisol, DOC, 11-deoxycorticosterone; B, corticosterone
THS, tetrahydro-deoxycortisol; THDOC, tetrahydro-deoxycorticosterone.

Table 5.9.1.2 Allele frequency of common 21OHD mutations in different populations

Ethnic group	C21OHD patients	NC21OHD patients	Unrelated alleles	% of 21OHD disease-causing alleles													Ref
				del/ conv	P30L	i2G	Δ8bp	I172N	E6 cluster	V281L	insT	Q318X	R356W	P453S	> 1 mut	Other	Total	
Anglo-Saxon	–	–	186	35.5	2.7	22.6	2.2	5.4	–	3.8	–	4.3	5.9	–	0	–	82.2	(Wilson *et al.*, 2007)
American[&]	166	47	364	30.5	0.8	23.4	0.5	12.6	1.1	12.6	0.3	3.3	3.6	0.5	3.7	6.6	99.5	(Finkielstain *et al.*, 2011)
American[&]	793	545	3005	20.0	2.6	22.9	2.1	8.2	2.1	23.9	–	3.5	3.6	–	–	–	88.9	(New *et al.*, 2013)
Argentinean[&]	237	217	866	11.2	0.7	20.6	0.8	8.2	2.0	26.2	–	6.7	4.2	1.4	3.8	2.2	88	(Marino *et al.*, 2011)
Ashkenazi Jews	–	–	310	12.6	0.3	13.5	0.3	1.9	–	60.0	–	1.6	1.3	–	0	–	91.6	(Wilson *et al.*, 2007)
Austrian	–	–	158	31.0	3.2	22.8	0	15.8	1.9	12.0	0	2.5	3.2	1.3	0	5.6	99.3	(Baumgartner-Parzer *et al.*, 2001)
Brazilian[&]	80	50	228	10.9	1.7	18.4	0.9	13.2	0.4	16.7	1.7	4.8	5.3	1.3	4.4	0.9	80.6	(Bachega *et al.*, 1998)
British[&]	–	–	306	30.7	2.9	24.5	1.6	14.4	0.3	7.2	0	2.9	4.9	2.0	2.3	3.9	97.7	(Krone *et al.*, 2013)
Dutch	–	–	370	33.8	0.3	28.1	4.3	12.4	3.0	2.2	0.3	3.5	8.4	0.5	0	3.2	100	(Stikkelbroeck *et al.*, 2003)
Finish	72	6	156	43.9	0	10.1	0	33.1	0*	2.7	0.7	1.4	0*	0*	3.4	2.6	97.9	(Jaaskelainen *et al.*, 1997)
French	75	54	258	18.9	–	20.5	2.7	8.9	5.0	16.7	1.2	3.9	–	0	3.9	5.0	86.7	(Barbat *et al.*, 1995)
German	144	11	310	27.4	2.6	30.3	1.6	19.7	1.0	2.9	0.3	4.8	4.5	0.3	1.6	1.6	98.7	(Krone *et al.*, 2000)
Iranian	44	0	88[§]	31.8	0*	14.8	0*	5.7	2.3	1.1	–	15.9	7.9	–	10.2	10.2	100	(Rabbani *et al.*, 2012)
Italian	60	13	146	27.4	2.7	19.9	0*	6.2	0	11.0	0*	8.2	0*	–	10.9	–	86.3	(Carrera *et al.*, 1996)
Portuguese	30	26	112	25.9	1.8	9.8	2.7	9.8	0*	25.9	4.4	6.3	1.8	0	2.6	8.0	99.1	(Friaes *et al.*, 2006)
Norwegian	63	0	126	45.3	1.6	19.8	0	21.4	0*	0.8	0*	0.8	0*	0.8	2.3	7.1	100	(Nermoen *et al.*, 2012)
Spanish	16	122	266	5.6	1.5	6.0	1.1	2.2	1.1	63.1	1.5	2.2	0.8	3.8	5.3	5.6	100	(Loidi *et al.*, 2006)
Swedish	104	18	186	29.8	1.6	27.7	1.2	20.8	0.5	5.4	0	0.5	3.8	0.5	4.8	2.6	99.2	(Wedell *et al.*, 1994)
Taiwanese	–	–	400	9.5	0	34.0	0.2	23.5	1.0	0.2	1.8	6.3	11.8	–	0	9.2	97.5	(Lee *et al.*, 2008)
Tunisian	36	14	100[§]	22.0	1.0	6.0	0	8.0	1.0	12.0	0	26.0	1.0	0	5.0	5.0	87.0	(Ben Charfeddine *et al.*, 2012)
Turkish	50	6	91	20.9	0	22.0	4.4	9.9	2.2	7.6	0	3.3	8.8	2.2	1.1	2.2	84.9	(Bas *et al.*, 2009)

* Only detected in alleles bearing more than 1 point mutation; [&] Heterogeneous population; [§] High degree of consanguinity; i2G, c.293-13A/C>G; Δ8bp, p.Gly110Val*fsX21*; insT, p.Leu307Phe*fsX6*.

homology, of about 98% and 96% at exon and intron level, and consist of 10 exons [6]. Both genes are arranged in tandem repeat with the *C4A* and *C4B* genes encoding the fourth complement factor. The *C4/CYP21A2* unit is flanked by the *RP1* (*STK19*) gene on the telomeric side and by the *TNXB* gene on the centromeric side, and their truncated pseudogenes, *RP2* and *TNXA*, forming the RCCX module (*RP-C4-CYP21-TNX*). The highly variable RCCX module spans about 30 kb. Most chromosomes harbour two copies of the modules with a *CYP21A1P* pseudogene in the telomeric module and a *CYP21A2* gene in the centromeric module. However, monomodular, trimodular, or even quadrimodular haplotypes have been described [11] (Figure 5.9.1.2). The reverse DNA strand from the *C4* and *CYP21A* genes encodes for the *TNXA* and *TNXB* genes with opposite transcriptional orientation. *TNXB* is a very large gene consisting of 43 exons and spans 68.2 kb of DNA encoding the extracellular matrix protein, tenascin (TNX). The *TNXA* pseudogene was truncated during the duplication of the ancestral RCCX module. The last exon of *TNXA* and *TNXB* is located in the 3-prime untranslated region of exon 10 of *CYP21A1P* and *CYP21A2*, respectively [11, 12].

Mutations Causing 21-Hydroxylase Deficiency

The majority of mutations causing 21OHD are the consequence of intergenic recombination within the RCCX module. The high recombination rate in the HLA III region along with the nucleotide identity shared across the RCCX locus promote unequal crossing-over and

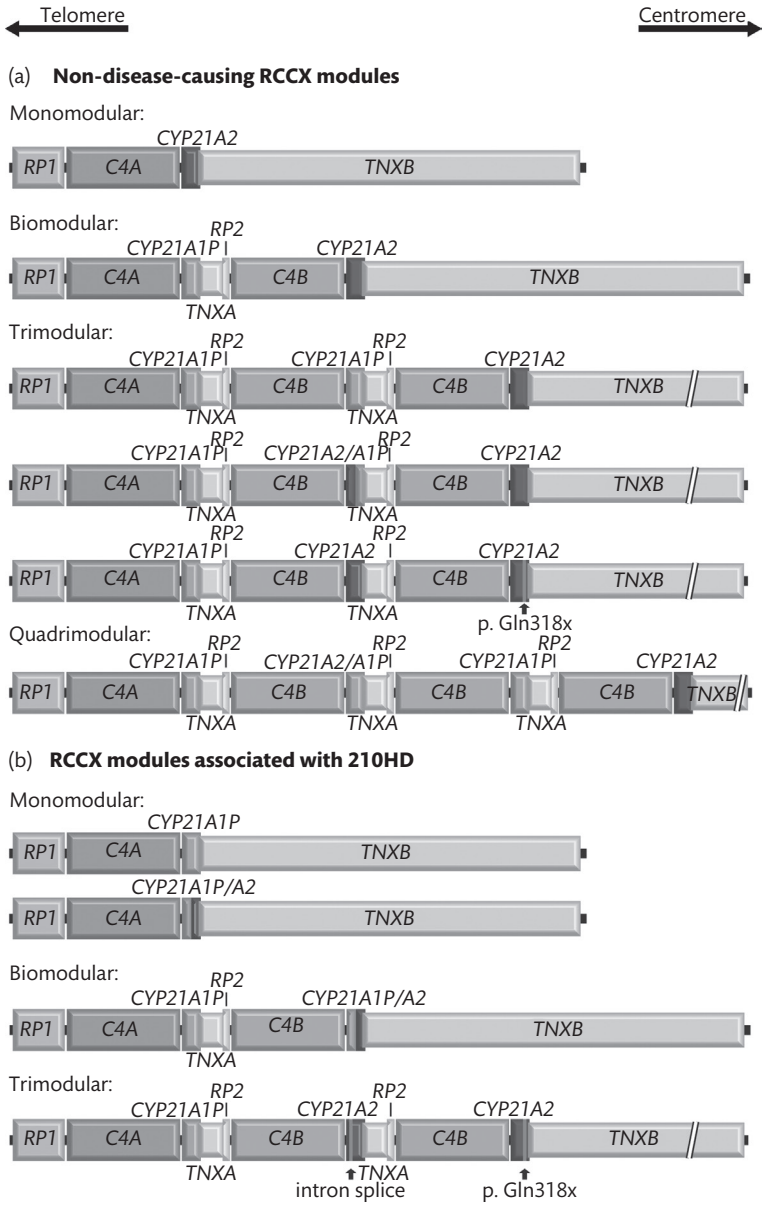


Figure 5.9.1.2 Organization of the RCCX module at chromosome 6p21. Representative Copy Number Variants at the RCCX locus (a) non-disease variants; (b) Configurations associated with 21-hydroxylase deficiency. *C4A* and *C4B*: complement component C4A and C4B genes; *CYP21A1P*: steroid 21-hydroxylase pseudogene; *CYP21A2:* steroid 21-hydroxylase gene; *TNXB:* tenascin-X gene. *TNXA*: tenascin-X pseudogene; *RP1*: serine/threonine kinase 19 gene (other names: *STK19*); *RP2*: serine/threonine kinase 19 pseudogene (other names: *STK19P*); intron splice, I2G, c.293-13A/C>G.

gene conversions between *CYP21A2* and its pseudogene. Unequal crossing overs with break points outside the *CYP21A* locus cause deletions of the *CYP21A2* gene. Chromosomes lacking the *CYP21A2* gene frequently carry a *TNXB/TNXA* chimeric gene. Misalignment of chromosomes carrying different RCCX copy numbers promote unequal crossing overs within the *CYP21A* locus leading to the formation of *CYP21A1P/CYP21A2* chimeric genes, which have in the past been regarded as large gene conversions. Deletions and chimeric genes account for approximately 20–25% of mutant alleles in 21OHD [6]. *CYP21A2* gene duplications are relatively frequent in some populations [8, 13]. The majority of these alleles carrying a *CYP21A2* gene duplication have a p.Gln318X mutation in the duplicated *CYP21A2* gene next to the *TNXB* gene, and a wild-type *CYP21A2* gene next to the *TNXA* pseudogene [8] (Figure 5.9.1.2a). Such alleles are non-disease-causing.

Approximately 75% of *CYP21A2* disease-causing mutations, including seven point mutations, a deletion of 8 base pairs in exon 3 and a cluster of three point mutations in exon 6 originate from the transfer of genetic material from the *CYP21A1P* pseudogene to the *CYP21A2* gene [6, 11] (Figure 5.9.1.3b). In addition, over 200 pseudogene-independent mutations are listed in the Human Gene Mutation Database (HGMD, http://www.hgmd.cf.ac.uk; August 2018) and the Pharmacogene Variation Consortium (https://www.pharmvar.org/gene/CYP21A2). Most of these rare mutations are sporadic. However, due to founder effects increased frequencies of some pseudogene-independent mutations are observed in some populations [11].

Pseudogene-derived mutations occur with similar frequencies in most populations (Table 5.9.1.2) [11]. Deletions and chimeric genes along with the splice site mutation in intron 2 (c.293-13A/C>G) and p.Ile172Asn are the most common mutations in most Caucasian populations [1, 14–19]. The p.Val281Leu is by far the most common mutation detected in NCCAH patients [6, 11]. Novel or rare mutations account for about 3–5% of detected mutations in large cohorts. The vast majority of these rare mutations has been identified in single families or small populations. Approximately 1–2% of *CYP21A2*-inactivating mutations arise *de novo* [6, 17]. If possible, carrier testing should be performed in the parents to confirm compound heterozygosity or homozygosity in the index case and estimate the reoccurrence risk.

Genotype-Phenotype Correlation in 21-Hydroxylase Deficiency

Mutation groups have been established based on the basis of the residual 21-hydroxylase activity of the mutant proteins as observed *in vitro* (Figure 5.9.1.3c). CYP21A2 deficiency is an autosomal recessive condition. About 65–75% of the CAH patients are compound heterozygous (e.g. they are affected), but carry different mutations on each chromosome. The clinical phenotype of CAH correlates well with the less severely mutated allele, and consequently with the allele encoding for the higher residual *in vitro* activity of 21-hydroxylase. This has major implications for genetic counselling in patients with NCCAH. If a patient with NCCAH is compound heterozygous for a mild and a severe mutation, the risk of having a child with classic CAH increases significantly to about 1 in 400 ($1/50 \times 1/2 \times 1/4$), assuming a heterozygous rate of 1/50 for classic mutations in the general population. The correlation with glucocorticoid and mineralocorticoid deficiency is strong (Figure 5.9.1.3c). However, divergence between genotype and phenotype occurs in some cases. Although a trend exists, the correlation between the genotype and degree of 46,XX DSD classified assessed by Prader genital stage is less pronounced (Figure 5.9.1.3c). This implies the importance of factors modifying clinical androgen effects. Several potential modifiers have been studied. However, no analysed candidate has sufficiently explained the phenotypic variability in 21OHD. Long-term health outcomes in adults do not necessarily show a good correlation with the genotype [20]. However, girls and women with CAH with more severe genotypes appear to have an increased risk for psychiatric disorders [21].

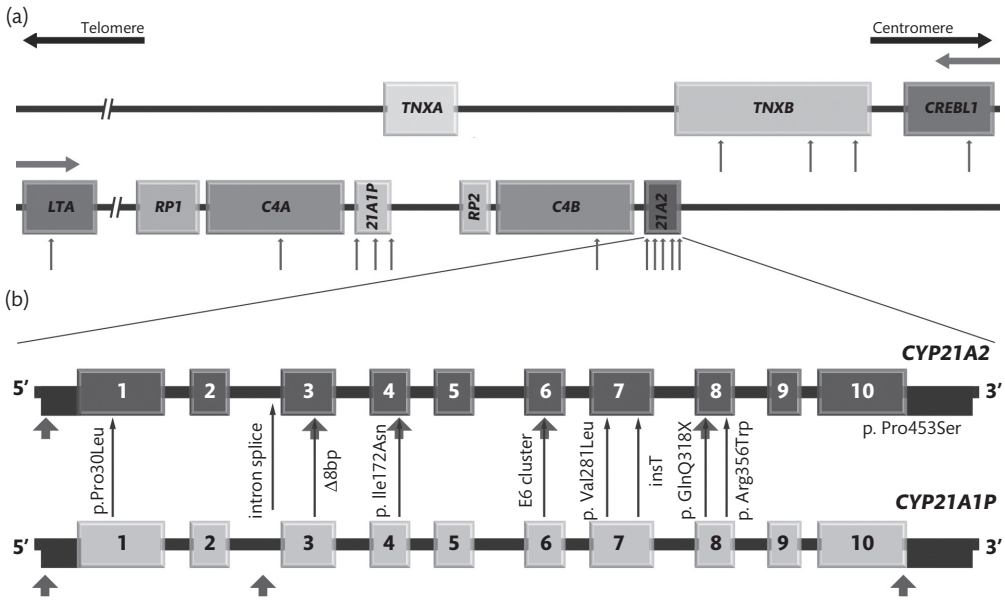
CYP21A2 gene deletions extending into *TNXB* lead to a contiguous gene syndrome consisting of CAH and Ehlers–Danlos syndrome, which is rarely observed in patients with severe SWCAH. Up to 10% of patients with SWCAH show haploinsufficiency of *TNX* and heterozygosity for *TNXB* mutations. This extended phenotype has been termed as CAH-X syndrome. CAH-X is associated with joint hypermobility, chronic arthralgia, joint subluxations, hernias, and cardiac defects [12]. Further investigations in patients with CAH-X should be considered.

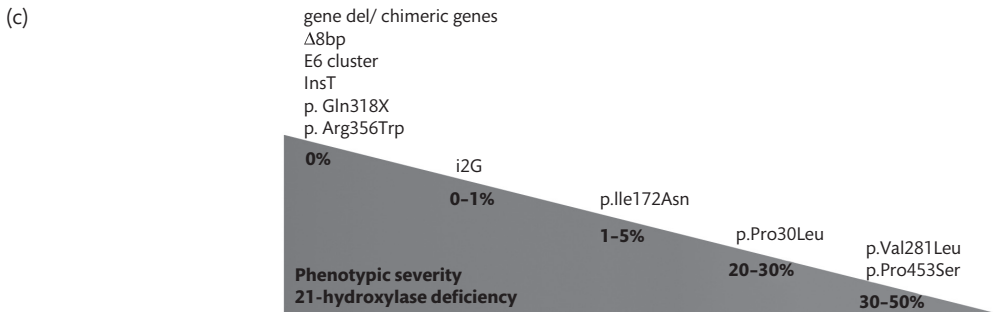
Novel insights into the underlying molecular pathology have been gained by the analysis of the CYP21A2 protein crystal structure. Null and A mutations commonly disrupt haeme- and/or substrate-binding domains, the anchoring of the protein to the membrane, or impair protein stability. Mutations categorized as group B partially impair membrane anchoring or affect conserved hydrophobic clusters within the protein. Milder mutation (group C) results in less severe alterations, often interfering with electron transfer from POR, salt-bridge and hydrogen-bonding networks, and non-conserved hydrophobic clusters [22].

Molecular Genetic Testing for *CYP21A2* Gene Mutations

Southern blot analysis has been the gold standard for the detection of *CYP21A2* gene deletions and *CYP21A1P*/*CYP21A2* chimeric genes. However, Southern Blot analysis requires relatively large amounts of high-quality DNA, is labour intensive and time consuming. In addition, *CYP21A1P* duplications and certain rare rearrangements at this locus may impede the detection of *CYP21A2* gene deletions or duplications [6, 11]. Several approaches for gene dosage detection have been developed to overcome the disadvantages of Southern Blot analysis [6]. The most widely used approach in current practice is multiplex ligation-dependent probe amplification (MLPA). MPLA requires only small amounts of DNA for detection of gene deletions, rearrangements and fusion genes [13, 23–25].

The design of *CYP21A2*-specific primers for PCR-based amplification of *CYP21A* is crucial to avoid amplification of the pseudogene and allele dropout by non-amplifying PCR fragments. This is challenging due to the high number of polymorphisms found across *CYP21A2* and the high homology with its pseudogene. A variety of targeted molecular genetic strategies for detecting mutations have been published. Direct sequencing of the amplified PCR products combined with a method for the detection of gene deletions and chimeric genes are the only available strategies that allow for the detection of 100% of *CYP21A2* mutated alleles and due to decreasing costs represents the method of choice in most routine laboratories.





Mutation group	Null	A	B	C
21OHD phenotype	Salt wasting		Simple virilizing	Non-classic
Positive predictive values				
Speiser et al., 1992	96%	85%	73%	63%
Krone *et al.*, 2000	100%	90%	74%	65%
Stikkelbroeck *et al.*, 2003	97%	96%	53%	100%
Finkielstain *et al.*, 2011	91%	92%	85%	98%
Marino *et al.*, 2011	100%	84%	87%	100%
New *et al.*, 2013	99%	91%	76%	94%
Prader genital stages				
Speiser *et al.*, 1992	II–IV (IV)	III–V (IV)	I–IV (III)	0–IV (0)
Wedell *et al.*, 1994	III–IV	II–V	I–IV	
Jaaskelainen *et al.*, 1997	II–V	0–V	II–V	0
Krone *et al.*, 2000	III–V (IV)	II–V (IV)	II–V (III)	0–IV (0)

Figure 5.9.1.3 *CYP21A2* gene locus, most common *CYP21A2* gene mutations and genotype-phenotype correlation. (a) Organization of the functional *CYP21A2* gene and its non-functional *CYP21A1P* pseudogene in the RCCX module Grey arrows indicate for the approximate position of MLPA probes. (b) Nine out of 10 common mutations are transferred by microconversions from the *CYP21A1P* gene into *CYP21A2*. Grey arrows show the five MLPA probes specific for the *CYP21A2* gene and the three probes specific for the *CYP21A1P* pseudogene. (c) Genotype-phenotype correlations in CAH due to 21-hydroxylase deficiency based on *in vitro* CYP21A2 activity. Mutation groups Null and A are associated with the salt-wasting (SW) form of 21-hydroxylase deficiency (21OHD), group B with the simple virilizing (SV) form, and group C with the non-classic (NC) form. Positive predictive values are calculated from the cited publications. The variability in the degree of virilization of the female external genitalia in the different mutation groups (grading according to Prader genital stages) is shown in the lower panel. Modal values are provided in brackets where possible. E6 cluster refers to the p.Ile236Asn, p.Val237Glu and p.Met239Leu mutation cluster at exon 6; intron splice refers to the c.293-13A/C>G mutation (other names: i2G, I2G, IVS2-13A/C>G): Δ8bp refers to the p.Gly110Val*fs*X21 mutation.

11β-Hydroxylase (CYP11B1) Deficiency

Approximately 5% of CAH cases are caused by 11β-hydroxylase (CYP11B1) deficiency (11OHD, OMIM + 202010) leading to impaired cortisol biosynthesis. Accumulation of the mineralocorticoid precursor 11-deoxycorticosterone leads to transactivation of the mineralocorticoid receptor resulting in arterial hypertension. Excess steroid hormone precursors are shunted into androgen synthesis and cause hyperandrogenism. Classic 11OHD occurs 1 in 100 000 to 1 in 200 000 live births.

A higher incidence has been reported in Israel (1 in 30 000 to 1 in 40 000 live births), in particular in Israeli Jews of Moroccan origin (1 in 5000 to 1 in 7000 live births) [26, 27]. The clinical presentation of classic 11OHD includes glucocorticoid deficiency, severe virilization of external genitalia in 46,XX neonates, precocious pseudopuberty in both sexes, and, in two-thirds of patients, hypertension [26, 27]. Non-classic 11OHD (NC11OHD) is much rarer than NC21OHD, but appears to be more frequent than previously thought [28] (Table 5.9.1.1). Patients with NC11OHD present with signs and symptoms of androgen excess during childhood [27, 28].

Mutations in the 11β-hydroxylase gene (*CYP11B1*; GeneID: 1584, GenbankID NC_000008.9) are the underlying cause of 11OHD [29]. *CYP11B1* is localized on chromosome 8q21, approximately 40 kb apart from the highly homologous aldosterone synthase gene (*CYP11B2*) [26]. The *CYP11B1* gene consists of nine exons (Figure 5.9.1.1) and encodes for a protein of 503 amino acids. CYP11B1-inactivating mutations are distributed over the entire coding region consisting of 9 exons. A cluster is reported in exons 2, 6, 7, and 8, but real hot spots do not exist. A broad variety of mutations have been reported to cause either classic or non-classic 11OHD. Most molecular analysis approaches amplify the *CYP11B1* gene in 3 fragments avoiding amplification of the highly homologous *CYP11B2* gene (Figure 5.9.1.1). Since no hotspots exist in the general population, direct DNA sequencing is the most suitable approach for molecular genetic analysis. Large rearrangements, such as large gene deletions and chimeric genes can be detected by Southern Blot analysis or MLPA [6, 26].

17α-Hydroxylase (CYP17A1) Deficiency

Steroid 17α-hydroxylase (CYP17A1) deficiency (17OHD, OMIM #202110) accounts for about 1% of all CAH cases. CYP17A1 facilitates two distinct enzymatic reactions, 17α-hydroxylase and 17,20-lyase, relying on electron transfer facilitated by the enzyme P450 oxidoreductase (POR) for its catalytic activity. Cytochrome b_5 (CYB5A) is required as an additional allosteric factor for an efficient 17,20-lyase reaction [1]. 17OHD results in both cortisol and sex steroid deficiency. In addition, the mineralocorticoid precursors DOC and corticosterone accumulate. The accumulation of the weak glucocorticoid corticosterone prevents adrenal crisis in the majority of 17OHD patients. Increased DOC concentrations lead to severe hypokalaemic hypertension. Deficient sex steroid synthesis results in male undervirlization (46,XY DSD) and in primary amenorrhea in 46,XX individuals. Absent pubertal development is observed in both sexes, with biochemical signs of hypergonadotropic hypogonadism [30] (Table 5.9.1.1). Isolated 17,20-lyase deficiency is a rare variant with largely preserved 17α-hydroxylase activity. This variant manifests with impaired sex steroid biosynthesis only, but neither with evidence of mineralocorticoid excess nor glucocorticoid deficiency [6, 30]. The *CYP17A1* gene (GeneID 1586; GenbankID NC_000010.9) is located on chromosome 10 (10q24.3) and consists of 8 exons (Figure 5.9.1.1) encoding a 508 amino acid protein. Over 70 *CYP17A1*-inactivating mutations exist. There is no mutational a hot spot in most large populations. Therefore, sequencing of the entire coding region is commonly necessary for molecular genetic diagnosis. Exceptions have been described (i.e. in Brazil) where the p.Met1Thr and p.Tyr329X mutations account for 82% of mutant alleles, in Canadian Mennonites and Dutch Frieslanders, as well as in Japan and in East Asia [1, 6, 30]. Mutations causing isolated 17,20-lyase deficiency result in amino acid substitutions located within the area of the CYP17A1 molecule interacting with the cofactor cytochrome b_5 [6, 30]. However, urine steroid metabolome analysis from patients carrying such CYP17A1 mutations provided evidence for impaired 17α-hydroxylation, which means that this condition does not represent true but apparent isolated 17,20 lyase deficiency. Apparently isolated 17,20-lyase deficiency might be mimicked by POR deficiency due to mutations in the *POR* gene [6] while mutations in CYB5A represent true 17,20 lyase deficiency [31].

3β-Hydroxysteroid Dehydrogenase Type 2 (HSD3B2) Deficiency

Two 3β-hydroxysteroid dehydrogenase isoforms, type 1 (HSD3B1) and type 2 (HSD3B2), exist. HSD3B2 is mainly expressed in adrenals and gonads while HSD3B1 is expressed in the placenta and peripheral tissues [32]. HSD3B2 deficiency (OMIM +201810) is a rare cause of CAH.

HSD3B2 is a nicotinamide adenine dinucleotide (NAD^+)-dependent enzyme anchored to the membrane of the endoplasmic reticulum and mitochondria. HSD3B2 catalyses the conversion of Δ5 to Δ4 steroids and HSD3B2 deficiency affects all three steroidogenic pathways (mineralocorticoids, glucocorticoids, and androgens). A broad phenotypic spectrum has been reported. Classic HSD3B2 deficiency presents with either salt-wasting or non-salt-wasting. Affected males show a variable impairment of gonadal steroidogenesis leading to different degrees of 46,XY DSD. Affected 46,XX individuals manifest only with adrenal symptoms and normal or mildly virilized external genitalia. Isolated premature pubarche and primary hypogonadism has been reported in infants and children of both sexes. A non-classic variant has also been described, manifesting with hirsutism and menstrual irregularities [32–34] (Table 5.9.1.1). The 3β-hydroxysteroid dehydrogenase type 1 and 2 isoforms, are encoded by the *HSD3B1* and *HSD3B2* genes, respectively. Mutations in the *HSD3B2* (GeneID 3284, GenbankID NC_000001.9) gene cause 3β-hydroxysteroid dehydrogenase deficiency. *HSD3B2* is located on chromosome 1p13.1, in close proximity to the highly homologous *HSD3B1* gene. It consists of one untranslated and 3 translated exons. Mutation analysis is performed by PCR amplification of *HSD3B2* of exons 2 to 4 covering intron-exon boundaries followed by direct DNA sequencing of the PCR products (Figure 5.9.1.1). Over 40 mutations have been reported so far, which are distributed over the entire genomic region. An overall good genotype-phenotype correlation has been described for the salt-loss phenotype, with completely inactivating mutations associated with the salt-wasting form and partly inactivating mutations allowing for some residual aldosterone synthesis. A poor correlation with gonadal steroidogenesis has been observed [32].

P450 Oxidoreductase (POR) Deficiency

P450 oxidoreductase (POR; OMIM *124015) deficiency (PORD; OMIM #201750) presents biochemically and clinically as apparent combined CYP17A1-CYP21A2 deficiency and with a malformation phenotype resembling the Antley–Bixler syndrome (ABS, OMIM #207410) [35, 36]. The incidence of PORD remains unclear, but it is likely the third most common CAH variant, after 21OHD and 11OHD. POR serves as an electron donor enzyme to CYP21A2, CYP17A1, and CYP19A1 (P450 aromatase). PORD impacts on both glucocorticoid and sex steroid synthesis and can result in mineralocorticoid excess with hypertension, albeit milder than in other CAH variants and often only manifesting in adulthood. The PORD phenotype usually consists of adrenal insufficiency, DSD or normal sex development in both sexes and presentations with or without skeletal and other malformations [35]. In contrast to other hyperandrogenic CAH forms, virilization does not progress after birth. Primary hypogonadism and lack or delayed pubertal development has been reported in both sexes. PORD is a multisystem disorders as it affects all microsomal CYP enzymes relying on POR for their enzymatic function (Table 5.9.1.1). The *POR* gene (GeneID 5447; GenbankID NC_000007.12) spans 32.9 kb and located on 7q11.2. It consists of 15 translated exons encoding for a 680 amino acid protein. An untranslated exon is located approximately 30.5 kb upstream of the first translated exon (Figure 5.9.1.1). *POR* is highly polymorphic with 140 genetic variants identified in a large cohort of healthy individuals [37] and over 60 disease-causing mutations have been described. No clear mutational hot spot exists for *POR*-inactivating mutations apart from two frequently detected mutations. The p.Ala287Pro mutation accounts for approximately 50% of mutated alleles in Caucasians, whereas the p.Arg457His mutation is the most common mutation (50–70% of alleles) in the Japanese population [35, 36]. Some studies have explored the genotype-phenotype correlation in PORD [38]. The phenotypic expression of a specific genotype is variable. Carriers of two mild mutations commonly present with a milder phenotype than compound heterozygotes for a mild and a severe mutation. Generally, mutations causing 46,XY DSD present with normal female external genitalia in 46,XX individuals and variants associated with 46,XX DSD lead to normal external male genitalia in 46,XY individuals. In carriers of severe genotypes, more severe skeletal and other complex malformation can be found [39]. The variable genotype-phenotype correlation is likely to be caused by the differential impact of POR mutations on the catalytic activity of different microsomal CYP enzymes; the malformation phenotype is explained by impairment of POR-dependent CYP enzymes involved in sterol biosynthesis and retinoic acid metabolism.

Steroid Acute Regulatory Protein (stAR) Deficiency–Congenital Lipoid Adrenal Hyperplasia

StAR actively transports cholesterol into the mitochondrion, where cholesterol is converted to pregnenolone by P450 side-chain cleavage enzyme (CYP11A1). Loss of StAR-dependent cholesterol transport causes CLAH, OMIM #201710), leading to an impairment of all three steroidogenic pathways. Classic CLAH is associated with salt-wasting and adrenal failure manifesting within the first month of life, female external genitalia irrespective of karyotype, and hypergonadotropic hypogonadism in both sexes. Non-classic Lipoid CAH may manifest later in life with primary adrenal insufficiency and mild 46,XY DSD or normal male genital development [1]. In some patients, hypocortisolism is the only clinical manifestation and non-classic CLAH might be misdiagnosed as familial glucocorticoid deficiency [40, 41] (Table 5.9.1.1).

The *StAR* gene (GeneID: 6770, NC_000008.10) is localized on chromosome 8p11.23 and consists of 7 exons (Figure 5.9.1.1) encoding for a protein of 285 amino acids. The N-terminus of the protein contains a mitochondrial signal and the C-terminus is involved in the interaction with cholesterol. Over 40 StAR mutations have been reported (HGMD, August 2018), nine of which are associated with non-classic CLAH. The majority of StAR-inactivating mutations are located in exons 5 to 7 encoding for the C-terminal protein. Recurrent mutations are detected in some populations. The p.Gln258X is found in over 80% of the CLAH-causing alleles in Korean and Japanese patients [1]. The p.Arg182Leu and the c.201_202delCT mutations are the most common mutations in Palestinian patients and the p.Arg182His is most frequently found in patients of Saudi Arabian origin [1]. In classic CLAH the genotype-phenotype correlation is good, whereas non-classic CLAH presents with phenotypic variability [1].

P450 Side-Chain Cleavage Enzyme (CYP11A1) Deficiency

CYP11A1 is a mitochondrial CYP type I enzyme and catalyses the first and rate-limiting step of steroidogenesis, the generation of pregnenolone from cholesterol. CYP11A1 deficiency (OMIM *118485) is rare and similar to CLAH impairs all three steroidogenic pathways in the adrenal and gonadal sex steroid synthesis. Its presentation resembles the CLAH phenotype. Affected patients present soon after birth or later in infancy with adrenal insufficiency and normal female external genitalia in both sexes [42]. A milder form has been described, presenting with different degrees of 46,XY DSD [42] or normal male genital development [43], and in some cases isolated adrenal insufficiency. In contrast to CLAH, all patients with CYP11A1 deficiency have small or normal-sized adrenals [42] (Table 5.9.1.1).

The *CYP11A1* gene (GeneID 1583, NC_000015.9) is located on chromosome 15q24.1 and consists of 9 exons (Figure 5.9.1.1). An overall good correlation of the genotype and onset and severity of adrenal phenotype has been reported. Severe mutations are invariably associated with female external genitalia irrespective of chromosomal sex. Current data suggest that the genotype in partially inactivating mutations cannot be used to predict DSD. In some populations, milder CYP11A1 deficiency appears to be a significant underlying cause for the clinical presentation of clinically diagnosed 'isolated' adrenal insufficiency.

Aldosterone Synthase (CYP11B2) Deficiency

Aldosterone synthase (CYP11B2) is a mitochondrial enzyme expressed in the zona glomerulosa. CYP11B2 deficiency causes isolated impairment of aldosterone synthesis in the context of normal cortisol and sex steroid production. Affected individuals present with salt-wasting crisis during the first days or weeks of life. Adults are usually asymptomatic, but might be more sensitive to episodes of severe salt loss than non-affected individuals [3]. Two biochemically distinct forms have been described, ASD1 (OMIM # 203400) and ASD2 (OMIM # 610600), which are both caused by CYP11B2 deficiency (**Table 5.9.1.1**). The *CYP11B2* gene (GeneID 1585, GenbankID NC_000008.10) lies on chromosome 18q24.3 approximately 40 kb apart from the highly homologous *CYP11B1* gene (**Figure 5.9.1.1**). More than 30 mutations have been described so far (HGMD August 2018). The p.[Arg181Trp+Val385Ala] double mutant is the most common ASD2 allele found in Iranian Jewish [3]. Homozygosity for only one of these mutants is not sufficient to cause the disease. The underlying molecular mechanism distinguishing the two ASD forms is not well understood. *In vitro* data suggests that mutations completely abolishing CYP11B2 activity are associated with ASD1, while mutations allowing for 11- and 18-hydroxylase activities appear to be associated with ASD2 [3].

REFERENCES

1. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*, 2011; **32**: 81–151.
2. Krone N, Dhir V, Ivison HE, Arlt W. Congenital adrenal hyperplasia and P450 oxidoreductase deficiency. *Clin Endocrinol*, 2007; **66**: 162–72.
3. White PC. Aldosterone synthase deficiency and related disorders. *Mol Cell Endocrinol*, 2004; **217**: 81–7.
4. Reisch N, Arlt W, Krone N. Health problems in congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Horm Res Paediatr*, 2011; **76**: 73–85.
5. Nandagopal R, Sinaii N, Avila NA, *et al.*, Phenotypic profiling of parents with cryptic nonclassic congenital adrenal hyperplasia: findings in 145 unrelated families. *Eur J Endocrinol*, 2011; **164**: 977–84.
6. Krone N, Arlt W. Genetics of congenital adrenal hyperplasia. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 181–92.
7. Zerah M, Ueshiba H, Wood E, *et al.* Prevalence of nonclassical steroid 21-hydroxylase deficiency based on a morning salivary 17-hydroxyprogesterone screening test: a small sample study. *J Clin Endocrinol Metab*, 1990; **70**: 1662–7.
8. Parajes S, Quinteiro C, Dominguez F, Loidi L. High frequency of copy number variations and sequence variants at CYP21A2 locus: implication for the genetic diagnosis of 21-hydroxylase deficiency. *PLoS One*, 2008; **3**: e2138.
9. Baumgartner-Parzer SM, Nowotny P, Heinze G, Waldhausl W, Vierhapper H. Carrier frequency of congenital adrenal hyperplasia (21-hydroxylase deficiency) in a Middle European Population. *J Clin Endocrinol Metab*, 2005; **90**: 775–8.
10. Fitness J, Dixit N, Webster D, et *al.* Genotyping of CYP21, linked chromosome 6p markers, and a sex-specific gene in neonatal screening for congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 1999; **84**: 960–6.
11. Parajes S, Krone N. Molecular genetics of 21-hydroxylase deficiency. *eLS*, 2014. doi:10.1002/9780470015902.a0023845
12. Miller WL, Merke DP. Tenascin-X, congenital adrenal hyperplasia, and the CAH-X syndrome. *Horm Res Paediatr*, 2018; **89**: 352–61.
13. Kleinle S, Lang R, Fischer GF, *et al.* Duplications of the functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect. *J Clin Endocrinol Metab*, 2009; **94**: 3954–8.
14. Marino R, Ramirez P, Galeano J, *et al.* Steroid 21-hydroxylase gene mutational spectrum in 454 Argentinean patients: genotype-phenotype correlation in a large cohort of patients with congenital adrenal hyperplasia. *Clin Endocrinol*, 2011; **75**: 427–35.
15. Loidi L, Quinteiro C, Parajes S, *et al.* High variability in CYP21A2 mutated alleles in Spanish 21-hydroxylase deficiency patients, six novel mutations and a founder effect. *Clin Endocrinol*, 2006; **64**: 330–6.
16. Krone N, Braun A, Roscher AA, Knorr D, Schwarz HP. Predicting phenotype in steroid 21-hydroxylase deficiency? Comprehensive genotyping in 155 unrelated, well defined patients from southern Germany. *J Clin Endocrinol Metab*, 2000; **85**: 1059–65.
17. Finkielstain GP, Chen W, Mehta SP, *et al.* Comprehensive genetic analysis of 182 unrelated families with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2011; **96**: E161–72.
18. Stikkelbroeck NM, Hoefsloot LH, de Wijs IJ, Otten BJ, Hermus AR, Sistermans EA. CYP21 gene mutation analysis in 198 patients with 21-hydroxylase deficiency in The Netherlands: six novel mutations and a specific cluster of four mutations. *J Clin Endocrinol Metab*, 2003; **88**: 3852–9.
19. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*, 2000; **21**: 245–91.
20. Krone N, Rose IT, Willis DS, *et al.* Genotype-phenotype correlation in 153 adult patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency—analysis of the United Kingdom Congenital adrenal Hyperplasia Adult Study Executive (CaHASE) cohort. *J Clin Endocrinol Metab*, 2013; **98**: E346–54.
21. Strandqvist A, Falhammar H, Lichtenstein P, *et al.* Suboptimal psychosocial outcomes in patients with congenital adrenal hyperplasia: epidemiological studies in a nonbiased national cohort in Sweden. *J Clin Endocrinol Metab*, 2014; **99**: 1425–32.
22. Pallan PS, Lei L, Wang C, Waterman MR, Guengerich FP, Egli M. Research resource: correlating human cytochrome P450 21A2 crystal structure and phenotypes of mutations in congenital adrenal hyperplasia. *Mol Endocrinol*, 2015; **29**: 1375–84.
23. Ghizzoni L, Cappa M, Vottero A, *et al.* Relationship of CYP21A2 genotype and serum 17-hydroxyprogesterone and cortisol levels in a large cohort of Italian children with premature pubarche. *Eur J Endocrinol*, 2011; **165**: 307–14.
24. Chan AO, But WM, Ng KL, *et al.* Molecular analysis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Hong Kong Chinese patients. *Steroids*, 2011; **76**: 1057–62.
25. Jang JH, *et al.* Multiplex ligation-dependent probe amplification assay for diagnosis of congenital adrenal hyperplasia. *Ann Clin Lab Sci*, 2011; **41**: 44–7.
26. White PC, Curnow KM, Pascoe L. Disorders of steroid 11 beta-hydroxylase isozymes. *Endocr Rev*, 1994; **15**: 421–38.

27. Nimkarn S, New MI. Steroid 11beta-hydroxylase deficiency congenital adrenal hyperplasia. *Trends Endocrinol Metab*, 2008; **19**; 96–9.
28. Parajes S, Loidi L, Reisch N, *et al.* Functional consequences of seven novel mutations in the CYP11B1 gene: four mutations associated with nonclassic and three mutations causing classic 11{beta}-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2010; **95**: 779–88.
29. Khattab A, Haider S, Kumar A, *et al.* Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11beta-hydroxylase deficiency. *Proc Natl Acad Sci U S A*, 2017; **114**: E1933–40.
30. Auchus RJ. The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol Metab Clin North Am*, 2001; **30**: 101–119, vii.
31. Idkowiak J, Randell T, Dhir V, *et al.* A missense mutation in the human cytochrome b5 gene causes 46,XY disorder of sex development due to true isolated 17,20 lyase deficiency. *J Clin Endocrinol Metab*, 2012; **97**, E465–75.
32. Simard J, Ricketts ML, Gingras S, Soucy P, Feltus FA, Melner MH. Molecular biology of the 3beta-hydroxysteroid dehydrogenase/delta5-delta4 isomerase gene family. *Endocr Rev*, 2005; **26**: 525–82.
33. Nordenstrom A, Forest MG, Wedell A. A case of 3beta-hydroxysteroid dehydrogenase type II (HSD3B2) deficiency picked up by neonatal screening for 21-hydroxylase deficiency: difficulties and delay in etiologic diagnosis. *Horm Res*, 2007; **68**: 204–8.
34. Johannsen TH, Mallet D, Dige-Petersen H, *et al.* Delayed diagnosis of congenital adrenal hyperplasia with salt wasting due to type II 3beta-hydroxysteroid dehydrogenase deficiency. *J Clin Endocrinol Metab*, 2005; **90**: 2076–80.
35. Arlt W, Walker EA, Draper N, *et al.* Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. *Lancet*, 2004; **363**: 2128–35.
36. Fluck CE, Miller WL. P450 oxidoreductase deficiency: a new form of congenital adrenal hyperplasia. *Curr Opin Pediatr*, 2006; **18**: 435–441.
37. Miller WL, Huang N, Agrawal V, Giacomini KM. Genetic variation in human P450 oxidoreductase. *Mol Cell Endocrinol*, 2009; **300**: 180–4.
38. Baranowski ES, Arlt W, Idkowiak J. Monogenic disorders of adrenal steroidogenesis. *Horm Res Paediatr*, 2018; **89**: 292–310.
39. Krone N, Reisch N, Idkowiak J, *et al.* Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*, 2012; **97**: E257–67.
40. Metherell LA, Naville D, Halaby G, *et al.* Nonclassic lipoid congenital adrenal hyperplasia masquerading as familial glucocorticoid deficiency. *J Clin Endocrinol Metab*, 2009; **94**: 3865–71.
41. Fluck CE, Pandey Av, Dick B, *et al.* Characterization of novel StAR (steroidogenic acute regulatory protein) mutations causing non-classic lipoid adrenal hyperplasia. *PLoS One*, 2011; **6**: e20178.
42. Hauffa B, Hiort O. P450 side-chain cleavage deficiency—a rare cause of congenital adrenal hyperplasia. *Endocr Dev*, 2011; **20**: 54–62.
43. Parajes S, Kamrath C, Rose IT, *et al.* A novel entity of clinically isolated adrenal insufficiency caused by a partially inactivating mutation of the gene encoding for P450 side chain cleavage enzyme (CYP11A1). *J Clin Endocrinol Metab*, 2011; **96**: E1798–806.

5.9.2 Modern Management of Congenital Adrenal Hyperplasia and Prospects for the Future

Richard J. Auchus

Introduction to Congenital Adrenal Hyperplasias

Genetic defects in the enzymes and cofactor proteins essential for cortisol production (Figure 5.9.2.1) lead to congenital adrenal hyperplasia (CAH). As the name implies, the defects are inborn errors, which typically lead to overgrowth of the adrenal glands. The lack of negative feedback from cortisol increases adrenocorticotropin (ACTH) secretion, adrenal growth, and the futile drive to produce the deficient cortisol. Importantly, these defects span the spectrum from complete or very severe ('classic CAH') to partial or mild ('non-classic CAH'). For the purposes of this chapter, classic CAH is defined as those patients with clinically manifest cortisol deficiency, imperative glucocorticoid replacement, and risk of adrenal crisis—with the exception of 17-hydroxylase deficiency (17OHD), as will be explained subsequently.

In addition to cortisol deficiency, cortisol precursors accumulate behind the enzymatic block—not only the substrate(s) for the deficient enzyme but also further upstream steroids. Often, one or more of these steroids and/or their peripheral metabolites has biologic activity, primarily mineralocorticoids or androgens. The loss of enzyme activity forces the shunting of steroids around the block and into alternative and ordinarily minor pathways, leading to a characteristic pattern of steroids for each defect. Corticosterone is the only alternative endogenous steroid to cortisol with glucocorticoid activity, and corticosterone substitutes for cortisol in 17OHD but no other form of CAH. Consequently, the goals of CAH treatment can be divided into replacing the deficient cortisol and other steroids decreased by CAH and reducing the production or attenuating the action of unwanted, CAH-related steroid excess particular for each form of CAH. In 17OHD and non-classic CAH, only the second goal is germane.

21-Hydroxylase Deficiency

Biochemistry of 21-Hydroxylase Deficiency

By far, 21-hydroxylase deficiency (21OHD) is the most common form of CAH, both classic at ~1:16 000 worldwide and non-classic, ~1:1000 or more prevalent in certain populations [1]. The genetics and diagnosis of 21OHD is reviewed in detail in

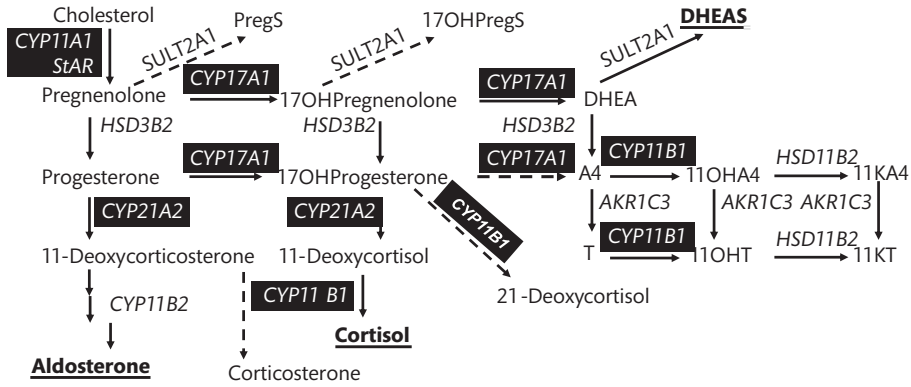


Figure 5.9.2.1. Composite diagram of adrenal steroid biosynthesis, highlighting the genetic defects causing the congenital adrenal hyperplasias (enzyme names in white on black boxes). Underlined steroid names in large font indicate major adrenal products: aldosterone, cortisol, and DHEAS. Minor pathways, either due to poor enzyme activity or limited substrate availability, are shown with dashed arrows. The 11-oxyandrogen pathways, which in part are due to conversions in peripheral tissues, are indicated at right.

Chapter 5.9.1. Upstream of the defective CYP21A2 enzyme, 17-hydroxyprogesterone (17OHP), progesterone, and other cortisol precursors accumulate, and some of these steroids are metabolized to androgens (Figure 5.9.2.2). Patients with classic 21OHD require cortisol replacement, and most also benefit from mineralocorticoid therapy, even those who make some aldosterone. In addition, treatment must also control the androgen excess to the extent that is clinically indicated. In non-classic 21OHD, treatment is employed to control androgen excess and reproductive dysfunction in children and women with these clinical manifestations. Based on conservative estimates from population studies, <10% of females and <1% of males with non-classic 21OHD are ever diagnosed with the condition, and of these roughly 70% are compound heterozygotes for classic and non-classic 21OHD alleles and thus carriers for classic 21OHD [2].

While 17OHP, androstenedione (A4), and testosterone (T) measurements have been traditionally used to diagnose and monitor therapy for 21OHD, these steroids also derive from the gonads, limiting their specificity. For example, diagnostic testing for non-classic 21OHD must be performed in the follicular phase avoid the contribution of ovarian 17OHP, and men with 21OHD in very poor control or very good control can both have a 'normal' male-range T. In addition, dehydroepiandrosterone sulphate (DHEAS), the most abundant steroid in the circulation through most of life, is elevated in non-classic 21OHD but normal or low in classic 21OHD, contrary to simple predictions by inspection of Figure 5.9.2.1.

The adrenals of patients with 21OHD, however, still contain the enzyme 11-hydroxylase (CYP11B1), and most of the 3-keto-Δ^4-steroids that accumulate in 21OHD are substrates for CYP11B1. For example, CYP11B1 converts 17OHP to the 11-oxygenated steroid 21-deoxycortisol (21dF), and 21dF is more specific than 17OHP for diagnosis of non-classic 21OHD [3] and for determination of carrier status [4]. The 3β-hydroxy-Δ^5-steroids pregnenolone and 17-hydroxypregnenolone, which also accumulate in 21OHD, are substrates for the sulfotransferase SULT2A1, and pregnenolone sulphate rises in classic 21OHD to about 40% the abundance of DHEAS [5]. More importantly, A4 is an excellent substrate for CYP11B1, and circulating concentrations of 11β-hydroxyandrostenedione (11OHA4) exceed those of A4 in treated patients with classic 21OHD by twofold [5]. Via two consecutive intra-adrenal and/or peripheral enzymatic steps, 11OHA4 is converted to 11-ketotestosterone (11KT), which is a

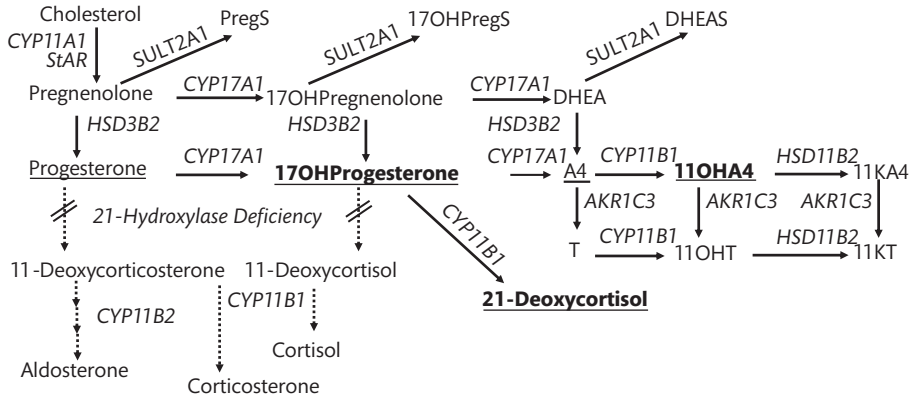


Figure 5.9.2.2. Altered steroidogenesis in classic 21OHD. The block at P450 21A2 (double crossed lines on dotted arrows) causes deficient synthesis of aldosterone and cortisol. Underlined steroid names in large font indicate major adrenal products in 21OHD: 17OHP, 21-deoxycortisol, and 11OHA4. Lack of substrate production also precludes downstream conversions, also indicated with dotted arrows. Note that pregnenolone sulphate (PregS) accumulates comparably to DHEAS in classic 21OHD, and progesterone (underlined in normal size font) is also elevated.

biologically relevant androgen, particularly in 21OHD. In teleost (bony) fishes and some reptiles, 11KT is the major gonadal androgen, and 11KT activates the human androgen receptor nearly as potently and effectively as T [6], depending on model system used. Furthermore, serum 11KT is roughly twice as high as T in treated patients with classic 21OHD, suggesting that 11KT is the dominant androgen in most patients with classic 21OHD [5, 7]. The 11KT correlates directly with T in women with classic 21OHD, as both steroids derive entirely or mostly from the adrenals. Conversely, 11KT correlates inversely with T in men with classic 21OHD, reflecting the testicular origin of T in well-controlled men and primarily adrenal origin of both T and 11KT in those with poor disease control.

Paediatric Care of Classic 21-Hydroxylase Deficiency

The diagnosis of a neonate with classic 21OHD is a medical emergency, whether the diagnosis is derived from newborn screening, clinical symptoms, or presence of genital ambiguity. Newborns with adrenal insufficiency of any cause are very vulnerable to crises due to several factors, including small body size and high surface-to-volume ratio (and propensity to fluid loss), resistance of the infant kidney to the action of mineralocorticoids, and low sodium content of the diet, particularly breast milk with 7 meq/L. Intentional yet modest overtreatment with glucocorticoid is common practice at first, to gain control of the de-repressed hypothalamic–pituitary–adrenal (HPA) axis and to begin to reverse the adrenal hyperplasia. If the dose is not excessive, the child will gradually 'grow into the dose', whereas growth suppression indicates overzealous treatment and mandates gradual dose reduction until growth normalizes. A typical dose of hydrocortisone will be about 25 mg/m^2/d in three divided doses [8]. Fludrocortisone acetate, used to replace aldosterone, is dosed at 0.1–0.4 mg/d—the same dose used in adults, to account for the poorly responsive infant kidney and tenuous volume status. Sometime in the first year of life, the hydrocortisone dose should be reduced to <17 mg/m^2/d, as higher doses impair growth [9].

During childhood, control of androgen excess is important to limit bone age advancement, which will compromise adult height. Most clinical androgen and oestrogen assays lack the accuracy to quantify the minute amounts needed to advance bone age, and 17OHP values are used as a surrogate for sex steroid measurements to gauge disease control. Growth rate, bone age, and weight are used with measurement of steroids, electrolytes, and plasma renin to adjust therapy. Controlled studies comparing various dosage regimens and monitoring schemes are lacking, which leads to considerable variations in practice style.

For girls with masculinized genitalia, clitoral reduction and vaginoplasty surgeries have been performed in the first year of life after stabilization of the medical condition. The purpose of the surgery is to establish functional anatomy for menses, voiding, and intercourse. For example, a urogenital sinus can cause vaginal voiding, the pooling of urine in the vaginal component, which leaks out upon standing. In recent years, patients, advocacy groups, and the medical community have broadly considered the factors that determine the optimal timing of reconstructive surgeries [10]. Evidence supporting early versus late timing are limited to retrospective studies. Many women who have had early surgeries endorse their parents' decision [11], while others report dissatisfaction and regret [12, 13]. Sexual activity and satisfaction are reduced women with classic 21OHD, and genital anatomy with or without reconstructive surgery is only one of many factors contributing to outcomes. Regardless of whether the child has surgery or not, control of androgen excess is important to limit further masculinization, and good control can lead to substantial clitoral regression during early childhood [14].

Adult Care for Classic 21-Hydroxylase Deficiency and Transitioning from Childhood

This chapter will focus on the management of adults with 21OHD, and adult care begins with the transition from paediatric care. For patients with 21OHD and many other chronic conditions, the transition process should be an organized series of step with the acquisition of skills and knowledge for greater self-management [15]. Unfortunately, the process is abrupt and haphazard for many patients. For example, a centre of excellence for CAH in Germany reported that less than half of young adults with 21OHD attended more than one appointment in the adult endocrinology clinic [16] and that many choose to discontinue therapy until suffering an adrenal crisis [17]. Guidance for coordinated care transition has been described [18, 19], yet logistic and financial barriers limit the widespread implementation of these strategies. Low income and lack of strong support structure are two documented risk factors for poor outcomes following transition to adult care for other chronic diseases arising in childhood [20]. At a minimum, adolescent should demonstrate a basic understanding of their disease and the reasons for treatment, the procedures to access medical care, and emergency protocols prior to transitioning to the adult clinic.

The natural history study of 21OHD from the National Institutes of Health [21], the CaHASE study from the United Kingdom [22], and other cohort studies from Europe [23] demonstrate that more adults with 21OHD suffer complications from supraphysiologic glucocorticoid exposure than from undertreatment. Common features include striae, obesity [22], adverse cardiovascular risk profiles [21, 23, 24], and reduced bone mineral density [25, 26] or occasionally osteoporosis [21, 22]. Consequently, unlike the child for whom normal growth and puberty are high priorities, the care for the adult with 21OHD must strongly consider strategies to mitigate the long-term health consequences of chronic glucocorticoid therapy.

Glucocorticoid Therapy

For motivated patients who have been consistently adherent to prescribed therapy with hydrocortisone during adolescence and show good disease control, there is no reason to significantly change treatment. Although surprisingly low doses of hydrocortisone such 10 mg BID sometimes achieves good control of adrenal androgens [27], most patients require more aggressive regimens of 5–10 mg TID with meals. The major advantages of hydrocortisone include the capacity to deliver a high peak morning glucocorticoid exposure as occurs in normal physiology and the low risk of serious complications from overtreatment, due to short half-life of about 1 hour. The short half-life of hydrocortisone is also its main disadvantage, as ACTH-driven adrenal steroid production tends to rise prior to the first morning dose when drug exposure and negative feedback are waning [28–30]. Consequently, the strategy when using hydrocortisone is to somewhat overtreat during the daytime in order to limit the ACTH rise in the early morning when exposure is poor.

For many adults, particularly those with compromised fertility or those whose schedules make midday dosing difficult, hydrocortisone therapy restores well-being without side effects but fails to control adrenal androgen production sufficiently to meet the patient's goals. One option is to continue the hydrocortisone dosing similar to conventional regimens for adrenal insufficiency—to replace the cortisol deficiency—and then add a small dose of a long-acting glucocorticoid at bedtime, to keep the ACTH suppressed until the first morning hydrocortisone dose. Although the use of two different glucocorticoids might seem paradoxical, the analogy would be a bedtime dose of long-acting insulin to control the fasting glucose rise from the dawn phenomenon. For example, a regimen might consist of hydrocortisone, 15 mg on arising and 5 mg in the early afternoon, plus 0.5–2 mg prednisolone, or 0.1–0.25 mg dexamethasone at bedtime. Prednisone, which is popular in the United States for its immunosuppressant properties, should be avoided, as prednisone is a prodrug that must be metabolized in the liver by 11β-hydroxysteroid dehydrogenase type 1 to prednisolone, yielding delayed and variable exposure to the active agent. 'Inverse diurnal rhythm' schedules, where the largest glucocorticoid dose is given at bedtime, should be avoided as well. These regimens do not replace the cortisol deficiency well and overtreat during the nighttime, when glucocorticoids are most likely to cause adverse effects such as weight gain and poor sleep. Instead, the dose of evening glucocorticoid should be smaller than those provided during the day and the minimum amount to maintain drug exposure until the first-morning dose [31].

Should the two-drug regimen be too complicated or insufficient to maintain adrenal androgens at goal, prednisolone as a single morning dose sometimes suffices to replace cortisol deficiency and control androgen production [32]. If disease control deteriorates, a second small bedtime dose of the same drug might be added, maintaining the larger dose in the morning for adrenal replacement. While dexamethasone is a potent, inexpensive, widely available, and long-acting glucocorticoid, its potency and duration of action render patients at high risk of adverse metabolic consequences during chronic therapy, and morning dexamethasone, due to its kinetics, does not replace the cortisol deficiency as well as other glucocorticoids. Dexamethasone also does not have any mineralocorticoid activity, which needs to be taken into account. Dexamethasone therapy is best limited to cautious combination with hydrocortisone and for androgen excess refractory to all other approaches. Doses of 0.25–0.75 mg once daily or rarely divided BID are typical, but even apparently modest doses can cause severe cushingoid manifestations when used over months to years. Whereas liquid or suspension preparations of hydrocortisone are no longer manufactured on an industrial scale due to poor consistency [33], liquid prednisolone and dexamethasone are generally available. Although in some ways inconvenient and messy, the liquid preparations when delivered precisely using syringes allow for continuous and careful dose titration to avoid overtreatment. A new multiparticulate granule preparation of hydrocortisone is in advanced stages of development [34]. Table 5.9.2.1 summarizes glucocorticoid options.

Regardless of glucocorticoid choice, patients are instructed to increase doses for significant physical stress and illness. For states of greater fluid loss than replacement, such as high fever, diarrhoea, and vomiting, the patient should double the dose until well for a full day.

Table 5.9.2.1 Glucocorticoid treatment options for CAH

Drug(s)	Advantages	Disadvantages
Hydrocortisone monotherapy or + bedtime prednisolone/ dexamethasone	Well tolerated Ideal cortisol replacement	Multiple daily doses Limited potency
Prednisolone/ methylprednisolone	Once/twice daily Good androgen control	Modest side effects Fair cortisol replacement
Dexamethasone	Once/twice daily Best androgen control Crosses placenta	High side effects Titration difficult Poor cortisol replacement

If the patient is too sick to take pills or to stand without assistance, the patient should receive 100 mg hydrocortisone hemisuccinate intramuscularly and taken to the hospital immediately. Patients who are vomiting are particularly vulnerable to adrenal crisis, and such illness should be approached with a low threshold for parenteral hydrocortisone and hospital transport (also see Chapter 5.8.2 regarding management of adrenal crisis).

Mineralocorticoid Replacement

Perhaps the most common error in treating all forms of adrenal insufficiency is inattention to mineralocorticoid replacement and volume status [35]. Most adults with 21OHD benefit from 9α-fludrocortisone acetate regardless of whether they were described as 'salt wasting' during childhood. The dose is typically 0.05–0.2 mg once daily, but some patients require 0.4–0.6 mg daily as a single dose or divided BID. Persistent fatigue unresponsive to higher glucocorticoid dosing often represents chronic volume depletion as a result of mineralocorticoid deficiency. Hydrocortisone provides some mineralocorticoid activity, particularly at higher doses, but prednisolone, methylprednisolone, and dexamethasone have little to no mineralocorticoid activity. Consequently, any patients converted from hydrocortisone to another glucocorticoid might need to increase their dose of 9α-fludrocortisone acetate to maintain euvolemia.

Special Considerations for Infertility in Women with 21OHD

Because classic 21OHD causes 46,XX disorders of sex development (DSD) and androgen excess in women, attention has been focused on infertility in women with this disease. A major impediment to fertility is simply the paucity of women with classic 21OHD who ever attempt pregnancy, estimated to be 25% overall and 10% of women with null alleles [36]. The fecundity rates for the minority who do attempt pregnancy, however, is more than 90%, equivalent to that of the general population—with proper treatment [37]. Glucocorticoid regimens that are adequate for replacing the cortisol deficiency, maintaining regular menses, and preventing unwanted acne and facial hair might require intensification to restore fertility. In particular, increased adrenal-derived progesterone, which accumulates upstream of 17OHP, is a major mechanism of impaired fertility in women with 21OHD. Sustained progesterone production through the follicular phase of the cycle causes changes in cervical mucus and thinning of the endometrial lining, similar to oral contraceptive pills. Thus, attention to suppression of

follicular-phase progesterone is critical for managing women with 21OHD attempting to conceive, and a bedtime dose of a long-acting glucocorticoid (but not dexamethasone, as it passes the placenta in pregnant women) is often necessary [37].

Special Considerations for Infertility in Men with 21OHD

Paradoxically, although men with classic 21OHD do not have 46,XY DSD, men appear to have greater problems with fertility than women with 21OHD [38–40]. One major mechanism of decreased sperm production is that high production of adrenal-derived androgens suppresses gonadotropin secretion [41, 42]. similar to exogenous androgens. The other major cause of infertility is testicular adrenal rest tumours (TART), which develop in 30–50% of adolescent and adult males with 21OHD when ultrasound is used for sensitive detection [21, 38]. TART appear in the posterior aspect of the testes and tend to be firm, irregular bilateral masses. The origin of TART and their mechanisms of formation are not known. Although most men with large TART have poor control of adrenal androgen synthesis, one study found no correlation of TART development with traditional laboratory parameters of disease control [43]. Originally, TART were thought to form from adrenal cortex cells, which migrated to the testis during embryonic development and later underwent hyperplasia under robust ACTH stimulation. Subsequently, TART cells were shown *in vitro* to produce the same pattern of steroids as the adrenal cortex of patients with 21OHD [44] and to express an overlapping set of enzymes unique to adrenal cortex and Leydig cells [45]. These data suggest that TART derive from Leydig progenitor cells, which have become partially transdifferentiated under the influence of elevated ACTH [38, 39, 46].

Because the testis is confined to a capsule, the growth of any testicular mass impairs blood flow to the normal cells of the testicular tubules. The chronically impaired perfusion leads to dysfunction and fibrosis of the normal testis. A high serum follicle-stimulating hormone (FSH) heralds the permanent loss of Sertoli cells and indicates chronic damage. As TARTs grow, the normal testis tissue suffers from compression, impaired blood flow, functional impairment, and fibrosis. To induce regression of TART, either dexamethasone (once or twice daily) or divided doses of hydrocortisone plus bedtime dexamethasone are required to keep ACTH suppressed all day long for weeks to months [47]. Sperm production and fertility are often restored with intensified therapy, but some TART do not regress with medical therapy, particularly when longstanding and already associated with fibrotic transformation [48]. Testis-sparing surgery effectively and usually permanently removes the TARTs, but fertility and testicular T production are rarely restored [38, 49]. Hence, testis surgery is usually employed only after the patient accepts the permanent loss of fertility in order to remove the mass effect. The presence of TART and an FSH >35 IU/L are poor prognostic signs for fertility in men with 21OHD [50]. One case report [51] describes TART regression and restoration of fertility in a man with 21OHD after 2 years of treatment with mitotane, an adrenolytic drug otherwise used in the context of adrenocortical carcinoma. Because the presence of TART is so common and its consequences so grave, boys with 21OHD should have a baseline testicular ultrasound upon reaching adulthood and periodically thereafter [52], as well as testicular exams at least annually.

Additional Treatment Options

Bilateral adrenalectomy for 21OHD patients was previously advocated as a means to eliminate adrenal-derived androgens without the use of supraphysiologic glucocorticoid therapy. Although short-term results appear successful [53] and fertility might even be restored [54, 55], long-term outcomes are not durable. One study found recurrent hyperandrogenaemia in 8 of 18 patients following bilateral adrenalectomy [56]. TART may subsequently develop in men [56] or adrenal rest tumours in the ovaries and pelvis in women [57, 58]. A pituitary corticotrope adenoma has been described in one patient with 21OHD following adrenalectomy [59], and patients appear to be more vulnerable to potentially fatal adrenal crises following bilateral adrenalectomy [55, 56]. For these reasons, bilateral adrenalectomy is rarely recommended any longer.

As with any form of androgen excess, mechanical epilation methods, including shaving, waxing, plucking, electrolysis, and laser hair removal, reduce the dependence on pharmacotherapy and thus might spare some glucocorticoid exposure. Topical application of 13.9% eflornithine cream is also a useful adjunct for controlling hirsutism, particularly for small areas of skin, but the cost can be prohibitive. The ethinyl oestradiol component of oral contraceptive pills increases circulating sex-hormone binding globulin concentrations and thus lowers the free and bioavailable fractions of T even without lowering total T. Androgen receptor antagonists might be employed as well, but spironolactone, which is commonly used for this purpose, is also a diuretic and mineralocorticoid receptor antagonist (MRA), which can cause serious volume depletion in women with classic 21OHD; its interference with the progesterone receptor can cause additional menstrual cycle irregularities. Terminal hairs never revert back to vellus hairs, and hair follicles cycle between growing, resting, and shedding phases. Medical treatments of androgen excess lead to thinner terminal hairs and more time in the resting phase. Electrolysis and laser epilation permanently destroy hair follicles, but only those in the growing phase, which are visible at the time of treatment.

Non-classic 21-Hydroxylase Deficiency

Patients with non-classic 21OHD generally do not have manifest cortisol (or aldosterone) deficiency but only androgen excess. However, partial adrenal insufficiency manifesting in times of stress has been described, as cortisol deficiency depends on the degree of enzymatic impairment, which is by definition on a continuous spectrum. When ascertained as children for premature androgen exposure and/or advanced bone age, hydrocortisone therapy is conventionally employed to attenuate the ACTH-mediated excess of adrenal-derived androgens. Logically, non-classic 21OHD is easier to manage than classic 21OHD, and lower doses of glucocorticoids can be used than in classic 21OHD. Upon reaching adulthood, treatment in boys is normally discontinued [8], except in some patients with genotypes at the more severe end of the non-classic 21OHD spectrum. In women diagnosed with non-classic 21OHD due to irregular menses and/or androgen excess and those diagnosed as children with persistent symptoms upon reaching adulthood, hydrocortisone treatment might be continued. More potent glucocorticoids are rarely necessary but can be used instead of hydrocortisone but sparingly, such as dexamethasone 0.25 mg at bedtime 3 days per week [60]. Similarly, mechanical epilation, oral

contraceptive pills, and antiandrogens including spironolactone are adjuncts or alternatives to glucocorticoids for these patients.

One area of confusion relates to pregnancy in women with non-classic 21OHD. The notion that women with non-classic CAH commonly suffer from reduced fecundity largely derives from ascertainment bias, as infertile women are more commonly tested for non-classic 21OHD than fertile women. Given that the prevalence of non-classic 21OHD based on allele frequencies is at least 1:1000 in the general population, most women with the disease are never diagnosed, and those who come to medical attention tend to have more severe disease, with 70% being compound heterozygotes for classic and non-classic 21OHD alleles [2]. In a French study, 83% of women with non-classic 21OHD became pregnant in 1 year with or without glucocorticoid therapy [61]. Of 139 pregnancies in a cohort from Israel of non-classic 21OHD women without male factor infertility, the mean time to conception was 4.0 months without and 3.3 months with glucocorticoid therapy, and only 17 pregnancies occurred after starting glucocorticoids following failure to conceive after an average of 10.2 months [62]. In contrast, the miscarriage rate in these women is as high as 26% in untreated women and, in retrospective series, is closer to that of the normal population (6%) in those treated with hydrocortisone therapy [61]. A more recent study showed a 23% rate of miscarriage without and 14% with glucocorticoid therapy, but the latter group included only six cases [62]. Conclusions should be drawn with caution, as the data are retrospective and not reproduced in other series. If hydrocortisone is used during pregnancy, the dose should be physiologic or slightly lower (10–20 mg/d) to prevent iatrogenic adrenal axis suppression.

Monitoring and Adjusting Treatment of 21OHD

The history taking in the adult with classic 21OHD includes prior therapy—changes, lapses in medication adherence, and frequency of adrenal crises. The patient should be asked about weight changes, salt craving, easy bruising, and orthostasis. For women, menstrual regularity and androgen excess features are important parameters, and women should be asked if they use epilation or other measures to control hirsutism and acne. Men should be asked about testicular pain and masses. The desire for current or future fertility should be explored, and it is alarming how many patients have been incorrectly counseled that fertility is not possible in 21OHD.

Physical examination should include sitting and standing blood pressure and heart rate to probe volume status and adequacy of mineralocorticoid replacement. Although physical exam findings are not very helpful for demonstrating adequacy of cortisol replacement, cushingoid features consistently develop with months of overtreatment. Weight gain frequently occurs with glucocorticoid therapy, yet weight gain is also common in untreated adults and difficult to ascribe to the glucocorticoids. Dermal atrophy is a very specific finding for glucocorticoid excess. For example, most adult treated with long-term dexamethasone develop dermal atrophy even in the absence of other cushingoid features. Dermal atrophy and capillary fragility lead to the easy bruising and purple, non-blanching striae in the abdomen but also flanks, thighs, and upper arms. Moon facies and plethora, proximal muscle weakness, and disproportionate fat accumulation in the head and neck area (supraclavicular and dorsocervical fat pads) are also quite specific for glucocorticoid excess. Women receive an assessment of acne and hirsutism, interpreted in the context of recent epilation, and annual gynaecologic exams. Men have at least annual testicular exams for atrophy and TART development.

Plasma renin activity (or mass concentration) correlates inversely with plasma volume, and reflects the adequacy of mineralocorticoid replacement. Low standing blood pressure, high serum potassium, and high plasma renin indicate inadequate fludrocortisone acetate dosing, and excessive dosing yields the opposite results. 17OHP is commonly used to titrate therapy in children, as normal 17OHP values assure that even small amounts of sex steroids capable of advancing bone age are not being produced. The utility of 17OHP in adults is limited, primarily because values fluctuate dramatically with time after each glucocorticoid dose, and adults do not experience detrimental consequences of small, brief rises in androgens relative to gonadal production. DHEAS is not elevated in classic 21OHD patients [63], and the only utility of DHEAS is to gauge adherence and exposure to glucocorticoid therapy, as treated patients should have suppressed DHEAS. A4, which is one step from T, is perhaps the best commonly available biomarker of adrenal androgen production, even though 11OHA4 is more abundant in 21OHD patients. A4 values fluctuate less than 17OHP throughout the day, and T values fluctuate even less than A4. For these reasons, A4 and T values are the primary laboratory parameters used for evaluating disease control and for titrating glucocorticoid dosing, but with some caution. Given that 11KT is typically more abundant than T in patients with classic 21OHD, T values tend to underestimate the production of adrenal-derived androgens.

For women with classic 21OHD who are not currently attempting to conceive, the main parameters of disease control are clinical assessments: regular menses, androgen excess features at goal, and absence of cushingoid features [64]. It is inappropriate to define goal values for A4 and T, but trends and correlates to good clinical features are valuable information. For example, if a woman with classic 21OHD suddenly develops amenorrhea, the question arises whether the reason is deterioration in disease control. If the A4 and T values are significantly increased from previous measurements when menses were regular, then poor control of 21OHD is likely the reason. If the laboratory data are unchanged, other reasons for amenorrhea should be sought.

On the other hand, regular menses alone does not necessarily indicate adequate disease control for women who desire to conceive. Upstream of 17OHP, progesterone (P4) accumulates as well in 21OHD, and chronically elevated P4 is a major cause of irregular menses and amenorrhea in women with 21OHD, because P4 withdrawal does not occurs [28]. Chronic elevation of P4 into luteal-phase values (>2 ng/ml or 6 nmol/L) thins the endometrial lining, alters cervical mucus, and interferes with menses, similar to progestin-only contraceptives. For women with 21OHD attempting to conceive, glucocorticoid therapy should be titrated to maintain the follicular-phase serum progesterone <0.6 ng/ml (<2 nmol/L) [37].

A quandary in managing men with 21OHD is that T can be in the normal male range for men in good control as well as those in poor control; the difference is whether the androgens derive from the adrenal or the testes [64]. Simultaneous A4 and luteinizing hormone (LH) values allow one to sort out the origin of most T. The Leydig cells of the testes are the only cells in postnatal human beings that normally contain the enzyme 17β-hydroxysteroid dehydrogenase type 3, which efficiently converts A4 to T. Thus, when T is produced

from the testes, the A4/T ratio is <0.5 and often <0.2. When the T is primarily of adrenal origin, the A4 exceeds T, and the A4/T ratio is >1 [31]. In addition, a suppressed LH is consistent with high adrenal androgen production. For example, a 20-year-old man with classic 21OHD and poor adherence through adolescence can have full secondary sexual characteristics, a T in the normal male range (~300 ng/dl or 10 nmol/L) yet have atrophic testes, undetectable LH, and A4 four times the T value. This man has never progressed through puberty and has been exposed only to adrenal-derived androgens. For post-pubertal men who later lapse into poor control, the LH can be suppressed with a T near the low end of normal (200 ng/dl or 7 nmol/L). One reason for this discrepancy between LH and T is that men with 21OHD in poor control have high 11KT. In men with 21OHD, 11KT correlates roughly inversely with T [5]. For men in poor control, the adrenal-derived 11KT is higher than T, and testicular T synthesis is negligible. For men in good control, testicular T production is normal, and little 11KT derives from the adrenal. As in women, a sex-hormone binding globulin measurement allows determination of free and bioavailable T. A high FSH value indicates severe testicular damage, and a semen analysis is indicated when evaluating for infertility.

Other Forms of Congenital Adrenal Hyperplasia

3β-Hydroxysteroid Dehydrogenase Deficiency (HSD3B2D)

This disease is the form of CAH most similar physiologically to 21OHD. Patients with classic HSD3B2D have cortisol and aldosterone deficiency and are vulnerable to adrenal crises as infants. The major difference is that the androgen excess is milder than 21OHD, such that girls have modest clitoral enlargement but generally little or no labioscrotal fusion [65]. Because the testicular Leydig cells and ovarian theca cells employ the enzyme in the A4 and T biosynthetic pathway, boys with HSD3B2D have undervirilization and often gynaecomastia. Thus, HSD3B2D is one of two forms of CAH that cause DSD in both 46,XX and 46,XY babies [66]. Boys often require T supplementation at the time of normal puberty and throughout adulthood. Nevertheless, one young man with severe deficiency was found to have a normal semen analysis, possibly due to aberrant compensation via up-regulation of the HSD3B1 enzyme in the Leydig cells [67]. There are no reports of fertility in women with HSD3B2D.

The diagnosis of HSD3B2D is best made from the 17-hydroxypregnenolone/cortisol ratio, which must be at least 6 (and typically >300) standard deviations above normal [68, 69]: typical values are a cortisol <5 μg/dl (<140 nmol/L) and a 17-hydroxypregnenolone of more than 6000 ng/dl (>160 nmol/L). Non-classic HSD3B2D was once thought to be common, but most women with mildly elevated delta-5 steroids such as DHEAS have polycystic ovary syndrome and no mutations in the *HSD3B2* gene [70]. One hypothesis is that these women have functional impairment of the enzyme due to high intra-adrenal delta-4 steroids, which also bind to the enzyme and reduce activity via product inhibition [71].

Monitoring for mineralocorticoid replacement in HSD3B2D is the same as for 21OHD. Glucocorticoid replacement goals in children and adults is likewise similar, and the androgen excess is milder than in classic 21OHD. Laboratory monitoring should not focus on delta-5 steroids, as some patients have persistently high DHEA despite adequate therapy [72]. The adrenal-derived delta-5 steroids are metabolized in the liver, skin, and other tissues one step beyond the enzymatic block to P4, 17OHP, and A4 (Figure 5.9.2.3). These delta-4 steroids do not significantly re-enter the adrenal steroidogenic pathways and progress to final steroid products (cortisol) but represent useful biomarkers for titrating therapy. Some A4 is converted to T via AKR1C3, and limited data suggest that 11OHA4 is low in patients with HSD3B2D [67].

17-Hydroxylase/17,20-lyase Deficiency (17OHD)

In many respects, 17OHD is the orthogonal disorder to 21OHD. Patients with 17OHD cannot produce androgens and oestrogens,

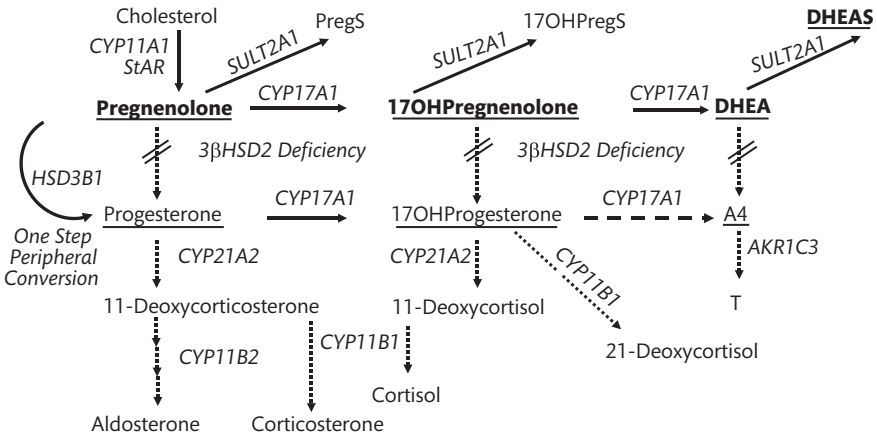


Figure 5.9.2.3. Altered steroidogenesis in classic HSD3B2 deficiency. The block at HSD3B2 (double crossed lines on dotted arrows) causes deficient synthesis of all delta-4 steroids in the adrenals and gonads; however, peripheral conversion via HSD3B1 affords increased circulating concentrations of the steroids one step past the block but not further (underlined steroid names in normal size font). Underlined steroid names in large font indicate major adrenal products in HSD3B2D: pregnenolone, 17-hydroxypregnenolone, and DHEA. Lack of substrate production also precludes downstream conversions, also indicated with dotted arrows. DHEAS is elevated in HSD3B2, and other delta-5 steroid sulphates might also be elevated (data are lacking). Limited data indicate that 11-oxyandrogens are not produced in HSD3B2.

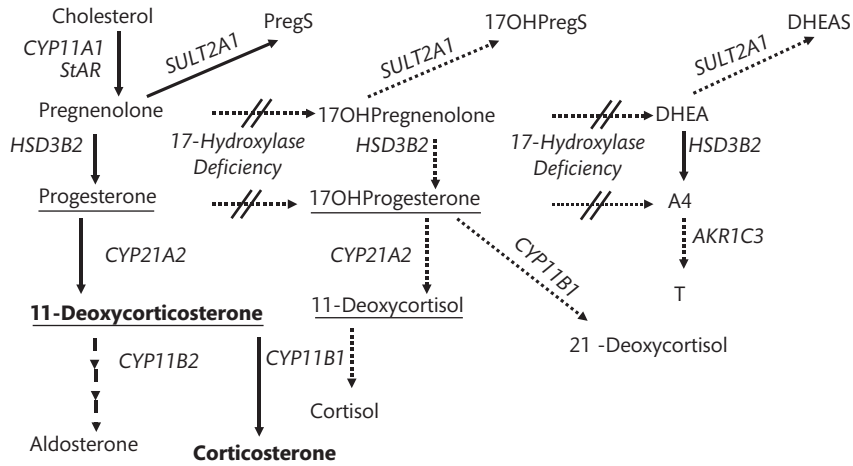


Figure 5.9.2.4. Altered steroidogenesis in 17OHD. The block at P450 17A1 (double crossed lines on dotted arrows) causes deficient synthesis of 17-hydroxysteroids and 19-carbon steroids. Underlined steroid names in large font indicate major adrenal products in 17OHD: 11-deoxycorticosterone and corticosterone. Lack of substrate production also precludes downstream conversions, also indicated with dotted arrows. Note that progesterone is also elevated in 17OHD (underlined in normal size font), and pregnenolone sulphate (PregS) might also be elevated (data are lacking). Aldosterone production (dashed arrows) is low due to renin suppression.

because the same *CYP17A1* gene, which encodes the CYP17A1 enzyme, is expressed in the adrenal and gonads [73] and is required for both cortisol and sex steroid synthesis. As previously mentioned, 17OHD is the only form of CAH without adrenal insufficiency, because the 17-deoxysteroid corticosterone, which is also a glucocorticoid, accumulates and compensates for the lack of cortisol [74]. Patients with 17OHD are rarely diagnosed as infants, unless an inguinal hernia or discordant prenatal 46,XY karyotype in a phenotypically female newborn is discovered. Upstream of corticosterone, 11-deoxycorticosterone (DOC) and P4 also accumulate (Figure 5.9.2.4), and DOC has mineralocorticoid activity. Consequently, hypertension and hypokalaemia often but not always develop during adolescence or early adulthood [75]. Thus, the typical presentation of 17OHD is pubertal failure in a phenotypic female with early-onset hypertension [66]. There is considerable variation in the clinical spectrum depending on the severity of the enzymatic defect. Because androgen biosynthesis requires the sequential 17-hydroxylase and 17,20-lyase activities of CYP17A1—two conversions catalysed by the same enzyme—the mutations necessarily impair androgen production greater than cortisol, and some elevation of 17-deoxysteroids is always present even in the absence of clinical manifestations. A non-classic form of 17OHD has not been described, but missense mutations that preferentially impair 17,20-lyase activity and cause isolated 17,20-lyase deficiency are known [76–78].

The diagnosis of 17OHD is established when cortisol and sex steroids are low and the steroids above the enzymatic block (P4, DOC, corticosterone) are all elevated. Infants and children with normal blood pressure and serum potassium for age do not require treatment, even when the diagnosis is established early in life, because corticosterone provides glucocorticoid activity. If and when hypertension and/or hypokalaemia develop, glucocorticoid therapy the suppress DOC formation is the traditional therapy. The dose is titrated primarily on the basis of blood pressure, serum potassium, and plasma renin, although renin can remain low for years after normalization of blood pressure. When using glucocorticoids, the patient should not be rendered cushingoid or treated with supraphysiologic doses, as alternatives to glucocorticoids are available. MRA, including spironolactone and eplerenone, are effective alternatives to glucocorticoids and act by blocking the action of DOC at the receptor [79]. Combination therapy with a small dose of hydrocortisone and as much MRA as necessary to normalize blood pressure and serum potassium is also a good option [74].

Sex steroid replacement is needed at the time of puberty or upon diagnosis if delayed until adulthood. Most patients are phenotypic females, and oestrogen replacement alone is sufficient in patients with 46,XY karyotype who lack a uterus. Gonadectomy should be considered in 46,XY patients with intra-abdominal testes, as gonadoblastoma may develop, although the absolute risk is not known. Patients with 46,XX karyotype have a uterus and therefore require periodic withdrawal bleeding with a progestin. A single case of term pregnancy has been reported in a patient with incomplete but classic 17OHD [80]. The treatment strategy was to suppress P4 with dexamethasone, prepare the uterine lining with oestradiol valerate, induce ovulation with gonadotropins, and implant embryos in a subsequent cycle after intracytoplasmic sperm injection. The amount of enzyme activity needed to enable pharmacologic induction of ovulation is not known, and gonadotropins should be used with extreme caution. Most women with 17OHD already have chronically elevated gonadotropins and chronic anovulation, which tends to cause the development of large ovarian cysts [81]. Undervirilized males with partial 17OHD or isolated 17,20-lyase deficiency with male gender identity require testosterone therapy and sometimes surgical correction of hypospadias. Fertility in men with 17OHD has not been described.

11β-Hydroxylase Deficiency (11OHD)

In 11OHD, the defect in CYP11B1 disrupts the final step of cortisol biosynthesis (Figure 5.9.2.5), and the phenotype is somewhat a combination of 17OHD and 21OHD. Similar to 17OHD, DOC but not corticosterone accumulates upstream of the enzymatic

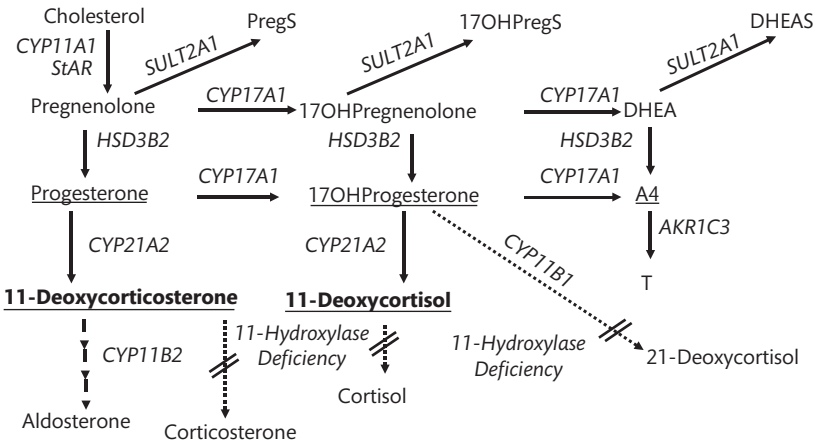


Figure 5.9.2.5. Altered steroidogenesis in classic 11OHD. The block at P450 11B1 (double crossed lines on dotted arrows) causes deficient synthesis of corticosterone, 21-deoxycortisol, and cortisol. Underlined steroid names in large font indicate major adrenal products in 11OHD: 11-deoxycorticosterone and 11-deoxycortisol. Progesterone, 17OHP, and A4 (underlined in normal size font) are also elevated in 11OHD, and aldosterone production (dashed arrows) is low due to renin suppression.

block, and patients with 11OHD tend to develop hypertension and/or hypokalaemia during adolescence—yet blood pressure does not correlate well with serum DOC [82]. Unlike 17OHD and similar to 21OHD, the steroid diversion leads to androgen excess as well, and girls are born with variable virilization of the external genitalia, which can be severe. Also like 21OHD, infants with 11OHD produce neither cortisol nor corticosterone and are prone to develop adrenal crises during illness if not treated. Thus, glucocorticoid therapy to replace the cortisol deficiency is required at birth or upon diagnosis.

Also like 21OHD, androgens, P4, and 17OHP are all elevated in 11OHD, but 17OHP is not elevated as much as in 21OHD. The diagnostic steroid is 11-deoxycortisol, which is the steroid immediately above the enzymatic block and is typically >10 000 ng/dL (>3000 nmol/L) in classic 11OHD [83]. Non-classic 11OHD is very rare, and the diagnosis should be considered when androgen excess presents in childhood and the 17OHP is elevated but not sufficient to diagnose non-classic 21OHD [84, 85]. Typically, the cortisol is normal, but the 11-deoxycortisol is >1800 ng/dl (>500 nmol/L).

The treatment of 11OHD throughout childhood and adulthood is similar to that of 21OHD, with divided doses of hydrocortisone but without the need for fludrocortisone acetate. More potent glucocorticoids such as prednisolone and dexamethasone are rarely necessary. In fact, MRA can be added to limit the dose of glucocorticoid used and still control blood pressure and serum potassium [66]. For 46,XX patients with 11OHD, spironolactone at doses of 50–200 mg/d functions as both an MRA and an antiandrogen, which compensates for mild residual elevations of DOC and T when physiologic doses of hydrocortisone are used. For 46,XY patients with 11OHD, eplerenone is a specific MRA that will not interfere with T action. Also like 21OHD, men are prone to developing TART and infertility [86], which would require intensification of glucocorticoid therapy. Women require glucocorticoid intensification to lower follicular-phase P4 when attempting to conceive. Only two cases of term pregnancies have been described in women with classic 11OHD [87].

P450 Oxidoreductase Deficiency (PORD)

PORD is the last form of CAH for which the molecular genetics and pathophysiology was determined [88, 89], although patients with this condition were first described clinically many years previously [90]. PORD is the most complex and variable form of CAH, in part because all patients thus far identified have incomplete deficiencies of POR function. POR mutations can alter the activities of any microsomal cytochrome P450 (CYP) enzyme, including the steroidogenic enzymes CYP21A2, CYP17A1, and CYP19A1 (aromatase); however, the influence on each enzymatic reaction is uniquely different for each missense mutation with some retained activity. Consequently, generalized statements about diagnosis, treatment, and outcomes are difficult to develop. The genetics and biochemistry of PORD is reviewed in detail elsewhere [91].

In general, P4, the only steroid upstream of all potential enzymatic defects, is always and often markedly elevated in PORD (Figure 5.9.2.6). Steroids past one impaired enzymatic step—including 17OHP and corticosterone—can be high or normal. Steroids past two impaired enzymatic steps—including cortisol, A4, and T—will be low and often very low. Oestrogen biosynthesis requires three enzymatic steps that utilize POR, and oestrogens are always very low, including maternal estriol in mothers carrying a fetus affected with PORD. Although androgen biosynthesis is always impaired, girls are born with genital virilization due to small but biologically significant amounts of 5α-reduced androgens of adrenal origin produced via the alternative pathway [88, 92]. With such complex and variable biochemical manifestations, PORD is best diagnosed with mass spectrometry profiling of urinary steroids [91].

Most patients with PORD have clinically meaningful cortisol deficiency and are prone to adrenal crises. These patients should be given replacement doses of hydrocortisone upon diagnosis and continued throughout life, and the dose is clinically titrated solely for cortisol replacement as in patients with absent adrenal function. Other patients might require stress dose hydrocortisone coverage only. Mineralocorticoid replacement is rarely required, even though CYP21A2 is required for aldosterone biosynthesis. In fact, the production of DOC and corticosterone is often increased in patients

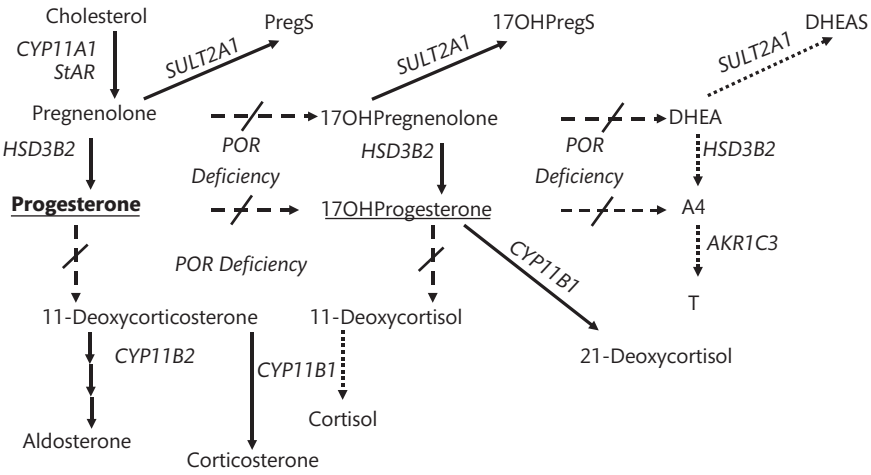


Figure 5.9.2.6. Altered steroidogenesis in PORD. The multiple and variable partial blocks at P450 21A2 and P450 17A1 (single crossed lines on dashed arrows) causes deficient synthesis of cortisol and androgens, which require two impeded transformations. Progesterone (underlined in large font) is upstream of all blocks and always accumulates in PORD; 17OHP and corticosterone (underlined in normal size font) variably accumulate depending on the specific mutation(s). Lack of substrate production precludes downstream conversions (dotted arrows).

with PORD but not to the same extent as in 17OHD. Patients with PORD rarely if ever develop early-onset hypertension despite DOC elevations, so adjunctive therapy with MRA is not required, either; however, in adulthood, hypertension regularly manifests in PORD patients. Both male and female patients require sex steroid replacement at the time of puberty and throughout adulthood, concordant with gender identity. PORD might be underdiagnosed in women with infertility, and one example of term pregnancy using ovulation induction and *in vitro* fertilization has been described [93].

Lipoid Congenital Adrenal Hyperplasia (LCAH)

LCAH derives from mutations in the *STAR* gene encoding the steroidogenic acute regulatory protein (StAR). StAR delivers cholesterol from the outer mitochondrial membrane to the cholesterol side-chain-cleavage enzyme (CYP11A1) on the inner mitochondrial membrane, where its conversion to pregnenolone occurs. Patients with classic LCAH are born with cholesterol ester-laden adrenals and make no steroids. Thus, similar to 21OHD and HSD3B2D, treatment with glucocorticoid and mineralocorticoid therapy at birth is critical for survival. More akin to PORD, these treatments are purely to replace the deficiencies, as misdirection of steroidogenesis and adrenal-derived hormone excess does not occur. Monitoring and dose titration is limited to adequacy of cortisol replacement, absence of cushingoid features, serum electrolytes, plasma renin, and standing blood pressure. StAR is also essential for normal gonadal androgen and oestrogen biosynthesis, so LCAH patients are all are phenotypically female and raised as girls. Intra-abdominal testes in 46,XY patients are non-functional and at risk for developing malignancies, and surgical orchiectomy is therefore prudent.

LCAH patients with 46,XX karyotype might have some ovulatory cycles and breast development at puberty [94], but ovarian failure typically ensues as cholesterol esters accumulate [95]. Consequently, oestrogen replacement therapy with or without progestin cycling depending on presence of a uterus is commenced in adolescence and continued throughout adulthood. Despite the severe adrenal insufficiency and pubertal failure, a few remarkable cases of pregnancy in 46,XX women with classic LCAH have been reported [96]. The first case conceived with clomiphene citrate and maintained two pregnancies to term with progesterone support. The second case involved gonadotropin stimulation, *in vitro* fertilization, and implantation of frozen embryos following uterine preparation with oestrogen [97].

Non-classic LCAH is well described, and these patients primarily show cortisol deficiency, simply because a higher circulating cortisol concentration is required to meet physiologic needs than for all other classes of steroid hormones [98, 99]. Thus, treatment of these patients is largely limited to cortisol replacement, although T supplementation is sometimes need in affected boys starting at puberty. A similar biochemical pattern to classic LCAH is observed in rare patients with P450 11A1 deficiency due to *CYP11A1* mutations, but the latter patients do not have cholesterol ester-filled adrenals [100]. Patients with P450 11A1 deficiency are managed similar as patients with classic LCAH, and limited information about long-term outcomes is available.

Experimental Therapies

In recent years, interest in improved treatment for classic 21OHD has grown [101–109]. These treatments can be grouped as improved glucocorticoid delivery approaches and glucocorticoid-sparing therapies to mitigate androgen excess. Small series of patients treated with continuous subcutaneous hydrocortisone infusion have been reported [101, 102], using an insulin pump with hydrocortisone hemisuccinate. This approach appears to be most useful for patients with rapid hydrocortisone metabolism [103] or patients with impaired absorption via the gut, but is labour-intensive and costly. Alternatively, modified-release oral hydrocortisone has been tested in once daily [104] and twice daily regimens [105], the latter yielding significantly better lowering of adrenal-derived androgens compared to conventional therapy [109]. A phase III trial of one preparation (Chronocort) has been completed primarily in Europe (NCT03062280), and one is planned for the United States (NCT03532022).

Conceptually, glucocorticoid-sparing regimens for 21OHD could add drugs to block androgen action, inhibit androgen biosynthetic enzymes, or reduce ACTH stimulation of steroidogenesis. The first such approach employed the antiandrogen flutamide and the first-generation aromatase inhibitor testolactone to reduce hydrocortisone exposure and to optimize adult height in children [106]. Subsequently, the potent CYP17A1 inhibitor abiraterone acetate, which is FDA-approved for the treatment of prostate cancer (at 1000 mg/d), was studied in six adult women with classic 21OHD. This short-term study demonstrated robust reductions is adrenal-derived androgens with 100–250 mg/d of abiraterone acetate when added to 20 mg/d hydrocortisone [27]. A phase I clinical trial of abiraterone acetate in children with 21OHD is now in progress (NCT 02574910).

Alternatively, direct suppression of ACTH secretion might be possible using small-molecule inhibitors of the corticotropin-releasing hormone type 1 receptor (CRHR1). A single-dose, dose-escalation study in eight women with classic 21OHD demonstrated significant reductions in first morning ACTH and 17OHP in half of patients with the CRHR1 antagonist NBI-77860 given at bedtime [107]. Currently, two short-term multiple-dose phase II clinical trials of CRHR1 antagonists are ongoing, crinecerfont (NCT 03525886) and tildacerfont (NCT03257462). The sterol-O-acyltransferase inhibitor ATR-101 was also studied in a dose-escalation trial that included eight men and women with classic 21OHD (NCT02804178), and an extended trial is now planned in Europe.

REFERENCES

1. Speiser PW, White PC. Congenital adrenal hyperplasia. *N Engl J Med*, 2003; **349**: 776–88.
2. Finkielstain GP, Chen W, Mehta SP, *et al*. Comprehensive genetic analysis of 182 unrelated families with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2011; **96**: E161–E72.
3. Costa-Barbosa FA, Carvalho VM, Nakamura OH, Bachega TA, Vieira JG, Kater CE. Zona fasciculata 21-hydroxysteroids and precursor-to-product ratios in 21-hydroxylase deficiency: further characterization of classic and non-classic patients and heterozygote carriers. *J Endocrinol Invest*, 2011; **34**: 587–92.
4. Costa-Barbosa FA, Tonetto-Fernandes VF, Carvalho VM, *et al*. Superior discriminating value of ACTH-stimulated serum 21-deoxycortisol in identifying heterozygote carriers for 21-hydroxylase deficiency. *Clin Endocrinol (Oxf)*, 2010; **73**: 700–6.
5. Turcu AF, Nanba AT, Chomic R, *et al*. Adrenal-derived 11-oxygenated 19-carbon steroids are the dominant androgens in classic 21-hydroxylase deficiency. *Eur J Endocrinol*, 2016; **174**: 601–9.
6. Rege J, Turcu AF, Kasa-Vubu JZ, *et al*. 11-Ketotestosterone is the dominant circulating bioactive androgen during normal and premature adrenarche. *J Clin Endocrinol Metab*, 2018; **103**: 4589–98.
7. Turcu AF, Mallappa A, Elman MS, *et al*. 11-oxygenated androgens are biomarkers of adrenal volume and testicular adrenal rest tumors in 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2017; **102**: 2701–10.
8. Speiser PW, Azziz R, Baskin LS, *et al*. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2010; **95**: 4133–60.
9. Bonfig W, Pozza SB, Schmidt H, Pagel P, Knorr D, Schwarz HP. Hydrocortisone dosing during puberty in patients with classical congenital adrenal hyperplasia: an evidence-based recommendation. *J Clin Endocrinol Metab*, 2009; **94**: 3882–8.
10. Lee PA, Nordenström A, Houk CP, *et al*. Global disorders of sex development update since 2006: perceptions, approach and care. *Horm Res Paediatr*, 2016; **85**: 158–80.
11. Binet A, Lardy H, Geslin D, Francois-Fiquet C, Poli-Merol ML. Should we question early feminizing genitoplasty for patients with congenital adrenal hyperplasia and XX karyotype? *J Pediatr Surg*, 2016; **51**: 465–8.
12. Mouriquand PD, Gorduza DB, Gay CL, *et al*. Surgery in disorders of sex development (DSD) with a gender issue: if (why), when, and how? *J Pediatr Urol*, 2016; **12**: 139–49.
13. Baratz AB. Re: 'Surgery in disorders of sex development (DSD) with a gender issue: If (why), when, and how?'. *J Pediatr Urol*, 2016; **12**: 442–3.
14. Bougneres P, Bouvattier C, Cartigny M, Michala L. Deferring surgical treatment of ambiguous genitalia into adolescence in girls with 21-hydroxylase deficiency: a feasibility study. *Int J Pediatr Endocrinol*, 2017; **2017**: 3.
15. Auchus RJ, Quint EH. Adolescents with disorders of sex development (DSD)--lost in transition? *Horm Metab Res*, 2015; **47**: 367–74.
16. Gleeson H, Davis J, Jones J, O'Shea E, Clayton PE. The challenge of delivering endocrine care and successful transition to adult services in adolescents with congenital adrenal hyperplasia: experience in a single centre over 18 years. *Clin Endocrinol (Oxf)*, 2013; **78**: 23–8.
17. Reisch N, Willige M, Kohn D, *et al*. Frequency and causes of adrenal crises over lifetime in patients with 21-hydroxylase deficiency. *Eur J Endocrinol*, 2012; **167**: 35–42.
18. Auchus RJ, Witchel SF, Leight KR, *et al*. Guidelines for the development of comprehensive care centers for congenital adrenal hyperplasia: guidance from the CARES foundation initiative. *Int J Pediatr Endocrinol*, 2010; **2010**: 275213.
19. Godbout A, Tejedor I, Malivoir S, Polak M, Touraine P. Transition from pediatric to adult healthcare: assessment of specific needs of patients with chronic endocrine conditions. *Horm Res Paediatr*, 2012; **78**: 247–55.
20. Lotstein DS, Kuo AA, Strickland B, Tait F. The transition to adult health care for youth with special health care needs: do racial and ethnic disparities exist? *Pediatrics*, 2010; **126** Suppl 3129–36.
21. Finkielstain GP, Kim MS, Sinaii N, *et al*. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2012; **97**: 4429–38.
22. Arlt W, Willis DS, Wild SH, *et al*. Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 patients. *J Clin Endocrinol Metab*, 2010; **95**: 5110–21.
23. Falhammar H, Filipsson H, Holmdahl G, *et al*. Metabolic profile and body composition in adult women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2007; **92**: 110–16.
24. Kim MS, Merke DP. Cardiovascular disease risk in adult women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Semin Reprod Med*, 2009; **27**: 316–21.
25. Falhammar H, Filipsson H, Holmdahl G, *et al*. Fractures and bone mineral density in adult women with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2007; **92**: 4643–9.
26. Sciannamblo M, Russo G, Cuccato D, Chiumello G, Mora S. Reduced bone mineral density and increased bone metabolism

rate in young adult patients with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2006; **91**: 4453–8.
27. Auchus RJ, Buschur EO, Chang AY, *et al.* Abiraterone acetate to lower androgens in women with classic 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2014; **99**: 2763–70.
28. Rosenfield RL, Bickel S, Razdan AK. Amenorrhea related to progestin excess in congenital adrenal hyperplasia. *Obstet Gynecol*, 1980; **56**: 208–15.
29. Winterer J, Chrousos GP, Loriaux DL, Cutler GB, Jr. Effect of hydrocortisone dose schedule on adrenal steroid secretion in congenital adrenal hyperplasia. *J Pediatr*, 1985; **106**: 137–42.
30. Charmandari E, Matthews DR, Johnston A, Brook CG, Hindmarsh PC. Serum cortisol and 17-hydroxyprogesterone interrelation in classic 21-hydroxylase deficiency: is current replacement therapy satisfactory? *J Clin Endocrinol Metab*, 2001; **86**: 4679–85.
31. Auchus RJ. Management considerations for the adult with congenital adrenal hyperplasia. *Mol Cell Endocrinol*, 2015; **408**: 190–7.
32. Caldato MC, Fernandes VT, Kater CE. One-year clinical evaluation of single morning dose prednisolone therapy for 21-hydroxylase deficiency. *Arq Bras Endocrinol Metabol*, 2004; **48**: 705–12.
33. Merke DP, Cho D, Calis KA, Keil MF, Chrousos GP. Hydrocortisone suspension and hydrocortisone tablets are not bioequivalent in the treatment of children with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2001; **86**: 441–5.
34. Neumann U, Whitaker MJ, Wiegand S, *et al.* Absorption and tolerability of taste-masked hydrocortisone granules in neonates, infants and children under 6 years of age with adrenal insufficiency. *Clin Endocrinol (Oxf)*, 2018; **88**: 21–9.
35. Esposito D, Pasquali D, Johannsson G. Primary adrenal insufficiency: managing mineralocorticoid replacement therapy. *J Clin Endocrinol Metab*, 2018; **103**: 376–87.
36. Hagenfeldt K, Janson PO, Holmdahl G, *et al.* Fertility and pregnancy outcome in women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Hum Reprod*, 2008; **23**: 1607–13.
37. Casteràs A, De Silva P, Rumsby G, Conway GS. Reassessing fecundity in women with classical congenital adrenal hyperplasia (CAH): normal pregnancy rate but reduced fertility rate. *Clin Endocrinol (Oxf)*, 2009; **70**: 833–7.
38. Claahsen-van der Grinten HL, Otten BJ, Hermus AR, Sweep FC, Hulsbergen-van de Kaa CA. Testicular adrenal rest tumors in patients with congenital adrenal hyperplasia can cause severe testicular damage. *Fertil Steril*, 2008; **89**: 597–601.
39. Reisch N, Flade L, Scherr M, *et al.* High prevalence of reduced fecundity in men with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2009; **94**: 1665–70.
40. Bouvattier C, Esterle L, Renoult-Pierre P, *et al.* clinical outcome, hormonal status, gonadotrope axis, and testicular function in 219 adult men born with classic 21-hydroxylase deficiency. A French national survey. *J Clin Endocrinol Metab*, 2015; **100**: 2303–13.
41. Tiitinen A, Välimäki M. Primary infertility in 45-year-old man with untreated 21-hydroxylase deficiency: successful outcome with glucocorticoid therapy. *J Clin Endocrinol Metab*, 2002; **87**: 2442–5.
42. Engels M, Gehrmann K, Falhammar H, *et al.* Gonadal function in adult male patients with congenital adrenal hyperplasia. *Eur J Endocrinol*, 2018; **178**: 285–94.
43. Reisch N, Rottenkolber M, Greifenstein A, *et al.* Testicular adrenal rest tumors develop independently of long-term disease control: a longitudinal analysis of 50 adult men with congenital adrenal hyperplasia due to classic 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2013; **98**: E1820–6.
44. Bouman A, Hulsbergen-van de Kaa C, Claahsen-van der Grinten HL. Prevalence of testicular adrenal rest tissue in neonates. *Horm Res Paediatr*, 2011; **75**: 90–3.
45. Smeets EE, Span PN, van Herwaarden AE, *et al.* Molecular characterization of testicular adrenal rest tumors in congenital adrenal hyperplasia: lesions with both adrenocortical and Leydig cell features. *J Clin Endocrinol Metab*, 2015; **100**: E524–30.
46. Cabrera MS, Vogiatzi MG, New MI. Long term outcome in adult males with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2001; **86**: 3070–8.
47. Mouritsen A, Juul A, Jorgensen N. Improvement of semen quality in an infertile man with 21-hydroxylase deficiency, suppressed serum gonadotropins and testicular adrenal rest tumours. *Int J Androl*, 2010; **33**: 518–20.
48. Claahsen-van der Grinten HL, Otten BJ, Sweep FC, Hermus AR. Repeated successful induction of fertility after replacing hydrocortisone with dexamethasone in a patient with congenital adrenal hyperplasia and testicular adrenal rest tumors. *Fertil Steril*, 2007; **88**: 705, e5–8.
49. Claahsen-van der Grinten HL, Otten BJ, Takahashi S, *et al.* Testicular adrenal rest tumors in adult males with congenital adrenal hyperplasia: evaluation of pituitary-gonadal function before and after successful testis-sparing surgery in eight patients. *J Clin Endocrinol Metab*, 2007; **92**: 612–15.
50. King TF, Lee MC, Williamson EE, Conway GS. Experience in optimizing fertility outcomes in men with congenital adrenal hyperplasia due to 21 hydroxylase deficiency. *Clin Endocrinol (Oxf)*, 2016; **84**: 830–6.
51. Bry-Gauillard H, Cartes A, Young J. Mitotane for 21-hydroxylase deficiency in an infertile man. *N Engl J Med*, 2014; **371**: 2042–4.
52. Avila NA, Premkumar A, Merke DP. Testicular adrenal rest tissue in congenital adrenal hyperplasia: comparison of MR imaging and sonographic findings. *AJR Am J Roentgenol*, 1999; **172**: 1003–6.
53. Van Wyk JJ, Gunther DF, Ritzén EM, *et al.* The use of adrenalectomy as a treatment for congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 1996; **81**: 3180–9.
54. Scaroni C, Favia G, Lumachi F, *et al.* Unilateral adrenal tumor, erectile dysfunction and infertility in a patient with 21-hydroxylase deficiency: effects of glucocorticoid treatment and surgery. *Exp Clin Endocrinol Diabetes*, 2003; **111**: 41–3.
55. Ogilvie CM, Rumsby G, Kurzawinski T, Conway GS. Outcome of bilateral adrenalectomy in congenital adrenal hyperplasia: one unit's experience. *Eur J Endocrinol*, 2006; **154**: 405–8.
56. Van Wyk JJ, Ritzén EM. The role of bilateral adrenalectomy in the treatment of congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2003; **88**: 2993–8.
57. Tiosano D, Vlodavsky E, Filmar S, Weiner Z, Goldsher D, Bar-Shalom R. Ovarian adrenal rest tumor in a congenital adrenal hyperplasia patient with adrenocorticotropin hypersecretion following adrenalectomy. *Horm Res Paediatr*, 2010; **74**: 223–8.
58. Crocker MK, Barak S, Millo CM, *et al.* Use of PET/CT with cosyntropin stimulation to identify and localize adrenal rest tissue following adrenalectomy in a woman with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2012; **97**: E2084–E9.

59. Charmandari E, Chrousos GP, Merke DP. Adrenocorticotropin hypersecretion and pituitary microadenoma following bilateral adrenalectomy in a patient with classic 21-hydroxylase deficiency. *J Pediatr Endocrinol Metab*, 2005; **18**: 97–101.
60. Witchel SF, Azziz R. Nonclassic congenital adrenal hyperplasia. *Int J Pediatr Endocrinol*, 2010; **2010**: 625105.
61. Bidet M, Bellanne-Chantelot C, Galand-Portier MB, *et al*. Fertility in women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2010; **95**: 1182–90.
62. Eyal O, Ayalon-Dangur I, Segev-Becker A, Schachter-Davidov A, Israel S, Weintrob N. Pregnancy in women with nonclassic congenital adrenal hyperplasia: time to conceive and outcome. *Clin Endocrinol (Oxf)*, 2017; **87**: 552–6.
63. Rezvani I, Garibaldi LR, Digeorge AM, Artman HG. Disproportionate suppression of dehydroepiandrosterone sulfate (DHEAS) in treated patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Pediatr Res*, 1983; **17**: 131–4.
64. Auchus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2013; **98**: 2645–55.
65. Simard J, Moisan AM, Morel Y. Congenital adrenal hyperplasia due to 3β-hydroxysteroid dehydrogenase/Δ^5-Δ^4 isomerase deficiency. *Semin Reprod Med*, 2002; **20**: 255–76.
66. Auchus RJ. The classic and nonclassic concenital adrenal hyperplasias. *Endocr Pract*, 2015; **21**: 383–9.
67. Donadille B, Houang M, Netchine I, Siffroi JP, Christin-Maitre S. Human 3β-hydroxysteroid dehydrogenase deficiency associated with normal spermatic numeration despite a severe enzyme deficit. *Endocr Connect*, 2018; **7**: 395–402.
68. Moisan AM, Ricketts ML, Tardy V, *et al*. New insight into the molecular basis of 3β-hydroxysteroid dehydrogenase deficiency: identification of eight mutations in the HSD3B2 gene in eleven patients from seven new families and comparison of the functional properties of twenty-five mutant enzymes. *J Clin Endocrinol Metab*, 1999; **84**: 4410–25.
69. Lutfallah C, Wang W, Mason JI, *et al*. Newly proposed hormonal criteria via genotypic proof for type II 3β-hydroxysteroid dehydrogenase deficiency. *J Clin Endocrinol Metab*, 2002; **87**: 2611–22.
70. Zerah M, Rheaume E, Mani P, *et al*. No evidence of mutations in the genes for type I and type II 3β-hydroxysteroid dehydrogenase (3βHSD) in nonclassical 3βHSD deficiency. *J Clin Endocrinol Metab*, 1994; **79**: 1811–17.
71. Topor LS, Asai M, Dunn J, Majzoub JA. Cortisol stimulates secretion of dehydroepiandrosterone in human adrenocortical cells through inhibition of 3βetaHSD2. *J Clin Endocrinol Metab*, 2011; **96**: E31–9.
72. Benkert AR, Young M, Robinson D, Hendrickson C, Lee PA, Strauss KA. Severe salt-losing 3β-hydroxysteroid dehydrogenase deficiency: treatment and outcomes of HSD3B2 c.35 G>A homozygotes. *J Clin Endocrinol Metab*, 2015; **100**: E1105–15.
73. Chung BC, Picado-Leonard J, Haniu M, *et al*. Cytochrome P450c17 (steroid 17α-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. *Proc Natl Acad Sci U S A*, 1987; **84**: 407–11.
74. Auchus RJ. Steroid 17-hydroxylase and 17,20-lyase deficiencies, genetic and pharmacologic. *J Steroid Biochem Mol Biol*, 2017; **165**(Pt A): 71–8.
75. Costa-Santos M, Kater CE, Auchus RJ. Two prevalent CYP17 mutations and genotype-phenotype correlations in 24 Brazilian patients with 17-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2004; **89**: 49–60.
76. Geller DH, Auchus RJ, Mendonça BB, Miller WL. The genetic and functional basis of isolated 17,20 lyase deficiency. *Nature Genet*, 1997; **17**: 201–5.
77. Sherbet DP, Tiosano D, Kwist KM, Hochberg Z, Auchus RJ. CYP17 mutation E305G causes isolated 17,20-lyase deficiency by selectively altering substrate binding. *J Biol Chem*, 2003; **278**: 48563–9.
78. Van Den Akker EL, Koper JW, Boehmer AL, *et al*. Differential inhibition of 17α-hydroxylase and 17,20-lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. *J Clin Endocrinol Metab*, 2002; **87**: 5714–21.
79. Mantero F, Opocher G, Rocco S, Carpene G, Armanini D. Long-term treatment of mineralocorticoid excess syndromes. *Steroids*, 1995; **60**: 81–6.
80. Bianchi PH, Gouveia GR, Costa EM, *et al*. Successful live birth in a woman with 17α-hydroxylase deficiency through ivf frozen-thawed embryo transfer. *J Clin Endocrinol Metab*, 2016; **101**: 345–8.
81. ten Kate-Booij MJ, Cobbaert C, Koper JW, de Jong FH. Deficiency of 17,20-lyase causing giant ovarian cysts in a girl and a female phenotype in her 46,XY sister: case report. *Hum Reprod*, 2004; **19**: 456–9.
82. Rösler A, Leiberman E, Sack J, *et al*. Clinical variability of congenital adrenal hyperplasia due to 11β-hydroxylase deficiency. *Horm Res*, 1982; **16**: 133–41.
83. Khattab A, Haider S, Kumar A, *et al*. Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11β-hydroxylase deficiency. *Proc Natl Acad Sci U S A*, 2017; **114**: E1933–E40.
84. Joehrer K, Geley S, Strasser-Wozak EM, *et al*. CYP11B1 mutations causing non-classic adrenal hyperplasia due to 11β-hydroxylase deficiency. *Hum Mol Genet*, 1997; **6**: 1829–34.
85. Reisch N, Högler W, Parajes S, *et al*. A diagnosis not to be missed: nonclassic steroid 11β-hydroxylase deficiency presenting with premature adrenarche and hirsutism. *J Clin Endocrinol Metab*, 2013; **98**: E1620–5.
86. Chihaoui M, Kanoun F, Chaker F, *et al*. Testicular adrenal rest tumours in young adult males with congenital adrenal hyperplasia: prevalence and impact on testicular function. *Andrologia*, 2016; **48**: 45–50.
87. Simm PJ, Zacharin MR. Successful pregnancy in a patient with severe 11β-hydroxylase deficiency and novel mutations in CYP11B1 gene. *Horm Res*, 2007; **68**: 294–7.
88. Arlt W, Walker EA, Draper N, *et al*. Congenital adrenal hyperplasia caused by mutant P450 oxidoreductase and human androgen synthesis: analytical study. *Lancet*, 2004; **363**: 2128–35.
89. Flück CE, Tajima T, Pandey AV, *et al*. Mutant P450 oxidoreductase causes disordered steroidogenesis with and without Antley-Bixler syndrome. *Nat Genet*, 2004; **36**: 228–30.
90. Peterson RE, Imperato-McGinley J, Gautier T, Shackleton C. Male pseudohermaphroditism due to multiple defects in steroid-biosynthetic microsomal mixed-function oxidases. A new variant of congenital adrenal hyperplasia. *N Engl J Med*, 1985; **313**: 1182–91.
91. Krone N, Reisch N, Idkowiak J, *et al*. Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *J Clin Endocrinol Metab*, 2012; **97**: E257–E67.

92. Homma K, Hasegawa T, Nagai T, *et al*. Urine steroid hormone profile analysis in cytochrome P450 oxidoreductase deficiency: implication for the backdoor pathway to dihydrotestosterone. *J Clin Endocrinol Metab*, 2006; **91**: 2643–9.
93. Song T, Wang B, Chen H, Zhu J, Sun H. *In vitro* fertilization-frozen embryo transfer in a patient with cytochrome P450 oxidoreductase deficiency: a case report. *Gynecol Endocrinol*, 2018; **34**: 385–8.
94. Bose HS, Pescovitz OH, Miller WL. Spontaneous feminization in a 46,XX female patient with congenital lipoid adrenal hyperplasia caused by a homozygous frame-shift mutation in the steroidogenic acute regulatory protein. *J Clin Endocrinol Metab*, 1997; **82**: 1511–15.
95. Bose HS, Sugawara T, Strauss JF, III, Miller WL. The pathophysiology and genetics of congenital lipoid adrenal hyperplasia. *N Engl J Med*, 1996; **335**: 1870–8.
96. Khoury K, Barbar E, Ainmelk Y, Ouellet A, Lehoux JG. Gonadal function, first cases of pregnancy, and child delivery in a woman with lipoid congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2009; **94**: 1333–7.
97. Khoury K, Barbar E, Ainmelk Y, Ouellet A, Lavigne P, LeHoux JG. Thirty-eight-year follow-up of two sibling lipoid congenital adrenal hyperplasia patients due to homozygous steroidogenic acute regulatory (STARD1) protein mutation. Molecular structure and modeling of the STARD1 L275P Mutation. *Front Neurosci*, 2016; **10**: 527.
98. Baker BY, Lin L, Kim CJ, *et al*. Nonclassic congenital lipoid adrenal hyperplasia: a new disorder of the steroidogenic acute regulatory protein with very late presentation and normal male genitalia. *J Clin Endocrinol Metab*, 2006; **91**: 4781–5.
99. Metherell LA, Naville D, Halaby G, *et al*. Nonclassic lipoid congenital adrenal hyperplasia masquerading as familial glucocorticoid deficiency. *J Clin Endocrinol Metab*, 2009; **94**: 3865–71.
100. Tee MK, Abramsohn M, Loewenthal N, *et al*. Varied clinical presentations of seven patients with mutations in CYP11A1 encoding the cholesterol side-chain cleavage enzyme, P450scc. *J Clin Endocrinol Metab*, 2013; **98**: 713–20.
101. Hindmarsh PC. The child with difficult to control Congenital Adrenal Hyperplasia: is there a place for continuous subcutaneous hydrocortisone therapy. *Clin Endocrinol (Oxf)*, 2014; **81**: 15–18.
102. Nella AA, Mallappa A, Perritt AF, *et al*. A phase 2 study of continuous subcutaneous hydrocortisone infusion in adults with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2016; **101**: 4690–8.
103. Bryan SM, Honour JW, Hindmarsh PC. Management of altered hydrocortisone pharmacokinetics in a boy with congenital adrenal hyperplasia using a continuous subcutaneous hydrocortisone infusion. *J Clin Endocrinol Metab*, 2009; **94**: 3477–80.
104. Verma S, Vanryzin C, Sinaii N, *et al*. A pharmacokinetic and pharmacodynamic study of delayed- and extended-release hydrocortisone (Chronocort) vs. conventional hydrocortisone (Cortef) in the treatment of congenital adrenal hyperplasia. *Clin Endocrinol (Oxf)*, 2010; **72**: 441–7.
105. Mallappa A, Sinaii N, Kumar P, *et al*. A phase 2 study of Chronocort, a modified-release formulation of hydrocortisone, in the treatment of adults with classic congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2015; **100**: 1137–45.
106. Merke DP, Keil MF, Jones JV, Fields J, Hill S, Cutler GB, Jr. Flutamide, testolactone, and reduced hydrocortisone dose maintain normal growth velocity and bone maturation despite elevated androgen levels in children with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2000; **85**: 1114–20.
107. Turcu AF, Spencer-Segal JL, Farber RH, *et al*. Single-dose study of a corticotropin-releasing factor receptor-1 antagonist in women with 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2016; **101**: 1174–80.
108. Storbeck KH, Bloem LM, Africander D, Schloms L, Swart P, Swart AC. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: a putative role in castration resistant prostate cancer? *Mol Cell Endocrinol*, 2013; **377**: 135–46.
109. Jones CM, Mallappa A, Reisch N, *et al*. Modified-release and conventional glucocorticoids and diurnal androgen excretion in congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2017; **102**: 1797–806.

SECTION 6

Neuroendocrine Tumours and Inherited Endocrine Tumour Syndromes

6.1

Overview and Pathophysiology of Neuroendocrine Neoplasms

Rajaventhan Srirajaskanthan and Guido Rindi

Introduction

The first description of a neuroendocrine tumour was in 1867 by T. Langhans, who described a submucosal tumour that resembled poorly differentiated glandular tissue arranged in 'nests' with a rich, thick fibrous stroma [1]. Thereafter in 1888, O. Lubarsch reported the post-mortem identification of multiple ileal tumours, which he was reluctant to classify as 'carcinomas' due to a benign growth pattern appearance [2]. However, it was Siegfried Oberndorfer who published his seminal paper *Carcinoid Tumours of the Small Intestine* in 1907 and recognized their unique nature; finally defining the lesions as a neoplasm distinct from carcinoma. The term 'karzinoide' (meaning carcinoma-like) was initially introduced by Oberndorfer in 1907 [3]. Although, the term 'carcinoid tumour' has historically been used, with advances in the understanding of the tumour biology, and the recent World Health Organization (WHO) classifications, the term neuroendocrine tumour (NET) is considered more appropriate. In consideration of the current classification terminology, the definition of neuroendocrine neoplasm (NEN) is now adopted to embrace well and poorly differentiated types. Originally, Pearse proposed that tumours develop from migration of cells from the neural crest; however, it is now thought that the tumour cells are derived from multipotent stem cells [4]. NENs are believed to originate from neuroendocrine or neuroendocrine-committed cells, which are present in organs throughout the body.

Incidence and Prevalence

Over the last 10 years there has been a number of studies investigating the true incidence and prevalence [5, 6]. SEER data demonstrated the incidence of 5–7 per 100 000 population. Further studies have demonstrated similar prevalence in Europe [7, 8]. Interestingly, however, while the overall incidence is around 7 per 100 000 population [6, 9], there is an ethnic variation with a higher incidence rectal NENs seen in African-Caribbean and Asian populations [10]. In addition, there is an ethnic variation in different countries [11]. Gender difference between male or female varies according to the anatomical site of neoplasm occurrence.

The prevalence of NENs in a recent study from the USA, using the SEER database is reported at 48 per 100 000 population, this excludes the appendiceal NETs [9]. The incidence in Europe and the rest of the world is reported between 4 and 6 cases per 100 000 [7, 8, 12, 13].

There is a significant increase in incidence of NENs over the last 40 years. This may represent with increasing incidence of these NENs in addition to increase detection with endoscopic screening and use of cross-sectional imaging [8, 10]. This trend has been demonstrated worldwide and affecting both low- and high-grade neuroendocrine neoplasms [14]. Incidence of carcinoid syndrome has been reported at 19% of all NENs and this give an overall incidence of 1: 100 000 population [5]. Little information is available as for risk factors for NEN. A recent metanalysis of published information indicated several variables including family history of cancer, alcohol consumption, smoking habits, and diabetes mellitus as potential risk factors with variable significance at different anatomical places [14].

Clinical Features and Diagnostic Investigations

Signs and symptoms of neuroendocrine neoplasms can be separated into non-functional and functional. Approximately 60% of neoplasms are non-functional and therefore, these often present as incidental findings, or as result of mass effect of the tumour. The

Table 6.1.1 The clinical features of neuroendocrine tumours

Site	Clinical features	Main cell type	MEN 1
Pancreas			
Non-functional	Symptoms related to mass effect	various	60%
Insulinoma	Hypoglycaemia, Whipple's triad, clammy, sweating, weight gain	B cell	5–10%
Gastrinoma	Zollinger–Ellison syndrome	G cells	25%
VIPoma	Werner–Morrison syndrome, watery diarrhoea	VIP	10%
Glucagonoma	Diabetes mellitus, necrolytic migratory erythema	A cell	5–10%
Somatostatinoma	Gallstones, diabetes mellitus, steatorrhoea	D cells	5–10%
Bronchial	Majority non-functional, 8% carcinoid syndrome, 'atypical' flushing		
Stomach	Majority non-functional, associated with chronic atrophic gastritis; rare 'atypical' carcinoid syndrome		
Duodenum	Majority non-functional; rare gastrinoma syndrome	G cells	20%
Small bowel	Majority non-functional, 40% develop carcinoid syndrome	Serotonin-producing Enterochromaffin EC cells	
Colorectal	Usually non-functional, however tumours may secrete somatostatin, other peptides, and occasionally carcinoid syndrome may occur	PYY/glicentin-producing cells	

functional tumours can lead to a number of distinct clinical syndromes which are summarized in Table 6.1.1. A number of chapters in this section of the textbook will deal with the functional tumours in much more detail (see Chapters 6.3–6.10).

Diagnostic Investigations

Diagnosis of NENs requires biochemical, topographical imaging, and, importantly, histological diagnosis. Efforts should be made to identify the primary tumour site, which can be difficult since some primary lesions are small and not detected by conventional cross-sectional imaging.

Biochemical Tests

Patients with suspected NENs should undergo biochemical testing, including fasting gut hormones (glucagon, vasoactive intestinal peptide, somatostatin (SST), and gastrin), chromogranin A, and pancreatic polypeptide. Urinary or plasma 5-HIAA measurement should also be performed in all patients with suspected carcinoid syndrome [15, 16]. Table 6.1.2 shows the different biochemical tests that are used for diagnosis of NENs.

Histology

Histology remains the gold standard for diagnosing NENs (see Figure 6.1.1). Along the diagnostic work-up it is advised to provide histological rather than cytological samples in order to achieve an accurate diagnosis. Specimens should be immunostained with a panel of antibodies to general neuroendocrine markers to demonstrate the neuroendocrine nature of the neoplasm. These include chromogranin A and synaptophysin; other general markers such as PGP9.5, neuron-specific enolase and CD56 N-CAM may also be of help specially to elucidate the neuroendocrine nature of poorly differentiated NENs. In addition, the sample should be stained with an antibody to the Ki-67 protein, since the Ki-67 proliferation index is necessary for grading neoplasms and has important significance for prognosis and treatment (see Figure 6.1.1). If required and upon the clinician's request, the sample may be stained for specific hormones, in order to support the clinically observed hyperfunction. The histological characteristics of NENs vary according to the degree of differentiation. Low-grade, well-differentiated NENs originating from the gut were previously termed 'carcinoids' and are now defined as NETs. These tumours had classic histological architecture of solid islet, trabecular, or ribbon-like cell patterns, with little or no cellular pleomorphism, occasional mitoses and absent or only spotty necrosis [17]. Usually NETs show a delicate stroma rich in vessels. According to the current classification by WHO and AJCC [18–21]. NETs are graded in three tiers as grade 1, grade 2, and grade 3 based on mitotic count and Ki-67%. The morphology of NETs is rather similar between grades, while only NETs G3 may show a somehow more solid structure with focal/spotty necrosis, a stroma more evident and often thick, evident mitoses, usually below 20 per 2 mm^2, and Ki-67% above 20%. Due to their non-typical morphology this type of NETs were previously defined as 'atypical carcinoid', a term which is still retained for NETs of the thoracic region [22]. High-grade, poorly differentiated NENs are currently

Table 6.1.2 General biochemical plasma markers raised in neuroendocrine tumour-dependent on anatomical site

Type of tumour	Plasma marker	Urinary marker
NEN	Chromogranin A, chromogranin B, neuron-specific enolase	5-Hydroxyindoloacetic acid
Phaeochromocytoma	Chromogranin A, chromogranin B, Neuron-specific enolase, β-human chorionic gonadotropin, neuropeptide Y, plasma-free metanephrines, α-fetoprotein	Fractionated metanephrines
Pancreatic NETs	Chromogranin A, chromogranin B, neuron-specific enolase, gastrin, pancreatic polypeptide, somatostatin, VIP, PTHrP	

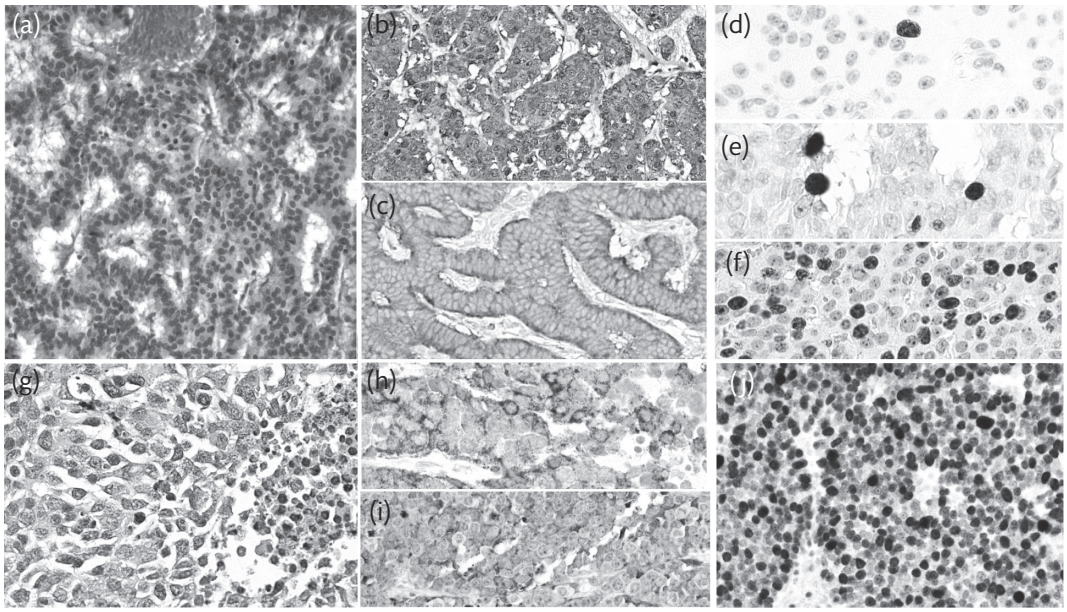


Figure 6.1.1 Histology of digestive neuroendocrine neoplasms (NENs): neuroendocrine tumours (NETs) are shown in pictures (a–f); neuroendocrine carcinomas (NECs) in pictures (g–j) (a) and (g) haematoxylin and eosin; (b–f) and (h–j) immunoperoxidase. (a) Colon non-functioning L-type cell NET showing a trabecular structure rich in delicate vessels and low cell atypia; (b) chromogranin A diffuse staining of a non-functioning pancreas NET with solid islet structure; (c) synaptophysin diffuse staining in a non-functioning colon trabecular L-type NET; d.-f. Ki67 nuclear labelling in NET G1 (pancreas insulinoma, **d**), G2 (pancreas non-functioning, **e**) and G3 (colon non-functioning, **f**); (g) NEC large cell type of the duodenum showing severe atypia and necrosis (right part of the micrograph); (h) chromogranin A faint but diffuse staining in a NEC of the gall-bladder, large cell type; (i) synaptophysin diffuse and intense staining in a colon NEC, large cell type; (j) diffuse nuclear labelling for Ki67 in a gastric NEC, small cell type.

defined as neuroendocrine carcinomas (NEC), by default G3. These tumours display solid or organoid structure with extensive necrosis, a usually thick/desmoplastic stroma, severe cellular atypia with evident mitoses, often atypical. The WHO classification gives clear parameters for categorizing NENs into the two categories of NET and NEC and the three NET classes described earlier. A minimal diagnosis of NEN should comprise the definition of the tumour category, whether NET or NEC, the definition of the grade for NETs and, for surgical specimens only, the stage definition. NETs are staged according to a site-specific NEN-dedicated staging system [23, 24]. NECs are staged following the site-corresponding adenocarcinoma/squamous cell carcinoma staging scheme.

Imaging

Cross-sectional imaging is usually with contrast computed tomography (CT), including arterial phase enhancement, of the abdomen, chest, and pelvis [25]. MRI is the most sensitive modality for liver metastases [26]. Studies of CT in carcinoid tumours show an overall sensitivity of 80% in detecting lesions [15, 25]. The sensitivity and specificity of CT and MRI alone are lower than the combination of ^{111}in-octreotide scan with CT or MRI.

Nuclear Medicine

Nuclear medicine imaging is important in staging of disease and determining suitability for therapy with somatostatin SST analogues and peptide receptor. The main nuclear imaging technology modalities are octreotide scintigraphy based and positron emission tomography (PET) scanning. Somatostatin receptor scintigraphy used to primarily be based on ^{111}In-Octreotide scintigraphy, however, more recently Tc-99m-tektrotyd scintigraphy has demonstrated similar if not superior sensitivity and specificity [27]. However, this is still inferior in terms of sensitivity and specificity to Ga-68 DOTATATE PET imaging [28].

There are five different SST receptors (SST_s) subtypes, all of which have strong affinity for SST [29]. Octreotide is an SST analogue which has a strong affinity for SST_2 and, to a lesser extent, SST_5 receptors [29]. NENs predominantly express SST_2 and this expression is mostly restricted to well-differentiated low-grade NETs. Synthetic radiolabelled SST analogues, such as ^{111}In-pentetreotide, enable SST_s scintigraphy to be performed [30]. The sensitivity of SST_s scintigraphy for the detection of GEP NETs has been well studied. The sensitivity has been reported to be between 67% and 100%, with no significant difference in carcinoid tumours from foregut, midgut, or hindgut origin. With pancreatic NETs, sensitivity of SST_s scintigraphy is dependent on the type of functional tumour. Gastrinomas detection has a sensitivity between 56 and 80%, VIPoma is 60–70%, and insulinoma lower at 50% due to a lower expression of SST_2. With pheochromocytomas, SST_s scintigraphy is often negative and other imaging modalities, such as MIBG, should be used. Medullary thyroid cancer express SST_1 therefore may be negative on SST_s scintigraphy. False-positive scans can be seen in patients with chronic inflammation and granulomatous disease. SST_s scintigraphy detection is also affected by the size of NETs and will often not detect lesions less than 1 cm.

MIBG has been used for two decades to visualize pheochromocytoma and paraganglioma. MIBG shares the same method of uptake as noradrenaline and is not dependent on SST receptor expression. In pheochromocytomas, MIBG has sensitivity of 87% and specificity of 99% [31].

FDG-PET scanning in other malignancies is well established; however, its role for NENs is still evolving. [^{18}F]2-fluoro-2-deoxyglucose (FDG)-PET is only suitable for intermediate (G2) and high-grade tumours (NET G3 and NEC) and is of minimal use in low-grade tumours due to their slow glucose turnover. Studies with Ga-68 DOTA-octreotide had a greater sensitivity than conventional SST_s scintigraphy [32]. Ga-68 DOTATATE PET has become standing functional imaging investigation in a number of centres [33–35].

Management of Neuroendocrine Neoplasms

Therapies for NENs incorporate those required for control of symptoms due to hormonal secretion from tumours, and also antiproliferative therapies. The management of NENs requires the use of a number of different therapies including surgery, biotherapy, chemotherapy, peptide receptor targeted therapy, and tumour embolization. The best way to provide the most appropriate management plan for patients is through a multidisciplinary approach for a patient-tailored therapy [15, 36]. Different therapies may be required for different grade NETs and for different clinical stages. In patients with indolent disease and mild symptoms merely symptomatic relief may be all that is required for some years.

Surgery remains the only method of cure and this should be offered to all patients when appropriate. If there is no evidence of metastatic disease, then resection of the primary should be considered to give the best outcomes and also for high-grade NENs [17, 36]. However, if there is metastatic disease the role of surgery is less clear, generally surgery can be considered if significant debulking can be achieved. There is also evidence suggesting the benefit of resection of small bowel primary tumours in the context of metastatic disease if there are causing symptoms [38, 39]. The ENETS guidelines recommend palliative resection of the primary tumour and mesenteric metastases in the setting of stage IV NET disease for symptomatic patients with intestinal obstruction, ischaemia, or tumour bleeding [16].

Non-Surgical Therapy

Liver Directed Loco-Regional Therapy

The most common site of metastatic disease from a NEN is the liver and current experience is almost restricted to NETs. There are a number of different liver-directed therapies that can be performed to reduce tumour volume and control disease progression [40, 41].

Hepatic Artery Embolization

NET liver metastases are highly vascular with an arterial supply thats if occluded will lead to ischaemia and necrosis. By performing hepatic artery embolization of the tumours, normal liver tissue is supplied from the portal vein and preserved during embolization of hepatic arteries. Hence the tumour blood supply can be embolized without significant damage to the surrounding liver. The process of embolization involves passing a catheter to the hepatic artery or branch and material (gelfoam powder, microembospheres, and polyvinyl alcohol particles) released to occlude the vessel in bland embolization. In chemoembolization, cytotoxics (like cisplatin, miriplatin, gemcitabine, doxorubicin, streptozocin, and 5-FU) are injected prior to arterial embolization in order to achieve higher concentrations and prolonged action in necrotic tissue. Contraindications to embolization include occlusion of the portal vein, severe liver dysfunction, and presence of biliary anastomosis. Relative contraindications include tumour burden, renal impairment, and heart disease (including carcinoid heart disease) [42, 43]. Median progression-free survival (PFS) rates of around 18 months after transarterial embolization (TAE) or chemoembolization (TACE) in patients with liver metastases [42, 44, 45]. Clinical response rates in terms of reducing in symptoms from carcinoid syndrome have been reported as high as 90%.

Selective internal radiation therapy (SIRT) otherwise known as radioembolization is performed using ^{90}Ytrrium resin microspheres that are injected through a percutaneously placed hepatic artery catheter via the femoral artery; this leads to cell death and tumour shrinkage [46, 47]. The contraindications to SIRT are similar to that of bland embolization. Long-term radiologic and biological responses can be achieved with radioembolization with partial or complete response seen in 63% [44, 45]. Median survival varies from 36 to 70 months [41, 48]. To date there is no randomized evidence that radiologic and symptom response rates following SIRT are different from those seen with TACE or TAE [43, 49].

Radiofrequency ablation of liver metastases is a well-established technique for management of liver metastases from a number of different tumours. The criteria are usually lesions less than 5 cm in maximum dimension and less than eight in number. The procedure involves electrical energy delivered to the tumour via a catheter that is inserted percutaneously or laparoscopically. This electrical energy leads to heating and cell death. High symptomatic response rates of 70–80% have been reported in patients with functional tumours such as those with carcinoid syndrome [50].

Systemic Medical Therapy for NENs

For patients with extensive metastatic disease a more systemic approach to treatment is usually considered. There are number of therapies available for treatment of NENs depending on grade, stage, and the site of the primary tumour [51]. A brief overview of these therapies is described next, with more detail regarding specific therapies in the forthcoming chapters.

Somatostatin Analogues

Over 70% of NENs express somatostatin receptors (SSRs) on their cell surface. SSRs expression is usually restricted to well-differentiated low-grade neoplasm (NETs) though up to 25% of NET G3 and NEC may also express SSTRs [52]. These can be targeted by synthetic somatostatin analogues. Initial, studies demonstrated a significant improvement in terms of controlling hormonal secretion and symptoms in patients with functional tumours, especially carcinoid syndrome. Over the past 10 years two randomized control trials (RCTs) have been published that have also demonstrated an antiproliferative effect of somatostatin analogues [53, 54]. The CLARINET trial was an RCT in which lanreotide was compared with placebo, was associated with significantly prolonged PFS (median not reached vs. median of 18.0 months, P <0.001). The estimated rates of PFS at 24 months were 65.1% (95% CI, 54.0–74.1)

in the lanreotide group and 33.0% (95% CI, 23.0–43.3) in the placebo group.

Chemotherapy

There have been a number of trials regarding chemotherapy in GEP NENs, most evidence is in treatment of pancreatic NENs with streptozocin-based chemotherapy for well-differentiated tumours and platinum-based chemotherapy for poorly differentiated neoplasms [55–57]. There is increasing data for the use of temozolomide monotherapy or in combination with capecitabine in pancreatic NENs [56]. The role of chemotherapy in PanNETs will be discussed in more detail in a later chapter. In the ENETS guidelines there is a recommendation for chemotherapy in PanNETs as a second-line therapy in patients with metastatic disease [51]. In small bowel NETs the evidence for chemotherapy is limited, and recent ENETS guidelines suggest chemotherapy for NECs only [16, 51]. There are no specific trials studying chemotherapy in bronchial NETs, however, patients with typical and atypical bronchial NETs are often treated with chemotherapy. Temozolomide is advocated as palliative treatment in bronchial NETs as it has been the most widely studied in the subgroup of lung NET and has an acceptable safety profile [58].

Molecular Targeted Therapy

There are two molecularly targeted therapies in management of metastatic NETs. Everolimus is an oral inhibitor of the mammalian target of rapamycin (mTOR) pathway and has activity against pancreatic NETs through a mechanism of cellular apoptosis and antiangiogenesis [59]. Median PFS is significantly longer in those treated with everolimus over placebo (11 vs. 4.6 months) [59]. Severe adverse events like hyperglycaemia and anaemia were rare, with stomatitis, diarrhoea, and fatigue are more commonly seen. Everolimus is also licenced for patients with pulmonary NETs and non-functional gastrointestinal NETs following the positive results obtained from the RADIANT-4 trial [60]. This trial demonstrated everolimus improved median PFS by 7.1 months in patients with advanced, progressive, well-differentiated (grade 1 or grade 2), non-functional lung or gastrointestinal NETs vs. placebo (hazard ratio, 0.48; 95% confidence interval [CI], 0.35–0.67; P <0.00001). In a subsequent subgroup analysis of the RADIANT-4 data. There were 90 of the 302 patients enrolled in the study had primary lung NET (everolimus, n = 63; placebo, n = 27). Median PFS (95% CI) by central review was 9.2 (6.8–10.9) months in the everolimus arm vs. 3.6 (1.9–5.1) months in the placebo arm (hazard ratio, 0.50; 95% CI, 0.28–0.88). More patients who received everolimus (58%) experienced tumour shrinkage compared with placebo (13%) [61].

Sunitinib, a multitargeted tyrosine kinase inhibitor, has activity against a range of molecular targets, including vascular endothelial growth factor (VEGF) receptors and platelet-derived growth factor receptors, and has been shown to have antitumour activity in pancreatic NETs. Median PFS is significantly longer in patients treated with sunitinib over placebo (11.4 vs. 5.5 months) [62]. Frequent adverse events encountered include diarrhoea, nausea, vomiting, asthenia, and fatigue. The long-term outcome is a 10 month longer survival in the treatment arm, but this did not reach statistical significance, likely because of the high cross over rate [63].

Peptide Receptor Radionuclide Therapy

Somatostatin receptor subtype 2 is expressed in the majority of NETs and confirmed through uptake in octreotide scintigraphy or somatostatin-based PET imaging. This offers a therapeutic option by attaching a β-emitting radioisotope to a molecule that will bind to somatostatin receptor-2 then particle binds to neuroendocrine tumour cells and leads to cell death. The two commonly used radioisotopes are Lu-177-DOTATATE which is a beta and gamma emitter and Y-90 DOTATATE which is primarily a beta emitter. Therapy is usually administered over four cycles spaced 8–12 weeks apart. Contraindications to therapy are significantly impaired renal function, severely suppressed bone marrow. There have a been a number of reports from single centres regarding peptide receptor radionuclide therapy (PRRT) demonstrating a partial response rate up to 30% with stabilization of disease in more than 60% of patients. The average time to progression of 40 months from commencing therapy [64]. One study demonstrated that from the point of diagnosis, there is a survival benefit of 40–72 months compared to historical controls [65]. Adverse events include bone marrow and liver toxicity as well as radiation-induced loss of renal function and gastrointestinal disturbance from the use of renoprotective agents (amino acid infusions). A recent phase 3 RCTs comparing Lutathera, a commercially developed Lu-177-DOTATATE, against high-dose Sandostatin LAR (60 mg every 28 days) in patients with progressive metastatic functional midgut NENS. This demonstrated a significant response rate of 18% in the 177Lu-Dotatate group versus 3% in the control group (P <0.001). In the planned interim analysis of overall survival, 14 deaths occurred in the 177Lu-Dotatate group and 26 in the control group (P = 0.004). The estimated rate of PFS at month 20 was 65.2% (95% confidence interval [CI], 50.0–76.8) in the 177Lu-Dotatate group and 10.8% (95% CI, 3.5–23.0) in the control group.

Prognosis of NENs

Survival in NENs has been improved over the last few years [13]. Recent data from the SEER database suggests there has been an improvement in survival over the last 30 years [9]. The overall median survival of all NENs in the SEER database was 9.3 years (112 months). Localized NETs had better median OS (>30 years) compared with regional NETs (10.2 years) and distant NETs (12 months) (P<0.001). The overall survival is dependent on stage, grade, and site of primary. In terms of factors that determine prognosis these are also dependent on primary site, grade, and tumour, and tumour burden.

REFERENCES

1. Langhans T. Ueber einen Drüsenpolyp im Ileum. *Virchows Arch A Pathol Anat Histol*, 1867; **38**: 550–60.
2. Lubarsch. Ueber dem primären Krebs des Ileum nebst Bemerkungen über das gleichzeitige Vorkommen von Krebs und Tuberculose. *Virchows Arch A Pathol Anat Histol,* 1888; **111**: 280–317.
3. Oberndorfer S. Karzinoide tumoren des dünndarms. *Frankf Z Pathol*, 1907; **1**: 426–32.
4. Stachura J, Urban A, Bigaj M, Szczudrawa J, Wysocki A. Histochemical and ultrastructural observation of endocrine

cells in pathological gastric mucosa. *Folia Histochem Cytochem (Krakow)*, 1978; **16**: 287–98.
5. Halperin DM, Shen C, Dasari A, *et al*. Frequency of carcinoid syndrome at neuroendocrine tumour diagnosis: a population-based study. *Lancet Oncol*, 2017; **18**: 525–34.
6. Yao JC, Hassan M, Phan A, *et al*. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, 2008; **26**: 3063–72.
7. Hallet J, Law CHL, Cukier M, Saskin R, Liu N, Singh S. Exploring the rising incidence of neuroendocrine tumors: a population-based analysis of epidemiology, metastatic presentation, and outcomes. *Cancer*, 2015; **121**: 589–97.
8. Sandvik OM, Søreide K, Gudlaugsson E, Kvaløy JT, Søreide JA. Epidemiology and classification of gastroenteropancreatic neuroendocrine neoplasms using current coding criteria. *Br J Surg*, 2016; **103**: 226–32.
9. Dasari A, Shen C, Halperin D, *et al*. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*, 2017; **3**: 1335–42.
10. Modlin IM, Lye KD, Kidd M. A 5-decade analysis of 13,715 carcinoid tumors. *Cancer*, 2003; **97**: 934–59.
11. Hauso O, Gustafsson BI, Kidd M, *et al*. Neuroendocrine tumor epidemiology: contrasting Norway and North America. *Cancer*, 2008; **113**: 2655–64.
12. Fraenkel M, Kim M, Faggiano A, de Herder WW, Valk GD, Knowledge NETwork. Incidence of gastroenteropancreatic neuroendocrine tumours: a systematic review of the literature. *Endocr Relat Cancer*, 2014; **21**: R153–63.
13. Boyar Cetinkaya R, Aagnes B, Myklebust TÅ, Thiis-Evensen E. Survival in neuroendocrine neoplasms; a report from a large Norwegian population-based study. *Int J Cancer*, 2018; **142**: 1139–47.
14. Leoncini E, Boffetta P, Shafir M, Aleksovska K, Boccia S, Rindi G. Increased incidence trend of low-grade and high-grade neuroendocrine neoplasms. *Endocrine*, 2017; **58**: 368–79.
15. Ramage JK, Ahmed A, Ardill J, *et al*. Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours (NETs). *Gut*, 2012; **61**: 6–32.
16. Niederle B, Pape UF, Costa F, *et al*. ENETS consensus guidelines update for neuroendocrine neoplasms of the jejunum and ileum. *Neuroendocrinology*, 2016; **103**: 125–38.
17. Soga J, Tazawa K. Pathologic analysis of carcinoids. Histologic reevaluation of 62 cases. *Cancer*, 1971; **28**: 990–8.
18. Rindi G, Arnold R, Capella C, *et al*. *World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of the Digestive System*. Lyon: IARC Press, 2010: pp. 10–12.
19. Amin MB. *AJCC Cancer Staging Manual*. VIII. New York: Springer-Verlag, 2017.
20. Klöppel G, Couvelard A HR, *et al*. Introduction. In: Lloyd RV, Osamura RY, Klöppel G, Rosai J, eds. *WHO Classification of Tumours of the Endocrine Organs*. Lyon: IARC Press, 2017: pp. 211–14.
21. WHO Classification of Tumours Editorial Board: Digestive System Tumours, vol. 1, 5th Edition ed. WHO Classification of Tumours. IARC Press, Lyon (2019).
22. Travis WD, Brambilla E, Burke E, Marx A, Nicholson AG. *Pathology and Genetics of Tumours of Lung, Pleura, Thymus and Heart*, 4th edn. Lyon: IARC Press, 2015.
23. Rindi G, Klöppel G, Alhman H, *et al*. TNM staging of foregut (neuro)endocrine tumors: a consensus proposal including a grading system. *Virchows Arch*, 2006; **449**: 395–401.
24. Rindi G, Kloppel G, Couvelard A, *et al*. 2007 TNM staging of midgut and hindgut (neuro) endocrine tumors: a consensus proposal including a grading system. *Virchows Arch*, **451**: 757–62.
25. Rockall AG, Reznek RH. Imaging of neuroendocrine tumours (CT/MR/US). Vol. 21. *Best Pract Res Clin Endocrinol Metab*, 2007; **21**: 43–68.
26. Chong S, Lee KS, Chung MJ, Han J, Kwon OJ KT. No Title. *Radiographics*, 2006; **26**: 41–57.
27. Artiko V, Sobic-Saranovic D, Pavlovic S, *et al*. The clinical value of scintigraphy of neuroendocrine tumors using (99m)Tc-HYNIC-TOC. *J BUON*, 2012; **17**: 537–42.
28. Kunikowska J, Lewington V, Krolicki L. Optimizing somatostatin receptor imaging in patients with neuroendocrine tumors: the impact of 99mTc-HYNICTOC SPECT/SPECT/CT Versus 68Ga-DOTATATE PET/CT upon clinical management. *Clin Nucl Med*, 2017; **42**: 905–11.
29. Oberg K, Casanovas O, Castaño JP, *et al*. Molecular pathogenesis of neuroendocrine tumors: implications for current and future therapeutic approaches. *Clin Cancer Res*, 2013; **19**: 2842–9.
30. Forrer F, Valkema R, Kwekkeboom DJ, de Jong M, Krenning EP. Neuroendocrine tumors. Peptide receptor radionuclide therapy. *Best Pract Res Clin Endocrinol Metab*, 2007; **21**: 111–29.
31. van der Lely AJ, de Herder WW. Carcinoid syndrome: diagnosis and medical management. *Arq Bras Endocrinol Metabol*, 2005; **49**: 850–60.
32. Srirajaskanthan R, Kayani I, Quigley AM, Soh J, Caplin ME, Bomanji J. The Role of 68 Ga-DOTATATE PET in patients with neuroendocrine tumors and negative or equivocal findings on 111 In-DTPA-octreotide scintigraphy. *J Nucl Med*, 2010; **51**: 875–82.
33. Kayani I, Conry BG, Groves AM, *et al*. A comparison of 68Ga-DOTATATE and 18F-FDG PET/CT in pulmonary neuroendocrine tumors. *J Nucl Med*, 2009; **50**: 1927–32.
34. Srirajaskanthan R, Kayani I, Quigley AM, Soh J, Caplin ME, Bomanji J. The role of 68Ga-DOTATATE PET in patients with neuroendocrine tumors and negative or equivocal findings on 111In-DTPA-octreotide scintigraphy. *J Nucl Med*, 2010; **51**: 875–82.
35. Desai K, Watkins J, Woodward N, *et al*. Use of molecular imaging to differentiate liver metastasis of colorectal cancer metastasis from neuroendocrine tumor origin. *J Clin Gastroenterol*, 2011; **45**: e8–11.
36. Falconi M, Eriksson B, Kaltsas G, *et al*. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
37. Falconi M, Eriksson B, Kaltsas G, *et al*. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
38. Daskalakis K, Karakatsanis A, Hessman O, *et al*. Association of a prophylactic surgical approach to stage IV small intestinal neuroendocrine tumors with survival. *JAMA Oncol*, 2018; **4**: 183–9.
39. Ahmed A, Turner G, King B, *et al*. Midgut neuroendocrine tumours with liver metastases: results of the UKINETS study. *Endocr Relat Cancer*, 2009; **16**: 885–94.
40. Basuroy R, Srirajaskanthan R, Ramage JK. A multimodal approach to the management of neuroendocrine tumour liver metastases. *Int J Hepatol*, 2012; **2012**: 819193.

41. Kennedy AS. Hepatic-directed therapies in patients with neuroendocrine tumors. *Hematol Oncol Clin North Am*, 2016; **30**: 193–207.
42. Ho AS, Picus J, Darcy MD, *et al*. Long-term outcome after chemoembolization and embolization of hepatic metastatic lesions from neuroendocrine tumors. *AJR Am J Roentgenol*, 2007; **188**: 1201–7.
43. Maire F, Lombard-Bohas C, O'Toole D, *et al*. Hepatic arterial embolization versus chemoembolization in the treatment of liver metastases from well-differentiated midgut endocrine tumors: a prospective randomized study. *Neuroendocrinology*, 2012; **96**: 294–300.
44. Strosberg JR, Choi J, Cantor AB, Kvols LK. Selective hepatic artery embolization for treatment of patients with metastatic carcinoid and pancreatic endocrine tumors. *Cancer Control*, 2006; **13**: 72–8.
45. King J, Quinn R, Glenn DM, *et al*. Radioembolization with selective internal radiation microspheres for neuroendocrine liver metastases. *Cancer*, 2008; **113**: 921–9.
46. Elf A-K, Andersson M, Henrikson O, *et al*. Radioembolization versus bland embolization for hepatic metastases from small intestinal neuroendocrine tumors: short-term results of a randomized clinical trial. *World J Surg*, 2018; **42**: 506–13.
47. Saxena A, Chua TC, Bester L, Kokandi A, Morris DL. Factors predicting response and survival after yttrium-90 radioembolization of unresectable neuroendocrine tumor liver metastases: a critical appraisal of 48 cases. *Ann Surg*, 2010; **251**: 910–16.
48. Kennedy AS, Dezarn WA, McNeillie P, *et al*. Radioembolization for unresectable neuroendocrine hepatic metastases using resin 90Y-microspheres: early results in 148 patients. *Am J Clin Oncol*, 2008; **31**: 271–9.
49. Cao CQ, Yan TD, Bester L, Liauw W, Morris DL. Radioembolization with yttrium microspheres for neuroendocrine tumour liver metastases. *Br J Surg*, 2010; **97**: 537–43.
50. Frilling A, Modlin IM, Kidd M, *et al*. Recommendations for management of patients with neuroendocrine liver metastases. *Lancet Oncol*, 2014; **15**: e8–21.
51. Pavel M, O'Toole D, Costa F, *et al*. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. *Neuroendocrinology*, 2016; **103**: 172–85.
52. Konukiewitz B, Schlitter AM, Jesinghaus M, *et al*. Somatostatin receptor expression related to TP53 and RB1 alterations in pancreatic and extrapancreatic neuroendocrine neoplasms with a Ki67-index above 20. *Mod Pathol*, 2017; **30**: 587–98.
53. Caplin ME, Pavel M, Ćwikła JB, *et al*. Lanreotide in Metastatic Enteropancreatic Neuroendocrine Tumors. *N Engl J Med*, 2014; **371**: 224–33.
54. Rinke A, Müller H-H, Schade-Brittinger C, *et al*. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*, 2009; **27**: 4656–63.
55. Wong MH, Chan DL, Lee A, *et al*. Systematic review and meta-analysis on the role of chemotherapy in advanced and metastatic neuroendocrine tumor (NET). *PLoS One*, 2016; **11**: e0158140.
56. Strosberg JR, Fine RL, Choi J, *et al*. First-line chemotherapy with capecitabine and temozolomide in patients with metastatic pancreatic endocrine carcinomas. *Cancer*, 2011; **117**: 268–75.
57. Turner NC, Strauss SJ, Sarker D, *et al*. Chemotherapy with 5-fluorouracil, cisplatin and streptozocin for neuroendocrine tumours. *Br J Cancer*, 2010; **102**: 1106–12.
58. Caplin ME, Baudin E, Ferolla P, *et al*. Pulmonary neuroendocrine (carcinoid) tumors: European Neuroendocrine Tumor Society expert consensus and recommendations for best practice for typical and atypical pulmonary carcinoids. *Ann Oncol*, 2015; **26**: 1604–20.
59. Yao JC, Shah MH, Ito T, *et al*. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 514–23.
60. Yao JC, Fazio N, Singh S, *et al*. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2016; **387**: 968–77.
61. Fazio N, Buzzoni R, Delle Fave G, *et al*. Everolimus in advanced, progressive, well-differentiated, non-functional neuroendocrine tumors: RADIANT-4 lung subgroup analysis. *Cancer Sci*, 2018; **109**: 174–81.
62. Raymond E, Dahan L, Raoul J-L, *et al*. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 501–13.
63. Faivre S, Niccoli P, Castellano D, *et al*. Sunitinib in pancreatic neuroendocrine tumors: updated progression-free survival and final overall survival from a phase III randomized study. *Ann Oncol Off J Eur Soc Med Oncol*, 2017; **28**: 339–43.
64. Brabander T, van der Zwan WA, Teunissen JJM, *et al*. Long-term efficacy, survival, and safety of [177Lu-DOTA0,Tyr3]octreotate in patients with gastroenteropancreatic and bronchial neuroendocrine tumors. *Clin Cancer Res*, 2017; **23**: 4617–24.
65. Kwekkeboom DJ, de Herder WW, Kam BL, *et al*. Treatment with the radiolabeled somatostatin analog [177 Lu-DOTA 0,Tyr3] octreotate: toxicity, efficacy, and survival. *J Clin Oncol*, 2008; **26**: 2124–30.

6.2

Neuroendocrine Tumour Markers

Waljit Dhillo and Paul Bech

Neuroendocrine Cells

In 1938, Friedrich Feyrter identified the presence of pale cells (Helle Zellen) distributed throughout the body and with greater preponderance in the intestinal tract. As these cells could secrete chemical messages, Feyter proposed the diffuse endocrine system. With advances in histochemistry and electron microscopy, understanding of the cell type progressed. In 1969, Everson Pears proposed the amine precursor uptake and decarboxylation (APUD) to describe the production of biogenic amines and polypeptide hormones [1]. Neuroendocrine (NE) cells make up the diffuse NE system and occur either singly or in small groups in a variety of tissues and organs. Although morphologically and embryologically diverse, they are now defined by the production of neuropeptides and neurotransmitters, and the presence of dense core secretory vesicles in the cytoplasm from which neuropeptides are released into the circulation by exocytosis. However, unlike neurones, NE cell do not possess synapses or axons. NE cells can be identified by the use protein components of the secretory granules unique to this cell.

Neuroendocrine Tumours (NETs)

Classification

Siegfried Oberndorfer was the first to investigate tumours in the bowel that were rare and indolent. He used the term '*Karzinoid*' ('carcinoid') (carcinoma-like) to describe the small tumours he found in the submucosa of the gut. The classification of NETs has changed over the years. In 1960 the World Health Organization (WHO) published a classification for carcinoids of the diffuse endocrine system, and subdivided the carcinoid group into enterochromaffin cell carcinoids, gastrin cell carcinoids, and other carcinoids. A second WHO classification was published in 2000, it uses the term NET, and classification is based on size, proliferation, localization, and differentiation. Distinction is made between well-differentiated NET (benign behaviour or uncertain malignant potential), well-differentiated NE carcinomas (low-grade malignancy), and poorly differentiated (usually small cell) NE carcinomas of high-grade malignancy. The latest WHO classification in 2010 defined NETs as neuroendocrine neoplasia and gastrointestinal NETs are classed according to their proliferative index and mitotic count. The most recent WHO classification published in 2017 introduced a new category of well-differentiated neoplasms, neuroendocrine tumours G3, in addition to the previous categories of neuroendocrine tumours G1 and G2 [2]. Despite the reclassifications older terms such as carcinoid are still used (see Chapter 6.1).

Types of NETs

Cells of the diffuse endocrine system are particularly prominent in the gastrointestinal tract, pancreas, C cells of the thyroid, adrenal medulla, parathyroid tissue, respiratory tract, skin, and genitourinary system. NETs are known to occur in all these tissues and, therefore, constitute a very heterogenous group of tumours.

Despite the large number of peptides in the gut, only a few when secreted in excess by NETs, result in clinical syndromes. The majority of gut-hormone secreting tumours originate in the pancreatic islet cells and in the midgut. Gastroenteropancreatic neuroendocrine tumours can be divided into functioning and non-functioning tumours. Functioning tumours include insulinomas, gastrinomas, VIPomas, glucagonomas, and somatostatinomas. The symptoms are caused by the excess peptide hormone secretion. The hormone secreted may be eutopic or ectopic.

Diagnosis

Historically, the diagnosis of NET was determined by histological findings. The significantly worse prognosis in advanced disease and the availability of multiple therapeutic options have highlighted the need for robust tumour markers that can be used both for diagnosis and follow-up. Currently, a number of normal and abnormal forms of peptides, biogenic amines, and hormones, secreted by NETs are routinely measured as markers of disease.

An ideal tumour biomarker would be one that is secreted exclusively by the tumour cells. Biomarkers can be useful (i) for screening and differential diagnosis of NET, (ii) as a prognostic indicator, (iii)

as an estimate of tumour burden, and (iv) as a surveillance tool. Although none of the currently available markers completely fit the paradigm for an ideal tumour marker, when measured in conjunction with each other, they are useful not only for making a diagnosis but also for monitoring response to therapy and in surveillance post-remission. Consensus on biomarkers for NETs have advocated the use of three biomarkers groups originally identified for clinical drug discovery. Type 0 biomarkers should indicate the natural history of the disease, type 1 biomarkers should show the effect of intervention and type 2 biomarkers are surrogates for clinical endpoints [3].

General Neuroendocrine Markers

NETs often express and secrete peptides that are common to most neuroendocrine cells and to cells that have undergone neuroendocrine differentiation.

Chromogranins

Dense core secretory vesicles, present throughout the neuroendocrine system, store and release chromogranins. Chromogranins are a group of acidic, soluble secretory proteins which are thought to play a role in the formation of vesicles.

Chromogranins (A, B, and C) form a major constituent of these granules [4] and are widely distributed in the neuroendocrine system. As a result, they are excellent tissue and serum markers for neuroendocrine tumours.

Chromogranin A

Chromogranin A (CgA) was the first of the 'granins' to be identified and has the widest distribution [4, 5]. It is quantitatively the major constituent of the secretory granules in the NE cells and is expressed in the cells of the anterior pituitary, C cells of the thyroid, chief cells of the parathyroid, islet cells of the pancreas and the chromaffin cells of the adrenal medulla. It is also widely distributed in the neuroendocrine cells of the bronchial and gastropancreatic systems and the skin.

The human form of CgA is an acidic 439 amino acid protein that is preceded by an 18 amino acid signal peptide. Both the N and C terminals are well conserved between species. The CgA molecule contains a number of mono and dibasic amino acid sites, thus implying fairly extensive and varied post-translational processing. This suggests that it may be a precursor to a number of other peptides. Although the function of CgA has not yet been fully elucidated, it is often cosecreted with the neuroendocrine hormones and peptides and therefore thought to play a role in the processing, packaging, and secretion of neuropeptide precursors and hormones [6].

The sensitivity and specificity of circulating CgA measurement in the biochemical diagnosis of NET have been extensively studied and values calculated range from 56% to 100%. This variation in values can be attributed to the heterogeneous composition of NET subtypes used in the patient groups, the varying number of subjects used in the patient and control groups, the different reference intervals used, the different methods to obtain the reference intervals and the different CgA assays used.

To investigate the difference between CgA assays, a comparison of four kits demonstrated no significant difference in sensitivity between the assays, however, the sensitivity could be increased significantly when a combination of two non-cross reacting assays were used [7].

A meta-analysis of 13 studies meeting the inclusion criteria out of 78 potential studies revealed a sensitivity and specificity of 73% and 95%, respectively. Furthermore, the area under the curve was 0.896 indicating the usefulness of CgA [8]. Therefore, most NETs are associated with increased circulating levels of CgA. Even tumours that produce hormones with no identifiable clinical features or have lost the ability to synthesize peptides, continue to express CgA [4] As a result, CgA is routinely used as a marker for both diagnosis and monitoring of NETs. It is particularly useful (i) when existing cell-specific markers (explained later in the chapter) are either unstable, rapidly fluctuating or inconvenient for clinical use; (ii) to confirm the neuroendocrine origin of a tumour; and (iii) as a general marker of disease when the neoplastic disease involves multiple neuroendocrine tissues (e.g. multiple endocrine neoplasia) [4].

CgA has been shown to correlate with tumour burden and extent of metastases. Increasing concentrations of CgA can also indicate worse prognosis. However, high circulating CgA concentrations have been measured in gastrinoma patients despite the small primary tumour and absence of metastases.

CgA can be used to monitor recurrence of disease. A comparison of CgA, urinary 5HIAA and radiological measurements in patients with a midgut NET found that elevated plasma CgA was the first to indicate disease recurrence.

There are, however, some limitations to using CgA as a tumour biomarker. An increase in circulating CgA levels due to a tumour often goes undetected until it has reached a size capable of producing appreciably increased amounts of CgA. Further, CgA cannot differentiate between different subtypes of NET and it is not equally expressed in all NETs. Some only weakly express CgA (e.g. small cell carcinoma of the lung) [9] Concentrations are only minimally elevated in patients with insulinomas. Most patients with metastatic foregut and midgut carcinoids have increased (tumour burden dependent) levels of CgA. Gastrinomas may show increased levels of CgA even in patients with very limited disease. This is probably because chronic hypergastrinemia causes hyperplasia of the enterochromaffin cells [10].

The biological variation of circulating CgA levels has been studied extensively and levels can vary by 30% from day to day in healthy controls and patients alike. In addition, there are a number of non-neoplastic conditions are associated with high CgA levels. Patients on proton pump inhibitors (PPI) have raised CgA levels. Renal impairment, and to a lesser extent, liver failure can also result in increased levels [11].

CgA undergoes post-translational changes before release. Tumours may, therefore, release different molecular forms of CgA [12]. As a result, a number of different forms of CgA are released into circulation. Thus, levels measured are dependent on the antiserum used. Some assays measure only the whole CgA, while others measure the whole CgA plus fragments that contain the specific epitope to which the antisera was developed. Thus reference ranges are assay-specific. A lack of standardization between the various commercially available immunoassays makes comparison difficult. However, sequential measurement with the same assay can reliably be used to monitor disease progression in a patient.

Pancreastatin is a 49-amino acid (CgA 240–288) peptide produced by dibasic cleavage of CgA. Pancreastatin assays that use

antisera raised to the mid-molecule cross-react strongly to CgA and can be used to measure both pancreastatin and CgA. Assays using antisera raised to the N or C terminals of pancreastatin are, however, specific for pancreastatin. Pancreastatin concentrations are shown to correlate well with extent of liver involvement. Specific pancreastatin assays may therefore be used to assess and monitor the extent liver involvement in patients with NET [13]. The advantage of assays using antibodies specific for pancreastatin is that they are unaffected by PPIs [14].

Chromogranin B

Chromogranin B (657 amino acid peptide) (CgB) or GAWK (a partial sequence of CgB 420–493) coexists with CgA in the secretory granules and bears a strong homology to CgA in the terminal regions. Like CgA it is an acidic protein [15]. The relative abundance of CgA and CgB is cell specific. CgA is the dominant granin in the pancreatic endocrine tumours and in the serotonin secreting tumours in the ileum and appendix. However, in rectal carcinoids, where CgA is virtually absent, CgB is the most abundant granin [16]. Elevated CgB concentrations have been detected in some NET patients when CgA is not raised [17]. It is, therefore, measured complementary to CgA in some centres, and improves diagnostic sensitivity. CgB has also been shown to be a more sensitive marker of PNETs than CgA [18]. CgB assays are, however, associated with the same problems of standardization as the CgA assays [17, 19]. The combined measurement of CgA and CgB has a sensitivity of 89% for NETs [20]. A polyclonal antiserum with cross reactivity to both CgA and B has been developed by Eriksson *et al.* [21].

Circulating levels of CgB have only been shown in phaeochromocytoma to correlate with tumour bulk [22] and there is little evidence to suggest that CgB can be used to predict disease recurrence. Unlike CgA, CgB is not affected by PPI treatment. It is, however, like CgA effected by renal impairment [23].

Chromogranins are best measured in plasma. Fasting samples are not required. Samples must be centrifuged, plasma aliquoted, and stored at –20°C immediately after collection, to prevent degradation of the peptide. Patients on acid suppressive therapy should be advised to come off treatment before CgA is measured (see summary).

Neuron-Specific Enolase

α-γ and γ-γ isomers of the glycolytic enzyme phosphopyruvate hydratase, 2-phospho-D-glycerate hydrolase occur mainly in the neuronal and neuroectodermal tissue and are collectively known as neuron-specific enolase (NSE). NSE is expressed in a number of primary neuroendocrine tumours and in tumour with neuroendocrine differentiation [24, 25]. It is specifically used in the diagnosis and follow-up of patients with neuroblastoma and small cell carcinoma of lung. This is particularly helpful as CgA is poorly expressed in small cell lung carcinoma [26]. Although the specificity of serum NSE in the diagnosis of neuroendocrine tumours is lower than that of CgA, the combination of both markers has a higher sensitivity than both markers separately [27].

Other Markers

Pancreatic polypeptide (PP) is a 36 amino acid peptide secreted by the normal pancreas. Levels increase in response to food. Although its function has not been fully elucidated, PP is known to slow gastric motility and is thought to play a role in the initiation of satiety [28]. Its levels are increased in 74% of gastropancreatic tumours and roughly 50% of carcinoids. The term PPoma is used to describe tumours secreting particularly high levels of PP. Although no specific clinical symptoms have so far been attributed to elevated PP levels [12, 21] these tumours are sometimes associated with non-specific gastrointestinal symptoms. Due to a significant response to food, PP is best measured in plasma samples taken after an overnight fast. Samples must be centrifuged, plasma aliquoted, and stored at –20°C immediately after collection, to prevent degradation of the peptide.

Somatostatin secreting tumours (somatostatinoma), although rare, have been identified for some time and are associated with diabetes mellitus, jaundice, diarrhoea, steatorrhoea, weight loss, and gall bladder disease) [29]. Other very rare secreting tumours include ghrelin secreting tumours [30, 31] which are not associated with features of acromegaly and GLP-1 secreting tumours which have symptoms of reactive hypoglycaemia [32]. Most recently a cholecystokinin secreting tumour has been identified; the symptoms included severe weight loss, non-watery diarrhoea, gall bladder, and peptic ulcer disease despite consistently low plasma gastrin concentrations [33]. Due to significant changes in circulating concentrations after food, all of these hormones are best measured in samples taken after an overnight fast. Samples must be centrifuged and plasma aliquoted and stored at –20°C immediately after collection, to prevent degradation of the peptide.

Neurokinin A and substance P are members of the tachykinin family and are expressed in midgut carcinoid cells. They act on lymphocytes and mast cell degranulation and cause vasodilatation and flushing and effect gastrointestinal motility. Levels of neurokinin A and its fragment neurokinin K are increased in 46% of patients with midgut carcinoids and serve as good prognostic markers as levels correlate well with tumour burden and therapeutic response. Increases in neurokinin A can predict worse survival [34]. Tachykinins can be measured in either plasma or serum. As tachykinins do not show a significant rise post-prandially, fasting samples are not required [35].

Alpha subunit of the glycoprotein hormones and/or β-HCG are elevated in about 25% of patients with neuroendocrine tumours like carcinoids, islet cell tumours, medullary thyroid cancer, and small cell lung cancer. Locally invasive NETs are associated with higher prevalence of increased β-HCG and alpha subunit expression and circulating levels [36].

Carcinoembryonic antigen (CEA) is not specific to neuroendocrine tumours but is found to be elevated in some NETs, such as medullary thyroid carcinoma. When used in conjunction with calcitonin, the specific marker for medullary thyroid carcinoma, CEA is a good prognostic marker as increased levels are associated with greater tumour aggressiveness and poorer prognosis [37].

Ki-67 is a proliferation antigen, which is expressed in G1, S, G2, and M phases of the mitotic cycle. Cells in G0 phase (resting phase) of the mitotic cycle do not express Ki67. As a result, actively proliferating cells are seen to have a higher expression of Ki67 on immunocytochemistry (ICC). Ki67 proliferation index is a measure of its expression in the cell based on the intensity of staining on ICC. Non-functioning and metastatic NETs are seen to have a higher Ki67 staining index. Increased expression of Ki67 is associated with a poorer prognosis. Levels, however, cannot be measured in circulation, thus limiting the use of the marker to ICC [38].

A number of other peptides including adrenomedullin [39] and cocaine-and-amphetamine-regulated-transcript (CART) [40] have been proposed as candidate markers for NETs. The highest levels of adrenomedullin are found in neuroendocrine tumours of bronchial, midgut, and unknown origin. Levels seem to be predictive of progressive disease, thus suggesting a role for adrenomedullin as a prognostic marker. CART has been found to be a specific tumour marker in patients with a range of neuroendocrine tumours and a more useful marker than either CgA or CgB in patients with phaeochromocytoma or paraganglioma [18]. Used in combination with CgA, CART, measurement has the potential to improve sensitivity in diagnosis and follow-up of neuroendocrine tumours, in particular progressive pancreatic neuroendocrine tumours. CART expression in small bowel tumours taken from 97 patients has shown that it is associated with worse survival [41].

Cell-Specific Markers

In addition to the markers just mentioned, NE cells are often associated with the synthesis of specific hormones/peptides. Excessive production and/or release of hormones by tumours arising from these cells may result in clearly identifiable clinical syndromes (e.g. Zollinger–Ellison syndrome in patients with gastrinomas and recurrent hypoglycaemia's in patients with insulinomas). Other NETs may present with unusual clinical symptoms as a result of ectopic secretion of hormones such as growth hormone or adrenocorticotropic hormone (ACTH). In such tumours (also known as functioning tumours), the specific hormone can be used as a marker for diagnosis, monitoring, and surveillance of disease. Some of these tumours and their markers are listed in Table 6.2.1. Each of these tumours is discussed in detail (see specific chapters in section 6).

Measurement of serum hormone concentrations can also be useful in the diagnosis of NETs in which the hormonal products produce a few non-specific or no clinical symptoms. Hormones such as calcitonin, PP, somatostatin, and pro-hormones like pro-opio-melanocortin (POMC) and calcitonin-gene-related-peptide (CGRP) are examples of this.

Most peptides, hormones, and neuropeptides secreted by neuroendocrine cells are first synthesized as precursors or pro-hormones. In normal NE cells they are then processed into mature hormones by sequence-specific and tissue-specific, post-translational modifications including glycosylation, amidation, phosphorylation, and sulphation. However, these processes are often defective in NETs. As a result, tumour cells may secrete a number of heterogeneous unprocessed or incorrectly/incompletely processed forms of hormones/peptides. A classic example is POMC, which is cleaved to ACTH and β-endorphin in an ordered manner in the normal pituitary. But in the presence of a tumour, high-molecular-weight forms of ACTH, as well as POMC are secreted into the circulation. Other common examples of pro-hormones secreted by tumours include CGRP, pro-gastrin, and pro-insulin. Thus, the presence of incompletely or incorrectly processed pro-hormones is suggestive of the presence of a tumour.

This fact has particular implications when validating assays for measuring hormones/peptides as tumour markers. The antibody used, must cross-react with as many forms of the hormone/peptide as possible, in order to avoid false negative results.

Dynamic function tests may sometimes improve the diagnostic sensitivity of hormone measurements. Common examples include the pentagastrin stimulation test for medullary thyroid carcinoma and 72-hour fasting test for insulinomas.

Table 6.2.1 Cell-specific markers for some of the NETs

Tumour/syndrome	Marker (peptide(s)/hormone(s))
Gastrinoma	Gastrin *(elevated fasting plasma levels, off acid suppressive treatment)*
Insulinoma	Insulin, pro-insulin, and C-peptide
	(inappropriately non-suppressed in the presence of hypoglycaemia: plasma glucose <2.2 mmol/L)
Glucagonoma	Glucagon *(elevated fasting plasma levels)*
VIPoma	Vasointestinal peptide *(elevated fasting plasma levels)*
Carcinoids	5-Hydroxyindoleacetic acid (5-HIAA)
	(elevated 24-hour urinary)
	(elevated in midgut, occasionally in foregut, and rarely in hindgut carcinoids)
Pheochromocytoma, carcinoids *(occasionally)*	Metanephrines, *(elevated plasma and 24 hr urinary)*
	Catecholamines *(elevated plasma and 24 hr urinary)*
Hypercalcaemia	Parathyroid hormone (PTH) related peptide
	(elevated in association with raised serum calcium and suppressed PTH)
Medullary thyroid carcinoma	Calcitonin
	(elevated plasma calcitonin)
Acromegaly	Growth Hormone releasing Hormone
	(elevated plasma levels due to ectopic secretion)
	Growth Hormone
	(elevated plasma levels due to ectopic secretion)
	Insulin like Growth Factor
	(elevated plasma levels due to ectopic secretion)
Cushing's syndrome	Adrenocorticotropic Hormone
	(elevated plasma levels due to ectopic secretion)
	Cortisol
	(elevated plasma levels due to high ACTH)

Glycine Extended Peptides and Considerations for New Assays

Most peptide hormones secreted from NE cells are α-amidated and this constitutes the final hormonal processing step. Many of these hormones promote cellular growth and often secretion increases in NETs and therefore, may act as tumour growth factors. The preceding precursor to the α-amidated peptide is the glycine extended peptide, whereby the glycine residue donates the amide for the C-terminus of the mature peptide. As the glycine extended precursor is the penultimate step in the pathway it is often released with mature α-amidated peptide. Early studies showed that the glycine extended peptide, may also be a tumour growth factor and have its own receptor. This prompted further studies to corroborate the potential significance of an alternative receptor for glycine extended peptides. However, failure to confirm the initial findings in the subsequent years has left researchers to conclude that there are no such receptors

[42]. What it does demonstrate, however, is that care should be taken to develop NET biomarker assays which do not distinguish between the two forms so that maximum sensitivity can be achieved. This is further highlighted with gastrin measurement. Gastrin is processed into various different forms with 14, 17, 34, and 71 amino acids. The 17 and 34 amino acid forms predominate in the circulation and can be α-amidated or glycine extended, furthermore, gastrin can be sulphated at a tyrosine residue. Therefore, to increase detection of gastrin all these forms should be detectable by the assay [43].

Summary

CgA is currently the best general available NET marker. However, none of the currently available tumour markers are specific or sensitive enough to be used as a single definitive marker and for most NETs, diagnosis by means of plasma level estimations of one hormone is not always clear cut. Therefore, two or more hormones are used for the diagnosis of most NET tumours.

Apart from the cell-specific hormones associated with functioning tumours, it is useful to monitor CgA (+/– Cg B) as a marker of disease in almost all NETs. NSE is particularly useful in monitoring small cell carcinoma lung and neuroblastomas.

A panel of tumour markers is used to investigate both functioning and non-functioning gastropancreatic tumours. The markers most commonly measured include PP, somatostatin, CgA, CgB, VIP, gastrin, glucagon, and insulin.

Of note, a number of these hormones (PP, somatostatin, VIP, gastrin, glucagon, and insulin) show a significant change in levels in response to food. Samples to measure these hormones must therefore be collected after an overnight fast.

Before collecting samples, it is always advisable to confirm sample requirements from the assaying laboratory, as requirements and reference ranges may sometimes differ. Guidelines offered by the supraregional assay service for gut hormone measurement in the United Kingdom (Imperial College Healthcare NHS Trust, London) suggest that all of these hormones, including CgA and CgB, are best measured in plasma samples collected in bottles containing ethylenediaminetetraacetic acid (EDTA) or LiHeparin (+ aprotinin). Neurotensin can only be measured in plasma samples collected in Li Heparin (+ aprotinin) bottles.

The peptides are prone to degradation. Samples must therefore be centrifuged, aliquoted, and frozen at –20°C, immediately after collection. Both CgA and gastrin levels are significantly affected by acid suppressive therapy. Unless contraindicated in the patient with clinical Zollinger–Ellison syndrome, all patients must stop PPIs for two weeks, histamine-2 receptor blocker therapy for three days and any other antacid therapy for one day before the measurement of CgA or gastrin [44]. (Further details on sample requirements can be found at pathology.imperial.nhs.uk.)

It is important to remember that, as with all other tumour markers, NET markers are only valuable when measured in the context of corroborative clinical and radiological findings.

It is hoped that future advances in microarrays and proteomics will lead to the discovery of more specific markers that will not only be more accurate diagnostic and prognostic indicators, but will also enable the use of highly effective, targeted, and individualised treatments for patients with NETs.

REFERENCES

1. de Herder WW, Rehfeld JF, Kidd M, Modlin IM. A short history of neuroendocrine tumours and their peptide hormones. *Best Pract Res Clin Endocrinol Metab*, 2016; **30**: 3–17.
2. Oberg K. Management of functional neuroendocrine tumors of the pancreas. *Gland Surg*, 2018; **7**: 20–7.
3. Oberg K, Modlin IM, De Herder W, *et al*. Consensus on biomarkers for neuroendocrine tumour disease. *Lancet Oncol*, 2015; **16**: e435–e46.
4. Nobels FR, Kwekkeboom DJ, Bouillon R, Lamberts SW. Chromogranin A: its clinical value as marker of neuroendocrine tumours. *Eur J Clin Invest*, 1998; **28**: 431–40.
5. Blaschko H, Comline RS, Schneider FH, Silver M, Smith AD. Secretion of a chromaffin granule protein, chromogranin, from the adrenal gland after splanchnic stimulation. *Nature*, 1967; **215**: 58–9.
6. Helle KB, Corti A, Metz-Boutigue MH, Tota B. The endocrine role for chromogranin A: a prohormone for peptides with regulatory properties. *Cell Mol Life Sci*, 2007; **64**: 2863–86.
7. Ramachandran R, Bech P, Murphy KG, *et al*. Chromogranin A assays: comparison of diagnostic accuracy. *Regul Pept*, 2010; **164**: 12.
8. Yang X, Yang Y, Li Z, *et al*. Diagnostic value of circulating chromogranin a for neuroendocrine tumors: a systematic review and meta-analysis. *PloS One*, 2015; **10**: e0124884.
9. Lloyd RV, Cano M, Rosa P, Hille A, Huttner WB. Distribution of chromogranin A and secretogranin I (chromogranin B) in neuroendocrine cells and tumors. *Am J Pathol*, 1988; **130**: 296–304.
10. Granberg D, Stridsberg M, Seensalu R, *et al*. Plasma chromogranin A in patients with multiple endocrine neoplasia type 1. *J Clin Endocrinol Metab*, 1999; **84**: 2712–17.
11. O'Connor DT, Pandlan MR, Carlton E, Cervenka JH, Hslao RJ. Rapid radioimmunoassay of circulating chromogranin A: *in vitro* stability, exploration of the neuroendocrine character of neoplasia, and assessment of the effects of organ failure. *Clin Chem*, 1989; **35**: 1631–7.
12. de Herder WW. Biochemistry of neuroendocrine tumours. *Best Pract Res Clin Endocrinol Metab*, 2007; **21**: 33–41.
13. Stronge RL, Turner GB, Johnston BT, *et al*. A rapid rise in circulating pancreastatin in response to somatostatin analogue therapy is associated with poor survival in patients with neuroendocrine tumours. *Ann Clin Biochem*, 2008; **45**(Pt 6): 560–6.
14. Raines D, Chester M, Diebold AE, *et al*. A prospective evaluation of the effect of chronic proton pump inhibitor use on plasma biomarker levels in humans. *Pancreas*, 2012; **41**: 508–11.
15. Benedum UM, Lamouroux A, Konecki DS, *et al*. The primary structure of human secretogranin I (chromogranin B): comparison with chromogranin A reveals homologous terminal domains and a large intervening variable region. *EMBO J*, 1987; **6**: 1203–11.
16. Angelsen A, Syversen U, Haugen OA, Stridsberg M, Mjolnerod OK, Waldum HL. Neuroendocrine differentiation in carcinomas of the prostate: do neuroendocrine serum markers reflect immunohistochemical findings? *Prostate*, 1997; **30**: 1–6.
17. Stridsberg M, Eriksson B, Fellstrom B, Kristiansson G, Tiensuu Janson E. Measurements of chromogranin B can serve as a complement to chromogranin A. *Regul Pept*, 2007; **139**: 80–3.
18. Ramachandran R, Bech P, Murphy KG, *et al*. Comparison of the utility of cocaine- and amphetamine-regulated transcript (CART), chromogranin A, and chromogranin B in

neuroendocrine tumor diagnosis and assessment of disease progression. *J Clin Endocrinol Metab*, 2015; **100**: 1520–8.
19. Sekiya K, Ghatei MA, Salahuddin MJ, *et al*. Production of GAWK (chromogranin-B 420–493)-like immunoreactivity by endocrine tumors and its possible diagnostic value. *J Clin Invest*, 1989; **83**: 1834–42.
20. Bech P, Winstanley V, Murphy KG, *et al*. Elevated cocaine- and amphetamine-regulated transcript immunoreactivity in the circulation of patients with neuroendocrine malignancy. *J Clin Endocrinol Metab*, 2008; **93**: 1246–53.
21. Eriksson B, Arnberg H, Lindgren PG, *et al*. Neuroendocrine pancreatic tumours: clinical presentation, biochemical and histopathological findings in 84 patients. *J Int Med*, 1990; **228**: 103–13.
22. Stridsberg M, Husebye ES. Chromogranin A and chromogranin B are sensitive circulating markers for phaeochromocytoma. *Eur J Endocrinol*, 1997; **136**: 67–73.
23. Bech PR, Ramachandran R, Dhillo WS, Martin NM, Bloom SR. Quantifying the effects of renal impairment on plasma concentrations of the neuroendocrine neoplasia biomarkers chromogranin A, chromogranin B, and cocaine- and amphetamine-regulated transcript. *Clin Chem*, 2012; **58**: 941–3.
24. Gerbitz KD, Summer J, Schumacher I, Arnold H, Kraft A, Mross K. Enolase isoenzymes as tumour markers. *J Clin Chem Clin Biochem*, 1986; **24**: 1009–16.
25. Pahlman S, Esscher T, Bergvall P, Odelstad L. Purification and characterization of human neuron-specific enolase: radioimmunoassay development. *Tumour Biol*, 1984; **5**: 127–39.
26. Bajetta E, Ferrari L, Martinetti A, *et al*. Chromogranin A, neuron specific enolase, carcinoembryonic antigen, and hydroxyindole acetic acid evaluation in patients with neuroendocrine tumors. *Cancer*, 1999; **86**: 858–65.
27. Nobels FR, Kwekkeboom DJ, Coopmans W, *et al*. Chromogranin A as serum marker for neuroendocrine neoplasia: comparison with neuron-specific enolase and the alpha-subunit of glycoprotein hormones. *J Clin Endocrinol Metab*, 1997; **82**: 2622–8.
28. Gardiner JV, Jayasena CN, Bloom SR. Gut hormones: a weight off your mind. *J Neuroendocrinol*, 2008; **20**: 834–41.
29. Garbrecht N, Anlauf M, Schmitt A, *et al*. Somatostatin-producing neuroendocrine tumors of the duodenum and pancreas: incidence, types, biological behavior, association with inherited syndromes, and functional activity. *Endocr Relat Cancer*, 2008; **15**: 229–41.
30. Tsolakis AV, Portela-Gomes GM, Stridsberg M, *et al*. Malignant gastric ghrelinoma with hyperghrelinemia. *J Clin Endocrinol Metab*, 2004; **89**: 3739–44.
31. Tsolakis AV, Stridsberg M, Grimelius L, *et al*. Ghrelin immunoreactive cells in gastric endocrine tumors and their relation to plasma ghrelin concentration. *J Clin Gastroenterol*, 2008; **42**: 381–8.
32. Todd JF, Stanley SA, Roufosse CA, *et al*. A tumour that secretes glucagon-like peptide-1 and somatostatin in a patient with reactive hypoglycaemia and diabetes. *Lancet*, 2003; **361**: 228–30.
33. Rehfeld JF, Federspiel B, Bardram L. A neuroendocrine tumor syndrome from cholecystokinin secretion. *N Engl J Med*, 2013; **368**: 1165–6.
34. Ardill JE, McCance DR, Stronge WV, Johnston BT. Raised circulating Neurokinin A predicts prognosis in metastatic small bowel neuroendocrine tumours. Lowering neurokinin A indicates improved prognosis. *Ann Clin Biochem*, 2016; **53**(Pt 2): 259–64.
35. Turner GB, Johnston BT, McCance DR, *et al*. Circulating markers of prognosis and response to treatment in patients with midgut carcinoid tumours. *Gut*, 2006; **55**: 1586–91.
36. Grossmann M, Trautmann ME, Poertl S, *et al*. Alpha-subunit and human chorionic gonadotropin-beta immunoreactivity in patients with malignant endocrine gastroenteropancreatic tumours. *Eur J Clin Investig*, 1994; **24**: 131–6.
37. Machens A, Ukkat J, Hauptmann S, Dralle H. Abnormal carcinoembryonic antigen levels and medullary thyroid cancer progression: a multivariate analysis. *Arch Surg*, 2007; **142**: 289–93; discussion 94.
38. Vilar E, Salazar R, Perez-Garcia J, Cortes J, Oberg K, Tabernero J. Chemotherapy and role of the proliferation marker Ki-67 in digestive neuroendocrine tumors. *Endocr Relat Cancer*, 2007; **14**: 221–32.
39. Pavel ME, Hoppe S, Papadopoulos T, *et al*. Adrenomedullin is a novel marker of tumor progression in neuroendocrine carcinomas. *Hormone Metab Res*, 2006; **38**: 112–18.
40. Bech PR, Martin NM, Ramachandran R, Bloom SR. The biochemical utility of chromogranin A, chromogranin B and cocaine- and amphetamine-regulated transcript for neuroendocrine neoplasia. *Ann Clin Biochem*, 2014; 51(Pt 1): 8–21.
41. Landerholm K, Shcherbina L, Falkmer SE, Jarhult J, Wierup N. Expression of cocaine- and amphetamine-regulated transcript is associated with worse survival in small bowel carcinoid tumors. *Clin Cancer Res*, 2012; **18**: 3668–76.
42. Rehfeld JF. The long goodbye to glycine extended peptide hormones. *Int J Med Oncol*, 2017; **4**: 3.
43. Rehfeld JF, Gingras MH, Bardram L, Hilsted L, Goetze JP, Poitras P. The Zollinger–Ellison syndrome and mismeasurement of gastrin. *Gastroenterology*, 2011; **140**: 1444–53.
44. Dhillo WS, Jayasena CN, Lewis CJ, *et al*. Plasma gastrin measurement cannot be used to diagnose a gastrinoma in patients on either proton pump inhibitors or histamine type-2 receptor antagonists. *Ann Clin Biochem*, 2006; **43**(Pt 2): 153–5.

6.3

Carcinoid Syndrome

Dominique Clement, Raj Srirajaskanthan, and Martyn E. Caplin

Introduction

Neuroendocrine tumours (NETs) are derived from cells of the diffuse neuroendocrine system, which are present in organs throughout the body. The term 'Karzinoide' (meaning carcinoma like) was initially introduced by Siegfried Oberndorfer in 1907 [1]. The term carcinoid tumour has historically been used; however, with advances in the understanding of the tumour biology, the term NET is considered more appropriate. The term carcinoid syndrome arose from the clinical syndrome associated with carcinoid tumours. NETs commonly arise from the gastroenteropancreatic tract. The majority of these tumours are non-functional tumours, meaning that they do not secrete proteins that can cause a clinical syndrome. Around 30–45% can cause a functional syndrome and the most common clinical syndrome is carcinoid syndrome [2]. This most commonly presents with symptoms of diarrhoea, abdominal pain, and flushing. It is thought to arise due to the production of serotonin and kinins from the tumour cells which lead to the symptoms of carcinoid syndrome. The small bowel is the most common primary site that leads to carcinoid syndrome. Treatment strategies are aimed at reducing tumour load and decreasing hormone secretion from the tumour. This can be achieved through a number of different therapies and modalities including medical therapy, surgery, liver-directed therapy, and radiotargetted therapy. Carcinoid crisis is a life treating condition which can appear spontaneously or precipitated by surgery or anaesthesia. A number of systemic complications can arise secondary to serotonin or kinin secretion and are commonly seen in carcinoid syndrome. These include carcinoid heart disease which can develop with tricuspid and pulmonary valve regurgitation or stenosis. Another complication of the carcinoid syndrome can be mesenteric fibrosis with ischaemia or small bowel obstruction. Nutritional deficiencies can also occur such as pellagra, which can develop as result of niacin deficiency due to serotonin overproduction. In this chapter, we will discuss the prevalence of carcinoid syndrome, its pathophysiology, and discuss treatment options and systemic complications seen in carcinoid syndrome.

Prevalence

Carcinoid syndrome was first described in the literature by Thorsson *et al.* in 1954, they described a series of 16 patients with a new syndrome [3]. Based on clinical findings diarrhoea, skin flushing, oedema, and ascites. Autopsies performed in these cases confirmed a small intestine carcinoid tumour with liver metastases, valvular disease of right side of the heart with septal defects.

Since 1954 the prevalence of carcinoid syndrome has been extensively studied. For metastatic NETs the prevalence was reported between 8% [4] and 72% [5]. More recent European data show a prevalence of carcinoid syndrome in around 20% of NET patients [6, 7]. A recent article by Halperin *et al.* analysed the Surveillance, Epidemiology, and End Results (SEER) database in the USA, demonstrated 19% of patients with NETs had carcinoid syndrome [8].

Signs and Symptoms

The first symptoms Thorsson described with the carcinoid syndrome are diarrhoea, skin flushing, oedema, and ascites [3]. The most common symptom is diarrhoea which is reported in around 80% of patients [2, 9, 10]. Diarrhoea is the secretory type and due to gastrointestinal motility factors. The frequency is defined as at least three times a day [11, 12]. Skin flushing appears in 50–85% of

patients and is described as red appearance of skin face, neck, and upper part of chest. A flush can last from seconds to several minutes and has characteristic appearances. Flushing episodes can be infrequent, however, in untreated patients can occur at least once a day but often more frequent [2, 11–14]. Flushing may appear spontaneously but can be provoked by emotional stress, alcohol, or tyramine containing foods such as cheese, coffee, chocolate, nuts, avocado, bananas, and red wine [11, 14–16]. Individuals may not be aware of the flushing episodes, however, often family members or friends note the flushing episodes.

Abdominal pain was described in around 40% of patients, peripheral oedema in 50–60%, heart lesions in 20%, and pellagra in around 1% [2, 9, 17].

Other studies describe wheezing or bronchospasm in 10–20% of patients [8, 11, 18–20] and telangiectasia around 25% of patients [11, 21–24]. Cardiac palpitations are a common symptom often reported as occurring spontaneously or following a stressful episode. The palpitations can be short in duration of persist for long periods of time around 20–30 minutes [11, 21]. In patients with carcinoid heart disease at presentation they may have clinical signs of heart failure, usually right sided heart failure. This can include peripheral oedema, ascites, elevated jugular venous pressure, and cardiac murmurs [23].

Patients with carcinoid syndrome experience worse quality of life and more anxiety, based on hospital anxiety and depression scores, compared to NET patients without carcinoid syndrome [24, 25].

Diagnosis

Diagnosis of carcinoid syndrome is based on clinical symptoms, biochemical, imaging, and histological diagnosis. Biochemical tests for diagnosis of carcinoid syndrome are primarily the 24-hour urine collection for 5-hydroxyindolacetic acid (5-HIAA) this is a breakdown product of serotonin and has a sensitivity and specificity around 90% for NETs [26]. Certain serotonin-rich foods (bananas, avocados, plums, eggplant, tomatoes, plantain, pineapples, kiwis, and nuts) and medications (analgesic acetaminophen, cough syrups, and warfarin) can increase urinary 5-HIAA levels and should be avoided 24 hours before and during specimen collection. More recently, there has been development of plasma 5-HIAA assays, however, these are not currently widely available [27]. Chromogranin A is another useful biomarker of NET, however it cannot be used in isolation to make the diagnosis [10, 14, 16, 23].

Cross-sectional imaging with a dual-phase staging computed tomography (CT) scan of the chest, abdomen, and pelvis may be appropriate if any hormonal tests are abnormal. Somatostatin receptor scintigraphy (Octreoscan), a nuclear medicine imaging modality, should be considered in any patients with histologically proven NETs or in patients with suspected NETs and lesions identified on CT. More recently ^{68}Ga-DOTATATE PET scan is becoming the new standard of functional imaging since it is more sensitive than Octreoscan [28].

Pathophysiology

NETs normally produce and secrete several different amines like serotonin, kinins, kallikrein, and prostaglandins. However, when the primary site is within the gut, the secreted amines are degraded by the liver and symptoms do not generally occur. In the presence of hepatic metastases, metastatic lesions secrete amines mainly serotonin that will exceed and bypass the degradation capacity of the liver and then reach the systemic circulation. In these cases, carcinoid syndrome will develop. Exceptions for this pathway are ovary, testicular, retroperitoneal, bronchial primary NETs, or metastases because they can directly access the systemic circulation [10, 22, 29–32].

In healthy people, serotonin is inactivated and metabolized to 5-HIAA by the hepatic monoamine oxidase and which is cleared by the kidneys. In these people, approximately 99% of tryptophan is used for the synthesis of nicotinic acid and less than 1% is converted to serotonin. However, in patients with carcinoid tumours there is a shift toward the production of serotonin [33].

Medical Management of Carcinoid Syndrome

Therapy for the carcinoid syndrome is aimed at lowering the secretion of serotonin and other amines/peptides by reducing hormone secretion or by reducing tumour load by surgical intervention or non-surgical therapies.

The mainstay of medical therapy is focused around use of somatostatin analogues to reduce hormone secretion in carcinoid syndrome. In 1986 Kvols *et al.* reported on 25 patients with carcinoid syndrome treated with the somatostatin analogue octreotide subcutaneous administered three times a day. Nearly 90% of patients reported significant improvement of symptoms, 72% of patients demonstrated a 50% reduction of 24 hr urine 5-HIAA level [34]. Since then octreotide and later developed lanreotide are considered cornerstones in the treatment of carcinoid syndrome with good evidence demonstrating significant reduction in carcinoid syndrome symptoms following commencement of somatostatin analogues [35].

Somatostatin analogues are also used for tumour control. The PROMID trial investigated the effect of long-acting octreotide (LAR) 30 mg intramuscular every 28 days, compared to placebo in patients with inoperable or metastatic NET. After 6 months 67% of patients treated with long-acting octreotide showed stable disease versus 37% of patients treated with placebo. The median time to progression for long-acting octreotide was 14.3 months versus 6 months in placebo patients [36]. The Clarinet trial investigated the effect of long-acting lanreotide 120 mg subcutaneous every 28 days compared to placebo in patients with inoperable or metastatic non-functional NETs. At 18 months treatment progression free survival for lanreotide median was not reached, for placebo was 18 months [37].

Tachyphylaxis can occur during treatment with long-acting octreotide or lanreotide after 6–18 months, dose escalation can result in regain of symptom control [19, 38].

There are five somatostatin receptors on the surface of NET cells and while octreotide and lanreotide have a good affinity for somatostatin receptor-2 and modest affinity for somatostatin receptor-5, they have a weaker affinity for other receptors. Pasireotide (SOM 230) was developed as new multireceptor somatostatin analogue, with high affinity for somatostatin receptor 1, 2, 3, and -5, for patients with inadequately symptom control on long-acting octreotide. A phase III trial randomized 110 patients between

pasireotide 60 mg or long-acting octreotide 40 mg every 28 days. During interim analysis pasireotide and long-acting octreotide showed similar rates of symptom control 20.9% and 26.7%, respectively. Progression free survival of 11.8 months with pasireotide and 6.8 months with octreotide was noticed. Within this interim analysis pasireotide was not expected to show superiority over long-acting octreotide in terms of symptom control and therefore the trail was halted [39, 40]. Currently, pasireotide does not have a licence for use in NETs.

Recently a new therapy has been licensed for use in carcinoid syndrome. Telotristat ethyl (Xermelo®) is an oral systemically available inhibitor of tryptophan hydroxylase. The amino acid tryptophan is converted into serotonin or niacin by tryptophan hydroxylase. Therefore, telotristat can reduce the production of serotonin. The TELESTAR trial was a phase III double-blind placebo-controlled trial in which in which a 135 participants were randomized to receive 250 mg telotristat ethyl tablet, 500 mg telotristat ethyl tablet or placebo three times daily for 12 weeks in the double-blind treatment period, followed by a 36 week open-label extension period. There was a statistically significant reduction in bowel movements in 42–44% of patients (250 mg or 500 mg) compared to 20% reduction in placebo patients, in addition to a significant reduction in urine-5 HIAA level and improvement of quality of life scores [41]. A companion trial (TELECAST) looking at patients with carcinoid syndrome and less than four bowel motions a day also demonstrated good safety and efficacy data [42].

Surgical Treatment for Tumour Control and Managing Carcinoid Syndrome

Surgical resection is the only curative treatment option for NETs. In patients with carcinoid syndrome surgical resection of the primary tumour and any metastatic disease should be considered if patients are fit enough to undergo surgery [43]. Liver debulking surgery should be considered when greater than 70% debulking threshold can be achieved, including the use of parenchyma sparing techniques with ablation or enucleation of the tumour [44, 45]. There is also a role for debulking tumour in selected cases to enable better control of carcinoid syndrome. In patients with florid carcinoid syndrome it is recommended they are established on somatostatin analogues and given intravenous octreotide cover prior to and during surgery [46].

Liver-Directed Therapy

Metastases from NETs are often isolated to the liver and depend on hepatic artery blood supply while the normal liver has portal vein blood supply. Therefore, embolization of the liver can result in necrosis of tumour tissue and consequent decrease in hormonal secretion. Embolization is commonly performed radiologically and can be performed with particles or chemoembolization. Contraindications to performing hepatic artery embolization include portal vein thrombosis, liver failure, and biliary reconstruction.

Symptomatic response is seen in 40–80% of cases and radiological response in 11–80% of cases [47, 48]. A biochemical response for hepatic embolization of 7–75%, and 12–75% for hepatic chemoembolization is seen [47, 49]. In the latter study Gupta *et al.* demonstrated no additional benefit of chemotherapy to transarterial hepatic embolization in metastatic midgut tumours [50]. Complications post procedure include ileus, portal vein thrombosis, hepatic abscess, hepatic fistula, encephalopathy, and renal insufficiency.

There is increasing evidence for the role of selective intrahepatic radioembolization. This technique involves injecting ^{90}Yttrium microspheres in to the liver. Yttrium is a high-energy, pure β-emitter with a half-life of 64 hours and maximum tissue penetration of 11 mm, which makes it very suitable for treatment of liver tumours. The microspheres are approximately 35 μm diameter, which means that they become permanently trapped at the arteriolar end of the capillary bed. The number of microspheres administered is such that the procedure has little or no devascularizing component. Systemic response is seen in about 70% of patients with carcinoid syndrome, radiological response in 22–66% of cases [51, 52]. There are no large studies comparing embolization to radioembolization, choice of treatment is based on patients' characteristics like number of liver lesions vascularization and proliferation index as well on local expertise [53].

New Treatments

Well-differentiated NETs express high levels of somatostatin receptors which is a target for radiolabelled somatostatin analogue therapy. Lutetium-177-DOTATATE (^{177}Lu) is a beta- and gamma-emitting radionuclide with a maximum particle range of 2 mm and a half-life of 160 hours and was studied in the NETTER 1 study. In this study 221 patients with metastasized or locally advanced midgut NET were randomized between ^{177}-Lu-DOTATATE- or octreotide LAR 60 mg. The interim analysis after 20 months showed progression free survival for 65.2% of patients treated with ^{177}-Lu-DOTATATE and 10.8% of patients treated with octreotide LAR. The partial response rate was 18% for ^{177}-Lu-DOTATATE and 3% for octreotide LAR patients [54].

One abstract shows improvement of symptoms in 62% patients with carcinoid syndrome treated with ^{177}-Lu-DOTATATE [55].

Retrospective analysis of long-term follow-up of patients treated with ^{177}-Lu-DOTATATE shows a progression free survival of 29 months and overall survival of 63 months [56].

Quality of life is studied with before and 6 weeks after treatment with ^{177}-Lu-DOTATATE showed significant improvement of functional and quality of life scores in 50 patients with metastatic NETs [57].

The Radiant 2 trial randomized 429 patients with low or intermediate grade metastatic or unresectable NET with carcinoid syndrome to octreotide LAR 30 mg with everolimus or placebo. Patients treated with octreotide LAR and everolimus showed progression free survival for 16.4 months versus progression free survival 11.3 months for patients treated with octreotide LAR and placebo. Octreotide LAR with everolimus demonstrated more reduction in 24h u-5 HIAA levels compared to octreotide LAR with placebo. Effect of everolimus on carcinoid related symptoms was no part of the study design [58]. The trial did not meet its primary endpoint and currently everolimus does not have a license for use in patients with carcinoid syndrome.

Prognosis

In general patients with metastatic well-differentiated GEP-NET have a good prognosis with an overall 5-year survival rate of 67%, depending tumour grading, staging, and site of origin [2]. Looking at details, patients with small bowel NETs have a prognosis of 168 months (in case of local disease only), 145 months (in case of regional lymph node metastases), and 70 months in case of distant metastases [19, 59, 60].

Patients with carcinoid syndrome have an overall survival of 4.7 years, compared to 7.1 years in patients without symptoms of carcinoid syndrome [8]. This data is from the SEER database and therefore, cause of death is not reported.

Potential Complications in Carcinoid Syndrome

Carcinoid Crisis

Carcinoid crisis is an acute and life-threatening complication of carcinoid syndrome. Symptoms are severe flushing, bronchospasm, profound hypotension, and arrhythmias or hypertension, central nervous system disfunction (stupor, confusion), diarrhoea. It is a result of rapid release of vasoactive hormones by in carcinoid tumour cells. Carcinoid crisis appears spontaneously but can be precipitated by surgery, anaesthesia, chemotherapy, lutetium therapy, radiological procedures, and stress, with a prevalence reported in retrospective studies of around 30%. Lack of appropriate treatment can result in death. Octreotide administered intravenously is advised to prevent and treat the symptoms. A commonly used preventive protocol is starting of octreotide infusion 50–100 μg/h at least 12 hours preoperatively. The octreotide infusion should be continued until 48 hours after the procedure [30, 46, 61]. If carcinoid crisis develops treatment is with extra intravenous boluses of octreotide and a higher infusion rate. If the syndrome is still refractory and/or bronchospasm is present then intravenous steroids may be required [46].

Carcinoid Heart Disease

Patients with carcinoid syndrome can develop carcinoid heart disease. The prevalence of carcinoid heart disease is estimated in modern series to be around 20% of patients with carcinoid syndrome, but this is higher in older series, with rates up to 60% [62, 63].

Carcinoid heart disease develops as a result of fibrous plaques depositing on cardiac valves and endocardial thickening usually in the right side of the heart. This plaque formation leads to retraction and fixation of the leaflets of the tricuspid and pulmonary valve. Main findings can be tricuspid regurgitation or stenosis and pulmonary stenosis, and both valves can be involved in 31% of patients. Left side heart disease can occur in less than 10% of patients mainly because of patent foramen ovale [10, 61, 64, 65]. Carcinoid heart disease may be asymptomatic but may present as fatigue and progressive exertional dyspnoea. Progressive development of heart failure can occur and eventually severe heart failure presenting with oedema, weight gain, ascites, and right upper abdominal pain as signs of right sided heart failure can be the next symptoms. Distension of the external jugular vein can be a clinical marker [10, 61, 64, 66].

Diagnosis is made by echocardiogram as golden standard to evaluate valve leaflets, valves, atrial, or ventricular dilatation and endocardial thickening [23, 61, 66]. The ENETS guideline advises annual echocardiographic screening in patients with carcinoid syndrome and signs of carcinoid heart disease [67]. The UKINETS guideline advises screening with serum N-terminal pro-brain natriuretic peptide (NT-proBNP), a surrogate marker of carcinoid heart disease [43]. Patients should be referred for echocardiography with a serum NTproBNP concentration >260 picogram/ml [17, 43, 66, 68]. Carcinoid heart disease is associated with limited 3-year survival of 31% compared to patients with metastatic NET without carcinoid heat disease of 68% [66, 69, 70]. Risk factors for development or progression of carcinoid heart disease are 5-HIAA levels ≥300 micromol/24 hours, ≥3 flushing episodes per day or treatment with cytotoxic chemotherapy [62, 71].

Valve replacement surgery or valvuloplasty is only definitive treatment and should be planned carefully and early in the course of the disease [66, 72]. The perioperative mortality since 2000 is around 5% [73, 74]. After surgery the 1-year survival is around 70% and after 5 year around 40% [73, 74]. It is hypothesized that reducing serotonin levels may prevent the development of carcinoid heart disease [61, 62, 67].

Mesenteric Fibrosis

Small bowel NETs frequently metastasize to lymph nodes in the mesentery. These lymph nodes can produce hormones and growth factors which can result in mesenteric fibrosis [29, 75]. The exact mechanism of fibrosis formation is not elucidated but growth factors such as transforming growth factor B, connective tissue growth factor, platelet derived growth factor, insulin like growth factors, epithelial growth factor, and fibroblast growth factor may play a role [75, 76]. The fibrosis causes shrinkage and fixation of the mesentery and mesenteric root to the retroperitoneum leading to small bowel obstruction. The mesenteric vessels may also become entrapped or occluded with resulting venous stasis and ischaemia and occasionally impairment of the arterial and venous circulation. The fibrosis can extend to the retroperitoneum and cause stenosis of the ureters and hydronephrosis. Because of the risk of ischaemia or small bowel obstruction when possible the removal of the primary tumour and mesenteric mass is advised [29, 61, 67, 77]. Two recent studies [76, 78] investigated the effect of mesenteric fibrosis diagnosed on CT scans and survival. Blažević *et al.* [78] showed mesenteric fibrosis was present in up to 41.4% of patients with biopsy-proven small bowel NET visible on CT scan. Mesenteric fibrosis was not a prognostic factor of overall survival with a median survival of 8.7 year and 5-year survival of 71%. They showed no survival benefit for palliative surgery, metastasectomy of mesenteric mass, or prophylactic surgery. Although Laskaratos *et al.* [76] showed primary resection was associated with a longer overall survival. In both studies older age, higher 24-hour urine 5-HIAA levels and chromogranin A levels were associated with developing of mesenteric fibrosis.

Pellagra

In carcinoid syndrome pellagra can develop with symptoms of dermatitis, diarrhoea, and dementia. It develops because of niacin deficiency. The amino acid tryptophan is the precursor of both niacin and serotonin. The uncontrolled production of serotonin by

carcinoid cells may lead to a depletion of tryptophan and impossibility to produce niacin leading to a niacin deficiency. The prevalence of pellagra in patients with carcinoid syndrome is estimated around 5% [16, 22, 61]. The recommended treatment is niacin supplementation 25–50 mg once daily [79].

Special Circumstances: Anaesthesia

Anaesthesia can provoke a carcinoid crisis therefore specific measures and advice are given to prevent a crisis. It starts with preoperative optimization with fluid and electrolyte resuscitation in patients with diarrhoea, adequate dose of somatostatin analogues and benzodiazepines to avoid stress. Followed by an octreotide protocol (as mentioned in the section carcinoid crisis).

Special advice regarding anaesthetic medication is to avoid possible histamine-releasing drugs such as thiopental, mivacurium, atracurium, cis-atracurium, succinylcholine, meperidine, and morphine. Induction of anaesthesia can be achieved with etomidate and propofol, while for muscle relaxation rocoronium or vecuronium can be used. Anaesthesia can be successfully maintained using a high-dose opioid infusion or inhalation anaesthetics such as isoflurane. First choices for analgesia are short acting opioids such as fentanyl, remifentanil, sufentanil, and alfafentanyl [30, 46, 64].

Summary

Carcinoid syndrome occurs in 20% of patients with NETs, primarily of the small bowel and bronchial primary sites. The overall outcomes are similar to those with non-functional tumours. However, management is focused on symptomatic management of the syndrome in addition to obtaining tumour control and growth. There are complications associated with carcinoid syndrome: carcinoid heart disease is an independent prognostic indicator for worse survival and should be routinely screened for in patients with carcinoid syndrome. In addition, development of mesenteric fibrosis can significantly impact on patients' quality of life. Further research is underway to manage these complications and control hormone secretion in carcinoid syndrome.

REFERENCES

1. Oberndorfer S. Karzinoide Tumoren des Dünndarms. *Frankf Z Pathol*, 1907; **1**: 426–32.
2. Fottner C, Ferrata M, Weber MM. Hormone secreting gastro-entero-pancreatic neuroendocrine neoplasias (GEP-NEN): when to consider, how to diagnose? *Rev Endocr Metab Disord*, 2017; **18**: 393–410.
3. Thorson A, Biorck G, Bjorkman G, Waldenstrom J. Malignant carcinoid of the small intestine with metastases to the liver, valvular disease of the right side of the heart (pulmonary stenosis and tricuspid regurgitation without septal defects), peripheral vasomotor symptoms, bronchoconstriction, and an unusual type of cyanosis; a clinical and pathologic syndrome. *Am Heart J*, 1954; **47**: 795–817.
4. Soga J. Carcinoids and their variant endocrinomas. An analysis of 11842 reported cases. *J Exp Clin Cancer Res*, 2003; **22**: 517–30.
5. Janson ET, Holmberg L, Stridsberg M, *et al*. Carcinoid tumors: analysis of prognostic factors and survival in 301 patients from a referral center. *Ann Oncol*, 1997; **8**: 685–90.
6. Garcia-Carbonero R, Capdevila J, Crespo-Herrero G, *et al*. Incidence, patterns of care and prognostic factors for outcome of gastroenteropancreatic neuroendocrine tumors (GEP-NETs): results from the National Cancer Registry of Spain (RGETNE). *Ann Oncol*, 2010; **21**: 1794–803.
7. Ploeckinger U, Kloeppel G, Wiedenmann B, Lohmann R. The German NET-registry: an audit on the diagnosis and therapy of neuroendocrine tumors. *Neuroendocrinology*, 2009; **90**: 349–63.
8. Halperin DM, Shen C, Dasari A, *et al*. Frequency of carcinoid syndrome at neuroendocrine tumour diagnosis: a population-based study. *Lancet Oncol*, 2017; **18**: 525–34.
9. Bax NDS, Woods HF, Batchelor A, Jennings M. Clinical manifestations of carcinoid disease. *World J Surg*, 1996; **20**: 142–6.
10. Ito T, Lee L, Jensenc RT. Carcinoid-syndrome: recent advances, current status and controversies. *Curr Opin Endocrinol Diabetes Obes*, 2018; **25**: 22–35.
11. Kaltsas GA, Besser GM, Grossman AB. The diagnosis and medical management of advanced neuroendocrine tumors. *Endocr Rev*, 2004; **25**: 458–511.
12. Ruszniewski P, Ish-Shalom S, Wymenga M, *et al*. Rapid and sustained relief from the symptoms of carcinoid syndrome: results from an open 6-month study of the 28-day prolonged-release formulation of lanreotide. *Neuroendocrinology*, 2004; **80**: 244–51.
13. McCormick D. Carcinoid tumors and syndrome. *Gastroenterol Nurs*, 2002; **25**: 105–11.
14. Vinik AI, Thompson N, Eckhauser F, Moattari AR. Clinical features of carcinoid syndrome and the use of somatostatin analogue in its management. *Acta Oncol*, 1989; **28**: 389–402.
15. Huguet I, Grossman A. Management of endocrine disease: flushing: current concepts. *Eur J Endocrinol*, 2017; **177**: R219–29.
16. van der Lely AJ, de Herder WW. Carcinoid syndrome: diagnosis and medical management. *Arq Bras Endocrinol Metabol*, 2005; **49**: 850–60.
17. Bhattacharyya S, Toumpanakis C, Caplin ME, Davar J. Analysis of 150 patients with carcinoid syndrome seen in a single year at one institution in the first decade of the twenty-first century. *Am J Cardiol*, 2008; **101**: 378–81.
18. Maasberg S, Knappe-Drzikova B, Vonderbeck D, *et al*. Malnutrition predicts clinical outcome in patients with neuroendocrine neoplasia. Neuroendocrinology, 2017; **104**: 11–25.
19. Öberg KE. The management of neuroendocrine tumours: current and future medical therapy options. *Clin Oncol*, 2012; **24**: 282–93.
20. Cheung VTF, Khan MS. A guide to midgut neuroendocrine tumours (NETs) and carcinoid syndrome. *Frontline Gastroenterol*, 2015; **6**: 264–9.
21. Massironi S, Sciola V, Peracchi M, Ciafardini C, Spampatti MP, Conte D. Neuroendocrine tumors of the gastro-entero-pancreatic system. *World J Gastroenterol*, 2008; **14**: 5377–84.
22. Schnirer II, Yao JC, Ajani JA. Carcinoid: a comprehensive review. *Acta Oncol (Madr)*, 2003; **42**: 672–92.
23. Davar J, Connolly HM, Caplin ME, *et al*. Diagnosing and managing carcinoid heart disease in patients with neuroendocrine tumors. *J Am Coll Cardiol*, 2017; **69**: 1288–304.
24. Lewis AR, Wang X, Magdalani L, *et al*. Health-related quality of life, anxiety, depression and impulsivity in patients with advanced gastroenteropancreatic neuroendocrine tumours. *World J Gastroenterol*, 2018; **24**: 671–9.
25. Pearman TP, Beaumont JL, Cella D, Neary MP, Yao J. Health-related quality of life in patients with neuroendocrine tumors: an

investigation of treatment type, disease status, and symptom burden. *Support Care Cancer*, 2016; **24**: 3695–703.
26. Tormey WP, FitzGerald RJ. The clinical and laboratory correlates of an increased urinary 5-hydroxyindoleacetic acid. *Postgrad Med J*, 1995; **71**: 542–5.
27. Adaway JE, Dobson R, Walsh J, *et al*. Serum and plasma 5-hydroxyindoleacetic acid as an alternative to 24-h urine 5-hydroxyindoleacetic acid measurement. *Ann Clin Biochem*, 2016; **53**(Pt 5): 554–60.
28. Srirajaskanthan R, Kayani I, Quigley AM, Soh J, Caplin ME, Bomanji J. The Role of 68 Ga-DOTATATE PET in patients with neuroendocrine tumors and negative or equivocal findings on 111 In-DTPA-octreotide scintigraphy. *J Nucl Med*, 2010; **51**: 875–82.
29. Basson MD, Ahlman H, Wängberg B, Modlin IM. Biology and management of the midgut carcinoid. *Am J Surg*, 1993; **165**: 288–97.
30. Castillo J, Silvay G, Weiner M. Anesthetic management of patients with carcinoid syndrome and carcinoid heart disease: the Mount Sinai algorithm. *J Cardiothorac Vasc Anesth*, 2018; **32**: 1023–31.
31. Kulke MH, Mayer RJ. Carcinoid tumors. *N Engl J Med*, 1999; **340**: 858–68.
32. Khan MS, El-Khouly F, Davies P, Toumpanakis C, Caplin ME. Long-term results of treatment of malignant carcinoid syndrome with prolonged release lanreotide (somatuline autogel). *Aliment Pharmacol Ther*, 2011; **34**: 235–42.
33. Caplin ME, Buscombe JR, Hilson AJ, Jones AL, Watkinson AF, Burroughs AK. Carcinoid tumour. *Lancet*, 1998; **352**: 799–805.
34. Kvols LK, Moertel CG, O'Connell MJ, Schutst AJ, Rubin J, Hahn RG. Treatment of the malignant carcinoid syndrome. Evaluation of a long-acting somatostatin analogue. *N Engl J Med*, 1986; **315**: 663–6.
35. Toumpanakis C, Garland J, Marelli L, *et al*. Long-term results of patients with malignant carcinoid syndrome receiving octreotide LAR. *Aliment Pharmacol Ther*, 2009; **30**: 733–40.
36. Rinke A, Müller HH, Schade-Brittinger C, *et al*. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID study group. *J Clin Oncol*, 2009; **27**: 4656–63.
37. Caplin ME, Pavel M, Ćwikła JB, *et al*. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*, 2014; **371**: 224–33.
38. Strosberg JR, Benson AB, Huynh L, *et al*. Clinical benefits of above-standard dose of octreotide lar in patients with neuroendocrine tumors for control of carcinoid syndrome symptoms: a multicenter retrospective chart review study. *Oncologist*, 2014; **19**: 930–6.
39. Wolin EM, Jarzab B, Eriksson B, *et al*. Phase III study of pasireotide long-acting release in patients with metastatic neuroendocrine tumors and carcinoid symptoms refractory to available somatostatin analogues. *Drug Des Devel Ther*, 2015; **9**: 5075–86.
40. Kvols LK, Öberg KE, O'Dorisio TM, *et al*. Pasireotide (SOM230) shows efficacy and tolerability in the treatment of patients with advanced neuroendocrine tumors refractory or resistant to octreotide LAR: results from a phase II study. *Endocr Relat Cancer*, 2012; **19**: 657–66.
41. Kulke MH, Hörsch D, Caplin ME, *et al*. Telotristat ethyl, a tryptophan hydroxylase inhibitor for the treatment of carcinoid syndrome. *J Clin Oncol*, 2017; **35**: 14–23.
42. Pavel M, Gross DJ, Benavent M, Perros P, Srirajaskanthan R, Warner RRP. Telotristat ethyl in carcinoid syndrome: safety and efficacy in the TELECAST phase 3 trial. *Endocr Relat Cancer*, 2018; **25**: 309–22.
43. Ramage JK, Ahmed A, Ardill J, *et al*. Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours (NETs). *Gut*, 2012; **61**: 6–32.
44. Nave H, Mössinger E, Feist H, Lang H, Raab HR. Surgery as primary treatment in patients with liver metastases from carcinoid tumors: a retrospective, unicentric study over 13 years. *Surgery*, 2001; **129**: 170–5.
45. Howe JR, Cardona K, Fraker DL, *et al*. The surgical management of small bowel neuroendocrine tumors. *Pancreas*, 2017; **46**: 715–31.
46. Kaltsas G, Caplin M, Davies P, *et al*. ENETS consensus guidelines for the standards of care in neuroendocrine tumors: pre- and perioperative therapy in patients with neuroendocrine tumors. *Neuroendocrinology*, 2017; **105**: 245–54.
47. Toumpanakis C, Meyer T, Caplin ME. Cytotoxic treatment including embolization/chemoembolization for neuroendocrine tumours. *Best Pract Res Clin Endocrinol Metab*, 2007; **21**: 131–44.
48. Kennedy A, Bester L, Salem R, Sharma RA, Parks RW, Ruszniewski P. Role of hepatic intra-arterial therapies in metastatic neuroendocrine tumours (NET): guidelines from the NET-Liver-Metastases Consensus Conference. *Hpb*, 2015; **17**: 29–37.
49. Eriksson BK, Larsson EG, Skogseid BM, *et al*. Liver embolizations of patients with malignant neuroendocrine gastrointestinal tumors. *Cancer*, 1998; **83**: 2293–301.
50. Gupta S, Johnson MM, Murthy R, *et al*. Hepatic arterial embolization and chemoembolization for the treatment of patients with metastatic neuroendocrine tumors: variables affecting response rates and survival. *Cancer*, 2005; **104**: 1590–602.
51. Jia Z, Wang W. Yttrium-90 radioembolization for unresectable metastatic neuroendocrine liver tumor: a systematic review. *Eur J Radiol*, 2018; **100**: 23–9.
52. Kennedy AS, Dezarn WA, McNeillie P, *et al*. Radioembolization for unresectable neuroendocrine hepatic metastases using resin 90Y-microspheres: early results in 148 patients. *Am J Clin Oncol Cancer Clin Trials*, 2008; **31**: 271–9.
53. Pavel M, O'Toole D, Costa F, *et al*. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. *Neuroendocrinology*, 2016; **103**: 172–85.
54. Strosberg J, El-Haddad G, Wolin E, *et al*. Phase 3 trial of 177 lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med*, 2017; **376**: 125–35.
55. Koffas A, Popat R, Dimitriou G, Quigley AM, Navalkissoor S, Caplin MTC. Efficacy of lutetium-177 DOTA octreotate peptide receptor radionuclide therapy in patients with advanced neuroendocrine tumours and carcinoid syndrome refractory to somatostatin analogues. 13th Annual ENETS conference, ENETS, 2016.
56. Brabander T, Van Der Zwan WA, Teunissen JJM, *et al*. Long-term efficacy, survival, and safety of [177Lu-DOTA0,Tyr3] octreotate in patients with gastroenteropancreatic and bronchial neuroendocrine tumors. *Clin Cancer Res*, 2017; **23**: 4617–24.
57. Teunissen JJM, Kwekkeboom DJ, Krenning EP. Quality of life in patients with gastroenteropancreatic tumors treated with [177Lu-DOTA0,Tyr3]octreotate. *J Clin Oncol*, 2004; **22**: 2724–9.

58. Pavel ME, Hainsworth JD, Baudin E, *et al*. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2011; **378**: 2005–12.
59. Dasari A, Shen C, Halperin D, *et al*. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*, 2017; **3**: 1335–42.
60. Yao JC, Hassan M, Phan A, *et al*. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, 2008; **26**: 3063–72.
61. Mota JM, Sousa LG, Riechelmann RP. Complications from carcinoid syndrome: review of the current evidence. *Ecancermedicalscience*, 2016; **10**: 1–17.
62. Bhattacharyya S, Toumpanakis C, Chilkunda D, Caplin ME, Davar J. Risk factors for the development and progression of carcinoid heart disease. *Am J Cardiol*, 2011; **107**: 1221–6.
63. Lundin L, Landelius J. Pathophysiology and natural history carcinoid heart disease: relationship of circulating vasoactive substances to ultrasound-detectable cardiac abnormalities. *Circulation*, 1985; **77**; 264–70.
64. Dierdorf SF. Carcinoid tumor and carcinoid syndrome. *Curr Opin Anaesthesiol*, 2003; **16**: 343–7.
65. Dobson R, Burgess MI, Pritchard DM, Cuthbertson DJ. The clinical presentation and management of carcinoid heart disease. *Int J Cardiol*, 2014; **173**: 29–32.
66. Dobson R, Valle JW, Burgess MI, Poston GJ, Cuthbertson DJ. Variation in cardiac screening and management of carcinoid heart disease in the UK and Republic of Ireland. *Clin Oncol*, 2015; **27**: 741–6.
67. Niederle B, Pape UF, Costa F, *et al*. ENETS consensus guidelines update for neuroendocrine neoplasms of the jejunum and ileum. *Neuroendocrinology*, 2016; **103**: 125–38.
68. Korse CM, Taal BG, De Groot CA, Bakker RH, Bonfrer JMG. Chromogranin-A and N-terminal pro-brain natriuretic peptide: an excellent pair of biomarkers for diagnostics in patients with neuroendocrine tumor. *J Clin Oncol*, 2009; **27**: 4293–9.
69. Pellikka PA, Tajik AJ, Khandheria BK, *et al*. Carcinoid heart disease. Clinical and echocardiographic spectrum in 74 patients. *Circulation*, 1993; **87**: 1188–96.
70. Westberg G, Wängberg B, Ahlman H, Bergh CH, Beckman-Suurküla M, Caidahl K. Prediction of prognosis by echocardiography in patients with midgut carcinoid syndrome. *Br J Surg*, 2001; **88**: 865–72.
71. Møller JE, Connolly HM, Rubin J, Seward JB, Modesto K, Pellikka PA. Factors associated with progression of carcinoid heart disease. *N Engl J Med*, 2003; **348**: 1005–15.
72. Hart EA, Meijs TA, Meijer RCA, *et al*. Carcinoid heart disease: a guide for screening and timing of surgical intervention. *Netherlands Hear J*, 2017; **25**: 471–8.
73. Connolly HM, Schaff HV, Abel MD, *et al*. Early and late outcomes of surgical treatment in carcinoid heart disease. *J Am Coll Cardiol*, 2015; **66**: 2189–96.
74. Mokhles P, van Herwerden LA, de Jong PL, *et al*. Carcinoid heart disease: outcomes after surgical valve replacement. *Eur J Cardio-Thoracic Surg*, 2012; **41**: 1278–83.
75. Modlin IM, Shapiro MD, Kidd M. Carcinoid tumors and fibrosis: an association with no explanation. *Am J Gastroenterol*, 2004; **99**: 2466–78.
76. Laskaratos F-M, Diamantopoulos L, Walker M, *et al*. Prognostic factors for survival in patients with small bowel neuroendocrine tumours associated with mesenteric desmoplasia. *Neuroendocrinology*, 2018; **106**: 366–80.
77. Daskalakis K, Karakatsanis A, Stålberg P, Norlén O, Hellman P. Clinical signs of fibrosis in small intestinal neuroendocrine tumours. *Br J Surg*, 2017; **104**: 69–75.
78. Blažević A, Hofland J, Hofland LJ, Feelders RA, De Herder WW. Small intestinal neuroendocrine tumours and fibrosis: an entangled conundrum. *Endocr Relat Cancer*, 2018; **25**: R115–30.
79. Liang V, Go W, Srihari P, Anne L, Burns K. Nutrition and gastroenteropancreatic neuroendocrine tumors. *Endocrinol Metab Clin North Am*, 2010; **39**: 827–37.

6.4

Lung Neuroendocrine Tumours

Kok Haw Jonathan Lim, Juan W. Valle, and Wasat Mansoor

Epidemiology

Lung neuroendocrine tumours (NETs) account for 25–30% of all NETs (**Figure 6.4.1**a) [1–3], constituting the largest subgroup of NETs outside the gastroenteropancreatic (GEP) tract. In general, they represent 1% of all cases discussed at lung multidisciplinary team (MDT) meetings, and altogether they form up to 2% of all lung cancers [2, 4–6]. Incidence has dramatically increased over the past three decades from 0.3 per 100 000 in 1973 [7] to most recently 1.49 per 100 000 reported in the United States (US) Surveillance, Epidemiology and End Results (SEER) 18 (2000–2012) (**Figure 6.4.1**b) [8], making it the most rapidly increasing subgroup of NETs by approximately 6% per year [1]. It is believed that the disease remains largely undiagnosed in the general population. Thus, it is predicted to continue rising with improved awareness among respiratory physicians who are often the first point of referral and also with better diagnostic tools now available for this disease entity [9].

The peak incidence for lung NETs is observed in the fourth to sixth decade of life [4]. There is a slight female preponderance and they tend to be more common among white Caucasians, and are relatively rare in other ethnicities including among Blacks, Asians, or Hispanics [10]. Importantly, lung NETs are the most common primary lung neoplasms diagnosed in children and late adolescents, particularly typical carcinoid (TC) [11–15]. Lung NETs are not usually associated with a history of smoking, except in large cell neuroendocrine carcinoma (LCNEC) and small cell lung carcinoma (SCLC) which develop in heavy smokers [16].

Histological Classification

Lung NETs originate from specialized epithelial cells called pulmonary neuroendocrine cells (PNECs) which can be found interspersed among other cells lining the entire respiratory tract, from the nasal epithelium to the terminal bronchioles [17–19]. PNECs exist as individual cells or as a cluster of cells, also known as neuroepithelioid bodies (NEBs) [20]. They are thought to be key in regulating fetal and neonatal airway development and play a particularly significant role as chemoreceptors for hypoxia [21–23]. Throughout development, it is hypothesized that PNECs and NEBs exert a paracrine function in regulating the growth of surrounding epithelial cells, and continue their function as oxygen-sensitive chemoreceptors until adulthood [24, 25].

There is still no clear understanding in the pathogenesis and stepwise malignant transformation of these PNECs and NEBs to lung NETs. One postulated theory is the concept of a putative precursor lesion called diffuse idiopathic neuroendocrine cell hyperplasia (DIPNECH) [26–28]. DIPNECH is a rare abnormal growth or hyperplasia in the PNECs and NEBs, resulting in the formation of multiple carcinoid 'tumourlets', which are usually <0.5 mm nodules, and are often associated with a surrounding fibrotic process in the lung parenchyma (**Figure 6.4.2**) [27, 29, 30]. It is predominantly a benign condition confined to the bronchial epithelium, but there is compelling evidence to suggest that it can be potentially premalignant with higher prevalence of DIPNECH often observed in the background when lung NET is diagnosed [30, 31]. This is currently being investigated further through longer term observational studies. Intriguingly, DIPNECH is about ten times more common in females compared to males [32].

According to the World Health Organization (WHO) criteria [33], there are four known main subtypes of lung NETs,

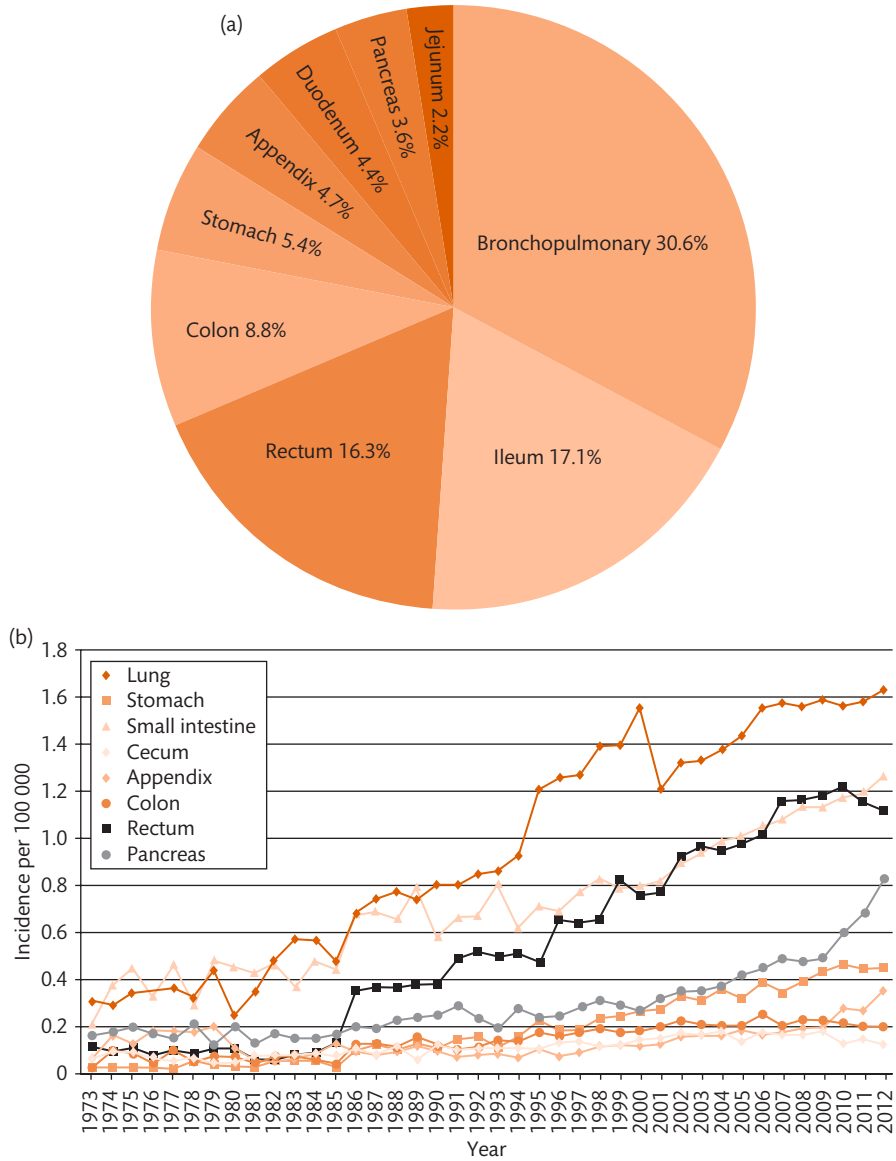


Figure 6.4.1 Epidemiological representation of neuroendocrine tumours (NETs). (a) Incidence of NETs by anatomical distribution. (b) Graph showing progressive increase in incidence of NETs, particularly lung, small intestine, and rectum NETs.

Panel (a) reproduced with permission from Bodei L, Ćwikla JB, Kidd M, Modlin IM. The role of peptide receptor radionuclide therapy in advanced/metastatic thoracic neuroendocrine tumors. *J Thorac Dis*. 2017;9(Suppl 15):S1511-S23. Panel (b) reproduced with permission from Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, *et al*. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol*. 2017;3(10):1335–42. Copyright © 2017, American Medical Association.

characterized based on their morphology and pathological characteristics, as summarized in Table 6.4.1 and Figure 6.4.3. Historically, well-differentiated tumours are termed lung carcinoids, and consist of typical and atypical lung carcinoids (TC and AC). There is a stronger link between DIPNECH to the formation of TC, and less is known of its association to AC [27]. Apart from morphological appearance, the key differentiating factor between TC and AC is predominantly based on the proportion of actively dividing cells, quantified by the mitotic rate, and presence or absence of necrosis [2, 6]. TC has a mitotic rate of less than 2 per x10 high power field (HPF) and necrosis is not usually observed. In the commonly used three-tier tumour grading classification system, TC would be considered a low grade (grade 1) tumour. AC tends to be intermediate grade (grade 2) and has 2–10 mitoses per ×10 HPF and is often associated with focal areas of necrotic cells. Meanwhile the more aggressive poorly differentiated lung NETs are formed of LCNEC and SCLC which are almost entirely different entities, and have no known association with DIPNECH. In contrast to lung carcinoids, both subtypes of lung neuroendocrine carcinomas (NECs) (LCNEC and SCLC) are poorly differentiated (grade 3) tumours frequently found with evidence of extensive necrosis and will have more than 10 mitoses

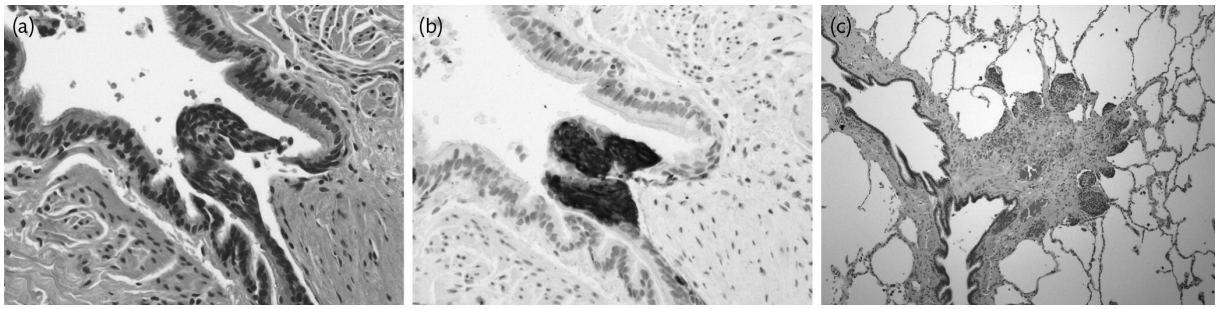


Figure 6.4.2 Morphological characteristics of neuroendocrine cell hyperplasia and carcinoid 'tumourlets'. (a) Early evidence of neuroendocrine cell hyperplasia illustrated with H&E staining with finger-like intraluminal projections from underlying stroma, ×200. (b) Similar cross-sectional view as (a), but with synaptophysin staining for neuroendocrine cell hyperplasia, ×200. (c) Low power (×100) overview of carcinoid 'tumourlets' with surrounding fibrotic changes in the stroma.

Table 6.4.1 World Health Organization (WHO) classification of lung neuroendocrine tumours (NETs)

	Premalignant	Lung Carcinoids (9%)		Lung NECs (91%)	
	DIPNECH	TC	AC	LCNEC	SCLC
Morphology	Well-differentiated	Well-differentiated	Well-differentiated	Poorly differentiated	Poorly differentiated
Grade	Low (Grade 1)	Low (Grade 1)	Intermediate (Grade 2)	High (Grade 3)	High (Grade 3)
Mitotic rate (×10 HPF)	None	<2	2–10	>10 (median 70)	>10 (median 80)
Necrosis	None	None	Punctate (Often)	Diffuse (Often)	Diffuse (Often)

per ×10 HPF (median mitotic rate of 70 per ×10 HPF for LCNEC and 80 per ×10 HPF for SCLC, respectively). In general, TC is 8 to 10 times more prevalent than AC, and together TC and AC constitutes 9% of all lung NETs, with the remaining 91% being high-grade NECs [1, 34].

Most lung NETs should demonstrate immunoreactivity to chromogranin A, and synaptophysin and/or CD56 on the immunohistochemistry panel [2, 16]. Cytokeratins, thyroid-transcription factor-1 (TTF-1) and the homeobox transcription factor CDX-2 may also be useful in the initial test to distinguish between primary lung NETs and other potential primary NETs with metastatic lung lesions [10]. Currently, the proliferation index Ki-67 which is widely used as a diagnostic tool in GEP-NETs is not yet assimilated as part of the classification criteria in lung NETs [10]. There have been several retrospective studies suggesting its utility as a prognostic tool [35–37], and there is a strong proposal for it to be incorporated into the next version update of the WHO classification to further guide clinical management.

Recently, there has also been interesting Next Generation Sequencing (NGS) data where targeted sequencing of 418 genes was performed on 53 TC, 35 AC, 27 LCNEC and 33 SCLC cases [38]. It appeared that NETs and carcinomas shared most of the mutated genes, but had different prevalence. There is a significant enrichment of mutations in TP53, RB1, cycle regulation genes and PI3K/Akt/mTOR in the carcinomas compared to NETs [38]. MEN1 alterations were exclusively a feature of well-differentiated NETs [38]. Intriguingly, novel mutations in chromatin-remodelling genes were also found in 45% of NETs and 50% of carcinomas, and were postulated by the authors to play a major part of the pathogenesis of NETs [38].

Clinical Features

The majority of lung NETs are found incidentally or diagnosed postoperatively. ACs are often peripherally located, and majority of patients are therefore asymptomatic [39]. On the other hand, TCs tend to be more centrally located and thus may report obstructive symptoms [39]. Despite this, lung carcinoids can occur anywhere in the lung parenchyma and in general, and remain difficult to diagnose. In most instances, patients report non-specific symptoms such as dyspnoea, wheeze, chest discomfort, haemoptysis, and recurrent chest infections [9, 10]. As a result, many patients tend to have advanced disease at diagnosis [40]. A recent study estimated

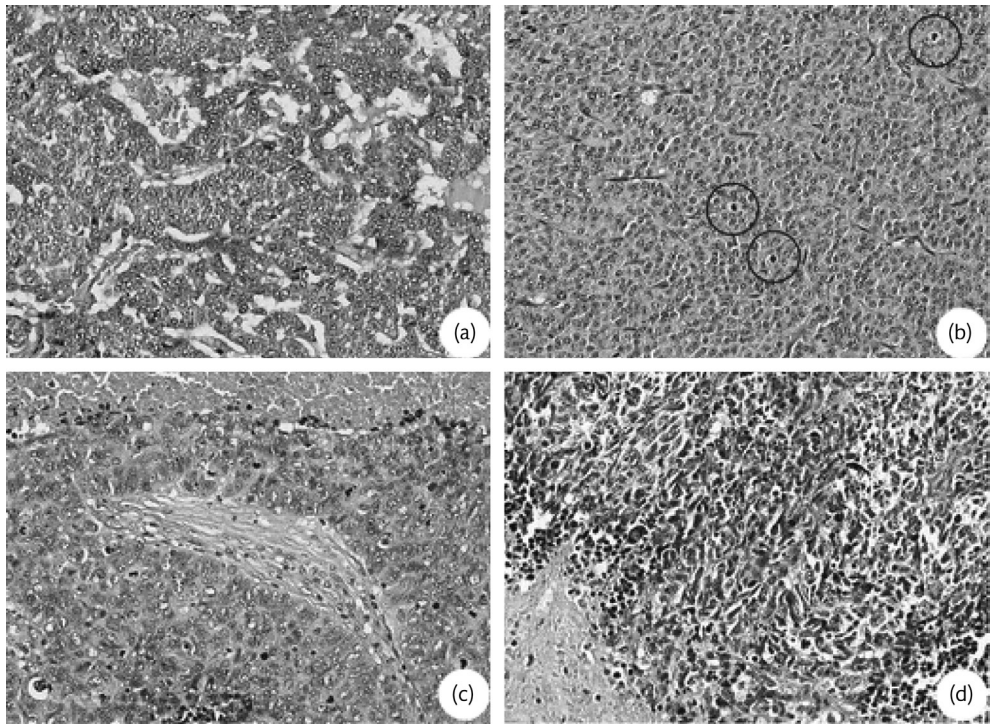


Figure 6.4.3 Morphological appearance and description of the four subgroups of lung neuroendocrine tumours (NETs). (a) Typical carcinoid (TC) is characterized by bland, polygonal, and uniform tumour cells with round nuclei with no areas of necrosis. (b) Atypical carcinoid is similar to TC but has higher mitotic activity and focal areas of necrosis (*circles*). (c) Large cell neuroendocrine carcinoma (LCNEC) consists of large tumour cells arranged in organoid, trabecular, or palisading patterns with prominent nuclei, granular chromatin, and diffuse necrosis. (d) Small cell lung carcinoma (SCLC) consists of small round/oval and angulated cells with dispersed 'salt and pepper' chromatin and extensive necrosis.

the median time to diagnosis of lung NET from being initially symptomatic was reported to be 2 years [41].

It is important to be aware that while uncommon, a small proportion (1–5%) of cases may present with carcinoid syndrome, and this can be independent of the presence of liver metastases [9, 10]. In even rarer circumstances, lung NETs may be discovered as part of other syndromes such as Cushing's syndrome (~40% of ectopic adrenocorticotropic hormone (ACTH) production are from lung carcinoids) (1–6%), multiple endocrine neoplasia type 1 (MEN-1) associated forms (less than 5%) especially among those with a strong family history, and acromegaly due to ectopic release of growth hormone-releasing hormone (GHRH) or insulin-like growth factor-1 (IGF-1) (very rare) [9, 10].

Diagnosis

Initial investigations will be guided by thorough history taking and vigilant clinical examination. In most patients who are relatively asymptomatic, baseline biochemical tests performed should include assessment of renal function, calcium, glucose, and serum chromogranin-A levels [10]. Chromogranin-A is raised in up to 75% of lung carcinoids and 60% of SCLC [42]. Neuron-specific enolase (NSE) may be raised in high-grade NETs [43]. Specific tests such as serum/urinary 5-hydroxyindoleacetic acid (5-HIAA), serum/urinary cortisol, serum ACTH, and serum GHRH or IGF-1 should only be requested if there is clinical suspicion of paraneoplastic syndromes [10]. Serum calcium and parathyroid hormone (PTH) levels can be simple screening tests for MEN-1, and further investigations should be undertaken by specialist services with appropriate genetic counselling. Currently, there is no validated molecular tests used and therefore is not routinely recommended outside clinical trials.

Plain imaging with chest X-ray may be useful as a screening tool (up 40% of incidental cases [44]) although smaller lesions are likely to be missed. All patients suspected to have NETs should have a contrast-enhanced full staging computed tomography (CT) scan of the thorax, abdomen, and pelvis as the standard of care [10]. Note that DIPNECH are best diagnosed using high resolution CT (HRCT) scan with an additional expiration study [10]. The current gold standard investigation in the diagnosis of lung carcinoids (both TC and AC) is ^{68}Gallium (^{68}Ga)-DOTA (DOTATOC or DOTATATE)-PET CT scan (not indicated in DIPNECH), and is recommended as a baseline imaging for all patients including those who had curative resection for their disease and also in advanced lung carcinoids [45].

The main indication for performing an ^{18}F-fluorodeoxyglucose (FDG)-PET CT scan is in high-grade (grade 3) tumours [46], although a proportion of patients might have already had this as part of their lung cancer preoperative work-up. This is generally only helpful in tumours with Ki-67 of more than 10% [47]. Other nuclear medicine imaging modalities such as somatostatin receptor scintigraphy (SRS) with ^{111}Indium pentetreotide and OctreoScan can be useful alternatives in instances when ^{68}Ga-DOTA-PET CT scan is not available or funded [10]. In other specific situations, MRI scan with contrast is often useful in characterizing indeterminate liver lesions when metastatic disease is suspected. Another simple alternative to MRI when contraindicated is a triple-phase contrast-enhanced ultrasound scan of the liver. In cases where bone metastases are suspected, a whole-body nuclear medicine bone scan is standard practice [10]. MRI scan can also be useful to guide surgical or radiotherapy interventions in circumstances when patients are symptomatic of their specific areas of bone metastases.

As tissue biopsy is the gold standard for definitive diagnosis, both rigid and flexible bronchoscopy can be helpful to obtain tissue sample in centrally located tumours. This must be performed with caution in view of risks of bleeding. Most patients with peripheral tumours commonly have a CT-guided biopsy instead. In addition, it is also good practice to perform a baseline echocardiogram to ensure no evidence of carcinoid heart disease, particularly in symptomatic patients. Where patients are frail or elderly, there may be safety concerns in pursuing a biopsy to obtain a tissue diagnosis. In this situation, it is reasonable to assume the diagnosis if the lung lesion is negative on FDG-PET but avid on a ^{68}Ga-DOTA-PET CT scan or OctreoScan.

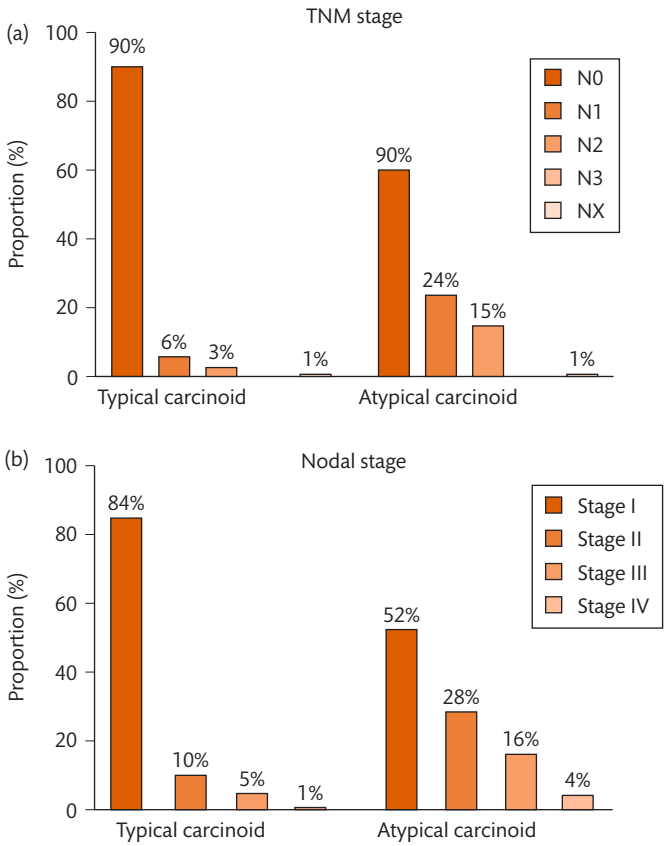


Figure 6.4.4 Bar graph comparing the nodal and AJCC TNM stage distribution between typical carcinoid (TC) and atypical carcinoid (AC) at presentation.

Data sourced from Filosso PL, Ferolla P, Guerrera F, Ruffini E, Travis WD, Rossi G, *et al.* Multidisciplinary management of advanced lung neuroendocrine tumors. *J Thorac Dis.* 2015;7(Suppl 2):S163–71.

Staging

The clinical staging of lung NETs follows the current American Joint Committee on Cancer (AJCC) TNM (**T**umour, lymph **N**odes affected, **M**etastases) staging system for lung cancer (8th edition) [48]. More than 90% of TC are node-negative, and therefore a majority (84%) of patients present with stage 1 disease, compared to about 60% of AC being node-negative (52% stage 1) (Figure 6.4.4) [49].

Management

The only curative treatment for lung NETs is by surgery. Complete anatomical resection either in the form of lobectomy or segmentectomy with adequate lymph node dissection (minimum of six nodes/stations, including three mediastinal and subcarinal stations) is recommended for R0 resection [10, 50]. In oligometastatic disease, metastasectomy should also be strongly considered particularly in syndromic patients with liver metastases (including complex cases as long as >90% of tumour burden can be removed) [10, 51].

For patients who are poor surgical candidates due to frailty or age, then more conservative approaches have been adopted. In these situation debulking surgery has been shown to be beneficial in delaying progression using endoluminal therapies such as diathermy, localized laser ablative therapy or radiofrequency ablation (RFA) [52].

Despite complete surgical resection of early disease, there is a risk of relapse and this is often with distant metastases particularly in the liver (55%), lung (25%), and bone (20%) [53, 54]. In a large prospective observational study of 337 patients with resected TC and AC, after a median follow-up time of 3.5 years, the recurrence rate is 6%, of which 95% were distant metastases [53]. Interestingly, most recurrences (76%) were symptomatic and not detected as part of scheduled imaging surveillance [53]. Recurrence rates vary widely across studies. There is a tendency for AC to relapse earlier, within the first 2 years of follow-up, with median time to recurrence of 1.8 years [53]. The median time to recurrence in TC is 4 years [53]. In both, late recurrences have also been reported, which therefore warrants the need for long-term surveillance [10].

To date, there is still no good quality prospective randomized control trials providing sufficient evidence base to advocate any

adjuvant treatment. Therefore, while this divides opinions, adjuvant chemotherapy or radiotherapy is not currently considered standard of care following curative surgery [10]. There are a number of small retrospective cohort studies which investigated the utility of adjuvant chemotherapy with or without radiotherapy following curative surgery. All of these studies investigated the role of combination chemotherapy with cisplatin and etoposide only in AC. In one of the largest cohort studied, a total of 39 node-positive AC patients received adjuvant chemotherapy out of 629 eligible patients identified, and did not demonstrate a statistical benefit in 5-year overall survival between having adjuvant chemotherapy versus surgery alone (69.7% vs. 80.9%, $P = 0.096$) [55]. Similarly, in another study with 89 node-positive AC out of 581 eligible patients, the 5-year overall survival is 47.6% in the adjuvant chemotherapy arm versus 67% in the observation arm ($P = 0.46$) [56]. In this study, there was no difference among those with node-positive disease or not. In fact, North American Neuroendocrine Tumor Society (NANETS) does not recommend any adjuvant treatment due to the lack of supportive evidence [57]. Meanwhile, in the National Comprehensive Cancer Network (NCCN) guidelines, stage II and III AC may be considered for adjuvant treatment [58]. Similarly, European Neuroendocrine Tumour Society (ENETS) suggests selective use of adjuvant chemotherapy in patients with AC who are node-positive and have a high proliferation index [10]. Adjuvant treatment is generally not recommended for TC [10].

There is also no robust imaging surveillance schedule currently recommended in guidelines, and this tends to vary between centres and treating physicians. The general consensus is that postoperative follow-up in high-volume centres is important and the intensity of active surveillance is determined predominantly by the nodal status at outset. The European Society for Medical Oncology (ESMO) published the ENETS expert consensus statement recommending: for resected TC—standard CT imaging at 3, 6, and 18 months followed by annual chest X-ray and 3-yearly CT scan thereafter; and a more intense schedule for resected AC—standard CT imaging at 3 months, followed by 6-monthly for up to 5 years [10].

There have been promising advances in recent years on the management of advanced or metastatic lung carcinoids. As this is a really heterogenous group of rare disease, treatment decisions can differ between individual treating physicians and in view of this, the key recommendation is to centralize treatment in high-volume centres with access to a specialist MDT input. Treatment options are guided by the aims of treatment. In about a third of patients who present with carcinoid syndrome, the first priority would be to control hormone-related symptoms with somatostatin analogues. Both ocreotide LAR 30 mg given intramuscularly (phase III PROMID trial [59]) and lanreotide autogel 120 mg subcutaneously (phase III CLARINET trial [60]) given 4-weekly, respectively, have been shown to be clinically beneficial and have now been licensed by Food and Drug Administration (FDA) and European Medicines Agency (EMA) for control of carcinoid syndrome. Both forms of somatostatin analogues have also been shown to have some antitumour or disease-modifying effect, and therefore are reasonable options in the first-line setting, particularly in patients with slowly growing TC or AC.

More recently, the phase III RADIANT-4 trial [61] led to the FDA and EMA approval of everolimus in non-functional lung NETs. Everolimus is an oral targeted therapy which inhibits the mammalian target of rapamycin (mTOR) pathway which is activated in some lung NETs [62]. In this randomized placebo-controlled prospective trial, there were 90 lung NETs patients included out of a total of 302 patients with non-functioning midgut NETs [61]. Treatment with everolimus 10 mg orally per day led to an superior median progression-free survival of 11.0 months (95% CI 9.2–13.3) compared to 3.9 months (95% CI 3.6–7.4) in the placebo arm, and was associated with a 52% risk reduction of progression or death, hazard ratio(HR) 0.48 (95% CI 0.35–0.67) (P <0.00001) [61]. The main grade 3 or 4 adverse events included stomatitis (9% vs. 0%), diarrhoea (7% vs. 2%), infections (7% vs. 0%), anaemia (4% vs. 1%), fatigue (3% vs. 1%) and hyperglycaemia (3% vs. 0%) [61]. Previously, the phase III RADIANT-2 trial [63, 64] did not demonstrate statistically significant overall survival benefit of everolimus in functional lung NETs following independent analysis of a small subgroup of patients, and therefore has not been recommended in functional lung NETs.

Treatment with palliative chemotherapy is mainly considered in AC which are 'actively progressive' [10]. There is no robust prospective clinical trial data to support specific regimes, but the most explored chemotherapy is the combination containing either 5-fluorouracil (5-FU) or oral capecitabine with streptozotocin or temozolomide [65–67]. This experience is built on small cohort of lung NETs analysed retrospectively. Other chemotherapy regimens potentially used include monotherapy with 5-fluorouracil (5-FU) or oral capecitabine, dacarbazine, or temozolomide [57]. Generally, it is estimated that response rates to chemotherapy range from 0 to 32%, with stabilization of disease in up to 70% of patients, and an estimated progression-free survival in the region of 5 to 10 months [68]. Prospective series are still lacking. Immunomodulating agent with interferon 2a can be considered in a selected group [57].

Peptide receptor radionuclide therapy (PPRT) is a promising alternative in patients with advanced disease following systemic treatment. In an Italian single-centre cohort study with 114 patients with advanced lung carcinoids, PRRT had been found to delay disease progression by a median of 28.0 months with objective response observed in 26.5% of patients [69]. This treatment tends to be well tolerated. The use of PRRT in lung NETs await outcomes from prospective studies and is not currently licensed for the treatment of advanced lung NETs in the United Kingdom. See **Box 6.4.1** and **Figure 6.4.5**.

There are several ongoing trials investigating further treatment options for lung NETs (**Table 6.4.2**). The Phase II LUNA trial [71] investigating treatment with pasireotide vs. everolimus vs. combination had enrolled 121 lung NET patients and recently met its primary endpoints of efficacy and safety in all three cohorts. The Phase III SPINET trial (lanreotide vs. placebo) also specifically for lung NETs will potentially provide the first robust evidence to support the use of somatostatin analogues in this subgroup of

Box 6.4.1 Factors to consider when deciding management plans

Factors when considering treatment

- Pathology
- Proliferation (mitotic index, Ki67)
- Somatostatin receptor expression
- Growth rate
- Stage
- Associated syndromes
- Performance status

Reproduced with permission from Talbot D. Thoracic Neuroendocrine Tumours. ESO-ESMO-RCE Clinical Update on Rare Adult Solid Cancers; Milan, Italy 2017.

NET. The phase II ATLANT trial will examine the combination of lanreotide and temozolomide. Other promising antiangiogenesis agents are also being investigated including sunitinib, axitinib, and cabozantinib. There are also many upcoming signal-finding studies with immunotherapy being explored, including treatment with monotherapy anti-PD1 agent pembrolizumab (within KEYNOTE-058 trial) or PDR001, or administered in combination anti-PD1 and anti-CTLA4: nivolumab and ipilimumab (DART trial) or durvalumab (MEDI4736) and tremelimumab (DUNE trial). Interestingly, there is also a pilot study investigating the combination of lanreotide and metformin (MetNET-2 trial).

Prognosis

Generally, patients with TC have better survival outcomes compared to those with AC. The 5-year median overall survival in TC is 92–100% compared to 61–88% in AC [4, 49, 72]. Table 6.4.3 shows the overall survival of lung NETs according to stage at diagnosis. In comparison, the 5-year median overall survival in LCNEC is 16–57% and SCLC had the worst outcome with only 2–5% of patients alive at 5 years [72]. The strongest predictor for recurrence remains the nodal status and potentially extent of surgery (R1 and R2 resections at higher risk compared to R0) [73].

Future Work

Lung NET is a heterogeneous group of rare tumours whereby further understanding of the biological basis of its pathogenesis and natural history will be pivotal in driving improvement in management options available. Prospective trials particularly in the adjuvant setting are much needed to improve success in curative treatments. There should be a systematic approach to personalize treatment for each patient following treatment line failure in the metastatic setting, including consideration of basket and umbrella trials with progress made in further understanding molecular pathways and genetics in lung NETs.

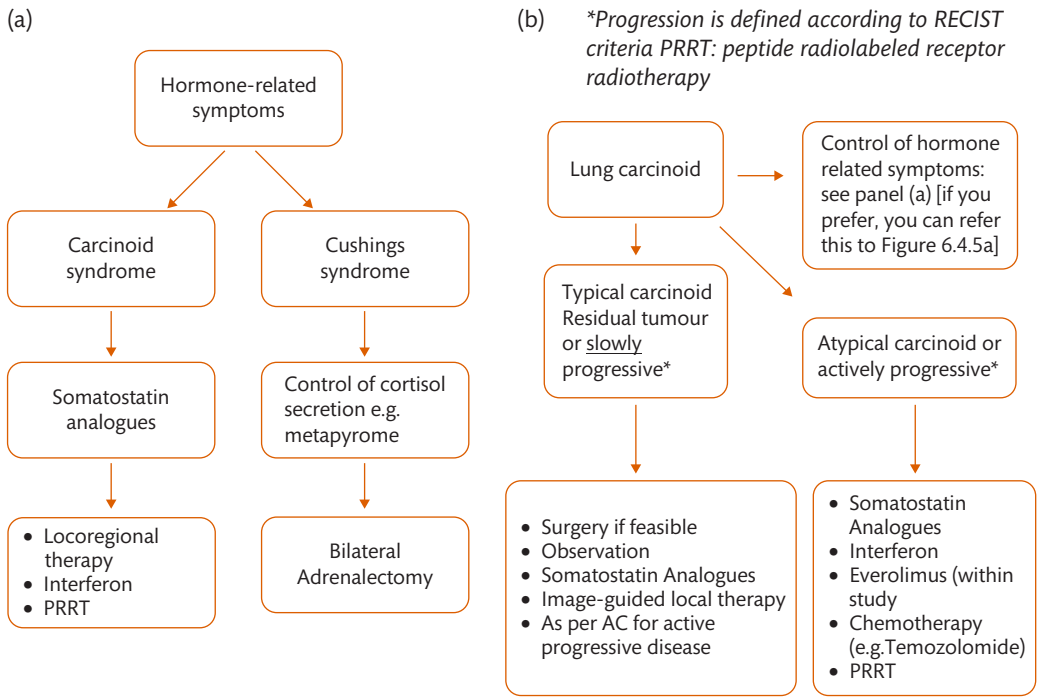


Figure 6.4.5 Flow chart summarizing the ENETS management guidelines for lung neuroendocrine tumours (NETs).

Reproduced with permission from Caplin ME, Baudin E, Ferolla P, Filosso P, Garcia-Yuste M, Lim E, *et al*. Pulmonary neuroendocrine (carcinoid) tumors: European Neuroendocrine Tumor Society expert consensus and recommendations for best practice for typical and atypical pulmonary carcinoids. *Ann Oncol*. 2015;26(8):1604–20.

Table 6.4.2 Repertoire of clinical trials relevant to lung neuroendocrine tumours (NETs) currently recruiting or awaiting report

Study name	Phase	Study location	Experimental arm	Control arm	Study population	Estimated sample size	Primary end point	Primary completion date
NCT02683941 SPINET	III	Multicentre; Austria, Canada, Denmark, France, Germany, Italy, Netherlands, Poland, Spain, United Kingdom, United States	Lanreotide Autogel/ Depot	Placebo	Unresectable or metastatic grade 1 or 2 lung NET, previously treated ≤2 prior systemic antiproliferative agents	216	PFS	August 2019
NCT02698410 ATLANT	II	Single centre; Italy	Lanreotide Autogel + Temozolomide	–	Unresectable or metastatic, progressive, grade 1 or 2 lung/ thymic NET, ≤3 lines of previous treatment	40	DCR	May 2019
NCT02823691 MetNET-2	I	Single centre; Italy	Lanreotide Autogel + Metformin	–	Unresectable or metastatic, grade 1 or 2 lung NET, pretreated	20	Incidence of SAE/ AEs	April 2022
NCT01563354 LUNA	II	Multicentre; Denmark, France, Germany, Greece, Italy, Netherlands, Spain, Sweden, United Kingdom	Pasireotide LAR vs. Pasireotide LAR + everolimus	Everolimus	Advanced, progressive grade 1 or 2 lung/thymic NET, treatment-naïve, or pretreated	124	PFR at 9 mo	November 2017
NCT01253161	II	Single centre; United States	Pasireotide LAR	–	Unresectable or metastatic, progressive, grade 1 or 2 lung or pancreatic NET, no prior systemic antiproliferative agents	29 (total)	PFS	December 2018
NCT02588170 SANET-EP	III	Multicentre; China	Sulfatinib	Placebo	Advanced grade 1 or 2 unresectable or metastatic extrapancreatic NET, previously treated ≤2 prior systemic antiproliferative agents	273 (total)	PFS	December 2018
NCT01744249 AXINET	II/III	Multicentre; Germany, Italy, Spain, United Kingdom	Axitinib + Sandostatin LAR	Placebo	Unresectable or metastatic grade 1 or 2 extrapancreatic NET, previously treated ≤2 prior systemic antiproliferative agents	253 (total)	PFS	November 2019
NCT01466036	II	Single centre; United States	Cabozantinib	–	Cohort 2: Unresectable or metastatic, progressive, grade 1 or 2 lung NET, pretreated	35	ORR	October 2018
NCT03375320 CABINET	III	Multicentre; United States	Cabozantinib	Placebo	Unresectable or metastatic grade 1 or 2 NET, following previous everolimus	395 (total)	PFS	January 2021
NCT02259725	II	Single centre; United States	Regorafenib	–	Unresectable or metastatic, progressive, grade 1 or 2 lung or GEP- NET, no prior targeted or antiangiogenic therapy	48 (total)	PFS	August 2019
NCT02399215	II	Single centre; United States	Nintedanib	–	Unresectable or metastatic grade 1 or 2 non-pancreatic NET, on stable dose of Ocreotide LAR or Lanreotide for 3 mo	30 (total)	PFS	July 2018

NCT02318784	II	Multicentre; United States	Carfilzomib	–	Unresectable or metastatic grade 1 or 2 NET (excluding ACs)	62 (total)	ORR	February 2019
NCT02575300	II	Single centre; United States	Ibrutinib	–	Cohort 1: Unresectable or metastatic grade 1 or 2, progressive, GI, lung, or unknown primary carcinoid	30	ORR	December 2018
NCT02936323	I/2a	Multicentre; United Kingdom, United States	PEN-221	–	Unresectable or metastatic all grades of NET including LCNEC and SCLC, progressive, pretreated	90 (total)	AEs, ORR, DOR	December 2018
NCT02628067 KEYNOTE-158	II	Multicentre; Australia, Brazil, Canada, Colombia, Denmark, France, Germany, Israel, Italy, Japan, Korea, Mexico, Netherlands, Norway, Philippines, Russia, South Africa, Spain, Taiwan, United Kingdom, United States	Pembrolizumab	–	Cohort 3: Unresectable or metastatic grade 1 or 2 lung, appendix, small intestine, colon, rectum, and pancreas NET, pretreated	1350 (total)	ORR	August 2023
NCT02955069	II	Multicentre; Australia, Austria, Belgium, Canada, France, Germany, Israel, Italy, Japan, Netherlands, Spain, Sweden, Switzerland, United Kingdom, United States	PDR001	–	Unresectable or metastatic, progressive, grade 1 or 2 GI, pancreatic, lung or thymic NET (non-functional), and GEP-NEC, pretreated	110 (total)	ORR	December 2019
NCT02834013 DART	II	Multicentre; United States	Nivolumab + Ipilimumab	–	Cohort 23: Neuroendocrine carcinoma including lung NET	707 (total)	ORR	August 2020
NCT03095274 DUNE	II	Single centre; Spain	Durvalumab (MEDI4736) + Tremelimumab	–	Cohort 1: Unresectable or metastatic grade 1 or 2 lung NET, previously treated ≤2 prior systemic antiproliferative agents	126 (total)	CBR	March 2018

NET, neuroendocrine tumour; GI, gastrointestinal; GEP, gastroenteropancreatic; ACs, atypical carcinoids; LCNEC, large cell neuroendocrine carcinoma; SCLC, small cell lung carcinoma; LAR, long-acting repeatable; AE, adverse event; SAE, serious adverse event; PFS, progression-free survival; PFR, progression-free rate; ORR, objective response rate; DOR, duration of response; DCR, disease control rate; CBR, clinical benefit rate.

Table 6.4.3 Table summarizing Kaplan–Meier estimates of overall survival in lung neuroendocrine tumours (NETs) according to stage at diagnosis

Stage at diagnosis	Median overall survival	5-year overall survival
Stage I	*Not reached*	86%
Stage II	*Not reached*	69%
Stage III	3.4 years	45%
Stage IV	5.1 years	56%

Data sourced from Filosso PL, Ferolla P, Guerrera F, Ruffini E, Travis WD, Rossi G, *et al*. Multidisciplinary management of advanced lung neuroendocrine tumors. *J Thorac Dis*. 2015;7(Suppl 2):S163–71.

REFERENCES

1. Modlin IM, Lye KD, Kidd M. A 5-decade analysis of 13,715 carcinoid tumors. *Cancer*, 2003; **97**: 934–59.
2. Rekhtman N. Neuroendocrine tumors of the lung: an update. *Arch Pathol Lab Med*, 2010; **134**: 1628–38.
3. Bodei L, Ćwikla JB, Kidd M, Modlin IM. The role of peptide receptor radionuclide therapy in advanced/metastatic thoracic neuroendocrine tumors. *J Thorac Dis*, 2017; **9**(Suppl 15): S1511–S23.
4. Skuladottir H, Hirsch FR, Hansen HH, Olsen JH. Pulmonary neuroendocrine tumors: incidence and prognosis of histological subtypes. A population-based study in Denmark. *Lung Cancer*, 2002; **37**: 127–35.
5. Naalsund A, Rostad H, Strøm EH, Lund MB, Strand TE. Carcinoid lung tumors--incidence, treatment and outcomes: a population-based study. *Eur J Cardiothorac Surg*, 2011; **39**: 565–9.
6. Travis WD, Brambilla E, Noguchi M, *et al*. International association for the study of lung cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*, 2011; **6**: 244–85.
7. Yao JC, Hassan M, Phan A, *et al*. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, 2008; **26**: 3063–72.
8. Dasari A, Shen C, Halperin D, *et al*. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*, 2017; **3**: 1335–42.
9. Öberg K, Hellman P, Ferolla P, Papotti M, Group EGW. Neuroendocrine bronchial and thymic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 2012; 23 Suppl 7: vii120–3.
10. Caplin ME, Baudin E, Ferolla P, *et al*. Pulmonary neuroendocrine (carcinoid) tumors: European Neuroendocrine Tumor Society expert consensus and recommendations for best practice for typical and atypical pulmonary carcinoids. *Ann Oncol*, 2015; **26**: 1604–20.
11. Broaddus RR, Herzog CE, Hicks MJ. Neuroendocrine tumors (carcinoid and neuroendocrine carcinoma) presenting at extra-appendiceal sites in childhood and adolescence. *Arch Pathol Lab Med*, 2003; **127**: 1200–3.
12. Dishop MK, Kuruvilla S. Primary and metastatic lung tumors in the pediatric population: a review and 25-year experience at a large children's hospital. *Arch Pathol Lab Med*, 2008; **132**: 1079–103.
13. Lal DR, Clark I, Shalkow J, Downey RJ, Shorter NA, Klimstra DS, *et al*. Primary epithelial lung malignancies in the pediatric population. *Pediatr Blood Cancer*, 2005; **45**: 683–6.
14. Rizzardi G, Marulli G, Calabrese F, *et al*. Bronchial carcinoid tumours in children: surgical treatment and outcome in a single institution. *Eur J Pediatr Surg*, 2009; **19**: 228–31.
15. Yu DC, Grabowski MJ, Kozakewich HP, *et al*. Primary lung tumors in children and adolescents: a 90-year experience. *J Pediatr Surg*, 2010; **45**: 1090–5.
16. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*, 4th edn. Lyon: IARC, 2015.
17. Warburton D, Schwarz M, Tefft D, Flores-Delgado G, Anderson KD, Cardoso WV. The molecular basis of lung morphogenesis. *Mech Dev*, 2000; **92**: 55–81.
18. Linnoila RI. Functional facets of the pulmonary neuroendocrine system. *Lab Invest*, 2006; **86**: 425–44.
19. Li Y, Linnoila RI. Multidirectional differentiation of Achaete-Scute homologue-1-defined progenitors in lung development and injury repair. *Am J Respir Cell Mol Biol*, 2012; **47**: 768–75.
20. Gosney JR, Sissons MC, Allibone RO. Neuroendocrine cell populations in normal human lungs: a quantitative study. *Thorax*, 1988; **43**: 878–82.
21. Aguayo SM, Schuyler WE, Murtagh JJ, Roman J. Regulation of lung branching morphogenesis by bombesin-like peptides and neutral endopeptidase. *Am J Respir Cell Mol Biol*, 1994; **10**: 635–42.
22. Sunday ME. Pulmonary neuroendocrine cells and lung development. *Endocr Pathol*, 1996; **7**: 173–201.
23. Cutz E, Yeger H, Pan J. Pulmonary neuroendocrine cell system in pediatric lung disease-recent advances. *Pediatr Dev Pathol*, 2007; **10**: 419–35.
24. Gould VE, Linnoila RI, Memoli VA, Warren WH. Neuroendocrine components of the bronchopulmonary tract: hyperplasias, dysplasias, and neoplasms. *Lab Invest*, 1983; **49**: 519–37.
25. Noguchi M, Sumiyama K, Morimoto M. Directed migration of pulmonary neuroendocrine cells toward airway branches organizes the stereotypic location of neuroepithelial bodies. *Cell Rep*, 2015; **13**: 2679–86.
26. Aguayo SM, Miller YE, Waldron JA, *et al*. Brief report: idiopathic diffuse hyperplasia of pulmonary neuroendocrine cells and airways disease. *N Engl J Med*, 1992; **327**: 1285–8.
27. Gosney JR. Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia as a precursor to pulmonary neuroendocrine tumors. *Chest*, 2004; **125**(5 Suppl): 108S.
28. Lantuéjoul S, Salameire D, Salon C, Brambilla E. Pulmonary preneoplasia--sequential molecular carcinogenetic events. *Histopathology*, 2009; **54**: 43–54.
29. Ferolla P, Daddi N, Urbani M, *et al*. Tumorlets, multicentric carcinoids, lymph-nodal metastases, and long-term behavior in bronchial carcinoids. *J Thorac Oncol*, 2009; **4**: 383–7.
30. Miller RR, Müller NL. Neuroendocrine cell hyperplasia and obliterative bronchiolitis in patients with peripheral carcinoid tumors. *Am J Surg Pathol*, 1995; **19**: 653–8.
31. Rizvi SM, Goodwill J, Lim E, *et al*. The frequency of neuroendocrine cell hyperplasia in patients with pulmonary neuroendocrine tumours and non-neuroendocrine cell carcinomas. *Histopathology*, 2009; **55**: 332–7.
32. Nassar AA, Jaroszewski DE, Helmers RA, Colby TV, Patel BM, Mookadam F. Diffuse idiopathic pulmonary neuroendocrine cell

hyperplasia: a systematic overview. *Am J Respir Crit Care Med*, 2011; **184**: 8–16.
33. IARC. *WHO Classification of Tumours of Endocrine Organs*, Vol. 10, 4th edn. Lloyd RV OR, Klöppel G, Rosai J, eds., 2017.
34. Soga J, Yakuwa Y. Bronchopulmonary carcinoids: an analysis of 1,875 reported cases with special reference to a comparison between typical carcinoids and atypical varieties. *Ann Thorac Cardiovasc Surg*, 1999; **5**: 211–19.
35. Rindi G, Klersy C, Inzani F, *et al*. Grading the neuroendocrine tumors of the lung: an evidence-based proposal. *Endocr Relat Cancer*, 2014; **21**: 1–16.
36. Skov BG, Holm B, Erreboe A, Skov T, Mellemgaard A. ERCC1 and Ki67 in small cell lung carcinoma and other neuroendocrine tumors of the lung: distribution and impact on survival. *J Thorac Oncol*, 2010; **5**: 453–9.
37. Walts AE, Ines D, Marchevsky AM. Limited role of Ki-67 proliferative index in predicting overall short-term survival in patients with typical and atypical pulmonary carcinoid tumors. *Mod Pathol*, 2012; **25**: 1258–64.
38. Simbolo M, Mafficini A, Sikora KO, *et al*. Lung neuroendocrine tumours: deep sequencing of the four World Health Organization histotypes reveals chromatin-remodelling genes as major players and a prognostic role for TERT, RB1, MEN1 and KMT2D. *J Pathol*, 2017; **241**: 488–500.
39. Marty-Ané CH, Costes V, Pujol JL, Alauzen M, Baldet P, Mary H. Carcinoid tumors of the lung: do atypical features require aggressive management? *Ann Thorac Surg*, 1995; **59**: 78–83.
40. Singh S, Granberg D, Wolin E, *et al*. Patient-reported burden of a neuroendocrine tumor (NET) diagnosis: results from the first global survey of patients with NETs. J Glob Oncol, 2017; 3(1): 43–53.
41. Granberg D, Sissons M, Kolarova T, Goldstein G, Leyden J. Lung neuroendocrine tumor (NET) patient (pt)-reported experience: results from the first global NET pt survey—a collaboration between the international neuroendocrine cancer alliance (INCA) and Novartis pharmaceuticals. *J Clin Oncol*, 2015; **33**: 15_suppl, e17739–e17739.
42. Modlin IM, Gustafsson BI, Moss SF, Pavel M, Tsolakis AV, Kidd M. Chromogranin A—biological function and clinical utility in neuro endocrine tumor disease. *Ann Surg Oncol*, 2010; **17**: 2427–43.
43. Baudin E, Gigliotti A, Ducreux M, *et al*. Neuron-specific enolase and chromogranin A as markers of neuroendocrine tumours. *Br J Cancer*, 1998; **78**: 1102–7.
44. Jeung MY, Gasser B, Gangi A, *et al*. Bronchial carcinoid tumors of the thorax: spectrum of radiologic findings. *Radiographics*, 2002; **22**: 351–65.
45. Lamarca A, Pritchard DM, Westwood T, *et al*. 68Gallium DOTANOC-PET imaging in lung carcinoids: impact on patients' management. *Neuroendocrinology*, 2018; **106**: 128–38.
46. Kayani I, Conry BG, Groves AM, *et al*. A comparison of 68Ga-DOTATATE and 18F-FDG PET/CT in pulmonary neuroendocrine tumors. *J Nucl Med*, 2009; **50**: 1927–32.
47. Abgral R, Leboulleux S, Déandreis D, *et al*. Performance of (18)fluorodeoxyglucose-positron emission tomography and somatostatin receptor scintigraphy for high Ki67 (≥10%) well-differentiated endocrine carcinoma staging. *J Clin Endocrinol Metab*, 2011; **96**: 665–71.
48. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. *Chest*, 2017; **151**: 193–203.
49. Filosso PL, Ferolla P, Guerrera F, *et al*. Multidisciplinary management of advanced lung neuroendocrine tumors. *J Thorac Dis*, 2015; **7**(Suppl 2): S163–71.
50. Goldstraw P. *International Association for the Study of Lung Cancer Staging Manual in Thoracic Oncology*. North Fort Myers, FL: Editorial Rx Press, 2009.
51. Pavel M, O'Toole D, Costa F, *et al*. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. *Neuroendocrinology*, 2016; **103**: 172–85.
52. Detterbeck FC. Management of carcinoid tumors. *Ann Thorac Surg*, 2010; **89**: 998–1005.
53. Lou F, Sarkaria I, Pietanza C, *et al*. Recurrence of pulmonary carcinoid tumors after resection: implications for postoperative surveillance. *Ann Thorac Surg*, 2013; **96**: 1156–62.
54. Filosso PL, Rena O, Donati G, *et al*. Bronchial carcinoid tumors: surgical management and long-term outcome. *J Thorac Cardiovasc Surg*, 2002; **123**: 303–9.
55. Nussbaum DP, Speicher PJ, Gulack BC, *et al*. Defining the role of adjuvant chemotherapy after lobectomy for typical bronchopulmonary carcinoid tumors. *Ann Thorac Surg*, 2015; **99**: 428–34.
56. Anderson KL, Mulvihill MS, Speicher PJ, *et al*. Adjuvant chemotherapy does not confer superior survival in patients with atypical carcinoid tumors. *Ann Thorac Surg*, 2017; **104**: 1221–30.
57. Phan AT, Öberg K, Choi J, *et al*. NANETS consensus guideline for the diagnosis and management of neuroendocrine tumors: well-differentiated neuroendocrine tumors of the thorax (includes lung and thymus). *Pancreas*, 2010; **39**: 784–98.
58. NCCN. NCCN clinical practice guidelines in oncology: neuroendocrine tumors of the GI, lung and thymus (carcinoid tumors), 2016. Available at: https://www.nccn.org/professionals/physician_gls/default.aspx
59. Rinke A, Müller HH, Schade-Brittinger C, *et al*. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*, 2009; **27**: 4656–63.
60. Caplin ME, Pavel M, Ruszniewski P. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*, 2014; **371**: 1556–7.
61. Yao JC, Fazio N, Singh S, *et al*. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2016; **387**: 968–77.
62. Phan AT, Dave B. The pivotal role of mammalian target of rapamycin inhibition in the treatment of patients with neuroendocrine tumors. *Cancer Med*, 2016; **5**: 2953–64.
63. Pavel ME, Hainsworth JD, Baudin E, *et al*. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2011; **378**: 2005–12.
64. Pavel ME, Baudin E, Öberg KE, *et al*. Efficacy of everolimus plus octreotide LAR in patients with advanced neuroendocrine tumor and carcinoid syndrome: final overall survival from the randomized, placebo-controlled phase 3 RADIANT-2 study. *Ann Oncol*, 2017; **28**: 1569–75.

65. Sun W, Lipsitz S, Catalano P, Mailliard JA, Haller DG, Group ECO. Phase II/III study of doxorubicin with fluorouracil compared with streptozocin with fluorouracil or dacarbazine in the treatment of advanced carcinoid tumors: Eastern Cooperative Oncology Group Study E1281. *J Clin Oncol*, 2005; **23**: 4897–904.
66. Turner NC, Strauss SJ, Sarker D, *et al*. Chemotherapy with 5-fluorouracil, cisplatin and streptozocin for neuroendocrine tumours. *Br J Cancer*, 2010; **102**: 1106–12.
67. Wong MH, Chan DL, Lee A, *et al*. Systematic review and meta-analysis on the role of chemotherapy in advanced and metastatic neuroendocrine tumor (NET). *PLoS One*, 2016; **11**: e0158140.
68. Koumarianou A, Kaltsas G, Kulke MH, *et al*. Temozolomide in advanced neuroendocrine neoplasms: pharmacological and clinical aspects. *Neuroendocrinology*, 2015; **101**: 274–88.
69. Mariniello A, Bodei L, Tinelli C, *et al*. Long-term results of PRRT in advanced bronchopulmonary carcinoid. *Eur J Nucl Med Mol Imaging*, 2016; **43**: 441–52.
70. Talbot D. Thoracic neuroendocrine tumours. ESO-ESMO-RCE clinical update on rare adult solid cancers. Milan, Italy, 2017.
71. Ferolla P, Brizzi MP, Meyer T, *et al*. Efficacy and safety of long-acting pasireotide or everolimus alone or in combination in patients with advanced carcinoids of the lung and thymus (LUNA): an open-label, multicentre, randomised, phase 2 trial. *Lancet Oncol*, 2017; **18**: 1652–64.
72. Fisseler-Eckhoff A, Demes M. Neuroendocrine tumors of the lung. *Cancers (Basel)*, 2012; **4**: 777–98.
73. Rea F, Rizzardi G, Zuin A, *et al*. Outcome and surgical strategy in bronchial carcinoid tumors: single institution experience with 252 patients. *Eur J Cardiothorac Surg*, 2007; **31**: 186–91.

6.5

Non-Functioning Pancreatic Neuroendocrine Tumours

Kok Haw Jonathan Lim, Juan W. Valle, and Wasat Mansoor

Epidemiology

Pancreatic neuroendocrine tumours (panNET) are a heterogeneous group of NETs which account for 3–5% of all NETs (see Figure 6.4.1a in Chapter 6.4) and 18% of gastroenteropancreatic (GEP) NETs [1, 2]. They constitute approximately 3% of all neoplasms in the pancreas, making panNET the second commonest after exocrine pancreatic adenocarcinoma (~90%) [3]. Like all NETs, the incidence of panNET has continued to rise with improved awareness and diagnostic tools, gradually increasing from 0.18 to 0.36 per 100 000 from 1973 to 2003 [1], but has been exponentially increasing over the past decade with its incidence up to 0.86 per 100 000 in 2012 (see Figure 6.4.1b in Chapter 6.4) [2]. The peak incidence is in the sixth to seventh decades of life and there is a slight preponderance for males over females [4]. Pancreatic NETs appear to predominantly affect the white population (84%), with a much lower reported incidence among Blacks (8%), Asians (5%), and others [4].

In health, the endocrine cells of the pancreas produce hormones including insulin, gastrin, glucagon, vasoactive intestinal peptide (VIP), and somatostatin. Pancreatic NETs can, therefore, take on this phenotype and are broadly categorized into functioning (F-panNET) or non-functioning (NF-panNET), by virtue of the presence of syndromic features secondary to excess secretion of any of these hormones [5]. Non-functioning pancreatic NETs form 60–90% of all panNETs, and most patients (>80%) present with advanced or metastatic disease due to the asymptomatic nature of this subgroup of panNET [4, 6, 7]. Within the remit of this chapter, only NF-panNET will be discussed and the other specific subtypes of F-panNET (gastrinoma, insulinoma, glucagonoma, VIPoma, somatostatinoma) will be covered in detail in subsequent chapters.

Histological Classification and Staging

The endocrine pancreas forms only 1–2% of volume of the adult pancreas, and is made up of clusters of cells called islets of Langerhans [8]. These pancreatic islets are formed of specialized cells which secrete hormones into the bloodstream through tightly regulated feedback loops: alpha (α) cells (glucagon), beta (β) cells (insulin), pancreatic polypeptide (PP) cells (pancreatic polypeptide and somatostatin), and delta (δ) cells which regulate other pancreatic cells. The pathophysiology and abnormal proliferation of these cells which eventually lead to the formation panNET remains poorly understood [9].

While the majority of panNETs (60%) have unknown aetiology or are sporadic, about 10-30% are diagnosed as part of an inherited syndrome [7]. Of these, multiple endocrine neoplasia type 1 (MEN-1) is the most common whereby it has been reported that 20-80% of patients with MEN-1 (see Chapter 6.11.2) will have panNETs, a large majority of whom were NF-panNETs [10]. About 10–17% of patients with von Hippel–Lindau (VHL) disease may have panNETs, almost all of which will be NF-panNETs (>98%) [10]. Meanwhile, panNETs are generally very rare in patients with Recklinghausen's syndrome (neurofibromatosis type 1, NF1), or tuberous sclerosis [7, 10].

Recently, Scarpa *et al.* [11] performed whole-genome sequencing on 102 primary panNETs, and described four 'core pathways' that may be implicated in the pathogenesis of panNETs. These may be categorized into somatic mutations involving the following cellular processes: chromatin remodelling (inactivation of *MEN1, SETD2, ARID1A* and *MLL3*); DNA damage repair (*MUTYH,*

CHEK2, and *BRCA2*); activation of mTOR signalling (inactivation of *PTEN, TSC2, TSC1*, and *DEPDC5*); and telomerase maintenance (upregulation of *TERT*, and inactivation of *DAXX* or *ATRX*) (Figure 6.5.1). Intriguingly, it is also known that overall panNET has more than three times less mutational burden compared to pancreatic ductal adenocarcinoma (mean 0.82 vs. 2.64 per megabase) [11, 12]; thus highlighting the very different nature of these entities.

Irrespective of whether panNETs are functioning or non-functioning, most demonstrate immunoreactivity to synaptophysin and chromogranin-A on histopathological analysis [13, 14]. Other non-specific neuroendocrine markers, for example Neuron-specific enolase (NSE) and CD56, may also be positive [15]. Additional pancreas-specific staining for ISl1 and PAX8 may be useful in cases where the primary cannot be easily determined at outset [16, 17]. Unlike in lung NETs, the proliferation index Ki-67 has been validated for use and should be reported as standard of care as it is a vital component of the classification criteria and also serves as a significant prognostic factor of the natural history of the patient's disease [7, 18].

Over the years, there have been several systems used to classify panNETs, and this has been progressively refined with improved understanding of the disease biology and physicians' growing clinical experience. There are currently two widely used systems as per the most recent ENETS consensus statement [7]. The main one is the recently updated World Health Organization (WHO) classification of panNETs which, as previously, recognizes clear morphological distinction between well- and poorly differentiated malignancies and also makes an assessment of grade based on the Ki-67 (Table 6.5.1) [19]. Well-differentiated grade 1 and 2 panNETs are distinguished by a mitotic count of <2 per 10× HPF, and Ki-67 of <3% versus mitotic count of 2–20% and Ki-67 of 3–20%, respectively [19]. Most patients with grade 3 neoplasms have poorly differentiated neuroendocrine carcinomas. However, the observation that some neoplasms with a Ki-67 of >20% also have a well-differentiated morphology has led to the addition of a well-differentiated grade 3 subgroup to include these patients [19]. A relatively rarer subgroup composed of a combination of both exocrine and endocrine components has now been renamed mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN), previously known as mixed adeno-neuroendocrine carcinoma (MANEC) [19, 20], in recognition that the non-neuroendocrine carcinomatous component can be other than adenocarcinoma (e.g. squamous cell carcinoma). The ENETS TNM classification is also widely used in parallel to the WHO classification (Table 6.5.2) [21].

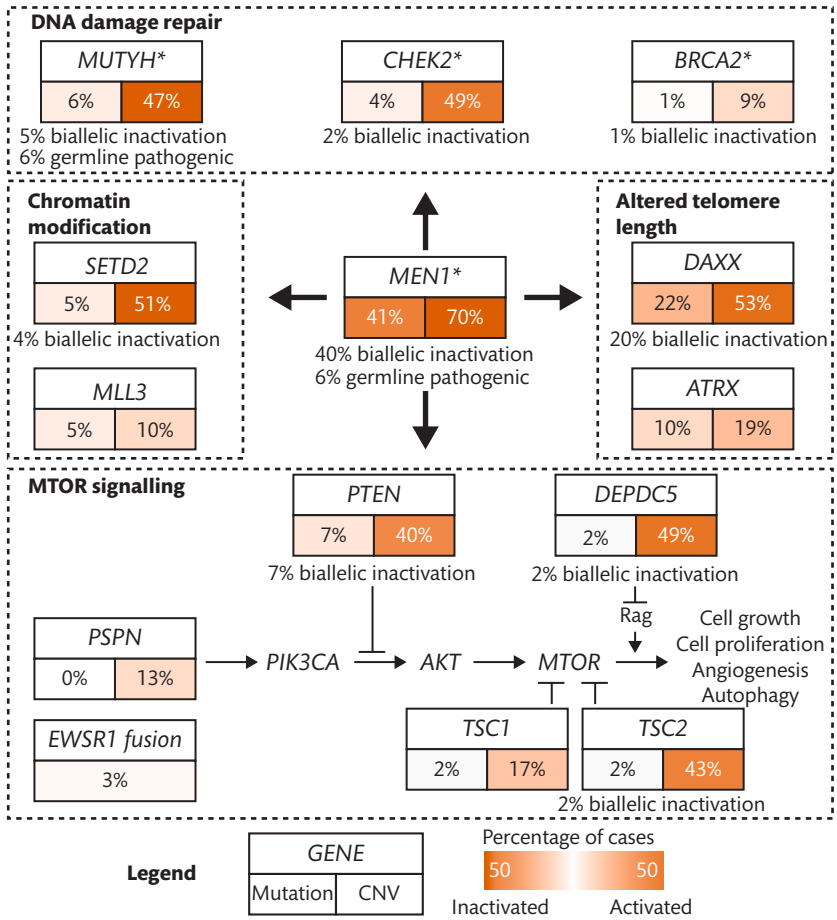


Figure 6.5.1 Core pathways in the pathogenesis of panNETs.

Reproduced with permission from Scarpa A, Chang DK, Nones K, Corbo V, Patch AM, Bailey P, *et al*. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*. 2017;543(7643):65–71. Copyright © 2017, Springer Nature.

Table 6.5.1 World Health Organization (WHO) 2017 classification of pancreatic neuroendocrine tumours (panNETs)

Morphology	Grade	Mitotic count (per 10x HPF)	Ki-67 index	Nomenclature
Well-differentiated	1	<2	<3%	G1-2 neuroendocrine tumour (NET)
	2	2 – 20	3 – 20%	
	3	>20	>20%	G3 neuroendocrine tumour (NET)
Poorly differentiated	3	>20	>20%	G3 neuroendocrine carcinoma (NEC)
Mixed neuroendocrine non-neuroendocrine neoplasm	A combination of malignant exocrine and neuroendocrine components, each of which accounts for at least 30% of the tumour mass			MiNEN*

*previously known as MANEC – mixed adeno-neuroendocrine carcinoma [20].

Reproduced with permission from Lloyd R, Osamura R, Klöppel G, Rosai J. *WHO Classification of Tumours of Endocrine Organs*. 4th ed. Lyon: IARC WHO Classification of Tumours; 2017.

Table 6.5.2 European Neuroendocrine Tumour Society (ENETS) TNM Staging for pancreatic neuroendocrine tumours (panNETs)

Tumour (T)	
Tx	Statement concerning the primary tumour not possible
T0	Primary tumour not verifiable
Tis	Carcinoma in situ
T1	Limited to the pancreas, ≤2 cm
T2	Limited to the pancreas, >2 cm – ≤4 cm
T3	>4 cm or with invasion of duodenum and common bile duct, limited to the pancreas
T4	Infiltration of proximal organs (stomach, spleen, colon, adrenal gland) or vascular invasion (coeliac trunk, superior mesenteric artery)

Node (N)	
Nx	Statement concerning the regional lymph node metastases not possible
N0	Regional lymph node metastases not verifiable
N1	Regional lymph node metastases

Metastases (M)	
Mx	Statement concerning the distant metastases not possible
M0	No distant metastases
M1	Distant metastases

Stage	
I	T1, N0, M0
IIa	T2, N0, M0
IIb	T3, N0, M0
IIIa	T4, N0, M0
IIIb	Any T, N1, M0
IV	Any T, any N, M1

Reproduced from Rindi, G., Klöppel, G., Alhman, H., *et al*., TNM staging of foregut (neuro)endocrine tumors: a consensus proposal including a grading system, *Virchows Archiv*, Vo. 449, No 4, pp. 395–401 with permission from Springer Nature under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)

Clinical Features

A majority of patients demonstrate no signs or symptoms of their disease and their panNET is found incidentally. This is partly attributed to the fact that about two-thirds of primary panNETs are located in the body and tail of the pancreas [22, 23], thereby unlikely to present with obstructive jaundice. As NF-panNET patients are also not syndromic, patients often present late with advanced disease with constitutional symptoms such as anorexia, weight loss, abdominal pain, and nausea. Fewer patients, particularly those with their primary lesions in the head of pancreas, may present with symptoms of mass-effect including jaundice (often painless) and an abdominal mass.

In patients with a strong family history and suspicion of an inherited syndrome, other symptoms and signs associated with MEN-1 and VHL disease as already described should be taken into consideration.

Diagnosis

Following thorough history taking and a physical examination, a standard panel of blood tests including full/complete blood count, biochemistry, liver function tests, clotting screen, and specifically urinary/serum chromogranin-A, 5-hydroxyindoleacetic acid (5HIAA), and pancreatic polypeptide (PP) are recommended [5, 7]. While these neuroendocrine biomarkers may not be necessarily elevated, if they indeed are, they can be utilized to monitor response to treatment or disease progression [7].

In order to identify and characterize the primary tumour better and also the disease distribution, cross-sectional imaging is necessary. The standard of care should be a contrast-enhanced full staging computed tomography (CT) scan of the thorax, abdomen and pelvis; in addition the gold-standard functional imaging modality with the best sensitivity and specificity for panNETs (especially in NF-panNETs), like in lung NETs, is a ^{68}Gallium (^{68}Ga)-DOTA (DOTATOC or DOTATATE)-PET CT scan [7]. This ^{68}Ga-DOTA-PET CT scan is recommended as a baseline imaging for all patients and is now preferred to conventional somatostatin-receptor scintigraphy (SRS) with single-photon emission computed tomography (SPECT) or

OctreoScan [7]. An MRI scan of the liver and/or pancreas may be useful if closer examination of the anatomical distribution of the disease is required, for instance for the purpose of surgical planning.

Once a primary lesion is uncovered, a definitive diagnosis can only be made with histological confirmation of panNET. In early-stage disease, endoscopic ultrasound (EUS) has been shown in many studies to improve diagnostic confidence and further characterize the extent of primary tumour in relation to its surrounding structures, and is also a useful diagnostic tool to obtain biopsy samples from the pancreas [24, 25]. In metastatic disease, biopsy samples may be more easily accessible from the liver, for example.

In view of the heterogeneity of panNETs and the various treatment options available, all patients and the results of investigations performed should be discussed in a specialized multidisciplinary team (MDT) meeting in a centre experienced in treatment of NETs [7].

Management

There are several important prognostic factors to be considered when deciding on the management plan for a patient with panNET including the location of the primary tumour, extent of disease, tumour stage, degree of proliferation/proliferation index, tumour grade, patient's age and performance status [1]. The only curative-intent management option for NF-panNETs is surgery [7]. This is strongly indicated particularly in early disease when complete staging reveals no evidence of distant metastasis (M0 disease), and the primary tumour is more than 2 cm and technically resectable [7].

The preferred oncologic resection choice is based on the anatomical location of the tumour, and is usually in the form of an enucleation (particularly in the setting of MEN-1-related panNET, as other lesions are likely to develop over time and sparing of the pancreatic parenchyma is preferred in these, often young, patients), distal pancreatic resection with or without splenectomy, or in some selected cases pancreaticoduodenal resection or Whipple's surgery [15]. In all cases, it is important to achieve sufficient lymph node clearance, particularly when enucleation is performed for NF-panNET [15, 26]. It is the consensus among all the ESMO, ENETS, and NANETS guidelines that laparoscopic procedures instead of laparotomy are not recommended due to the need for adequate lymphadenectomy and complete surgical exploration intra-abdominally [7, 15, 26].

The recommended management for small primary NF-panNET ≤2 cm remains an area of contention. In a large retrospective Italian study, outcomes were categorically compared by their tumour sizes in 177 patients with NF-panNET who underwent R0 resection: 90 patients (51%) ≤2 cm, 46 patients (26%) >2 and ≤4 cm, and 41 patients (23%) >4 cm [27]. In their multivariable analysis, a NF-panNET of >2 cm was an independent predictor of malignancy, and this study proposed that incidental NF-panNETs ≤2 cm can be conservatively managed as they had excellent long-term outcomes [27]. In contrast, in a more recently published retrospective study of 1854 patients with all NF-panNETs ≤2 cm from the US National Cancer Database (1998–2011), the 5-year overall survival was significantly inferior among those who did not have surgery (28% vs. 72–86%, P <0.01), and outcomes were better for those who had surgery irrespective of the type of surgery performed or adequacy of lymphadenectomy [28]. In the absence of a prospective study, a conservative approach is considered safe and has been recommended in the ENETS consensus guidelines [7]. A systematic review has now confirmed the safety of watchful waiting in those with small (≤2 cm), asymptomatic, sporadic NF-panNETs [29]. There is now also a large prospective study of Asymptomatic Small Pancreatic Endocrine Neoplasms (ASPEN) (NCT03084770) currently open for recruitment which will provide further concrete guidance on the management of this subgroup.

In patients with locally advanced NF-panNET (grades 1 and 2) and those with liver-only metastases, surgery with curative intent may still be pursued as long as this is technically possible [30, 31]. Interventional radiological options may be considered as bridging or downstaging approaches with a view of surgery later, and these include radiofrequency ablation (RFA), laser-induced thermotherapy (LiTT), transarterial chemoembolization (TACE), and selective internal radiation therapy (SIRT) (**Figure 6.5.2**) [31]. Most surgeons will now also consider cytoreductive surgery for metastatic disease when more than 70% of patient's total tumour bulk can be excised [15, 32, 33]. In carefully selected cases, liver transplantation may be considered but these are few and far in between as most patients will still relapse with metastatic disease [34, 35]. Therefore, liver transplantation for panNET is still not standard of care in most countries in Europe [31].

Generally, following radical surgery, there is currently no clear evidence for adjuvant systemic treatments for grades 1 and 2 NF-panNETs. However, in patients with G3 NF-panNET or panNEC with a known high risk of relapse despite surgery, adjuvant chemotherapy may be considered [36]. The most commonly recommended regimen is the combination of cisplatin or carboplatin with etoposide [36].

In most NF-panNET patients who present with metastatic disease at outset, it is important to consider the patients' condition carefully. It is not unreasonable to take the watchful waiting approach in patients with grades 1 or 2 NF-panNET who are relatively asymptomatic and has slow-growing disease if surgery is not considered feasible or is contraindicated for other reasons [30]. Alternatively, antiproliferative agents in the form of somatostatin analogues have been shown to be effective in prolonging progression-free survival in NETs. In the Phase III double-blind placebo-control PROMID trial, 42 patients (n = 42/85, 49%) with non-functioning midgut NETs who were randomized to receive octreotide LAR 30 mg given intramuscularly every 28 days had median time-to-progression of 28.8 months compared to 5.9 months in the placebo arm, HR 0.25 (95% CI 0.10-0.59), P = 0.0008 [37]. This trial allowed crossover, and 88% eventually went on to receive octreotide [37]. There was no statistically significant difference in the octreotide and placebo arm after long-term follow-up [37]. Note that PROMID did not include any patients with panNETs. Meanwhile in the Phase III double-blind placebo-control CLARINET trial, 101 patients (n = 101/204, 50%) who received lanreotide autogel 120 mg subcutaneously also had superior progression-free survival of 38.5 months versus 18 months, P <0.001 [38, 39]. It is important to note that both trials included slightly different cohorts of patients, which may account for the stark difference in outcomes. In PROMID, 98% had grade 1 intestinal NET, while CLARINET had a Ki-67 cut-off of <10% and included grades 1 and 2 midgut and hindgut NETs, with 96% of patients having no tumour progression 3–6 months prior to randomization [37, 38]. In CLARINET, there were 91 NF-panNET

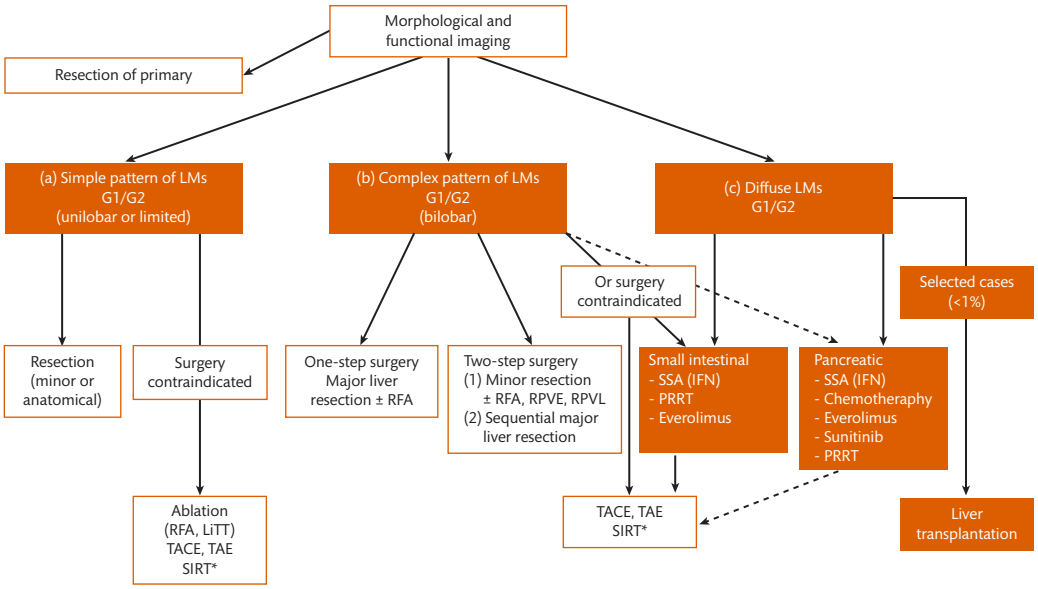


Figure 6.5.2 ENETS consensus guidelines/algorithm for the management of liver metastases in G1/G2 neuroendocrine tumours (NETs) in those without extrahepatic disease.

patients included (n = 91/204, 45%), with 42 and 49 patients receiving lanreotide LAR versus placebo, respectively [38, 39]. In the exploratory subgroup analysis, these patients derived a similar magnitude of benefit compared to other GI NETs, affirming the utility of somatostatin analogues, lanreotide in particular, in panNETs.

Chemotherapy remains the mainstay of systemic treatment for patients with high volume and/or symptomatic and/or rapidly growing (in ≤6–12 months) grade 2 NF-panNET and grade 3 NF-panNET or panNEC (15, 31). First-line standard cytotoxic option for metastatic grade 2 or 3 NF-panNETs is the combination of streptozotocin and 5-fluorouracil (5-FU) or doxorubicin with a 35–40% response rate [15, 31, 40–42]. Alternative chemotherapy options include temozolomide +/– capecitabine, and oxaliplatin-based chemotherapy + 5-FU or capecitabine with retrospective studies suggesting a response rate of 40–70% [15, 31, 43, 44]. Meanwhile in grade 3 NF-panNEC, carboplatin and etoposide combination regime is now standard first-line, with estimated response rate between 30 and 70%, conferring a median progression-free survival of 5 months and median overall survival of 15 months with treatment [31, 45]. FOLFOX (containing folinic acid/leucovorin, 5-FU and oxaliplatin; also known as oxaliplatin and modified de Gramont, OxMdG) and FOLFIRI (containing folinic acid/leucovorin, 5-FU, and irinotecan; also known as irinotecan and modified de Gramont, IrMdG) are reasonable second-line options following failure to carboplatin and etoposide in G3 NF-panNEC [31, 46, 47].

Targeted therapies have also shown to be efficacious in NF-panNETs. In the Phase III RADIANT-3 randomized placebo-control trial, 410 patients with radiologically proven progressive well- to moderately differentiated (grade 1 or 2) locally advanced or metastatic panNETs were randomized to continuous everolimus 10 mg orally once a day or placebo [48]. The 207 patients (n = 207/410; 50%) receiving everolimus had a median progression-free survival of 11.0 months compared to 4.6 months in the placebo arm, HR 0.35 (95% CI 0.27–0.45), P <0.001 [48]. Everolimus was well-tolerated with anaemia (6% vs. 0%) and hyperglycaemia (5% vs. 2%) being the main reported grade 3 or 4 adverse events (48). Similarly, there was a Phase III randomized placebo-control trial of sunitinib in 171 patients with radiologically proven progressive well-differentiated (grade 1 or 2) advanced panNET [49]. Sunitinib conferred a gain in median progression-free survival of 11.4 months compared to 5.5 months in the placebo arm, HR 0.42 (95% CI 0.26–0.66), P <0.001 [49]. The main reported grade 3 or 4 toxicities of sunitinib are neutropenia (12% vs. 0%), hypertension (10% vs. 1%), palmar-plantar erythrodysesthesia (6% vs. 0%) and stomatitis (4% vs. 0%) [49]. Currently, both everolimus and sunitinib are widely available and recommended after progression to either somatostatin analogue and/or chemotherapy (**Figure 6.5.3**) [31].

An important question often facing clinicians remains the most effective sequencing or combination of treatments. There is already evidence that targeted therapies remain beneficial for patients irrespective of whether they have received prior chemotherapy or somatostatin analogues [48]. The SEQTOR trial (NCT02246127) is an innovative Phase III study aimed at addressing how to best sequence chemotherapy and targeted therapy, and is currently recruiting grade 1 or 2 advanced progressive panNET patients to receive combination chemotherapy with streptozotocin/5-FU upfront followed by everolimus upon progression, or in the reverse order. There is also evidence from the Phase III RADIANT-2 trial to support the use of everolimus in combination with octreotide patients with high volume and/or symptomatic and/or rapidly growing grades 1 or 2 NF-panNET [50]. More work is now needed to personalize the optimal treatment pathway for each individual patient.

Evidence is emerging in the use of liver-directed therapies in the form of peptide receptor radionuclide therapy (PRRT) in grade 1 and 2 midgut NETs. Preliminary data from the recently published NETTER-1 trial reported that the use of ^{177}Lu-Dotatate PRRT after progression on first-line octreotide was beneficial in patients with

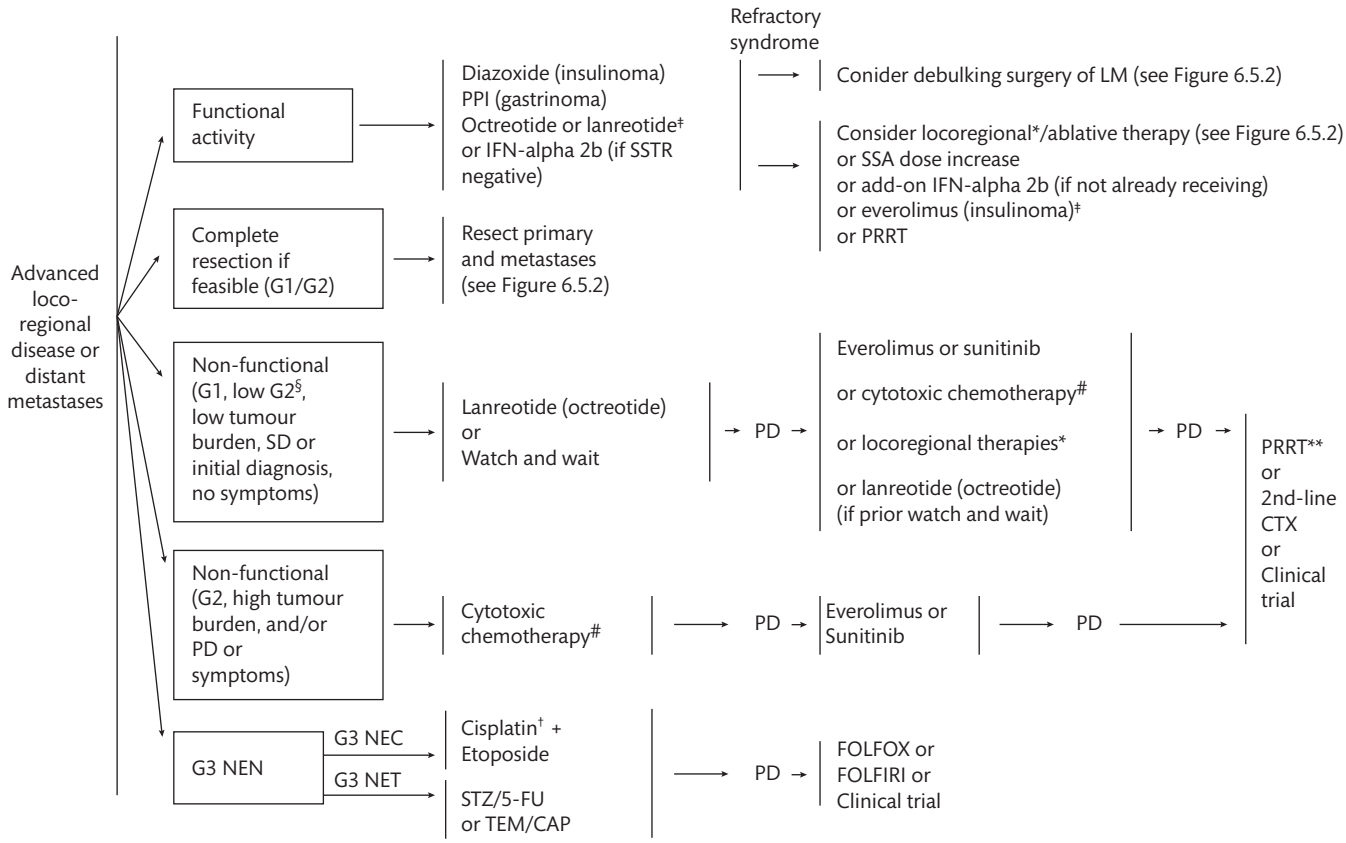


Figure 6.5.3 ENETS consensus guidelines/algorithm for the management of locally advanced or metastatic pancreatic neuroendocrine tumours (panNETs).

somatostatin-receptor positive midgut NETs (no patients with panNET were included) compared to those who received an escalated dose of octreotide [51]. The median progression-free survival in the PRRT arm was not reached after a follow-up of 20 months (65.2%) versus 10.8% in the control group [51]. There appears to be an overall survival advantage, and further mature data are awaited [51]. Non-randomized, retrospective data from a large Dutch cohort, which included a subgroup of 112 panNET patients (133 panNETs in total), demonstrated clinical benefit in 85% of patients (complete response/partial response/stable disease) with a median progression-free survival of 30 months and median overall survival of 69 months [52]. The magnitude of benefit was comparable between patients with panNETs and those with intestinal NETs.

Prognosis

Compared to F-panNET, NF-panNET has been found to have an inferior prognosis, with median overall survival of 26 months (95% CI 23–30) compared to 54 months (95% CI 37–75) [HR 0.71 (95%CI 0.57–0.84), $P = 0.004$] [4]. This is likely to be due to the lead-time bias of patients presenting with earlier disease due to hormone-related symptoms, but may also be potentially attributed to a difference in disease biology.

In general, the 5-year survival rate of panNET is favourable particularly when patients are diagnosed with early disease which may remain indolent for a number of years. In fact, the estimated 5-year survival in well-differentiated localized disease is between 60% and 100%, compared to around 40% for locally advanced or regional disease [6, 15]. Patients who undergo successful metastasectomies may achieve a 5-year overall survival of more than 60%, and if R0 resection is performed, this may even be up to 80% [53]. The outcomes are much worse in those with poorly differentiated tumours (NETs or NECs) with their 5-year survival estimated to be around 12% [6]. With modern treatment options available, it is predicted that more than 25% of those with metastatic panNETs on the whole will be alive at 5 years [15, 31].

Future Work

NF-panNET is a heterogenous group of GEP NETs and the challenge remain in the early disease whereby more prospective studies are needed to identify better predictors of the disease in order to de-escalate treatment in those who may not need it at all and at the same time guide timely treatments for those who require earlier intervention. There also remain questions on the most efficacious sequence and/or combinations of various treatment modalities which are now approved and available for use in treating NF-panNETs, some of which will be hopefully addressed by the several ongoing and upcoming clinical trials. Immunotherapy agents may be promising in well selected groups, although more preclinical data and better biomarkers are necessary to inform more rational use of these drugs. There will be a rapidly expanding use

of various radionuclide-based theranostics which will then need longer term clinical experience to understand if there are any potential late effects of treatment.

REFERENCES

1. Yao JC, Hassan M, Phan A, *et al*. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, 2008; **26**: 3063–72.
2. Dasari A, Shen C, Halperin D, *et al*. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*, 2017; **3**: 1335–42.
3. Fesinmeyer MD, Austin MA, Li CI, De Roos AJ, Bowen DJ. Differences in survival by histologic type of pancreatic cancer. *Cancer Epidemiol Biomarkers Prev*, 2005; **14**: 1766–73.
4. Halfdanarson TR, Rabe KG, Rubin J, Petersen GM. Pancreatic neuroendocrine tumors (PNETs): incidence, prognosis and recent trend toward improved survival. *Ann Oncol*, 2008; **19**: 1727–33.
5. Falconi M, Bartsch DK, Eriksson B, *et al*. ENETS Consensus Guidelines for the management of patients with digestive neuroendocrine neoplasms of the digestive system: well-differentiated pancreatic non-functioning tumors. *Neuroendocrinology*, 2012; **95**: 120–34.
6. Scarpa A, Mantovani W, Capelli P, *et al*. Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. *Mod Pathol*, 2010; **23**: 824–33.
7. Falconi M, Eriksson B, Kaltsas G, *et al*. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
8. Bardeesy N, DePinho RA. Pancreatic cancer biology and genetics. *Nat Rev Cancer*, 2002; **2**: 897–909.
9. Zhang J, Francois R, Iyer R, Seshadri M, Zajac-Kaye M, Hochwald SN. Current understanding of the molecular biology of pancreatic neuroendocrine tumors. *J Natl Cancer Inst*, 2013; **105**: 1005–17.
10. Jensen RT, Berna MJ, Bingham DB, Norton JA. Inherited pancreatic endocrine tumor syndromes: advances in molecular pathogenesis, diagnosis, management, and controversies. *Cancer*, 2008; **113**(7 Suppl): 1807–43.
11. Scarpa A, Chang DK, Nones K, *et al*. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*, 2017; **543**: 65–71.
12. Waddell N, Pajic M, Patch AM, *et al*. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*, 2015; **518**: 495–501.
13. Klöppel G, Couvelard A, Perren A, *et al*. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: towards a standardized approach to the diagnosis of gastroenteropancreatic neuroendocrine tumors and their prognostic stratification. *Neuroendocrinology*, 2009; **90**: 162–6.
14. Klimstra DS, Modlin IR, Coppola D, Lloyd RV, Suster S. The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas*, 2010; **39**: 707–12.
15. Öberg K, Knigge U, Kwekkeboom D, Perren A, Group EGW. Neuroendocrine gastro-entero-pancreatic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*, 2012; **23** Suppl 7: vii124–30.
16. Schmitt AM, Riniker F, Anlauf M, *et al*. Islet 1 (Isl1) expression is a reliable marker for pancreatic endocrine tumors and their metastases. *Am J Surg Pathol*, 2008; **32**: 420–5.
17. Sangoi AR, Ohgami RS, Pai RK, Beck AH, McKenney JK. PAX8 expression reliably distinguishes pancreatic well-differentiated neuroendocrine tumors from ileal and pulmonary well-differentiated neuroendocrine tumors and pancreatic acinar cell carcinoma. *Mod Pathol*, 2011; **24**: 412–24.
18. Panzuto F, Boninsegna L, Fazio N, *et al*. Metastatic and locally advanced pancreatic endocrine carcinomas: analysis of factors associated with disease progression. *J Clin Oncol*, 2011; **29**: 2372–7.
19. Lloyd R, Osamura R, Klöppel G, Rosai J. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. Lyon: IARC WHO Classification of Tumours, 2017.
20. Bosman FT, Carneiro F, Hruban RH, Theise ND. *WHO Classification of Tumours of the Digestive System*, 4th edn. Lyon: IARC, 2010.
21. ENETS TNM Classification. Available at: https://www.enets.org/1578441600045.d.f.426.pdf (accessed: 26 March 2020).
22. Vagefi PA, Razo O, Deshpande V, *et al*. Evolving patterns in the detection and outcomes of pancreatic neuroendocrine neoplasms: the Massachusetts General Hospital experience from 1977 to 2005. *Arch Surg*, 2007; **142**: 347–54.
23. Zerbi A, Falconi M, Rindi G, *et al*. Clinicopathological features of pancreatic endocrine tumors: a prospective multicenter study in Italy of 297 sporadic cases. *Am J Gastroenterol*, 2010; **105**: 1421–9.
24. Khashab MA, Yong E, Lennon AM, *et al*. EUS is still superior to multidetector computerized tomography for detection of pancreatic neuroendocrine tumors. *Gastrointest Endosc*, 2011; **73**: 691–6.
25. James PD, Tsolakis AV, Zhang M, *et al*. Incremental benefit of preoperative EUS for the detection of pancreatic neuroendocrine tumors: a meta-analysis. *Gastrointest Endosc*, 2015; **81**: 848–56.e1.
26. Kulke MH, Anthony LB, Bushnell DL, *et al*. NANETS treatment guidelines: well-differentiated neuroendocrine tumors of the stomach and pancreas. *Pancreas*, 2010; **39**: 735–52.
27. Bettini R, Partelli S, Boninsegna L, *et al*. Tumor size correlates with malignancy in nonfunctioning pancreatic endocrine tumor. *Surgery*, 2011; **150**: 75–82.
28. Gratian L, Pura J, Dinan M, Roman S, Reed S, Sosa JA. Impact of extent of surgery on survival in patients with small non-functional pancreatic neuroendocrine tumors in the United States. *Ann Surg Oncol*, 2014; **21**: 3515–21.
29. Partelli S, Cirocchi R, Crippa S, *et al*. Systematic review of active surveillance versus surgical management of asymptomatic small non-functioning pancreatic neuroendocrine neoplasms. *Br J Surg*, 2017; **104**: 34–41.
30. Pavel M, Baudin E, Couvelard A, *et al*. ENETS Consensus Guidelines for the management of patients with liver and other distant metastases from neuroendocrine neoplasms of foregut, midgut, hindgut, and unknown primary. *Neuroendocrinology*, 2012; **95**: 157–76.
31. Pavel M, O'Toole D, Costa F, *et al*. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. *Neuroendocrinology*, 2016; **103**: 172–85.
32. Graff-Baker AN, Sauer DA, Pommier SJ, Pommier RF. Expanded criteria for carcinoid liver debulking: maintaining survival

and increasing the number of eligible patients. *Surgery*, 2014; **156**: 1369-76; discussion 76–7.
33. Maxwell JE, Sherman SK, O'Dorisio TM, Bellizzi AM, Howe JR. Liver-directed surgery of neuroendocrine metastases: what is the optimal strategy? *Surgery*, 2016; **159**: 320–33.
34. Gedaly R, Daily MF, Davenport D, *et al*. Liver transplantation for the treatment of liver metastases from neuroendocrine tumors: an analysis of the UNOS database. *Arch Surg*, 2011; **146**: 953–8.
35. Le Treut YP, Grégoire E, Klempnauer J, *et al*. Liver transplantation for neuroendocrine tumors in Europe-results and trends in patient selection: a 213-case European liver transplant registry study. *Ann Surg*, 2013; **257**: 807–15.
36. Garcia-Carbonero R, Sorbye H, Baudin E, *et al*. ENETS consensus guidelines for high-grade gastroenteropancreatic neuroendocrine tumors and neuroendocrine carcinomas. *Neuroendocrinology*, 2016; **103**: 186–94.
37. Rinke A, Müller HH, Schade-Brittinger C, *et al*. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*, 2009; **27**: 4656–63.
38. Caplin ME, Pavel M, Ćwikła JB, *et al*. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*, 2014; **371**: 224–33.
39. Wolin E, Pavel M, Ćwikła J, *et al*. Final progression-free survival (PFS) analyses for lanreotide autogel/depot 120 mg in metastatic enteropancreatic neuroendocrine tumors (NETs): the CLARINET extension study. *J Clin Oncol*, 2017; **35**(Suppl 15): 4089.
40. Moertel CG, Lefkopoulo M, Lipsitz S, Hahn RG, Klaassen D. Streptozocin-doxorubicin, streptozocin-fluorouracil or chlorozotocin in the treatment of advanced islet-cell carcinoma. *N Engl J Med*, 1992; **326**: 519–23.
41. Kouvaraki MA, Ajani JA, Hoff P, *et al*. Fluorouracil, doxorubicin, and streptozocin in the treatment of patients with locally advanced and metastatic pancreatic endocrine carcinomas. *J Clin Oncol*, 2004; **22**: 4762–71.
42. Fjallskog ML, Janson ET, Falkmer UG, Vatn MH, Öberg KE, Eriksson BK. Treatment with combined streptozotocin and liposomal doxorubicin in metastatic endocrine pancreatic tumors. *Neuroendocrinology*, 2008; **88**: 53–8.
43. Strosberg JR, Fine RL, Choi J, *et al*. First-line chemotherapy with capecitabine and temozolomide in patients with metastatic pancreatic endocrine carcinomas. *Cancer*, 2011; **117**: 268–75.
44. Fine RL, Gulati AP, Krantz BA, *et al*. Capecitabine and temozolomide (CAPTEM) for metastatic, well-differentiated neuroendocrine cancers: The Pancreas Center at Columbia University experience. *Cancer Chemother Pharmacol*, 2013; **71**: 663–70.
45. Sorbye H, Welin S, Langer SW, *et al*. Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the NORDIC NEC study. *Ann Oncol*, 2013; **24**: 152–60.
46. Hadoux J, Malka D, Planchard D, *et al*. Post-first-line FOLFOX chemotherapy for grade 3 neuroendocrine carcinoma. *Endocr Relat Cancer*, 2015; **22**: 289–98.
47. Hentic O, Hammel P, Couvelard A, *et al*. FOLFIRI regimen: an effective second-line chemotherapy after failure of etoposide-platinum combination in patients with neuroendocrine carcinomas grade 3. *Endocr Relat Cancer*, 2012; **19**: 751–7.
48. Yao JC, Shah MH, Ito T, *et al*. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 514–23.
49. Raymond E, Dahan L, Raoul JL, *et al*. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 501–13.
50. Pavel ME, Hainsworth JD, Baudin E, *et al*. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2011; **378**: 2005–12.
51. Strosberg J, El-Haddad G, Wolin E, *et al*. Phase 3 trial of 177Lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med*, 2017. **376**: 125–35.
52. Brabander T, van der Zwan WA, Teunissen JJM, *et al*. Long-term efficacy, survival, and safety of [177Lu-DOTA0,Tyr3] ocreotate in patients with gastroenteropancreatic and bronchial neuroendocrine tumors. *Clin Cancer Res*, 2017; **23**(16): 4617–24.
53. Modlin IM, Lye KD, Kidd M. A 5-decade analysis of 13,715 carcinoid tumors. *Cancer*, 2003; **97**: 934–59.

6.6

Gastrinoma

Christos Toumpanakis and Martyn E. Caplin

Introduction

Gastrin is a gastrointestinal hormone, produced predominantly by the G cells of the gastric antrum and duodenum, although small amounts of gastrin have been isolated in the pituitary and some vagal nerve fibres. The biologically active forms of gastrin include carboxy-amidated-gastrin-17, and also carboxy-amidated-gastrin-34, which bind mainly to cholecystokinin (CCK)-2 receptor. The main role of amidated gastrin is the stimulation of gastric acid secretion by regulating histamine release from the gastric enterochromafin-like (ECL) cells, while it may also have a trophic effect on gastric mucosa. There is evidence that the precursor forms of gastrin, such as progastrin and glycine-extended gastrin are also of biological importance, binding to a separate 'CCK-C' receptor. They may induce cellular and tumour growth and these precursors are implicated in several cancers, such as colon and pancreatic adenocarcinomas.

Gastrinomas represent a group of functional predominantly duodenal and pancreatic neuroendocrine tumours, characterized by autonomous release of gastrin by the tumour cells, which results in symptoms not only due to the tumour growth *per se*, but also due to gastric acid hypersecretion.

In 1955 at the annual meeting of American Surgical Association, Dr Robert M. Zollinger and Dr Edwin H. Ellison presented a study entitled 'Primary Peptic Ulcerations of the Jejunum Associated with Islet Cell Tumour of the Pancreas' [1]. They proposed a new clinical syndrome of: (a) ulceration in unusual locations in the upper gastrointestinal tract or recurrent ulcerations; (b) gastric acid hypersecretion; and (c) non-beta islet cell tumours of the pancreas. However, the potent gastric secretagogue for the 'Zollinger–Ellison syndrome' (ZES) was not identified until 1960, when Rodney Gregory and Hilda Tracy of the University of Liverpool discovered that, the extract from the pancreas of a patient with ZES was the hormone 'gastrin'. Thus, these pancreatic tumours have been called 'gastrinomas'.

Epidemiology

The incidence of gastrinomas is 0.5–21.5/million population/year. There is a slight male predominance with an average age of diagnosis in the mid-forties, although patients may well have had symptoms for 5–6 years [2].

Genetics

Gastrinomas can either be sporadic or can be associated with multiple endocrine neoplasia type 1 (MEN-1) syndrome in 25% of cases [3]. This syndrome should always be considered and excluded in every patient with gastrinoma, as it may have significant implications in patient management and also prognosis. MEN-1 is an autosomal-dominant disorder due to mutations in MENIN, with high penetrance (see Chapter 6.11.1). Approximately, 23–29% of patients with gastrinomas associated with MEN-1, may develop gastric neuroendocrine (carcinoid) tumours type II. Recently, it was found that somatic *MEN-1* mutations are found also in 37% of sporadic gastrinomas, which indicates that *MEN-1* may be involved in the pathogenesis of both familial and sporadic cases. Other genetic alterations associated with sporadic gastrinomas include those in $p16^{INK4a}$ gene on chromosome 9p21 and in the *HER2/neu* gene. Mutations of p53 gene are not common in gastrinomas [4].

Pathophysiology and Molecular Pathogenesis

Gastrinomas secrete mainly the carboxy-amidated-gastrin-17, which is associated with gastric acid hypersecretion and relevant

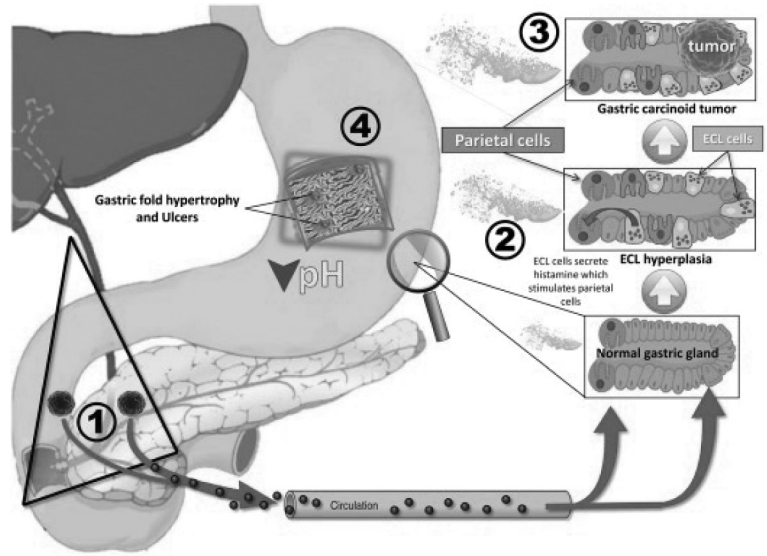


Figure 6.6.1 Pathophysiology of type II gastric carcinoids within the context of gastrinomas associated with MEN-1 syndrome. Gastrinomas usually located inside the gastrinoma triangle (duodenum or pancreas) secrete excess amount of gastrin (1). Gastrin stimulates parietal cells to secrete gastric acid (2). Gastrin also stimulates mucosal growth in the stomach causing ECL cell hyperplasia (3), which can then lead to the formation of a neuroendocrine tumour (gastric carcinoid type II).

clinical features. This autonomous hypergastrinaemia induces ECL cell hyperplasia, which may lead to the development of gastric neuroendocrine tumours (Figure 6.6.1) [5]. Apart from amidated gastrin, these tumours may also secrete precursor forms, progastrin, and glycine-extended gastrin [6]. Gastrinomas overexpress various growth factor receptors (epidermal growth factor receptor, insulin growth factor 1 receptor and hepatocyte growth factor receptor) but among these, only epidermal growth factor receptor (EGFR) is known to correlate with angioinvasion. The activation of CCK-2 receptor by carboxy-amidated-gastrin-17 initiates multiple signal transduction pathways, including activating EGFR.

Clinicopathological Features

Although gastrinomas were previously reported to be located predominantly in the pancreas, recent series have shown that the duodenum (especially the first and the second part) is the most common location for both sporadic and MEN-1 associated gastrinomas (50–88% and 70–100% of cases, respectively). In cases of pancreatic gastrinomas, the tumour is usually located (70%) in the pancreatic body/tail. At surgery, 70–85% of all gastrinomas are found in the so-called gastrinoma triangle, an anatomical area that is defined, superiorly by the confluence of the cystic and common bile duct; inferiorly by the junction of the second and third portions of the duodenum; and medially, by the junction of the neck and body of the pancreas. Rarely (10%), gastrinomas can be found in other abdominal (stomach, liver, bile duct, ovary) or extra-abdominal (heart, small lung cancer) sites.

At presentation, pancreatic gastrinomas are usually large lesions (mean size 3.8 cm), whereas the duodenal ones, especially those associated with MEN-1, are small (1–20 mm). The latter are usually multiple, while sporadic gastrinomas are predominantly (80%) solitary tumours.

Gastrinomas may metastasize to the liver, lymph nodes, and rarely to the bones. At presentation, the majority of patients have either localized (36%) or locally advanced disease with only lymph node metastases (29%). The development of metastases is independent of tumour size, as even small (<5 mm) duodenal gastrinomas may have a high malignant potential. However, liver metastases seem to be more common in pancreatic rather than duodenal gastrinomas (22–35% vs. 0–10%, respectively) [7].

Gastrinomas share the same histopathological features with the other gastroenteropancreatic neuroendocrine tumours (Figure 6.6.2). Additionally, they predominantly express gastrin (Figure 6.6.3), and can produce other gastrointestinal peptides as well, such as insulin. However, such additional peptides are either not released in the systemic circulation or they are released in small quantities and thus, they are not usually of any clinical significance. Other

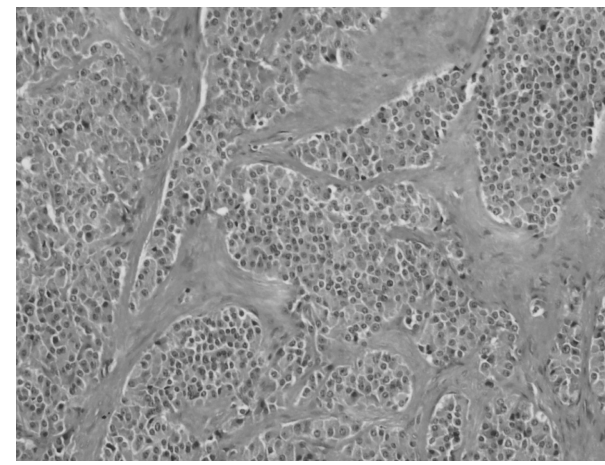


Figure 6.6.2 Well-differentiated, grade 1, pancreatic gastrinoma.

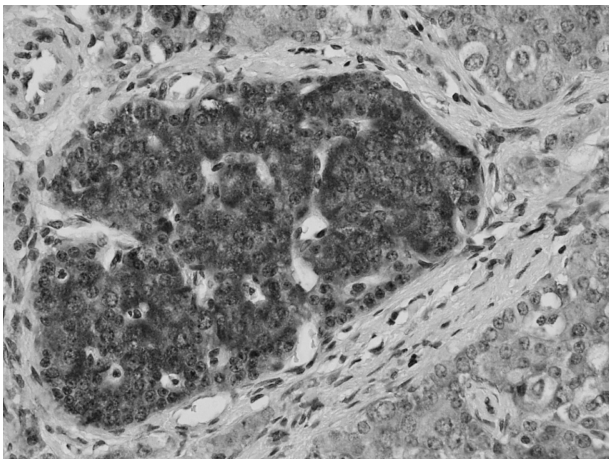


Figure 6.6.3 Gastrin immunostaining in the above patient with pancreatic gastrinoma.

neuroendocrine tumours may express gastrin, but they are not considered gastrinomas if clinical and biochemical features are lacking.

According to the recent (2017) World Health Organization (WHO) for pancreatic neuroendocrine neoplasms, gastrinomas are divided into: (1) well-differentiated neuroendocrine tumours of grade(G) 1, G2, or G3; and (2) poorly differentiated neuroendocrine carcinomas (either small-cell or large-cell type) with high-grade (G3) malignant behaviour. Criteria for categorization of these tumours include general morphologic description, mitotic rate (two or more mitoses/mm^2), proliferation index (as assessed by nuclear Ki67 expression), tumour size, and evidence of invasion of blood vessels/nerves/adjacent organs. The majority of gastrinomas are well-differentiated neuroendocrine neoplasms.

In 2010, the European Neuroendocrine Tumour Society suggested the TNM system for staging of pancreatic neuroendocrine tumours, including gastrinomas, which also included a grading system based on mitotic rate and proliferation index. That grading system was updated in 2017, as part of the new WHO classification [8]. The TNM system as well as, the 2017 grading system are summarized in Box 6.6.1 and Table 6.6.1.

Box 6.6.1 TNM classification for neuroendocrine tumours of the pancreas

T—primary tumour

Tx: primary tumour cannot be assessed

T0: no evidence of primary tumour

T1: tumour limited to the pancreas with size <2 cm

T2: tumour limited to the pancreas with size 2–4 cm

T3: tumour limited to the pancreas with size >4 cm or invading duodenum or bile duct

T4: tumour invading adjacent organs or the wall of large vessels

N—regional lymph nodes

Nx: regional lymph nodes cannot be assessed

No: no regional lymph node metastases

N1: regional lymph node metastases

M—distant metastases

Mx: distant metastases cannot be assessed

M0: no distant metastases

M1: distant metastases

Stage I: T1, N0, M0

Stage IIa: T2, N0, M0

Stage IIb: T3, N0, M0

Stage IIIa: T4, N0, M0

Stage IIIb: any T, N1, M0

Stage IV: any T, any M, M1

Diagnosis

Clinical Presentation

Most of the symptoms in patients with gastrinomas are associated with gastric acid hypersecretion. The latter can lead to the development of peptic ulcers, erosive oesophagitis, and chronic diarrhoea. Symptoms associated with peptic ulcers and their complications (bleeding, perforation, pyloric stenosis) are the most common presenting clinical features in patients with gastrinoma (80%). Peptic ulcers in these patients are often multiple, located in unusual anatomic sites and resistant to treatment. They are less associated to *H. pylori* infection compared to idiopathic peptic ulcers (24–48% vs. >90%, respectively) and are not associated with use of non-steroidal anti-inflammatory drugs (NSAIDS). Erosive oesophagitis causing heartburn and potentially dysphagia occur in 50–60% of these patients. Finally, chronic diarrhoea is a result of inactivation of pancreatic enzymes (especially lipases), and damage of the intestinal mucosa, due to acid hypersecretion. Gastrinoma-related diarrhoea is usually watery, may be associated with malabsorption, while it is the only diarrhoea that responds dramatically to proton pump inhibitors (PPIs). It occurs in 40–70% of patients with gastrinoma, and may be the only symptom in 20% of cases [9].

Clinical suspicion for a gastrinoma associated with MEN-1 syndrome is raised when one or more of the aforementioned clinical features coexist with hyperparathyroidism or any other MEN-1 related endocrinopathies, and when there is a family history of MEN-1 syndrome.

In patients with sporadic gastrinomas, the mean age at the onset of symptoms is 48–55 years, while patients with MEN-1 present usually at an earlier age (32–35 years). The frequency of most symptoms in similar between these two groups of patients, although it seems that diarrhoea is less common in MEN-1 patients.

Less than 10% of gastrinoma patients develop features of Cushing's syndrome due to ectopic production of adrenocorticotrophic hormone (ACTH), which represents a poor prognostic sign.

Table 6.6.1 WHO 2017 Grading for pancreatic neuroendocrine neoplasms

	Mitotic count (10 HPF)	Ki67 index (%)
Well-differentiated Neuroendocrine Neoplasms		
Neuroendocrine tumour (G1)	<2	<3
Neuroendocrine tumour (G2)	2–20	3–20
Neuroendocrine tumour (G3)	>20	>20
Poorly differentiated neuroendocrine neoplasms		
Neuroendocrine carcinoma (G3)	>20	>20

Table 6.6.2 Clinical features indicating gastrinoma

Clinical feature	% of patients
Peptic ulcers resistant to treatment, multiple, located in unusual anatomic sites, resistant to treatment, less associated to *H. pylori*, and not associated with use of NSAIDS	80
Erosive oesophagitis causing heartburn and potentially dysphagia, resistant to treatment	50
Diarrhoea responding to proton pump inhibitors	40–70
The above features in combination with other endocrinopathies	20
The above features in combination with family history of neuroendocrine tumours	25

NSAIDS, non-steroidal anti-inflammatory drugs.

Suspicious clinical features that may indicate gastrinoma are summarized in Table 6.6.2.

Laboratory Tests

The biochemical confirmation, following a clinical suspicion, of a gastrinoma requires a significant elevation of **fasting serum gastrin**, in combination with hyperchlorhydria. The presence of the latter is very important, as hypergastrinaemia alone can be result of chronic hypochlorhydria/achlorhydria, that is associated with chronic fundus atrophic gastritis, chronic use of PPIs, as well as vagotomy.

A fasting serum gastrin level of >10-fold the upper normal limit in the presence of gastric p H<2 or basal acid output (BAO)>15 mmol/h is considered as diagnostic of a gastrinoma. If possible, PPIs should be discontinued preferably at least ten days prior to serum gastrin estimation, while a discontinuation of histamine-2 receptor (H2R) antagonists for only 48–72 hours prior to the test seems to be adequate. However, discontinuation of PPIs should be performed with caution in patients who are at risk for complications (peptic ulcer bleeding or perforation, severe diarrhoea) and those patients should ideally be referred to specialist units with experience in diagnosis and management of gastrinomas. Moderately elevated serum gastrin levels (<10-fold the upper normal limit) and hyperchlorhydria may occur in 66% of gastrinoma patients, but in this situation other clinical disorders need to be excluded, such as: *H. pylori* infection, gastric outlet obstruction, antral G cell hyperplasia, short bowel syndrome, retained antrum, or renal failure. In this situation, a provocative test with intravenous (IV) administration of secretin is indicated [10]. After an overnight fast an IV bolus of secretin (2 U/kgr) is given to the patient. A rise of serum gastrin concentration of 120 pg/ml, noted within 10 minutes of secretin administration, can establish the diagnosis of gastrinoma (with sensitivity and specificity of 94% and 100%, respectively), whereas in the aforementioned non-tumour-related causes the serum gastrin level remains unchanged. A diagnostic algorithm for fasting serum gastrin levels is summarized in Figure 6.6.4.

Another biochemical test that contribute to the diagnosis is the estimation of *Chromogranin-A (CgA)* (see Chapter 6.2). However, recent studies in patients with gastrinomas suggest that there is not always precise correlation between CgA levels and tumour burden, as in these patients CgA may also be produced by the enterochromaffin-like cells of stomach in response to hypergastrinaemia.

As soon as the biochemical diagnosis of gastrinoma is established, it is important for all patients (even those without any suspicious clinical features) to be screened for MEN-1, with a baseline estimation of serum parathyroid hormone (PTH) levels, calcium levels (preferably ionized calcium or albumin-corrected calcium), as well as prolactin. In presence of suspicious clinical or biochemical data, a MEN-1 germline mutation DNA test should be performed, and if it is positive, genetic counselling is offered to all patients' kindreds after their first decade of life.

Rarely, patients with gastrinoma may have hypercalcaemia with normal PTH levels, which may be a result of parathyroid hormone-related peptide (PTH-r P) secreted by these tumours. Finally, in patients who develop clinical signs of Cushing's syndrome biochemical evaluation is needed (see Chapter 5.7).

Tumour Localization

Precise localization of the primary tumour lesion as well as metastatic deposits has a significant impact on the patient's management. Invasive and non-invasive localization studies are used in order to identify the primary lesion and determine the extent of resection

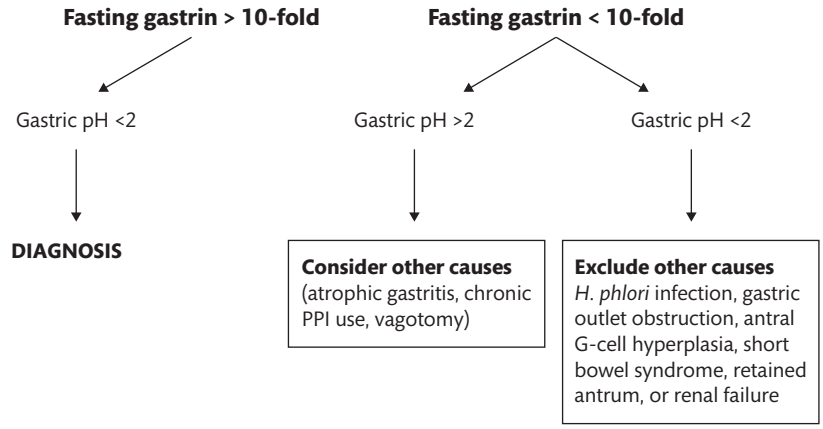


Figure 6.6.4 Differential diagnosis of hypergastrinemia (ULN: upper limit of normal).

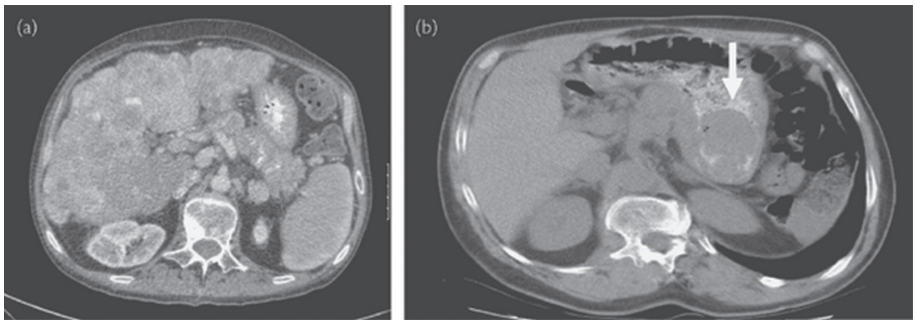


Figure 6.6.5 (a) CT in metastatic sporadic gastrinoma. (b) CT in a patient with MEN-1 gastrinoma and type II gastric neuroendocrine tumour (arrow).

when surgery is planned, and also to assess the tumour extent in patients with advanced disease.

Non-Invasive Techniques

Among the conventional radiological studies, transabdominal ultrasound has the lowest sensitivity for detection of the primary site and hepatic metastases (20% and 45% respectively), but its specificity may be greater than 90%. Spiral computed tomography (CT) and MRI have better sensitivities for the primary lesion (59–78% and 40–85%, respectively) and for distant metastases (45–70% and 70–80%, respectively), while their specificity is also above 90% (Figure 6.6.5).

Indium-111-diethylentriamine penta-acetic acid (DTPA)-octreotide (OctreoScan) was previously recognized as the most sensitive modality for gastroenteropancreatic neuroendocrine tumours' imaging (Figure 6.6.6). Recently the introduction of positron emission tomography (PET) molecular imaging with 68Ga-labelled somatostatin analogues provides clear advantages over conventional OctreoScan with better sensitivity and specificity and is logistically easier as it is a single scan compared to the requirement with OctreoScan to scan at 4 hours, 24 hours, and 48 hours post injection of isotope. ^{68}Ga-somatostatin analogue PET imaging is now considered the gold standard for nuclear medicine imaging of NETs. Multiple peptides have been used with in conjunction with ^{68}Ga. Somatostatin analogues such as DOTA-[Tyr3]-octreotide (TOC), DOTA-[Tyr3,Thr8]-octreotide (TATE), and DOTA-[Nal3]-octreotide (NOC) have recently been used with PET imaging. 68Ga-DOTA PET studies, that can be completed within 2–3 hours and visualize lesions as small as 7 mm, have a sensitivity of 97% and specificity of 92% (Figure 6.6.7). It provides not only an accurate localization of the primary and metastatic lesions, but also may detect unsuspected lesions, not shown by the previous conventional studies, which is crucial when surgery is planned. If, the aforementioned non-invasive techniques (including ^{68}Ga-DOTA PET) have failed to identify the primary tumour site, especially when surgery is planned, a variety of invasive localization studies can be performed.

Invasive Techniques

Oesophagogastroduodenoscopy (OGD) needs to be performed in every patient with gastrinoma. OGD not only corroborates the clinical suspicion by demonstrating severe oesophagitis, peptic ulcers in uncommon locations and prominent gastric folds, but also may reveal small duodenal gastrinomas, as well as gastric polyps that may represent gastric neuroendocrine tumours type II in patients with MEN-1 gastrinomas (Figure 6.6.8).

Endoscopic ultrasound (EUS) should be considered in all patients who are due to have an operation and especially in those with MEN-1 gastrinomas. In experienced hands, it has a sensitivity of 95% for pancreatic gastrinomas, while it also provides useful staging information. EUS can precisely estimate tumour size, distinguish a pancreatic tumour from a peripancreatic lymph node, and also enables histology samples via fine-needle aspiration. However, EUS has a lower sensitivity for duodenal gastrinomas (50%).

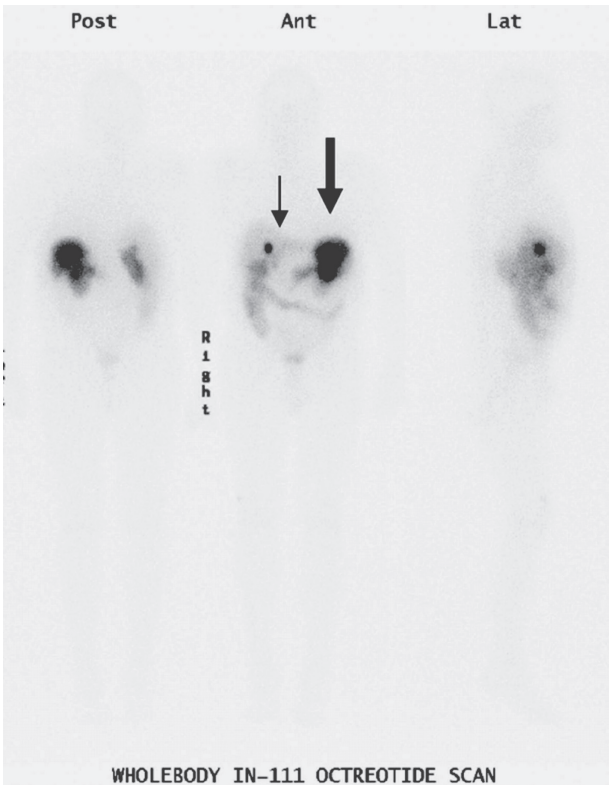


Figure 6.6.6 OctreoScan in patient with metastatic MEN-1 gastrinoma. Solitary hepatic metastatic deposit (thin arrow), gastric neuroendocrine tumour (thick arrow).

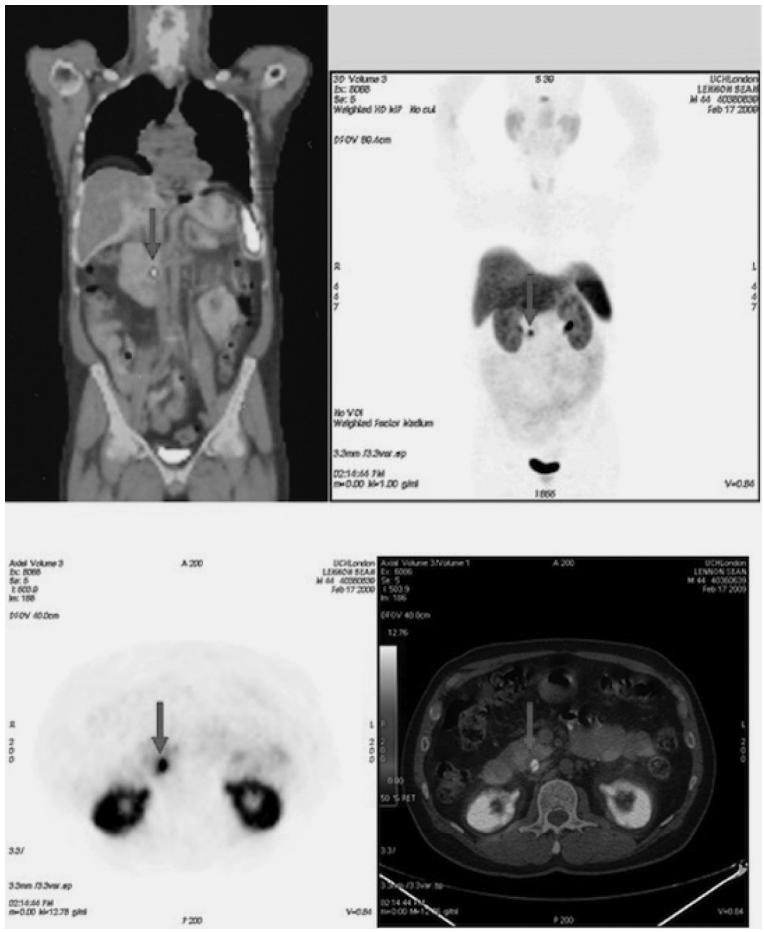


Figure 6.6.7 ^{68}Ga-DOTA PET in patient with sporadic gastrinoma and unknown primary, despite all previous imaging studies. A foci of uptake was demonstrated at the area of second part of duodenum thick red arrow).

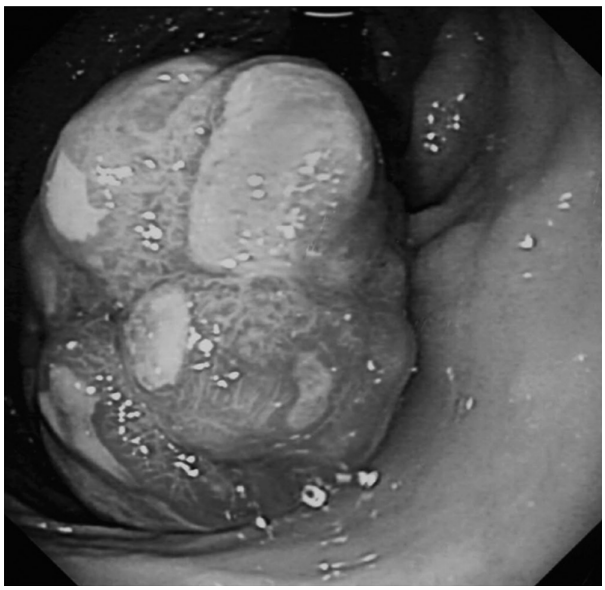


Figure 6.6.8 Gastric neuroendocrine tumour type II in a patient with MEN-1 gastrinoma (same patient as in Figure 6.5.3b).

Angiography is less commonly used nowadays, and usually includes angiography with secretin or calcium stimulation and hepatic venous sampling. During this test, small doses of secretin are injected into the splenic, hepatic, gastroduodenal, and then the superior mesenteric arteries, following a selective cannulation. Then, blood samples are taken from the right hepatic vein, at baseline, 20, 40, 60, and 120 seconds for gastrin estimation.

In patients, where all the aforementioned imaging techniques have failed to detect the primary lesion preoperatively, an intraoperative transillumination of the duodenum in combination with intraoperative ultrasound should always be performed. These tests should always be followed by duodenotomy, which is able to detect up to 50% of duodenal tumours.

Management

Once the diagnosis of Zollinger–Ellison syndrome is established, and even before the completion of localization studies, it is important to start medical treatment in order control gastric acid hypersecretion and prevent its complications.

Control of Symptoms

In the early years, the treatment of choice was total gastrectomy. However, as soon as the antisecretory medications, initially H_2-receptor antagonists and subsequently PPIs,) became widely available, effective control of gastric acid secretion was achieved, and thus, total gastrectomy is no longer necessary.

PPIs are superior to H_2-receptor antagonists, in terms of effectiveness in gastric acid control and long-term action, and therefore they are considered as first-choice medical treatment for these patients. The initial dose that is usually used is an equivalent dose to 60 mg of omeprazole/day in sporadic gastrinomas [11]. Patients with MEN-1 gastrinomas and hypercalcaemia, patients with severe gastro-oesophageal reflux symptoms as well as those with previous Bilroth-II resection may require higher initial doses (40–60 mg bd). The maintenance dose tends to be lower than the initial dose in 40–80% of the patients. PPIs are considered as safe long-term treatment with patients being treated for more than 15 years with no significant adverse effects and no evidence of tachyphylaxis. It has been, however, recommended that long-term treatment with PPIs may be associated with B_{12} malabsorption, and thus B_{12} levels need to be monitored once a year. There may also be an increased predisposition to osteoporosis.

Patients who cannot tolerate oral PPIs can be treated with high dose of H_2-receptor antagonists and the dose requirements tends to increase with time.

The use of somatostatin analogues (octreotide and lanreotide preparations) for antiacid control, although potentially effective, is not a first-line option and should be reserved only for patients with difficult to control hyperacidity and those who had developed gastric neuroendocrine tumours.

Gastric surgery (vagotomy, gastrectomy) for control of gastric acid hypersecretion is very rarely required in the era of antisecretory medications, and is only indicated in patients who cannot or will not take any antisecretory medications. Parathyroidectomy in patients with MEN-1 gastrinomas and parathyroid adenomas is usually indicated at an early stage, as results in a decrease in the basal acid output and fasting gastrin levels and therefore the required dosage of antisecretory medications as well.

Surgical Therapy

Sporadic Gastrinomas

The aim of surgery in sporadic gastrinomas patients is a long-term curative resection, in order to decrease the risk of distant metastases development, as well as to completely control of the hormonal symptoms.

Pancreatic head/body tumours may be enucleated, in combination with regional lymph node dissection (peripancreatic, periduodenal, and hepatoduodenal ligament), while duodenotomy is always required to detect small duodenal gastrinomas. Whipple's pancreaticoduodenectomy may be required for large gastrinomas. Tumours in pancreatic tail may be enucleated, or for larger lesions a distal pancreatectomy is performed [11, 12].

The overall 15-years survival following a curative resection in sporadic gastrinoma patients is between 80% and 100%.

MEN-1 Gastrinomas

The benefit of surgery in gastrinomas associated with MEN-1 syndrome is controversial. Tumours are usually small, multiple, and located in various sites in pancreas and duodenum. Surgical exploration, especially if the tumours have not been visualized in the preoperative imaging studies or their size is <2 cm, does not seem to be worthwhile, as the curative potential in these cases is low, metastases are usually regional, and the long-term prognosis is favourable. Surgical intervention should be considered in presence of tumour size >2 cm. The concept of this approach is to reduce the possibility of subsequent liver metastases, although its efficacy remains controversial [11–13].

Management of Advanced Disease

All patients with advanced disease, consisting mainly of hepatic metastases, should have antitumour treatment in order to prolong the survival rates. Treatment options include invasive procedures or systemic medical treatment.

Invasive Treatment

Cytoreductive surgery may be an option when hepatic metastases are confined to one lobe, and also in cases where more than 90% of disease can be removed surgically. Liver resection can also be combined with radiofrequency ablation (RFA) at the same time, for tumour lesions that cannot be resected.

In patients with disease predominantly in the liver, transarterial hepatic embolization (TAE) or chemoembolization (TACE) may be considered, especially when a clinical or radiological progression is noted. TAE or TACE cannot be performed in patients with occluded portal vein and poor performance status.

Finally, liver transplantation can be considered rarely in selective cases, mainly young patients with G1 tumours, no significant comorbidities, and no evidence of extrahepatic disease [14].

Non-Invasive Treatment

In patients unsuitable for surgery, systemic medical treatment should be considered in order to control tumour growth and improve survival.

Biological agents, such as somatostatin analogues and less commonly interferon, are considered as antitumour treatment in patients with G1/G2 tumour and slowly progressive disease [15]. A recent large randomized trial with Lanreotide Autogel (long acting preparation of lanreotide) demonstrated increased progression-free survival in a series of gastro-entero-pancreatic neuroendocrine tumours, including gastrinoma patients [16]. Overall, disease stabilization can be achieved in up to 50% of patients being treated with somatostatin analogues, whereas the objective response rate (decrease in tumour size >20%) with those medications is approximately 5–10%.

Novel molecular targeted treatments, such as oral everolimus (mTOR inhibitor) and oral sunitinib (tyrosine-kinase-inhibitor) have shown increased progression-free survival in large randomized studies of advanced pancreatic neuroendocrine tumours, which included patients with gastrinomas. Objective tumour responses with those agents are noted less than 10% of patients [17, 18]. Serum gastrin levels have been decreased, following commencement of everolimus, and that became more prominent when everolimus was combined with somatostatin analogues [19].

Cytotoxic chemotherapy, using combination regimens based on streptozotocin, and more recently temozolomide +/– capecitabine, may be an option in patients with well-differentiated G2/G3

tumours and disease progression despite treatment with biological or molecular targeted agents, and in patients with extensive tumour load and significant symptomatic deterioration. Objective responses and disease stabilization have been reported in 30–40% and 70% of patients, respectively. Cytotoxic chemotherapy, mainly platinum-based, remains the first treatment option in poorly differentiated, G3, neuroendocrine carcinomas [14, 20].

Peptide receptor radionuclide therapy, represent an alternative option to cytotoxic chemotherapy and can be considered in patients with avid tumour uptake in OctreoScan or in ^{68}Ga-DOTA somatostatin analogue PET. The concept of this treatment is to transfer cytotoxic radiation directly to the tumour cells, by using somatostatin analogues (octreotide or octreotate) radiolabelled with isotopes such as yttrium-90 or more commonly lutetium-177. Although data from large randomized trials are lacking for pancreaticoduodenal neuroendocrine tumours, peptide receptor radionuclide therapy, especially Lutetium-177 DOTATATE, has demonstrated high objective response rate, increased progression-free survival, as well as overall survival, based on several large retrospective series [21].

The optimal sequence of the aforementioned systemic treatments, in advanced gastroenteropancreatic neuroendocrine tumours, is currently being evaluated into clinical trials, which include patients with gastrinomas.

Prognostic Factors–Survival

The 15-year survival rate in patients either sporadic or MEN-1 gastrinomas and localized disease (including lymph nodes metastasis) is >90%, whereas in those with hepatic metastases the 5-year survival is between 50% and 70%. Other poor prognostic factors include: short disease history; female gender; pancreatic primary location; primary tumour size >3 cm; absence of MEN-1 syndrome; extremely high serum gastrin levels; ectopic Cushing syndrome; and presence of bone metastases [11]. Histological features such as: angioinvasion, perineural invasion, mitotic rate > 2/20HPF, Ki67>2% are associated with poor prognosis, as is the overexpression of EGFR and *HER2/neu* gene expression.

REFERENCES

1. Zollinger RM, Ellison EH. Primary peptic ulcerations of the jejunum associated with islet cell tumors of the pancreas. *Ann Surg*, 1955; **142**: 709–23.
2. Caplin ME. Zollinger–Ellison syndrome. In: Modlin I, ed. *From Gastrin to GERD: A Century of Acid Suppression*. Hanover: Felsenstein, 2006.
3. Jensen RT. Gastrinomas: advances in diagnosis and management. *Neuroendocrinology*, 2004; **80** Suppl 1: 23–7.
4. Toumpanakis CG, Caplin ME. Molecular genetics of gastroenteropancreatic neuroendocrine tumors. *Am J Gastroenterol*, 2008; **103**: 729–32.
5. Grozinsky-Glasberg S, Alexandraki KI, *et al*. Gastric carcinoids. *Endocrinol Metab Clin North Am*, 2018; **47**: 645–60.
6. Ellison EC. Zollinger–Ellison syndrome: a personal perspective. *Am Surg*, 2008; **74**: 563–71.
7. Ito T, Igarashi H, Jensen RT. Zollinger–Ellison syndrome: recent advances and controversies. *Curr Opin Gastroenterol*, 2013; **29**: 650–61.
8. Perren A, Couvelard A, Scoazec JY, *et al*. ENETS consensus guidelines for the standards of care in neuroendocrine tumors: pathology: diagnosis and prognostic stratification. *Neuroendocrinology*, 2017; **105**: 196–200.
9. Mendelson AH, Donowitz M. Catching the zebra: clinical pearls and pitfalls for the successful diagnosis of Zollinger–Ellison syndrome. *Dig Dis Sci*, 2017; **62**: 2258–65.
10. Ramage JK, Ahmed A, Ardill J, *et al*. Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours (NETs). *Gut*, 2012; **61**: 6–32.
11. Jensen RT, Niederle B, Mitry E, *et al*. Gastrinoma (duodenal and pancreatic). *Neuroendocrinology*, 2006; **84**: 173–82.
12. Fendrich V, Langer P, Waldmann J, Bartsch DK, Rothmund M. Management of sporadic and multiple endocrine neoplasia type 1 gastrinomas. *Br J Surg*, 2007; **94**: 1331–41.
13. Alexakis N, Neoptolemos JP. Pancreatic neuroendocrine tumours. *Best Pract Res Clin Gastroenterol*, 2008; **22**: 183–205.
14. Pavel M, O'Toole D, Costa F, *et al*. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. *Neuroendocrinology*, 2016; **103**: 172–85.
15. Shah T, Caplin M. Endocrine tumours of the gastrointestinal tract. Biotherapy for metastatic endocrine tumours. *Best Pract Res Clin Gastroenterol*, 2005; **19**: 617–36.
16. Caplin ME, Pavel M, Ćwikła JB, *et al*. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*, 2014; **371**: 224–33.
17. Yao JC, Shah MH, Ito T, *et al*. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 514–23.
18. Raymond E, Dahan L, Raoul JL, *et al*. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 501–13.
19. Pavel ME, Chen D, He W, *et al*. Everolimus effect on gastrin and glucagon in pancreatic neuroendocrine tumors. *Pancreas*, 2017; **46**: 751–7.
20. Falconi M, Eriksson B, Kaltsas G, *et al*. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
21. Brabander T, van der Zwan WA, Teunissen JJM, *et al*. Long-term efficacy, survival, and safety of [177lu-dota0,tyr3]octreotate in patients with gastroenteropancreatic and bronchial neuroendocrine tumors. *Clin Cancer Res*, 2017; **23**: 4617–24.

6.7

Insulinoma and Hypoglycaemia

Ingrid Y.F. Mak and Ashley B. Grossman

Key Points

1. Hypoglycaemia is not uncommon in both hospitalized and ambulatory patients, diabetic and non-diabetic individuals. This condition is potentially life-threatening if left untreated, hence it is crucial to identify and treat hypoglycaemia and its underlying cause as early as possible.
2. Insulinoma is the most common cause of endogenous hyperinsulinaemic hypoglycaemia in otherwise well patients. It is small, benign, solitary, and intrapancreatic in most cases, and may be associated with the MEN1 syndrome.
3. The diagnosis of insulinoma should be suspected in patients having symptomatic hypoglycaemia typical of Whipple's triad, spontaneously or during a 48–72 h fast, with elevated blood insulin and C-peptide levels in the absence of sulphonylureas in the blood or urine and autoantibodies to insulin or insulin receptors. These biochemical test results, together with careful history taking and imaging studies, are useful in differentiating between the various causes of endogenous hyperinsulinaemic hypoglycaemia.
4. Tumour localization for diagnostic confirmation and guidance of surgical treatment is usually possible by non-invasive techniques with good diagnostic performances such as CT, MRI, endoscopic ultrasound, and functional imaging. Selective arterial calcium stimulation with hepatic venous sampling, the more sensitive but invasive approach, is currently regarded as the gold standard for preoperative localization of insulinoma and should be considered in cases with diagnostic difficulty, although it may soon be superseded by radionuclide functional imaging.
5. Surgical excision, particularly laparoscopic enucleation, with the use of diazoxide for preoperative glycaemic control, is the treatment of choice for insulinomas. For inoperable or metastatic cases, other treatment options include everolimus, somatostatin analogues, peptide receptor radionuclide therapy, chemotherapy, and liver-debulking procedures.

Insulinomas

Introduction and Epidemiology

Hypoglycaemia is an important and life-threatening condition which may occur in many circumstances, and especially in patients with significant disease in intensive care units, and in patients with type 1 and type 2 diabetes mellitus. However, in the previously-well patient with bouts of confirmed hypoglycaemia, the major differential diagnoses will be either an insulinoma or factitious hypoglycaemia. Insulinomas, although rare, are the most common hormone-producing neuroendocrine tumour (NET) of the gastrointestinal tract, with an estimated annual incidence of 0.5–1 per million of the population per year (similar to that of phaeochromocytomas). Insulinomas are four times more common in females. They can occur at any age, but present most commonly in middle age, with a median age at diagnosis of 47 years for sporadic cases and 23 years in patients with multiple endocrine neoplasia type 1 (MEN1) syndrome [1]. Most insulinomas are solitary (90%), benign (90%), small (65% with diameter <1.5 cm), intrapancreatic (99%), and are distributed equally throughout the pancreas: 16% are associated with the MEN1 syndrome and these tumours are bigger, multiple, malignant in 25% of cases, and more resistant to treatment with a higher recurrence rate. Features such as young age of onset (<40 years), multifocal disease, and/or a suspicious family history should prompt screening for the MEN1 syndrome.

Clinical Presentation

Hypoglycaemia is the hallmark of an insulinoma. Patients usually present with non-specific, episodic hypoglycaemic symptoms, which fall into two categories: neuroglycopaenic or autonomic (adrenergic/neurogenic). Neuroglycopaenia occurs when plasma glucose levels fall below 2.2 mmol/L and is due to glucose deprivation in the central nervous system. These symptoms include dizziness, confusion, fatigue, difficulty in speaking or concentrating, headache, changes in vision, seizures, and loss of consciousness.

Autonomic symptoms caused by sympathoadrenal activation consist of sweating, hunger, paraesthesiae, tremor, anxiety, palpitations, and nausea. All these symptoms usually resolve with the ingestion of carbohydrate or injection of glucose. Most patients with insulinomas are exclusively symptomatic in the fasting state (73%), some in the post-prandial period (6%), or both (21%) [2]. They may accommodate by eating regular, high-sugar snacks, exacerbating weight gain (although the hyperinsulinaemia may also contribute to this). Diagnosis is often delayed due to the non-specific presentation.

Diagnosis

The symptoms of hypoglycaemia are often non-specific. Objective measurement of hypoglycaemia during a symptomatic period with relief following administration of glucose (Whipple's triad) is strongly suggestive of endogenous hyperinsulinaemia. During endogenous insulin synthesis by pancreatic β-cells, the precursor hormone proinsulin requires post-translational proteolytic cleavage of C-peptide to release the active hormone insulin. Insulin and C-peptide are thus secreted in equimolar amounts and C-peptide acts as a surrogate marker of endogenous insulin production. In the presence of excessive insulin actions, blood ketones and β-hydroxybutyrate would be suppressed indicating inappropriate inhibition of ketogenesis during fasting. Without any drug exposure, the combination of documented Whipple's triad with simultaneous biochemical hypoglycaemia (blood glucose <3 mmol/L), C-peptidaemia (>0.2 nmol/L), hyperinsulinaemia (>3 uU/ml), and suppressed serum β-hydroxybutyrate (<2.7 mmol/L) is usually pathognomonic of an insulinoma (Table 6.7.1), with a sensitivity and specificity of higher than 90% and 70%, respectively [2]. Some advocate a lower glucose value of less than 2.2 mmol/L as the cut-off, and it has to be admitted that there is a 'grey area' in the range 2.2–3.0 mmol/L. Provocation tests such as a prolonged fasting test and a mixed-meal test should be carried out if there is no hypoglycaemia spontaneously or after an overnight fast.

The 72-h Fast

The prolonged (72-h) supervised fast is considered the gold-standard test. According to previous reports, 65% of patients with insulinoma developed hypoglycaemia at 24 h, 84% at 36 h, 93% at 48 h, and 99% at 72h [3]. Food is withheld and intravenous access is established, although the patient may consume water or non-caloric beverages. Blood glucose is checked every 4–6 hours and when the patient develops symptoms suggestive of hypoglycaemia, and the frequency should be increased when the glucose level becomes relatively low (e.g. <3.5 mmol/L). When the laboratory-measured glucose concentration falls below 3.0 mmol/L, simultaneous serum samples for insulin, C-peptide and β-hydroxybutyrate, as well as blood and/or urine sulphonylurea screening, should be sent for analysis before carbohydrate replacement. If the patient is asymptomatic at the end of the test, a short period of exercise (10–15 min) may provoke hypoglycaemia. In those diseases with abundant insulin or insulin-like growth factor (IGF) levels, blood glucose concentration would increase by more than 1.4 mmol/L after intravenous glucagon reflecting adequate hepatic glycogen stores (Table 6.7.1). However, in our experience the glucagon test is rarely required. In some centres the fast is only continued for 48 hours. It should be emphasized that the patient should be under regular nursing supervision throughout, and if they become stuporous or even develop seizures, then intravenous 5% glucose should be immediately administered (a basal blood sample should always be taken if possible just before the glucose is given). The test is terminated whenever significant hypoglycaemia is confirmed on a venous sample, or if the patient develops their typical symptomatology and the glucose is shown to be normal.

Mixed-Meal Test

A small proportion of insulinoma patients (5%) who do not develop hypoglycaemia during prolonged fasting may have a positive mixed-meal test instead, especially those presenting with postprandial symptoms [4]. During the test, a mixed meal prepared from commercial nutritional supplement formulary or food reported by patient as the possible culprit for inducing hypoglycaemic symptoms is given after an overnight fast. Multiple blood samples are taken regularly (e.g. every 30 minutes until 5 h). The interpretation of test findings is similar to that under fasting conditions.

Tumour Localization

Most insulinomas (>80%) are benign and solitary, rendering surgical cure with complete resection possible. Once a biochemical diagnosis of insulinoma has been established, preoperative localization, and definition of the anatomy is necessary for an optimal surgical outcome. Virtually all insulinomas arise from within the pancreas and therefore localization techniques should be directed to this organ. Some advocate that intraoperative bimanual palpation and ultrasonography by an experienced surgeon may be more sensitive than preoperative localization, but this method requires an open surgical approach and may subject those with other causes of endogenous hyperinsulinism to unnecessary operation. Various modalities such as cross-sectional imaging, endoscopic ultrasonography, functional imaging techniques, and selective intra-arterial injection of calcium with hepatic venous sampling have been adopted for preoperative localization, with different degrees of invasiveness and diagnostic accuracy.

Computer Tomography (CT) and Magnetic Resonance Imaging (MRI)

Multidetector CT (MDCT) is the most popular preoperative imaging for insulinomas. It allows rapid acquisition of images in multiple contrast-enhancement phases with excellent spatial resolution. As insulinomas are usually well-circumscribed oval hypervascular tumours, they in general exhibit more avid contrast enhancement than the background normal pancreatic tissue on CT images in the arterial or pancreatic phase, producing good visual contrast for lesion detection [5]. During the venous phase, there is in general a reduction in the difference in attenuation of the hypervascular tumour and background pancreatic parenchyma, resulting in lower visual conspicuity of the tumour.

MRI at 1.5 T or 3.0 T has become more widely performed nowadays for further lesion characterization due to its high soft tissue contrast. On MRI, insulinomas are typically hypointense on T_1-weighted images and hyperintense on T_2-weighted images compared with the background pancreatic parenchyma, and are more conspicuous on fat-suppressed T_1-weighted images [5]. As in CT, there is often marked enhancement after the intravenous administration of gadolinium with the maximal visual conspicuity in

Table 6.7.1 Interpretation of various biochemical parameters in symptomatic patients with confirmed hypoglycaemia (blood glucose <3.0 mmol/L)

Diagnosis	Insulin	C-peptide	β-HBA	Glucose increase after glucagon	Other suggestive features
Normal response	↓	↓	↑	↓	Malnutrition and/or acute illnesses
Insulinoma	↑	↑	↓	↑	Fasting hypoglycaemia (more common), pancreatic tumour on imaging, MEN1 syndrome
Primary nesidioblastosis	↑	↑	↓	↑	Postprandial hypoglycaemia
PGBH (post gastric bypass)	↑	↑	↓	↑	Postprandial hypoglycaemia, history of bypass surgery
Insulin autoimmune syndrome	↑↑	↑↑	↓	↑	Postprandial hypoglycaemia (more common), presence of other autoimmune diseases or culprit drugs, autoantibodies to insulin
Exogenous insulin	↑↑	↓	↓	↑	History or evidence of prior insulin injection or ready access to insulin
Oral hypoglycaemic agents	↑	↑	↓	↑	Ready access to insulin secretagogues, positive sulphonylurea screen
IGF-mediated	↓	↓	↓	↑	Presence of mesenchymal tumour, ↑pro-IGF-2 and free IGF-2, ↓IGF-1, IGF-2/IGF-1 >3:1 (typically >10:1)

the arterial phase. As diffusion-weighted imaging (DWI) detects changes in molecular diffusion of water in tissues, it may be a promising tool for detecting small insulinomas, especially those without a hypervascular pattern, and liver metastases. Our own study showed the clear advantage of MRI over CT [6], while a recent prospective study demonstrated that contrast-enhanced CT, MRI without DWI, and MRI with DWI yielded sensitivities of around 70%, 80%, and 90% respectively upon analysis of 47 confirmed cases of insulinoma with 51 lesions in total [7]. CT and MRI also allow better delineation of the anatomical location of the tumour and its relationship with vascular structures and the ductal system for surgical planning, in addition to the extent of regional involvement, especially liver metastases.

Endoscopic Ultrasonography (EUS)

Transabdominal ultrasound has a limited role in detecting pancreatic tumours and sonographic visualization of the pancreas due to partial obscuration by stomach and bowel gas, with a sensitivity of only 9–64% [8]. Although EUS is relatively more invasive, it demonstrates high diagnostic accuracy and possesses the benefit of obtaining simultaneous cytology using fine needle aspiration. A recent meta-analysis showed that the pooled sensitivity and specificity of EUS in detecting pancreatic NETs were 87% and 98%, respectively [9]. In another analysis, the mean detection rates of preoperative EUS and intraoperative ultrasound for insulinoma were both found to be 92% [10]. EUS should be considered when cytological diagnosis is desired or when other non-invasive imaging studies are negative or inconclusive. However, this procedure is highly operator-dependent and may miss lesions that are very small, isoechoic, or located at the pancreatic tail. Contrast-enhanced EUS using gas-filled microbubbles as contrast agents is one of the latest advances that may potentially improve lesion conspicuity.

Somatostatin Receptor (SSTR) and Glucagon-Like Peptide-1 Receptor (GLP-1R) Scans

Nuclear medicine makes use of the functional aspects of the tumour rather than simple dimensions alone, thereby providing valuable information when other anatomical studies have failed to localize an occult insulinoma before proceeding to the more invasive approach. In addition, whole body functional scintigraphy allows screening for ectopic lesions and distant metastases.

The importance of somatostatin receptor scintigraphy in the management of both non-functioning and other functioning NETs such as gastrinomas is well-acknowledged in recent literature. Although earlier studies have questioned its utility in localizing insulinoma as its sensitivity could be as low as 24% [11], malignant insulinomas often express SSTR particularly type 2 (SSTR2) [12], so paradoxically this functional technique may be more useful for the more malignant tumours. However, with the advent of newer techniques using high-affinity radioligands targeting SSTR2 coupled with PET/CT e.g. ^{68}Gallium-DOTATATE-PET/CT, there may be more of a role for SSTR-based imaging in the localization of insulinoma. A recent retrospective study demonstrated that ^{68}Gallium-DOTATATE-PET/CT correctly identified the culprit lesion in 9 out of 10 patients with insulinomas (median tumour size of 1.5 cm), in one of which it was the only positive imaging [13]. Apart from localization of insulinoma and preoperative staging of disease, having a positive SSTR scan can also suggest the feasibility of using peptide receptor radionuclide therapy (PRRT) as a therapeutic option in those with metastatic disease (see next).

The vast majority (>90%) of benign insulinomas, as opposed to the malignant and metastatic ones, strongly express GLP-1R with a high-density distribution even when lacking SSTR [14]. These tumours are often smaller and more indolent in nature, and therefore more challenging to identify precisely before surgery. In the recent decade, some researchers have recognized the emerging role of GLP-1R-based imaging in detecting occult insulinomas, using ^{111}Indium- or ^{99m}Tc-labelled scintigraphy and ^{68}Gallium-labelled positron emission tomography (PET)/CT. ^{111}Indium-DTPA-exendin-4 single photon emission computer tomography (SPECT)/CT was found to have a much higher sensitivity (95%) when compared to CT/MRI (47%) in an early study including 23 patients with histologically-confirmed insulinoma [15]. The most recent prospective study of 38 patients having histology-proven benign insulinoma with a median size of 1.3 cm showed that ^{68}Gallium-DOTA-exendin-4 PET/CT had a significantly higher accuracy and

sensitivity (94% and 95%, respectively) than ^{111}Indium-DOTA exendin-4 SPECT/CT (68% and 58%) and contrast-enhanced MRI (68% and 69%), influencing the surgical plan in 40% of patients [16]. The diagnostic superiority of GLP-1R PET/CT over SPECT/CT can be attributed to its greater spatial resolution, scanner sensitivity, and tumour-to-background ratio, as well as lower radiation dose and shorter acquisition time. However, these scans are often costly and not widely available, and the physiological high renal uptake may compromise the detection of lesions over the tail of pancreas which is in close proximity to the left kidney.

A prospective study compared the aforementioned imaging techniques and demonstrated that ^{68}Ga-NOTA-exendin-4 PET/CT correctly detected lesions in 42 of 43 patients with mostly benign insulinomas, resulting in a sensitivity of 98%, whereas ^{99m}Tc-labelled somatostatin receptor scintigraphy only yielded a sensitivity of less than 20%; the sensitivities of CT, MRI and EUS in this study were 74%, 56%, and 84% respectively [17]. ^{68}Ga-exendin-4 PET/CT scanning was also shown to be specifically beneficial in detecting MEN1-associated insulinomas [18]. However, GLP-1R and SSTR scans may be complementary to each other in tumour localization as all insulinomas are likely to express either one or both of these receptors, and the combination of these two peptide receptor imaging techniques could yield a sensitivity of 100% [12], facilitating a minimally invasive surgical approach.

Selective Intra-Arterial Calcium Stimulation with Hepatic Venous Sampling

Selective arteriography alone was formerly considered the gold-standard method, but has been superseded by selective arterial calcium stimulation with hepatic venous sampling. This has been proposed as the most sensitive preoperative localization technique. An angiogram is carried out by selective cannulation of the splenic (supplying the pancreatic body and tail), superior mesenteric (supplying the uncinate process) and gastroduodenal artery (supplying the head of the pancreas). Calcium gluconate, a potent β-cell secretagogue, is injected locally into each respective artery. Blood samples are taken from the right hepatic vein at 0, 30, 60, and 120-s post-injection to measure insulin. A twofold increase in insulin in the 30- and/or 60-s sample confirms the diagnosis [19]. This technique has a reported sensitivity of over 90% (range 87.5–100%) in the **regionalization** of pancreatic insulinomas, but is invasive and not routinely available. It is therefore only considered when there is strong clinical suspicion of an insulinoma, but diagnosis with other tests has proved elusive [20]. This technique also has the advantage that it confirms functionality of any lesion seen on cross-sectional imaging. However, it is likely that in due course functional radionuclide imaging will replace its use.

Management

Surgical Management

As most insulinomas are benign and solitary tumours, surgical resection remains the mainstay of treatment which offers a definitive cure together with the elimination of hormonal effects. The surgical decision depends upon patient's fitness, stage at presentation, tumour location, and limits of resection. After successful surgical removal, prognosis is good, with a 10-year survival of 88% [19]. Surgery may be laparoscopic or open and may involve enucleation, pancreatic resection, and metastasectomy. Enucleation is the preferred method of removal, ideally laparoscopically. Almost all insulinomas possess a pseudocapsule with a clear plane of dissection between the tumour and the surrounding soft pancreatic parenchyma. Enucleation is sufficient if the lesion is small, clearly localized before surgery, at or near the pancreatic surface, and easily defined intraoperatively [21]. Resection is indicated when the lesion is sizeable with considerable ductal or vascular invasion, or where malignancy is suspected with a hard, infiltrating tumour and puckering of the surrounding soft tissue, distal dilatation of the pancreatic duct, or lymph node involvement. Regional lymphadenectomy should be routinely carried out during pancreatectomy for pancreatic NETs of >2 cm due to the higher risk of lymph node metastases. Splenic preservation is ideal to minimize postoperative complications, but may not always be possible. Histological analysis is important to confirm the diagnosis of insulinoma, tumour grade and proliferation rate (Ki-67 index and/or mitotic count), and adequacy of resection.

Hepatic metastases indicate a poor prognosis. Malignant insulinomas with minimal extrahepatic disease and satisfactory performance status are managed with pancreatectomy, metastasectomy, and/or adjunctive treatment such as hepatic artery embolization, and/or chemotherapy.

With advances in laparoscopic techniques, both laparoscopic enucleation and resection have been performed successfully. Although the conversion rate to open surgery is 14%, this most probably represents the learning curve for this procedure [22]. Surgical treatment of insulinomas is effective and safe; reported success rates lie between 77% and 100% and mortality and morbidity rates are 2% and 26%, respectively, including postoperative infection (particularly post-splenectomy) and pancreatic abscess, pseudocysts, and fistula formation. After resection, the risk of recurrence is greater in patients with MEN 1 (21% at 20 years) than those without MEN 1 (5% at 10 years and 7% at 20 years).

For insulinomas as well as other functioning pancreatic NETs that are unsuitable for surgery, there is an increasing literature on the use of EUS-guided ablation techniques (such as ethanol or radiofrequency ablation) to reduce the symptoms due to hormone excess [23]. More data are certainly required to confirm their role in these unresectable cases.

Medical Management

Patients may be medically managed while awaiting surgery or medical therapies may be used alone for patients deemed unsuitable for surgery due to a high anaesthetic risk or inoperable disease related to local invasion or distant metastases, or for those with unsuccessful surgical outcomes such as persistent symptoms post-resection and failure to localize tumour in the theatre. The goal of medical management is to correct hypoglycaemia, prevent disease progression where possible, preserve quality of life, and avoid comorbidities from detrimental side effects.

Dietary and Lifestyle Advice

Patients are encouraged to eat frequent, small meals to avoid hypoglycaemia. Regular consumption of complex carbohydrates and foods with a low glycaemic index is recommended for the maintenance of blood glucose level while high glycaemic index foods are reserved for relieving acute, symptomatic hypoglycaemic attacks.

Guar gum has been shown to reduce insulin secretion in patients with insulinoma. Guar gum is an indigestible saccharide that delays gastric emptying and thus reduces the peak glucose load presented to the small intestine, which acts as a stimulus for insulin secretion, thereby slowing the rate of glucose absorption. Those with great fluctuations in glucose levels should refrain from driving, excessive alcohol intake, operation of dangerous machinery, and strenuous exercise.

Diazoxide

Diazoxide is an effective first-line medical treatment for alleviating hypoglycaemic episodes. It works directly on β-cells to suppress insulin secretion via activation of ATP-dependent potassium channels that leads to cell membrane hyperpolarization and a decrease in calcium influx. Significant side effects occur in 10–50% of patients, including oedema, weight gain, hirsutism, and hypokalaemia. Serious side effects, such as cardiomyopathy, cardiac arrhythmia, and myelosuppression, warrant close monitoring and cessation of therapy, but are rare. The usual initiating dose is 50 mg tds but dosing up to 100 mg tds may be used.

Somatostatin Analogues

In light of the impressive results from clinical trials such as *PROMID* and *CLARINET*, octreotide and lanreotide have become established systemic treatments for enteropancreatic NETs. These synthetic somatostatin analogues bind strongly to SSTR2, and to a lesser extent SSTR5, and are useful in controlling symptoms related to hormonal excess as well as tumour cell proliferation. Their therapeutic value is controversial in benign insulinoma since hypoglycaemia is usually manageable with diazoxide prior to surgical resection and there is a variable expression of SSTR2 on benign tumour cells. However, somatostatin analogues may be considered as first-line option in unresectable low-grade malignant insulinoma for tumour stabilization [24]. Although somatostatin may improve hypoglycaemia through inhibition of autonomous insulin secretion from tumour tissues, they may also paradoxically **worsen** hypoglycaemia by suppressing counter-regulatory hormones such as glucagon or growth hormone. Therefore, blood glucose levels should always be monitored closely upon treatment initiation preferably with administration of subcutaneous short-acting octreotide, and this should be initiated as an in-patient. Pasireotide, a novel somatostatin analogue which binds to SSTR 1, 2, 3, 5 with high affinity, can be potentially more effective in insulinoma due to its greater hyperglycaemic effects and the strong expression of SSTR5 in malignant insulinoma. Well-designed therapeutic trials are warranted to explore the antihormonal and antiproliferative effects of somatostatin analogues on insulinoma. Patients who are selected for these medications require positive somatostatin receptor scintigraphy. Side effects are generally tolerable, such as nausea, flatulence, diarrhoea, or steatorrhoea related to pancreatic enzyme insufficiency, fat-soluble vitamin deficiencies, and the development of gallstones.

Other older therapies include calcium antagonists such as verapamil and membrane stabilizers such as phenytoin, but these are rarely used nowadays.

Everolimus

Everolimus is registered in many countries for treating advanced, progressive, low, or intermediate grade non-functional NETs of gastrointestinal or lung origin. It is an oral mammalian/mechanistic target of rapamycin (mTOR) inhibitor which can potentially produce direct antitumour effects and impact on glucose homeostasis via growth factor signalling (cell growth, proliferation, and angiogenesis) and insulin-mediated pathways, thereby controlling both hypoglycaemic symptoms and tumour proliferation in malignant insulinoma. It probably works principally on modulating insulin release rather than blocking the action of insulin as it is ineffective in insulin-like growth factor 2 (IGF-2)-mediated hypoglycaemia [25]. A recent report described rapid changes in serum insulin and glucose levels following initiation, discontinuation, and re-administration of everolimus, with further *in-vitro* experiments confirming its direct inhibitory effect on insulin secretion independent of tumour regression [26]. The landmark study RADIANT-3 involving 410 patients with advanced progressive pancreatic NETs, including 24% with functioning lesions, showed that everolimus modestly prolonged progression-free survival (11.0 months with everolimus vs. 4.6 months with placebo) while stabilizing the disease in 73% of patients. Among patients with aggressive insulinoma and refractory hypoglycaemia despite multiple lines of treatment evaluated after receiving everolimus, immediate and substantial normalization of glucose levels and disease stabilization were achieved in most of them [27–28], and the median symptom-free survival with everolimus and placebo were 6.5 months and less than 1 month, respectively [27]. Stomatitis, skin rash, and diarrhoea are common side effects, whereas important adverse reactions include myelosuppression and pneumonitis.

Peptide Receptor Radionuclide Therapy (PRRT)

PRRT is a relatively new technique, which allows radiolabelled peptide hormone agonists, with ^{111}indium, ^{90}yttrium, and ^{177}lutetium being the most popular radionuclides, to specifically bind to somatostatin receptors on the NET cell surface, thereby delivering focused radioactivity. There is a growing body of evidence, including the first phase-III randomized controlled *NETTER-1* trial, that clearly demonstrate the efficacy and durability of PRRT in terms of improvement in symptom control and quality of life as well as tumour response and disease stabilization in well-differentiated midgut NETs, with median progression-free survival reaching 3–4 years and an objective response rate of around 30% [29]. Several case reports also illustrated its potential therapeutic value in malignant insulinoma with uncontrolled hypoglycaemia despite the use of multiple treatment modalities. In a case series of five patients with metastatic insulinoma requiring continuous glucose infusion, all of them achieved stable disease for a mean period of 27 months without any recurrence of hypoglycaemic episodes after PRRT despite tumour progression in three patients, suggesting additional mechanisms for glycaemic control (e.g. induction of tumour de-differentiation) [30]. PRRT should be considered in well-differentiated but inoperable cases with multiorgan involvement who have failed first-line therapy. Sufficient tumour uptake greater than the liver on somatostatin receptor imaging is mandatory. Such therapy is usually well tolerated, with ^{177}Lutetium-DOTATATE, the first PRRT agent that has been recently approved for the treatment of somatostatin receptor-positive enteropancreatic NETs, yielding a more favourable toxicity profile. Theoretically, it could cause transient hypoglycaemia due to the release of peptide hormones during rapid tumour necrosis, hence careful monitoring of glucose

level together with adjustment of concomitant therapies is necessary. Other acute side effects include nausea and vomiting, while myelotoxicity and renal impairment are rare but important complications in the longer term.

Chemotherapy

Specific data on the chemotherapy in the treatment of insulinoma are lacking, while regimens containing streptozotocin and 5-fluorouracil (5-FU) have been extensively studied in pancreatic NETs with promising results. Although these potent chemotherapeutic agents carry a considerable risk of drug toxicity, they may be useful in the palliative setting for rapidly progressive or highly proliferative malignant insulinoma. Capecitabine generally replaces 5-FU nowadays, while temozolomide can be tried in place of streptozotocin.

Liver-Directed Therapies

Liver-debulking surgery should always be considered in patients without extrahepatic disease if the majority of the liver metastases can be resected. For patients with high tumour burden in the liver who are not suitable candidates for surgical treatment, locoregional therapies such as transarterial bland embolization, chemoembolization, or radioembolization, and radiofrequency ablation may be reasonable alternatives based on individual tumour characteristics. For instance, hepatic artery embolization to deprive tumour tissues of their blood supply may be an option in a patient with multiple vascularized liver metastases and refractory hypoglycaemia but acceptable residual liver function and a patent portal vein.

Other Causes of Hypoglycaemia

Glucose is a critical fuel for all body organs and especially the brain, and therefore it is imperative to recognize the possibility of significant hypoglycaemia and reverse it promptly, and treat the underlying disease. Hypoglycaemia is usually a consequence of inadequate glucose intake or production relative to glucose utilization from the circulation (Table 6.7.2). Iatrogenic or factitious hypoglycaemia from exogenous administration of antidiabetic agents (insulin or insulin secretagogues) is the most common cause of hypoglycaemia among both diabetic and non-diabetic individuals. In a hospital setting, hypoglycaemia due to critical illnesses such as major organ failure, sepsis and malignancy are frequently encountered, particularly when there is reduced food intake or concomitant use of glucose-lowering medications.

Besides insulinoma, endogenous hyperinsulinaemic hypoglycaemia may also be due to β-cell stimulation by autoantibodies or β-cell hyperplasia associated with nesidioblastosis. These diseases typically demonstrate inappropriately raised insulin and C-peptide levels in response to hypoglycaemia [31]. Hypoglycaemia can also be caused by hypersecretion of abnormal IGF-2 from non-islet cell tumours.

Iatrogenic or Factitious Hypoglycaemia

Hypoglycaemia is a well-recognized side effect of insulin treatment and may account for 4% of deaths in patients with type 1 diabetes as they are vulnerable due to hypoglycaemic unawareness and impaired catecholamine responses. With the advent of insulin pump therapy, the frequency of severe hypoglycaemia in type 1 diabetes was shown to be markedly reduced (for continuous subcutaneous insulin infusion versus multidose insulin injections, a mean 2.9-fold reduction for randomized controlled trials, 4.3-fold for before/after studies, and 4.2-fold for all studies) despite superior glycaemic control [32]. Examples of insulin secretagogues include sulphonylureas and meglitinides, but these have become less popular in the treatment of type 2 diabetes considering the availability of other newer agents that are less likely to cause hypoglycaemia. Elderly patients with cognitive or renal impairment and those using multiple or more potent agents are more prone to hypoglycaemia.

Factitious hypoglycaemia in diabetic patients can be secondary to accidental overdose or inappropriate self-manipulation of these medications. As for non-diabetic individuals, unexpected administration of hypoglycaemic agents as in dispensing errors, and in situations where these drugs are easily accessible, surreptitious, or even malicious, hypoglycaemia may not be evident upon initial evaluation. Meticulous history taking together with thorough review of drug records is thus essential. Patients receiving insulin injections typically have very high blood insulin but low C-peptide levels. However, the biochemical pictures in the use of oral hypoglycaemic agents and insulinoma are similar: elevated blood insulin

Table 6.7.2 Causes of hypoglycaemia in adults

Endogenous hyperinsulinism	Hormonal deficiencies	Acute illnesses	Non-islet cell tumour hypoglycaemia	Drugs	
				Classification	Examples
Insulinoma Functional β-cell disorders (CHH, NIPHS, PGBH) Autoimmune hypoglycaemia	Cortisol Glucagon and epinephrine (in type 1 & advanced type 2 diabetes) Growth hormone	Liver, renal, or heart failure Sepsis Starvation or malnutrition	Mesenchymal tumours Malignant epithelial tumours (e.g. hepatocellular carcinoma, renal cell carcinoma & gastrointestinal tumours)	Antidiabetic (Iatrogenic or factitious hypoglycaemia)	Insulin Insulin secretagogues
				Alcohol	Salicylates
				Anti-inflammatory	Indomethacin Quinolones
				Antimicrobial	Quinine Chloroquine Pentamidine Trimethoprim-sulfamethoxazole
				Antihypertensive	ACE inhibitors β-adrenoceptor blockers
				Others	Disopyramide Haloperidol

and C-peptide levels. Demonstration of detectable sulphonylurea level in the serum and/or urine is therefore a crucial clue in such a scenario, assuming the relevant agent can be assayed.

Alcohol and Other Drugs

Alcohol ingestion influences glucose metabolism in several ways. Ethanol inhibits gluconeogenesis (but has no effect on hepatic glycogenolysis) and the counter-regulatory hormones of insulin—glucagon, adrenaline, growth hormone, and cortisol. Furthermore, patients may not be aware of and compensate for hypoglycaemic symptoms during mild alcohol intoxication. Classically, they present with hypoglycaemia with a normal ethanol level a few days after binge drinking with reduced food intake, when the hepatic glycogen store is depleted, and gluconeogenesis becomes the main source of fuel.

Some other medications are reported to cause hypoglycaemia (Table 6.7.2) [31, 33].

Post-Prandial Syndrome

Postprandial syndrome (formerly called 'reactive hypoglycaemia') refers to a functional and idiopathic condition, observed mostly in young and female patients, with mild symptoms of sympathetic overactivity (2–4 hr after a meal) caused by hyperinsulinaemia in response to a high carbohydrate intake. Although some have proposed the use of an oral glucose tolerance test (OGTT) for diagnosis, the diagnostic glucose value is difficult to define due to a considerable overlap between normal subjects and patients, and there is lack of correlation between symptoms and low glucose levels during the OGTT. Indeed, some have questioned the existence of the syndrome. Nonetheless, the symptoms are usually self-limiting and should resolve after dietary modification with frequent, small meals.

Primary Functional β-Cell Disorders (Nesidioblastosis)

Congenital hyperinsulinaemic hypoglycaemia (CHH) is characterized by inappropriate insulin secretion and nesidioblastosis with islet cell hypertrophy ± hyperplasia in infants and children due to genetic defects involved in the regulation of insulin release from β-cells (channelopathies or metabolopathies) [34]. While some forms tend to abate later in life, the more severe ones, usually caused by mutations in *ABCC8* and *KCNJ11* genes that encode the sulfonylurea receptor (SUR1) and inwardly rectifying potassium channel (Kir6.2) proteins respectively, are invariably resistant to medication and often require operative intervention. Non-insulinoma pancreatogenous hypoglycaemic syndrome (NIPHS) has a similar histopathology causing β-cell hyperfunction and endogenous hyperinsulinaemic hypoglycaemia in adults with a male predominance, but its genetic mechanisms are poorly understood [34–35].

These rare functional β-cell disorders are biochemically indistinguishable from insulinomas, but in contrast to insulinomas they typically give rise to post-prandial hypoglycaemia and negative radiological localization studies [31]. Although both NIPHS and insulinoma may produce positive responses to the selective intra-arterial calcium stimulation test, it was observed that the maximum insulin concentration as well as the extent of insulin increase from baseline during hepatic venous sampling were significantly higher in insulinoma [36]. Medical treatment with diazoxide and/or octreotide is preferred for both CHH and NIPHS. Surgery is only considered when medical options fail and should be restricted to removal of mostly 80% of the pancreas, which can be guided by the selective intra-arterial calcium stimulation test, and usually results in symptomatic improvement.

Post-Gastric Bypass Hypoglycaemia (PGBH)

Some patients develop postprandial hypoglycaemia after Roux-en-Y gastric bypass surgery due to nesidioblastosis, which is possibly a result of dysregulation of gut hormones and insulin after gastroduodenal surgery and significant weight loss, but the exact mechanisms are uncertain [37, 38]. It may also occur following the simpler sleeve gastrectomy. Unlike NIPHS, PGBH has a female predominance and can coexist with insulinoma. Most patients improve with frequent, small meals containing slowly absorbed, complex carbohydrates. Medical therapies such as diazoxide, α-glucosidase inhibitors and octreotide can also be effective. We have reported one patient with the 'dumping syndrome' due to over-rapid transport of glucose to the small intestine and consequent hyperinsulinaemia following oesophagectomy for oesophageal carcinoma: the assumed excessive responses to GLP-1 and GIP could be inhibited by pasireotide, and the patient remains well on this therapy [39]. Partial pancreatectomy is no longer recommended after bypass surgery or sleeve gastrectomy as many patients experienced recurrent hypoglycaemia after the procedure [38].

Non-Islet Cell Tumour Hypoglycaemia (NICTH)

NICTH is a rare paraneoplastic phenomenon characterized by the excessive secretion of aberrant pro-IGF-2 related molecules (or 'big IGF-2'), which are incompletely processed precursors of IGF-2 with potent insulin-like effects. These precursors cannot readily bind with the usual IGF-binding proteins such as BP3 and ALS, and thus the fragment is able to cross blood vessels and gain access to the insulin receptor of target tissues causing hypoglycaemia. Large mesenchymal tumours (e.g. solitary fibrous tumour, haemangiopericytoma and mesothelioma) located in retroperitoneal, intra-abdominal, or intra-thoracic regions are the most common tumours causing this condition. It is occasionally caused by epithelial tumours such as hepatocellular carcinoma, renal cell carcinoma and tumours arising from the gastrointestinal tract [40].

These big IGF-2 molecules inhibit the secretion of both insulin and growth hormone via negative feedback mechanisms which in turn reduces the serum concentrations of IGF-1 and IGF-binding protein. Therefore, this further decreases binding of IGF-2 and contributes to a high proportion of free IGF-2 [40]. Circulating total IGF-2 level may be raised or normal depending on the assay, as these IGF-2 variants can only be detected by special chromatography-based methods [41]. Hence, an IGF-2:IGF-1 ratio of greater than 3:1 (often >10:1) may be helpful in diagnosing NICTH. Apart from IGF-2 mediated hypoglycaemia, other possible mechanisms for tumour-related hypoglycaemia include extensive tumour load in the liver or adrenal glands, emaciation with depleted glycogen reserve, and insulin secretion from the tumour.

Debulking surgery is the treatment of choice in reversing the metabolic derangements associated with NICTH. However, surgery may not be always immediately feasible due to various patient and disease factors. Glucagon infusion is useful in short-term alleviation of hypoglycaemia. Glucocorticoids are known to be an effective

maintenance therapy as they stimulate hepatic gluconeogenesis, suppress the production, and increase the clearance of IGF-2, but their chronic use may be limited by systemic adverse effects. Recombinant human growth hormone can be used alone at high doses, or preferably in combination with glucocorticoids [40]. Diazoxide and octreotide are not useful.

Autoimmune Hypoglycaemia

Insulin autoimmune syndrome (IAS) is a rare disease which was first reported and is most prevalent in Japan, and is defined as endogenous hyperinsulinaemic hypoglycaemia associated with the presence of insulin autoantibodies (IAA) in patients without a prior history of insulin treatment. Sometimes it is associated with autoimmune diseases such as Graves' disease or rheumatoid arthritis, or drugs containing a sulfhydryl group such as methimazole, carbimazole, penicillamine, tiopronin, and glutathione [42]. The suggested mechanisms for hypoglycaemia in IAS are the release of free insulin from the large buffering reservoir of insulin-IAA complexes, which can dissociate unpredictably, and potentiation of insulin bioactivity [43]. Symptoms usually occur in the late postprandial rather than the fasting period because the initial binding of insulin to IAA after food ingestion further stimulates the secretion of free insulin, the insulin-IAA complexes then dissociate spontaneously several hours after eating to release more free insulin leading to hypoglycaemia. In patients with autoimmune disease, hypoglycaemia caused by autoantibodies against insulin receptors which activate insulin receptors regardless of circulating insulin concentrations have also been described. An increase in blood insulin with low C-peptide levels in these rare cases is probably related to the decrease in insulin clearance.

Insulin-IAA complexes often interfere with insulin assays to produce spuriously high insulin levels. Total insulin levels in IAS can be up to 100-fold higher than free insulin. Laboratory confirmation can be obtained by directly measuring IAA in the blood or using polyethylene glycol immunoprecipitation and gel filtration chromatography to confirm the presence of autoantibodies [44]. In contrast to hyperinsulinaemia due to exogenous insulin or insulin receptor autoantibodies, the C-peptide level is not suppressed.

Treatment of IAS consists of small frequent meals with low contents of simple sugars, discontinuation of culprit drugs, diazoxide, plasmapheresis, and corticosteroid therapy. Plasmapheresis removes IAA mechanically, whereas steroids decrease IAA production and alters their properties (binding and dissociation capacities). Most cases resolve spontaneously within one year but may recur if a culprit drug is re-administered.

Critical Illnesses

Besides those with diabetes mellitus receiving hypoglycaemic agents, some hospitalized patients are susceptible to hypoglycaemia as a result of multiple factors such as failure of major organs involved in gluconeogenesis, sepsis, and chronic malnutrition. Regular blood glucose monitoring is essential in these patients who may be too weak to report their symptoms.

Major Organ Failure

The liver is largely responsible for gluconeogenesis and glycogenolysis. However, due to a large hepatic glycogen reserve, hypoglycaemia only ensues after hepatectomy or liver failure caused by extensive and rapid hepatic damage (such as fulminant viral or drug-induced hepatitis) in addition to defective peripheral glucose production. Key mechanisms leading to hypoglycaemia in renal failure include impaired renal gluconeogenesis, reduced insulin clearance, and coexisting heart failure, while hepatic congestion and hypoxia together with diminished gluconeogenesis may be responsible for hypoglycaemia associated with severe cardiac failure. Polypharmacy, sepsis, and poor nutrition may also contribute to hypoglycaemia in these conditions.

Sepsis

Patients with severe infection are prone to hypoglycaemia, which is known to increase mortality because of decreased hepatic glucose output, increased glucose utilization, and associated organ failure. Therefore, tight glycaemic control with the use of intensive insulin therapy in severe sepsis is not currently advocated [45].

Hormone Deficiencies

The insulin-antagonizing hormones play an extremely important role in maintaining glucose homeostasis. Subjects may be more vulnerable to hypoglycaemia when glucagon and catecholamines are insufficient in the context of insulin-deficient diabetes [31], but it is generally not a feature of isolated catecholamine deficiency resulting from bilateral adrenalectomy with glucocorticoid replacement or pharmacological catecholamine blockade, although postoperative hypoglycaemia due to a dramatic fall in catecholamine levels with rebound hyperinsulinism has been observed after resection of phaeochromocytoma. In the presence of intact catecholamine responses, glucagon deficiency would not be expected to cause hypoglycaemia.

Glucocorticoids have profound effects on glucose metabolism. Inadequate endogenous glucose production, escalated glucose disposal and oxidation, increased insulin sensitivity in the liver and peripheral tissues, and anorexia are the likely explanations for hypoglycaemia in adrenal insufficiency.

Hypoglycaemia is not a feature of growth hormone deficiency in adults. In contrast, young children are more susceptible to hypoglycaemia especially in cases of hypopituitarism and prolonged fasting, with diminished hepatic glucose output, depleted glycogen stores, increased glucose oxidation and consumption, and reduced fat mobilization, and lipolysis as contributing mechanisms.

Inborn Errors of Metabolism (IEM)

In rare occasions, fasting hypoglycaemia in adults can be a presenting feature of an IEM, such as glycogen storage disease (particularly Type I and III), defect of fatty acid oxidation and ketogenesis, or gluconeogenesis disorders [46]. These conditions should therefore be considered in patients having other cardinal features (e.g. lactic acidosis, hepatomegaly and myopathy, early onset of symptoms, a positive family history, and otherwise unexplained hypoglycaemia).

ACKNOWLEDGEMENTS

We would like to express our gratitude and appreciation to Puja Mehta and Jeannie Todd, authors of the chapter in the second edition, for their valuable and inspirational contributions.

REFERENCES

1. Service FJ, McMahon MM, O'Brien PC, Ballard DJ. Functioning insulinoma incidence, recurrence, and long-term survival of patients: a 60-year study. *Mayo Clin Proc*, 1991; **66**: 711–19.
2. Placzkowski KA, Vella A, Thompson GB, *et al.* Secular trends in the presentation and management of functioning insulinoma at the Mayo Clinic, 1987–2007. *J Clin Endocrinol Metab,* 2009; **94**: 1069–73.
3. Service FJ, Natt N. The prolonged fast. *J Clin Endocrinol Metab*, 2000; **85**: 3973–4.
4. Kar P, Price P, Sawers S, Bhattacharya S, Reznek RH, Grossman AB. Insulinomas may present with normoglycaemia after prolonged fasting but glucose-stimulated hypoglycaemia. *J Clin Endocrin Metab*, 2006; **91**: 4733–6.
5. Dromain C, Déandréis D, Scoazec JY, *et al.* Imaging of neuroendocrine tumours of the pancreas. *Diagn Interv Imaging*, 2016; **97**: 1241–57.
6. Druce MR, Muthuppalaniappan VM, O'Leary B, *et al.* Diagnosis and localisation of insulinoma: the value of modern MRI in conjunction with calcium stimulation catheterization. *Eur J Endocrinol*, 2010; **162**: 971–8.
7. Zhu L, Xue H, Sun Z, *et al.* Prospective comparison of biphasic contrast-enhanced CT, volume perfusion CT, and 3 Tesla MRI with diffusion-weighted imaging for insulinoma detection. *J Magn Reson Imaging,* 2017; **46**: 1648–55.
8. Galiber AK, Reading CC, Charboneau JW, *et al.* Localization of pancreatic insulinoma: comparison of pre- and intraoperative US with CT and angiography. *Radiology*, 1988; **166**: 405–8.
9. Puli SR, Kalva N, Bechtold ML, *et al.* Diagnostic accuracy of endoscopic ultrasound in pancreatic neuroendocrine tumours: a systematic review and meta-analysis. *World J Gastroenterol,* 2013; **19**: 3678–84.
10. Sundin A, Vullierme MP, Kaltsas G, Plöckinger U. ENETS consensus guidelines for the standards of care in neuroendocrine tumours: radiological examinations. *Neuroendocrinology,* 2009; **90**: 167–83.
11. Vezzosi D, Bennet A, Rochaix P, *et al.* Octreotide in insulinoma patients: efficacy on hypoglycaemia, relationships with octreoscan scintigraphy and immunostaining with anti-sst2A and anti-sst5 antibodies. *Eur J Endocrinol,* 2005; **152**: 757–67.
12. Wild D, Christ E, Caplin ME, *et al.* Glucagon-like peptide-1 versus somatostatin receptor targeting reveals 2 distinct forms of malignant insulinomas. *J Nucl Med*, 2011; **52**: 1073–8.
13. Nockel P, Babic B, Millo C, *et al.* Localisation of insulinoma using 68Ga-DOTATATE PET/CT scan. *J Clin Endocrinol Metab,* 2017; **102**: 195–9.
14. Körner M, Christ E, Wild D, Reubi JC. Glucagon-like peptide-1 receptor overexpression in cancer and its impact on clinical applications. *Front Endocrinol (Lausanne),* 2012; **3**: 158.
15. Christ E, Wild D, Ederer S, *et al.* Glucagon-like peptide-1 receptor imaging for the localization of insulinomas: a prospective multicentre imaging study. *Lancet Diabetes Endocrinol*, 2013; **1**: 115–22.
16. Antwi K, Fani M, Heye T, *et al.* Comparison of glucagon-like peptide-1 receptor (GLP-1R) PET/CT, SPECT/CT and 3T MRI for the localisation of occult insulinomas: evaluation of diagnostic accuracy in a prospective crossover imaging study. *Eur J Nucl Med Mol Imaging*, 2018; **45**: 2318–27.
17. Luo Y, Pan Q, Yao S, *et al.* Glucagon-like peptide-1 receptor PET/CT with ^{68}Ga-NOTA-exendin-4 for detecting localized insulinoma: a prospective cohort study. *J Nucl Med*, 2016; **57**: 715–20.
18. Antwi K, Nicolas G, Fani M, *et al.* ^{68}Ga-exendin-4 PET/CT scanning specifically detects insulinomas in patients suffering from hypoglycaemia due to multiple endocrine neoplasia type 1. *J Clin Endocrinol Metab*, 2019; **104**: 5843–52.
19. Tucker ON, Crotty PL, Conlon KC. The management of insulinoma. *Br J Surg*, 2006; **93**: 264–75.
20. Morganstein DL, Lewis DH, Jackson J, *et al.* The role of arterial stimulation and simultaneous venous sampling in addition to cross-sectional imaging for localisation of biochemically proven insulinomas. *Eur Radiol*, 2009; **19**: 2467–73.
21. Ramage JK, Ahmed A, Ardill J, *et al.* Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours. *Gut*, 2005; **54**: iv1–16.
22. Isla AM, Arbuckle JD, Kekis PB, *et al.* Laparoscopic management of insulinomas. *Br J Surg*, 2009; **96**: 185–90.
23. Signoretti M, Valente R, Repici A, Delle Fave G, Capurso G, Carrara S. Endoscopy-guided ablation of pancreatic lesions: technical possibilities and clinical outlook. *World J Gastrointest Endosc*, 2017; **9**: 41–54.
24. Baudin E, Caron P, Lombard-Bohas C, *et al.* Malignant insulinoma: recommendations for characterization and treatment. *Ann Endocrinol (Paris)*, 2013; **74**: 523–33.
25. Korevaar T, Ragazzoni F, Weaver A, Karavitaki N, Grossman AB. IGF2-induced hypoglycaemia unresponsive to everolimus. *QJM*, 2014; **107**: 297–300.
26. Suzuki L, Miyatsuka T, Himuro M, *et al.* Everolimus directly suppresses insulin secretion independently of cell growth inhibition. *J Endocr Soc*, 2018; **2**: 589–96.
27. Bernard V, Lombard-Bohas C, Taquet MC, *et al.* Efficacy of everolimus in patients with metastatic insulinoma and refractory hypoglycaemia. *Eur J Endocrinol,* 2013; **168**: 665–74.
28. Kulke MH, Bergsland EK, Yao JC. Glycaemic control in patients with insulinoma treated with everolimus. *N Engl J Med*, 2009; **360**: 195–7.
29. Strosberg J, El-Haddad G, Wolin E, *et al.* Phase 3 trial of ^{177}Lu-DOTATATE for midgut neuroendocrine tumours. *N Engl J Med*, 2017; **376**: 125–35.
30. van Schaik E, van Vliet EI, Feelders RA, *et al.* Improved control of severe hypoglycaemia in patients with malignant insulinomas by peptide receptor radionuclide therapy. *J Clin Endocrinol Metab,* 2011; **96**: 3381–9.
31. Cryer PE, Axelrod L, Grossman AB, *et al.* Evaluation and management of adult hypoglycaemic disorders: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab,* 2009; **94**: 709–28.
32. Pickup JC, Sutton AJ. Severe hypoglycaemia and glycaemic control in type 1 diabetes: meta-analysis of multiple daily insulin injections compared with continuous subcutaneous insulin infusion. *Diabet Med*, 2008; **25**: 765–74.
33. Vue MH, Setter SM. Drug-induced glucose alterations part 1: drug-induced hypoglycaemia. *Diabetes Spectr*, 2011; **24**: 171–7.
34. Nessa A, Rahman SA, Hussain K. Hyperinsulinaemic hypoglycaemia—the molecular mechanisms. *Front Endocrinol (Lausanne)*, 2016; **7**: 29.
35. Service FJ, Natt N, Thompson GB, *et al.* Noninsulinoma pancreatogenous hypoglycaemia: a novel syndrome of hyperinsulinaemic hypoglycaemia in adults independent of mutations in Kir6.2 and SUR1 genes. *J Clin Endocrinol Metab*, 1999; **84**: 1582–9.

36. Thompson SM, Vella A, Thompson GB, *et al*. Selective arterial calcium stimulation with hepatic venous sampling differentiates insulinoma from nesidioblastosis. *J Clin Endocrinol Metab*, 2015; **100**: 4189–97.
37. Service GJ, Thompson GB, Service FJ, Andrews JC, Collazo-Clavell ML, Lloyd RV. Hyperinsulinaemic hypoglycaemia with nesidioblastosis after gastric-bypass surgery. *N Engl J Med,* 2005; **353**: 249–54.
38. Shantavasinkul PC, Torquati A, Corsino L. Post-gastric bypass hypoglycaemia: a review. *Clin Endocrinol (Oxf)*, 2016; **85**: 3–9.
39. Mohammadi A, Sulaiman R, Grossman AB. Pasireotide and octreotide in the treatment of severe late dumping syndrome. *Clin Case Rep*, 2017; **5**: 1608–11.
40. Bodnar TW, Acevedo MJ, Pietropaolo M. Management of non-islet-cell tumour hypoglycaemia: a clinical review. *J Clin Endocrinol Metab*, 2014; **99**: 713–22.
41. Miraki-Moud F, Grossman AB, Besser M, Monson JP, Camacho-Hubner C. A rapid method for analysing serum pro-insulin-like growth factor-II (pro-IGF-II) in patients with non-islet cell tumor hypoglycaemia. *J Clin Endocrinol Metab*, 2005; **90**: 3819–23.
42. Uchigata Y, Eguchi Y, Takayama-Hasumi S, Omori Y. Insulin autoimmune syndrome (Hirata disease): clinical features and epidemiology in Japan. *Diabetes Res Clin Pract*, 1994; **22**: 89–94.
43. Taylor SI, Barbetti F, Accili D, Roth J, Gorden P. Syndromes of autoimmunity and hypoglycaemia. Autoantibodies directed against insulin and its receptor. *Endocrinol Metab Clin North Am,* 1989; **18**: 123–43.
44. Church D, Cardoso L, Bradbury S, *et al*. Diagnosis of insulin autoimmune syndrome using polyethylene glycol precipitation and gel filtration chromatography with ex vivo insulin exchange. *Clin Endocrinol (Oxf)*, 2017; **86**: 347–53.
45. Ling Y, Li X, Gao X. Intensive versus conventional glucose control in critically ill patients: a meta-analysis of randomized controlled trials. *Eur J Intern Med*, 2012; **23**: 564–74.
46. Douillard C, Mention K, Dobbelaere D, Wemeau JL, Saudubray JM, Vantyghem MC. Hypoglycaemia related to inherited metabolic diseases in adults. *Orphanet J Rare Dis*, 2012: **7**: 26.

6.8

Glucagonoma

Karim Meeran

Introduction

Glucagon is a peptide formed by alternative splicing of the pre-pro glucagon gene. Alpha cells of the islets of Langerhans secrete glucagon in response to hypoglycaemia.

Glucagonomas are neuroendocrine tumours arising from the alpha cells of the islets of Langerhans, which result in excessive secretion of glucagon and peptides derived from preproglucagon. Post-translational modification of proglucagon is tissue specific and results in various glucagon peptides [1]. It is the ratio of insulin to glucagon that controls the balance of gluconeogenesis and glycogenolysis in the liver. Glucagon stimulates hepatic gluconeogenesis and inhibits both glycolysis and glycogen synthesis. It increases production of free fatty acids from triglyceride breakdown by activating hormone-sensitive lipase; these undergo fatty oxidation in the liver via acetyl CoA, forming ketone bodies. The increase in free fatty acids from lipolysis inhibits hepatic lipogenesis. Glucagon also increases muscle proteolysis, resulting in an increase in amino acid supply to the liver.

Epidemiology

The annual incidence of glucagonomas is estimated to be only 1 per 20 million [2, 3] so that we would expect three new cases in the United Kingdom each year. The majority of glucagonomas are sporadic, with only 3% associated with multiple endocrine neoplasia type 1 (MEN 1) [4]. The current data in the United Kingdom is either based upon single case reports or a few small series of cases. Table 6.8.1 illustrates the clinical features of 192 patients between 1998 and 2016 [2] in China. The median age of presentation is 52 years.

Clinical Features

A series of cases of glucagon-secreting pancreatic tumours associated with necrolytic migratory erythema, weight loss, diabetes mellitus, and stomatitis was described in 1974 when the term glucagonoma syndrome was first coined [5]. Not all glucagonomas are symptomatic and they may be identified solely through screening of patients with MEN 1. The nutrient deficiencies arising due to hyperglucagonaemia and the secretion from the tumour of glucagon-like peptides 1 and 2, as well as cosecretion of other hormones such as pancreatic polypeptide, give rise to a spectrum of clinical features [6].

Necrolytic Migratory Erythema (NME)

NME is characterized by erythematous, well-demarcated plaques that are pruritic and painful, and often involve the intertriginous areas, perineum, and buttocks (Figure 6.8.1). This painful pruritic rash is a typical feature of the glucagonoma syndrome and one of the most common presenting signs, occurring in 82% of patients [3, 7]. Suspicion of the diagnosis by dermatologists, who may be the first people to see these patients has been the most common route of referral. An added clue is recent onset diabetes. Although characteristic for the glucagonoma syndrome, NME is not pathognomonic. The initial lesions of NME are erythematous plaques, which may be associated with bullae. These lesions form erosions and crusts, which eventually heal to leave central areas of hyperpigmentation and induration. The lesions demonstrate the Koebner phenomenon, i.e. occurring at sites of trauma [5]. Skin biopsy histology reveals necrolysis of the upper dermis and vacuolization of keratinocytes [8]. The occurrence of NME does not correlate with metastases and has been noted in 60% of patients with benign glucagonoma.

The exact pathogenesis of NME remains unclear. There have been several postulated theories; the condition appears to be a multifactorial disease caused by a combination of zinc, amino acid, and fatty acid deficiencies [9]. The glucagonoma syndrome shares a number of clinical features with vitamin B_2, B_3, B_6, and B_{12} deficiency. Indeed, vitamin B deficiency may arise as a result of hyperglucagonaemia

Table 6.8.1 Presenting clinical features in 192 patients with glucagonoma (1998–2016)

Clinical feature	Males (%) N = 85	Females (%) N = 107	Total (%) N = 192
Necrolytic migratory erythema	64 (75.3)	88 (82)	152 (79)
Diabetes mellitus or IFG	51 (60)	69 (64)	120 (62.5)
Weight loss	50 (58.8)	54 (50.5)	104 (54.2)
Anaemia	34 (40)	50 (46.7)	84 (43.8)
Angular chelitis	8 (9.4)	17 (15.9)	25 (13)
Glossitis	16 (18.8)	28 (26.2)	44 (23)
Diarrhoea	11 (12.9)	8 (7.5)	19 (10)
Psychiatric symptoms	9 (10.6)	8 (7.5)	17 (9)
Thrombosis	4 (4.7)	5 (4.7)	9 (4.6)

[6]. NME has been reported following intravenous glucagon treatment, suggesting that NME may be a direct consequence of glucagon action on the skin [10, 11]. NME has also been associated with conditions other than glucagonoma (e.g. coeliac disease and cirrhosis); both of these conditions may have raised glucagon or glucagon-like peptides levels [7].

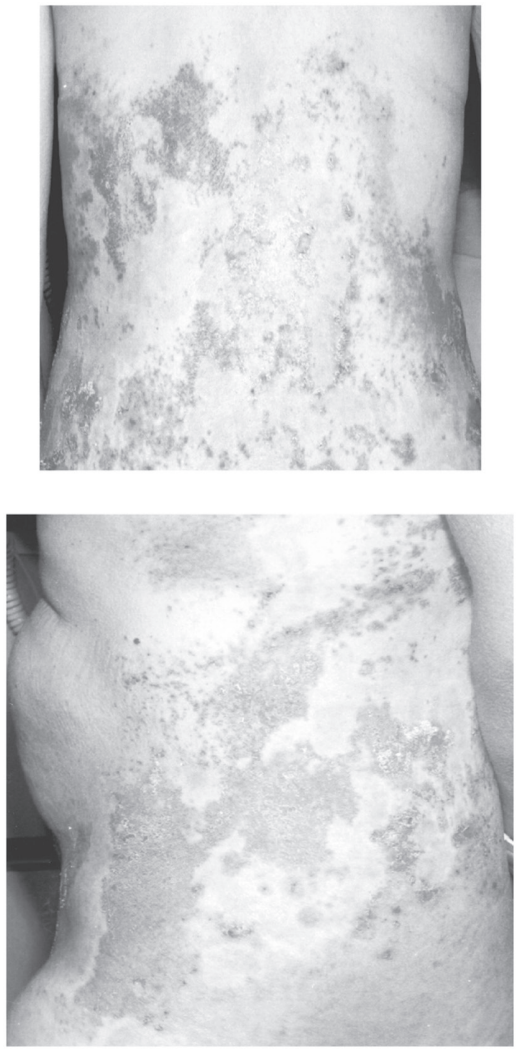


Figure 6.8.1 Necrolytic migratory erythema on the back and trunk of patient with malignant glucagonoma. (See also Plate 36.)

Diabetes Mellitus

Diabetes mellitus is common in sporadic glucagonomas, occurring in 55% of patients at presentation and eventually developing in 75% of cases. Of these three-quarters require insulin therapy [12]. Although rare, diabetic ketoacidosis has been reported [13].

Other Clinical Features

A review from the Mayo clinic of 21 patients with glucagonoma [12] suggested that weight loss or cachexia is a common presenting complaint, occurring in 71% of patients with metastatic glucagonoma. Similar rates of weight loss have been noted in the Hammersmith series [7] with 72% in patients with metastases and 40% with local disease. Normocytic normochromic anaemia was noted in approximately a third of patients and is probably the result of direct bone marrow suppression by glucagon [7]. Diarrhoea occurs in approximately one-fifth of patients with glucagonoma [12]. Of these, half also have elevated gastrin and pancreatic polypeptide levels [7]. Involvement of the mucous membranes may lead to the development of stomatitis, glossitis, and chelitis in a third of cases. Psychiatric symptoms occur in 20% of patients and may vary from depression to paranoid delusions. Thromboembolism is a major source of morbidity and mortality in the glucagonoma syndrome and occurs in up to 11% of cases [7] and may account for up to 50% of deaths [14].

Metastases and Site of Primary Tumour

Over 80% of patients with sporadic tumours have metastases at presentation [7, 12]. Hepatic metastases usually involve both lobes of the liver and are multiple in two-thirds of cases; of the single hepatic metastases 75% occur in the right lobe. Of primary tumours, 41% are confined to the tail of the pancreas, 14% involve the head and body, 14% occur in the head alone, and 9% in the body alone. The primary site of glucagonomas is the pancreas in 100% of cases. Sensitive imaging modalities and hepatic angiography may allow an increased detection of the primary tumour site [7].

Investigations

Biochemistry

Because of the rarity of glucagonomas, it is very important to minimize false positives by not screening patients who have a low prior probability of having a glucagonoma. For example patients with diabetes alone are very unlikely to have a glucagonoma, so this is a group who should not be screened, unless one is measuring glucagon for other reasons. Raised fasting plasma glucagon immunoreactivity is the basis for diagnosis. The reference range at the Hammersmith Clinical Chemistry Gut Hormone Laboratory is fasting plasma glucagon level below 50 pmol/L. False-positive results may occur due to other causes of a raised fasting plasma glucagon such as renal or hepatic failure, drugs, and prolonged fasting [15]. Plasma glucagon levels may be elevated to various degrees ranging from only 1.5 to 150 times the upper limit of normal [7]. Thus, if clinically supported, a marginally elevated plasma glucagon level may still be suspicious [16].

A raised fasting plasma gastrin level is noted in one-fifth of patients at presentation and may be associated with the Zollinger–Ellison syndrome [17]. Plasma gastrin levels should be monitored regularly since they may rise up to 6 years after initial diagnosis. Other hormones may also be elevated, for example insulin, 5-hydroxyindoleacetic acid, human pancreatic polypeptide, chromogranin, and vasoactive intestinal peptide. It is therefore important an annual assessment of fasting gut hormone profile should be undertaken [7].

Biochemical investigations may reveal hypoproteinaemia, hypoalbuminaemia, and hypocholesterolaemia [14]. Specific nutritional deficiencies (e.g. zinc deficiency), should be screened for, although plasma levels of trace elements may not reflect tissue levels [7]. Patients with MEN-1 should be screened with fasting gut hormones, including glucagon levels on an annual bases [18].

Imaging

Gallium (68)-labelled octreotide scanning [19], contrast-enhanced computed tomography (CT), and visceral angiography are the imaging modalities of choice and are more sensitive than abdominal ultrasound in tumour detection [20] (Figure 6.8.2) (see also Chapter 6.11). Gallium scanning can be coregistered with a low dose CT scan, so that areas of increased uptake can be correlated with anatomical landmarks. Using colour for increased uptake superimposed on a black and white CT scan is most useful. MRI also has a high sensitivity [21] both in detecting the primary tumour and in detecting metastases. Repeated MRI scanning is preferable to repeated CT scans when screening individuals with MEN1 as repeated CT scans can cause a significant cumulative radiation exposure over time [7]. Nuclear medicine scanning with indium 111 or gallium 68 labelled octreotide reveals the presence of somatostatin receptors on these tumours [22].

Treatment

To ensure that all aspects of treatment are considered at the outset, a multidisciplinary approach to treatment is required. Review by a team that includes an endocrinologist, a pancreatic and upper gastrointestinal (GI) surgeon, an interventional radiologist, a histopathologist, a nuclear medicine physician, and an oncologist is required. Nutritional assessment, correction of any deficiencies, and implementation of weight maintenance strategies is imperative. Where required, diabetic control should be optimized. Anticoagulation therapy should be instigated although there are currently no guidelines regarding which anticoagulant to use or the extent of anticoagulation required.

Surgical resection, either of the tumour itself or distal pancreatectomy and splenectomy in local disease, is the treatment of choice and offers 5-year survival rates of over 66% [7]. However, 90% of patients have metastases at presentation and these commonly extend beyond lymph node metastases. Nevertheless, surgical resection or debulking of the tumour or distal pancreatectomy and splenectomy may offer good symptom relief. Unfortunately, symptoms return in a quarter of patients by 1 year, and 1-year survival rates in those in this subgroup are only 50% [7].

Because the liver has a dual blood supply, and because liver metastases secrete angiogenetic factors that derive a blood supply exclusively from the hepatic artery, occlusion of the hepatic artery in a patient with hepatic metastases results in preferential death of metastatic cells, while hepatocytes survive on the portal vein supply. The metastases recur, and derive blood supply from other anastomotic blood vessels of the coeliac axis, so that hepatic artery occlusion cannot be repeated. However hepatic artery embolization with microspheres allows selective devascularization of hepatic metastases with minimal effect on the liver parenchyma. Careful angiography is required, and insertion of the catheter as far into a metastasis as possible minimizes hepatic damage. Such hepatic artery embolization allows symptomatic relief in 80% of patients; this may not correlate with a fall in plasma glucagon [23]. Prior to the procedure, the patency of the portal vein must be established to ascertain whether there is adequate supply to the normal liver parenchyma. Complications of the procedure include massive peptide release, the effects of this may be minimized with the use of octreotide. Vasodilating peptides and contrast load may lead to severe hypotension, thus optimal fluid balance must be maintained both pre-and postembolization. Additionally, there are risks of infection in the necrotic tissue and of hepatic abscess formation [7].

Somatostatin analogues such as octreotide are the mainstay of medical therapy and provide rapid symptomatic relief, especially of NME [24], although they are less effective in control of weight loss and diabetes [25]. Somatostatin inhibits growth hormone and other pituitary and pancreatic hormones. It has been demonstrated to both reduce plasma glucagon levels and shrink tumour size, but its use was restricted by its 2-minute plasma half-life. Octreotide is a somatostatin analogue with a longer plasma half-life of 2 hours after intravenous administration. Lanreotide is a longer-acting somatostatin analogue, which can be administered every 2 weeks, and has been shown to be effective in controlling the rash [26]. Octreotide LAR and Lanreotide Autogel are slow-release preparations that are administered monthly, and control glucagon levels [27] and control the NME rash in patients with glucagonomas. It is not known whether these drugs prolong life in patients with glucagonomas, but these drugs are required to control the NME rash. Patients may require increasing doses of octreotide or lanreotide after 6 months to control symptoms [7]. Patients should be monitored for symptoms of gall stone formation as cholestasis is noted in 50% of patients [7].

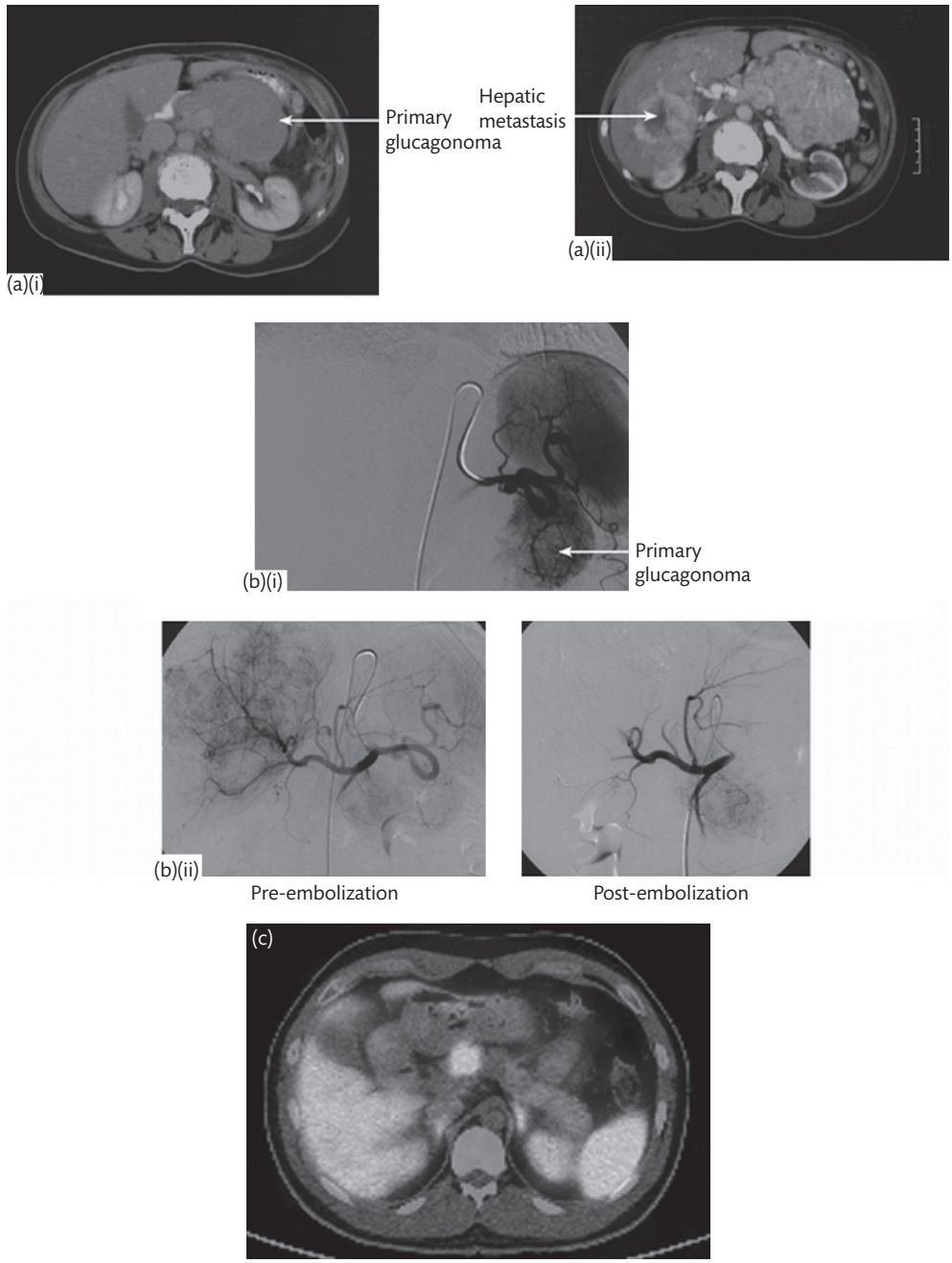


Figure 6. 8. 2 (a) Abdominal CT scans showing (i) primary glucagonoma, (ii) progression of primary glucagonoma, and hepatic metastases after 4 years. (b) (i) visceral angiogram showing cannulation of the splenic artery and vascular blush of primary glucagonoma, (ii) visceral angiogram showing cannulation of the hepatic artery and vascular blush of hepatic metastases, pre- and post-hepatic artery embolization. (c) Gallium (68) octreotate (colour) scan coregistered with a black and white CT scan showing a primary glucagonoma in the pancreas. Somatostatin receptors are bright yellow or white. The spleen has many somatostatin receptors and is thus bright white in all normal patients. The adrenal glands and liver have some receptors and are yellow in this scan. The glucagonoma expresses somatostatin receptors and is bright white on this scan. These receptors explain why the secretion of the hormone that causes the rash can be reduced by therapy with somatostatin analogues.

The PROMID study (Placebo controlled, Double-Blind, Prospective, Randomized study of the Effect of Octreotide LAR in the Control of Tumor Growth in Patients with Metastatic Neuroendocrine MIDgut Tumors) did not investigate glucagonomas, but its results probably apply to them. It was the first randomized trial to investigate the antitumor effect of octreotide LAR Depot [28, 29]. Eighty-five treatment-naïve patients with locally inoperable or metastatic NET were randomized to receive 30 mg octreotide LAR Depot or placebo. The median time to progression was 14.3 months in the treatment group versus 6 months in the placebo group (P = 0.000072) [28]. The European Society for Medical Oncology recommends octreotide LAR or lanreotide LA for managing patients with recurrent, inoperable, or metastatic GI NETs.

Chemotherapeutic agents such as streptozocin and 5-fluorouracil have been used in the treatment of glucagonoma. Streptozocin is a nitrosourea antibiotic with selective toxicity to pancreatic B cells,

demonstrated in animals, and 5-fluorouracil inhibits DNA synthesis. Chemotherapy has a very limited role in management in those with symptoms persisting at 6 months and patients tend to survive for less than a year [7].

Treatment with systemic radiolabelled lutetium (177) has been shown to help with remission for up to two years in a patient with a glucagonoma [22].

Combining the technology of hepatic embolization with yttrium-90 radiolabelled beads is also effective for hepatic metastases and is known as selective intrahepatic radiolabelled therapy (SIRT).

Prognosis

Patients with benign disease have an 85% survival rate at a mean follow-up of 4.7 years. Those with malignant disease treated with combination therapy have a 60% survival rate with a mean follow-up of 4.8 years.

REFERENCES

1. Holst JJ. Enteroglucagon. *Ann Rev Physiol*, 1997; **59**: 257–71.
2. Krejs GJ. Gastrointestinal endocrine tumors. *Am J Med*, 1987; **82**): 1–3.
3. Song X, Zheng S, Yang G, *et al*. Glucagonoma and the glucagonoma syndrome. *Oncology Letters*, 2018; **15**: 2749–55.
4. Wermers RA, Fatourechi Vand Kvols LK. Clinical spectrum of hyperglucagonemia associated with malignant neuroendocrine tumors. *Mayo Clinic Proceedings*, 1996; **71**: 1030–8.
5. Mallinson CN, Bloom SR, Warin AP, Salmon PR, Cox B. A glucagonoma syndrome. *Lancet*, 1974; **2**: 1–5.
6. van Beek AP, de Haas, Ellen RM, *et al*. The glucagonoma syndrome and necrolytic migratory erythema: a clinical review. *Eur J Endocrinol*, 2004; **151**: 531–7.
7. Frankton S, Bloom SR. Gastrointestinal endocrine tumours. Glucagonomas. *Bailliere's Clinical Gastroenterology*, 1996; **10**: 697–705.
8. Binnick AN, Spencer SK, Dennison WL, Horton ES. Glucagonoma syndrome. Report of two cases and literature review. *Arch Dermatol*, 1977; **113**: 749–54.
9. Tierney EP, Badger J. Etiology and pathogenesis of necrolytic migratory erythema: review of the literature. *MedGenMed*, 2004; **6**: 4.
10. Mullans EA, Cohen PR. Iatrogenic necrolytic migratory erythema: a case report and review of nonglucagonoma-associated necrolytic migratory erythema. *J Am Acad Dermatol*, 1998; **38**(5 Pt 2): 866–73.
11. Wald M, Lawrenz K, Luckner D, Seimann R, Mohnike K, Schober E. Glucagon therapy as a possible cause of erythema necrolyticum migrans in two neonates with persistent hyperinsulinaemsic hypoglycaemia. *Eur J Pediatr*, 2002; **161**: 600–3.
12. Wermers RA, Fatourechi V, Wynne AG, Kvols LK, Lloyd RV. The glucagonoma syndrome. Clinical and pathologic features in 21 patients. *Medicine*, 1996; **75**: 53–63.
13. Domen RE, Shaffer MB, Finke J, Sterin WK, Hurst CB. The glucagonoma syndrome. Report of a case. *Arch Int Med*, 1980; **140**: 262–3.
14. Chastain MA. The glucagonoma syndrome: a review of its features and discussion of new perspectives. *Am J Med Sci*, 2001; **321**: 306–20.
15. Bloom SR, Polak JM. Glucagonoma syndrome. *Am J Med*, 1987; **82**: 25–36.
16. Edney JA, Hofmann S, Thompson JS and Kessinger A. Glucagonoma syndrome is an underdiagnosed clinical entity. *Am J Surg*, 1990; **160**: 629.
17. White A, Tan K, Gray C, Roberts I, Ratcliffe JG. Multiple hormone secretion by a human pancreatic glucagonoma in culture. *Regulatory Peptides*, 1985; **11**: 335–45.
18. Thakker RV, Newey PJ, Walls GV, *et al*. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*, 2012; **97**: 2990–3011.
19. Al-Nahhas A, Win Z, Szyszko T, *et al*. Gallium-68 PET: a new frontier in receptor cancer imaging. *Anticancer Res*, 2007; **27**: 4087–94.
20. Hammond PJ, Jackson JA, Bloom SR. Localization of pancreatic endocrine tumours. *Clinical Endocrinology*, 1994; **40**: 3–14.
21. Armbruster M, Sourbron S, Haug A, *et al*. Evaluation of neuroendocrine liver metastases: a comparison of dynamic contrast-enhanced magnetic resonance imaging and positron emission tomography/computed tomography. *Invest Radiol*, 2014; **49**: 7–14.
22. Makis W, McCann K, Riauka TA, McEwan AJB. Glucagonoma pancreatic neuroendocrine tumor treated with 177Lu DOTATATE induction and maintenance peptide receptor radionuclide therapy. *Clin Nucl Med*, 2015; **40**: 877–9.
23. Ajani JA, Carrasco CH, Charnsangavej C, Samaan NA, Levin B, Wallace S. Islet cell tumors metastatic to the liver: effective palliation by sequential hepatic artery embolization. *Ann Int Med*, 1988; **108**: 340–4.
24. Anderson JV, Bloom SR. Neuroendocrine tumours of the gut: long-term therapy with the somatostatin analogue SMS 201-995. *Scand J Gastroenterol Suppl*, 1986; **119**: 115–28.
25. Schmid R, Allescher HD, Schepp W, *et al*. Effect of somatostatin on skin lesions and concentrations of plasma amino acids in a patient with glucagonoma-syndrome. *Hepato-Gastroenterology*, 1988; **35**: 34–7.
26. Tomassetti P, Migliori M, Gullo L. Slow-release lanreotide treatment in endocrine gastrointestinal tumors. *Am J Gastroenterol*, 1998; **93**: 1468–71.
27. Jockenhövel F, Lederbogen S, Olbricht T, *et al*. The long-acting somatostatin analogue octreotide alleviates symptoms by reducing posttranslational conversion of prepro-glucagon to glucagon in a patient with malignant glucagonoma, but does not prevent tumor growth. *Clin Investig*, 1994; **72**: 127–33.
28. Rinke A, Muller HH, Schade-Brittinger C, *et al*. PROMID Study Group Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*, 2009; **27**: 4656–63.
29. Rinke A, Wittenberg M, Schade-Brittinger C, *et al*. Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuro-endocrine midgut tumors (PROMID): results of long-term survival. *Neuroendocrinology*, 2017; **104**: 26–32

6.9

Vasointestinal Polypeptide Secreting Tumours

Alia Munir

Introduction

Vasointestinal polypeptide (VIP) secreting tumours (VIPomas) are rare functioning neuroendocrine tumours. The first case was reported in 1957. This patient presented with diarrhoea, islet cell tumour, peptic ulceration, and hypokalaemia [1, 2]. This was thought initially to be a variant of the Zollinger–Ellison syndrome. However, not until Verner and Morrison in 1958, described their eponymous syndrome, was this recognized as having an alternative hormonal aetiology [3]. They described a syndrome comprising of profuse, refractory watery diarrhoea, with severe hypokalaemia and dehydration, in association with non-insulin-secreting tumours of the pancreatic islets. VIP as the cause of the diarrhoea was confirmed by Kane in 1983 [4]. Bloom *et al.* suggested the term VIPoma syndrome in 1973 [5]. The absence of gastric hypersecretion and even achlorhydria was noted. Later termed 'pancreatic cholera' due to the resemblance to the severe diarrhoea caused by *Vibrio cholera* [6]. The acronym WDHA (watery diarrhoea hypokalaemia achlorhydria) has also been used to describe the VIP syndrome [6]. Notably, bicarbonate wasting is the cause of the associated systemic acidosis. Other tumour sites have been reported to secrete VIP including: in adults: lung cancer, colorectal cancer, phaeochromocytoma, hepatoma, hypernephroma, cutaneous mastocytoma and adrenal tumours [7–11]; in children: VIPomas may occur in sympathetic ganglia (ganglioneuroblastoma) and the adrenal glands [12].

Epidemiology

VIPomas are rare, with the reported annual incidence of 1 per 10 million in the general population (0.05–0.2 new cases per million per year) [13, 14]. In adults, 90% arise from the pancreas, but 10% are reported to arise from neural tissue, sympathetic ganglia, colon, bronchi, and adrenals. Some 40–70% are malignant and 5% occur as part of MEN 1 [13–15]. They are usually solitary and occur in the tail of the pancreas in up to 75% of patients, with up to 60–70% having already metastasized at the time of diagnosis. In children, VIPomas can occur along the autonomic chain and in the adrenal medulla, as ganglioneuromas and ganglioblastomas [16]. In adults, the typical age of presentation is 30–50 years and in children between 2 and 4 years [13, 14].

Pathogenesis

Recent whole-genome sequencing of 102 primary pancreatic neuroendocrine tumours (pNETs) has demonstrated that sporadic pNETs contain a larger than expected proportion of germline mutations [17]. Mutations in DNA repair genes *MUTYH, CHEK 2*, and *BRCA 2* (previously unreported), together with mutations in *MENIN* and *VHL*, have been shown to occur in 17% of patients. Point mutations and gene fusions were found in genes involved in the four main pathways including chromatin remodelling, DNA damage repair, activation of the mammalian target of rapamycin (mTOR) signalling (including previously undescribed EWSR1 fusions) and telomere maintenance [18]. A hypoxia and HIF subgroup was also identified in the gene analysis [19].

Pathophysiology

Pre-proVIP is a 170 amino acid precursor which is post-translationally modified to the 28 amino acid polypeptide, VIP, which binds to high affinity receptors on intestinal epithelial cells [20]. Other peptide fragments including peptide histidine methionine are also produced [21–23]. VIP is structurally similar to secretin and normally functions as a neurotransmitter within neurones. The cloned receptors are G-protein coupled receptors of the secretin family. It is a potent vasodilator and stimulates gastrointestinal and pancreatic secretion. It can inhibit gastric acid secretion. VIP binds to receptors on epithelial cells in all segments of the intestine and activation of cellular adenylate cyclase and the production of cAMP results in net fluid and electrolyte secretion (Na, K, Cl, HCO3 and water) into the lumen causing diarrhoea and hypokalaemia [24]. Hypercalcaemia and hyperglycaemia may also occur. Other agents that maybe implicated in the diarrhoea syndrome include gastrin, glucagon, pancreatic polypeptide, thyrocalcitonin, prostaglandins, or other peptides from pre-proVIP.

Diagnosis and Clinical Features

Diagnosis can sometimes be challenging due to overlap with other causes of diarrhoea. A raised fasting VIP, with profuse secretory diarrhoea (of no other aetiology) that occurs despite fasting and characteristic symptoms are usually diagnostic. This occurs in nearly all patients (90–100%), with hypokalaemia (often less than 2.5 mmol/L) present in 80–100% and dehydration in 83%. In a review of 55 patients with diarrhoea and hypokalaemia, the most dominant feature was high volume profuse cholera-like diarrhoea, often present for 3–4 years prior to diagnosis [15, 18, 25, 26]. The common features of the VIP syndrome are shown in Table 6.9.1 and the causes of secretory diarrhoea are shown in Box 6.9.1.

In VIPoma, the diarrhoea will not disappear with 48–72 hours of fasting, and has the appearance of dilute tea. Stool volumes frequently exceed 6–8 litres of stool per 24 hours [15, 18, 25]. The average potassium loss from stool is of the order of 300 mmol/24 hours. The presence of a secretory diarrhoea may be confirmed by finding an osmotic gap of <50 mOsmol/kg between serum and stool osmolality: serum osmolality (or 290) minus the product of twice the stool potassium plus sodium concentrations (290–2(Na+K) mOsm/Kg) [27].

Early in the disease the diarrhoea may be intermittent and episodic, but as the tumour enlarges it becomes continuous and life-threatening electrolyte abnormalities ensue. Other notable clinical and biochemical features include: hypercalcaemia, stimulation of glycogenolysis causing hyperglycaemia, and gallbladder distension. The cause of hypercalcaemia is unknown, but maybe attributed to diarrhoea, parathyroid hormone-related peptide secretion (PTH-rp) or coincidental, and coexistent primary hyperparathyroidism in the context of MEN 1. Only 8% of VIPomas are associated with facial flushing, which when it occurs, tends to be patchy and urticarial. Hypomagnesaemia has also been associated and maybe responsible for tetany in some patients. Abdominal pain is mild or absent and symptoms relate to dehydration, hypokalaemia, and include lethargy, nausea, vomiting, muscle weakness, and muscle cramps.

Box 6.9.1 Causes of secretory diarrhoea

- Infection
  - Cholera
  - *E. coli*
- Villous adenoma of the rectum
- Laxative abuse
- IgA deficiency
- Congenital
  - Dysautonomia
  - Chloridorrhoea
  - Structural enteric abnormalities
- Neuroendocrine tumours
  - VIPoma
  - Carcinoid
  - Gastrinoma (Zollinger–Ellison syndrome)
  - Medullary carcinoma of the thyroid
- Miscellaneous
  - Systemic mastocytosis
  - Basophilic leukaemia
- Idiopathic

Reproduced with permission from Wynne K. VIPomas. In: Wass J, Semple R, Arlt W, eds. *Oxford Textbook of Endocrinology and Diabetes*. 2nd ed. Oxford: Oxford University Press; 2011.

Table 6.9.1 Features of the VIPoma syndrome

VIP action	Clinical/biochemical feature
Intestinal secretion of Na^+, Cl^-, and HCO^{3-}	Secretory diarrhoea Dehydration Weight loss Metabolic acidosis
Intestinal secretion of K^+ and hyperaldosteronism secondary to hypovolaemia	Hypokalaemia
Increased bone reabsorption Acidosis Tumour secretion of PTHrP Hyperparathyroidism secondary to MEN 1	Hypercalcaemia
Increased glycogenolysis	Hyperglycaemia
Vasodilation	Flushing

PTHrP, parathyroid hormone-related protein; MEN 1, multiple endocrine neoplasia type 1.

Reproduced with permission from Wynne K. VIPomas. In: Wass J, Semple R, Arlt W, eds. *Oxford Textbook of Endocrinology and Diabetes*. 2nd ed. Oxford: Oxford University Press; 2011.

Localization

Numerous localization studies are recommended in an effort to accurately localize, stage, and plan treatment in VIPoma, especially to determine whether the tumour is amenable to surgical resection with curative intent. Conventional imaging studies, such as contrast computerized tomography (CT) scan are standard first-line imaging tools. Magnetic resonance imaging (MRI) and endoscopic ultrasound scan (EUS) [28] (useful for fine-needle aspiration) may be used to ascertain liver involvement and to confirm a tissue diagnosis, respectively. Functional imaging with somatostatin receptor scintigraphy (SRS) has been superseded, where available,

by ^{68}Gallium-DOTATATE/TOC due to the higher sensitivity and specificity for diagnosis and staging of neuroendocrine tumours. Fluorodeoxyglucose (FDG) positron emission tomography (PET) CT may be positive in tumours with higher grade. Nuclear imaging modalities may also be accurate predictors of octreotide responsiveness and guide future therapies. Other isotopes also used include ^{11}C-5HTP and ^{18}F DOPA, but availability of these may be limited.

Treatment

The treatment of VIPoma, as with other functional pancreatic neuroendocrine tumours, has two main goals: (1), control of clinical symptoms resulting from the hormonal excess; (2), control of tumour growth. Given the high incidence of metastatic disease at diagnosis, combinations of therapies are frequently employed. It is recognized that cytoreductive surgery can facilitate the management of hormonal excess and should be considered where 70–90% of the visible disease is potentially resectable [19]. There is, however, a paucity of control studies, but the ENETS guidelines recommend surgery with curative intent wherever possible [18]. Liver-directed therapies such as transarterial chemoembolization (TACE) or radioembolization, selective internal radiotherapy (SIRT) as well as radiofrequency ablation (RFA) can be employed where appropriate [29] but data specifically in VIPoma is limited.

Management

Initial management is focussed on correction of electrolyte abnormalities and restoration of renal function and maintenance of cardiac stability, and high dependency or intensive care monitoring is usually needed. Parenteral crystalloid fluids (3–6 litres per day) of with potassium replacement (up to 350 mmol/day) may be required at initial presentation. The next aim is to slow the diarrhoea. This frequently requires somatostatin analogue therapy to block the secretion of VIP [30]. In addition, somatostatin analogues have been shown to have a tumour antiproliferative effect with significantly increased progression free survival [31, 32] (see Chapters 6.4 and 6.5). Glucocorticoids (such as prednisolone 10–30 mg/day) may be effective at reducing the volume of diarrhoea. Glycaemic control frequently requires insulin. In selected cases ongoing parenteral nutrition/fluid maybe needed to facilitate discharge from hospital and promote medical fitness for ongoing treatments.

Surgery

In non-metastatic disease, surgery can offer cure in up to 50%, and a large adult series of 241 case reports of VIPoma showed a 5-year survival of 89% with pancreatic VIPoma and 68.5% in neurogenic VIP producing tumours [15, 33]. Debulking surgery may also be of benefit. Perioperative proton-pump inhibition is recommended to prevent rebound gastric acid secretion. Careful monitoring of fluid balance is imperative, both peri and postoperatively, as fluid overload may occur as VIP levels fall after surgery.

Pharmacotherapy

Somatostatin analogues provide a symptomatic response in 80–90% of neuroendocrine tumours within days of initiation and are therefore drug of choice to control diarrhoea in VIPomas [34, 35]. Dosing can vary, depending on clinical need, but initiating at 100 mcg subcutaneously every 8 hours is a reasonable starting dose, with increments of 50 mcg to 200 mcg as needed [36]. Intravenous octreotide may need to be considered if control poor. Response should be measured by stool output and not plasma VIP levels, initially. Loperamide and codeine phosphate may useful to help control diarrhoea. Somatostatin analogue-induced steatorrhoea is responsive to oral pancreatic enzyme supplementation. Mild glucose intolerance and asymptomatic gallstones may occur in up to a quarter of patients on somatostatin analogues [37]. Once over the initial acute presentation patients can be switched to long-acting somatostatin analogues [38]. Escape of control may occur quite frequently in VIPomas and more definitive treatments are usually required in the longer term. Frequently the addition of glucocorticoids gives symptomatic relief and when used in combination with somatostatin analogues can prove effective. This may necessitate more intense glycaemic control, usually with insulin therapy. Consideration may be given to other drugs which enhance sodium absorption from the jejunum or inhibit intestinal secretion such as clonidine and opiates, respectively, but these may have limited long-term impact.

Metastatic Disease

There is a wide range of available treatments in patients with metastatic disease, and patients should be discussed in a neuroendocrine multidisciplinary meeting (NET MDT) by multiple specialists within the field in order to optimize outcomes [13].

Limited hepatic metastases may be considered for resection. Debulking the disease may have a positive effect on hormonal control and tumour load. Liver transplantation may be indicated in a limited small number of patients in whom life-threatening hormonal symptoms persist despite maximal medical therapy and standard therapy is not feasible. This is only undertaken in the context of discussion in a specialist NET MDT [13].

Hepatic artery embolization may be used in symptomatic liver metastases not amenable to surgery [39]. Infection, hormonal secretory crisis, and post-procedure hyperuricaemia must be proactively avoided in such cases [40]. Hepatic radiofrequency ablation, cryotherapy, and laser therapy are regarded as less invasive therapies but are limited in general to smaller sized metastases (4–5 cm).

The recommended sequencing of therapies has not been evaluated in VIPomas, but there is a suggestion that PRRT should be considered prior to embolization of the hepatic artery to maximize delivery of the therapy to liver metastases.

Chemotherapy, Targeted Therapies, and Peptide Receptor Radionuclide Therapy (PRRT)

The efficacy of chemotherapy in VIPomas is difficult to determine as very small numbers of patients have been treated with

chemotherapy and in most studies all types of pancreatic neuroendocrine have been included. Chemotherapy agents have been well established in the treatment of pancreatic neuroendocrine tumours in general (see Chapters 6.4 and 6.5). The use of streptozocin, temozolomide, capecitabine, fluorouracil, and cisplatin has been published, usually in combination with somatostatin analogues for symptom control [41–43]. In a randomized multicentre study of 80 patients with metastatic NET, there was no significant difference between those given lanreotide autogel, interferon alfa or both [44]. Interferon alfa has been used in VIPomas with refractory symptoms showing some benefit in terms of symptom control [45].

There have also been reports of positive use of multitargetted tyrosine kinase inhibitors (TKI) (Sunitinib) [46, 47] and mTOR inhibitors (Everolimus) in VIPomas, with an increase in progression free survival by up to 18 months.

The effect of the tyrosine kinase inhibitor, Sunitinib was studied recently in a randomized control trial of 171 patients with advanced well differentiated pancreatic neuroendocrine tumours, including VIPomas. The progression free survival rate was 11.4 months in the sunitinib group compared to 5.5 months in the placebo group. Less toxicity was reported when compared to systemic chemotherapy [48]. Other TKI such as pazopanib and cabozantinib have also shown some activity in NET demonstrating antiproliferative and antihormonal effects (see also Chapters 6.4 and 6.5).

Everolimus is now licenced for use worldwide for use in pNETS, GI and lung neuroendocrine tumours (NETS) with advanced progressive disease supported by data from recent placebo-controlled randomized controlled trials [49, 50]. In RADIANT 1, a phase II study, everolimus alone stabilized disease in 67.8% of patients and in combination with octreotide LAR, stabilized disease in 80% [51]. In RADIANT III, 410 patients, with radiological progression, were randomized to everolimus or usual therapy. Progression free survival was 11 months compared to 4.6 months with usual therapy [49].

Peptide receptor radiotherapy (PRRT) using radiolabelled somatostatin analogues have been approved for the treatment of metastatic neuroendocrine tumours [52, 53], and can be highly effective in metastatic disease and in controlling the functional aspects in pNETS [54, 55]. High-dose β radiation is delivered to tumours following receptor binding and internalization of a somatostatin analogue labelled with a radionuclide (^{177}lutetium- or ^{90}yttrium). There have been case reports of success with reduced tumour burden and improved symptoms following PRRT in patient with VIPomas [56].

Somatostatin analogues, interferon alfa, and PRRT have been shown to improve quality of life but formal data on overall tumour response are limited [44].

Prognosis

Given the rarity of VIPomas accurate survival data are limited. It is known, however, that smaller tumours without metastases have a more favourable prognosis. In metastatic disease electrolyte stabilization and use of more modern therapies may be of benefit. A small study of pancreatic VIPoma has reported a 5-year survival of up to 68% [33].

Although there has been great progress in the management of NETs, VIPoma remains a rare and challenging neuroendocrine tumour to diagnose and manage. In particular, VIPoma demands meticulous management of electrolytes, acid-base status, and glycaemia. With respect to the antitumour and hormonal control, somatostatin analogues remain the mainstay of treatment, with newer targeted agents showing great promise. Ideally a worldwide study for this rare functioning neuroendocrine tumour would be required to evaluate and assess treatment sequencing, quantify tumour and hormonal response, and measure long-term survival.

REFERENCES

1. Priest WM, Alexander MK. Islet cell tumour of the pancreas with peptic ulceration, diarrhoea, and hypokalaemia. *Lancet*, 1957; **273**: 1145–7.
2. Murray JS, Paton RR, Pope CE, 2nd. Pancreatic tumor associated with flushing and diarrhea. Report of a case. *N Engl J Med*, 1961; **264**: 436–9.
3. Verner JV, Morrison AB. Islet cell tumor and a syndrome of refractory watery diarrhea and hypokalemia. *Am J Med*, 1958; **25**: 374–80.
4. Kane MG, O'Dorisio TM, Krejs GJ. Production of secretory diarrhea by intravenous infusion of vasoactive intestinal polypeptide. *N Engl J Med*, 1983; **309**: 1482–5.
5. Bloom SR, Polak JM, Pearse AG. Vasoactive intestinal peptide and watery-diarrhoea syndrome. *Lancet*, 1973; **2**: 14–16.
6. Marks IN, Bank S, Louw JH. Islet cell tumor of the pancreas with reversible watery diarrhea and achylorhydraia. *Gastroenterology*, 1967; **52**: 695–708.
7. Wesley JR, Vinik AI, O'Dorisio TM, Glaser B, Fink A. A new syndrome of symptomatic cutaneous mastocytoma producing vasoactive intestinal polypeptide. *Gastroenterology*, 1982; **82**(5 Pt 1): 963–7.
8. Tiedemann K, Pritchard J, Long R, Bloom SR. Intractable diarrhoea in a patient with vasoactive intestinal peptide-secreting neuroblastoma. Attempted control by somatostatin. *Eur J Pediatr*, 1981; **137**: 217–19.
9. Hamilton I, Reis L, Bilimoria S, Longet RG. A renal vipoma. *Br Med J*, 1980; **281**: 1323–4.
10. Tiedemann K, Long RG, Pritchard J, Bloom SR. Plasma vasoactive intestinal polypeptide and other regulatory peptides in children with neurogenic tumours. *Eur J Pediatr*, 1981; **137**: 147–50.
11. Modlin IM, Bloom SR. VIPomas and the watery diarrhoea syndrome. *S Afr Med J*, 1978. **54**: 53–6.
12. Blair AW, Ahmed S. Presacral vipoma in a 16-month-old child. *Acta Paediatr Belg*, 1981; **34**: 89–91.
13. Falconi M, Eriksson B, Kaltsas G, *et al*. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. *Neuroendocrinology*, 2016; **103**: 153–71.
14. Jensen RT, Cadiot G, Brandi ML, *et al*. ENETS Consensus Guidelines for the management of patients with digestive neuroendocrine neoplasms: functional pancreatic endocrine tumor syndromes. *Neuroendocrinology*, 2012; **95**: 98–119.
15. Smith SL, Branton SA, Avino AJ, *et al*. Vasoactive intestinal polypeptide secreting islet cell tumors: a 15-year experience and review of the literature. *Surgery*, 1998; **124**: 1050–5.
16. Long RG, Bryant MG, Mitchell SJ, Adrian TE, Polak JM, Bloom SR. Clinicopathological study of pancreatic and

ganglioneuroblastoma tumours secreting vasoactive intestinal polypeptide (vipomas). *Br Med J (Clin Res Ed)*, 1981; **282**: 1767–71.
17. Scarpa A, Chang DK, Nones K, *et al.* Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*, 2017; **543**: 65–71.
18. Partelli S, Bartsch DK, Capdevila J, *et al.* ENETS Consensus guidelines for standard of care in neuroendocrine tumours: surgery for small intestinal and pancreatic neuroendocrine tumours. *Neuroendocrinology*, 2017; **105**: 255–65.
19. Oberg K. Management of functional neuroendocrine tumors of the pancreas. *Gland Surg*, 2018; **7**: 20–7.
20. Itoh N, Obata K, Yanaihara N, Okamoto H. Human preprovasoactive intestinal polypeptide contains a novel PHI-27-like peptide, PHM-27. *Nature*, 1983; **304**: 547–9.
21. Tatemoto K, Mutt V. Isolation of two novel candidate hormones using a chemical method for finding naturally occurring polypeptides. *Nature*, 1980; **285**: 417–18.
22. Ghiglione M, Christofides ND, Yiangou Y, Uttenthal LO, Bloom SR. PHI stimulates intestinal fluid secretion. *Neuropeptides*, 1982; **3**: 79–82.
23. Said SI, Mutt V. Potent peripheral and splanchnic vasodilator peptide from normal gut. *Nature*, 1970; **225**: 863–4.
24. Said SI, Mutt V. Polypeptide with broad biological activity: isolation from small intestine. *Science*, 1970; **169**: 1217–18.
25. Verner JV, Morrison AB. Endocrine pancreatic islet disease with diarrhea. Report of a case due to diffuse hyperplasia of nonbeta islet tissue with a review of 54 additional cases. *Arch Intern Med*, 1974; **133**: 492–9.
26. Ghaferi AA, Chojnacki KA, Long WD, Cameron JL, Yeo CJ. Pancreatic VIPomas: subject review and one institutional experience. *J Gastrointest Surg*, 2008; **12**: 382–93.
27. Donowitz M, Kokke FT, Saidi R. Evaluation of patients with chronic diarrhea. *N Engl J Med*, 1995; **332**: 725–9.
28. Anderson MA, Carpenter S, Thompson NW, Nostrant TT, Elta GH, Scheiman JM. Endoscopic ultrasound is highly accurate and directs management in patients with neuroendocrine tumors of the pancreas. *Am J Gastroenterol*, 2000; **95**: 2271–7.
29. Leavey SF, Holloway H, Courtney MG, *et al.* Pancreatic vipomas: spectrum of presentation and evolution of diagnostic and therapeutic modalities. *Ir J Med Sci*, 1995; **164**: 37–9.
30. Song S, Shi R, Li B, Liu Y. Diagnosis and treatment of pancreatic vasoactive intestinal peptide endocrine tumors. *Pancreas*, 2009; **38**: 811–14.
31. Caplin ME, Pavel M, Ruszniewski P. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*, 2014; **371**: 1556–7.
32. Rinke A, Müller HH, Schade-Brittinger C, *et al.* Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID Study Group. *J Clin Oncol*, 2009; **27**: 4656–63.
33. Soga J, Yakuwa Y. Vipoma/diarrheogenic syndrome: a statistical evaluation of 241 reported cases. *J Exp Clin Cancer Res*, 1998; **17**: 389–400.
34. O'Dorisio TM, Mekhjian HS, Gaginella TS. Medical therapy of VIPomas. *Endocrinol Metab Clin North Am*, 1989; **18**: 545–56.
35. Kraenzlin ME, Ch'ng JL, Wood SM, Carr DH, Bloom SR. Long-term treatment of a VIPoma with somatostatin analogue resulting in remission of symptoms and possible shrinkage of metastases. *Gastroenterology*, 1985; **88**(1 Pt 1): 185–7.
36. Lamberts SW, van der Lely AJ, de Herder WW, Hofland LJ. Octreotide. *N Engl J Med*, 1996; **334**: 246–54.
37. Newman CB, Melmed S, Snyder PJ, *et al.* Safety and efficacy of long-term octreotide therapy of acromegaly: results of a multicenter trial in 103 patients--a clinical research center study. *J Clin Endocrinol Metab*, 1995; **80**: 2768–75.
38. Oberg K, Kvols L, Caplin M, *et al.* Consensus report on the use of somatostatin analogs for the management of neuroendocrine tumors of the gastroenteropancreatic system. *Ann Oncol*, 2004; **15**: 966–73.
39. Ajani JA, Carrasco CH, Charnsangavej C, Samaan NA, Levin B, Wallace S. Islet cell tumors metastatic to the liver: effective palliation by sequential hepatic artery embolization. *Ann Intern Med*, 1988; **108**: 340–4.
40. Clouse ME, Lee RG. Management of the posthepatic artery embolization syndrome. *Radiology*, 1984; **152**: 238.
41. Moertel CG, Lefkopoulo M, Lipsitz S, Hahn RG, Klaassen D. Streptozocin-doxorubicin, streptozocin-fluorouracil or chlorozotocin in the treatment of advanced islet-cell carcinoma. *N Engl J Med*, 1992; **326**: 519–23.
42. Pignal F, René E, Chayvialle JA, Rigaud D, Lehy T, Bonfils S. Streptozotocin treatment in pancreatic cholera (Verner-Morrison) syndrome. *Digestion*, 1982; **24**: 176–82.
43. Vinik A. Vasoactive intestinal peptide tumor (VIPoma). In: De Groot LJ, Chrousos G, Dungan K, *et al.*, eds. *Endotext*. South Dartmouth, MA: MDText.com, 2000.
44. Faiss S, Pape UF, Böhmig M, *et al.* Prospective, randomized, multicenter trial on the antiproliferative effect of lanreotide, interferon alfa, and their combination for therapy of metastatic neuroendocrine gastroenteropancreatic tumors—the International Lanreotide and Interferon Alfa Study Group. *J Clin Oncol*, 2003; **21**: 2689–96.
45. Oberg K. Interferon in the management of neuroendocrine GEP-tumors: a review. *Digestion*, 2000; **62 Suppl 1**: 92–7.
46. de Mestier L, Walter T, Brixi H, Lombard-Bohas C, Cadiot G. Sunitinib achieved fast and sustained control of VIPoma symptoms. *Eur J Endocrinol*, 2015; **172**: K1–3.
47. Bourcier ME, Vinik AI. Sunitinib for the treatment of metastatic paraganglioma and vasoactive intestinal polypeptide-producing tumor (VIPoma). *Pancreas*, 2013; **42**: 348–52.
48. Raymond E, Dahan L, Raoul J-L, *et al.* Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 501–13.
49. Yao JC, Shah MH, Tetsuhide I, *et al.* Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*, 2011; **364**: 514–23.
50. Yao JC, Lombard-Bohas C, Baudin E, *et al.* Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *Lancet*, 2016; **387**: 968–77.
51. Yao JC, Lombard-Bohas C, Baudin E, *et al.* Daily oral everolimus activity in patients with metastatic pancreatic neuroendocrine tumors after failure of cytotoxic chemotherapy: a phase II trial. *J Clin Oncol*, 2010; **28**: 69–76.
52. Hicks RJ, Kwekkeboom DJ, Krenning E, *et al.* ENETS consensus guidelines for the standards of care in neuroendocrine neoplasia: peptide receptor radionuclide therapy with radiolabeled somatostatin analogues. *Neuroendocrinology*, 2017; **105**: 295–309.
53. Strosberg J, El-Haddad G, Wolin E, *et al.* Phase 3 trial of (177)lu-dotatate for midgut neuroendocrine tumors. *N Engl J Med*, 2017; **376**: 125–35.

54. Kwekkeboom DJ, Kam BL, van Essen M, *et al*. Somatostatin-receptor-based imaging and therapy of gastroenteropancreatic neuroendocrine tumors. *Endocr Relat Cancer*, 2010; **17**: R53–73.
55. van Schaik E, van Vliet EI, Feelders RA, *et al*. Improved control of severe hypoglycemia in patients with malignant insulinomas by peptide receptor radionuclide therapy. *J Clin Endocrinol Metab*, 2011; **96**: 3381–9.
56. Yachin M, Quigley AM, Meyer T, Caplin M, Toupanakis C. Salvage treatment with peptide receptor radionuclide therapy (PRRT) in patients with advanced VIPomas and severe Verner-Morrison syndrome. In: 12th Annual ENETS Conference, 2015.

6.10

Somatostatinoma

John A.H. Wass

Introduction

Somatostatin was isolated in 1973 by Paul Brazeau in Roger Guillemin's laboratory. It was found to have a widespread distribution, not only in the hypothalamus and brain but also in the gastrointestinal tract. Sixty-five per cent (65%) of the body's somatostatin is in the gut, mostly in the D cells of the gastric and intestinal epithelium. It is also present in the myometric and submucosal plexuses. The highest concentration is in the antrum of the stomach and there is a gradual decrease of concentrations down the gastrointestinal tract. Five per cent (5%) of the body's somatostatin is in the pancreas.

Infused somatostatin, which has a short half-life of 3 minutes, has a large number of actions on the pituitary gland, the endocrine and exocrine pancreas, gastrointestinal tract, other hormones, and on the nervous system (Box 6.10.1). Among its various actions of importance in the gastrointestinal tract is the inhibition of gastrin and cholecystokinin (CCK). In the pancreas, insulin and glucagon are inhibited. Non-endocrine actions include inhibition of gastric acid secretion, pancreatic exocrine function, gall bladder contraction, and intestinal motility. Intestinal absorption of nutrients, including glucose, triglycerides, and amino acids, is also inhibited [1].

Somatostatin and its long-acting analogues, lanreotide, and octreotide are also potent inhibitors of tumour growth including that of neuroendocrine tumours.

Somatostatin exists in two main forms, as a 14-amino acid peptide (somatostatin 14) present mainly in the pancreas and the stomach, and as a 28-amino acid peptide present mainly in the intestine. Somatostatin 14 is the peptide present in enteric neurons.

Somatostatin receptors are present on many cell types, including the parietal cells of the stomach, G cells, D cells themselves, and cells of the exocrine and endocrine pancreas. A large number of tumours also have somatostatin receptors and these include pituitary adenomas, endocrine pancreatic tumours, carcinoid tumours, paragangliomas, phaeochromocytomas, small cell lung carcinomas, lymphomas, and meningiomas. This fact aids detection of these tumours by PET-CT scanning (^{68}Ga-DOTATATE). Five different somatostatin receptors (SSTRs) have been cloned (SSTR1–SSTR5) and all are on different chromosomes. These have a varying affinity for somatostatin 14 and somatostatin 28 and a varying

Box 6.10.1 Actions of exogenously administered somatostatin on endocrine and exocrine secretion

Endocrine secretion—inhibits the secretion of:

Pituitary
- Growth hormone
- Thyroid-stimulating hormone

Gastrointestinal tract
- Gastrin
- Cholecystokinin
- Secretin
- Vasoactive intestinal polypeptide
- Gastrin-inhibiting peptide
- Motilin
- Enteroglucagon
- Pancreatic polypeptide
- Insulin
- Glucagon
- Somatostatin

Other peptides
- Renin

Exocrine secretion—inhibition of:
- Gastric acid secretion
- Gastric emptying rate
- Pancreatic exocrine function: volume, electrolytes, and enzyme content
- Gall bladder contraction
- Intestinal motility
- Intestinal absorption of nutrients
- Splanchnic blood flow
- Renal water reabsorption
- Activity of some central nervous system neurons

tissue distribution with SSTR2 and 5 being predominant in the pituitary [2].

Somatostatin can act either as an endocrine hormone or in a paracrine or autocrine way. It probably also has luminal effects in the gastrointestinal tract. Lastly, it can act as a neurotransmitter [3].

Somatostatinoma

Somatostatinomas are neuroendocrine tumours [4, 5]. They are rare, with an estimated incidence of about 1 in 40 million. In total, over 200 have been described. They may be sporadic (90%) or familial (10%). Two main types exist: pancreatic somatostatinomas (56%), which are large tumours often associated with features of somatostatin excess; and duodenal tumours (44%), which are usually small and more amenable to surgical resection [6]. They have also been described in the jejunum and cystic duct. The two types are compared in Table 6.10.1. They are infrequently associated with multiple endocrine neoplasia type 1 syndrome (7%), neurofibromatosis type 1, or Von Hippel–Lindau syndrome. The somatostatinoma syndrome is a triad of diabetes mellitus, diarrhoea, and gallstones [7].

Pancreatic Somatostatinoma

Somatostatinoma syndrome was first described in 1977 [8]. Over 100 such cases have now been reported with features as in Box 6.10.2. The syndrome consists of cholelithiasis, the cause of which is multifactorial, including suppression of CCK production which results in impaired gallbladder contractility. High levels of somatostatin also inhibit bowel transit, which alters bowel flora, thus increasing bile acid reabsorption and this is associated with super saturated bile [9]. Mild diabetes occurs and has often been present for many years before diagnosis. It is probably due to suppression of insulin secretion. Diarrhoea and steatorrhoea also occur and relate to the inhibition of pancreatic exocrine function. Hypochlorhydria relates to the inhibition of gastric acid secretion and gastrin. Anaemia, abdominal pain, and weight loss are also present and are non-specific. They are probably related to the size of the tumour, which is usually large, and also to the fact that it is malignant. Those tumours are often diagnosed late and distant metastases may be present in lymph nodes, liver, or bone (55% are in the head of the pancreas).

Plasma and tissue levels of somatostatin are elevated and levels are higher in pancreatic as opposed to duodenal somatostatinomas. These somatostatin-secreting cells often also secrete adrenocorticotrophic hormone (ACTH), calcitonin, insulin, or some other peptides. This means that Cushing's syndrome, flushing, or hypoglycaemia (if there is cosecretion of insulin) may be present [10].

Table 6.10.1 Comparison of pancreatic and extrapancreatic somatostatinomas

Feature	Pancreatic	Extrapancreatic (duodenal)
Number of patients	81	81
Inhibitory syndrome (%)	18.5	2.5
von Recklinghausen's disease (%)	1.2	43.2
Large tumour (>20 mm) (%) (NFI)	85.5	41.4
Multisecretorary activity (%)	33.3	16.3 No differences
Metastatic rate and malignancy		
5-year survival	75.2% overall 59.9% with metastases 100% without metastases	

Box 6.10.2 Features of pancreatic somatostatinoma

- Hyperglycaemia 95%
- Cholelithiasis 68%, if inhibitory syndrome present
- Steatorrhoea 47%
- Hypochlorhydria
- Diarrhoea 60% with pancreatic; 11% with duodenal
- Abdominal pain 40%
- Weight loss 25%
- Anaemia 14%
- Elevated plasma and tissue somatostatin
- Histologically malignant, may be associated with ACTH, calcitonin, and insulin secretion

Duodenal Somatostatinoma

Duodenal somatostatinomas tend to be smaller and present earlier. The vast majority occur near the ampulla of Vater where they tend to cause obstructive biliary disease (NFI) (39%). Some are associated with neurofibromatosis type 1 and some are occasionally associated with phaeochromocytoma. Radiologically they can be difficult to diagnose. This may need endoscopic techniques. At presentation paraduodenal lymph nodes are involved because there is a high malignancy rate, although this is usually low grade. None of the duodenal somatostatinoma patients have developed the full-blown somatostatinoma syndrome but diabetes and gall stones have been noted in some cases.

Recently gain of function hypoxia-inducible factor 2 (HIF 2A) post-zygotic somatic mutations (mosaicism) were detected in ampullary somatostatinoma associated with multiple paraganglioma and polycythaemia [11, 12]. Although originally described only in females it has also been described in males [13].

Histologically these are psammomatous tumours. Treatment is with surgery if this is feasible, chemotherapy, and, if necessary, hepatic embolization. Somatostatin analogues may lower somatostatin levels and improve symptoms (such as diarrhoea) of both types of somatostatinoma if metastases are present.

Localization

The radiopharmaceutical ^{68}Ga-DOTATATE combined with CT or MRI is superior to previous imaging methods and is now the gold standard in localization of NETs.

Treatment

Surgical treatment provides the best chance of cure. If this is not possible, surgical debulking will provide significant relief and

prolongs survival rate. Liver resection can be considered with solitary metastasis.

Somatostatic analogues relieve symptoms and inhibit growth. Lanreotide may prolong progression free survival [13].

Radiotherapy with 177 lutetium (177-lu-octreotate) causes tumour response in those with a high uptake on the OctreoScan; capecitabine can also be used in combination. In ENETS grade 1 and 2 tumours, everolimus or sunitinib may be useful for progressive disease.

Treatment Outcomes

There is a high prevalence of metastatic disease in somatostatinomas, likely a consequence of late diagnosis. Patients with somatostatinoma tend to have a long survival rate and complete removal of sporadic or hereditary somatostatinoma is usually effective and ensures prolonged survival. Most duodenal somatostatinomas survive and for pancreatic somatostatinoma 10-year survival is around 70% [12].

REFERENCES

1. Schultz A. Somatostatin: physiology and clinical application. *Clin Endocrinol Metabol*, 1994; **8**: 215–36.
2. Farooqi S, Bevan JS, Sheppard MC, Wass JAH. The therapeutic value of somatostatin and its analogues. *Pituitary*, 1999; **2**: 79–88.
3. Schonbrunn A. Somatostatin in endocrinology. In: De Groot LJ, Jameson JL, eds. Philadelphia, PA: WB Saunders and Co., 2005: 427–77.
4. Nesi G, Marcucci T, Rubio CA, Brandi ML, Tonelli F. Somatosta tinoma: clinicopathological features of three cases and literature reviewed. *J Gastroenterol Hepatol*, 2008; **23**: 521–6.
5. Oberg K, Eriksson B. Endocrine tumours of the pancreas. *Best Pract Res Clin Gastroenterol*, 2005; **19**: 753–81.
6. Krejs GJ, Orci L, Conlon JM, *et al*. Somatostatinoma syndrome. Biochemical, morphological and clinical features. *N Engl J Med*, 1979; **301**: 285–92.
7. Soga J, Yakuwa Y. Somatostatinoma/inhibitory syndrome: a statistical evaluation of 173 reported cases as compared to other pancreatic endocrinomas. *J Exp Clin Cancer Res*, 1999; **18**: 13–22.
8. Ganda OP, Weir GC, Soeldner JS, *et al*. 'Somatostatinoma'; the somatostatin containing tumour of the endocrine pancreas. *N Engl J Med*, 1977; **296**: 963–7.
9. Dowling RH, Hussaini SH, Murphy GM, Besser GM, Wass JAH. Gallstones during octreotide therapy. *Clin Exp Metabol*, 1992; **41**: 22–33.
10. Wright J, Abolfathr A, Penman E, Marks V. Pancreatic somatostatinoma presenting with hypoglycaemia. *Clin Endocrinol*, 1980; **12**: 603–8.
11. Zhuang Z, Yang C, Lorenzo F, *et al*. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. *N Engl J Med*, 2012; **367**: 922–30.
12. Därr R, Nambuba J, Del Rivero J, *et al*. Novel insights into the polycythemia-paraganglioma-somatostatinoma syndrome. *Endocr Relat Cancer*, 2016; **23**: 899–908.
13. Toyoda H, Hirayama J, Sugimoto Y, *et al*. Polycythemia and paraganglioma with a novel somatic HIF2A mutation in a male. *Pediatrics*, 2014; **133**: e1787–91.

6.11

Imaging Neuroendocrine Tumours of the Gastrointestinal Tract/ Gastroenteropancreatic Neuroendocrine Tumours (GEP-NET)

Prakash Manoharan

Introduction

The incidence of neuroendocrine neoplasms (NENs) has increased over the last decades [1]. NENs constitute a heterogeneous group of malignancies originating from cells of the endocrine (hormonal) and nervous systems. Due to the heterogeneity, these tumours have a varied clinical presentation and proliferation. The heterogeneous nature of these tumours creates a challenge with regards to diagnosis, staging, and management. Within the gastrointestinal system, NENs can be broadly defined as tumours arising from the hollow viscus such as the stomach, small bowel, appendix, colorectal regions (extrapancreatic), and from the solid organs, essentially and exclusively the pancreas (pancreatic). NENS arising from other solid organs of the gastrointestinal region are exceedingly rare [2]. They are further subdivided into non-functioning tumours, which have a higher incidence and functioning NENs producing peptides that can cause distinct clinical symptoms such as flushing and diarrhoea (carcinoid syndrome), hypoglycaemia (insulinoma), gastric ulcers (gastrinoma) or skin rash (glucagonoma). Collectively these NENs are termed gastroenteropancreatic neuroendocrine tumours [3, 4] (GEP-NETs). The key factor differentiating NENs from other types of malignancies is based on pathological findings (see Chapter 6.1). Therefore, investigating and imaging GEP-NETs is dependent on a variety of factors, including symptom profile, site of tumour, functioning versus non-functioning (hormonal profile) and tumour pathology. Tumour pathology especially the grade of the tumour (grades 1 to 3) plays a crucial role with regards to the direction and choice of imaging modality.

Imaging Modalities

Imaging continues to undergo rapid and sustained technological advances. Imaging modalities can be divided into non-invasive and invasive techniques. Non-invasive imaging modalities consist of morphological modalities such as multidetector computed tomography (MDCT), which is a cross-sectional modality and molecular imaging modalities that image biological activities such as positron emission tomography (PET) techniques. Invasive techniques include endoscopic modalities (e.g. endoscopic ultrasound) and interventional radiology (e.g. vascular angiography). More recently a process of hybridization of imaging modalities has created several new hybrid modalities that have combined traditional morphological and molecular imaging techniques to create a host of emerging and evolving modalities such as PET-MDCT. Imaging plays a key role in investigating GEP-NET patients. Due to the complexity and heterogeneity of this disease process, a multimodality approach is now routinely utilized.

Non-Invasive Morphological Imaging Modalities

With regards to conventional radiological imaging modalities, ultrasound (US), MDCT and magnetic resonance imaging (MRI), are the most routinely employed modalities in abdominal imaging and this also applies to GEP-NETs.

Ultrasound (US)

Transabdominal ultrasonography (US) is a non-ionizing, readily available non-invasive technique. The role of US in GEP-NET

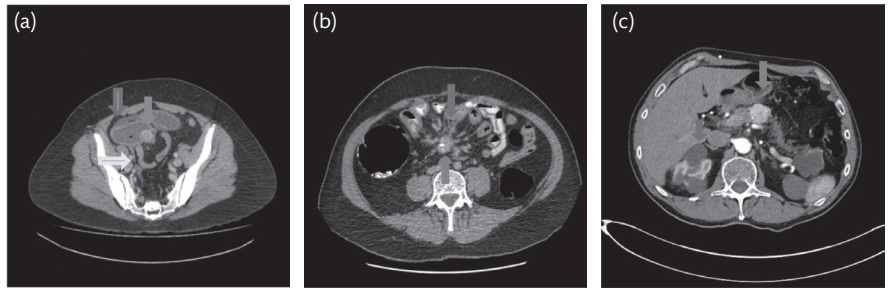


Figure 6.11.1 (a) Arterially enhancing tumour (red arrow) involving the ileal small bowel causing bowel obstruction- dilated proximal bowel loops (blue arrow) and collapsed distal small bowel loops (yellow arrow). (b) Classic example of a mesenteric nodal mass with internal calcification (red arrow) and fibrotic bands (blue arrow). (c) Arterially enhancing pancreatic body mass (red arrow) and multiple benign renal cysts (blue arrow) in a patient with Von Hippel–Lindau.

patients is limited, although it is sometimes the modality through which abnormalities in symptomatic GEP-NET are incidentally detected in patients at initial presentation. In certain circumstances, US has a role in the management of younger patients (under 16 years) and patients who are pregnant due to its ease of access and non-ionizing properties. It is not utilized as a staging tool due to difficulties in imaging the entire body and limitations in assessing abdominal organs due to overlying bowel gas. Classically, primary and metastatic GEP-NETs appear as a hypoechoic space occupying lesions surrounded by a hyperechoic halo. US, however, has a significant role in image-guided biopsies to obtain pathology specimen including sampling of more superficial sites such as lymph nodes.

MDCT

MDCT is a cross-sectional imaging modality and is considered the 'work horse' of oncology imaging modalities with high spatial and temporal resolution. MDCT is not only important in identifying a primary site of GEP-NETs but also enables whole-body assessment of tumour burden in a number of seconds. On the whole, patients find investigation with MDCT acceptable, particularly due to the speed, ease, and availability of this technique. MDCT also allows multiphasic contrast assessment and multiplanar reconstruction of images increasing the likelihood of detecting the extent of disease and aiding interventional/surgical planning. Typically, primary and metastatic GEP-NETs appear as hypervascular lesions (**Figure 6.11.1**—vascular primary lesions) [5, 6]. These lesions also demonstrate a varying degree of calcification and fibrotic changes that are usually detectable on MDCT (**Figure 6.11.1b**—calcified and fibrotic mesenteric mass). Bone metastases appear sclerotic and can demonstrate increased uptake of somatostatin cell surface receptors (SSTR) on hybrid imaging (**Figures 6.11.2**). Despite these advantages, MDCT holds some limitations depending on the grade and

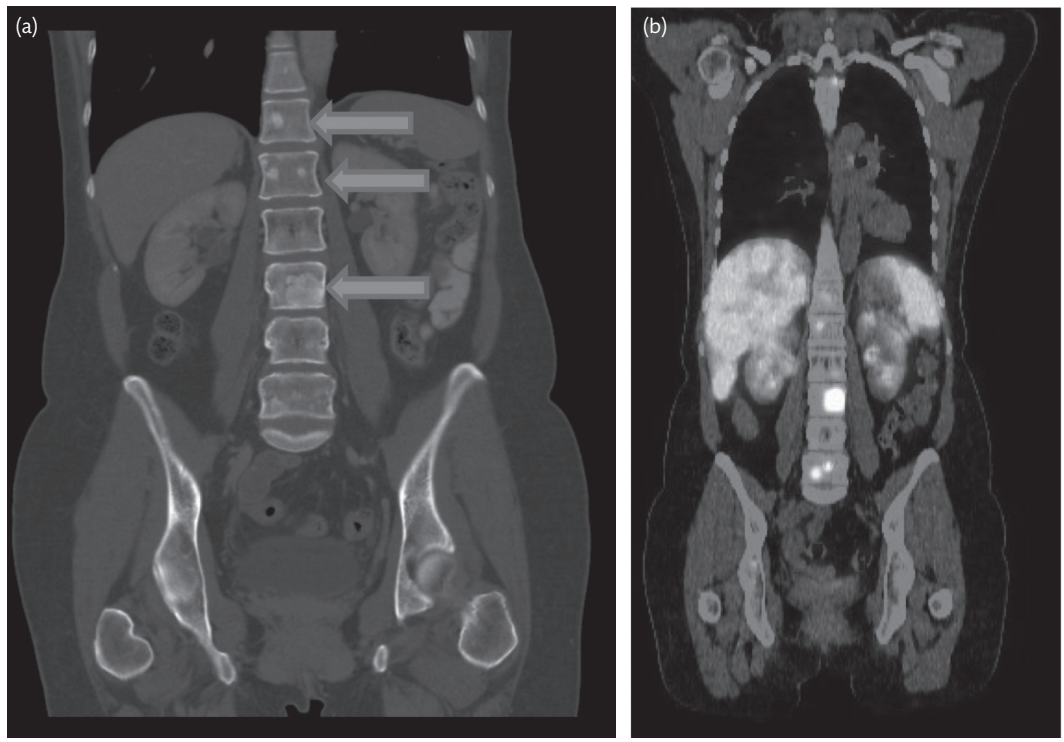


Figure 6.11.2 (a) Multiple metastatic sclerotic bone lesions (blue arrows). (b) Multiple metastatic sclerotic bone lesions demonstrating increased uptake of tracer on somatostatin cell surface receptors (SSTR) on hybrid PET-MDCT imaging.

Table 6.11.1 Cross-sectional morphological MDCT and MRI imaging protocols for GEP-NETs.

(a) MDCT, multiphasic iodinated contrast-enhanced protocol for GEP-NETS

Iodinated contrast-enhanced MDCT with pump bolus injection trigger utilizing 4 ml/s flow rate (sec) triggered by bolus tracking of the descending aorta		
Arterial	Portal venous	Pancreatic
12 sec	60–90 sec	30–45 sec

(b) MRI multiphasic contrast-enhanced/multiparametric protocol for GEP-NETS

Morphological		Diffusion-weighted imaging (DWI) clinical	Gadolinium contrast-enhanced utilizing either extracellular (hepatic and extra hepatic) or Hepatobiliary (hepatic)
T_1-weighted (T1W) imaging with multiplanar acquisition (axial, coronal, and sagittal)	T_2-weighted (T2W) imaging with multiplanar acquisition (axial, coronal, and sagittal)	A minimum of 2 b values (s/mm^2) is recommended to calculate the Apparent diffusion coefficient (ADC) ideally b values (50, 200, 400, 800)	Dynamic contrast-enhanced volumetric, fat-saturated (FS), gradient-echo T1W sequences with multiplanar reconstruction

site of tumour; mainly lack of specificity and sensitivity especially at sites involving the nodes, liver, and bones. Table 6.11.1 details a detailed multiphasic iodinated contrast-enhanced MDCT imaging protocol of GEP-NETs.

MRI

As an imaging modality MRI has superior intrinsic soft tissue contrast when compared to other technologies. This, coupled with the multiplanar and multiphasic contrast-enhanced acquisition capabilities, makes it a very important tool to assess GEP-NETs [7, 8]. Most protocols utilize a combination of T_1- and T_2-weighted sequences combined with or without fat saturation. Newer sequences such as diffusion-weighted imaging (DWI) have been incorporated into the majority of protocols. DWI sequences are particularly useful in detecting small liver and pancreatic lesions non-invasively that are occult in other cross-sectional imaging modalities (Figures 6.11.3).

MRI is a non-ionizing radiation modality. It is generally utilized as a problem-solving tool and is usually used as a more focused examination of a site of abnormality/clinical concern. Part of the limitation of MRI is that even with recent software and hardware improvements, MRI image acquisition still takes longer than MDCT. Depending on the site of examination it usually takes between 30 and 60 minutes to complete an examination. Although a non-ionizing modality, there are restrictions to MRI due to magnetic radiation safety issues [9]. In addition, access to MRI technology is more limited due to scarcity of equipment and it is also a more expensive modality. Patients tend to tolerate MRI less well when compared to MDCT due to a longer scan time and longer bore of the equipment. Patient cooperation with the scanning protocol is also quite crucial to achieve better quality images.

GEP-NETs have a varied appearance on MRI. The MRI appearance is dictated by the grading and tumour biology. Majority are of low signal intensity on T_1-weighted MR images and have a varying hyperintensity on T_2-weighted images [7, 8]. Cystic tumour changes which are a feature of a spectrum of GEP-NETs are better detected on MRI due to enhanced inherent tissue contrast. As per MDCT, on contrast-enhanced MR images, these tumours are often hypervascular in the arterial phase and show varying degrees of contrast enhancement (Figure 6.11.4).

For a detailed multiphasic gadolinium contrast-enhanced and multiparametric MRI imaging protocol please refer to Table 6.11.1.

Non-Invasive Molecular Imaging Modalities (Hybrid Molecular Imaging Modalities)

Hybrid molecular imaging is a specialized field of imaging where biological activity is imaged utilizing labelled radioactive tracers. This requires a nuclear medicine department with access to a radiopharmacy, physicist, and specialized imaging experts to interpret the scans. There are also usually stringent national and international legislative requirements in place due to the handling and administration of radioactive particles. These factors coupled with fewer scanners make this modality fairly specialized and expensive.

Single Photon Emission Tomography/Multidetector Computed Tomography (SPECT-MDCT)

SPECT-MDCT is a hybrid molecular imaging modality combining detection of biological activity after administration of an intravenous labelled gamma emitting nuclear medicine tracer with a scintillation camera and cross-sectional MDCT. Most modern nuclear medicine departments utilize these hybrid imaging systems replacing traditional stand-alone SPECT camera. With SPECT-MDCT, the sensitivity, specificity, and accuracy of detecting the biological activity including in GEP-NET is improved when compared to SPECT [10]. Therefore, in the current era, this is the nuclear medicine modality of choice and most large healthcare centres will have access to this modality. GEP-NETs have cell surface receptors and biological activities that can be detected with SPECT-MDCT. This modality is utilized for further characterization and assessment of GEP-NETs; especially grade I and II tumours. Detection of tumour or metastatic site is dependent on the size, location, biological activity, and concentration of the biological activity (e.g. receptor density) and if positive, it is seen as a focal 'hot spot' when compared to the background. There is usually a long period for the background activity to reduce to improve detection of pathology. Typically in GEP-NETs, the imaging protocol requires imaging at 4 hours post injection of a

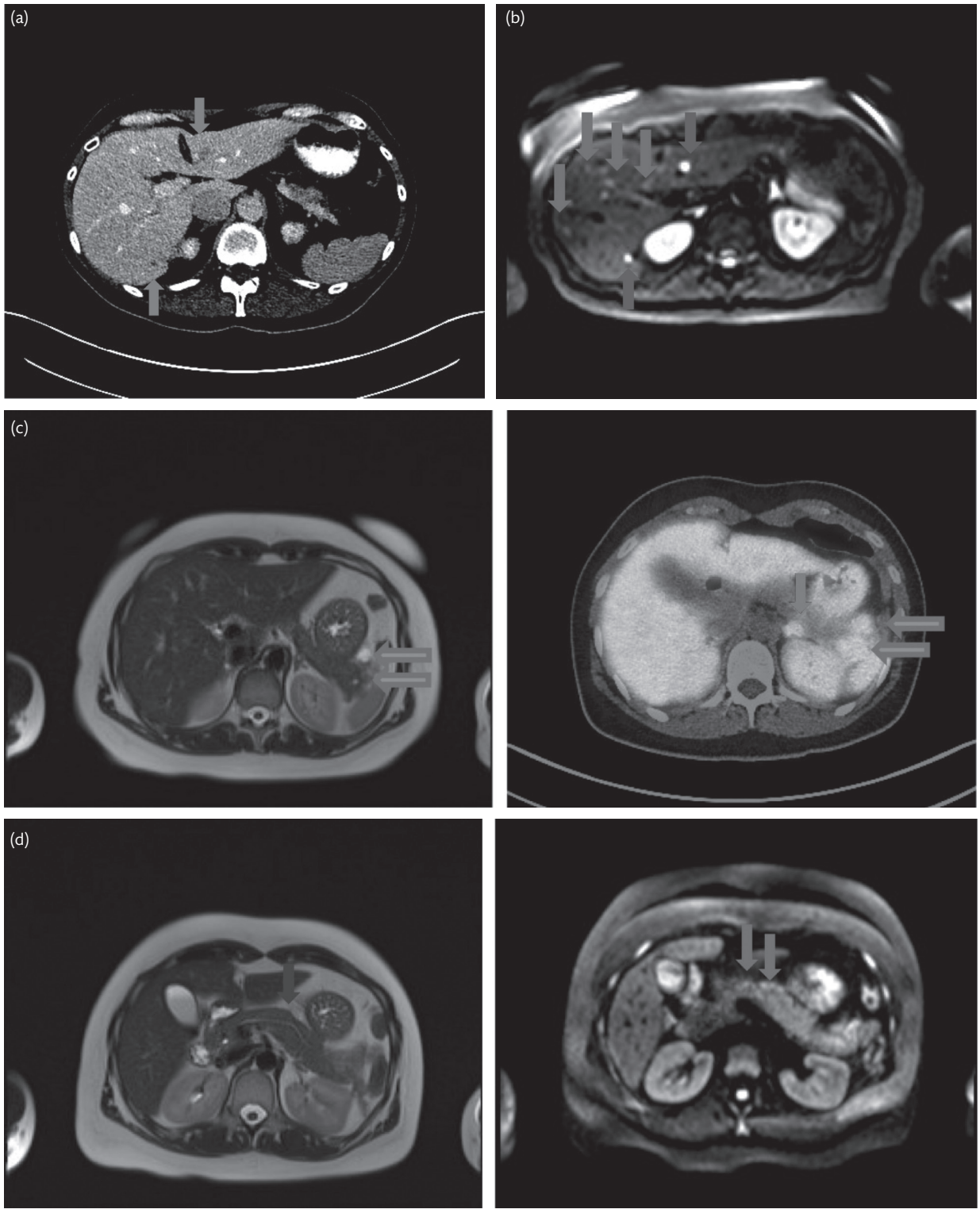


Figure 6.11.3 The IV contrast CT scan (a) detects two liver metastases (red arrows). The corresponding MRI DWI image (b) demonstrates the two liver metastases detected on the CT scan (red arrows) and multiple other liver metastases (blue arrows). (c) Patient with MEN 1. On the left, the T2W MRI of the pancreas has two partly cystic pancreatic tail pancreatic NET (blue arrows). On the right, these tumours in the tail demonstrate intense uptake of tracer on the Ga68 DOTATOC SSTR PET-CT, demonstrated here on the fused image (blue arrows). No further areas of abnormal uptake detected (intense uptake labelled with red arrow is physiological left adrenal gland uptake). (d) Patient with MEN 1 as per case 3c. On the left, T2W MRI of the pancreatic body demonstrate normal pancreatic signal (red arrow). On the right, the MRI DWI image of the body demonstrates multiple tumours non-invasively (blue arrows) which was confirmed on EUS.

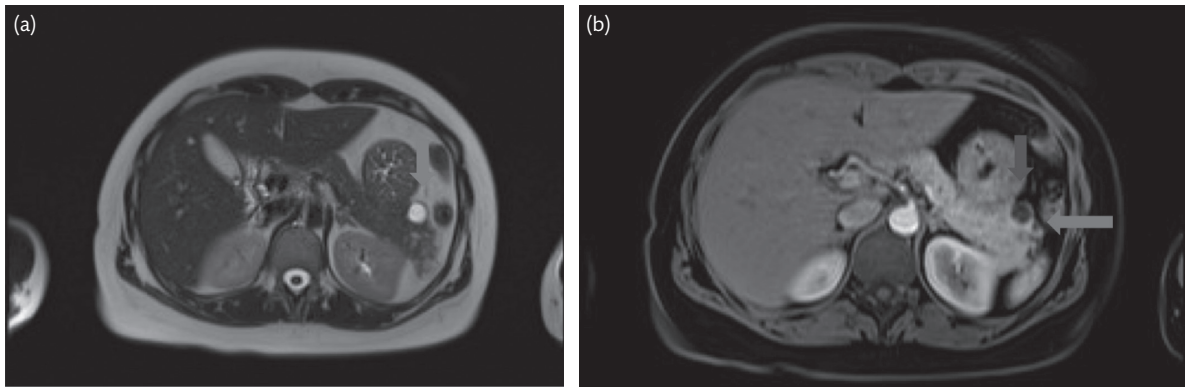


Figure 6.11.4 (a) Cystic component of the tail of pancreas NET 'bright' on T2W MRI (red arrow). (b) Post-gadolinium-enhanced T1W fat supressed MRI demonstrating non-enhancing cystic component (red arrow) and rim of arterial hypervascular enhancement (blue arrow).

radioactive tracer and then a further imaging event at 24 hours post injection of tracer.

PET-MDCT

PET-MDCT is a hybrid molecular imaging modality combining detection of biological activity after administration of an intravenous labelled positron emitting radiotracer with a PET camera and cross-sectional MDCT. Again, increase in sensitivity, specificity, and accuracy is achieved with a hybrid PET-MDCT when compared to a PET alone system [11, 12]. GEP-NETs have cell surface receptors and biological activities that can be detected with PET-MDCT. This modality is utilized for further characterization and assessment of GEP-NETs; especially grade 1 and 2 tumours (and to a lesser extent grade 3 tumours). Detection of tumour or metastatic site is dependent on the size, location, biological activity, and concentration of the biological activity (e.g. receptor density) and if positive is seen as a focal 'hot spot' when compared to the background. However, when compared to SPECT-MDCT, in the imaging of GEP-NET patients, all PET-MDCT tracer-based imaging is completed on the same day usually within 180 minutes. In general, PET-MDCT has a much higher sensitivity and specificity for like for like tracers when compared to SPECT-MDCT (Figures 6.11.5). It is considered the 'gold standard' molecular imaging modality but is reasonably expensive and thus has restricted availability.

GEP-NETs Radioactive Tracers (SPECT-MDCT and PET-MDCT)

The radiotracers utilized in imaging GEP-NETs broadly target SSTR, amine precursor uptake and decarboxylation (APUD) and glucose metabolism [13, 14]. Specific tracers and mechanism of actions including experimental novel tracers are detailed in Table 6.11.2.

Invasive Imaging Modalities: Endoscopy and Endoscopic Ultrasound (EUS)

Endoscopic Techniques

Endoscopic techniques allow the direct visualization and biopsy of tumours in hollow viscus (stomach, small bowel, and colon/rectum). It is the predominant modality to assess the stomach, proximal small bowel, and the colorectal regions. Small bowel enteroscopy, either with an endoscope or a capsule allows visualization of portions of the small bowel but it is limited by the length of small bowel and by lesions that obstruct the lumen. These modalities are operator depended but crucial in the assessment of the hollow viscus in GEP-NETs.

EUS can depict small lesions not visible on other imaging methods. This modality utilizes ultrasound probes with high frequency (7.5–12-MHz), which are positioned in proximity to the organ of interest (stomach, pancreas, duodenum, and rectum). This allows an opportunity to directly visualize, characterize, and help obtain pathological specimen via guided endoscopic biopsy [15, 16]. This modality plays an important role in assessing GEP-NETs especially in the pancreatic region.

Interventional Radiology (IR)

One of the key roles of IR is obtaining tissue for pathological assessment utilizing image guidance. There is also a growing role in therapeutic IR in GEP-NET patients. The role of IR in the diagnostic pathway of GEP-NETs however has reduced significantly especially with the advent and progress in molecular imaging. Occasionally, angiographic techniques are required to make a diagnosis of GEP-NETs (please refer to pancreas section).

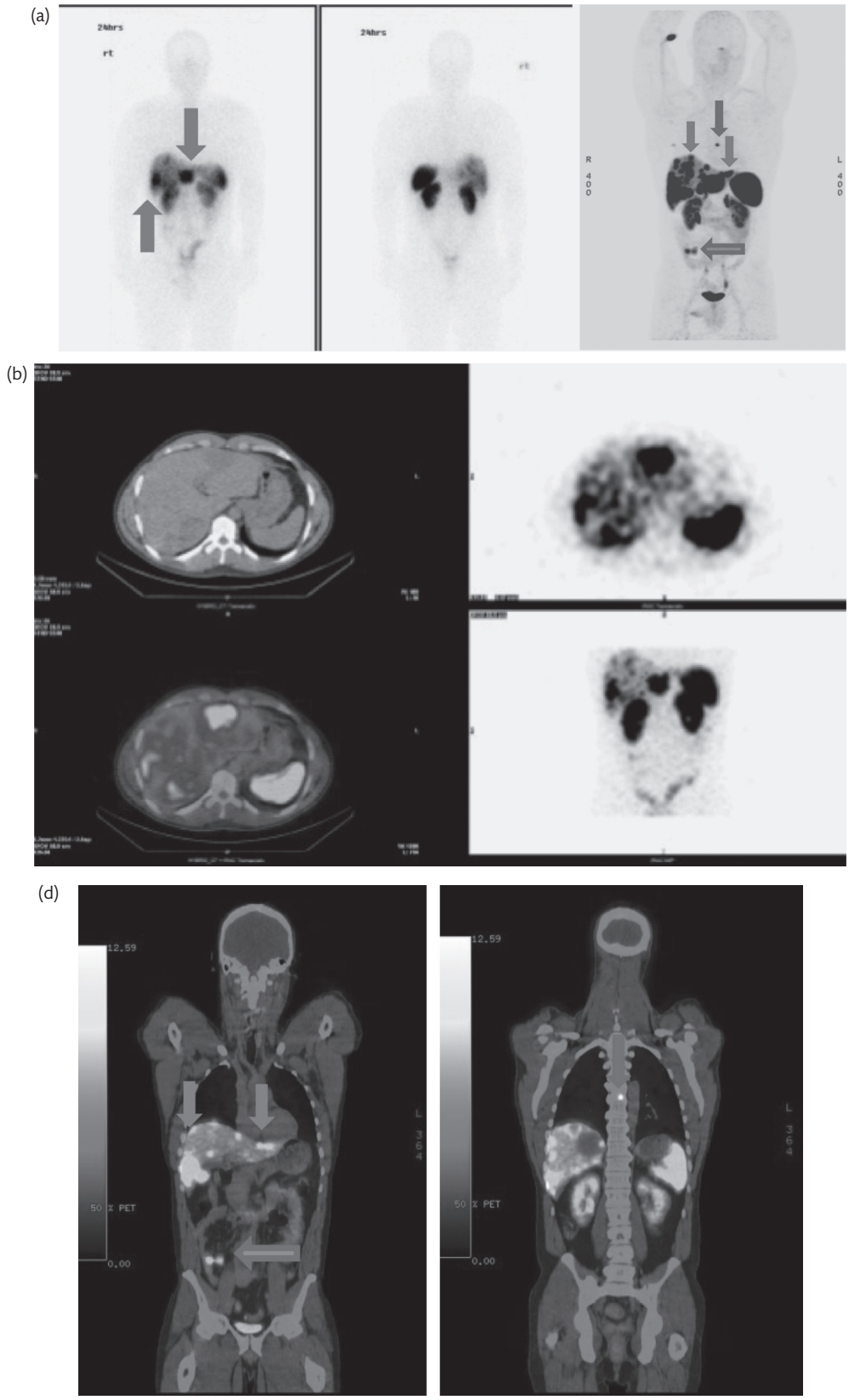
Role of Imaging Modalities in the Diagnosis, Staging, and Management of GEP-NETs

Extrapancreatic GEP-NETs

There are functioning and non-functioning extrapancreatic GEP-NETs. Functioning tumours classically present in the metastatic state with 'carcinoid syndrome'. The choice of the appropriate imaging strategy for detection and localization of GEP-NETs will largely depend on the clinical presentation however there are general strategies in setting out an imaging assessment.

Imaging/Detection of Primary Extrapancreatic GEP-NETs

Assessing and detecting suspected primary extrapancreatic GEP-NET will depend on the site of disease. Lesions in the stomach, duodenum, colon, and rectum are usually detected incidentally during endoscopic assessment in symptomatic/asymptomatic



Figures 6.11.5 Patient with grade II small bowel GEP-NET. 24-hour SPECT-MDCT (a, planar) and (b, with fused) demonstrating multiple liver metastases with increased uptake of tracer (Octreoscan®) with no further sites of abnormal uptake (red arrows). Same patient re-staged with Ga68 DOTATOC. 90 minutes PET-MDCT (c) and (d, fused) with areas of increased uptake of tracer ^{68}Ga-DOTATOC identifying primary small bowel tumour with mesenteric node (horizontal blue arrow), higher number of liver metastases (red arrows) and bone metastases (vertical blue arrow).

Table 6.11.2 Molecular imaging tracers for GEP-NETs (including novel tracers)

Tracer (trade name)	Mechanism	Clinical utility GEP-NET
(a) SPECT-MDCT		
111In-DTPA-octreotide (Octreoscan®)	SSTR agonist imaging	Grade I, II, and to a lesser extent III
99mTc-HYNIC-TOC	SSTR agonist imaging	Grade I, II, and to a lesser extent III
123I-Metaiodobenzylguanidine: 123 I-MIBG	APUD uptake	Has lower sensitivity and specificity. Not recommended for routine utility
Novel ^{111}In-DOTA-exendin-4	GLP-1R imaging	Benign insulinoma
(b) PET-MDCT		
^{68}Ga-DOTA-SSTR Peptides ^{68}Ga-DOTATOC (SomaKit TOC™) ^{68}Ga-DOTATATE (NETSPOT®)	SSTR agonist imaging	Grade I, II, and to a lesser extent III
^{18}F-Fluorodeoxyglucose (^{18}F-FDG)	Glucose metabolism	Grade III. Less of a role in grade I and II
^{18}F-Dihydroxyphenylalanine (^{18}F-DOPA)	APUD uptake	Insulinoma, grade I, and II NET
Novel ^{68}Ga-DOTA-exendin-4	GLP-1R imaging	Benign insulinoma
^{68}Ga-NODAGA-JR11	SSTR antagonist imaging	Grade I, II, and to a lesser extent III

patients. US has a very limited if any role in the wider context of imaging of primary extrapancreatic GEP-NETs. The sites of primary GEP-NETs as discussed are predominantly in gas filled viscus and therefore are difficult to assess with US. In contrast, endoscopic/EUS is essential to localize the primary lesion, assess mucosal invasion, and make a histopathologic diagnosis. Even though the appendix is one of the most common sites of extrapancreatic GEP-NETs, the vast majority of lesions in the appendix are detected incidentally in post-appendicectomy specimens in patients suspected of having appendicitis. These appendix GEP-NET lesions are also usually quite small measuring less than 1 cm. Cross-sectional and molecular imaging therefore play a very small or no role at all in the detection of primary lesions at these sites (stomach, duodenum, colon, rectum and appendix [15, 17–21]).

On the other hand, cross-sectional (MDCT/MRI) and molecular imaging (SPECT/PET-MDCT) play a crucial role in detecting extrapancreatic GEP-NET primary of ileal and jejunal small bowel origin. As discussed earlier, in general, these lesions share a number of similar features including arterial enhancement on cross-sectional imaging (MDCT/MRI). The primary disease without metastatic site can appear as mucosal, submucosal, and multifocal lesions. Not infrequently the primary tumour may also manifest on MDCT scans as asymmetric or concentric regions of mural thickening. The mural thickening occurs usually due to a combination of desmoplastic submucosal fibrosis and infiltrating tumour. This leads to thickening and stiffening of the intestinal wall, thus producing mural and irregular fold thickening (Figure 6.11.6a). There is usually associated mesenteric fibrosis and calcified nodal mass (Figure 6.11.1b). Patients present with a spectrum of clinical scenario, extending from non-specific

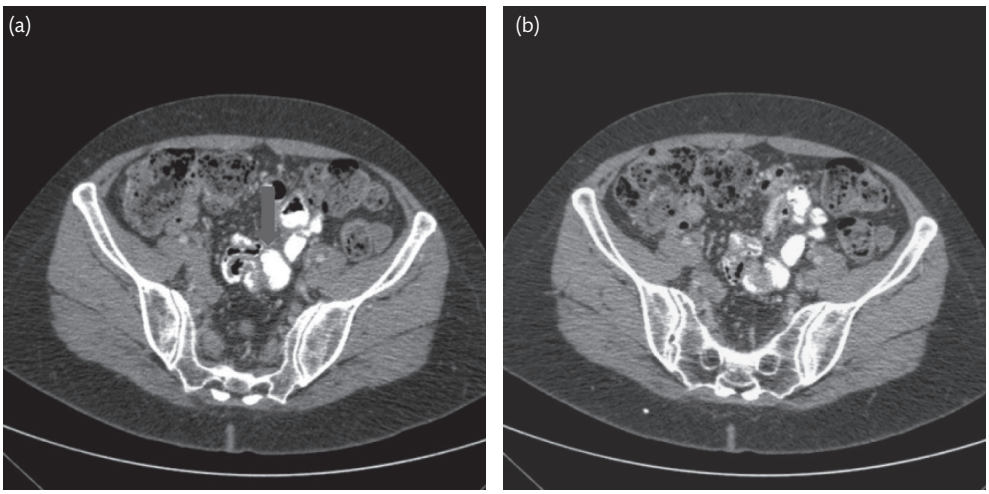


Figure 6.11.6 Patient with grade I small bowel GEP-NET. Mural thickening of primary tumour (a; red arrow) with early small bowel obstruction (b, blue arrow). 68Ga-DOTATOC MIP (c) with increased uptake of tracer in the primary small bowel site, mesenteric nodal, and liver metastases.

abdominal discomfort to intussusception and small bowel obstruction (Figure 6.11.6b). Even in cases of bowel symptomology and mechanical bowel obstruction, detection of the primary site in extrapancreatic GEP-NET with MDCT/MRI can be challenging due to the variable size of the primary disease; which can range between 43% and 50% [22]. However, utilizing MDCT and MRI cross-sectional modality but incorporating an enteroclysis protocol with intravenous contrast enhancement leads to a higher detection rate of tumours of the small bowel with reported sensitivity of detecting a primary lesion in the range of 85–94%. However, with respect to GEP-NETS, as the tumours in many cases are small nodular lesions, it may be very difficult to detect the primary lesion even with enteroclysis [23–25]. Molecular imaging fills this gap in the diagnostic pathway in the detection of occult/unknown primary extrapancreatic GEP-NETs. Molecular imaging has a higher sensitivity, specificity, and accuracy compared to cross-sectional imaging [26, 27]. As already discussed, molecular imaging detects cell surface receptors and cellular biological activity. With regards to grade I and II extrapancreatic GEP-NETs, the most sensitive molecular imaging modality to detect the primary site is somatostatin receptor imaging [26]. Most GEP-NETs have a varying amount of expression of cell surface SSTR. The expression of SSTR is dependent on the grade of the GEP-NET. Grades 1 and 2 have SSTR expression between 80% and 100%. There are currently up to five different SSTR subtypes that have been cloned and characterized (subtypes 1–5). Subtype 2 is the most expressed subtype. Therefore, in grade 1 and 2, GEP-NETs SSTR receptor detection is an integral part of the management especially with regards to detection of the primary lesion/diagnosis. It also plays an important role in staging and planning management including assessment for labelled molecular targeted radiotherapy. In the last decade, several clinical studies have compared the diagnostic accuracy of different ^{68}Ga-DOTA-SSTR Peptides (using PET-MDCT image acquisition) to ^{111}In-pentetreotide (Octreoscan®) (using SPECT-MDCT image acquisition) in patients diagnosed with NETs/GEP-NETs. Universally the results favour the use of these ^{68}Ga-DOTA-SSTR peptides in: (1) detecting small tumours or tumours bearing only a low density of SSTR; (2) offering excellent imaging properties and very high tumour/background ratios; (3) better intrinsic spatial resolution; and (4) detecting additional lesions and altering clinical management [28–30]. Therefore in the imaging and detection of extrapancreatic GEP-NETs grade I and II primary, ^{68}Ga-DOTA- SSTR Peptides PET-MDCT is currently the gold standard modality (Figure 6.11.5) (see also section on imaging metastatic extrapancreatic GEP-NETs, next). Recent advances in the synthesis and manufacturing of these tracers has led to two licensed ^{68}Ga-DOTA- SSTR Peptides; SomaKit TOC™ (Europe) and NETSPOT® (USA) [31].

Imaging Metastatic Extrapancreatic GEP-NETs

MDCT is the modality of choice in the initial imaging assessment of suspected metastatic extrapancreatic GEP-NET. This widely available modality can potentially detect and stage the entire extent of disease utilizing a multiphasic contrast-enhanced protocol (Table 6.11.1). MRI is reserved for focused assessment of anatomical sites with suspected abnormality on MDCT or as part of imaging due to risk stratification and high-risk sites of metastatic disease (bone and liver). However, these morphological cross-sectional modalities have reduced sensitivity, specificity and accuracy when compared to SSTR molecular imaging modalities with regards to detecting metastatic disease in lymph nodes, bones and liver [32] (please also see the later section on special imaging considerations). Cross-sectional modalities (MDCT and to a lesser extent MRI) are also limited with regards to assessing the biology of metastatic extrapancreatic GEP-NETs.

In the vast majority of clinical sites, the molecular imaging modality of choice would be SPECT-MDCT with Octreoscan® (partly due to the limited availability and cost of ^{68}Ga-DOTA- SSTR Peptides PET-MDCT). As well as being a superior staging modality, SSTR molecular imaging modality is the modality of choice to assess the biology of extrapancreatic GEP-NETs. The role of ^{18}F-FDG PET-MDCT is more limited but it can be instructive in the more histologically aggressive grade 3 extrapancreatic GEP-NETs. It has a growing role in the prognostic assessment of extrapancreatic GEP-NETs [33]. Also, there is growing evidence of a histological spectrum of metastatic disease that is influenced by the primary site and different metastatic sites of extrapancreatic GEP-NETs [34, 35]. This heterogeneity in biology may warrant the need to stage patients with different molecular imaging tracers. Imaging with two molecular tracers (^{18}F-FDG and ^{68}Ga-DOTA- SSTR Peptides) utilizing PET-MDCT is gaining traction. This strategy has an associated cost and is relatively scarce; thus, is best utilized for patients with atypical clinical presentation and where there is variable clinical metastatic behaviour (Figure 6.11.7).

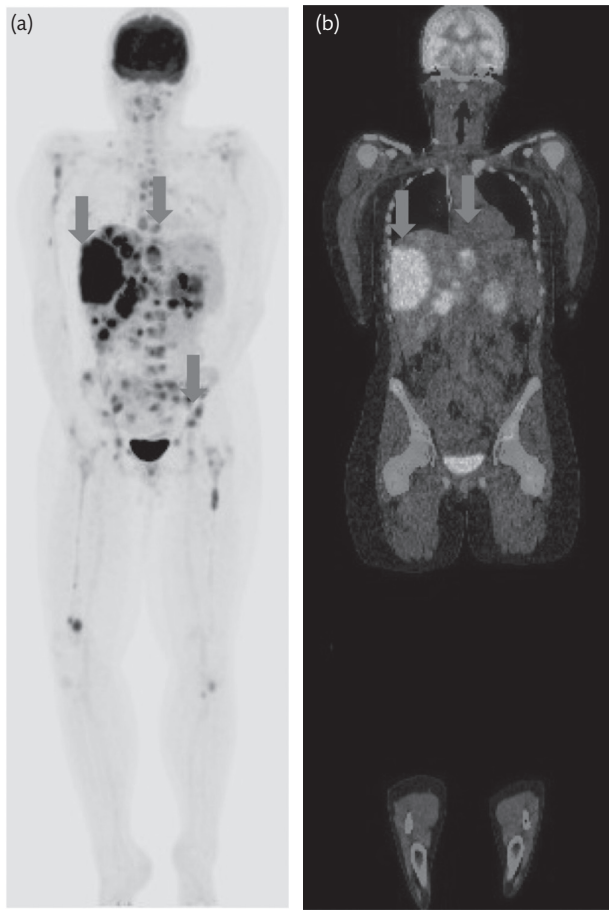


Figure 6.11.7 Patient with Grade III unknown primary GEP-NET. ^{18}F-FDG PET-MDCT (a, MIP) and (b, fused) with areas of increased uptake of tracer involving the liver, nodes, and bones (red arrows). ^{68}Ga-DOTATOC (c, MIP) and (d, fused) do not demonstrate increased uptake of tracer at metastatic sites identified on ^{18}F-FDG.

With the advent of hybrid imaging, more imaging departments are offering a 'one-stop shop' solution by combining contrast-enhanced MDCT and SPECT-MDCT/PET-MDCT in one visit. This allows the combination of high-resolution anatomical assessment along with molecular imaging, which improves sensitivity, specificity, and accuracy of staging metastatic extrapancreatic GEP-NETs [36]. With regards to this, intravenous multiphasic iodinated contrast-enhanced PET-MDCT with ^{68}Ga-DOTA- SSTR Peptides PET is currently the gold standard modality for assessing grade 1 and 2 GEP-NETs.

Pancreatic GEP-NETs

Pancreatic GEP-NETs are a heterogeneous group and typically demonstrate varied clinical presentation. Broadly, they can be divided into functioning and non-functioning tumours (each constituting approximately 50%). Functioning tumours such as insulinomas, gastrinomas, VIPomas, glucagonomas, and somatostatinomas classically present with different clinical syndromes dictated by secretion of vasoactive hormones. These are also related to hereditary and familial syndromes. Non-functioning tumours present with non-specific symptoms and are usually detected as 'incidental' findings during the course of investigating abdominal symptoms, which could be unrelated. Imaging pancreatic GEP-NETs is dictated by the clinical presentation and as per extrapancreatic GEP-NETs tumour grading, plays a crucial role in the selection of the imaging modality to further investigate this disease.

Imaging Primary Pancreatic GEP-NETs

Morphologically primary pancreatic GEP-NETs (functioning and non-functioning) share some common features on cross-sectional imaging modalities. These are usually arterially enhancing hypervascular lesions (Figure 6.11.8a is an example of an insulinoma with hypervascular blush) on MDCT/MRI, occasionally demonstrating cystic changes, can be multifocal, cause pancreatic ductal obstruction and atrophy (but to a lesser extent than pancreatic adenocarcinoma). On MRI scans these tumours demonstrate higher T2W signal than adenocarcinoma. Functioning primary pancreatic GEP-NETs tend to present earlier and are at times difficult to visualize. Two of the commonest tumours are insulinomas and gastrinomas. Insulinomas are the commonest, usually benign tumours, and are typically small at presentation, measuring less than 2 cm at time of diagnosis, and can be distributed throughout the pancreas (Figure 6.11.8a). There is a site predilection to some of the functioning tumours, as, for example, 80% of gastrinomas are found in the 'gastrinoma triangle' defined by the neck and body of the pancreas medially, confluence of the cystic and common bile duct superiorly and the second/third portions of the duodenum inferiorly. Again, gastrinomas are also typically small at presentation measuring between 1 cm and 3 cm.

Non-functioning tumours usually present as larger bulky tumours but do not exhibit as much adjacent vascular and structure invasion when compared to pancreatic adenocarcinoma. The first line modality of choice to investigate suspected tumours of the pancreas including pancreatic GEP-NETs is MDCT (Table 6.11.1). If this fails to detect a lesion, MRI is the next cross-sectional modality of choice. An EUS of the pancreas is usually, nearly always also performed for further characterization of the tumour, also providing tissue for pathological diagnosis. Furthermore, this can permit 'screening' of the pancreas as there may be multiple lesions especially in a syndromic patient. Molecular imaging has a function in detecting pancreatic GEP-NETs primary but plays an equally important role in the comprehensive staging of this disease, discussed next.

Imaging Metastatic Pancreatic GEP-NETs

There is variability with regards to the metastatic status of pancreatic GEP-NETs. Functioning tumours have differing metastatic potential. As an example, most insulinomas are 'benign' with solitary lesions but most gastrinomas are metastatic at presentation (Figure 6.11.9 demonstrates multifocal gastrinoma recurrence SSTR PET-CT imaging). As non-functioning tumours present late, there is a higher likelihood of nodal and liver metastases at diagnosis. MDCT is the modality of choice to investigate metastatic pancreatic GEP-NETs. However, MDCT underestimates metastatic disease especially metastases to nodal and liver sites. MRI has a higher detection rate of metastatic sites particularly in the liver (please refer to special imaging consideration section) but it does not provide a comprehensive whole-body staging and is utilized for focused assessment of anatomical sites.

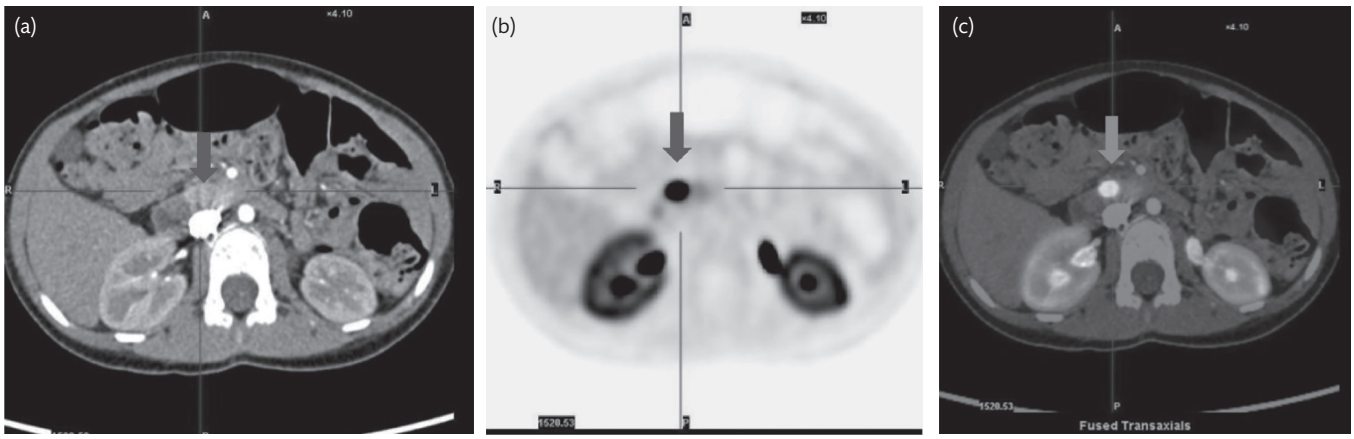


Figure 6.11.8 Patient with severe hypoglycaemia. MDCT (a) with a focus of increased enhancement in the head of the pancreas (red arrow). ^{18}F-DOPA PET-MDCT (b, PET) and (c, fused) with an area of intense focus of increased uptake of tracer at the site of increased enhancement (red arrow). Surgically proven insulinoma.

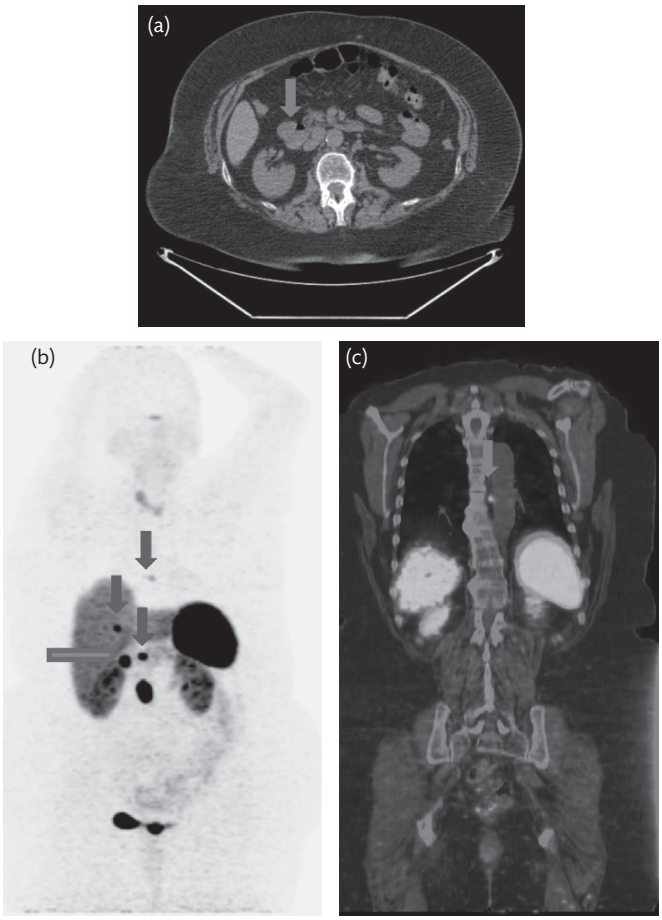


Figure 6.11.9 Patient with resected gastrinoma grade II. Rising gastrin level, presurgery MDCT detected enlarged periduodenal node (a, red arrow). ^{68}Ga-DOTATOC PET-MDCT (b) MIP confirms enlarged periduodenal node (horizontal blue arrow) and identified three further areas of metastases (red arrows): solitary liver metastasis and two metastatic nodes (b, MIP; c, fused).

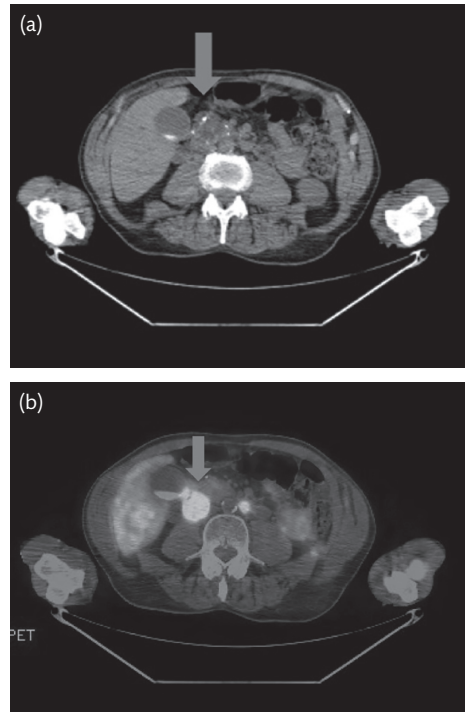


Figure 6.11.10 Patient with grade II non-functioning pancreatic GEP-NET. Planning MDCT accurately identified pancreatic head mass and no further sites of disease. Unenhanced MDCT (a) confirmed pancreatic head mass (red arrow). Fused ^{68}Ga-DOTATOC PET –MDCT (b) has concurrent finding with regards to pancreatic head mass (red arrow) but also identifies a further focus of abnormal uptake consistent with a metastatic peripancreatic node (blue arrow).

Molecular imaging has a significant and unique input into the management of primary and metastatic pancreatic GEP-NETs. The performance of the molecular tracers to detect and stage disease differs according to the type of tumour and grade. Somatostatin tracer imaging utilizing Octreoscan® and ^{68}Ga-DOTA-SSTR Peptides utilizing SPECT-MDCT and PET-MDCT, respectively, have good diagnostic sensitivity (80–100%) for nearly all grade I and II pancreatic GEP-NETs apart from insulinoma, which has poorer expression of somatostatin receptors. Pancreatic GEP-NETs apart from insulinomas imaged with ^{68}Ga-DOTA- SSTR Peptides PET-MDCT have been shown to have much better diagnostic accuracy for the detection of small primary tumours and metastatic disease (Figure 6.11.10) [37, 38]. There are also reports indicating that ^{68}Ga-DOTA- SSTR Peptides PET-MDCT is as accurate as EUS [39]; however, this finding requires further studies to confirm.

With regards to insulinomas, ^{18}F-DOPA PET-MDCT can be useful to detect benign insulinomas owing to the variable expression of SSTR in these tumours, however it has a variable sensitivity and specificity. This tracer is not available in most countries hence the scarcity of its availability (Figure 6.11.8b and c).

Despite recent advances in morphological and molecular imaging, due to the challenges in the detection of insulinomas, patients may still require an IR study. This angiographic assisted study is performed in highly specialized centres and involves selective pancreatic intra-arterial calcium stimulation of insulin release with hepatic venous sampling. This procedure has a reported localization rate of between 65% and 100% [14].

Special Imaging Consideration in GEP-NETs

Imaging Patients with Inherited/Genetic Preponderance for GEP-NETs

A subgroup of inherited genetic diseases such as MEN-1, NF1, Von Hippel-Lindau syndrome (VHL), and tuberous sclerosis complex (TSC) have a preponderance of GEP-NETs. These patients also develop GEP-NETs at an earlier age. Therefore, assessing and following up these patients requires a more nuanced utility of imaging modalities. The patients require multiorgan assessment and multitime point follow-up. As these patients present earlier and younger, repeated imaging with ionizing radiation presents a long-term risk of stochastic effects, potentially leading to a second tumour related to the radiation exposure. As per all GEP-NETs, a multimodal imaging approach is required. As with other non-familial GEP-NETs patients, ^{68}Ga-DOTA- SSTR Peptides PET-MDCT with iodinated contrast-enhanced MDCT obtained at the same time is a key imaging tool for grade I and II patients. However, after the initial diagnosis and intervention, follow-up should ideally be performed

with as little as or no exposure to ionizing radiation. Whole-body imaging MRI (WBMRI) is gradually gaining popularity for longitudinal imaging of these patients. Due to the rarity of these inherited familial diseases, comprehensive long-term strategy thus far has been lacking. A small number of recent studies have tried to address this lack of evidence-based practice, summarized in the paper by Ito *et al.* [40]. A recent multicentre, prospective study [41] found similar sensitivity between MRI and EUS in detecting pNETs. Both techniques had issues, and with regards to MRI, missed small lesions. Follow-up of this patient group remains difficult and as it currently stands, majority of the follow-up proposals are based on opinion based clinical practice [42]. A pragmatic way forward could be to utilize WBMRI for follow-up and EUS/^{68}Ga-DOTA- SSTR peptides PET-MDCT when the patient experiences biochemical or symptomatic changes.

Detecting and Staging Metastatic Liver Disease

The different modalities discussed here have differing detection capabilities depending on the region of interest. The liver is one of the commonest sites for metastatic disease in GEP-NETs, and GEP-NETs patients with liver metastases have a poorer prognosis. The problem is that due to the differing capability of the modalities to detect metastatic liver disease, the extent of disease can be easily underestimated. The gold standard for assessing metastatic liver disease is gadolinium contrast-enhanced MRI. It outperforms MDCT and the hybrid molecular imaging modalities, especially with the addition of DWI sequences [43, 44]. Therefore, a multimodal approach to imaging high-risk GEP-NETs patients is recommended. These patients should ideally be staged with ^{68}Ga-DOTA- SSTR Peptides PET-MDCT with iodinated contrast-enhanced MDCT and contrast-enhanced liver MRI with DWI (**Figure 6.11.11** illustrates the different sensitivities of the modalities in the same patient). This would serve as a reference baseline. With alterations in serum tumour biomarkers, a repeat MRI of the liver can be obtained to accurately and comprehensively identify interval changes to this organ, and thus inform management plan accordingly.

Assessing Cardiac Complications of GEP-NET

Patients with metastatic GEP-NETs (liver and nodes) often have cardiac complications secondary to systemic release of unmetabolized vasoactive 5-HIIAA. This leads to progressive valvular sclerosis (tricuspid) leading to right heart failure. Therefore, the heart is a target organ in the assessment of metastatic GEP-NETs. NT-pro-BNP and a baseline imaging with echocardiogram in a valvular heart specialist centre is recommended. Cardiac MRI might be required depending on the findings of the echocardiogram. Serial NT-pro-BNP serum levels can enable monitoring of cardiac status [45].

Follow-Up and Response Assessment

GEP-NET patients show a spectrum of disease phenotypes and trajectories, from the potentially curative disease, to those with a high burden of metastatic disease but slow progression, to a patient with

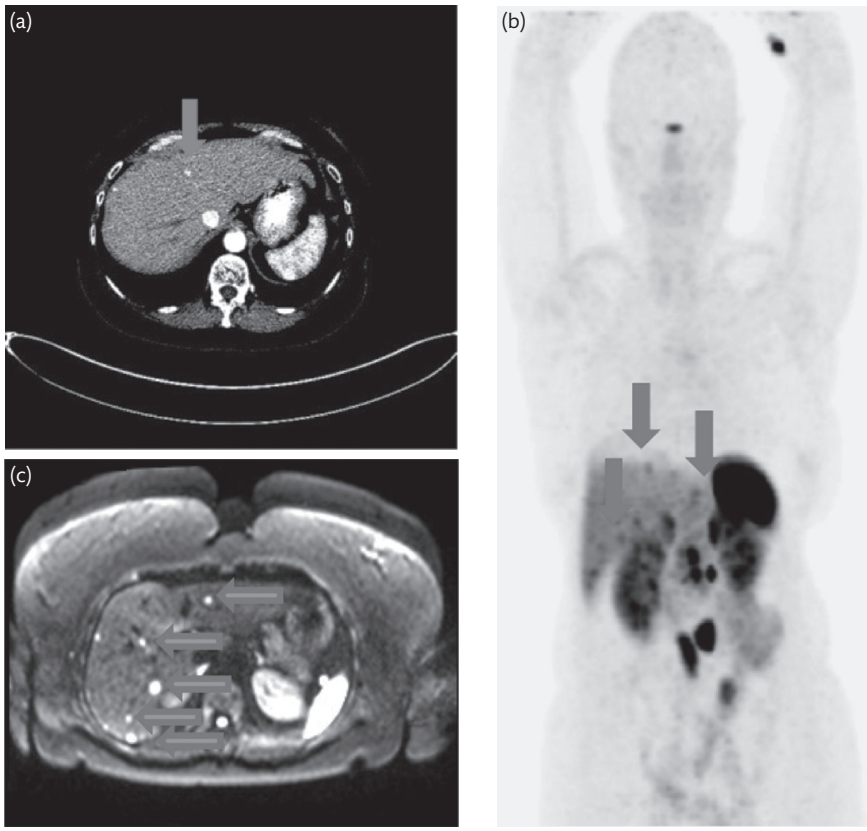


Figure 6.11.11 Patient with small bowel grade II GEP-NET. Arterial phase MDCT (a) demonstrating at least one arterially enhancing metastatic liver lesion (red arrow). Maximum intensity projection image (MIP) (b) of ^{68}Ga-DOTA-SSTR Peptides (^{68}Ga-DOTATOC) with a number of small areas of metastatic increased uptake within the liver (red arrows). DWI b (400) image (c) of the liver demonstrates vastly more restricted diffusion lesions (areas of focal high signal) consistent with multiple metastatic bilobar disease (blue horizontal arrows).

rapid progression based on tumour biology. Tailoring of imaging follow-up and response assessment is thus complex and challenging. Tumour biology drives the follow-up imaging matrix. In the vast majority of patients that have had a baseline multimodal approach including molecular imaging modalities, follow-up with MDCT is recommended. Molecular imaging, MRI, and EUS can be utilized more judiciously depending on the patient's clinical symptoms and biochemical status. This strategy is acceptable due to the exceedingly high cost and limited resource of some of these imaging modalities [46].

Future Directions: Novel Molecular Imaging Tracers and Hybrid Imaging

Glucagon-Like Peptide-1 Receptor (GLP-1R)

Almost all benign insulinomas overexpress GLP-1R. This overexpression therefore can be imaged with molecular imaging utilizing labelled GLP-1R. As already discussed labelled SSTR imaging to detect benign insulinoma has low sensitivity due to lower expression of SSTR receptors (50–60%) [13, 47, 48]. In contrast, malignant insulinoma, has lower expression of GLP-1R and express higher SSTR [49]. Radiolabelled GLP-1R analogues for SPECT/CT imaging have been developed and evaluated in patients with benign insulinomas (Table 6.11.2a) [50]. These tracers demonstrate a high benign insulinoma detection rate. Recently, it has been shown that PET-MDCT GLP-1R tracers might be more effective than GLP-1R labelled SPECT/MDCT and morphological imaging [51]. These tracers are still undergoing clinical trial assessments (Table 6.11.2b).

Somatostatin Receptor Imaging Using Antagonists

Currently all somatostatin-based radiotracers used in the clinic for diagnostic purposes are SSTR agonists. *In-vitro* and *in-vivo* data suggest SSTR antagonists have higher tumour uptake and retention [52]. The antagonist SSTR radiotracers (Table 6.11.2b) also exhibit higher number of potential binding sites on the receptor which might be responsible for higher tumour uptake [52, 53]. These tracers are currently undergoing clinical trials and appear promising [54].

Hybrid Imaging Technological Development

PET-MRI is the latest of the iterations of molecular imaging hybrid equipment. The theoretical and potential advantage of this imaging modality combining molecular imaging and exquisite MRI tissue contrast imaging is very appealing but is still under investigation. With regards to grade I/II GEP-NETs patients, recent publications of small series of patients demonstrate variation in findings compared to PET-CT [55]. As it currently stands this modality remains in the research realm.

Summary

Effective imaging of GEP-NETs relies on effective input from clinicians due to the varied clinical presentation. Optimal selection of imaging modality is driven by histopathological findings especially the grading of the tumour. While conventional cross-sectional modalities (MDCT/MRI) provide an important first step in tumour characterization, molecular imaging techniques can refine this further and thus inform management strategies. Hybrid molecular imaging (PET-MDCT) in the form of a 'one-stop' PET with iodinated intravenous contrast-enhanced MDCT together with ^{68}Ga-DOTA- SSTR Peptides represents the gold standard in GEP-NET imaging.

REFERENCES

1. Dasari A, Shen C, Halperin D, *et al*. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol*, 2017; **3**: 1335–42.
2. Yao JC, Hassan M, Phan A, *et al*. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*, 2008; **26**: 3063–72.
3. Garcia-Carbonero R, Sorbye H, Baudin E, *et al*. ENETS consensus guidelines for high-grade gastroenteropancreatic neuroendocrine tumors and neuroendocrine carcinomas. *Neuroendocrinology*, 2016; **103**: 186–94.
4. Niederle B, Pape UF, Costa F, *et al*. ENETS consensus guidelines update for neuroendocrine neoplasms of the jejunum and ileum. *Neuroendocrinology*, 2016; **103**: 125–38.
5. Dumortier J, Ratineau C, Roche C, Lombard-Bohas C, Chayvialle JA, Scoazec JY. Angiogenesis and endocrine tumors [in French]. *Bull Cancer*, 1999; **86**: 148–53.
6. Procacci C, Carbognin G, Accordini S, *et al*. Nonfunctioning endocrine tumors of the pancreas: possibilities of spiral CT characterization. *Eur Radiol*, 2001; **11**: 1175–83.
7. Tan EH, Tan CH. Imaging of gastroenteropancreatic neuroendocrine tumors. *World J Clin Oncol*, 2011; **2**: 28–43.
8. Bader TR, Semelka RC, Chiu VC, Armao DM, Woosley JT. MRI of carcinoid tumors: spectrum of appearances in the gastrointestinal tract and liver. *J Magn Reson Imaging*, 2001; **14**: 261–9.
9. Shellock FG. *MRI Bioeffects, Safety, and Patient Management*. Los Angeles: Biomedical Research Publishing Group.
10. Perri M, Erba P, Volterrani D, *et al*. Octreo-SPECT/CT imaging for accurate detection and localization of suspected neuroendocrine tumors. *Q J Nucl Med Mol Imaging*, 2008; **52**: 323–33.
11. Cerfolio RJ, Ojha B, Bryant AS, Raghuveer V, Mountz JM, Bartolucci AA. The accuracy of integrated PET-CT compared with dedicated PET alone for the staging of patients with nonsmall cell lung cancer. *Ann Thorac Surg*, 2004; **78**: 1017–23.
12. Pelosi E, Messa C, Sironi S, *et al*. Value of integrated PET/CT for lesion localisation in cancer patients: a comparative study. *Eur J Nucl Med Mol Imaging*, 2004; **31**: 932–9.
13. Baumann T, Rottenburger C, Nicolas G, Wild D. Gastroenteropancreatic neuroendocrine tumours (GEP-NET)-Imaging and staging. *Best Pract Res Clin Endocrinol Metab*, 2016; **30**: 45–57.
14. de Herder WW. Functional localisation and scintigraphy in neuroendocrine tumours of the gastrointestinal tract and pancreas (GEP-NETs). *Eur J Endocrinol*, 2014; **170**: R173–83.
15. Sahani DV, Bonaffini PA, Castillo CF-D, Blake MA. Gastroenteropancreatic neuroendocrine tumors: role of imaging in diagnosis and management. *Radiology*, 2013; **266**: 38–61.
16. Tan EH, Tan CH. Imaging of gastroenteropancreatic neuroendocrine tumors. *World J Clin Oncol*, 2011; **2**: 28–43.
17. Shapiro R, Eldar S, Sadot E, Papa MZ, Zippel DB. Appendiceal carcinoid at a large tertiary center: pathologic findings and long-term follow-up evaluation. *Am J Surg*, 2011; **201**: 805–8.

18. Chang S, Choi D, Lee SJ, *et al*. Neuroendocrine neoplasms of the gastrointestinal tract: classification, pathologic basis, and imaging features. *RadioGraphics*, 2007; **27**: 1667–79.
19. Ramage JK, Ahmed A, Ardill J, *et al*. Guidelines for the management of gastroenteropancreatic neuroendocrine (including carcinoid) tumours (NETs). *Gut*, 2012; **61**: 6–32.
20. Delle Fave G, O'Toole D, Sundin A, *et al*. ENETS consensus guidelines update for gastroduodenal neuroendocrine neoplasms. *Neuroendocrinology*, 2016; **103**: 119–24.
21. Pape UF1, Niederle B, Costa F, *et al*. ENETS consensus guidelines for neuroendocrine neoplasms of the appendix (excluding goblet cell carcinomas). *Neuroendocrinology*, 2016; **103**: 144–52.
22. Chiti A, Fanti S, Savelli G, *et al*. Comparison of somatostatin receptor imaging, computed tomography and ultrasound in the clinical management of neuroendocrine gastro-entero-pancreatic tumours. *Eur J Nucl Med*, 1998; **25**: 1396–403.
23. Soyer P, Dohan A, Eveno C, *et al*. Carcinoid tumors of the small-bowel: evaluation with 64-section CT-enteroclysis. *Eur J Radiol*, 2013; **82**: 943–50.
24. Kamaoui I, De-Luca V, Ficarelli S, *et al*. Value of CT enteroclysis in suspected small-bowel carcinoid tumors. *AJR Am J Roentgenol*, 2010; **194**: 629–33.
25. Masselli G, Polettini E, Casciani E, *et al*. Small-bowel neoplasms: prospective evaluation of MR enteroclysis. *Radiology*, 2009; **251**: 743–50.
26. Rufini V, Baum RP, Castaldi P, *et al*. Role of PET/CT in the functional imaging of endocrine pancreatic tumors. *Abdom Imaging*, 2012; **37**: 1004–20.
27. Treglia G, Castaldi P, Rindi G, Giordano A, Rufini V. Diagnostic performance of Gallium-68 somatostatin receptor PET and PET/CT in patients with thoracic and gastroenteropancreatic neuroendocrine tumours: a meta-analysis. Endocrine, 2012; **42**: 80–7.
28. Hofmann M, Maecke H, Borner R, *et al*. Biokinetics and imaging with the somatostatin receptor PET radioligand (68)Ga-DOTATOC: preliminary data. *Eur J Nucl Med*, 2001; **28**: 1751–7.
29. Buchmann I, Henze M, Engelbrecht S, *et al*. Comparison of 68Ga-DOTATOC PET and 111In-DTPAOC (Octreoscan) SPECT in patients with neuroendocrine tumours. *Eur J Nucl Med Mol Imaging*, 2007; **34**: 1617–26.
30. Srirajaskanthan R, Kayani I, Quigley AM, Soh J, Caplin ME, Bomanji J. The role of 68Ga-DOTATATE PET in patients with neuroendocrine tumors and negative or equivocal findings on 111In-DTPA-octreotide scintigraphy. *J Nucl Med*, 2010; **51**: 875–82.
31. Manoharan P, Lamarca A, Navalkissoor S, *et al*. Safety and tolerability of 'ready-to-use' (SOMAKIT TOC®) 68Ga-DOTA0-Tyr3-octreotide (68Ga-DOTATOC) for injection in patients with proven gastro-entero-pancreatic neuroendocrine tumours (GEP-NETs). *ESMO Open*, 2020; **5**: e000650.
32. Naswa N, Sharma P, Kumar A, *et al*. Gallium-68-DOTA-NOC PET/CT of patients with gastroenteropancreatic neuroendocrine tumors: a prospective single-center study. *AJR Am J Roentgenol*, 2011; **197**: 1221–8.
33. Ilett EE, Langer SW, Olsen IH, Federspiel B, Kjær A, Knigge U. Neuroendocrine carcinomas of the gastroenteropancreatic system: a comprehensive review. *Diagnostics*, 2015; **5**: 119–76.
34. Fitzgerald TL, Mosquera C, Lea CS, McMullen M. Primary site predicts grade for gastroenteropancreatic neuroendocrine tumors. *Am Surg*, 2017; **7**: 799–803.
35. Grillo F, Albertelli M, Brisigotti MP, *et al*. Grade increases in gastroenteropancreatic neuroendocrine tumor metastases compared to the primary tumor. *Neuroendocrinology*, 2016; **103**: 452–9.
36. Berzaczy D, Giraudo C, Haug AR, *et al*. Whole-body 68Ga-DOTANOC PET/MRI versus 68Ga-DOTANOC PET/CT in patients with neuroendocrine tumors. *Clin Nucl Med*, 2017; **42**: 669–74.
37. Baum RP, Prasad V. PET and PET-CT imaging of neuroendocrine tumors. In: Wahl R, ed. *Principles and Practice of PET and PET/CT*. Philadelphia, PA: Lippincott Williams & Wilkins, 2008: 411–37.
38. Schreiter NF, Brenner W, Nogami M, *et al*. Cost comparison of 111In-DTPA-octreotide scintigraphy and 68Ga-DOTATOC PET-CT for staging enteropancreatic neuroendocrine tumours. *Eur J Nucl Med*, 2012; **39**: 72–82.
39. Versari A, Camellini L, Carlinfante G, *et al*. Ga-68 DOTATOC PET, endoscopic ultrasonography, and multidetector CT in the diagnosis of duodenopancreatic neuroendocrine tumors. *Clin Nucl Med*, 2010; **35**: 321–8.
40. Ito T, Jensen RT. Imaging in multiple endocrine neoplasia type 1: recent studies show enhanced sensitivities but increased controversies. *Int J Endocr Oncol*, 2016; **3**: 53–66.
41. Barbe C, Murat A, Dupas B, *et al*. Magnetic resonance imaging versus endoscopic ultrasonography for the detection of pancreatic tumours in multiple endocrine neoplasia type 1. *Dig Liver Dis*, 2012; **44**: 228–34.
42. Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. *Lancet Diabetes Endocrinol*, 2015; **3**: 895–905.
43. Giesela FL, Kratochwila C, Mehndirattab A, Wulferta S, Moltze JH, Zechmanna CM. Comparison of neuroendocrine tumor detection and characterization using DOTATOC-PET in correlation with contrast enhanced CT and delayed contrast enhanced MRI. *Eur J Radiol*, 2012; **81**: 2820–5
44. Flechsig P, Zechmann CM, Schreiweis J, *et al*. Qualitative and quantitative image analysis of CT and MR imaging in patients with neuroendocrine liver metastases in comparison to (68)Ga-DOTATOC PET. *Eur J Radiol*, 2015; **84**: 1593–600.
45. Davar J, Connolly HM, Caplin ME, *et al*. Diagnosing and managing carcinoid heart disease in patients with neuroendocrine tumors: an expert statement. *J Am Coll Cardiol*, 2017; **69**: 1288–304.
46. Sundin A, Arnold R, Baudin E, *et al*. ENETS consensus guidelines for the standards of care in neuroendocrine tumors: radiological, nuclear medicine & hybrid imaging. *Neuroendocrinology*, 2017; **105**: 212–44.
47. Reubi JC, Waser B. Concomitant expression of several peptide receptors in neuroendocrine tumours: molecular basis for *in vivo* multireceptor tumour targeting. *Eur J Nucl Med Mol Imaging*, 2003; **30**: 781–93.
48. Krenning EP, Kwekkeboom DJ, Bakker WH, *et al*. Somatostatin receptor scintigraphy with [111In-DTPA-D-Phe1]- and [123I-Tyr3]-octreotide: the Rotterdam experience with more than 1000 patients. *Eur J Nucl Med*, 1993; **20**: 716–31.
49. Wild D, Christ E, Caplin ME, *et al*. Glucagon-like peptide-1 versus somatostatin receptor targeting reveals 2 distinct forms of malignant insulinomas. *J Nucl Med*, 2011; **52**: 1073–8.
50. Christ E, Wild D, Ederer S, *et al*. Glucagon-like peptide-1 receptor imaging for the localisation of insulinomas: a prospective multicentre imaging study. *Lancet Diabetes Endocrinol*, 2013; **1**: 115–22.

51. Antwi K, Fani M, Nicolas G, *et al*. Localization of hidden insulinomas with (6)(8)Ga-DOTA-Exendin-4 PET/CT: a pilot study. *J Nucl Med*, 2015; 56: 1075–8.
52. Ginj M, Zhang H, Waser B, *et al*. Radiolabeled somatostatin receptor antagonists are preferable to agonists for *in vivo* peptide receptor targeting of tumors. *Proc Natl Acad Sci U S A*, 2006; **103**: 16436–41.
53. Wadas TJ, Eiblmaier M, Zheleznyak A, *et al*. Preparation and biological evaluation of 64Cu-CB-TE2A-sst2-ANT, a somatostatin antagonist for PET imaging of somatostatin receptor-positive tumors. *J Nucl Med*, 2008; **49**: 1819–27.
54. Wild D, Fani M, Behe M, *et al*. First clinical evidence that imaging with somatostatin receptor antagonists is feasible. *J Nucl Med*, 2011; **52**: 1412–17.
55. Sawicki LM, Deuschl C, Beiderwellen K, *et al*. Evaluation of 68Ga-DOTATOC PET/MRI for whole-body staging of neuroendocrine tumours in comparison with 68Ga-DOTATOC PET/CT. *Eur Radiol*, 2017; **27**: 4091–9.

6.11.1 Multiple Endocrine Neoplasia Type 1

Rajesh V. Thakker

Introduction

Multiple endocrine neoplasia [1, 2] is characterized by the occurrence of tumours involving two or more endocrine glands within a single patient. The disorder has previously been referred to as multiple endocrine adenopathy (MEA) or the pluriglandular syndrome. However, glandular hyperplasia and malignancy may also occur in some patients and the term multiple endocrine neoplasia (MEN) is now preferred. Four major forms of MEN are recognized and referred to as types 1–4 (MEN 1–4). All these forms of MEN may either be inherited as autosomal dominant syndromes or they may occur sporadically (i.e. without a family history). However, this distinction between sporadic and familial cases may sometimes be difficult as in some sporadic cases the family history may be absent because the parent with the disease may have died before developing symptoms. Each form of MEN is characterized by the development of tumours within specific endocrine glands (Table 6.11.1.1). Thus, MEN 1 which is also referred to as Wermer's syndrome, is characterized by the combined occurrence of tumours of the parathyroid glands, the pancreatic islet cells, and the anterior pituitary, and is due to mutations of the tumour suppressor gene, *MEN 1* that encodes the protein menin [1]. MEN 2a (or MEN 2) and is also called Sipple's syndrome, is characterized by the occurrence of medullary thyroid carcinoma (MTC) in association with phaeochromocytomas, and parathyroid tumours. However, some patients and families may develop only MTC, and this variant is referred to as MTC-only (Table 6.11.1.1). MEN 2b (or MEN 3) is characterized by the occurrence of MTC and phaeochromocytomas, in association with: a Marfanoid habitus; mucosal neuromas of the lips, tongue, and eyelids; medullated corneal nerve fibres; and intestinal autonomic ganglion dysfunction leading to multiple diverticulae and megacolon. MEN 2a (MEN 2), MTC-only and MEN 2b (MEN 3) are due to mutations of the rearranged during transfection (*RET*) proto-oncogene, which encodes a receptor tyrosine kinase (RTK) (Table 6.11.1.1). Some patients with MEN1-associated tumours, such as parathyroid adenomas, pituitary adenomas, and pancreatic neuroendocrine tumours (NETs) who do not have mutations of the MEN 1 gene, but instead have mutations of the gene encoding the cyclin-dependent kinase inhibitor (CK1) p27kip (*CDNK1B*), are designated to have MEN 4 (Table 6.11.1.1). MEN 4 patients may also develop gonadal, adrenal, renal, and thyroid tumours. In this chapter, the main clinical features and molecular genetics of the MEN 1 syndrome will be discussed.

Clinical Features of MEN 1

Parathyroid, pancreatic, and pituitary tumours constitute the major components of MEN 1. In addition to these tumours adrenal cortical, carcinoid, facial angiofibromas, collagenomas, and lipomatous tumours may also occur in some patients [1–3]. Patients with MEN 1 have a decreased life expectancy, predominantly due to the pancreatic islet cell tumours, and the prognosis for MEN 1 patients may be improved by presymptomatic tumour detection with earlier and specific treatment for MEN 1 tumours [1].

Parathyroid Tumours

Primary hyperparathyroidism is the most common feature of MEN 1 and occurs in more than 95% of all MEN 1 patients [1]. Patients may present with asymptomatic hypercalcaemia, or nephrolithiasis, or osteitis fibrosa cystica, or vague symptoms associated with hypercalcaemia, for example polyuria, polydipsia, constipation, malaise, or occasionally with peptic ulcers. Biochemical investigations reveal hypercalcaemia, usually in association with raised circulating parathyroid hormone (PTH) concentrations. The hypercalcaemia is usually mild, and severe hypercalcaemia resulting in crisis or parathyroid carcinoma are rare occurrences [4]. Additional differences in the primary hyperparathyroidism of MEN 1 patients from that in non-MEN 1 patients include an earlier age of onset (20–25 years vs. 55 years), and an equal male:female ratio (1:1 versus 1:3) [1, 5]. Primary hyperparathyroidism in MEN 1 patients is unusual before the age of 15 years, and the age of conversion from being unaffected to affected has been observed to be between 20 and 21 years in some individuals

Table 6.11.1.1 Multiple endocrine neoplasia syndromes and their characteristic tumours and associated genetic abnormalities

Type (chromosome location)	Tumours (estimated penetrance)	Gene; Most frequently mutated codons
MEN1 (11q13)	Parathyroid adenoma (95%) Entero-pancreatic tumour (30–80%) • Gastrinoma (40%) • Insulinoma (10%) • Non-functioning & PPoma (20–55%) • Glucagonoma (<1%) • VIPoma (<1%) Pituitary adenoma (30–40%) • Prolactinoma (20%) • Somatotropinoma (10%) • Corticotropinoma (<5%) • Non-functioning (<5%) Associated tumours: • Adrenal cortical tumour (40%) • Phaeochromocytoma (<1%) • Bronchopulmonary NET (2%) • Thymic NET (2%) • Gastric NET (10%) • Lipomas (30%) • Angiofibromas (85%) • Collagenomas (70%) • Meningiomas (<10%)	*MEN1* 83/84, 4-bp del (≈4%) 119, 3-bp del (≈3%) 209–211, 4-bp del (≈8%) 418, 3-bp del (≈4%) 514–516, del or ins (≈7%) Intron 4 ss (≈10%)
MEN2 (10 cen-10q11.2)		
MEN2 (also known as MEN2A)	MTC (90%) Phaeochromocytoma (50%) Parathyroid adenoma (20–30%)	*RET* 634, missense e.g. Cys→Arg (~85%)
MTC-only	MTC (100%)	*RET* 618, missense (>50%)
MEN3 (also known as MEN2B)	MTC (>90%) Phaeochromocytoma (40–50%) Associated abnormalities (40–50%) Mucosal neuromas Marfanoid habitus Medullated corneal nerve fibres Megacolon	*RET* 918, Met→Thr (>95%)
MEN4 (12p13)	Parathyroid adenoma[a] Pituitary adenoma[a] Reproduction organ tumours[a] (e.g. testicular cancer, neuroendocrine cervical carcinoma) ?Adrenal + renal tumours[a]	*CDKN1B* No common mutations identified to date

Autosomal dominant inheritance of the MEN1 syndrome has been established. del, deletion; ins, insertion; PPoma, pancreatic polypeptide-secreting tumour; VIPoma, vasoactive intestinal polypeptide-secreting tumour; NET, neuroendocrine tumour; MTC, medullary thyroid cancer, [a] insufficient numbers reported to provide prevalence information.

Adapted from Thakker RV *et al*: Clinical Practice Guidelines for Multiple Endocrine Neoplasia Type 1 (MEN1). *J Clin Endocrinol Metab*, 2012, 97:2990–3011.

[1]. Surgical removal of the abnormally overactive parathyroids is the definitive treatment [1, 6]. However, in MEN 1 patients, all four parathyroid glands are usually affected with multiple adenomas or hyperplasia, although this histological distinction may be difficult. Thus, there is a controversy regarding the type and extent of parathyroid surgery. One opinion is to undertake total parathyroidectomy with the resultant lifelong hypocalcaemia being treated with oral calcitriol (1,25 dihydroxyvitamin D_3) while the other option is to perform a subtotal parathyroidectomy (e.g. removal of 3.5 glands) [1, 6]. Minimally invasive parathyroidectomy is not recommended because all four parathyroid glands are likely to have neoplastic changes [1]. For the timing of surgery, it is recommended that parathyroid surgery should be reserved for the symptomatic hypercalcaemic patient with MEN 1, and that the asymptomatic MEN1 patient with mild hypercalcaemia (total serum calcium <3.00 mmol/L) should not have parathyroid surgery but have regular assessments for the onset of symptoms and complications, when appropriate parathyroidectomy should be undertaken [1]. Calcimimetics (e.g. cinacalcet), which act via the calcium-sensing receptor, have been used to treat primary hyperparathyroidism in some patients when surgery is unsuccessful or contraindicated [7].

Pancreatic Tumours

The incidence of pancreatic islet cell tumours, which are NETs, in MEN 1 patients varies from 30% to 80% in different series [1–3, 8].

Many of these tumours produce excessive amounts of hormone, for example gastrin, insulin, glucagon, or vasoactive intestinal polypeptide (VIP), and are associated with distinct clinical syndromes [1, 3, 8, 9], although ~40% of these pancreatic NETs are non-secreting (i.e. non-functional). Current treatments for pancreatic NETs in MEN 1 patients are based on extrapolations from outcomes in non-MEN 1 patients, as clinical trials have not been formally undertaken in MEN 1 patients [1, 3, 8].

Gastrinoma

These gastrin-secreting tumours represent over ~40% of all pancreatic islet cell tumours in MEN 1 and approximately 20% of patients with gastrinomas will have MEN 1. Gastrinomas are the major cause of morbidity and mortality in MEN 1 patients. This is due to the recurrent, severe multiple peptic ulcers which may perforate. This association of recurrent peptic ulceration, marked gastric acid production, and non-β-islet cell tumours of the pancreas is referred to as the Zollinger–Ellison syndrome. Additional prominent clinical features of this syndrome include diarrhoea and steatorrhoea. The diagnosis is established by demonstration of a raised fasting serum gastrin concentration in association with an increased basal gastric acid secretion [1]. Medical treatment of MEN 1 patients with the Zollinger–Ellison syndrome is directed to reducing basal acid output to less than 10 mmol/L, and this may be achieved by parietal cell H^+-K^+-ATPase inhibitors (e.g. omeprazole). The ideal treatment for a non-metastatic gastrinoma is surgical excision of the gastrinoma (1). However, in patients with MEN 1 the gastrinomas are frequently multiple or extrapancreatic and the role of surgery has been controversial [8, 10, 11]. For example, in one study [10], only 16% of MEN 1 patients were free of disease immediately after surgery, and at 5 years this had declined to 6%; the respective outcomes in non-MEN 1 patients were better at 45 and 40%; however, a more recent study has reported that following surgery >75% of hypergastrinaemic MEN 1 patients were engastrinaemic, with a negative secretion provocation test at 6-month follow-up [12], and further studies are required to identify whether surgery achieves long-term remission and an improvement in survival. The treatment of disseminated gastrinomas is difficult and hormonal therapy with human somatostatin analogues (e.g. octreotide), chemotherapy with streptozotocin and 5-fluoroaracil, hepatic artery embolization, and removal of all resectable tumour have all occasionally been successful [1].

Insulinoma

These β-islet cell tumours secreting insulin represent ~10% of pancreatic tumours in MEN 1 patients [1, 3]. Insulinomas also occur in association with gastrinomas in 10% of MEN 1 patients, and the two tumours may arise at different times. Insulinomas occur more often in MEN 1 patients who are below the age of 40 years, and many of these arise in individuals before the age of 20 years [1], whereas in non-MEN1 patients insulinomas generally occur in those above the age of 40 years. Insulinomas may be the first manifestation of MEN 1 in 10% of patients and <5% of patients presenting with insulinoma will have MEN 1. Patients with an insulinoma present with hypoglycaemic symptoms, which develop after a fast or exertion and improve after glucose intake. Biochemical investigations reveal raised plasma insulin concentrations in association with hypoglycaemia. Circulating concentrations of C-peptide and proinsulin, which are also raised, may be useful in establishing the diagnosis. It is important to demonstrate the absence of sulphonylureas in plasma and urine samples obtained during investigation of hypoglycaemia. Medical treatment, which consists of frequent carbohydrate feeds and diazoxide, may be useful in the short-term, with surgery being the definitive treatment. Preoperative localization with computed tomography scanning, coeliac axis angiography, and preoperative percutaneous transhepatic portal venous sampling is difficult and success rates have varied. Surgical treatment, which ranges from enucleation of a single tumour to a distal pancreatectomy or partial pancreatectomy, has been curative in some patients [1]. For metastatic disease, chemotherapy, which consists of streptozotocin, 5-fluorouracil and doxorubicin, and hepatic artery embolization, has been used with success in some patients [1, 3, 8].

Glucagonoma

These α-islet cell, glucagon-secreting pancreatic tumours occur in less than 1% of MEN 1 patients [1, 3]. The characteristic clinical manifestations of a skin rash (necrolytic migratory erythema), weight loss, anaemia, and stomatitis may be absent and the presence of the tumour may be indicated only by glucose intolerance and hyperglucagonaemia [1]. The tail of the pancreas is the most frequent site for glucagonomas and surgical removal of these is the treatment of choice. However, treatment may be difficult as 50% of patients have metastases at the time of diagnosis. Medical treatment of these with somatostatin analogues, or with streptozotocin and 5-fluorouracil, and hepatic artery embolization has been successful in some patients [1, 3, 8] (see Chapter 6.8, 'Glucagonoma').

Vasoactive Intestinal Peptide (VIP) Tumours (VIPomas)

VIPomas have been reported in only a few patients with MEN 1, who develop watery diarrhoea, hypokalaemia, and achlorhydria, referred to as the watery diarrhoea, hypokalaemia, and achlorhydria (WDHA) syndrome. This clinical syndrome has also been referred to as the Verner–Morrison syndrome or the VIPoma syndrome. The diagnosis is established by excluding laxative and diuretic abuse, confirming a stool volume in excess of 0.5–1.0 L/day during a fast, and documenting a markedly raised plasma VIP concentration [1]. Surgical management of VIPomas, which are mostly located in the tail of the pancreas, has been curative. However, in patients with unresectable tumour, treatment with somatostatin analogues, streptozotocin, corticosteroids, indomethicin, metoclopramide, and lithium carbonate has proved beneficial, along with hepatic artery embolization for treatment of metastases [1, 3, 8] (see Chapter 6.9, 'VIPomas').

Pancreatic Polypeptide-Secreting Tumours (PPomas) and Non-Functioning Pancreatic NETs

PPomas are found in ~40% of patients with MEN 1 [1, 9]. No pathological sequelae of excessive pancreatic polypeptide secretion are apparent and the clinical significance of PP is unknown. Many PPomas have been unrecognized or classified as non-functioning

pancreatic NETs, which likely represents the most common enteropancreatic NET in MEN 1 [1, 3, 8, 9]. The absence of both a clinical syndrome and specific biochemical abnormalities may result in a delayed diagnosis of non-functioning pancreatic NETs, which are associated with a worse prognosis than other functioning tumours, including insulinoma and gastrinoma [1, 8, 9].

The management of non-functioning pancreatic NETs is controversial, as the patients are usually asymptomatic [9, 13, 14]. One recommendation is to undertake surgery irrespective of tumour size, after biochemical assessment is complete. However, most centres recommend surgery based on tumour size, using either tumour diameters of >1 cm or >2 cm as the criteria for surgery [1]. Pancreatic duodenal surgery is successful in removing tumours in 80% of patients, but >40% of patients develop complications including diabetes mellitus, frequent steatorrhoea, dumping syndromes, and other gastrointestinal symptoms, although 50–60% of non-MEN 1 patients treated surgically survive >5 years [1, 8, 15]. However, occult metastatic disease is likely to be present in these and MEN 1 patients at the time of presentation [16], and inhibitors of RTKs (e.g. sunitinib) and of the mammalian target of rapamycin (mTOR) signalling pathway (e.g. everolimus) have been reported to be effective in treating such pancreatic NETs in non-MEN 1 patients and in doubling the progression-free survival time [17, 18].

Treatment of Advanced and Metastatic Pancreatic NETs

The choice of optimal antitumour therapies for advanced metastatic pancreatic NETs in MEN 1 patients remain a challenge, as such therapies have not been formally evaluated in MEN 1 patients [3, 8]. The treatments comprise locoregional and systemic therapies and are indicated for hormonal hypersecretion, symptoms directly related to metastases, and tumour progression [3, 8]. Locoregional therapies consist of cytoreductive surgery, radiofrequency ablation, and transarterial embolization/chemoembolization [3, 8]. Systemic therapies include: somatostatin analogues (e.g. octreotide and lanreotide); chemotherapy (usually combination therapy) using cytotoxic drugs (e.g. streptozotocin, 5-fluorouracil, doxorubicin, cisplatin, etoposide) that target different components of the cell cycle); RTK inhibitors (e.g. sunitinib); mTOR pathway inhibitors (e.g. everolimus); and peptide receptor radionuclide therapy, in which a somatostatin analogue (e.g. octreotide, octreotate, DOTATATE and DOTATOC) is labelled with a β-emitting nuclide such as lutetium-177 (^{177}Lu) or yttrium-90 (^{90}Y) [3, 8].

Pituitary Tumours

The incidence of pituitary tumours in MEN 1 patients varies from 15 to 90% in different series [1, 3]. Approximately 60% of MEN 1 associated pituitary tumours secrete prolactin, less than 25% secrete growth hormone, 5% secrete adrenocorticotrophic hormone (ACTH), and the remainder appear to be non-functioning. Prolactinomas may be the first manifestation of MEN 1 in less than 10% of patients and somatotropinomas occur more often in patients over the age of 40 years [1]. Less than 3% of patients with anterior pituitary tumours will have MEN 1. The clinical manifestations depend upon the size of the pituitary tumour and its product of secretion. Enlarging pituitary tumours may compress adjacent structures such as the optic chiasm or normal pituitary tissue and cause bitemporal hemianopia or hypopituitarism, respectively. The tumour size and extension are assessed radiologically by MRI, and if this is contraindicated by computed tomography (CT) scanning. Treatment of pituitary tumours in MEN 1 patients is similar to that in non-MEN 1 patients and consists of medical therapy (e.g. bromocriptine or cabergoline for prolactinoma; or octreotide or lanreotide for somatotropinoma) or selective hypophysectomy by the transphenoidal approach if feasible, with radiotherapy being reserved for residual unresectable tumour [1].

Associated Tumours

Patients with MEN 1 may have tumours involving glands other than the parathyroids, pancreas, and pituitary. Thus NETs, adrenal cortical, facial angiofibromas, collagenomas, thyroid, and lipomatous tumours have been described in association with MEN 1 (1).

Other NETs Associated with MEN 1

Other NETs occur in more than 3% of patients with MEN 1, and may be located in the bronchi, gastrointestinal tract, pancreas, or thymus [1]. Bronchial NETs (carcinoids) in MEN 1 patients predominantly occur in women (M:F = 1:4) whereas thymic carcinoids predominantly occur in men, with cigarette smokers having a higher risk of developing tumours. Most patients are asymptomatic and do not suffer from the flushing attacks and dyspnoea associated with the carcinoid syndrome, which usually develops after the tumour has metastasized to the liver. Somatostatin analogues have been successfully used to treat symptoms and may in some patients result in regression of gastric carcinoids [1, 19].

Adrenal Cortical Tumours

The incidence of asymptomatic adrenal cortical tumours in MEN 1 patients has been reported to be as high as 40% [1]. The majority of these tumours are non-functioning. However, functioning adrenal cortical tumours in MEN 1 patients have been documented to cause hypercortisolaemia and Cushing's syndrome, and primary hyperaldosteronism, as in Conn's syndrome [1].

Lipomas

Lipomas may occur in more than 30% of patients [1], and frequently they are multiple. In addition, pleural or retroperitoneal lipomas may also occur in patients with MEN 1.

Thyroid Tumours

Thyroid tumours consisting of adenomas, colloid goitres, and carcinomas have been reported to occur in over 25% of MEN 1 patients [1]. However, the prevalence of thyroid disorders in the general population is high and it has been suggested that the association of thyroid abnormalities in MEN 1 patients may be incidental and not significant.

Facial Angiofibromas and Collagenomas

Multiple facial angiofibromas, which are similar to those observed in patients with tuberous sclerosis, have been observed in >85% of

MEN 1 patients [1] and collagenomas have been reported in >70% of MEN 1 patients [1].

Molecular Pathology and Function of the *MEN 1* Gene

The gene causing MEN 1 is located on chromosome 11q13 and encodes a 610 amino acid protein called menin [1, 2]. The *MEN 1* gene, which is a tumour suppressor gene, consists of 10 exons with a 1830-bp coding region, and to date over 1300 germline and 200 somatic mutations of the *MEN 1* gene have been identified [2, 20]. The majority (>70%) of these *MEN 1* mutations are inactivating, and are consistent with its role as a tumour suppressor gene [2, 20]. These mutations are diverse in their types and approximately 25% are nonsense mutations, approximately 40% are frameshift deletions or insertions, approximately 5% are in-frame deletions or insertions, approximately 10% are splice site mutations, approximately 20% are missense mutations, and less than 1% are whole or partial gene deletions. More than 10% of the *MEN 1* mutations arise *de novo* and may be transmitted to subsequent generations [1, 2]. It is also important to note that between 5% and 10% of MEN 1 patients may not harbour mutations in the coding region of the *MEN 1* gene [1, 2], and that these individuals may have mutations in the promoter or untranslated regions, which remain to be investigated. The mutations are not only diverse in their types but are also scattered throughout the 1830-bp coding region of the *MEN 1* gene with no evidence for clustering as observed in MEN 2a (MEN 2) (Table 6.11.1.1 and Chapter 6.11.2, 'Multiple Endocrine Neoplasia Type 2a and 2b'). Correlations between the *MEN 1* mutations and the clinical manifestations of the disorder appear to be absent [1, 2]. Tumours from MEN 1 patients and non-MEN 1 patients have been observed to harbour the germ line mutation together with a somatic loss of heterozygosity (LOH) involving chromosome 11q13, as expected from Knudson's model and the proposed role of the *MEN 1* gene as a tumour suppressor [1, 2]. Menin, which has been shown to have three nuclear localization sites (NLSs) and to be located predominantly in the nucleus, is a ubiquitously expressed protein that functions as a nuclear scaffold protein with roles in transcriptional regulation, genome stability, cell division, cell cycle control and epigenetic regulation [1–3].

MEN 1 Mutational Analysis in Clinical Practice

MEN 1 mutational analysis is helpful in clinical practice in several ways that include: (1) confirmation of the clinical diagnosis; (2) identification of family members who harbour the *MEN 1* mutation and require screening for tumour detection and early/appropriate treatment; and (3) identification of the 50% of family members who do not harbour the familial germline *MEN 1* mutation and can therefore be reassured and alleviated of the anxiety burden of developing future tumours [1]. This latter aspect cannot be overemphasized as it helps to reduce the cost to the individuals and their children, and also to the health services in not having to undertake unnecessary biochemical and radiological investigations (Table 6.11.1.2) [1].

Indications for *MEN 1* Mutational Analysis

The guidelines [1] recommend that *MEN 1* mutational analysis should be undertaken in: (1) an index case with two or more MEN 1-associated endocrine tumours (i.e. parathyroid, pancreatic or pituitary tumours); (2) asymptomatic first-degree relatives of a known *MEN1* mutation carrier; (3) a first-degree relative of an *MEN 1* mutation carrier expressing familial MEN 1 (i.e. having symptoms, signs, biochemical, or radiological evidence for one or more MEN 1-associated tumours); or (4) in patients with suspicious or atypical MEN 1, which includes individuals with parathyroid adenomas occurring before the age of 30 years or multigland parathyroid disease, gastrinoma, or multiple pancreatic NETs at any age, or individuals who have two or more MEN 1-associated tumours that are not part of the classical triad of parathyroid, pancreatic islet, and anterior pituitary tumours (e.g. parathyroid tumour plus adrenal tumour). Such mutational analysis may be undertaken in children within the first decade because children with MEN 1-tumours have been reported by the age of 10 years, and appropriate intervention in the form of biochemical testing or treatment or both has been considered [1, 9, 21].

A DNA test identifying an individual, who may be an asymptomatic relative of a patient with MEN 1, as a mutant gene carrier

Table 6.11.1.2 Suggested biochemical and radiological screening in individuals at high risk of developing MEN 1

Tumour	Age to begin (yr)	Biochemical test (plasma or serum) annually	Imaging test (time interval)
Parathyroid	8	Calcium, PTH	None
Pancreatic NETs			
Gastrinoma	20	Gastrin (± gastric pH)	None
Insulinoma	5	Fasting glucose, insulin	None
Other Pancreatic NET	<10	Chromogranin A; pancreatic polypeptide, glucagon, vasoactive intestinal peptide	MRI, CT, or EUS (annually)
Anterior pituitary	5	Prolactin, IGF-1	MRI (every 3 years)
Adrenal	<10	None unless symptoms or signs of functioning tumour and/or tumour >1 cm identified on imaging	MRI or CT (annually with pancreatic imaging)
Thymic and Bronchial carcinoid	15	None	CT or MRI (every 1–2 years)

PTH, parathyroid hormone; NETs, neuroendocrine tumours; IGF-1, insulin-like growth factor 1; CT, Computed tomography; MRI, magnetic resonance imaging; EUS, endoscopic ultrasound.

Adapted from Thakker RV *et al.*: Clinical Practice Guidelines for Multiple Endocrine Neoplasia Type 1 (MEN1). *J Clin Endocrinol Metab*, 2012, 97:2990–3011.

is likely to lead to earlier and more frequent biochemical and radiologic screening rather than to immediate medical or surgical treatment. In contrast, those relatives who do not harbour the *MEN 1* mutation have their risk of developing MEN 1-associated endocrine tumours markedly decreased from 1 in 2 for an autosomal dominant disorder, to that of the general population, thereby freeing these relatives without the *MEN 1* mutation from the requirement for further repeated clinical investigations (Figure 6.11.1.1). Thus, identification of *MEN 1* mutations may be of help in the clinical management of patients and their families with this disorder. Finally, *MEN 1* mutational analysis in a symptomatic family member (i.e. an individual already showing a clinical manifestation of MEN 1), from a family with a known *MEN 1* mutation, has been challenged as being unnecessary to establish the diagnosis of MEN 1. However, studies have reported that 5–10% of MEN1 kindreds have the occurrence of phenocopies, which may confound the diagnosis, and therefore the guidelines suggest that MEN 1 family members with one MEN 1-associated tumour should be offered *MEN 1* mutation analysis [1, 22, 23].

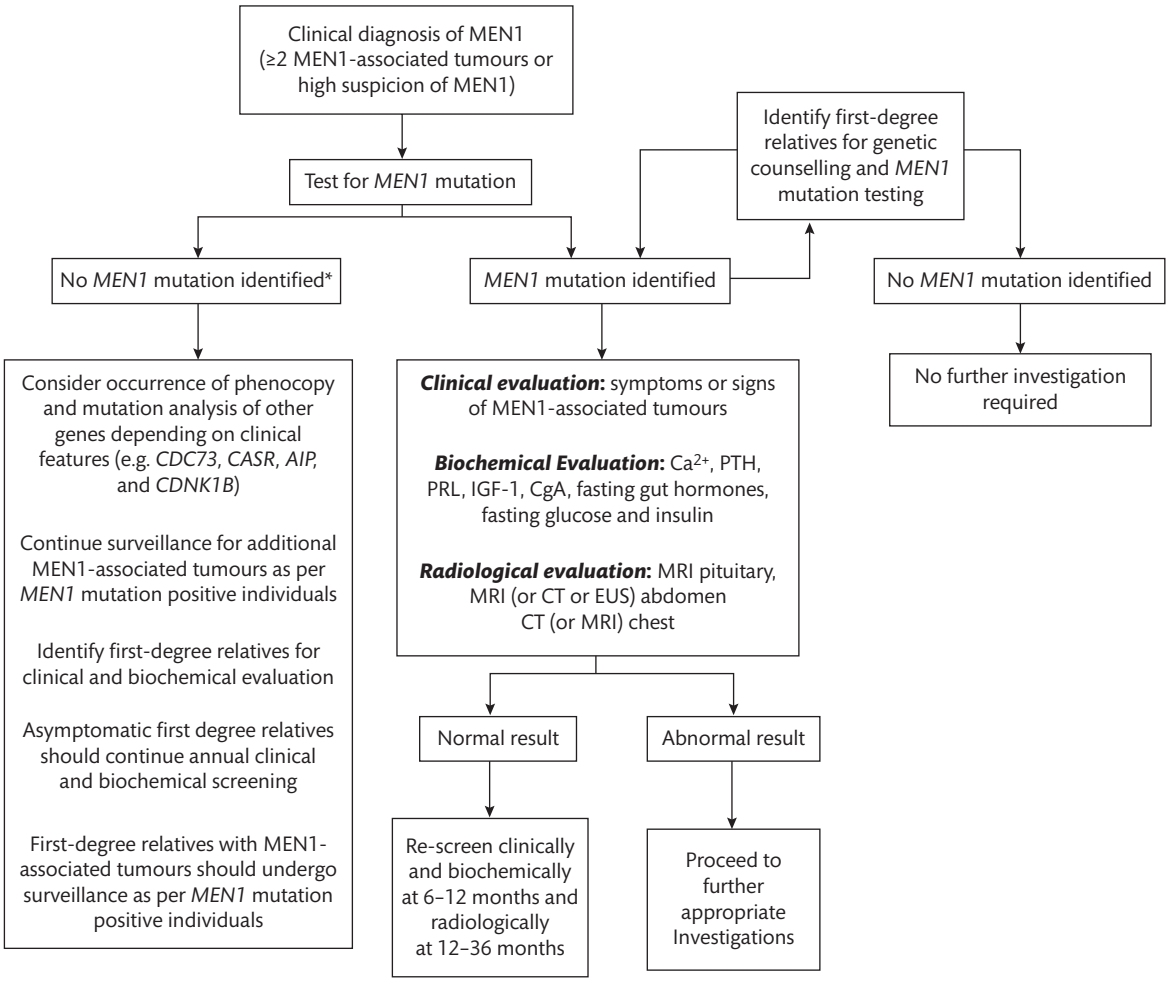


Figure 6.11.1.1 An approach to screening in MEN 1. Index cases, or individuals in whom there is a high suspicion of clinical MEN 1 (e.g. multigland parathyroid disease, parathyroid + adrenal tumour), should be offered genetic counselling and *MEN 1* mutation testing. Mutation testing should also be offered to those with familial MEN 1 (i.e. individual with one MEN 1-associated tumour and a first-degree relative with a known *MEN 1* mutation). The identification of a germline MEN 1 mutation should prompt entry into a periodic clinical, biochemical, and radiological screening programme. At the same time, first-degree relatives should be identified and offered genetic counselling and *MEN 1* mutation testing. Individuals who have inherited the *MEN 1* mutation should enter periodic screening, even if asymptomatic. First-degree relatives who have not inherited the MEN 1 mutation require no further follow-up and may be alleviated of the anxiety associated with the development of MEN 1-associated tumours. For index cases, in whom a *MEN 1* mutation, which includes testing for partial or whole gene deletions (asterisked), is not identified, additional genetic testing may be required depending on the specific clinical features. This may include examination for mutations in genes associated with familial endocrine tumour syndromes including: *CDC73* associated with the hyperparathyroidism-jaw tumour syndrome (HPT-JT); the calcium-sensing receptor (*CASR*) associated with familial benign hypercalciuric hypercalcaemia (FBHH); cyclin-dependent kinase 1B (*CDKN1B*) associated with MEN 4; and aryl hydrocarbon receptor interacting protein (*AIP*) associated with familial isolated pituitary adenomas (FIPA). Up to 10% of kindreds with clinical MEN 1 may harbour phenocopies emphasizing the importance of accurate genetic evaluation. For MEN 1 kindreds in whom no *MEN 1* mutation is identified a pragmatic approach is to offer clinical, biochemical and radiological screening to those with clinical manifestations of disease and to offer annual clinical and biochemical screening to asymptomatic first-degree relatives. Ca^{2+}, calcium; PTH, parathyroid hormone; PRL, prolactin; IGF-1, insulin-growth-factor-1; CgA, chromogranin A; MRI, magnetic resonance imaging; CT, computer tomography; EUS, endoscopic ultrasound.

Screening for MEN 1 Tumours

Biochemical screening for the development of MEN 1 tumours in asymptomatic members of families with MEN 1 is likely to be of benefit in as much as earlier diagnosis and treatment of these tumours may help reduce morbidity and mortality (Figure 6.11.1.1) [1]. Age-related penetrance (i.e. the proportion of gene carriers manifesting symptoms or signs of the disease by a given age) has been ascertained, and the mutation appears to be non-penetrant in those younger than 5 years [21]. Thereafter, the mutant *MEN 1* gene has a high penetrance, >50% penetrant by 20 years of age and >95% by 40 years [21]. Screening for MEN 1 tumours may be difficult because clinical and biochemical manifestations in members of any one family are not uniformly similar [1, 2]. Attempts to screen for development of MEN 1 tumours in the asymptomatic relatives of an affected individual have depended largely on measuring serum concentrations of calcium, gastrointestinal hormones (e.g. gastrin), prolactin and insulin-like growth factor (IGF-1), as well as on abdominal and pituitary imaging (Table 6.11.1.2). Parathyroid overactivity causing hypercalcaemia is almost invariably the first manifestation of the disorder and has become a useful and easy biochemical screening investigation [1]. In addition, hyperprolactinemia, which may be asymptomatic, may represent the first manifestation in ~15% of patients and may thus also be a helpful and an easy biochemical screening investigation [1]. Pancreatic involvement in asymptomatic individuals has been detected by measuring fasting plasma concentrations of gastrin, PP, glucagon, and chromogranin A and by abdominal imaging [1]. The guidelines suggest that individuals at high risk for MEN 1 (i.e. mutant gene carriers) undergo biochemical screening (Figure 6.11.1.1) at least once per annum and also have baseline pituitary and abdominal imaging (e.g. MRI or CT), which should then be repeated at 1- to 3-year intervals (Table 6.11.1.2) [1]. Screening should possibly commence in early childhood because the disease has developed in some individuals by the age of 5 years, and it should be repeated throughout life because the disease may not manifest in some individuals until the eighth decade [1].

Conclusions

Multiple endocrine neoplasia type 1 (MEN 1) is characterized by the occurrence of parathyroid, pancreatic islet, and anterior pituitary tumours. Some patients may also develop carcinoid tumours, adrenal cortical tumours, meningiomas, facial angiofibromas, collagenomas, and lipomas. MEN 1 is inherited as an autosomal dominant disorder that is due to mutations in the tumour suppressor gene *MEN 1*, which encodes a 610-amino acid protein, menin. Patients with MEN 1 have a decreased life expectancy. Moreover, the finding of MEN 1 in a patient also has important implications for family members, because first-degree relatives have a 50% risk of developing the disease, and these at-risk relatives can be identified by *MEN 1* mutational analysis. The prognosis of MEN 1 patients and that of the at-risk relatives may be improved by presymptomatic detection of the tumours and with implementation of earlier and specific treatments for the MEN 1 tumours.

ACKNOWLEDGEMENTS

I am grateful to Mrs Tracey Walker for her expert secretarial assistance.

REFERENCES

1. Thakker RV, Newey PJ, Walls GV, *et al*. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*, 2012; **97**: 2990–3011.
2. Lemos M, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat*, 2008; **29**: 22–32.
3. Frost M, Lines KE, Thakker RV. Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol*, 2018; **14**: 216–27.
4. Cardoso L, Stevenson M, Thakker RV. Molecular genetics of syndromic and non-syndromic forms of parathyroid carcinoma. *Hum Mutat*, 2017; **38**: 1621–48.
5. Eller-Vainicher C, Chiodini I, Battista C, *et al*. Sporadic and MEN1-related primary hyperparathyroidism: differences in clinical expression and severity. *J Bone Miner Res*, 2009; **24**: 1404–10.
6. Norton JA, Venzon DJ, Berna MJ, *et al*. Prospective study of surgery for primary hyperparathyroidism (HPT) in multiple endocrine neoplasia-type 1 and Zollinger-Ellison syndrome: long-term outcome of a more virulent form of HPT. *Ann Surg*, 2008; **247**: 501–10.
7. Moyes VJ1, Monson JP, Chew SL, Akker SA. Clinical use of cinacalcet in MEN1 hyperparathyroidism. *Int J Endocrinol*, 2010; **2010**: 906163.
8. Yates CJ, Newey PJ, Thakker RV. Challenges and controversies in management of pancreatic neuroendocrine tumours in patients with MEN1. *Lancet Diabetes Endocrinol*, 2015; **3**: 895–905.
9. Newey PJ, Jeyabalan J, Walls GV, *et al*. Asymptomatic children with multiple endocrine neoplasia type 1 mutations may harbour nonfunctioning pancreatic neuroendocrine tumours. *J Clin Endocrinol Metab*, 2009; **94**: 3640–6.
10. Norton JA, Fraker DL, Alexander R, *et al*. Surgery to cure the Zollinger-Ellison syndrome. *N Engl J Med*, 1999; **341**: 635–44.
11. Anlauf M, Garbrecht N, Henopp T, *et al*. Sporadic versus hereditary gastrinomas of the duodenum and pancreas: distinct clinico-pathological and epidemiological features. *World J Gastroenterol*, 2006; **12**: 5440–6.
12. Imamura M, Komoto I, Ota S, *et al*. Biochemically curative surgery for gastrinoma in multiple endocrine neoplasia type 1 patients. *World J Gastroenterol*, 2011; **17**: 1343–53.
13. Bartsch DK, Fendrich V, Langer P, Celik I, Kann PH, Rothmund M. Outcome of duodenopancreatic resections in patients with multiple endocrine neoplasia type 1. *Ann Surg*, 2005; **242**: 757–64.
14. Triponez F, Goudet P, Dosseh D, *et al*. Is surgery beneficial for MEN1 patients with small (< or = 2 cm), nonfunctioning pancreaticoduodenal endocrine tumour? An analysis of 65 patients from the GTE. *World J Surg*, 2006; **30**: 654–62.
15. You YN, Thompson GB, Young WF Jr, *et al*. Pancreatoduodenal surgery in patients with multiple endocrine neoplasia type 1: operative outcomes, long-term function, and quality of life. *Surgery*, 2007; **142**: 829–36.

16. Anlauf M, Schlenger R, Perren A, *et al*. Microadenomatosis of the endocrine pancreas in patients with and without the multiple endocrine neoplasia type 1 syndrome. *Am J Surg Pathol*, 2006; **30**: 560–74.
17. Raymond E, Dahan L, Raoul JL, *et al*. Sunitinib malate for the treatment of pancreatic neuroendocrine tumours. *N Engl J Med*, 2011; **364**: 501–13.
18. Yao JC, Shah MH, Ito T, *et al*. Everolimus for advanced pancreatic neuroendocrine tumours. *N Engl J Med*, 2011; **364**: 514–23.
19. Tomassetti P, Migliori M, Caletti GC, Fusaroli P, Corinaldesi R, Gullo L. Treatment of type II gastric carcinoid tumours with somatostatin analogues. *N Engl J Med*, 2000; **343**: 551–4.
20. Concolino P, Costella A, Capoluongo E. Multiple endocrine neoplasia type 1 (MEN1): an update of 208 new germline variants reported in the last nine years. *Cancer Genet*, 2016; **209**: 36–41.
21. Bassett JHD, Forbes SA, Pannett AAJ, *et al*. Characterisation of mutations in patients with multiple endocrine neoplasia type 1 (MEN1). *Am J Hum Genet*, 1998; **62**: 232–44.
22. Burgess JR, Nord B, David R, *et al*. Phenotype and phenocopy: the relationship between genotype and clinical phenotype in a single large family with multiple endocrine neoplasia type 1 (MEN1). *Clin Endocrinol*, 2000; **53**: 205–11.
23. Turner JJ, Christie PT, Pearce SH, Turnpenny PD, Thakker RV. Diagnostic challenges due to phenocopies: lessons from multiple endocrine neoplasia type1 (MEN1). *Hum Mutat*, 2010; **31**: E1089–101.

6.11.2 Multiple Endocrine Neoplasia Type 2a and 2b

Electron Kebebew, Douglas Wiseman, and Mustapha El Lakis

Introduction

Multiple endocrine neoplasia (MEN) type 2a (also referred to as type 2) and 2b (also referred to as type 3) are autosomal dominant inherited multiglandular cancer syndrome, resulting from mutations in the *RET* (REarranged during Transfection) proto-oncogene. It is the only inherited cancer syndrome that results from a germline-activating mutation in one allele of a proto-oncogene. All of the other inherited cancer syndromes with known susceptibility genes result from germline mutations in tumour suppressor genes in one allele and a second hit in the other allele in the target organ. Together, MEN 2a and 2b have an estimated worldwide prevalence of 1 in 35 000 individuals [1].

MEN 2a (or MEN 2) is characterized by the development of medullary thyroid carcinoma (MTC), phaeochromocytoma (PHEO), and primary hyperparathyroidism (PHPT). MEN 2a (MEN 2) was first described by Sipple in 1961 when he reported a patient with PHEO and thyroid cancer [2]. Seven years later, Steiner and colleagues described a family with MTC, PHEO, PHPT, and Cushing's syndrome and suggested the entity characterized by MEN 2a (MEN 2) [3]. MEN 2b (or MEN 3) is characterized by MTC, PHEO, and distinct physical phenotypes such as mucosal neuromas and Marfanoid habitus. It was first described in 1966 by Williams and Pollak [4]. In 1986, familial medullary thyroid carcinoma (FMTC) was described as a syndrome distinguished by the presence of hereditary MTC but none of the extrathyroidal manifestations of MEN 2a or 2b. FMTC was then classified as a variant of MEN 2a along with three other MEN 2a variants: (1) Classical MEN 2a; (2) MEN 2a and cutaneous lichen amyloidosis (CLA); and (3) MEN 2a and Hirschsprung's disease (HD) (Table 6.11.2.1) [5, 6]. In this chapter, the genetics, diagnosis, and management of the various MEN 2a and 2b syndromes are discussed.

Multiple Endocrine Neoplasia Type 2a

MEN 2a (or MEN 2) accounts for more than 90% of combined MEN 2a and 2b cases [1]. Classical MEN 2a is the most common variant, and 95% of individuals test positive for *RET* germline mutations [7]. In this chapter, we will refer to classical MEN 2a (or MEN 2) as 'MEN 2a'. Virtually all individuals affected with MEN 2a who did not undergo prophylactic thyroidectomy will have biochemical signs of MTC by age 35, while up to 50% will develop PHEOs, and approximately 30% will develop PHPT [1].

MTC, a neuroendocrine tumour that originates from the calcitonin-secreting parafollicular (C-cells) of the thyroid gland, is the defining component of all patients with MEN 2a (MEN 2). It is usually the first to manifest with almost complete penetrance. The incidence of MTC peaks in the third decade of life and usually presents with a neck mass or neck pain in probands or in patients

Table 6.11.2.1 Types of multiple endocrine neoplasia type 2a

Subtype	Associated Diseases				
	MTC	PHPT	PHEO	CLA	HD
1. Classical MEN 2a	+	+	+	–	–
2. MEN 2a and CLA	+	+	+	+	–
3. MEN 2a and HD	+	+	+	–	+
4. Familial MTC	+	–	–	–	–

MEN 2, multiple endocrine neoplasia type 2a; MTC, medullary thyroid carcinoma; PHPT, primary hyperparathyroidism; PHEO, phaeochromocytoma; CLA, cutaneous lichen amyloidosis; HD, Hirschsprung's disease.

unaware of their genetic predisposition. At the time of diagnosis, up to 70% of patients have cervical lymph node metastases if they present with clinically evident MTC.

PHEO, a neuroendocrine tumour originating from the adrenal medulla chromaffin cells, generally presents in MEN 2a (MEN 2) after or concurrently with MTC, but it can be the first presenting tumour in 15% of patients with MEN 2a [8]. PHEO is the second-most-common tumour in MEN 2. The penetrance and age at diagnosis of PHEO in MEN 2a (MEN 2) is dependent on the specific underlying *RET* mutation (often referred to as *RET* genotype-phenotype associations). For instance, in patients with *RET* codon 634 mutations, PHEO penetrance is 25% by age 30, 52% by age 50, and 88% by age 77 [9]. Compared to individuals without MEN 2a-associated PHEOs, patients with MEN 2a are diagnosed at a younger age and present with less-obvious symptoms. MEN 2a–associated PHEOs have a high tendency to develop synchronous or metachronous bilateral adrenal tumours in up to approximately 50% of patients by the age of 50. MEN 2a–associated PHEOs are rarely malignant, with only two reported cases out of 563 patients in a large multinational observational study [8]. In addition, the impact of MEN 2a–associated PHEO on mortality is lower than expected, but this may be due to active surveillance for PHEO in patients with MEN 2a.

PHPT usually presents years after the diagnosis of MTC, with an average age at presentation of 34 years [10]. PHPT penetrance in MEN 2a is dependent on the specific *RET* mutation [10]. The highest predisposition for PHPT is in kindred with cysteine–arginine mutations at codon 634, exon 11 (up to 30%). On the other hand, *RET* mutations in codons 609, 611, 618, and 620 are associated with a penetrance between 2% and 12% [11, 12]. The clinical course of MEN 2a-related PHPT is usually mild and associated with few if any symptoms. MEN 2a-related PHPT can be due to a single enlarged parathyroid gland but is also reportedly due to multiple enlarged glands in up to 82% of patients [13]. Parathyroid carcinoma is rare, and only a single case has been reported in the literature [14].

MEN 2a with CLA is a rare variant of the MEN 2a syndrome and was first described by Gagel and associates [15]. CLA is a rare skin condition usually located in the interscapular region (dermatomes T2–6) or on the extensor surfaces of the extremities and is a pruritic, scaly, papular, and pigmented lesion. The pruritus usually starts during childhood and predates the development of MTC by years. Pruritus improves with sun exposure and worsens during periods of stress. Repeated scratching of the pruritic zone causes epidermal hyperplasia and pigmentation [16].

MEN 2a (or MEN 2) with HD is also a rare variant of MEN 2a. One of the first reports highlighting the co-occurrence of HD in a large kindred with MEN 2a was published in 1982 [17]. HD clinically presents as either failure to pass meconium within 48 hours of birth, intractable constipation, or small bowel obstruction. HD, also known as aganglionic megacolon, occurs due to the congenital absence of parasympathetic neuronal ganglia in a segment of the hindgut. HD can precede C-cell hyperplasia or MTC, which is why children with HD should be tested for germline *RET* mutations.

Multiple Endocrine Neoplasia Type 2b

MEN 2b (or MEN 3) is very rare, with an estimated incidence of two per million live births annually. This is because most cases of MEN 2b (or MEN 3) are *de novo*, as patients typically die of their MTC before being able to have offspring. MEN 2b (MEN 3) is associated with the early onset of MTC and PHEO in approximately 50% of patients as well as phenotypical features such as ophthalmologic abnormalities (thickened and everted eyelids, inability to tear while crying, and prominent corneal nerves), skeletal malformations (Marfanoid body habitus, pes cavus, pectus excavatum, poor dentition, high-arched palate or nodules in the palate, scoliosis, and slipped capital femoral epiphyses), and a generalized ganglioneuromatosis (**Figure 6.11.2.1**). Neuromas can also occur anywhere along the gastrointestinal tract, resulting in loss of normal bowel tone, distension, segmental dilatation, and megacolon. Intestinal involvement usually presents early in childhood with abdominal pain and constipation before other extrathyroidal manifestations are perceived. In MEN 2b (MEN 3), MTC can occur in the first year of life, with the youngest patient reported at 9 weeks old [18].

Genetics of MEN 2a and 2b

MEN 2a and 2b result from germline gain-of-function mutations in the *RET* proto-oncogene located on chromosome 10q11.2, which contains 20 exons spanning a region of more than 55KB [19, 20]. *RET* was first discovered in 1985 by Takahashi and colleagues via transfection of murine NIH 3T3 cells with human lymphoma deoxyribonucleic acid (DNA) [21]. In 1993, mutations in *RET* were linked to MEN 2 and FMTC [22]. The following year, germline sequencing of DNA from a MEN 2b (or MEN 3) cohort revealed a point mutation in *RET* [23]. Later, in 1994, the International RET Mutation Consortium met for the first time in Stockholm, Sweden, where genetic data submitted from 361 MEN 2a and 2b families from Asia, Australia, Europe, and North America showed that 93–100% of MEN 2a families, 94% of MEN 2b families, and 87% of FMTC kindred had *RET* germline mutations, solidifying the relationship between MEN 2a syndromes and *RET* mutations [24]. The *RET* gene codes for a single-pass transmembrane receptor tyrosine kinase in cells of neural crest origin (**Figure 6.11.2.1**). The extracellular portion of the RET protein, which includes the N-terminus, has four cadherin-like domains as well as a cysteine-dense region that forms intermolecular disulphide bonds during homodimerization [25–27]. The transmembrane domain of the RET protein contains 22 amino acids. The C-terminus and the receptor tyrosine kinase

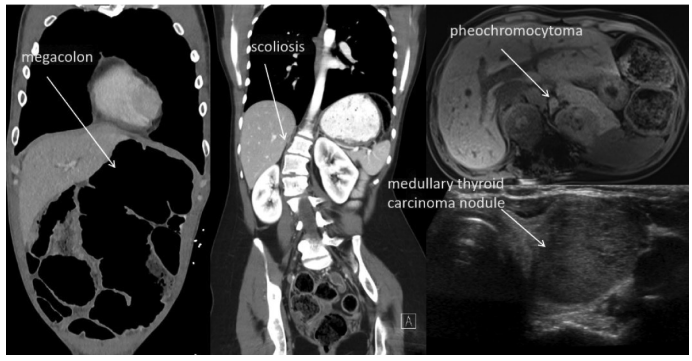


Figure 6.11.2.1 Imaging showing the manifestations of MEN 2b including megacolon, scoliosis, phaeochromocytoma, and medullary thyroid carcinoma.

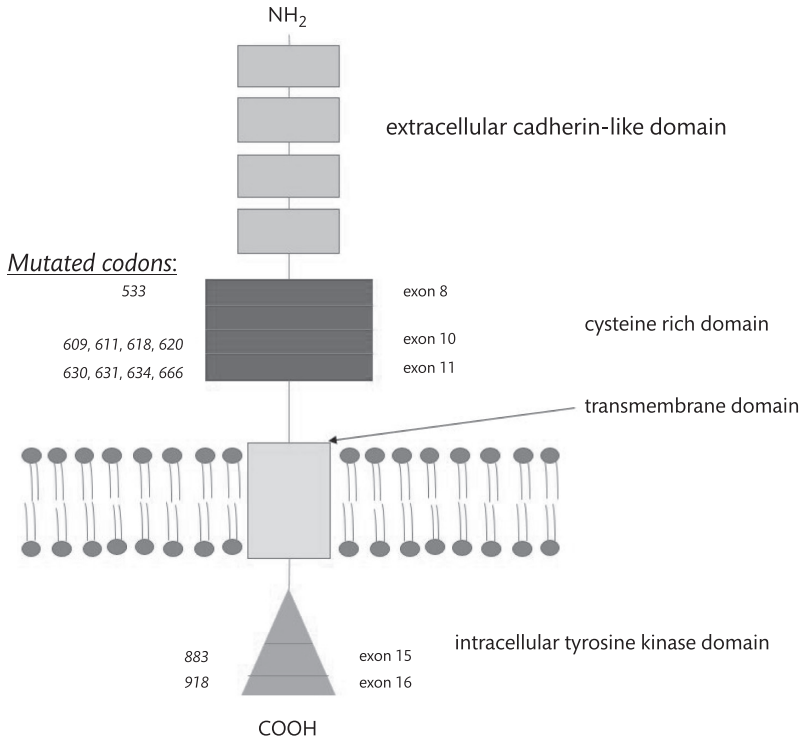


Figure 6.11.2.2 Schematic of the RET protein illustrating protein domains and codon mutations within the corresponding exons.

domain are positioned within the cytoplasm. RET protein plays a role in cell development and migration, and the mechanism of RET activation is via glial cell-derived neurotrophic factor (GDNF) ligands [28]. Pathogenic mutations in RET cause constitutive activation of the RET protein.

Ninety-five percent of MEN 2b (or MEN 3) patients have a missense mutation at codon 918 (M918T), located in the intracellular kinase region of the RET protein, while the other 5% of MEN 2b (MEN 3) mutations occur at codon 883 (A883F) [29]. Codon 918 (C918) mutations are characterized by aggressive MTC with onset in infancy, often with metastases to cervical lymph nodes and distant sites such as the mediastinum, lung, and/or bone. Fifty percent of MEN 3 patients will develop PHEOs. Other physical attributes characteristic of MEN 2b (MEN 3) described earlier are present in patients with either C918 or C883 mutations, and the patients have been described as indistinguishable on physical examination [30]. However, the C883 mutation may be associated with the less-aggressive MTC [31]. The current American Thyroid Association (ATA) guidelines place the MEN 2b C918 mutation in the highest-risk category and the MEN 2b C883 mutation in the high-risk category in regard to the aggressiveness of MTC.

Genotype-Phenotype Association

Eighty-five percent of patients with classical MEN 2a (MEN 2) have a missense mutation of *RET* in codon 634 (C634R), which is in the extracellular domain of the RET protein. Other mutations in the classical variant of MEN 2 involve codons 609, 611, 618, or 620 [9]. An uncommon missense mutation (V292M) in MEN 2a involving the extracellular domain of the RET protein has been documented [32]. All classical MEN 2a patients develop MTC. The 2015 ATA guidelines stratify MTC risk into three categories—moderate, high, and highest—for the aggressiveness and age of onset of MTC based on the specific *RET* mutation (Table 6.11.2.1). The penetrance of PHEO and PHPT in classical MEN 2a is dependent upon the specific *RET* mutation. A higher penetrance of PHEO and PHPT is associated with the C634R mutation, with decreasing penetrance for *RET* mutations in codons 609, 611, 618, and 620 (Table 6.11.2.2) [11, 12].

Genetic Testing

Germline *RET* mutations in MEN 2a (MEN 2) syndromes are driver mutations, and genetic testing and counselling is an important part of disease management. Genetic testing and counselling should be offered to all patients with MEN 2a and 2b. The ATA guidelines for the management of MTC recommend a single or tiered approach to genetic mutational testing in patients with MEN 2a. Single or

Table 6.11.2.2 *RET* mutations and MTC aggressiveness with PHEO and PHPT penetrance

RET mutation type	MTC aggressiveness [6]	PHEO penetrance	PHPT penetrance
M918T	highest	~50%	0%
C634F/G/R/S/W/Y, A883F	high	~50%	0–30%
All others	moderate	0–30%	0–10%

MTC, medullary thyroid carcinoma; PHEO, phaeochromocytoma; PHPT, primary hyperparathyroidism.

multitiered testing should be offered to search for *RET* mutations in exons 10 and 11 as well as exons 8, 13, 14, 15, and 16. Additional *RET* genetic testing should proceed if no mutations are found in the aforementioned coding regions of *RET*. However, with advances in sequencing technology and the decreasing cost of genetic testing, such tiered testing is probably unnecessary.

MEN 2b (MEN 3) patients should be offered genetic testing focusing on C918 and C883. If these two codons are free of mutation, all 20 exons should be analysed for possible mutations. Because MEN 2a and 2b (or MEN 2 and 3) syndromes are inherited in an autosomal dominant fashion, first-degree relatives of patients with known *RET* mutations should undergo *RET* mutational screening. Patients with exon 10 mutations and HD as well as patients with CLA should proceed with *RET* mutation testing. In patients with MEN 2a and exon 10 mutations, a search for HD mutations would be recommended. Finally, all patients with MTC should have genetic testing for germline *RET* mutations, as up to 8% of patients with apparently sporadic MTC (no family history) will be found to have a germline *RET* mutation.

Diagnosis and Management

The diagnosis of MEN 2a and 2b (MEN 2 and 3) syndromes begins with a complete history and physical and is confirmed with high accuracy via genetic testing for *RET* germline mutations. All patients presenting with MTC, PHEO, and C-cell hyperplasia (CCH) should be tested for *RET* germline mutations. In patients with a positive family history of MEN 2a or 2b, genetic testing should be pursued. A negative family history does not guarantee the absence of MEN 2. Up to 40% of patients with *RET* mutations will not develop clinically detectable disease, hence genetic testing should proceed if the patient has features suggestive of MEN 2a or 2b. At institutions or countries where genetic testing is not available, MEN 2a can be diagnosed clinically by the presence of at least two of the classical features (MTC, PHEO, and PHPT).

Medullary Thyroid Carcinoma

MTC is the most prevalent neoplastic feature of all MEN 2a and 2b syndromes and is usually the first manifestation, if not the only manifesting feature. It is also the most frequent disease-specific cause of mortality in MEN 2a and 2b patients. Surgery is the mainstay of disease management and offers the only viable option for remission. The goal of curative surgical therapy for patients with hereditary MTC is complete surgical resection prior to malignant transformation or spread beyond the thyroid gland.

MEN 2a patients with known *RET* mutations are stratified into moderate or high-risk ATA categories for the aggressiveness of MTC based on specific codon mutations (Table 6.11.2.2). Children with MEN 2a in the moderate-risk category often develop MTC at a later age compared to the high-risk category [33]. For moderate-risk category patients, screening, and surveillance for MTC should begin at age 5. Surveillance includes annual physical exam, ultrasound (US) of the thyroid, and measurement of serum basal and/or stimulated calcitonin (CTN) and carcinoembryonic antigen (CEA). Surveillance may last for several years before surgery is indicated, and parents may choose to forego a prolonged period of surveillance and proceed with prophylactic total thyroidectomy before elevation of biomarkers (CTN, CEA) are noted. It is recommended that prophylactic total thyroidectomy be pursued when serum basal CTN is 40 pg/ml or higher (Figure 6.11.2.3). Follow-up after prophylactic total thyroidectomy includes US of the neck and basal or stimulated serum CTN and CEA every 6 months for 1 year. If imaging is negative and biomarkers remain undetectable or normal, basal, or stimulated serum CTN and CEA should be performed annually. Any suspicious findings on surveillance US warrant a fine needle aspiration (FNA) biopsy.

For children with MEN 2a in the MTC ATA high-risk category, surveillance begins at age 3. Thyroidectomy is recommended at or before age 5, and the extent of surgery is guided by CTN levels [34]. Postoperative follow-up for high-risk MTC is similar to the moderate-risk category [6].

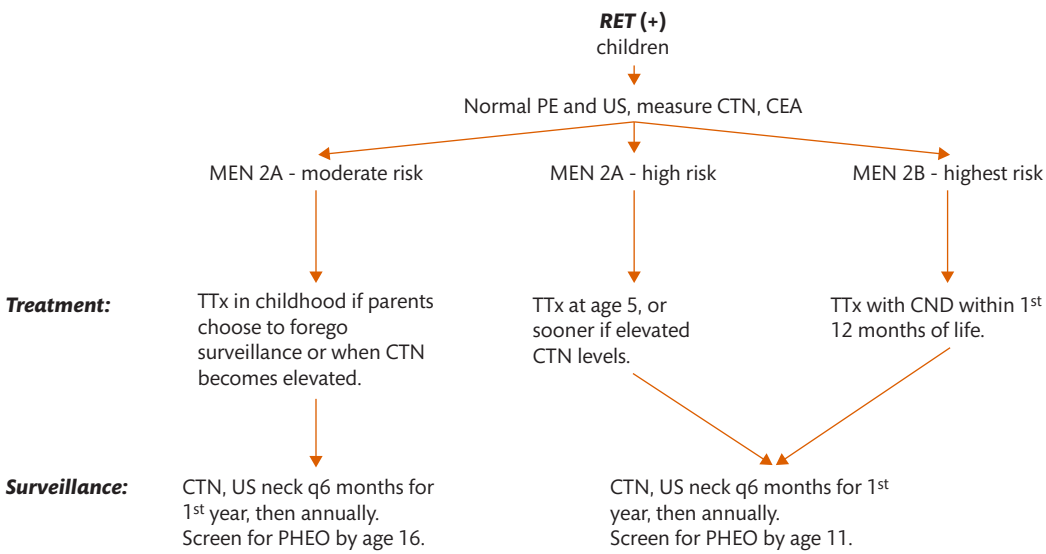


Figure 6.11.2.3 Flowchart for medullary thyroid carcinoma (MTC) management in *RET*-positive children. PE, physical exam; US, ultrasound, CTN, calcitonin; CEA, carcinoembryonic antigen; TTx, total thyroidectomy; CND, central neck dissection; q6, every six; PHEO, phaeochromocytoma.

The ATA highest-risk category for MTC is comprised of only MEN 3 patients with a *RET* codon 918 mutation. These patients can develop an aggressive variant of MTC within the first year of life. Patients can either inherit the 918 *RET* mutation or develop the mutation *de novo* [35]. Patients of a known MEN 2b (MEN 3) kindred should have prophylactic total thyroidectomy early, which may be in the first months of life and at least by 1 year of age [6]. Both high basal CTN levels in paediatric patients persisting to three years of age and higher values in males as compared to females have been noted [36]; therefore, in these young MEN 2b patients, CTN levels may not be useful. The decision to proceed with a central lymph node dissection (level VI) in the highest-risk patients should be determined by nodal status. Alternatively, central neck dissection may be performed prophylactically if the parathyroid glands can be identified and preserved *in situ* or reimplanted to avoid a complication of hypoparathyroidism. Patients with *de novo RET* codon 918 mutations typically present with a thyroid nodule [35, 37]. Age at identification in one cohort of *de novo* 918 mutations was from 1 to 31 years of age [38]. Because these patients typically present with advanced disease, extent of surgery is guided by imaging findings and biomarker levels (CTN, CEA). Foregoing a level VI lymph node dissection is not recommended [6].

Patients with clinically evident disease that is proven to be MTC by FNA cytology or histology should have serum CEA and CTN levels measured. MTC histology demonstrates amyloid staining of the stroma and parafollicular or C-cells. All patients with MTC should have genetic testing for *RET* mutations. The work-up for *RET*-positive patients with MTC includes a thyroid and cervical soft tissue US. Patients presenting with extensive cervical disease and/or serum basal CTN levels greater than 500 pg/ml should have a complete metastatic work-up including computed tomography (CT) of the neck and chest with intravenous contrast, contrast-enhanced magnetic resonance imaging (MRI) of the liver, MRI of the bony spine, and a bone scan to evaluate for distant metastases. Serum CTN levels have been correlated with locoregional and distant metastatic disease. Serum CTN levels above 500 pg/ml also indicate the need for additional imaging beyond US searching for metastases [6]. Functional imaging is not currently recommended [6] to evaluate a patient for MTC metastases due to lower sensitivity when compared to anatomic imaging [39]. However, literature exists to support tandem use of 18-F-DOPA PET/CT and 18-F-FDG PET/CT with improved sensitivities when CTN is above 150 pg/ml [40]. 18-F-FDG PET/CT use in some MTC cohorts has increased sensitivity of metastatic lesion detection for CTN >1000 pg/ml [41].

The extent of surgery for MTC is guided by imaging findings and biomarker levels. Prior to surgical management of MTC in a patient with MEN2/3, PHEO and PHPT should be excluded biochemically. A MEN2/3 patient presenting with MTC limited to the thyroid and CTN <20 pg/ml should undergo a total thyroidectomy (TTx) and level VI lymph node dissection. In the absence of gross locoregional disease, CTN levels may correlate with occult lymph node metastasis and subsequently will have implication for the initial surgical treatment. For instance, in patients with CTN levels of 50–200 pg/ml, the surgery should include TTx with level VI and ipsilateral levels II–V nodal dissection, while CTN above 200 pg/ml would include a TTx and bilateral levels II–V and level VI lymph node dissection for a patient with preoperative disease limited to the thyroid. However, the contralateral lateral neck dissection (levels II–V) may be performed in a staged fashion or not at all if postoperative CTN levels are undetectable.

All patients should have levothyroxine replacement after thyroidectomy with a goal of maintaining thyroid-stimulating hormone (TSH) in the normal range. TSH should be checked 6 to 8 weeks after thyroidectomy and the initiation of replacement levothyroxine therapy.

CTN and CEA are biomarkers used preoperatively and postoperatively to follow disease. Postoperatively, undetectable levels of serum biomarkers should be observed to indicate remission [42]. CTN and CEA should be checked three months after operation [43, 44] and should be checked every 6 months for 1 year and then annually thereafter if they remain undetectable or within the normal range [6]. Three-month postoperative CTN less than 150 pg/ml warrants a US of the neck and physical examination; if negative, the surveillance should be repeated at 6-month intervals [6]. Persistent or recurrent disease is indicated by elevated or rising serum levels, while the normalization of CEA and CTN has a favourable prognosis. It has also been shown that the doubling time of CTN and CEA over a period of less than two years is associated with a worse prognosis [7].

American Joint Committee on Cancer (AJCC) stage at diagnosis, nodal metastases, and age remain the main predictive factors for survival on multivariate analysis [45]. Torresan and colleagues reported on their single institution experience with MTC from 1980 to 2015, including 255 patients with sporadic and hereditary (34.9%) MTC. In their cohort, the cure rate was 56.8% in all patients after operation. Deaths occurred only in patients with stages III–IV, with 6% of MEN 2a and 66.7% of MEN 2b patients succumbing to disease and 17.7% of all patients having persistent or recurrent disease [45]. A French group presented a study of 899 patients with MTC published in 1998, in which 96% of patients were undergoing surgery and 43% of the cohort had familial forms of MTC [46]. Forty-three per cent (43%) of operated patients were biochemically cured. Overall survival (OS) was 85% at 5 years and 78% at 10 years. Multivariate analysis demonstrated that age and stage were independent predictive factors of survival while gender, type of surgery, and type of familial form were only predictive in univariate analysis. Biochemical cure predicted an OS rate of 98% at 10 years. Recurrence, defined as subsequent elevation of CTN after postoperative normalization, was found in 5% of their cohort. In patients with biochemical evidence of disease (57%), OS was 80% and 70% at 5 and 10 years, respectively. Biochemical cure was associated with stage in the French study.

MTC can spread to multiple organs including the liver, brain, bone, skin, lung, and mediastinum. Patients with isolated brain metastases are candidates for surgical resection or external beam radiation. External beam radiation is an option among others such as thermoablation and surgery for bone fractures secondary to metastasis. When lung and liver metastasis are solitary, surgical resection is a viable option. For disseminated liver metastasis, chemoembolization could be considered if there is significant disease progression. Palliation of patients with advanced MTC should be considered when metastatic lesions are causing pain, mechanical compression, or signs and symptoms of hormonal excess. Patients with bone metastasis to the spinal cord with compression symptoms require urgent treatment with glucocorticoid therapy and surgical decompression. Space-occupying metastases that cause bronchial

and oesophageal obstruction can be palliated by surgical resection, external beam radiation, or the administration of systemic therapy.

For patients with metastatic MTC, the rate of tumour progression and quality of life should be taken into consideration against the limited efficacy and potential toxicities of local and systemic therapies. For instance, systemic therapy is not recommended for patients with stable low-volume metastatic disease or patients in whom imaging finds no metastatic disease despite elevation in CTN. In recent years, targeted therapy with tyrosine kinase inhibitors (TKIs) has been evaluated in phase I, II, and III clinical trials in patients with advanced MTC. Currently, there are two Food and Drug Administration (FDA)–approved TKIs for the management of metastatic MTC: vandetanib and cabozantinib. In the two phase III trials, vandetanib [47] and cabozantinib [48] have shown the potential to provide high rates of disease control with durable responses and a highly significant improvement of progression-free survival. Thus, vandetanib and cabozantinib should be considered as single-agent first-line systemic therapies in patients with advanced progressive disease. Patients treated with TKIs for advanced MTC require careful monitoring because they are at increased risk for developing hypothyroidism and other toxicities.

Single-agent or combination cytotoxic chemotherapeutic regimens have shown low response rates (15–20%) of short duration. The most effective regimens are combination therapies with either doxorubicin and epirubicin or 5-fluorouracil and dacarbazine [49].

Phaeochromocytoma

Up to 50% of MEN2/3 patients will develop PHEO. All patients with MEN 2a/2b diagnosed with MTC should have biochemical testing for concurrent PHEO with plasma-free metanephrines or fractionated urine metanephrines before undergoing thyroidectomy and also during surveillance. Patients in the ATA highest- and high-risk *RET* mutations categories should be tested for PHEO beginning at age 11. Those in the moderate-risk group should be tested beginning at age 16 [6]. Once PHEO is diagnosed biochemically, anatomical imaging such as contrast CT or MRI of the adrenals should be obtained for localization. The patient should then be started on alpha blockade for 7 to 14 days prior to adrenalectomy. Adrenalectomy should precede any neck surgery for MTC to avoid intraoperative hemodynamic instability and to minimize associated complications such as cerebral vascular accident (CVA) or myocardial infarction (MI). Laparoscopic or retroperitoneoscopic adrenalectomy is the surgical procedure of choice for PHEO. If possible, patients with unilateral PHEO may be offered a partial or cortical-sparing adrenalectomy (**Figure 6.11.2.4**) [8]. We prefer this approach, as the malignancy risk is minimal, and it aims to retain needed adrenal cortical tissue. This may circumvent the need for exogenous steroids replacement, as the patient is likely to develop a contralateral PHEO within 10 years of the first PHEO and require additional adrenal surgery [6]. Bilateral PHEOs may also be amenable to minimally invasive resection techniques and should be evaluated for possible bilateral partial adrenalectomy. Generally, it is recommended that an adrenal remnant of at least one-third of the normal adrenal volume with the adrenal vein left *in situ* is needed to maintain adequate adrenal function. Postoperatively, an adrenocorticotrophic hormone (ACTH) stimulation test is usually carried out to test for the residual adrenal cortex function, and it may help guide the need for steroid replacement therapy. Patients undergoing partial adrenalectomy should be screened annually for PHEO recurrence with biochemical testing.

Hyperparathyroidism

PHPT is diagnosed biochemically with elevated serum calcium and inappropriately elevated intact parathormone (PTH). In patients with known MEN 2a, the initiation of surveillance for PHPT mirrors the initiation for surveillance of PHEOs. The penetration of PHPT is related to the specific *RET* mutations, with codon 634 mutation demonstrating the highest penetration (30%) [6]. PHPT is usually addressed surgically at the time of thyroidectomy for MTC, if present. Most MEN 2a/2b patients present with multiple enlarged parathyroid glands. The goal is to achieve a euparathyroid state with surgical intervention. Surgical management in a patient with MEN 2 syndrome and PHPT with no prior neck surgery should have a bilateral neck exploration with excision of only enlarged glands. If

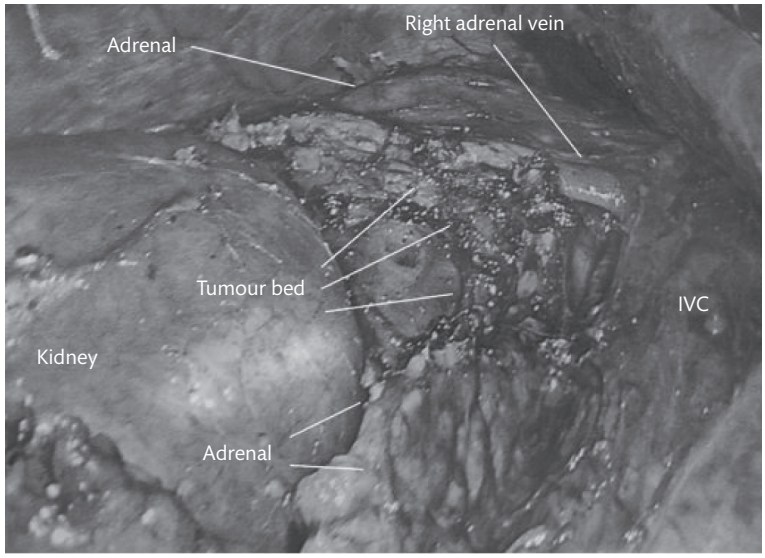


Figure 6.11.2.4 Partial adrenalectomy for a 2 cm phaeochromocytoma in a MEN 2a patient.

PHPT presents after MTC and thyroidectomy, localization studies (US, sestamibi, or CT) and laryngoscopy (to assess vocal cord function) should be performed prior to neck exploration, and all enlarged glands should be excised.

Conclusions

MEN 2a and 2b patients have distinct phenotypes that a clinician should be able to recognize. Genetic testing is an important part of disease management as it will allow for risk stratification of the patient's MTC and guide subsequent therapy. CTN and CEA are accurate biomarkers used to follow MTC for life in MEN2/3 patients. Early diagnosis and intervention are paramount to improving survival in this cohort of patients. All patients with MEN 2a/2b should have biochemical screening and surveillance for PHEO and PHPT. The body of knowledge regarding MEN 2a/2b syndromes continues to grow. As our understanding of the disease at the molecular level increases, we have changed practice in regard to diagnosis, management, and surveillance.

REFERENCES

1. Moline J, Eng C. Multiple endocrine neoplasia type 2: an overview. *Genet Med*, 2011; **13**: 755–64.
2. Sipple J. The association of phaeochromocytoma and thyroid cancer. *Am J Med*, 1961; **31**: 163–6.
3. Steiner AL, Goodman AD, Powers SR. Study of a kindred with phaeochromocytoma, medullary thyroid carcinoma, hyperparathyroidism and Cushing's disease: multiple endocrine neoplasia, type 2. *Medicine (Baltimore)*, 1968; **47**: 371–409.
4. Williams ED, Pollock DJ. Multiple mucosal neuromata with endocrine tumours: a syndrome allied to von Recklinghausen's disease. *J Pathol Bacteriol*, 1966; **91**: 71–80.
5. Farndon JR, Leight GS, Dilley WG, *et al*. Familial medullary-thyroid carcinoma without associated endocrinopathies—a distinct clinical entity. *Br J Surg*, 1986; **73**: 278–81.
6. Wells SA, Jr., Asa SL, Dralle H, *et al*. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*, 2015; **25**: 567–610.
7. Raue F, Frank-Raue K. Genotype-phenotype correlation in multiple endocrine neoplasia type 2. *Clinics*, 2012; **67**: 69–75.
8. Castinetti F, Qi XP, Walz MK, *et al*. Outcomes of adrenal-sparing surgery or total adrenalectomy in phaeochromocytoma associated with multiple endocrine neoplasia type 2: an international retrospective population-based study. *Lancet Oncology*, 2014; **15**: 648–55.
9. Imai T, Uchino S, Okamoto T, *et al*. High penetrance of phaeochromocytoma in multiple endocrine neoplasia 2 caused by germ line RET codon 634 mutation in Japanese patients. *Eur J Endocrinol*, 2013; **168**: 683–7.
10. Schuffenecker I, Virally-Monod M, Brohet R, *et al*. Risk and penetrance of primary hyperparathyroidism in multiple endocrine neoplasia type 2A families with mutations at codon 634 of the RET proto-oncogene. *J Clin Endocrinol Metab*, 1998; **83**: 487–91.
11. Frank-Raue K, Rybicki LA, Erlic Z, *et al*. Risk profiles and penetrance estimations in multiple endocrine neoplasia type 2A caused by germline RET mutations located in exon 10. *Human Mutation*, 2011; **32**: 51–8.
12. Herfarth KKF, Bartsch D, Doherty GM, Wells SA, Lairmore TC. Surgical management of hyperparathyroidism in patients with multiple endocrine neoplasia type 2A. *Surgery*, 1996; **120**: 966–73.
13. Gertner ME, Kebebew E. Multiple endocrine neoplasia type 2. *Curr Treat Options Oncol*, 2004; **5**: 315–25.
14. Jenkins PJ, Satta MA, Simmgen M, *et al*. Metastatic parathyroid carcinoma in the MEN2A syndrome. *Clin Endocrinol (Oxf)*, 1997; **47**: 747–51.
15. Gagel RF, Levy ML, Donovan DT, Alford BR, Wheeler T, Tschen JA. Multiple endocrine neoplasia type-2a associated with cutaneous lichen amyloidosis. *Ann Int Med*, 1989; **111**: 802–6.
16. Scapineli JO, Ceolin L, Punales MK, Dora JM, Maia AL. MEN 2A-related cutaneous lichen amyloidosis: report of three kindred and systematic literature review of clinical, biochemical and molecular characteristics. *Fam Cancer*, 2016; **15**: 625–33.
17. Verdy M, Weber AM, Roy CC, Morin CL, Cadotte M, Brochu P. Hirschsprung's disease in a family with multiple endocrine neoplasia type 2. *J Pediatr Gastroenterol Nutr*, 1982; **1**: 603–7.
18. Shankar RK, Rutter MJ, Chernausek SD, Samuels PJ, Mo JQ, Rutter MM. Medullary thyroid cancer in a 9-week-old infant with familial MEN 2B: implications for timing of prophylactic thyroidectomy. *Int J Pediatr Endocrinol*, 2012; **2012**: 25.
19. Ceccherini I, Bocciardi R, Luo Y, *et al*. Exon structure and flanking intronic sequences of the human ret protooncogene. *Biochem Biophys Res Commun*, 1993; **196**: 1288–95.
20. Pasini B, Hofstra RM, Yin L, *et al*. The physical map of the human RET proto-oncogene. *Oncogene*, 1995; **11**: 1737–43.
21. Takahashi M, Ritz J, Cooper GM. Activation of a novel human transforming gene, ret, by DNA rearrangement. *Cell*, 1985; **42**: 581–8.
22. Donis-Keller H, Dou S, Chi D, *et al*. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Hum Mol Genet*, 1993; **2**: 851–6.
23. Carlson KM, Dou S, Chi D, *et al*. Single missense mutation in the tyrosine kinase catalytic domain of the RET protooncogene is associated with multiple endocrine neoplasia type 2B. *Proc Natl Acad Sci U S A*, 1994; **91**: 1579–83.
24. Mulligan LM, Marsh DJ, Robinson BG, *et al*. Genotype-phenotype correlation in multiple endocrine neoplasia type 2: report of the International RET Mutation Consortium. *J Intern Med*, 1995; **238**: 343–6.
25. Anders J, Kjar S, Ibanez CF. Molecular modeling of the extracellular domain of the RET receptor tyrosine kinase reveals multiple cadherin-like domains and a calcium-binding site. *J Biol Chem*, 2001; **276**: 35808–17.
26. Airaksinen MS, Saarma M. The GDNF family: signalling, biological functions and therapeutic value. *Nat Rev Neurosci*, 2002; **3**: 383–94.
27. Kjaer S, Kurokawa K, Perrinjaquet M, Abrescia C, Ibanez CF. Self-association of the transmembrane domain of RET underlies oncogenic activation by MEN2A mutations. *Oncogene*, 2006; **25**: 7086–95.
28. Roskoski R, Jr., Sadeghi-Nejad A. Role of RET protein-tyrosine kinase inhibitors in the treatment RET-driven thyroid and lung cancers. *Pharmacol Res*, 2018; **128**: 1–17.
29. Oberg K. The genetics of neuroendocrine tumors. *Semin Oncol*, 2013; **40**: 37–44.
30. Smith DP, Houghton C, Ponder BA. Germline mutation of RET codon 883 in two cases of de novo MEN 2B. *Oncogene*, 1997; **15**: 1213–17.

31. Jasim S, Ying AK, Waguespack SG, *et al*. Multiple endocrine neoplasia type 2B with a RET proto-oncogene A883F mutation displays a more indolent form of medullary thyroid carcinoma compared with a RET M918T mutation. *Thyroid*, 2011; **21**: 189–92.
32. Castellone MD, Verrienti A, Magendra Rao D, *et al*. A novel de novo germ-line V292M mutation in the extracellular region of RET in a patient with phaeochromocytoma and medullary thyroid carcinoma: functional characterization. *Clin Endocrinol (Oxf)*, 2010; **73**: 529–34.
33. Szinnai G, Meier C, Komminoth P, Zumsteg UW. Review of multiple endocrine neoplasia type 2A in children: therapeutic results of early thyroidectomy and prognostic value of codon analysis. *Pediatrics*, 2003; **111**: E132–9.
34. Machens A, Gimm O, Hinze R, Hoppner W, Boehm BO, Dralle H. Genotype-phenotype correlations in hereditary medullary thyroid carcinoma: oncological features and biochemical properties. *J Clin Endocrinol Metab*, 2001; **86**: 1104–9.
35. Wray CJ, Rich TA, Waguespack SG, Lee JE, Perrier ND, Evans DB. Failure to recognize multiple endocrine neoplasia 2B: more common than we think? *Ann Surg Oncol*, 2008; **15**: 293–301.
36. Basuyau JP, Mallet E, Leroy M, Brunelle P. Reference intervals for serum calcitonin in men, women, and children. *Clin Chem*, 2004; **50**: 1828–30.
37. Brauckhoff M, Machens A, Hess S, *et al*. Premonitory symptoms preceding metastatic medullary thyroid cancer in MEN 2B: an exploratory analysis. *Surgery*, 2008; **144**: 1044–50; discussion 1050–43.
38. Waguespack SG, Rich TA. Multiple endocrine neoplasia [corrected] syndrome type 2B in early childhood: long-term benefit of prophylactic thyroidectomy. *Cancer*, 2010; **116**: 2284.
39. Giraudet AL, Vanel D, Leboulleux S, *et al*. Imaging medullary thyroid carcinoma with persistent elevated calcitonin levels. *J Clin Endocrinol Metab*, 2007; **92**: 4185–90.
40. Verbeek HH, Plukker JT, Koopmans KP, *et al*. Clinical relevance of 18F-FDG PET and 18F-DOPA PET in recurrent medullary thyroid carcinoma. *J Nucl Med*, 2012; **53**: 1863–71.
41. Ong SC, Schoder H, Patel SG, *et al*. Diagnostic accuracy of 18F-FDG PET in restaging patients with medullary thyroid carcinoma and elevated calcitonin levels. *J Nucl Med*, 2007; **48**: 501–7.
42. Engelbach M, Gorges R, Forst T, *et al*. Improved diagnostic methods in the follow-up of medullary thyroid carcinoma by highly specific calcitonin measurements. *J Clin Endocrinol Metab*, 2000; **85**: 1890–4.
43. Elisei R, Pinchera A. Advances in the follow-up of differentiated or medullary thyroid cancer. *Nat Rev Endocrinol*, 2012; **8**: 466–75.
44. Pellegriti G, Leboulleux S, Baudin E, *et al*. Long-term outcome of medullary thyroid carcinoma in patients with normal postoperative medical imaging. *Br J Cancer*, 2003; **88**: 1537–42.
45. Torresan F, Cavedon E, Mian C, Iacobone M. Long-term outcome after surgery for medullary thyroid carcinoma: a single-center experience. *World J Surg*, 2018; **42**: 367–75.
46. Modigliani E, Cohen R, Campos JM, *et al*. Prognostic factors for survival and for biochemical cure in medullary thyroid carcinoma: results in 899 patients. The GETC Study Group. Groupe d'etude des tumeurs a calcitonine. *Clin Endocrinol (Oxf)*, 1998; **48**: 265–73.
47. Wells SA, Jr., Robinson BG, Gagel RF, *et al*. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol*, 2012; **30**: 134–41.
48. Elisei R, Schlumberger MJ, Muller SP, *et al*. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol*, 2013; **31**: 3639–46.
49. Bajetta E, Rimassa L, Carnaghi C, *et al*. 5-Fluorouracil, dacarbazine, and epirubicin in the treatment of patients with neuroendocrine tumors. *Cancer*, 1998; **83**: 372–8.

6.12

Familial Syndromes and Genetic Causes of Paraganglioma and Phaeochromocytoma

Eamonn R. Maher and Ruth T. Casey

Introduction

A phaeochromocytoma is a tumour arising from the adrenal medulla and a paraganglioma refers to its extra-adrenal counterpart, which can develop from sympathetic or parasympathetic tissue anywhere from the skull base to the pelvis. Phaeochromocytomas and paragangliomas (PPGL) are considered to be the most heritable tumours as up to 40% of patients who develop these tumours have a hereditary predisposition. This chapter provides an update on the clinical and molecular genetics of PPGL and related syndromes and offers a guideline for genetic testing and surveillance of those individuals identified as carriers for a known PPGL predisposition gene.

The Discovery Timeline of Phaeochromocytoma and Paraganglioma (PPGL) Predisposition Genes

Prior to the turn of this century, it was widely accepted that a phaeochromocytoma followed a rule of 'tens': (i) 10% were bilateral; (ii) 10% were malignant; (iii) 10% were extra-adrenal (now termed paraganglioma (PGL) arising from sympathetic ganglia anywhere from the thorax to pelvis); and finally and (iv) 10% were hereditary.

Familial occurrence of phaeochromocytoma (PC) was first recognized by Chase *et al.* in 1937 [1] and almost 50 years later the co-occurrence of PC and its extra-adrenal counterpart PGL in the same kindred was reported [2]. Those tumours that were familial were associated with hereditary syndromes such as neurofibromatosis type 1, caused by germline mutations in the neurofibromin 1 gene (*NF1*), multiple endocrine neoplasia type 2 (MEN2) caused by germline mutations in the *RET* proto-oncogene and Von Hippel–Lindau syndrome (VHL), caused by germline mutations in the *VHL* tumour suppressor gene. Each of these syndromes predispose to the development of bilateral phaeochromocytoma but extra-adrenal tumours are rarely encountered and typically each of these syndromes are associated with other characteristic phenotypic features.

Between the years 2000 and 2010, nine further phaeochromocytoma predisposition genes were identified and this era of discovery completely dispelled the contention that only 10% were hereditary. This era of discovery was triggered by the seminal finding that some families affected by a head and neck paragangliomas (HNPGL) such as a chemodectoma or glomus jugulare tumour, carried a mutation in the succinate dehydrogenase subunit D gene (*SDHD*) [3]. Between 2000 and 2011, all four of the succinate dehydrogenase subunit genes (*SDHA, SDHB, SDHC* and *SDHD*) were implicated in the development of familial PPGL[4, 5], as well as the *SDHAF2* gene which encodes its namesake protein responsible for the flavination of the SDHA protein [6].

In 2010, an integrated genomics approach led to the finding that a gene encoding a transmembrane protein (TMEM127) localized to chromosome 2, was implicated in familial PPGL [7]. This transmembrane protein normally functions as a modulator of the mTOR pathway and notably patients with mutations in the *TMEM127* gene were found to present at an older age compared to patients with other syndromic causes of PPGL [7]. Subsequently, whole-exome sequencing identified recurrent mutations in the *MAX* (MYC associated factor X) gene in three unrelated patients with hereditary PPGL [8]. Though all of these genes are inherited in an autosomal

dominant manner, *SDHD, SDHAF2*, and *MAX* are associated with predisposition to PPGL, almost always only after paternal transmission of a mutation [8].

Finally, mutations in another citric acid cycle enzyme encoding gene fumarate hydratase (*FH*) were implicated in hereditary phaeochromocytoma [9], in addition to its previously accepted role in hereditary leiomyomatosis and renal cell cancer.

Importantly, PPGLs are now considered to be the most heritable of all tumours with 40% of patients having a genetic predisposition in one of 11 major susceptibility genes routinely screened for in clinical practice (*SDHA, SDHB, SDHC, SDHD, SDHAF2, NF1, RET, TMEM127, MAX, FH, VHL*) [10]. Although several additional predisposition genes have been reported in the past five years (e.g. malate dehydrogenase 2 (*MDH2*) gene [11], *SLC25A11* gene, which encodes the mitochondrial 2-oxoglutarate/malate carrier and *EPAS1* [12], which encodes the hypoxia inducible factor-2 alpha (HIF-2alpha)), these are a rare cause of PPGL and not commonly analysed in first-line genetic testing.

Neurofibromatosis Type 1 (OMIM 162200)

Neurofibromatosis type 1 was first described by Von Recklinghausen in 1882 and is an autosomal dominant syndrome characterized by neuro-ectodermal abnormality and variable systemic manifestations affecting bone, nervous system, eyes, and other sites. The *NF1* gene maps to chromosome 17 and is 57 exons in length [13]. Generally, a diagnosis of neurofibromatosis type 1 is made on clinical criteria (including café-au-lait macules, neurofibromas, axillary or inguinal freckling, optic gliomas, Lish nodules in the iris, or a first-degree relative with a diagnosis of NF1[14]) and genetic testing is not routinely performed owing to the large size of the gene. Many mutations occur *de novo* with no preceding family history [13, 15]. However, recent reductions in the cost of next generation sequencing, have facilitated genetic testing in patients with suspected NF1 [15]. PC is a rare feature, affecting 0.1–6% of all patients with NF1 [16]. The median age of onset is approximately 41 years and most patients with NF1 will not have PC as the presenting feature. Extra-adrenal and malignant tumours are rarely associated with NF1 but up to a third of patients in one large study developed bilateral disease [16].

Multiple Endocrine Neoplasia 2 (MEN2) Syndrome (OMIM 171400)

In the early 1970s, Sizemore and his colleagues first described the syndrome of MEN2 and subclassified the syndrome into MEN2A and MEN2B. The MEN2A syndrome is an autosomal dominantly inherited disorder, characterized by a predisposition to medullary thyroid carcinoma, phaeochromocytoma, and hyperparathyroidism. MEN2B predisposes to the same tumour types but is also associated with mucosal neuromas and mesodermal abnormalities [17]. In 1993, activating mutations in the *RET* proto-oncogene on chromosome 11 were identified as the molecular drivers of MEN2 syndrome [18].

Increasing clinical data has led to the realization that very specific genotype–phenotype correlations exist in MEN2 syndrome. Overall, inherited mutations in the *RET* gene predict a 30–50% lifetime risk of developing a PC [19] and mutations in codon 634 and 918 of the *RET* gene are most commonly associated with the development of PC. The median age of onset is ~35 years and patients with a codon 634 mutation in *RET* have a 30% lifetime risk of developing bilateral tumours [20, 21]. Typically, patients with *RET* gene mutations will present with medullary thyroid carcinoma as the presenting disease but PC can be the presenting feature in ~5% of patients [21]. Similar to NF1, extra-adrenal and malignant tumours are very rarely associated with MEN2 [21] and the secretory pattern of *RET* mutated PC is predominately adrenergic [22].

Von Hippel–Lindau Disease (OMIM 193300)

VHL is a dominantly inherited familial cancer syndrome with multisystem involvement caused by germline inactivating mutations in the *VHL* tumour suppressor gene [23]. The most frequent features are retinal and central nervous system haemangioblastomas, renal cell carcinoma (RCC), and renal, pancreatic, and epididymal cysts [23]. The most important endocrine complications are phaeochromocytoma and non-functioning pancreatic islet cell neuroendocrine tumours. There are marked interfamilial differences in PC frequency in VHL disease. Thus in some families PC is the most common manifestation, but in others it is rare. These differences reflect genotype–phenotype correlations and the high risk of phaeochromocytoma associated with certain *VHL* missense mutations [24]. Large deletions, protein truncating mutations, and missense mutations that disrupt protein stability are associated with a high risk of retinal angioma, cerebral hemangioblastoma, and RCC, but a low risk of PC (type 1 VHL phenotype). On the other hand, missense mutations affecting amino acids on the VHL protein (pVHL) surface, predominate in VHL patients with PC (type 2 VHL) [24].

The clinical presentation of PC in VHL disease is similar to that in sporadic cases except that there is a higher frequency of bilateral or multiple tumours and, on average, an earlier onset (mean approximately 30 years) in VHL disease. As with sporadic tumours, PPGL in VHL disease may be sometimes be extra-adrenal and, in about 5% of cases, malignant. Interestingly, a particular secretory PC phenotype is associated with *VHL* gene mutations, namely a predominant noradrenergic secretory pattern related to down regulation of the phenylethanolamine N-methyltransferase (PNMT) enzyme in *VHL* mutated PC [22]. The PNMT enzyme normally acts to convert noradrenaline to adrenaline and therefore a predominant noradrenergic secretory phenotype in a PC can point towards an underlying *VHL* mutation [22].

Early detection of tumours in VHL disease can improve management and prognosis and affected individuals and at-risk relatives enter into an annual surveillance programme (see **Box 6.12.1**). All VHL patients and at-risk individuals are offered screening for a PC irrespective of whether there is a family history of PC but those with a positive family history of PC or a PC-associated missense mutation are given special attention. Patients with apparently non-syndromic familial or bilateral PC, or PC at a young age may have a germline *VHL* gene mutation [21] and should be offered *VHL* mutation analysis. In such cases the nature of the *VHL* mutation identified will indicate the risk of other types of VHL-related tumours. See **Figure 6.12.1**.

Box 6.12.1 Example of surveillance protocol for Von Hippel–Lindau disease in asymptomatic affected patients and at-risk relatives

1 Affected adult patient:
- Annual physical examination, and direct and indirect ophthalmoscopy
- MRI brain scan every 3 years to age 50 years and every 5 years thereafter
- Annual abdominal MRI (or ultrasonography if MRI is unavailable) for renal, adrenal, and pancreatic masses
- Annual 24-h urine collection for urinary fractionated metanephrines by mass spectrometric methods or plasma free metanephrines by mass spectrometric methods [45]

2 At-risk relative:
- Annual physical examination, and direct and indirect ophthalmoscopy from age 5 years until age 60 years
- MRI brain scan every 3 years to from age 15 to 40 years and then every 5 years until age 60 years
- Annual renal MRI or ultrasonography from age 16 years to age 65 years
- Annual 24-h urine collection for urinary fractionated metanephrines by mass spectrometric methods or plasma free metanephrines by mass spectrometric methods [45]

Succinate Dehydrogenase Subunit (*SDHA*, *SDHB* *SDHC*, and *SDHD*) Genes

The heterotetrameric succinate dehydrogenase enzyme complex (SDH) functions as a key enzyme coupling the oxidation of succinate to fumarate in the citric acid cycle and the reduction of ubiquinone to ubiquinol in the electron transport chain. Therefore, SDH facilitates the cellular metabolism of lipids, glucose, and amino acids, and feeds into the mitochondrial respiratory chain to generate cellular energy. The SDH complex comprises two hydrophilic subunits (SDHA and SDHB), and two hydrophobic subunits (SDHC and SDHD). The SDHA subunit is a flavoprotein, which covalently binds a flavine adenine dinucleotide (FAD) cofactor and contains the binding site for the succinate metabolite, coupling the oxidation of succinate to fumarate and reduction of FAD to FADH2. The SDHB subunit contains three iron clusters which are employed in the transfer of electrons from FADH2 to ubiquinone. Two LYR motifs are essential binding sites to enable the incorporation of the three Fe-S clusters within the final structure of complex II. The SDHC and SDHD subunits anchor the complex to the inner mitochondrial membrane and facilitate the binding of ubiquinone. Each subunit of the SDH complex is encoded by its namesake gene, the *SDHA* gene (OMIM 600857) maps to the short arm of chromosome 5 (5p15), *SDHB* (OMIM 185470) to 1p36.13, *SDHC* (OMIM 602413) to 1q23.3 and *SDHD* to 11q23.

Inactivation of the SDH complex results in accumulation of the metabolite succinate as oxidative dehydrogenation of succinate to fumarate in the citric acid cycle is no longer feasible [25]. Succinate accumulation is postulated to be a key driver in tumourigenesis and succinate is now referred to as an 'oncometabolite'[25]. Succinate accumulation inhibits 2-oxoglutarate-dependent dioxygenases, including DNA and histone demethylase enzymes and hypoxic gene response regulators. As a consequence, SDH-deficient tumours demonstrate epigenetic abnormalities, namely hypermethylation and an activated hypoxic gene response [26, 27]. See Figure 6.12.2.

To date >445 germline mutations have been described in SDH subunit genes and the majority of these have been described in association with PPGL [28]. *SDH* mutations are heterozygous mutations inherited in an autosomal dominant fashion and tumourigenesis is initiated by bi-allelic loss as per the Knudson 'two hit' hypothesis. The most common type of mutation affecting the *SDH* genes are missense mutations [28].

Mutations in one of the four *SDHx* genes account for 30–40% of hereditary PPGL cases and the majority of germline *SDHx* mutations reported to date have been identified in the *SDHB* gene (52%), followed by mutations in *SDHD* (35%) [28]. Mutations in the *SDHx* genes predispose to multiple, synchronous PGL, and bilateral PC. Similar to the secretory phenotype associated with *VHL* mutated PC, *SDHx* gene mutations also down regulate the intratumoural PNMT enzyme, resulting in a predominant noradrenergic secretory phenotype [22].

The strongest existing genotype–phenotype correlation for *SDHx*-mutated PPGL, is the risk of malignancy associated with *SDHB* mutated PPGL. Germline *SDHB* gene mutations have been detected in up to 50% of patients with a malignant PPGL and associated 5-year survival is less than 50% [29]. Extra-adrenal PGLs in the abdomen and pelvis as well as adrenal PC are more commonly associated with *SDHB* mutations than other *SDHx* gene subunits but tumours can arise in the neck and thorax [29–31].

Mutations in the *SDHC* and *SDHD* genes are most commonly associated with head and neck paraganglioma (HNPGL, derived from parasympathetic ganglia) than PC or PGL. *SDHD* is the most common of the *SDHx* subunit genes to be implicated in HNPGL and tumours are frequently multifocal with a lower malignant potential compared to *SDHB* [32]. Until 2010, *SDHA* gene mutations were only associated with autosomal recessive metabolic encephalopathy

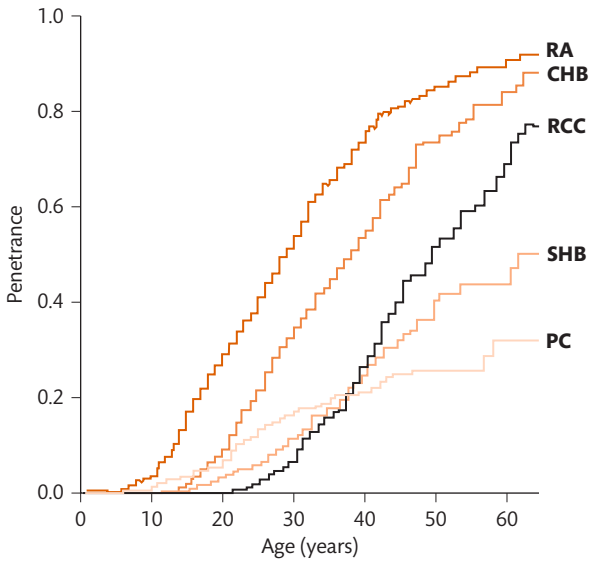


Figure 6.12.1 Age-related risks for the five major manifestations of von Hippel–Lindau disease. RA, retinal angioma; CHB, cerebellar haemangioblastoma; SHB, spinal haemangioblastoma; RCC, renal cell carcinoma; PC, phaeochromocytoma.

Reproduced with permission from Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype–phenotype correlations in von Hippel–Lindau disease. *Hum Mutat*, 2007; 28: 143–9. Copyright 2006 © Wiley Periodicals, Inc.

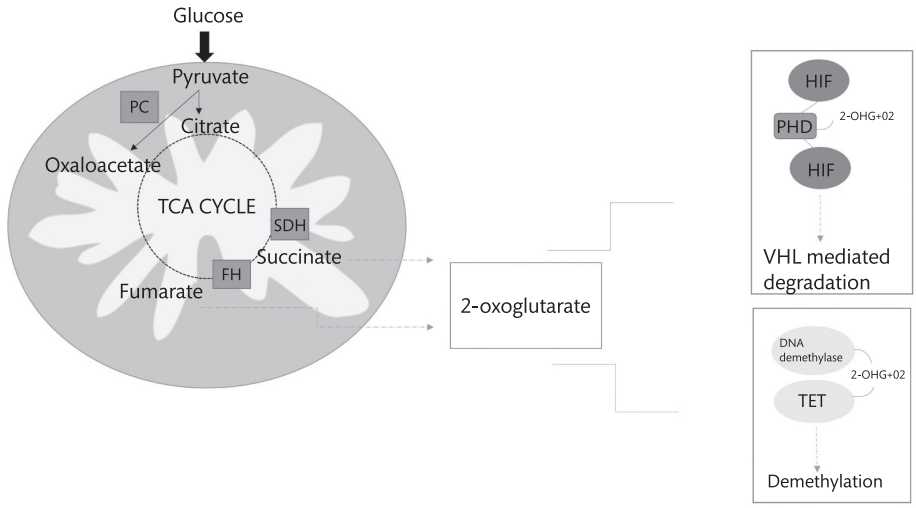


Figure 6.12.2 Pathways implicated in SDH-deficient tumorigenesis precipitated by accumulation of the 'oncometabolite' succinate.
Adapted with permission from Morin A, Letouzé E, Gimenez-Roqueplo A-P, Favier J. Oncometabolites-driven tumorigenesis: From genetics to targeted therapy. *Int J Cancer* [Internet]. 2014 Nov 15;135(10):2237–48. Available from: http://doi.wiley.com/10.1002/ijc.29080. Copyright 2014 © UICC.

syndrome (Leigh syndrome) but *SDHA* mutations are now also recognized as rare cause of hereditary PPGL [33, 34].

Mutations in the *SDHx* complex genes are also associated with the development of other tumour types including gastrointestinal stromal tumours (GIST), RCC, and pituitary tumours. GIST is the second most common tumour phenotype after PPGL [28]. Most GIST tumours in adults harbour somatic mutations in the *KIT* or *PDGFRA* genes and respond to the tyrosine kinase inhibitor Imatinib. Wild-type GIST (wtGIST), refer to tumours that are negative for *KIT* and *PDGFRA* gene mutations and account for 15% of adult and 85% of paediatric GIST tumours [35]. Almost 90% of wtGIST are associated a germline mutation in the *SDHx* complex genes (75%) or a somatic epigenetic inactivation of the *SDHC* gene [35]. Interestingly *SDHA* mutations are more frequent than other SDH subunit genes [28, 34]. Non-SDH associated wtGIST are associated with germline mutations in *NF1* or less commonly somatic *BRAF* mutations. Differences in clinical phenotype, tumour location and histology can help guide genetic testing further in wtGIST (Figure 6.12.3).

Germline *SDHx* gene mutations have been described in the context of renal and PPGL tumour syndrome (RAPTAS) [36, 37]. SDH-deficient RCC account for approximately 0.2% of all cases of RCC and to date 83% of the germline mutations identified in SDH-deficient RCCs occurred in *SDHB* [28, 37]. SDHB-deficient renal carcinoma was accepted as a provisional entity in the 2013 International Society of Urological Pathology (ISUP) Vancouver Classification and a unique morphology consisting of solid architecture, intracytoplasmic inclusions and intratumour mast cells is characteristic of SDHB-deficient RCC [38]. Pituitary tumours have also occasionally been associated with *SDHx* gene mutations and the estimated incidence is 0.3% [28].

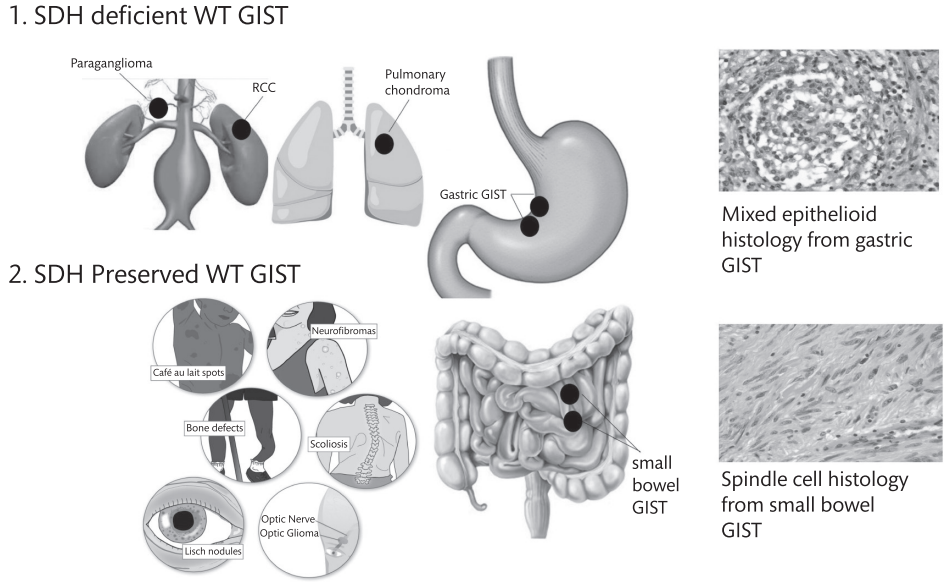


Figure 6.12.3 Differences in clinical presentation and histology between SDH-deficient and SDH preserved (e.g. *NF1*-related) wild-type GIST.

The *SDHAF2* gene (OMIM 613019) responsible for the flavination of the SDHA protein has very rarely been implicated in hereditary PPGL since its discovery in 2009 and a recent study of over 900 patients, has reported the prevalence of *SDHAF2* mutations in patients presenting with PPGL as 0.1% [39].

TMEM127, *MAX*, and *FH* Gene Mutations

The *TMEM127* gene (OMIM 613403) is localized to chromosome 2 and functions as a tumour suppressor gene. Although the precise molecular function of *TMEM127* remains poorly understood, this gene product participates in the mTor signalling pathway [7]. The largest studies to date investigating the role of *TMEM127* in hereditary PPGL, have estimated the incidence of mutations in this gene at approximately 2% [39, 40]. Mutations in *TMEM127* predispose to the development of PC, with a later age of onset (41.2 years) compared to other gene mutations such as *VHL* or *SDHB*. Extra-adrenal, malignant and bilateral tumours are rare with *TMEM127* [7, 39, 40], as are cases of RCC or co-occurring cases of RCC and PPGL [36, 41].

Biallelic loss of the *MAX* tumour suppressor gene was first implicated in hereditary PPGL in 2011 [8]. The *MAX* gene (OMIM 154950) localizes to chromosome 14 and the MAX protein contains a basic helix loop helix zipper that is commonly involved in a complex formation and sequences in the promoter region of hundreds of genes encoding for proteins essential in cellular metabolism and angiogenesis [42]. The estimated incidence of *MAX* mutations in PPGL is less than 2% and typically associated with adrenal PC and rarely with extra-adrenal or malignant PPGL [8, 39] or RCC [36, 42].

The most recent PPGL predisposition gene to be discovered and routinely screened for in clinical practice, is *FH* (OMIM 136850), which is localized to chromosome 1 and encodes its namesake enzyme in the citric acid cycle. Mutations in *FH* cause accumulation of the 'oncometabolite' fumarate which, similar to succinate accumulation in SDH-deficient tumours, drives tumourigenesis through inhibition of 2-oxoglutarate-dependent dioxygenases and promoting a pseudohypoxic and hypermethylation phenotype [27, 43]. *FH* mutations are most commonly associated with renal cell carcinoma (type 2 papillary RCC) and cutaneous and uterine leiomyomas (hereditary leiomyomatosis and renal cell cancer syndrome, HLRCC) and are rarely associated with PPGL [9, 44].

The Role of Next Generation Sequencing in PPGL

The Endocrine Society guidelines on the management of PPGL recommended that genetic screening should be considered in all patients with PPGL and that *SDHx* mutations in particular should be screened for in patients with extra-adrenal PGL [45]. Next generation sequencing (NGS) has facilitated a cost effective and time efficient transition from previous methods of targeted, sequential, analysis of individual susceptibility genes in hereditary PPGL. It has been estimated that a second-generation sequencing test for a 'panel' of nine susceptibility genes predisposing to PPGL (*MAX, RET, SDHA, SDHB, SDHC, SDHD, SDHAF2, TMEM127*, and *VHL*), can achieve a 70% cost reduction compared to standard targeted sequencing methods of individual candidate genes [10]. More recent NGS panels also include *FH*. It is advised that genetic testing is offered within the framework of best clinical practice and that patients are offered genetic counselling before and after genetic testing [45]. Furthermore, it is very important that each identified variant is correctly interpreted before presymptomatic genetic counselling of relatives is performed or a personalized management approach for the patient is adopted [45].

SDHB Immunohistochemistry in Clinical Practice

Biallelic inactivation of any of the *SDH* subunit genes, will usually result in destabilization of the SDH enzyme complex, which can be detected by loss of staining for the SDHB protein on immunohistochemistry (IH) [46]. Negative immunostaining for SDHB is now a validated marker for germline mutations in *SDHA*/*SDHB*/*SDHC*/*SDHD* and provides a reasonable assessment of pathogenicity of an identified novel *SDHx* variant [46]. SDHB IH has been adapted into clinical practice and serves a number of important clinical applications including: (i) the early diagnosis of SDH deficiency in specific phenotypes such as PPGL, wild-type GIST and specific histological subtypes of RCC; (ii) additional prognostic information regarding malignant risk associated with PPGL [47]; (iii) a cost effective functional tool for the assessment of *SDHx* variant pathogenicity; and (iv) SDHB IH can also identify those tumours that are SDH-deficient due to a somatic mutation or epigenetic inactivation.

Surveillance Protocol for PPGL Predisposition Syndromes

The cumulative evidence to date would suggest that surveillance of asymptomatic gene carriers can lead to early tumour detection and therefore surveillance protocols should be designed to maximize the benefit of early tumour detection and minimize risks associated with ionizing radiation exposure and anxiety generated by false positive test results. At present, given the low incidence of PPGL associated with *NF1* and the diverse clinical phenotype, the current recommendation is for annual clinical review in dedicated specialist clinics and testing only if patients are clinically symptomatic [16].

The most robust evidence and experience for surveillance protocols exist for VHL and MEN2A. See **Box 6.12.1**.

Surveillance for *RET* mutation carriers differs depending on the exact genotype due to the well-defined genotype–phenotype correlations [19]. It is recommended that those patients identified with a highest risk mutation (such as mutations in codon 918), undergo a prophylactic thyroidectomy in the first year of life and that screening for PC is commenced at the age of 11 years [48] using the same biochemical tests as described for *VHL* mutation carriers in **Box 6.12.1**. For those carriers with a high risk *RET* mutation (e.g. codon 634) a review and prophylactic thyroidectomy is recommended before the age of five years and that screening for PC is commenced at the age of 11 years. For moderate or low-risk *RET* mutations, a review and surgery are advised from the age of 5 years and screening for PC is advised to commence from the age of 16 years [48]. Annual clinical review and biochemical measurements of serum calcitonin

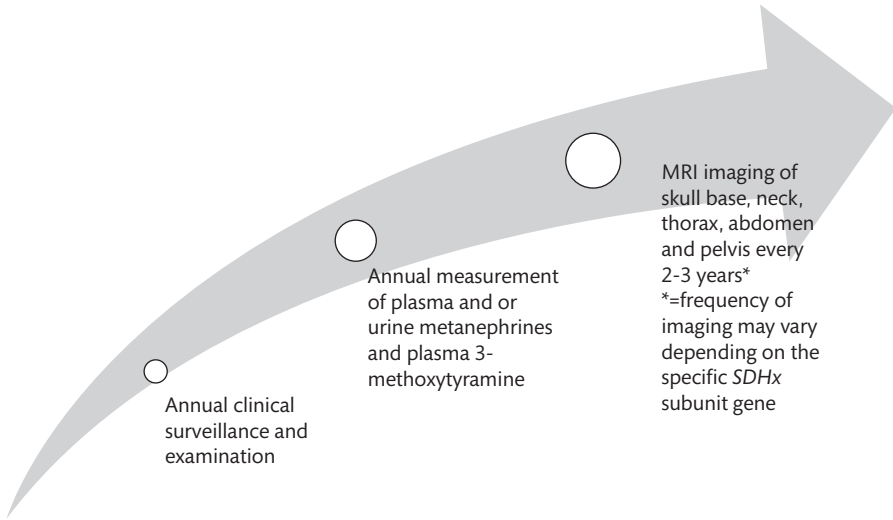


Figure 6.12.4 Demonstrates an acceptable clinical, biochemical, and radiological surveillance strategy for *SDH* gene carriers.

following thyroidectomy and calcium testing from the onset of PC screening, is also recommended [48].

Unlike VHL and MEN2 disease, there is relatively less experience of the utility of surveillance in *SDHx* gene carriers or carriers of the rare predisposition genes (*TMEM127, MAX, FH, SDHAF2*).

At present surveillance for *SDHx* gene carriers is advised to begin at the age of 5 years for *SDHB* carriers and 10 years for carriers of a pathogenic variant in *SDHA*/*SDHC*/*SDHD* [49, 50] and to continue lifelong, if acceptable to the patient [51]. Although regular surveillance for *SDHx* mutation carriers is a widely accepted clinical practice, there is no consensus on the frequency of surveillance or on the optimal radiological modality. The current recommendation is that annual clinical review is performed in addition to measurement of plasma metanephrines or urine metanephrines (Figure 6.13.4) [44]. However the frequency of cross-sectional imaging and the choice of radiological modality differs between practices and although the Endocrine Society guidelines recommend cross-sectional imaging for *SDHx* carriers, to diagnosis non-secretory tumours, they do not specify appropriate time intervals between imaging studies or a preferred imaging modality [45]. In recent years, as updated information regarding disease penetrance of the *SDHx* mutations has become available, concern regarding unnecessary radiation exposure in this group has mounted and more centres are recommending MRI alone as the imaging modality of choice [50]. This approach is supported by a recent study which identified that MRI was as sensitive as 18F-fluorodeoxyglucose (FDG) PET CT in the diagnosis of occult tumours in *SDHx* carriers [51]. Finally the interval frequency at which radiological screening is carried out remains contentious but it is widely accepted that the first radiological screen is the most important and yields the highest diagnoses of tumours in this cohort [50, 51]. At present, studies would support baseline MRI imaging of skull base, neck, thorax, abdomen, and pelvis at first clinic visit for all asymptomatic *SDHx* carriers, followed by two to three yearly interval MRI scanning. An individual approach to radiological surveillance based on the individual subunit gene is now recommended (e.g. *SDHB* carriers require more frequent imaging compared to *SDHA*/*SDHC*/*SDHD* carriers [52]; see Figure 6.12.4). There is a need for additional long-term surveillance data to better inform this challenging area of practice.

Finally, data is lacking to guide surveillance protocols for the less common predisposition genes (*TMEM127, MAX, FH, SDHAF2*) and at present annual clinical review, annual biochemical testing with urinary or plasma metanephrines and judicious, individual case-based use of cross-sectional imaging is advised.

REFERENCES

1. Chase WH. Familial and bilateral tumours of the carotid body. *J Pathol Bacteriol*, 1933; **36**: 1–12.
2. Pritchett JW. Familial concurrence of carotid body tumor and pheochromocytoma. *Cancer*, 1982; **49**: 2578–9.
3. Baysal BE, Ferrell RE, Willett-Brozick JE, *et al.* Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science*, 2000; **287**: 848–51.
4. Burnichon N, Brière J-J, Libé R, *et al.* SDHA is a tumor suppressor gene causing paraganglioma. *Hum Mol Genet*, 2010; **19**: 3011–20.
5. Niemann S, Müller U. Mutations in SDHC cause autosomal dominant paraganglioma, type 3. *Nat Genet*, 2000; **26**: 268–70.
6. Hao H-X, Khalimonchuk O, Schraders M, *et al.* SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science*, 2009; **325**: 1139–42.
7. Qin Y, Yao L, King EE, *et al.* Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nat Genet*, 2010; **42**: 229–33.
8. Comino-Méndez I, Gracia-Aznárez FJ, Schiavi F, *et al.* Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet*, 2011; **43**: 663–7.
9. Clark GR, Sciacovelli M, Gaude E, *et al.* Germline FH mutations presenting with pheochromocytoma. *J Clin Endocrinol Metab*, 2014; **99**: E2046–50.
10. Rattenberry E, Vialard L, Yeung A, *et al.* A comprehensive next generation sequencing–based genetic testing strategy to improve diagnosis of inherited pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*, 2013; **98**: E1248–56.

11. Cascón A, Comino-Méndez I, Currás-Freixes M, *et al*. Whole-exome sequencing identifies mdh2 as a new familial paraganglioma gene. *JNCI J Natl Cancer Inst*, 2015; **107**: pii: djv053.
12. Zhuang Z, Yang C, Lorenzo F, *et al*. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. *N Engl J Med*, 2012; **367**: 922–30.
13. Neurofibromatosis. Conference statement. National Institutes of Health Consensus Development Conference. *Arch Neurol*, 1988; **45**: 575–8.
14. National Institutes of Health Consensus Development Conference Statement: neurofibromatosis. Bethesda, MD, July 13–15, 1987. *Neurofibromatosis*, 1988; **1**: 172–8.
15. Tadini G, Milani D, Menni F, Pezzani L, Sabatini C, Esposito S. Is it time to change the neurofibromatosis 1 diagnostic criteria? *Eur J Intern Med*, 2014; **25**: 506–10.
16. Bausch B, Borozdin W, Mautner VF, *et al*. Germline NF1 mutational spectra and loss-of-heterozygosity analyses in patients with pheochromocytoma and neurofibromatosis type 1. *J Clin Endocrinol Metab*, 2007; **92**: 2784–92.
17. Carney JA. Familial multiple endocrine neoplasia: the first 100 years. *Am J Surg Pathol*, 2005; **29**: 254–74.
18. Mulligan LM, Kwok JB, Healey CS, *et al*. Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature*, 1993; **363**: 458–60.
19. Mulligan LM, Eng C, Healey CS, *et al*. Specific mutations of the RET proto-oncogene are related to disease phenotype in MEN 2A and FMTC. *Nat Genet*, 1994; **6**: 70–4.
20. Siqueira DR, Ceolin L, Ferreira CV, *et al*. Role of RET genetic variants in MEN2-associated pheochromocytoma. *Eur J Endocrinol*, 2014; **170**: 821–8.
21. Neumann HPH, Bausch B, McWhinney SR, *et al*. Germ-line mutations in nonsyndromic pheochromocytoma. *N Engl J Med*, 2002; **346**: 1459–66.
22. Eisenhofer G, Klink B, Richter S, Lenders JW, Robledo M. Metabologenomics of phaeochromocytoma and paraganglioma: an integrated approach for personalised biochemical and genetic testing. *Clin Biochem Rev*, 2017; **38**: 69–100.
23. Maher ER, Kaelin WG. von Hippel–Lindau disease. *Medicine (Baltimore)*, 1997; **76**: 381–91.
24. Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype-phenotype correlations in von Hippel–Lindau disease. *Hum Mutat*, 2007; **28**: 143–9.
25. Morin A, Letouzé E, Gimenez-Roqueplo A-P, Favier J. Oncometabolites-driven tumorigenesis: from genetics to targeted therapy. *Int J Cancer*, 2014; **135**: 2237–48.
26. López-Jiménez E, Gómez-López G, Leandro-García LJ, *et al*. Research resource: transcriptional profiling reveals different pseudohypoxic signatures in SDHB and VHL-related pheochromocytomas. *Mol Endocrinol*, 2010; **24**: 2382–91.
27. Letouzé E, Martinelli C, Loriot C, *et al*. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 2013; **23**: 739–52.
28. Evenepoel L, Papathomas TG, Krol N, *et al*. Toward an improved definition of the genetic and tumor spectrum associated with SDH germ-line mutations. *Genet Med*, 2015; **17**: 610–20.
29. Gimenez-Roqueplo A-P, Favier J, Rustin P, *et al*. Mutations in the SDHB gene are associated with extra-adrenal and/or malignant phaeochromocytomas. *Cancer Res*, 2003; **63**: 5615–21.
30. Benn DE, Gimenez-Roqueplo A-P, Reilly JR, *et al*. Clinical presentation and penetrance of pheochromocytoma/paraganglioma syndromes. *J Clin Endocrinol Metab*, 2006; **91**: 827–36.
31. Timmers HJLM, Kozupa A, Eisenhofer G, *et al*. Clinical presentations, biochemical phenotypes, and genotype-phenotype correlations in patients with succinate dehydrogenase subunit B-associated pheochromocytomas and paragangliomas. *J Clin Endocrinol Metab*, 2007; **92**: 779–86.
32. Ricketts CJ, Forman JR, Rattenberry E, *et al*. Tumor risks and genotype-phenotype-proteotype analysis in 358 patients with germline mutations in SDHB and SDHD. *Hum Mutat*, 2010; **31**: 41–51.
33. Casey RT, Challis BG, Marker A, *et al*. A case of a metastatic SDHA mutated paraganglioma re-presenting twenty-three years after initial surgery. *Endocr Relat Cancer*, 2017; **24**: L69–71.
34. Casey RT, Ascher DB, Rattenberry E, *et al*. SDHA related tumorigenesis: a new case series and literature review for variant interpretation and pathogenicity. *Mol Genet Genomic Med*, 2017; **5**: 237–50.
35. Boikos SA, Pappo AS, Killian JK, *et al*. Molecular subtypes of KIT/PDGFRA wild-type gastrointestinal stromal tumors. *JAMA Oncol*, 2016; **2**: 922.
36. Casey RT, Warren AY, Martin JE, *et al*. Clinical and molecular features of renal and pheochromocytoma/paraganglioma tumor association syndrome (RAPTAS): case series and literature review. *J Clin Endocrinol Metab*, 2017; **102**: 4013–22.
37. Ricketts C, Woodward ER, Killick P, *et al*. Germline SDHB mutations and familial renal cell carcinoma. *JNCI J Natl Cancer Inst*, 2008; **100**: 1260–2.
38. Srigley JR, Delahunt B, Eble JN, *et al*. The International Society of Urological Pathology (ISUP) Vancouver classification of renal neoplasia. *Am J Surg Pathol*, 2013; **37**: 1469–89.
39. Bausch B, Schiavi F, Ni Y, *et al*. Clinical characterization of the pheochromocytoma and paraganglioma susceptibility genes SDHA, TMEM127, MAX, and SDHAF2 for gene-informed prevention. *JAMA Oncol*, 2017; **3**: 1204.
40. Yao L, Schiavi F, Cascón A, *et al*. Spectrum and prevalence of FP/TMEM127 gene mutations in pheochromocytomas and paragangliomas. *JAMA*, 2010; **304**: 2611.
41. Qin Y, Deng Y, Ricketts CJ, *et al*. The tumor susceptibility gene TMEM127 is mutated in renal cell carcinomas and modulates endolysosomal function. *Hum Mol Genet*, 2014; **23**: 2428–39.
42. Korpershoek E, Koffy D, Eussen BH, *et al*. Complex MAX rearrangement in a family with malignant pheochromocytoma, renal oncocytoma, and erythrocytosis. *J Clin Endocrinol Metab*, 2016; **101**: 453–60.
43. Isaacs JS, Jung YJ, Mole DR, *et al*. HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell*, 2005; **8**: 143–53.
44. Castro-Vega LJ, Buffet A, De Cubas AA, *et al*. Germline mutations in FH confer predisposition to malignant pheochromocytomas and paragangliomas. *Hum Mol Genet*, 2014; **23**: 2440–6.
45. Lenders JWM, Duh Q-Y, Eisenhofer G, *et al*. Pheochromocytoma and paraganglioma: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 1915–42.
46. Papathomas TG, Oudijk L, Persu A, *et al*. SDHB/SDHA immunohistochemistry in pheochromocytomas and paragangliomas: a multicenter interobserver variation analysis using virtual microscopy: a multinational study of the European

Network for the Study of Adrenal Tumors (ENS@T). *Mod Pathol*, 2015; **28**: 807–21.
47. Kimura N, Takayanagi R, Takizawa N, *et al*. Pathological grading for predicting metastasis in phaeochromocytoma and paraganglioma. *Endocr Relat Cancer*, 2014; **21**: 405–14.
48. Wells SA, Asa SL, Dralle H, *et al*. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid*, 2015; **25**: 567–610.
49. Wong MY, Andrews KA, Challis BG, *et al*. Clinical practice guidance: surveillance for phaeochromocytoma and paraganglioma in paediatric succinate dehydrogenase gene mutation carriers. *Clin Endocrinol (Oxf)*, 2019; **90**: 499–505.
50. Tufton N, Shapiro L, Srirangalingam U, *et al*. Outcomes of annual surveillance imaging in an adult and paediatric cohort of succinate dehydrogenase B mutation carriers. *Clin Endocrinol (Oxf)*, 2017; **86**: 286–96.
51. Lepoutre-Lussey C, Caramella C, Bidault F, *et al*. Screening in asymptomatic SDHx mutation carriers: added value of 18F-FDG PET/CT at initial diagnosis and 1-year follow-up. *Eur J Nucl Med Mol Imaging*, 2015; **42**: 868–76.
52. Tufton N, Sahdev A, Drake WM, Akker SA. Can subunit-specific phenotypes guide surveillance imaging decisions in asymptomatic SDH mutation carriers? *Clin Endocrinol (Oxf)*, 2019; **90**: 31–46.

6.13

Carney's Complex

Constantine A. Stratakis and Fabio R. Faucz

Introduction

Carney's complex (CNC) is an autosomal dominant disorder, which was described in 1985 as 'the complex of myxomas, spotty pigmentation, and endocrine overactivity' in 40 patients [1]. Since then, several hundred index cases have been reported, including the first molecularly confirmed CNC patient seen by Dr Harvey Cushing in 1914, resulting in better definition of the disease and the establishment of diagnostic criteria [2–4]. As implied from the initial description, CNC is not only a multiple neoplasia syndrome, but it is also associated with pigmented lesions of the skin and mucosae, as well as other non-endocrine tumours [5]. Before the characterization of CNC as a syndrome, several patients described under the acronyms NAME (nevi, atrial myxomas, and ephelides) and LAMB (lentigines, atrial myxomas, and blue nevi) can now be probably considered CNC patients [6, 7]. Thus, lentigines, blue nevi, café-au-lait spots, and cutaneous tumours, such as myxomas, fibromas, and others, are major features of the disease [5, 8–11]. However, we cannot say that CNC is completely understood. Interestingly, in 2017, two cases of corticotropinoma in CNC patients have been published, adding a new uncommon feature [12, 13].

The clinical characteristics of CNC have been reviewed and are presented in **Box 6.13.1** [3, 10]. A definite diagnosis of CNC is given if two or more major manifestations are present [5, 10, 14, 15]. Other related manifestations may accompany or suggest the presence of CNC but are not considered diagnostic of the disease (**Box 6.13.1**). Cutaneous manifestations constitute three of the major disease manifestations: (1) spotty skin pigmentation with a typical distribution (lips, conjunctiva, and inner or outer canthi, genital mucosa); (2) cutaneous or mucosal myxoma; and (3) blue nevi (multiple) or epithelioid blue nevus. Suggestive or associated with CNC findings but not diagnostic are: (1) intense freckling (without darkly pigmented spots or typical distribution); (2) multiple blue nevi of common type; (3) café-au-lait spots or other 'birthmarks'; and (4) multiple skin tags or other skin lesions, including lipomas and angiofibromas.

The relationship between the cutaneous and non-cutaneous manifestations of CNC appears to be an essential clue to the molecular aetiology of the disease. According to the latest reports, more than half of CNC patients present with both characteristic dermatological and endocrine signs; however, a significant number of patients present with skin lesions that are only 'suggestive' and not characteristic of CNC [10]. A recent classification based on both dermatological and endocrine markers has subgrouped CNC patients as: multisymptomatic (with extensive endocrine and skin signs); intermediate (with few dermatological and endocrine manifestations); and, paucisymptomatic (with isolated primary pigmented nodular adrenocortical disease (PPNAD) alone and no cutaneous signs) [10]. A recommended clinical surveillance of patients with CNC is presented in **Box 6.13.2.**

Skin Manifestations in CNC

Skin lesions are consistently reported in the majority of the CNC patients (above 80%), the most common being lentigines (in 70–75% of cases). Other pigmented lesions, most frequently blue nevi and café-au-lait spots, with or without lentigines, are seen in approximately 50% of CNC patients. The effort to systemize the knowledge on the cutaneous lesions in CNC patients is driven by their high diagnostic value—presented early in life and easily recognizable, the skin manifestations are an early sign that directs dermatologists' attention towards underlying endocrine or other pathology. In an attempt to outline the most specific and sensitive skin abnormalities in CNC, several research groups have published exhaustive analyses that add to an improved diagnostic and preventive approach [10, 11, 16]. The major challenge appears to be in distinguishing the disease-associated prominent lesions from the more common non-CNC-specific, age- or sun-related skin alterations.

Lentigo is a hamartomatous melanocytic lesion, clinically similar but histologically different from freckles [17]. Morphologically, lentigines are flat, poorly circumscribed, brown-to-black macules,

Box 6.13.1 Diagnostic criteria for Carney's complex

Major diagnostic criteria for Carney's complex

1. Spotty skin pigmentation with typical distribution (lips, conjunctiva, and inner or outer canthi, vaginal and penile mucosal)
2. Myxoma[a] (cutaneous and mucosal)
3. Cardiac myxoma[a]
4. Breast myxomatosis[a] or fat-suppressed magnetic resonance imaging findings suggestive of this diagnosis
5. Primary pigmented nodular adrenocortical disease[a] or paradoxical positive response of urinary glucocorticosteroid excretion to dexamethasone administration during Liddle's test[b]
6. Acromegaly due to growth hormone-producing adenoma[a]
7. Large-cell calcifying Sertoli cell tumour[a] or characteristic calcification on testicular ultrasound
8. Thyroid carcinoma[a] or multiple, hypoechoic nodules on thyroid ultrasound in a young patient
9. Psammomatous melanotic schwannomas[a]
10. Blue naevus, epithelioid blue naevus[a]
11. Breast ductal adenoma[a]
12. Osteochondromyxoma[a]

Supplementary criteria

1. Affected first-degree relative
2. Inactivating mutation of the *PRKAR1A* gene

Findings suggestive of or possibly associated with Carney's complex, but not diagnostic for the disease

1. Intense freckling (without darkly pigmented spots or typical distribution)
2. Blue naevus, common type (if multiple)
3. Café-au-lait spots or other birthmarks
4. Elevated insulin-like growth factor-1 levels, abnormal glucose tolerance test, or paradoxical growth hormone response to thyrotropin-releasing hormone testing in the absence of clinical acromegaly
5. Cardiomyopathy
6. Corticotropinoma
7. Pilonidal sinus
8. History of Cushing's syndrome, acromegaly, or sudden death in extended family
9. Multiple skin tags or other skin lesions; lipomas
10. Colonic polyps (usually in association with acromegaly)
11. Hyperprolactinaemia (usually mild and almost always combined with clinical or subclinical acromegaly)
12. Single, benign thyroid nodule in a young patient; multiple thyroid nodules in an older patient (detected on ultrasonography)
13. Family history of carcinoma, in particular of the thyroid, colon, pancreas, and ovary; other multiple benign or malignant tumours

[a] After histological confirmation.
[b] It has been shown that patients with primary pigmented nodular adrenocortical disease exhibit a paradoxical increase in cortisol secretion in response to Liddle's test (administration of dexamethasone at doses of 2 mg/d for 2 days followed by 8 mg/d for 2 days); this abnormal cortisol response is now used as a criterion for the diagnosis of the disease.

Box 6.13.2 Recommended clinical surveillance of patients with CNC

1. Post-pubertal paediatric and adult patients:
   - Echocardiogram (annually or biannually for adolescent patients with a history of excised myxoma)
   - Testicular ultrasound (annually)
   - Thyroid ultrasound (baseline examination; it may be repeated, as needed)
   - Transabdominal ultrasound of the ovaries (baseline examination; it may be repeated, as needed)
   - Urinary free cortisol levels (annually)
   - Serum IGF-1 levels (annually)
2. Prepubertal paediatric patients:
   - Echocardiogram (annually; *biannually* for patients with a history of excised myxoma)
   - Testicular ultrasound for boys; close monitoring of growth rate and pubertal staging (annually)
3. Further evaluation of patients of all age groups, as needed:

   For **primary pigmented nodular adrenocortical disease**, in addition to urinary free cortisol levels:
   - Diurnal cortisol levels (11.30 p.m., 12.00MN and 7.30 a.m., 8.00 a.m. sampling)
   - Dexamethasone-stimulation test (modified Liddle's test, as per Stratakis *et al.* [10])
   - Adrenal computed tomography

   For **gigantism/acromegaly**, in addition to serum IGF-1 levels:
   - Pituitary magnetic resonance imaging
   - 3-hour oral glucose tolerance test (oGTT)
   - 90-minute thyrotropin-releasing hormone (TRH) testing

   For psammomatous melanotic schwannoma:
   - Magnetic resonance imaging (brain, spine, chest, abdomen, retroperitoneum, pelvis)

usually less than 0.5 cm in diameter, but these may differ in different ethnic groups. In African Americans, for example, lentigines may be slightly raised, dark papules, similar to nevi [17]. In contrast to the common freckles, on histological examination lentigines show basal cell layer hyperpigmentation associated with an increased number of melanocytes (hyperplasia), the majority of which appear hypertrophic. This distinguishes them from freckles (ephelides), which present with a regular number of melanocytes and are pigmented as a result of melanin disposition in the surrounding keratinocytes.

Lentiginosis is one of the manifestations of CNC that can occur early; lentigines usually acquire their typical intensity and distribution during the peripubertal period [10, 11, 18]. They typically involve the centrofacial area, including the vermilion border of the lips, and the conjunctiva, especially the lacrimal caruncle and the conjunctival semilunar fold; intraoral pigmented spots have also been reported [19]. In contrast to age-related skin lesions, CNC-associated lentigines tend to fade after the fourth decade of life, but may be detectable as late as the eighth decade [10, 18].

The next very common skin manifestation in CNC is a lesion known as blue nevus, which is infrequent in the general population. Blue nevi can be seen as small (usually <5 mm), blue to black-coloured marks with a circular or star-shaped appearance. Their distribution is variable; most often they occur on the face, trunk, and limbs, and less frequently on the hands or feet.

An interesting subtype of blue nevus, which is exceedingly rare as a sporadic lesion in the general population but is sometimes seen in patients with CNC, is the epithelioid blue nevus [19, 20]. Epithelioid blue nevus usually presents with intensive pigmentation and poorly circumscribed proliferative regions containing two cell types: heavily pigmented globular and fusiform cells; and lightly pigmented, polygonal spindle melanocytes with a single prominent nucleolus [19]. In contrast to blue nevi, epithelioid blue nevi

display no dermal fibrosis [19]. After comprehensive comparative analysis and based on the fact the epithelioid blue nevi have also been reported in patients with none of the other features of CNC, epithelioid blue nevi are not considered pathognomonic for CNC but simply associated with the disease [10, 19].

Blue nevi and lentigines in CNC are often accompanied by café-au-lait spots, which are otherwise rarely present as an isolated skin manifestation of CNC. Like lentigines, café-au-lait spots can be present at birth. In general, café-au-lait spots in CNC are less intensely pigmented than those seen in McCune–Albright syndrome and they are more similar to those seen in the neurofibromatosis syndromes.

The third most common skin manifestation of CNC—cutaneous myxoma—is reported in between 30% and 55% of the studied patients [5, 10, 11]. Cutaneous myxomas rarely exceed 1 cm in diameter and often affect the eyelids, ears, and nipples, but may also be seen on other areas of the face, ears, trunk, and perineum. They usually appear as asymptomatic, sessile, small, opalescent, or dark pink papules and large, finger-like, pedunculated lesions. They are typically diagnosed early in life, most often during the teenage years (mean age, 18 years). In the majority of patients (>70%) cutaneous myxomas show multiple appearance and a tendency to recur. The frequency of myxoma may be underestimated because clinical diagnosis may be uncertain at times; hence, histological examination is strongly recommended when in doubt. Histopathologically, myxomas are characterized by a location in the dermis or, occasionally, more superficially in the subcutaneous tissues, sharp circumscription (sometimes encapsulation), relative hypocellularity with abundant myxoid stroma, prominent capillaries, lobulation (larger lesions), and occasional presence of an epithelial component. It is estimated that approximately 80% of CNC patients with life-threatening cardiac myxoma present with cutaneous myxoma earlier in life; therefore, cutaneous myxoma can serve as good marker for the disease with high prognostic significance [5, 10, 11].

Other CNC-related skin abnormalities include melanocytic and atypical nevi, and the so-called Spitz nevus. Occasionally, depigmented lesions can be present at birth or, more often, develop in early childhood. These manifestations, although usually not considered specific, may be suggestive for the disease or may accompany other CNC signs of importance for the diagnosis.

Molecular Genetics

Most cases of CNC are caused by inactivating mutations in the gene encoding one of the subunits of the protein kinase A (PKA) tetrameric enzyme, namely regulatory subunit type 1 α (*PRKAR1A*), located at 17q22–24 [5]. Although a second locus (2p16) has been implicated, sequencing of the region in the linked families did not reveal alterations in known coding sequences [21].

PRKAR1A extends to a total genomic length of approximately 21 kb and consists of 11 exons, encoding a total of 381 amino acids, with a dimerization/ docking domain, and two cAMP binding domains, A and B. Since the identification of *PRKAR1A* mutations in CNC, more than 140 disease-causing pathogenic sequence changes have been reported (http://prkar1a.nichd.nih.gov/hmdb/intro.html); they are spread over the entire coding sequence of the gene, without a notable preference for a region or exon. Structurally, the majority of the mutations consists of base substitutions, small deletions, and insertions or combined rearrangements, involving up to 15 bp [5, 22]; although rare, large *PRKAR1A* deletions have been reported [23, 24].

Mutations in *PRKAR1A* are seen in more than 70% of the patients with classical CNC and, in most cases, they lead to complete inactivation of one of the *PRKAR1A* alleles as a result of premature stop codon generation and subsequent nonsense-mediated mRNA decay (NMD) [5, 11, 22]. In its inactive form, PKA is a tetramer composed of two regulatory and two catalytic subunits [25]. The decreased cellular concentration of regulatory subunits results in a balance shift between the formation and the disassembly of the PKA tetramer, towards the release of the catalytic subunits. The free catalytic subunits, which are active serine–threonine kinases, further phosphorylate a series of targets that regulate downstream effectors and transcription of specific genes, mediating cell growth and differentiation [26]. Thus, functionally, the mechanism by which *PRKAR1A* haploinsufficiency causes CNC is through excess cellular cAMP signalling in affected tissues [27]. CNC lesions frequently show loss-of-heterozygosity, suggesting a tumour suppressor function for PRKAR1A [4, 5].

Although significantly less frequent, mutations that escape NMD and lead to the expression of an abnormal, defective PRKAR1A protein have been reported [24, 28–30]. These expressed mutations may lead to a characteristic phenotype that reflects the location and the type of the genetic change. Examples include a germline in-frame deletion of exon 3 which results in severe expression of the majority of the CNC manifestations—a phenotype illustrating the importance of exon 3 in linking the dimerization/docking and the first cAMP binding domain [24]. In contrast, another in-frame variant—a splice-site deletion that eliminates exon 7—is seen associated mostly with lentiginosis and the adrenal component of CNC, PPNAD. Just as lentiginosis is the most common non-endocrine CNC manifestation, PPNAD is the most frequently observed endocrine tumour of the disease. Thus, the presence of only two features of CNC, the most common ones, with this splice-site variant is consistent with the anticipation of a milder phenotype associated with certain splice mutations, due to their incomplete penetrance at the mRNA level (i.e. not all DNA molecules harbouring the splice variant result in mRNA species lacking exon 7) [24, 28–31].

Apart from the above-mentioned, expressed mutant PRKAR1A protein isoforms, several other expressed isoforms that result from single amino acid substitutions have been reported [29, 30]. Detailed *in vitro* analysis of their effects on protein function have revealed important PRKAR1A domain features [30, 32]. The six naturally occurring missense substitutions examined by this study (p.Ser9Asn, p.Arg74Cys, p.Arg146Ser, p.Asp183Tyr, p.Ala213Asp, p.Gly289Trp) are spread over all the functional domains of the protein. Although, as mentioned before, the low number of individuals affected by each of these mutations prevented detailed phenotype–genotype analysis, these studies support the previous suggestion that the alteration of PRKAR1A function alone (and not only its complete loss) is sufficient to increase PKA activity, leading to CNC.

Until recently, no genotype–phenotype correlations had been found for the different stop codon mutations, which are expected to uniformly lead to lack of the *PRKAR1A* mutant allele's protein product in cells. This was because most of the mutations were identified in single patients only and only two (c.491–492delTG/

p.Val164fsX4, and c.709(-7-2) del6(TTTTTA)) had been seen in more than three kindreds [5, 28]. The first study to explore all *PRKAR1A* mutations found to date against all CNC phenotypes was recently completed; 353 individuals, 258 of whom (73%) were positive for a *PRKAR1A* mutation, were studied [11]. Several features that distinguish *PRKAR1A* mutation carriers from mutation-negative CNC patients were identified; the former presented more frequently and earlier in life with pigmented skin lesions, myxomas, thyroid, and gonadal tumours. In addition, essential correlations between certain genetic defects and the severity and type of CNC manifestations were found. Bertherat *et al.* [11] outlined subgroups of patients; the first group presented with isolated PPNAD, in some cases accompanied with lentiginosis. In this group the following tendencies were observed: (1) patients diagnosed before 8 years of age were rarely carriers of *PRKAR1A* mutations; and (2) most of the patients with isolated PPNAD and the presence of *PRKAR1A* mutation were carriers of either the c.709(-7-2) del6(TTTTTA) mutation (P <0.0001) or the c.1A>G/p.Met1Val substitution affecting the initiation codon of the protein. These observations were in line with previously published reports [5, 28] and both mutations are rather rare. Although the molecular mechanism of the p.Met1Val substitution is not completely clear, it is the only mutation that alters the protein initiation site, and may, in theory, result in alternative initiation [33]; c.709(-7-2) del6(TTTTTA) is a splice variant that is expected to result in an exon skip, frame shift, and premature stop codon generation. However, since it does not affect the two immediate nucleotides on either site of the splice junction, it is expected to lead to splicing in less than 100% of the molecules that harbour it, and thus, presumably, to lead to a milder phenotype. The fact that a milder phenotype involves only the adrenal and skin is suggestive of their high sensitivity to changes in PKA activity.

The second group of CNC patients that was suggested to have a particular genotype–phenotype correlation comprised individuals with myxomas (affecting all locations—skin, heart, and breast), PMS, thyroid tumours, and large-cell calcifying Sertoli cell tumours (LCCSCT). In these patients, *PRKAR1A* mutations were seen substantially more often. Related to this is the recognition that certain tumours present at a significantly younger age in *PRKAR1A* mutation carriers: cardiac myxomas (P = 0.02), thyroid tumours (P = 0.03), and LCCSCTs (P = 0.04) [11]. Another finding in these patients was that mutations that escaped NMD and led to an alternate, usually shorter, protein were associated with an overall higher total number of CNC manifestations (P = 0.04).

In terms of pigmented skin lesions in CNC, two important correlations have been observed: (1) lentigines (as well as PMS, acromegaly, and cardiac myxomas), were seen significantly more often in CNC patients with exonic *PRKAR1A* mutations, compared to those with intronic ones (P = 0.04); and (2) lentigines (as well as cardiac myxoma and thyroid tumours) were significantly associated with the hot spot c.491–492delTG mutation compared to all other *PRKAR1A* defects (P = 0.03). These data add greatly to the understanding of the molecular mechanisms of the involvement of *PRKAR1A* in endocrine and other tumourigenesis and, thus, for genetic counselling and prognosis in CNC families.

Interestingly, a 2.3-Mb deletion in chromosome band 17q24.2–q24.3, which involved *PRKAR1A* together with another 13 genes, resulted in a number of clinical features, including posterior laryngeal cleft, growth restriction, microcephaly, and moderate mental retardation. The only CNC manifestation was numerous freckles and lentigines at a young age [34]; the authors called the observed phenotype 'CNC plus'.

To date, the molecular causes underlying the formation of pigmented skin lesions in CNC are not fully understood. A possible mechanism involves the PKA-mediated activation of pathways downstream of the melanocortin receptors (MCRs), which form a subfamily of the G protein-coupled receptors (GPCRs) and regulate a wide variety of processes, including skin pigmentation [35–37]. The melanocortin 1 receptor (MC1R) is expressed preferentially in epidermal melanocytes and is known to be the key regulator of mammalian pigmentation [36, 38]. MC1R is stimulated by the proopiomelanocortin-derived melanocyte-stimulating hormone and adrenocorticotrophic hormone (ACTH) and, in turn, activates the rate-limiting enzyme in melanin synthesis, tyrosinase. As a GPCR, MC1R is positively coupled with adenylate cyclase, and its actions are mainly mediated by PKA, in coordination with other signalling molecules involving protein kinase C (PKC) and MAPKs [39–41].

Relationship to Other Syndromes

CNC shares clinical features and molecular pathways with several other familial lentiginosis syndromes, such as McCune–Albright syndrome (OMIM #174800), Peutz–Jeghers (OMIM #175200), LEOPARD (OMIM #151100), Noonan's (OMIM #163950), Cowden's disease (OMIM #158350), and Bannayan–Ruvalcaba–Riley syndrome (OMIM #153480). In all of these conditions skin lesions accompany underlying endocrine and/or other abnormalities, which, as in CNC, are considered an important diagnostic sign.

Probably the closest, at least in terms of a molecular pathway link, to CNC is McCune–Albright syndrome. Patients with this condition have characteristic lesions that affect predominantly three systems: the skin, the endocrine system, and the skeleton. The café-au-lait spots in McCune–Albright syndrome patients are similar to those observed in CNC, but tend to be more intensely pigmented. McCune–Albright syndrome is caused by postzygotic, activating, somatic mutations of *GNAS,* located on 20q13, which encodes the adenylate cyclase-stimulating G α protein (Gsa) of the heterotrimeric G protein [42]. G proteins couple hormone receptors to adenylyl cyclase and are therefore required for hormone-stimulated cAMP synthesis. Because of the somatic nature of the genetic defect, the presentation of the disease is mosaic and the level of clinical involvement of any tissue is highly variable. The mutations in *GNAS* are always missense substitutions at the critical sites for the GTPase inactivation (amino acid positions p.Arg201 and p.Gln227), and, in contrast to *PRKAR1A* defects, lead to constant protein activation and increased cAMP production.

We have reported another endocrine lesion that is associated with increased tissue levels of cAMP, isolated micronodular adrenocortical hyperplasia (iMAD). In these patients, inactivating mutations in the genes encoding phosphodiesterases types 11A (*PDE11A*) and 8B (*PDE8B*) have been reported [43–45]. iMAD patients were initially considered CNC patients, but it soon became clear that iMAD is not the same as PPNAD [46].

Peutz–Jeghers syndrome, another autosomal dominant familial lentiginosis syndrome, is characterized by melanocytic macules of the lips, buccal mucosa, and digits, multiple gastrointestinal hamartomatous polyps, and an increased risk of various neoplasms.

The lentigines observed in patients with Peutz–Jeghers syndrome shows similar density and distribution to the ones in CNC [17]. Peutz–Jeghers syndrome has been elucidated at the molecular level [47, 48]; the disease was first mapped to chromosome 19p13.3 and, soon after, the gene encoding the serine–threonine kinase 11 (*STK11* also known as *LKB1*) was found to be mutated in most patients [47–49]. The proposed mechanism of the disease is through elimination of the kinase activity of the STK11/LKB1 tumour suppressor protein.

LEOPARD is an acronym for the manifestations of the syndrome comprising: multiple lentigines, electrocardiographic conduction abnormalities, ocular hypertelorism, pulmonic stenosis, abnormal genitalia, retardation of growth, and sensorineural deafness [50]. LEOPARD is allelic to Noonan's syndrome; both diseases are linked to mutations in *PTPN11* (12q24), the gene encoding the non-receptor tyrosine phosphatase Shp-2 [51, 52]. The protein encoded by this gene is a member of the protein tyrosine phosphatase family, proteins that are known to regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.

Cowden's disease and Bannayan–Ruvalcaba–Riley syndrome share clinical characteristics, including mucocutaneous lesions, hamartomatous polyps of the gastrointestinal tract, and increased risk of developing neoplasms. They are described extensively elsewhere in this book (see Chapter 6.15). Both conditions are caused by mutations in the *PTEN* gene [53–55]. *PTEN* is located on 10q23.31 and encodes phosphatidylinositol-3, 4, 5-trisphosphate 3-phosphatase. The gene was recognized as a tumour suppressor gene and has been found to be mutated in a number of tumours [56]. It contains a tensin-like domain as well as a catalytic domain similar to that of the dual-specificity protein tyrosine phosphatases. Unlike most of the protein tyrosine phosphatases, PTEN preferentially dephosphorylates phosphoinositide substrates. It negatively regulates intracellular levels of phosphatidylinositol-3, 4, 5-trisphosphate in cells and its tumour suppressor effect is expressed by inhibition of the AKT/PKB signalling pathway.

The overlapping clinical manifestations of these syndromes, which are caused by distinct molecular defects, suggest crosstalk between the involved pathways. Indeed, PRKAR1A inactivation leads to phosphorylation of mTOR and ERK1/2 [57, 58], LKB1 is phosphorylated by PKA [59], and PTEN expression is positively regulated by transcription factor Egr-1 in a PKA-dependent manner [60].

ACKNOWLEDGEMENTS

Studies on CNC and related syndromes have been supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development, NIH, intramural project Z01-HD-000642–04 (to Dr C.A. Stratakis).

REFERENCES

1. Carney JA, Gordon H, Carpenter PC, *et al.* The complex of myxomas, spotty pigmentation, and endocrine overactivity. *Medicine (Baltimore),* 1985; **64**: 270–83.
2. Tsay CJ, Stratakis CA, Faucz FR, *et al.* Harvey Cushing treated the first known patient with Carney complex. *J Endocr Soc*, 2017; **1**: 1312–21.
3. Boikos SA, Stratakis CA. Carney complex: pathology and molecular genetics. *Neuroendocrinology*, 2006; **83**: 189–99.
4. Boikos SA, Stratakis CA. Carney complex: the first 20 years. *Curr Opin Oncol*, 2007; **19**: 24–9.
5. Stratakis CA, Kirschner LS, Carney JA. Clinical and molecular features of the Carney complex: diagnostic criteria and recsommendations for patient evaluation. *J Clin Endocrinol Metab*, 2001; **86**: 4041–6.
6. Atherton DJ, Pitcher DW, Wells RS, *et al.* A syndrome of various cutaneous pigmented lesions, myxoid neurofibromata and atrial myxoma: the NAME syndrome. *Br J Dermatol*, 1980; **103**: 421–9.
7. Rhodes AR, Silverman RA, Harrist TJ, *et al.* Mucocutaneous lentigines, cardiomucocutaneous myxomas, and multiple blue nevi: the 'LAMB' syndrome. *J Am Acad Dermatol*, 1984; **10**: 72–82.
8. Carney JA, Headington JT, Su WP. Cutaneous myxomas. A major component of the complex of myxomas, spotty pigmentation, and endocrine overactivity. *Arch Dermatol*, 1986; **122**: 790–8.
9. Jabbour SA, Davidovici BB, Wolf R. Rare syndromes. *Clin Dermatol*, 2006; **24**: 299–316.
10. Mateus C, Palangie A, Franck N, *et al.* Heterogeneity of skin manifestations in patients with Carney complex. *J Am Acad Dermatol*, 2008; **59**: 801–10.
11. Bertherat J, Horvath A, Groussin L, *et al.* Mutations in regulatory subunit type 1A of cyclic adenosine 5'-monophosphate-dependent protein kinase (PRKAR1A): phenotype analysis in 353 patients and 80 different genotypes. *J Clin Endocrinol Metab*, 2009; **94**: 2085–91.
12. Hernandez-Ramirez LC, Tatsi C, Lodish MB, *et al.* Corticotropinoma as a component of Carney complex. *J Endocr Soc*, 2017; **1**: 918–25.
13. Kiefer FW, Winhofer Y, Iacovazzo D, *et al.* PRKAR1A mutation causing pituitary-dependent Cushing disease in a patient with Carney complex. *Eur J Endocrinol*, 2017; **177**: K7–K12.
14. Bertherat J. Carney complex (CNC). *Orphanet J Rare Dis*, 2006; **1**: 21.
15. Sandrini F, Stratakis C. Clinical and molecular genetics of Carney complex. *Mol Genet Metab*, 2003; **78**: 83–92.
16. Bauer AJ, Stratakis CA. The lentiginoses: cutaneous markers of systemic disease and a window to new aspects of tumourigenesis. *J Med Genet*, 2005; **42**: 801–10.
17. Stratakis CA. Genetics of Peutz–Jeghers syndrome, Carney complex and other familial lentiginoses. *Horm Res*, 2000; **54**: 334–43.
18. Young WF, Jr., Carney JA, Musa BU, *et al.* Familial Cushing's syndrome due to primary pigmented nodular adrenocortical disease. Reinvestigation 50 years later. *N Engl J Med*, 1989; **321**: 1659–64.
19. Carney JA, Ferreiro JA. The epithelioid blue nevus. A multicentric familial tumor with important associations, including cardiac myxoma and psammomatous melanotic schwannoma. *Am J Surg Pathol*, 1996; **20**: 259–72.
20. Zembowicz A, Carney JA, Mihm MC. Pigmented epithelioid melanocytoma: a low-grade melanocytic tumor with metastatic potential indistinguishable from animal-type melanoma and epithelioid blue nevus. *Am J Surg Pathol*, 2004; **28**: 31–40.
21. Stratakis CA, Carney JA, Lin JP, *et al.* Carney complex, a familial multiple neoplasia and lentiginosis syndrome. Analysis of 11 kindreds and linkage to the short arm of chromosome 2. *J Clin Invest*, 1996; **97**: 699–705.
22. Horvath A, Bertherat J, Groussin L, *et al.* Mutations and polymorphisms in the gene encoding regulatory subunit type 1-alpha of protein kinase A (PRKAR1A): an update. *Hum Mutat*, 2010; **31**: 369–79.

23. Salpea P, Horvath A, London E, *et al*. Deletions of the PRKAR1A locus at 17q24.2-q24.3 in Carney complex: genotype-phenotype correlations and implications for genetic testing. *J Clin Endocrinol Metab*, 2014; **99**: E183–8.
24. Horvath A, Bossis I, Giatzakis C, *et al*. Large deletions of the PRKAR1A gene in Carney complex. *Clin Cancer Res*, 2008; **14**: 388–95.
25. Tasken K, Skalhegg BS, Tasken KA, *et al*. Structure, function, and regulation of human cAMP-dependent protein kinases. *Adv Second Messenger Phosphoprotein Res*, 1997; **31**: 191–204.
26. Shabb JB. Physiological substrates of cAMP-dependent protein kinase. *Chem Rev*, 2001; **101**: 2381–411.
27. Robinson-White A, Meoli E, Stergiopoulos S, *et al*. PRKAR1A Mutations and protein kinase A interactions with other signaling pathways in the adrenal cortex. *J Clin Endocrinol Metab*, 2006; **91**: 2380–8.
28. Groussin L, Horvath A, Jullian E, *et al*. A PRKAR1A mutation associated with primary pigmented nodular adrenocortical disease in 12 kindreds. *J Clin Endocrinol Metab*, 2006; **91**: 1943–9.
29. Veugelers M, Wilkes D, Burton K, *et al*. Comparative PRKAR1A genotype-phenotype analyses in humans with Carney complex and prkar1a haploinsufficient mice. *Proc Natl Acad Sci U S A*, 2004; **101**: 14222–7.
30. Greene EL, Horvath AD, Nesterova M, *et al*. *In vitro* functional studies of naturally occurring pathogenic PRKAR1A mutations that are not subject to nonsense mRNA decay. *Hum Mutat*, 2008; **29**: 633–9.
31. Bonnet-Serrano F, Bertherat J. Genetics of tumors of the adrenal cortex. *Endocr Relat Cancer*, 2018; **25**: R131–52.
32. Horvath A, Giatzakis C, Tsang K, *et al*. A cAMP-specific phosphodiesterase (PDE8B) that is mutated in adrenal hyperplasia is expressed widely in human and mouse tissues: a novel PDE8B isoform in human adrenal cortex. *Eur J Hum Genet*, 2008; **16**: 1245–53.
33. Kirschner LS, Sandrini F, Monbo J, *et al*. Genetic heterogeneity and spectrum of mutations of the PRKAR1A gene in patients with the carney complex. *Hum Mol Genet*, 2000; **9**: 3037–46.
34. Blyth M, Huang S, Maloney V, *et al*. A 2.3 Mb deletion of 17q24.2-q24.3 associated with 'Carney Complex plus'. *Eur J Med Genet*, 2008; **51**: 672–8.
35. Butler AA, Cone RD. The melanocortin receptors: lessons from knockout models. *Neuropeptides*, 2002; **36**: 77–84.
36. Abdel-Malek ZA. Melanocortin receptors: their functions and regulation by physiological agonists and antagonists. *Cell Mol Life Sci*, 2001; **58**: 434–41.
37. Gantz I, Fong TM. The melanocortin system. *Am J Physiol Endocrinol Metab*, 2003; **284**: E468–74.
38. Kadekaro AL, Kanto H, Kavanagh R, *et al*. Significance of the melanocortin 1 receptor in regulating human melanocyte pigmentation, proliferation, and survival. *Ann N Y Acad Sci*, 2003; **994**: 359–65.
39. Busca R, Ballotti R. Cyclic AMP a key messenger in the regulation of skin pigmentation. *Pigment Cell Res*, 2000; **13**: 60–9.
40. Tsatmali M, Ancans J, Yukitake J, *et al*. Skin POMC peptides: their actions at the human MC-1 receptor and roles in the tanning response. *Pigment Cell Res*, 2000; **13** Suppl 8: 125–9.
41. Busca R, Abbe P, Mantoux F, *et al*. Ras mediates the cAMP-dependent activation of extracellular signal-regulated kinases (ERKs) in melanocytes. *EMBO J*, 2000; **19**: 2900–10.
42. Weinstein LS, Shenker A, Gejman PV, *et al*. Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *N Engl J Med*, 1991; **325**: 1688–95.
43. Horvath A, Giatzakis C, Robinson-White A, *et al*. Adrenal hyperplasia and adenomas are associated with inhibition of phosphodiesterase 11A in carriers of PDE11A sequence variants that are frequent in the population. *Cancer Res*, 2006; **66**: 11571–5.
44. Horvath A, Boikos S, Giatzakis C, *et al*. A genome-wide scan identifies mutations in the gene encoding phosphodiesterase 11A4 (PDE11A) in individuals with adrenocortical hyperplasia. *Nat Genet*, 2006; **38**: 794–800.
45. Horvath A, Mericq V, Stratakis CA. Mutation in PDE8B, a cyclic AMP-specific phosphodiesterase in adrenal hyperplasia. *N Engl J Med*, 2008; **358**: 750–2.
46. Gunther DF, Bourdeau I, Matyakhina L, *et al*. Cyclical Cushing syndrome presenting in infancy: an early form of primary pigmented nodular adrenocortical disease, or a new entity? *J Clin Endocrinol Metab*, 2004; **89**: 3173–82.
47. Hemminki A, Markie D, Tomlinson I, *et al*. A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature*, 1998; **391**: 184–7.
48. Jenne DE, Reimann H, Nezu J, *et al*. Peutz–Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet*, 1998; **18**: 38–43.
49. Meserve EE, Nucci MR. Peutz–Jeghers syndrome: pathobiology, pathologic manifestations, and suggestions for recommending genetic testing in pathology reports. *Surg Pathol Clin*, 2016; **9**: 243–68.
50. Gorlin RJA, Anderson RC, Blaw M. Multiple lentigenes syndrome. *Am J Dis Child*, 1969; **117**: 652–62.
51. Jamieson CR, van der Burgt I, Brady AF, *et al*. Mapping a gene for Noonan syndrome to the long arm of chromosome 12. *Nat Genet*, 1994; **8**: 357–60.
52. van der Burgt I, Berends E, Lommen E, *et al*. Clinical and molecular studies in a large Dutch family with Noonan syndrome. *Am J Med Genet*, 1994; **53**: 187–91.
53. Nelen MR, Padberg GW, Peeters EA, *et al*. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet*, 1996; **13**: 114–16.
54. Liaw D, Marsh DJ, Li J, *et al*. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet*, 1997; **16**: 64–7.
55. Marsh DJ, Dahia PL, Zheng Z, *et al*. Germline mutations in PTEN are present in Bannayan-Zonana syndrome. *Nat Genet*, 1997; **16**: 333–4.
56. Yin Y, Shen WH. PTEN: a new guardian of the genome. *Oncogene*, 2008; **27**: 5443–53.
57. Robinson-White A, Hundley TR, Shiferaw M, *et al*. Protein kinase-A activity in PRKAR1A-mutant cells, and regulation of mitogen-activated protein kinases ERK1/2. *Hum Mol Genet*, 2003; **12**: 1475–84.
58. Mavrakis M, Lippincott-Schwartz J, Stratakis CA, *et al*. Depletion of type IA regulatory subunit (RIalpha) of protein kinase A (PKA) in mammalian cells and tissues activates mTOR and causes autophagic deficiency. *Hum Mol Genet*, 2006; **15**: 2962–71.
59. Collins SP, Reoma JL, Gamm DM, *et al*. LKB1, a novel serine/threonine protein kinase and potential tumour suppressor, is phosphorylated by cAMP-dependent protein kinase (PKA) and prenylated in vivo. *Biochem J*, 2000; **345** Pt 3: 673–80.
60. Fernandez S, Garcia-Garcia M, Torres-Aleman I. Modulation by insulin-like growth factor I of the phosphatase PTEN in astrocytes. *Biochim Biophys Acta*, 2008; **1783**: 803–12.

6.14

Molecular and Clinical Characteristics of the McCune–Albright Syndrome

Michael A. Levine and Steven A. Lietman

Introduction

Heterotrimeric guanine nucleotide binding proteins (G proteins) couple a diverse spectrum of extracellular receptor proteins to intracellular effector enzymes and ion channels. The observation that alterations in G protein-coupled signalling pathways can impact cellular function and proliferation, and cause human disease, has stimulated investigation into the molecular and pharmacological regulation of G protein expression and action. The most well-characterized models for altered G protein expression defects have been based on naturally occurring mutations in *GNAS*, a complex imprinted gene at 20q13 that encodes the α subunit of Gs, the G protein that stimulates adenylyl cyclase (AC). Somatic mutations in GNAS that activate Gαs (the *gsp* oncogene) are present in a subset of endocrine tumours, fibrous dysplasia, and in patients with the McCune–Albright syndrome, a sporadic disorder characterized by increased hormone production and/or cellular proliferation of many tissues [1]. The clinical phenotype is therefore highly variable, and depends upon the location and timing of the mutation during embryologic development as well as the parental origin of *GNAS* allele that carries the mutation.

By contrast, germline mutations of the *GNAS* gene that decrease expression or function of $G\alpha_s$ are present in subjects with Albright hereditary osteodystrophy (AHO), a heritable disorder associated with a constellation of developmental defects and, in many patients, reduced responsiveness to multiple hormones that signal through receptors that require $G\alpha_s$ to activate AC (i.e. pseudohypoparathyroidism type 1a) [2]. MAS and AHO represent contrasting gain of function and loss of function mutations in the *GNAS* gene, respectively. Clinical and biochemical analyses of subjects with these syndromes have extended our understanding of the developmental and functional consequences of dysfunctional G protein action, and have provided unexpected insights into the importance of cAMP as a regulator of the growth and/or function of many tissues. This chapter will focus on the clinical implications of activating mutations of *GNAS* as the basis for MAS.

G Protein Structure and Signalling

G proteins share a common heterotrimeric structure consisting of an α subunit and a tightly coupled βγ dimer. The α subunit interacts with detector and effector molecules, binds guanosine triphosphate (GTP), and possesses intrinsic GTPase activity. There are 16 genes in mammals that encode some 20 different α chains. The α subunits associate with a smaller group of β (at least 5) and γ (more than 12) subunits [3]. The Gα subunits are categorized in four classes, and include Gsα (G stimulatory); Giα (G inhibitory) and Goα (G other); Gq/11α; and G12/13α. They behave differently in the recognition of the effector, but share similar structures and mechanism of activation. The Gα subunits consist of two domains: a GTP-binding domain and a helical insertion domain. The GTP-binding domain is homologous to Ras-like small GTPases, and includes switch regions I and II, which change conformation during activation. The switch regions are loops of alpha-helices with conformations sensitive to guanine nucleotides. The helical insertion domain is inserted into the GTP-binding domain before switch region I and is unique to heterotrimeric G

proteins. This helical insertion domain functions to sequester the guanine nucleotide at the interface with the GTP-binding domain and must be displaced to enable nucleotide dissociation.

Combinatorial specificity in the associations between various G protein subunits provides the potential for enormous diversity, and may allow distinct heterotrimers to interact selectively with only a limited number of effector proteins and G protein-coupled receptor (GPCR) proteins. The GPCRs are encoded by more than 1% of the genome of vertebrates, and represent a large protein family of receptors that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. GPCRs, also termed 7TM receptors, heptahelical receptors, and serpentine receptors, contain alpha helical regions that pass through the cell membrane seven times. The human genome encodes roughly 950 GPCRs, which detect photons (light), hormones, growth factors, drugs, and other endogenous ligands. Approximately 150 of the GPCRs found in the human genome have unknown functions. GPCRs are activated by a wide variety of signals, including proteins, nucleotides, amino acid residues, Ca^{2+}, light photons, and odorants, and it is postulated that ligand-binding alters the conformation of transmembrane domains and intracellular loops, increasing the affinity of the receptor for specific heterotrimeric G proteins.

G protein coupled signalling is regulated by a mechanism in which the binding and hydrolysis of GTP acts as a molecular timing switch (Figure 6.14.1). In the basal (inactive) state, G proteins exist in the heterotrimeric form with GDP bound to the α chain. The interaction of a ligand-bound receptor with a G protein facilitates the release of tightly bound GDP and the subsequent binding of cytosolic GTP. The binding of GTP to the α chain induces conformational changes that facilitate the dissociation of the α-GTP chain from the βγ dimer and the receptor. The free α-GTP chain assumes an active conformation in which a new surface is formed that enables the α chain to interact with target enzymes and ion channels with 20- to 100-fold higher affinity than in the GDP bound state. The βγ dimers also participate in downstream signalling events through interaction with an ever-widening array of targets, including certain forms of AC and phospholipase C, potassium channels, and GPCR kinases.

G protein signalling is terminated by the hydrolysis of α-GTP to α-GDP by an intrinsic GTPase. The GTPase reaction is a high-energy transition state that requires association of the γ-phosphorus atom with the oxygen of a water molecule. To catalyse this reaction, the γ-phosphate of GTP must be stabilized so that a straight line, perpendicular to the plane of the γ-phosphate, connects the water, the γ-phosphorus, and the oxygen molecule leaving the β-phosphate. In $G\alpha_s$, amino acids arginine^{201} and glutamine^{227} function as 'fingers' to position the γ-phosphate of GTP. With hydrolysis of GTP to GDP, the α-GDP chain re-associates with the βγ dimer and the heterotrimeric G protein is capable of participating in another cycle of receptor-activated signalling (Figure 6.14.1).

The GTPase of the Gα chain provides a molecular timing switch that controls the duration, and thereby the intensity, of the signalling event. The intrinsic rates of GTP hydrolysis by G protein α chains differ widely, and interactions that influence the rate of the GTPase reaction can have profound consequences. Several factors, termed **GAPs** for 'GTPase activating proteins', can interact directly with specific α chains to accelerate the slow intrinsic rate of GTP hydrolysis. One important class of GAP's is represented by the evolutionarily conserved superfamily of RGS proteins, for 'regulators of G protein signalling', that can stimulate a 40-fold increase in the catalytic rate of GTP hydrolysis, and thus can markedly accelerate the termination of G protein signalling. On the other hand, inhibition of intrinsic GTPase by modification or replacement of key

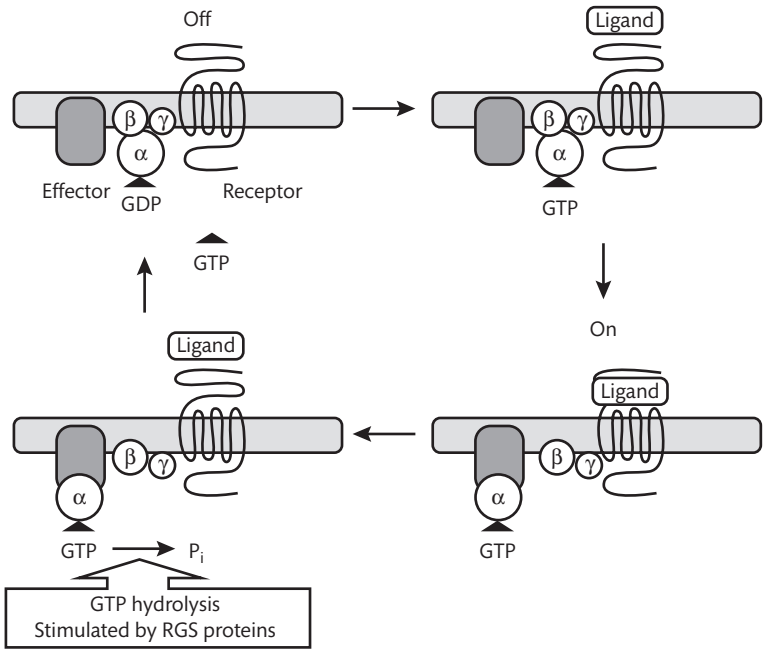


Figure 6.14.1 The cycle of hormone-dependent GTP binding and hydrolysis that regulates heterotrimeric G protein signal transduction. In the non-stimulated, basal (Off) state, GDP is tightly bound to the α chain of the heterotrimeric G protein. Binding of an agonist (Ligand) to its receptor (depicted with seven transmembrane spanning domains) induces a conformational change in the receptor, and enables it to activate the G protein. The G protein now releases GDP and binds GTP present in the cytosol. The binding of GTP to the α chain leads to dissociation of the α-GTP from the βγ dimer, and each of these molecules is now free to regulate downstream effector proteins. The hydrolysis of GTP to GDP by the intrinsic GTPase of the α chain promotes reassociation of α-GDP with βγ and the inactive state is restored. The heterotrimeric G protein is ready for another cycle of hormone-induced activation.

amino acid residues (e.g. arginine^{201} or glutamine^{227} in $G\alpha_s$) can delay termination of the signal transduction process, and cause persistent and excessive signalling. For example, exotoxins secreted by *Vibrio cholerae* and some strains of *E. coli* catalyse the addition of an ADP-ribose moiety to the side chain of arginine^{201} in $G\alpha_s$. This covalent modification markedly reduces GTP hydrolysis and maintains $G\alpha_s$ in its active GTP-bound form, thus resulting in persistent stimulation of AC [4]. The subsequent accumulation of cAMP in intestinal epithelial cells stimulates secretion of salt and water into the intestine and produces, in part, the watery diarrhoea associated with cholera. A drug screening assay has been developed to determine the efficacy of small molecule inhibitors for the constitutively active signalling with *gsp* mutations in MAS [5]. This assay could identify small molecule inhibitors for the treatment of MAS.

Activating Mutations of the *GNAS* Gene Induce Cellular Proliferation

Activity of AC is under dual regulatory control through receptors that interact with either G_s to stimulate AC or with G_i to inhibit AC. Increased intracellular cAMP stimulates proliferation of many cell types, and can increase synthesis and secretion of endogenous hormones and neurotransmitters. Both germline and somatic mutations in *GNAS* that lead to a gain of function in $G\alpha_s$ produce constitutive (i.e. hormone independent) activation of AC. Vallar *et al.* initially described a subset of human growth hormone secreting pituitary tumours in which basal AC activity *in vitro* was very high and failed to increase further with addition of growth hormone releasing hormone [6]. Subsequent studies showed that these somatotropic tumours contained unusual forms of $G\alpha_s$ that lacked GTPase activity due to somatic mutations in *GNAS* that replace either arginine^{201} or glutamine^{227} and thereby convert *GNAS* into the *gsp* oncogene. Arginine^{201} corresponds to the site of choleragen modification of $G\alpha_s$, whereas glutamine^{227} in $G\alpha_s$ corresponds to the cognate amino acid Gln^{61} in the low molecular weight GTP-binding protein $p21^{ras}$. Naturally occurring Gln^{61} mutations convert $p21^{ras}$ into an oncogene that plays a role in the development of a variety of human tumours [7]. Replacement of either arginine^{201} or glutamine^{227} in $G\alpha_s$ enable the protein to remain in the active, GTP-bound state, and the consequent increase in cAMP leads to cellular proliferation and excessive hormone secretion [8, 9]. Such activating mutations occur in approximately 40% of somatotropic tumours (Table 6.14.1). In addition to growth hormone secreting pituitary tumours, *gsp* mutations are also present in a small number

Table 6.14.1 Clinical manifestations of McCune–Albright syndrome. Clinical data compiled from over 250 cases of MAS reported in the literature (summarized in [36, 40, 44, 77, 84, 98, 112]). Evaluations include clinical and biochemical data; other rarely described manifestations include metabolic acidosis, nephrocalcinosis, intellectual disability, thymic and splenic hyperplasia, and colonic polyps

	% of all patients affected	% of males affected	% of females affected	Age at diagnosis, years (range)	Comments
Fibrous dysplasia	98	96	98	7.7 (0–52)	Polyostotic more common than mono-ostotic
Cafe'-au-lait lesions	85	92	82	7.7 (0–52)	Variable size and number of lesions, irregular border ('coast of Maine')
Precocious puberty	52	15	70	4.9 (0.3–9)	Common initial manifestation
Acromegaly/Gigantism	27	38	21	14.8 (0.2–42)	65% with adenoma on MRI/CT
Hyperprolactinaemia	15	17	13	16.0 (0.2–42)	55% of acromegalics with ↑ PRL
Hyperthyroidism	19	13	22	14.4 (0.5–37)	Euthyroid goitre is common; ultrasound generally shows nodules
Hypercortisolism	6	8	5	4.4 (0.2–17)	All primary adrenal
Myxomas	5	6	5	34 (17–50)	Extremity myxomas
Osteosarcoma	2	2	3	36 (34–37)	At sites of fibrous dysplasia, not related to prior radiation therapy
Rickets/Osteomalacia	3	2	3	27.3 (8–52)	Responsive to phosphorus plus calcitriol; most likely due to excess section of FGF23 from fibrous dysplasia
Cardiac abnormalities	11	15	9	(0.1–66)	Arrhythmias and CHF reported
Hepatic abnormalities	10	11	10	1.9 (0.3–4)	Neonatal icterus is most common
Testicular abnormalities	81				Hyperechoic lesions (49%), hypoechoic lesions (30%), microlithiasis (30%), heterogeneity (47%), and focal calcifications (11%). Precocious puberty in 21%
Gastrointestinal abnormalities	56%	52	48	35.1 years	Radiographic GI abnormalities: 46% with intraductal pancreatic papillary mucinous neoplasms; also, gastric heterotopia/metaplasia, gastric hyperplastic polyps, fundic gland polyps, and hamartomatous polyp
Hearing loss	22.4% of ears in patients with temporal bone fibrous dysplasia	55	45	Median age 19.6 years; range, 4.6–80.3 years	Conductive in 65.9%, sensorineural in 29.3%, and mixed in 4.9%. Hearing loss was mild and non-progressive in most participants

Abbreviations: PRL, prolactin; CHF, congestive heart failure.

Box 6.14.1 Clinical syndromes associated with activating mutations of *GNAS*

Missense mutations of *GNAS* at Arg^{201} and Gln^{227} which cause constitutive activation of AC and the cAMP signalling cascade have been identified in patients with McCune–Albright syndrome (MAS) and subsets of a variety of endocrine tumours.

- McCune–Albright syndrome (100%)
- Pituitary adenomas (4–50%)
  - Growth hormone-secreting adenomas (35–40%)
  - ACTH-secreting adenomas (4–9%)
  - Clinically non-functioning adenomas (rare)
- Thyroid neoplasms (3–70%)
  - Hyperfunctioning and non-functioning follicular adenomas
  - Papillary and follicular carcinomas
- Parathyroid neoplasms (<5%)
  - Parathyroid adenomas
- Adrenocortical disorders (<5%)
  - Aldosterone-producing adenomas
  - Adrenal hyperplasia
  - Phaeochromocytoma
- Leydig cell and ovarian neoplasms (66%)

of ACTH-secreting pituitary tumours [10, 11], a subset of thyroid neoplasms, and testicular and ovarian stromal Leydig tumours [12], but are rare in other endocrine tumours (Box 6.14.1). Moreover, *gsp* mutations have been described in ovarian cysts that cause isosexual gonadotropin-independent precocious puberty [13, 14], in intramuscular myxomas [15] and in isolated fibrous dysplasia of the bone [16].

Molecular Basis for the McCune–Albright Syndrome

In 1937, McCune and Bruch [17] and Albright and associates [18] independently described a sporadic syndrome characterized by the clinical triad of polyostotic fibrous dysplasia, *café-au-lait* skin lesions, and endocrine hyperfunction now known as McCune–Albright syndrome (MAS) (Figure 6.14.2). Despite excessive activity of endocrine tissues, serum levels of the relevant regulatory or tropic hormones were either normal or decreased, suggesting autonomous function. Based on the observation that the cutaneous hyperpigmentation in MAS typically follows the developmental lines of Blaschko, Happle proposed that the underlying genetic abnormality might be a dominantly acting somatic mutation that occurs early in development and which leads to a mosaic pattern of distribution of mutant cells [19]. Similarly, a lack of documented heritability of MAS has been interpreted as evidence that germline transmission of the mutation would be lethal [19].

The molecular basis for MAS is a somatic mutation in exon 8 of *GNAS* that replaces the residue arginine at position 201 by histidine or cysteine [20, 21]. Very infrequently, arginine is replaced by serine, glycine, or leucine [22]. Although missense mutations that replace the nearby glutamine at position 227 have been identified in solitary endocrine tumours, they have not been described in patients with MAS.

Consistent with a postzygotic somatic mutation, cells containing the *gsp* mutation are not present in all tissues of patients with MAS. Rather, cells containing a mutant *GNAS* gene are distributed in a mosaic pattern, with the greatest number of *gsp*-containing cells present in the most abnormal areas of affected tissues (Figure 6.14.3) [20, 21, 23]. In some cases, *gsp* alleles may be present in only some cell types within tissues that are derived from different embryological precursors. For example, a 3-year-old male MAS patient with macro-orchidism but no precocious puberty was reported to have a R201H *gsp* allele present only in Sertoli cells, resulting in isolated Sertoli cell hyperfunction, evidenced by increased Anti-Müllerian hormone (AMH) expression and cell hyperplasia leading to prepubertal macro-orchidism. There were no signs of Leydig cell activation, and no evidence of excess androgen action [24, 25]. The different early embryologic origin of precursors contributing to Sertoli and Leydig cell lineages may underlie the differential existence of the mutated *GNAS* gene.

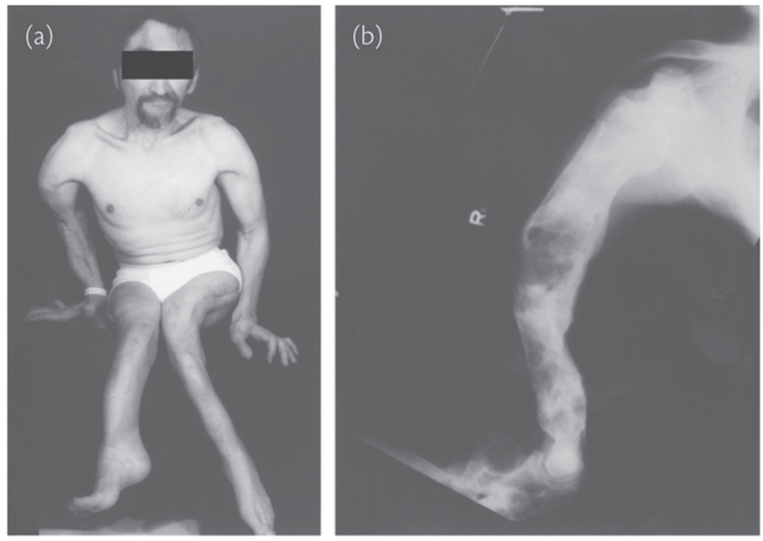


Figure 6.14.2 Patient with McCune–Albright syndrome. This patient demonstrates the complete clinical triad of McCune–Albright syndrome, with café-au-lait, polyostotic fibrous dysplasia, and excessive endocrine function (hyperthyroidism). The fibrous dysplasia has affected his skull and long bones and led to progressive and debilitating deformity. The classic features (right panel) of fibrous dysplasia are illustrated in this radiograph of his right upper extremity, which reveals expansile, lytic lesions with a 'ground glass' pattern and a scalloped border secondary to endosteal erosion.

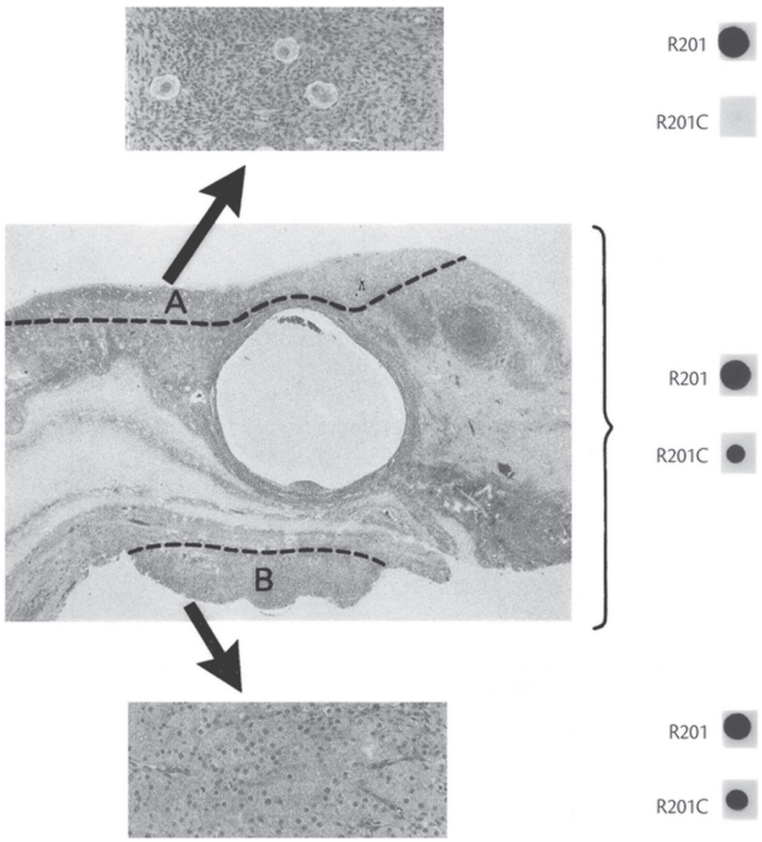


Figure 6.14.3 Correlation of the abundance of mutant alleles with the pathological abnormalities in ovarian tissue from a young girl with MAS and precocious puberty. A cross-section from a paraffin-embedded section of ovary from a patient (patient 1 in [21]) with MAS is shown in the centre (×50). The two outlined areas, shown at ×120, were dissected and analysed independently; area A shows ovarian cortex containing primordial follicles, and area B shows follicular cyst lining containing stimulated luteinized theca. On the right are blots showing the results of allele-specific oligonucleotide (ASO) hybridization of DNA with wild-type (R201) or mutant (R201C) radioactively labelled primers after PCR amplification; DNA was isolated from total ovary (centre) or specific regions as shown.

The variable involvement of different tissues in patients with MAS, as well as the clinical heterogeneity among affected patients, is assumed to be a result of several unique features. First, the number of tissues in which the *gsp* is present, and the proportion and distribution of affected cells in a tissue, will be determined by the timing of the mutational event. Thus, mutations that arise early in embryogenesis are likely to affect several cell lineages and produce a more severe phenotype than mutational events that occur later. For example, acquisition of a *gsp* mutation months or even years after birth could explain the development of a solitary endocrine tumour or a single fibrous dysplasia lesion in some patients.

Second, epigenetic and/or microenvironmental factors that regulate *GNAS* expression can influence the MAS phenotype. For example, stochastic effects, such as allelic imbalance, may favour expression of the mutant allele in some tissues [26], thus exaggerating the effect of a *gsp* mutation. Even more importantly, tissue-specific imprinting of *GNAS* can exert a discrete effect on expression of *gsp* alleles. *GNAS* transcripts that encode $G\alpha_s$ are preferentially expressed from the maternal allele in some cells (e.g. renal proximal tubule cells, thyroid follicular cells, and pituitary somatotrophs) [27, 28]. Theoretically, in those cells in which $G\alpha_s$ is expressed predominately if not exclusively from the maternal allele, only somatic of the maternal allele will have pathophysiological consequences. This is the case for sporadic GH-secreting pituitary adenomas as well as patients with MAS who have GH-secreting pituitary adenomas, where activating mutations of $G\alpha_s$ almost always occur on the maternal allele [1, 29]. By contrast, the parental origin of a *gsp* allele will be far less important in cells and tissues where both $G\alpha_s$ alleles are expressed.

Additional transcripts are generated by *GNAS* using alternative first exons that are spliced to exons 2–13, but the effect of these proteins on the MAS phenotype remains uncertain. Exon 1A, is located approximately 2.5 kb upstream of exon 1. Transcripts beginning with exon 1A are expressed only from the paternal allele, and are probably untranslated [30]. Further upstream are two additional alternative first exons; one encodes the amino terminus of the $XL\alpha_s$ protein, which is expressed only from paternal alleles. $XL\alpha_s$ shares carboxyl terminal sequences with $G\alpha_s$, and functions in G-protein coupled signal transduction. *gsp* mutations in $XL\alpha_s$ can affect signal transduction *in vitro* [31] and recent data suggests that constitutive $XL\alpha_s$ activity adds to the molecular pathogenesis of MAS [32]. The third alternative first exon is approximately 52 kb upstream of exon 1 and is expressed exclusively from the maternal allele. This exon contains the entire coding sequence for the neurosecretory protein NESP55 [33–35], a chromogranin-like protein that is present in secretory granules and shares no protein homology with $G\alpha_s$. Thus, activating mutations in exon 8 of *GNAS* would not be present in NESP55.

Third, the clinical and endocrinological features of MAS will be influenced by the effects of cAMP in a specific cell type. For example, a *gsp* oncogene will produce the most significant consequences in those tissues in which cAMP stimulates cellular proliferation and/or hormone secretion rather than differentiation. Cyclic adenosine monophosphate (AMP) is not mitogenic in all cell types, and in some cell types cAMP can actually inhibit growth. Moreover, even in cells in which cAMP is a strong growth stimulator, changes in the expression of other genes or induction of counter-regulatory responses (such as increased cAMP phosphodiesterase activity) could mitigate or even reverse the effects of the *gsp* oncogene.

Clinical Manifestations of McCune–Albright Syndrome

Comprehensive reviews of the clinical spectrum of MAS have extended our appreciation of this unusual disorder [36, 37] (Figure 6.14.2). The mean age at the time of clinical diagnosis of MAS is 5.7 years, with a range of 0.7 to 11 years. Almost all patients who ultimately manifest the complete clinical triad of pigmented skin lesions, excessive endocrine function, and fibrous dysplasia will have evidence of café-au-lait skin lesions at birth. There is a 50% likelihood of precocious puberty in females by age 4 years, and a 50% likelihood of bone lesions by age 8 years.

Fibrous Dysplasia

Fibrous dysplasia of the skeleton occurs in nearly all (98%) patients with MAS, and may manifest as either a solitary lesion (mono-ostotic, 13%) or multiple lesions (poly-ostotic, 87%) [35].

Fibrous dysplasia typically develops during the first decade of life (Table 6.14.1), and seems to progress over time in most patients, with an increase in both the extent and number of bone lesions. The femur and pelvis are most commonly involved, and the shepherd's crook deformity of the femur is a pathognomonic lesion. Spinal involvement, with progressive scoliosis, is apparently more common than originally thought [38, 39]. Bisphosphonates have not been effective in decreasing the progression of spine curvature [40].

Most MAS patients will experience at least one fracture (peak age 7–12 years) and many patients will have multiple fractures. Radiographs of affected bones reveal long expansile (long lesions in long bones), lytic lesions with a 'ground glass' (as in the ground glass stoppers for glass bottles) pattern and a scalloped cortical bone border secondary to endosteal erosion (Figure 6.14.2) [38, 39, 41]. The marrow cavity, which usually has a cellular fatty tissue, is replaced by fibro-osseous tissue. Bone histology discloses three primary, but distinct, histological patterns, defined as Chinese writing type, sclerotic/Pagetoid type, and sclerotic/hypercellular type, which are characteristically associated with the axial/appendicular skeleton, cranial bones, or gnathic bones, respectively [42]. Craniofacial involvement occurs in many patients, and should be evaluated with both CT and MRI in order to demonstrate the extent of disease, and potential compressive complications of fibrous dysplasia such as optic neuropathy [43]. While hearing loss appears to be common [44, 45], ocular complications are infrequent even with involvement of the optic nerve canals [45].

The basis for the unusual cellular changes in fibrous dysplasia is poorly understood. Recent evidence indicates that the fibrotic areas consist of an excess of pre-osteogenic cells, whereas the bone formed *de novo* within fibrotic areas is produced by mature but abnormal osteoblasts [42]. It is likely that at least some of the phenotypic changes in affected osteogenic cells result from cAMP-induced increases in expression of interleukin-6 and the *c-fos* proto-oncogene [22, 46, 47]. The mosaic distribution of lesions in fibrous dysplasia may also play an important pathogenic role, as close contact between transplanted normal bone cells and osteogenic cells containing the *gsp* mutation is necessary to reproduce the fibrous dysplasia lesion in mice [48]. Using mouse models that allow the conditional expression of a *gsp* mutation in the skeletal stem cell (SSC) lineage, Zhao *et al.* have shown that *GNAS* activation is sufficient for initiation of fibrous dysplasia, as cells containing the mutation proliferated along the osteogenic lineage but failed to differentiate to mature osteoblasts. Remarkably, silencing the *gsp* mutation resulted in regression of the fibrous dysplasia-like lesion, suggesting that targeting *gsp* might represent a strategy for cure of fibrous dysplasia [49].

Sarcomatous degeneration (e.g. osteosarcoma, fibrosarcoma, and chondrosarcoma, in descending order of frequency) occurs as a rare complication of fibrous dysplasia in MAS patients (mean age of 36 years) [50], and F-18 fluorodeoxyglucose positron emission tomography may be able to disclose early malignant transformation of fibrous dysplasia lesions [43, 51].

Treatment for fibrous dysplasia is principally based on surgery, with fracture repair and reconstruction of craniofacial deformity the most common procedures. Regrowth or recurrence of fibrous dysplasia after skull surgery is common (66%), and is more frequent after operations in subjects with MAS who have growth hormone excess (88%) than without growth hormone excess (58%) [52]. Therefore, patients with extensive fibrous dysplasia should be evaluated carefully for the presence of MAS in general, and growth hormone excess in particular. By contrast, outcomes are favourable for aneurysmal bone cysts and biopsies [52].

No medical treatment for fibrous dysplasia is entirely satisfactory. Most but not all studies have demonstrated that bisphosphonates can relieve bone pain, decrease bone resorption, and improve the radiological appearance (e.g. filling of lytic lesions and/or thickening of cortices) of bone lesions in about 50% of patients. Bone mineral density in affected sites is also significantly increased after treatment with pamidronate, a potent second-generation bisphosphonate that is administered intravenously. In a series of nine patients on long-term pamidronate treatment who became resistant to this medication, a switch to intravenous zoledronic acid did not produce any substantial improvement [53]. Recent case reports describe early and encouraging experiences with denosumab as a treatment for fibrous dysplasia in adults [54–56]. Studies in children are even more limited. Denosumab treatment of one 9-year-old boy with extremely painful and severe fibrous dysplasia led to a clinically significant reduction of osteolytic expansion and fibrous dysplasia-related bone pain, but was associated with worrisome disturbances of mineral metabolism both while on treatment and after discontinuation, limiting enthusiasm for use of this drug in children [56].

Anecdotally, 30 days after intranasal calcitonin administration a wheelchair bound patient had improved pain and at 90 days was able to walk without assistance [57]. Therefore, calcitonin may be

considered a short-term therapeutic option in cases of severe and refractory bone pain [57]. Finally, in patients with a high skeletal burden in fibrous dysplasia, consistent with MAS, addressing maladaptive illness perceptions may also improve quality of life [58].

Café-au-Lait Skin Lesions

Patients with MAS typically have one or more pigmented macules, termed *café-au-lait* lesions, that have irregular borders (coast of Maine) (Figure 6.14.4). By contrast, *café-au-lait* skin lesions that occur in patients with neurofibromatosis (Von Recklinghausen's syndrome) have a smooth border (coast of California) (Figure 6.14.5). The distribution of skin lesions in MAS is also characteristic (Figure 6.14.4), consisting of an S-shaped pattern on the chest, a V-shaped pattern on the back, and a linear distribution on the extremities, that conforms to the embryological lines of ectodermal migration (i.e. lines of Blashko) and reflects the dorsoventral outgrowth of two populations of cells [19]. Lesions rarely extend beyond the midline and in most patients the skin lesions tend to be on the same side of the body as the skeletal lesions. They occur most commonly on the buttocks and lumbosacral regions.

Endocrine Abnormalities

Autonomous endocrine function is common in MAS (Table 6.14.1) [59]. Precocious puberty is the most common endocrine disorder in MAS, and has been reported in over 60% of patients. Precocious puberty is a common initial manifestation of MAS in girls, and characteristically presents as thelarche and/or vaginal bleeding in a girl under 5 years of age. Vaginal bleeding may occur in the absence of significant breast development or pubarche. Some young girls will have seemingly regular menses and progressive pubertal development, including rapid advancement of bone age, whereas others will have irregular or intermittent bleeding that is associated with relatively normal rates of growth. The production of oestrogen appears related to the growth and involution of small ovarian cysts, and is typically not associated with follicular maturation or ovulation. Ovarian activity can undergo a spontaneous remission in some cases. Large, benign ovarian cysts may also occur [13, 14], and surgical excision may result in regression of secondary sexual characteristics until the onset of normal pubertal development. Patients typically have low or suppressed levels of serum LH and follicle-stimulating hormone (FSH), which fail to increase significantly after administration of gonadotropin-releasing hormone (GnRH), a characteristic of gonadotropin-independent precocious puberty (i.e. precocious 'pseudopuberty'). Testing may be normal during intervals of apparent ovarian inactivity, however. Given the episodic nature of oestrogen production, and the poor performance characteristic of many clinical assays for oestradiol, serum concentrations of this steroid are often not elevated. Of interest, after several years of excessive sex steroid exposure some girls experience a transition to central precocious puberty, particularly those whose bone age is 11 years or greater [60–62]. As adults, women with a past history of gonadotropin-independent precocious puberty may have irregular menses and reduced fertility due to continued autonomous production of oestrogen.

Treatment of precocious puberty in girls with MAS is challenging. Therapy with GnRH analogues and superagonists is not effective unless there has been a progression to central precocious puberty [62]. Treatment with first generation aromatase inhibitors was generally disappointing [61, 63], but recent long-term follow-up studies of girls treated with letrozole have been more promising [64], with sustained beneficial effects on skeletal maturation, growth velocity, and predicted adult height [65]. The efficacy of compounds with antioestrogenic activity, such as the selective oestrogen receptor modulators (SERMs) tamoxifen or raloxifene, also appears to be useful [65–67].

Contrary to prevailing thinking, the incidence of gonadal pathology in MAS is equal in males and females. The predominant histopathological finding is variable degrees of seminiferous tube development with Leydig cell hyperplasia, which carries a low risk of malignant transformation and can be managed conservatively [68]. Testicular enlargement is generally bilateral but can be unilateral [69]. Although testicular enlargement is usually associated with excessive production of testosterone and precocious puberty, occasionally the enlargement is limited to autonomous hyperfunction of Sertoli cells with no activation of Leydig cells [24]. Most boys with MAS will show testicular microlithiasis on ultrasound, which can

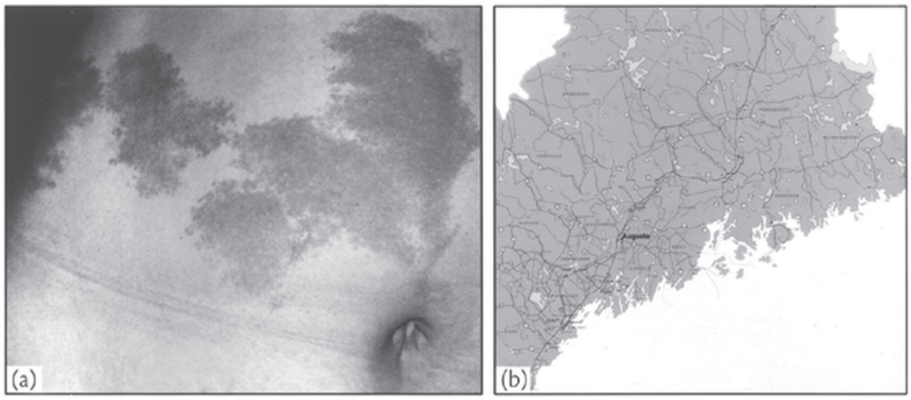


Figure 6.14.4 Café-au-lait lesions in McCune–Albright syndrome. The pigmented lesions follow the embryological lines of Blashko, and are typically ipsilateral to and near the skeletal lesions of fibrous dysplasia. The pigmented macules have irregular margins (left) that resemble the 'coast of Maine' (right).

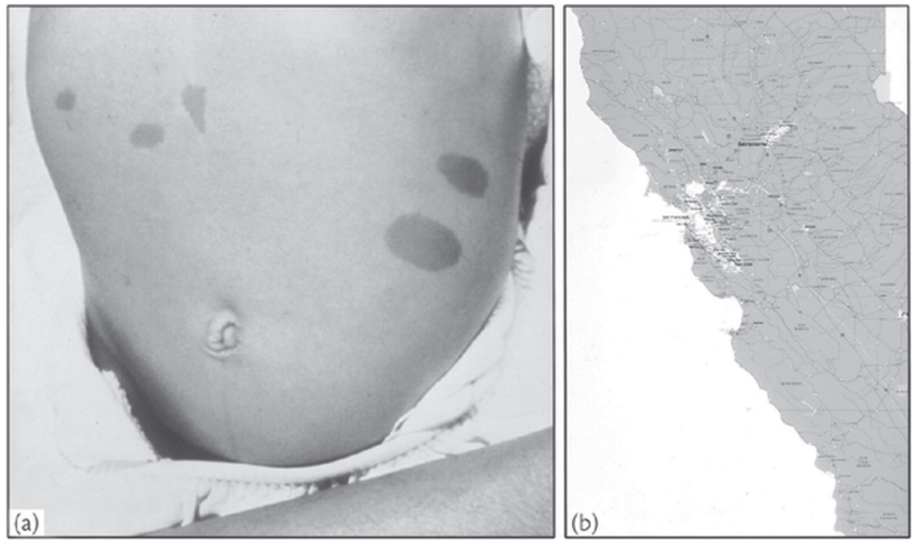


Figure 6.14.5 Café-au-lait lesions in neurofibromatosis. The pigmented macules have smooth margins (left) that resemble the 'coast of California' (right).

be a very useful screening test [68]. Treatment is similar to that for familial male precocious puberty due to activating mutations of the LH receptor (i.e. testitoxicosis) [70], and consists of the combination of an aromatase inhibitor plus an androgen receptor blocker [66, 68, 69, 71]. In those cases where gonadotropin-independent precocious puberty leads to early activation of central puberty, the addition of a GnRH analogue may be required to arrest further pubertal development and to allow boys to achieve a normal adult height [68, 72, 73].

Growth hormone (GH) excess is common in MAS [74], and may produce either gigantism or acromegaly [75, 76]. The biochemical behaviour of growth hormone-producing pituitary tumours in patients with MAS appears indistinguishable from that of sporadic tumours with and without *gsp* mutations. GH secretion is stimulated by TRH, GHRH, and sleep, and is incompletely suppressed by glucose administration. However, only 65% of MAS patients with GH excess have radiographic evidence of a pituitary tumour, a much lower incidence than in sporadic cases of acromegaly (99%) [77]. In addition, hyperprolactinemia occurs in over 50% of MAS patients with elevated GH levels, a frequency that is somewhat greater than occurs in patients with sporadic pituitary tumours (40%) [77]. The indications for surgical treatment (e.g. transphenoidal tumour excision) of GH-secreting pituitary tumours are similar for those in patients without MAS. Improved outcomes may be achieved with neuronavigational guidance and preoperative treatment with long-acting ostreotide analogues [78].

Medical therapy with bromocriptine may reduce tumour size and hormonal secretion in many, but not all patients [10, 79]. However, other medical treatments, such as long-acting octreotide and pegvisomant [80–82], appear more useful. Insulin growth factor 1 (IGF1) hyperactivity increases risk of morbidity in MAS presumably thru increased GH. Early medical therapy, for example, with long-acting octreotide (with or without pegvisomant) therapy is effective in normalizing IGF1 in most patients, and early treatment is associated with decreased risk of optic neuropathy and decreased growth of GH-secreting adenomas [83].

Hyperthyroidism and/or autonomous thyroid nodules have been identified in approximately 33% of MAS patients who underwent thyroid evaluation [77, 84–86]. Radioactive iodine ablation or surgery has been used to treat thyroid nodules. The degree of hyperthyroidism is variable, and serum concentrations of thyroid-stimulating hormone (TSH) are typically low and thyroid-stimulating immunoglobulins are undetectable. The thyroid gland will often appear normal by physical exam, but nodules are nearly always detectable by sonography.

Patients with MAS occasionally develop autonomous function of the adrenal gland and primary hypercortisolism at a young age (mean age of 4.4 years) [77]. Adrenal gland histopathology reveals either nodular hyperplasia or solitary adenoma [76].

Other Features

Recent analyses have documented the occurrence of additional non-endocrine features in patients with MAS that extend the clinical spectrum of the disorder. These include hypophosphataemia, hepatobiliary disease, and cardiac disease. Hypophosphataemia and/or decreased renal tubular reabsorption of phosphate occur in over 50% of subjects with MAS, and may lead to the development of rickets or osteomalacia [87]. A similar syndrome of hypophosphataemic rickets has been described in patients with fibrous dysplasia who lack other features of MAS, as well as in other patients who have various mesenchymal tumours, and appears due to secretion of circulating phosphaturic factors termed 'phosphatonins'. FGF23 is the best characterized of the phosphatonins, and is produced by the abnormal osteogenic precursors present in fibrous dysplasia lesions. The concentration of circulating FGF23 correlates with the extent of fibrous dysplasia throughout the skeleton [88, 89]. Based on the success of a fully humanized monoclonal antibody against FGF23, burosumab, for the treatment of adults and children with X-linked hypophosphataemic rickets [90, 91] and tumour-induced osteomalacia [92, 93], burosumab may emerge as a successful medical

therapy for the FGF23-dependent hypophosphataemic rickets and/or osteomalacia in patients with fibrous dysplasia [93, 94]. An alternative explanation for hypophosphataemia in patients with MAS is the presence of the *gsp* oncogene in the proximal renal tubule, where it induces increased cAMP production and an intrinsic defect in reabsorption of phosphate [95].

While neonatal jaundice in patients with MAS typical resolves, liver function enzymes typically remain mildly elevated. Liver histology varies from near normal to discrete portal fibrosis to giant cell hepatitis [96]. Liver disease is due to the presence of the *gsp* mutation in hepatic tissue [86, 96, 97], and the degree of histological abnormality correlates with the relative amount of abnormal Gαs protein and adenylyl cyclase activation [97]. Another unusual manifestation of MAS is cardiac disease [86]. Cardiac involvement in patients with MAS ranges from tachycardia and hypertension to cardiomegaly and sudden death [86]. Affected cardiac tissue contains cells with the *gsp* mutation [86], and it is likely that elevated levels of cAMP account directly for the abnormal cardiac function.

There are abnormalities in the gastrointestinal tract and pancreas in patients with MAS and patients with MAS should be evaluated for gastrointestinal pathology as well [98]. Intraductal papillary mucinous neoplasms (IPMNs), which are the most common cystic precursor lesions of invasive pancreatic cancer, *GNAS* mutations at codon 201 have been identified. Histological subtype of IPMNs correlates with *GNAS* mutational frequency [99].

Thus, the clinical MAS phenotype is highly variable, depending upon the location and timing of the mutation during embryologic development.

Diagnosis

The diagnosis of MAS remains a clinical exercise, and is straightforward when all three features are present in a candidate patient. However, many patients with MAS lack some features at the time of initial presentation, and it would therefore be desirable to have a molecular test for the disorder. The mosaic distribution of cells bearing the *GNAS* mutation, and the variable number of affected cells in a tissue, makes it technically difficult to detect mutant *GNAS* alleles even in affected tissues, as they may represent only a small proportion of the *GNAS* alleles present in a DNA sample. Detection of a *gsp* mutant in DNA samples can be greatly enhanced by protocols that enrich the relative abundance of mutant alleles as polymerase chain reaction (PCR) targets and thereby facilitate selective amplification. These techniques have relied upon either multiple rounds of PCR and restriction endonuclease digestion of wild-type amplicons [100] or inclusion of a peptide nucleic acid (PNA) in the PCR to block amplification of wild-type *GNAS* targets [101, 102]. The sensitivity of nested PCR and PNA clamping appears comparable, but the nested PCR method requires more time and expense than PNA clamping [37]. A recent improvement over standard PNA clamping uses a labelled PNA hybridization probe and fluorescence resonance energy transfer (FRET) to allow for the direct and rapid quantification of *gsp* alleles with a sensitivity that allows detection in tissues that contain as few as 5% mutant cells [103]. While analysis of DNA from lesional tissue affords greatest sensitivity, it is neither practical nor expedient to biopsy affected tissue(s) in all patients. Both nested PCR and PNA clamping have been used to detect *gsp* mutations in peripheral blood samples [37, 104]. Next generation sequencing can analyse millions of PCR amplicons. When PNA clamping and next generation sequencing are combined, the Arg201 mutation can be detected in virtually all affected tissues and in leukocytes of up to 75% of individuals [105], making this the most the most sensitive detection method.

The detection of a *gsp* mutation in circulating cells from a patient with FD or an isolated endocrinopathy (e.g. growth hormone-producing pituitary tumour, ovarian cysts) does not necessarily imply that the patient has MAS, however. Even with molecular demonstration of a *gsp* mutation, additional studies and clinical interpretation will be needed to distinguish between MAS and an isolated lesion.

On the other hand, identification of a *gsp* mutation can distinguish between fibrous dysplasia and similar lesion such as osteofibrous dysplasia [106], and may assist in distinguishing between atypical forms of MAS and Carney complex [107–109] or Mazabraud syndrome [110, 111].

Treatment

Each specific feature of MAS requires intensive management that follows the usual recommendations for an isolated lesion, with occasional special caveats based on the potential confounding interactions of other endocrine defects. The management of the various clinical characteristics of MAS is previously described in each preceding section.

Conclusion

The diagnosis of MAS remains a clinical one, and requires a careful integration of physical findings, biochemical evaluation, and radiological examination. The disorder can present as a form fruste, and identification of a specific *GNAS* mutation in DNA from affected tissues and in many cases peripheral blood cells can confirm a clinical diagnosis of MAS. Finally, the genetic basis for MAS, mosaicism of a somatic *gsp* mutation, provides new insights into the role of imprinting as a modulator of human disease.

ACKNOWLEDGEMENTS

This work was supported in part by United States Public Health Service Grant R01 DK34281 from the NIDDK and The Children's Hospital of Philadelphia Research Institute.

REFERENCES

1. Hayward BE, Barlier A, Korbonits M, *et al.* Imprinting of the G(s)alpha gene GNAS1 in the pathogenesis of acromegaly. *J Clin Invest*, 2001; **107**: R31–R6.
2. Haldeman-Englert CR, Hurst ACE, Levine MA. Disorders of *GNAS* inactivation. In: Adam MP, Ardinger HH, Pagon RA, *et al.*, eds. GeneReviews [Internet]. Seattle, WA: University of Washington, 2017.

3. Clapham DE, Neer EJ. G protein beta gamma subunits. *Annu Rev Pharmacol Toxicol*, 1997; **37**: 167–203.
4. Kahn RA, Gilman AG. ADP-ribosylation of Gs promotes the dissociation of its alpha and beta subunits. *J Biol Chem*, 1984; **259**: 6235–40.
5. Bhattacharyya N, Hu X, Chen CZ, *et al*. A high throughput screening assay system for the identification of small molecule inhibitors of gsp. *PloS One*, 2014; **9**: e90766.
6. Vallar L, Spada A, Giannattasio G. Altered Gs and adenylate cyclase activity in human GH-secreting pituitary adenomas. *Nature*, 1987; **330**: 566–8.
7. Conti CJ. Mutations of genes of the ras family in human and experimental tumors. *Prog Clin Biol Res*, 1992; **376**: 357–78.
8. Landis CA, Masters SB, Spada A, Pace AM, Bourne HR, Vallar L. GTPase inhibiting mutations activate the alpha chain of Gs and stimulate adenylyl cyclase in human pituitary tumours. *Nature*, 1989; **340**: 692–6.
9. Lyons J, Landis CA, Griffith H, *et al*. Two G protein oncogenes in human endocrine tumors. *Science*, 1990; **249**: 655–9.
10. Barlier A, Gunz G, Zamora AJ, *et al*. Pronostic and therapeutic consequences of Gs alpha mutations in somatotroph adenomas. *J Clin Endocrinol Metab*, 1998; **83**: 1604–10.
11. Spada A, Lania A, Ballare E. G protein abnormalities in pituitary adenomas. *Mol Cell Endocrinol*, 1998; **142**: 1–14.
12. Fragoso MC, Latronico AC, Carvalho FM, *et al*. Activating mutation of the stimulatory G protein (gsp) as a putative cause of ovarian and testicular human stromal Leydig cell tumors. *J Clin Endocrinol Metab*, 1998; **83**: 2074–8.
13. Pienkowski C, Lumbroso S, Bieth E, Sultan C, Rochiccioli P, Tauber M. Recurrent ovarian cyst and mutation of the Gs alpha gene in ovarian cyst fluid cells: what is the link with McCune-Albright syndrome? *Acta Paediatr*, 1997; **86**: 1019–21.
14. Rodriguez-Macias KA, Thibaud E, Houang M, Duflos C, Beldjord C, Rappaport R. Follow up of precocious pseudopuberty associated with isolated ovarian follicular cysts. *Arch Dis Child*, 1999; **81**: 53–6.
15. Okamoto S, Hisaoka M, Ushijima M, Nakahara S, Toyoshima S, Hashimoto H. Activating Gs(alpha) mutation in intramuscular myxomas with and without fibrous dysplasia of bone. *Virchows Arch*, 2000; **437**: 133–7.
16. Alman BA, Greel DA, Wolfe HJ. Activating mutations of Gs protein in monostotic fibrous lesions of bone. *J Orthop Res*, 1996; 14(2): 311–15.
17. McCune DJ, Bruch H. Osteodystrophia fibrosa. *Am J Dis Child*, 1937; **54**: 806–48.
18. Albright F, Butler AM, Hampton AO, Smith P. Syndrome characterized by osteitis fibrosa disseminata, areas of pigmentation and endocrine dysfunction, with precocious puberty in females. *N Engl J Med*, 1937; **216**: 727–41.
19. Happle R. The McCune–Albright syndrome: a lethal gene surviving by mosaicism. *Clin Genet*, 1986; **29**: 321–4.
20. Schwindinger WF, Francomano CA, Levine MA. Identification of a mutation in the gene encoding the alpha subunit of the stimulatory G protein of adenylyl cyclase in McCune-Albright syndrome. *Proc Natl Acad Sci U S A*, 1992; **89**: 5152–6.
21. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating mutations of the stimulatory G protein in the McCune-Albright syndrome. *N Engl J Med*, 1991; **325**: 1688–95.
22. Candeliere GA, Glorieux FH, Prud'homme J, St-Arnaud R. Increased expression of the c-fos proto-oncogene in bone from patients with fibrous dysplasia. *N Engl J Med*, 1995; **332**: 1546–51.
23. Gorelov VN, Gyenes M, Neser F, Roher HD, Goretzki PE. Distribution of Gs-alpha activating mutations in human thyroid tumors measured by subcloning. *J Cancer Res Clin Oncol*, 1996; **122**: 453–7.
24. Coutant R, Lumbroso S, Rey R, *et al*. Macroorchidism due to autonomous hyperfunction of Sertoli cells and G(s)alpha gene mutation: an unusual expression of McCune-Albright syndrome in a prepubertal boy. *J Clin Endocrinol Metab*, 2001; **86**: 1778–81.
25. Rey RA, Venara M, Coutant R, *et al*. Unexpected mosaicism of R201H-GNAS1 mutant-bearing cells in the testes underlie macro-orchidism without sexual precocity in McCune-Albright syndrome. *Hum Mol Genet*, 2006; **15**: 3538–43.
26. Michienzi S, Cherman N, Holmbeck K, *et al*. GNAS transcripts in skeletal progenitors: evidence for random asymmetric allelic expression of Gs alpha. *Hum Mol Genet*, 2007; **16**: 1921–30.
27. Germain-Lee EL, Ding CL, Deng Z, *et al*. Paternal imprinting of Galpha(s) in the human thyroid as the basis of TSH resistance in pseudohypoparathyroidism type 1a. *Biochem Biophys Res Commun*, 2002; **296**: 62–72.
28. Liu J, Erlichman B, Weinstein LS. The stimulatory G protein alpha-subunit Gs alpha is imprinted in human thyroid glands: implications for thyroid function in pseudohypoparathyroidism types 1A and 1B. *J Clin Endocrinol Metab*, 2003; **88**: 4336–41.
29. Mantovani G, Bondioni S, Locatelli M, *et al*. Biallelic expression of the Gsalpha gene in human bone and adipose tissue. *J Clin Endocrinol Metab*, 2004; **89**: 6316–19.
30. Swaroop A, Agarwal N, Gruen JR, Bick D, Weissman SM. Differential expression of novel Gsα signal transduction protein cDNA species. *Nucleic Acids Res*, 1991; **17**: 4725–9.
31. Linglart A, Mahon MJ, Kerachian MA, *et al*. Coding GNAS mutations leading to hormone resistance impair *in vitro* agonist- and cholera toxin-induced adenosine cyclic 3',5'-monophosphate formation mediated by human XLalphas. *Endocrinology*, 2006; **147**: 2253–62.
32. Mariot V, Wu JY, Aydin C, *et al*. Potent constitutive cyclic AMP-generating activity of XLalphas implicates this imprinted GNAS product in the pathogenesis of McCune-Albright syndrome and fibrous dysplasia of bone. *Bone*, 2011; **48**: 312–20.
33. Hayward BE, Kamiya M, Strain L, *et al*. The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc Natl Acad Sci U S A*, 1998; **95**: 10038–43.
34. Hayward BE, Moran V, Strain L, Bonthron DT. Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc Natl Acad Sci U S A*, 1998; **95**: 15475–80.
35. Benhamou J, Gensburger D, Messiaen C, Chapurlat R. Prognostic factors from an epidemiologic evaluation of fibrous dysplasia of bone in a modern cohort: the FRANCEDYS study. *J Bone Miner Res*, 2016; **31**: 2167–72.
36. Robinson C, Collins MT, Boyce AM. Fibrous dysplasia/McCune-Albright syndrome: clinical and translational perspectives. *Curr Osteoporos Rep*, 2016; **14**: 178–86.
37. Lietman SA, Schwindinger WF, Levine MA. Genetic and molecular aspects of McCune-Albright syndrome. *Pediatr Endocrinol Rev*, 2007; **4** Suppl 4: 380–5.
38. Leet AI, Magur E, Lee JS, Wientroub S, Robey PG, Collins MT. Fibrous dysplasia in the spine: prevalence of lesions and association with scoliosis. *J Bone Joint Surg Am*, 2004; **86-A**: 531–7.

39. Mancini F, Corsi A, De Maio F, Riminucci M, Ippolito E. Scoliosis and spine involvement in fibrous dysplasia of bone. *Eur Spine J*, 2009; **18**: 196–202.
40. Berglund JA, Tella SH, Tuthill KF, *et al*. Scoliosis in fibrous dysplasia/McCune-albright syndrome: factors associated with curve progression and effects of bisphosphonates. *J Bone Miner Res*, 2018; **33**: 1641–8.
41. Leet AI, Chebli C, Kushner H, *et al*. Fracture incidence in polyostotic fibrous dysplasia and the McCune-Albright syndrome. *J Bone Miner Res*, 2004; **19**: 571–7.
42. Riminucci M, Robey PG, Bianco P. The pathology of fibrous dysplasia and the McCune-Albright syndrome. *Pediatr Endocrinol Rev*, 2007; **4** Suppl 4: 401–11.
43. Bulakbasi N, Bozlar U, Karademir I, Kocaoglu M, Somuncu I. CT and MRI in the evaluation of craniospinal involvement with polyostotic fibrous dysplasia in McCune-Albright syndrome. *Diagn Interv Radiol*, 2008; **14**: 177–81.
44. Boyce AM, Brewer C, DeKlotz TR, *et al*. Association of hearing loss and otologic outcomes with fibrous dysplasia. *JAMA Otolaryngol Head Neck Surg*, 2018; **144**: 102–7.
45. Boyce AM, Glover M, Kelly MH, *et al*. Optic neuropathy in McCune-Albright syndrome: effects of early diagnosis and treatment of growth hormone excess. *J Clin Endocrinol Metab*, 2013; **98**: E126–34.
46. Riminucci M, Fisher LW, Shenker A, Spiegel AM, Bianco P, Gehron RP. Fibrous dysplasia of bone in the McCune-Albright syndrome: abnormalities in bone formation. *Am J Pathol*, 1997; **151**: 1587–600.
47. Shenker A, Weinstein LS, Sweet DE, Spiegel AM. An activating Gsα mutation is present in fibrous dysplasia of bone in McCune-Albright syndrome. *J Clin Endocrinol Metab*, 1994; **79**: 750–5.
48. Bianco P, Kuznetsov SA, Riminucci M, Fisher LW, Spiegel AM, Robey PG. Reproduction of human fibrous dysplasia of bone in immunocompromised mice by transplanted mosaics of normal and Gsalpha-mutated skeletal progenitor cells. *J Clin Invest*, 1998; **101**: 1737–44.
49. Zhao X, Deng P, Iglesias-Bartolome R, *et al*. Expression of an active Galphas mutant in skeletal stem cells is sufficient and necessary for fibrous dysplasia initiation and maintenance. *Proc Natl Acad Sci U S A*, 2018; **115**: E428–E37.
50. Ruggieri P, Sim FH, Bond JR, Unni KK. Osteosarcoma in a patient with polyostotic fibrous dysplasia and Albright's syndrome. *Orthopedics*, 1995; **18**: 71–5.
51. Berrebi O, Steiner C, Keller A, Rougemont AL, Ratib O. F-18 fluorodeoxyglucose (FDG) PET in the diagnosis of malignant transformation of fibrous dysplasia in the pelvic bones. *Clin Nucl Med*, 2008; **33**: 469–71.
52. Boyce AM, Burke A, Cutler Peck C, DuFresne CR, Lee JS, Collins MT. Surgical management of polyostotic craniofacial fibrous dysplasia: long-term outcomes and predictors for postoperative regrowth. *Plast Reconstruct Surg*, 2016; **137**: 1833–9.
53. Chapurlat RD. Medical therapy in adults with fibrous dysplasia of bone. *J Bone Miner Res*, 2006; **21** Suppl 2: 114–19.
54. Ganda K, Seibel MJ. Rapid biochemical response to denosumab in fibrous dysplasia of bone: report of two cases. *Osteoporosis Int*, 2014; **25**: 777–82.
55. Eller-Vainicher C, Rossi DS, Guglielmi G, *et al*. Prompt clinical and biochemical response to denosumab in a young adult patient with craniofacial fibrous dysplasia. *Clin Cases Miner Bone Metab*, 2016; **13**: 253–6.
56. Boyce AM, Chong WH, Yao J, *et al*. Denosumab treatment for fibrous dysplasia. *J Bone Miner Res*, 2012; **27**: 1462–70.
57. Fighera TM, Spritzer PM. Effect of intranasal calcitonin in a patient with McCune-Albright syndrome, fibrous dysplasia, and refractory bone pain. *Case Rep Endocrinol*, 2017; **2017**: 7898713.
58. Majoor BCJ, Andela CD, Quispel CR, *et al*. Illness perceptions are associated with quality of life in patients with fibrous dysplasia. *Calcified Tissue Int*, 2018; **102**: 23–31.
59. Collins MT, Singer FR, Eugster E. McCune-Albright syndrome and the extraskeletal manifestations of fibrous dysplasia. *Orphanet J Rare Dis*, 2012; **7** Suppl 1: S4.
60. Boepple PA, Frisch LS, Wierman ME, Hoffman WH, Crowley WF, Jr. The natural history of autonomous gonadal function, adrenarche, and central puberty in gonadotropin-independent precocious puberty. *J Clin Endocrinol Metab*, 1992; **75**: 1550–5.
61. Feuillan PP, Jones J, Cutler GB, Jr. Long-term testolactone therapy for precocious puberty in girls with the McCune-Albright syndrome. *J Clin Endocrinol Metab*, 1993; **77**: 647–51.
62. Schmidt H, Kiess W. Secondary central precocious puberty in a girl with McCune-Albright syndrome responds to treatment with GnRH analogue. *J Pediatr Endocrinol Metab*, 1998; **11**: 77–81.
63. Mieszczak J, Lowe ES, Plourde P, Eugster EA. The aromatase inhibitor anastrozole is ineffective in the treatment of precocious puberty in girls with McCune-Albright syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 2751–4.
64. Feuillan P, Calis K, Hill S, Shawker T, Robey PG, Collins MT. Letrozole treatment of precocious puberty in girls with the McCune-Albright syndrome: a pilot study. *J Clin Endocrinol Metab*, 2007; **92**: 2100–6.
65. Estrada A, Boyce AM, Brillante BA, Guthrie LC, Gafni RI, Collins MT. Long-term outcomes of letrozole treatment for precocious puberty in girls with McCune-Albright syndrome. *Eur J Endocrinol*, 2016; **175**: 477–83.
66. Neyman A, Eugster EA. Treatment of girls and boys with McCune-Albright syndrome with precocious puberty—update 2017. *Pediatr Endocrinol Rev*, 2017; **15**: 136–41.
67. Eugster EA, Rubin SD, Reiter EO, *et al*. Tamoxifen treatment for precocious puberty in McCune-Albright syndrome: a multicenter trial. *J Pediatr*, 2003; **143**: 60–6.
68. Boyce AM, Chong WH, Shawker TH, *et al*. Characterization and management of testicular pathology in McCune-Albright syndrome. *J Clin Endocrinol Metab*, 2012; **97**: E1782–90.
69. Arrigo T, Pirazzoli P, de Sanctis L, *et al*. McCune-Albright syndrome in a boy may present with a monolateral macroorchidism as an early and isolated clinical manifestation. *Horm Res*, 2006; **65**: 114–19.
70. DiMeglio LA, Pescovitz OH. Disorders of puberty: inactivating and activating molecular mutations. *J Pediatr*, 1997; **131**(1: Pt 2): S8–12.
71. Mieszczak J, Eugster EA. Treatment of precocious puberty in McCune-Albright syndrome. *Pediatr Endocrinol Rev*, 2007; **4** Suppl 4: 419–22.
72. Leschek EW, Cutler GB, Jr. Familial male precocious puberty. *Curr Ther Endocrinol Metab*, 1997; **6**: 343–5.
73. Leschek EW, Jones J, Barnes KM, Hill SC, Cutler GB, Jr. Six-year results of spironolactone and testolactone treatment of familial male-limited precocious puberty with addition of deslorelin after central puberty onset. *J Clin Endocrinol Metab*, 1999; **84**: 175–8.
74. Salenave S, Boyce AM, Collins MT, Chanson P. Acromegaly and McCune-Albright syndrome. *J Clin Endocrinol Metab*, 2014; **99**: 1955–69.

75. Lee PA, Van Dop C, Migeon CJ. McCune-Albright syndrome. Long-term follow-up. *JAMA*, 1986; **256**: 2980–4.
76. Mauras N, Blizzard RM. The McCune-Albright syndrome. *Acta Endocrinol Suppl (Copenh)*, 1986; **279**: 207–17.
77. Ringel MD, Schwindinger WF, Levine MA. Clinical implications of genetic defects in G proteins. The molecular basis of McCune-Albright syndrome and Albright hereditary osteodystrophy. *Medicine*, 1996; **75**: 171–84.
78. Yao Y, Liu Y, Wang L, *et al*. Clinical characteristics and management of growth hormone excess in patients with McCune-Albright syndrome. *Eur J Endocrinol*, 2017; **176**: 295–303.
79. Barlier A, Pellegrini-Bouiller I, Gunz G, Zamora AJ, Jaquet P, Enjalbert A. Impact of gsp oncogene on the expression of genes coding for Gsalpha, Pit-1, Gi2alpha, and somatostatin receptor 2 in human somatotroph adenomas: involvement in octreotide sensitivity. *J Clin Endocrinol Metab*, 1999; **84**: 2759–65.
80. Akintoye SO, Kelly MH, Brillante B, *et al*. Pegvisomant for the treatment of gsp-mediated growth hormone excess in patients with McCune-Albright syndrome. *J Clin Endocrinol Metab*, 2006; **91**: 2960–6.
81. Tajima T, Tsubaki J, Ishizu K, Jo W, Ishi N, Fujieda K. Case study of a 15-year-old boy with McCune-Albright syndrome combined with pituitary gigantism: effect of octreotide-long acting release (LAR) and cabergoline therapy. *Endocr J*, 2008; **55**: 595–9.
82. Almeida JP, Albuquerque LA, Ferraz CL, Mota I, Gondim J, Ferraz TM. McCune-Albright syndrome and acromegaly: hormonal control with use of cabergoline and long-acting somatostatin—case report. *Arq Bras Endocrinol Metabol*, 2009; **53**: 102–6.
83. Tessaris D, Boyce AM, Zacharin M, *et al*. Growth hormone (GH)—insulin like growth factor 1 (IGF-1) axis hyperactivity on bone fibrous dysplasia in McCune-Albright Syndrome. *Clin Endocrinol*, 2018; **89**: 56–64.
84. de Sanctis C, Lala R, Matarazzo P, *et al*. McCune-Albright syndrome: a longitudinal clinical study of 32 patients. *J Pediatr Endocrinol Metab*, 1999; **12**: 817–26.
85. Mastorakos G, Mitsiades NS, Doufas AG, Koutras DA. Hyperthyroidism in McCune-Albright syndrome with a review of thyroid abnormalities sixty years after the first report. *Thyroid*, 1997; **7**: 433–9.
86. Shenker A, Weinstein LS, Moran A, *et al*. Severe endocrine and nonendocrine manifestations of the McCune-Albright syndrome associated with activating mutations of stimulatory G protein GS. *J Pediatr*, 1993; **123**: 509–18.
87. Lala R, Matarazzo P, Andreo M, Defilippi C, de Sanctis C. Impact of endocrine hyperfunction and phosphate wasting on bone in McCune-Albright syndrome. *J Pediatr Endocrinol Metab*, 2002; **15** Suppl 3: 913–20.
88. Imel EA, Hui SL, Econs MJ. FGF23 concentrations vary with disease status in autosomal dominant hypophosphatemic rickets. *J Bone Miner Res*, 2007; **22**: 520–6.
89. Riminucci M, Collins MT, Fedarko NS, *et al*. FGF-23 in fibrous dysplasia of bone and its relationship to renal phosphate wasting. *J Clin Invest*, 2003; **112**: 683–92.
90. Carpenter TO, Imel EA, Ruppe MD, *et al*. Randomized trial of the anti-FGF23 antibody KRN23 in X-linked hypophosphatemia. *J Clin Invest*, 2014; **124**: 1587–97.
91. Imel EA, Zhang X, Ruppe MD, *et al*. Prolonged correction of serum phosphorus in adults with X-linked hypophosphatemia using monthly doses of KRN23. *J Clin Endocrinol Metab*, 2015; **100**: 2565–73.
92. Minisola S, Peacock M, Fukumoto S, *et al*. Tumour-induced osteomalacia. *Nat Rev Dis Primers*, 2017; **3**: 17044.
93. Kinoshita Y, Fukumoto S. X-linked hypophosphatemia and FGF23-related hypophosphatemic diseases—prospect for new treatment. *Endocr Rev*, 2018; **39**: 274–91.
94. Kutilek S. Burosumab: a new drug to treat hypophosphatemic rickets. *Sudan J Paediatr*, 2017; **17**: 71–3.
95. Zung A, Chalew SA, Schwindinger WF, *et al*. Urinary cyclic adenosine 3',5'-monophosphate response in McCune-Albright syndrome: clinical evidence for altered renal adenylate cyclase activity. *J Clin Endocrinol Metab*, 1995; **80**: 3576–81.
96. Silva ES, Lumbroso S, Medina M, Gillerot Y, Sultan C, Sokal EM. Demonstration of McCune-Albright mutations in the liver of children with high gamma GT progressive cholestasis. *J Hepatol*, 2000; **32**: 154–8.
97. Schwindinger WF, Yang SQ, Miskovsky EP, Diehl AM, Levine MA. An activating Gsα mutation in McCune-Albright syndrome increases hepatic adenylyl cyclase activity. *The Endocrine Society Program and Abstracts*, 1993; **75**: 517.
98. Wood LD, Noe M, Hackeng W, *et al*. Patients with McCune-Albright syndrome have a broad spectrum of abnormalities in the gastrointestinal tract and pancreas. *Virchows Archiv*, 2017; **470**: 391–400.
99. Molin MD, Matthaei H, Wu J, *et al*. Clinicopathological correlates of activating GNAS mutations in intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Ann Surg Oncol*, 2013; **20**: 3802–8.
100. Candeliere GA, Roughley PJ, Glorieux FH. Polymerase chain reaction-based technique for the selective enrichment and analysis of mosaic arg201 mutations in G alpha s from patients with fibrous dysplasia of bone. *Bone*, 1997; **21**: 201–6.
101. Bianco P, Riminucci M, Majolagbe A, *et al*. Mutations of the *GNAS1* gene, stromal cell dysfunction, and osteomalacic changes in non-McCune-Albright fibrous dysplasia of bone. *J Bone Miner Res*, 2000; **15**: 120–8.
102. Lietman SA, Ding C, Levine MA. A highly sensitive polymerase chain reaction method detects activating mutations of the GNAS gene in peripheral blood cells in McCune-Albright syndrome or isolated fibrous dysplasia. *J Bone Joint Surg Am*, 2005; **87**: 2489–94.
103. Karadag A, Riminucci M, Bianco P, *et al*. A novel technique based on a PNA hybridization probe and FRET principle for quantification of mutant genotype in fibrous dysplasia/McCune-Albright syndrome. *Nucleic Acids Res*, 2004; **32**: e63.
104. Kalfa N, Philibert P, Audran F, *et al*. Searching for somatic mutations in McCune-Albright syndrome: a comparative study of the peptidic nucleic acid versus the nested PCR method based on 148 DNA samples. *Eur J Endocrinol*, 2006; **155**: 839–43.
105. Narumi S, Matsuo K, Ishii T, Tanahashi Y, Hasegawa T. Quantitative and sensitive detection of GNAS mutations causing McCune-Albright syndrome with next generation sequencing. *PloS One*, 2013; **8**: e60525.
106. Sakamoto A, Oda Y, Iwamoto Y, Tsuneyoshi M. A comparative study of fibrous dysplasia and osteofibrous dysplasia with regard to Gsalpha mutation at the Arg201 codon: polymerase chain reaction-restriction fragment length polymorphism analysis of paraffin-embedded tissues. *J Mol Diagn*, 2000; **2**: 67–72.
107. DeMarco L, Stratakis CA, Boson WL, *et al*. Sporadic cardiac myxomas and tumors from patients with Carney complex are

not associated with activating mutations of the Gs alpha gene. *Hum Genet*, 1996; **98**: 185–8.
108. Libe R, Mantovani G, Bondioni S, *et al*. Mutational analysis of PRKAR1A and Gs(alpha) in sporadic adrenocortical tumors. *Exp Clin Endocrinol Diabetes*, 2005; **113**: 248–51.
109. Mantovani G, Corbetta S, Bondioni S, *et al*. Analysis of GNAS and PRKAR1A gene mutations in human cardiac myxomas not associated with multiple endocrine disorders. *J Endocrinol Invest*, 2009; **32**: 501–4.
110. Tagliafico A, Succio G, Martinoli C, Serafini G. Clinical overlap between Mazabraud and McCune-Albright syndromes. *J Ultrasound Med*, 2009; **28**: 397–9.
111. Zoccali C, Teori G, Prencipe U, Erba F. Mazabraud's syndrome: a new case and review of the literature. *Int Orthop*, 2009; **33**: 605–10.
112. Robinson C, Boyce AM, Estrada A, *et al*. Bone marrow failure and extramedullary hematopoiesis in McCune-Albright syndrome. *Osteoporosis Int*, 2018; **29**: 237–41.

6.15

Cowden Syndrome

Lamis Yehia, Shreya Malhotra, and Charis Eng

Introduction

Cowden syndrome (CS; OMIM 158350), named after the first-recognized patient Rachel Cowden, is an autosomal dominant inherited cancer predisposition syndrome characterized by multiple hamartomas involving organ systems derived from all three germ cell layers and a risk of breast, thyroid and other cancers [1–4]. Endocrinologists may make the diagnosis of CS when they are presented with these patients' endocrine lesions, chief of which are multinodular goitre, thyroid adenomas, and epithelial thyroid cancer. The first identified CS susceptibility gene, *PTEN*, is located on chromosome sub-band 10q23.3 [5, 6]. Because of the variable, protean, and often subtle external manifestations of CS, many cases remain undiagnosed. It is estimated that CS has an incidence of 1:200 000 individuals, although this is likely an underestimate [4].

Aetiology and Pathogenesis

Genetics

Inheritance patterns in families with CS implicate an autosomal dominant pattern. While expressivity is variable, CS is a highly penetrant disorder. Based on the earliest population-based clinical epidemiology study, some believe that the penetrance is 90% after the age of 20 years [5]. The precise penetrance was clarified after further study of the susceptibility gene within families and affected individuals. CS was mapped to 10q22–23, without genetic heterogeneity [5]. Further germline and somatic genetic analysis helped place the candidate predisposition gene between markers D10S215 and D10S541, a region of less than 1 cM [5].

PTEN

A candidate tumour suppressor gene *PTEN/MMAC1/TEP1* was located precisely in the CS critical interval [5, 7]. The gene comprises 1209 coding base pairs encoding a 403-amino acid protein. The protein has a tyrosine phosphatase domain and homology to tensin and auxilin [7, 8]. Hence, this new gene was dubbed *PTEN* for phosphatase and tensin homologue deleted on chromosome ten. *In vitro, PTEN* acts as a dual specificity phosphatase, meaning it is both a lipid phosphatase and a protein phosphatase as well as a tyrosine phosphatase and serine–threonine phosphatase [9, 10]. Subsequent *in vivo* work in mouse models suggested that PTEN plays a role in the PI3Kinase/AKT cell survival/apoptosis pathway [11–13]. Because of its homology to the focal adhesion molecules tensin and auxilin, it was hypothesized that PTEN may also play a role in cell migration and focal adhesion. When PTEN was overexpressed in NIH 3T3 cells, it appeared that cell migration was inhibited while antisense PTEN enhanced migration [14]. Evidence for PTEN interaction with focal adhesion kinase (FAK) was given when integrin-mediated cell spreading and focal adhesion formation were down-regulated by wild-type but not mutant PTEN; PTEN must interact with FAK to reduce its tyrosine phosphorylation [14]. This leads to the hypothesis that PTEN functions as a phosphatase by negatively regulating cell interactions with the extracellular matrix.

Although the genetic evidence that points to PTEN as a tumour suppressor—broad spectrum of mutations scattered throughout the gene, truncating mutations, and location of the gene in a region of loss of heterozygosity—is strong, functional demonstration was still required. When functional wild-type PTEN was transfected into a series of glioma cell lines, which carry endogenous *PTEN* mutations or are PTEN null, growth suppression was observed [15–17]. Multiple *in vitro* studies where wild-type and mutant PTEN were overexpressed in a broad variety of cancer cell lines, including those of the breast, thyroid, and prostate, demonstrate that PTEN-phosphatase- and PI3K-dependent G1 cell cycle arrest and/or apoptosis result in growth suppression [18]. This is *in vitro* functional evidence that PTEN acts as a tumour suppressor. Further studies demonstrated that PTEN has phosphatase-independent functions and critical roles within the nucleus [19]. As such, nuclear PTEN is involved in double-stranded break repair by upregulating RAD51 and genomic stability by interacting with centromere protein C

(CENP-C) [19]. In the nucleolus, PTEN regulates ribosomal DNA (rDNA) transcription and cellular proliferation [20].

PTEN is the Cowden Syndrome Gene

The experimental data, its putative function as suggested by structural motifs, and its location within 10q23.3 all argued strongly that *PTEN* was an ideal candidate for the Cowden syndrome susceptibility gene [6]. Therefore, to determine if germline *PTEN* mutations could be aetiological for CS, five families, with a high prior probability of having mutations, were chosen for initial analysis [5, 6]. Two families had nonsense mutations, Arg233X and Glu157X, while two unrelated families shared an identical missense mutation, Gly129Glu, which is a nonconservative amino acid alteration occurring in one of the conserved glycines of the phosphatase signature motif. No unaffected family member carried these mutations. In each family, the family-specific germline *PTEN* mutation segregated with disease but not in unaffected family members nor normal controls. Given these data, *PTEN* was believed to be the most likely susceptibility gene for CS. Early studies on families with CS and individuals with full-blown phenotypic features identified that germline *PTEN* mutations accounted for up to 85% of CS [21, 22]. Subsequent analysis including prospective community-accrued CS and CS-like (not meeting full diagnostic criteria) probands estimated that ~25% of patients who met more relaxed diagnostic criteria harboured pathogenic *PTEN* mutations [23].

PTEN is Also the Bannayan–Riley–Ruvalcaba Gene

Bannayan–Riley–Ruvalcaba syndrome (BRRS; OMIM 158350) is a rare autosomal dominant congenital disorder classically characterized by macrocephaly in combination with intestinal hamartomatous polyposis, vascular malformations, lipomas, haemangiomas, and genital freckling [24]. In addition to the cardinal clinical features, other reported phenotypes include high birth weight, developmental delay, mild to severe mental retardation, delayed psychomotor development, muscle hypotonia, lipid storage myopathy, joint hyperextensibility, pectus excavatum, and scoliosis [24, 25]. Endocrinologists might encounter BRRS as paediatric or young adult patients with benign thyroid manifestations such as Hashimoto thyroiditis, and/or with differentiated thyroid cancers [24, 26]. Clinically, differential diagnoses include other disorders with overlapping phenotypes of macrocephaly, gastrointestinal polyposis, and tumours, such as juvenile polyposis syndrome, Peutz–Jeghers syndrome, and neurofibromatosis type 1. BRRS is allelic to CS, as subsets of both disorders have been associated with germline mutations in *PTEN*, the latter excluded as a candidate locus for the other differential diagnoses except for juvenile polyposis of infancy [27, 28]. Multiple germline *PTEN* mutations have now been described in both familial and isolated cases of BRRS, such that ~60% of BRRS individuals harbour intragenic *PTEN* mutations [21, 29, 30]. Among those without intragenic mutations, another 10% harbour large deletions encompassing or including *PTEN* [22]. Since identical mutations (e.g. Arg233X) have been found in CS as well as BRRS, genetic and non-genetic modifiers must play a role in helping dictate the ultimate phenotype. BRRS patients with germline pathogenic *PTEN* mutations are thought to have the same lifetime cancer risks as *PTEN* mutation-positive CS patients. It is unclear whether these risks apply to BRRS patients who do not have *PTEN* alterations [4].

Concept of the *PTEN* Hamartoma Tumour Syndrome (PHTS)

In addition to CS and BRRS, germline *PTEN* mutations were found in variable subsets of several seemingly unrelated clinical syndromes. For example, up to 20% of individuals with Proteus syndrome have germline *PTEN* mutations [31]. Approximately 10–20% of individuals with autism spectrum disorder (ASD) and macrocephaly harbour germline *PTEN* mutations [32–36]. Single cases of VATER and megalencephaly and hemimegencephaly have been reported to carry germline *PTEN* mutations as well [37, 38]. The concept of PHTS to encompass any clinical disorder with germline *PTEN* mutation was proposed because it is clinically useful [4, 39]. A diagnosis of PHTS is established upon the identification of a pathogenic germline *PTEN* mutation on molecular genetic testing, regardless of phenotype. Finding a germline *PTEN* mutation should trigger cancer risk management and genetic counselling similar to those used for CS-*PTEN*. Additionally, healthcare providers benefit from understanding the clinical spectrum of PHTS in relation to other hereditary cancer syndromes (Table 6.15.1).

Table 6.15.1 Clinical characteristics of *PTEN* hamartoma tumour syndrome (PHTS)

General features
- Multiple hamartomas
- Cancer predisposition
- Autosomal dominant inheritance
- 10–44% caused by de novo *PTEN* mutation
- Extreme intrafamilial variability common
- Penetrance close to 100% by adulthood

Benign neoplasias
- Dermatologic
  - Palmoplantar keratoses
  - Trichilemmomas
  - Papillomatous papules
  - Lipomas
  - Fibromas
  - Freckling of the glans penis
- Vascular anomalies/haemangiomas
- Lhermitte-Duclos disease
- Genitourinary tumours/malformations
- Colorectal polyposis
- Mucosal lesions
- Thyroid goitre/nodules
- Hashimoto thyroiditis
- Proliferative breast changes

Malignancies
- Breast cancer
- Epithelial thyroid cancer
- Renal cancer
- Endometrial cancer
- Colorectal cancer
- Melanoma

Neurodevelopmental
- Macrocephaly
- Autism spectrum disorder
- Developmental delay
- Dysmorphic features (especially with large *PTEN* deletions)
- Dolichocephaly

Adapted with permission from Mester, J. and C. Eng, When overgrowth bumps into cancer: the PTEN-opathies. *Am J Med Genet C Semin Med Genet*, 2013. 163C(2): p.114–21.

PTEN Mutation Spectrum and Genotype–Phenotype Correlations

Pathogenic germline *PTEN* mutations have been reported in all 9 exons of the gene [23, 40, 41]. Exons 5, 7, and 8 are overrepresented in the *PTEN* germline mutation spectrum, likely reflecting encoded domain functions [23]. Exon 5 is a hotspot for germline mutations due to its biological significance as the catalytic core motif of PTEN [23, 30]. Additionally, although limited data are available for *PTEN* promoter variants, some pathogenic promoter mutations have been shown to affect *PTEN* transcription and translation, the latter due to altered RNA secondary structure [22]. Particular *PTEN* intronic variants can cause exon skipping, alternative splicing or the use of cryptic splice sites [42].

With the genotypic and phenotypic heterogeneity observed in PHTS, it became prudent to explore whether genotype–phenotype correlations exist. Earlier exploratory genotype–phenotype analyses identified two potential associations [21, 39]. The first is the association between germline *PTEN* mutations and the presence of malignant breast disease. The second is the association between missense mutation and/or position of mutation within the phosphatase core motif and the development of multiorgan disease. Expanded series of PHTS individuals revealed that germline *PTEN* frameshift mutations are overrepresented, but not absolute, in thyroid cancer [43], nonsense mutations in colorectal cancer [40], promoter mutations in breast cancer [40], and missense mutations in individuals with ASD [44]. In addition to cancer risk, research efforts have also identified differences in *PTEN* missense mutations impacting the three-dimensional dynamics and stability of PTEN protein structure to influence cancer versus ASD phenotypes [45]. Relatedly, multiple studies demonstrated that ASD-associated *PTEN* mutations tend to be less severe than non-ASD-associated mutations [46–48]. Conceivably, independent validation of consistent *PTEN* genotype–phenotype correlations is vital. If proven true, these preliminary associations will help in tailoring medical management with regard to surveillance. It is also suspected that with larger series, other associations might be found as well.

A proof-of-principle study established that other genetic modifiers modulate CS-related cancer risks and tumour histology [49]. Approximately 6–10% of CS/CS-like individuals harboured germline variants in succinate dehydrogenase subunit (mitochondrial complex II) genes, *SDHx*, on top of their pathogenic *PTEN* mutations [49]. The coexistence of a *PTEN* mutation with *SDHx* variants was associated with a 2.5–7-fold increase in breast cancer prevalence compared to individuals with either gene variant alone. Interestingly, while thyroid cancer prevalence was not elevated in individuals with both a *PTEN* mutation and an *SDHx* variant, the histology was papillary for all tumours in this group, in contrast to the follicular histology overrepresented in individuals with *PTEN* mutations alone [43].

PTEN Mutation-Negative CS and BRRS

Approximately 15% of classic CS and ~95% of CS-like individuals remain without detectable *PTEN* mutations [40]. Additionally, germline *PTEN* mutations exist in ~60% of BRRS patients [21, 22, 29, 39]. Therefore, non-*PTEN* aetiologies exist in *PTEN* wild-type patients. Studies have identified multiple susceptibility genes for *PTEN* wild-type CS and BRRS, including *SDHB-D, KLLN, AKT1, PIK3CA, SEC23B, USF3* and *TTN* [50]. More recently, known genes such as *BRCA1, BRCA2, RET* and others were associated with a rare subset of *PTEN* wild-type CS/CS-like and BRRS/BRRS-like patients [51]. The identification of other CS/CS-like and BRRS/BRRS-like susceptibility genes reflects the phenotypic heterogeneity in these patients. Validating these and other gene discoveries is important because while *PTEN* mutation-negative patients can be diagnosed clinically, they do not have the benefit of specific gene-informed genetic counselling, predictive testing of family members, precise risk assessment, and subsequent management.

Clinical Aspects

Diagnostic Criteria

Clinical diagnosis of CS is often challenging, particularly because of the protean manifestations that may occur in isolation and sporadically in the general population. Because of the lack of uniform diagnostic criteria for CS prior to 1995, a group of individuals, the International Cowden Consortium [52], interested in systematically studying this syndrome arrived at a set of consensus operational diagnostic criteria (Table 6.15.2). These guidelines form the basis for the US-based National Comprehensive Cancer Network (NCCN) guidelines. CS usually presents by the late 20s. It has variable expression and age-related penetrance. By the third decade, 99% of affected individuals would have developed the mucocutaneous stigmata although any of the features could be present already. The most commonly, but not uniformly, reported manifestations are macrocephaly (specifically, megalencephaly), mucocutaneous lesions, thyroid abnormalities, fibrocystic disease and carcinoma of the breast, gastrointestinal hamartomas, multiple, early-onset uterine leiomyoma, and developmental delay [3, 4, 53]. Pathognomonic mucocutaneous lesions are believed to exist in 100% of CS patients by age 30.

As a clinical decision support tool, a nomogram-based clinical predictor named Cleveland Clinic *PTEN* Risk Calculator, was developed to evaluate the pretest probability of harbouring a germline *PTEN* mutation [23]. Based on relevant age-adjusted phenotypic features, a weighted sum referred to as the *PTEN* Cleveland Clinic score (CC score) is generated. The questionnaire-based clinical decision tool is available online to assist clinicians at the point of patient care (http://www.lerner.ccf.org/gmi/ccscore/). In adults, a CC score of 10, indicating a pretest probability of 3% and a sensitivity of 90%, was a recommended threshold for referral of patients to genetic specialists for *PTEN* genetic testing [23]. A CC score of 15, corresponding to a 10% a priori risk of positive *PTEN* mutation status, is the most cost-effective cut-off to refer CS-like patients for *PTEN* germline testing [54]. For paediatric patients (<18 years), distinct criteria were developed to guide selection for germline *PTEN* mutation testing [23]. Macrocephaly is a major criterion based on 100% prevalence at diagnosis. Neurodevelopmental and dermatologic features were also present in 100% of PHTS paediatric patients (Table 6.15.3). However, since dermatologic features are often overlooked and difficult to recognize, less prevalent features such as vascular malformations, thyroid goitre, gastrointestinal polyps and early-onset cancers also warrant referral of the paediatric patient for

Table 6.15.2 International Cowden Consortium (ICC) operational diagnostic criteria

Pathognomonic	Major	Minor
• Adult Lhermitte-Duclos disease (LDD) • Mucocutaneous lesions • Trichilemmomas, facial • Acral keratoses • Papillomatous papules • Mucosal lesions	• Breast carcinoma • Thyroid carcinoma (non-medullary), especially follicular thyroid carcinoma • Macrocephaly (occipital frontal circumference ≥97th percentile) • Endometrial carcinoma	• Other thyroid lesions (e.g. adenoma, multinodular goitre) • Intellectual disability (i.e. IQ ≤75) • Gastrointestinal hamartomas • Fibrocystic breast disease • Lipomas • Fibromas • Genitourinary tumours (especially renal cell carcinoma) • Genitourinary malformations • Uterine fibroids
Operational diagnosis in an individual		
Any of following: • Mucocutaneous lesions alone, if ≥ six facial papules (three of which must be trichilemmomas) • Cutaneous facial papules and oral mucosal papillomatosis • Oral mucosal papillomatosis and acral keratoses • ≥ Six palmoplantar keratoses • ≥ Two major criteria (one of which must be macrocephaly or LDD) • One major and ≥ three minor criteria • ≥ Four minor criteria		
Operational diagnosis in a family where one individual is diagnostic for CS		
• Any one pathognomonic criterion • Any one major criterion ± minor criteria • Two minor criteria • History of Bannayan–Riley–Ruvalcaba syndrome (BRRS)		

Adapted with permission from Eng, C., Will the real Cowden syndrome please stand up: revised diagnostic criteria. *J Med Genet*, 2000. 37(11): p.828–30.

genetics evaluation for consideration of *PTEN* testing. Importantly, early diagnosis is vital to minimize morbidity and mortality.

Pathology

Like other inherited cancer syndromes, multifocality and bilateral involvement is the rule. Hamartomas are the hallmark of CS. These are classic hamartomas in general and are benign tumours comprising at least two (strictly three) elements of a particular organ but growing in a disorganized fashion. Of note, the hamartomatous polyps found in this syndrome are different in histomorphology from Peutz–Jeghers polyps, which have a distinct appearance. Hence, caution must be taken when the polyp histology is not read by a dedicated gastrointestinal pathologist as histological diagnoses are often incorrect when compared to genetic classification [55]. Importantly, the majority (>90%) of *PTEN* mutation-positive patients who had colonoscopy performed as part of clinical care, had colorectal polyps typically with a mix of histologic subtypes, including ganglioneuromas, hamartomatous, juvenile and adenomatous polyps [56]. Patients who developed colorectal cancer also tended to have pre-/coexisting colonic polyposis.

Table 6.15.3 Paediatric criteria for consideration of PHTS

Required criterion	Secondary criteria
Macrocephaly (≥2 standard deviations)	At least one of the following criteria should be present: • Autism spectrum disorder or developmental delay • Dermatologic features (lipomas, oral papillomas, trichilemmomas, penile freckling) • Vascular features (arteriovenous malformations or haemangiomas) • Gastrointestinal polyps • Paediatric-onset thyroid cancer or germ cell tumours

Reproduced with permission from Tan, M.H., J. Mester, C. Peterson, Y. Yang, J.L. Chen, L.A. Rybicki, *et al.*, A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *Am J Hum Genet*, 2011. 88(1): p.42–56.

PTEN-Related Cancer Risks

The most serious complication of CS is the increased risk of cancer. Even before the discovery of *PTEN* as the first CS susceptibility gene, it was recognized that CS patients have elevated lifetime risks for breast and thyroid cancers [2, 57]. The identification of germline *PTEN* mutations in individuals with CS [6] subsequently led to the identification of other organ-specific cancer risk estimates [40]. Female breast cancer lifetime risk is the most pronounced, beginning at around age 30 and reaching an estimated lifetime risk of 85%. Although male breast cancer had been reported in CS cases [21, 58], an increased lifetime risk in *PTEN* mutation-positive males was not noted in this prospectively accrued series of >3000 patients [40]. Also noted are elevated lifetime risks of 35% for epithelial thyroid cancer, 28% for endometrial cancer, 34% for renal cell cancer, 9% for colorectal cancer and 6% for melanoma. Elevated cancer risks have been independently replicated by other groups [59, 60]. Individuals with germline *PTEN* mutations also have a 7-fold increased risk of developing a second malignant (primary) neoplasm as compared to the US general population [41]. Collectively, these observations resulted in tailoring clinical surveillance and medical management for PHTS patients (Table 6.15.4). As relevant to endocrinology clinic, earlier clinical surveillance may be advisable because Hashimoto thyroiditis and nodules are seen by the time

Table 6.15.4 Screening and management recommendations for PHTS

	Screening/surgical guidelines*	Age to start	Frequency
Breast	Breast awareness and self-exam: report changes to healthcare provider	18	Consistent
	Clinical breast exam	25**	Every 6–12 months
	Mammogram with consideration of tomosynthesis and breast MRI with contrast	30–35**	Every 12 months
	Discuss mastectomy	Personalized	As needed
Thyroid	Thyroid ultrasound	Time of PHTS diagnosis, including childhood	Every 12 months
	Discuss thyroidectomy	Personalized	As needed
Kidney	Consider renal ultrasound	40	Every 1–2 years
Endometrium	Encourage patient education and prompt response to symptoms (e.g. abnormal bleeding)	Not applicable	Not applicable
	Consider random endometrial biopsies and/or transvaginal ultrasound	30–35	Every 12 months
	Discuss hysterectomy with completion of childbearing	Personalized	As needed
Colon	Colonoscopy	35** unless symptomatic	Every 5 years or more frequently depending if patient is symptomatic or polyps are found
Dermatologic	Dermatologic exam	Personalized	Clinician's recommendation
Developmental	Consider psychomotor assessment in children	Time of PHTS diagnosis	Clinician's recommendation
	Brain MRI if symptomatic	Time of PHTS diagnosis	Clinician's recommendation

* Annual comprehensive physical exam starting at age 18 y or 5 y before the youngest age of diagnosis of a component cancer in the family (whichever comes first), with particular attention to thyroid exam. Encourage patient education regarding the signs and symptoms of cancer.
** Cancer screening should begin 5–10 years before the earliest known component cancer in the family or according to the ages listed in the above table, whichever comes first.

patients reach adolescence, and cancer diagnosis occurs on average 14 years earlier than expected, with the earliest thyroid cancer diagnosis reported at age 6 [23]. Furthermore, the thyroid cancer risks justify prophylactic total thyroidectomy in PHTS patients undergoing surgery for benign thyroid disease [4]. Relatedly, risk reduction surgery is considered for patients with ASD or cognitive deficits who do not permit ultrasound without sedation [61]. The dilemma in these groups of patients, however, is whether daily replacement thyroid hormone is feasible as they may have difficulties swallowing pills.

Role of Endocrinologists in Recognizing Clinical 'Red Flags'

There are several ways in which CS patients can come to the attention of endocrinologists or endocrine surgeons. Sometimes, an individual with known CS is referred for management of their endocrine problems, chief of which are multinodular goitre, thyroid adenomas, and epithelial thyroid carcinomas. More commonly, such patients are not previously diagnosed and seek endocrine-related care because of abnormal thyroid function or a thyroid mass. Over two-thirds of CS patients have thyroid problems, which may occur at any age. However, finding multifocal lesions, especially in young individuals, should raise suspicion. Endocrinologists and endocrine surgeons should be especially mindful of the differential diagnosis of CS should they see patients with these thyroid lesions. A careful history and physical examination, as well as a meticulous family history to look for other component symptoms and signs of CS, are warranted.

Rarely, CS individuals present with uncommon features such as hyperparathyroidism or parathyroid adenomas. When these occur together with 'a thyroid cancer', the initial diagnosis that endocrinologists might think of is multiple endocrine neoplasia (MEN) type 2a (also referred to as type 2) (see Chapter 6.11.2, 'Multiple Endocrine Neoplasia Type 2a and 2b') [3, 62]. However, it would be prudent to pursue the histology of the thyroid cancer as this might turn out to be a CS patient and not a MEN 2a (or MEN 2) case. Even more unusual, CS can present with ganglioneuromas of the gut and are referred to the endocrinologist for MEN 2B (also referred to as MEN 3). However, in general, MEN 2B (or MEN 3) and CS, are clinically and genetically distinct [3, 62]. A few MEN 2B (MEN 3) cases can present with apparently isolated intestinal ganglioneuromatosis without the other classic stigmata of MEN 2B, yet all were found to have the MEN 2B-defining germline *RET* mutation p.Met918Thr and all developed medullary thyroid carcinoma [63]. In contrast, follicular thyroid carcinoma and follicular variant of papillary thyroid carcinoma tend to be overrepresented in PHTS [43]. PHTS patients may also present with metabolic dysregulation in the form of significantly lower fasting plasma insulin levels, higher glucose infusion rate, and increased measures of obesity [64]. This insulin hypersensitivity phenotype reflects an apparently divergent effect of

PTEN mutations, with increased risks of obesity and cancer, but decreased risk of type 2 diabetes.

Genetic Counselling, Medical Management, and Therapeutic Approaches

The key to proper genetic counselling in CS is recognition of the syndrome. Families with CS should be counselled as for any autosomal dominant trait with high penetrance. What is unclear, however, is the variability of expression between and within families. We suspect that there are CS families who have nothing but trichilemmomas and, therefore, never come to medical attention. As an autosomal dominant disorder, children of an affected person have a 50% probability to inherit the gene mutation and relatives are at an increased risk as well. In up to 44% of cases, patients have a de novo *PTEN* mutation, which means the mutation is not shared by any of the parents [65]. If a family-specific mutation is known, then screening for that particular mutation in unaffected family members would yield results which are 100% accurate, barring administrative error. During the genetic testing process, genetic counselling is recommended by the American Society of Clinical Oncology and many other professional societies. Genetic counsellors can guide the patient to appropriate specialists and help them understand management. A relationship between the patient, primary care provider, and genetics team can be crucial for patient care. The key to successful management of PHTS/CS patients and their families is a multidisciplinary team (such an example can be viewed at my.clevelandclinic.org). There should always be a primary care provider, who orchestrates the care of such patients, some of whom will need the care of surgeons, endocrinologists, gynaecologists, dermatologists, gastroenterologists, oncologists and high-risk cancer specialists, neurologists, and geneticists.

As relevant to targeted therapeutics, because *PTEN* alterations result in enhanced PI3K/AKT/mTOR signalling, the latter represents a rational pathway to target in PHTS. A phase II open label clinical trial (NCT00971789) utilized the mTOR inhibitor sirolimus in adult (>18 years) PHTS patients, with ongoing follow-up data and analysis (https://theoncologist.onlinelibrary.wiley.com/doi/full/10.1634/theoncologist.2019-0514). Another mTOR inhibitor trial is currently accruing paediatric, adolescent, and young adult patients with germline pathogenic *PTEN* mutations and ASD [4]. It is warranted to explore the clinical utility of pharmacologically inhibiting other upstream components of the PTEN pathway such as PI3K and AKT. From a biological perspective, since PTEN regulates a vast array of cellular processes independent of PI3K/AKT/mTOR signalling, it is also critical to explore targeting such vulnerabilities. For example, since nuclear PTEN participates in maintaining genomic integrity, PARP inhibitors would make sense in individuals with mutations that disrupt nuclear PTEN function [66].

ACKNOWLEDGEMENTS

We are grateful to our patients and families who have participated in our research for over two decades. LY is an Ambrose Monell Foundation Cancer Genomic Medicine Fellow at the Cleveland Clinic Genomic Medicine Institute. CE is the Sondra J. and Stephen R. Hardis Chair of Cancer Genomic Medicine at the Cleveland Clinic and an American Cancer Society Clinical Research Professor.

REFERENCES

1. Lloyd KM, 2nd, Dennis M. Cowden's disease. A possible new symptom complex with multiple system involvement. *Ann Intern Med*, 1963; **58**: 136–42.
2. Starink TM, van der Veen JP, Arwert F, *et al*. The Cowden syndrome: a clinical and genetic study in 21 patients. *Clin Genset*, 1986; **29**: 222–33.
3. Zbuk KM, Eng C. Cancer phenomics: RET and PTEN as illustrative models. *Nat Rev Cancer*, 2007; **7**: 35–45.
4. Eng C. PTEN hamartoma tumor syndrome. 2001 Nov 29 [Updated 2016 Jun 2]. In: Adam MP, Ardinger HH, Pagon RA, *et al*., eds. GeneReviews® [Internet]. Seattle (WA): University of Washington: 1993–2018.
5. Nelen MR, Padberg GW, Peeters EA, *et al*. Localization of the gene for Cowden disease to chromosome 10q22-23. *Nat Genet*, 1996; **13**: 114–16.
6. Liaw D, Marsh DJ, Li J, *et al*. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. *Nat Genet*, 1997; **16**: 64–7.
7. Li J, Yen C, Liaw D, Podsypanina K, *et al*. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, 1997; **275**: 1943–7.
8. Li DM, Sun H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. *Cancer Res*, 1997; **57**: 2124–9.
9. Myers MP, Stolarov JP, Eng C, *et al*. P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. *Proc Natl Acad Sci U S A*, 1997; **94**: 9052–7.
10. Maehama T, Dixon JE. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J Biol Chem*, 1998; **273**: 13375–8.
11. Suzuki A, de la Pompa JL, Stambolic V, *et al*. High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Curr Biol*, 1998; **8**: 1169–78.
12. Stambolic V, Suzuki A, de la Pompa JL, *et al*. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*, 1998; **95**: 29–39.
13. Dahia PL, Aguiar RC, Alberta J, *et al*. PTEN is inversely correlated with the cell survival factor Akt/PKB and is inactivated via multiple mechanismsin haematological malignancies. *Hum Mol Genet*, 1999; **8**: 185–93.
14. Tamura M, Gu J, Danen EH, Takino T, Miyamoto S, Yamada KM. PTEN interactions with focal adhesion kinase and suppression of the extracellular matrix-dependent phosphatidylinositol 3-kinase/Akt cell survival pathway. *J Biol Chem*, 1999; **274**: 20693–703.
15. Furnari FB, Lin H, Huang HS, Cavenee WK. Growth suppression of glioma cells by PTEN requires a functional phosphatase catalytic domain. *Proc Natl Acad Sci U S A*, 1997; **94**: 12479–84.
16. Cheney IW, Johnson DE, Vaillancourt MT, *et al*. Suppression of tumorigenicity of glioblastoma cells by adenovirus-mediated MMAC1/PTEN gene transfer. *Cancer Res*, 1998; **58**: 2331–4.

17. Li DM, Sun H. PTEN/MMAC1/TEP1 suppresses the tumorigenicity and induces G1 cell cycle arrest in human glioblastoma cells. *Proc Natl Acad Sci U S A*, 1998; **95**: 15406–11.
18. Weng LP, Brown JL, Eng C. PTEN coordinates G(1) arrest by down-regulating cyclin D1 via its protein phosphatase activity and up-regulating p27 via its lipid phosphatase activity in a breast cancer model. *Hum Mol Genet*, 2001; **10**: 599–604.
19. Planchon SM, Waite KA, Eng C. The nuclear affairs of PTEN. *J Cell Sci*, 2008; **121**(Pt 3): 249–53.
20. Liang H, Chen X, Yin Q, *et al.* PTENbeta is an alternatively translated isoform of PTEN that regulates rDNA transcription. *Nat Commun*, 2017; **8**: 14771.
21. Marsh DJ, Coulon V, Lunetta KL, *et al.* Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. *Hum Mol Genet*, 1998; **7**: 507–15.
22. Zhou XP, Waite KA, Pilarski R, *et al.* Germline PTEN promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant PTEN protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway. *Am J Hum Genet*, 2003; **73**: 404–11.
23. Tan MH, Mester J, Peterson C, *et al.* A clinical scoring system for selection of patients for PTEN mutation testing is proposed on the basis of a prospective study of 3042 probands. *Am J Hum Genet*, 2011; **88**: 42–56.
24. Gorlin RJ, Cohen MM, Jr., Csondon LM, Burke BA. Bannayan-Riley-Ruvalcaba syndrome. *Am J Med Genet*, 1992; **44**: 307–14.
25. DiLiberti JH, D'Agostino AN, Ruvalcaba RH, Schimschock JR. A new lipid storage myopathy observed in individuals with the Ruvalcaba-Myhre-Smith syndrome. *Am J Med Genet*, 1984; **18**: 163–7.
26. Peiretti V, Mussa A, Feyles F, *et al.* Thyroid involvement in two patients with Bannayan-Riley-Ruvalcaba syndrome. *J Clin Res Pediatr Endocrinol*, 2013; **5**: 261–5.
27. Marsh DJ, Roth S, Lunetta KL, *et al.* Exclusion of PTEN and 10q22-24 as the susceptibility locus for juvenile polyposis syndrome. *Cancer Res*, 1997; **57**: 5017–21.
28. Delnatte C, Sanlaville D, Mougenot JF, *et al.* Contiguous gene deletion within chromosome arm 10q is associated with juvenile polyposis of infancy, reflecting cooperation between the BMPR1A and PTEN tumor-suppressor genes. *Am J Hum Genet*, 2006; **78**: 1066–74.
29. Marsh DJ, Dahia PL, Zheng Z, *et al.* Germline mutations in PTEN are present in Bannayan-Zonana syndrome. *Nat Genet*, 1997; **16**: 333–4.
30. Eng C. PTEN: one gene, many syndromes. *Hum Mutat*, 2003; **22**: 183–98.
31. Zhou X, Hampel H, Thiele H, *et al.* Association of germline mutation in the PTEN tumour suppressor gene and Proteus and Proteus-like syndromes. *Lancet*, 2001; **358**: 210–11.
32. Butler MG, Dasouki MJ, Zhou XP, *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet*, 2005; **42**: 318–21.
33. Herman GE, Butter E, Enrile B, Pastore M, Prior TW, Sommer A. Increasing knowledge of PTEN germline mutations: two additional patients with autism and macrocephaly. *Am J Med Genet A*, 2007; **143A**: 589–93.
34. Varga EA, Pastore M, Prior T, Herman GE, McBride KL. The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. *Genet Med*, 2009; **11**: 111–17.
35. McBride KL, Varga EA, Pastore MT, *et al.* Confirmation study of PTEN mutations among individuals with autism or developmental delays/mental retardation and macrocephaly. *Autism Res*, 2010; **3**: 137–41.
36. Hansen-Kiss E, Beinkampen S, Adler B, *et al.* A retrospective chart review of the features of PTEN hamartoma tumour syndrome in children. *J Med Genet*, 2017; **54**: 471–8.
37. Reardon W, Zhou XP, Eng C. A novel germline mutation of the PTEN gene in a patient with macrocephaly, ventricular dilatation, and features of VATER association. *J Med Genet*, 2001; **38**: 820–3.
38. Merks JH, de Vries LS, Zhou XP, *et al.* PTEN hamartoma tumour syndrome: variability of an entity. *J Med Genet*, 2003; **40**: e111.
39. Marsh DJ, Kum JB, Lunetta KL, *et al.* PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. *Hum Mol Genet*, 1999; **8**: 1461–72.
40. Tan MH, Mester JL, Ngeow J, Rybicki LA, MS Orloff, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res*, 2012; **18**: 400–7.
41. Ngeow J, Stanuch K, Mester JL, Barnholtz-Sloan JS, Eng C. Second malignant neoplasms in patients with cowden syndrome with underlying germline PTEN mutations. *J Clin Oncol*, 2014; **32**: 1818–24.
42. Chen HJ, Romigh T, Sesock K, Eng C. Characterization of cryptic splicing in germline PTEN intronic variants in Cowden syndrome. *Hum Mutat*, 2017; **38**: 1372–7.
43. Ngeow J, Mester J, Rybicki LA, Ni Y, Milas M, Eng C. Incidence and clinical characteristics of thyroid cancer in prospective series of individuals with Cowden and Cowden-like syndrome characterized by germline PTEN, SDH, or KLLN alterations. *J Clin Endocrinol Metab*, 2011; **96**: E2063–71.
44. Frazier TW, Embacher R, Tilot AK, Koenig K, Mester J, Eng C. Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism. *Mol Psychiatry*, 2015; **20**: 1132–8.
45. Smith IN, Thacker S, Jaini R, Eng C. Dynamics and structural stability effects of germline PTEN mutations associated with cancer versus autism phenotypes. *J Biomol Struct Dyn*, 2019; **37**: 1766–82.
46. Spinelli L, Black FM, Berg JN, Eickholt BJ, Leslie NR. Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. *J Med Genet*, 2015; **52**: 128–34.
47. Leslie NR, Longy M. Inherited PTEN mutations and the prediction of phenotype. *Semin Cell Dev Biol*, 2016; **52**: 30–8.
48. Mighell TL, Evans-Dutson S, O'Roak BJ. A saturation mutagenesis approach to understanding PTEN lipid phosphatase activity and genotype-phenotype relationships. *Am J Hum Genet*, 2018; **102**: 943–55.
49. Ni Y, He X, Chen J, *et al.* Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. *Hum Mol Genet*, 2012; **21**: 300–10.
50. Yehia L, Eng C. One gene—many endocrine and metabolic syndromes: PTEN-opathies and precision medicine. *Endocr Relat Cancer*, 2018; **25**: T121–40.
51. Yehia L, Ni Y, Sesock K, Niazi F, *et al.* Unexpected cancer-predisposition gene variants in Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome patients without underlying germline PTEN mutations. *PLoS Genet*, 2018; **14**: e1007352.

52. Eng C. Will the real Cowden syndrome please stand up: revised diagnostic criteria. *J Med Genet*, 2000; **37**: 828–30.
53. Hanssen AM, Fryns JP. Cowden syndrome. *J Med Genet*, 1995; **32**: 117–19.
54. Ngeow J, Liu C, Zhou K, Frick KD, Matchar DB, Eng C. Detecting germline PTEN mutations among at-risk patients with cancer: an age- and sex-specific cost-effectiveness analysis. *J Clin Oncol*, 2015; **33**: 2537–44.
55. Sweet K, Willis J, Zhou XP, *et al*. Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. *JAMA*, 2005; **294**: 2465–73.
56. Heald B, Mester J, Rybicki L, Orloff MS, Burke CA, Eng C. Frequent gastrointestinal polyps and colorectal adenocarcinomas in a prospective series of PTEN mutation carriers. *Gastroenterology*, 2010; **139**: 1927–33.
57. Brownstein MH, Wolf M, Bikowski JB. Cowden's disease: a cutaneous marker of breast cancer. *Cancer*, 1978; **41**: 2393–8.
58. Fackenthal JD, Marsh DJ, Richardson AL, *et al*. Male breast cancer in Cowden syndrome patients with germline PTEN mutations. *J Med Genet*, 2001; **38**: 159–64.
59. Bubien, V., F. Bonnet, V. Brouste, *et al*. High cumulative risks of cancer in patients with PTEN hamartoma tumour syndrome. *J Med Genet*, 2013; **50**: 255–63.
60. Nieuwenhuis MH, Kets CM, Murphy-Ryan M, *et al*. Cancer risk and genotype-phenotype correlations in PTEN hamartoma tumor syndrome. *Fam Cancer*, 2014; **13**: 57–63.
61. Milas M, Mester J, Metzger R, *et al*. Should patients with Cowden syndrome undergo prophylactic thyroidectomy? *Surgery*, 2012; **152**: 1201–10.
62. Eng C. Seminars in medicine of the Beth Israel Hospital, Boston. The RET proto-oncogene in multiple endocrine neoplasia type 2 and Hirschsprung's disease. *N Engl J Med*, 1996; **335**: 943–51.
63. Gordon C, Majzoub JA, Marsh DJ, *et al*. Four cases of mucosal neuroma syndrome: multiple endocrine neoplasm 2B or not 2B? *J Clin Endocrinol Metab*, 1998; **83**: 17–20.
64. Pal A, Barber TM, Van de Bunt M, *et al*. PTEN mutations as a cause of constitutive insulin sensitivity and obesity. *N Engl J Med*, 2012; **367**: 1002–11.
65. Mester J, Eng C. Estimate of de novo mutation frequency in probands with PTEN hamartoma tumor syndrome. *Genet Med*, 2012; **14**: 819–22.
66. Dillon LM, Miller TW. Therapeutic targeting of cancers with loss of PTEN function. *Curr Drug Targets*, 2014; **15**: 65–79.

Index

Tables, figures, and boxes are indicated by an italic *t*, *f*, and *b* following the page number.

Oxford Textbook of

Endocrinology and Diabetes

Oxford Textbook of

Endocrinology and Diabetes

THIRD EDITION

Volume 2: Sections 7–15

EDITED BY

John A.H. Wass

Professor of Endocrinology, Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Wiebke Arlt

William Withering Chair of Medicine, Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

Robert K. Semple

Professor of Translational Molecular Medicine, University of Edinburgh Centre for Cardiovascular Science, Queen's Medical Research Institute, Edinburgh, UK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2002

Second Edition published in 2011

Third Edition published in 2022

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2021951543

ISBN 978–0–19–880232–7
ISBN 978–0–19–887018–0

DOI: 10.1093/med/9780198802327.001.0001

Printed in Great Britain by
Bell & Bain Ltd., Glasgow

Preface

It is now over a hundred and seventy years since Arnold Berthold demonstrated that endocrine glands convey their effects via the bloodstream and a hundred and ten years since the word 'hormone' was coined by Ernest Starling. Amazingly, it is just under a hundred years since one of the nascent specialty's most spectacular Nobel Prize-winning triumphs, the purification and use of insulin to transform Type I diabetes from a rapidly fatal wasting disease of childhood into a manageable chronic condition. Similarly, seventy years ago the Nobel Prize recognised the life-saving impact of glucocorticoids both for patients with adrenal insufficiency and as immunosuppressive therapy for chronic inflammatory disease. Since then endocrinology has continued to ride the crest of a wave of technological advances, most prominent recently in areas such as molecular genetics, imaging, immunotherapy and rational drug design.

The first edition of the *Oxford Textbook of Endocrinology & Diabetes* was published in 2002. The landscape of endocrine disease has evolved enormously since then, driven largely by burgeoning obesity and by population ageing. The spectrum of endocrinology has been further broadened by the increasing use of immunomodulatory therapies with endocrine complications, evolving patterns of recreational drug use impacting endocrine systems, and widespread exposure to endocrine disrupting chemicals. The therapeutic armamentarium of the endocrinologist has also expanded at pace, not only through development of novel small molecule and biological therapies, but also through step changes in e-technology and their application to chronic disease management and "precision" medicine.

We now comprehensively update the second edition of the *Textbook* published in 2012. The expanded first part of the book is devoted to overview chapters focusing on principles underpinning the science and practice of endocrinology. Elsewhere, other sections have also been enlarged to capture exciting developments in subspeciality practice, with separate sections now devoted to endocrine disease in pregnancy and transgender endocrinology.

The diabetes section has been extensively reorganised to reflect rapid advances in understanding the molecular pathogenesis of diabetes, step changes in the sophistication of technologies used for metabolic monitoring and insulin delivery, and innovations in immunotherapy, behaviour-focused and cell-based therapies. New chapters are devoted to current urgent responses to diabetes as a public health and economic challenge.

Fascination in the science underlying endocrinology continues to endure, and clinical endocrinologists have an ever more sophisticated ability to transform the length and quality of life of those with hormone-related diseases. This book aims both to illuminate the emerging scientific concepts that underlie endocrinology, and to provide an accessible and authoritative account of cutting edge endocrine practice.

We are very grateful indeed to our national and international colleagues who have kindly, expertly and cerebrally contributed to all sections. We are proud of this *magnum opus* and thank them sincerely for their significant efforts.

We also thank Claire Brankin, James Oates and Helen Liepman from Oxford University Press who have expertly and efficiently guided us through this whole process.

This book should be available to every endocrinologist, trainee and researcher and we hope that it provides as much enjoyment and intellectual satisfaction in the reading as it did in putting it together.

John Wass
Robert Semple
Wiebke Arlt

Contents

Volume 1

SECTION 4

Parathyroid, Calcium and Bone Metabolism Disorders

Section editor: John Bilezikian

SECTION 5

Adrenal Diseases

Section editor: Wiebke Arlt

Volume 2

SECTION 7
Disorders of Growth, and Development and Transition

Section editor: Peter Clayton

SECTION 8
Female Reproductive Endocrine Disorders

Section editor: Bulent Okan Yildiz

SECTION 9
Endocrine Disorders of Pregnancy

Section editors: Kristien Boelaert and Cathy Williamson

SECTION 10
Male Reproductive Endocrine Disorders

Section editors: Frederick Wu and Mathis Grossmann

SECTION 11
Management of the Transgender Patient

Section editor: Guy T'Sjoen

SECTION 12
Endocrine Responses to Systemic Diseases or Substance Misuse

Section editor: Ken Ho

SECTION 13
Endocrinology of Cancer

Section editor: David Ray

Symbols and Abbreviations

AAS	androgenic anabolic steroids
AC	adenylate cyclase (*also* arachnoid cysts)
ACE	American College of Endocrinology
ACE	angiotensin-converting enzyme
ACTH	adrenocorticotrophic hormone
AD	autosomal dominant
ADH	autosomal dominant hypocalcaemia
ADIS	agonist-driven insertional signalling
ADT	androgen deprivation therapy
AF	atrial fibrillation
AGHDA	assessment of GH deficiency in adults
AgRP	agouti-related peptide
AHA	Anterior hypothalamic area
AHO	Albright's hereditary osteodystrophy
AI	alcohol-induced (*also* aromatase inhibitors)
AIDS	acquired immunodeficiency syndrome
AIDS	advanced HIV disease
AIP	aryl hydrocarbon receptor-interacting protein
AIRE	autoimmune regulator
AKAP	A-kinase-anchoring proteins
ALS	acid-labile subunit
ALSPAC	Avon Longitudinal Study of Parents and Children
ANCA	antineutrophil cytoplasmic antibodies
ANF	atrial natriuretic factor
AP	anterior pituitary
APECED	autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy
APT	aggressive pituitary tumours
AR	androgen receptor
ARC	arcuate nucleus
ARE	androgen response elements
ART	assisted reproduction technologies
ASD	autism spectrum disorders
ATA	American Thyroid Association
ATD	antithyroid drug
ATMA	antithyroid microsomal antibody
ATOR	Australian thyroid-associated orbitopathy research
AVP	arginine-vasopressin (*also* antidiuretic hormone)
BMAD	bone mineral apparent density
BMD	bone mineral density
BMI	body mass index
BP	blood pressure
BST	bed nucleus of the stria terminalis
CAH	congenital adrenal hyperplasia
CAIS	complete androgen insensitivity syndrome
CAPTEM	capecitabine and temozolomide
CART	cocaine- and amphetamine-related transcript
CAS	clinical activity score
CAS	Court of Arbitration for Sport
CaSR	calcium-sensing receptor
CBG	corticosteroid-binding globulin
CBG	cortisol-binding globulin
CC	clivus chordoma (*also* clomiphene citrate)
CCCR	calcium-to-creatinine clearance ratio
CCH	central congenital hypothyroidism
CDI	central diabetes insipidus
CDK	cyclin-dependent kinase
CDP	constitutional delay of puberty
CE	calcium excretion
CEA	carcinoembryonic antigen
CG	chorionic gonadotrophin
CGTT	clinical genetics think tank
CH	cavernous haemangiomas (*also* congenital hypopituitarism)
CHH	congenital hypogonadotropic hypogonadism
CIRCI	critical illness-related corticosteroid insufficiency
CKD	chronic kidney disease
CLIP	corticotropin-like intermediate peptide
CLOCK	circadian locomotor output cycles kaput
CME	clathrin-mediated endocytosis
CNS	central nervous system
CNV	copy number variant
COPD	chronic obstructive pulmonary disease
CPAP	continuous positive airway pressure
CPHD	combined pituitary hormone deficiency
CPI	checkpoint inhibitors
CRE	CAMP-response elements
CREB	CAMP-Response Element Binding
CREB	cyclic AMP-response element binding
CRH	corticotropin-releasing hormone
CRP	C-reactive protein
CRPC	castration-resistant prostate cancer
CSF	cerebrospinal fluid
CSHI	continuous subcutaneous hydrocortisone infusion
CSM	cavernous sinus meningiomas
CT	computed tomography
DA	dopamine agonists
DAA	direct-acting antivirals
DALY	disability-adjusted life year
DBD	DNA-binding domain
DCV	dense-cored vesicles
DDT	dichloro-diphenyl-trichloroethane

DEXA	dual-energy X-ray absorptiometry
DHAES	dehydroepiandrosterone sulphate
DI	diabetes insipidus
DMD	Duchenne muscular dystrophy
DMH	dorsomedial hypothalamus
DON	dysthyroid optic neuropathy
DS	diaphragma sellae
DSA	digital subtraction angiography
DSD	disorder of sex development
dSPZ	dorsal subparaventricular zone
DXA	dual-energy X-ray absorptiometry
ED	endocrine disruptor (*also* erectile dysfunction)
EDC	endocrine disrupting compounds
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
EMA	European Medicines Agency
EMAS	European Male Ageing Study
ENIGI	European Network for the Investigation of Gender Incongruence
EPP	ectopic posterior pituitary
EQA	external quality assurance
ER	endoplasmic reticulum
ERR	excess relative risk
ESE	European Society of Endocrinology
ESPGHAN	European Society of Paediatric Gastroenterology, Hepatology and Nutrition
ESPR	European Society of Paediatric Research
ESR	erythrocyte sedimentation rate
ESRD	end stage renal disease
ETA	European Thyroid Association
EUGOGO	European Group on Graves' Orbitopathy
FAP	familial amyloidotic polyneuropathy
FDA	US Food and Drug Administration
FDH	familial dysalbuminaemic hyperthyroxinaemia
FEO	food-entrainable oscillator
FHH	familial hypocalciuric hypercalcaemia
FHPP	familial hypokalaemic periodic paralysis
FIH	familial isolated hyperparathyroidism
FISH	fluorescent *in situ* hybridization
FN	follicular neoplasm
FNA	fine-needle aspiration
FNAB	fine-needle aspiration biopsy
FSH	follicle-stimulating hormone
GABA	gamma aminobutyric acid
GAD	glutamic acid decarboxylase
GAH	gender-affirming hormones
GBD	gracile bone dysplasia
GD	gender dysphoria (*also* Graves' disease)
GDNF	glial cell-derived neurotrophic factor
GDR	German Democratic Republic
GFR	glomerular filtration rate
GH	growth hormone
GHBP	growth hormone-binding protein
GHD	growth hormone deficiency
GHR	growth hormone receptor
GHRH	growth hormone-releasing hormone
GHS	growth hormone secretagogues
GK	Gamma Knife
GNAS	guanine nucleotide-binding protein G(S)
GnRH	gonadotropin-releasing hormone
GO	Graves' orbitopathy
GPCR	G-protein-coupled receptor
GREAT	Graves' Recurrent Events After Therapy
GRP	gastrin-releasing peptide
GTR	gross total resection
GTV	gross tumour volume
GWAS	genome-wide association studies
H&E	haematoxylin & eosin
HCG	human chorionic gonadotropin
HDL	high-density lipoprotein
HDR	hypoparathyroidism, deafness, and renal
HGPIN	high-grade prostatic intraepithelial neoplasia
HH	hypogonadotropic hypogonadism (*also* hypothalamic hamartoma)
HIF	hypoxia-inducing factor
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HOMA	homeostasis model assessment
HOS	hypo-osmotic swelling
HP	hypothalamo–pituitary
HPA	hypothalamic–pituitary–adrenal
HPF	high power field
HPT	hypothalamic-pituitary-testicular
HPT	hypothalamic–pituitary–thyroid
HPV	human papilloma virus
HRE	hormone response element
HRT	hormone replacement therapy
HSP	heat shock protein
HT	Hashimoto's thyroiditis
HyOb	hypothalamic obesity
IAD	isolated ACTH deficiency
IAPP	islet amyloid polypeptide
ICSI	intracytoplasmic sperm injection
ICU	intensive care unit
IDD	iodine deficiency disorders
IGF	insulin-like growth factor
IGF-1	insulin-like growth factor 1
IGF1R	insulin-like growth factor 1 receptor
IGHC	integrated growth hormone concentration
IGT	impaired glucose tolerance
IHH	idiopathic hypothalamic hypogonadism
IIH	idiopathic infantile hypercalcaemia
IJV	internal jugular vein
ILP	interstitial laser photocoagulation
IMRT	intensity-modulated radiotherapy
INSR	insulin receptor
IOC	International Olympic Committee
IOM	Institute of Medicine
IPEX	immunodysregulation polyendocrinopathy enteropathy X-linked
IPSC	induced pluripotent stem cells
IQ	intelligence quotient
IR	insulin resistance
IRAE	immune-related adverse effects
IRD	inner ring deiodination

ISE	ion-specific electrode
ISRS	International Stereotactic Radiosurgery Society
ITT	insulin tolerance test
IUI	intrauterine insemination
IVF	*in-vitro* fertilization
JNK	Jun N-terminal kinase
KS	Kallmann syndrome (*also* Klinefelter syndrome)
LAR	long-acting release
LATS	long-acting thyroid stimulator
LBD	ligand-binding domain
LC	liquid chromatography
LC	locus coeruleus
LCH	Langerhans cell histiocytosis
LDL	low-density lipoprotein
LDT	laterodorsal tegmental nucleus
LEPR	leptin receptor
LGBT	lesbian, gay, bisexual, and transgender
LH	luteinizing hormone
LHA	lateral hypothalamic area
LHRH	luteinizing hormone-releasing hormone
LIF	leukaemia inhibitory factor
LINAC	linear accelerator
LN	lymph node
LOH	local osteolytic hypercalcaemia
LOH	loss of heterozygosity
LS	Lugol's solution
LS	Lynch syndrome
LVMI	left ventricular mass index
MACE	major adverse cardiovascular events
MAI	mycobacterium avium intracellulare
MAP	mitogen-activated protein
MAPK	mitogen-activated protein kinase
MC2R	melanocortin-2 receptor
MCH	melanin-concentrating hormone
MCT	monocarboxylate transporter
MDD	major depressive disorder
MDT	multidisciplinary team
MEN	multiple endocrine neoplasia
MES	mineralocorticoid excess syndrome
MFB	medial forebrain bundle
MFS	metastasis-free survival
MHC	major histocompatibility complex
MHRA	Medicines and Healthcare products Regulatory Agency
MI	myocardial infarction
MIF	migration inhibition factor
MIP	minimally invasive parathyroidectomy
MMI	methyl-mercapto-imidazole
MMSE	Mini-Mental State Examination
MNG	multinodular goitre
MPB	male pattern baldness
MPN	medial preoptic nucleus
MPO	medial preoptic area
MR	magnetic resonance
MRI	magnetic resonance imaging
MSH	melanocyte-stimulating hormone
MTC	medullary thyroid cancer
MTC	medullary thyroid carcinoma
MTP	mitochondrial trifunctional protein
NAFLD	non-alcoholic fatty liver disease
NCD	non-communicable diseases
NCRP	National Council of Radiation Protection
NEFA	non-esterified fatty acids
NFPA	non-functioning pituitary adenomas
NGS	next generation sequencing
NHL	non-Hodgkin's lymphoma
NICE	National Institute for Clinical Excellence
NLS	nuclear localization sequence
NNRTI	non-nucleoside reverse transcriptase inhibitor
NOD	non-obese diabetic
NOGG	National Osteoporosis Guideline Group
NRTI	nucleoside reverse transcriptase inhibitors
NTCP	Na-taurocholate cotransporting polypeptide
NTD	N-terminal domain
NTD	N-terminal transcriptional regulation domain
NTE	neuropathy target esterase
NTG	non-toxic goitre
NTI	non-thyroidal illness
NYHA	New York Heart Association
OAR	organ at risk
OATP	organic anion transporting polypeptide
ODS	osmotic demyelination syndrome
OF	orbital fibroblasts
OGTT	oral glucose tolerance test
OHG	optico-hypothalamic gliomas
OI	osteogenesis imperfecta
OIS	oncogene-included senescence
ONH	optic nerve hypoplasia
OPG	optic pathway gliomas
OR	odds ratio
ORD	outer ring deiodination
OS	overall survival
OSA	obstructive sleep apnoea
OXT	oxytocin
PA	pituitary adenomas
PASS	pheochromocytoma of the adrenal gland scaled score
PC	pituitary carcinoma (*also* prohormone convertase)
PCB	polychlorinated biphenyls
PCOS	polycystic ovary syndrome
PCR	polymerase chain reaction
PCS	petroclival chondrosarcomas
PDX	patient-derived xenograft
PEARS	Parathyroid Epidemiology and Audit Research Study
PEG	polyethylene glycol
PET	positron emission tomography
PeVN	periventricular nucleus
PFS	progression-free survival
PH	pleckstrin homology
PI	protease inhibitor
PIA	proliferative inflammatory atrophy
PIC	pars intermedia cysts
PIN	prostatic intraepithelial neoplasia
PLAP	placental alkaline phosphatase
PPT	pedunculopontine tegmental nucleus
PREGO	presentation of Graves' orbitopathy

PROTAC	PROteolysis TArgeting Chimera
PRRT	peptide receptor radionuclide therapy
PRV	planning risk volume
PSA	prostate-specific antigen
PSIS	pituitary stalk interruption syndrome
PSMA	prostate-specific membrane antigen
PTC	papillary thyroid carcinoma
PTH	parathyroid hormone
PTHrP	parathyroid hormone-related protein
PTM	post-translational modification
PTPR	papillary tumours of the pineal region
PTSD	post-traumatic stress disorder
PTTG	pituitary tumour-transforming gene
PTV	planning target volume
PVH	periventricular hypothalamus
PVN	paraventricular nucleus
QC	quality control
QoL	quality of life
RAAS	renin angiotensin aldosterone system
RANKL	RANK ligand
RCC	Rathke's cleft cysts
RCOG	Royal College of Obstetrics and Gynaecology
RCT	randomized clinical trial
REM	rapid eye movement
REMS	risk evaluation and mitigation strategy
RL	RAS-like
RNI	reference nutrient intake
RR	relative risk
RRM	RNA recognition motifs
RS	radiosurgery
RTH	resistance to thyroid hormone
RTK	receptor tyrosine kinases
RXR	retinoid X receptor
SAH	subarachnoid haemorrhage
SCCM	Society of Critical Care Medicine
SCN	suprachiasmatic nucleus
SD	standard deviation
SDR	spontaneous dwarf rats
SDS	standard deviation score
SE	spin echo
SERM	selective oestrogen receptor modulator
SGK	serum glucocorticoid-regulated kinase
SH	subclinical hyperthyroidism
SHBG	sex-hormone-binding globulin
SIADH	syndrome of inappropriate antidiuretic hormone
SIBO	small intestinal bacterial overgrowth
SINE	selective inhibitors of nuclear export
SMR	standardized mortality ratio
SNP	single-nucleotide polymorphism
SOCS	suppressor of cytokine signalling
SOD	septo-optic dysplasia
SON	supraoptic nucleus
SPECT	single photon emission computed tomography
SRIF	somatotropin release inhibiting factor
SRL	somatostatin receptor ligands
SRS	sex reassignment surgery
SSA	senile systematic amyloidosis
SSRI	selective serotonin reuptake inhibitors
SST	short synacthen test
StAR	steroidogenic acute regulatory
SWS	slow-wave sleep
TAO	thyroid-associated orbitopathy
TBI	traumatic brain injury
TBS	trabecular bone score
TC	thyroid cancer
TCR	T-cell receptor
TDF	Tenofovir disoproxil fumarate
TED	thyroid eye disease
TGCC	testicular germ cell cancer
TGF	transforming growth factor
TH	thyroid hormone
THDP	thyroid hormone distributor protein
THR	thyroid hormone receptor
TKI	tyrosine kinase inhibitor
TMC	total motile count
TMD	transmembrane domains
TMN	tuberomammillary nucleus
TPP	thyrotoxic periodic paralysis
TR	thyroid hormone receptor
TR	thyroid receptor
TRH	thyrotropin-releasing hormone
TRT	testosterone replacement therapy
TSH	thyroid-stimulating hormone
TSHR	thyroid-stimulating hormone receptor
TSM	tuberculum sellae meningioma
TSS	transcriptional start site
TSS	transsphenoidal surgery
TVDT	tumour volume doubling time
UFC	urine-free cortisol
USI	universal salt iodization
VDDR	vitamin D-dependent rickets
VDR	vitamin D receptor
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
VEP	visual evoked potential
VFA	vertebral fracture assessment
VIP	vasoactive intestinal peptide
VLPO	ventrolateral preoptic area
VMH	ventromedial hypothalamus
VMN	ventromedial nucleus
vSPZ	ventral subparaventricular zone
VTA	ventral tegmental area
WADA	World Anti-Doping Agency
WBS	whole-body scan
WGS	whole genome sequencing
WHO	World Health Organization
WHR	waist/hip ratio
WT	Wilm's tumour
XLH	X-linked hypophosphataemia

Section Editors

Wiebke Arlt Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

John P. Bilezikian International Education and Research, Metabolic Bone Diseases Unit; Department of Medicine, Division of Endocrinology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Kristien Boelaert University Hospital Birmingham NHS Foundation Trust, Birmingham, UK

Peter Clayton Faculty of Biology, Medicine & Health, University of Manchester; Royal Manchester Children's Hospital, Manchester University NHS Foundation Trust, Machester, UK

Mathis Grossmann Department of Endocrinology Austin Health, The University of Melbourne, Heidelberg, Victoria, Australia

Ken K.Y. Ho The Garvan Institute of Medical Research and St. Vincent's Hospital, Darlinghurst, Sydney, Australia

Desmond Johnston Department of Diabetes Endocrinology and Metabolic Medicine, Imperial College London, UK

John Newell- Price Department of Oncology and Metabolism, University of Sheffield; Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

David Ray Division of Diabetes, Endocrinology & Gastroenterology, Manchester Institute for Collaborative Research on Ageing, Lydia Becker Institute of Immunology and Inflammation, University of Manchester, UK

Robert K. Semple Professor of Translational Molecular Medicine, University of Edinburgh Centre for Cardiovascular Science, Queen's Medical Institute, Edinburgh, UK

James Shaw Newcastle Diabetes Centre, Newcastle University, Freeman Hospital, UK

Guy T'Sjoen Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

John A.H. Wass Professor of Endocrionolgy, Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Wilmar M. Wiersinga Department of Endocrinology and Metabolism, University of Amsterdam, Amsterdam, The Netherlands

Catherine Williamson Guy's and St Thomas' NHS Foundation Trust; King's College London, UK

Frederick Wu Division of Diabetes, Endocrinology & Gastroenterology, Faculty of Biology, Medicine and Health, University of Manchester, UK

Bulent Okan Yildiz Division of Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, Hacettepe University, Turke

Contributors

Rochan Agha-Jaffar Consultant, Imperial College Healthcare NHS Trust, London, UK

Shahzada Ahmed Department of Ear, Nose and Throat, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

S. Faisal Ahmed University of Glasgow, Royal Hospital for Sick Children, Glasgow, UK

Salma R. Ali Developmental Endocrinology Research Group, School of Medicine, Dentistry and Nursing, University of Glasgow; and Office for Rare Conditions, Royal Hospital for Children and Queen Elizabeth University Hospital, Glasgow, UK

Zain Al-Safi Department Obstetrics & Gynecology, University of California, Los Angeles, USA

Stephanie A. Amiel Department of Diabetes, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, UK

Nobuyuki Amino Amino Thyroid Research Laboratory, Osaka, Japan

Panagiotis Anagnostis Unit of Reproductive Endocrinology, First Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Vivian Anastasiou School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Bradley D. Anawalt Professor and Vice Chair of Medicine, University of Washington, Seattle, WA, USA

Stig Andersen Department of Geriatrics and Clinical Medicine, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

Bo Angelin Metabolism Unit and Integrated CardioMetabolic Centre, Department of Endocrinology, Metabolism & Diabetes, Department of Medicine, Karolinska Institute at Karolinska University Hospital Huddinge, Stockholm, Sweden

Leen Antonio Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven; Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium

Naoko Arata Division of Internal Medicine, National Center for Child Health and Development, Tokyo, Japan

Jon Arcelus Nottingham Centre for Transgender Health, Nottingham University School of Medicine, Nottingham, UK

Laleh Ardeshirpour Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Wiebke Arlt William Withering Chair of Medicine, Institute of Metabolism and Systems Research, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, Birmingham, UK

Eleni Armeni ENETS Centre of Excellence, Royal Free Hospital, London, UK

Baris Ata Department of Obstetrics and Gynaecology, School of Medicine, Koc University, Ankara, Turkiye

Richard J. Auchus Division of Metabolism, Endocrinology, and Diabetes, Department of Internal Medicine, Department of Pharmacology, and the Program for Disorders of Sexual Development, University of Michigan, Ann Arbor, MI, USA

Anna Aulinas Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona. Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Ricardo Azziz Center for Androgen Related Disorders, Cedars-Sinai Medical Center, and Department of Obstetrics and Gynecology and Department of Medicine, The David Geffen School of Medicine at UCLA, Los Angeles, USA

Muriel Babey Diabetes, Endocrinology and Metabolism Fellow, Division of Endocrinology and Metabolism, Department of Medicine, University of California, San Francisco, CA, USA

Corin Badiu Department of Endocrinology, C. I. Parhon National Institute, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania

Michael N. Badminton Department of Medical Biochemistry and Immunology, University Hospital of Wales; School of Medicine, Cardiff University, Wales, UK

Clifford J. Bailey Aston Pharmacy School, School of Life and Health Sciences, Aston University, Birmingham, UK

Adam H. Balen Leeds Teaching Hospitals Leeds Fertility, Seacroft Hospital, Leeds, UK

Stephen G. Ball The Medical School, Newcastle University, and Newcastle Hospitals NHS Trust, Newcastle, UK

Irina Bancos Mayo Clinic, Rochester, MN, USA

Indraneel Banerjee Royal Manchester Children's Hospital, University of Manchester, UK

María-José Barahona Department of Endocrinology/Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona; Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Edouard Bardou-Jacquet Liver Unit, Pontchaillou University Hospital, Rennes, France

Luigi Bartalena Department of Medicine and Surgery, University of Insubria, Endocrine Unit, ASST dei Sette Laghi, Varese, Italy

Shehzad Basaria Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Waiel Bashari Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Ayat Bashir Northern Institute Forfor Cancer Research, Newcastle-upon-Tyne, UK

Fulvio Basolo Department of Surgery, Medical, Molecular, and Critical Area Pathology, University of Pisa, Pisa, Italy

J.H. Duncan Bassett Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Rachel L. Batterham Centre for Obesity Research, Department of Medicine, University College London; Bariatric Centre for Weight Management and Metabolic Surgery, University College London Hospital, NHS Foundation Trust; National Institute for Health Research, Biomedical Research Centre, University College London Hospital, London, UK

Adrian Bauman Sydney School of Public Health, Sydney University, Australia

Paul Bech Department of Investigative Medicine, Hammersmith Hospital, Imperial College, London, UK

Finn N. Bennedbæk Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen; Department of Endocrinology, Herlev and Gentofte Hospital, Herlev, Denmark

Emelie Benyi Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Sarah L. Berga Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT, USA

Jérôme Bertherat Endocrinology Department, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris; Endocrinology, Diabetes and Metabolism Department, INSERM U1016, Institut Cochin and Université Paris Descartes, Paris, France

Rudi Beschorner Diagnostic and Interventional Neuroradiology, Department of Radiology, University of Tuebingen, Tuebingen, Germany

Rachel Besser Consultant in Paediatric Diabetes and Endocrinology at Oxford University Hospitals NHS Foundation Trust; and Honorary Senior Clinical Lecturer at the University of Oxford, Oxford, UK

Ragini Bhake Department of Diabetes and Endocrinology, University Hospital of Leicester NHS Trust, Leicester Royal Infirmary, Leicester, UK

N.R. Biermasz Department of Medicine, Division of Endocrinology, and Center for Endocrine Tumors, Leiden, The Netherlands

John P. Bilezikian International Education and Research, Metabolic Bone Diseases Unit; Department of Medicine, Division of Endocrinology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Anita Boelen Department of Endocrinology and Metabolism, Academic Medical Center, University of Amsterdam, AZ Amsterdam, Netherlands

Steen Joop Bonnema Professor and Consultant, University of Southern Denmark, Department of Endocrinology, Odense University Hospital, Odense, Denmark

Fidéline Bonnet-Serrano Hormonal Biology Department, Cochin Hospital, APHP; Paris Descartes University, Paris, France

Jean-François Bonneville Department of Endocrinology, CHU Sart-Tilmann, Liège, Belgium

Charlotte Boughton University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Sheerazed Boulkroun INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Andrew J.M. Boulton Division of Endocrinology, Diabetes & Gastroenterology, Faculty of Medicine, Biology and Health, University of Manchester; Consultant Physician, Manchester Royal Infirmary, Manchester, UK; Visiting Professor, Diabetes Research Institute, University of Miami, Miami, FL, USA

Walter Pierre Bouman Nottingham Centre for Transgender Health, Nottingham University School of Medicine, Nottingham, UK

Frank Lee Bowling Podiatric Consultant Surgery, Division of Surgery, Manchester Royal Infirmary and Reader in Translational Medicine; Division of Endocrinology, Diabetes & Gastroenterology, Faculty of Medicine, Biology and Health, University of Manchester, Manchester, UK

Gurkan Bozdag Department of Obstetrics and Gynaecology, School of Medicine, Hacettepe University, Istanbul, Turkiye

Georg Brabant FRCP Exp. & Clin Endocrinology, University of Lübeck, Germany

Glenn D. Braunstein Department of Medicine, Cedars-Sinai Medical Center and University of California, Los Angeles, CA, USA

Gemma V. Brierley University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science; National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Augustin Brooks Bournemouth Diabetes and Endocrine Centre, Royal Bournemouth Hospital, Bournemouth, UK

Morris J. Brown Barts and the London School of Medicine; St Bartholomew's Hospital, London, UK

Gary Butler Consultant in Paediatric and Adolescent Endocrinology, University College London Hospital, Honorary Clinical Professor in Child and Adolescent Health, UCL Great Ormond Street Institute of Child Health, London, UK

Marilia Calcia Diabetes Psychiatry and Psychology Service, King's College Hospital, London, UK

Chris Callaghan Guy's and St Thomas' NHS Foundation Trust, London, UK

Fiona Campbell Children's Diabetes Centre, Leeds Children's Hospital, Leeds Teaching Hospitals, Leeds, UK

Ernesto Canalis Departments of Orthopaedic Surgery and Medicine, and The UConn Musculoskeletal Institute, UConn Health, Farmington, CT, USA

Martyn E. Caplin Department of Gastroenterology, Royal Free Hospital, London, UK

Cesare Carani University of Modena and Reggio Emilia, Modena, Italy

Jean-Claude Carel Pediatric Endocrinology Diabetology Department, Reference Center for Growth and Development Endocrine Diseases, Université de Paris; Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1141, Assistance Publique-Hôpitaux de Paris, Robert-Debré University Hospital, Paris, France

Thomas O. Carpenter Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Paul V. Carroll Consultant Endocrinologist and Honorary Senior Lecturer, Department of Endocrinology, Guy's & St. Thomas' NHS Foundation Trust; Faculty of Life Sciences & Medicine, King's College London, UK

Ruth T. Casey Department of Medical Genetics, University of Cambridge, Cambridge, UK

Manuela Cerbone University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Filomena Cetani Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Ghassan Chamseddine School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Li F. Chan Centre for Endocrinology, William Harvey Research Institute, London, UK

R. Jeffrey Chang Division of Reproductive Endocrinology and Infertility, University of California, San Diego, CA, USA

Krishna Chatterjee Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, UK

Steven Chernausek Department of Pediatrics, Section of Diabetes and Endocrinology, University of Oklahoma Health Science Center, College of Medicine, Oklahoma City, OK, USA

C.L. Cheung Department of Therapeutics, The University of Hong Kong, Hong Kong, China

Jennifer Chin Department of Obstetrics, Gynecology, and Women's Health, University of Hawaii, Honolulu, HI, USA

Luca Chiovato Department of Endocrinology, University of Pisa, Pisa, Italy

Yeoungjee Cho Department of Nephrology, Princess Alexandra Hospital, Brisbane, Australia

Pratik Choudhary Diabetes Department, King's College London, London, UK

Frank Claessens Department of Cellular and Molecular Medicine, Laboratory of Molecular Endocrinology, KU Leuven, Leuven, Belgium

Orlo H. Clark Endocrine Surgery Section, Department of Surgery, University of California, San Francisco, CA, USA

Robert Clarke Department of Epidemiology and Public Health, University of Oxford, Oxford, UK

Bart L. Clarke Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, Minnesota, USA

Philip Clarke Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford; School of Population Health, University of Melbourne, Melbourne, UK

Dominique Clement Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Kings College Hospital, London, UK

Stephen Colagiuri Boden Institute of Obesity, Nutrition, Exercise & Eating Disorders, University of Sydney, Sydney, Australia

Trevor Cole Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

Anthony P. Coll University Lecturer/Honorary Consultant Physician, University of Cambridge Metabolic Research Laboratories, MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Cambridge, UK

Cristina Colom Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona; Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Sarah J. Conner Centre for Human Reproductive Science, IMSR, College of Medical and Dental Sciences, University of Birmingham; ICS, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK Consultant Gastroenterologist, Kings Health Partners ENETS Centre of Excellence, London, UK

Gerard S. Conway Professor of Reproductive Endocrinology, Institute for Women's Health, University College London, UK

Giovanni Corona Endocrinology Unit, Medical Department, AziendaUsl Bologna Maggiore-Bellaria Hospital, Bologna, Italy

Fernanda A. Correa Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular (LIM/42) do Hospital das Clinicas, Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Jeremy Cox Department of Endocrinology and Diabetes, St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Richard M. Cubbon Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT laboratories, University of Leeds, Leeds, UK

Janine A. Danks Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Australia, Australia

Mehul T. Dattani University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Chantal Daumerie Department of Endocrinology, Université catholique de Louvain, Cliniques Universitaires Saint-Luc, Brussels, Belgium

Terry F. Davies Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Melanie J. Davies Diabetes Research Centre, University of Leicester, Leicester General Hospital, Leicester, UK

Julian Davis Department of Endocrinology, Manchester Royal Infirmary; School of Medical Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK

Colin Dayan Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK

Chloë De Roo Department of Reproductive Medicine, Ghent University Hospital, Corneel Heymanslaan, Ghent, Belgium

Sunita M.C. De Sousa Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Justine Defreyne Department of Endocrinology, Center for Sexology and Gender, Ghent University Hospital, Corneel Heymanslaan, Ghent, Belgium

Yves Deugnier Liver Unit, Pontchaillou University Hospital, Rennes, France

Ketan Dhatariya Consultant in Diabetes, Endocrinology and General (Internal) Medicine Honorary Professor, Norwich Medical School, Elsie Bertram Diabetes Centre, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK

Whaljit Dhillo Department of Investigative Medicine, Hammersmith hospital, Imperial College, London, UK

Evanthia Diamanti-Kandarakis Department of Endocrinology and Diabetes, HYGEIA Hospital, Athens, Greece

Ruxandra Dobrescu Parhon National Endocrinology Institute in Bucharest, Bucharest, Romania

Adrian Dobs Divisions of Endocrinology, Metabolism and Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Martin Drage Guy's and St Thomas' NHS Foundation Trust, Great Ormond Street Hospital, London, UK

William M. Drake Department of Endocrinology, St Bartholomew's Hospital; Barts and the London School of Medicine, London, UK

Daniel Dumesic Department Obstetrics & Gynecology, University of California, Los Angeles, USA

Richard Eastell Academic Unit of Bone Metabolism, Metabolic Bone Centre, Northern General Hospital, Sheffield, UK

David A. Ehrmann The University of Chicago, Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Chicago, IL, USA

Mustapha El Lakis Endocrine Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Rossella Elisei Department of Endocrinology and Metabolism, University of Pisa, Pisa, Italy

Jackie Elliott Sheffield Teaching Hospitals Foundation Trust, University of Sheffield, Sheffield, UK

Cara E. Ellis Laboratory of Molecular and Cellular Medicine, Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Charis Eng Genomic Medicine Institute, Lerner Research Institute; Taussig Cancer Institute, Cleveland Clinic; Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine; Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Cleveland, OH, USA

Ulrike Ernemann Department of Neuropathology, University of Tuebingen, Tuebingen, Germany

Mohammed Eslam Storr Liver Centre, The Westmead Institute for Medical Research, University of Sydney and Westmead Hospital, Westmead, Australia

I. Sadaf Farooqi University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Maritza T. Farrant University of Washington School of Medicine, Seattle, WA

Alan P. Farwell Division of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts, USA

Fabio R. Faucz Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Ulla Feldt-Rasmussen Department of Medical Endocrinology and Metabolism, Rigshospitalet, Copenhagen University, Copenhagen, Denmark

Fabio Fernandes-Rosa INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité, Paris, France

Luca Ferraro School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Naomi Fersht Department of Oncology, University College London Hospital, London, UK

Nicholas Finer National Centre for Cardiovascular Prevention and Outcomes, Institute of Cardiovascular Science, University College London, London, UK; Senior Principle Clinical Scientist, Global Medical Affairs Management, Novo Nordisk A/S

Sarah B. Fisher Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Anneliese Flatt Newcastle University, Newcastle-upon-Tyne, UK

Karin Frank-Raue Endokrinologische Gemeinschaftspraxis, Heidelberg, Germany

Steve Franks Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College; St. Mary's and Hammersmith Hospitals, London, UK

Bernard Freudenthal Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Dagmar Führer Professor of Medicine, Department of Endocrinology, Diabetes and Metabolism, University Hospital Essen, University of Duisburg-Essen, Germany

John W. Funder Prince Henry's Institute of Medical Research, Monash Medical Centre, Clayton, Australia

John S. Fuqua Professor of Clinical Pediatrics, Division of Pediatric Endocrinology and Diabetology, Indiana University School of Medicine, Riley Hospital for Children, Indianapolis, IN, USA

Thiago Gagliano-Jucá Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Sonya Galcheva Department of Paediatrics, Varna Medical University/University Hospital 'St. Marina', Varna, Bulgaria

Alessandra Gambineri Endocrinology Unit, Department of Medical and Surgical Sciences, St Orsola-Malpighi Hospital, Alma Mater University of Bologna, Bologna, Italy

Hoong-Wei Gan University College London, Great Ormond Street Institute of Child Health and Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Chris Garrett Diabetes and Mental Health Group, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

Amalia Gastaldelli Institute of Clinical Physiology, CNR, Pisa, Italy

Minu George Department of Pediatrics, Section of Diabetes and Endocrinology, University of Oklahoma Health Science Center, College of Medicine, Oklahoma City, OK, USA

Jacob George Storr Liver Centre, The Westmead Institute for Medical Research, University of Sydney and Westmead Hospital, Westmead, Australia

Nitin Narayan Gholap Consultant in Diabetes and Endocrinology, Departments of Diabetes, Endocrinology, and Metabolism, University Hospitals Coventry and Warwickshire, Coventry, UK; Diabetes Research Centre, University of Leicester, Leicester, UK

James Gibney Tallaght University Hospital, Dublin, Ireland

Helena K. Gleeson Department of Endocrinology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; Diabetes Research Centre, University of Leicester, Leicester, UK

Luigi Gnudi School of Cardiovascular Medicine & Science, King's College London, London, UK

David Goltzman Metabolic Disorders and Complications (MeDiC), Room # EM1.3220, RI-McGill University Health Centre, Glen Site, 1001 Décarie Blvd, Montréal, QC H4A 3J1, Canada

Phillip Gorden National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

Dimitrios G. Goulis Unit of Reproductive Endocrinology, First Department of Obstetrics and Gynecology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Hans Graf Professor of Endocrinology at the Federal University of Paraná, Brazil

Antonio R.M. Granata Department of Medical Specialties, Azienda Ospedaliero-Universitaria of Modena, Modena, Italy

Claus H. Gravholt Department of Endocrinology, Aarhus University, Aarhus, Denmark

Alice Greenhalgh Institute of Cancer Sciences, University of Manchester, Manchester, UK

Louise C. Gregory Genetics and Genomic Medicine Programme, UCL Great Ormond Street Institute of Child Health, London, UK

Steven K. Grinspoon Professor of Medicine, Harvard Medical School; MGH Program in Nutritional Metabolism and Neuroendocrine Unit, Boston, MA, USA

Ashley B. Grossman Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Royal Free Hospital, London, UK

Mathis Grossmann Professor, Department of Medicine (Austin Health), The University of Melbourne, Department of Endocrinology, Austin Health, Heidelberg, Australia

Peter Guest X-ray Department, Radiology, University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital Medical Centre, Birmingham, UK

Jan Gunst Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven, Belgium

Mark Gurnell Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Stefanie Hahner Department of Medicine, Endocrinology and Diabetology, University Hospital Würzburg, Germany

Peter Hammond Harrogate and District NHS Foundation Trust, Lancaster Park Road, Harrogate, North Yorkshire, UK

David J. Handelsman ANZAC Research Institute, and Department of Andrology, Concord Hospital, Sydney, Australia

Fadil M. Hannan Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, Faculty of Health & Life Sciences, University of Liverpool, Liverpool; Academic Endocrine Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Andrew Hattersley Professor of Molecular Genetics, College of Medicine and Health, University of Exeter; and Honorary Consultant in Diabetes and Endocrinology, The Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Laszlo Hegedüs Professor and Consultant, University of Southern Denmark, Department of Endocrinology, Odense University Hospital, Odense, Denmark

Simon Heller School of Medicine and Biomedical Sciences, University of Sheffield, Sheffield, UK

Christel Hendrieckx School of Psychology, Deakin University, Geelong; The Australian Centre for Behavioural Research in Diabetes, Diabetes Victoria, Melbourne, Victoria, Australia

Geoffrey N. Hendy† Calcium Research Laboratory, Royal Victoria Hospital, Montreal, Canada

David Henley Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Perth, Australia

Claire E. Higham Consultant Endocrinologist, Christie Hospital NHS Foundation Trust; University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

Olaf Hiort Division of Paediatric Endocrionolgy and Diabetes, Department of Paediatric and Adolescent Medicine, University of Lübeck, Germany

Ken K.Y. Ho The Garvan Institute of Medical Research, St. Vincent's Hospital and the University of New South Wales, Sydney, Australia

Anita C.S. Hokken-Koelega Erasmus University Medical Center, Rotterdam, The Netherlands

Richard I.G. Holt Human Development and Health Academic Unit, Faculty of Medicine, University of Southampton Faculty of Medicine, Southampton, UK

Jürgen Honegger Department of Neurosurgery, University of Tuebingen, Tuebingen, Germany

Roman Hovorka University of Cambridge Metabolic Research Laboratories, Wellcome Trust- MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Wenyu Huang Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Ilpo Huhtaniemi Institute of Reproductive and Developmental Biology, Imperial College London; Institute of Biomedicine, Research Centre for Integrative Physiology and Pharmacology, University of Turku

Khalid Hussain Department of Paediatric Medicine, Division of Endocrinology, Sidra Medicine, Doha, Qatar

Daniela Ibarra-Gasparini Endocrinology Unit, Department of Medical and Surgical Sciences, St Orsola-Malpighi Hospital, Alma Mater University of Bologna, Bologna, Italy

Radha Indusekhar Department of Obstetrics and Gynaecology, North Staffordshire Hospital, The University Hospital of North Midlands NHS Trust, Royal Stoke University Hospital, Staffordshire, UK

Khalida Ismail Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

Peter Jacob Diabetes Department, King's College London, London, UK

Susie Jacob Leeds Centre for Reproductive Medicine, Seacroft Hospital, Leeds, UK

Barbara Jarzab M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Holger Jäschke Department of Endocrinology, Diabetes and Metabolism, University Hospital Essen, University of Duisburg-Essen, Germany

Friedrich C. Jassil Centre for Obesity Research, Department of Medicine, University College London; Bariatric Centre for Weight Management and Metabolic Surgery, University College London Hospital, NHS Foundation Trust, London, UK

Channa Jayasena Departments of Medicine and Andrology, Imperial College London, Hammersmith Hospital; Department of Obstetrics & Gynecology, St Mary's Hospital, London, UK

David W. Johnson Department of Nephrology, Princess Alexandra Hospital, Brisbane, Australia

Peter M. Jones Department of Diabetes, King's College London, London, UK

Alexander A.L. Jorge Unidade de Endocrinologia Genética, Laboratório de Endocrinologia Celular e Molecular (LIM/25), Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

Jens O.L. Jørgensen Department of Endocrinology and Internal Medicine, Aarhus University Hospital, Aarhus, Denmark

Anne Jouinot Oncology Department, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris; Endocrinology, Diabetes and Metabolism Department, INSERM U1016, Institut Cochin and Université Paris Descartes, Paris, France

Eleni A. Kandaraki Department of Endocrinology and Diabetes, HYGEIA Hospital, Athens, Greece

Bliss Kaneshiro Department of Obstetrics, Gynecology and Women's Health, University of Hawaii, Honolulu, HI, USA

Niki Karavitaki Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Jean-Marc Kaufman Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

Mark T. Kearney Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT laboratories, University of Leeds, Leeds, UK

Electron Kebebew Department of Surgery and Stanford Cancer Institute, Stanford University, Stanford, CA, USA

Brian Keevil Clinical Biochemistry, Manchester University Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Bernard Khoo ENETS Centre of Excellence, Royal Free Hospital, London, UK

Kamlesh Khunti Diabetes Research Centre; Leicester Clinical Trials Unit, University of Leicester, Leicester Diabetes Centre; NIHR Collaborations for Leadership in Applied Health Research and Care, Leicester, UK

Timothy J. Kieffer Laboratory of Molecular and Cellular Medicine, Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Na Ri Kim Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven, Leuven, Belgium

Jackson C. Kirkman-Brown Centre for Human Reproductive Science, IMSR, College of Medical and Dental Sciences, University of Birmingham; Birmingham Women's Fertility Centre, Birmingham Women's & Children's NHS Foundation Trust, Birmingham, UK

Anne Klibanski Neuroendocrine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Daniel Klink ZNA Queen Paola Children's Hospital Antwerp, Antwerp, Belgium

Joanna Klubo-Gwiezdzinska Metabolic Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

Karen E. Knudsen Departments of Cancer Biology, Medical Oncology, Urology, Radiation Oncology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Josef Köhrle Institut für Experimentelle Endokrinologie und Endokrinologisches Forschungs-Centrum der Charité EnForCé, Charité Universitätsmedizin Berlin, CVK, Berlin, Germany

Márta Korbonits Centre for Endocrinology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London, UK

Tim I.M. Korevaar Department of Internal Medicine, Academic Center for Thyroid Diseases, Erasmus Medical Center, Rotterdam, The Netherlands

Olympia Koulouri Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, University of Cambridge and National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Kalman Kovacs Department of Laboratory Medicine, Division of Pathology; The Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

Jolanta Krajewska M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Jelena Kravarusic Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Nils P. Krone Academic Unit of Child Health, Department of Oncology and Metabolism, University of Sheffield, Sheffield Children's Hospital, Sheffield, UK

Annie W.C. Kung Department of Medicine, The University of Hong Kong, Hong Kong, China

Robin H. Lachmann Charles Dent Metabolic Unit, National Hospital for Neurology and Neurosurgery, London, UK

Lies Langouche Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium

Michael Laßmann Klinik und Poliklinik für Nuklearmedizin, Universitätsklinikum Würzburg, Würzburg, Germany

Francesco Latrofa Department of Clinical and Experimental Medicine, Endocrinology Unit I, University of Pisa; University Hospital of Cisanello, Pisa, Italy

Deepthi Lavu Department of Obstetrics and Gynaecology, Health Education West Midlands, Birmingham, UK

Edward R. Laws Jr. Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

John H. Lazarus Thyroid Research Group, Cardiff University, Cardiff, Wales, UK

Juliane Léger Pediatric Endocrinology Diabetology Department, Reference Center for Growth and Development Endocrine Diseases, Université de Paris; Institut National de la Santé et de la Recherche Médicale (INSERM), UMR 1141, Assistance Publique-Hôpitaux de Paris, Robert-Debré University Hospital, Paris, France

Michael A. Levine Professor of Pediatrics and Medicine, Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia and Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

Miles J. Levy Department of Diabetes and Endocrinology, University Hospital of Leicester NHS Trust, Leicester Royal Infirmary, Leicester, UK

Rossella Libè Endocrinology Department, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France

Steven A. Lietman Department of Orthopaedic Surgery, Geisinger Medical Center, Danville, PA, USA

Stafford L. Lightman Henry Wellcome Laboratories for Integrative Neuroscience & Endocrinology, University of Bristol, Bristol, UK

Kok Haw Jonathan Lim Department of Medical Oncology, The Christie NHS Foundation Trust, Manchester, UK

Siew Lim Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Jonathan Z.M. Lim University of Liverpool, Liverpool, UK

Kirstie Lithgow Clinical Research Fellow, Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Lynn Loriaux Department of Medicine, Oregon Health Sciences University, Portland, Oregon, USA

Sandra Lowe Royal Hospital for Women; School of Women's and Children's Health, University of New South Wales, Sydney, Australia

Thomas Lung The George Institute for Global Health, University of New South Wales, Kensington, Australia

Markus Luster Department of Nuclear Medicine, University Hospital Marburg, Marburg Germany

Djuro Macut Clinic of Endocrinology, Diabetes and Metabolic Diseases, and Faculty of Medicine, University of Belgrade, Belgrade, Serbia

Mario Maggi Andrology and Sexual Medicine Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy

Eamonn R. Maher Department of Medical Genetics, University of Cambridge, Cambridge, UK

Ingrid Y.F. Mak Neuroendocrine Tumour Unit, ENETS Centre of Excellence, Royal Free Hospital, London, UK

Noemi Malandrino National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

Shreya Malhotra Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Crystal Man Ying Lee Boden Institute of Obesity, Nutrition, Exercise & Eating Disorders, University of Sydney, Sydney; School of Psychology and Public Health, La Trobe University, Melbourne, Australia

Prakash Manoharan Department of Radiology, The Christie NHS Foundation Trust, Manchester, UK

Wasat Mansoor Department of Medical Oncology, The Christie NHS Foundation Trust, Manchester, UK

Claudio Marcocci Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

John C. Marshall Center for Research in Reproduction, University of Virginia, Charlottesville, VA, USA

Sally M. Marshall Translational and Clinical Research Institute (Diabetes), Faculty of Clinical Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

George Mastorakos Endocrine Unit, University of Athens, Medical School, 'Aretaieion' Hospital, Athens, Greece

Alvin M. Matsumoto Department of Medicine, Division of Gerontology and Geriatric Medicine, University of Washington, Seattle VA Puget Sound Health Care System, Seattle, Washington, USA

Gherardo Mazziotti Endocrine Unit, ASST Carlo Poma, Mantua, Italy

Christopher R. McCartney Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, VA, USA

Ann McCormack Department of Endocrinology, St Vincent's Hospital; Garvan Institute of Medical Research; University of New South Wales, Darlinghurst, NSW, Australia

Karim Meeran Department of Investigative Medicine, Imperial College London, London, UK

Shlomo Melmed Pituitary Center, Cedars-Sinai Medical Center, Los Angeles, California

Francesca Menconi Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Radu Mihai Department of Endocrine Surgery, Churchill Cancer Centre, Oxford University Hospitals Foundation Trust, Oxford, UK

Paul D. Miller Distinguished Clinical Professor of Medicine, University of Colorado Health Sciences Center, Medical Director, Colorado Center for Bone Research at Panorama Orthopedics and Spine Center, Golden, Colorado, USA

Karen K. Miller Neuroendocrine Unit, Massachussetts General Hospital; Faculty of Medicine at Harvard Medical School, Boston, MA, USA

Mark E. Molitch Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

Carla Moran Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, UK

Lisa Moran National Health and Medical Research Council Centre for Research Excellence in PCOS, University of Adelaide, Adelaide; Monash Centre for Health Research and Implementation, Monash Public Health and Preventive Medicine, Monash University and Monash Health, Melbourne, Vic., Australia

John F. Morris Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford, UK

Laura Morrison Faculty of Medicine, Cancer Center, Yale School of Medicine, New Haven, CT, USA

Calum Moulton Department of Psychological Medicine, King's College London, UK

Aya Mousa Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Sezcan Mumusoglu Department of Obstetrics and Gynecology, School of Medicine, Hacettepe University, Ankara, Turkey

Alia Munir Sheffield Teaching Hospitals NHS Foundation Trust, University of Sheffield, Sheffield, UK

Helen R. Murphy Cambridge University Hospitals NHS Foundation Trust, Cambridge; Women's Health Academic Centre, Division of Women's and Children's Health, King's College London, London; Norwich Medical School, University of East Anglia, Norwich, UK

Robert D. Murray Consultant Endocrinologist, Leeds Teaching Hospital NHS Trust; Honorary Associate Professor, Leeds Institute for Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, UK

Sonia Nagi Faculty of Medicine, National Institute of Neurology, University of Tunis El Manar, Tunis-Tunisia

Radhika Narla Department of Medicine, Division of Metabolism, Endocrinology and Nutrition, University of Washington, Seattle VA Puget Sound Health Care System, Seattle, Washington, USA

Melissa Nataatmadja Department of Nephrology, Sunshine Coast University Hospital and Health Service; Faculty of Medicine, University of Queensland, Brisbane, Australia

John Newell-Price Department of Oncology and Metabolism, University of Sheffield; Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Markella Nezi Endocrine Unit, University of Athens, Medical School, 'Aretaieion' Hospital, Athens, Greece

Vania Nosé Professor of Pathology, Harvard Medical School, Director of Anatomic and Molecular Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA

Birte Nygaard Department of Clinical Medicine, Herlev-Gentofte Hospital, University of Copenhagen, Denmark

Shaughn O'Brien Department of Obstetrics and Gynaecology, Keele University, Keele; Nuffield Health North Staffordshire Hospital, Newcastle-under-Lyme, UK

Stephen O'Rahilly Department of Endocrinology, Beaumont Hospital Dublin and Royal College of Surgeons in Ireland (RCSI), Ireland

Michael W. O'Reilly Department of Endocrinology, Beaumont Hospital Dublin and Royal College of Surgeons in Ireland (RCSI), Ireland

Kutluk Oktay Innovation Institute for Fertility Preservation and In Vitro Fertilization, New York, NY; Laboratory of Molecular Reproduction and Fertility Preservation, Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Nick Oliver Department of Metabolism, Digestion and Reproduction, Imperial College Healthcare NHS Trust, London, UK

Richard A. Oram Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK

Jacques Orgiazzi Université Claude-Bernard Lyon-1, and Hospices Civils de Lyon, Service d'Endocrinologie-Diabétologie-Maladies Métaboliques, Groupement Hospitalier Sud, Pierre-Bénite, France

Henrik Oster Institute for Neurobiology, University of Lübeck, CBBM, Germany

Katharine R. Owen Associate Professor, University of Oxford, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, Oxford, UK

Catherine J. Owen Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, UK

Furio Pacini Departments of Internal Medicine, Endocrinology and Metabolism, and Biochemistry, University of Siena, Siena, Italy

Stephanie T. Page University of Washington School of Medicine, Seattle, WA

Fausto Palazzo Hammersmith Hospital and Imperial College London, UK

Maria Papagianni Pediatric Endocrinology Unit, Third Department of Pediatrics, 'Hippokrateion' General Hospital of Thessaloniki, Aristotle University of Thessaloniki School of Medicine, Thessaloniki, Greece

Socrates E. Papapoulos Department of Endocrinology and Metabolic Diseases, Leiden University Medical Centre, Leiden, The Netherlands

Thomas G. Papathomas Clinician Scientist in Molecular Pathology, Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

Paolo Parini Division of Clinical Chemistry, Department of Laboratory Medicine, Metabolism Unit, Department of Endocrinology, Metabolism & Diabetes, Department of Medicine, Karolinska Institute at Karolinska University Hospital and Inflammation and Infection Theme, Karolinska University Hospital Huddinge, Stockholm, Sweden

Stavroula A. Paschou Assistant Professor of Endocrinology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

Peysh A. Patel Leeds Institute for Cardiovascular and Metabolic Medicine, LIGHT laboratories, University of Leeds, Leeds, UK

Kashyap A. Patel Wellcome Turst Clinical Research Career Development Felllow, College of Medicine and Health, University of Exeter; and Honorary Consultant in Diabetes and Endocrinology, The Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

Gilberto Paz-Filho Medical & Scientific Affairs, Janssen Australia and New Zealand

Michael Pazianas Senior Research Associate, Institute of Musculoskeletal Sciences, Oxford University, Oxford, UK

Elizabeth N. Pearce Division of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts, USA

Simon H.S. Pearce Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, UK

Inge Bülow Pedersen Department of Endocrinology and Medicine, Aalborg Hospital, Århus University Hospital, Aalborg, Denmark

Robin P. Peeters Department of Internal Medicine, Academic Center for Thyroid Diseases, Erasmus Medical Center, Rotterdam, The Netherlands

David L. Penn Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Nancy D. Perrier Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Shanta J. Persaud Department of Diabetes, King's College London, London, UK

John R. Petrie Professor of Diabetic Medicine, Institute of Cardiovascular and Medical Sciences, BHF Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK

John Pickup Diabetes Research Group, King's College London School of Medicine, Guy's Hospital, London, UK

Duarte Pignatelli Faculty of Medicine of Porto and I3S - Institute of Investigation and Innovation in Health, University of Porto, and Centro Hospitalar Universitário S João, Porto, Portugal

Terhi Piltonen Department of Obstetrics and Gynecology, PEDEGO Research Unit, Medical Research Center, Oulu University Hospital, University of Oulu, Oulu, Finland

Biljana Popovic-Todorovic Centre for Reproductive Medicine, UZ Brussels, Belgium

Kris Poppe Endocrine Unit Centre Hospitalier Universitaire Saint Pierre; Université Libre de Bruxelles (ULB), Brussels, Belgium

Iulia Potorac Endocrinology, CHU Sart-Tilmann, Liège, Belgium

L.D.K.E. Premawardhana Thyroid Research Group, Cardiff University, Cardiff, Wales, UK

Jackie Price Public Health Sciences Section, Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, UK

Hermione Price Research and Development Department, Tom Rudd Unit, Moorgreen Hospital, Southampton, UK

Margit G. Proescholdt University of Basel, Psychiatric Hospital, Division of Substance Use Disorders, Switzerland

Dalal Qanaq School of Life Course Sciences, Department of Diabetes, King's College London, London, UK

Shwetha Ramachandrappa Centre for Endocrinology, William Harvey Research Institute, London, UK

Giulia Rastrelli Andrology and Sexual Medicine Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy

Friedhelm Raue Endokrinologische Gemeinschaftspraxis, Heidelberg, Germany

David W. Ray Division of Digestion, Endocrinology and Metabolism, The University of Manchester, Manchester; NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital; Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

Gerry Rayman The Ipswich Diabetes Centre and Research Unit, Ipswich Hospital NHS Trust, Suffolk, UK

Narendra Reddy Department of Endocrinology, University Hospital of Leicester NHS Trust, Leicester, UK

Caroline S. Repetti Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Samantha J. Richardson School of Science and School of Health and Biomedical Sciences, RMIT University, Melbourne, Australia

Guido Rindi Section of Anatomic Pathology, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Roma, Italia; Section of Anatomic Pathology, Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italia; Roma European NeuroEndocrine Tumor Society (ENETS) Center of Excellence, Roma, Italia

Rachel E. Roberts Department of Obstetrics and Gynaecology, Queen Charlotte's and Chelsea Hospital, Imperial College Healthcare NHS Trust, London, UK

Cemre Robinson Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA; Assistant Professor of Pediatrics, Division of Endocrinology and Diabetes, Icahn School of Medicine at Mount Sinai, New York, USA

Stephen Robinson Department of Endocrinology and Diabetes, St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Vincenzo Rochira Unit of Endocrinology, Department of Biomedical, Metabolic, and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy; Department of Medical Specialties, Azienda Ospedaliero-Universitaria of Modena, Modena, Italy

Alan D. Rogol Emeritus Professor of Pediatrics, Section of Pediatric Endocrinology and Diabetology, University of Virginia School of Medicine, Charlottesville, VA, USA

Stefano Romeo Institute of Medicine, Sahlgrenska Academy, Department of Molecular and Clinical Medicine, University of Gothenburg; Cardiology Department, Sahlgrenska University Hospital, Gothenburg, Sweden; Clinical Nutrition Unit, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy

Richard Ross Department of Oncology and Metabolism, University of Sheffield, UK

Fabio Rotondo Department of Laboratory Medicine, Division of Pathology; The Keenan Research Centre for Biomedical Science at the Li Ka Shing Knowledge Institute, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

Francesco Rubino School of Life Course Sciences, Department of Diabetes, King's College London; Bariatric and Metabolic Surgery, London, UK

Dagmara Rusinek M. Sklodowska-Curie Institute-Oncology Center Gliwice Branch, Gliwice, Poland

Susan Sam The University of Chicago, Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Chicago, IL, USA

Ferruccio Santini Department of Endocrinology and Metabolism, University of Pisa, Pisa, Italy

Massimo Santoro Dipartimento di Medicina Molecolare e Biotecnologie Mediche, University 'Federico II', Naples, Italy

Federica Saponaro Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Naveed Sattar Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, Scotland, UK

David B. Savage Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Lars Sävendahl Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Clark Sawin† Veterans Health Administration, Washington, DC, USA

Renata C. Scalco Unidade de Endocrinologia Genética, Laboratório de Endocrinologia Celular e Molecular (LIM/25), Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo; Disciplina de Endocrinologia da Faculdade de Ciências Médicas da Santa Casa de São Paulo, São Paulo, Brazil

Peter H. Scanlon Gloucestershire and Oxford Eye Units; Harris Manchester College, University of Oxford; University of Gloucestershire, UK

Matthew J. Schiewer Department of Cancer Biology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Danja Schulenburg-Brand Department of Medical Biochemistry and Immunology, University Hospital of Wales; School of Medicine, Cardiff University, Wales, UK

Peter Senior Division of Endocrinology and Metabolism, Department of Medicine, University of Alberta, Edmonton, Canada

Amna Sheri Breast Unit, Royal Marsden Hospital, London, UK

Dolores M. Shoback Professor of Medicine, Endocrine Research Unit, Department of Medicine, San Francisco Veterans Affairs Medical Center, University of California, San Francisco, CA, USA

Soulmaz Shorakae Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia

Angela Shore Institute of Biomedical and Clinical Science, Peninsula Medical School, University of Exeter, Hevitree, UK

Kenneth Siddle University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science; National Institute for Health Research Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge, UK

Craig Sineath Rollins School of Public Health, School of Medicine, Emory University, Atlanta, Georgia, USA

Timothy C. Skinner La Trobe University, Bendigo, Victoria, Australia; Institute for Psychology, University of Copenhagen; Steno Diabetes Centre Copenhagen, Denmark

Francesca Soldà Department of Oncology, University College London Hospital, London, UK

Peter Sonksen St Thomas' Hospital and King's College, London; University of Southampton Faculty of Medicine, Southampton, UK

Jane Speight School of Psychology, Deakin University, Geelong; The Australian Centre for Behavioural Research in Diabetes, Diabetes Victoria, Melbourne, Victoria, Australia

Raj Srirajaskanthan Department of Gastroenterology, Kings College Hospital, London

Marietta Stadler Diabetes Department, King's College London Faculty of Life Sciences and Medicine, London, UK

Irinel Stanciu Clinical Assistant Professor of Medicine, University of Colorado Health Sciences Center Panorama Orthopedics and Spine Center

Takara L. Stanley Associate Professor in Pediatrics, Harvard Medical School, Boston, MA, USA

Anna Stears Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

Constantine A. Stratakis Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA

Agnieszka Święcicka Department of Endocrinology, Manchester Royal Infirmary; School of Medical Sciences, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK

Luis V. Syro Department of Neurosurgery, Hospital Pablo Tobon Uribe and Clinica Medellin, Medellin, Colombia

Guy T'Sjoen Department of Endocrinology, Ghent University Hospital, Ghent, Belgium

Vin Tangpricha School of Medicine, Emory University, Atlanta, Georgia, USA

Juha Tapanainen Department of Obstetrics and Gynecology, University of Helsinki, Helsinki, Finland

Danijela Tatovic North Bristol NHS Trust, University of Bristol, Bristol; Cardiff University School of Medicine, Cardiff, UK

Enes Taylan Department of Obstetrics, Gynecology and Reproductive Sciences, Yale University School of Medicine, New Haven, CT, USA

Roy Taylor Professor of Medicine and Metabolism, Magnetic Resonance Centre, Institute of Cellular Medicine, Newcastle University, UK

Peter N. Taylor Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK

Shu Teng Chai Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham; Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners; Department of Endocrinology, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

Massimo Terzolo Department of Clinical and Biological Sciences, San Luigi Gonzaga Hospital, University of Turin, Orbassano, Italy

Solomon Tesfaye Consultant Diabetologist and Honorary Professor of Diabetic Medicine University of Sheffield, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

Rajesh V. Thakker Academic Endocrine Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

Henri Timmers Department of Internal Medicine, Division of Endocrinology, Radboud University Medical Centre, Nijmegen, The Netherland

Yaron Tomer Division of Endocrinology, Metabolism and Bone Diseases, Icahn School of Medicine at Mount Sinai, the James J. Peters VA Medical Center; Department of Medicine, Albert Einstein School of Medicine, New York, NY, USA

Jeremy W. Tomlinson Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, UK

Massimo Tonacchera Department of Endocrinology, University of Pisa, Pisa, Italy

Jorma Toppari Institute of Biomedicine, Research Centre for Integrative Physiology and Pharmacology, University of Turku; Department of Pediatrics, Turku University Hospital, Turku, Finland

David J. Torpy Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Christos Toumpanakis Neuroendocrine Tumour Unit, Centre for Gastroenterology, Royal Free Hospital, London, UK

Herman Tournaye Centre for Reproductive Medicine, UZ Brussels, Belgium

Peter J. Trainer Department of Endocrinology, The Christie NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Nicholas A. Tritos Neuroendocrine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

A.S. Paul van Trotsenburg Department of Pediatric Endocrinology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Engin Türkgeldi Department of Obstetrics and Gynaecology, School of Medicine, Koc University, Istanbul, Turkiye

Clara Ugolini Department of Laboratory Medicine, Section of Pathology Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy

David Unuane Department of Endocrinology, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels, Belgium

Thomas Upton University of Bristol, Bristol, UK; University of Otago, Dunedin, New Zealand

Bijay Vaidya Department of Endocrinology, Royal Devon & Exeter Hospital and University of Exeter Medical School, Exeter, UK

Jonathan Valabhji St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

Juan W. Valle Department of Medical Oncology, The Christie NHS Foundation Trust; Division of Molecular and Clinical Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

Greet Van den Berghe Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven, Belgium

Wouter R. van Furth Department of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands

Mark P.J. Vanderpump Consultant Physician and Endocrinologist, OneWelbeck Endocrine and Diabetes Partners, The Physicians' Clinic, London, UK

Dirk Vanderschueren Department of Chronic Diseases, Metabolism and Ageing (CHROMETA), Laboratory of Clinical and Experimental Endocrinology, KU Leuven; Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium

Irina A. Vasilevskaya Department of Cancer Biology and The Sidney Kimmel Cancer Center, Jefferson University, Philadelphia, PA, USA

Flora Veltri Endocrine Unit Centre Hospitalier Universitaire Saint Pierre; Université Libre de Bruxelles (ULB), Brussels, Belgium

Bruno Vergès Endocrinology-Diabetology Department, University-Hospital and INSERM, Medicine University; Service Endocrinologie, Diabétologie et Maladies Métaboliques CHU-Dijon, Dijon, France

Russell M. Viner Great Ormond Street Institute of Child Health, University London College Hospital, London, UK

W. Edward Visser The Rotterdam Thyroid Center and Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

Paolo Vitti Department of Clinical and Experimental Medicine, Endocrinology Unit I, University of Pisa, Italy

Robert Volpé† Department of Medicine, The Wellesley Hospital, Toronto, Canada

Mark Walker Institute of Cellular Medicine, Newcastle University, UK

Marc Walter University of Basel, Psychiatric Hospital, Division of Substance Use Disorders, Switzerland

Nicholas J. Wareham MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK

Leonard Wartofsky Department of Medicine, Washington Hospital Center, Washington, DC, USA

John A.H. Wass Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, Oxford, UK

Laura M. Watts Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Susan M. Webb Department of Endocrinology/ Medicine and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBER-ER, Unidad 747), ISCIII, Barcelona. Hospital Sant Pau, Universitat Autònoma de Barcelona, Spain

Anthony P. Weetman Faculty of Medicine, Dentistry, and Health, The University of Sheffield Medical School, Sheffield, UK

Ralf Werner Division of Paediatric Endocrionolgy and Diabetes, Department of Paediatric and Adolescent Medicine, and Institute for Molecular Medicine, University of Lübeck, Germany

Philip J. Weston Diabetes Centre, Royal Liverpool University Hospital, Liverpool, UK

Clare Whicher Research and Development Department, Tom Rudd Unit, Moorgreen Hospital, Southampton, UK

Billy White Great Ormond Street Institute of Child Health, University London College Hospital, London, UK

Michael G. White Institute of Cellular Medicine, Newcastle Medical School, Newcastle University, Newcastle Upon Tyne, UK

Michael P. Whyte Center for Metabolic Bone Disease and Metabolic Research, Shriners Hospital for Children, and Division of Bone and Mineral Diseases, Departments of Medicine, Pediatrics, and Genetics, Washington University School of Medicine, St Louis, Missouri, USA

Wilmar M. Wiersinga Department of Endocrinology and Metabolism, University of Amsterdam, Amsterdam, The Netherlands

Sarah Wild Usher Institute, University of Edinburgh, Edinburgh, UK

John P. Wilding Department of Obesity and Endocrinology, University of Liverpool, UK

Graham R. Williams Molecular Endocrinology Laboratory, Department of Medicine, Imperial College London, Hammersmith Campus, London, UK

Bryan Williams University of Leicester School of Medicine, and Leicester Blood Pressure Clinic, University Hospitals of Leicester NHS Trust, Leicester, UK

Peter Winocour ENHIDE, East and North Herts Institute of Diabetes and Endocrinology, East and North Hertfordshire NHS Trust, UK

Douglas Wiseman Endocrine Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Gary Wittert Professor and Head, Discipline of Medicine, Director, Freemasons Foundation Centre for Men's Health, University of Adelaide, Senior Consultant Endocrinologist, Royal Adelaide Hospital, Adelaide, Australia

F. Susan Wong Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

Rohana J. Wright Department of Diabetes and Endocrinology, St John's Hospital, Livingston, UK

Jing Wu Department of Endocrinology, Xiangya Hospital, Central South University, Changsha, China

Swaytha Yalamanchi Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, The Johns Hopkins University, Baltimore, Maryland, USA

Jennifer M. Yamamoto Department of Medicine, Division of Endocrinology, University of Calgary, Alberta, Canada

Shunichi Yamashita Faculty of Medicine, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

Ephia Yasmin Consultant Gynaecologist, sub-specialist in Reproductive Medicine, Institute for Women's Health, University College London, UK

Bu B. Yeap Professor, The Medical School, University of Western Australia, Endocrinologist, Department of Endocrinology and Diabetes, Fiona Stanley Hospital, Perth, Western Australia

Lamis Yehia Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Bulent Okan Yildiz Division of Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, Hacettepe University, Turkey

Dorina Ylli Faculty of Medicine, University of Tirana, Radiology and Clinical Semeiotics, Tirana, Albania

Xuefei Yu Institute of Cellular Medicine, Newcastle University, UK

Sagen Zac-Varghese ENHIDE, East and North Herts Institute of Diabetes and Endocrinology, East and North Hertfordshire NHS Trust, UK

Maria Christina Zennaro INSERM, UMRS_970, Paris Cardiovascular Research Center; Université Paris Descartes, Sorbonne Paris Cité; Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Service de Génétique, Paris, France

Michael B. Zimmermann The Human Nutrition Laboratory, ETH Zürich, Switzerland

Michael Zitzmann University Clinics of Muenster, Center for Clinical Andrology, Domagkstrasse, Münster, Germany

Nitash Zwaveling-Soonawala Department of Pediatric Endocrinology, Amsterdam University Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Free personal online access for five years

Individual purchasers of this book are also entitled to free personal access to the online edition for five years via oxfordmedicine.com/otendocrinology3. Please refer to the access token for instructions on token redemption and access.

Accessing this content online allows you to print, save, cite, email, and share content; download high-resolution figures as PowerPoint® slides; save often-used books, chapters, or searches; annotate; and quickly jump to other chapters or related material on a mobile-optimised platform.

SECTION 7

Disorders of Growth, and Development and Transition

7.1

Growth and Its Disorders

7.1.1 Recognizing Normal and Disordered Growth

Gary Butler

Introduction

The growth of a human being from a single cell to a fully mature individual is a remarkable process and something that is subject to a large number of influences across the whole age spectrum. Growth before birth is proportionately the most rapid and probably the least understood phase. The size of an infant at birth, therefore, is dependent on a number of factors which are primarily maternal such as the well-being of the feto-placental unit and its level of functioning. This unit is markedly affected in maternal undernutrition, which gives significant deleterious effects on fetal growth. Additionally maternal size is critical. A physically small mother will constrain growth even when the fetus has a large genetic potential size. Lastly, fetal factors are important themselves. Genetic or endocrine disturbances may constrain fetal growth.

In paediatric practice we are concerned about postnatal growth. It is useful to think of growth in three separate phases: infancy, childhood, and puberty [1]. The infancy phase is largely nutrition dependent and lasts for one to two years. After this, the childhood phase, which is predominantly growth hormone driven takes over and continues until the pubertal or adolescent phase. This final phase is under the influence of sex steroids and the speed of this phase determines the timing and rate of acceleration of the pubertal growth spurt and the cessation of growth. It is very helpful to consider the different influences on each phase when presented with the diagnostic challenge of a child with abnormal growth [2, 3]. However, the process of assessing growth variations begins with the need to make accurate and reproducible measurements.

Assessment of Growth

Equipment

As with any routine clinical assessment, using the correct equipment is important. This is certainly true for the assessment of growth where one of the biggest sources of error is broken, incorrectly calibrated, or out of date equipment. The other source of errors is poor measuring technique.

Although in specialist clinics very accurate equipment such as the Harpenden stadiometer is used, it requires regular calibration and maintenance. There are other types of stadiometer, including electronic ones which once installed accurately do not require recalibration and are suitable for routine use in clinics, wards, and surgeries. There is also relatively cheap portable equipment such as the Leicester height measure which, when used with the correct measurement technique can give an accurate height reading.

Weighing scales are often poorly maintained and calibrated and do not give an accurate readout across the whole range. Children's weight is best measured with correctly calibrated and zeroed class III electronic scales. Head circumference is best measured with a specially designed measure, such as a metal anthropometric tape or the plastic Lasso-o (Figure 7.1.1.1). If that is not available then a paper tape folded lengthways to reduce its width can be used.

Measuring Technique—Length and Height

It is now recommended that length should be measured in children up to the age of two years. The correct technique requires two people—one to hold the infant's head against the headboard with the eyes facing forwards and in the Frankfurt plane (the outer canthus of the eye in the same plane as the upper margin of the pinna). The other observer straightens the legs having removed the nappy beforehand and holds the legs extended and brings the board against both heels and the length is read to the nearest 0.1 cm (Figure 7.1.1.1). Height should be measured in children

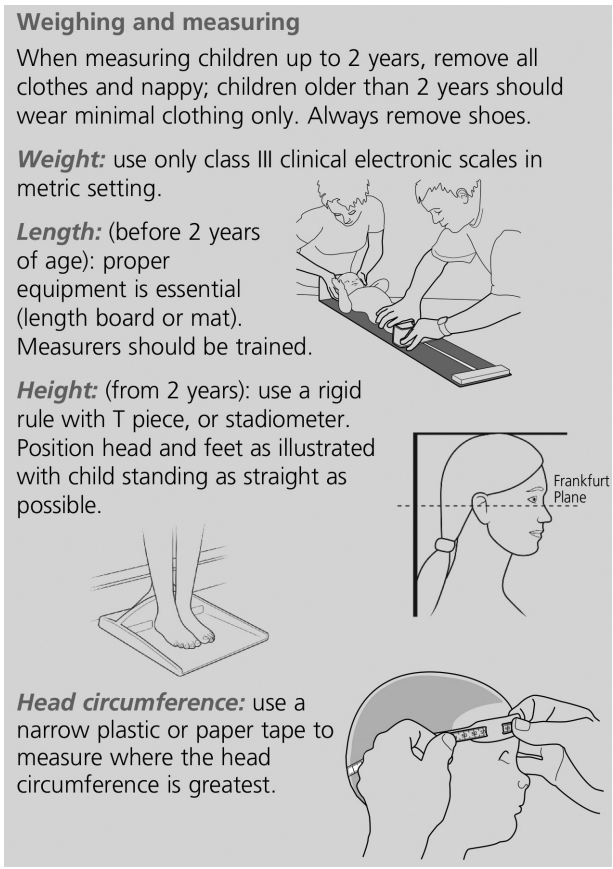


Figure 7.1.1.1 Recommended approach to measurement of supine length, weight, and occipito-frontal head circumference.

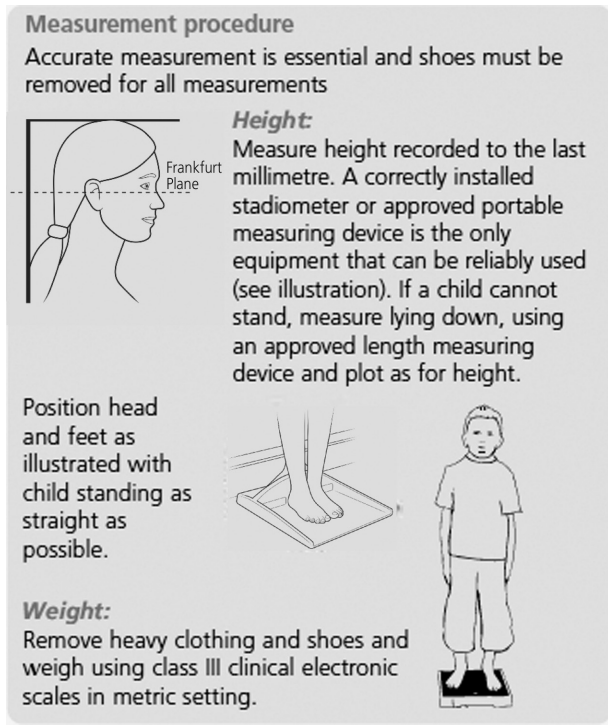


Figure 7.1.1.2 Recommended approach to measuring standing height indicating the Frankfurt plane, and weight.

over 2 years of age, using the appropriate equipment with the child standing as tall as possible, back touching the wall or backplate of the stadiometer, feet together, ankles together, and heels placed against the foot restraint. The head is placed in the Frankfurt plane and the measuring board is brought down on the top of the head (Figure 7.1.1.2). Shoes need to be removed and hair arrangements that prevent the measuring arm resting on the top of the head should be undone. The practice of stretching, putting upward pressure under the mastoid processes while the child takes a deep breath and then exhales is sometimes recommended to attempt to compensate for the up to 2 cm loss in height that occurs gradually throughout the day on account of spinal compression (diurnal variation). The difficulty here is that interobserver variations in stretching are difficult to control for, but the key to accuracy is that consistency of technique is applied, whether stretching is used or not. Instruction in measurement techniques and videos are available on the UK Royal College of Paediatrics and Child Health website [4].

Reliability and Reproducibility

With appropriately calibrated equipment the reproducibility of the measurement of length or height is usually within 0.5 cm. This variability is known as the measurement error. With a trained auxologist using a calibrated stadiometer this can be lower than 0.2 cm. Training is therefore vital and the recognition of measurement error is important as this is magnified when height velocity is calculated from two height values. For this reason, height velocity measurements at intervals of less than one year are likely to contain a greater proportion of measurement error. Furthermore, height velocity reference charts are only valid for calculation of the growth increment over a whole year. This also has the benefit that it cancels out any seasonal variation in growth which is well described, the timing of peaks and troughs in velocity varying between individuals.

Timing of Measurements

There is considerable debate as to what are the optimal ages for children to be measured. Weight, head circumference, and ideally length should be measured at birth. The infant should be weighed again within the first two weeks to assess postnatal weight loss. After that, it is recommended that infants are weighed at 2, 3, and 4 months coinciding with the UK vaccination schedule and subsequently at all other points of interaction with healthcare services. It is recommended that length is measured when required and certainly whenever there is concern about weight gain. However, it is good practice for an infant to have their length measured within the first year, then height in the toddler years and then also preschool to assess for any growth deviation. The UK National Child Measuring Programme includes height and weight estimation at school entry at age 4–5 years and on primary school exit at age 10–11 years. This was initiated as a childhood obesity surveillance scheme, and not primarily for screening for individual growth abnormalities [5].

Growth During Infancy

Growth during infancy is the most rapid phase of postnatal growth. In the first year of life, weight triples from 3 kg to approximately 10 kg on average and there is a 50% increase in stature from 50 cm to 75 cm. There is also a further 33% increase in head circumference especially within the initial 6 months of life which reflects the very rapid maturation and growth of the brain during this first year. Even in the presence of good nutrition and infant care there can be a normal drifting of measurements upwards or downwards across as many as two centile lines for length or weight, and one for head circumference over the first year, and this represents the infant determining their own genetic growth trajectory. Most of the parameters of birth size, as mentioned earlier, are determined by the intrauterine environment. Thus, the correlation between birth weight, birth length, head circumference, and the equivalent adult body proportions is actually very low.

Growth During Childhood

The variability in height velocity during the childhood phase is quite considerable, values of between 5 cm and 15 cm per year being quite normal. Although the overall trend is for the mean rate of growth to gradually decline, this is dependent upon a number of factors. First is the obvious fact that children with a taller height potential need to grow more quickly in the same time frame to reach their adult height. The converse is true for short children. Consequently, it can be problematic to determine the lowest acceptable limit of height velocity when presented with a short and slowly growing child. Calculating height velocity over intervals less than one year will be less accurate on account of a greater contribution of measurement error and subject to seasonal influences in growth rate. There are also the endogenous rhythms or cycles in growth to take into consideration [6, 7]. Normal children have phases of growth acceleration and deceleration every two to three years known as the prepubertal growth spurts (Figure 7.1.1.3), the most well-recognized and constant of these being the mid-childhood growth spurt, depicted on some national height velocity standards (e.g. US and Irish). Hence repeat height measurements to determine serial height velocity measurements are the best pointer to a growth disorder.

Given the inherent challenges of growth measurement accuracy and the knowledge about the variable pattern of childhood growth, what then is the best method for recognizing abnormal growth? In a review of practice from the Netherlands, Finland, and the UK, the most accurate parameters are deflection in height and distance from mid-parental target height [8]. Thus, there is a significantly greater chance of pathology in a child who is tall or short for their family, whose height increases or decreases by 0.5 SD over 1 year or more than 0.7 SD over 2 years. This, on the UK RCPCH growth charts, equates to a shift of over half a centile bandwidth over 1 year and across greater than one intercentile space (0.67 SD) over 2 years.

Puberty and Pubertal Growth

The staging described by Tanner is now the universally accepted method of describing this transitional phase of life [9]. Although

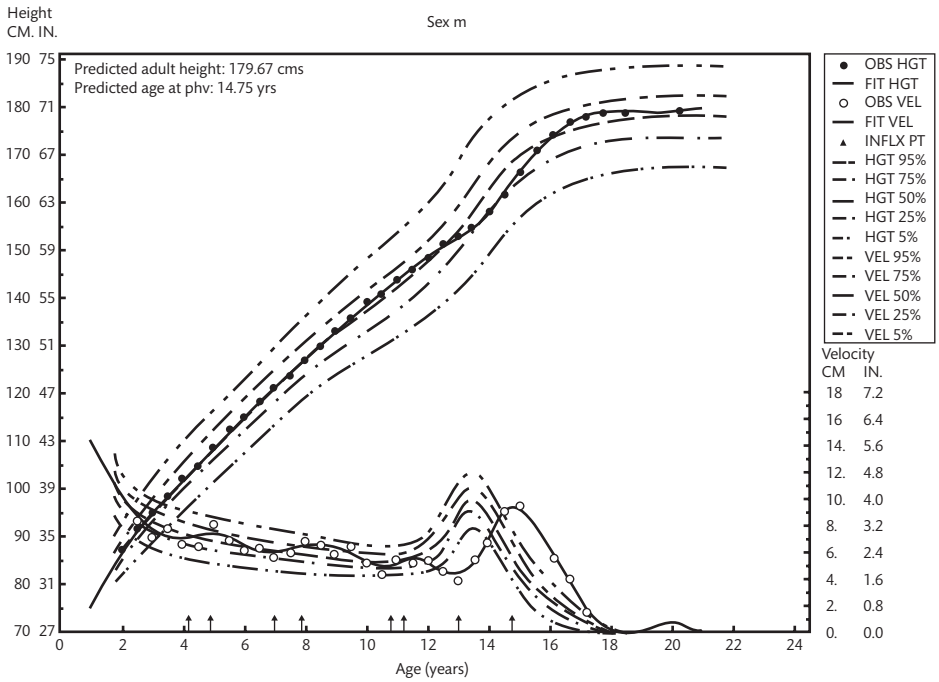


Figure 7.1.1.3 Height and height velocity curves for one boy on the background of standard height and height velocity curves. The curves are derived by mathematical modelling with raw height data. The height curve demonstrates the normal centile crossing that can occur when puberty is different from the mean. The height velocity curve demonstrates three prepubertal spurts in growth. PHV, peak height velocity; OBS HGT, observed height values; FIT HGT, fitted height curve; OBS VEL, observed height velocity values; FIT VEL, fitted height velocity curve; INFLX PTS, points of inflexion—peaks and troughs in velocity.

Table 7.1.1.1 Key milestones of the adolescent growth spurt in boys and girls indicating how rapidly puberty progresses around the peak stage of development, at what age in the process this occurs and what is the range of each parameter (given in SD). Height increase is given in cm/yr, puberty parameters in Tanner stages/yr and testis volume in ml/yr [10].

Sex	Marker	Age at peak velocity (yr)		Range of peak velocity (units/yr)	
		Mean	SD	Mean	SD
Boys	Height (cm)	14.1	1.03	9.7	0.83
	Genital stage	13.0	0.80	1.2	0.35
	Pubic hair stage	13.7	0.94	1.7	0.68
	Testis volume (ml)	13.3	1.13	6.3	2.7
Girls	Height (cm)	11.9	1.01	7.6	0.91
	Breast stage	12.0	1.09	1.5	0.39
	Pubic hair stage	12.6	0.97	1.8	0.68
	Menarche	13.1	1.21	–	–

pubertal development is continuous, an attempt to understand the process in defined stages can be very helpful. The details are well known and found in all reference texts. However, the description of five distinct stages can lead to misinterpretation unless what happens at each stage is fully understood. As stage 1 is that of the prepubertal child, and stage 5 that of the fully mature individual, there are in reality only three transitional phases of puberty: stage 2—early, stage 3—middle, and stage 4—late. The timing of the key events is what differs between the sexes. In early puberty, defined as testicular enlargement in boys and areolar growth in girls, external signs are negligible in boys but in girls, growth accelerates fast. By mid-puberty girls have reached peak height velocity and boys' growth is beginning to speed up. At stage 4 or late puberty most of the external changes in boys such as lowering of the voice, the development of pubic, axillary and facial hair, and rapid growth become apparent and testis volume increases to around 12–15 ml, whereas in girls the key event is menarche. Table 7.1.1.1 details the mean age and range of the key features of the pubertal growth spurt [10].

Tanner staging requires conducting a clinical examination and training of the clinician to ensure accuracy. In day-to-day clinical practice, this is not always feasible, possible, or desirable, yet the complete evaluation of growth in a girl over 8 years and a boy over 9 years necessitates knowing what is happening to pubertal development. Thus the use of the puberty 'phases' system may be more appropriate. The division of puberty into three phases, **prepuberty**, **in puberty**, and **completing puberty** allows ascertainment by history alone if examination is not possible (Table 7.1.1.2), and comparison with the vertical **puberty lines** on the 2–18 year or 2–20 year Childhood and Puberty Close Monitoring RCPCH growth charts can be made [11]. These lines depict the 99.6th and 0.4th centiles for the onset and completion of puberty and contain vertical reminders of the status of each puberty line (Figures 7.1.1.4 and 7.1.1.5).

The sequence of pubertal events does not normally vary between individuals. It is the timing of the onset of pubertal changes and the rate of progress which does. The childhood component of growth will continue until superseded by the pubertal phase, and variations in growth can usually be explained by what progress has or has not occurred in puberty. Although markedly variable from person to person, total pubertal growth in girls is usually 15–25 cm and in boys 20–30 cm, with smaller gains in later maturing adolescents. Hence the importance of pubertal assessment in the evaluation of growth during adolescence.

Weight and Other Related Measures

Weight and Body Composition

Weight in itself is a poor guide to complete well-being as it is compounded by many factors. Although it is the most commonly measured parameter in the first few postnatal weeks and months, the

Table 7.1.1.2 The phases of puberty as assessed by history

	Prepuberty (Tanner stage 1) If all of the following:	In puberty (Tanner stage 2–3) If any of the following:	Completing puberty (Tanner stage 4–5)
Girls	• No signs of nipple or breast development • No pubic hair	• Any breast enlargement so long as nipples also enlarged • Any pubic or axillary (armpit) hair growth	If all of the following: Started periods (menarche) with breast, pubic, and axillary hair development
Boys	• High voice • No growth of testes or penis • No pubic hair	• Slight voice deepening • Reddening of scrotum with growth of the testes • Early testicular or penile enlargement • Early pubic or axillary hair growth	If any of the following: • Voice fully changed (broken) • Adult size of penis with pubic and axillary hair growth • Moustache and early facial hair growth

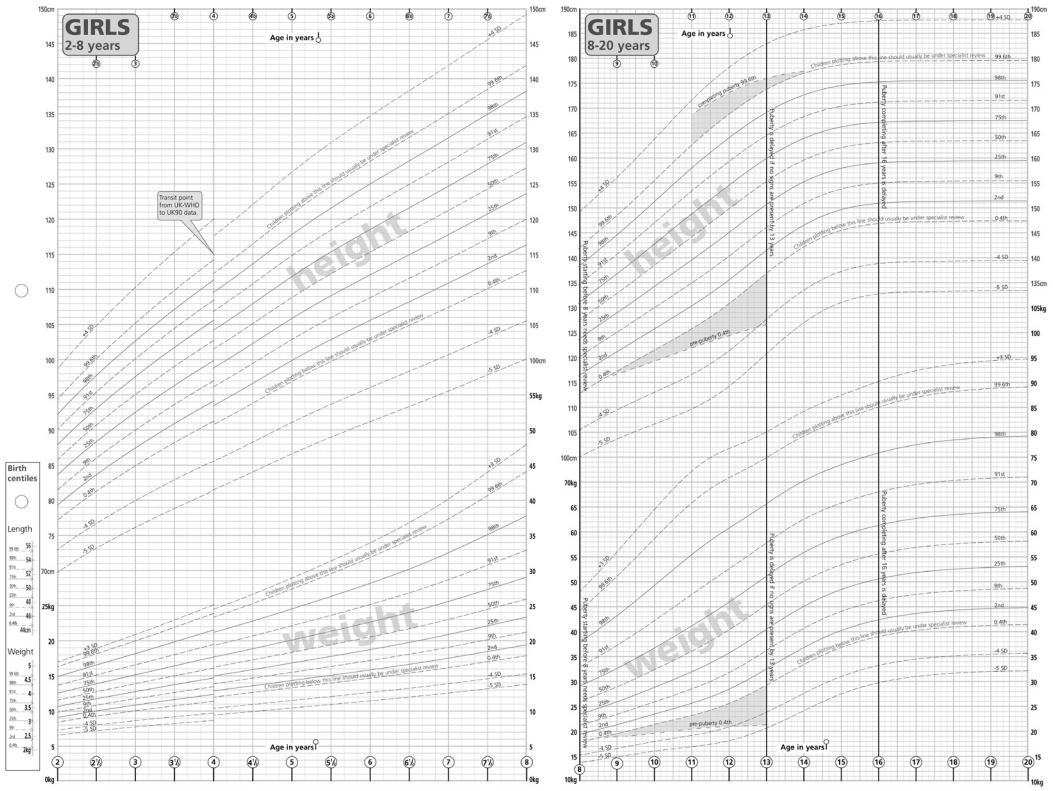


Figure 7.1.1.4 Childhood and puberty close monitoring 2–20 years growth chart for girls.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Royal College of Paediatrics and Child Health.

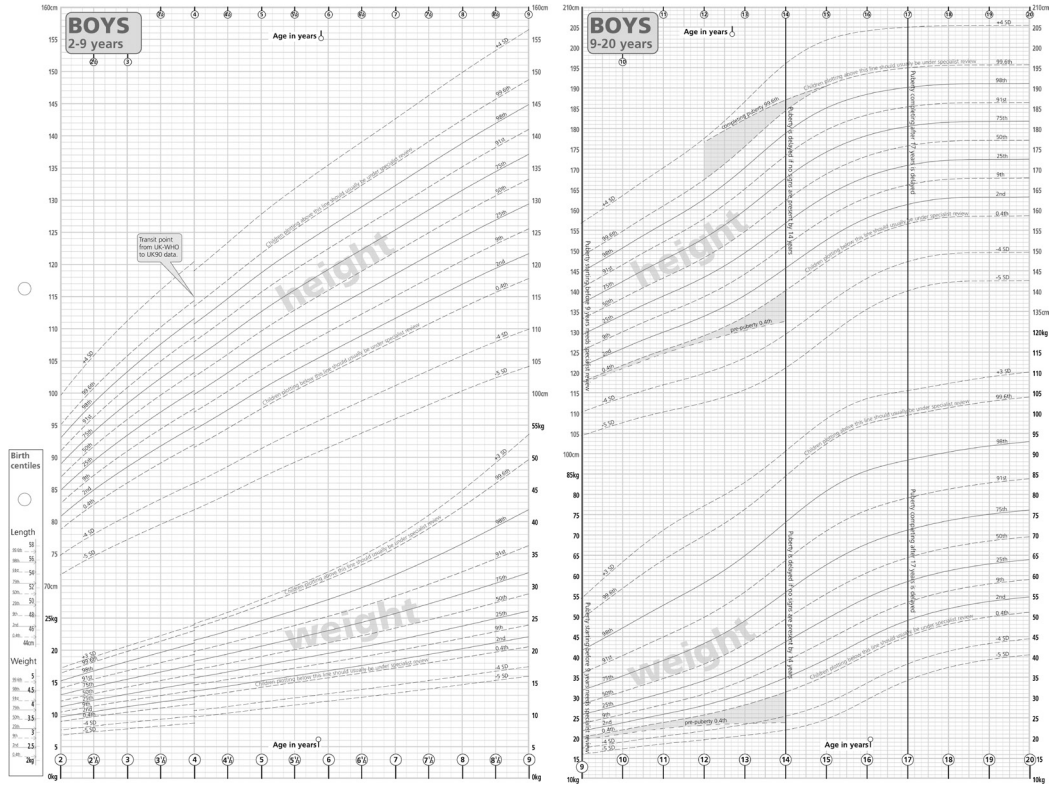


Figure 7.1.1.5 Childhood and puberty close monitoring 2–20 years growth chart for boys.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Royal College of Paediatrics and Child Health.

total weight gain and the relative importance of the degree of gain needs to be compared with growth in length and measurement of the head as well.

It is difficult to comment accurately on gain in weight without knowing which component of weight has increased. Many attempts have been made to assess body composition, and most of these depend largely on the methodology available and are part of research studies. Cross-sectional CT scanning is the most accurate way of assessing fat and fat-free mass, but it is inconvenient and exposes the subject to considerable doses of radiation. Electronic measures of biometric impedance which give a proxy measurement of body composition are easily obtained in routine clinical practice such as with the Tanita® scale, but it is necessary to compare values obtained to reference standards supplied by the manufacturer of the equipment to interpret the data appropriately. Subcutaneous skinfold thicknesses can be measured, but there is inherent inaccuracy in the technique and many children find this unpleasant. Waist and hip circumferences may be measured but their value has yet to be fully determined as a predictor of cardiometabolic disease in paediatric practice. Mid-upper arm circumference is a very accurate and well described method of assessment of body composition in adults but is largely used for determination of the degree of undernutrition of children in developing countries [12].

Body Mass Index

This ratio of weight in kilogrammes over height in metres squared has become the benchmark of assessment for underweight and weight excess.

$$\text{BMI} = \frac{\text{weight (kg)}}{\left(\text{height (m)}\right)^2}$$

It is most easily calculated on a smartphone by entering weight in *kg* then dividing twice by height in *m*. BMI charts for childhood show a pattern of rapid gain in the first year of life, followed by decline over the preschool years until about 6 years of age, after which there is the normal pattern of adiposity rebound [11] (Figure 7.1.1.6 and 7.1.1.7). An Increase in BMI continues onwards and upwards in girls, whereas in adolescent boys under the influence of testosterone, there is a slowing down and then a decline from mid-puberty onwards. Overweight is defined as a BMI on the 91st centile or above, whereas obesity is classified as BMI on or above the 98th centile. Underweight is classified as low BMI below the 9th centile and very thin below the 2nd centile. The new UK-WHO charts contain a simple nomogram which enables the BMI to be looked up quickly from height and weight centiles without having to perform any calculations [13] (Figure 7.1.1.8).

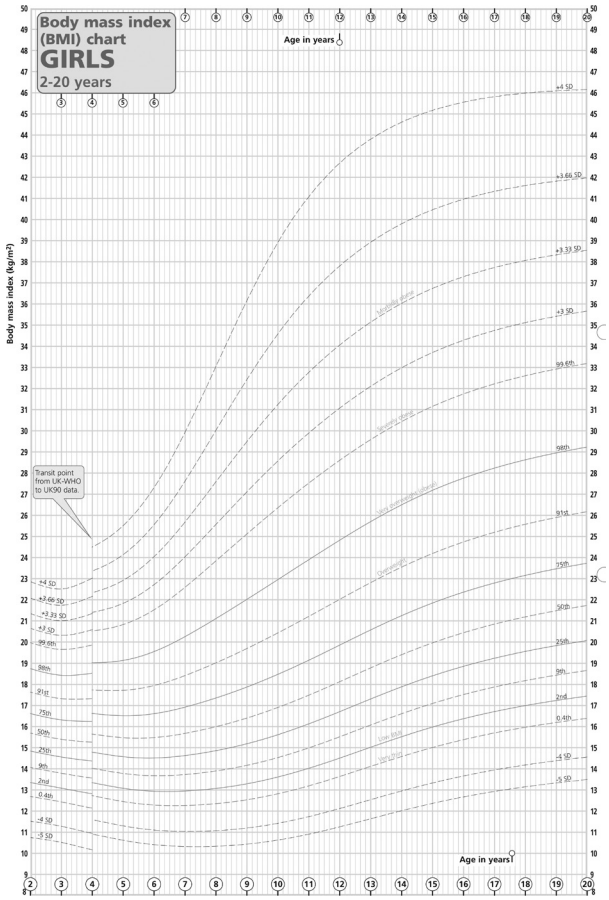


Figure 7.1.1.6 BMI chart for girls.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Royal College of Paediatrics and Child Health.

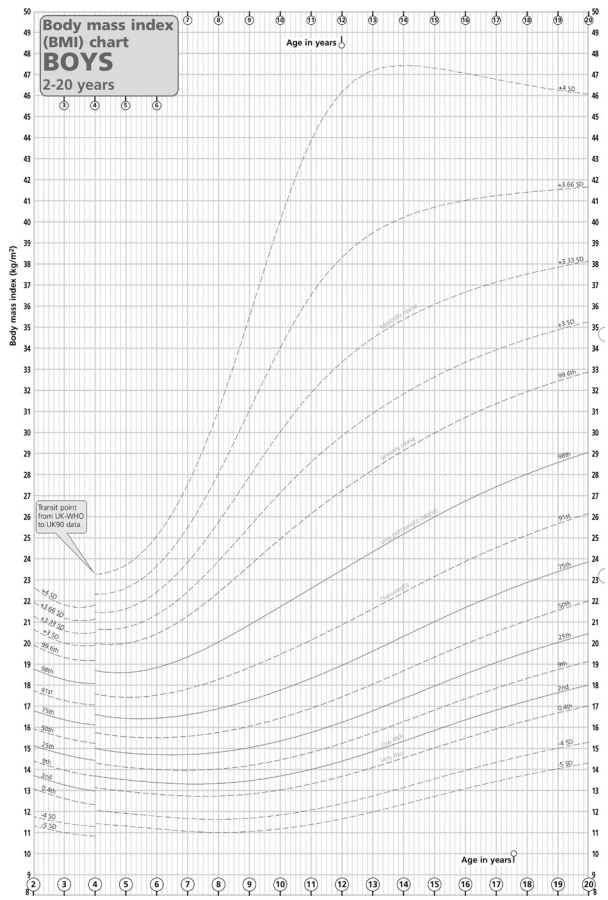


Figure 7.1.1.7 BMI chart for boys.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Royal College of Paediatrics and Child Health.

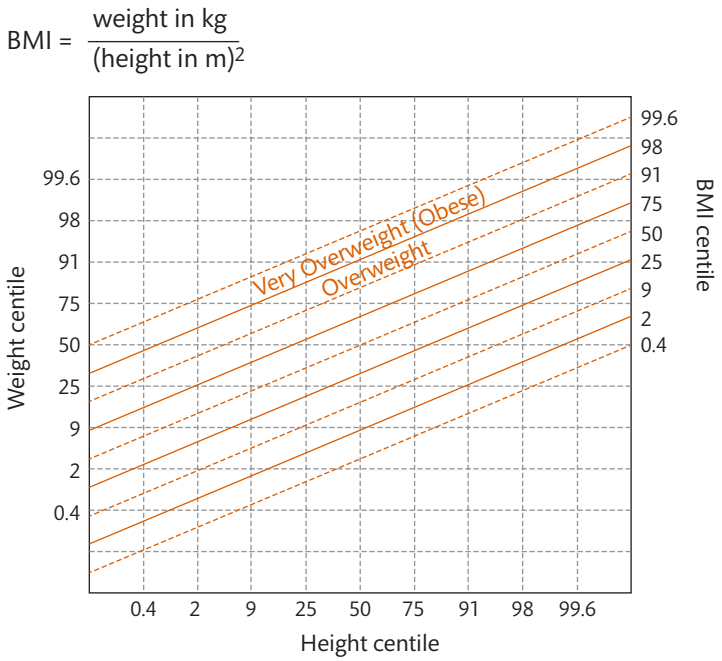


Figure 7.1.1.8 Weight–height to BMI conversion chart: the BMI look-up. Read off the weight and height centiles from the growth chart. Plot the weight centile (left axis) against the height centile (bottom axis). Read off the corresponding BMI centile from the slanting lines. Record centile with date in the data box. Accurate to ¼ centile space.

Body Surface Area

Body surface area (BSA) is preferred in the calculation of drug dose and regimens for infants. Calculating per unit of weight gives a relatively steeper increase in administered dose across all ages than by BSA. Infants have a large surface area in comparison with older children on account of the size of the head and neck, and so consequently when BSA is used to calculate fluid and drug regimens, infants will receive relatively larger amounts than using a weight-based guide. The converse is true in adolescence. BSA can be calculated by either looking up nomograms in paediatric reference manuals, apps, or read off from precalculated tables as a function of weight. Surface area can also be calculated as follows:

$$\text{BSA}\ \left(\text{m}^2\right) = \frac{\sqrt{\text{weight}\ (\text{kg}) \times \text{length}\ (\text{cm})}}{3600}$$

Types of Growth Charts and Their Use in Clinical Practice

Around the world various growth charts are in use, many of which are from locally derived populations and attempt to describe the normal variants in growth within that population. However, although there is a clear difference in physical shape and size of different ethnic groups among the human race, there has been an attempt to look at the mode of growth and pattern of gain in height, weight, and head circumference among different children of different racial backgrounds.

WHO Charts

The publication of the WHO growth charts has begun a new era in the way growth standards are constructed and presented. These have been based on prospectively collected data which are intended to represent how children should grow optimally rather than just a simple representation of how children in the population do grow—the basis behind existing growth references. The WHO growth study followed a cohort of selected babies from six countries around the world: Brazil, Oman, United States, Norway, India, and Ghana. These babies were born to non-smoking mothers, were breastfed exclusively for six months after which they were weaned and were reared in optimal social circumstances. They were followed up for five years. The complete longitudinal data for weight, head circumference, supine length until age 2 years and thereafter standing height have been used to generate these new standards. They are available on the WHO website [14] and have been combined for international usage at age 5 years with the existing United States National Centre for Health Statistics growth standards [15]. The Royal College of Paediatrics and Child Health and the Government Department of Health in the United Kingdom as well as the Government Health Departments of many countries worldwide decided to adopt the WHO standards as they best represent the growth of breastfed babies which should be regarded as the norm and represent the ideal trajectory of infant growth. These standards are now used in the UK charts from 2 weeks of age until four years of age for weight, length, and head circumference (Figures 7.1.1.9 and 7.1.1.10) [11]. There is a changeover from supine length to standing measurements at 2 years of age, a prospective decision taken in the WHO study, and therefore there is a small step downwards on the chart to account for the slight drop

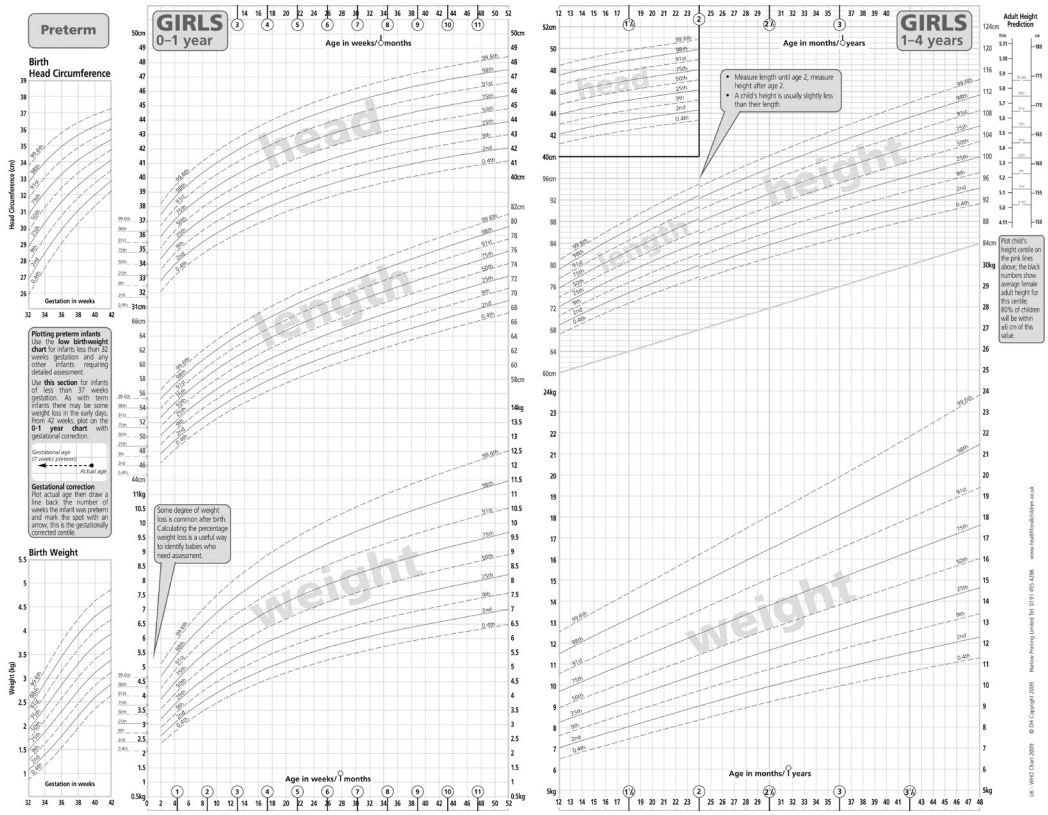


Figure 7.1.1.9 UK-WHO 0–4 years growth chart for girls.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Department of Health.

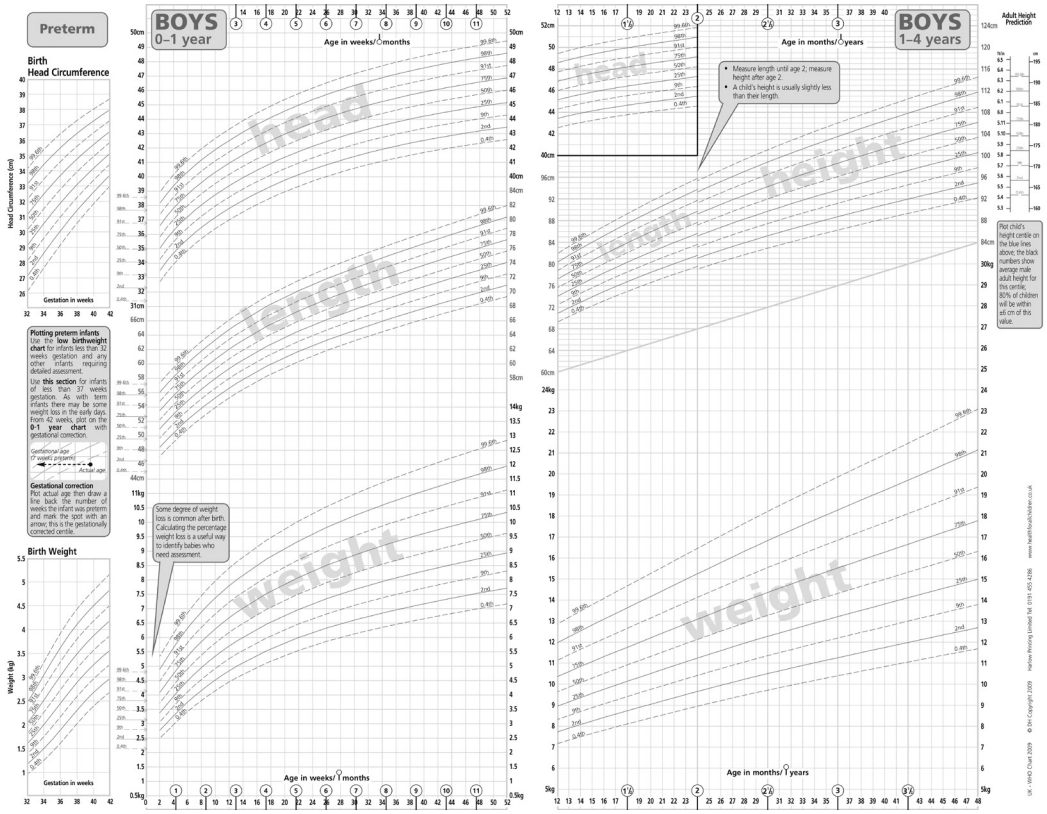


Figure 7.1.1.10 UK-WHO 0–4 years growth chart for boys.
Reproduced with kind permission from RCPCH and Harlow Printing Limited. Copyright 2019 © Department of Health.

from length to height measurement that occurs on account of the change in positioning. The birth data centiles were taken from the existing UK norms [16] as British babies are somewhat heavier than those in the WHO study. Thus, there is a gap for the first 2 weeks of life between the birth centiles on the charts and the start of the WHO centiles. This gap reminds us that the percentage weight loss in the first two weeks of life should be calculated which is usually no more than 10%. Community health staff are instructed to refer urgently any baby whose weight loss is greater than this amount as there is a significant chance of pathology if that is the case. After four years of age, children's growth can be plotted on the 2–18 years school age or 2–20 years Childhood and Puberty Close Monitoring RCPCH growth charts (Figures 7.1.1.4 and 7.1.1.5) [11].

The WHO growth study charts are designed to help identify growth problems and the mean and normal distribution is presented in either standard deviation score (SDS) or centile format. National practice determines which format is preferred, but the UK charts' centile lines are spaced exactly 0.67 standard deviations apart so conversion to SDS is straightforward, the normal range being plus and minus two standard deviations around the mean, this being represented by the 98th and 2nd centiles respectively. Specialist charts such as the 2–20-year Childhood and Puberty Close Monitoring Chart incorporate high and low reading lines of minus 4 to plus three standard deviations (Figures 7.1.1.4 and 7.1.1.5) [11].

Preterm Growth Charts

The UK-WHO charts have separate sections for the measurements of infants from 32 weeks' gestation until term. Weight and head circumference centiles for preterm infants from 32 weeks' gestation through to 42 weeks are included in separate boxes on these growth charts and in the parent/child health record kept by families. These centiles should be used over that period of time for any infant born before 37 completed weeks of pregnancy, which is the definition of preterm. Once the equivalent of 42 weeks postnatal age is reached, measurements are then plotted on the 0–4 years UK-WHO charts but with gestational age adjustment. The recommended approach to clarify whether or not correction has taken place is as follows: plot the measurement at the actual number of weeks of postnatal age and also at the gestationally corrected age. Link the two by a dotted line and an arrow as per the illustration (Figure 7.1.1.11). For infants who are of extremely low birth weight and who are also preterm (from 23 weeks gestational age), the Neonatal and Infant Close Monitoring charts are available [11]. These are derived from UK birthweight data and represent the size at birth of infants of that particular gestational age—head circumference, length, and weight. They do not describe how preterm infants grow postnatally as this pattern is very variable. It is quite usual for preterm infants, many of whom have severe respiratory disease, to show quite considerable weight loss, sometimes two centile spaces or more, and it may take several weeks or months to regain birth weight. Growth in length may also be affected by illness. Indeed, there are no standards currently available which reflect the growth of preterm infants on account of this variability. An additional feature of the preterm charts are centile lines showing minus 3, minus 4, and minus 5 standard deviations below the mean for all three measure to make allowance for the infant born small for gestational age after severe intrauterine growth retardation. These low birthweight charts continue up until two years of age, after which the 2–18-year school age, or 2–20 year Childhood and Puberty Close Monitoring RCPCH growth charts should be used (Figures 7.1.1.4 and 7.1.1.5) [11].

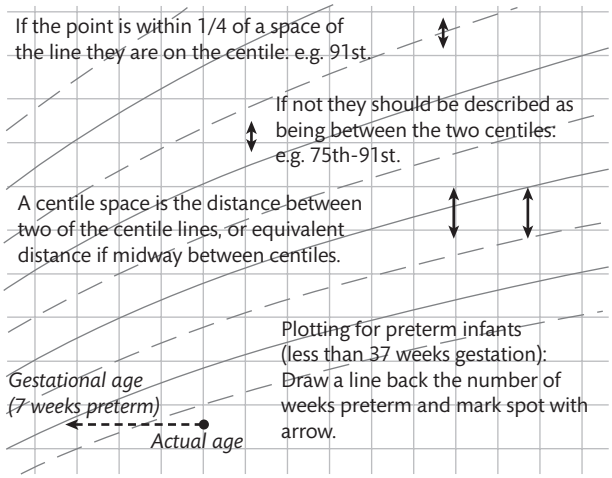


Figure 7.1.1.11 UK-WHO current recommendations for centile terminology and adjustment for prematurity.

Longitudinal Growth Charts

The majority of growth charts in routine use in countries worldwide, including the UK RCPCH [11] and US NCHS [15] charts are constructed from cross-sectional data. To produce such references many thousands of children of different ages are measured, usually only once and centile lines are constructed usually by mathematical means through the data set [16]. Such charts state the position of any child's growth parameters at a particular age with reference to the population distribution. They do not directly indicate normality or otherwise. Longitudinal charts, like the WHO charts, are derived from serial measurements of a smaller number of children from birth until maturity and the resulting values are used to construct the charts. The pattern of growth described is different, especially in early infancy and in the pubertal years and is closer to the course taken by an individual child. Longitudinal charts are preferred by some clinicians for follow-up of children with chronic conditions or growth disorders. The RCPCH charts are built around the pattern of the growth of around 190 children in the Edinburgh Growth Study, which remains the only pure longitudinal birth to maturity growth study ever conducted in the UK [10].

Individual Variability and Growth: Recognizing Abnormality

Growth charts and reference standards are constructed from the measurements of children who are growing within the normal range or close to it. There is enormous interindividual variation in growth and recognition of normal patterns of growth, so recognizing true deviations can be challenging. When a growth chart demonstrates an apparently abnormal pattern of growth it is important to be able to recognize easily that something is wrong. Before puberty, and after the postnatal period of *ex-utero* adjustment, most children

grow fairly steadily within one centile bandwidth. So much so that any deviation, up or down, may be considered as an indication of potential abnormality, especially if the child is tall or short for their target mid-parental height.

Growth Assessment During Puberty

The unpredictability of the onset of puberty and its subsequent pattern makes the development and presentation of a way of accurately assessing growth during puberty difficult. Current growth charts only represent the average amount of growth taking place for average children going through puberty at the average time, consequently, and paradoxically the charts do not represent the growth curve of any individual child or the wide variation in the timing and the intensity of growth spurt during puberty. The UK 2–18 year and 2–20-year charts have, for the first time, begun to address the issue of assessing pubertal growth [11].

Puberty Growth Charts

There have been several attempts to be able to design growth charts to allow for pubertal growth variations. There are four challenges that need to be met:

1. Age at the onset of puberty
2. The proportion of growth completed at the onset of puberty
3. The distance from target or genetic height
4. The pace of puberty

Additionally, the assessment of the degree of development during puberty either using Tanner stages or **puberty phases** has always been regarded as a separate assessment. The RCPCH UK charts for school age children and the Childhood and Puberty Close Monitoring specialist charts have attempted to combine these assessment by firstly depicting the normal ranges for beginning and completing puberty with vertical puberty lines creating zones on the growth chart to help interpret the appropriateness of growth (Figure 7.1.1.12). As in most cases, normally developing and growing adolescents will not cause concern. It is those that are at the extremes, the tall and advanced in maturing and the short and delayed in puberty who require some form of judgement as to whether their pattern of growth is acceptable. As a result of which additional lines at the extremes, 0.4th centile and 99.6th centile have been added which depict the upper ranges of normal growth, i.e. 99.6th centile for those adolescents who have completed puberty and a lower 0.4th centile for those adolescents who have not yet entered into puberty (Figure 7.1.1.13). Thus, for any young person whose height and weight plots within those ranges, consideration needs to be given as to what has happened to puberty and whether this is indeed normal; this may help to differentiate between pathology and normal variation. However, it is important to note that if a child's height falls within the shaded zone then their growth could still be abnormal if the child's height is below their target mid-parental centile range.

Syndrome-Specific Growth Charts

For children with recognized growth conditions there are a number of specialist growth charts available for children with Downs [17], Turner [18], Prader–Willi [19], Williams [20] syndromes, and achondroplasia [21]. These reference charts confer the advantage of being able to see whether a child is growing adequately, not only in comparison to the usual pattern of growth, but also with that particular diagnosis as children with these named conditions grow differently to the population. It also allows recognition of whether an intervention to ameliorate growth has had a significant effect compared with the spontaneous trajectory. They are also useful for reassurance of parents that their child is growing appropriately for their condition.

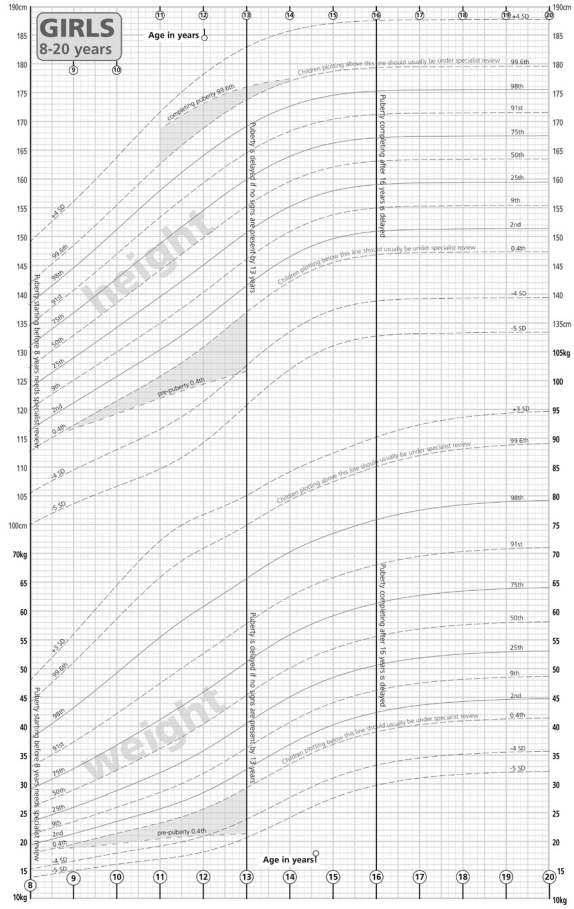


Figure 7.1.1.12 Puberty zones: vertical puberty lines depict the normal ranges for beginning and completing puberty. The shaded puberty zone marks area where 0.4th centile varies with **phase** of puberty. Heights in the shaded area below the 0.4th centile mark: (1) Prepuberty: If within two centiles of mid-parental height = **within** the normal range; and (2) in or completing puberty = **below** normal range.

Prediction of Height Gain

Prediction of height gain and adult height is always of interest to parents even when their child is growing normally, and even more so if there is a deviation from the normal pattern of growth. How, therefore can one estimate the potential growth of a child? Difficulty with accurate prediction arises on account of a number of variables involved. Firstly, there is only a relatively low correlation between a child's height and that of any one parent. Attempts to overcome this have been made by calculating a value and a

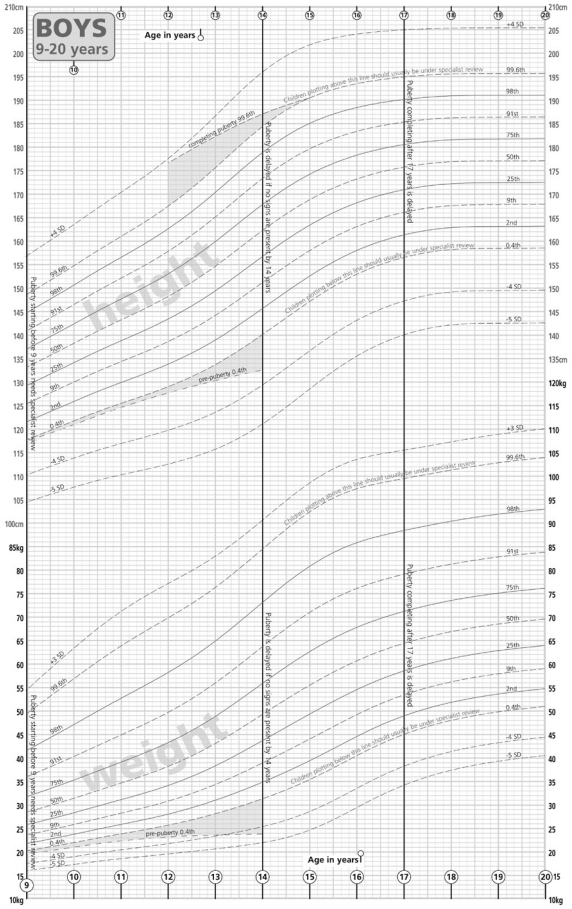


Figure 7.1.1.13 Lower 0.4th centile for prepuberty phase only. Children whose height lies within the shaded area may be normal if they are not short for their family. **Puberty lines** mark boundaries of normal pubertal development. Shaded zone marks area where 0.4th centile varies with **phase** of puberty.

range by combining both parents' heights (mid-parental height and target height range, see next). Even in the absence of pathology and illness, children show wide variation in their rate of growth and maturity. This is often referred to as the 'tempo of growth'. Children whose tempo is fast are advanced maturers and so will be tall for their age, usually with early normal puberty, but not necessarily tall as adults on account of earlier cessation of growth. The opposite is true for those with a slow tempo (the later maturers) and it is these children who often present to healthcare services with concerns over their short stature and slow growth. The variation in maturation may be estimated by taking a bone age (see next).

Mid-Parental Height (Target Height)

Since there is a reasonable association between the centile position of a child growing normally and the measured heights of *both* biological parents, there are several methods suggested for the calculation of mid-parental (or target) height. The RCPCH UK charts have a graphical mid-parental centile comparator where both parents' measured heights are plotted and when a line is drawn between these two points, the intersection identifies the mid-parental centile. Ninety-one per cent (91%) of children's adult heights will lie within two centile bandwidths of this point (i.e. plus or minus 1.33 standard deviations) and 98% will lie within plus or minus three bandwidths (plus or minus two standard deviations) (Figure 7.1.1.14).

A mathematical calculation can also be used. The difference between the 50th centile heights of male and female adults in the UK population is 14 cm, this sex difference when applied to the individual child is corrected for by adding 14 cm onto the mother's height for a boy and subtracting 14 cm from the father's height for a girl and then calculating the simple mean of both adjusted parents' heights. This value is the target height and 95% of children's adult heights will fall plus or minus 1.6 standard deviations of this value (10 cm in boys and 8.5 cm in girls). This simple approach also does not control for secular trend (i.e. the population increase in stature from one generation to the next). The method of Hermanussen and

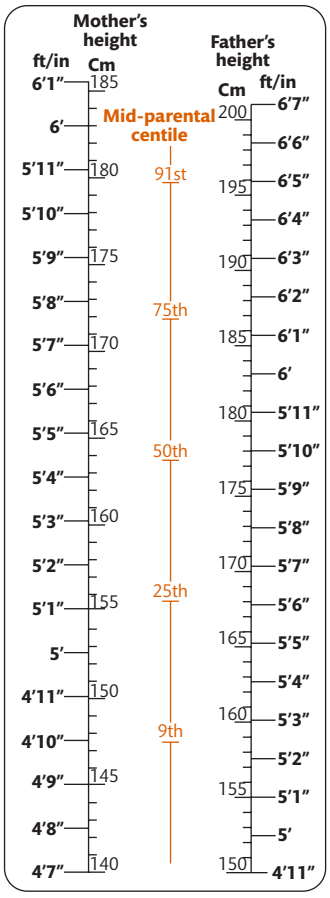


Figure 7.1.1.14 Mid-parental centile comparator where both parents' measured heights are plotted and when a line is drawn between these two points, the intersection identifies the mid-parental centile. 91% of children's adult heights will lie within two centile bandwidths of this point (i.e. plus or minus 1.33 standard deviations) and 98% will lie within plus or minus three bandwidths (plus or minus two standard deviations). Mark mother's height on the left-hand scale and the father's height on the right scale using arrows. Draw a line between arrowheads and read off mid-parental centile where this crosses the central line. Regression adjustment means that children of very short or tall parents have mid-parental centile nearer to average than expected.

Cole makes an attempt to control for this and may provide a slightly more accurate evaluation of parental height but requires the calculation of SDSs for parents' heights [22]. The authors state that by calculating parents' height SDS from growth standards (which are often constructed from measurements of children in a generation contemporary to the parents themselves), a simple correction factor can be applied to predict the target height in the child's current generation.

Calculating the target height and target centile range is useful in describing the expected range of adult stature. It can help to determine whether a child with extremes of stature is appropriately tall or short for their family, and if not, then set the clinician thinking about alternative reasons for this variation in growth.

Predictive Adult Height Calculator

An alternative approach to height prediction is the 'adult height look-up' which is incorporated into the UK-WHO charts [23]. Here the child's adult height centile can be predicted from the child's current height centile with a given accuracy of plus or minus 6 cm in 80% of cases (Figure 7.1.1.15). This method has been derived from a series of regression analyses predicting a child's actual adult height in comparison with their height at various ages beforehand as longitudinal data. There is usually a reasonably close correlation at most ages except at the time of puberty on account of the wide variation in the timing and magnitude of the adolescent growth spurt.

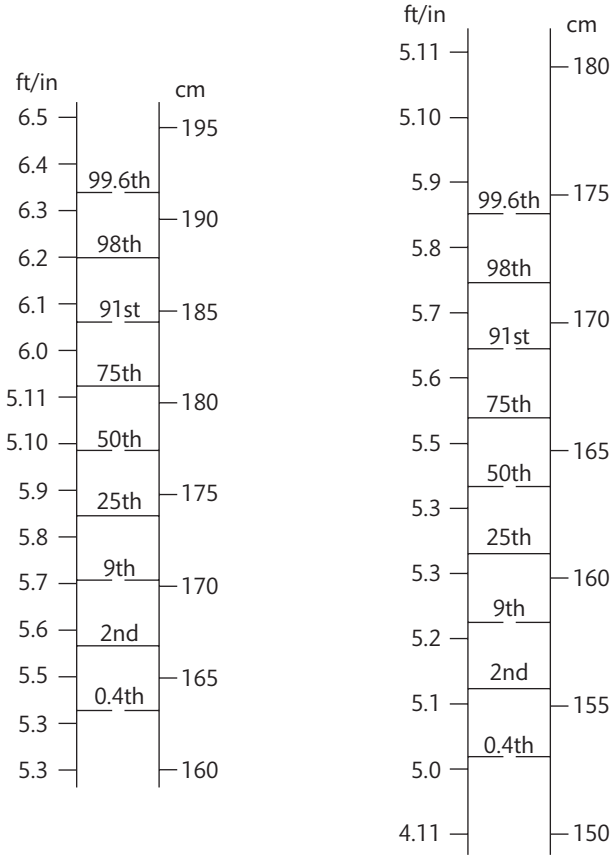


Figure 7.1.1.15 Adult height predictor. Plot the child's height centile on the lines (left-boy; right-girl). The values show the average male or female height (respectively) for this centile. 80% of children will have an adult height within +/– 6 cm of this value.

Bone Age

The bone age is the recognized standard method of assessing biological maturity. It is usually assessed on the left hand (by convention, even in left-handed children) and the X-ray should be taken in a standard radiographic position to view the wrist bones, carpals, the short bones of the hand. The accuracy in predicting maturity compared with biological status is less good than calculating the dental age from a dental radiograph, but the latter is not routinely done as the dose of radiation required is considerably higher. Bone age may be derived by a number of methods.

The most common method used in routine clinical practice is that of Greulich and Pyle, where the child's radiograph is compared with standards from an atlas compiled in 1959 derived from socially advantaged US children in the 1920s [24]. This has more recently been re-validated for Dutch children and therefore is appropriate to use in contemporary clinical practice [25]. The Tanner Whitehouse-3 (TW3) method is a different system which assesses the development of the radius, ulna, and the short bones (metacarpals and phalanges) separately and maturity scores are given to each of these bones [26]. The bone age is calculated by the summation of these maturity scores. Carpal bone maturity was scored in previous versions but as it did not contribute to improved height prediction it is no longer incorporated in height prediction. A computerized method, BoneXpert, has been derived which actually calculates the bone maturity in a three-phase process, analysing the size, shape, and density of the long and short bones [27]. This has the advantage of significantly reduced interobserver variability in rating and also reproducibility of bone age assessment especially for clinical research studies. It has also been shown to be reliable in routine clinical practice. None of the systems are compatible with each other and give different maturity ratings on the same X-ray images, so it is necessary to stick with one approach for internal consistency [28].

Uses of Bone Age

As with growth standards, bone age standards have been derived from radiographs in children growing at normal rates mostly within the normal range of stature. The index groups include few children whose growth is abnormal or who show different growth trajectories. In children growing normally, the bone age is mostly within plus or minus two years of their chronological age (equivalent to plus or minus two standard deviations from the mean). A bone age estimation within this range may not be useful for positive diagnostic purposes but may help confirm the absence of a major health issue. It is most helpful when a child has a significant pathology and the bone age is severely delayed or advanced. It may sometimes be useful for monitoring a response to treatment such as in precocious sexual maturation.

Bone Age and Prediction of Adult Height

The prediction of adult height is regarded by many as the main reason for assessing bone age. Predicted adult height may be

calculated by a number of methods. The simplest is the Bayley-Pinneau tables which, based on the Greulich and Pyle bone age measurement, are able to give an estimate of the percentage of height attained for a given child's bone age and consequently their height prognosis (i.e. their remaining growth) [29]. Although based on American children, this still has a degree of accuracy worldwide and appears to be consistently the most accurate way of predicting growth outcomes for short and tall children, although with notable under- and overprediction errors [28]. The TW3 method formulae describe residual growth for certain degrees of maturity and allows for the well-recognized acceleration in skeletal maturity that occurs during puberty. This has been described as the most accurate for the UK population. All of these methods need to be used with caution in pathological situations, even when a child has a known growth abnormality and an awareness of the variability in predictions in each pathological condition is required [28].

Conclusion

If a child's well-being is reflected in the normality of their growth, then knowing how to evaluate growth is crucial. It will require a knowledge of measuring techniques, selection of an appropriate growth chart and then determining whether a problem already exists due to abnormal size, or whether one is developing over time requiring repeated assessment. It is vital to recognize changes in growth trajectory and how they vary between infancy, childhood, and adolescence, as recognizing abnormal patterns of under- or overgrowth will aid in the more rapid diagnosis of disorders of growth and puberty.

REFERENCES

1. Karlberg J. On the construction of the infancy–childhood–puberty growth standard. *Acta Paediatrica Scandinavica Suppl*, 1989; **356**: 26–37.
2. Cheetham T, Davies JH. Investigation and management of short stature. *Arch Dis Child*, 2014; **99**: 767–71.
3. Hannema SE, Sävendahl L. The evaluation and management of tall stature. *Horm Res Paediatr*, 2016; **85**: 347–52.
4. RCPCH. UK-WHO growth charts—guidance for health professionals. Available at: https://www.rcpch.ac.uk/resources/uk-world-health-organisation-growth-charts-guidance-health-professionals (accessed: 30 March 2020).
5. Gov.uk. National child measurement programme. Available at: https://www.gov.uk/government/collections/national-child-measurement-programme (accessed: 30 March 2020).
6. Butler GE, McKie M, Ratcliffe, SG. The cyclical nature of prepubertal growth. *Ann Hum Biol*, 1990; **17**: 177–98.
7. Bock RD. Multiple prepubertal growth spurts in children of the Fels Longitudinal Study: comparison with results from the Edinburgh Growth Study. *Ann Hum Biol*, 2004; **31**: 59–74
8. Stalman SE, Hellinga I, Dommelen P, *et al*. Application of the Dutch, Finnish and British screening guidelines in a cohort of children with growth failure. *Horm Res Paediatr*, 2015; **84**: 376–82.
9. Tanner JM. *Growth at Adolescence*. Oxford: Blackwell, 1962.
10. Cole TJ, Pan H, Butler GE. A mixed effects model to estimate timing and intensity of pubertal growth from height and secondary sexual characteristics. *Ann Hum Biol*, 2014; **41**: 76–83.
11. RCPCH. Growth charts. Available at: https://www.rcpch.ac.uk/resources/growth-charts (accessed: 30 March 2020).
12. World Health Organization. Child growth standards. Available at: http://www.who.int/childgrowth/standards/ac_for_age/en/ (accessed: 30 March 2020).
13. Cole TJ. A chart to link child centiles of body mass index, weight and height. *Eur J Clin Nutr*, 2002; **56**: 1194–9.
14. World Health Organization. The WHO child growth standards. Available at: http://www.who.int/childgrowth/standards/en/ (accessed: 30 March 2020).
15. Centers for Disease Control and Prevention. CDC Growth Charts. Available at: https://www.cdc.gov/growthcharts/cdc_charts.htm (accessed: 30 March 2020).
16. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med*, 1998; **17**: 407–29.
17. The Down Syndrome Medical Interest Group. Growth charts. Available at: https://www.dsmig.org.uk/information-resources/growth-charts/ (accessed: 30 March 2020).
18. Available at: http://www.healthforallchildren.com/shop-base/shop/growth-charts/turner-syndrome-growth-charts-100-pack/ (accessed: 30 March 2020).
19. Harlow Healthcare. Turner syndrome growth charts. Available at: http://pediatrics.aappublications.org/content/135/1/e126 (accessed: 30 March 2020).
20. Williams Syndrome Association. Growth charts. Available at: https://williams-syndrome.org/growth-charts/growth-charts (accessed: 30 March 2020).
21. Horton WA, Rotter JI, Rimoin DL, *et al*. Standard growth curves for achondroplasia. *J Pediatr*, 1978; **93**: 435–8.
22. Hermanussen M, Cole TJ. The calculation of target height reconsidered. *Horm Res*, 2003; **59**: 180–3.
23. Cole TJ, Wright CM, Butler G. An improved tool to predict adult height and relate to parental height for the new UK-WHO growth charts. *Arch Dis Child*, 2009; **94**: A8.
24. Greulich WW, Pyle SI. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. Stanford, CA: Stanford University Press, 1959.
25. van Rijn RR, Lequin MH, Robben SG, Hop WC, van Kuijk C. Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today? *Pediatr Radiol*, 2001; **31**: 748–52.
26. Tanner JM, Healy MJR, Goldstein H, Cameron N. *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3) Method*, edn 3. Singapore: Saunders, 2001.
27. Thodberg HH. An automated method for the determination of bone age. *J Clin Endocrinol Metab*, 2009; **94**: 2239–44.
28. Martin DD, Wit JM, Hochberg Z, *et al*. The use of bone age in clinical practice· *Horm Res Paediatr*, 2011; **76**: 10–16.
29. Bayley N, Pinneau SR. Tables for predicting adult height from skeletal age revised for Greulich and Pyle standards. *J Pediatr*, 1952; **40**: 423–41.

7.1.2 Disorders of the GH-IGF Axis

Alexander A.L. Jorge, Fernanda A. Correa, and Renata C. Scalco

Introduction

Disorders of the growth hormone-insulin-like growth factor (GH-IGF) axis include a large heterogeneous group of conditions with distinct phenotypes (Figure 7.1.2.1, Box 7.1.2.1 and Table 7.1.2.1). Each of the defects identified in this axis is responsible for a rare and specific condition that endocrinologists need to be familiar with. Traditionally, clinical and hormonal evaluations are the cornerstones of diagnosis of these disorders but, in recent years, molecular genetic tests have become an important tool for the investigation of GH-IGF axis disorders. In this chapter, we will focus on genetic causes and highlight their main characteristics and the diagnostic approaches for each of them. If available, a number prefix that facilitates access to 'Online Mendelian Inheritance in Man' database (OMIM—https://www.omim.org/) will be provided, expediting the access to additional information for each disease.

The determination of an IGF-1 level and GH secretion status are usually the first step to assess the GH-IGF axis [1, 2] (Table 7.1.2.1). IGF-1 should be evaluated according to sex and chronological age [2, 3]. Because of the pulsatile nature of GH secretion, sufficiency or deficiency of GH is usually determined after a pharmacological stimulation test [2]. However, in many conditions described in this chapter, basal GH is already markedly elevated, excluding GH deficiency [4]. Based on clinical findings and IGF-1/GH levels, further evaluation of other components of this axis can be included: insulin-like growth factor-binding protein 3 (IGFBP-3), acid-labile subunit (ALS), insulin-like growth factor 2 (IGF-2), growth hormone-binding protein (GHBP). Additionally, the acute IGF-1 and IGFBP-3 response after exogenous GH therapy (the IGF Generation test) can give valuable information to better characterize the GH-IGF axis defect. Molecular genetic tests can then be used to define the precise defect in the GH-IGF axis [5].

GH deficiency is by far the most common defect in GH-IGF axis; its frequency is around 1–2% in non-selected groups of children with short stature [6]. In the last decade, several studies have suggested that *IGF1R* defects leading to variable degrees of IGF-1 insensitivity can be present in 2% of short children born small for gestational age (SGA) [7].

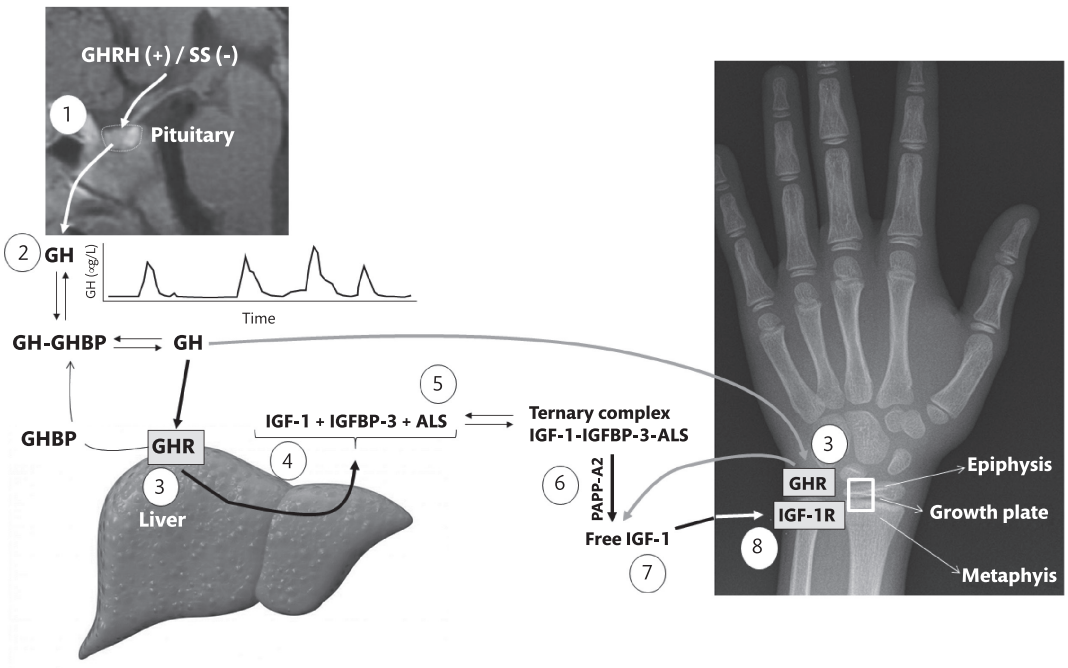


Figure 7.1.2.1 Schematic representation of the GH-IGF axis indicating the main disorders in this system: (1) growth hormone deficiency (GHD); (2) bioinactive GH; (3) GH insensitivity; (4) IGF-1 deficiency; (5) acid-labile subunit deficiency; (6) defects in the proteolytic cleavage of IGFBPs; (7) bioinactive IGF-1; and (8) IGF insensitivity.

Box 7.1.2.1 Disorders of the GH-IGF axis

GH deficiency
- Idiopathic
- Acquired (craniopharyngioma, pituitary tumours, autoimmune diseases, granulomatous diseases, central nervous system infections, post-radiotherapy, head trauma)
- Congenital: associated with structural defects
- Genetic
  - GH secretion (*GH1* and *GHRHR* genes)
  - Pituitary cells differentiation (*POU1F1* and *PROP1* genes)
  - Pituitary development (*HESX1*, *GLI2*, *OTX2*, *LHX3*, *LHX4*, and *SOX3* genes)

Bioinactive GH (Kowarski syndrome, OMIM 262650)

GH insensitivity (GHI)
- Complete GHI (OMIM 262500)
- Partial GHI (OMIM 604271)
- GHI associated to immune dysfunction (OMIM 245590)
- Atypical GHI
- Secondary or acquired GHI (anti-GH antibodies, malnutrition, liver disorders, poorly controlled diabetes mellitus, uraemia)

Ternary complex defects
- Acid-labile subunit deficiency (OMIM 615961)
- Defects on proteolytic cleavage of IGFBPs (*PAPPA2* gene)

Primary IGFs deficiency
- IGF1 defects (OMIM 608747)
- IGF2 defects (OMIM 616489)

Bioinactive IGF-1

IGF insensitivity (OMIM 270450)

GH, growth hormone; IGF-1, insulin-like growth factor type 1; IGFBP-3. insulin-like growth factors binding protein 3; ALS, acid-label subunit; IGFs, insulin-like growth factors.

GH-IGF Axis

Growth Hormone (GH) and Its Receptor (GHR)

GH is produced by somatotrophs in the anterior pituitary under the regulation of two main hypothalamic peptides: growth hormone-releasing hormone (GHRH) and somatostatin (SST) [8]. GHRH stimulates the synthesis and release of GH, whereas SST inhibits its release and regulates the pulsatile GH secretion. Besides GHRH, ghrelin, which is a peptide secreted mostly by gastric cells, can stimulate GH synthesis and release through its action on the GH secretagogue receptor (GHSR), located in the hypothalamus and pituitary. GH is secreted by the pituitary as monomeric and oligomeric isoforms, 22kDa being the main one, followed by 20kDa, both with biological activity. Approximately 50% of circulating GH is bound to GHBP, a high affinity binding protein that represents the GHR extracellular domain and is produced through GHR proteolysis.

GH acts by binding to growth hormone receptor (GHR; Figure 7.1.2.2), a transmembrane receptor that is a member of the class I cytokine receptor family. It has only one transmembrane domain and no intrinsic enzymatic activity, needing to be coupled to a tyrosine kinase called Janus kinase 2 (JAK2) to trigger the intracellular signal transduction. When GH binds to a pre-assembled receptor dimer, there is a conformational change in its intracytoplasmic region that allows JAK2 to become phosphorylated and then to phosphorylate tyrosine residues in GHR, which become binding sites for signalling molecules [8, 9]. In this manner, GHR and JAK2 activate signalling pathways that are common to other receptors, such as the mitogen-activated protein kinase (MAPK) and the

Table 7.1.2.1 Phenotype according to GH-IGF axis defects

	Gene	Inheritance	Growth disorder		Hormone measurement			Additional features
			Prenatal	Postnatal	GH	IGF-1	IGFBP-3	
GH deficiency	several	AR / AD	Nl	↓ to ↓↓↓	↓ to ↓↓↓	↓ to ↓↓↓	↓ to ↓↓↓	See Table 7.1.2.2
Bioinactive GH	*GH1*	AR / AD	Nl or ↓	↓↓↓	↑↑	↓↓	↓↓	Mild phenotype of Laron syndrome (*). IGF-1 increases after use of exogenous GH
GH insensitivity (GHI)	*GHR*	AR > AD	Nl or ↓	↓↓ to ↓↓↓	Nl to ↑↑	↓ to ↓↓↓	↓ to ↓↓↓	Phenotype of Laron syndrome in complete GHI. Low GHBP in 70% of cases. IGF-1 does not change after exogenous GH
GHI associated to immune dysfunction	*STAT5b*	AR > AD	Nl or ↓	↓↓↓	↑↑	↓↓	↓↓	Phenotype similar to Laron syndrome + immune dysfunction and elevated PRL in AR forms
Ternary complex defects	*IGFALS*	AR	Nl or ↓	↓	Nl to ↑	↓↓	↓↓↓	Laboratory abnormality disproportionate to mild short stature, pubertal delay
	PAPPA2	AR	Nl or ↓	↓ to ↓↓	↑↑	↑↑↑	↑↑↑	Mild microcephaly. Low free IGF-1 (*)
IGF1 defects	*IGF1*	AR/AD	↓↓↓	↓↓↓	↑↑	↓↓↓↓	Nl or ↑	Microcephaly, sensorineural deafness, developmental delay, intellectual disability, insulin resistance
IGF2 defects	*IGF2*	ADp	↓↓↓	↓↓ to ↓↓↓	Nl or ↑	Nl or ↑	Nl or ↑	Phenotype resembling Silver-Russell syndrome and low IGF-2 levels (*)
Bioinactive IGF-1	*IGF1*	AR	↓↓↓	↓↓↓	↑↑	↑↑↑↑	Nl	Microcephaly, sensorineural deafness, developmental delay, intellectual disability, insulin resistance (*)
IGF insensitivity	*IGF1R*	AD > AR	↓ to ↓↓↓	↓ to ↓↓↓	Nl to ↑↑	Nl to ↑↑	Nl to ↑↑	Variable clinical findings in the dominant forms. Microcephaly, developmental delay, and intellectual disability may be present

AR, autosomal recessive; AD, autosomal dominant; ADp, AD with paternal transmission; Nl, normal; ↑, increase; ↓, decrease; PRL, prolactin. *, phenotype based on a limited number of patients/families.

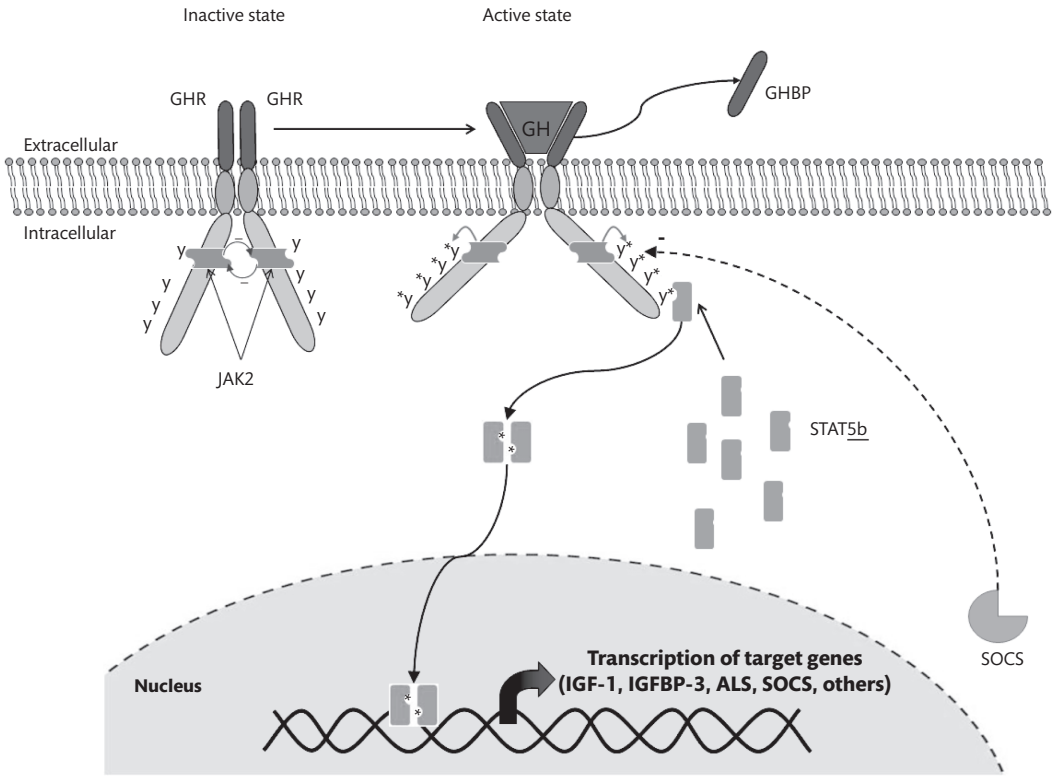


Figure 7.1.2.2 Schematic representation of GH-GHR signalling through JAK2-STAT5B pathway. In the inactive state, two GHR molecules are predimerized in the cell membrane. Each JAK2 protein bound to the intracytoplasmic portion of one receptor inhibits the catalytic activity of the other. Once GH binds to the two receptors, a change of the conformational structure takes place, moving apart the two JAK2 molecules that can then phosphorylate sites of tyrosine in themselves as well as in GHR. Tyrosine-phosphorylation creates docking sites for signalling proteins such as STAT5Bs that become phosphorylated, allowing their dimerization, transport to the nucleus and activation of transcription of target genes (IGF-1, IGFBP-3, ALS, and others). This signalling pathway also stimulates the synthesis of SOCS, which ultimately inhibits this signalling process.

inositol-1,4,5-trisphosphate 3-kinase (IP-3K) pathways, which are important to the metabolic and proliferative effects of GH. Besides these pathways, cytokine receptors have a unique signalling pathway that uses cytoplasmic proteins known as signal transducer and activator of transcription (STAT). Among the seven STAT proteins currently known, the most important for GHR signal transduction in humans is STAT5B. This protein, when phosphorylated, dimerizes, and translocates into the nucleus, where its binding to GH responsive elements in DNA stimulates the transcription of genes such as IGF-1, IGFBP3, and ALS [8, 9]. This signalling pathway also stimulates the synthesis of suppressors of cytokine signalling (SOCS), which ultimately inhibit this signalling process.

- **Insulin-like growth factor type 1 (IGF-1) and its receptor (IGF-1R)**

Most of the growth-promoting actions of GH are mediated by IGF-1 [10]. It has an essential role in pre- and postnatal growth, brain, and inner ear development. *In utero*, its secretion is independent of GH and regulated mainly by nutritional status while after birth its circulating fraction is mainly produced by the liver in response to GH. It also has a paracrine/autocrine action, being secreted by most tissues. The role of circulating versus locally produced IGF-1 in growth is still a matter of debate [11]. Circulating IGF-1 is bound to high affinity binding proteins, the insulin-like growth factor-binding proteins (IGFBP). Currently, six IGFBPs are well characterized. These proteins regulate IGF-1's availability to bind to its receptor IGF-1R and also have independent roles in the modulation of autocrine, paracrine, and endocrine IGF-1 actions. Some 70–80% of circulating IGF-1 is found in a ternary complex with IGFBP-3 and ALS, around 20% is bound to others IGFBPs and less than 5% is found in the free form. IGF-1 found in the ternary complex is unable to reach the extravascular compartment and consequently to act on target tissues. Through this mechanism, the ternary complex increases IGF-1 half-life from minutes to several hours and modulates its biological activity. Another IGF, IGF-2, is the most important factor in the regulation of intrauterine growth. Interestingly, the *IGF2* gene is located in an imprinting region where the paternally inherited allele is preferentially expressed [12].

Most of IGF-1 and IGF-2 actions are mediated by IGF-1R, a tyrosine kinase receptor structurally similar to the insulin receptor (Figure 7.1.2.3). When IGF-1 or IGF-2 binds to a homodimer of IGF-1R, there is a conformational change in the intracytoplasmic region that allows trans-autophosphorylation and creates docking sites for signalling molecules [13]. The main protein activated by IGF-1R is insulin receptor substrate 1 (IRS-1), a hydrophilic phosphoprotein that recruits and regulates the activity of other intracellular proteins. IRS-1 acts as an adaptor molecule and activates signalling cascades associated with tyrosine kinase receptors. One of the most important signalling pathways of IGF-1R is MAPK, which is responsible for the proliferative effects of IGF. Another

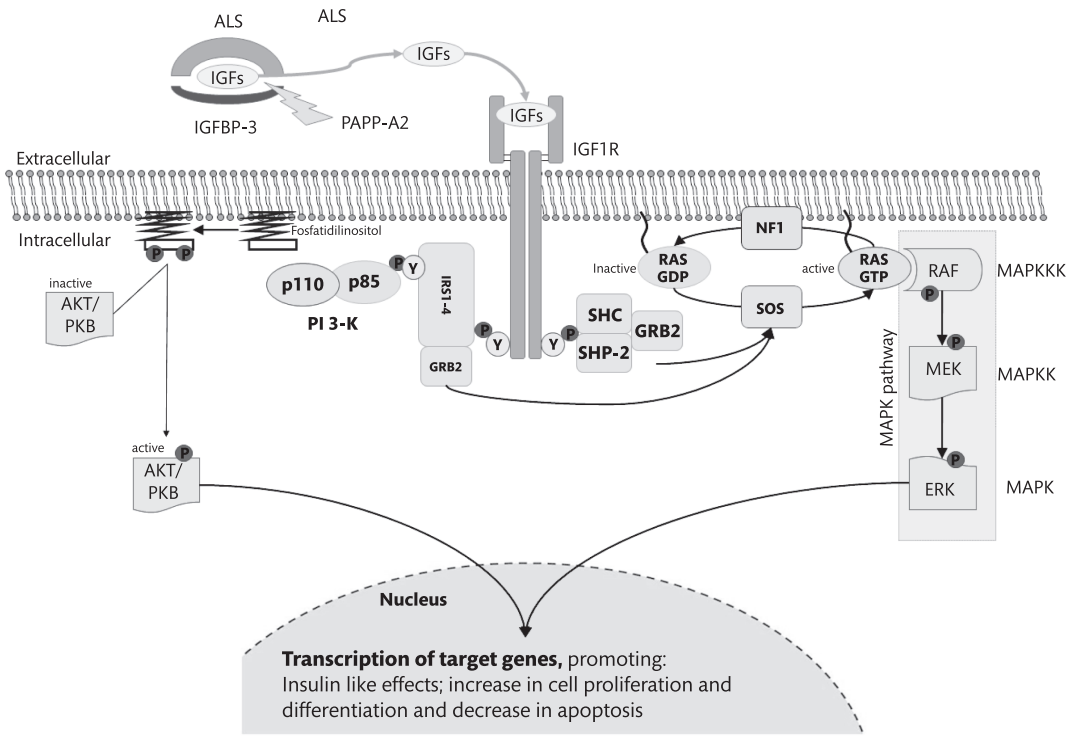


Figure 7.1.2.3 Schematic representation of IGF-IGF-1R signalling through RAS/MAPK and PI3K/PBK pathways. Once IGF-1 or IGF-2 binds to a homodimer of IGF-1R, there is a conformational change in the intracytoplasmic region that allows trans-autophosphorylation and creates docking sites for signalling molecules. Several proteins bind to these phosphorylated tyrosine sites and become activated. This initiates a cascade of events that activate RAS/MAPK and PI3K/PBK pathways that promote the genomic effects triggered by the IGFs.

signalling pathway is PI-3K, responsible for the metabolic effects of IGF, which are similar to insulin effects.

Growth Hormone Deficiency (GHD)

GHD can be either isolated (IGHD) or combined with other pituitary hormones deficiencies (CPHD). It can be congenital, usually due to a genetic defect, or acquired, due to tumours in the hypothalamic-pituitary region, neurosurgery, radiotherapy, autoimmune, and granulomatous diseases and even brain trauma. It is not unusual that patients with congenital/genetic causes present with associated complex phenotypes. The most common are: structural changes of the hypothalamic-pituitary region and/or the central nervous system (CNS); craniofacial malformations; septo-optic dysplasia (SOD); and developmental delay/intellectual disability [14]. Considering the genetic defects that cause hypopituitarism, we can classify them based on the main mechanism of pituitary dysfunction (Table 7.1.2.2).

- **Genetic defects in GH secretion leading to isolated growth hormone deficiency (IGHD)**

Classically, familial IGHD has been classified into four types depending on the inheritance pattern: autosomal recessive (types IA, OMIM 262400 and IB, OMIM 612781), autosomal dominant (type II, OMIM 173100), and X-linked (type III, OMIM 307200) [8]. Mutations in *GH1* can cause IGHD types IA, IB, and II, while IGHD type IB can also be caused by mutations in growth hormone-releasing hormone receptor (*GHRHR*). In relation to IGHD type III, a X-linked GHD associated with agammaglobulinemia has been described: although the *BTK* (Bruton tyrosine kinase) gene, a key regulator of B cell development, has been associated with this condition, its genetic aetiology remains unknown. Additionally, homozygous and heterozygous mutations in the GHSR, also known as ghrelin receptor, have been associated with partial IGHD (OMIM 615925). On neuroimaging, regardless of the genetic defect, patients have normal or hypoplastic anterior pituitary with a normally sited posterior pituitary 'bright' signal [15].

This traditional classification was made before the exact genetic defects were known. In this chapter, we have chosen to focus on the most well characterized genetic defects and mentioned the former classification for completeness.

GH1

Mutations in the *GH1* (growth hormone 1) gene are an excellent example of how different genetic defects and modes of inheritance lead to distinct phenotypes. Homozygous deletions of the *GH1* gene locus ranging from 6.7 to 45 kB lead to severe forms of IGHD. Patients usually present within the first 6 months of life with typical clinical features associated with complete GH deficiency: severe growth failure (height standard deviation score (SDS) <–3), low growth velocity, hypoglycaemia, prolonged jaundice, mid-facial hypoplasia, protruding forehead, saddle nose, high-pitched voice, and truncal obesity [8, 16]. Typically, they have undetectable IGF-1 and GH concentrations. These patients tend to develop high titres of anti-GH antibodies after hormone replacement therapy is initiated, which can neutralize GH actions and therefore lead to failure of response to recombinant human GH (r-hGH) therapy. Microdeletions

Table 7.1.2.2 Genes associated with growth hormone deficiency either isolated (IGHD) or combined (CPHD) and their clinical features

Gene	Inheritance	Hormonal deficiencies	Anterior pituitary	Posterior pituitary	Other features
GH1	Recessive	IGHD	NL or ↓	Topic	Severe IGHD, depending on the genetic defect anti-GH antibodies leads to ineffective treatment with r-hGH
	Dominant	IGHD, may evolve with other deficiencies	NL or ↓	Topic	Variable phenotypes regarding height and time of presentation. patients may develop additional pituitary hormone deficiencies
GHRHR	Recessive	IGHD	↓	Topic	Severe IGHD
PROP1	Recessive	GH+TSH+PRL+LH/FSH may evolve with ACTH deficiency	↑, NL or ↓	Topic	Hormonal deficiencies can emerge asynchronously
POU1F1	Recessive/dominant	GH+TSH+PRL	NL or ↓	Topic	
GLI2	Dominant	IGHD or CPHD	↓	Ectopic/NV	Polydactyly, HPE, craniofacial malformations
HESX1	Recessive/dominant	IGHD or CPHD	A, NL or↓	Topic or ectopic	SOD
LHX3	Recessive	CPHD	A, ↑, NL or ↑	Topic	Deafness, Limited neck rotation, elevated shoulders
LHX4	Dominant	CPHD	↓	Topic or ectopic	Chiari Malformation
SOX2	Dominant	HH + GHD	↓	Topic	Anophthalmia, forebrain defects, oesophagus atresia, deafness, learning difficulties
SOX3	X-linked	IGHD or CPHD	↓	Ectopic	Intellectual disability, infundibular hypoplasia, corpus callosum abnormalities
FGF8	Recessive	HH, IGHD, or CPHD	↑ or NL	Topic	SOD, HPE, KS, Moebius syndrome
FGFR1	Dominant	HH or CPHD	NL or ↓	Ectopic or topic	SOD, midline craniofacial malformations, corpus callosum abnormalities
PAX6	Dominant	IGHD or CPHD	↓	Topic	midline craniofacial malformations, ophthalmologic abnormalities
GLI3	Dominant	IGHD or CPHD	A or ↓	Topic	Pallister Hall syndrome
ARNT2	Recessive	CPHD	↓	Topic	Brain, eye, kidney, and urinary tract abnormalities, corpus callosum abnormalities
CDON	Dominant	CPHD	A	Topic or ectopic	HPE, maternal ethanol exposure worsens the phenotype
GPR161	Recessive	IGHD or CPHD	↓	Ectopic	
IGSF1	X-linked	CPHD	NL	Topic	Macro-orchidism, undetectable prolactin
PROKR2	Dominant or recessive	HH, IGHD, or CPHD	NL or ↓	Topic or Ectopic	SOD, Hirschsprung disease
TGIF1	Dominant	CPHD	↓	Ectopic	HPE, midline craniofacial malformations

IGHD, isolated growth hormone deficiency; CPHD, combined pituitary hormone deficiencies; HH, hypogonadotropic hypogonadism; A, aplastic; NL, normal; ↓, small; ↑, large; NV, non-visualized; r-hGH, recombinant human growth hormone; SOD, septo-optic dysplasia, HPE, holoprosencephaly, KS, Kallmann syndrome, DD/ID, developmental delay/intellectual disability.

and point mutations have also been reported to cause this phenotype, known as IGHD type IA [17]. In contrast, other homozygous mutations—splice-site, frameshift, or nonsense—cause a less severe phenotype also with marked short stature, delayed bone age, and low GH concentrations but with a good response to r-hGH without neutralizing antibody formation. This form is known as IGHD type IB.

In the autosomal dominant mode of inheritance, known as IGHD type II, patients present with low but usually detectable GH, a variable height deficit and time of presentation and they may or may not have anterior pituitary hypoplasia on magnetic resonance imaging (MRI) (38–50%) [18]. These patients may develop additional pituitary hormone deficiencies (ACTH, prolactin, TSH, and/or gonadotropin deficiencies) needing lifelong follow-up. Most mutations have been shown to affect the correct splicing of *GH1*, including single base mutations in the first six nucleotides of intron-3, mutations in the exonic splice enhancers ESE1 and ESE2 or disruption of sequences downstream of the consensus splicing sites that affect intronic splice enhancers. The result is the skipping of exon 3 and the production of a 17.5 kDa isoform that exerts a dominant-negative effect on the secretion of the 22 KDa molecule. In transgenic mice studies the overexpression of 17.5 kD isoform showed defective GH secretory vesicles with anterior pituitary hypoplasia, loss of somatotropes, and invasion by macrophages. The most severely affected animal developed other anterior pituitary cell line deficiencies [19]. There are also missense mutations with complex mechanisms of action which disturb GH secretion or reduce the binding affinity for the GH receptor.

GHRHR

The *GHRHR* gene encodes a G-protein coupled receptor. Its expression is upregulated by POU class 1 homeobox 1 (*POU1F1*, formerly called PIT1) and is required for the proliferation of somatotropes.

Patients with pathogenic variants in this gene present with early and severe growth failure, typical facial features, and have a good response to r-hGH therapy; other clinical findings such as neonatal hypoglycaemia and micropenis are absent [20]. On MRI, most patients have anterior pituitary hypoplasia with a normally placed posterior pituitary signal; however, there are some reports of a normal-sized anterior pituitary. These patients are also classified as having IGHD type IB (OMIM 612781). Homozygous or compound heterozygous mutations have been reported (missense, nonsense, splice-site, deletions, or regulatory mutations) and patients are usually of consanguineous pedigrees or certain ethnic backgrounds—Indian subcontinent and Brazil. There is a report of a heterozygous mutation in the signal peptide leading to a dominant form of inheritance with variable penetrance [15].

- **Genetic defects in pituitary cell differentiation leading to CPHD**

PROP1

PROP1 (Prophet of *POU1F1*) encodes a 226-amino acid paired-like homeodomain transcription factor which activates *POU1F1* expression and pituitary organogenesis. Currently, bi-allelic *PROP1* mutations are the most frequently recognized genetic cause of CPHD worldwide [21, 22]. In general, short stature is the first symptom reported, probably due to combined GH and TSH deficiency. Growth failure usually develops within the first year of life (height −1.5 ± 0.9 SDS at 1.5 years of age) and becomes more prominent later in childhood between the ages of 1.5 and 3 years (−3.6 ± 1.3 SDS at 3 years of age), when parents search for medical assistance [23]. At diagnosis, bone age is usually severely delayed (−4.0 SDS; 2.5 y) [24]. Gonadotropins deficiency may be manifested as a lack of pubertal development or a failure to complete puberty [25]. The hormonal deficiencies can emerge asynchronously over time, even among individuals carrying the same genotype. Some patients develop corticotrophin deficiency, so all patients with *PROP1* pathogenic variants deserve lifetime clinical surveillance [26]. Furthermore, as GH replacement can increase cortisol metabolism, it is necessary to be aware of the signs of an unveiled adrenal insufficiency [27]. On neuroimaging the pituitary stalk is normal and the posterior lobe is in the normal position. The anterior lobe in patients with *PROP1* mutations is usually hypoplastic, but may be normal or even enlarged, which may be misdiagnosed as a tumour. The pituitary can wax and wane in size before undergoing complete involution for reasons as yet undetermined [28].

PROP1 mutations lead to CPHD in an autosomal recessive inheritance pattern, therefore the prevalence of these mutations is higher in familial cases and in patients born to consanguineous parents. Different types of molecular defects have been described disrupting *PROP1* function, varying from complete gene deletion to frameshift, small deletions and insertions, and point mutations including missense, nonsense, splicing variants, and mutations affecting the initiation codon [22].

POU1F1

POU class 1 homeobox 1 (formerly called PIT1) is a member of the POU family of transcription factors that regulate mammalian development and is important for the development of GH, PRL, and TSH lineages. Usually GH and PRL deficiencies are detected at first presentation and severe short stature is the first complaint. TSH deficiency can be present from infancy and such patients may present with severe mental retardation, or it can emerge later in life [29]. On MRI, the anterior pituitary can be small or normal and the posterior pituitary is eutopic [30]. The first reports were about homozygous mutations but heterozygous mutations with dominant-negative effects have also been described, thus recessive and dominant inheritance modes are possible [31].

Genetic Defects in Pituitary Development Leading to IGHD or CPHD

Congenital hypopituitarism can be associated with abnormalities in the development of the anterior craniofacial midline, forebrain, and eye. This clinical picture and the growing knowledge of the genetic cascade of signalling molecules and transcription factors involved in hypothalamic-pituitary embryogenesis has led to the discovery of many genetic defects in early developmental genes. As a consequence, overlap with other conditions such as isolated hypogonadotropic hypogonadism and holoprosencephaly is seen in many patients [8]. A well-established clinical syndrome is the pituitary stalk interruption syndrome (PSIS) characterized by congenital hypopituitarism, absent or thin stalk, anterior pituitary hypoplasia, and ectopic posterior pituitary [32]. After a period of extensive study of individual developmental genes, the advent of large parallel sequencing methods like whole-exome sequencing (WES) has led to the discovery of novel genetic factors involved in hypopituitarism and its associated phenotypes. There is an ever-increasing list of candidate genes that is updated on a monthly basis. Here we describe in more detail three genes classically associated with hypopituitarism with defects in early developmental factors. Other genes are cited in Table 7.1.3.3 and Fang *et al.* [32].

GLI2

The GLI family zinc finger 2 is a mediator of Sonic hedgehog (Shh) signalling and plays a role during embryogenesis. Mutations in *GLI2* were at first associated with holoprosencephaly but the range of possible phenotypes has expanded and includes IGHD or CPHD with or without diabetes insipidus. Other complex phenotypes are developmental delay, seizures, polydactyly, midline craniofacial malformations, and cryptorchidism. On MRI there is a small anterior pituitary with a non-visualized or ectopic posterior pituitary. The mode of inheritance is autosomal dominant with incomplete penetrance. The variants classified as pathogenic are: complete gene deletions; nonsense or frameshift mutations that result in protein truncation; mutations in the universal splicing sites; and mutations within the zinc finger region [33].

HESX1

HESX homeobox 1 encodes a transcriptional repressor in the developing forebrain and pituitary gland. Homozygous mutations (missense, frameshift, splice donor site) in *HESX1* result in CPHD usually with associated phenotypes: SOD, forebrain midline malformations (absent septum pellucidum and agenesis of the corpus callosum), and ophthalmic abnormalities. Heterozygous mutations have incomplete penetrance and lead to a less severe phenotype with IGHD or CPHD and various degrees of SOD. The hypothalamic-pituitary

region shows hypoplasia or aplasia of the anterior pituitary with normal, non-visualized or ectopic posterior pituitary [31, 34].

OTX2

Orthodenticle homeobox 2 encodes a transcription factor that plays a role in brain, craniofacial, and sensory organ development. Heterozygous mutations in *OTX2* were initially associated with anophthalmia or microphthalmia, later it was observed that some patients had IGHD or CPHD. On neuroimaging the anterior pituitary can be normal or hypoplastic and the posterior pituitary can be normal or ectopic. Additionally, there are reports of absent or severely hypoplastic infundibulum [31, 34].

Bioinactive GH (Kowarski Syndrome, OMIM 262650)

The existence of defects in the GH molecule generating a biologically inactive but immunoreactive GH was first proposed by Kowarski *et al.* in 1978 [35]. These authors evaluated two children with proportional short stature, normal GH secretion and very low somatomedin (IGF-1) level. Contrary to what would be expected in a patient with GH insensitivity, these patients normalized their growth rate and IGF-1 levels with exogenous GH treatment. Although the molecular basis in these first patients was not elucidated, Takahashi *et al.* published two other patients at the end of the 1990s with similar phenotypes associated with heterozygous pathogenic variants in the *GH1* gene [36, 37].

After these initial reports, a small number of other patients with bioinactive GH have been reported [38, 39]. These patients had a common phenotype characterized by severe short stature (usually height SDS <–3), very low levels of IGF-1/IGFBP-3 and normal or elevated GH levels (basal or after a stimulation test). Some patients presented with mild clinical characteristics observed in severe GHD (prominent forehead and saddle nose). All patients showed some improvement in growth rate and IGF-1 concentrations with r-hGH treatment. In one patient, this improvement was suboptimal, possibly due to a marked dominant-negative effect of mutated GH [36]. The majority of the cases had an autosomal dominant inheritance pattern, but an autosomal recessive inheritance was observed in one family [38].

The mutated GH molecules were capable of binding to the receptor with lower, similar, or even with higher affinity than wild-type GH, but all of them were unable to properly activate the signal transduction by GH receptor. Mutations associated with an autosomal dominant inheritance had a dominant-negative effect on wild-type GH action by competing for binding to the receptor and thus preventing the wild-type GH from activating the receptor [36, 37]. This is analogous to the mechanism of action of pegvisomant, a GH receptor (GHR) antagonist used in the treatment of conditions of GH excess [40].

GH Insensitivity (GHI)

In the 1960s, Laron *et al.* [41] described three siblings from a consanguineous Jewish family with severe short stature and a physical appearance that resembled patients with complete GH deficiency. After the development of a methodology to measure circulating GH, it was observed that these patients had markedly elevated GH levels. In the subsequent years, other patients with a similar phenotype were identified by the same group. Multiple and sequential investigations allowed them to demonstrate that the circulating GH in these patients had normal biological activity, but ^{125}I-labelled GH was unable to bind to the liver membrane from affected patients, suggesting that the defect was located in GHR [41]. Only in 1989, after *GHR* was cloned and characterized, homozygous defects in this gene were demonstrated to cause severe GH insensitivity (GHI) [42, 43] and this presentation became known as Laron syndrome (OMIM 262500). In the last three decades, more than 250 patients with GHI have been described, expanding our knowledge on phenotype variability and associated molecular defects.

Complete GH Insensitivity (OMIM 262500)

Defects in *GHR* cause variable degrees of insensitivity to GH. The classical form of GHI, also known as Laron syndrome, is characterized by severe short stature observed during childhood (height SDS usually <–4) associated with dysmorphic features also presented in patients with complete congenital GH deficiency (craniofacial disproportion with a relatively small face, depressed nasal bridge, high-pitched voice, truncal obesity, and micropenis) [4, 41, 44]. The adult height of non-treated patients is around 40 centimetres or more below the population mean. These patients have very low levels of IGF-1, IGFBP-3, ALS, and GHBP, but elevated basal and stimulated GH levels. This condition is inherited in an autosomal recessive pattern, usually involving consanguineous parents or families from inbred communities.

Important insights on longevity of patients with complete GHI have been published in recent years. Despite obesity and a decrease in lean body mass due to a lifetime without GH action, these patients are protected from diabetes and cancer and have a longevity similar to controls [45]. In a similar manner, GHI animal models demonstrate an improvement of insulin sensitivity over time and extension of lifespan [46, 47].

Since the first description of homozygous defects in *GHR* causing complete GHI, more than 70 pathogenic variants in this gene have been identified [4], involving several types of mutations: gross deletions (6%); non-synonymous single nucleotide variants (41%) or premature stop codons (20%); small deletions causing a frameshift (14%); and nucleotide changes that directly affect the splicing process (19%). The majority of these mutations are located in exons that encode the extracellular domain of GHR (exons 4 to 7). The missense mutations usually disrupt the binding sites for GH and/or impair the transport of the receptor to the plasma membrane. Usually the diagnosis of these patients is not difficult and molecular studies add little additional information to the clinical assessment.

Partial GH Insensitivity (OMIM 604271)

After characterization of complete forms of GHI, several cases have been described with a milder phenotype. The clinical spectrum of partial GHI may vary from patients who are initially classified as having idiopathic short stature (ISS), to patients with typical laboratory findings of GHI who lack typical facial and dysmorphic features associated with Laron syndrome [4]. Two main molecular bases for partial GHI are described [4]: patients homozygous or compound heterozygous for hypomorphic mutations and patients heterozygous

for dominant-negative *GHR* mutations. In both situations, some residual GHR activity is retained causing variable partial phenotypes regarding the severity of growth impairment and hormonal profile. Additionally, GHBP is frequently detectable and even elevated in some cases. The absence of a typical phenotype raises a challenge for establishing the diagnosis of this condition based only on clinical and laboratory data. In this scenario, the use of genetic studies that analyse, besides *GHR*, several other genes described in this chapter can lead to a precise molecular diagnosis [48].

GHI Associated with Immune Dysfunction (OMIM 245590)

In 2003, Kofoed *et al.* described a girl with severe short stature (height SDS −7.5 with 16 years old), recurrent infections, and progressive respiratory failure. GH levels after a stimulation test with insulin were very high (53.8 μg/L), while IGF-1, IGFBP-3, and ALS were markedly low, even with r-hGH treatment, and GHBP was normal. The *GHR* gene was sequenced but it was normal. GHR signalling pathways were then analysed and a homozygous missense mutation was found in *STAT5B* gene, which prevented STAT5B activation by GH [49].

Since then, a total of ten patients have been reported harbouring seven different homozygous *STAT5B* mutations. These patients presented with severe postnatal growth failure (height SDS −3.0 to −9.9), normal to elevated GH levels and abnormally low IGF-1, IGFBP-3, and ALS, similar to Laron syndrome, but also with manifestations of immune dysregulation, including increased susceptibility for opportunistic infections and autoimmune-associated disorders like lymphocytic interstitial pneumonia and eczema [50]. The immune dysregulation is explained by the fact that other molecules use the STAT5B signalling pathway, including some interleukins. The defect in IL2 action is particularly important, since this interleukin participates in the activation of T lymphocytes and in the development of regulatory T lymphocytes [51]. Other unique features are the elevation of prolactin levels since its negative feedback also depends on STAT5B signalling and the normal levels of GHBP because the defect is intracellular and consequently does not affect the GHR extracellular domain [50]. Heterozygous carriers were approximately 4 cm shorter than non-carrier relatives. They also had lower IGF-1 and IGFBP-3 levels and a higher frequency of eczema than their wild-type relatives [52].

Besides these cases in which genetic inheritance was autosomal recessive, recently patients from three unrelated families with dominant-negative heterozygous *STAT5B* variants were reported to have postnatal growth failure, eczema, and elevated IgE but no severe immune diseases, expanding the variability of clinical and genetic presentations [53]). Moreover, there are some reports on patients with partial GHI and immune dysfunction who have *STAT3* gain of function variants associated with diminished *STAT5B* transcriptional activity [54].

Atypical GHI

The term atypical GHI refers to conditions that cause GHI, usually partial, due to mechanisms that do not directly involve genes on GHR-JAK2-STAT5B transduction pathway. This broad concept encompasses both acquired and congenital conditions. Several mechanisms of atypical GHI have been described: an increase of RAS/MAPK pathway activity, as observed in Noonan syndrome (OMIM 163950) [55]; an increase in NF-κB activity [56]; an increase in FGF21-FGFR1 signalling; and an increase of SOCS associated with chronic inflammatory state [56]. All these conditions are characterized by growth impairment associated with low IGF-1, normal GH secretion, and poor hormonal and growth responses to exogenous GH therapy. The study of these mechanisms will allow a better understanding of the interactions between the GH-IGF axis and several conditions that impair children's growth.

Ternary Complex Defects

Acid-Labile Subunit Deficiency (OMIM 615961)

In 2004, Domené *et al.* described an adolescent boy with mild short stature (height SDS—2.05), pubertal delay, and hyperinsulinemia with normal glucose levels, who had markedly reduced IGF-1 (5.3 SDS below the mean) and IGFBP-3 (9.7 SDS below the mean) and undetectable ALS [57]. A homozygous frameshift variant was found in the *IGFALS* gene, preventing the synthesis of this protein. Since then, more than 20 patients have been described, with mild to moderate short stature (height SDS −3.6 to −0.4, mean −2.3), undetectable or very low serum levels of ALS and very low serum levels of IGF-1 and IGFBP-3, with IGFBP-3 being more affected than IGF-1. Some of them also presented with variable reduction in birth size, pubertal delay, variable degrees of insulin insensitivity and/or low bone mineral density.

It was hypothesized that growth restriction is mild, despite the marked reduction in IGF-1 and IGFBP-3 levels (compatible with severe GHD or GHI) because of the preserved expression of locally produced IGF-1 [58]. The aetiology of hyperinsulinemia is still unknown; however, some authors have suggested that the reduced IGF-1 availability could impair glucose sensing and insulin sensitivity [59]. Genetic inheritance is autosomal recessive and patients are homozygous or compound heterozygous for ALS mutations. Heterozygous carriers have a mild reduction in height (height SDS lower by 0.9) in relation to their non-carrier relatives [60].

Defects in the Proteolytic Cleavage of IGFBPs (*PAPPA2* Gene)

Pregnancy-associated plasma protein A2 (PAPP-A2) is a protease highly specific for IGFBP-3 and -5, liberating IGF-1 from IGFBP-3 and ALS, and allowing free IGF-1 to exert its actions on target tissues. In 2016, two families were described with homozygous variants in *PAPPA2* gene. Their affected members presented with progressive growth failure (with normal or mildly affected birth length), microcephaly and thin long bones. They had a marked elevation of total IGF-1, IGFBP-3, IGFBP-5, and ALS concentrations, but low free IGF-1. GH is also elevated in the patients, probably due to decreased negative feedback secondary to low free IGF-1 concentrations. Genetic inheritance was autosomal recessive in both families.

Primary IGF Deficiency

IGF1 Defects (OMIM 608747)

IGF-1 deficiency was first described in humans in 1996, in a patient with severe intrauterine and postnatal growth failure, microcephaly,

sensorineural deafness, intellectual deficit and severe reduction in vertebral bone mass who had a homozygous partial deletion in the *IGF1* gene [61]. His basal GH and his peak after stimulation with clonidine were elevated, and IGFBP-3 and ALS were normal.

Since then, seven other index cases with *IGF1* variants were described, three of them with homozygous variants, inherited as an autosomal recessive trait. Besides the characteristics described in the first patient, they can also present with insulin resistance (secondary to GH excess), partial gonadal dysfunction, or obesity. One of the homozygous patients had a bioinactive IGF-1 (see next) and another had partial IGF-1 deficiency, presenting with pre- and postnatal growth retardation and microcephaly but with mild developmental delay and normal hearing [62]. Heterozygous parents of homozygous patients have a height SDS below the population mean, with some of them presenting with a height SDS lower than -2, compatible with a dose-dependent effect. Patients harbouring heterozygous mutations, inherited in an autosomal dominant manner, have a milder clinical presentation, with significant short stature but with variable presence of cognitive delay or hearing impairment [63–65].

IGF2 Defects (OMIM 616489)

In 2015, Begemann *et al*. described the first *IGF2* gene variant associated with growth restriction in humans [66]. Four members from one family presented with severe intrauterine (birth length SDS –4.2 to –4.9) and variable postnatal growth restriction. They also had clinical signs compatible with Silver-Russell syndrome (OMIM 180860), such as normal head circumference at birth, triangular face and micrognathia, and delayed psychomotor development. The IGF-2 level was low and IGF-1, IGFBP-3, and GH levels were normal to elevated. A heterozygous stop-gain variant, which introduces a premature termination codon, was found in the *IGF2* gene in seven family members, but only those who inherited the variant through paternal transmission were affected, a finding consistent with the maternal imprinting of the *IGF2* gene.

In 2018, Habib *et al* investigated a population of 192 patients with intrauterine growth retardation, suspected of Silver-Russell syndrome for whom a genetic or epigenetic diagnosis was not found. They found two *de novo* heterozygous variants in the *IGF2* gene, two heterozygous variants in *HMGA2* gene (one *de novo* and the other probably paternally inherited) and two heterozygous variants in *PLAG1* gene (one sporadic and one familiar case). *PLAG1* and *HMGA2* are oncogenes; PLAG1 activates *IGF2* through a promoter and HMGA2 is an upstream regulator of *PLAG1*. Both were associated with growth in mice models and in genome-wide association studies (GWAS) meta-analysis studies in humans [67].

Bioinactive IGF-1

In 2005, Walenkamp *et al*. described an adult patient who exhibited severe pre and postnatal growth failure (birth length –4.3 SDS and adult height –8.5 SDS), microcephaly, sensorineural deafness, mental retardation (IQ less than 40), and osteoporosis [68]. GH levels were very elevated in response to insulin-induced hypoglycaemia and IGF-1 was also markedly elevated (+7.3 SDS). They identified a missense variant in the *IGF1* gene and functional analysis showed a 90-fold reduced affinity of the mutated protein for the IGF-1 receptor. Heterozygous carriers had significantly lower birth weight, adult height (–1.0 height SDS, or 0.6 SDS below non-carrier relatives) and head circumference and higher fasting insulin than non-carriers.

IGF Insensitivity (OMIM 270450)

Various groups have searched for **IGF1 receptor (IGF1R)** variants in patients born SGA who did not show catch-up growth, because of the importance of this receptor for growth regulation and because of the abundance of mutations described in the insulin receptor, which is structurally closely related to IGF-1R. In 2003, Abuzzahab *et al*. [69] described the first two patients with intrauterine and postnatal growth failure associated with *IGF1R* gene defects. Since then, there have been many reports of other patients with *IGF1R* variants, the vast majority being heterozygous.

Patients with biallelic (homozygous or compound heterozygous) *IGF1R* defects have a clinical phenotype similar to the phenotype of patients with homozygous *IGF1* mutations. In contrast with this condition, which is rare and has a relatively homogeneous clinical presentation, heterozygous *IGF1R* variants are relatively frequent and manifest considerable clinical variability. In one study, two patients from a group of 100 SGA children without catch-up growth had heterozygous *IGF1R* variants [7]. The extent of growth restriction is variable as well as the degrees of psychomotor and mental development and the absence or presence of microcephaly. Bone age is delayed and IGF-1, IGFBP-3, and GH range from normal to high [13]. There are also reports of patients with abnormalities in chromosome 15 in which the *IGF1R* gene is deleted, with phenotypes varying according to the type and extent of the abnormality. As a rule, patients with biallelic *IGF1R* defects or heterozygous nonsense mutations have a more severe clinical presentation than patients heterozygous for missense mutations, reflecting a spectrum related to the residual activity of IGF-1 [13]. Due to the significant clinical variability, diagnosis based only on clinical and laboratory findings is not possible, reinforcing the importance of molecular genetic diagnosis for these patients.

Growth Promotion Treatment for Patients with GH-IGF Axis Disorders

In GHD, r-hGH is well established as an effective therapy for normalization of height during childhood [16]. Treatment should be initiated as soon as the diagnosis is made and is based on daily subcutaneous injections. The dose is usually adjusted based on body weight (25–50 μg/kg/day) every 3–6 months. In known CPHD, additional hormonal deficiencies should also be adequately replaced. In adult patients with persistent GHD, the treatment goals for r-hGH therapy are to improve body composition, bone mineral density and lipid metabolism [70].

Patients with GHI have, by definition, an inability to respond to r-hGH therapy. In complete GHI, r-hIGF-1 therapy is recommended. Nonetheless, the height gain achieved is inferior to the height gain seen in GHD children receiving r-hGH treatment [71, 72]. This lesser response could be explained in the following ways: a decrease in IGF-1 half-life due to the lack of other components of the ternary complex (i.e. IGFBP-3 and ALS); a reduction of IGF-1

generation in peripheral tissues, which compromises paracrine and autocrine IGF-1 effects; and the absence of GH dependent (IGF-1 independent) growth-promoting actions [71]. Treatment with r-hIGF-1 is administered as twice-daily subcutaneous injections after a carbohydrate-containing meal to reduce the risk of hypoglycaemia. For patients with partial GHI, a trial with r-hGH is advocated, which, if effective, is preferable to r-hIGF-1 for the reasons given earlier, and for safety and patient convenience [72]. In GHI associated with immune dysfunction, there has been only one report of treatment with r-hGH and r-hIGF-1, both without significant response [73]. This lack of response to r-hIGF-1 therapy could be partially explained by the impact of chronic illness.

Patients with *IGFALS* defects who have been treated with r-hGH [74] usually present a suboptimal growth response. There are no studies evaluating the response to r-hIGF-1 treatment, since these patients usually have mild short stature and can reach normal adult height spontaneously. In defects in the proteolytic cleavage of IGFBPs (*PAPPA2* gene), there is a report on two siblings who received r-hIGF-1, one with a modest growth response and improvement of insulin resistance and bone mass [75].

For patients with homozygous *IGF1* defects, treatment with r-hIGF-1 is the logical approach. Because of the rarity of these cases, this treatment was demonstrated in only one case with increased height velocity and markers of bone formation, and improved insulin sensitivity [76]. Patients with heterozygous defects in *IGF1* may show improvement in growth velocity and IGF-1 levels with r-hGH treatment, especially in high doses [64, 65]. Similar responses to r-hGH therapy have been reported in patients with heterozygous *IGF2* defects [66] and *IGF1R* haploinsufficiency [7, 77]. This last group of patients typically present with high IGF-1 levels during r-hGH treatment.

Future Perspectives

Traditionally, the diagnosis of disorders of the GH-IGF-1 axis was based on clinical and hormonal evaluations. In recent years, molecular genetic tests are progressively making these diagnoses easier and more accurate. Moreover, the use of next-generation sequencing techniques will probably allow the discovery of new genes associated with these disorders and the identification of genetic defects in GH-IGF-1 axis in atypical or mild cases. Additionally, the participation of genetic variants in this axis in the pathophysiology of common short stature should be further evaluated. The expansion of our knowledge in this field may lead to treatment based on the molecular defect.

REFERENCES

1. Blum W, Alherbish A, Alsagheir A, *et al*. The growth hormone-insulin-like growth factor-I axis in the diagnosis and treatment of growth disorders. *Endocr Connect*, 2018; **7**: R212–22.
2. Sizonenko PC, Clayton PE, Cohen P, Hintz RL, Tanaka T, Laron Z. Diagnosis and management of growth hormone deficiency in childhood and adolescence. Part 1: diagnosis of growth hormone deficiency. *Growth Horm IGF Res*, 2001; 11: 137–65.
3. Shen Y, Zhang J, Zhao Y, Yan Y, Liu Y, Cai J. Diagnostic value of serum IGF-1 and IGFBP-3 in growth hormone deficiency: a systematic review with meta-analysis. *Eur J Pediatr*, 2015; 174: 419–27.
4. David A, Metherell LA, Clark AJ, Camacho-Hubner C, Savage MO. Diagnostic and therapeutic advances in growth hormone insensitivity. *Endocrinol Metab Clin North Am*, 2005; 34: 581–95, viii.
5. Dauber A, Rosenfeld RG, Hirschhorn JN. Genetic evaluation of short stature. *J Clin Endocrinol Metab*, 2014; 99: 3080–92.
6. Fayter D, Nixon J, Hartley S, *et al*. Effectiveness and cost-effectiveness of height-screening programmes during the primary school years: a systematic review. *Arch Dis Child*, 2008; 93: 278–84.
7. Ester WA, van Duyvenvoorde HA, de Wit CC, *et al*. Two short children born small for gestational age with insulin-like growth factor 1 receptor haploinsufficiency illustrate the heterogeneity of its phenotype. *J Clin Endocrinol Metab*, 2009; 94: 4717–27.
8. Ranke MB, Wit JM. Growth hormone—past, present and future. *Nat Rev Endocrinol*, 2018; 14: 285–300.
9. Waters MJ, Brooks AJ. Growth hormone receptor: structure function relationships. *Horm Res Paediatr*, 2011; 76 Suppl 1: 12–16.
10. Wang J, Zhou J, Cheng CM, Kopchick JJ, Bondy CA. Evidence supporting dual, IGF-I-independent and IGF-I-dependent, roles for GH in promoting longitudinal bone growth. *J Endocrinol*, 2004; 180: 247–55.
11. Cannata D, Vijayakumar A, Fierz Y, LeRoith D. The GH/IGF-1 axis in growth and development: new insights derived from animal models. *Adv Pediatr*, 2010; 57: 331–51.
12. Eggermann T, Eggermann K, Schonherr N. Growth retardation versus overgrowth: Silver-Russell syndrome is genetically opposite to Beckwith-Wiedemann syndrome. *Trends Genet*, 2008; 24: 195–204.
13. Klammt J, Kiess W, Pfäffle R. IGF1R mutations as cause of SGA. *Best Pract Res Clin Endocrinol Metab*, 2011; 25: 191–206.
14. Correa FA, Jorge AA, Nakaguma M, *et al*. Pathogenic copy number variants in patients with congenital hypopituitarism associated with complex phenotypes. *Clin Endocrinol (Oxf)*, 2018; 88: 425–31.
15. Alatzoglou KS, Webb EA, Le Tissier P, Dattani MT. Isolated growth hormone deficiency (GHD) in childhood and adolescence: recent advances. *Endocr Rev*, 2014; 35: 376–432.
16. Growth Hormone Research S. Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. GH Research Society. *J Clin Endocrinol Metab*, 2000; 85: 3990–3.
17. Phillips JA, Hjelle BL, Seeburg PH, Zachmann M. Molecular basis for familial isolated growth hormone deficiency. *Proc Natl Acad Sci U S A*, 1981; 78: 6372–5.
18. Hamid R, Phillips JA, 3rd, Holladay C, *et al*. A molecular basis for variation in clinical severity of isolated growth hormone deficiency type II. *J Clin Endocrinol Metab*, 2009; 94: 4728–34.
19. Salemi S, Yousefi S, Lochmatter D, *et al*. Isolated autosomal dominant growth hormone deficiency: stimulating mutant GH-1 gene expression drives GH-1 splice-site selection, cell proliferation, and apoptosis. *Endocrinology*, 2007; 148: 45–53.
20. Wajnrajch MP, Gertner JM, Harbison MD, Chua SC, Leibel RL. Nonsense mutation in the human growth hormone-releasing hormone receptor causes growth failure analogous to the little (lit) mouse. *Nat Genet*, 1996; 12: 88–90.

21. Wu W, Cogan JD, Pfäffle RW, *et al*. Mutations in PROP1 cause familial combined pituitary hormone deficiency. *Nat Genet*, 1998; 18: 147–9.
22. Madeira JL, Nishi MY, Nakaguma M, *et al*. Molecular analysis of brazilian patients with combined pituitary hormone deficiency and orthotopic posterior pituitary lobe reveals eight different PROP1 alterations with three novel mutations. *Clin Endocrinol (Oxf)*, 2017; 87: 725–32.
23. Lebl J, Vosáhlo J, Pfaeffle RW, *et al*. Auxological and endocrine phenotype in a population-based cohort of patients with *PROP1* gene defects. *Eur J Endocrinol*, 2005; 153: 389–96.
24. Dusatkova P, Pfäffle R, Brown MR, *et al*. Genesis of two most prevalent PROP1 gene variants causing combined pituitary hormone deficiency in 21 populations. *Eur J Hum Genet*, 2016; 24: 415–20.
25. Pavel ME, Hensen J, Pfäffle R, Hahn EG, Dörr HG. Long-term follow-up of childhood-onset hypopituitarism in patients with the PROP-1 gene mutation. *Horm Res*, 2003; 60: 168–73.
26. Mendonca BB, Osorio MG, Latronico AC, Estefan V, Lo LS, Arnhold IJ. Longitudinal hormonal and pituitary imaging changes in two females with combined pituitary hormone deficiency due to deletion of A301, G302 in the PROP1 gene. *J Clin Endocrinol Metab*, 1999; 84: 942–5.
27. Weaver JU, Thaventhiran L, Noonan K, *et al*. The effect of growth hormone replacement on cortisol metabolism and glucocorticoid sensitivity in hypopituitary adults. *Clin Endocrinol (Oxf)*, 1994; 41: 639–48.
28. Nascif SO, Vieira TC, Ramos-Dias JC, Lengyel AM, Abucham J. Waxing and waning of a pituitary mass in a young woman with combined pituitary hormone deficiency (CPHD) due to a PROP-1 mutation. *Pituitary*, 2006; 9: 47–52.
29. Pfäffle RW, DiMattia GE, Parks JS, *et al*. Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science*, 1992; 257: 1118–21.
30. Turton JP, Reynaud R, Mehta A, *et al*. Novel mutations within the POU1F1 gene associated with variable combined pituitary hormone deficiency. *J Clin Endocrinol Metab*, 2005; 90: 4762–70.
31. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson IC, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. *Endocr Rev*, 2009; 30: 790–829.
32. Wang CZ, Guo LL, Han BY, Su X, Guo QH, Mu YM. Pituitary stalk interruption syndrome: from clinical findings to pathogenesis. *J Neuroendocrinol*, 2017; 29: doi: 10.1111/jne.12451.
33. Arnhold IJ, Franca MM, Carvalho LR, Mendonca BB, Jorge AA. Role of GLI2 in hypopituitarism phenotype. *J Mol Endocrinol*, 2015; 54: R141–50.
34. Kelberman D, Dattani MT. Role of transcription factors in midline central nervous system and pituitary defects. *Endocr Dev*, 2009; 14: 67–82.
35. Kowarski AA, Schneider J, Ben-Galim E, Weldon VV, Daughaday WH. Growth failure with normal serum RIA-GH and low somatomedin activity: somatomedin restoration and growth acceleration after exogenous GH. *J Clin Endocrinol Metab*, 1978; 47: 461–4.
36. Takahashi Y, Kaji H, Okimura Y, Goji K, Abe H, Chihara K. Brief report: short stature caused by a mutant growth hormone. *N Engl J Med*, 1996; 334: 432–6.
37. Takahashi Y, Shirono H, Arisaka O, *et al*. Biologically inactive growth hormone caused by an amino acid substitution. *J Clin Invest*, 1997; 100: 1159–65.
38. Besson A, Salemi S, Deladoey J, *et al*. Short stature caused by a biologically inactive mutant growth hormone (GH-C53S). *J Clin Endocrinol Metab*, 2005; 90: 2493–9.
39. Petkovic V, Miletta MC, Boot AM, *et al*. Short stature in two siblings heterozygous for a novel bioinactive GH mutant (GH-P59S) suggesting that the mutant also affects secretion of the wild-type GH. *Eur J Endocrinol*, 2013; 168: K35–43.
40. Tritos NA, Biller BM. Pegvisomant: a growth hormone receptor antagonist used in the treatment of acromegaly. *Pituitary*, 2017; 20: 129–35.
41. Laron Z. Laron syndrome (primary growth hormone resistance or insensitivity): the personal experience 1958–2003. *J Clin Endocrinol Metab*, 2004; 89: 1031–44.
42. Amselem S, Duquesnoy P, Attree O, *et al*. Laron dwarfism and mutations of the growth hormone-receptor gene. *N Engl J Med*, 1989; 321: 989–95.
43. Godowski PJ, Leung DW, Meacham LR, *et al*. Characterization of the human growth hormone receptor gene and demonstration of a partial gene deletion in two patients with Laron-type dwarfism. *Proc Natl Acad Sci U S A*, 1989; 86: 8083–7.
44. Savage MO, Attie KM, David A, Metherell LA, Clark AJ, Camacho-Hubner C. Endocrine assessment, molecular characterization and treatment of growth hormone insensitivity disorders. *Nat Clin Pract Endocrinol Metab*, 2006; 2: 395–407.
45. Guevara-Aguirre J, Balasubramanian P, Guevara-Aguirre M, *et al*. Growth hormone receptor deficiency is associated with a major reduction in pro-aging signaling, cancer, and diabetes in humans. *Sci Transl Med*, 2011; 3: 70ra13.
46. Fontana L, Partridge L, Longo VD. Extending healthy life span—from yeast to humans. *Science*, 2010; 328: 321–6.
47. Basu R, Qian Y, Kopchick JJ. Mechanisms in endocrinology: lessons from growth hormone receptor gene-disrupted mice: are there benefits of endocrine defects? *Eur J Endocrinol*, 2018; 178: R155–R81.
48. Shapiro L, Chatterjee S, Ramadan DG, *et al*. Whole-exome sequencing gives additional benefits compared to candidate gene sequencing in the molecular diagnosis of children with growth hormone or IGF-1 insensitivity. *Eur J Endocrinol*, 2017; 177: 485–501.
49. Kofoed EM, Hwa V, Little B, *et al*. Growth hormone insensitivity associated with a STAT5b mutation. *N Engl J Med*, 2003; 349: 1139–47.
50. Hwa V, Nadeau K, Wit JM, Rosenfeld RG. STAT5b deficiency: lessons from STAT5b gene mutations. *Best Pract Res Clin Endocrinol Metab*, 2011; 25: 61–75.
51. Cohen AC, Nadeau KC, Tu W, *et al*. Cutting edge: decreased accumulation and regulatory function of CD4+ CD25(high) T cells in human STAT5b deficiency. *J Immunol*, 2006; 177: 2770–4.
52. Scalco RC, Hwa V, Domené HM, *et al*. STAT5B mutations in heterozygous state have negative impact on height: another clue in human stature heritability. *Eur J Endocrinol*, 2015; 173: 291–6.
53. Klammt J, Neumann D, Gevers EF, *et al*. Dominant-negative STAT5B mutations cause growth hormone insensitivity with short stature and mild immune dysregulation. *Nat Commun*, 2018; 9: 2105.
54. Gutierrez M, Scaglia P, Keselman A, *et al*. Partial growth hormone insensitivity and dysregulatory immune disease associated with de novo germline activating STAT3 mutations. *Mol Cell Endocrinol*, 2018; 473: 166–77.
55. De Rocca Serra-Nedelec A, Edouard T, *et al*. Noonan syndrome-causing SHP2 mutants inhibit insulin-like growth factor 1 release via growth hormone-induced ERK hyperactivation, which contributes to short stature. *Proc Natl Acad Sci U S A*, 2012; 109: 4257–62.

56. Wit JM, de Luca F. Atypical defects resulting in growth hormone insensitivity. *Growth Horm IGF Res*, 2016; 28: 57–61.
57. Domené HM, Bengolea SV, Martinez AS, *et al.* Deficiency of the circulating insulin-like growth factor system associated with inactivation of the acid-labile subunit gene. *N Engl J Med*, 2004; 350: 570–7.
58. Domené HM, Hwa V, Argente J, *et al.* Human acid-labile subunit deficiency: clinical, endocrine and metabolic consequences. *Horm Res*, 2009; 72: 129–41.
59. Heath KE, Argente J, Barrios V, *et al.* Primary acid-labile subunit deficiency due to recessive IGFALS mutations results in postnatal growth deficit associated with low circulating insulin growth factor (IGF)-I, IGF binding protein-3 levels, and hyperinsulinemia. *J Clin Endocrinol Metab*, 2008; 93: 1616–24.
60. Fofanova-Gambetti OV, Hwa V, Wit JM, *et al.* Impact of heterozygosity for acid-labile subunit (IGFALS) gene mutations on stature: results from the international acid-labile subunit consortium. *J Clin Endocrinol Metab*, 2010; 95: 4184–91.
61. Woods KA, Camacho-Hubner C, Savage MO, Clark AJ. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med*, 1996; 335: 1363–7.
62. Netchine I, Azzi S, Le Bouc Y, Savage MO. IGF1 molecular anomalies demonstrate its critical role in fetal, postnatal growth and brain development. *Best Pract Res Clin Endocrinol Metab*, 2011; 25: 181–90.
63. Batey L, Moon JE, Yu Y, *et al.* A novel deletion of IGF1 in a patient with idiopathic short stature provides insight Into IGF1 haploinsufficiency. *J Clin Endocrinol Metab*, 2014; 99: E153–9.
64. van Duyvenvoorde HA, van Setten PA, Walenkamp MJ, *et al.* Short stature associated with a novel heterozygous mutation in the insulin-like growth factor 1 gene. *J Clin Endocrinol Metab*, 2010; 95: E363–7.
65. Netchine I, Azzi S, Houang M, *et al.* Partial primary deficiency of insulin-like growth factor (IGF)-I activity associated with IGF1 mutation demonstrates its critical role in growth and brain development. *J Clin Endocrinol Metab*, 2009; 94: 3913–21.
66. Begemann M, Zirn B, Santen G, *et al.* Paternally inherited IGF2 mutation and growth restriction. *New Engl J Med*, 2015; 373: 349–56.
67. Abi Habib W, Brioude F, Edouard T, *et al.* Genetic disruption of the oncogenic HMGA2-PLAG1-IGF2 pathway causes fetal growth restriction. *Genet Med*, 2018; 20: 250–8.
68. Walenkamp MJ, Karperien M, Pereira AM, *et al.* Homozygous and heterozygous expression of a novel insulin-like growth factor-I mutation. *J Clin Endocrinol Metab*, 2005; 90: 2855–64.
69. Abuzzahab MJ, Schneider A, Goddard A, *et al.* IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med*, 2003; 349: 2211–22.
70. Molitch ME, Clemmons DR, Malozowski S, Merriam GR, Vance ML, Endocrine S. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; 96: 1587–609.
71. Backeljauw PF, Kuntze J, Frane J, Calikoglu AS, Chernausek SD. Adult and near-adult height in patients with severe insulin-like growth factor-I deficiency after long-term therapy with recombinant human insulin-like growth factor-I. *Horm Res Paediatr*, 2013; 80: 47–56.
72. Grimberg A, DiVall SA, Polychronakos C, *et al.* Guidelines for growth hormone and insulin-like growth factor-i treatment in children and adolescents: growth hormone deficiency, idiopathic short stature, and primary insulin-like growth factor-I deficiency. *Horm Res Paediatr*, 2016; 86: 361–97.
73. Pugliese-Pires PN, Tonelli CA, Dora JM, *et al.* A novel STAT5B mutation causing GH insensitivity syndrome associated with hyperprolactinemia and immune dysfunction in two male siblings. *Eur J Endocrinol*, 2010; 163: 349–55.
74. Domené HM, Hwa V, Argente J, *et al.* Human acid-labile subunit deficiency: clinical, endocrine and metabolic consequences. *Horm Res*, 2009; 72: 129–41.
75. Dauber A, Munoz-Calvo MT, Barrios V, *et al.* Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. *EMBO Mol Med*, 2016; 8: 363–74.
76. Woods KA, Camacho-Hubner C, Bergman RN, Barter D, Clark AJ, Savage MO. Effects of insulin-like growth factor I (IGF-I) therapy on body composition and insulin resistance in IGF-I gene deletion. *J Clin Endocrinol Metab*, 2000; 85: 1407–11.
77. Walenkamp MJ, de Muinck Keizer-Schrama SM, de Mos M, *et al.* Successful long-term growth hormone therapy in a girl with haploinsufficiency of the insulin-like growth factor-I receptor due to a terminal 15q26.2->qter deletion detected by multiplex ligation probe amplification. *J Clin Endocrinol Metab*, 2008; 93: 2421–5.

7.1.3 Short Stature in Children Born Small for Gestational Age

Anita C.S. Hokken-Koelega

Introduction

Small-for-gestational-age (SGA) is defined as a birth weight and/or length <–2 standard deviation scores (SDS) [1, 2]. As the aetiology of SGA is multifactorial and includes maternal lifestyle and obstetric factors, placental dysfunction, and numerous (epi)genetic abnormalities, SGA-born children comprise a heterogeneous group. The majority of SGA-born infants show catch-up growth to a normal stature, but 10% remains short [3]. For more than 30 years, studies have been performed in short children after SGA birth, including children with Silver–Russell syndrome (SRS). Studies have generally excluded short SGA children with major dysmorphic features

or a (suspected) syndrome, primordial dwarfism, or DNA repair disorder. Thus present knowledge and management, particularly on growth hormone (GH) treatment, are based on the results in non-syndromic short SGA/SRS children.

This chapter presents our current knowledge of the (epi)genetic causes of short stature for those born SGA, the health consequences of SGA, and the diagnostic approach and management of short SGA-born children, including the efficacy and safety of GH treatment.

Definition of Small for Gestational (SGA)

SGA is not a diagnosis but refers to the size at birth, defined as a birth weight and/or length of at least two SDS below the median for gestational age [1, 2]. Accurate information on gestational age, birth weight, and birth length and appropriate reference data are required to determine if a child is born SGA.

The term intrauterine growth retardation (IUGR) is often used synonymously with the term SGA. IUGR is, however, a prenatal diagnosis referring to a deceleration of intrauterine growth, based on at least two fetal size measurements by ultrasound. Being born SGA does not necessarily mean that IUGR has occurred. Conversely, infants with documented IUGR are not inevitably born SGA. Neonates with SGA can be born either term or preterm.

SGA can be the result of maternal and paternal factors, placental insufficiency or fetal (epi)genetic factors [1]. The aetiology of SGA is unknown in 60% of cases, but precise diagnoses will expand due to the rapid advancement in genetic testing.

Early Growth

Most children born SGA show sufficient postnatal catch-up growth and have a normal height at the age of 2 years, but approximately 10% remain short (height <–2 SDS) [3]. Short stature after SGA accounts for approximately 20% of all cases of short stature [4]. Most of the catch-up growth occurs during the first year and is near completion by the age of 2 years, but it can take up to the age of 4 years in those born very prematurely [3].

Preterm infants might show extrauterine growth restriction (EUGR), due to a complicated neonatal course. When they have a length or weight <–2 SDS at the corrected age of 3 months and fail to catch-up subsequently, their risk of short adult stature is similar to term SGA infants [5].

Low birth size has been associated with an increased risk for cardiovascular disease, hypertension, and type 2 diabetes mellitus in adult life [6]. However, it has been shown that these risks are mainly associated with a rapid postnatal gain in weight for length [7].

Endocrine Consequences

GH-IGF Axis

Approximately 60% of short SGA children show reduced spontaneous GH secretion during a 24-hour sampling period and 25% a low GH peak during GH provocation tests, while serum IGF-I and IGFBP-3 levels are lower than in healthy controls, but not below –2 SDS [8]. This indicates that the majority of short SGA children do not have severe GH deficiency. Some short SGA children have normal or high serum IGF-I levels, suggestive of IGF-I resistance.

Pubertal Development

Puberty starts in SGA children slightly earlier than average but pubertal progression is normal [9], and adrenarche is age-appropriate [10].

Bone Mineral Density

Bone mineral density (BMD), corrected for height, is in the low normal range in short SGA children [11] but similar in SGA adults with or without short stature [12].

Gonadal Function

In prepubertal boys without cryptorchidism, serum inhibin B levels are not different between those born small for gestational age (SGA) or appropriate for gestational age (AGA) [13]. In adulthood, no differences in serum levels of inhibin B, testosterone, LH, and FSH were found between men born SGA or AGA [14]. Young SGA women have normal serum AMH levels, indicating that they do not have a smaller follicle pool size than AGA women [15].

Metabolic and Cardiovascular Consequences

Body Composition

In SGA children, total and abdominal fat mass at 4 years of age is closely related to the rate of weight gain in the first 2 years of life [16]. Short SGA adults have a similar fat mass percentage as AGA adults, while SGA adults who had spontaneous catch-up have a higher fat mass percentage than AGA adults. The lean body mass percentage is lower in all SGA adults compared to AGA adults, regardless of degree of catch-up growth [17], suggesting that in SGA-born subjects, lean body mass is reprogrammed during fetal life, with long-lasting effects on body composition.

Insulin Resistance and Type 2 Diabetes Mellitus

Low birth weight is associated with a higher risk of insulin resistance and type 2 DM in adulthood [6]. Short SGA children have a 38% reduction in insulin sensitivity adjusted for BMI compared to short AGA children, with higher insulin levels [18]. At the age of 4 years, SGA children with spontaneous catch-up growth are more insulin-resistant than AGA children [16]. Rapid fat accumulation in the first months of life is associated with a lower insulin sensitivity in early adulthood [7]. Insulin sensitivity is similar in short SGA and AGA adults, but lower in SGA adults with spontaneous catch-up growth to a normal height [19].

Cardiovascular Risk Factors

Birth weight has an inverse association with risk factors for hypertension and cardiovascular disease in adults [6]. Blood pressure is increased in short children born SGA, mainly in those born preterm [20]. Short SGA adults have a higher systolic and diastolic

blood pressure compared to AGA adults, irrespective of whether they had remained short or had caught up [21].

Metabolic and transcriptomic profiles in SGA children aged 4–9 years showed increased cardiometabolic risks [22]. However, serum lipid levels are within the normal range in short SGA children and adults [17, 23, 24].

Carotid intima media thickness (cIMT), a marker for generalized atherosclerosis, is higher in SGA children with spontaneous catch-up growth than in AGA children of a similar height, weight, and BMI [25]. Similarly, cIMT is higher in SGA adults with spontaneous catch-up growth compared to short adults born either SGA or AGA [26].

Neurodevelopment, Psychosocial Consequences, and Health-Related Quality of Life

SGA children have an IQ within the normal range but on average lower than AGA children [27]. Among SGA children in epidemiologic studies, those born smallest had the lowest IQ and more school failure and problem behaviour (attention deficit and social behaviour), particularly children with persistent short stature [28]. Short SGA children also have lower scores on tests of social functioning and health-related quality of life compared to reference populations [29].

(Epi) genetic Diagnoses in Short Children Born SGA

Non-syndromic children with short stature after SGA form a heterogeneous group, with a wide spectrum of clinical symptoms and responses to GH treatment. For many years, only very limited (epi) genetic causes were known, but the increasing use of next-generation sequencing, whole exome sequencing (WES), chromosomal microarrays, RNA sequencing, and methylation arrays has resulted in the discovery of novel (epi)genetic causes of short stature after SGA birth. Some monogenic disorders are now being found in children previously considered to have idiopathic short stature (ISS). Most currently known monogenic primordial growth disorders present with short stature after an SGA birth as one of their clinical features. For clinical purposes, first the monogenic disorders with a normal head circumference and short stature (either proportionate or disproportionate) are presented (Table 7.1.3.1), followed by monogenic primordial disorders with microcephaly (Table 7.1.3.2) and finally the currently known imprinting disorders and methylation disturbances (Table 7.1.3.3). This division is arbitrary, because some genetic aberrations were first reported for children with a very characteristic phenotype (including microcephaly), while over time milder phenotypes have been recognized. The genetic disorders in short SGA children can also be presented according to a pathophysiological classification (see recent review [30]).

Monogenic Disorders with Normal Head Circumference and Proportionate Short Stature

Children with GH deficiency due to a *GH-1 gene* mutation or GH insensitivity as result of an inactivating mutation of *GHR*, or *PIK3R1* or an activating *STAT3* mutation, have a lower mean birth weight and birth length, but most of them do not fulfil the criteria for SGA. However, in the case of a very low serum IGF-I and low GH peak during a provocative test or in the case of lack of growth response during GH treatment in a short SGA child, one may consider testing for these genes, because short SGA does not exclude an abnormality in the GH-IGF pathway.

Heterozygous *IGFALS* mutations are found in short SGA children [31], but also reported in children with ISS. Serum IGF-I and IGFBP-3 levels are low. GH treatment may increase growth in children with a heterozygous *IGFALS* mutation [32]. Absence of acid-labile subunit (ALS) due to homozygous *IGFALS* mutations (OMIM 615961) results in low IGF-I and very low IGFBP-3, birth weights varying from −3.7 to −0.1 SDS [31], while the effects of GH treatment are unknown.

3-M syndrome is characterized by pre- and postnatal growth failure and is caused by mutations in *CUL7* (OMIM 273750), *OBSL1* (MIM 610991) or *CCDC8* (MIM 614205) [33]. It is associated with reduced *IGF2* expression and increased *H19* expression, as also found in Silver–Russell syndrome. Growth response to GH treatment is disappointing.

Recently, a paternal IGF2 gene mutation was reported which leads to features resembling Silver–Russell syndrome. GH treatment is likely to be as effective as in other genetic variants of SRS [34].

Floating–Harbor syndrome occurs due to heterozygous mutations in the *SRCA gene (*(OMIM 136140). Only 26% of such children are born SGA and the phenotype is often mild [35]. Data on the effects of GH treatment are very limited.

Mulibrey nanism (MIM 253250) is caused by biallelic *TRIM37* mutations. Most children are born SGA and remain short but the progressive cardiomyopathy and other clinical problems are much more important. There are no data on GH treatment.

Monogenic Disorders with Normal Head Circumference and Disproportionate Short Stature

In most genetic disorders associated with disproportionate short stature, the genes are already abnormally expressed during fetal life which often results in a lower birth length SDS than birth weight SDS. The more severe forms of skeletal dysplasia can be easily diagnosed in young children, but clinical features can be so mild that many children are initially labelled as short SGA of unknown origin or ISS. Body disproportion and skeletal abnormalities can become more abnormal when the child becomes older. In each short SGA child, the clinician should therefore regularly assess body proportions, particularly in the case of a relatively lower birth length than birth weight. When an SGA child has disproportionate short stature (sitting/height >2 SDS and/or arm-span width > height), genetic testing can be an efficient diagnostic approach. A description of all skeletal dysplasias and their underlying genetic aberrations is beyond the scope of this chapter, but the most frequent ones in relation to SGA and short stature are given in Table 7.1.3.1.

Turner syndrome (TS), either due to 45XO or mosaicism, may explain disproportionate short stature after SGA. Birthweight is lower than normal, but only 30% are born SGA. It is nevertheless important to exclude TS in all short SGA girls when the height is a considerable distance below target height (>1.6 SDS, based on national growth references). The disturbed prenatal and postnatal growth is caused by short-stature-homeobox (*SHOX)* haploinsufficiency [36]. A full description of the medical care for girls with TS is beyond the

Table 7.1.3.1 Monogenic disorders in short children born SGA with normal head circumference

Syndrome [OMIM]	Genetic defect	Inheritance	Incidence	Mean BW/ BL SDS	Clinical features	Laboratory data	Treatment
Monogenic disorders with normal head circumference and proportionate short stature							
GH deficiency	*GH1, GHRHR, BTK*	AR, AD X-linked	1:5.000	−0.9 /−0.6	Wide variation in height deficit	↓GH peak during GH stimulation test, ↓IGF-I, ↓IGFBP-3, ↓ALS	GH registered
Laron syndrome [262500]	*GHR*	AR or rarely AD	>300 cases	−0.6 /−1.6	Wide variation in height, midfacial hypoplasia	↑GH, ↓IGF-I, ↓IGFBP-3, ↓ALS, variable GHBP	IGF-I treatment moderately effective
Multisystem infantile-onset autoimmune disease [615952]	Act *STAT3*	AD	19 cases	−1.7/−1.8	Early-onset multiorgan autoimmune disease	↓IGF-I, ↑TSH, abnormal immunoglobulins	No data
SHORT syndrome [269880]	*PIK3R1*	AD	32 cases	−3.3 85% SGA	Hyperextensibility, ocular depression, teething delay, lipoatrophy, insulin resistance	↓IGF-I non-responding to GH injections	No data
ALS deficiency [615961]	*IGFALS*	AR	>65 cases	−2.2	Mild to moderate short stature	↓IGF-I, ↓IGFBP-3, ↓ALS	No data
3-M syndrome [273750, 612921, 614205]	*CUL7, OBSL1, CCDC8*	AR	≈200 cases	−3.1	Facial features, normal mental development, long and slender tubular bones, reduced AP diameter of vertebral bodies, delayed bone age	NI GH, IGF-I, IGFBP-3	Effect of GH considered insufficient
Silver–Russell variant [616489]	*IGF2*	AD, paternal	8 cases	−3.9/−4.6	Dysmorphic features like SRS, fulfilling the Netchine–Harbison criteria for SRS, including relative macrocephaly	↑/normal GH, normal IGF-1, ↑/normal IGFBP-3, ↓IGF-2	GH treatment likely as effective as in other genetic variants of SRS
Floating–Harbor syndrome [136140]	*SCRAP*	AD	>52 cases	−2.5	Proportionate short stature, delayed bone age and speech, triangular face. deep-set eyes, long eyelashes, bulbous nose, wide columella, short philtrum, thin lips	NI GH, IGF-I, IGFBP-3	Insufficient data
Mulibrey nanism [253250]	*TRIM37*	AR	>110 cases	−2.8	Progressive cardiomyopathy, characteristic facial features, hypo- or hypergonadotropic hypogonadism, type 2 DM, predisposition to Wilms tumour	NI GH, IGF-I, IGFBP-3	Insufficient data
Monogenic disorders with normal head circumference and disproportionate short stature							
SHOX-associated short stature [300582]	*SHOX*	AD	2–17% of short stature	−0.4/−1.1	Short forearm and lower leg, bowing of forearm and tibia, dislocation of ulna at elbow, Madelung deformity, muscular hypertrophy, radiologic signs at wrist and forearm	NI GH, IGF-I, IGFBP-3	GH has similar efficacy as in Turner syndrome; registered in many countries
Achondroplasia [100800]	Act *FGFR3*	AD	1:15 000–40 000	−0.7/−1.0	Rhizomelic limb shortening, frontal bossing, midface hypoplasia, exaggerated lumbar lordosis, limitation of elbow extension, genu varum, trident hand	NI GH, IGF-I, IGFBP-3	Effects of GH considered insufficient
Hypochondroplasia [146000]	Act *FGFR3*		1:15 000–40 000	–	Rhizomelic limb shortening, limitation of elbow extension, brachydactyly, relative macrocephaly, generalized laxity, specific radiologic features	NI GH, IGF-I, IGFBP-3	Effect of GH considered insufficient
Short stature with mild disproportion [616255]	*NPR2*	AD	1–2% of short SGA and ISS	−0.8/−2.3	↑ sitting height/height ratio, phenotypic or radiographic indicators of *SHOX* HI (but no Madelung deformity)	NI GH, IGF-I, IGFBP-3	Insufficient data
Brachydactyly type A1 [1112500]	*IHH*	AD	1.6% of short SGA and ISS	−/ −1.4	↑ sitting height/height ratio, shortening of middle phalanx of 2nd and 5th fingers with cone-shaped epiphyses	NI GH, IGF-I, IGFBP-3	Preliminary data GH treatment positive
Short stature with early-onset osteoarthritis or osteochondritis [165800]	*ACAN*	AD	1–2% of short SGA and ISS	−0.7/−1.5	Proportionate or disproportionate short stature, with or without advanced bone age, brachydactyly, early-onset osteoarthritis	NI GH, IGF-I, IGFBP-3	Insufficient data
Noonan syndrome [163950]	*PTPN11*	AD	1:1000–2500	−1.0	Short stature, facial dysmorphism, wide spectrum of congenital heart defects, coagulation defect	NI GH, IGF-I, IGFBP-3	GH registered in US only; effect on adult height?

Abbreviations: act, activating; AD, autosomal dominant; AR, autosomal recessive; BL, birth length; BW, birth weight.

Table 7.1.3.2 Monogenic disorders in short children born SGA with microcephaly

Syndrome [OMIM]	Genetic defect	Inheritance	Incidence	Mean BW/BL SDS	Clinical features	Laboratory data	Treatment
Monogenic disorders with microcephaly							
IGF-I deficiency [608747]	*IGF1*	AR, AD	4 homozyg 7 heterozyg	hom −3.7 het −1.9	Microcephaly, deafness, reduced mental development	↑GH, variable IGF-I, ↑IGFBP-3,	IGF-I treatment moderately effective
Resistance to IGF-I [1270450]	*IGF1R*	AD	1–2% of short SGA	−2.1/−2.7	Microcephaly, reduced mental development	↑/normal GH, ↑/ normal IGF-I, ↑/ normal IGFBP-3	GH treatment moderately effective
PAPP-A2 deficiency	*PAPPA2* [homozyg]	AR	5 cases	−1.6/−1.3	Microcephaly, skeletal abnormalities	↑GH, ↑IGF-I, ↑IGFBP-3, ↑IGFBP-5, ↑ALS	IGF-I treatment possibly effective
Primordial dwarfism with microcephaly							
Cornelia de Lange syndrome (1–5) [122470]	*NIPBL*, *SMC1A*, *SMC3*, *RAD21*, *HDAC8*	AD	1/40 000	−3.4/−	Low anterior hairline, connected eyebrows, ante-verted nares, maxillary prognathism, long philtrum, thin lips, 'carp' mouth	Not available	No data on GH, likely ineffective
Meier-Gorlin syndrome (1–5) [224690]	*ORC1*, *ORC4*, *ORC6*, *CDT1*, *CDC6*	AR	>67 cases	−3.8/−	Bilateral microtia, aplasia, or hypoplasia of the patellae, normal intelligence	Not available	No data on GH, likely ineffective
MOPD I [210710]	*RNU4ATAC*	AR	<1/ 1 000 000	Extremely low	Neurologic abnormalities, including intellectual disability, brain malformations, ocular or auditory sensory deficits	Not available	No data on GH, likely ineffective
MOPD II [210720]	*PCNT*	AR	Extremely rare	−3.9	No or mild mental impairment, truncal obesity, DM, moyamoya disease, small loose teeth, radiologic abnormalities	Not available	No data on GH, likely ineffective
Seckel syndrome (1–8) [210600]	*ATR*, *RBBP8*, *CENPJ*, *CEP152*, *CEP63*, *NIN*, *DNA2*, *ATRIP*	AR	<1/ 1 000 000	Extremely low	Microcephaly, intellectual disability, characteristic 'bird-headed' face (receding forehead and micrognathia)	Not available	No data on GH, likely ineffective
Smit-Lemli-Opitz [270400]	*DHCR7*	AR	1/50 000 (mainly in Caucasians)	Extremely low	Microcephaly, moderate to severe intellectual disability, dysmorphic features, and organ malformations (heart, palate, syndactyly 2nd and 3rd toes, underdeveloped genitalia boys)	Not available	No data on GH, likely ineffective
DNA repair defects with microcephaly							
Bloom syndrome [210900]	*RECQL3*	AR	1/48 000 (Ashk Jews, but also others)	−4.7/−4.8	Microcephaly, sun-sensitivity, telangiectatic, hypo- and hyperpigmented skin lesions, predisposition to cancer, maturity-onset DM	Not available	GH contraindicated
Fanconi anaemia [many]	*FANCA* and 14 other genes	AR or X-linked	1/160 000	−1.8/−2.1	Microcephaly, genomic instability, hypo- and hyperpigmentation, skin lesions abnormalities in major organ systems, bone marrow failure, predisposition to cancer	Not available	GH contraindicated
Nijmegen breakage Syndrome [#251260]	*NBN*	AR	Extremely rare	−1.8/−2.2	Microcephaly, mild to moderate intellectual disability, immunodeficiency, predisposition to cancer	Not available	GH contraindicated
LIG 4 syndrome [606593]	*LIG4*	AR	Extremely rare	−3.0/−3.8	Microcephaly, sun-sensitive, combined immunodeficiency	Not available	GH contraindicated
XRCC4 syndrome	*XRCC4*	AR	Extremely rare	−1.6/−2.5	Microcephaly, progressively short, hypergonadotropic hypogonadism, multinodular goitre, diabetes mellitus	Not available	GH contraindicated

Table 7.1.3.3 Imprinting disorders and methylation disturbances in short children born SGA

Syndrome [OMIM]	Epigenetic defect	Incidence	Mean BW/BL SDS	Clinical features besides SGA	Treatment
Silver–Russell syndrome (SRS) [180860]	11p15 LOM (30–60%), upd(7)mat (5–10%), upd(20)mat, upd(16)mat act *CDKN1C*, *HMAG2*, *PLAG1*, and CNVs # Paternal *IGF2* mutation (see Table 7.1.3.1)	1:30 000–100 000	11p15: –3.2/–4.5 UPD7: –2.3/–2.5 Clinical: –2.7/–1.8	Relative macrocephaly, protruding forehead, body asymmetry, feeding problems, and/or low BMI	GH effective (in label of GH treatment for short SGA)
Temple syndrome [616222]	14q32 abnormalities: upd(14) mat, paternal microdeletions, hypomethylation of DLK1/GTL2 IG-DMR	>50 cases	–1.9/–1.6	Postnatal growth failure, hypotonia, delayed development of motor skills, feeding problems in infancy, early puberty, broad forehead, short nose with wide nasal tip, small hands, and feet	Insufficient data
IMAGe syndrome [614732]	Maternally inherited activating mutations in *CDKN1C*	>15 cases	–2.0 to –4.0	Relative macrocephaly at birth, no or mild intellectual disability, frontal bossing, low-set ears, flat nasal bridge, short nose, congenital adrenal hypoplasia, metaphyseal, and/or epiphyseal dysplasia, male genital anomalies, early-onset type 1 DM	Insufficient data
Prader–Willi syndrome [176270]	Paternal 15q11.2q13 deletion (60%), upd(15)mat (40%), or imprinting centre mutation (1–3%). Loss of *SNRPN* and *NDN* expression	1:16 000	–1.2/–1.1	Diminished fetal activity, obesity, muscular hypotonia, intellectual disability, short stature, hypogonadotropic hypogonadism, small hands, and feet	GH registered for PWS
Pseudohypoparathyroidism type 1a/c * [103580]	Heterozygous *GNAS1* mutation inherited from the mother	1:150 000	–0.6/–1.1 35% born SGA	Resistance to PTH and other hormones (TSH, LH, FSH, and GHRH), Albright hereditary osteodystrophy (short stature, obesity, round face, subcutaneous ossifications, brachydactyly, mild intellectual disability)	Insufficient data
Pseudopseudohypoparathyroidism [612463])	Heterozygous *GNAS1* mutation inherited from the father	1:150 000	–2.7/–3.0 95% born SGA	Albright hereditary osteodystrophy without multiple hormone resistance and no hypocalcaemia	Insufficient data

* Pseudohypoparathyroidism type 1b [OMIM 603233] is associated with normal or increased birth weight and overgrowth in childhood.

Abbreviations: act, activating; BL, birth length; BW, birth weight; PTH, parathyroid hormone.

scope of this chapter, but GH treatment at higher doses than recommended for growth hormone deficiency (GHD) children improves growth and adult height. The younger age at GH start, the better the growth response and adult height will be [36].

Mutations or deletions in the *SHOX* gene can be present in short SGA children. These are located at the tip of the X and Y chromosome, and transmitted in a pseudoautosomal fashion [36]. Bi-allelic inactivating *SHOX* mutations cause the severe Langer mesomelic dysplasia (OMIM 249700), while heterozygous mutations or deletions of *SHOX* or its enhancers (or even duplications) cause a milder skeletal dysplasia, Leri-Weill dyschondrosteosis with a Madelung deformity of the wrist (OMIM 127300) [37]. This mutation can present as SGA or ISS, with minor or no dysmorphic features, and no or mild body disproportion (OMIM 300582), particularly *SHOX* enhancer deletions. Birth length and weight show wide variations (–4.3 to 1.5 and –3.3 to 3.2 SDS, resp.) [38]. GH treatment results in a similar growth response and adult height gain as seen in girls with TS syndrome [39].

Heterozygous activating mutations in *FGFR3* lead to a wide range of disorders, including achondroplasia (OMIM 100800), hypochondroplasia (MIM 146000) and even proportionate short stature. Neonates with achondroplasia have a birth length around –1 SDS, but SGA is uncommon. The effect of GH treatment on adult height results in a modest increase of +0.5 SDS [40]. GH is only registered for this condition in Japan.

Children with hypochondroplasia (OMIM 146000) have rhizomelic limb shortening, limitation of elbow extension, brachydactyly, and relative macrocephaly. It is a relatively frequent genetic mutation in children with SGA and ISS. GH treatment results in better growth in the first years, but there are limited data on adult height [41].

Heterozygous carriers of *NPR2* mutations (OMIM 616255) show a similar phenotype to *SHOX* haploinsufficiency, with short forearms and short lower legs (mesomelia), but without Madelung deformity. *NPR2* mutations explain ~2% of cases with short stature due to SGA or ISS [42]. No data on GH treatment are available.

Heterozygous mutations of *Indian hedgehog (IHH)* are associated with brachydactyly type A1 (OMIM 1112500), but may also have mild disproportion [43]. Most children are born SGA for length, and 50% have shortening of the middle phalanx of the second and fifth fingers with cone-shaped epiphyses. No data on GH treatment are available.

A heterozygous mutation of *ACAN* (encoding aggrecan) leads to abnormal cartilage matrix formation, with mild skeletal dysplasia, spondyloepiphyseal dysplasia (OMIM 608361), or short stature without radiographic skeletal dysplasia (OMIM 165800) [44]. Approximately 30–40% of cases are born SGA and 14% of short SGA children with a bone age advancement of 6 or more months, had a heterozygous ACAN mutation [45]. However, bone age can be normal. Birth length SDS is always lower than birth weight SDS and patients often have early-onset osteoarthritis and/or osteochondritis dissecans. At present, there are only limited data on the effects of GH treatment, either alone or combined with a gonadotropin-releasing hormone (GnRH) analogue and/or an aromatase inhibitor [44].

Noonan syndrome may also explain disproportionate short stature after an SGA birth. Activation of the Ras/Mitogen activated protein kinase (MAPK) signalling pathway results in a number of overlapping syndromes, so-called rasopathies, including Noonan (OMIM 163950), Leopard (OMIM 151100), Costello (OMIM 218040), Cardiofaciocutaneous (OMIM 115150) and Neurofibromatosis-Noonan syndromes (OMIM 601321). All have a varying degree of postnatal growth failure and sometimes there are no obvious clinical features [46]. Twenty-four per cent (24%) are born SGA. GH treatment improves height SDS during childhood and can improve adult height [47]. GH is registered for Noonan syndrome in the USA, but not in Europe and Japan.

Monogenic Disorders with Microcephaly and Short Stature

Children with a *IGF1* mutation (OMIM-608747) are born with a very low birth weight and length and microcephaly, although some have a milder phenotype. Recombinant IGF-I treatment is moderately effective [47, 48]. Children with complete loss-of-function mutations also have sensorineural deafness (OMIM-608747) [49].

Children with a heterozygous *IGF1R* mutation (MIM 270450) have a wide range of birth weight (–3.5 to –1.5 SDS), birth length (–5.0 to 0.3 SDS) and head circumference (–3.0 to 0 SDS [48]. The prevalence of heterozygous *IGF1R* mutations or deletions is estimated at 1–2% in short SGA children [50]. Other features include normal baseline serum IGF-I levels and very high IGF-I levels during GH treatment, early feeding problems, variable delay in psychomotor development, and mild dysmorphic features. GH treatment results in a moderate growth response [50]. Homozygous mutations have a more severe phenotype [51].

Terminal 15q deletion with allelic loss of *IGF1R* leads to pre- and postnatal growth retardation and cardiac symptoms, intellectual disability, diaphragmatic hernia, hearing problems, aortic root dilatation, neonatal lymphedema, and aplasia cutis [52]. Limited data on GH treatment are available [53].

Homozygous pregnancy-associated-plasma-protein-A2 (*PAPPA2)* mutations were recently found in short children with ISS or born SGA [53]. They had progressive growth failure, moderate microcephaly, thin long bones, mildly decreased bone density, and *elevated* levels of serum IGF-I, IGFBP-3, IGFBP-5, ALS, and IGF-2. Lack of PAPP-A2 protein decreases the liberation of IGF-I from the ternary complex and likely results in lower IGF bioavailability. Two years of biosynthetic IGF-I treatment resulted in a height gain of +0.4 to 1.0 SDS [54].

Primordial dwarfism is a group of rare genetic disorders characterized by severe IUGR and SGA, extreme short stature, and distinct microcephaly, which occur as a result of disorganized molecular and genomic changes during embryonic development. Most children can be easily recognized, like those with Cornelia de Lange syndrome (OMIM 122470), Meier-Gorlin syndrome (OMIM 224690), microcephalic osteodysplastic primordial dwarfism (MOPD) types I (OMIM 210710) and II (OMIM 210720), and Seckel syndrome (OMIM 210600)] [55]. There are no data on GH treatment, which is likely to be ineffective.

Smith-Lemli-Opitz syndrome (OMIM 270400) is an autosomal recessive disorder characterized by SGA, short stature, microcephaly, dysmorphic features, mild to severe mental retardation, and multiple malformations. Patients have decreased serum cholesterol levels due a deficient cholesterol synthesis as result of mutations of the 3beta-hydroxysterol-delta7 reductase gene (*DHCR7*). There are no data on GH, which is likely to be ineffective.

Disorders of DNA repair or genomic instability are frequently associated with SGA, short stature, and microcephaly. Bloom syndrome (OMIM 210900) is a DNA repair disorder caused by a mutation in the gene encoding DNA helicase RecQ protein-like-3 (*RECQL3*) and features include skin hypersensitivity to sunlight, particularly recognizable on the face. Fanconi syndrome (OMIM 227650) has genomic instability and clinical features include an irregular pattern of skin pigmentations which increases over time, abnormalities in major organ systems, early-onset bone marrow failure, and a high predisposition to cancer. Other syndromes in this category are Nijmegen breakage syndrome (OMIM#251260), *LIG4* (MIM 606593) and *XRCC4* mutations (OMIM 616541), Cockayne syndrome (OMIM 216400) and Rothmund-Thomson syndrome (OMIM 268400). GH treatment is *contraindicated* in these conditions, given the unknown long-term effects on cell division and their predisposition to cancer,

Imprinting Disorders (ID) and Methylation Disturbances

IDs are characterized by molecular disturbances affecting genomically imprinted chromosomal regions and genes. Genomic imprinting describes the expression of specific genes in a parent-of-origin specific manner (i.e. they are expressed only from the maternal or from the paternal gene copy) [56]. Silver–Russell syndrome (SRS) is well-known, but there are several other imprinting disorders associated with short SGA [57]. It was recently reported that methylation disturbances can also occur in non-imprinted genes [58].

Silver–Russell Syndrome (SRS)

Children with SRS (OMIM 180860) are mostly born SGA, have short stature and relative macrocephaly, a triangular shaped face with frontal bossing, clinodactyly, and asymmetry of face and/or extremities [34]. Severe feeding difficulties are often present, especially during infancy. Mean adult height is around –4 SDS. SRS is primarily a clinical diagnosis, which can be established with the Netchine–Harbison clinical scoring system [34].

In around 60% of children with SRS, an underlying cause can be identified. Fifty percent are caused by a loss of methylation (LOM) of the telomeric domain in the 11p15.5 region, resulting in downregulation of paternal *IGF2* expression [34]. The 11p15.5 region has two imprinting control regions, which affect imprinted genes involved in the regulation of pre- and postnatal growth. Two rare genetic causes of SRS were found in the same 11p15 region: (1) an increased expression of *CDKN1C* (by a gain-of-function mutation or maternal duplication) [59], which can be associated with the IMAGe syndrome (OMIM 614732); the latter is characterized by pre- and postnatal growth failure, metaphyseal dysplasia, congenital adrenal hypoplasia, genital anomalies and early-onset DM; and (2) an *IGF2* loss-of-function mutation (OMIM 616489) resulting in a paternally transmitted SRS (Table 7.1.3.1) [60]. Disruptions in the HMGA2-PLAG1-IGF2 pathway can also cause an SRS phenotype [61], while copy number variants (CNVs) involving the 11p15.5 region can result in a specific SRS phenotype depending on CNV size, location, and parental origin [62].

Between 5% and 10% of children with SRS have a maternal uniparental disomy of chromosome 7 (upd(7)mat), with potential causative genes including *GRB10* (7p12.1) and *MEST* (7q32) [34]. In 35%, the genetic cause is yet unknown (i.e. clinical SRS). Patients with 11p15 LOM have a more 'classic' SRS phenotype with a higher prevalence of asymmetry than those with upd(7)mat, but the latter ones carry a higher risk of behavioural problems [34]. More widespread use of methylation studies will probably uncover more epigenetic disorders in short SGA children, as shown by a study in patients with suspected SRS or unexplained short stature/IUGR, in whom 37% showed abnormal methylation in eleven imprinted loci [63].

A typical feature of children with SRS is their very poor appetite, severely reduced food intake, and very slim appearance. Most children have a rapid bone age advancement around the age of 8–9 years, a relatively young age at onset of puberty albeit within the normal range [56]. Males with SRS have an increased risk of genital abnormalities such as cryptorchidism and hypospadias, which could be associated with reproductive problems later in life. In females, Mayer–Rokitansky–Küster–Hauser syndrome, characterized by hypoplasia or aplasia of the uterus and upper part of the vagina, has been reported [64].

Long-term GH treatment has proven to be safe and effective in improving adult height in SRS, with similar results as non-SRS subjects born SGA [65]. As the rapid bone maturation during puberty may compromise adult height, it is recommended to add (2 years of) GnRH agonist treatment in GH-treated SRS patients with a poor adult height prognosis at the onset of puberty [34]. The metabolic health profile of GH-treated young adults with SRS was similar to that of GH-treated non-SRS subjects born SGA [66].

Other Imprinting Disorders

Besides upd(7)mat (SRS), there are several other IDs associated with SGA and short stature. Maternal UPD of chromosome 20, upd(20)mat, causes SGA, short stature, and prominent feeding difficulties with failure to thrive, but the patients differ from SRS because there is no asymmetry, prominence of the forehead and relative macrocephaly [66]. Upd(16)mat has also been associated with SGA [34]. Temple syndrome (OMIM 616222) due to by maternal UPD 14 (upd14mat), paternal deletion, distal 14q duplication, or LOM at the intergenic differentially methylated region, is characterized by hypotonia, early puberty, short stature and sometimes SGA [67, 68]. Prader–Willi syndrome (OMIM 176270) is a genetic disorder resulting from the absence of locus q11-q13 (PWS region) on the paternally derived chromosome 15, mostly due to a paternal deletion or maternal uniparental disomy (upd(15)mat) and leads to a characteristic phenotype including muscular hypotonia, short stature, hypogonadism, obesity, psychomotor delay, behavioural problems, and cognitive impairment. Long-term GH treatment improves body composition, growth, physical strength, psychomotor development, and cognitive functioning [69]. Fifty per cent (50%) of children with PWS have a birth weight below the 10th percentile and almost all remain short. Recently, quantitative DNA methylation analyses found abnormal methylation of several non-imprinted genes in SGA [70].

Management of Short Children Born SGA

Follow-Up of Young Children

An SGA-born child should have measurements of length, weight, and head circumference every 3 months during the first year of life and every 6 months thereafter [2]. Children with microcephaly

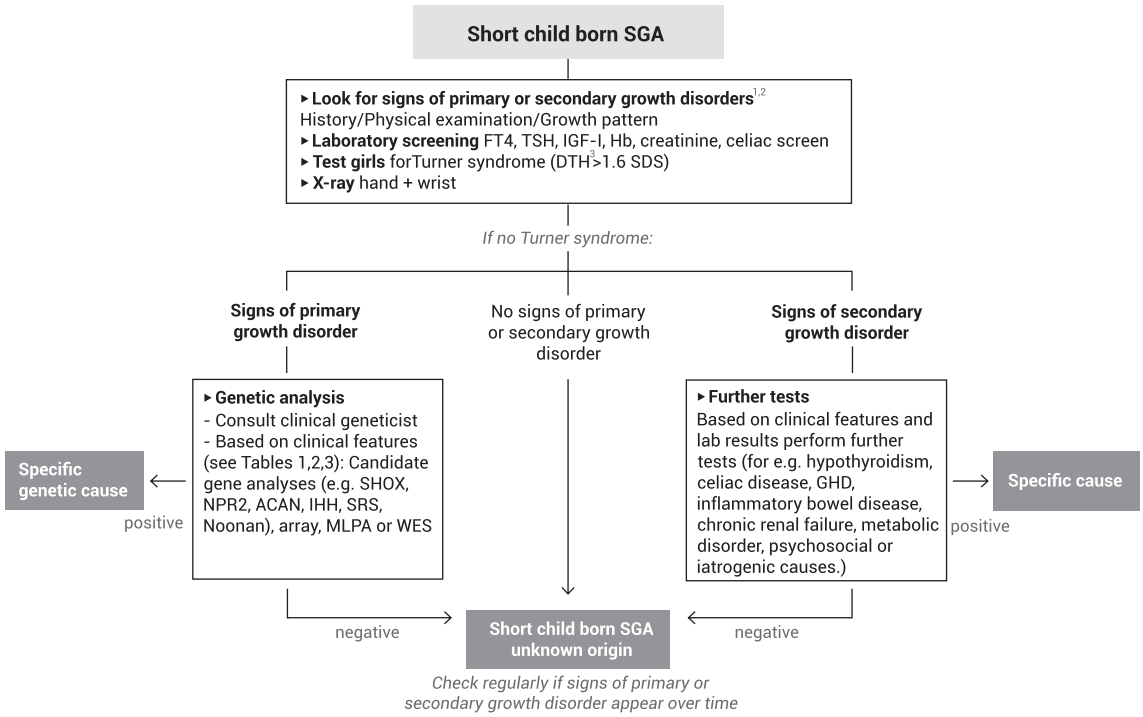
might have a syndrome (Table 7.1.3.2) and those who do not show significant catch-up in length in the first year and are short by the age of 2 years, may have an underlying disorder that limits growth. These children should be identified and referred to a paediatric endocrinologist with experience in short stature after SGA birth.

Diagnostic Approach in Short SGA Children

Figure 7.1.3.1 presents a stepwise diagnostic approach for short children born SGA. As a first step, the clinician needs to identify diagnostic signs from the medical history, physical examination, and growth pattern, followed by a screening laboratory panel and an X-ray of the left hand and wrist to assess skeletal age and search for anatomic abnormalities associated with genetic disorders like *SHOX, NPR2, ACAN, IHH*. In addition, in all girls whose height is well below target height (>1.6 SDS), TS should be excluded, by array analysis (SNP array or CGH array) or a karyotype. The diagnostic power of an array to detect TS is similar to that of a karyotype [71], but with a SNP array CNVs and uniparental isodisomy can be detected. When there are signs of a secondary growth disorder, conditions like hypothyroidism, celiac disease, GH-IGF-axis disturbance, or chronic disease should be evaluated. Only when serum IGF-I level is low for age and gender, is a GH stimulation test indicated. When there are signs of a primary growth disorder and TS is excluded, further genetic testing could be performed. When there are dysmorphic features and/or microcephaly, one may consider evaluation by a clinical geneticist. Repeatedly high serum IGF-I levels (> 1 SDS) requires genetic testing of the *IGF1R* gene. If there is a strong suspicion of a specific genetic syndrome, DNA sequencing combined with an MLPA (Multiplex Ligation-dependent Probe Amplification) test of one specific gene can be performed ('candidate gene approach'), but one may also decide to perform a growth-specific gene panel or a specific exome-based gene panel targeted to growth-related genes [72]. When the child has a positive Netchine–Harbison score [34], genetic testing for Silver–Russell syndrome is indicated. In the case of dysmorphic signs or (some) suspicion of a primordial disorder, an array analysis is probably more efficient to find a genetic cause. As last step, RNA sequencing or a methylation array could be considered [58]. In special cases (e.g. if a novel monogenic disorder is suspected), WES in a 'trio' (patient and both parents, or including siblings) can be performed. The latter is mostly done after consultation with a clinical geneticist. Presently, the yield of genetic diagnoses after such analyses is disappointingly low in non-syndromic short SGA children. However, as genetic techniques are extended, a cause will be identified in an increasing number of short SGA children and it might turn out that many cases of SGA are caused by a combination of multiple (epi)genetic variants [58].

Growth Hormone Treatment of Short Children Born SGA

Biosynthetic GH is officially approved for short SGA children by the Food and Drug Administration in USA since 2001, the European Agency for the Evaluation of Medicinal Products since 2003 and in Japan since 2008. The criteria at which GH treatment can be started are somewhat different between Europe, USA, and Japan [2]. Short



1. Signs for primary growth disorder: Medical history: alcohol abuse mother, developmental delay, behavior problems, dominant pattern of short stature, early-onset osteoarthritis and parental body disproportion; Physical examination: body disproportion, dysmorphic features, short upper or forearm, microcephaly, cryptorchidism, muscular hypertrophy; Growth: DTH[3] >1.6 SDS, parental short stature; X-hand/wrist: anatomic abnormalities.
2. Signs for secondary growth disorder: Medical history: decreasing or increasing BMI SDS, anorexia, fatigue, polyuria, polydipsia, medication (corticosteroids, methylphenidate), symptoms of increased intracranial pressure, vomiting, disturbed vision; Physical examination: low or high BMI SDS, Cushingoid appearance, hypertension; Growth: decreasing height SDS, decreasing BMI SDS (IBD) or increasing BMI SDS (Cushing, hypothyroidism, GHD); X-hand/wrist: delayed bone age
3. DTH = Distance to target height

Figure 7.1.3.1 A stepwise diagnostic approach of a short child born SGA.

children with a dysmorphic syndrome, except those with Silver–Russell syndrome, tend to show poor growth to GH [2].

The registered starting dose of GH is 0.033 mg/kg/day (~1 mg/m²/day) in Europe and Japan, while in the USA the recommended dose ranges from 0.033 to 0.067 ug/kg/day (~1–2 mg/m²/day) [2].

Effect on Growth and Adult Height

GH treatment effectively induces catch-up growth and improves adult height in most short children born SGA [73, 74]. A systematic review reported that the mean height gain was on average 1.25 SDS [75].

The growth response to GH is, however, highly variable, which is probably associated with multiple gene variants. Prediction models could explain 52% of the variability of the growth response in the first year and 40% of adult height [76, 77], but it is impossible to reliably predict the individual growth response both before and during treatment [77]. The growth response to GH treatment does not differ between those born preterm or term [78].

No difference in adult height SDS was found between children treated with a GH dose of 0.033 mg/kg/day and 0.067 mg/kg/day [73], while the higher GH dose led to an average serum IGF-I of +2 SDS. For that reason, it is recommended to treat short SGA children with a GH dose of 0.033 mg/kg/day, and to sonly increase the GH dose when the growth response is unsatisfactory and other causes of a poor growth response are ruled out [2]. High serum IGF-I levels >2 SDS during GH treatment with 0.033 mg/kg/day might be due to a heterozygous *IGF1R* defect. Consideration should be given to undertaking genetic testing for this diagnosis. These children require a higher GH dose to obtain an acceptable growth response, and an elevated serum IGF-I on GH treatment may be accepted due to their partial IGF-I insensitivity. Persistently high serum IGF-I levels are also found in short SGA children with syndromes like Bloom and Fanconi, which may not have been identified due to lack of clear dysmorphic features. Again, genetic testing would be helpful to achieve a specific diagnosis. If such a syndrome is identified, then GH should be discontinued. Children with SRS (15p11) can also show high IGF-I levels, likely due to a relative IGF-I resistance, but until this has become clear, it is recommended to lower the GH dose. GH dosing based on IGF-I titration results in a poorer growth response than treatment based on a fixed GH dose, and is therefore not recommended [79]. Adult height is greater when treatment is started at least 2 years before the onset of puberty [74]. However, some short SGA children come to medical attention when they have just started puberty. When their predicted adult height is <–2.5 SDS at onset of puberty, their adult height can improve by combined treatment of GH and 2 years of a GnRHa [80].

Effect on Pubertal Development

On average, GH-treated SGA children start puberty at a similar age as the normal population, although some start relatively early. GH treatment has no influence on the onset and progression of puberty compared to AGA controls, regardless of GH dose (0.033 vs. 0.067 mg/kg/day) [9].

Effect on Metabolism, Insulin Sensitivity, and Cardiovascular Health

GH has lipolytic, anabolic, and insulin-antagonistic effects and leads in short SGA children to an increase in lean body mass and decrease in fat mass [81]. Short SGA children have reduced insulin sensitivity before the start of GH treatment, a further reduction in insulin sensitivity with a compensatory increase in insulin secretion during GH [18], but a complete normalization after cessation of GH. Assessment of the impact of long-term GH treatment in large study groups has shown that none of the GH-treated SGA-born subjects developed β-cell dysfunction or type 2 DM [17].

Blood pressure and cholesterol levels decline during GH treatment, and become significantly lower than in untreated SGA children [21]. The few studies reporting on the carotid intima media thickness (cIMT) in GH-treated SGA children showed no effect of GH treatment [26, 82].

Treatment with combined GH/GnRHa in SGA children results in similar body composition, insulin sensitivity, β-cell function, blood pressure, and lipid levels at adult height as those treated with GH only [83].

Effect on Bone Mineral Density

During long-term GH treatment, the total-body BMD improves from –1.00 SDS to –0.44 SDS, and lumbar-spine BMD corrected for height, from –0.48 SDS to –0.14 SDS [84].

Effect on Cognitive Functioning and Health-Related Quality of Life

Short SGA children have on average lower cognitive functioning. One study showed that long-term GH treatment improved performance IQ and attention while the estimated total IQ scores significantly increased by 5–10 points, to the same range as the normal population [27].

Long-term GH treatment improves health-related quality of life (HRQoL) in short SGA children [29].

Safety

Since the early nineties, multiple large cohort studies in short children born SGA have been conducted, showing that GH treatment is well tolerated, and that serious adverse effects are uncommon [75].

Diagnosing syndromes can be challenging, particularly when children have mild symptoms, but identifying the genetic aetiology is important for health prognosis, genetic counselling, and treatment. GH treatment is contraindicated in several conditions, such as chromosomal breakage syndromes and DNA repair disorders [85].

Monitoring During GH Treatment

Height, weight, and Tanner pubertal stage should be monitored regularly. It is recommended to determine the serum IGF-I level 3 to 6 months after the start of treatment, to evaluate whether the GH dose requires adjustment and/or further evaluation. Thereafter, it is advised to determine the serum IGF-I level annually. When IGF-I levels remain high even with a relatively low GH dose, one should consider performing *IGF1R* mutational analysis, and to examine for the presence of a possible underlying dysmorphic syndrome like Bloom syndrome [85]. As serum-free T_4 levels decrease to levels just below the norm in 14% of GH-treated SGA children but thyroid-stimulating hormone (TSH) levels remained normal [86], it is recommended to annually assess serum-free T_4 and TSH.

As bone maturation in short children born SGA is highly variable and unreliable for growth prediction, radiological investigations are not very informative [2]. Since SGA birth is associated with a

higher risk of the metabolic syndrome [6], one could argue that metabolic syndrome parameters should be checked annually. On the other hand, many studies have shown that GH has very limited effects on cardiometabolic parameters during treatment and several years thereafter. For that reason, it is recommended to only monitor blood pressure, BMI, abdominal waist circumference, fasting glucose, and lipid levels. Periodical measurement of fasting insulin is not recommended for clinical care because of the absence of criteria to differentiate normal from abnormal. However, additional monitoring and testing should be performed when indicated (e.g. in the case of a positive family history for DM or cardiometabolic diseases) [2].

Conclusions

GH treatment in short SGA children increases longitudinal growth and adult height, and has positive effects on cardiometabolic health and cognitive functioning. The addition of GnRHa for 2 years in early pubertal children with a low predicted adult height can improve adult height. The safety profile of GH treatment alone or combined with GnRHa is good.

REFERENCES

1. Lee PA, Chernausek SD, Hokken-Koelega AC, Czernichow P. International Small for Gestational Age Advisory Board consensus development conference statement: management of short children born small for gestational age, April 24–October 1, 2001. *Pediatrics*, 2003; **111**: 1253–61.
2. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. *J Clin Endocrinol Metab*, 2007; **92**: 804–10.
3. Hokken-Koelega AC, De Ridder MA, Lemmen RJ, *et al*. Children born small for gestational age: do they catch up? *Pediatr Res*, 1995; **38**: 267–71.
4. Karlberg J, Albertsson-Wikland K. Growth in full-term small-for-gestational-age infants: from birth to final height. *Pediatr Res*, 1995; **38**: 733–73.
5. Finken MJ, Dekker FW, de Zegher F, Wit JM. Long-term height gain of prematurely born children with neonatal growth restraint: parallellism with the growth pattern of short children born small for gestational age. *Pediatrics*, 2006; **118**: 640–3.
6. Barker DJ, Hales CN, Fall CH, Osmond C, Phipps K, Clark PM. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia*, 1993; **36**: 62–7.
7. Leunissen RW, Kerkhof GF, Stijnen T, Hokken-Koelega A. Timing and tempo of first-year rapid growth in relation to cardiovascular and metabolic risk profile in early adulthood. *JAMA*, 2009; **301**: 223–42.
8. De Waal WJ, Hokken-Koelega AC, Stijnen T, de Muinck Keizer Schrama SM, Drop SL. Endogenous and stimulated GH secretion, urinary GH excretion, and plasma IGF-I and IGF-II levels in prepubertal children with short stature after intrauterine growth retardation. *Clin Endocrinol*, 1994; **41**: 621–30.
9. Boonstra V, van Pareren Y, Mulder P, Hokken-Koelega A. Puberty in growth hormone-treated children born small for gestational age (SGA). *J Clin Endocrinol Metab*, 2003; **88**: 5753–8.
10. Boonstra VH, Mulder PG, de Jong FH, Hokken-Koelega AC. Serum dehydroepiandrosterone sulfate levels and pubarche in short children born small for gestational age before and during growth hormone treatment. *J Clin Endocrinol Metab*, 2004; **89**: 712–17.
11. Willemsen RH, Arends NJ, Bakker-van Waarde WM, *et al*. Long-term effects of growth hormone (GH) treatment on body composition and bone mineral density in short children born small-for-gestational-age: six-year follow-up of a randomized controlled GH trial. *Clin Endocrinol (Oxf)*, 2007; **67**: 485–92.
12. Leunissen RW, Stijnen T, Boot AM, Hokken-Koelega AC. Influence of birth size and body composition on bone mineral density in early adulthood: the PROGRAM study. *Clin Endocrinol (Oxf)*, 2008; **69**: 386–92.
13. Boonstra VH, Weber RF, de Jong FH, Hokken-Koelega AC. Testis function in prepubertal boys and young men born small for gestational age. *Horm Res*, 2008; **70**: 357–63.
14. Kerkhof GF, Leunissen RW, Willemsen RH, de Jong FH, Stijnen T, Hokken-Koelega AC. Influence of preterm birth and birth size on gonadal function in young men. *J Clin Endocrinol Metab*, 2009; **94**: 4243–50.
15. Kerkhof GF, Leunissen RW, Willemsen RH, *et al*. Influence of preterm birth and small birth size on serum anti-Mullerian hormone levels in young adult women. *Eur J Endocrinol*, 2010; **163**: 937–44.
16. Ibanez L, Ong K, Dunger DB, de Zegher F. Early development of adiposity and insulin resistance after catch-up weight gain in small-for-gestational-age children. *J Clin Endocrinol Metab*, 2006; **91**: 2153–8.
17. Van der Steen M, Smeets CC, Kerkhof GF, Hokken-Koelega AC. Metabolic health of young adults who were born small for gestational age and treated with growth hormone, after cessation of growth hormone treatment: a 5-year longitudinal study. *Lancet Diabetes Endocrinol*, 2017; **5**: 106–16.
18. Arends NJ, Boonstra VH, Duivenvoorden HJ, *et al*. Reduced insulin sensitivity and the presence of cardiovascular risk factors in short prepubertal children born small for gestational age (SGA). *Clin Endocrinol (Oxf)*, 2005; **62**: 44–50.
19. Leunissen RW, Oosterbeek P, Hol LK, Hellingman AA, Stijnen T, Hokken-Koelega AC. Fat mass accumulation during childhood determines insulin sensitivity in early adulthood. *J Clin Endocrinol Metab*, 2008; **93**: 445–51.
20. Willemsen RH, de Kort SW, van der Kaay DC, Hokken-Koelega AC. Independent effects of prematurity on metabolic and cardiovascular risk factors in short small-for-gestational-age children. *J Clin Endocrinol Metab*, 2008; **93**: 452–8.
21. Van Dijk M, Bannink EM, van Pareren YK, Mulder PG, Hokken-Koelega AC. Risk factors for diabetes mellitus type 2 and metabolic syndrome are comparable for previously growth hormone-treated young adults born small for gestational age (sga) and untreated short SGA controls. *J Clin Endocrinol Metab*, 2007; **92**: 160–5.
22. Stevens A, Bonshek C, Whatmore A, *et al*. Insights into the pathophysiology of catch-up compared with non-catch-up growth in children born small for gestational age: an integrated analysis of metabolic and transcriptomic data. *Pharmacogenomics J*, 2014; **14**: 376–84.

23. Leunissen RW, Kerkhof GF, Stijnen T, Hokken-Koelega AC. Fat mass and apolipoprotein E genotype influence serum lipoprotein levels in early adulthood, whereas birth size does not. *J Clin Endocrinol Metab*, 2008; **93**: 4307–14.
24. Chiavaroli V, Marcovecchio ML, de Giorgis T, Diesse L, Chiarelli F, Mohn A. Progression of cardio-metabolic risk factors in subjects born small and large for gestational age. *PLoS One*, 2014; **9**: e104278.
25. Sebastiani G, Diaz M, Bassols J, *et al*. The sequence of prenatal growth restraint and post-natal catch-up growth leads to a thicker intima-media and more pre-peritoneal and hepatic fat by age 3–6 years. *Pediatr Obes*, 2016; **11**: 251–7.
26. Van der Steen M, Kerkhof GF, Smeets CC, Hokken-Koelega AC. Cardiovascular risk factors and cIMT in young adults born small for gestational age after cessation of GH treatment: a 5-year longitudinal study. *Lancet Diabetes Endocrinol*, 2017; 975–85.
27. Van Pareren YK, Duivenvoorden HJ, Slijper FS, Koot HM, Hokken-Koelega AC. Intelligence and psychosocial functioning during long-term growth hormone therapy in children born small for gestational age. *J Clin Endocrinol Metab*, 2004; **89**: 5295–302.
28. Lindstrom L, Wikstrom AK, Bergman E, Lundgren M. Born small for gestational age and poor school performance—how small is too small? *Horm Res Paediatr*, 2017; **88**: 215–23.
29. Bannink EM, van Pareren YK, Theunissen NC, Raat H, Mulder PG, Hokken-Koelega AC. Quality of life in adolescents born small for gestational age: does growth hormone make a difference? *Horm Res*, 2005; **64**: 166–74.
30. Finken MJJ, van der Steen M, Smeets CCJ, *et al*. Children born small for gestational age: differential diagnosis, moleclar-genetic evaluation and implications. *Endocr Rev*, 2018; **39**: 851–94.
31. Isik E, Haliloglu B, van Doorn J, *et al*. Clinical and biochemical characteristics and bone mineral density of homozygous, compound heterozygous and heterozygous carriers of three novel IGFALS mutations. *Eur J Endocrinol*, 2017; **176**: 657–67.
32. Grandone A, Miraglia del Giudice E, Cirillo G, *et al*. Clinical features of a new acid-labile subunit (IGFALS) heterozygous mutation: anthropometric and biochemical characterization and response to growth hormone administration. *Horm Res Paediatr*, 2014; **81**: 67–72.
33. Clayton PE, Hanson D, Magee L, *et al*. Exploring the spectrum of 3-M syndrome, a primordial short stature disorder of disrupted ubiquitination. *Clin Endocrinol (Oxf)*, 2012; **77**: 335–42.
34. Wakeling EL, Brioude F, Lokulo-Sodipe O, *et al*. Diagnosis and management of Silver-Russell syndrome: first international consensus statement. *Nat Rev Endocrinol*, 2017; **13**:105–24.
35. Nikkel SM, Dauber A, de Munnik S, *et al*. The phenotype of Floating–Harbor syndrome: clinical characterization of 52 individuals with mutations in exon 34 of SRCAP. *Orphanet J Rare Dis*, 2013; **8**: 63.
36. Gravholt CH, Andersen NH, Conway GS, *et al*. International Turner Syndrome Consensus G. Clinical practice guidelines for the care of girls and women with Turner syndrome: proceedings from the, 2016 Cincinnati International Turner Syndrome Meeting. *Eur J Endocrinol*, 2017; **177**: G1–G70.
37. Marchini A, Ogata T, Rappold GA. A track record on SHOX: from basic research to complex models and therapy. *Endocr Rev*, 2016; **37**: 417–48.
38. Donze SH, Meijer CR, Kant SG, *et al*. The growth response to GH treatment is greater in patients with SHOX enhancer deletions compared to SHOX defects. *Eur J Endocrinol*, 2015; **173**: 611–21.
39. Benabbad I, Rosilio M, Child CJ, *et al*. Safety outcomes and near-adult height gain of growth hormone-treated children with SHOX deficiency: data from an observational study and a clinical trial. *Horm Res Paediatr*, 2017; **87**: 42–50.
40. Harada D, Namba N, Hanioka Y, *et al*. Final adult height in long-term growth hormone-treated achondroplasia patients. *Eur J Pediatr*, 2017; **176**: 873–79.
41. Massart F, Miccoli M, Baggiani A, Bertelloni S. Height outcome of short children with hypochondroplasia after recombinant human growth hormone treatment: a meta-analysis. *Pharmacogenomics*, 2015; **16**: 1965–73.
42. Wang SR, Jacobsen CM, Carmichael H, *et al*. Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature. *Hum Mutat*, 2015; **36**: 474–81.
43. Vasques GA, Funari MFA, Ferreira FM, *et al*. IHH gene mutations causing short stature with nonspecific skeletal abnormalities and response to growth hormone therapy. *J Clin Endocrinol Metab*, 2018; **103**: 604–14.
44. Gkourogianni A, Andrew M, Tyzinski L, *et al*. Clinical characterization of patients with autosomal dominant short stature due to aggrecan mutations. *J Clin Endocrinol Metab*, 2017; **102**: 460–9.
45. Van der Steen M, Pfundt R, Maas S, *et al*. ACAN gene mutations in short children born sga and response to growth hormone treatment. *J Clin Endocrinol Metab*, 2017; **102**: 1458–67.
46. Cessans C, Ehlinger V, Arnaud C, *et al*. Growth patterns of patients with Noonan syndrome: correlation with age and genotype. *Eur J Endocrinol*, 2016; **174**: 641–50.
47. Tamburrino F, Gibertoni D, Rossi C, *et al*. Response to long-term growth hormone therapy in patients affected by RASopathies and growth hormone deficiency: patterns of growth, puberty and final height data. *Am J Med Genet A*, 2015; **167A**: 2786–94.
48. Walenkamp MJ, Karperien M, Pereira AM, *et al*. Homozygous and heterozygous expression of a novel insulin-like growth factor-I mutation. *J Clin Endocrinol Metab*, 2005; **90**: 2855–64.
49. Woods KA, Camacho-Hubner C, Savage MO, Clark AJ. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med*, 1996; **335**: 1363–7.
50. Ester WA, van Duyvenvoorde HA, de Wit CC, *et al*. Two short children born small for gestational age with insulin-like growth factor 1 receptor haploinsufficiency illustrate the heterogeneity of its phenotype. *J Clin Endocrinol Metab*, 2009; **94**: 4717–27.
51. Abuzzahab MJ, Schneider A, Goddard A, *et al*. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med*, 2003; **349**: 2211–22.
52. O'Riordan AM, McGrath N, Sharif F, *et al*. Expanding the clinical spectrum of chromosome 15q26 terminal deletions associated with IGF-1 resistance. *Eur J Pediatr*, 2017; **176**: 137–42.
53. Ho SC, Clayton P, Vasudevan P, *et al*. Recombinant human growth hormone therapy in children with chromosome 15q26 deletion. *Horm Res Paediatr*, 2015; **83**: 424–30.
54. Dauber A, Munoz-Calvo MT, Barrios V, *et al*. Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. *EMBO Mol Med*, 2016; **8**: 363–74.
55. Hawkins-Carranza FG, Munoz-Calvo MT, Martos-Moreno GA, *et al*. rhIGF-1 treatment increases bone mineral density and trabecular bone structure in children with PAPP-A2 deficiency. *Horm Res Paediatr*, 2018; **89**: 200–4.
56. Khetarpal P, Das S, Panigrahi I, Munshi A. Primordial dwarfism: overview of clinical and genetic aspects. *Mol Genet Genomics*, 2016; **291**: 1–15.

57. Eggermann T, Perez de Nanclares G, Maher ER, *et al*. Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin Epigenetics*, 2015; **7**: 123.
58. Stalman SE, Solanky N, Ishida M, *et al*. Genetic analyses in small for gestational age newborns. *J Clin Endocrinol Metab*, 2018; **103**: 917–25.
59. Brioude F, Oliver-Petit I, Blaise A, *et al*. CDKN1C mutation affecting the PCNA-binding domain as a cause of familial Russell Silver syndrome. *J Med Genet*, 2013; **50**: 823–30.
60. Begemann M, Zirn B, Santen G, *et al*. Paternally Inherited IGF2 Mutation and Growth Restriction. *N Engl J Med*, 2015; **373**: 349–56.
61. Abi Habib W, Brioude F, Edouard T, *et al*. Genetic disruption of the oncogenic HMGA2-PLAG1-IGF2 pathway causes fetal growth restriction. *Genet Med*, 2018; **20**: 250–8.
62. Begemann M, Spengler S, Gogiel M, *et al*. Clinical significance of copy number variations in the 11p15.5 imprinting control regions: new cases and review of the literature. *J Med Genet*, 2012; **49**: 547–53.
63. Turner CL, Mackay DM, Callaway JL, *et al*. Methylation analysis of 79 patients with growth restriction reveals novel patterns of methylation change at imprinted loci. *Eur J Hum Genet*, 2010; **18**: 648–55.
64. Goedegebuure WJ, Smeets CC, Renes CC, de Rijke YB, Hokken-Koelega AC. Gonadal function and pubertal development in patients with Silver-Russell syndrome. *Hum Reproduction*, 2018; **33**: 2122–30.
65. Smeets CC, Zandwijken GR, Renes JS, Hokken-Koelega AC. Long-term results of GH treatment in Silver-Russell syndrome (SRS): do they benefit the same as non-SRS short-SGA? *J Clin Endocrinol Metab*, 2016; **101**: 2105–12.
66. Smeets CC, Renes JS, van der Steen M, Hokken-Koelega AC. Metabolic health and long-term safety of growth hormone treatment in Silver-Russell syndrome. *J Clin Endocrinol Metab*, 2017; **102**: 983–91.
67. Mulchandani S, Bhoj EJ, Luo M, *et al*. Maternal uniparental disomy of chromosome 20: a novel imprinting disorder of growth failure. *Genet Med*, 2016; **18**: 309–15
68. Ioannides Y, Lokulo-Sodipe K, Mackay DJ, Davies JH, Temple IK. Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. *J Med Genet*, 2014; **51**: 495–501.
69. Bakker NE, Kuppens RJ, Siemensma EP, *et al*. Eight years of growth hormone treatment in children with Prader–Willi syndrome: maintaining the positive effects. *J Clin Endocrinol Metab*, 2013; **98**: 4013–22.
70. Bens S, Haake A, Richter J, *et al*. Frequency and characterization of DNA methylation defects in children born SGA. *Eur J Hum Genet*, 2013; **21**: 838–43.
71. Prakash S, Guo D, Maslen CL, *et al*. Single-nucleotide polymorphism array genotyping is equivalent to metaphase cytogenetics for diagnosis of Turner syndrome. *Genet Med*, 2014; **16**: 53–9.
72. de Bruin C, Dauber A. Insights from exome sequencing for endocrine disorders. *Nat Rev Endocrinol*, 2015; **11**: 455–64.
73. Van Pareren Y, Mulder P, Houdijk M, Jansen M, Reeser M, Hokken-Koelega A. Adult height after long-term, continuous growth hormone (GH) treatment in short children born small for gestational age: results of a randomized, double-blind, dose-response GH trial. *J Clin Endocrinol Metab*, 2003; **88**: 3584–90.
74. Dahlgren J, Wikland KA, Swedish Study Group for Growth Hormone T. Final height in short children born small for gestational age treated with growth hormone. *Pediatr Res*, 2005; **57**: 216–22.
75. Maiorana A, Cianfarani S. Impact of growth hormone therapy on adult height of children born small for gestational age. *Pediatrics*, 2009; **124**: 519–31.
76. Ranke MB, Lindberg A, Board KI. Prediction models for short children born small for gestational age (SGA) covering the total growth phase. Analyses based on data from KIGS (Pfizer International Growth Database). *BMC Med Inform Decis Mak*, 2011; **11**: 38.
77. De Ridder MA, Stijnen T, Hokken-Koelega AC. Prediction model for adult height of small for gestational age children at the start of growth hormone treatment. *J Clin Endocrinol Metab*, 2008; **93**: 477–83.
78. De Kort SW, Willemsen RH, van der Kaay DC, Duivenvoorden HJ, Hokken-Koelega AC. Does preterm birth influence the response to growth hormone treatment in short, small for gestational age children? *Clin Endocrinol (Oxf)*, 2009; **70**: 582–7.
79. Jensen RB, Thankamony A, O'Connell SM, *et al*. A randomised controlled trial evaluating IGF1 titration in contrast to current GH dosing strategies in children born small for gestational age: the North European Small-for-Gestational-Age Study. *Eur J Endocrinol*, 2014; **171**: 509–18.
80. Lem AJ, van der Kaay DC, de Ridder MA, *et al*. Adult height in short children born SGA treated with growth hormone and gonadotropin releasing hormone analog: results of a randomized, dose-response GH trial. *J Clin Endocrinol Metab*, 2012; **97**: 4096–105.
81. Breukhoven PE, Kerkhof GF, van Dijk M, Hokken-Koelega AC. Long-term impact of GH treatment during childhood on body composition and fat distribution in young adults born SGA. *J Clin Endocrinol Metab*, 2011; **96**: 3710–16.
82. de Arriba A, Dominguez M, Labarta JI, *et al*. Metabolic syndrome and endothelial dysfunction in a population born small for gestational age relationship to growth and Gh therapy. *Pediatr Endocrinol Rev*, 2013; **10**: 297–307.
83. Van der Steen M, Lem AJ, van der Kaay DC, *et al*. Metabolic health in short children born small for gestational age treated with growth hormone and gonadotropin-releasing hormone analog: results of a randomized, dose-response trial. *J Clin Endocrinol Metab*, 2015; **100**: 3725–34
84. Smeets CCJ, van der Steen M, Renes JS, Hokken-Koelega ACS. Bone mineral density after cessation of gh treatment in young adults born SGA: a 5-year longitudinal study. *J Clin Endocrinol Metab*, 2017; **102**: 3508–16.
85. Renes JS, Willemsen RH, Wagner A, Finken MJ, Hokken-Koelega AC. Bloom syndrome in short children born small for gestational age: a challenging diagnosis. *J Clin Endocrinol Metab*, 2013; **98**: 3932–8.
86. Lem AJ, de Rijke YB, van Toor H, de Ridder MA, Visser TJ, Hokken-Koelega AC. Serum thyroid hormone levels in healthy children from birth to adulthood and in short children born small for gestational age. *J Clin Endocrinol Metab*, 2012; **97**: 3170–8.

7.1.4 Growth Disorders with No Defined Aetiology

Minu George and Steven Chernausek

Introduction

Short stature or slowing of linear growth are frequent problems presenting to providers of healthcare for children. While a comprehensive history and physical examination, with review of the growth record, and limited laboratory testing occasionally reveals an aetiology, more commonly the cause is not obvious and the patient is classified as having either idiopathic short stature, familial short stature, or constitutional delay of growth and adolescence. These are largely diagnoses of exclusion where one relies on characteristic clinical features and the absence of signs of a defined condition. The 2008 Consensus Workshop on the Diagnosis and Treatment of Children with Idiopathic Short Stature (ISS) defined ISS as a condition where the height is below –2 standard deviations for age and sex when compared to appropriate normative data for the population, where there is no evidence for 'systemic, endocrine, nutritional, or chromosomal abnormalities', and in whom birth weight is normal [1]. Individuals are further subcategorized as to whether the short stature has a familial basis (i.e. those whose current stature is within the range expected for parents' height, and those with delayed bone age, which is assumed a harbinger of delayed adolescence).

Previous consensus statements have addressed idiopathic short stature and the short child born small for gestational age (SGA) separately [1, 2]. Although nearly 90% of individuals born small for gestational age will experience accelerated growth following birth and reach heights in the normal range by age 2 years, those who do not show such 'catch-up growth' constitute a significant proportion of individuals seen by healthcare providers for short stature. Similar to idiopathic short stature, the reason short stature persists in these individuals remains an enigma. Thus, separation of the short child born SGA from those classified as isolated idiopathic short stature is arbitrary and in part reflects the licensing and approval processes for recombinant human growth hormone (rhGH) therapy as each have separate indications. In this chapter, we briefly describe the epidemiology and clinical presentation of children with growth deficits. We discuss the impact of the disorder on the patients' well-being and adult height and propose an approach to diagnosis that incorporates newer available genetic testing. Lastly, we review the risks and benefits of pharmacologic interventions that would stimulate growth and/or increase adult height. We occasionally include data concerning short children born small for gestational age along with those classified as ISS according to consensus guideline when relevant. The reasons for a broad-based approach are the overlap of phenotypes, the lack of evidence that these entities are entirely physiologically distinct, and the fact that the issues concerning prognosis and response to therapy are shared.

Demographics

It is disconcerting to parents and healthcare providers alike that the majority of children who stand below –2 SDS for height fall into the category of ISS. Both would like an explanation for the condition and clarity with regards to expected adult stature. From population studies it is estimated that approximately 5% of short children have a pathological condition [3–5]. In another 15% a low birth weight or length for gestational age (SGA) is found [6]. The 2008 consensus statement for ISS estimated that about 60–80% of children presenting to a paediatric endocrinologist with short stature would have ISS. Since that time there has been tremendous progress in applying molecular genetic approaches to evaluation of short stature. As such, we now suggest that 25–40% of children diagnosed with idiopathic short stature could receive a molecular diagnosis [7]. The shorter the stature, the greater the prevalence of organic disease, and therefore the better chance of finding a pathological cause with advanced genetic testing.

Pathophysiology

Normal somatic growth requires the secretion of growth-controlling hormones, along with the appropriate response of growing tissues via cognate receptors and signal transduction pathways, and the expression of genes regulating cell division and protein synthesis, all in the context of adequate nutrition and oxygenation. Thus, the causes of aberrant growth are many and result from a wide variety of hormonal, genetic, or nutritional disturbances. It seems an oxymoron to consider 'causes' of something 'idiopathic', but in fact the term translates to 'one's own suffering' (Gr. idios + pathos), describing a condition intrinsic to an individual. There is always an explanation for a growth deficit; it may just not be obvious at the outset. In fact, the mechanisms involved in the pathophysiology of idiopathic short stature are frequently in play in more pronounced disorders recognized as distinct diseases. The various conditions can be categorized as follows (see Table 7.1.4.1).

Defects in the Growth Hormone/Insulin-Like Growth Factor (GH/IGF) Axis

For the purposes of this discussion the GH/IGF axis will consist of GH and the IGFs (1 and 2), their respective receptors, the factors directly controlling their secretion (e.g. GHRH), bioactivity (e.g. IGF binding proteins) and the intracellular signalling pathways that follow receptor activation (e.g. STAT5B for GHR, AKT for IGF1R). The relevance of this axis to ISS is based on several

Table 7.1.4.1 Idiopathic short stature and related conditions: examples of genes involved

Classification	Gene	Classic disease	Non-classic/hypomorphic	Est ISS prevalence	Refs
GH/IGF Axis	*GH1, GHRHR GHSR*	Isolated GH deficiency	Normal prenatal growth, slow postnatal growth	Rare	[44, 75]
	IGF1	IGF deficiency with severe pre and postnatal growth failure. Microcephaly. Sensorineural deafness	Both pre and postnatal growth affected, low circulating IGF1	Rare	[10, 12, 75]
	IGF1R	IGF resistance syndrome, severe pre an postnatal growth failure	Both pre and postnatal growth affected, normal or increased circulating IGF1	1–2%	[10, 12, 75]
	IGFALS	Modest reduction in postnatal growth—low circulating IGF1	Unknown	2–10%	[16, 75]
	PAPPA-2	Modest short stature, increased circulating IGF1, Skeletal abnormalities	Unknown	Unknown	[76]
Growth Plate/Skeleton	*ACAN*	Spondyloepiphyseal or metaphyseal dysplasias-	Short stature with advanced bone age	2%	[21, 75, 77, 78]
	FGR3	Achondroplasia	Hypochondroplasia, idiopathic short stature	1–2%	[75]
	NPR2	Acromesomelic dysplasia	Idiopathic short stature +/– disproportionate	2–6%	[79–82]
	SHOX	Leri-Weill dyschondrosteosis, Turner syndrome	Idiopathic short stature +/– disproportionate	3–15%	[23, 83, 84]
Intrinsic Cellular Growth	*BLM*	Bloom syndrome	Prenatal growth failure—syndromic features may be subtle—propensity for malignancy	Rare	[85]
	FANCA	Fanconi anaemia	Prenatal growth failure, GH deficiency in some, propensity for malignancy	Rare	[86]
	CUL7, OBSL1, CCDC8	3-M syndrome	Probable pre and postnatal growth failure +/– syndromic features	unknown	[44]
	KRAS, MAPK2, PTNP11	Noonan spectrum	Not SGA, ISS +/– syndromic features	~ 1–2%	[44]

Genes noted here are only representative of a much larger number that have been implicated in short stature/growth failure. For a more extensive listing of genes associated with monogenic disorders, see Wit *et al*. [87].

factors. First, the GH/IGF axis is the major controller of growth in mammals. Studies in mice show that up to 70% of adult size is determined by the combination of GH and IGF1 action [8]. Genetic mutations in humans confirm that human growth is also profoundly influenced by GH and IGF [9]. Second, there is evidence that subtle defects in GH or IGF secretion or action are frequently found in children with idiopathic short stature and related conditions [10]. Third, rhGH hormone and rhIGF1 are presently the main therapeutic options for treating individuals with short stature of any cause.

Several lines of evidence suggest that impairment of GH action or IGF secretion is restraining growth in a significant proportion of ISS children. Up to 50% of individuals classified as ISS have subnormal circulating concentrations of IGF1 even when indices of GH release are normal [10]. In addition, ISS patients with low IGF1 levels appear less sensitive to GH administration in terms of both growth stimulation and IGF1 production [11]. These patients could be considered to have 'Non-classical' GH/IGF defects. Such data has provoked examination of relevant genes in patients with ISS, revealing that mutations in *SHOX* and *IGF1-R* are more frequent in children with ISS [12], although still not explaining the condition in most patients.

There is also a potential causative role of genetic variants in the type 1 IGF receptor (IGF1R) for short stature of undetermined aetiology. Patients with definitive molecular defects characteristically manifest with substantial pre- and postnatal growth failure accompanied by normal or elevated circulating concentrations of IGF1 [13]. *IGF1R* mutations are found in up to 2% of short children born SGA [14]. The IGF1R is a critical element in the pathway of cellular growth control and heterozygous variations that are less disruptive are present in the normal population and appear to impact growth [15].

Recent studies have identified novel defects in the IGF axis that influence stature. The vast majority of IGF1 in the bloodstream circulates as a ternary complex composed of IGFBP-3, the acid labile subunit (ALS), and IGF1 itself. Genetic defects in *IGFALS* lead to low circulating levels of IGF1 and generally modest short stature [16]. A growth deficit is also present in patients with mutations in pregnancy-associated plasma protein A2 (PAPPA2). PAPPA2 selectively cleaves IGF binding proteins 3 and 5 and thereby facilitates the release of IGF1 from the IGFBPs, which is required for IGF receptor interaction. Individuals with *PAPPA2* mutations have short stature in addition to other skeletal findings. Mutations in *IGFALS* have been reported in surveys of individuals with ISS. The frequency of *PAPPA2* variations is unknown.

Intrinsic Defects in Growing Tissues

Certain growth disorders are the result of aberrations in basic mechanisms of cellular growth and consequently affect all growing tissues. Conditions that retard the cell cycle or increase apoptosis would be expected to result in global reduction in growth and be expressed by both pre- and postnatal growth failure. Fanconi anaemia and Bloom syndrome are examples of such disorders. They are characterized by DNA instability and are manifest by low birthweight, slow growth, and a propensity to malignancy. Other conditions in this category are the RASopathies (RAS/MAPK pathway syndromes [17]) which include Noonan and Noonan-like syndromes and the 3-M syndrome (typically mutation of *CUL7, OBSL1* or *CCDC8*) [18]. Though uncommon, mutations in these pathways can be found in individuals in whom common genetic and endocrine conditions have been excluded.

Disorders of the Growth Plate and Skeleton

Growth in stature is the sole result of division, maturation, and expansion within the growth plates of the axial skeleton. Thus, it is not surprising that the cause of short stature in many, including those with ISS, involve defects in chondrocyte response to hormonal stimuli or its division and maturation within the growth plate itself [19]. Genome-wide association studies have identified 700 distinct loci that determine adult stature [20]. A large number of these involve genes that are active within the growth plate. In addition, genetic defects that cause severe short stature and obvious skeletal dysplasia in some present with a much more modest phenotype when the mutation results in a lesser change in function. Conditions associated with aberrations in *ACAN* and *SHOX* are illustrative. Mutations in *ACAN* (encoding a matrix protein) are the basis for spondyloepiphyseal and spondyloepimetaphyseal dysplasia but are also found in approximately 2% of individuals with ISS. Those affected typically show modest changes in body segment proportion and an advanced bone age, which is characteristic of the syndrome [21, 22]. SHOX is a transcription factor located on the X chromosome and is the main reason why girls with Turner syndrome are short. The severe form of SHOX haplo-insufficiency results in Leri-Weill dyschondrosteosis but SHOX defects are also found in approximately 5% of individuals with ISS [23, 24].

Psychological Aspects of Short Stature

Shortness during childhood could be perceived as a disadvantage and might interfere with normal psychological development. No clear-cut relationships have been established between reduced stature and psychological distress. The question of psychosocial morbidity as a justification for treatment in ISS has been asserted by proponents of rhGH treatment in non-GH deficient children but continues to be a contentious subject. There is a (mis)perception among children and adults that taller is better and that short people who are assertive have a 'complex' that attempts to compensate for their size.

Psychosocial Stress and Adjustment

The belief that short stature is an impediment to social adjustment due to more negative social experiences, including teasing, being treated younger than your age (juvenilization), having fewer friends, and less social acceptance is pervasive [25]. Studies of patients referred for short stature (normal variants as well as those with pathological conditions) showed slightly more than half were regularly teased about being short. This is about the same proportion as juvenilization [26]. As adults, those with short stature also experience more problems driving a car, have less income, hold fewer higher-level academic degrees, and have problems with dating and marriage; the latter being particularly the case for shorter men [27]. Therefore, it is not surprising that children and adults of both genders desire to be taller [28]. There are reports that suggest short children are more likely to be bullied than their taller peers [5]. Paediatric endocrinologists also share these beliefs; more than 50% report the emotional well-being of children and adults with heights less than third percentile is sometimes impaired and more than a quarter believe it is always or often impaired [29]. Therefore these children and adolescents may experience more psychosocial stress. However, not all short children are equally affected and furthermore the stressors are not directly related to the degree of short stature [26].

In contrast to studies of children or families seeking medical care, community-based studies reach different conclusions. A large, longitudinal, non-interventional study of 7–9-year-old children conducted in a single region in the UK (the Wessex study) showed no evidence that stature of less than third percentile was associated with significant psychosocial or academic disadvantages [30]. Despite the desire to be taller and reports of bullying there was no measurable effect on self-esteem or school performance. A similar study in the United States of 11–18-year-olds showed height was unrelated to popularity, total number of friends, or whether friendships were reciprocated. Although children were often judged younger than they actually were, this was unrelated to social acceptance among peers. In addition, the findings did not vary based on gender. [31].

Educational Attainment and Economic Status

The effects of adult height on educational attainment and the type of employment and income level achieved have been explored is some detail [32]. Studies of stature and income frequently report that taller men and women earn more than their shorter peers do [33]. Upon further scrutiny, it appears that taller people earn more because of stronger cognitive abilities throughout their lives. Height is associated with better mental and physical health and cognitive function in adult life [34]. Secondary analysis supports the hypothesis that stronger cognitive ability explains the difference in occupational status and income and that height only serves as a proxy measure for an attribute of greater importance (i.e. cognitive ability). Thus, height is a marker of early childhood health and nutrition, both of which affect cognitive development. This suggests that childhood nutrition and environment contribute to cognitive development which subsequently influences educational attainment, school achievement and, later, occupational choice. All of these may in turn affect health and cognitive function in old age [34].

Health-Related Quality of Life

Examination of health-related quality of life (HrQoL) is a novel analytical approach to provide a better understanding of the psychosocial consequences of short stature and can help clinicians identify vulnerable groups of paediatric patients that could benefit from further assessment and referral. Use of the instrument in a study of 345

children aged 8–18 years showed that children with short stature, particularly those with current height below –2 SD, are at increased risk for internalizing problems. Children/adolescents with short stature presented more parent-reported internalizing problems and lower self- and parent-reported condition-specific HrQoL, compared to patients with an achieved height above –2 SD. GH treated children self-reported better HrQoL than the GH untreated group [35].

Functional Impairment

An association between short stature and functional impairment (e.g. decreased intelligence, poor academic achievement, and behaviour problems) has long been postulated, but there is no consensus as to whether such a cause-and-effect relationship exists. In the Wessex Growth Study short children up to 13 years of age achieved significantly lower scores on measures of intelligence quotient, reading attainment, and basic number skills, and were less satisfied with their height. However, they did not differ significantly from the control children of average stature on measures of self-esteem, self-perception, parents' perception, or behaviour. Furthermore, the short children more frequently came from working class homes and therefore their attainment scores were predicted by class and IQ together rather than by height alone. These results provide only limited support for the hypothesis that short children are disadvantaged [36]. A systematic review of the association of short stature and functional impairments showed most short children scored within the normal range in functional tests; however, within studies, short stature was often associated with decreased intelligence, academic achievement, and visual-motor skills. On average, children with short stature score lower than their peers on functional tests but few short children scored outside the normal range. Among six studies examining academic achievement in children with ISS or constitutional delay of growth and adolescence, all short children scored at least within approximately 1 SD of normal scores. However, most of these same studies found that children with short stature were more likely to have lower academic achievement scores than children of average height. These studies show no evidence regarding whether treatment of short stature improves functional impairment adaptation [37]. Therefore, while treatment of short stature has been justified by potential improvements in functional impairment, quality of life, and economic status due to increased stature, the data remain inconclusive, and the subject controversial.

Approach to the Patient

Certain diagnoses must be excluded to categorize a patient as having ISS (see Table 7.1.4.2). Those listed are conditions that have the potential to slow growth with few outward manifestations. In addition, the growth pattern for ISS typically has an onset of the growth attenuation either before birth or within the first years of life. A growth chart that displays normal growth for several years after which the growth rate markedly slows is atypical for ISS and indicates an acquired cause of growth failure should be pursued.

Defining short stature as 2 SD or more below the mean for age and sex will identify many children with healthy variations in height and lacking underlying pathology. The condition is simply benign normal variation. A smaller number of children will have a pathologic condition affecting growth and perhaps other aspects of their health. It can be a challenge for the clinician to identify those and therefore many children receive the nebulous diagnosis of idiopathic or familial short stature after routine studies. Thus, assuming the conditions listed in Table 7.1.4.2 have been eliminated, several questions must be addressed. To what extent is the patient's stature compromising quality of life? What is the likelihood that the patient will reach an acceptable adult stature? Should additional testing be conducted in an attempt to reach a definitive diagnosis? Is treatment of the growth deficit warranted and, if so, with what modality?

Table 7.1.4.2 Diagnoses to exclude before classification as ISS and related conditions

Condition	Standard Testing	Comment
Growth hormone deficiency	IGF1, IGFBP3, GH stimulation	
Hypothyroidism	T4, TSH	Thyroid dysgenesis may mimic ISS
Turner syndrome in girls	Karyotype or sex chromosome FISH	May present with short stature as the only obvious clinical finding
Bone dysplasias & chondrodystrophies	Anthropometry, consider skeletal survey	Probably underdiagnosed
Occult GI disease (e.g. inflammatory bowel disease), malnutrition	CBC, ESR, or CRP, celiac screen	
Renal disease	Electrolytes, BUN, Cr, UA	Rare—prenatal onset can be SGA
Other chronic systemic diseases (e.g. asthma, cyanotic congenital heart disease)	Specific tests prompted by history and physical examination	In most cases, cause would be obvious

A foreknowledge of future growth and ultimate height as an adult would be helpful in addressing those questions. Assessment of growth velocity over 6–12 months is an essential element in the evaluation. It is especially relevant for those born SGA as the majority have accelerated growth following birth and reach a normal size by age 2 years [38]. A below average growth velocity after age 2 years would be a concern for any patient.

Accurately predicting adult height in a young child is difficult. Most methods rely primarily on bone age determination and were derived from individuals with variants of normal stature, not from those with pathologic conditions [39]. However, as a group, about two-thirds of children with ISS ultimately reach an adult stature within the lower normal range [40], the increase in SD score from childhood to final height being attributed to a delay in pubertal onset that prolongs the growing period. Thus, those with modest height deficits early in life (e.g. –2 to –2.5 SD) and who maintain near normal growth rates may need no treatment other than reassurance and perhaps androgen therapy for those with constitutional delay of growth and adolescence.

Figure 7.1.4.1 illustrates an approach to pursue specific diagnoses that might not be evident following standard testing in the short child. Because of the tremendous technical advances allowing the detection of meaningful variations in the human genome at reasonable cost, we believe that a genetic approach to uncovering the cause short stature is warranted for many [7]. Five distinct clinical scenarios that encompass the majority of patients with the initial diagnosis of ISS are described in Figure 7.1.4.1. Each scenario

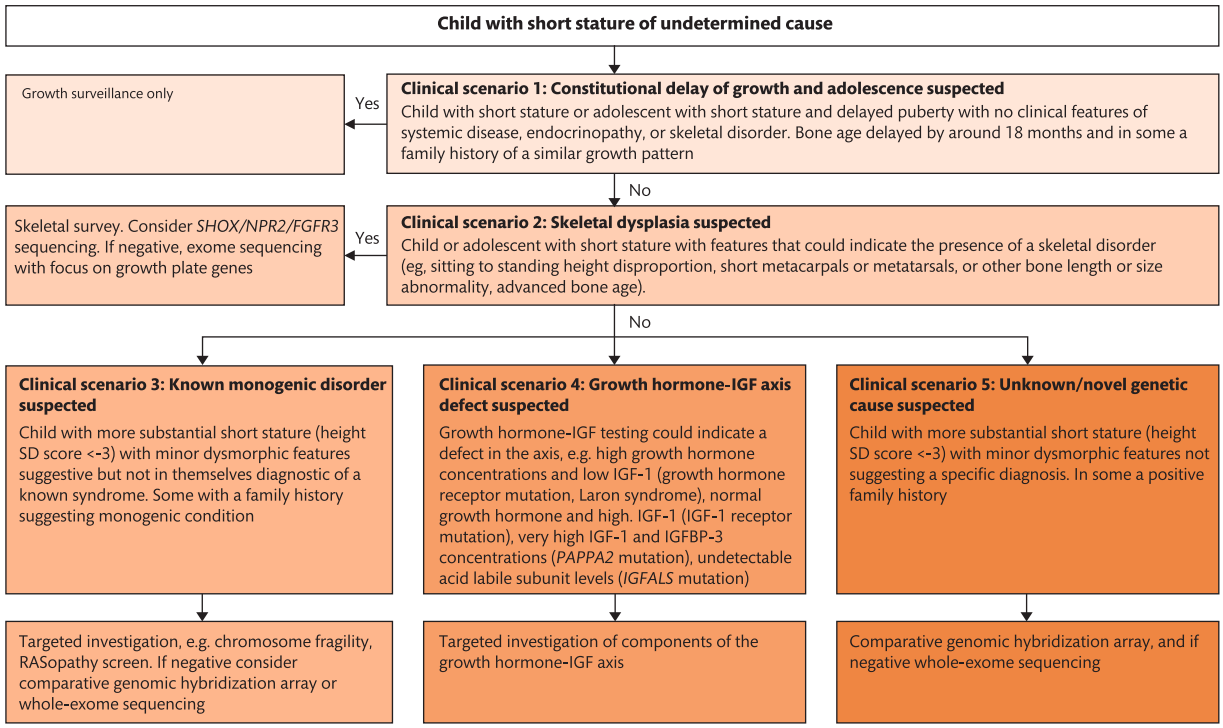


Figure 7.1.4.1 Proposed schema for molecular diagnosis in a child with short stature of undetermined cause. Each box contains a potential clinical scenario leading to an investigative strategy. Short stature of undetermined cause is defined as short stature more than −2 SDs below the mean, with or without being born small for gestational age, with no readily recognizable syndrome diagnosis, no substantial microcephaly, normal screening investigations (e.g. karyotype [in females], routine blood testing to exclude occult systemic disease, and thyroid function), and no growth hormone deficiency by standard growth hormone stimulation testing. SHOX-short stature homeobox. IGF-insulin-like growth factor.

corresponds roughly to the pathophysiologic groups described previously and is associated with specific diagnostic steps. For example, patients with modest short stature and a history and physical examination consistent with constitutional delay of growth and adolescence (scenario 1) are simply observed. Those showing subtle features of skeletal dysplasia would have assessment of select genes such as *SHOX* and *NPR2,* as mutations in those are consistent with that phenotype. Box 7.1.4.1 delineates features that would increase the likelihood of attaining a genetic diagnosis.

Another group includes those patients with evidence for defects in the GH/IGF axis based on measures of circulating IGF1 and the GH response to stimuli. The increase in circulating IGF1 following GH administration (the IGF generation test) has been suggested as a way to identify intermediate forms of GH resistance. Although attractive in concept, it has not proved sufficiently sensitive and is of limited value [41]. Instead, a molecular genetic approach is indicated, with the hormonal measures dictating which genes are analysed. Low levels of IGF1 in the face of normal or elevated GH release are consistent with defects in *GHR, IGF1, IGFALS*, whereas if both IGF1 and GH are increased, defects in *IGF1R* are more likely. A recent report supports the value of this approach by identifying genetic defects for half the cases when a problem with GH/IGF axis is suspected on clinical/hormonal grounds [42].

The fifth scenario addresses those individuals who appear to have a pathologic condition affecting growth by virtue of dysmorphic features or severe short stature, but for whom no specific diagnosis or pathway is suspected. Those individuals would undergo an untargeted survey of their genome starting with comparative genome hybridization and ending with whole exome sequencing. Diagnostic yields may approach one-third depending on selection of subjects [43, 44].

Reaching a specific diagnosis has several benefits. Foremost, it ends the diagnostic odyssey and can provide the family with additional information regarding prognosis, the potential value of treatment, and the risk for relatives and offspring. However, there remain drawbacks and limitations to pursuing specific diagnoses in this way. In many cases involving novel or modest changes in genetic structure, solid information about height prognosis or response to growth-promoting agents is lacking. In countries where ISS is an approved indication for rhGH therapy, it is possible that reaching a specific diagnosis removes the patient from the ISS category and

Box 7.1.4.1 Factors favouring genetic studies in ISS

- Severe growth deficit (Ht. <−3 SD)
- Height much below target ht. (e.g. mid-parental height + 2 SD, patient −2SD)
- Pedigree suggestive of monogenic disorder
- Dysmorphic features
- Developmental delay
- Disproportionate short stature

makes acquiring treatment more difficult. In addition, there are circumstances where the genetic basis of the condition will not be uncovered using whole exome sequencing. Epigenetic changes or mutations in non-coding regulatory regions will not be detected. In certain cases short stature will be the result of a combination of genetic variations each of which alone would have a modest effect on stature, but when combined interact to yield a more pronounced phenotype. The current approaches typically seek a monogenic explanation and are inadequate to deal with more complex situations [45]. Despite this, the advantages of reaching a specific diagnosis outweigh the disadvantages.

Treatment to Increase Stature

Medical treatments that influence stature either involve increasing the growth rate (e.g. rhGH, rhIGF1), prolonging the growth period (e.g. GnRH agonists, aromatase inhibition), or combine the two effects (e.g. rhGH and aromatase inhibition). At the current time, the only treatment approved by national regulatory agencies for ISS or the short child born SGA is rhGH. Moreover, approvals are not uniform across the continents. For example, rhGH is approved for ISS by USFDA, but not EMEA and the criteria for treatment of the short SGA patient also differs by agency. Data relevant to ISS for both approved and off-label therapies are discussed in the following section. The reader should review current literature, regulatory agency publications, and professional society guidelines for specific prescribing information and recommendations.

Human Growth Hormone

Human growth hormone has been prescribed since 1957 initially to treat GH deficiency and later for short stature due to other causes. It was first isolated from cadaveric pituitary glands, requiring arduous and costly collection of glands, followed by extraction and purification of the hormone. This limited supply restricted its use to children with severe GH deficiency who were treated with low doses and suboptimal schedules. The use of pituitary-derived GH was discontinued in 1985 following an outbreak of Creutzfeldt-Jakob disease and all subsequent treatment has been with recombinant hormone (r-hGH). The development of recombinant DNA-derived GH, allowed the production of virtually unlimited amounts of GH, leading to the approval as therapy for additional childhood conditions characterized by short stature.

The story of GH treatment in children for ISS begins in 1983 when 15 non-GH deficient short children were treated with cadaveric GH for 6 months and increased their growth rates by more than 2 cm per year [46]. Subsequent trials (see Table 7.1.4.3) demonstrated efficacy and led to the 2003 US FDA approval for GH treatment of children with ISS with heights at or less than −2.25 standard deviations (≤1.2 percentile) below the mean for age and sex and pretreatment growth rates unlikely to permit attainment of adult height in the normal range. The effectiveness of GH therapy in ISS was affirmed at the 2008 ISS Consensus Workshop which concluded that children with ISS whose heights were less than −2 SDS for age and sex and were more than 2 SDS below their mid-parental target height or had a predicted height less than -2 SDS warrant consideration for treatment [1].

Controversy persists despite the consensus statement as to the effectiveness and value of GH treatment for children with ISS [47]. A detailed meta-analysis of the rhGH trials examined the impact of rhGH therapy of children with ISS [48]. Out of the 19 GH ISS trials to reach adult height (mean duration of 5.4 years), 10 met the criteria for controlled trials; of which three were randomized

Table 7.1.4.3a Summary of results from the randomized controlled trials of GH therapy in children with ISS treated to final height

Study & group	N	GH dose (mg/kg/day)	Mean age at start of therapy (yrs)	Mean (SD) years of therapy (yrs)	Height at baseline (SD score)	Mean adult height (SD score)	Mean height gain (SD score)	Difference between case and control in adult height (SD score)
McCaughey *et al.* 1998 [88]								
Treated	8	0.04	6.24 (0.38)	6.2	−2.52 (0.26)	−1.14 (1.06)	1.38 (0.7)	1.23
Untreated	6	–	6.14 (0.64)	6.14	−2.55 (0.32)	−2.37 (0.46)	0.18 (0.4)	–
Leschek *et al.* 2004 [89]								
Treated	22	0.03	12.5 (1.6)	4.4	−2.7 (0.6)	−1.77 (0.80)	0.93 (0.75)	0.57
Untreated	11	–	12.9 (1.1)	4.1	−2.8 (0.6)	−2.34 (0.56)	0.46 (0.23)	–
Albertsson-Wikland *et al.* 2008 [90]								
Treated with 0.033 mg/kg/day	18	0.033	11.5 (1.3)	–	–	-1.7 (0.68)	1.20 (0.82)	0.5
Treated with 0.067 mg/kg/day	31	0.067	11.5 (1.3)	–	–	-1.5 (0.84)	1.30 (0.73)	0.7
Combined treatment groups	49	(0.0330–0.067)	11.5 (1.3)	5.64	−2.84 (0.56)	-1.6 (0.68)	1.24 (0.82)	0.6
Untreated	19	–	12 (1.6)	–	−2.76 (0.39)	-2.2 (0.75)	0.40 (0.62)	–

Table 7.1.4.3b Summary of results from the non-randomized controlled trials of GH therapy in children with ISS treated to final height

Study & group	N	GH dose (mg/kg/day)	Mean age at start of therapy yrs. (SD)	Mean years of therapy	Height at baseline SD score (SD)	Mean adult height SD score (SD)	Mean height gain SD score (SD)	Difference of treated vs. untreated mean height gain SD score
Wit *et al.* 1995 [91]								
Treated	12	0.02	9.2 (1.6)	5.7	−3.8 (0.7)	−2.4 (0.9)	1.4 (0.9)	0.6
Untreated	27	–	10.5 (1.2)		−3.2 (0.3)	−2.4 (0.7)	0.8 (0.5)	–
Hindmarsh *et al.* 1996 [92]								
Treated	16	0.02–0.04	8.35 (1.88)	7.5	−2.17 (0.58)	−1.33 (0.94)	0.84 (0.76)	0.38
Untreated	10	–	7.62 (1.50)	–	−2.34 (0.61)	−1.88 (0.57)	0.46 (0.6)	–
Buchlis *et al.* 1998 [93]								
Treated	36	0.04	11.9 (2.8)	3.5	−2.9 (0.6)	−1.5(0.8)	1.4 (0.7)	0.6
Untreated	58	–	12.5 (2.5)	–	−2.9 (0.6)	−2.1(1.0)	0.8 (0.8)	–
Lopez-Siguero *et al.* 2000 [94]								
Treated	35	0.02	11.1 (1.4)	5.3	−2.78 (0.5)	−1.31 (0.7)	1.47 (0.6)	0.9
Untreated	42	–	10.8 (2.2)		−2.4 (0.4)	−2.03 (0.7)	0.37 (0.8)	
Coutant *et al.* 2001 [95]								
Treated	32	0.02	11.7 (2.0)	3.9	−3.0 (0.67)	−2.1 (0.76)	0.9 (0.57)	0.24
Untreated	51	–	12.1 (2.8)		−2.74 (0.64)	−2.08 (1.01)	0.66 (0.89)	
Wit *et al.* 2002 [96]								
Treated	30	0.03–0.04	10.7 (2.2)	5.9	−3.3 (0.5)	−1.9 (0.9)	1.4 (0.9)	0.8
Untreated	64		10.9 (1.1)		−3.0 (0.5)	−2.4 (0.8)	0.6 (0.6)	

trials and seven were non-randomized. The three randomized controlled trials included 115 children, 79 cases, and 36 controls whereas the seven non-randomized control included 477 children, 181 cases, and 296 controls. Analysis of the randomized trials showed that the adult height of the GH-treated exceeded that of untreated children by 0.65 SDs, or about 4 cm, after 5.4 years of rhGH therapy [48]. In the non-randomized controlled studies, the adult height of the rhGH-treated patients exceeded that of controls by only 0.45 SDS or about 3 cm on average. The relationship between height gain and quality of life or psychosocial outcomes was not addressed in this analysis. These authors concluded that: (1) no single, high quality evidence RCT has been carried out up to the achievement of adult height; (2) the overall magnitude of rhGH effects on reducing the adult height deficit in children with ISS is on average less that achieved with other conditions for which rhGH was licensed; and (3) the individual response to therapy is highly variable.

GH treatment of the short child born small for gestational age. The definition of 'small for gestational age' has varied from less than the 10th percentile for birth weight to less than −2 SDs for weight or length. The latter definition is more commonly accepted as indicating prenatal growth restraint and is the definition applied in a 2007 consensus statement [2]. Children born SGA represent a heterogeneous group at risk for multiple long-term complications including short adult stature along with other metabolic and endocrine disorders. For the 10–15% of SGA children who do not experience catch up growth within the first few years, rhGH is effective in increasing adult height, particularly when started early in life and at a sufficient dosing. In addition, rhGH treatment also positively affects body composition with an increase in lean mass leading to a decreased risk of metabolic complications in SGA [49]. Additional information concerning children born small for gestational age can be found in Chapter 7.1.3.

Safety during rhGH treatment. Human growth hormone is a safe drug when prescribed as recommended with adverse events rarely observed (78 per 100 000 treatment years for patients with ISS) [50]. Adverse effects are related to the stimulation of skeletal growth (e.g. slipped capital femoral epiphysis and scoliosis), the water retentive action of growth hormone (e.g. intracranial hypertension) or the effect of GH on glucose metabolism (e.g. decreased insulin sensitivity) [51]. See Box 7.1.4.2. In children, the frequency of adverse effects depends on the underlying disorder and is similar or lower in patients with ISS when compared to other rhGH-treated conditions [52]. At dosages of 0.24–0.37 mg/kg/week, rhGH treatment in children with ISS does not cause hyperglycaemia. At dosages ≥0.3 mg/kg/week, a dose-dependent increase in fasting and stimulated insulin levels is observed [50]. Current evidence derived from 'on-treatment' surveillance studies in GH deficient children indicates that rhGH does not increase the risk for new malignancies nor is there an increased risk of recurrence of primary cancer in cancer survivors. For the latter group there is suggestive evidence that the development second malignancies may be increased by GH treatment [53]. Thus, the safety profile of rhGH at doses ≤0.37 mg/kg/week for the treatment of children with ISS is similar to or better

Box 7.1.4.2 Effects of growth hormone treatment in children

Known adverse effects of growth hormone treatment in children

- Intracranial hypertension
- Slipped capital femoral epiphysis
- Progression of scoliosis

Potential/theoretic adverse effects of growth hormone treatment in children

- Pancreatitis
- Increased risk for adverse cardiovascular events during adulthood
- Increased risk of second malignancy in cancer survivors

than that seen in other rhGH-treated conditions and is not associated with any predictable adverse events [50].

Long-term safety of rhGH. With the expanded use of rhGH in children and the higher doses employed, concerns about the potential for adverse effects developing long after the completion of GH therapy have arisen. Studies that address this problem have been extremely limited until recently. The SAGhE (Safety and Appropriateness of Growth Hormone Treatments in Europe) cohort is the largest and longest follow-up study of GH-treated patients independent of industry [54, 55]. This ongoing study assesses the cause-specific mortality and cancer incidence of now over 24 000 patients from eight European countries treated during childhood with GH since 1984. As such, it is a major resource for investigating cancer and mortality risks in rhGH patients.

The first reports from SAGhE were discordant, with some indicating increased mortality and risk for stroke [56, 57] while others found no excess deaths due to cancer or cardiovascular events [58]. The overall all cause standardized mortality ratios were noted to be increased in a subsequent systematic review, but malignancy and cardiovascular standardized mortality ratios were not significantly increased [59]. The validity of these early SAGhE data was questioned by many, especially since in the cohort from Belgium, Netherlands and Sweden there was no convincing evidence for the induction of new cancers or relapse of pre-existing malignancies following GH treatment [58]. Most recently a combined registry analysis of the long term safety of rhGH has shown no increased risk of mortality in children or adults treated with rhGH. No evidence of increased risk of stroke, new malignancy, leukaemia, non-leukaemic extracranial tumours or recurrence of malignancy in patients without predisposing risk factors was seen [60]. In children and adults it can be stated with certainty that rhGH given by the current dose ranges and indication especially the idiopathic short stature indication for this medication is considered remarkably safe [53].

Patient selection and dosing for rhGH. Treatment of ISS with rhGH assumes that the patient will benefit from having a stature that approximates that of peers during childhood and as an adult. The benefits may be psychosocial or functional in those with severe short stature, but there are no data that indicate GH-treated patients are otherwise healthier. Thus, it can be a challenge to determine which patients are most likely to benefit given the variability in response and concerns about long-term safety. The 2016 guidelines developed by the Pediatric Endocrine Society suggest 'a shared decision-making approach to pursuing GH treatment for a child with ISS. The decision can be made on a case-by-case basis after assessment of physical and psychological burdens, and discussion of risks and benefits'. These guidelines also recommend a starting dose of 34 µg/kg daily, increasing up to 67 µg/kg daily if needed. However, studies comparing dose regimens are limited. Typically the response to treatment is best in the first year, so for some patients a higher initial dose may be warranted. Lastly, patient's growth should be monitored closely given the unpredictability of response to treatment and GH discontinued in those who do not show a meaningful increase in growth rate above baseline. The response required to justify continued treatment may differ among providers, but a growth velocity of at least +1.0 SD for age and sex or an increase of at least 3 cm per year above baseline have been suggested as reasonable cut points [1].

Recombinant Human Insulin-Like Growth Factor 1 (rhIGF1)

Regulatory agencies within the United States, Japan, and Europe have approved rhIGF1 for treatment of short stature due to severe primary IGF1 deficiency, defined as height and circulating IGF1 concentrations 3 SD or more below means for age and sex with adequate GH secretion. Most reports of treatment largely reflect results from patients with GH insensitivity due to GH receptor deficiency (Laron syndrome) [61]. However, studies that are more recent have demonstrated short-term growth response to rhIGF1 alone and in combination with hGH in less severely affected individuals [62]. Because of the generally suppressed levels of IGF1 in ISS [10, 63], some children will qualify for treatment and could potentially respond. In ISS children who do not respond to GH treatment, rhIGF1 therapy is a theoretical option; however, data are lacking regarding efficacy and safety in this population [64].

Sex Steroids

Testosterone esters are commonly prescribed for boys with constitutional delay of growth and puberty (CDGP) as both short stature and lack of secondary sexual characteristics are of frequent concern to the patient. Treatment will accelerate growth and induce a masculine phenotype and body composition [65]. Testosterone esters at doses of 50–100 mg as intramuscular injections per month are effective in stimulating growth and puberty without attenuating adult height when used in boys older than 13 years with no signs of puberty and a predicted adult height in the normal range [66]. A 3–6 month course of these low doses of testosterone may suffice to induce spontaneous development of puberty and subsequent psychological well-being in this subgroup of CDGP. The oral preparation testosterone undecanoate (40 mg/day) can also be used, although its effect is considered more variable than that of intramuscular preparations. No data are available for newer cutaneous forms of androgens, for example gels or patches.

Oxandrolone is an anabolic androgenic steroid, which has a theoretical benefit of having less androgenic effects than testosterone

and additionally does not aromatize to oestrogen. It has been shown to increase height velocity in the short term in several studies but does not lead to increase in predicted or measured final adult height [67]. While it does offer the advantage of oral administration over IM for testosterone, it does have the disadvantages of being only weakly androgenic and carrying a small risk for hepatotoxicity. While European and US regulatory agencies have approved oxandrolone for treatment of short stature in girls with Turner syndrome, approval is currently lacking specifically for ISS. Given the lack of an effect on adult height and the scarcity of data for use in the population, its place in the management of ISS is uncertain.

Characteristics of growth in girls with CDGP is less well known. A proportion of CDGP girls do not attain adult height consistent with their mid-parental target height. Those with a lower height SDS in childhood have a reduction in final adult height. Low-dose oestrogen treatment, which might seem a logical approach, does not seem to influence adult height in girls with CDGP [68].

Aromatase inhibitors. Males with mutations in the aromatase enzyme CYP19A1 exceed their adult target heights because they cannot synthesize oestrogens from androgens which, in turn, delays growth plate maturation. This observation gave rise to the notion that aromatase inhibitors used during puberty could augment adult height by blocking oestrogen production and thereby prolonging growth during adolescence. Small controlled trials have been conducted in adolescent boys, the first of which was with letrozole in constitutional delay in growth and puberty with adult height increase [69] and afterwards in children with ISS alone [70] or in combination with rhGH [71]. Combination therapy with AI/GH for up to 36 months increases height in pubertal boys with ISS more than either GH or AI alone. Height gains from baseline stature for those treated for 36 months were: AI, mean, +23.8 cm; GH, +26.7 cm; and AI/GH, +30.7 cm. These gains in height compare favourably with the expected height gain from 14.1 to 17.4 years of +13.0 cm for boys with a height SDS of −2.0 [71]. AIs alone and in combination with GH were well tolerated without adverse effects for up to 3 years. They concluded that AIs are an alternative treatment to enhance linear growth in adolescent boys with ISS, particularly in combination with GH [71]. This treatment modality is not used in girls due to a risk for ovarian cyst development and the potential for subsequent ovarian torsion. Concerns about adverse effects such as aberrant bone turnover leading to vertebral anomalies, reduced HDL cholesterol, increased insulin resistance, impairment of cognitive function, and long-term effects on spermatogenesis remain. These concerns coupled with a lack of larger studies put the use of aromatase inhibitors in the category of experimental drugs; they should be used cautiously and ideally in clinical trials that continue to evaluate safety and efficacy.

Gonadotropin Releasing-Hormone Analogues

Gonadotropin releasing-hormone analogues (GnRHa) block luteinizing hormone and follicle-stimulating hormone secretion, thereby decreasing gonadal production of androgens and oestrogens resulting in prolongation of growth. Although GnRHa treatment effectively delays epiphyseal fusion, it renders patients hypogonadal at a critical time of development. When administered to children with ISS, 3.5 years of GnRHa resulted in approximately 4 cm of increased adult height compared to controls [72]. The modest height gain compared to controls is due in part to the slowing of growth rate that accompanies the treatment. In an attempt to counteract this, rhGH and GnRHa have been administered in combination to patients with ISS and short children born SGA. Adolescents with Tanner stage 2–3, age and bone age less than 12 years for girls or less than 13 years for boys, height SD less than −2.0 SDS or between −1.0 and −2.0 SD plus a predicted adult height less than −2.0 SD were randomly allocated to receive GH plus GnRHa or no treatment for 3 years. FH was assessed at the age of 18 years or older in girls or 19 years or older in boys. Initial growth rates are increased by the rhGH addition, but the effect on final height is either modest or insignificant after stopping treatment as 50% of the predicted height gain at treatment withdrawal was lost at treatment withdrawal [73]. Most recently, a randomized control study treating patients with ISS at puberty onset with either GH with leuprorelin in combination or GH alone. Although promising, a new treatment adverse event emerged as more frequent bone fractures occurred in combination than GH [74]. These modest increases in adult height probably do not outweigh the disadvantages of delaying puberty. Therefore, while GnRHa are the standard treatment for central precocious puberty and are safe, its use alone or in combination with GH is controversial and not generally recommended for treatment of ISS when pubertal onset is at a normal age.

In summary, administration of rhGH continues to be the principal treatment to address compromised stature in children with ISS. In recent years, long-term benefits of the growth achieved from rhGH has been questioned because of modest or variable height gains and limited evidence of psychological benefit. That said, studies to date have not always used optimal doses and were conducted in relatively older children (i.e. 11+ years) for whom efficacy and benefit may be constrained. Therefore, for each child with substantial non-GH deficient short stature a risk to benefit ratio assessment must be made and a decision reached in conjunction with parent and patient wishes. Many ethical concerns such as the duty to provide appropriate assent and re-assent at critical junctions such as pubertal onset, protecting children from unnecessary treatment, considering fairness to non-treated children, and allocating healthcare resources responsibly should also be considered. Less expensive and invasive treatments or perhaps no growth enhancing treatment at all along with counselling remain viable options.

Conclusions

The slow-growing, short child of undetermined aetiology continues to challenge healthcare providers. A decade ago most were categorized simply as ISS after testing that typically included assessments of GH secretory status. In some cases, GH was administered to improve stature. Although the value of GH treatment for ISS is a matter of persistent debate, it is clear that the response is highly variable among individuals, with some showing remarkable response. Certainly, treatment may be warranted in children with severe short stature to alleviate restrictions on activities of daily living [75].

The availability of newer genetic tests and the appreciation for a much wider range of conditions attenuating growth has improved

our ability to make specific diagnoses and reduced the population of those labelled ISS. However, uncovering novel diagnoses presents additional challenges as we frequently lack information on the prognosis for adult height or response to growth-promoting therapies. Despite this, we believe that reaching a specific diagnosis is worthwhile and that with time a better understanding of the distinct conditions will lead to appropriate management and therapy.

REFERENCES

1. Cohen P, Rogol AD, Deal CL, *et al.* Consensus statement on the Diagnosis and Treatment of Children with Idiopathic Short Stature: a summary of the Growth Hormone Research Society, the Lawson Wilkins Pediatric Endocrine Society, and the European Society for Paediatric Endocrinology Workshop. *J Clin Endocrinol Metab*, 2008; **93**: 4210–17.
2. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. *J Clin Endocrinol Metab*, 2007; **92**: 804–10.
3. Ahmed ML, Allen AD, Sharma A, Macfarlane JA, Dunger DB. Evaluation of a district growth screening programme: the Oxford Growth Study. *Arch Dis Child*, 1993; **69**: 361–5.
4. Lindsay R, Feldkamp M, Harris D, Robertson J, Rallison M. Utah Growth Study: growth standards and the prevalence of growth hormone deficiency. *J Pediatr*, 1994; **125**: 29–35.
5. Voss LD, Mulligan J, Betts PR, Wilkin TJ. Poor growth in school entrants as an index of organic disease: the Wessex growth study. *BMJ*, 1992; **305**: 1400–2.
6. Grote FK. Assessment of short stature in children: auxological screening and diagnostic work-up, 2007. Available at: https://openaccess.leidenuniv.nl/handle/1887/11971 (accessed: 31 March 2020).
7. Murray PG, Clayton PE, Chernausek SD. A genetic approach to evaluation of short stature of undetermined cause. *Lancet Diabetes Endocrinol*, 2018; **6**: 564–74.
8. Lupu F, Terwilliger JD, Lee K, Segre GV, Efstratiadis A. Roles of growth hormone and insulin-like growth factor 1 in mouse postnatal growth. *Dev Biol*, 2001; **229**: 141–62.
9. David A, Hwa V, Metherell LA, *et al.* Evidence for a continuum of genetic, phenotypic, and biochemical abnormalities in children with growth hormone insensitivity. *Endocr Rev*, 2011; **32**: 472–97.
10. Clayton P, Bonnemaire M, Dutailly P, *et al.* Characterizing short stature by insulin-like growth factor axis status and genetic associations: results from the prospective, cross-sectional, epidemiogenetic EPIGROW study. *J Clin Endocrinol Metab*, 2013; **98**: E1122–30.
11. Cohen P, Germak J, Rogol AD, Weng W, Kappelgaard AM, Rosenfeld RG. Variable degree of growth hormone (GH) and insulin-like growth factor (IGF) sensitivity in children with idiopathic short stature compared with GH-deficient patients: evidence from an IGF-based dosing study of short children. *J Clin Endocrinol Metab*, 2010; **95**: 2089–98.
12. Caliebe J, Broekman S, Boogaard M, *et al.* IGF1, IGF1R and SHOX mutation analysis in short children born small sfor gestational age and short children with normal birth size (idiopathic short stature). *Horm Res Paediatr*, 2012; **77**: 250–60.
13. Klammt J, Kiess W, Pfaffle R. IGF1R mutations as cause of SGA. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 191–206.
14. Ester WA, van Duyvenvoorde HA, de Wit CC, *et al.* Two short children born small for gestational age with insulin-like growth factor 1 receptor haploinsufficiency illustrate the heterogeneity of its phenotype. *J Clin Endocrinol Metab*, 2009; **94**: 4717–27.
15. Kansra A, Martin LJ, Deka R, Dolan LM, Chernausek SD, eds. IGF receptor gene variants in normal adolescents: effect on postpubertal stature. The Endocrine Society's 89th Annual Meeting, 2007; Toronto, Canada.
16. Domene HM, Scaglia PA, Martinez AS, *et al.* Heterozygous IGFALS gene variants in idiopathic short stature and normal children: impact on height and the IGF system. *Horm Res Paediatr*, 2013; **80**: 413–23.
17. Aoki Y, Niihori T, Inoue S, Matsubara Y. Recent advances in RASopathies. *J Hum Genet*, 2016; **61**: 33–9.
18. Clayton PE, Hanson D, Magee L, *et al.* Exploring the spectrum of 3-M syndrome, a primordial short stature disorder of disrupted ubiquitination. *Clin Endocrinol (Oxf)*, 2012; **77**: 335–42.
19. Baron J, Savendahl L, De Luca F, *et al.* Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol*, 2015; **11**: 735–46.
20. Wood AR, Esko T, Yang J, *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*, 2014; **46**: 1173–86.
21. Hauer NN, Sticht H, Boppudi S, *et al.* Genetic screening confirms heterozygous mutations in ACAN as a major cause of idiopathic short stature. *Sci Rep*, 2017; **7**: 12225.
22. van der Steen M, Pfundt R, Maas SJ, Bakker-vanWaarde WM, Odink RJ, Hokken-Koelega AC. ACAN gene mutations in short children born SGA and response to Growth Hormone treatment. *J Clin Endocrinol Metab*, 2016: jc20162941.
23. Benito-Sanz S, Barroso E, Heine-Suner D, *et al.* Clinical and molecular evaluation of SHOX/PAR1 duplications in Leri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS). *J Clin Endocrinol Metab*, 2011; **96**: E404–12.
24. Hirschfeldova K, Florianova M, Kebrdlova V, Urbanova M, Stekrova J. Detection of SHOX gene aberrations in routine diagnostic practice and evaluation of phenotype scoring form effectiveness. *J Hum Genet*, 2017; **62**: 253–7.
25. Meyer-Bahlburg HF. Short stature: psychological issues. In: Lifshitz F, ed. *Pediatric Endocrinology*. **2**. New York, NY: Marcel Dekker, 1990: 173–96.
26. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the multidimensional scale of perceived social support. *J Pers Assess*, 1990; **55**: 610–17.
27. Underwood LE. The social cost of being short: societal perceptions and biases. *Acta Paediatr Scand Suppl*, 1991; **377**: 3–8.
28. Rekers-Mombarg LT, Busschbach JJ, Massa GG, Dicke J, Wit JM. Quality of life of young adults with idiopathic short stature: effect of growth hormone treatment. Dutch Growth Hormone Working Group. *Acta Paediatr*, 1998; **87**: 865–70.
29. Silvers JB, Marinova D, Mercer MB, Connors A, Cuttler L. A national study of physician recommendations to initiate and discontinue growth hormone for short stature. *Pediatrics*, 2010; **126**: 468–76.
30. Voss LD, Bailey BJ, Mulligan J, Wilkin TJ, Betts PR. Short stature and school performance—the Wessex Growth Study. *Acta Paediatr Scand Suppl*, 1991; **377**: 29–31; discussion 2.
31. Sandberg DE, Bukowski WM, Fung CM, Noll RB. Height and social adjustment: are extremes a cause for concern and action? *Pediatrics*, 2004; **114**: 744–50.

32. Ekwo E, Gosselink C, Roizen N, Brazdziunas D. The effect of height on family income. *Am J Hum Biol*, 1991; **3**: 181–8.
33. Judge TA, Cable DM. The effect of physical height on workplace success and income: preliminary test of a theoretical model. *J Appl Psychol*, 2004; **89**: 428–41.
34. Case A, Paxson C. Height, health, and cognitive function at older ages. *Am Econ Rev*, 2008; **98**: 463–7.
35. Quitmann JH, Bullinger M, Sommer R, Rohenkohl AC, Bernardino Da Silva NM. Associations between psychological problems and quality of life in pediatric short stature from patients' and parents' perspectives. *PLoS One*, 2016; **11**: e0153953.
36. Downie AB, Mulligan J, Stratford RJ, Betts PR, Voss LD. Are short normal children at a disadvantage? The Wessex growth study. *BMJ*, 1997; **314**: 97–100.
37. Ross JL, Sandberg DE, Rose SR, *et al*. Psychological adaptation in children with idiopathic short stature treated with growth hormone or placebo. *J Clin Endocrinol Metab*, 2004; **89**: 4873–8.
38. Albertsson-Wikland K, Karlberg J. Natural growth in children born small for gestational age with and without catch-up growth. *Acta Paediatr Suppl*, 1994; **399**: 64–70; discussion 1.
39. Spadoni GL, Cianfarani S. Bone age assessment in the workup of children with endocrine disorders. *Horm Res Paediatr*, 2010; **73**: 2–5.
40. Ranke MB, Grauer ML, Kistner K, Blum WF, Wollmann HA. Spontaneous adult height in idiopathic short stature. *Horm Res*, 1995; **44**: 152–7.
41. Coutant R, Dorr HG, Gleeson H, Argente J. Diagnosis of endocrine disease: limitations of the IGF1 generation test in children with short stature. *Eur J Endocrinol*, 2012; **166**: 351–7.
42. Shapiro L, Chatterjee S, Ramadan DG, *et al*. Whole-exome sequencing gives additional benefits compared to candidate gene sequencing in the molecular diagnosis of children with growth hormone or IGF-1 insensitivity. *Eur J Endocrinol*, 2017; **177**: 485–501.
43. Guo MH, Shen Y, Walvoord EC, *et al*. Whole exome sequencing to identify genetic causes of short stature. *Horm Res Paediatr*, 2014; **82**: 44–52.
44. Hauer NN, Popp B, Schoeller E, *et al*. Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. *Genet Med*, 2018; **20**: 630–8.
45. Dauber A, Rosenfeld RG, Hirschhorn JN. Genetic evaluation of short stature. *J Clin Endocrinol Metab*, 2014; **99**: 3080–92.
46. Van Vliet G, Styne DM, Kaplan SL, Grumbach MM. Growth hormone treatment for short stature. *N Engl J Med*, 1983; **309**: 1016–22.
47. Ranke MB, Lindberg A. The basis for optimising growth with growth hormone usage in children with idiopathic short stature: analysis of data from KIGS (Pfizer International Growth Study Database). *Horm Res Paediatr*, 2011; **76** Suppl 3: 48–50.
48. Deodati A, Cianfarani S. Impact of growth hormone therapy on adult height of children with idiopathic short stature: systematic review. *BMJ*, 2011; **342**: c7157.
49. van der Steen M, Hokken-Koelega AC. Growth and metabolism in children born small for gestational age. *Endocrinol Metab Clin North Am*, 2016; **45**: 283–94.
50. Allen DB. Safety of growth hormone treatment of children with idiopathic short stature: the US experience. *Horm Res Paediatr*, 2011; **76** Suppl 3: 45–7.
51. Bell J, Parker KL, Swinford RD, Hoffman AR, Maneatis T, Lippe B. Long-term safety of recombinant human growth hormone in children. *J Clin Endocrinol Metab*, 2010; **95**: 167–77.
52. Wilton P. Adverse events reported in KIGS. In: Ranke MB, Price DA, Reiter EO, eds. *Growth Hormone Therapy in Pediatrics: 20 Years of KIGS*. Basel: Karger; 2007: 432–41.
53. Allen DB, Backeljauw P, Bidlingmaier M, *et al*. GH safety workshop position paper: a critical appraisal of recombinant human GH therapy in children and adults. *Eur J Endocrinol*, 2016; **174**: P1–9.
54. Swerdlow AJ, Cooke R, Albertsson-Wikland K, *et al*. Description of the SAGhE cohort: a large European study of mortality and cancer incidence risks after childhood treatment with recombinant growth hormone. *Horm Res Paediatr*, 2015; **84**: 172–83.
55. Swerdlow AJ, Cooke R, Beckers D, *et al*. Cancer risks in patients treated with growth hormone in childhood: the SAGhE European Cohort Study. *J Clin Endocrinol Metab*, 2017; **102**: 1661–72.
56. Carel JC, Ecosse E, Landier F, *et al*. Long-term mortality after recombinant growth hormone treatment for isolated growth hormone deficiency or childhood short stature: preliminary report of the French SAGhE study. *J Clin Endocrinol Metab*, 2012; **97**: 416–25.
57. Poidvin A, Touze E, Ecosse E, *et al*. Growth hormone treatment for childhood short stature and risk of stroke in early adulthood. *Neurology*, 2014; **83**: 780–6.
58. Savendahl L, Maes M, Albertsson-Wikland K, *et al*. Long-term mortality and causes of death in isolated GHD, ISS, and SGA patients treated with recombinant growth hormone during childhood in Belgium, the Netherlands, and Sweden: preliminary report of 3 countries participating in the EU SAGhE study. *J Clin Endocrinol Metab*, 2012; **97**: E213–17.
59. Deodati A, Ferroli BB, Cianfarani S. Association between growth hormone therapy and mortality, cancer and cardiovascular risk: systematic review and meta-analysis. *Growth Horm IGF Res*, 2014; **24**: 105–11.
60. Stochholm K, Kiess W. Long-term safety of growth hormone-A combined registry analysis. *Clin Endocrinol (Oxf)*, 2018; **88**: 515–28.
61. Chernausek SD, Backeljauw PF, Frane J, Kuntze J, Underwood LE. Long-term treatment with recombinant insulin-like growth factor (IGF)-I in children with severe IGF-I deficiency due to growth hormone insensitivity. *J Clin Endocrinol Metab*, 2007; **92**: 902–10.
62. Backeljauw PF, Miller BS, Dutailly P, *et al*. Recombinant human growth hormone plus recombinant human insulin-like growth factor-1 coadministration therapy in short children with low insulin-like growth factor-1 and growth hormone sufficiency: results from a randomized, multicenter, open-label, parallel-group, active treatment-controlled trial. *Horm Res Paediatr*, 2015; **83**: 268–79.
63. Blair JC, Camacho-Hubner C, Miraki Moud F, *et al*. Standard and low-dose IGF-I generation tests and spontaneous growth hormone secretion in children with idiopathic short stature. *Clin Endocrinol (Oxf)*, 2004; **60**: 163–8; discussion 1–2.
64. Grimberg A, DiVall SA, Polychronakos C, *et al*. Guidelines for growth hormone and insulin-like growth factor-i treatment in children and adolescents: growth hormone deficiency, idiopathic short stature, and primary insulin-like growth factor-I deficiency. *Horm Res Paediatr*, 2016; **86**: 361–97.
65. De Luca F, Argente J, Cavallo L, *et al*. Management of puberty in constitutional delay of growth and puberty. *J Pediatr Endocrinol Metab*, 2001; **14** Suppl 2: 953–7.
66. Kelly BP, Paterson WF, Donaldson MD. Final height outcome and value of height prediction in boys with constitutional delay

in growth and adolescence treated with intramuscular testosterone 125 mg per month for 3 months. *Clin Endocrinol (Oxf)*, 2003; **58**: 267–72.
67. Tse WY, Buyukgebiz A, Hindmarsh PC, Stanhope R, Preece MA, Brook CG. Long-term outcome of oxandrolone treatment in boys with constitutional delay of growth and puberty. *J Pediatr*, 1990; **117**: 588–91.
68. Wehkalampi K, Pakkila K, Laine T, Dunkel L. Adult height in girls with delayed pubertal growth. *Horm Res Paediatr*, 2011; **76**: 130–5.
69. Wickman S, Sipila I, Ankarberg-Lindgren C, Norjavaara E, Dunkel L. A specific aromatase inhibitor and potential increase in adult height in boys with delayed puberty: a randomised controlled trial. *Lancet*, 2001; **357**: 1743–8.
70. Hero M, Norjavaara E, Dunkel L. Inhibition of estrogen biosynthesis with a potent aromatase inhibitor increases predicted adult height in boys with idiopathic short stature: a randomized controlled trial. *J Clin Endocrinol Metab*, 2005; **90**: 6396–402.
71. Mauras N, Ross JL, Gagliardi P, *et al.* Randomized trial of aromatase inhibitors, growth hormone, or combination in pubertal boys with idiopathic, short stature. *J Clin Endocrinol Metab*, 2016; **101**: 4984–93.
72. Yanovski JA, Rose SR, Municchi G, *et al.* Treatment with a luteinizing hormone-releasing hormone agonist in adolescents with short stature. *N Engl J Med*, 2003; **348**: 908–17.
73. van Gool SA, Kamp GA, Visser-van Balen H, *et al.* Final height outcome after three years of growth hormone and gonadotropin-releasing hormone agonist treatment in short adolescents with relatively early puberty. *J Clin Endocrinol Metab*, 2007; **92**: 1402–8.
74. Benabbad I, Rosilio M, Tauber M, *et al.* Growth hormone in combination with leuprorelin in pubertal children with idiopathic short stature. *Endocr Connect*, 2018; **7**: 708–18.
75. Wheeler PG, Bresnahan K, Shephard BA, Lau J, Balk EM. Short stature and functional impairment: a systematic review. *Arch Pediatr Adolesc Med*, 2004; **158**: 236–43.

7.1.5 Tall Stature

Lars Sävendahl and Emelie Benyi

Introduction

Tall stature can be defined as a height above the 97th percentile (corresponding to a standard deviation score (SDS) of +1.88). Extreme tall stature is commonly defined as a height exceeding +3 SDS. Even though tall stature is constitutional in most cases, growth disorders causing excessive growth are important to rule out since they can be associated with other health implications. There are several physical and developmental signs to look out for which could be associated with a growth disorder. These can be either primary or secondary, a division that is further detailed next.

Growth and Growth Plate Regulation

Many signalling pathways, local and systemic, have been demonstrated to regulate longitudinal bone growth [1, 2]. Systemic factors include growth hormone (GH), insulin-like growth factors (IGFs), insulin, thyroid hormone, leptin, sex steroids, vitamin D, and glucocorticoids. There are also local regulators in the growth plate such as fibroblast growth factors (FGFs), and bone morphogenetic proteins (BMPs). Dysregulation of these signalling pathways may cause excessive growth and tall stature.

Growth Hormone and IGF-I

Growth hormone (GH) is a single polypeptide chain produced by the pituitary gland. Secretion is regulated by the hypothalamic peptides, growth hormone-releasing hormone (GHRH) and the inhibitory hormone somatostatin. It is also stimulated by ghrelin produced in the stomach while IGF-1 exerts negative feedback control. Growth hormone stimulates longitudinal bone growth both via direct stimulation of the growth plate, and indirectly by stimulating the hepatic production of IGF-1 [3]. IGF-1 is also ubiquitously expressed in many other tissues such as fat, muscle and in the growth plate. It has systemic effects on amino acid uptake, bone mineral density, muscle mass, and lipolysis [4]. In addition, it has local effects in the growth plate under the influence of GH stimulating both proliferation and hypertrophy of chondrocytes as well as ossification by affecting the osteoblasts [5, 6]. In a growing individual, a GH secreting pituitary adenoma will cause accelerated growth and tall stature which is referred to as pituitary gigantism.

Sex Steroids

Sex steroids including oestrogen and testosterone are mainly produced by the gonads. Oestrogens play an important role for growth in both girls and boys. In girls, oestrogens are mainly produced in the ovaries. In boys, the main production occurs in peripheral tissues through aromatization of androgens. Oestrogens affect growth by reducing IGF-1 mediated negative feedback which regulates the effect of GH and its secretion [7]. Oestrogens stimulate growth during puberty but are also the factor that finally leads to growth plate fusion at the end of puberty in both genders [8]. Testosterone increases the effect of GH on IGF-1 secretion [7]. Androgen receptors are expressed in the human growth plate but any local direct effects of androgens on longitudinal bone growth have not been established [9].

Oestrogens act by binding two different nuclear receptors in the growth plate, oestrogen receptor (ER)α and ERβ, and a more recently discovered G protein-coupled oestrogen receptor 1 (GPER1) [10, 11]. ERα and ERβ are both expressed throughout the different zones of the growth plate [12]. The abundance of both receptors increases as the cells differentiate [13]. Oestrogens stimulate osteoblasts and inhibit osteoclasts but the exact mechanisms through which they promote bone growth are not clear [14]. Open growth plates at 28 years of age were observed in a 204 cm tall male who had an ERα inactivating mutation [15]. Patients with aromatase deficiency exhibit similar phenotypes with a lack of pubertal growth spurt and growth plates which remain open into adulthood leading to tall stature. Patients with aromatase deficiency respond to oestrogen treatment which improves bone mineralization and closes the growth plates. Such therapy has no effect in an ERα-mutated patient suggesting that ERα has an important role for growth plate fusion in humans.

Local Regulators of Growth in the Growth Plate

One type of local regulator in the growth plate are fibroblast growth factors (FGFs) which have important roles at every stage of endochondral bone formation. Fibroblast growth factor-1 and FGF-3 are growth-inhibiting whereas FGF-2 is growth promoting [2, 16]. Bone morphogenetic proteins (BMPs) are also present throughout the endochondral ossification process and seem to have antagonizing effects to FGFs in some of the stages [17]. Parathyroid hormone-related peptide (PTHrP) has a stabilizing role in the growth plate by maintaining chondrocytes in a proliferative stage and also inhibiting hypertrophy [18]. C-type natriuretic peptide (CNP) is a local positive regulator of growth [19]. Activating mutations or overexpression of CNP can cause tall stature [20, 21].

Diagnosis of Tall Stature

Even though tall stature is constitutional in the vast majority of cases, growth disorders causing excessive growth are important to rule out in patients seeking medical attention for tall stature since they can be associated with other health issues. There are several physical and developmental signs to look out for which could be associated with a growth disorder. To give some examples, an increased birth length may be a sign of Beckwith-Wiedemann syndrome. A large head circumference, on the other hand, may indicate an underlying syndrome such as Sotos syndrome or Weaver syndrome. Developmental problems are common in individuals with homocystinuria and several syndromes including Klinefelter, Triple X, Fragile X, Loeys–Dietz, Sotos, and Weaver syndrome. There is a common division into primary and secondary growth disorders which are summarized in Table 7.1.5.1. Primary growth disorders affect the growth process itself and may have a prenatal onset. Secondary growth disorders are caused by external factors and affect growth indirectly.

Constitutional Tall Stature

Constitutional tall stature is also known as familial tall stature. It is genetic, hence dependent on the height of the mother and father of

Table 7.1.5.1 Overview of growth disorders associated with tall stature

	Major clinical findings (in addition to tall stature)	Diagnostic testing
Primary growth disorders		
Marfan syndrome	Long extremities, aorta dilatation, scoliosis, flexible joints, ectopia lensis	Ghent criteria, genetic testing for FBN-1 mutation, echocardiography, and ophthalmology evaluation
Loeys–Dietz syndrome	Long extremities, aorta dilatation, scoliosis, flexible joints, cleft palate	Genetic testing for mutations of the TGFß signalling pathway
Homocystinuria	Long extremities, scoliosis, ectopia lensis, developmental delay	Blood test for homocysteine concentration
Multiple endocrine neoplasia 2B	Long face, scoliosis, joint laxity, alacrima, mucosal neuromas on the tongue, lips, oral cavity, and conjunctivae	Genetic testing for mutations in the *RET* gene
Klinefelter syndrome	Small testes, muscle weakness, cognitive disabilities, infertility, gynaecomastia	Chromosome karyotyping (XXY or variants with ≥2 X chromosomes)
Fragile X	Protruding forehead and large ears, autism, developmental delay	Genetic testing for FMR1 mutation
Sotos syndrome	Large skull, cognitive disabilities	Genetic testing for NSD1 mutation
Weavers syndrome	Dysmorphic facial features, learning difficulties	Genetic testing for EZH2 mutation
Beckwith–Wiedemann syndrome	Neonatal hypoglycaemia, macroglossia, hemihyperplasia, embryonic tumours	Clinical assessment and genetic testing
Secondary growth disorders		
Pituitary gigantism	Frontal bossing, prominent jaw, large hands and feet, delayed puberty, muscle weakness	Blood test for basal GH and IGF-1; GH suppression test
Precocious puberty (central and peripheral)	Early onset of puberty and pubertal growth spurt (tall stature in childhood)	Blood test for LH, FSH, oestrogen and testosterone, bone age
Familial glucocorticoid deficiency	Failure to thrive, hypoglycaemia	Blood test for plasma ACTH and cortisol. Genetic testing

the child. Height is highly heritable in a polygenic manner. It has been estimated in a sibling study that the heritability of height is 80% [22]. Through genome-wide association studies (GWAS) over 400 loci associated with height have been identified, explaining approximately 20% of the phenotypic and 25% of the genetic height variation [23, 24]. Elevated levels of GH and IGF-1 have been observed in studies of individuals with constitutional tall stature but the results are inconclusive [25–28]. One study showed higher peak GH concentrations in tall children when compared to those of short and normal height [25]. Another study showed higher IGF-2 levels in tall individuals when compared to controls whereas no difference in IGF-1 levels was found [26]. IGF-1, IGF-2, and GH levels were compared between tall and normal/short army recruits (six in each group) and even though a tendency towards higher levels was seen in taller individuals, differences were not statistically significant which can however relate to a small number of study participants [28].

Constitutionally tall children often grow more rapidly in the first 4 years of life, after which growth drops to more normal rates [29]. Dickerman *et al.* studied the growth patterns from birth to 9 years of age in 65 children diagnosed with constitutional tall stature. Their mean birth lengths were 52.1 and 53.5 cm in girls and boys, respectively, corresponding to the 75th percentile. At 4 years of age, their heights had increased to 111.7 and 113.0 cm, respectively, corresponding to 2.5 SD above the mean. Between four and nine years, the growth curves were parallel with the 50th percentile, but 2.75 SD above on average. The authors suggest that the fact that the birth length does not deviate from normal indicates that factors leading to tall stature are postnatal.

Primary Growth Disorders

Marfan Syndrome

Marfan syndrome is a connective tissue disorder caused by a mutation in the fibrillin-1 gene (FBN1) located on chromosome 15 (15q21.1). This gene encodes a glycoprotein called fibrillin-1 which is essential for connective tissue formation. Marfan syndrome affects approximately two to three per 10 000 individuals [30, 31]. It is usually inherited in an autosomal dominant pattern although approximately 25% of affected individuals have a de novo mutation. The most serious abnormality associated with this syndrome is dilatation of the aorta which can lead to aortic dissection and rupture, a fatal condition. Other clinical features include ectopia lentis, scoliosis, crowded teeth as well as flexible and painful joints. Children with Marfan syndrome may grow extremely tall with especially long arms, legs, fingers, and toes. In 1986, an expert panel developed a set of diagnostic criteria, called the Berlin nosology, aimed to facilitate accurate diagnosis [32]. This was later revised due to overdiagnosis and the commonly used Ghent criteria were developed [33], with further revisions in 2010 [34]. These diagnostic criteria are mainly focused on the family history of Marfan, ectopia lentis, aortic dilatation, and FBN-1 mutations. Additional criteria include body disproportions and musculoskeletal abnormalities.

Loeys–Dietz Syndrome

Loeys–Dietz syndrome is an autosomal dominant genetic disorder affecting the collagen production in the connective tissue caused by a mutation of the transforming growth beta signalling pathway. It has similarities to Marfan syndrome such as an increased risk of aorta dissection due to an aneurysm. Other clinical features beside overgrowth are hypertelorism and cleft palate [35].

Homocystinuria

Homocystinuria is an autosomal recessive disorder caused by a mutation in the cystathionine-beta-synthase (CBS) gene on chromosome 21 (21q22.3) affecting approximately one in 6000 newborn children [36]. Individuals with homocystinuria have a deficient elimination of homocysteine and its metabolites. There are two types, B_6-responsive and B_6-nonresponsive homocystinuria, the first usually being milder than the second. Homocystinuria affects connective tissues in the body, including fibrillin. Affected individuals may exhibit clinical features similar to those in Marfan syndrome including tall stature with long legs and arms, scoliosis, and ectopia lentis. They might also suffer developmental delay (especially the B_6-nonresponsive type) and thromboembolic complications which can lead to an early death [37].

Multiple Endocrine Neoplasia Type 2B

Another differential diagnosis in patients with Marfanoid habitus is multiple endocrine neoplasia type 2B (MEN2B) which is a rare autosomal dominant disorder caused by mutations in the *RET* gene. Clinical manifestations include mucosal neuromas on the tongue, lips, oral cavity, and conjunctivae which are white-yellow papules a few millimetre in size and may be present at birth and can usually be found by age 10 years [38]. Gastrointestinal complaints caused by diffuse gastrointestinal ganglioneuromatosis may be present from infancy or early childhood. Affected individuals are at high risk of developing medullary thyroid cancer at a very young age and have an increased risk of pheochromocytoma.

Klinefelter Syndrome

Klinefelter syndrome, also called 47,XXY, affects one to two boys per 1000 [39]. Due to an extra X-chromosome, these patients have excess expression of certain genes. One of these is the short stature homeobox (SHOX) gene, which has been shown to influence the phenotype seen in Klinefelter syndrome [40, 41]. Clinical features include tall stature, small external genitalia, muscle weakness, cognitive disabilities, behavioural problems, infertility, and gynecomastia. Affected boys also have a higher risk of developing diabetes mellitus, cancer, and cardiovascular disease. Klinefelter syndrome is often diagnosed late, when affected men fail to conceive [42].

Fragile X

Fragile X affects approximately one in 5000 males and one in 10 000 females and is caused by a mutation in the fragile X mental retardation 1 (FMR1) gene located on the X-chromosome. Symptoms include tall stature, intellectual disability, developmental delay, autism and a characteristic appearance with a protruding forehead, and large ears [43].

Sotos Syndrome

Sotos syndrome affects one in 10 000 to 14 000 newborns and is caused by mutations in the nuclear receptor binding SET domain protein 1 (NSD1) gene in 90% of the cases. The skull is large and affected individuals have an increased height velocity during the first 4 years of life and an increased adult height. Cognitive disabilities are common [44].

Nevo Syndrome

Nevo syndrome is caused by a NSD1 deletion and is inherited in an autosomal recessive pattern. It shares clinical features with the more common Sotos syndrome described earlier [45].

Weaver Syndrome

Mutations in the oncogene EZH2 causes Weaver syndrome which is associated with tall stature, dysmorphic facial features, and learning difficulties [46]. Due to the similar clinical picture it is often mistaken for Sotos syndrome.

Trisomy X

Trisomy X is a sex chromosome aneuploidy affecting approximately 1 in 1000 girls. It is associated with tall stature as well as behavioural problems and learning difficulties [47].

Simpson–Golabi–Behmel Syndrome

The overgrowth condition Simpson–Golabi–Behmel syndrome (SGBS) is caused by a mutation in the X-linked SGBS gene. In addition to tall stature affected individuals may exhibit supernumerary nipples, muscular hypotonia and an increased risk of tumours in infancy has been seen [48].

Beckwith–Wiedemann Syndrome

Most affected individuals with Beckwith–Wiedemann syndrome (BWS) have an imprinting disorder of one or two gene clusters that contain the *IGF2* and *H19* genes, under control of chromosome 15 imprinting centre 1 (IC1), and the *CDKN1C* and *KCNQ1OT1* genes, under control of imprinting centre 2 (IC2). Approximately 65% of children with BWS have a birth weight above the 90th percentile [49]. They are tall in childhood and mean adult height has been reported to be 1.8 ± 1.2 SDS above normal [50]. Congenital anomalies such as macroglossia, ear creases or pits, naevus flammeus, and abdominal wall defects are characteristic of BWS as is hemihyperplasia [49]. Affected individuals may develop hypoglycaemia due to hyperinsulinism in the neonatal period and also have an increased risk of developing embryonic tumours, especially Wilms tumour.

PTEN Hamartoma Tumour Syndromes

This group of growth disorders includes Cowden disease and Bannayan–Riley–Ruvalcaba syndrome and affects approximately one in 200 000–250 000 [51]. They are caused by a mutation of the PTEN tumour suppressor genes and are inherited in an autosomal dominant pattern.

Perlman Syndrome

Perlman syndrome is a rare autosomal recessive genetic disorder caused by a mutation in the *DIS3L2* gene. It is characterized by overgrowth, dysmorphic facial features, renal dysplasia, and an increased risk of developing Wilms tumour. It has a poor prognosis with a high neonatal mortality.

Secondary Growth Disorders

Pituitary Gigantism

Pituitary gigantism causes abnormal growth in childhood due to excessive secretion of GH. The most common cause is a benign pituitary tumour but the incidence is unknown [52]. In some cases, underlying causes are Carney complex, McCune–Albright syndrome (MAS), multiple endocrine neoplasia type 1 (MEN-1), and neurofibromatosis [53, 54]. Pituitary gigantism causes bones as well as muscles and internal organs to grow excessively. Other physical features associated include frontal bossing, a prominent jaw, large hands, and feet. It can also cause delayed puberty, headache, muscle weakness, double vision, and sleeping problems. Different treatment methods are available including transsphenoidal surgery, somatostatin analogues, and GH antagonists [55].

Obesity

Obese children often experience tall stature due to early adrenarche, early puberty, and advanced skeletal maturation. The mechanism is thought to be related to hyperinsulinemia. Insulin binds to the IGF-1 receptor and increases height velocity. It also increases free IGF-1 in the circulation [55]. However, adult height is usually not affected due to a more blunted pubertal growth spurt and earlier growth plate closure [56].

Hyperthyroidism

Thyrotoxicosis in childhood may cause an acceleration of skeletal maturation and linear growth. It is unclear whether this acceleration affects the final height or if it is compensated by an accelerated bone maturation [57, 58].

Precocious Puberty

Children are considered to experience premature puberty if onset is before 8 years of age for girls and 9 years for boys [59]. Precocious puberty affects approximately one in 500 girls and one in 2000 boys [60]. These girls and boys are tall during childhood, but due to the advancement in skeletal maturation and early growth plate fusion, adult height is usually unaffected or if left untreated sometimes can be reduced [39, 55].

Familial Glucocorticoid Deficiency

Familial glucocorticoid deficiency (FGD) is characterized by a failure of the adrenal cortex to produce glucocorticoids in spite of normal levels of ACTH [61]. There are different types of this condition caused by different mutations. FGD type 1 is caused by a mutation in the ACTH receptor. Mutations in accessory or regulatory proteins can give the same clinical picture with failure to thrive and hypoglycaemia in early childhood and later an association with tall stature where the mechanism is unclear [62].

Aromatase Deficiency

Aromatase deficiency is a rare autosomal recessive disease caused by a mutation of the gene *CPY19* affecting the aromatization of androgens into oestrogens. This leads to delayed skeletal maturation and open growth plates even after adolescence resulting in extreme tallness unless the patient is treated with oestradiol. Affected girls are born with ambiguous genitalia [63].

Psychosocial Aspects of Being Tall

The most common complaint with tall stature is psychosocial problems. Tall individuals may experience social difficulties at school and parents are sometimes concerned of their tall child's ability to

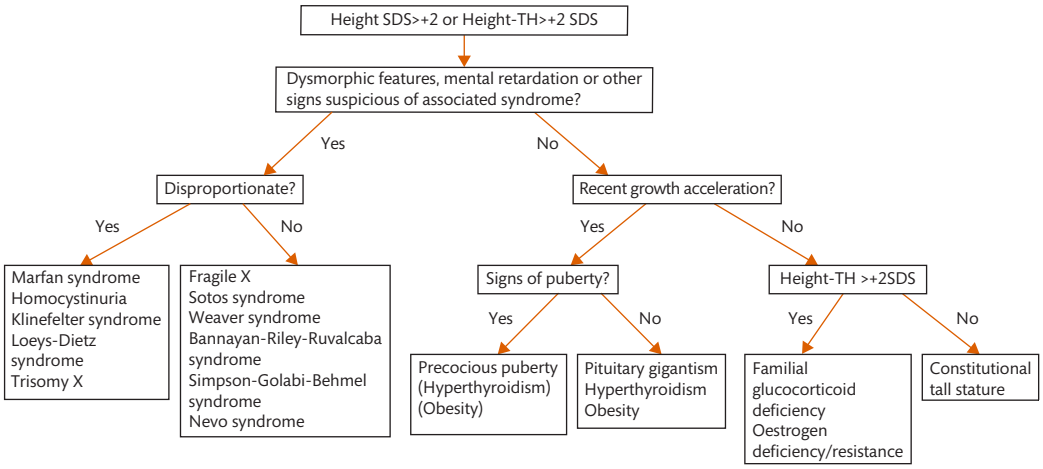


Figure 7.1.5.1 Flow chart for the assessment of a tall child.

find a partner. In an Australian cohort of 650 tall women (approximately one third were previously treated with high-dose sex steroids to reduce their heights), no difference in mental or physical health was found between treated and untreated individuals. However, both these groups had a significantly higher risk of developing major depression compared with population-based studies [64].

In a public school in the United States, social outcomes in relation to height were studied in 956 pupils in grade 6–12 in order to establish whether or not there were grounds for intervention in extremes of height. They found that there were minimal effects from height on social behaviour and popularity [65].

There are also studies reporting positive traits associated with tall stature. Higher income in tall individuals has been reported [66]. Another study in which the 45 210 study participants were divided into fifths according to their heights reported that the shortest 20% had a lower mood than the taller participants except for those financially disadvantaged. In this group the association between height and mood was surprisingly reversed [67].

A Swedish study of the risk of suicide in association with fetal and childhood growth reported that short adults are more prone to commit suicide and tall adult stature is protective [68].

Assessment of the Tall Child

The assessment of tall stature can be facilitated by using a flow chart (Figure 7.1.5.1). A thorough history and physical examination is essential as summarized in Box 7.1.5.1. It is important to assess the child's height against standard growth charts and to decide whether the child has normal body proportions or not. Where disproportionate growth occurs, it is more likely that there is an underlying genetic syndrome. Furthermore, it is necessary to establish whether one or even both parents are tall, and if so whether there is a dominant growth abnormality such as Marfan syndrome. Also, children with heights close to average might need investigation if their heights are well above their mid-parental target height or if they experience sudden growth acceleration. Some conditions can cause excessive growth during childhood with a normalization of height before adult age, such as obesity and precocious puberty.

Auxological measurements should include total height as well as sitting height and arm span. Sitting height percentage is usually calculated as the proportion of the upper body segment compared to total height (including the lower extremities). Patients with Klinefelter or Marfan syndrome usually have disproportionally long legs and arms. Even individuals with constitutional tall stature often have relatively long legs with a lower sitting height percentage than individuals of average height. Body mass index (BMI) should be calculated and pubertal staging performed including measurement of testicular volumes in boys. Bone age can be determined from an X-ray image of the left hand and wrist. This should not be done until the patient has entered puberty as final height predictions based on bone age are unreliable at prepubertal ages. Box 7.1.5.1 summarizes suggested initial investigations at the primary care level.

Of vital importance is to exclude potentially serious conditions which might require treatment, such as Marfan syndrome. If this is suspected, the patient should be referred for echocardiography to assess for possible aorta dilatation which can lead to aortic rupture. In some cases, clinical geneticists are consulted and specific genetic

Box 7.1.5.1 History and examination in assessment of tall children

History

- Pregnancy and birth information including birth weight, length, head circumference
- Early development (speech, motor skills, social development)
- Cognition and learning difficulties
- Family history including pubertal timing, growth, and heights, cardiovascular- and other disorders
- Medical history
- Nutrition

Physical examination and tests

- Auxology (plotted onto a growth chart) including height, weight, head circumference, sitting height, leg length, arm span
- Dysmorphic signs
- Pubertal stage
- Signs of thyroid disease
- Musculoskeletal abnormalities including laxity in joints, spinal deformation
- Neurological examination
- Cardiovascular examination
- Ophthalmology examination/consultation
- Bone age X-ray (left hand and wrist): only of value for adult height prediction once the patient has entered puberty
- Lab tests or genetic karyotyping if suspicion of specific disorders (see Table 7.1.5.1)

Box 7.1.5.2 Referral to specialist

Consider referral if:

- Marked height (but not weight) acceleration after 1 year of age
- Tall stature atypical for the family (height >2 SD over SD for target height)
- Patient/parental concern of extreme adult height

Information to include in referral:

Patient history

- Who initiated investigation?
- Growth chart including height and weight from birth until present
- Other health issues including any back or joint problems, previous lens dislocation
- Time of pubertal onset and menarche (girls)
- Any school difficulties

Hereditary factors

- Sudden deaths in the family
- Parental heights and timing of puberty

Physical examination

- Pubertal stage
- Testicle volumes (boys)

testing requested. Karyotyping is performed when, for example Klinefelter syndrome is suspected. Some diagnoses can be confirmed with blood tests, for example pituitary gigantism (GH and IGF-1) and homocystinuria (plasma homocysteine). Box 7.1.5.2 indicates when referral to a paediatric endocrinologist should be considered.

In the clear majority of cases, tall children will be diagnosed with constitutional tall stature. Reassurance is often sufficient in the management of constitutionally tall patients and their families. In rare cases, an intervention might be called for. Following an algorithm may facilitate the management of a tall child being considered for surgical treatment to reduce further growth (Figure 7.1.5.2).

Height Prediction in Tall Girls and Boys

As an individual's height is 80% heritable, calculations have been proposed to predict adult height based on parental heights (i.e. 'target height', TH). However, this method has a large margin of error and more reliable is a prediction based on skeletal maturation assessed by hand and wrist X-ray. The so called bone age can be determined with the atlas of Greulich and Pyle [69]. Based on the bone age, the remaining growth is then estimated with methods such as the tables of Bayley and Pinneau [70] or Tanner–Whitehouse [71]. De Waal *et al.* evaluated different methods of adult height prediction in tall individuals and concluded that the precision was clinically acceptable in girls but not in boys. In both tall girls and tall boys, the tables of Bayley–Pinneau overestimates while Tanner–Whitehouse underestimates final height. The absolute error was on average approximately 2 cm in girls and 3 cm in boys for each of these methods [72]. There are also automated options for bone age determination from X-ray images including BoneXpert (Visiana, Denmark) [73]. This software has been found to be reliable in several different patient groups, resulting in predictions similar to those made by experienced clinicians [74–76]. However, it has to our knowledge not been studied specifically in tall individuals.

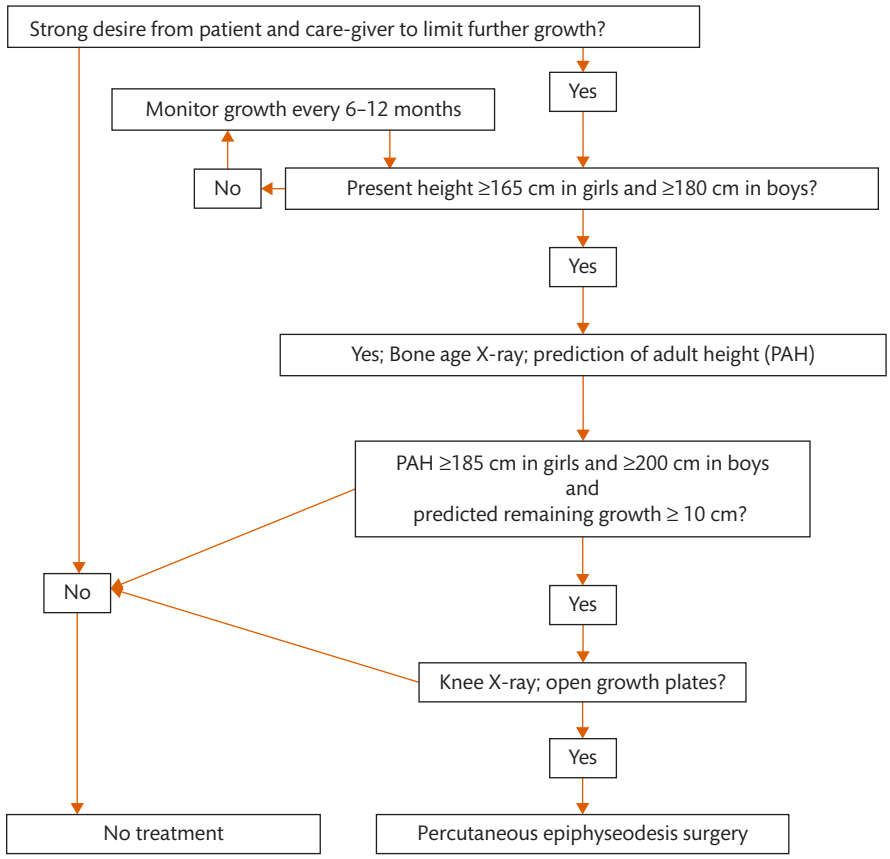


Figure 7.1.5.2 Flow chart for monitoring a tall child considered for surgical treatment.

Growth-Reducing Therapy

The use of growth reduction therapies is declining, possibly due to more acceptance of tall stature in the society and fear of side effects [77]. Historically, sex steroids have been the most common treatment method to reduce growth in tall girls and boys. Other therapies which have not reached the same popularity are somatostatin analogues (inhibits GH release from the pituitary gland) and bromocriptine (dopamine agonist) [78–82]. Surgical options have been introduced in the last decades which may replace medical methods of growth reduction [83–85].

High-Dose Sex Steroid Therapy in Girls

Oestrogens cause growth plate fusion during normal pubertal development. Sex steroid therapies to reduce growth in tall children were developed in the 1950s. It has been demonstrated that administration of high-dose oestradiol accelerates skeletal maturation which causes premature fusion and reduced adult height [86–88]. The daily doses of ethinyl oestradiol have gradually decreased from approximately 500 to 100 μg during the following decades, in some with the addition of medroxyprogesterone acetate.

Treatment efficacy varies greatly in literature which is partly explained by difficulties in final height prediction in tall individuals. A study of 56 treated girls showed a treatment effect of up to 6 cm height reduction [89]. This study included a control group to estimate the prediction error and found it to be 1.2 cm. Drop *et al.* reported 1.1–2.4 cm reductions in adult height compared to prediction in high-dose oestrogen treated girls depending on which method that was applied to predict adult height [90]. Another study showed a mean growth reduction in 159 treated women of between 1.7 and 4.2 cm depending on prediction method. They also found a linear relationship between bone age at start of treatment and the degree of effect. Treatment needed to be started before 14 years of age to have any effect at all [72]. In 279 tall girls treated with diethylstilboestrol or ethinyl oestradiol, their adult heights were reduced by 2.5 cm on average and the effect decreased by one cm per year of delay in treatment start [91]. A more recent Danish study of 60 girls treated with 17β-oestradiol girls showed a reduction in adult height of 1.6 cm [92].

Reported short-term side effects of high-dose oestrogen treatment include acne, weight gain, nausea, and headache [93–96]. With regards to long-term side effects, infertility in women treated with high-dose oestrogens for tall stature has been reported [97, 98]. The treated women had decreased fecundity, were more likely to have used fertility drugs and had more advanced aging of their ovaries compared to those untreated. There are also indications of possible risk increases in thrombosis and cancer [99, 100].

High-Dose Sex Steroids in Boys

Boys have also been treated with high-dose sex steroids to reduce adult height and the testosterone compounds used were generally long-acting esters such as testosterone propionate, enanthate and decanoate [90]. Testosterone doses of 250–500 mg were usually administered through intramuscular injections every second week and the treatment duration was typically 1.5 years [90, 101].

A final height reduction of 4.4 cm was reported in 33 treated boys [94]. Others reported a final height reduction of 4.8 cm when using the Bayley and Pinneau tables for height prediction but as low as 0.5 cm with the Tanner-Whitehouse prediction methods [72]. Height reductions of approximately seven centimetres were found in another study of 25 treated tall boys [102]. Other important factors affecting the efficacy were timing of initiation and duration of the treatment. Failure in optimizing these conditions appears in some cases to have led to an effect opposite to that desired. Drop *et al.* reported that when treatment was started at a bone age of 14 years or later, final height was increased instead of decreased [90]. Van den Bosch *et al.* reported that an addition of what the authors considered to be a low dose of ethinyl oestradiol (50 μg daily) at the beginning of testosterone therapy in tall boys caused an instant decrease in height velocity [103]. Decker *et al.*, on the other hand, found no additional effect of ethinyl oestradiol in this combination [104].

Reported short-term side effects of high-dose sex steroid therapy in boys include acne, weight gain, gynecomastia, and aggression [93–94]. Male infertility after testosterone treatment has not been shown even though lower serum testosterone levels have been reported [105, 106].

Surgical Methods

Epiphysiodesis is a surgical method to stop growth by destruction of the epiphyseal growth plate. It was first described by Phemister in 1933 for correcting leg length discrepancy [107]. With an open approach a block of bone was removed from the physis in the longer leg and put back in a rotated position, in which bony bridges formed and prevented further bone growth. In the 1980s a less invasive method was developed, so called percutaneous epiphysiodesis [108, 109].

Bilateral epiphysiodesis to reduce height in tall adolescents was first described in 1997 [85]. The Phemister method was used and they reported a height reduction of nine centimetres and no serious side effects in six treated boys. The first report on bilateral percutaneous epiphysiodesis to reduce height was published in 2006 and included 17 treated tall boys [84]. It was performed on the growth plates of the distal femur, proximal tibia, and proximal fibula. In boys, the procedure was found to be safe and reduced final heights by on average 7 cm (range 1.2–13.8 cm). Another study on both girls and boys showed a growth reduction of 4.1 and 6.4 cm, respectively, and no serious side effects were reported [83]. The authors recommend treatment when the patient has at least 15 cm of predicted remaining growth which likely results in a reduction of adult height by at least 5 cm. As an example, a girl with a predicted adult height of 190 cm should be treated before having reached a height of 175 cm and a boy with a predicted adult height of 205 cm should be treated before having reached a height of 190 cm. In recent years, radiofrequency application has been proposed as a new method for epiphysiodesis but has so far only been tested in rabbits [110, 111].

With a surgical approach to reduce height, many of the reported or potential side effects of pharmacological treatments mentioned earlier can be avoided [72, 94, 95, 98]. On the downside, surgery always brings a risk of infection and complications associated with anaesthesia. Also, skeletal deformities including leg length difference are other potential side effects.

Ethical Considerations Regarding Growth-Reducing Therapies

There are important ethical considerations to address before offering any treatment to reduce height. When a child presents for evaluation of tall stature it is important to establish whether it is

the child, the parents, or someone else who are most concerned. Other important questions are: what is the long-term value of the treatment? Is it possible that the psychosocial impact of being extremely tall is pronounced only during adolescent years due to peer pressure and reluctance to stand out, factors for which the importance diminishes with age? Treated individuals are generally healthy adolescents. Few side effects or long-term complications are acceptable under such circumstances. There are no known significant physical problems associated with height *per se* that could be avoided by reducing further growth potential. Sex steroid treatment in girls has been shown to impair future fertility and its use has since diminished. For surgical options, long-term complications are yet to be investigated. In the case of percutaneous epiphysiodesis, however, the procedure has been used within the field of paediatric orthopaedic surgery for many decades to treat leg length discrepancy and is regarded as safe. Also, the acceptance in society of extreme height seems to be increasing and any psychosocial impact of standing out because of being tall may be diminishing.

Summary

Tall stature is usually constitutional but might be caused by a growth disorder. Assessment of children and adolescents seeking medical attention for tall stature must be thorough and exclude any associated condition, such as Marfan syndrome, which might need medical treatment and monitoring. Other disorders associated with tall stature are Klinefelter syndrome, homocystinuria and the rare condition pituitary gigantism where there is an overproduction of growth hormone. In those few cases where an intervention is considered to reduce further growth, it is crucial to establish if it is the patient her- or himself who has a genuine desire to be treated and to carefully weigh up potential gains against risks. Any psychosocial advantage from such treatment has not been established, while the impact of height reduction on mental health has also not been adequately defined. In the past, tall boys and girls have been treated with high dose sex steroids but it has been shown to reduce fertility in girls and the use of this therapy has decreased markedly. New surgical methods are becoming more common including percutaneous epiphysiodesis. With this method the growth plate cartilage around the knee is destroyed surgically and the impact on height is certainly comparable, if not better than hormonal treatment. Overall, however, acceptance of tall stature is improving in society and in future it is likely that intervention for height reduction will rarely be required.

REFERENCES

1. Kronenberg HM. Developmental regulation of the growth plate. *Nature*, 2003; **423**: 332–6.
2. Wang X, Qi H, Wang Q, *et al.* FGFR3/fibroblast growth factor receptor 3 inhibits autophagy through decreasing the ATG12-ATG5 conjugate, leading to the delay of cartilage development in achondroplasia. *Autophagy*, 2015; **11**: 1998–2013.
3. Wu S, Yang W, De Luca F. Insulin-like growth factor-independent effects of growth hormone on growth plate chondrogenesis and longitudinal bone growth. *Endocrinology*, 2015; **156**: 2541–51.
4. Yakar S, Isaksson O. Regulation of skeletal growth and mineral acquisition by the GH/IGF-1 axis: lessons from mouse models. *Growth Horm IGF Res Off J Growth Horm Res Soc Int IGF Res Soc*, 2016; **28**: 26–42.
5. Baxter RC. Characterization of the acid-labile subunit of the growth hormone-dependent insulin-like growth factor binding protein complex. *J Clin Endocrinol Metab*, 1988; **67**: 265–72.
6. Shim KS. Pubertal growth and epiphyseal fusion. *Ann Pediatr Endocrinol Metab*, 2015; **20**: 8–12.
7. Meinhardt UJ, Ho KKY. Modulation of growth hormone action by sex steroids. *Clin Endocrinol (Oxf)*, 2006; **65**: 413–22.
8. Börjesson AE, Windahl SH, Karimian E, *et al.* The role of estrogen receptor-α and its activation function-1 for growth plate closure in female mice. *Am J Physiol Endocrinol Metab*, 2012; **302**: E1381–9.
9. Chagin AS, Vannesjö J, Sävendahl L. Androgen receptor modulation does not affect longitudinal growth of cultured fetal rat metatarsal bones. *Horm Res*, 2009; **71**: 219–27.
10. Thomas P, Pang Y, Filardo EJ, Dong J. Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology*, 2005; **146**: 624–32.
11. Chagin AS, Sävendahl L. GPR30 estrogen receptor expression in the growth plate declines as puberty progresses. *J Clin Endocrinol Metab*, 2007; **92**: 4873–7.
12. Nilsson O, Chrysis D, Pajulo O, *et al.* Localization of estrogen receptors-alpha and -beta and androgen receptor in the human growth plate at different pubertal stages. *J Endocrinol*, 2003; **177**: 319–26.
13. Simm PJ, Russo VC, Werther GA. The effect of selective oestrogen receptor antagonists in an in vitro model of growth plate chondrogenesis. *Endocrine*, 2011; **40**: 27–34.
14. Rochira V, Kara E, Carani C. The endocrine role of estrogens on human male skeleton. *Int J Endocrinol*, 2015; **2015**: 165215.
15. Smith EP, Specker B, Korach KS. Recent experimental and clinical findings in the skeleton associated with loss of estrogen hormone or estrogen receptor activity. *J Steroid Biochem Mol Biol*, 2010; **118**: 264–72.
16. Lazarus JE, Hegde A, Andrade AC, Nilsson O, Baron J. Fibroblast growth factor expression in the postnatal growth plate. *Bone*, 2007; **40**: 577–86.
17. Niswander L, Martin GR. FGF-4 and BMP-2 have opposite effects on limb growth. *Nature*, 1993; **361**: 68–71.
18. Späth S-S, Andrade AC, Chau M, Nilsson O. Local regulation of growth plate cartilage. *Endocr Dev*, 2011; **21**: 12–22.
19. Pejchalova K, Krejci P, Wilcox WR. C-natriuretic peptide: an important regulator of cartilage. *Mol Genet Metab*, 2007; **92**: 210–15.
20. Wang SR, Jacobsen CM, Carmichael H, *et al.* Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature. *Hum Mutat*, 2015; **36**: 474–81.
21. Hannema SE, van Duyvenvoorde HA, Premsler T, *et al.* An activating mutation in the kinase homology domain of the natriuretic peptide receptor-2 causes extremely tall stature without skeletal deformities. *J Clin Endocrinol Metab*, 2013; **98**: E1988–98.
22. Visscher PM, Medland SE, Ferreira MAR, *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*, 2006; **2**: e41.
23. Lango Allen H, Estrada K, Lettre G, *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 2010; **467**: 832–8.

24. Wood AR, Esko T, Yang J, *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*, 2014; **46**: 1173–86.
25. Albertsson-Wikland K, Rosberg S. Analyses of 24-hour growth hormone profiles in children: relation to growth. *J Clin Endocrinol Metab*, 1988; **67**: 493–500.
26. Garrone S, Radetti G, Sidoti M, Bozzola M, Minuto F, Barreca A. Increased insulin-like growth factor (IGF)-II and IGF/IGF-binding protein ratio in prepubertal constitutionally tall children. *J Clin Endocrinol Metab*, 2002; **87**: 5455–60.
27. Tauber M, Pienkowski C, Rochiccioli P. Growth hormone secretion in children and adolescents with familial tall stature. *Eur J Pediatr*, 1994; **153**: 311–16.
28. Stratakis CA, Mastorakos G, Magiakou MA, *et al*. 24-hour secretion of growth hormone (GH), insulin-like growth factors-I and -II (IGF-I, -II), prolactin (PRL) and thyrotropin (TSH) in young adults of normal and tall stature. *Endocr Res*, 1996; **22**: 261–76.
29. Dickerman Z, Loewinger J, Laron Z. The pattern of growth in children with constitutional tall stature from birth to age 9 years. A longitudinal study. *Acta Paediatr Scand*, 1984; **73**: 530–6.
30. Judge DP, Dietz HC. Marfan's syndrome. *Lancet Lond Engl*, 2005; **366**: 1965–76.
31. Groth KA, Hove H, Kyhl K, *et al*. Prevalence, incidence, and age at diagnosis in Marfan Syndrome. *Orphanet J Rare Dis*, 2015; **10**: 153.
32. Beighton P, de Paepe A, Danks D, *et al*. International Nosology of Heritable Disorders of Connective Tissue, Berlin, 1986. *Am J Med Genet*, 1988; **29**: 581–94.
33. De Paepe A, Devereux RB, Dietz HC, Hennekam RC, Pyeritz RE. Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet*, 1996; **62**: 417–26.
34. Loeys BL, Dietz HC, Braverman AC, *et al*. The revised Ghent nosology for the Marfan syndrome. *J Med Genet*, 2010; **47**: 476–85.
35. Van Laer L, Dietz H, Loeys B. Loeys-Dietz syndrome. *Adv Exp Med Biol*, 2014; **802**: 95–105.
36. Refsum H, Fredriksen A, Meyer K, Ueland PM, Kase BF. Birth prevalence of homocystinuria. *J Pediatr*, 2004; **144**: 830–2.
37. Picker JD, Levy HL. Homocystinuria caused by cystathionine beta-synthase deficiency. In: Pagon RA, Adam MP, Ardinger HH, *et al*., eds. GeneReviews(®) [Internet]. Seattle, WA: University of Washington, 1993.
38. Castinetti F, Moley J, Mulligan L, Waguespack SG. A comprehensive review on MEN2B. *Endocr Relat Cancer*, 2018; **25**: T29–39.
39. Davies JH, Cheetham T. Investigation and management of tall stature. *Arch Dis Child*, 2014; **99**: 772–7.
40. Groth KA, Skakkebæk A, Høst C, Gravholt CH, Bojesen A. Clinical review: Klinefelter syndrome—a clinical update. *J Clin Endocrinol Metab*, 2013; **98**: 20–30.
41. Tüttelmann F, Gromoll J. Novel genetic aspects of Klinefelter's syndrome. *Mol Hum Reprod*, 2010; **16**: 386–95.
42. Bar G, Lunenfeld E, Levitas E. [Klinefelter syndrome: genetic aspects, characteristics and reproduction—present and future]. *Harefuah*, 2014; **153**: 342–5, 366.
43. Saldarriaga W, Tassone F, González-Teshima LY, *et al*. Fragile X syndrome. *Colomb Médica Cali Colomb*, 2014; **45**: 190–8.
44. Sheth K, Moss J, Hyland S, Stinton C, Cole T, Oliver C. The behavioral characteristics of Sotos syndrome. *Am J Med Genet A*, 2015; **167A**: 2945–56.
45. Kanemoto N, Kanemoto K, Nishimura G, *et al*. Nevo syndrome with an NSD1 deletion: a variant of Sotos syndrome? *Am J Med Genet A*, 2006; **140**: 70–3.
46. Tatton-Brown K, Hanks S, Ruark E, *et al*. Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height. *Oncotarget*, 2011; **2**: 1127–33.
47. Freilinger P, Kliegel D, Hänig S, Oehl-Jaschkowitz B, Henn W, Meyer J. Behavioral and psychological features in girls and women with triple-X syndrome. *Am J Med Genet A*, 2018; **176**: 2284–91.
48. Neri G, Gurrieri F, Zanni G, Lin A. Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. *Am J Med Genet*, 1998; **79**: 279–83.
49. Maas SM, Vansenne F, Kadouch DJM, *et al*. Phenotype, cancer risk, and surveillance in Beckwith-Wiedemann syndrome depending on molecular genetic subgroups. *Am J Med Genet A*, 2016; **170**: 2248–60.
50. Brioude F, Lacoste A, Netchine I, *et al*. Beckwith-Wiedemann syndrome: growth pattern and tumor risk according to molecular mechanism, and guidelines for tumor surveillance. *Horm Res Paediatr*, 2013; **80**: 457–65.
51. Mester J, Charis E. PTEN hamartoma tumor syndrome. *Handb Clin Neurol*, 2015; **132**: 129–37.
52. Eugster EA, Pescovitz OH. Gigantism. *J Clin Endocrinol Metab*, 1999; **84**: 4379–84.
53. Christoforidis A, Maniadaki I, Stanhope R. McCune-Albright syndrome: growth hormone and prolactin hypersecretion. *J Pediatr Endocrinol Metab JPEM*, 2006; **19** Suppl 2: 623–5.
54. Rostomyan L, Daly AF, Beckers A. Pituitary gigantism: causes and clinical characteristics. *Ann Endocrinol*, 2015; **76**: 643–9.
55. Sotos JF, Argente J. Overgrowth disorders associated with tall stature. *Adv Pediatr*, 2008; **55**: 213–54.
56. He Q, Karlberg J. Bmi in childhood and its association with height gain, timing of puberty, and final height. *Pediatr Res*, 2001; **49**: 244–51.
57. Wong GW, Lai J, Cheng PS. Growth in childhood thyrotoxicosis. *Eur J Pediatr*, 1999; **158**: 776–9.
58. Wit JM, Camacho-Hübner C. Endocrine regulation of longitudinal bone growth. *Endocr Dev*, 2011; **21**: 30–41.
59. Berberoğlu M. Precocious puberty and normal variant puberty: definition, etiology, diagnosis and current management. *J Clin Res Pediatr Endocrinol*, 2009; **1**: 164–74.
60. Teilmann G, Pedersen CB, Jensen TK, Skakkebaek NE, Juul A. Prevalence and incidence of precocious pubertal development in Denmark: an epidemiologic study based on national registries. *Pediatrics*, 2005; **116**: 1323–8.
61. Meimaridou E, Hughes CR, Kowalczyk J, *et al*. Familial glucocorticoid deficiency: new genes and mechanisms. *Mol Cell Endocrinol*, 2013; **371**: 195–200.
62. Elias LL, Huebner A, Metherell LA, *et al*. Tall stature in familial glucocorticoid deficiency. *Clin Endocrinol (Oxf)*, 2000; **53**: 423–30.
63. Bulun SE. Aromatase and estrogen receptor α deficiency. *Fertil Steril*, 2014; **101**: 323–9.
64. Bruinsma FJ, Venn AJ, Patton GC, *et al*. Concern about tall stature during adolescence and depression in later life. *J Affect Disord*, 2006; **91**: 145–52.
65. Sandberg DE, Bukowski WM, Fung CM, Noll RB. Height and social adjustment: are extremes a cause for concern and action? *Pediatrics*, 2004; **114**: 744–50.

66. Rashad I. Height, health, and income in the US, 1984–2005. *Econ Hum Biol*, 2008; **6**: 108–26.
67. Osika W, Montgomery SM. Economic disadvantage modifies the association of height with low mood in the US, 2004: the disappointment paradox. *Econ Hum Biol*, 2008; **6**: 95–107.
68. Mittendorfer-Rutz E, Wasserman D, Rasmussen F. Fetal and childhood growth and the risk of violent and non-violent suicide attempts: a cohort study of 318,953 men. *J Epidemiol Community Health*, 2008; **62**: 168–73.
69. Pyle SI, Waterhouse AM, Greulich WW. Attributes of the radiographic standard of reference for the National Health Examination Survey. *Am J Phys Anthropol*, 1971; **35**: 331–7.
70. Bayley N, Pinneau SR. Tables for predicting adult height from skeletal age: revised for use with the Greulich-Pyle hand standards. *J Pediatr*, 1952; **40**: 423–41.
71. Tanner JM, Landt KW, Cameron N, Carter BS, Patel J. Prediction of adult height from height and bone age in childhood. A new system of equations (TW Mark II) based on a sample including very tall and very short children. *Arch Dis Child*, 1983; **58**: 767–76.
72. de Waal WJ, Greyn-Fokker MH, Stijnen T, *et al*. Accuracy of final height prediction and effect of growth-reductive therapy in 362 constitutionally tall children. *J Clin Endocrinol Metab*, 1996; **81**: 1206–16.
73. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging*, 2009; **28**: 52–66.
74. Khan KM, Gonzalez-Bolanos MT, Holm T, Miller BS, Sarafoglou K. Use of automated bone age for critical growth assessment. *Clin Pediatr (Phila)*, 2015; **54**: 1038–43.
75. Martin DD, Meister K, Schweizer R, Ranke MB, Thodberg HH, Binder G. Validation of automatic bone age rating in children with precocious and early puberty. *J Pediatr Endocrinol Metab JPEM*, 2011; **24**: 1009–14.
76. Martin DD, Heil K, Heckmann C, *et al*. Validation of automatic bone age determination in children with congenital adrenal hyperplasia. *Pediatr Radiol*, 2013; **43**: 1615–21.
77. Barnard ND, Scialli AR, Bobela S. The current use of estrogens for growth-suppressant therapy in adolescent girls. *J Pediatr Adolesc Gynecol*, 2002; **15**: 23–6.
78. Carel J-C, Blumberg J, Bougeard-Julien M, *et al*. Long-acting lanreotide in adolescent girls with constitutional tall stature. *Horm Res*, 2009; **71**: 228–36.
79. Brion DE, Murrieta D, Job JC. Bromocriptine treatment in tall adolescents: two years of clinical experience. *Horm Res*, 1985; **22**: 107–9.
80. Evain-Brion D, Garnier P, Blanco-Garcia M, Job JC. Studies in constitutionally tall adolescents. II. Effects of bromocriptine on growth hormone secretion and adult height prediction. *J Clin Endocrinol Metab*, 1984; **58**: 1022–6.
81. Schoenle EJ, Theintz G, Torresani T, Prader A, Illig R, Sizonenko PC. Lack of bromocriptine-induced reduction of predicted height in tall adolescents. *J Clin Endocrinol Metab*, 1987; **65**: 355–8.
82. Schwarz HP, Joss EE, Zuppinger KA. Bromocriptine treatment in adolescent boys with familial tall stature: a pair-matched controlled study. *J Clin Endocrinol Metab*, 1987; **65**: 136–40.
83. Benyi E, Berner M, Bjernekull I, *et al*. Efficacy and safety of percutaneous epiphysiodesis operation around the knee to reduce adult height in extremely tall adolescent girls and boys. *Int J Pediatr Endocrinol*, 2010; **2010**: 740629.
84. Odink RJ, Gerver WJ, Heeg M, Rouwé CW, van Waarde WMB, Sauer PJ. Reduction of excessive height in boys by bilateral percutaneous epiphysiodesis around the knee. *Eur J Pediatr*, 2006; **165**: 50–4.
85. Plaschaert VF, van der Eijken JW, Odink RJ, Delemarre HA, Caron JJ. Bilateral epiphysiodesis around the knee as treatment for excessive height in boys. *J Pediatr Orthop Part B*, 1997; **6**: 212–14.
86. Nilsson O, Weise M, Landman EBM, Meyers JL, Barnes KM, Baron J. Evidence that estrogen hastens epiphyseal fusion and cessation of longitudinal bone growth by irreversibly depleting the number of resting zone progenitor cells in female rabbits. *Endocrinology*, 2014; **155**: 2892–9.
87. Weise M, De-Levi S, Barnes KM, Gafni RI, Abad V, Baron J. Effects of estrogen on growth plate senescence and epiphyseal fusion. *Proc Natl Acad Sci U S A*, 2001; **98**: 6871–6.
88. Goldzieher MA. Treatment of excessive growth in the adolescent female. *J Clin Endocrinol Metab*, 1956; **16**: 249–52.
89. Svan H, Ritzén EM, Hall K, Johansson L. Estrogen treatment of tall girls: dose dependency of effects on subsequent growth and IGF-I levels in blood. *Acta Paediatr Scand*, 1991; **80**: 328–32.
90. Drop SL, De Waal WJ, De Muinck Keizer-Schrama SM. Sex steroid treatment of constitutionally tall stature. *Endocr Rev*, 1998; **19**: 540–58.
91. Venn A, Hosmer T, Hosmer D, *et al*. Oestrogen treatment for tall stature in girls: estimating the effect on height and the error in height prediction. *Clin Endocrinol (Oxf)*, 2008; **68**: 926–9.
92. Upners EN, Juul A. Evaluation and phenotypic characteristics of 293 Danish girls with tall stature: effects of oral administration of natural 17β-estradiol. *Pediatr Res*, 2016; **80**: 693–701.
93. de Waal WJ, Torn M, de Muinck Keizer-Schrama SM, Aarsen RS, Drop SL. Long-term sequelae of sex steroid treatment in the management of constitutionally tall stature. *Arch Dis Child*, 1995; **73**: 311–15.
94. Binder G, Grauer ML, Wehner AV, Wehner F, Ranke MB. Outcome in tall stature. Final height and psychological aspects in 220 patients with and without treatment. *Eur J Pediatr*, 1997; **156**: 905–10.
95. Radivojevic U, Thibaud E, Samara-Boustani D, Duflos C, Polak M. Effects of growth reduction therapy using high-dose 17beta-estradiol in 26 constitutionally tall girls. *Clin Endocrinol (Oxf)*, 2006; **64**: 423–8.
96. Weimann E, Bergmann S, Böhles HJ. Oestrogen treatment of constitutional tall stature: a risk-benefit ratio. *Arch Dis Child*, 1998; **78**: 148–51.
97. Venn A, Bruinsma F, Werther G, *et al*. Oestrogen treatment to reduce the adult height of tall girls: long-term effects on fertility. *Lancet Lond Engl*, 2004; **364**: 1513–18.
98. Hendriks AEJ, Laven JSE, Valkenburg O, *et al*. Fertility and ovarian function in high-dose estrogen-treated tall women. *J Clin Endocrinol Metab*, 2011; **96**: 1098–105.
99. Rask O, Nilsson KO, Berntorp E. Oestrogen treatment of constitutional tall stature in girls: is there a risk of thrombosis or bleeding? *Acta Paediatr Oslo Nor 1992*, 2008; **97**: 342–7.
100. Benyi E, Ritzen M, Savendahl L, *et al*. Risks of malignant and non-malignant tumours in tall women treated with high-dose oestrogen during adolescence. *Horm Res Paediatr Horm Res Paediatr*, 2014; **82**: 89–96.
101. Reinehr T, Gueldensupp M, Wunsch R, Bramswig JH. Treatment of tall stature in boys: comparison of two different treatment regimens. *Horm Res Pædiatrics*, 2011; **76**: 343–7.

102. Brämswig JH, Schellong G, Borger HJ, Breu H. [Testosteron treatment of excessively tall boys (author's transl)]. *Dtsch Med Wochenschr 1946*, 1981; **106**: 1656–61.
103. van den Bosch JS, Smals AG, Pieters GF, Valk IM, Kloppenborg PW. Instant growth inhibition by low dose oestrogens in excessively tall boys. *Acta Endocrinol (Copenh)*, 1982; **100**: 327–32.
104. Decker R, Partsch C-J, Sippell WG. Combined treatment with testosterone (T) and ethinylestradiol (EE2) in constitutionally tall boys: is treatment with T plus EE2 more effective in reducing final height in tall boys than T alone? *J Clin Endocrinol Metab*, 2002; **87**: 1634–9.
105. Hendriks AEJ, Boellaard WPA, van Casteren NJ, *et al*. Fatherhood in tall men treated with high-dose sex steroids during adolescence. *J Clin Endocrinol Metab*, 2010; **95**: 5233–40.
106. Lemcke B, Zentgraf J, Behre HM, Kliesch S, Bramswig JH, Nieschlag E. Long-term effects on testicular function of high-dose testosterone treatment for excessively tall stature. *J Clin Endocrinol Metab*, 1996; **81**: 296–301.
107. Phemister DB. Operative arrestment of longitudinal growth of bones in the treatment of deformities. *J Bone Jt Surg*, 1933; **15**: 1.
108. Bowen JR, Johnson WJ. Percutaneous epiphysiodesis. *Clin Orthop*, 1984; **190**: 170–3.
109. Canale ST, Russell TA, Holcomb RL. Percutaneous epiphysiodesis: experimental study and preliminary clinical results. *J Pediatr Orthop*, 1986; **6**: 150–6.
110. Ghanem I, El Hage S, Diab M, *et al*. Radiofrequency application to the growth plate in the rabbit: a new potential approach to epiphysiodesis. *J Pediatr Orthop*, 2009; **29**: 629–35.
111. Widmann RF, Amaral TD, Yildiz C, Yang X, Bostrom M. Percutaneous radiofrequency epiphysiodesis in a rabbit model: a pilot study. *Clin Orthop*, 2010; **468**: 1943–8.

7.2

Sex Development

7.2.1 Sex Determination and Differentiation

Physiology Leading to Male and Female Development

Olaf Hiort and Ralf Werner

Introduction

The biological sex of a human being is dependent on a genetic pathway that determines the differentiation of the gonads to develop either into a testis or an ovary. This pathway usually starts with the initial setting of either a 46,XX or a 46,XY karyotype. Subsequently, a tightly regulated cascade of both sex chromosome and autosome derived genetic expression is initiated. From the bipotent gonadal anlagen, ovarian or testicular cells develop, which set the stage for hormone synthesis and allow the shaping of the individual phenotype into either male or female. These processes occur in a precise step-wise and time-dependent manner and any disruption at sensitive time intervals will allow deviation from the usual pathways.

These conditions are described as 'disorders or differences of sex development (DSD)', and the clinical approach of professionals towards people affected by DSD is described in the next Chapter 7.2.2. DSD encompasses mostly rare to very rare genetically determined variants of a discrepancy between chromosomal, gonadal, and phenotypic sex. It is unclear at this time if external, non-genetic factors may disrupt the endocrine pathways and thus lead to DSD. Sex development is not only related to development of genital structures, but also includes sexually dimorphic development of brain structures. Therefore, DSD will affect not only the appearance of primary genital structures and subsequent sexual maturation at puberty, but also gender identity. Furthermore, the initial genetic components of gonadal development may be involved in differentiation and function of other organs, such as the adrenal, the heart, the skeletal system or even the neurological system. Therefore, patients with DSD conditions may have associated endocrine illnesses and other system disorders.

Chromosomal Sex

The chromosomal sex of an individual is determined with the fusion of egg and sperm at the time of conception. Usually, one X chromosome is inherited from the mother, while the sperm from the father may provide either an X or a Y chromosome to account for the 46,XX or 46,XY karyotype of the embryo. This genetic layout is seen as a starting point of sex development in the principally 'un-sexed' child. The chromosomes themselves may harbour sex-specific and sex-nonspecific, pseudoautosomal genes. Sex chromosome division may be altered either prior to conception or during the first cell divisions of the blastocyst, leading to numerical aberrations. Well-known are sex chromosome trisomies such as 47,XXY (Klinefelter syndrome) or monosomies such as 45,X (Turner syndrome), but also higher numbers of sex chromosomes (such as 48,XXXY, 49,XXXXY etc.) have been described in humans [1]. Mosaicism (e.g. 45,X/46,XY) arises from the loss of a sex chromosome during the first cell cycles. Also, chimeric alterations (e.g. 46,XX/46,XY) may occur. It is believed that the presence of Y-chromosomal material expressing *SRY* (*sex determining region Y*) in any 46,XX or mosaic individual may lead to testicular or partial testicular development of the gonadal structures and, hence to a range of phenotypes.

Development of the Bipotential Gonad

Most information regarding the development of the gonads and the genes involved in their development in mammals come from mouse genetics and humans presenting with DSD. The gonadal primordium is believed to arise as part of a common adrenogonadal primordium from the urogenital ridge, which develops as a paired structure at 10.5 days post-coitum (dpc) in mice and around the fifth week post-conception (p.c.) in humans. A gubernaculum forms at the cranial and caudal ends to allow the descent of the gonad during later development (Figure 7.2.1.1). The bipotential gonads harbour two distinct cell types, somatic cells and germline cells, that differ in their embryologic origin. The somatic part is composed of cells developing

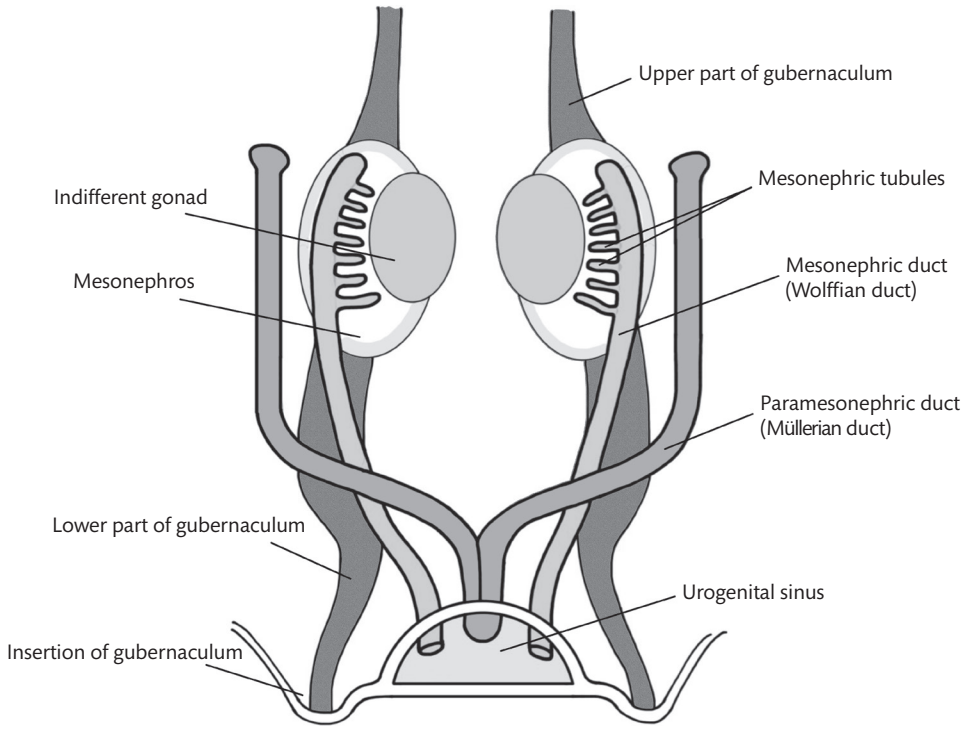


Figure 7.2.1.1 The indifferent or bipotent gonads with the gubernaculum and the mesonephric and paramesonephric ducts.
Reprinted from Hiort O, Danne T, Wabitsch M (eds.) *Pädiatrische Endokrinologie und Diabetologie*, 2nd edition, Springer 2019, reproduced with permission of SNCSC. Adapted by cand.med. Mirkka Hiort.

from the coelomic epithelium and the mesenchyme beneath. It originates from the intermediate mesoderm and partly from the mesonephros. Genital ridge formation starts with increased proliferation of epithelial cells on the surface of the mesonephros forming a dense and pseudostratified layer. Its underlying basal membrane becomes disintegrated and allows the epithelial cells to invade the mesonephros. The germ line, however, comes from the primordial germ cells which derive from the ectoderm. In the mouse, induction of primordial germ cells from the proximal epiblast is first visible as a cluster of about 45 cells around 6.25 dpc [2]. The primordial germ cells proliferate and migrate to the gonadal ridge after which they undergo epigenetic reprogramming and erasure of parental imprints [3]. In humans, at the end of the fifth week p.c. (10.5 dpc in mice), the primordial germ cells migrate to the surface of the gonadal ridge joining the somatic cells. The growing coelom epithelium is now in very close contact with the primordial germ cells, maintaining the integrity of germ cells and the somatic part of the gonadal ridge. The germ cells proliferate and no differences between male or female gonads are visible in mice up to 12 dpc, and the sixth week p.c. in humans (Figure 7.2.1.2).

Genes Involved in Formation of the Bipotent Gonad

Studies of knock out mice and people with gonadal dysgenesis provided information on genes involved in genital ridge development. A set of genes including *GATA* binding protein 4 (*Gata4*), nuclear receptor subfamily 5 group A member *1* (*Nr5a1,* also known as steroidogenic factor 1, Sf1), Wilms tumour 1 (*Wt1*), Lim homeobox protein 9 (*Lhx9*) and empty spiracles homeobox 2 (*Emx2*) have all been shown to be vital to the development of the genital ridge [4]. *Wt1* is expressed as early as 9.5 dpc in the whole urogenital ridge [5] and plays an important role in the development of the gonad as well as the kidneys. This zinc finger transcription factor has 24 isoforms. Alternative splicing leads to an insertion or omission of three amino acids (+/– KTS; lysine, threonine, serine) between the third and fourth zinc finger. The Wt1-KTS isoform seems to have a key role in gonadal ridge development since ablation of the *Wt1*-KTS leads to a lack of broadening of the genital ridge due to increased apoptosis [6].

GATA4 is a zinc finger transcription factor that is necessary for genital ridge initiation and is expressed in the genital ridge in an anterior-posterior direction preceding immediately the thickening of the coelomic epithelium. *Gata4* knock-out is embryonic lethal due to defects in early ventral morphogenesis [7]. Therefore, conditional *Gata4* knock-out mice have been created that display no gonadal initiation. Their coelomic epithelium remains a monolayer and the basement membrane underneath persists unfragmented preventing migration [4]. Knock out mice for *Emx2* also show gonadal dysgenesis and a decrease of epithelial cells migrating to the mesenchymal part of the gonadal ridge [4, 8, 9]. In humans, haploinsufficiency caused by deletions encompassing *EMX2* have been linked to a wide spectrum of DSD ranging from hypospadias to complete sex reversal [10]. *Lhx9* KO mice show agenesis of the gonads in both sexes. The genital ridge develops, but the coelomic epithelium fails to proliferate and gonads are not formed. Nevertheless, germ cells migrate normally to the presumptive gonadal region at the expected time [11]. Although mice lacking *Lhx9* do not exhibit further severe defects, to date no mutations in this

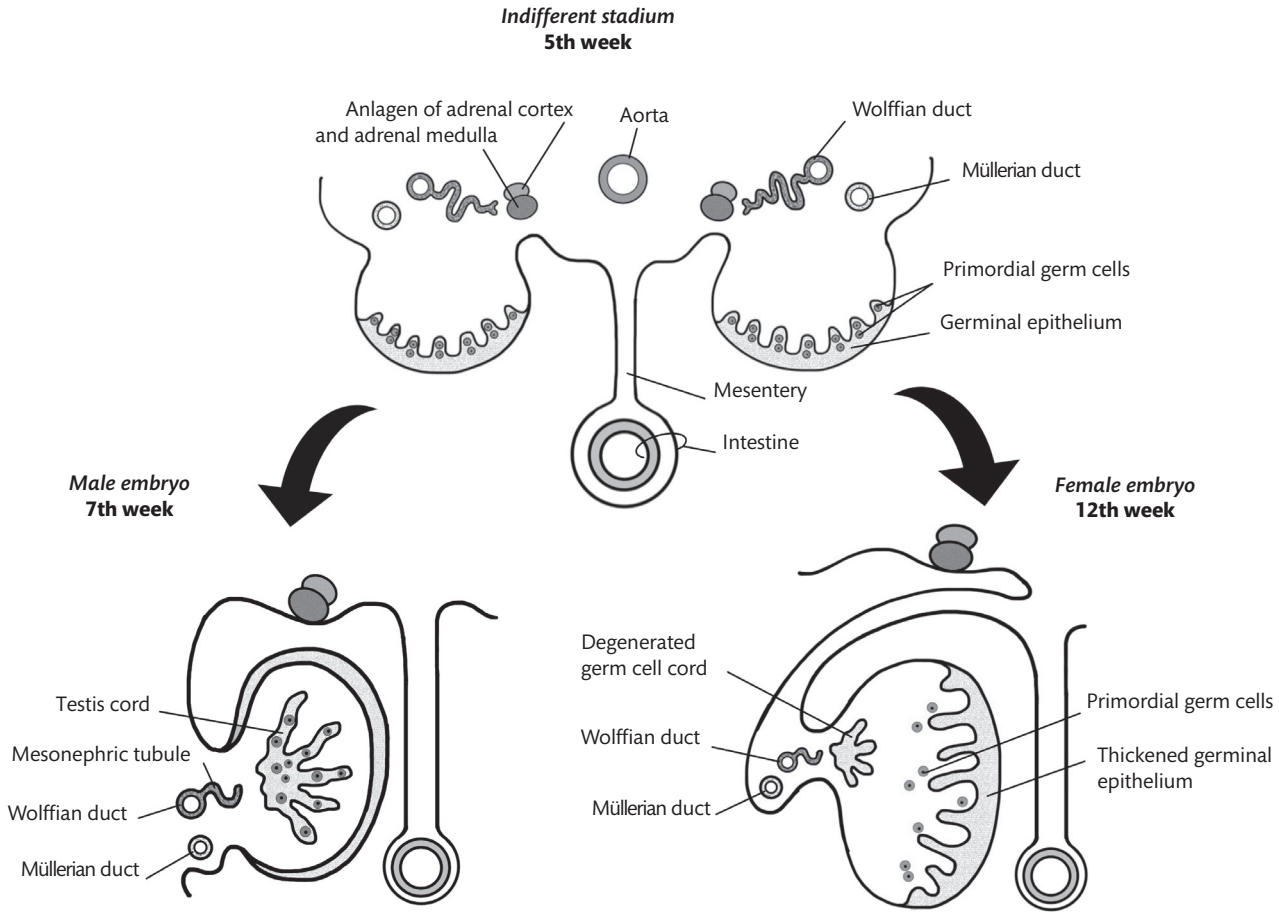


Figure 7.2.1.2 Differentiation of testis and ovary from the indifferent stage.

Reprinted from Hiort O, Danne T, Wabitsch M (eds.), *Pädiatrische Endokrinologie und Diabetologie*, 2nd edition, Springer 2019, reproduced with permission of SNCSC. Adapted by cand.med. Mirkka Hiort.

gene in humans have been described. *Nr5a1* expression follows *Gata4* expression and is restricted to the genital ridge and the adrenal primordium. Mice lacking *Nr5a1* have a similar phenotype to *Lhx9* KO mice. They lack adrenal glands and gonads and both male and female mice have female internal genitalia [12]. Interestingly, expression of *Nr5a1* in *Lhx9* –/– mice is greatly reduced, suggesting that LHX9 might be a regulator acting upstream of *Nr5a1*.

After fragmentation of the basement membrane some proliferating SF1 positive cells of the coelomic epithelium undergo epithelial-mesenchymal transformation (EMT) and migrate inwards into the mesenchyme forming the gonadal somatic precursor cells.

Testis Determination

The presence and expression of the master switch gene *Sry* (sex-determining region of the Y chromosome) leads to an upregulation of its target *Sox9* and drives the bipotential precursors into the Sertoli cell lineage leading to testis development. In mice, XY genital ridges with a deletion or malfunction of *Sry* or *Sox9* can differentiate into an ovary and forced expression of *Sry* or overexpression of *Sox9* in XX genital ridges can induce development of testes [13–15]. This demonstrates that the cell types in the genital ridge are equally capable of expressing both male and female pathway genes resulting in testis or ovary development with a high dependency on the presence of a Y-chromosome (including the *Sry* gene). *Sry* expression is the trigger to activate the testis determining pathway and to suppress the ovarian determining pathway [16]. In mice, *Sry* is expressed only in a small-time window from 10.5 dpc to 12.5 dpc starting in the centre of the gonad. The expression expands rapidly to its poles, reaches its maximum around 11.5 dpc and extinguishes from the centre to the poles. An accurate spatiotemporal expression of *Sry* is necessary since delayed or reduced *Sry* expression impairs testis development [17]. In humans *SRY* expression starts in the seventh week of gestation (41–44 days post ovulation, d.p.o.) in XY gonads but persists at low levels throughout the embryonic phase and beyond [18]. Studies in mice have shown numerous factors regulating *Sry* expression. Recently, these factors have been integrated into a model consisting of three regulatory modules that modify the activity of the three key transcription factors SF1, GATA4, and WT1 binding to the *Sry* promoter [17]. Mutations that affect the level of *SRY* expression or its timing during sex determination might also be responsible for some idiopathic cases of gonadal dysgenesis [19]. One of the earliest morphological changes after *Sry* expression is an increase in proliferation that results in an increase of the size of the XY gonad compared to the XX gonad. This increase in proliferation is dependent on SRY and occurred in two phases. In the first phase enhanced proliferation is observed in SF1 positive cells of the

coelomic epithelium contributing to the Sertoli-cell population, and in a second phase proliferation occurs in SF1 negative cells of the genital ridge that are not precursors of Sertoli cells [20].

The transcription factor SRY binds to and activates the testis specific enhancer (TES) on *Sox 9* [21, 22]. After *Sry* is downregulated, SOX9 maintains its expression in Sertoli cell precursors by binding to its own TES via the same binding motif as SRY [21]. SOX9 also directly or indirectly upregulates the proliferation-promoting factor FGF9 that maintains *Sox9* expression via its FGF receptor 2 (FGFR2). FGFR2 was detected at the plasma membrane of proliferating coelomic epithelial cells and in the nucleus of Sertoli precursors [23, 24]. In *Fgf9* or *Fgfr2* KO mice *Sox9* expression is reduced and all testis specific events in the developing gonad are disrupted such as cell proliferation, Sertoli cell differentiation or testis cord formation [24–27]. Recently, a first patient with a heterozygous mutation in *FGFR2* displaying craniosynostosis with 46,XY DSD due to gonadal dysgenesis has been reported [28] and gave a first indication that FGF9 levels may also play a role in human testis development. SRY and SOX9 also bind to the promoter of the *Ptgs* gene encoding prostaglandin D synthase mediating the synthesis of PGD2 [29, 30] initiating a paracrine feed-forward loop. PGD2 binds to its DP-receptor and activates an intracellular cascade that leads to activation of cAMP-dependent protein kinase A and phosphorylation of SOX9 protein. Phosphorylated SOX9 translocates into the nucleus and binds to its own promoter thereby maintaining its expression [31]. This pathway is a second SOX9 amplification pathway independent of the FGF9 pathway [32] contributing to Sertoli cell differentiation and may play a role in *Sox9* upregulation of neighbouring SRY negative cells to reach a threshold of SOX9 positive Sertoli precursors to force testis development.

One of the early target genes of SOX9 is anti-Müllerian hormone (AMH), a signalling molecule of the TGF (transforming growth factor) superfamily responsible for the regression of the Müllerian ducts in XY embryos (see next). In the genital ridge, *Amh* expression is limited to pre-Sertoli cells after *Sry* is expressed and testis cord formation becomes visible [33]. Once *Amh* expression is initiated its expression level continues at high levels during gestation. Several transcription factors that are also important for the development of the gonadal ridge and the first steps of sex determination such as WT1, GATA4, SF1, and SOX9 are recruited to the *Amh* promoter and control *Amh* transcription [34, 35].

Differentiation of Sertoli cells is the first step towards testis determination. Sertoli cells are the organizing centres in the developing testes. They influence testis cord formation and the differentiation of other important cell types of the testis, like peritubular myoid cells, fetal Leydig cells and endothelial cells [16]. Testis cord formation starts with the aggregation of pre-Sertoli cells around clusters of germ cells, but does not depend on the presence of germ cells, since cords devoid of germ cells develop normally. The cords tubularize and attach with both ends to a common junction, the rete testis. Peritubular myoid cells (PMC) form the outer layer of the seminiferous tubules and are in direct contact with the surface of Sertoli cells. In mice, this is a monolayer, but in humans it typically consists of multiple layers that are separated by an extracellular matrix, which is deposited by both PMC and Sertoli cells [36, 37]. PMC are first visible at 13.5 dpc in mice or 12 weeks in humans [38, 39].

In mice, Sertoli cells express *Desert Hedgehog* (*Dhh*) shortly after *Sry* and its expression persists into adulthood. DHh is one of three secreted mammalian hedgehog proteins (Sonic Hedgehog, SHh; Indian Hedgehog IHh, and DHh) that acts as a morphogenic ligand via its receptor patched1 to regulate Leydig cell and PMC differentiation. *Dhh* KO-mice show affected PMC and Leydig cell differentiation with disrupted testicular cord structure, fewer fetal Leydig cells in the embryo and absence of adult Leydig cells [40]. The phenotype of the XY mice depends on the genetic background and ranges from infertile males to complete sex reversal [41, 42]. In humans homozygous mutations in *DHH* are associated with partial or complete gonadal dysgenesis with and without polyneuropathy [43–45]. Interestingly, ectopic activation of the hedgehog pathway in SF1 positive somatic precursors of fetal mouse ovaries transformed these cells into fetal Leydig cells. These cells produce androgens and INSL3 (insulin-like growth factor 3) and lead to virilization of the female embryos and ovarian descent, but the female reproductive system is left intact [46]. Leydig cells appear at least in two phases (three phases in humans) as two distinct Leydig cell populations: In the fetal testis as fetal Leydig cells (FLC) and in the adult testis as adult Leydig cells (ALC). In humans, there is a third phase of Leydig cell amplification in the neonatal phase (neonatal Leydig cells, NLS) during minipuberty between one and six months of age [47]. It is believed that fetal Leydig cells decline in the postnatal testis and are replaced by another population of adult Leydig cells, but recent lineage tracing experiments in mice have shown that FLC may persist in the adult testis. In mice, fetal and adult Leydig cells differ in the androgens they produce. An example is the expression of *Hsd17b3*, which converts androstenedione to testosterone. *Hsd17b3* is expressed in ALC but not FLC, therefore the main androgens produced by FLC and ALC are androstenedione and testosterone, respectively. In the fetus, androstenedione is converted by Sertoli cells to testosterone required for masculinization of the embryo [48]. Another difference between FLC and ALC in rodents is the ability of FLC to produce glucocorticoids [49]. Nevertheless, recent studies support the hypothesis that FLC and ALC arise from stem Leydig cells (SLC) that have their origin in the fetal testis, but it is unclear if they arise from common progenitors. The initial steroidogenesis of FLC is independent of gonadotrophins, but human FLC become dependent on placental human chorionic gonadotrophin (hCG) and human LH/CG receptor (LHCGR), which is expressed at the 11th week of gestation. In humans, testosterone secretion peaks between week 12–14 and the concentration of hCG is 10-fold higher than that of LH. During the second trimester, hCG concentration falls significantly followed by a decrease of testosterone secretion while LH concentration remains constant [50]. Human mutations in the LH/CG-receptor show the importance of gonadotrophin stimulation of fetal Leydig cells, since 46,XY patients with homozygous inactivating mutations in the *LHCGR* gene present with Leydig cell hypoplasia or Leydig cell agenesis [51]. In contrast mutations in the *LHB* (luteinizing hormone beta polypeptide) gene in 46,XY individuals lead to normal virilization at birth, but cause pubertal delay and hypogonadism [52, 53] indicating that FLC are dependent on hCG but not LH.

Ovary Determination

Sex determination is regulated by a molecular antagonism between testis and ovary pathways. The cell fate decision in the supporting cells, that determine a Sertoli or granulosa cell line, is a key step in sex determination. Since *Sry* is absent in a XX gonosomal

background, *Sox9* is not upregulated in the gonadal ridge. Therefore, SF1 positive supporting cell precursors are not recruited to the Sertoli cell lineage. Instead the *Wnt/ß-catenin* pathway is activated by R-spondin1 (*Rspo1*), Wingless-related MMTV integration site 4 (*Wnt4*), and Forkhead box L2 (*Foxl2*) [54]. *Wnt4* and *Rspo1* are expressed in the early somatic precursor cells of the genital ridge and, if not downregulated by the testis pathway, maintained in the mouse ovary. RSPO1 is involved in the female-specific upregulation of *Wnt4* and therefore knockout of either *Wnt4* or *Rspo1* in mice leads to a similar phenotype with partial female to male sex reversal, associated with a presence of steroidogenic cells with ectopic androgen production and formation of coelomic vessels [55, 56]. In humans, a heterozygous *WNT4* missense mutation was detected in a 46,XX patient with regressed Müllerian structures and elevated androgen production suggesting partial gonadal dysgenesis [57]. Homozygous disruption of WNT4 is associated with SERKAL syndrome (46,XX **Se**x **R**eversal with dysgenesis of **K**idneys, **A**drenals and **L**ungs) including 46,XX DSD (OMIM # 611812) [58]. Disruption of *RSPO1* causes a syndrome characterized by 46,XX ovotesticular DSD (female to male sex reversal), palmoplantar hyperkeratosis and predisposition to squamous cell carcinoma of the skin (OMIM #610644) ([59, 60]. In contrast, a 46,XY patient with sex reversal carrying a chromosome 1p31–35 duplication including *WNT4* and *RSPO1* genes has been identified [61]. Signalling through the canonical WNT pathway by WNT4 and RSPO1 induces stabilization and nuclear translocation of ß-catenin and activation of target genes. ß-catenin is expressed in the gonads of both sexes, but specific ablation of ß-catenin in SF1 positive somatic cells of the gonadal ridge affects only ovarian development and the phenotype is similar to that seen in *Wnt4* or *Rspo1* knockouts [62]. Conditional knock-in mice that express a stabilized form of ß-catenin in Sf1 positive somatic cells of XY gonads also override the male programme and show a male to female sex reversal [63]. The Rspo1/Wnt4/ß-catenin and the Fgf9 signalling pathway repress each other [23]. Interestingly, *Sox9* and *Fgf9* are not upregulated in *Ctnnb1* (ß-catenin) XX knock out mice [62] demonstrating that masculinization is a direct effect of loss of ß-catenin and other genes must be repressed by ß-catenin. This is supported by several studies of double knock-out mice affecting male and female pathway genes like *Wnt4/Fgfr2, Wnt4/Fgf9* [64], *Rspo1/Sox9* [65], and *Ctnnb1/Sox9* [66]. While knock-out of *Sox9* or the *Fgf9* pathway could be rescued in XY-mice by additional knock-out of the *ß-catenin* pathway, possibly by expression of other SRY target genes like *Sox8*, XX-mice still show partial masculinization. These results suggested that *Sox9* might be dispensable and that some other pro-testis genes require repression by the Wnt/ß-catenin pathway [66].

Surprisingly, once the ovary or testis is determined, its maintenance needs further lifelong expression of pro-ovarian or pro-testis genes. Conditional ablation of the transcription factor *Foxl2* in the adult mouse ovary leads to somatic ovary to testis trans-differentiation with reprogramming of granulosa and theca cells into Sertoli-like and Leydig-like cells capable of testosterone production at a level consistent with normal males [67]. FOXL2 maintains the ovarian phenotype mainly by repression of *Sox9* together with oestrogen receptor 1/ESR1 [67]. In humans, heterozygous mutations in *FOXL2* cause the blepharophimosis/ptosis/epicanthus inversus syndrome (BPES, OMIM #110100) associated with ovarian failure [68]. Recently, BPES in conjunction with congenital heart failure was also detected in two of three children with 46,XX ovotesticular/testicular DSD (OTDSD/TDSD) harbouring a heterozygous frame shift mutation in Nuclear receptor subfamily 2 group F member 2 (*NR2F2,* also known as Chicken ovalbumin upstream promoter transcription factor 2, Coup-TFII) indicating that NR2F2 might be a new pro-ovary and anti-testis sex determining factor [69].

In contrast, loss of the double-sex and mab-3 related transcription factor 1 (*Dmrt1*) in XY mice activates *Foxl2* expression and re-programmes Sertoli cells into granulosa cells, even if ablated conditionally in adults [70]. Under the influence of transdifferentiated Sertoli cells, theca-like cells form and produce oestrogen and even germ cells are feminized in the mutant gonads [70]. This plasticity of Sertoli and granulosa cells after terminal differentiation demonstrates that the molecular antagonism between male and female sex determination continues into adulthood and indicates that DMRT1 and FOXL2 are essential for postnatal sex maintenance of the testis or ovary, respectively. Interestingly, XY *Dmrt1* mutant mice are born as males with testes and feminize only after birth [71]. However, chromosome 9p deletions in humans removing *DMRT1* or partial deletions of *DMRT1* can cause 46,XY DSD with feminization at birth (OMIM #154230) [72, 73].

Sex Differentiation

Sex differentiation refers to the development of the phenotypic features of sex, namely the development of the internal and external male or female structures. This is mostly a hormone driven process, depending on the presence or absence of either anti-Müllerian hormone (AMH) or androgenic sex steroids. This finally results in the male or female somatic sex.

While the female somatic sex at birth does not seem to be dependent on ovarian endocrine activity, male sex differentiation has to be seen as the active, hormone-dependent, and irreversible modification of the primarily bipotent tissue structures of the internal and external genital anlagen. Androgenic steroids such as testosterone will lead to irreversible downstream expression patterns of genes corresponding to induction of the male phenotype in cells of the genital tubercle. Thus, these androgenic steroids program the transcriptome and epigenome of the cells in a so-called male programming window in the developing embryo [74]. Clinically, the effects are seen as the anogenital distance, the development of the scrotum and penile differentiation and length. These effects are also likely to account for reproductive issues such as sperm count and long-term testosterone levels in the adult male. We have to presume that the androgenization effects are also major players in the determination of facets such as body composition (muscle:fat ratio) and bone mineral density.

Development of Internal Genital Structures

In both sexes the internal genital structures arise again from bipotent anlagen, the Wolffian ducts and the Müllerian ducts. Under the influence of AMH secreted from the Sertoli cells of the testes during the critical period, the Müllerian ducts regress. So in the typical male development, the differentiation and growth of the uterus, upper third of the vagina, and fallopian tubes is actively suppressed. AMH acts through the AMH II receptor in the Müllerian duct mesenchyme [75].

In contrast, the Wolffian ducts develop under the influence of testosterone, giving rise to the epididymis, the seminal vesicles, and some parts of the prostate. The human fetal testes synthesize testosterone from the eighth week of gestation. The timing of development for testosterone-dependent structures, however, is very distinct. The prostate forms around the tenth to thirteenth week of gestation, while the seminal vesicles develop later, around week 14–16 (for a review, see [74]). Testosterone initially stabilizes the Wolffian ducts and elongation and convolution of the cranial end leads to formation of the epididymis, the middle portion forms the vasa deferentia and the caudal part differentiates into the seminal vesicles. The prostate develops from the urogenital sinus, which arises in humans around the seventh week of gestation from the cloaca. Androgens, namely testosterone, will lead to differentiation into zonal subdivisions to form the prostate. Female development does not depend on hormonal influences to mediate differentiation. In the absence of AMH, the Müllerian duct will differentiate into the uterus, fallopian tubes, cervix, and upper vagina. The lack of androgens causes the Wolffian duct to regress.

Again, the male development is dominant and hormone dependent. In some types of DSD with high amounts of testosterone formed in the adrenal, such as 21-hydroxylase deficiency leading to congenital adrenal hyperplasia, there are residues of prostate tissue even in 46,XX individuals.

Development of External Genital Structures

The external genital anlagen is in a bipotent stage around the seventh week of gestation. If no sex steroids are acting on the tissue, female external structures will arise with a clitoris, labia minora and majora. Under the influence of androgens, namely dihydrotestosterone converted peripherally from testosterone through 5alpha-reductase type 2, the phallic structure arising from the genital tubercle will form into a penis with elongation of the urethra to the tip of the phallus. Likewise, the urogenital folds will fuse and form the scrotal folds, while the labia minora are equivalent to parts of the shaft of the phallus in the male. This differentiation will be finished by the end of the twelfth week of gestation and thereafter, high amounts of dihydrotestosterone are needed for the growth of the phallus and of the prostate. Interestingly, these differentiation processes can only take place during very defined and strict time intervals, which represent the previously described 'male programming window'. Elaborate animal model experiments demonstrate that the lack of androgens will lead to a female phenotype regardless of karyotype while androgenization leads to a male phenotype regardless of karyotype.

Thus, human sex development depends to a large extent on the systematic and local synthesis and action of specific hormones, which act in a very distinct spatial and time-dependent manner [76]. The precursor of sex steroids is cholesterol, which is metabolized through a variety of enzymes, which may in part be dependent on adequate co-enzymes and co-factors. The majority of steroidogenic enzymes are P450 (CYP) enzymes, which catalyse redox reactions relying on electron supply via specific electron transfer. Typically, these enzymes are localized in the mitochondria. This applies for instance to the first critical enzyme P450 side-chain cleavage (CYP11A1), 11beta-hydroxylase (CYP11B1), and aldosterone synthase (CYP11A1); in contrast 17alpha-hydroxylase (CPY17A1), 21-hydroxylase (CYP21A2), and aromatase (CYP19A1) are localized to the endoplasmic reticulum. The latter enzymes depend on P450 oxidoreductase (POR) for electron transfer.

The initial step of cholesterol conversion to pregnenolone is crucial and of special interest, because it not only relies on the P450 side-chain cleavage enzyme (P450scc), but in selected organs such as adrenal and testis it needs a specific transporter, the Steroidogenic Acute Regulatory (StAR) protein; StAR promotes a rapid trans-membranous crossing of cholesterol into the mitochondrium. Two very rare conditions have been described due to mutations of either P450scc or StAR. The description of deleterious mutations in P450scc came as a surprise, because it had been thought that this would also hamper severely the placental steroid production and thus not be compatible with life [77, 78]. However, recently mutations have been described which only partially affect P450scc function and lead to minor phenotypes in the DSD spectrum [79, 80]. StAR defects have initially been described in infants with grossly enlarged adrenals, 'congenital lipoid adrenal hyperplasia' [81], due to accumulation of cholesterol within the cytoplasm, but these findings have been challenged in part by other case descriptions [82].

CYP17A1 is an interesting protein with actually two enzymatic capacities, namely the 17alpha-hydroxylase activity and an additional 17,20 lyase activity. The first converts pregnenolone or progesterone to their 17alpha-hydroxylated products, the latter metabolizes these compounds into dehydroepiandrosterone (DHEA) or androstenedione, respectively. The differentiation of the two enzymatic capacities of CYP17A1 is important in the diagnosis of DSD, because of the presence or absence of alterations of the mineralo- and glucocorticoid pathways. Isolated 17,20 lyase deficiency is mostly due to defects in the cytochrome b5, which is an allosteric factor promoting the interaction with POR and CYP17A1 and thus enhancing the 17,20 lyase activity without influencing 17alpha-hydrolyse activity [83].

The 3beta-hydroxysteroid dehydrogenase type 2 (HSD3B2) enzyme is critical, because of its ability to convert pregnenolone and its 17alpha-hydroxylated form into progesterone and 17-OH progesterone, therefore playing an important role in the adrenal for mineralo- and glucocorticoid synthesis. In addition, HSD3B2 converts DHEA directly into androstenedione and thus controls the complete Δ4 steroid synthesis pathway. Mutations affecting HSD3B2 are associated with rare forms of congenital adrenal hyperplasia, but due to its capacity to convert DHEA to androstenedione it should be considered in the differential diagnosis of DSD.

Mainly in 46,XY DSD, defects of the down-stream cascade of testosterone synthesis play an important role. The enzymes are crucial and necessary for male phenotypic development. 17beta-hydroxysteroid dehydrogenase (HSD) type 3 (HSD17B3) synthesizes testosterone from androstenedione. This enzyme is mostly expressed in the testes and primarily acts during fetal development.

However, currently, a number of isoenzymes are known, for instance 17beta HSD type 5 (AKR1C3, aldo-keto reductase family 1 member C3) which may be involved in substantial testosterone synthesis in the testes postnatally, especially at the time of puberty [84]. This may support virilization in patients with 17beta HSD type 3 deficiency during pubertal development despite underlying deleterious mutations [85]. Conversion of testosterone to the more potent dihydrotestosterone (DHT) is mainly not localized in the testes, but directly in the androgen target tissues. Here 5-alpha reductase type 2 (SRD5A2) is the main enzyme, in particular during prenatal development. Lack of DHT due to mutation of *SRD5A2* will lead to

variable phenotypes of under-androgenization from almost completely female external genitalia to only slightly diminished penile size [86]. These patients have a high virilization potential at the time of puberty, mainly due to the effects of testosterone; however, some of DHT may also be synthesized via the 5-alpha reductase type 1 (*SRD5A1*) pathway in these patients. Genetic alterations of *HSD17B3* and *SRD5A2* play a role only in 46,XY male development; in females with homozygous or compound heterozygous mutations, no abnormality in sex development is seen and these women are presumably fertile.

It has recently been demonstrated that 5-alpha reduction of testosterone is not the only pathway of DHT synthesis. Via the so-called backdoor pathway DHT may be produced from androstanediol, a compound that is converted through several steps directly from 17-alpha-hydroxy progesterone and requires CYP17A1 and 17-beta HSD type 3 and type 5, with the latter also known as 3alpha hydroxysteroid dehydrogenase type 2. This alternative mode for DHT synthesis could explain some of the virilization in patients with congenital adrenal hyperplasia due to 21 hydroxylase deficiency, because of the very high substrate levels of 17-OH progesterone [87] (Figure 7.2.1.3).

In humans and other mammals, androgens act via a single androgen receptor (AR) in a very specific manner. Both sexes express the AR, therefore the appropriate levels of androgens promote male sex development of an individual. The AR is a single-copy gene localized on the X-chromosome at Xq11–12. Due to the hemizygous state of the gene in 46,XY individuals, mutations in the AR gene directly affect male sex development. The AR acts as a typical nuclear receptor via transcriptional regulation of defined target genes. The genetic and functional structure of the AR is of interest as numerous studies have been performed to explain a genotype-phenotype correlation in androgen insensitivity syndrome (OMIM #300068), but also in conditions relating to other functions of the AR, such as spinobulbar muscular atrophy (SBMA, OMIM #313200) or the role of the AR in prostate cancer. The gene is composed of 8 exons, where the first large exon encodes for the variable N-terminal domain, exons 2 and 3 for the DNA-binding domain, and exons 5–8 for the ligand-binding domain. The N-terminal domain contains two variable repeat regions, a CAG and a GGN repeat, which have been associated with aspects of androgen action. Elongation of the CAG-repeat has been associated with sub- or infertility [88], and very expanded CAG repeats may lead to spinal and bulbar muscular atrophy [89].

The AR initially resides in the cytoplasm and is bound to a complex of heat-shock proteins, chaperones, and co-chaperones. Ligand-binding induces a conformational switch, involving N- and C-terminal interaction, which eventually unmasks a nuclear localization signal. The ligand-bound AR translocates rapidly into the nucleus, which again may involve a number of heat-shock proteins and immunophilins. Within the nucleus, the AR forms a homodimer and binds to androgen-response elements within the target DNA. Some of these elements may also be recognized by other steroid receptors, but the AR can bind to selective androgen-response elements, also involving specific co-regulator proteins, which enhance

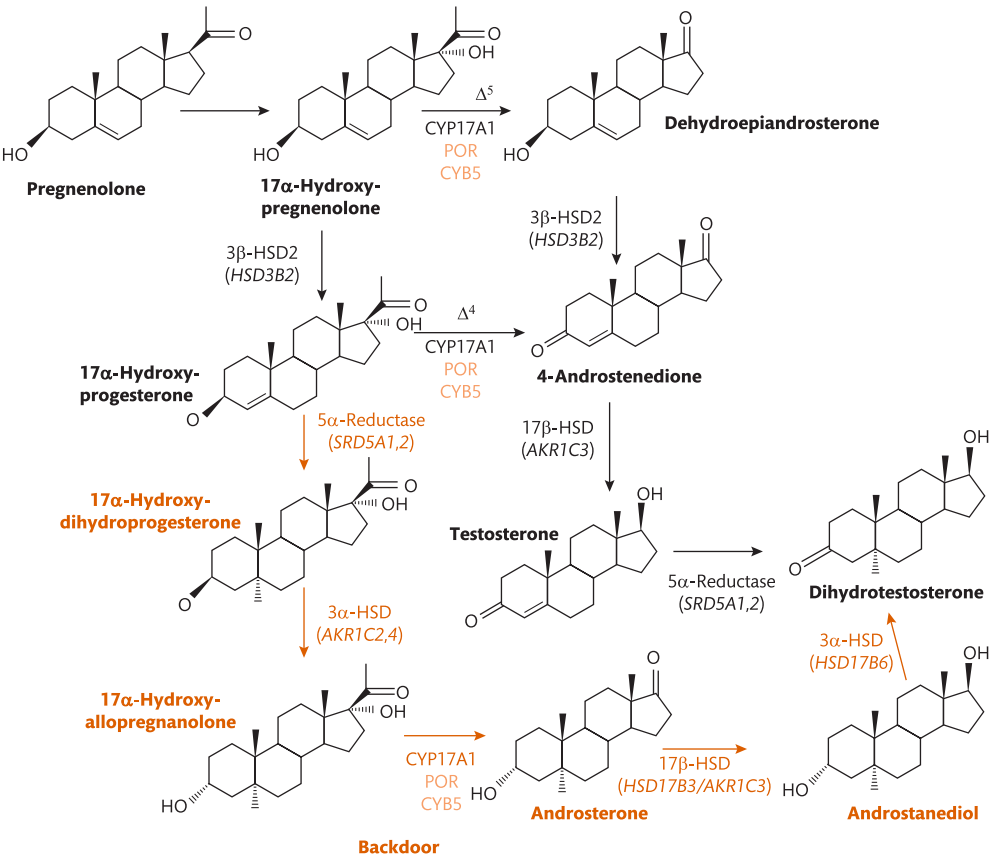


Figure 7.2.1.3 The 'backdoor' pathway of dihydrotestosterone synthesis.

or supress AR-dependent transcription of target genes. Many co-factors possess enzymatic activities that lead to post-translational modifications including phosphorylation, acetylation, methylation, ubiquitination, sumoylation, and ADP-ribosylation. More than 300 different AR-interacting proteins have been described, but their role in DSD conditions remains unclear. Analysis of AR-dependent transcription of target genes demonstrates that cell-specific classes of genes may be up- or downregulated.

Interestingly, different androgens can elicit different effects via the single AR, as has been shown in elegant cellular model studies [90]. This is also seen in normal sex development with the different actions of testosterone and DHT or even androstenedione and is exemplified by DSD conditions due to mutations in the aforementioned enzymatic pathways.

This may be due to the different binding capacities of the various androgens, but also due to cellular metabolism within the target cells, as demonstrated by DHT synthesis in cells harbouring the 5-alpha reductase enzyme. Furthermore, AR action is, in common with other nuclear receptors, dependent on cell- or tissue-specific coregulators, which may also be expressed in a time-dependent manner [91]. Hence, androgen-related genital and sex development will arise from very specific gene expression patterns regulated through the AR [92, 93]. This effect is irreversible, and genetic signatures will be present throughout life representing the sexual phenotype of an individual. Each individual may have a specific 'androgen response index', which may reflect the variability in all humans, but may be specifically altered in people with DSD conditions [76].

The Role of Oestrogens in Prenatal Development

Currently, little is known about the role oestrogens play in prenatal sex development. We know from patients with complete gonadal dysgenesis, who presumably have no gonadal function and therefore should have a lack of systemic oestrogen synthesis, that sex development is comparable to other 46,XX females. However, the fetus is usually exposed to oestrogenic steroids via the placenta and therefore should receive some general effects through these compounds. Some endocrine disrupting compounds are actively enriched in the fetus, such as bisphenol A, and we are now recognizing their effects both on fertility and sexual behaviour of exposed fetuses [94].

Brain Sexual Differentiation

Mammalian sex development is not restricted to genital tissues. Functional studies in mouse models demonstrate that up to three quarters of all genes will be transcribed in a sex-dimorphic expression pattern [95]. This includes also the brain, which is of course of high interest in DSD, because this will influence the psychological aspects of sex including gender identity, gender role behaviour and sexual orientation. The latter are developed through biological factors (genes and hormones) as well as through modification of psychological, social, and cultural factors. In part, a sex-specific gene expression signature is already present in the brain before gonadal differentiation sets in [96]; however, for more than 30 years the additional influence of perinatal androgens on the sex-specific development of certain brain regions has been recognized [97]. Sex differences may characterize specific neural regions, such as the pre-optic area, the bed nucleus of the stria terminalis, the medial nucleus of the amygdala, the ventromedial nucleus of the hypothalamus, the arcuate nucleus, and the cerebral cortex.

In general, sex steroids will play an activating and an organizational role in brain and behavioural development. While activating influences are transient and correspond to a waxing and waning of hormone levels (e.g. associated with sexual interest), organizational aspects are permanent and will, as in sex determination and development in general, not be changed throughout life. Significant differences in DNA methylation are encountered between males, females, and also masculinized females [98]. This aspect is also the core understanding of those increasing voices that describe people with DSD living as intersex persons, and has already led to the recommendation not to assign a usual sex to affected persons or even to develop further positions on variant sex assignment.

Perinatal hormone exposure may influence behavioural patterns. For instance, if female rats are treated with testosterone during the first days of life, they lose the typical female reproductive behaviour and instead develop male behavioural patterns. This demonstrates that early post-natal androgens may influence permanently organizational structures in the brain. In contrast, peripheral oestrogens may play only a minor role in female-typical behavioural patterns, with levels modulating qualitative behavioural aspects. However, there are suggestions that testosterone may pass the blood–brain barrier and may be aromatized in the brain to local oestrogens, which then promotes masculinizing brain effects [98].

Throughout life, the diminishing of sex hormone levels may lead to different peripheral and central patterns of change in stress response systems in male and female rats [99], thereby implying a sex difference for different mood disorders like depression.

Conclusion

Recently it has become increasingly clear that sex development is not a simple pathway leading to robust different poles of male and female. Rather, sex development starts from a common trunk allowing numerous alterations of phenotype through genetic and endocrine programming and is also susceptible to cultural and social influences in humans. While the identification of the effects of sex and gender factors from molecular to genomic sciences has led to recognition of male and female differences in health and disease, a general appreciation and acceptance of broad variability within and between the sexes is only partially acknowledged at this time. It will need further understanding of the developmental processes and a linkage with social sciences to comprehend the continuum of human sex development.

REFERENCES

1. Tartaglia N, Ayari N, Howell S, D'Epagnier C, Zeitler P. 48,XXYY, 48,XXXY and 49,XXXXY syndromes: not just variants of Klinefelter syndrome. *Acta Paediatr*, 2011; **100**: 85–60.
2. Spiller C, Koopman P, Bowles J. Sex determination in the mammalian germline. *Annu Rev Genet*, 2017; **51**: 265–85.
3. Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol*, 2005; **278**: 440–58.

4. Hu YC, Okumura LM, Page DC. Gata4 is required for formation of the genital ridge in mice. *PLoS Genet*, 2013; **9**: e1003629.
5. Armstrong JF, Pritchard-Jones K, Bickmore WA, Hastie ND, Bard JB. The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. *Mech Dev*, 1993; **40**: 85–97.
6. Hammes A, Guo JK, Lutsch G, *et al*. Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell*, 2001; **106**: 319–29.
7. Crispino JD, Lodish MB, Thurberg BL, *et al*. Proper coronary vascular development and heart morphogenesis depend on interaction of GATA-4 with FOG cofactors. *Genes Dev*, 2001; **15**: 839–44.
8. Kusaka M, Katoh-Fukui Y, Ogawa H, *et al*. Abnormal epithelial cell polarity and ectopic epidermal growth factor receptor (EGFR) expression induced in Emx2 KO embryonic gonads. *Endocrinology*, 2010; **151**: 5893–904.
9. Miyamoto N, Yoshida M, Kuratani S, Matsuo I, Aizawa S. Defects of urogenital development in mice lacking Emx2. *Development*, 1997; **124**: 1653–64.
10. Piard J, Mignot B, Arbez-Gindre F, *et al*. Severe sex differentiation disorder in a boy with a 3.8 Mb 10q25.3-q26.12 microdeletion encompassing EMX2. *Am J Med Genet A*, 2014; **164A**: 2618–22.
11. Birk OS, Casiano DE, Wassif CA, *et al*. The LIM homeobox gene Lhx9 is essential for mouse gonad formation. *Nature*, 2000; **403**: 909–13.
12. Luo X, Ikeda Y, Parker KL. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*, 1994; **77**: 481–90.
13. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. Male development of chromosomally female mice transgenic for Sry. *Nature*, 1991; **351**: 117–21.
14. Vidal VP, Chaboissier MC, de Rooij DG, Schedl A. Sox9 induces testis development in XX transgenic mice. *Nat Genet*, 2001; **28**: 216–7.
15. Barrionuevo F, Bagheri-Fam S, Klattig J, *et al*. Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod*, 2006; **74**: 195–201.
16. Svingen T, Koopman P. Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes Dev*, 2013; **27**: 2409–26.
17. Larney C, Bailey TL, Koopman P. Switching on sex: transcriptional regulation of the testis-determining gene Sry. *Development*, 2014; **141**: 2195–205.
18. Hanley NA, Hagan DM, Clement-Jones M, *et al*. SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development. *Mech Dev*, 2000; **91**: 403–7.
19. Wilhelm D, Washburn LL, Truong V, Fellous M, Eicher EM, Koopman P. Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice. *Mech Dev*, 2009; **126**: 324–36.
20. Schmahl J, Eicher EM, Washburn LL, Capel B. Sry induces cell proliferation in the mouse gonad. *Development*, 2000; **127**: 65–73.
21. Sekido R, Lovell-Badge R. Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, 2008; **453**: 930–4.
22. Knower KC, Kelly S, Ludbrook LM, *et al*. Failure of SOX9 regulation in 46XY disorders of sex development with SRY, SOX9 and SF1 mutations. *PloS One*, 2011; **6**: e17751.
23. Kim Y, Kobayashi A, Sekido R, *et al*. Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. *PLoS Biol*, 2006; **4**: e187.
24. Kim Y, Bingham N, Sekido R, Parker KL, Lovell-Badge R, Capel B. Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination. *Proc National Acad Sci U S A*, 2007; **104**: 16558–63.
25. Bagheri-Fam S, Sim H, Bernard P, *et al*. Loss of Fgfr2 leads to partial XY sex reversal. *Dev Biol*, 2008; **314**: 71–83.
26. Siggers P, Carre GA, Bogani D, *et al*. A novel mouse Fgfr2 mutant, hobbyhorse (hob), exhibits complete XY gonadal sex reversal. *PloS One*, 2014; **9**: e100447.
27. Colvin JS, Green RP, Schmahl J, Capel B, Ornitz DM. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell*, 2001; **104**: 875–89.
28. Bagheri-Fam S, Ono M, Li L, *et al*. FGFR2 mutation in 46,XY sex reversal with craniosynostosis. *Hum Mol Genet*, 2015; **24**: 6699–710.
29. Wilhelm D, Hiramatsu R, Mizusaki H, *et al*. SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development. *J Biol Chem*, 2007; **282**: 10553–60.
30. Wilhelm D, Martinson F, Bradford S, *et al*. Sertoli cell differentiation is induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. *Dev Biol*, 2005; **287**: 111–24.
31. Malki S, Nef S, Notarnicola C, *et al*. Prostaglandin D2 induces nuclear import of the sex-determining factor SOX9 via its cAMP-PKA phosphorylation. *EMBO J*, 2005; **24**: 1798–809.
32. Moniot B, Declosmenil F, Barrionuevo F, *et al*. The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. *Development*, 2009; **136**: 1813–21.
33. Miyamoto Y, Taniguchi H, Hamel F, Silversides DW, Viger RS. A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *BMC Mol Biol*, 2008; **9**: 44.
34. Arango NA, Lovell-Badge R, Behringer RR. Targeted mutagenesis of the endogenous mouse Mis gene promoter: in vivo definition of genetic pathways of vertebrate sexual development. *Cell*, 1999; **99**: 409–19.
35. Giuili G, Shen WH, Ingraham HA. The nuclear receptor SF-1 mediates sexually dimorphic expression of Mullerian inhibiting substance, in vivo. *Development*, 1997; **124**: 1799–807.
36. Skinner MK, Tung PS, Fritz IB. Cooperativity between Sertoli cells and testicular peritubular cells in the production and deposition of extracellular matrix components. *J Cell Biol*, 1985; **100**: 1941–7.
37. Georg I, Barrionuevo F, Wiech T, Scherer G. Sox9 and Sox8 are required for basal lamina integrity of testis cords and for suppression of FOXL2 during embryonic testis development in mice. *Biol Reprod*, 2012; **87**: 99.
38. Jeanes A, Wilhelm D, Wilson MJ, *et al*. Evaluation of candidate markers for the peritubular myoid cell lineage in the developing mouse testis. *Reproduction*, 2005; **130**: 509–16.
39. Ostrer H, Huang HY, Masch RJ, Shapiro E. A cellular study of human testis development. *Sex Dev*, 2007; **1**: 286–92.
40. Clark AM, Garland KK, Russell LD. Desert hedgehog (Dhh) gene is required in the mouse testis for formation of adult-type Leydig cells and normal development of peritubular cells and seminiferous tubules. *Biol Reprod*, 2000; **63**: 1825–38.

41. Bitgood MJ, Shen L, McMahon AP. Sertoli cell signaling by Desert hedgehog regulates the male germline. *Curr Biol*, 1996; **6**: 298–304.
42. Yao HH, Whoriskey W, Capel B. Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev*, 2002; **16**: 1433–40.
43. Umehara F, Tate G, Itoh K, *et al*. A novel mutation of desert hedgehog in a patient with 46,XY partial gonadal dysgenesis accompanied by minifascicular neuropathy. *Am J Hum Genet*, 2000; **67**: 1302–5.
44. Werner R, Merz H, Birnbaum W, *et al*. 46,XY Gonadal dysgenesis due to a homozygous mutation in desert hedgehog (DHH) identified by exome sequencing. *J Clin Endocrinol Metab*, 2015; **100**: E1022–9.
45. Tajouri A, Kharrat M, Hizem S, *et al*. In vitro functional characterization of the novel DHH mutations p.(Asn337Lysfs*24) and p.(Glu212Lys) associated with gonadal dysgenesis. *Hum Mutat*, 2018; **39**: 2097–109.
46. Barsoum IB, Bingham NC, Parker KL, Jorgensen JS, Yao HH. Activation of the Hedgehog pathway in the mouse fetal ovary leads to ectopic appearance of fetal Leydig cells and female pseudohermaphroditism. *Dev Biol*, 2009; **329**: 96–103.
47. Ye L, Li X, Li L, Chen H, Ge RS. Insights into the development of the adult leydig cell lineage from stem Leydig cells. *Front Physiol*, 2017; **8**: 430.
48. Shima Y, Miyabayashi K, Haraguchi S, *et al*. Contribution of Leydig and Sertoli cells to testosterone production in mouse fetal testes. *Mol Endocrinol*, 2013; **27**: 63–73.
49. Teerds KJ, Huhtaniemi IT. Morphological and functional maturation of Leydig cells: from rodent models to primates. *Hum Reprod Update*, 2015; **21**: 310–28.
50. Fowler PA, Bhattacharya S, Gromoll J, Monteiro A, O'Shaughnessy PJ. Maternal smoking and developmental changes in luteinizing hormone (LH) and the LH receptor in the fetal testis. *J Clin Endocrinol Metab*, 2009; **94**: 4688–95.
51. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev*, 2000; **21**: 551–83.
52. Weiss J, Axelrod L, Whitcomb RW, Harris PE, Crowley WF, Jameson JL. Hypogonadism caused by a single amino acid substitution in the beta subunit of luteinizing hormone. *N Engl J Med*, 1992; **326**: 179–83.
53. Potorac I, Rivero-Muller A, Trehan A, *et al*. A vital region for human glycoprotein hormone trafficking revealed by an LHB mutation. *J Endocrinol*, 2016; **231**: 197–207.
54. Windley SP, Wilhelm D. Signaling pathways involved in mammalian sex determination and gonad development. *Sex Dev*, 2015; **9**: 297–315.
55. Vainio S, Heikkila M, Kispert A, Chin N, McMahon AP. Female development in mammals is regulated by Wnt-4 signalling. *Nature*, 1999; **397**: 405–9.
56. Chassot AA, Ranc F, Gregoire EP, *et al*. Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. *Hum Mol Genet*, 2008; **17**: 1264–77.
57. Biason-Lauber A, Konrad D, Navratil F, Schoenle EJ. A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman. *N Engl J Med*, 2004; **351**: 792–8.
58. Mandel H, Shemer R, Borochowitz ZU, *et al*. SERKAL syndrome: an autosomal-recessive disorder caused by a loss-of-function mutation in WNT4. *Am J Hum Genet*, 2008; **82**: 39–47.
59. Parma P, Radi O, Vidal V, *et al*. R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat Genet*, 2006; **38**: 1304–9.
60. Tomaselli S, Megiorni F, De Bernardo C, *et al*. Syndromic true hermaphroditism due to an R-spondin1 (RSPO1) homozygous mutation. *Hum Mutat*, 2008; **29**: 220–6.
61. Jordan BK, Mohammed M, Ching ST, *et al*. Up-regulation of WNT-4 signaling and dosage-sensitive sex reversal in humans. *Am J Hum Genet*, 2001; **68**: 1102–9.
62. Liu CF, Bingham N, Parker K, Yao HH. Sex-specific roles of beta-catenin in mouse gonadal development. *Hum Mol Genet*, 2009; **18**: 405–17.
63. Maatouk DM, DiNapoli L, Alvers A, Parker KL, Taketo MM, Capel B. Stabilization of beta-catenin in XY gonads causes male-to-female sex-reversal. *Hum Mol Genet*, 2008; **17**: 2949–55.
64. Jameson SA, Lin YT, Capel B. Testis development requires the repression of Wnt4 by Fgf signaling. *Dev Biol*, 2012; **370**: 24–32.
65. Lavery R, Chassot AA, Pauper E, *et al*. Testicular differentiation occurs in absence of R-spondin1 and Sox9 in mouse sex reversals. *PLoS Genet*, 2012; **8**: e1003170.
66. Nicol B, Yao HH. Gonadal identity in the absence of pro-testis factor SOX9 and pro-ovary factor beta-catenin in mice. *Biol Reprod*, 2015; **93**: 35.
67. Uhlenhaut NH, Jakob S, Anlag K, *et al*. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell*, 2009; **139**: 1130–42.
68. Crisponi L, Deiana M, Loi A, *et al*. The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet*, 2001; **27**: 159–66.
69. Bashamboo A, Eozenou C, Jorgensen A, *et al*. Loss of function of the nuclear receptor NR2F2, encoding COUP-TF2, causes testis development and cardiac defects in 46,XX children. *Am J Hum Genet*, 2018; **102**: 487–93.
70. Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ, Zarkower D. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature*, 2011; **476**: 101–4.
71. Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev*, 2000; **14**: 2587–95.
72. Ledig S, Hiort O, Wunsch L, Wieacker P. Partial deletion of DMRT1 causes 46,XY ovotesticular disorder of sexual development. *Eur J Endocrinol*, 2012; **167**: 119–24.
73. Muroya K, Okuyama T, Goishi K, *et al*. Sex-determining gene(s) on distal 9p: clinical and molecular studies in six cases. *J Clin Endocrinol Metab*, 2000; **85**: 3094–100.
74. Welsh M, Suzuki H, Yamada G. The masculinization programming window. *Endocr Dev*, 2014; **27**: 17–27.
75. Staack A, Donjacour AA, Brody J, Cunha GR, Carroll P. Mouse urogenital development: a practical approach. *Differentiation*, 2003; **71**: 402–13.
76. Hiort O. The differential role of androgens in early human sex development. *BMC Med*, 2013; **11**: 152.
77. Miller WL. Why nobody has P450scc (20,22 desmoslase) deficiency. *J Clin Endocrinol Metab*, 1998; **83**: 1399–400.
78. Hiort O, Holterhus PM, Werner R, *et al*. Homozygous disruption of P450 side-chain cleavage (CYP11A1) is associated with prematurity, complete 46,XY sex reversal, and severe adrenal failure. J Clin Endocrinol Metab, 2005; **90**: 538–41.
79. Tee MK, Abramsohn M, Loewenthal N, *et al*. Varied clinical presentations of seven patients with mutations in CYP11A1

encoding the cholesterol side-chain cleavage enzyme, P450scc. *J Clin Endocrinol Metab*, 2013; **98**: 713–20.
80. Hauffa B, Hiort O. P450 side-chain cleavage deficiency--a rare cause of congenital adrenal hyperplasia. *Endocr Dev*, 2011; **20**: 54–62.
81. Prader A. [Genital findings in the female pseudo-hermaphroditism of the congenital adrenogenital syndrome; morphology, frequency, development and heredity of the different genital forms]. *Helv Paediatr Acta*, 1954; **9**: 231–48.
82. Korsch E, Peter M, Hiort O, *et al*. Gonadal histology with testicular carcinoma in situ in a 15-year-old 46,XY female patient with a premature termination in the steroidogenic acute regulatory protein causing congenital lipoid adrenal hyperplasia. *J Clin Endocrinol Metab*, 1999; **84**: 1628–32.
83. Idkowiak J, Randell T, Dhir V, *et al*. A missense mutation in the human cytochrome b5 gene causes 46,XY disorder of sex development due to true isolated 17,20 lyase deficiency. *J Clin Endocrinol Metab*, 2012; **97**: E465–75.
84. Werner R, Kulle A, Sommerfeld I, *et al*. Testosterone synthesis in patients with 17beta-hydroxysteroid dehydrogenase 3 deficiency. *Sex Dev*, 2012; **6**: 161–8.
85. Hiort O, Marshall L, Birnbaum W, *et al*. Pubertal development in17Beta-hydroxysteroid dehydrogenase type 3 deficiency. *Hormone Research in Paediatrics*, 2016; **87**. DOI: 10.1159/000453613.
86. Hiort O, Sinnecker GH, Willenbring H, Lehners A, Zollner A, Struve D. Nonisotopic single strand conformation analysis of the 5 alpha-reductase type 2 gene for the diagnosis of 5 alpha-reductase deficiency. *J Clin Endocrinol Metab*, 1996; **81**: 3415–8.
87. Kamrath C, Hochberg Z, Hartmann MF, Remer T, Wudy SA. Increased activation of the alternative 'backdoor' pathway in patients with 21-hydroxylase deficiency: evidence from urinary steroid hormone analysis. *J Clin Endocrinol Metab*, 2012; **97**: E367–75.
88. Hiort O, Werner R, Zitzmann M. Pathophysiology of the androgen receptor. In: Nieschlag E, Behre HM, eds. *Testosterone: Action, Deficiency, Substitution*, 4th edn. Cambridge, UK: Cambridge University Press, 2012: 33–59.
89. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 1991; **352**: 77–9.
90. Holterhus PM, Piefke S, Hiort O. Anabolic steroids, testosterone-precursors and virilizing androgens induce distinct activation profiles of androgen responsive promoter constructs. *J Steroid Biochem Mol Biol*, 2002; **82**: 269–75.
91. Bebermeier JH, Brooks JD, Deprimo SE, *et al*. Cell-line and tissue-specific signatures of androgen receptor-coregulator transcription. *J Mol Med*, 2006; **84**: 919–31.
92. Holterhus PM, Hiort O, Demeter J, Brown PO, Brooks JD. Differential gene-expression patterns in genital fibroblasts of normal males and 46,XY females with androgen insensitivity syndrome: evidence for early programming involving the androgen receptor. *Genome Biol*, 2003; **4**: R37.
93. Holterhus PM, Deppe U, Werner R, *et al*. Intrinsic androgen-dependent gene expression patterns revealed by comparison of genital fibroblasts from normal males and individuals with complete and partial androgen insensitivity syndrome. *BMC Genomics*, 2007; **8**: 376.
94. McLachlan JA. Environmental signaling: from environmental estrogens to endocrine-disrupting chemicals and beyond. *Andrology*, 2016; **4**: 684–94.
95. Yang X, Schadt EE, Wang S, *et al*. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res*, 2006; **16**: 995–1004.
96. Dewing P, Shi T, Horvath S, Vilain E. Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. *Brain Res Mol Brain Res*, 2003; **118**: 82–90.
97. Döhler KD, Coquelin A, Davis F, Hines M, Shryne JE, Gorski RA. Differentiation of the sexually dimorphic nucleus in the preoptic area of the rat brain is determined by the perinatal hormone environment. *Neurosci Lett*, 1982; **33**: 295–8.
98. McCarthy MM, Nugent BM, Lenz KM. Neuroimmunology and neuroepigenetics in the establishment of sex differences in the brain. *Nat Rev Neurosci*, 2017; **18**: 471–84.
99. Guo L, Chen YX, Hu YT, *et al*. Sex hormones affect acute and chronic stress responses in sexually dimorphic patterns: consequences for depression models. *Psychoneuroendocrinology*, 2018; **95**: 34–42.

7.2.2 Disorders of Sex Development (DSD) in the Newborn

S. Faisal Ahmed and Salma R. Ali

Introduction

The birth of a new baby is one of the greatest wonders of nature and one of the most exciting events known to humankind. The first question that is usually posed by the new parent is 'is it a boy or a girl?'; without this information the parents cannot even formulate the second question which is usually 'is he/she alright?'. It is no wonder that the birth of a child with an abnormality of genital development where the sex of rearing is uncertain at birth, presents difficult clinical and ethical issues. However, the recognition of genital ambiguity may depend on the expertise of the observer. Whilst the prevalence of genital anomalies at birth may be as high as 1 in 300 births [1], the prevalence of complex anomalies that may lead to true genital ambiguity may be as low as 1 in 5000 births [2]. Rather than treating every affected child as a medical emergency, it is paramount that such a child is assessed by an expert with adequate knowledge about the range of variation in the physical appearance of genitalia, the underlying pathophysiology of disorders of sex development and the strengths and weaknesses of the tests that can be performed in early infancy. This expert should ensure that the parents' needs for information are comprehensively addressed while appropriate investigations are performed in a timely fashion. This expert needs to have immediate access to the multidisciplinary team that is essential for the management of such a child. In the field of rare conditions, it is imperative that the clinician shares the experience with others through national and international clinical and research collaboration. The I-DSD registry is an example of how registries can facilitate international collaboration to address issues ranging from fundamental mechanisms to clinical outcomes [3].

Terminology

The use of terminology which is clear, easy to use and understood by all health professionals, patients and their families is fundamental to the management of affected newborns and children. In addition, terminology should respect the individual and avoid terms which might cause offense. The term 'intersex' has had variable connotations even within professionals; some employed it as a term that covered all affected newborns whilst at the other end of the spectrum, some believed that the term should only apply to those where there is complete mismatch between chromosomal and anatomic sex. The consensus reached in 2005 on management of these patients, stressed the importance of the aspect of terminology and recommended substitution of the term 'intersex' with 'disorder of sex development (DSD)' which is defined as any congenital condition in which development of chromosomal, gonadal or anatomic sex is atypical [4]. It also recommended the abandonment of terms such as 'pseudohermaphroditism' and 'true hermaphroditism'. Whilst, the new nomenclature (Table 7.2.2.1) is easier to use and understand, and helps the professional planning investigations, it will nevertheless evolve over time as our understanding of long-term outcome, as well as molecular aetiology, improves. Given that atypical genitalia may occur as commonly as 1 in 300 births and may not always be associated with a functional abnormality, some have advocated the use of 'differences' in preference to the term 'disorder' [5, 6]. The strength of the acronym 'DSD' is that it can be used to cover both differences and disorders of sex development. The likelihood of this difference existing as a disorder will depend on the functional implications of the condition which may be heavily influenced by the social and cultural framework within which the child exists [7].

General Principles of Management

Optimal clinical management of infants with DSD should comprise the following principles:

- All newborn infants should receive a male or female sex assignment.
- When there is any doubt about sex assignment, a hasty decision must be avoided prior to expert evaluation.
- Whilst all specialist neonatal units should be expected to be able to stabilize the critically unwell infant with a DSD, comprehensive evaluation and the development of a plan for long-term management must be performed at a specialist centre with an experienced multidisciplinary team.
- The specialist centre should be able to complete first-line investigations quickly which are sufficient for deciding sex assignment and excluding immediate medical concerns. The centre should be able to develop a plan for second-line investigations that will guide long-term management.
- Management should be patient-centred, holistic, and as far as possible, evidence-based. Decisions which are not evidence-based should be explained to the family.
- Patient and family concerns should be respected and addressed in strict confidence.
- Open communication with patients and families is essential and participation in decision-making is encouraged. The multidisciplinary specialist team should have the ability to arrange, or preferably, provide, long-term care from infancy to adulthood in the affected individual.

Communication

The initial contact with the parents of a child with a DSD is important as first impressions from these encounters often persist. A key point to emphasize is that the child with a DSD has the potential to become a well-adjusted, functional member of society. The use of the phrase 'differences in sex development' may be particularly beneficial in introducing the concept of the range of variation in sex development that can be encountered to those with little prior knowledge of the field. It should be emphasized that DSD is not shameful. In those cases where there are no doubts about sex assignment, it should not be assumed that the parents' need for information and psychological help are any less (Table 7.2.2.2). The parents' perception of risk may be quite different from the clinical perception of the severity of illness [8]. In those cases where there is true genital ambiguity, it should be explained to the parents that the best course of action may not initially be clear, but the healthcare team will work with the family to reach the best possible set of decisions in the circumstances. The healthcare team should discuss with

Table 7.2.2.1 The classification of DSD

Disorder of gonadal development	Disorder of androgen synthesis	Disorder of androgen action	Disorder of androgen excess	Leydig cell defect	Persistent Müllerian duct syndrome	Defects of Müllerian development	Non-specific disorder of undermasculinization	Other
Complete gonadal dysgenesis Partial gonadal dysgenesis Gonadal regression Ovotesticular DSD Testicular DSD Other	StAR def P450 scc def (CYP11A1) 3β-HSD def (HSD3B2) CYP17 def (P450CYP17) 17βHSD def (HSD17B3) 5α reductase def (SRD5A2) P450 oxidoreductase def (POR) Other	PAIS CAIS Other	21αhydroxylase def (CYP21A) 11βhydroxylase def (CYP11B1) Aromatase def (CYP19A1) P450 oxidoreductase def (POR) Maternal androgens Other	Leydig cell hypoplasia LH deficiency Other	AMH low AMH normal AMH not known	MURCS MRKH Uterine Didelphys Other	Isolated hypospadias Isolated bilateral cryptorchidism Isolated micropenis Anomalies EMS >8 Anomalies EMS 5–8 Anomalies EMS <5	Cloacal Anomaly Bladder exstrophy Smith Lemli Opitz Synd Other

the parents what information to share in the early stages with family members and friends. It is essential that the parents do not register the birth until the sex of rearing is established. Parents need to be informed about sex development; they should be provided with written information and directed to internet-based information (Box 7.2.2.1). Ample time and opportunity should be made for continued discussion with review of information previously provided.

The Multidisciplinary Team

Optimal care for children with DSD requires an experienced multidisciplinary team that is generally found in regional centres. The team may exist as a clinical network with links to other children's centres (Scottish DSD Network, https://www.sdsd.scot.nhs.uk). Ideally, the team includes paediatric subspecialists in endocrinology, surgery and/or urology, psychology/psychiatry, gynaecology, genetics, neonatology, nursing and, if possible, social work, and medical ethics. Core composition will vary according to DSD type, local resources, developmental context, and location. The team has a responsibility to educate other healthcare staff in the appropriate initial management of affected newborns and their families and should also have the ability to review and discuss its own performance through audit of clinical activity and attendance at joint clinics and education events. For new infants with a DSD, the team should develop a plan for clinical management with respect to diagnosis, gender assignment and treatment options before making any recommendations. Ideally, ongoing discussions with the family are conducted by one professional with appropriate communication skills. Transitional care should be organized with the multidisciplinary team operating in an environment comprising specialists with experience in both paediatric and adult practice. Support groups have an important role to play, and their contact details should be supplied to the parents, it is possible that affected parents may prefer to talk to local families

Table 7.2.2.2 Themes, subthemes, and percentage of parents who raised the theme during qualitative interview about the parents' own experience during the early years of their affected child's life

Theme & subthemes	%	Theme & subthemes	%
General experience		**Coping strategies**	
Suboptimal initial provision of information at birth	95	Relying on the clinical staff	63
Emotional vulnerability of the mother	68	Treatability of condition	58
Relief on talking to a consultant surgeon	16	'Moving the concern to the back of the mind'	53
		Comparison of child with another who is 'worse'	37
Handling the subject of genital anomalies		Lack of pain	32
Ridicule and stigma	68	Comparison of child with another in similar situation	32
Difficult subject to discuss with parents/relatives	63	'Getting on with it'	11
Difficult subject to discuss with friends	63	'Focusing on the positive'	11
Appropriate level of sensitivity as inpatients	26	Less of a concern if more serious anomalies present	11
Support for parents to discuss condition with child	21		
Professional need for sensitivity when teaching	16	**Impact of condition on child & family**	
		Non-specific concerns about anaesthetic & surgery	84
Concomitant stressors		Sexual function & fertility	84
Complications associated with delivery and prematurity	32	Unclear about post-operative appearance of genitalia	58
Problems with cognitive or social development	11	Special care after surgery	47
Other offspring with medical conditions	11	Need for more than one operation	37
Recent bereavements	11	Risk of recurrence	26
Marital disharmony	11	Pain following surgery	22
		Delay in surgery and likelihood of ridicule in school	11
Sources of support			
Relatives – helpful	74	**Suggestions on improving service**	
Consultant surgeons	74	Local network of affected families	42
Other healthcare staff	42	Information on cleaning genitalia	42
Relatives—not helpful	26	Information on postop care of urinary catheter	26
Parents of other affected children	16	Gradual and steady provision of information	32
Health visitors	16	Images of average outcomes	32
General practitioners	16	Recommended websites	32
		Link person for family support at presentation	21
		Pain control following surgery	11

Box 7.2.2.1 Some examples of online information on DSD for patients, parents, and professionals

General Information about sex development
- UK Intersex Association—http://www.ukia.co.uk
- Intersex Society of North America—https://www.isna.org
- Child physiology—https://www.sickkids.ca

Congenital adrenal hyperplasia
- Congenital Adrenal Hyperplasia Education & Support Network—https://www.congenitaladrenalhyperplasia.org
- Children Living with inherited Metabolic Diseases (CLIMB) Congenital Adrenal Hyperplasia UK Support Group— https://www.livingwithcah.com
- Adrenal Hyperplasia Network—http://www.ahn.org.uk

Androgen insensitivity syndrome
- Androgen Insensitivity Syndrome Support Group—http://www.aissg.org
- eMedicine—http://emedicine.medscape.com/article/924996-overview
- Complete androgen insensitivity syndrome—https://www.rch.org.au/endo/cais/Complete_androgen_insensitivity_syndrome/

XY/XO gonadal dysgenesis
- Turner Syndrome Support Society—http://tss.org.uk/

Hypospadias
- Hypospadias UK Trust—https://www.hypospadiasuk.co.uk

Clinical networks
- The Scottish DSD Network—https://www.sdsd.scot.nhs.uk
- Netzwerk Intersexualitat—http://www.uksh.de/kinderhormonzentrum-luebeck/Forschung/DSDnet.html

Registries
- I-DSD Registry—http://www.i-dsd.org
- I-CAH Registry—http://www.i-cah.org

Consensus views
- CAH Clinical Practice Guideline—The Endocrine Society https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2936060/
- Consensus Statement on Management of Intersex Disorders – https://www.pedsendo.org/education_training/healthcare_providers/consensus_statements/assets/DSDconsensusPediatrics2006.pdf

Medical and genetic overview
- Medline Plus—https://www.nlm.nih.gov/
- Genecard—https://www.genecards.org

affected in a similar way. The availability of such a local pool of voluntary helpers who had some support from the specialists would complete the composition of the multidisciplinary team.

Clinical Evaluation of the Infant with a Suspected DSD

The infant with a suspected DSD may need evaluation for a few reasons. Firstly, there may be a need to determine the sex of rearing. Secondly, there may be concerns about immediate, life-threatening metabolic conditions that are more likely to be associated with certain diagnoses that are, for instance, associated with adrenal insufficiency. Thirdly, an improved knowledge of the aetiology of the underlying condition may allow the development of a long-term management plan. Fourthly, continued evaluation over the longer-term will allow the affected individuals and their care-providers to understand issues such as fertility, sexual function and the risk of tumour development and help with informed disclosure of the diagnosis itself. Finally, the process of evaluation also allows the creation of rapport with the patient and the parents.

Initial Approach

It is very likely that the clinician from whom a specialist opinion is sought will encounter the infant and the parents after the family have already been seen by other health professionals. Their anticipation of meeting someone who can answer all their questions, provide them with reassurance and solve all the problems can be a daunting and impossible task for a single clinician, irrespective of their level of expertise. It is likely that this clinician will form a long-lasting relationship with this family and over time with the help of the multidisciplinary team will be able to address most of the issues mentioned. It is, therefore, very important to have a positive and systematic approach for the first encounter with the family with emphasis on the general well-being of the child.

History

An adequate history should concentrate particularly on:

Family history: Parental consanguinity, history of an infant with salt-losing, unexplained infant deaths, or DSD in relatives. These elements may indicate autosomal recessive genetic disorders associated with disturbed steroidogenesis (usually CAH). In contrast, an X-linked recessive mode of inheritance is suggestive of androgen insensitivity syndrome (AIS).

Antenatal history: Maternal ingestion of drugs which may cause fetal virilization (androgens), or signs of maternal androgen excess which may indicate a maternal androgen secreting tumour; Exposure to specific environmental factors able to inhibit virilization of the fetus. Some assisted conception techniques include progestogen-containing drugs and these methods increase the likelihood of male offspring with genital anomalies.

Information about antenatal counselling and results of prenatal tests: Knowledge of what has already been discussed with the parents by health professionals, and their understanding of the information, is essential.

Social history: An enquiry about the family social network, parents' general understanding of DSD and their current concerns.

General Examination

The general physical examination should determine whether there are any dysmorphic features and the general health of the baby. Affected infants, particularly those who have XY DSD, are more likely to be small for gestational age and may display other developmental anomalies [1, 9, 10]. In addition to a systematic examination, the affected infant should be examined for mid-line defects which may point towards an abnormality of the hypothalamo–pituitary axis. The state of hydration and blood pressure should be assessed as various forms of adrenal steroid biosynthetic defects can be associated with differing degrees of salt loss, varying degrees of masculinization in girls or under-masculinization

in boys, or hypertension. The cardiovascular collapse with salt loss and hyperkalaemia in congenital adrenal hyperplasia does not usually occur until the second week of life (with salt loss usually evident from day 4) and so will not be apparent at birth in a well neonate, however, it should be anticipated in a suspected case. Jaundice (both conjugated and unconjugated) may be observed in cases of hypopituitarism or cortisol deficiency. The urine should be checked for protein as a screen for any associated renal anomaly (e.g. Denys–Drash/Frasier syndromes) and a pre-feed blood glucose should be checked for hypoglycaemia (suggestive of hypopituitarism, or occasionally in CAH e.g. 3βHSD deficiency). Renal tract anomalies such as ureteropelvic junction obstruction, vesicoureteric reflux, pelvic or horseshoe kidney, crossed renal ectopia, and renal agenesis may occur in as many as 1% and 5% of cases with isolated distal and proximal hypospadias, respectively [11].

Examination of the External Genitalia

A detailed physical examination and documentation of the genitalia is necessary to evaluate the degree of genital anomaly. The first step is careful inspection and palpation. In a number of infants, gonads or swellings may be visible in the labioscrotal folds or the inguinal regions but they may disappear on palpation. In those presenting with apparently normal female external genitalia, bilateral hernias containing testes (and, rarely uterus or fallopian tubes) should be sought by palpation. In any case, if gonads are palpated externally, these will be testes (ovaries tend to remain in the pelvic position) or, rarely, ovotestes. A careful measurement of the phallus (stretched dorsal length) and comparison to published normative data [4] is recommended to assess the extent of deviation of the appearance from normal and to explain this difference to the parents. The presence or absence of a chordee should be noted; the location of two (urethral and vaginal) or one orifice (urethral or urogenital sinus) that opens on the dorsal (epispadias) or ventral surface (hypospadias) of the phallic structure should be noted. An epispadias is a very rare condition and is usually part of a spectrum of conditions (bladder and cloacal exstrophy) where there can be a failure of fusion of a number of lower abdominal and pelvic organs including external genitalia. Hypospadias is a much commoner condition where the location of the urethral orifice may be proximal and close to the perineum, mid-shaft or distal and close to the coronal sulcus or the glans. A shorter anogenital and anoscrotal distance may correlate with the severity of hypospadias [12]. The description of the degree of labioscrotal fold fusion, i.e. complete absence of scrotal fusion, a posterior fusion of labia majora, a partially fused hemiscrotum or completely fused scrotum is also very important. Finally, the nature of the skin of the genitalia and labioscrotal folds (texture and pigmentation) and the shape of the folds and whether they are sac-like provides helpful information on androgenization and the possibility of finding testes.

Although scoring systems such as the Prader scoring system for XX DSD [13] and modifications of this system for XY DSD [14, 15] may provide an integrated summary description of the genitalia, these scoring systems are not sufficiently discriminatory to portray the full spectrum of the variation encountered in the external genitalia. The external masculinization score (EMS) which scores external genitalia individually for scrotal fusion, microphallus, location of urethral meatus and location of each gonad may be a more objective method of describing the external appearance [16] (Figure 7.2.2.1).

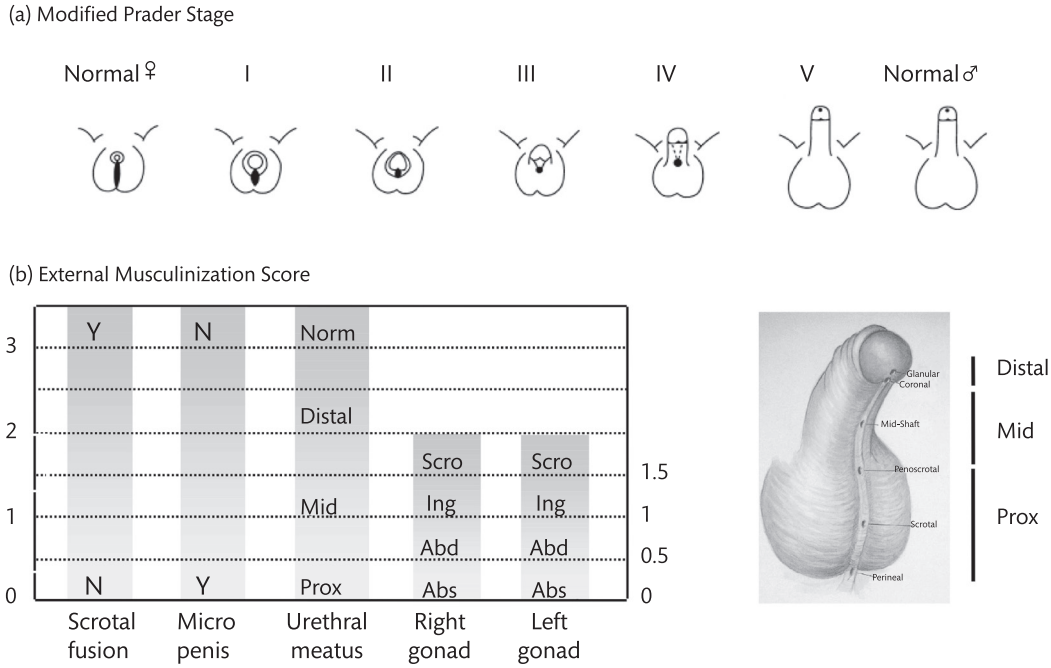


Figure 7.2.2.1 Scoring external genitalia.

Why Investigate

There are clear reasons for investigating an infant with genital anomalies and these include determination of sex of rearing, concerns about early medical problems, concerns about medical and surgical problems in later childhood and development of a long-term plan that anticipates future health issues such as sexual development and function, tumour risk and fertility. A clear knowledge of the underlying aetiology may also facilitate explanation of the condition to the parent and the older child. Thus, investigations should be performed with these different objectives in mind and should be split into first-line and second-line investigations. First-line investigations should, in most cases, be sufficient to guide sex of rearing, exclude early medical problems and provide an idea of the nature of the problem. In the newborn infant, detailed dynamic endocrine investigations should only be performed if they can alter the management of the child; in most cases these investigations can be performed after 3 months when many reproductive and adrenal-related hormones have reached a status quo and the results are easier to interpret. Furthermore, collecting blood samples may be simpler in the older child and collection of multiple blood samples from an otherwise well infant may exert unnecessary stress on the child's parents.

Which Infant Should be Investigated?

Most infants with a suspected DSD will present with:

- overt genital ambiguity
- a family history of DSD such as complete androgen insensitivity syndrome
- a discordance between genital appearance and a prenatal karyotype
- apparent female genitalia with an enlarged clitoris and posterior labial fusion
- apparent female genitalia with an inguinal/labial mass
- apparent male genitalia with bilateral undescended testes
- apparent male genitalia with a microphallus
- apparent male genitalia with proximal hypospadias
- apparent male genitalia with distal or mid-shaft hypospadias with undescended testis

The greatest amount of debate regarding the need for investigation involves the case of the boy presenting with hypospadias and/or cryptorchidism (i.e. the undermasculinized boy). Considerable variation exists about the extent to which these infants should be investigated [17]. Routine systematic examination of 423 consecutive newborn boys in one hospital revealed that 412 (98%) had an EMS of 12. The median (tenth centile) EMS for the group of 11 infants with an EMS of less than 12 was 11 [10, 16]. One infant with isolated micropenis had an EMS of 9; three infants with isolated glandular hypospadias had an EMS of 11; three infants with absent unilateral testis also had an EMS of 11; four infants with a unilateral inguinal testis had an EMS of 11.5. Thus, an EMS of less than 11 was only encountered in 1/423 boys [16]. These data are similar to population data suggesting that genital anomalies occur in about 1:300 total births and 75% of these patients have an associated hypospadias [1]. Population studies also suggest that approximately 50% of hypospadias cases affect the distal penis (glandular or coronal) [18]. The largest study to date of karyotype analysis in children with isolated cryptorchidism, isolated hypospadias or a combination of the two anomalies revealed chromosomal anomalies in 27 of the 916 patients with cryptorchidism (2.94%) and in 7 of the 100 with hypospadias (7%) and in 4 of the 32 with a combination of cryptorchidism and hypospadias (12.5%) [19]. The incidence of chromosome anomalies was 1.8% in cases of isolated cryptorchidism and 6.7% in those with other associated anomalies. In patients with hypospadias, abnormal karyotypes were only detected when there were additional congenital abnormalities. In one specialist centre, out of 63 unselected cases with proximal hypospadias (penoscrotal, scrotal, perineal) who were studied for all known causes of hypospadias with clinical as well as molecular biological techniques, an underlying aetiology was identified in 31% of cases. Of these 31% of cases, 17% were due to complex genetic syndromes, 9.5% were due to chromosomal anomalies and 1 involved the vanishing testes syndrome, the androgen insensitivity syndrome and 5 alpha-reductase type 2 deficiency, respectively [20]. Thus, infants who require further evaluation and investigation should include all children with EMS of less than 11 and all children with familial hypospadias. This will avoid detailed investigations of boys with isolated glandular hypospadias and boys with isolated inguinal testes.

First-Line Investigations

Typically, in the young infant with DSD, gonadal palpability combined with karyotyping, ultrasound examination for Müllerian structures and determination of 17-hydroxyprogesterone level should provide a reasonable guide for the initial practical management of the newborn with a DSD (Figure 7.2.2.2). Karyotype is indicated in all such cases of atypical genitalia and even when prenatal karyotype is available. The results of the karyotype and the ultrasound should be available within 48hrs of presentation. Whilst FISH or PCR analysis using X and Y specific probes is sufficient for initial management it is recommended that these tests are confirmed by a formal karyotype. It also needs to be borne in mind that any mosaicism that is evident may be tissue dependent. Ultrasound is the imaging modality of choice in newborns and if this is difficult to perform, then a laparoscopy should be considered especially as dysgenetic gonads are not clearly delineated by MRI. Genitograms are not as frequently carried out and have been replaced with cystoscopy/genitoscopy in most specialist centres [21, 22].

Due to the effects of stress of labour, and insufficient time for accumulation of hormone in CAH, a sample for 17-hydroxyprogesterone may be difficult to interpret in an infant who is less than 36 hours old. Besides 17-hydroxyprogesterone, biochemistry tests should include serum testosterone, anti-Müllerian hormone (AMH), cortisol, androstenedione, ACTH, gonadotropins, a sample for DNA extraction, and urine for analysis and steroid profile. It is likely that the biochemistry results will be available within a week. Any spare sample should be stored for analysis at a later date. Given that steroid hormones and gonadotrophins fluctuate over the first few weeks of life, serial measurements are particularly valuable. Infants with salt-losing forms of congenital adrenal hyperplasia may start to show biochemical signs of salt loss from day 4, with a rise in potassium being the earliest sign. Whilst it is safest to provide salt and mineralocorticoid where salt loss is suspected, it is also important to establish the diagnosis for long-term management. Urinary electrolytes are unhelpful. Sending a sample for plasma renin activity or measuring renin concentration prior to treatment can be helpful retrospectively,

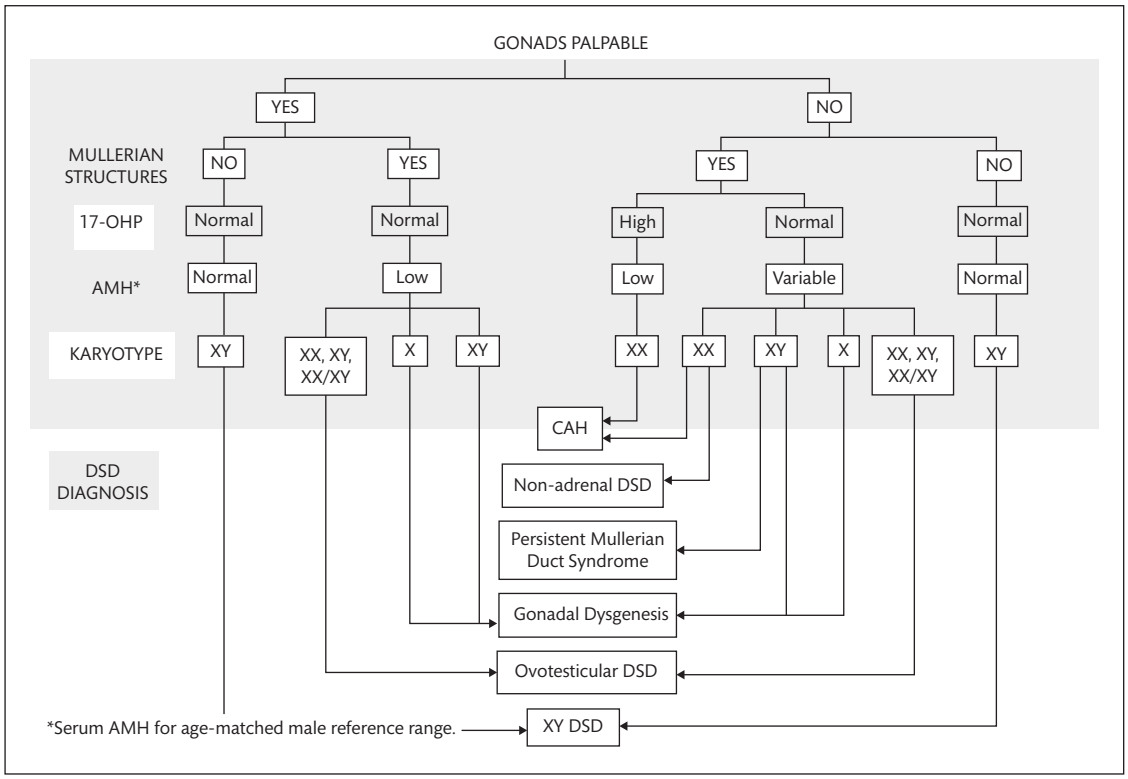


Figure 7.2.2.2 The use of first-line investigations in the newborn.

Reproduced with permission from Keir LS, O'Toole S, Robertson AL, Wallace AM, Ahmed SF. A 5-year-old boy with cryptorchidism and pubic hair: investigation and management of apparent male disorders of sex development in mid-childhood. *Horm Res.* 2009 Jan;71 Suppl 1:87–92. doi: 10.1159/000178046. Epub 2009 Jan 21.

and genetic analysis in CAH is informative. Monitoring weight is useful in any infant where there is a risk of salt loss.

Serum testosterone estimation has often been used as a marker of functioning testes as well as a sign of intact pathways for the synthesis of testosterone. However, given that many commercially available testosterone assays are non-specific in the early neonatal period [23] and can cross-react with other conjugated steroids, it is possible that for the newborn infant serum AMH level is a more diagnostically reliable marker of testes than serum testosterone. AMH is a useful tool to assess Sertoli function in suspected DSD and may have diagnostic utility in conditions associated with androgen deficiency or insensitivity [24]. There is a clear difference between AMH concentration in boys and girls, especially in early childhood. In boys under the age of 8 years, a serum AMH concentration of 200 pmol/L may be an appropriate cut-off to denote normality, given that it was the approximate tenth centile for this age range. However, AMH concentrations are generally higher in the young boy before they fall in late childhood (Figure 7.2.2.3). It should also be noted that AMH concentrations tend to rise over the first 3 months in some young infants [25] and may in some cases be lower than 200 pmol/L at initial

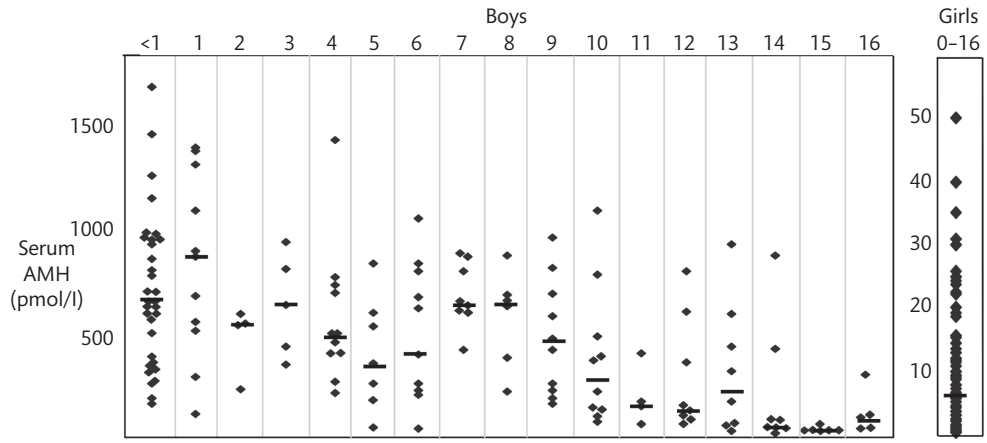


Figure 7.2.2.3 AMH concentration in boys and girls.

Reproduced with permission from Ahmed SF, Rodie M. Investigation and initial management of ambiguous genitalia. *Best Practice & Research Clinical Endocrinology & Metabolism*. Volume 24, Issue 2, April 2010, Pages 197–218. Copyright © 2009 Elsevier Ltd.

evaluation, although still above 25pmol/L which approximately represents the 90th centile for girls [26].

Second-Line Investigations

In most cases these tests are performed to investigate the underlying aetiology but are usually not necessary to determine the sex of rearing.

These investigations could include:

- Biochemistry to assess the gonadal and adrenal axes- hCG Stimulation to assess production of testosterone, androstenedione, dihydrotestosterone (DHT), 11-deoxycortisol and 17-hydroxypregnenolone. To detect abnormalities of the last three steroids, it may be more effective to analyse a urine steroid profile (spot or 24 hr) by gas chromatography-mass spectrometry (GC-MS). Other biochemical investigations that may need to be considered include LHRH, ACTH stimulation, renin, and aldosterone. Measurement of serum cholesterol and 7-dehydrocholesterol are indicated in the child who has features consistent with Smith–Lemli–Opitz syndrome.
- Imaging: ultrasound scan, MRI, genitogram.
- Internal surgical examination: cystourethroscopy, laparoscopy.
- Pathology: gonadal biopsy; there are, however, unresolved questions as to whether one biopsy represents the whole gonad. In addition, it is unclear as to what is the minimum amount of ovarian or testicular tissue that should be present to classify the gonad as an ovotestis.
- Genetics: high resolution karyotype, karyotype from different tissues (blood, skin, gonads), DNA for storage and analysis. Molecular genetic investigations are increasingly being performed at an earlier stage (Figure 7.2.2.4) in the diagnostic process [27].
- Functional studies of androgen sensitivity: a functional assessment of androgen sensitivity can be performed by assessing the clinical response of testosterone on the phallus. Secondly, androgen sensitivity can be assessed by measuring change in an androgen-responsive circulating protein such as sex hormone binding globulin (SHBG); SHBG levels should fall following androgen exposure and a failure to show this reduction may be indicative of androgen insensitivity [28]. The utility of this test in the young infant is unclear given that circulating SHBG is very variable in the young infant. Androgen binding studies involve evaluation of the concentration of androgen receptors; the number of receptors and their affinity for testosterone are measured on cultured genital skin fibroblasts. However, the results may depend on the site from which the skin is originally collected. Over 80% of cases with a phenotype consistent with complete androgen insensitivity syndrome and abnormal androgen binding may have a variant in the androgen receptor (*AR*) gene [29]. However, in cases consistent with a partial androgen insensitivity syndrome phenotype, only 50% of cases with abnormal binding may have a variant in the androgen receptor gene. Given that AR analysis may reveal a variant in over 80% of cases with a CAIS phenotype anyway, there probably is no need to perform androgen binding studies in this group of infants. However, the yield of AR variants in PAIS is much lower at less than 30%. Boys with PAIS and a variant in the *AR* may have poorer clinical outcomes, thus, routine genetic analysis of *AR* to confirm PAIS will inform long-term prognosis and management [30]. It is also possible that in some cases of AIS, the gene variant may exist beyond the *AR* coding region. Identification of such cases of AIS will be challenging and may require more complex tools such as measurement of androgen responsive proteins in genital skin fibroblasts [31] or measurement of androgen responsive peripheral transcriptome [32].

The hCG Stimulation Test

Although controversy exists regarding the optimal regimen, stimulation with human chorionic gonadotrophin (hCG) has been used to assess the presence of functioning testicular tissue and the detection of defects in testosterone biosynthesis and action for over 40 years [33]. In the UK, a number of different protocols are used for hCG stimulation but most use intramuscular hCG 1000–1500 units on 3 consecutive days for a standard test [17]. If there is a poor response, this test can be followed by prolonged hCG stimulation 1500 units on 3 consecutive days for the first week followed by 1500 units on 2 days a week for the 2 following weeks. The definition of a normal response may depend on the age of the child and the regimen itself. In infants and older children, who have a more active gonadotrophin axis, the Leydig cells may be more responsive to hCG stimulation and the shorter duration of hCG stimulation may be sufficient [34]. Recently, in an older group of children with suspected hypogonadotropic hypogonadism, cut-off for a normal testosterone response has been reported to be at 3.5 nmol/L after 3 days of stimulation and 9.5 nmol/L after the 3 week stimulation regimen [35]. Besides testosterone, other androgens that should be assessed include dihydrotestosterone and androstenedione. For these two metabolites, the day 4 sample is more important than the day 1 sample. There is no additional benefit of collecting a sample for these two metabolites on day 22.

Aetiology of XX DSD

46,XX DSD can be divided into disorders of ovarian development, disorders of androgen synthesis, disorders of Müllerian development and other conditions affecting sex development.

The most common cause of 46,XX DSD with ambiguous genitalia in newborns is congenital adrenal hyperplasia (CAH). It is characterized by androgen excess and a variable alteration in glucocorticoid and mineralocorticoid function and a specific profile of steroid hormones. This profile can identify enzyme defects including deficiency of 21α-hydroxylase (90–95% of cases), 11β-hydroxylase (4–8% cases), 3β-hydroxysteroid dehydrogenase type 2 (rare), and P450 oxidoreductase (unknown prevalence). Other causes of androgen excess leading to masculinization of XX individuals including P450 aromatase deficiency, maternal androgen excess, and androgenic tumours in pregnant mothers (e.g. luteoma) are relatively infrequent, however, these should be considered in the differential diagnosis of maternal/fetal virilization.

Disorders of Gonadal Development

46,XX Ovotesticular DSD (Formerly Known as 'True Hermaphrodites') and 46,XX Testicular DSD ('46,XX Males')

Rarely, the developing ovary may contain some testicular tissue (Ovotesticular DSD) or may develop as a functioning testis that secretes adequate amounts of testosterone for adequate virilization and AMH for regression of the Müllerian ducts (testicular

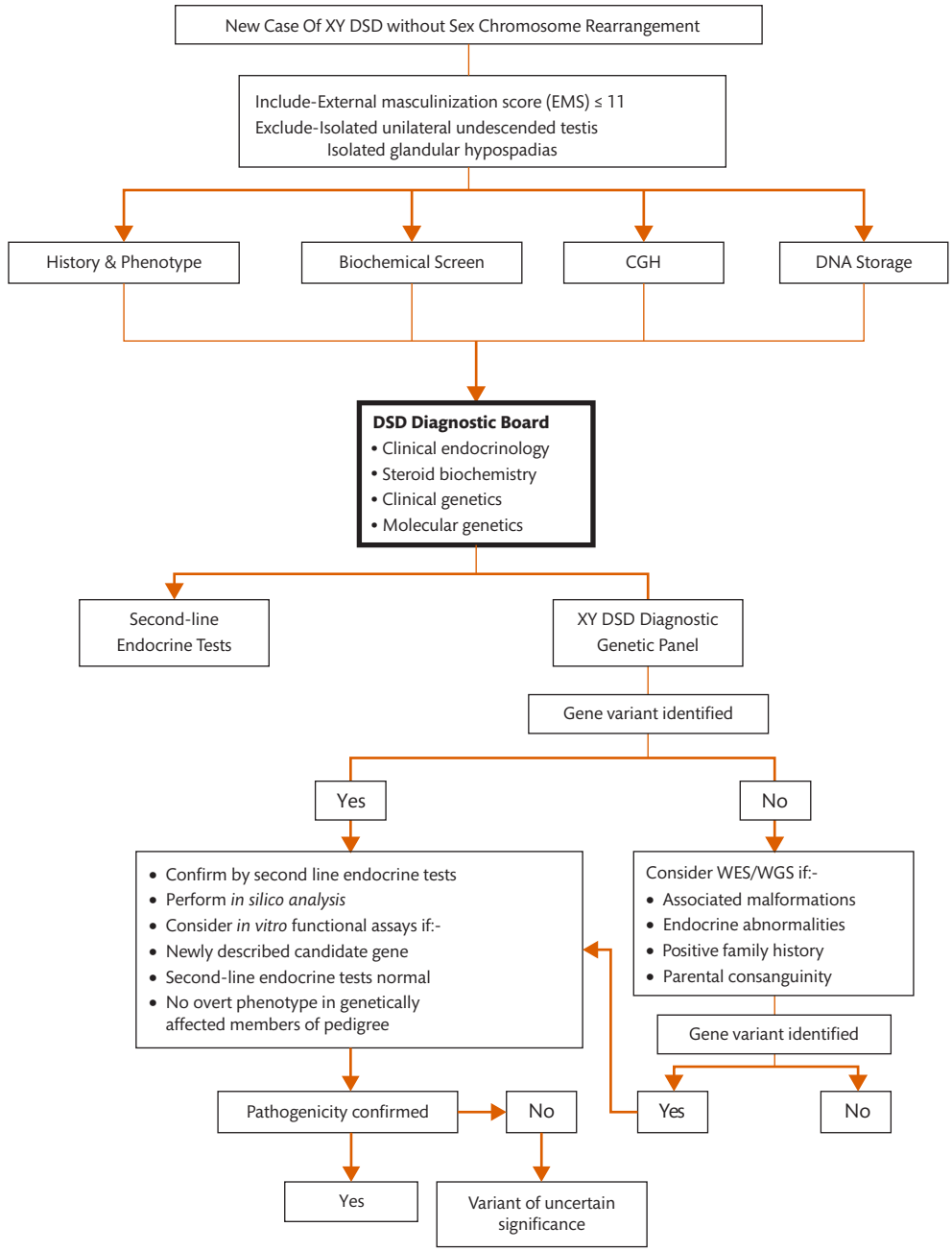


Figure 7.2.2.4 Suggested pathway for investigating cases of 46,XY disorder of sex development (DSD) through an integrated endocrine and genetic approach that relies on a joint DSD diagnostic board with expertise in endocrinology, steroid biochemistry, clinical genetics and molecular genetics.

Reproduced with permission from Alhomaidah D, McGowan R, Ahmed SF. The current state of diagnostic genetics for conditions affecting sex development. *Clin Genet*. 2017 Feb;91(2):157–62. doi: 10.1111/cge.12912.

DSD). The incidence of 46,XX testicular DSD is estimated to be 1 in 20 000. Ovotesticular DSD can be subclassified according to the type and location of the gonads. Lateral cases (20%) have a testis on one side and an ovary on the other. Bilateral cases (30%) have testicular and ovarian tissues present bilaterally as ovotestes. Unilateral cases (50%) have an ovotestis present on one side and a normal ovary or testis present on the other side. In ovotesticular DSD, the initial manifestations are ambiguous genitalia in almost all cases and the internal duct structures display gradations between male and female. There is often a urogenital sinus and a uterus or a hemi- or a rudimentary uterus on the side of the ovary or ovotestis. Breast development will occur in puberty and even menses may occur in a significant proportion when ovarian tissue is present. However, without removal of testicular tissue, these children will also proceed to virilize at puberty. Presence of functional testicular tissue can be investigated by checking AMH or testosterone levels following hCG stimulation. Assessment of functioning ovaries by biochemical markers has not been thoroughly explored and the utility of measuring oestradiol after repeat FSH stimulation or measurement of an ovarian specific marker such as Inhibin A requires further study. Two-thirds of affected children are raised as boys (see section, 'Sex Assignment in the Affected Newborn'). If the testicular components

are removed, serial AMH levels may allow adequate confirmation of complete removal of functioning testicular tissue. In contrast, 46XX testicular DSD is usually associated with a normal male phenotype or a relatively mild abnormality of the male genitalia, such as distal or mid-shaft hypospadias. In adulthood, although testosterone synthesis is not affected, spermatogenesis is usually severely affected.

Ovarian Dysgenesis

Ovarian dysgenesis is most frequently seen in association with sex chromosome aneuploidy such as Turner syndrome and related variants. However, these conditions do not present in infancy with physical abnormalities of sex development.

Disorders of Androgen Excess

21α-Hydroxylase (CYP21) Deficiency

Congenital adrenal hyperplasia due to 21-hydroxylase deficiency is the commonest cause of 46XX DSD, accounting for up to 95% of all cases. Consensus guidelines exist for the management of this condition in infancy as well as in the older child [36]. The newborn girl with this condition can be virilized to a varying extent, as illustrated by the Prader classification (Figure 7.2.2.1). High serum concentration of 17OH-progesterone (>300 nmol/L) after the first 48 hours of birth and high androstenedione and testosterone in the early neonatal period are the biochemical hallmarks of this condition. More than 75% of these infants will also be salt losers because of a deficiency of mineralocorticoid synthesis and the affected child will present with a salt losing crisis in the second or third week of life.

3β-Hydroxysteroid Dehydrogenase (HSD3B2) Deficiency

3β-hydroxysteroid dehydrogenase type 2 catalyses the conversion of Δ^5 steroids to Δ^4 steroids and a deficiency of this enzyme results in adrenal insufficiency as well as accumulation of pregnenolone, dehydroepiandrosterone (DHEA), and androstenediol. In peripheral tissues, as well as the placenta, the accumulating steroids, and particularly DHEA, can be converted to more potent androgens, such as testosterone, by the Type 1 isoenzyme. Most girls with this condition present with relatively mild signs of virilization such as clitoromegaly, associated with adrenal deficiency.

P450 Oxidoreductase (POR) Deficiency

Defects in P450 oxidoreductase can cause combined deficiencies of 21α-hydroxylase, 17α-hydroxylase, and aromatase enzymes and this can be associated with abnormal genital development in both girls and boys. Children with this condition usually have cortisol deficiency, but have normal mineralocorticoid function.

11β-Hydroxylase (CYP11B1) Deficiency

This is the second commonest cause of virilizing congenital adrenal hyperplasia accounting for approximately 5% of all cases. Apart from a DSD, this condition may also be associated with hypertension and hypokalaemia, but these abnormalities are not universally present, particularly, not in infancy. These abnormalities are due to the accumulation of 11-deoxycorticosterone which is a weak mineralocorticoid. They may be associated with a low renin. Children with this condition usually have cortisol deficiency.

Familial Glucocorticoid Resistance

This is a rare condition, usually due to a heterozygous mutation in the glucocorticoid receptor α gene. The partial end-organ insensitivity leads to high ACTH, cortisol, mineralocorticoids, and androgens. One case of a girl with a homozygous mutation in this gene and a coexisting heterozygous mutation in CYP21 has been described with marked virilization at birth.

P450 Aromatase Deficiency

Aromatase deficiency is inherited as a rare autosomal recessive condition caused by CYP19A1 gene mutation. There is often a history of maternal virilization after the second trimester of pregnancy coupled with elevated maternal androgen levels which resolve after the pregnancy. Urine from these mothers has elevated levels of 16α-hydroxyandrostenedione (16OH-D4A) and its metabolites. In infancy and subsequently during puberty, these girls have high serum androgens and low oestrogen concentrations as affected individuals cannot synthesize endogenous oestrogens. Girls show no signs of feminization and progressively virilize. Inadequate oestrogen supplementation may be associated with osteoporosis and a failure of timely epiphyseal fusion. Molecular testing of the *CY19A1* gene is available.

Maternal Androgen Excess

Any maternal source of elevated androgens can induce virilization of the female fetus. Ovarian tumours include luteoma of pregnancy, arrhenoblastoma, hilar-cell tumour, masculinizing ovarian stromal cell tumour, and Krukenberg tumour. Discrepancy between the marked virilization of the mother and the minimal androgen effect in female offspring can be explained by the placental aromatase activity, which converts androgens to oestrogens, or to the metabolism of androgen, which thus becomes less active. Apart from untreated maternal virilizing CAH, androgen-secreting adrenal tumour in the mother is rare. In both cases, investigation of abnormal androgen production by the mother must be performed immediately after delivery. Maternal ingestion of androgens, progestogens, or other drugs is another cause of fetal virilization. Exogenous steroids administered during the pregnancy may cause posterior fusion of the labia, clitoral enlargement, and even increased degrees of androgenization. In the past, several oral progestational compounds, given because of threatened abortion, have been implicated, such as 19-nor testosterone. Other drugs, like danazol or stilbestrol that are used in pregnancy, have also been associated with abnormalities of the genitalia.

Disorders of Müllerian Development

Abnormalities in uterine development can result in bicornuate uterus, uterine hemiagenesis, hypoplasia, or agenesis. These can be associated with renal, cardiac or spine abnormalities as part of the Mayer–Rokitansky–Küster–Hauser (MRKH) syndrome or MURCS (Müllerian agenesis, renal aplasia, cervical spine anomalies) syndrome. MRKHS (1:5000 females) is thought to be the second commonest cause of primary amenorrhoea after ovarian dysgenesis. The MURCS syndrome occurs in approximately one-third of all MRKH patients. On rare occasions, absence of Müllerian structures

and the presence of coexisting hyperandrogenaemia, has been associated with a mutation in the WNT4 gene [37]. Other conditions such as maturity onset diabetes of the young, the hand–foot–genital syndrome and Laurence–Moon–Biedl syndrome have also been associated with abnormalities of Müllerian development.

Other 46,XX DSD

Complex urogenital abnormalities, such as cloacal anomalies, can affect both sexes and require major reconstructive surgery.

Variations That May Present as DSD

Clitoral lengths are variable and when in doubt should be compared to published norms [4]. In addition, the clitoris may be enlarged in conditions such as neurofibromatosis. In any newborn girl, the labial folds may be very swollen and oedematous immediately after birth and may look like scrotal sacs. In premature babies, the lack of labial adipose tissue may make the relative size of the clitoris more pronounced so that it is mistaken for clitoromegaly. Labial adhesions and vaginal bleeding in the newborn are signs of the normal oestrogen surge in the newborn period.

Aetiology of XY DSD

46,XY DSD can be divided into disorders of testis development, disorders of androgen synthesis, disorders of androgen action and other conditions affecting sex development. Biochemically, based on AMH and the hCG stimulation test, these disorders can also be divided into conditions where (1) AMH levels are low and testosterone levels do not rise following hCG stimulation—abnormalities of testes development or maintenance; (2) AMH levels are normal and testosterone levels do not rise following hCG stimulation—abnormalities of testosterone synthesis; (3) AMH levels are normal and testosterone levels do rise following hCG stimulation—abnormalities of testosterone action, dihydrotestosterone synthesis, persistent Müllerian duct syndrome or non-specific disorder of masculinization; (4) AMH levels are low and testosterone levels do rise following hCG stimulation—persistent Müllerian duct syndrome (Figure 7.2.2.5). However, in many cases, the biochemical assessment will not clearly delineate the case into any of these four categories.

Disorders of Testis Development

These disorders can have a spectrum of phenotypes and presentations. In the most extreme case, complete testicular dysgenesis, infants raised as girls do not present until adolescence with primary amenorrhoea. These girls will have normal external female genitalia and Müllerian structures and this condition is often called Swyer's syndrome. Partial gonadal dysgenesis may be associated with a variable and internal phenotype even extending to a phenotype of simply male infertility. Accordingly, there will be a variable reduction in AMH and testosterone response to hCG stimulation. Given that several single gene disorders, as well as chromosomal rearrangements have been described to be associated with the clinical picture of gonadal dysgenesis, the latter should not necessarily be considered the final diagnosis. These disorders are often associated with abnormalities in other systems and a thorough clinical evaluation of the affected infant will prove very useful in directing appropriate genetic analysis that can lead to the correct diagnosis. Currently, a genetic diagnosis is only reached in approximately 30% of cases of gonadal dysgenesis. The importance of reaching a genetic diagnosis in these cases is highlighted by conditions such as those associated with a mutation in the steroidogenic factor (*SF1*) gene which may occasionally be associated with adrenal deficiency or a mutation of the Wilms' tumour-related gene-1 (*WT1*), where the DSD may be the first sign of conditions such as WAGR syndrome, Denys–Drash syndrome, and Frasier syndrome.

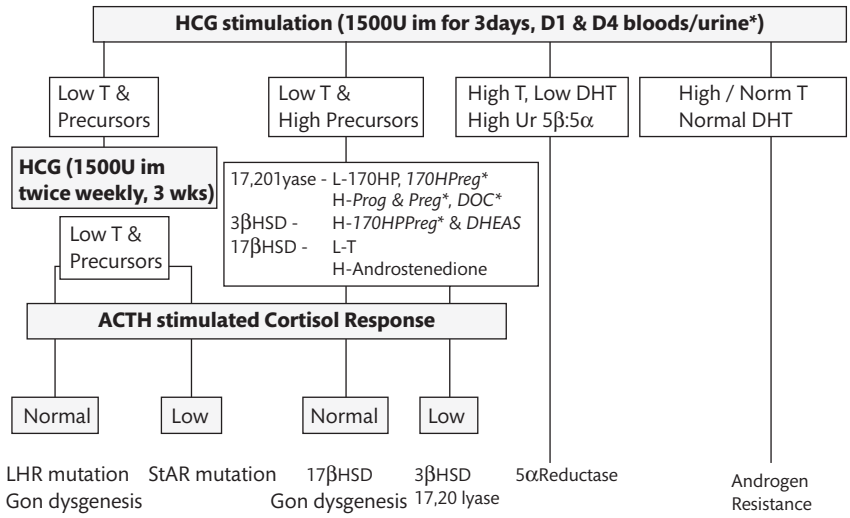


Figure 7.2.2.5 The HCG Test For Investigating XY DSD.

Reproduced with permission from Ahmed SF, Rodie M. Investigation and initial management of ambiguous genitalia. *Best Practice & Research Clinical Endocrinology & Metabolism*. Volume 24, Issue 2, April 2010, Pages 197–218. Copyright © 2009 Elsevier Ltd.

Disorders of Androgen Synthesis

Defects anywhere along the pathway of androgen synthesis and target organ action can result in an XY DSD.

Cholesterol Synthesis Defects

A deficiency of 7-dehydrocholesterol reductase (DHCR7) results in a failure of cholesterol synthesis and results in the Smith–Lemli–Opitz syndrome. This is associated with a wide range of clinical features including microcephaly, cardiac defects, micrognathia, cleft palate, polydactyly, and syndactyly. In childhood, these children may display intellectual disability and growth failure. The genitalia in the affected XY infant may range from hypospadias to completely normal female external genitalia with no Müllerian ducts. The condition is diagnosed by low levels of cholesterol and elevated levels of its precursor, 7-dehydrocholesterol, as well as an androgen deficiency and a normal AMH. Adrenal insufficiency may occur in some cases and needs evaluation. Mutational analysis of the *DHCR7* gene will provide further confirmation of the diagnosis.

Leydig Cell Hypoplasia

A defect of the luteinizing hormone-choriogonadotropin receptor (*LHCGR*) gene leads to impaired sensitivity to hCG and LH and Leydig cell agenesis or hypoplasia. The genitalia in the affected XY infant may range from isolated hypospadias or micropenis to completely normal female external genitalia with no Müllerian ducts. The biochemical picture may include high basal and luteinizing hormone-releasing hormone (LHRH)-stimulated luteinizing hormone (LH) and follicle-stimulating (FSH) levels. There is a poor response to hCG stimulation and the AMH levels should be normal. Histology of the testes in the prepubertal child will show a marked lack of Leydig cells. Mutational analysis of the LH/hCG receptor gene provides further confirmation of the diagnosis.

Congenital Lipoid Adrenal Hyperplasia

Defects in the steroidogenic acute regulatory (StAR) protein lead to deranged intracellular transport of cholesterol and abnormalities of steroid biosynthesis. Affected XY infants have severe adrenal failure and the external genitalia are unambiguously female with no Müllerian structures. The testes may be palpable in the labioscrotal folds but are usually undescended. CT or MRI imaging of the adrenal glands, as well as histology, may reveal lipid accumulation. This condition is commoner in Japan and Korea. A non-classical form of this condition also exists and is associated with progressive adrenal insufficiency in early childhood but without any overt abnormalities of androgen synthesis. Mutational analysis of the StAR gene will provide further confirmation of the diagnosis.

P450 Side Chain Cleavage Deficiency

Defects in the P450scc enzyme result in a failure of conversion of cholesterol to pregnenolone which is the first common step in steroid biosynthesis. Affected XY infants have a phenotype which is very similar to congenital lipoid adrenal hyperplasia due to defect of StAR protein. Mutational analysis of the P450scc gene (also called *CYP11A1*) will provide further confirmation of the diagnosis.

3β-Hydroxysteroid Dehydrogenase (3β-HSD) Type 2 Deficiency

Defects in 3β-HSD Type 2 results in a failure to convert Δ^5-steroids to Δ^4-steroids. The genitalia in the affected XY infant may range from isolated hypospadias or micropenis to more severe undermasculinization but not completely normal female external genitalia. There are no Müllerian ducts. Besides a poor androgen response to hCG stimulation, affected infants will have adrenal deficiency which may not necessarily include salt wasting. A urine steroid profile that shows high concentrations of Δ^5-steroids (e.g. 17OH-pregnenolone, pregnenolone, dehydroepiandrosterone) and low concentrations of Δ^4-steroids (e.g. progesterone and cortisol) is helpful. However, there is a need for careful analysis and interpretation of the steroid profile as extra-adrenal/gonadal 3β-HSD type 1 may raise the levels of some Δ^4-steroids such as androstenedione and 17OH-progesterone. In addition, raised 17-OHP can lead to an erroneous diagnosis of 21 hydroxylase deficiency. Mutational analysis of the 3β-HSD Type 2 gene (also called *HSD3B2*) will provide confirmation of the diagnosis.

17α-Hydroxylase/17,20-Lyase Deficiency

Defects of the P450c17 enzyme can lead to a variable extent of a combined deficiency of 17α-Hydroxylase and 17,20-Lyase activity. In the affected XY infant, this will be associated with a variable degree of undermasculinization ranging from mild abnormalities of the genitalia to unambiguously female external genitalia. These account for around 1% of cases of 46,XY DSD. Besides a poor androgen response to hCG stimulation, affected infants will have a poor cortisol response to adrenal stimulation but may not display adrenal insufficiency as they have highly raised deoxycorticosterone levels which may lead to a state of low renin hypertension and hypokalaemic alkalosis in the older child. Mutational analysis of the P450c17 gene (also called *CYP17*) will provide confirmation of the diagnosis.

P450 Oxidoreductase Deficiency

The P450 oxidoreductase enzyme is necessary for electron transfer from NADP to many P450 enzymes and its deficiency can affect the activity of a number of P450 enzymes. Infants with XY DSD and an abnormality of this enzyme tend to present with a clinical picture consistent with combined deficiency of 21α-hydroxylase deficiency and 17,20-lyase deficiency. The genitalia in the affected XY infant may range from isolated hypospadias or micropenis to more severe undermasculinization but not completely normal female external genitalia. There are no Müllerian ducts. Besides a poor testosterone response to hCG stimulation, affected infants will have adrenal deficiency which is usually restricted to glucocorticoid deficiency. The deficiency of this enzyme may be associated with a condition called Antley–Bixler syndrome which is a skeletal dysplasia classically characterized by radiohumeral stenosis and craniosynostosis. This syndrome is not universally associated with abnormalities of the P450 oxidoreductase enzyme. Mutational analysis of the P450 oxidoreductase gene (also called *POR*) may provide confirmation of the diagnosis.

17β-Hydroxysteroid Dehydrogenase (17β-HSD) Type 3 Deficiency

17β-HSD has 6 isoenzymes which convert androstenedione, DHEA and estrone to testosterone. Deficiency of 17β-HSD Type 3 is

associated with XY DSD and affected infants often present with female external genitalia or sometimes ambiguous genitalia. However, these children can undergo spontaneous virilization during puberty with a rise in testosterone levels, possibly due to increased activity of the other isoenzymes. Thus, early accurate diagnosis of this condition is important as the affected infant may need sex reassignment if initially raised as a girl (see section, 'Sex Assignment in the Affected Newborn'). These infants shall have a poor testosterone response to hCG but may have a relatively high level of serum androstenedione such that the testosterone:androstenedione ratio may be less than 0.8. However, this is not an invariable finding in this condition; furthermore, a low ratio may also be found in poorly functioning testes. Mutational analysis of the 17β-HSD type 3 gene (also called *HSD17B3*) will provide confirmation of the diagnosis.

Steroid 5α-Reductase (5α-RD) Type 2 Deficiency

5α-RD exists as two isoenzymes. Type 1 is expressed in skin and type 2 in the genitalia. In XY DSD infants may present with a variable phenotype ranging from micropenis or hypospadias to female external genitalia. This phenotype is due to reduced activity of 5α-RD type 2 and a failure to convert testosterone to dihydrotestosterone. The classical biochemical profile includes a high testosterone:dihydrotestosterone ratio following hCG stimulation (normal ratio is <25:1) although many affected cases have a lower ratio. An additional diagnostic feature is a urinary steroid profile which shows a decreased ratio for 5α:5α-reduced C_{21} and C_{19} steroids. It may not be possible to detect this abnormality in the urine until late infancy. Like 17β-HSD type 3 deficiency, these children can undergo spontaneous virilization during puberty with a rise in testosterone levels, possibly due to increased activity of the Type 1 isoenzymes. Thus, early accurate diagnosis of this condition is important as the affected infant may need sex reassignment if initially raised as a girl (see section, 'Sex Assignment in the Affected Newborn').

Disorders of Androgen Action

In XY DSD, a disorder of the androgen receptor leads to a phenotype that can range from a man with infertility through to a range of abnormalities of the genitalia in the newborn boy (partial androgen insensitivity syndrome, PAIS) to completely female external genitalia (Complete Androgen Insensitivity Syndrome, CAIS). The exact prevalence of PAIS and CAIS are unknown and estimates vary widely. Older estimates give prevalence rates of CAIS in 46,XY individuals of approximately 1:20 000 to 1:64 000. Children with AIS should have normal testosterone and dihydrotestosterone response to hCG stimulation and should have a normal urinary steroid profile. However, a number of children with a confirmed genetic diagnosis of AIS may have a poor response to hCG stimulation, and this may be related to associated abnormalities of the testes or the test itself. The AMH level should be normal; sometimes it has been shown to be somewhat high for age-matched standards. Similarly, LH levels may be high especially following LHRH stimulation. There are no Müllerian ducts. In the older infant, fixed treatment with testosterone may not be accompanied by changes in testosterone responsive effects, such as a fall in SHBG or change in the size of the phallus. Mutational analysis of the androgen receptor gene (also called *AR*) will provide confirmation of the diagnosis. Androgen binding studies may be helpful in directing mutational analysis, particularly in cases of PAIS. As the condition is inherited in an X-linked pattern, a consistent family history is very helpful. Furthermore, exploration of X-linked markers in affected and non-affected family members can indicate the likelihood of the condition. A number of cases of XY DSD are incorrectly labelled as 'PAIS' when no firm biochemical or genetic abnormalities are identified in gonadal function, androgen synthesis or androgen action. Strictly speaking, the term PAIS should be reserved for those children who have XY DSD and a genetic abnormality of the AR gene. The children without a genetic abnormality may be better described as 'XY DSD with a non-specific disorder of under-masculinization'.

Persistent Müllerian Duct Syndrome (PMDS)

AMH is secreted by the Sertoli cells from around 7 weeks gestation and subsequently acts through the AMH Type 2 receptor to lead to regression of the Müllerian ducts. PMDS is a rare condition occurring due to a variant of the AMH gene or its receptor. In XY infants with PMDS, boys are born with male external genitalia but have persistence of internal Müllerian structures. The diagnosis is usually suspected when a child has a repair of an inguinal hernia, orchidopexy, or coincidental intrabdominal surgery. There are two anatomic forms. In the commoner type, there is one inguinal hernia which contains the ipsilateral testis and the ipsilateral fallopian tube and the uterus. In some of these herniae, the contralateral testis may also be present. In the less-commoner form, all the structures including the testes are present in the pelvis. Affected children have a normal testosterone response to hCG but fertility and, sometimes Leydig cell function may be compromised in adulthood due to unsuccessful attempts at orchidopexy and anatomical abnormalities of the epididymis and the vas deferens. Surgical opinion about the timing of salpingectomy and hysterectomy vary.

Disorders of Testes Maintenance

A number of different terminologies (bilateral vanishing testes, embryonic testicular regression, rudimentary testes, congenital anorchia) are used to describe a group of conditions which are characterized in infants with a XY karyotype and absent or rudimentary testes. The syndrome entails the presence of testes which vanish during embryogenesis. The aetiology of this syndrome is unclear: regression of the testes *in utero* may be due to a genetic variant, a teratogen factor, or a bilateral torsion. Clinically, the syndrome encompasses a spectrum of phenotypes, ranging in severity from genital ambiguity to a male phenotype with an empty scrotum. The management of patients with a defect of testes maintenance is dictated by their position in the clinical spectrum of the disorder. Patients with rudimentary testes have a male phenotype with micropenis, small atrophic testis with pre-Sertoli and Leydig cells. Some patients present with perineal hypospadias and persistent Müllerian derivatives. Congenital anorchia is characterized by the complete absence of testicular tissue at birth, but normal male sexual differentiation without Müllerian derivatives.

Gender and Its Development

Unlike the sex categories, male and female, gender has several aspects: gender assignment, gender role, gender identity, gender attribution, and sexuality. In most societies, gender assignment occurs at birth, long before we have a say in the matter, marking the beginning of the process of gender socialization. The process of gender socialization also includes society's expectations of how males or females should behave, as expressed in their gender role behaviour. Gender identity is distinct from gender role behaviour and refers to the individual's perception of one's own gender and how it conforms to the male or female gender role in society. Gender attribution is what we all do when we meet someone and want to decide whether they are a man or a woman. This is often based on obtaining a number of cues which are symbolic manifestations of gender and that have traditionally included clothing, mannerisms, physical appearance, gait, and occupational choice. Finally, sexuality refers to erotic desires, sexual practices, or sexual orientation. In some cultures, individuals are often socially identified as homosexuals or heterosexuals as if a person's sexual orientation encapsulates the total personality and identity. For most people, their gender identity, gender role, and the symbolic gender manifestations are congruent and, in addition, they will be sexually attracted to the opposite sex. However, it is also possible that a man may have gender manifestations that do not completely converge with his male gender identity and remains sexually attracted to a member of the opposite sex; of course, a number of other permutations may also exist. Some aspects of gender, such as role, assignment, the symbolic manifestations, as well as the different types of sexuality, may differ markedly from one society to another and continue to evolve within respective societies. In some cultures, the distinction is becoming less absolute and it may be better to consider these aspects as a continuum, with female characteristics at one extreme and male ones at the other. The development of gender identity is the result of a complex interaction between genetic, prenatal, and postnatal endocrine influences and postnatal psychosocial and environmental experiences. Gender development consists of gender identity formation such as gender knowledge, self-perception, preferences (toy, playmate), and gender role behaviours [38]. By the end of the first year of life, infants may already be able to discriminate between the sexes, and some may be able to display sex related toy preferences. By 2–3 years of age children are able to correctly label themselves and others according to gender. By the age of 3 years, preference for one sex role has emerged with the child having a clear sense of whether he/she is a boy or girl. Children fix on cues such as clothing and hair in gender labelling exercises; even when genital cues are available they are used far less to make categorization decisions than these other cues, at least until the age of 8 years or so, possibly reflecting insufficient biological understanding of gender differences. By the age of 5 years, children learn that gender remains stable over time, becoming preoccupied with categorical differences between males and females. However, it is not until children have mastered the concept that gender remains constant (despite superficial changes in appearance), at the age of between 5 and 7, that many argue is when a gender identity has been fully attained. Theorists have suggested that once 'gender constancy' has been mastered, this becomes a motivator to shaping sex appropriate gender behaviour [38].

Sex Assignment in the Affected Newborn

Initial gender uncertainty is unsettling and stressful for families as well as the health professionals. Given that gender development is a relatively long-term process, clinical professionals involved in management need to be clear of the distinction between sex assignment and gender assignment; the latter cannot be achieved by the clinical team and should be considered intrinsic to the child's own development. However, expediting a thorough assessment and reaching a decision on sex assignment is required. Factors that influence sex assignment include the diagnosis, genital appearance, surgical options, need for lifelong replacement therapy, the potential for fertility, views of the family and sometimes, circumstances relating to cultural practices. More than 90% of 46,XX CAH patients and all 46,XY CAIS assigned female in infancy identify as females. Evidence supports the current recommendation to raise markedly virilized 46,XX infants with CAH as female. In the late presenting virilized 46,XX child who has been raised as a boy, there are cases where gender reassignment has not been undertaken and there is a need for long-term outcome studies in these cases as well as those where gender reassignment has occurred [39]. Approximately, 60% of 5α-reductase (5αRD2) deficient patients assigned female in infancy and virilizing at puberty (and all assigned male) live as males. In 5αRD2 and possibly 17β-hydroxysteroid dehydrogenase (17βHSD3) deficiencies, where the diagnosis is made in infancy, the combination of a male gender identity in the majority and the potential for fertility (documented in 5αRD2, but unknown in 17βHSD3) should be discussed when providing evidence for gender assignment. Among patients with PAIS, androgen biosynthetic defects, and incomplete gonadal dysgenesis, there is dissatisfaction with the sex of rearing in about 25% of individuals whether raised male or female; recent data have shown a trend towards these affected infants being raised as boys [40]. The decision on sex of rearing in ovotesticular DSD should consider the potential for fertility based on gonadal differentiation and genital development, and assuming the genitalia are, or can be made, consistent with the chosen sex. In the case of mixed gonadal dysgenesis (MGD), factors to consider include prenatal androgen exposure, testicular function at and after puberty, phallic development, and gonadal location. Individuals with cloacal exstrophy reared female show variability in gender identity outcome, but more than 65% appear to live as women.

Surgical Management

The surgeon has a responsibility to outline the surgical sequence and subsequent consequences from infancy to adulthood. Only surgeons with expertise in the care of children and specific training in the surgery of DSD should perform these procedures. Parents now appear to be less inclined to choose surgery. As orgasmic function and erectile sensation may be disturbed by clitoral surgery, the surgical procedure should be anatomically based to preserve erectile function and the innervation of the clitoris. Emphasis should be placed more on functional outcome, rather than a strictly cosmetic appearance. It is generally felt that surgery that is performed for cosmetic reasons in the first year of life relieves parental distress and improves attachment between the child and the parents. However, systematic evidence for this belief is lacking. It is anticipated that surgical reconstruction in infancy will need to be refined at the time of puberty.

Vaginal dilatation should not be undertaken before puberty. The surgeon must be familiar with a number of operative techniques in order to reconstruct the spectrum of urogenital sinus disorders. An absent or inadequate vagina (with rare exceptions) requires a vaginoplasty in adolescence when the patient is psychologically motivated and a full partner in the procedure. In the case of a DSD associated with hypospadias, standard techniques for surgical repair include chordee correction, urethral reconstruction, and the judicious use of testosterone supplementation. The magnitude and complexity of phalloplasty in adulthood should be taken into account during the initial counselling period. It should also be explained to parents that sexual contentment is not simply dependent on penetrative sex. Parents must not be given unrealistic expectations about penile reconstruction, including the use of tissue engineering. The testes in patients with CAIS and those with PAIS, raised female, need to be removed to prevent malignancy in adulthood, but this can be deferred until adolescence which allows spontaneous feminization and an opportunity for the patient to have a say in the timing of removal. The streak gonad in a patient with MGD raised male should be removed in early childhood. Bilateral gonadectomy is performed in early childhood in females (bilateral streak gonads) with gonadal dysgenesis and Y chromosome material. In patients with androgen biosynthetic defects raised female, gonadectomy should be performed before puberty. A scrotal testis in patients with gonadal dysgenesis remains at risk for malignancy and there is little consensus on screening besides regular palpation in adolescence and adulthood.

Psychosocial Management

Psychosocial care should be an integral part of management in order to promote positive adaptation and allow parents to express and resolve their concerns. The early concerns of parents may be less to do with the long-term implications of the condition and more to do with coping and adjustment during infancy (Table 7.2.2.2). Experienced healthcare staff including the mental healthcare services who work as part of a clinical network with access to others with more specialist knowledge and experience may be particularly valuable in providing generic psychosocial support. A common issue seems to be related to how the condition should be explained to friends and relatives [8]. This expertise can facilitate team decisions about gender assignment/reassignment, timing of surgery and sex hormone replacement. Psychosocial screening tools that identify families at risk for maladaptive coping with a child's medical condition should be considered [41]. Once the child is sufficiently developed for a psychological assessment of gender identity, such an evaluation must be included in discussions about gender reassignment. Gender identity development begins before the age of 3 years, but the earliest age at which it can be reliably assessed remains unclear. The generalization that the age of 18 months is the upper limit of imposed gender reassignment should be treated with caution and viewed conservatively. Atypical gender role behaviour is more common in children with DSD than in the general population, but should not be taken as an indicator for gender reassignment. It is important to emphasize the distinction between sex-typical behaviour, sexual orientation, and gender identity. Thus homosexual orientation (relative to sex of rearing) or strong cross-sex interest in an individual with DSD is not an indication of incorrect gender assignment. In the longer-term, most current studies suggest that affected individuals lead productive lives, but a small proportion may have functional problems and may also suffer from gender identity disorders [42]. Parents do need to be aware of these issues. It should be explained that the process of disclosure concerning facts about karyotype, gonadal status, and prospects for future fertility is a collaborative ongoing action which requires a flexible individual-based approach.

REFERENCES

1. Ahmed SF, Dobbie R, Finlayson AR, *et al*. Regional and temporal variation in the occurrence of genital anomalies amongst singleton births, 1988–1997, Scotland. *Arch Dis Child Fetal Neonatal Ed*, 2004; **89**: F149–51.
2. Thyen U, Lanz K, Holterhus PM, Hiort O. Epidemiology and initial management of ambiguous genitalia at birth in Germany. *Horm Res*, 2006; **66**: 195–203.
3. Ahmed SF, Bryce J, Hiort O. International networks for supporting research and clinical care in the field of disorders of sex development. *Endocr Dev*, 2014; **27**: 284–92.
4. Hughes IA, Houk C, Ahmed SF, Lee PA. Consensus statement on management of intersex disorders. *Arch Dis Child*, 2006; **91**: 554–63.
5. Diamond DA, Burns JP, Mitchell C, *et al*. Sex assignment for newborns with ambiguous genitalia and exposure to fetal testosterone: attitudes and practices of pediatric urologists. *J Pediatr*, 2006; **148**: 445–9.
6. Reis E. Divergence or disorder? The politics of naming intersex. *Perspect Biol Med*, 2007; **50**: 535–43.
7. Markosyan R, Ahmed SF. Sex assignment in conditions affecting sex development. *J Clin Res Pediatr Endocrinol*, 2017; **9**: 106–12.
8. Duguid A, Morrison S, Robertson A, *et al*., on behalf of the Scottish Genital Anomaly Network. The psychological impact of genital anomalies on the parents of affected children. *Acta Paediatr*, 2007; **96**: 348–52.
9. Poyrazoglu S, Darendeliler F, Ahmed SF, *et al*. Birth weight in different etiologies of disorders of sex development. *J Clin Endocrinol Metab*, 2017; **102**: 1044–1050.
10. Cox K, Bryce J, Jiang J, *et al*. Novel associations in disorders of sex development: findings from the I-DSD Registry. *J Clin Endocrinol Metab*, 2014; 99: E348–55.
11. Baskin LS. Hypospadias. *Adv Exp Med Biol*, 2004; **545**: 3–22.
12. Cox K, Kyriakou A, Amjad B, *et al*. Shorter anogenital and anoscrotal distances correlate with the severity of hypospadias: a prospective study. *J Pediatr Urol*, 2017; 13: 57.e1–57.e5.
13. Prader A. Der genitalbefund beim pseudohermaphroditismus femininus des kongenitalen androgenitalen syndroms. *Helvetica Paediatr Acta*, 1954; **3**: 231–48.
14. Quigley CA, De Bellis A, Marschke KB, *et al*. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev*, 1995; 16: 271.
15. Sinnecker GH, Hiort O, Nitsche EM, *et al*. Functional assessment and clinical classification of androgen sensitivity in patients with mutations of the androgen receptor gene. German Collaberative Intersex Study Group. *Eur J Pediatr*, 1997; **156**: 7–14.
16. Ahmed SF, Khwaja O, Hughes IA. The role of a clinical score in the assessment of ambiguous genitalia. *BJU Int*, 2000; **85**: 120–4.
17. Ahmed SF, Cheng A, Hughes IA. Biochemical evaluation of the gonadotrophin-gonadal axis in androgen insensitivity syndrome. *Arch Dis Child*, 1999; **80**: 324–9.
18. Pierik FH, Burdorf A, Nijman RJM, *et al*. A high hypospadias rate in The Netherlands. *Human Reprod*, 2002; **17**: 1112–15.
19. Moreno-Garcia M, Miranda EB. Chromosomal anomalies in cryptorchidism and hypospadias. *J Urol*, 2002; **168**: 2170–2.

20. Boehmer ALM, Nijman RJM, Lammers BAS, *et al.* Etiological studies of severe or familial hypospadias. *J Urol*, 2001; **165**: 1246–54.
21. Chavhan GB, Parra DA, Oudjhane K, *et al.* Imaging of ambiguous genitalia: classification and diagnostic approach. *Radiographics*, 2008; 28: 1891–904.
22. Denes FT, Cocuzza MA, Schneider-Monteiro ED, *et al.* The laparoscopic management of intersex patients: the preferred approach. *BJU*, 2005; **95**: 863–7.
23. Tomlinson C, Macintyre H, Dorrian C, Ahmed SF, Wallace AM. Testosterone measurements in early infancy. *Arch Dis Child*, 2004; **89**: F558–919.
24. Ahmed SF, Achermann JC, Arlt W, *et al.* Society for Endocrinology UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development (revised 2015). *Clin Endocrinol*, 2016; **84**: 771–88.
25. Lee MM, Donahoe KP, Silverman LB, *et al.* Müllerian inhibiting substance in humans. Normal levels from infancy to adulthood. *J Clin Endocrinol Metab*, 1996; **81**: 571–6.
26. Ahmed SF, Keir L, McNeilly J, Galloway P, O'Toole S, Wallace AM. The concordance between serum anti-Müllerian hormone and testosterone concentrations depends on duration of hCG stimulation in boys undergoing investigation of gonadal function. *Clin Endocrinol*, 2010; **72**: 814–19.
27. Alhomaidah D, McGowan R, Ahmed SF. The current state of diagnostic genetics for conditions affecting sex development. *Clin Genet*, 2017; **91**: 157–62.
28. Belgorosky A, Rivarola MA. Sex hormone binding globulin response to testosterone. An androgen sensitivity test. *Acta Endocrinol (Copenh)*, 1985; **109**: 130–8.
29. Ahmed SF, Cheng A, Dovey L, *et al.* Phenotypic features, androgen receptor binding, and mutational analysis in 278 clinical cases reported as androgen insensitivity syndrome. *J Clin Endocrinol Metab*, 2000; **85**: 658–65.
30. Lucas-Herald A, Bertelloni S, Juul A, *et al.* The long-term outcome of boys with partial androgen insensitivity syndrome and a mutation in the androgen receptor gene. *J Clin Endocrinol Metab*, 2016; 101: 3959–67.
31. Hornig NC, Ukat M, Schweikert HU, *et al.* Identification of an AR mutation-negative class of androgen insensitivity by determining endogenous AR activity. *J Clin Endocrinol Metab*, 2016; **101**: 4468–77.
32. Rodie ME, Mudaliar MAV, Herzyk P, *et al.* Androgen-responsive non-coding small RNAs extend the potential of HCG stimulation to act as a bioassay of androgen sufficiency. *Eur J Endocrinol*, 2017; **177**: 339–46.
33. Kolon T, Miller OF. Comparison of single versus multiple dose regimens for the human chorionic gonadotrophin stimulatory test. *J Urol*, 2003; **166**: 1451–4.
34. Dixon J, Wallace AM, O'Toole S, *et al.* Prolonged human chorionic gonadatrophin stimulation as a tool for investigating and managing undescended testes. *Clin Endo*, 2007; **67**: 816–21.
35. Segal TY, Mehta A, Anazodo A, Hindmarsh PC, Dattani MT. Role of gonadotropin-releasing hormone and human chorionic gonadotrophin stimulation tests in differentiating patients with hypogonadotropic hypogonadism from those with constitutional delay of growth and puberty. *J Clin Endocrinol Metab*, 2008; 94: 780–5.
36. Speiser PW, Azziz R, Baskin LS, *et al.* Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2010; **95**: 4133–60.
37. Sultan C, Biason-Lauber A, Philibert P. Mayer-Rokitansky-Kuster-Hauser syndrome: recent clinical and genetic findings. *Gynecol Endocrinol*, 2009; **25**: 8–11.
38. Ruble DR, Martin CL. Gender development. In: Daman W, Eisenberg N, eds. *The Handbook of Child Psychology*, Vol. **3**, edn 5. New York, NY: Wiley, 1998: 105–76.
39. Keir LS, O'Toole S, Robertson AL, Wallace AM, Ahmed SF. A 5-year-old boy with cryptorchidism and pubic hair: investigation and management of apparent male disorders of sex development in mid-childhood. *Horm Res*, 2009; 71 Suppl 1: 87–92.
40. Kolesinska Z, Ahmed SF, Niedziela M, *et al.* Changes over time in sex assignment for disorders of sex development. *Pediatrics*, 2014; 134: e710–15.
41. Macqueen Z, Sandberg DE, Gardner M, *et al.* Parent reported outcome measures for assessing clinical care in young children with conditions affecting somatic sex development. *Int J Pediatr Endocrinol*, 2020: **3**. DOI: 10.1186/s13633-020-0073-x.
42. Warne GL. Long term outcome of disorders of sex development. *Sex Dev*, 2008; **2**: 268–77.

7.3

Pubertal Disorders

7.3.1 Recognizing Normal and Disordered Pubertal Development

Alan D. Rogol and John S. Fuqua

Definition of Pubertal Maturation

Puberty (Latin *pubertas*, of ripe age, adult) is 'the condition of being or the period of becoming first capable of reproducing sexually that is brought on by the production of sex hormones and the maturing of the reproductive organs (such as the testes and ovaries), development of secondary sex characteristics (such as male facial hair growth and female breast development), and in humans and the higher primates by the first occurrence of menstruation in the female' [1].

We shall use pubertal maturation, a process, to define the physical and hormonal changes as the child becomes an adolescent and then an emerging adult. The psychological/behavioural changes associated with pubertal development will not be further discussed in this chapter.

Pre-pubertally the hypothalamic-pituitary-gonadal (HPG) axis is active during the second trimester of gestation but relatively quiescent following mini-puberty in the first few months of life and preceding pubertal maturation. The outward signs differ in boys and girls. One evaluates pubic hair (both sexes), breast maturation (females), and penile and testicular maturation (boys). These are noted in detail in Table 7.3.1.1 and shown in Figure 7.3.1.1. Individuals vary in the level of maturity attained at a given point in time (maturity status at a given age), in timing (when maturational events occur) and in tempo (rate of maturation) [2]. Thus, in individuals of the same sex and chronological age, wide variations of biological age occur as indicated by physical examination and bone age determination. Not only is there variability in the age of attainment of one or another stage, but there may be significant variability between the stages for breast and pubic hair or genital development and pubic hair within any individual (Tables 7.3.1.2 and 7.3.1.3)

Table 7.3.1.1 Pubertal maturation in males and females

Stage	Physical characteristics
Pubic hair in males and females	
1	Prepubertal
2	Sparse growth of long, slightly pigmented hairs at the base of the penis (males) or mons veneris/labia majora (females)
3	Further darkening and coarsening of hair, with spread over the symphysis pubis
4	Hair is adult in character, but not in distribution, has not spread to the lower abdomen (males) or to the medial surface of the thighs (males and females)
5	Hair is adult in distribution, with extension to the lower abdomen (males) and/or the medial surface of the thighs (males and females)
Breast development in females	
1	Prepubertal
2	Breast budding, widening of the areola with elevation of both breast and nipple as a small mound
3	Continued enlargement of both breast and areola, but without separation of their contours
4	Formation f the areola and nipple as a secondary mound projecting above the contour of the breast
5	Adult shape with the areola and nipple recessed to the contour of the breast
Genital development in males	
1	Prepubertal
2	Enlargement of the testes and scrotum, thinning and reddening of the scrotal skin, penis remains prepubertal
3	Further growth of testes and scrotum; enlargement of the penis, predominately in length
4	Further growth of testes and scrotum with pigmentation of the scrotal skin; further enlargement of the penis, especially in circumference, and development of the glans
5	Testes, scrotum, and penis are adult in size and shape

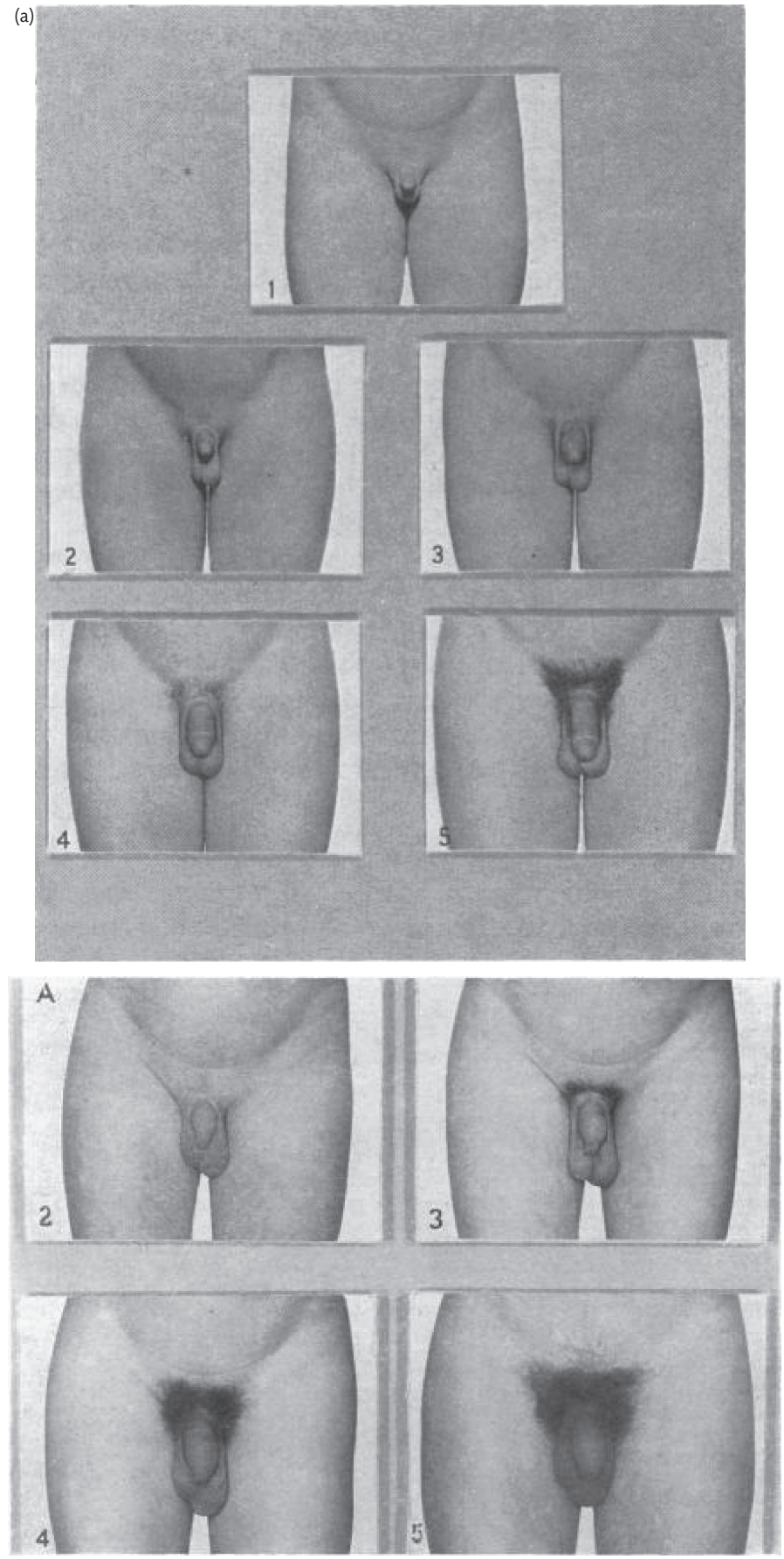


Figure 7.3.1.1 *Continued*

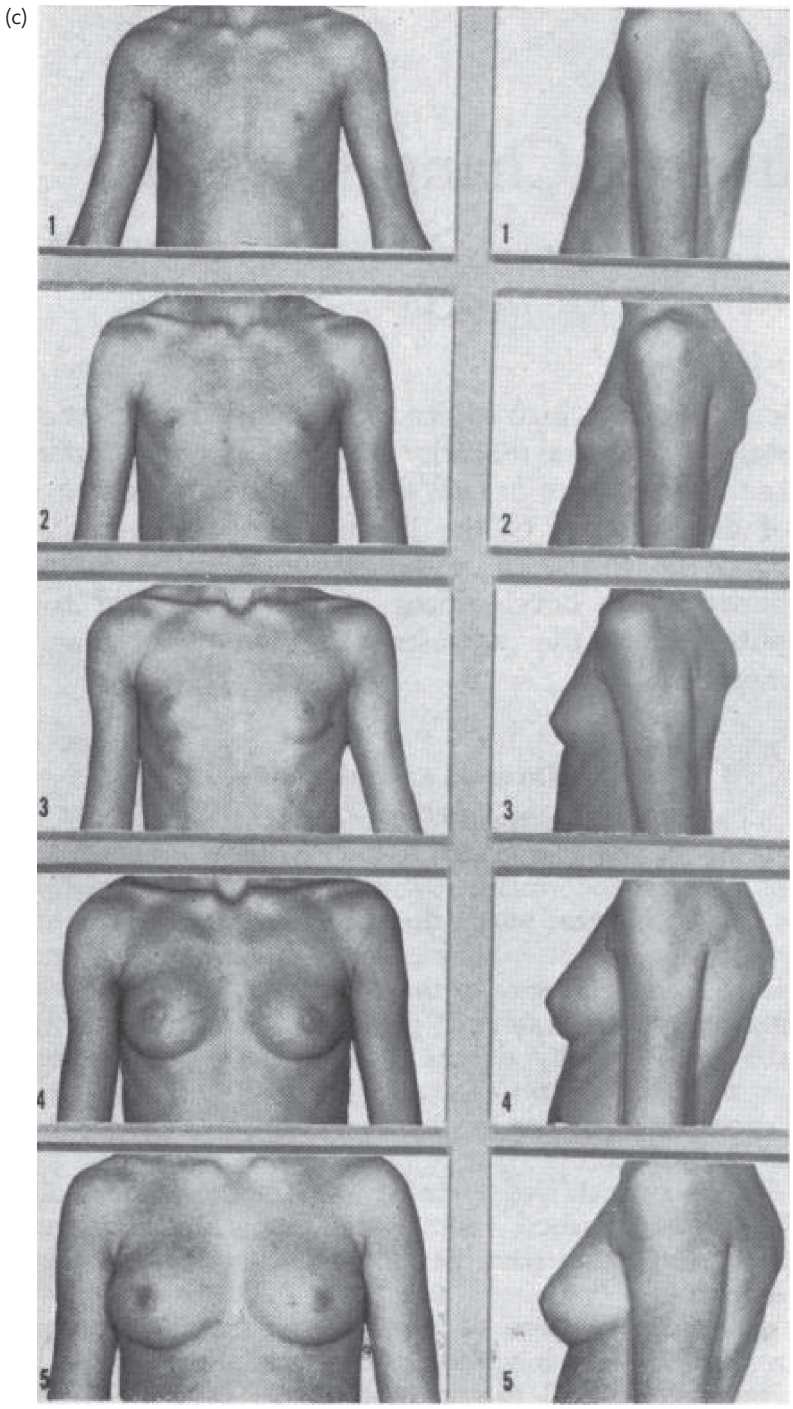


Figure 7.3.1.1 Tanner staging of (a) pubic hair, (b) genital development in boys, and (c) breast development in girls.

Panels (a) and (b) are reproduced with permission from Marshall WA and Tanner JM. Variations in the pattern of pubertal changes in boys. *Archives of Disease in Childhood*, 45:13–23, 1970. Panel (c) is reproduced with permission from Marshall WA and Tanner JM. Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*, 44:291–303, 1969. Copyright © 1970, BMJ Publishing Group Ltd and the Royal College of Paediatrics and Child Health.

[2–4]. Some boys complete the process of moving from G2 to G5 in two years and others may take up to 4.5 years. Boys may enter G2 only 6 months after girls are at stage 2 for breast development, but they differ by almost 2 years in age at peak height velocity (PHV), a maturational benchmark or biological anchor of pubertal maturation. This translates to girls having PHV often at B2 and boys in the latter part of G3 or G4.

Tables 7.3.1.2 and 7.3.1.3 show the variability of entry into a stage and progression on to the next stage. However, one of the limitations of the data is that they are cross-sectional and it requires a longitudinal study to evaluate the 'kinetics' or trajectory of an individual. Additionally, there are limitations of the sexual maturity rating (SMR) process, especially at the earliest stages when palpation is mandatory to distinguish breast budding from

Table 7.3.1.2 Percentages of Swiss girls in each stage of breast development when they reached each stage of pubic hair development (top); and in each stage of pubic hair development when they reached each stage of breast development (bottom)

Pubic hair stage	Percentage in each breast stage				
	B1	B2	B3	B4	B5
PH2	49	46	5	0	0
PH3	0	36	51	12	1
PH4	0	4	47	33	16
PH5	0	1	6	34	59
Breast stage	**Percentage in each pubic hair stage**				
	PH1	PH2	PH3	PH4	PH5
B2	16	67	16	1	0
B3	2	26	50	20	2
B4	0	4	28	43	25
B5	0	0	7	29	64

B, breast; PH, pubic hair.

adipomastia or as the testis changes from a prepubertal volume. The prepubertal testis has a volume ≤3 ml, or ≤2.5 cm in longest diameter, while at stage 5 the testis is 12–25 ml or 2.8–5 cm in longest diameter.

The sequence of pubertal progression is predictable and once entrained usually goes to completion. However, there is variability in the rate of progression just as there is in its onset. Thelarche is usually the first sign in more than 90% of girls, followed by sexual hair a year or more later. However, pubarche may occur first (and/or simultaneously with breast development) in a small percentage of girls [5].

There are very significant changes in the shape and composition of the adolescent body as it progresses through male or female maturation. These changes include the maturation of body tissues such as increases in haemoglobin concentration (much later in boys than girls, at SMR 4 or 5); alkaline phosphatase, especially the bone fraction as an indicator of bone osteoblastic activity during rapid bone growth; a small increase in blood pressure [2]; the amount and the distribution of adipose tissue; and both bone mass and fat-free lean tissue mass [6]. Significant increases in total energy requirement (boys greater than girls) accompany, and likely drive, the changes in body composition. Body measurements at any one time do not indicate the state of maturity; however, longitudinal data that span adolescence may be evaluated for specific growth parameters that do indicate maturity status: age of onset of the growth spurt and age at PHV [2]. Growth spurts in other dimensions also occur but not necessarily in synchrony, as may be seen in Table 7.3.1.4 [2, 7]. In both sexes the peak velocity in weight, lean mass, and bone mineral content occur after peak height velocity. These parameters reach their young adult values often in the late second or early third decade, especially for males. This pattern has physical performance implications for both sexes, as the strength of adolescent boys accelerates markedly compared to girls relatively early in pubertal maturation [8, 9] and to a greater degree in males. In any event it is more strongly correlated to the biological parameter of peak height velocity than to chronologic age [8]. The development of muscle strength not only depends on the neuroendocrine axes for growth hormone and the gonads, but also depend on neural and biomechanical factors. Sex differences emerge during puberty, with boys demonstrating accelerated increases in strength, but girls continuing to increase strength at a similar rate as in prepuberty, allegedly due to neuromuscular facilitation rather than increased muscle bulk [10]. By late puberty boys and girls may differ in strength by ~50% [11].

It should be remembered that there are consequences of pubertal body composition changes that remain as risk factors for common diseases in the adult, for example, cardiovascular disease. Some of these, especially increased fat mass, often track into adulthood [6].

Table 7.3.1.3 Percentages of Swiss boys in each stage of genital development when they reached each stage of pubic hair development (top); and in each stage of pubic hair development when they reached each stage of genital development (bottom)

Pubic hair stage	Percentages in each genital stage				
	G1	G2	G3	G4	G5
PH2	9	54	33	4	0
PH3	0	9	49	37	5
PH4	0	0	6	64	30
PH5	0	0	0	20	80
Genital stage	**Percentage in each pubic hair stage**				
	PH1	PH2	PH3	PH4	PH5
G2	63	36	1	0	0
G3	15	50	32	3	0
G4	0	9	42	44	5
G5	0	1	10	34	55

G, genital; PH, pubic hair.

Table 7.3.1.4 Ages at attainment and peak velocities of aspects of body composition in boys and girls

	Girls		Boys	
	Mean	SD	Mean	SD
Measurement				
Ages at peak velocity (yr)				
Stature	11.8	0.9	13.4	1.0
Lean mass	12.1	1.0	13.7	0.9
Weight	12.3	1.2	13.8	1.1
Fat mass	12.6	2.0	14.0	1.3
Bone mineral content	12.5	0.9	14.0	1.0
Peak velocities				
Stature (cm/yr)	8.6	1.1	10.4	1.2
Lean mass (kg/yr)	5.2	1.2	8.8	1.6
Weight (kg/yr)	8.7	1.4	10.3	1.9
Fat mass (kg/yr)	−0.4	1.8	−1.9	2.2
Bone mineral content (g/yr)	325	67	407	93

Hormones of Puberty

Regulation of Gonadal Steroid Production

Gonadotropin releasing hormone (GnRH) is secreted from the hypothalamus in a pulsatile fashion. A complex neuroendocrine regulatory mechanism involving multiple stimulatory and inhibitory inputs contributes to GnRH secretion. GnRH secretion in prepubertal children as measured by LH concentration is characterized by low and irregular pulses occurring every 3–4 hours, and serum concentrations of LH and FSH are low. As the child approaches puberty, both the amplitude and frequency of GnRH pulses increase during the night-time hours, with pulse frequency approximating once per hour and returning to baseline by morning (Figure 7.3.1.2). This phenomenon leads to a diurnal variation in testosterone secretion in boys and to a lesser extent, oestradiol concentrations in girls. As puberty progresses, the changes in GnRH pulses extend to the daytime hours, stabilizing at a period of every 90–120 minutes, with corresponding increases in afternoon concentrations of sex steroids [12]. However, the diurnal variation in testosterone remains noticeable in adult men throughout the lifespan. Changes in GnRH pulse frequency alter relative production of FSH and LH, with slower pulsatility increasing FSH production. In adult women, pulse frequency varies over the menstrual cycle, ranging from every 60–100 minutes [13].

An interesting phenomenon is the transient activation of GnRH, gonadotropin, and sex steroid secretion that occurs in the first few months of life, termed the 'minipuberty of infancy'. In boys, testosterone concentrations fall in the first weeks after birth. By 4 weeks of age, however, testosterone levels rise and peak in boys at 7.2 ± 2.3 nmol/L at about 1–3 months of age, approximating levels seen in early-mid puberty [14]. By 5–6 months of age, testosterone concentrations fall and reach otherwise normal prepubertal ranges in the second half of the first year. In girls, oestradiol levels do not have an obvious peak but fluctuate over time, decreasing by 6–12 months of age [15]. Gonadotropin levels in boys are sufficient to cause increases in testicular volume, and testosterone concentrations are high enough to cause penile growth and acne [16]. In girls, oestradiol levels during the minipuberty contribute to breast bud and uterine growth [17].

As puberty approaches, increases in GnRH pulsatility can be measured in both boys and girls as subtle increases in LH levels years before the clinical onset of puberty. These increases have been detected in serum as well as in urine [18, 19]. As clinically apparent

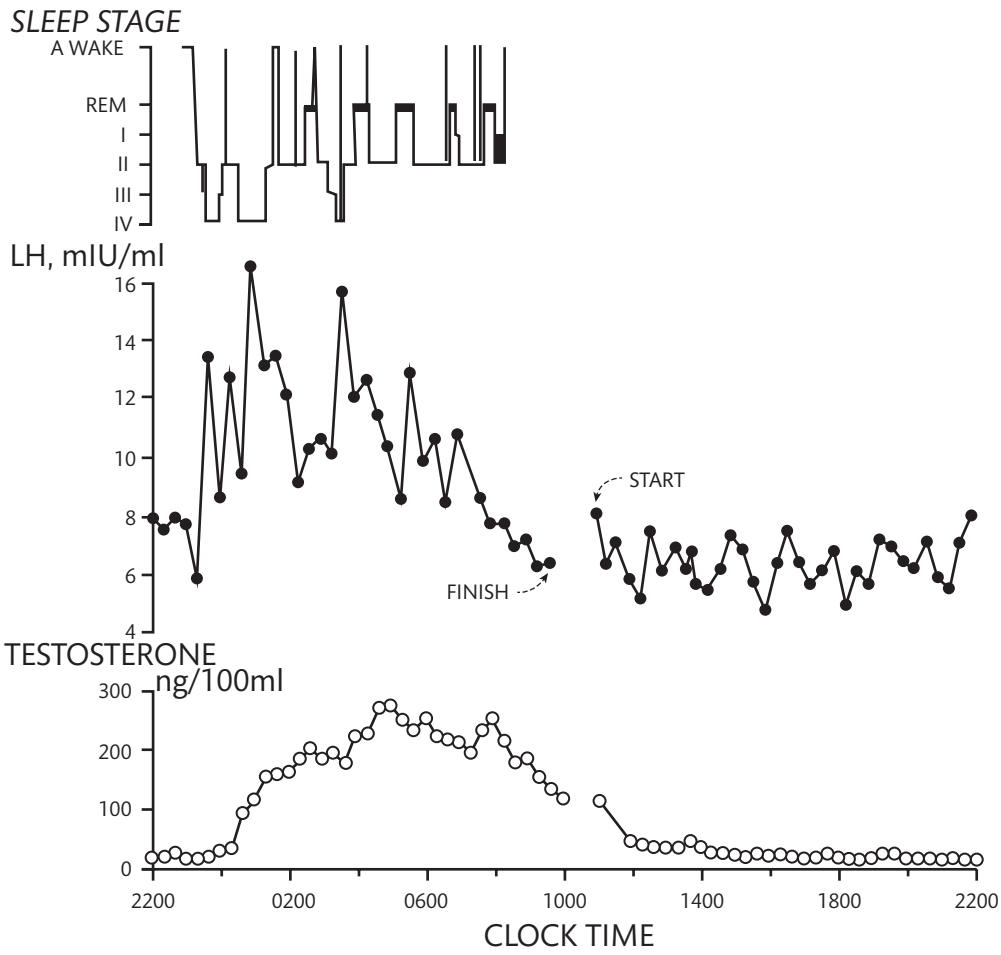


Figure 7.3.1.2 LH and testosterone secretion in a boy in the early stages of pubertal maturation demonstrating increased testosterone concentration and LH pulse amplitude and frequency during sleep, with decreases in both amplitude and frequency in the morning.

changes occur with pubertal maturation, concentrations of sex steroids increase concurrently [20] and eventually reach adult levels.

Gonadal Peptide Hormones

In addition to the sex steroids testosterone and oestradiol, other important products of the testes and ovaries are the peptide hormones inhibin and anti-Müllerian hormone (AMH). Both are products of the Sertoli cells of the testes and the ovarian granulosa cells. Inhibin exists in two major isoforms, inhibin B and inhibin A, with inhibin B produced primarily during the follicular phase and inhibin A during the luteal phase [21]. Inhibin secretion is stimulated by FSH and functions in both males and females as part of a negative feedback loop to suppress pituitary release of FSH. Anti-Müllerian hormone plays an important role in the development of the male internal genitalia in the early weeks of gestation. In the adult ovary, AMH appears to decrease the growth of primordial follicles. Measurement of AMH has clinical utility as a reflection of the number of primordial ovarian follicles and possibly in the assessment of polycystic ovary syndrome.

Serum concentrations of AMH and the inhibins change between infancy and adulthood. In boys, AMH concentrations are high at birth and gradually decline during prepubertal life, with levels averaging 90–100 ng/ml at age 7–10 years. At the onset of puberty, concentrations fall rapidly to reach levels similar to those seen in females. In girls, AMH concentrations are low at birth and remain modest in comparison to males. Levels increase slightly around the time of puberty, but average 2–3 ng/ml in adolescent girls [22]/

By contrast, concentrations of inhibin B gradually rise during the prepubertal years. Several investigators have studied this relationship of the rise in inhibin B to the clinical manifestations of puberty. In a prospective, longitudinal study of a small number of boys and girls, Crofton *et al*. found progressive increases in inhibin B from genital Tanner stages 1 through 3 in boys and increases in both inhibin A and B between breast Tanner stages 1 and 4 in girls [23]. In a series of much larger cross sectional studies of data derived from NHANES III, Lee and colleagues [24, 25] found a gradual increase of inhibin B in both boys and girls through late childhood and into early adolescence, with the majority of prepubertal girls having levels below the limit of detection. On a population basis in girls, inhibin B was modestly useful for predicting the onset of puberty as established on clinical grounds, with an optimal cut-point on receiver operating characteristic (ROC) analysis of 17.9 pg/ml, but relatively high false positive and negative rates [24, 25]. However, in boys the degree of interindividual variation rendered inhibin B alone unhelpful to establish the onset of puberty [20].

Growth Hormone/Insulin-like Growth Factor-1 (IGF-1)

A hallmark of puberty is acceleration of height velocity, often called the pubertal growth spurt. Growth hormone is released in a pulsatile fashion, with more than 75% of daily release occurring at night. The most reliable pulse of growth hormone secretion occurs 45–90 minutes after the onset of the first cycle of deep sleep, with a variable number of lesser pulses through the night [26]. With the onset of puberty and the rise in oestradiol and testosterone concentrations, the amount of growth hormone released increases by a factor of approximately three. Serum concentrations of IGF-1 increase by a similar amount [27]. In addition to directly stimulating growth hormone secretion, aromatizable androgens potentiate the effect of growth hormone on IGF-1 release. Additionally, growth hormone has a direct effect on ovarian granulosa cells to increase steroidogenesis, and IGF-1 acts to increase ovarian FSH receptors [28]. With progression of puberty, oestrogen concentrations rise in both sexes, ultimately resulting in maturation of the epiphyseal growth plates and cessation of linear growth. The higher concentration and earlier secretion of oestrogens in girls appears to account for the earlier occurrence of the pubertal growth spurt and the shorter average adult height in girls vs. boys [29].

Regulation of Pubertal Timing

The timing of puberty is a complex trait, with a Gaussian distribution in the general population. It is a dynamic interaction between genes and the environment, with influences including nutritional factors, environmental influences, and genetic input. It is estimated that between 50% and 80% of the variance of normal pubertal timing is explained by genetic factors [30, 31]. Efforts to understand the genetics of pubertal timing have led to the discovery of many genes that are clearly necessary for pubertal maturation. Many of these genes, when mutated, lead to specific syndromes of delayed or absent puberty. However, it is not clear that these genes individually or as a group explain much of the variability of the onset of **normal** puberty, and despite the identification of these involved genes, our understanding of the genetic trigger(s) for the onset of pubertal maturation remains incomplete. Genome-wide association studies and other high throughput technologies have shown promise in elucidating regulatory systems for complex traits. Additionally, systems biology approaches have demonstrated multi-layered networks of genetic control [32]. These networks comprise lower level primary mechanisms of GnRH secretory control and higher levels of transcriptional regulation, theorized as a system of network 'nodes' [33]. Primary GnRH secretory control is exerted through stimulatory and inhibitory neuronal inputs as well as excitatory glial-neuronal interactions. Higher level control networks seem to be largely inhibitory and may themselves be controlled in part by epigenetic mechanisms. Putative members of higher-level control networks include *IRF2BPL* (*EAP1*), *CADM1* (*SynCAM*), *POU2F2* (*OCT2*), and *TTF1* [33].

In 2003, the kisspeptin/GPR54 system was discovered to be a key regulatory factor in the initiation of increased GnRH pulsatility at the onset of puberty [34]. In the human, kisspeptin is produced in neurons in the infundibular nucleus and preoptic area of the hypothalamus and interacts with its receptor, GPR54, on GnRH secreting neurons. A subpopulation of kisspeptin neurons, termed KNDy neurons, coexpress neurokinin B and dynorphin, Multiple factors influence kisspeptin secretion, including neurokinin B and dynorphin, which are thought to be stimulatory and inhibitory factors, respectively [35]. Kisspeptin expression increases with the physiologic onset of puberty in several mammalian species. In multiple animal species, including humans, exogenous administration of kisspeptin increases the pulse amplitude of GnRH. Administration of kisspeptin to healthy male and female volunteers increases gonadotropin concentrations [36, 37]. In humans, inactivating mutations of this system have been associated with delays in puberty, while activating mutations lead to precocious puberty [38]. The kisspeptin/GPR54 system

may also be involved in the negative and positive feedback effects of sex steroids on GnRH secretion. Despite these lines of evidence, it remains unestablished if kisspeptin is the primary trigger of puberty or if it acts as an amplifier [35].

GnRH secretion is negatively regulated by makorin ring finger protein 3 (MKRN3), a hypothalamic factor that may act directly on GnRH secreting neurons or may act via the KNDy neurons [39]. Serum concentrations of MKRN3 decrease as normal puberty progresses in boys and girls [40]. *MKRN3* is an imprinted gene on chromosome 17q11–q13 that is silenced on the maternal allele. Thus, inactivating mutations lead to autosomal dominant precocious puberty, but only when paternally inherited. Methylation defects appear to be uncommon. *MKRN3*-related precocious puberty affects both girls and boys, with a median age at pubertal onset of 6 and 8.25 years, respectively [41].

Gamma-aminobutyric acid (GABA)-secreting neurons play an inhibitory role in puberty and may be involved in the juvenile pause, the temporary suppression of the hypothalamic-pituitary-gonadal axis that occurs during childhood. Suppression of GABAergic neuronal input may lead to the initiation of puberty [42]. Additional factors may include increased glutamate stimulation of N-methyl-D-aspartate (NMDA) receptors, which are stimulatory to GnRH release [43], and an decreased negative feedback effect of low levels of circulating sex steroids. Investigation of this phenomenon is difficult due to species specificity [44].

Epigenetics

Multiple lines of evidence suggest that epigenetic changes may play roles in the timing of puberty. Major epigenetic mechanisms include DNA methylation as well as acetylation and methylation of histones. In rodent models, all three mechanisms alter the activity of kisspeptin neurons and GnRH neurons. Alterations in epigenetic markers may occur during pubertal maturation, including increases in histone acetylation that increase kisspeptin mRNA transcription and methylation changes in promotor region of the *GNRH* gene in rodents and non-human primates [35].

MicroRNAs (miRNAs) provide an additional epigenetic mechanism for regulation of gene expression and translation. MicroRNAs are approximately 22 nucleotide-long non-coding RNAs that transcriptionally silence target genes by increasing mRNA degradation. Individual miRNAs may affect dozens of target genes, while target genes may each be controlled by several miRNAs, creating a diverse network of regulatory control. Evidence for miRNA control of pubertal timing comes primarily from studies of mice. Several distinct miRNAs either activate or suppress hypothalamic–pituitary–gonadal function, both at the level of the hypothalamus and by altering gonadal maturation [35]. There is limited evidence suggesting a role of miRNAs in human pubertal maturation. GWAS demonstrated an association between age at menarche and changes at chromosome 6q21 near the *LIN28B* gene, implicated in processing of the let-7 family of miRNAs, shown in animal studies to be a pubertal activator [45].

Environmental Effects

Because development and maintenance of reproductive competency is an energy-intensive process, it is unsurprising that puberty is metabolically gated. In general, factors that result in negative energy balance delay the onset or progression of pubertal maturation, while positive energy balance, such as occurs with obesity, tends to accelerate the process.

Studies of the relationship of obesity to the timing of pubertal maturation in girls have demonstrated earlier onset of breast development in obese girls. A recent meta-analysis reviewing 11 studies of a total of 4841 subjects showed a relative risk of 2.44 for obese girls to have early breast development. Interestingly, although some of the analysed studies showed a relationship between obesity and the age at menarche, this did not reach statistical significance in the combined meta-analysis [46]. Data in boys were less consistent, and the investigators were unable to demonstrate a relationship between obesity and genital development. This was due in part to the low number of male subjects in included studies. It is likely that in boys different degrees of obesity produce different effects on the timing of pubertal maturation.

A key factor that links body fat stores to reproductive maturation is leptin, an adipocyte-derived signalling molecule. Indeed, humans carrying mutations in either the gene encoding leptin or its receptor demonstrate disruption of puberty, and leptin replacement in deficient individuals helps to restore hypothalamic–pituitary–gonadal function [47]. Leptin appears to be a permissive factor for puberty in both males and females, in that threshold levels are required for pubertal initiation. However, leptin in and of itself is not sufficient. The mechanism of action for this permissive effect of leptin is not well-understood. Kisspeptin expression is increased by leptin [48], but whether this is due to a direct effect on KNDy neurons or is driven by intermediary cells is unclear. Such intermediates may include POMC neurons and GABAergic neurons [35]. Other possible links between energy stores and reproductive maturation include ghrelin, proposed to have inhibitory effects, and insulin, which may act within the CNS as a permissive or stimulatory factor.

In addition to classical endocrine and neuroendocrine systems, overfeeding may lead to earlier pubertal maturation by inducing epigenetic changes, especially during key points in development, such as fetal life, early post-natal life, and puberty. Evidence for this comes predominately from animal studies, with limited human data [49].

An interesting phenomenon is the higher prevalence of early puberty in girls who have been internationally adopted. Theories for why this occurs have included the effects of early nutritional deprivation with later catch up weight gain, a response to psychologic stress, lack of accurate age determination, or as a result of environmental endocrine disrupting chemical exposure [50–52].

Environmental Endocrine Disruptors

Endocrine disrupting chemicals (EDCs) are broadly defined as naturally occurring or synthetic chemicals that have stimulatory or inhibitory effects at hormone-responsive tissues leading to alterations in normal developmental or physiologic processes. From a reproductive standpoint, most EDCs have estrogenic or anti-estrogenic actions, anti-androgenic actions, or a combination and may have effects at the level of the brain, gonads, or target tissues. They have been linked to alterations in decreased sperm counts, infertility, gonadal cancers, and genital malformations. Although a single EDC may have minimal measurable effects at environmental concentrations, human exposures are typically multiple, and the possibility of additive or synergistic effects is a concern [35]. Most of our understanding of the estrogenic effects of EDCs comes from animal studies, although there are both prospectively and retrospectively collected human data.

Exposure to EDCs has been postulated to contribute to earlier age at onset of puberty [53]. Interpretation of human studies is challenging as a result of variations in the aspects of puberty that were assessed and characteristics of the exposures, such as the specific EDCs studied, times of exposure, exposure doses, and the effects of other EDCs, either those assessed or unknown exposures. In a study of 200 adolescents living in a polluted area of Belgium, delays in genital development and pubic hair growth in boys were associated with serum concentrations of polychlorinated biphenyls (PCBs), and delayed breast development in girls occurred when serum dioxin concentrations were high [54]. Similarly, Leijs *et al.* reported results from a prospective study that related prenatal and perinatal exposure to dioxin-like compounds with delayed onset of breast development in girls and a trend to delayed age at first ejaculation in boys [55]. Other investigators have been unable to demonstrate large effects of similar EDCs [56, 57].

Many of the action of EDCs are thought to occur through their direct effects on oestrogen receptors (ER), particularly for high-dose exposures. However, in animal studies very low-level exposure has been shown to cause abnormal development through alternative mechanisms. Bisphenol A (BPA) is a plasticizer found in many bottles, cans, and other items that come in contact with food. It is estimated that 90% of the US population has detectable BPA concentrations in urine [58]. BPA is a selective oestrogen receptor modulator (SERM) that acts both at ERα and ERβ. Low dose in utero BPA exposure in mice altered expression of DNA methyltransferases in the fetal CNS, enzymes involved in epigenetic gene methylation. These alterations were associated with changes in methylation of the ERα gene in the same brain regions, which led to altered ERα gene expression. Exposed female offspring demonstrated disrupted sexually dimorphic behaviours, indicating long-lasting effects even at low doses [59].

Secular Trends in Pubertal Maturation

Secular or long-term temporal trends for human puberty, menarche, and reproduction have been of interest for centuries. One should distinguish the issue of *timing*—the chronological ages at which certain events within the pubertal maturation scheme occur—from *tempo*, which refers to the time taken to pass through the various stages of pubertal maturation. In terms of an evolutionary/developmental model, data from skeletal ages of remains (e.g. limb length) suggest that in Palaeolithic women menarche occurred at ages between 7 and 13 years (slightly younger or comparable to present ages), early sexual maturation being an evolutionary trade-off of a reduced life expectancy [60]. There was an apparent stabilization at age 12–15 years through the Classical and Medieval times, which are not so different from the present era. During the mid-nineteenth century Industrial Revolution there was an increase to 15.5 to 17.4 years [61, 62], with adolescents from the higher socioeconomic strata at the lower end of that age range, perhaps reflecting good general health and nutritional status. The increase in the age at menarche during the mid-nineteenth century was considered to be related to deleterious living conditions in which hygiene deteriorated, the population density increased, nutrient availability was diminished, and infectious diseases were common.

Investigators have since shown a secular trend toward a decreasing age at menarche from the mid-19th century (~16–17 years) as sanitation and living standards improved and the burden of infectious diseases diminished, especially water and food borne illnesses [60]. This general trend occurred at a rate as high as 12 months per decade, decreasing to perhaps 3 months per decade in the early and mid-twentieth century. Over the last 4–6 decades, this trend has levelled off, except for continuing small decreases in age at menarche in some areas of the developing world [60]. A similar trend was noted in boys using the timing of voice breaking as the criterion for pubertal maturation.

More recently, there has been interest in other indicators of pubertal maturation besides menarche, including indicators for boys. Trends have been noted in adult height, with a gain of 1–3 cm per decade as well as in weight, although the latter has been more variable [63]. There is significant heterogeneity in these changes related to ethnic, geographical, and socioeconomic factors, including 'modernization' that has included decreases in family size and child/early adolescent labour. Stress and endocrine disrupting chemicals (EDC) have also been implicated [64]. It is likely that early nutritional alterations have also played a role. There is a significant influence of birth weight as an output variable of environmental factors acting *in utero* [63]. Especially strong is the influence of the rate of weight gain in the first 1000 days from conception (developmental origins of health and disease, DOHaD, or Barker hypothesis) [63, 65] on the timing and tempo of pubertal maturation and subsequent menarche in girls [66]. During these first 2 years the child maintains developmental plasticity [65] or the ability to modify structure and function in response to environmental cues, especially those present *in utero* [63]. These permanent effects are thought to occur by epigenetic mechanisms in response to the mismatch of the initial environment (e.g. poor nutrition *in utero)* to the later environment (e.g. a surfeit of energy availability). Some girls with lower birth weight and rapid early, post-natal weight gain enter puberty early and mature more rapidly than peers of normal birth weight, often resulting in short stature as adults [67].

Recent trends show slightly more than a year earlier start of pubertal maturation in girls, but a much smaller decrement in the age at menarche, indicating a slower tempo from the start of maturation to its completion [68]. Are there longer-term effects of this modern trend? There likely are, with respect to earlier menopause; reproductive system neoplasia, breast cancer; and the metabolic syndrome linked to the long-term consequences of obesity [64].

Secular changes can occur even over the span of several decades, especially within an isolated society that may undergo demographic and epidemiological shifts. Demographic shifts include a change from relatively low population growth and a small difference between birth and death rates to a rapid population growth and a divergence of births and deaths. Epidemiologic shifts include a change from high infectious disease mortality to an increase in mortality from degenerative (non-communicable) diseases in older adults [69]. This is precisely what has happened in the Valley of Oaxaca in southern Mexico, with a direct correlation between changes in maturation and demographic and epidemiological transitions within the Zapotec-speaking indigenous population as chronicled over 4 decades by Malina and his coworkers [69]. The data show changes in body size (height, weight, and BMI all increase) and in the age of menarche over the time of observation. The age at menarche (15 years) was stable in those born between 1896 and 1954 before the demographic change. However, there was a decrease of ~0.51 yr/decade for those born from 1955 to 1981 (14 years) at the

onset of the demographic transition and an accelerated decrement of 0.78 yr/decade for those born between 1983 and 1992, with subsequent menarche at approximately 13 years [69]. This insular and very well evaluated and reported population is likely a microcosm for the more broad-based populations for which there are data over centuries.

An interesting set of data have come from the Fels Longitudinal Study [6] in which two cohorts of females were followed longitudinally. One cohort was born in the years 1929–1954 and the second in 1955–1982. All had data obtained until at least age 21 years, and there was an average of 21 serial observations. The initial BMI 6 years before menarche was indistinguishable between the two cohorts. The final BMI 6 years beyond menarche was 1.2 kg/m^2 greater in the more recently born women. This difference was statistically significant. In fact, the later-born cohort began to have greater BMI four years before menarche. Yet at all times the heights at each year were indistinguishable between cohorts. This is consistent with those in the later cohort having more adipose tissue compared to those in the earlier cohort. There were no significant differences between the two cohorts (that is, no secular trend) in attaining ages at early childhood minimum height velocity (MHV), prepubertal MHV, PHV, or age at menarche. From a larger database, Chumlea and colleagues concluded that there is no secular trend in the age at menarche in US girls over the past few decades [70]. Thus, it seems in the general US population that the secular trend in menarche has vanished over the past 4 or more decades; however, that does not mean that in isolated populations, as noted by Malina *et al*. [69], rapid changes in socioeconomic variables can still support secular changes in pubertal maturation, including age at menarche.

Tempo of Pubertal Maturation

Pubertal tempo is a measure of the rate of individual change over time. It may refer to passage between individual stages or across multiple stages; for example, from the beginning (B2) to the end (B5 or menarche). The data of Marshall and Tanner showed that traversing genital stage 2 to stage 5 in boys required approximately 3 years, with a range of 1.86–4.7 years. In girls the transition from B2 to B5 took an average of 4 years, but with a striking range, 1.51 to nearly 9 years. It should be noted that some girls mature to B4 but not B5 and others may 'skip' B4 maturation stage [71, 72].

Flor and colleagues noted the synchrony of the magnitude of skeletal age maturation and the magnitude of pubertal acceleration or delay in their study of boys with sexual precocity due to congenital adrenal hyperplasia, familial male-limited precocious puberty, or delayed sexual maturation due to constitutional delay of growth and puberty [73]. The synchronous maturation was not related to weight, height nor BMI. The hypothesis, at least in boys, is that skeletal maturation influences the HPG axis maturation.

Quantitative data for girls were presented by Pantsiotou and colleagues, who noted differences in first to third quartile maturation compared to the slowest fourth quartile for passage through several stages of sexual maturation (Table 7.3.1.5) [74]. These data are concordant with those of Rosenfield and colleagues, who showed that those girls who began pubertal development early had a slower course through pubertal maturation to menarche [68].

Table 7.3.1.5 Rates of pubertal maturation in girls. Data presented as years (IQR)

Maturation stages	Early	Average	Late
B2 to PHV	1.6 (1.0–2.2)	1.1 (0.7–1.7)	0.9 (0.3–1.1)
PHV to menarche	0.8 (0.6–1.2)	0.8 (0.4–1.3)	0.9 (0.5–1.3)
B2 to menarche	2.3 (2.0–2.8)	2.1 (1.7–2.8)	1.6 (1.4–2.4)

B2, Breast Tanner stage 2; IQR, interquartile range; PHV, peak height velocity.

Activity in sport has been proposed as a modifier of growth and pubertal maturation; however, at least in some aesthetic sports, for example, artistic gymnastics, distance running, and ice skating, body size is apparently a selection criterion [75]. In small subsets of male and female adolescents in the Wroclaw, Poland Longitudinal Growth Study, Malina and Bielicki found that the heights and weights of boys who were active in sport and training followed the growth trajectory of early maturing boys [76]. The growth trajectories for girls who were active in sport and training were indistinguishable from those of the normal, non-training girls [76].

Distinction of Early Normal Puberty from Precocious Puberty

At times, the challenge for the clinician is to determine whether a patient's pubertal changes occur because of normal maturational processes that happen at the younger end of the age spectrum or because of pathological processes occurring later in childhood. Precocious puberty may take numerous forms, as discussed in Chapter 7.3.3, and an extensive discussion is beyond the scope of this chapter. However, there are a number of considerations when trying to distinguish pathologic vs. early normal pubertal maturation.

1. **Age at onset**. Much has been published about the younger end of the normal range of pubertal onset. Traditional teaching has been that it is abnormal for signs of puberty to occur before age 8 years in girls and 9 years in boys. However, the data upon which these ages are based were derived from relatively racially and ethnically homogenous populations and were obtained before the dramatic increase in the prevalence of childhood obesity. More recent studies of racially/ethnically diverse children show that puberty may begin earlier in African-American and Mexican-American children than in White children and that pubertal maturation in general may begin earlier [77]. While the younger end of normal is a subject for further debate, the clinician should have a high index of suspicion when pubertal changes occur before these ages.
2. **Pattern of maturation**. Physiological puberty usually follows a characteristic sequence in which breast development and testicular enlargement occur in temporal relationship to pubic hair and other signs of maturation. Physical changes occurring out of sequence or missing elements of maturation (e.g. pubic hair without testicular growth) should be viewed with suspicion.
3. **Tempo of maturation**. Pubertal changes that progress more rapidly than normal may indicate a pathological cause (e.g. adrenal tumour).

4. **Rate of linear growth**. There is a predictable relationship of the pubertal growth acceleration to the other physical changes in boys and girls, with peak height velocity occurring earlier in the maturational process in girls than in boys. When signs of pubertal maturation are noted without accompanying normal growth acceleration, a normal variant such as premature adrenarche or premature thelarche may be present.
5. **Skeletal maturation**. There is a close relationship between skeletal maturation (bone age) and the onset of the pubertal growth spurt. Both boys and girls with normal puberty begin the pubertal growth spurt when 85% of adult height is attained [78]. When translated into bone age, this corresponds to a bone age of 11.0 years for girls and 13.5 years for boys using the methods of Bayley and Pinneau and Greulich and Pyle [79]. Thus, an obvious linear growth acceleration at significantly younger bone ages should raise concern for abnormal puberty.
6. **Laboratory testing**. Endocrine testing may confirm the presence of pathologic causes of early puberty, such as suppression of LH secretion in response to GnRH analogue administration in cases of peripheral precocious puberty or high concentrations of 17-hydroxyprogesterone in the 21-hydroxylase deficiency variant of congenital adrenal hyperplasia. Of note, there is a large degree of overlap in gonadotropin concentrations between normal prepubertal children and those in the early stages of normally timed or precocious central puberty, making the measurement of LH and FSH levels an insensitive approach to assessing the onset of puberty. Additionally, demonstration of normal pubertal changes in endocrine testing will not by itself distinguish central precocious puberty from normal timing.
7. **Imaging**. Similar to laboratory testing, imaging of the CNS, adrenal glands, or pelvis may identify pathologic causes of puberty. Demonstration of ovarian enlargement on pelvic imaging suggests gonadotropin-mediated puberty, but will not distinguish central precocious puberty from normal timing.

Distinction of Late Puberty (Physiologic) from Pathologic Delay

When evaluating a child with delayed puberty, it is important to distinguish those with true pathology, for example, idiopathic hypogonadotropic hypogonadism, from those with 'late blooming' or constitutional delay of growth and puberty (CDGP). The latter is characterized by a *delay* in the timing and potentially the tempo of maturation. Height velocity appears slow as affected children continue to grow at their previous prepubertal rate or slower, compared to their peers who are undergoing the pubertal growth spurt. The evaluation should include determination of a child's bone age, which appears younger than chronological age [80]. When plotted on the growth chart against bone age rather than chronological age, the height of a child with CDGP appears normal and more consistent with mid-parental target height. A family history of CDGP is not uncommon, and children often reach their genetic potential *without* intervention. Spontaneous puberty occurs by 18 years of age; however, testosterone therapy has the potential to reduce some of the psychosocial distress that frequently accompanies delayed puberty.

Although some studies have shown little effect of CDGP on psychosocial well-being [81], others demonstrate an association with poor psychological outcomes. Some of these negative outcomes may persist into adolescence and emerging adulthood. The Oakland Growth Study, for example, found that late maturing boys were not only perceived as less mature and less attractive than peers, but also had more feelings of inadequacy, which persisted into emerging adulthood [82]. Delayed pubertal timing has also been associated with higher rates of depression, anxiety, and substance use in adolescence and young adulthood. Academic achievement may be adversely affected as well [81].

Delayed puberty may also occur in those with permanent hypogonadism, caused by failure to produce sufficient levels of testosterone, sperm, or both [83]. The differential diagnosis between these two categories, especially the latter, is complex and discussed in greater detail in other chapters (7.3.2 and 2.3.1) in this text. Clinical and laboratory findings to distinguish those with CDGP from those with idiopathic hypogonadotropic hypogonadism are noted in Table 7.3.1.6.

Nutritional and Other Non-Hormonal Causes of Delayed Puberty

Worldwide, malnutrition caused by inadequate intake of calories, protein, or both is by far the largest factor in poor growth, stunting, and delayed puberty. The greatest impact is on infants and toddlers, often with lifelong consequences of stunting and inadequate psychological and intellectual development. This may manifest by being born small-for-gestational age or as infectious diseases and diarrhoea during infancy. Although these factors may continue through childhood and manifest as delayed pubertal maturation (and development), there is less documentation.

Table 7.3.1.6 Clinical and laboratory findings in IHH and CDGP

Feature	IHH	CDGP
Hormonal profiling during 'mini-puberty' at 4–8 weeks of life	Lack of appropriate rise in LH and testosterone	Elevated LH and testosterone levels
Pubertal onset	Absent or pubertal arrest (less common)	Delayed; spontaneous progressive maturation by age 18
Puberty pattern	Absent or arrested gonadarche	Delayed pubarche and gonadarche
Growth	Prepubertal growth; +/– eunuchoid proportions (delayed growth plate fusion)	Normal growth for bone age
Family history of delayed puberty	+/–	+/–
Sense of smell	+/– Anosmia (30–50%)	Normal
Testicular size	Small testes (1–2 ml); +/– history of cryptorchidism	Normal for bone age
Stretched penile length	+/– micropenis	Normal for bone age
Possible associations	Bilateral synkinesia (mirror movements), dental agenesis, renal anomalies, hearing impairment, midline defects (cleft lip/palate)	None

In this section, we will highlight some acquired causes of delayed growth and maturation that are seemingly 'primary' as opposed to those resulting from chronic diseases and their treatments. Adolescent eating disorders include anorexia nervosa, an extreme dissatisfaction with one's body (size or shape or even certain parts). This dissatisfaction may lead to pathologic phobias and food aversion [84]. Bulimia nervosa reflects a fear of fatness and attempts to lose weight by fasting and binge eating but then purging to avoid weight gain. Patient with bulimia nervosa are usually not as underweight as those with anorexia nervosa. Both disorders are more commonly seen in adolescent girls.

Similarly, there is a large group of adolescents, mainly female, with perhaps a 'sub-clinical' form of anorexia nervosa, designated as 'fear of obesity'. These adolescents practice many of the same dietary strategies to remain thin, often with goals of a fraction of their ideal weight [85]. Slowing of growth and maturation is one consequence of this eating behaviour, which may occur in over 50% of 12th grade girls, especially in affluent areas [86].

'Excessive' Energy Expenditure

Energy availability and growth and pubertal maturation may be better understood through a dissection of the mechanism of relative energy deficiency in sport, formerly known as the female athlete triad, as this concept also affects males [87]. The triad originally referred to the relationship between three inter-related components: energy availability (EA), menstrual function, and bone health [88]. There is often a trajectory over time: healthy athlete to the opposite end of the spectrum—low EA, amenorrhea, and osteoporosis [87]. The broadened definition now includes a clinical syndrome that results from relative energy deficiency including alterations in metabolic rate, protein synthesis, menstrual function, bone health, immunity, and cardiovascular and psychological health [87]. The more direct connection to puberty was shown in two elegant reports by Loucks who noted alterations in LH pulsatility within 5 days of decreasing the daily energy balance from 45 kcal/kg lean body mass (LBM)/day to 30, 20, and 15 kcal/kg LBM/day. Although no significant slowing of the GnRH pulse generator occurred at 30 kcal/kg LBM/day, significant slowing—as defined by number of LH pulses—was noted at 20 kcal/kg/LBM/day, with further slowing at 10 kcal/kg LBM/day [89]. That this was relevant to the hypothalamic–pituitary–gonadal axis at puberty was shown by Loucks, noting that susceptibility to the effects of the energy deficit was attenuated as adolescent women 'aged', and was gone by 14 years post-menarche [90]. These data fit well with the known 'physiologic' disorders of menstrual cycles within the first 2 years following menarche. The data may be interpreted as an exquisite sensitivity to nutritional or psychologic stressors.

A similar mechanism may be operative in the male, as states of chronic undernutrition can cause profound suppression of reproductive function, including diminished pulsatile LH secretion. Using the male rhesus monkey as an example, Cameron and colleagues showed that a brief period of fasting suppressed LH secretion, commencing a few hours after a missed meal. This was considered to be due to an energy deficit and not non-specific 'stress', since naloxone, an opioid antagonist, did not reverse the alteration to the central gonadal axis [91].

An example in the human is that short stature of youth artistic gymnasts has been attributed to the effects of intensive training, especially from an early age. The data from a conference devoted to this issue are noted by Malina *et al.* [75], and may be summarized as follows:

1. Near adult or adult height of female and male artistic gymnasts is not compromised by intensive training at a young age or during the pubertal growth spurt
2. Gymnastics training does not attenuate growth of the upper (sitting height) or lower (legs) body segment lengths.
3. Gymnastics training does not appear to attenuate pubertal growth and maturation, including skeletal age, secondary sex characteristics, age at menarche, rate of growth, and timing and tempo of the growth spurt.
4. Present data are insufficient to address the issue of intensive gymnastics training and alterations within multiple endocrine systems
5. Though shorter and lighter than their average peers, gymnasts have an appropriate weight for height, making it less likely that malnutrition is an important factor.
6. Given the individuality of physical growth and biological maturation and the variety of factors known to influence these processes, it is difficult to specify effects attributable to systematic training in artistic gymnasts.
7. Artistic gymnasts show patterns of growth characteristics comparable to short, late-maturing youth with short parents.
8. Other variables that may affect growth and maturation should be considered and include: dietary intake, family size and related characteristics, and quantitative indicators of training time, intensity, and environment (e.g. 'toxic' stress).

Although these data were gathered and debated specifically for artistic gymnastics, their reach goes far beyond this very select group of adolescents; for the effects of some of the factors listed here, especially #8, may have profound effects on far larger groups of adolescents from many backgrounds.

Stress

In day-to-day activities the stress response is meant to be transient. Its accompanying anti-growth, anti-reproductive, catabolic, and immunosuppressive effects are temporarily beneficial and/or of no adverse consequences to the individual [92]. On the other hand, chronic activation of this system may affect growth, maturation, and reproduction due to its effects on the hypothalamic–pituitary–adrenal (HPA) axis. In the realm of growth and maturation, prolonged activation of the HPA axis may lead to:

- Suppression of GH secretion
- Glucocorticoid-induced effects on IGF-I action
- CRH-induced increases in somatostatin, indirectly inhibiting GH secretion [92]
- Inhibition of gonadotropin secretion

These hormonal effects are manifest in the clinical realm as the syndrome of psychosocial short stature, a term that was originally used to describe severely compromised height in children or adolescents due to emotional deprivation and/or physical/psychological abuse [93]. The original description included a syndrome simulating idiopathic hypopituitarism, with apparent GH deficiency or secondary adrenal failure. What was of great surprise was that removal from the inciting environment led very quickly to reversal of the signs and symptoms.

Within a month of convalescent 'therapy' away from the initial environment, a number of the children grew at an accelerated rate and had reversal of their abnormal pituitary stimulation tests [94, 95].

The syndrome is seemingly far less common in the adolescent, but one of us (ADR) followed such a patient through late adolescence and into young adulthood [96]. Growth failure and delayed sexual maturation developed in an adolescent male after severe emotional trauma at age 12 years. He had an eating disorder, was profoundly malnourished for several months, and then resumed eating spontaneously. When evaluated at age 17 years 8 months, he was entirely prepubertal, tested multiple times as GH deficient, and had low gonadotropin levels. The hormonal deficits persisted for many years despite good nutrition and lack of abuse, a course inconsistent with psychosocial short stature or effects of malnutrition *per se.* In his third decade the patient spontaneously grew to a normal adult stature compatible with his mid-parental target height and had full maturation of primary and secondary sexual characteristics. The conclusion was that psychologic trauma induced a deranged hormonal state that persisted for almost a decade [96].

Disorders (Differences) of Sex Development (DSD)

Most individuals with DSDs present at birth or during infancy. However, some of the milder DSDs, especially those that change at adolescence, present with differences in the development or progression of pubertal maturation. Chapter 7.2.2 ('Disorders of Sex Development (DSD) in the Newborn') presents the subject in depth. In summary, the following groups of adolescents may present with a suspected DSD:

1. Girls with primary amenorrhea, with or without breast development
2. Girls who virilize at puberty
3. Boys with pubertal delay (very small percentage of the far more prevalent group with CDGP)

Summary and Conclusions

Puberty is the time for major changes in sexual maturation and psychosocial development. We emphasize the physical changes and their hormonal and regulatory concomitants as the immature, prepubertal boy and girl traverse multiple adolescent stages to become emerging adults at nearly full adult height and sexual maturation. Fully developed body composition usually does not occur until the third decade. Genetic and environmental factors, including the social environment, nutrition, and endocrine disrupting chemicals, are noted as modifiers or modulators of these physiologic processes, with emphasis on the timing, tempo, and longer-term secular trends. We finish with describing some common variants of pubertal maturation, which are more fully described in Chapters 7.3.2 and 7.3.3.

REFERENCES

1. Merriam-Webster; Puberty. Available at: https://www.merriam-webster.com/dictionary/puberty (accessed: 1 April 2020).
2. Malina RM, Bouchard C, Bar-Or O. *Growth, Maturation and Physical Activity*, 2nd edn. Champaign, IL: Human Kinetics Press, 2004.
3. Largo RH, Prader A. Pubertal development in Swiss girls. *Helv Paediatr Acta*, 1983; **38**: 229–43.
4. Largo RH, Prader A. Pubertal development in Swiss boys. *Helv Paediatr Acta*, 1983; **38**: 211–28.
5. Biro FM, Huang B, Daniels SR, Lucky AW. Pubarche as well as thelarche may be a marker for the onset of puberty. *J Pediatr Adolesc Gynecol*, 2008; **21**: 323–8.
6. Siervogel RM, Demerath EW, Schubert C, *et al*. Puberty and body composition. *Horm Res*, 2003; **60**(Suppl 1): 36–45.
7. Iuliano-Burns S, Mirwald RL, Bailey DA. Timing and magnitude of peak height velocity and peak tissue velocities for early, average, and late maturing boys and girls. *Am J Hum Biol*, 2001; 13: 1–8.
8. Beunen GP, Malina RM, Van't Hof MA, *et al*. *Adolescent Growth and Motor Performance: A Longitudinal Study of Belgian Boys*. Champaign, IL: Human Kinetics Press, 1988.
9. Handelsman DJ. Sex differences in athletic performance emerge coinciding with the onset of male puberty. *Clin Endocrinol (Oxf)*, 2017; **87**: 68–72.
10. Bergeron MF, Mountjoy M, Armstrong N, *et al*. International Olympic Committee consensus statement on youth athletic development. *Br J Sports Med*, 2015; **49**: 843–51.
11. De Ste. Croix MBA. Muscle strength. In: Armstrong N, Van Mechelen W, eds. *Paediatric Exercise Science and Medicine*, 2nd edn. Oxford: Oxford University Press, 2008: 199–226.
12. Marshall JC, Kelch RP. Gonadotropin-releasing hormone: role of pulsatile secretion in the regulation of reproduction. *N Engl J Med*, 1986; **315**: 1459–68.
13. Reame N, Sauder SE, Kelch RP, Marshall JC. Pulsatile gonadotropin secretion during the human menstrual cycle: evidence for altered frequency of gonadotropin-releasing hormone secretion. *J Clin Endocrinol Metab*, 1984; **59**: 328–37.
14. Forest MG, Sizonenko PC, Cathiard AM, Bertrand J. Hypophyso-gonadal function in humans during the first year of life. 1. Evidence for testicular activity in early infancy. *J Clin Invest*, 1974; **53**: 819–28.
15. Bidlingmaier F, Strom TM, Dorr HG, Eisenmenger W, Knorr D. Estrone and estradiol concentrations in human ovaries, testes, and adrenals during the first two years of life. *J Clin Endocrinol Metab*, 1987; **65**: 862–7.
16. Kuiri-Hanninen T, Haanpaa M, Turpeinen U, Hamalainen E, Dunkel L, Sankilampi U. Transient postnatal secretion of androgen hormones is associated with acne and sebaceous gland hypertrophy in early infancy. *J Clin Endocrinol Metab*, 2013; 98: 199–206.
17. Kuiri-Hanninen T, Haanpaa M, Turpeinen U, *et al*. Postnatal ovarian activation has effects in estrogen target tissues in infant girls. *J Clin Endocrinol Metab*, 2013; **98**: 4709–16.
18. Apter D, Cacciatore B, Alfthan H, Stenman UH. Serum luteinizing hormone concentrations increase 100-fold in females from 7 years to adulthood, as measured by time-resolved immunofluorometric assay. *J Clin Endocrinol Metab*, 1989; **68**: 53–7.
19. Demir A, Voutilainen R, Juul A, *et al*. Increase in first morning voided urinary luteinizing hormone levels precedes the physical onset of puberty. *J Clin Endocrinol Metab*, 1996; 81: 2963–7.
20. Lee PA, Gollenberg AL, Hediger ML, Himes JH, Zhang Z, Louis GM. Luteinizing hormone, testosterone and inhibin B levels in the peripubertal period and racial/ethnic differences among boys aged 6–11 years: analyses from NHANES III, 1988–1994. *Clin Endocrinol (Oxf)*, 2010; **73**: 744–51.

21. Roberts VJ, Barth S, el-Roeiy A, Yen SS. Expression of inhibin/activin subunits and follistatin messenger ribonucleic acids and proteins in ovarian follicles and the corpus luteum during the human menstrual cycle. *J Clin Endocrinol Metab*, 1993; **77**: 1402–10.
22. Lee MM, Donahoe PK, Hasegawa T, *et al.* Mullerian inhibiting substance in humans: normal levels from infancy to adulthood. *J Clin Endocrinol Metab*, 1996; **81**: 571–6.
23. Crofton PM, Illingworth PJ, Groome NP, *et al.* Changes in dimeric inhibin A and B during normal early puberty in boys and girls. *Clin Endocrinol (Oxf)*, 1997; **46**: 109–14.
24. Addo OY, Miller BS, Lee PA, Hediger ML, Himes JH. Age at hormonal onset of puberty based on luteinizing hormone, inhibin B, and body composition in preadolescent U.S. girls. *Pediatr Res*, 2014; **76**: 564–70.
25. Sims EK, Addo OY, Gollenberg AL, Himes JH, Hediger ML, Lee PA. Inhibin B and luteinizing hormone levels in girls aged 6–11 years from NHANES III, 1988–1994. *Clin Endocrinol (Oxf)*, 2012; **77**: 555–63.
26. Martha PM, Jr., Gorman KM, Blizzard RM, Rogol AD, Veldhuis JD. Endogenous growth hormone secretion and clearance rates in normal boys, as determined by deconvolution analysis: relationship to age, pubertal status, and body mass. *J Clin Endocrinol Metab*, 1992; **74**: 336–44.
27. Roemmich JN, Clark PA, Mai V, *et al.* Alterations in growth and body composition during puberty: III. Influence of maturation, gender, body composition, fat distribution, aerobic fitness, and energy expenditure on nocturnal growth hormone release. *J Clin Endocrinol Metab*, 1998; **83**: 1440–7.
28. Barreca A, Artini PG, Del Monte P, *et al.* In vivo and in vitro effect of growth hormone on estradiol secretion by human granulosa cells. *J Clin Endocrinol Metab*, 1993; **77**: 61–7.
29. Klein KO, Martha PM, Jr., Blizzard RM, Herbst T, Rogol AD. A longitudinal assessment of hormonal and physical alterations during normal puberty in boys. II. Estrogen levels as determined by an ultrasensitive bioassay. *J Clin Endocrinol Metab*, 1996; **81**: 3203–7.
30. Palmert MR, Boepple PA. Variation in the timing of puberty: clinical spectrum and genetic investigation. *J Clin Endocrinol Metab*, 2001; 86: 2364–8.
31. Wohlfahrt-Veje C, Mouritsen A, Hagen CP, *et al.* Pubertal onset in boys and girls is influenced by pubertal timing of both parents. *J Clin Endocrinol Metab*, 2016; **101**: 2667–74.
32. Ojeda SR, Lomniczi A, Mastronardi C, *et al.* Minireview: the neuroendocrine regulation of puberty: is the time ripe for a systems biology approach? *Endocrinology*, 2006; **147**: 1166–74.
33. Lomniczi A, Wright H, Ojeda SR. Epigenetic regulation of female puberty. *Front Neuroendocrinol*, 2015; **36**: 90–107.
34. Seminara SB, Messager S, Chatzidaki EE, *et al.* The GPR54 gene as a regulator of puberty. *N Engl J Med*, 2003; **349**: 1614–27.
35. Avendano MS, Vazquez MJ, Tena-Sempere M. Disentangling puberty: novel neuroendocrine pathways and mechanisms for the control of mammalian puberty. *Hum Reprod Update*, 2017; **23**: 737–63.
36. Dhillo WS, Chaudhri OB, Patterson M, *et al.* Kisspeptin-54 stimulates the hypothalamic-pituitary gonadal axis in human males. *J Clin Endocrinol Metab*, 2005; **90**: 6609–15.
37. Dhillo WS, Chaudhri OB, Thompson EL, *et al.* Kisspeptin-54 stimulates gonadotropin release most potently during the preovulatory phase of the menstrual cycle in women. *J Clin Endocrinol Metab*, 2007; **92**: 3958–66.
38. Teles MG, Bianco SD, Brito VN, *et al.* A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*, 2008; **358**: 709–15.
39. Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol*, 2016; **4**: 254–64.
40. Busch AS, Hagen CP, Almstrup K, Juul A. Circulating MKRN3 levels decline during puberty in healthy boys. *J Clin Endocrinol Metab*, 2016; **101**: 2588–93.
41. Abreu AP, Macedo DB, Brito VN, Kaiser UB, Latronico AC. A new pathway in the control of the initiation of puberty: the MKRN3 gene. *J Mol Endocrinol*, 2015; **54**: R131–9.
42. Terasawa E, Fernandez DL. Neurobiological mechanisms of the onset of puberty in primates. *Endocr Rev*, 2001; **22**: 111–51.
43. Gamba M, Pralong FP. Control of GnRH neuronal activity by metabolic factors: the role of leptin and insulin. *Mol Cell Endocrinol*, 2006; 254–5: 133–9.
44. Sisk CL, Foster DL. The neural basis of puberty and adolescence. *Nat Neurosci*, 2004; **7**: 1040–7.
45. Elks CE, Perry JR, Sulem P, *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet*, 2010; **42**: 1077–85.
46. Li W, Liu Q, Deng X, Chen Y, Liu S, Story M. Association between obesity and puberty timing: a systematic review and meta-analysis. *Int J Environ Res Public Health*, 2017; **14**: pii: E1266.
47. Farooqi IS, Matarese G, Lord GM, *et al.* Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*, 2002; **110**: 1093–103.
48. Smith JT, Acohido BV, Clifton DK, Steiner RA. KiSS-1 neurones are direct targets for leptin in the ob/ob mouse. *J Neuroendocrinol*, 2006; **18**: 298–303.
49. Almstrup K, Johansen ML, Busch AS, *et al.* Erratum: pubertal development in healthy children is mirrored by DNA methylation patterns in peripheral blood. *Sci Rep*, 2016; **6**: 30664.
50. Hayes P. International adoption, 'early' puberty, and underrecorded age. *Pediatrics*, 2013; **131**: 1029–31.
51. Krstevska-Konstantinova M, Charlier C, Craen M, *et al.* Sexual precocity after immigration from developing countries to Belgium: evidence of previous exposure to organochlorine pesticides. *Hum Reprod*, 2001; **16**: 1020–6.
52. Soriano-Guillen L, Corripio R, Labarta JI, *et al.* Central precocious puberty in children living in Spain: incidence, prevalence, and influence of adoption and immigration. *J Clin Endocrinol Metab*, 2010; **95**: 4305–13.
53. Mouritsen A, Aksglaede L, Sorensen K, *et al.* Hypothesis: exposure to endocrine-disrupting chemicals may interfere with timing of puberty. *Int J Androl*, 2010; **33**: 346–59.
54. Den Hond E, Roels HA, Hoppenbrouwers K, *et al.* Sexual maturation in relation to polychlorinated aromatic hydrocarbons: Sharpe and Skakkebaek's hypothesis revisited. *Environ Health Perspect*, 2002; **110**: 771–6.
55. Leijs MM, Koppe JG, Olie K, *et al.* Delayed initiation of breast development in girls with higher prenatal dioxin exposure; a longitudinal cohort study. *Chemosphere*, 2008; **73**: 999–1004.
56. Gladen BC, Ragan NB, Rogan WJ. Pubertal growth and development and prenatal and lactational exposure to polychlorinated biphenyls and dichlorodiphenyl dichloroethene. *J Pediatr*, 2000; **136**: 490–6.
57. Wolff MS, Britton JA, Boguski L, *et al.* Environmental exposures and puberty in inner-city girls. *Environ Res*, 2008; **107**: 393–400.

58. Calafat AM, Ye X, Wong LY, Reidy JA, Needham LL. Exposure of the U.S. population to bisphenol A and 4-tertiary-octylphenol: 2003–2004. *Environ Health Perspect*, 2008; **116**: 39–44.
59. Kundakovic M, Gudsnuk K, Franks B, *et al*. Sex-specific epigenetic disruption and behavioral changes following low-dose in utero bisphenol A exposure. *Proc Natl Acad Sci U S A*, 2013; **110**: 9956–61.
60. Papadimitriou A. The evolution of the age at menarche from prehistorical to modern times. *J Pediatr Adolesc Gynecol*, 2016; **29**: 527–30.
61. Brundtland GH, Walloe L. Menarchal age in Norway: halt in the trend towards earlier maturation. *Nature*, 1973; **241**: 478–9.
62. Malina RM. Secular changes in size and maturity: causes and effects. *Monographs of the Society for Research in Child Development*, 1979; **44**: 59–102.
63. Ong KK, Ahmed ML, Dunger DB. Lessons from large population studies on timing and tempo of puberty (secular trends and relation to body size): the European trend. *Mol Cell Endocrinol*, 2006; **254–5**: 8–12.
64. Walvoord EC. The timing of puberty: is it changing? Does it matter? *J Adolesc Health*, 2010; **47**: 433–9.
65. Gluckman PD, Hanson MA. Living with the past: evolution, development, and patterns of disease. *Science*, 2004; **305**: 1733–6.
66. Gluckman PD, Hanson MA. Evolution, development and timing of puberty. *Trends Endocrinol Metab*, 2006; **17**: 7–12.
67. Lee Y, Styne D. Influences on the onset and tempo of puberty in human beings and implications for adolescent psychological development. *Horm Behav*, 2013; **64**: 250–61.
68. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*, 2009; **123**: 84–8.
69. Malina RM, Little BB, Pena Reyes ME. Secular trends are associated with the demographic and epidemiologic transitions in an indigenous community in Oaxaca, Southern Mexico. *Am J Phys Anthropol*, 2018; **165**: 47–64.
70. Chumlea WC, Schubert CM, Roche AF, *et al*. Age at menarche and racial comparisons in US girls. *Pediatrics*, 2003; **111**: 110–13.
71. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*, 1969; **44**: 291–303.
72. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*, 1970; **45s**: 13–23.
73. Flor-Cisneros A, Leschek EW, Merke DP, *et al*. In boys with abnormal developmental tempo, maturation of the skeleton and the hypothalamic-pituitary-gonadal axis remains synchronous. *J Clin Endocrinol Metab*, 2004; **89**: 236–41.
74. Pantsiotou S, Papadimitriou A, Douros K, Priftis K, Nicolaidou P, Fretzayas A. Maturational tempo differences in relation to the timing of the onset of puberty in girls. *Acta Paediatr*, 2008; **97**: 217–20.
75. Malina RM, Baxter-Jones AD, Armstrong N, *et al*. Role of intensive training in the growth and maturation of artistic gymnasts. *Sports Med*, 2013; **43**: 783–802.
76. Malina RM, Bielicki T. Retrospective longitudinal growth study of boys and girls active in sport. *Acta Paediatr*, 1996; **85**: 570–6.
77. Euling SY, Herman-Giddens ME, Lee PA, *et al*. Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*, 2008; **121** Suppl 3: S172–91.
78. Sanders JO, Qiu X, Lu X, *et al*. The uniform pattern of growth and skeletal maturation during the human adolescent growth spurt. *Sci Rep*, 2017; **7**: 16705.
79. Greulich WW, Pyle SI. *Radiographic Atlas of Skeletal Development of the Hand and Wrist*, 2nd edn. Stanford, CA: Stanford University Press, 1959.
80. Styne DM, Grumbach MM. Puberty: ontogeny, neuroendocrinology, physiology, and disorders. In: Melmed S, Polonsky KS, Larsen PR, Kronenberg HM, eds. *Williams Textbook of Endocrinology*, 12th edn. Philadelphia, PA: Elsevier Saunders; 2011: 1054–201.
81. Zhu J, Chan YM. Adult consequences of self-limited delayed puberty. *Pediatrics*, 2017; **139**: pii: e20163177.
82. Fuqua JS, Rogol AD. Neuroendocrine alterations in the exercising human: implications for energy homeostasis. *Metabolism*, 2013; **62**: 911–21.
83. Basaria S. Male hypogonadism. *Lancet*, 2014; **383**: 1250–63.
84. Herpertz-Dahlmann B. Adolescent eating disorders: update on definitions, symptomatology, epidemiology, and comorbidity. *Child Adolesc Psychiatr Clin N Am*, 2015; **24**: 177–96.
85. Lifshitz F, Moses N. Nutritional dwarfing: growth, dieting, and fear of obesity. *J Am Coll Nutr*, 1988; **7**: 367–76.
86. Neumark-Sztainer D, Hannan PJ. Weight-related behaviors among adolescent girls and boys: results from a national survey. *Arch Pediatr Adolesc Med*, 2000; **154**: 569–77.
87. Mountjoy M, Sundgot-Borgen J, Burke L, *et al*. The IOC consensus statement: beyond the Female Athlete Triad—Relative Energy Deficiency in Sport (RED-S). *Br J Sports Med*, 2014; **48**: 491–7.
88. Nattiv A, Loucks AB, Manore MM, *et al*. American College of Sports Medicine position stand. The female athlete triad. *Med Sci Sports Exerc*, 2007; **39**: 1867–82.
89. Loucks AB, Thuma JR. Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. *J Clin Endocrinol Metab*, 2003; **88**: 297–311.
90. Loucks AB. The response of luteinizing hormone pulsatility to 5 days of low energy availability disappears by 14 years of gynecological age. *J Clin Endocrinol Metab*, 2006; **91**: 3158–64.
91. Cameron JL, Helmreich DL, Schreihofer DA. Modulation of reproductive hormone secretion by nutritional intake: stress signals versus metabolic signals. *Hum Reprod*, 1993; **8** Suppl 2: 162–7.
92. Charmandari E, Kino T, Souvatzoglou E, Chrousos GP. Pediatric stress: hormonal mediators and human development. *Horm Res*, 2003; **59**: 161–79.
93. Powell GF, Brasel JA, Blizzard RM. Emotional deprivation and growth retardation simulating idiopathic hypopituitarism. I. Clinical evaluation of the syndrome. *N Engl J Med*, 1967; **276**: 1271–8.
94. Powell GF, Brasel JA, Raiti S, Blizzard RM. Emotional deprivation and growth retardation simulating idiopathic hypopituitarism. II. Endocrinologic evaluation of the syndrome. *N Engl J Med*, 1967; **276**: 1279–83.
95. Skuse D, Albanese A, Stanhope R, Gilmour J, Voss L. A new stress-related syndrome of growth failure and hyperphagia in children, associated with reversibility of growth-hormone insufficiency. *Lancet*, 1996; **348**: 353–8.
96. Magner JA, Rogol AD, Gorden P. Reversible growth hormone deficiency and delayed puberty triggered by a stressful experience in a young adult. *Am J Med*, 1984; **76**: 737–42.

7.3.2 Pubertal Delay and Hypogonadism

Alan D. Rogol and John S. Fuqua

Introduction

Puberty may be defined as the physiologic process resulting in the attainment of sexual maturity and reproductive capacity. An assessment of puberty is an integral component of the evaluation and treatment of endocrine disorders in children and adolescents. Not only does it impact sexual maturation, but it has other effects with lifelong consequences, including linear growth, changes in body composition, and skeletal mineralization. Patients with disorders of puberty, including precocious and delayed puberty, make up a large percentage of the children and adolescents who consult paediatric endocrinologists. An understanding of delayed or absent puberty requires a foundation in the normal processes regulating the onset of puberty and factors essential for its progression and completion. In this chapter, we will first review normal growth and puberty, particularly with regard to their interdependence. We shall then discuss the differential diagnosis of delayed or absent puberty and present diagnostic algorithms for hypergonadotropic and hypogonadotropic hypogonadism, emphasizing some gender-specific aspects.

Physical Changes of Puberty

Concurrent with the secretion of sex steroids during puberty, major physical changes, physiologic adaptations, and social and emotional challenges occur. The measurement and assessment of these changes are critical for determining when pubertal maturation is progressing normally or not and to monitor the efficacy of treatment.

Puberty in Boys

In boys, the earliest physical change associated with puberty is testicular enlargement, although a subset of boys has pubic hair growth due to adrenal androgens prior to testicular enlargement. Testicular size is commonly assessed by using a series of calibrated, testis-shaped ellipsoids (beads) called the Prader orchidometer. If this is not available, the long axis of the testis can be measured using simple callipers or an ordinary tape measure. Prepubertal testes are <4 ml in volume and less than or equal to 2.5 cm in length. As puberty ensues, the testes gradually enlarge, mainly due to increases in volume of the seminiferous tubule content, and eventually reach the adult volume of 15–25 ml or length of 4–5 cm. Physical changes accompanying testicular enlargement include thinning of the scrotal skin, apocrine sweating and odour, and the growth of sexual hair. Additional changes present in boys include an increase in muscular size and strength and body hair growth in a typical adult male pattern. Deepening of the voice occurs during the second half of pubertal maturation.

Genital development in boys is often assessed using the method of Tanner. Two rating scales are used in males: one for pubic hair growth and another for enlargement of the testes, penis, and scrotum. Tanner stages for boys are reviewed in Table 7.3.1.1 in Chapter 7.3.1. Briefly, pubic hair growth starts as fine, straight, lightly pigmented hairs generally located on the pubic symphysis at the base of the penis. As puberty progresses, the hair becomes coarser and curly, with darker pigmentation. At Tanner stage 5, the growth extends down the medial thighs and up the lower abdomen. Genital Tanner stages are somewhat more subjective. The earliest stage of puberty consists of testicular enlargement only, followed by gradual enlargement of the penis, first in length and then in circumference, and enlargement of the testes to reach full adult maturity [1].

Puberty in Girls

In girls, the first clinically apparent sign of puberty is breast development, although pubic hair growth precedes breast development by up to 6 months in approximately 30%. It is common for one breast to grow for several months before the other, and mild asymmetry is often present. Pubic hair growth usually begins within 6 months of the onset of breast development. As oestrogen levels increase, changes occur in the vaginal mucosa. In the prepubertal girl, the vaginal epithelium is thin with a dark red colour and consists mainly of basal and parabasal cells. With advancing pubertal maturation, the epithelium proliferates and thickens. Intermediate and superficial cornified squamous cells overlay the parabasal cells and give the mucosa a pink, opalescent appearance. This is often accompanied by a physiologic vaginal discharge.

The progression of puberty in girls is also assessed with Tanner staging. Pubic hair in girls is assessed using the same hallmarks as in boys, although normal girls with Tanner stage 5 pubic hair do not have extension up the lower abdomen (Table 7.3.1.1 in Chapter 7.3.1). Some physicians add a Tanner stage 6 in girls to describe those with pubic hair extension both to the medial thighs and in the midline of the lower abdomen. Breast development is commonly measured using Tanner staging as well, and this is also depicted in Table 7.3.1.1 in Chapter 7.3.1. Breast development begins at Tanner stage 2 with budding, in which there is a firm palpable disk of tissue not larger than the areola. In stage 3, the diameter of the tissue exceeds the areola but does not have stage 4 morphology. Stage 3 encompasses a large range of development, from very early in puberty up to the later stages of maturation. With stage 4, there is a 'double contour' in which the profile of the areola is distinct from the profile of the breast. Although stage 5 is considered to be full adult development, some normal adult women do not progress beyond stage 4, and some never develop a double contour, skipping stage 4 altogether [2].

Growth and Pubertal Maturation

The clinical hallmark of puberty as it relates to body size is the pubertal growth spurt. In boys, the peak of the growth spurt is timed to Tanner stage III–IV, whereas in girls the peak occurs earlier in puberty, typically at Tanner stages II–III. The average peak growth velocity in boys is 9.5 cm/year, and in girls it is somewhat less, 8.3 cm/year. The later onset of puberty, the later occurrence of the growth spurt within puberty in boys, and the greater magnitude of the growth spurt result in an average height difference between the sexes of 13 cm.

Puberty also has effects on skeletal maturation, and the timing of puberty is more closely correlated with the bone age than with chronologic age. In European girls, the average bone age at the time of thelarche is 10.5 years (95% confidence limits 8.5–13.2), and at the time of menarche it is 12.8 years (11.3–13.6). In girls, the peak height velocity occurs at a bone age of 12 years. In boys, the average bone age at stage 2 of puberty is 11.5 years (9.0–14.2). At peak height velocity in boys, the bone age ranged between 11.9 and 15.5 years in one study and between 12.5 and 16 years in another study [3, 4].

Children with advanced bone ages typically enter puberty earlier than their peers, while those with delayed bone ages enter puberty later. Factors that alter the bone age also alter the timing of puberty, and these may include sex steroids and the GH/IGF-1 system [4].

Age at Onset of Puberty

There is a great deal of disagreement about the age at which pubertal maturation is normal. Most of this disagreement relates to the lower age limit of normal. There is evidence that the age of onset of puberty has decreased in the last several decades in both girls and boys. For boys, data show that the average age at Tanner stage 2 maturation is between 11.2 and 12.4 years. The normal range of attainment of stage 2 puberty in boys is commonly considered to extend from 9 years up to 13.5 years. For girls, there is more disagreement. Historically, the normal range of onset of puberty has been between 8 and 13 years, but the lower limit of normal may extend down to 7 years for white girls and 6 years for black girls. An analysis of data from the Third National Health and Nutrition Examination Survey (NHANES) in the United States revealed that body mass index (BMI) is an independent predictor of the age of onset of puberty, with earlier occurrence of breast and pubic hair development and earlier menarche in girls with BMI above the 85th percentile [5]. The upper limit of 13 years for girls remains generally accepted. The **tempo** of pubertal maturation is also important to consider. Girls starting puberty earlier than average tend to have a slower progression to menarche than girls starting puberty later. Hence, the age at menarche has less variability than the age at thelarche and is about 12.6 years. This phenomenon also probably occurs in males, but the lack of a clearly definable event such as menarche in male puberty makes study difficult [6–9].

Hypothalamic–Pituitary–Gonadal (HPG) Axis Maturation

A detailed discussion of HPG axis maturation is beyond the scope of this chapter but is discussed in more detail in Chapter 7.3.1. 'Recognizing Normal and Disordered Pubertal Development'. Briefly, the hypothalamic-pituitary-gonadal axis is active *in utero*, with peak secretion of gonadotropin releasing hormone (GnRH), luteinizing hormone (LH), and follicle stimulating hormone (FSH) occurring between 20 and 24 weeks' gestation. During later pregnancy, levels drop as the negative feedback effects of gonadal hormones intensify. There is a brief 'minipuberty of infancy' that occurs early after birth, but by 5–6 months of age GnRH secretion, LH and FSH levels, and gonadal steroid levels fall to their prepubertal levels. This cessation of activity is known as the juvenile pause. Over the course of the next several years, the juvenile pause persists, reaching the greatest degree of axis suppression in females at about age 6 years.

When measured using immunoassays, which lack sensitivity at low concentrations, the morning increases in testosterone and oestradiol may be difficult to detect during the very earliest stages of puberty. However, modern techniques, such as liquid chromatography/tandem mass spectrometry (MS) can usually detect even the lower prepubertal levels. In early puberty, gonadotropin levels also may be difficult to distinguish from normal prepubertal levels, in part due to the pulsatile nature of their secretion and in part due to the low amplitude of secretion during the daytime. These factors make the routine laboratory evaluation of delayed puberty difficult, because randomly obtained gonadotropin and sex steroid levels do not differentiate a patient who is nearing a normal but delayed puberty from one who will never enter puberty due to a pathological condition. However, as GnRH pulsatility increases in early puberty, pituitary stores of LH also increase. These stores may be released following acute stimulation by GnRH. This is the basis for the GnRH stimulation test, which may be positive even before physical changes of puberty become clinically apparent.

Over the course of puberty in girls, LH levels increase 25–40-fold relative to levels present before puberty. Oestradiol concentrations increase from <5 pg/ml (<18.4 pmol/L) before puberty up to >100 pg/ml (>367 pmol/L) in a post-menarchal girl. Testosterone concentrations in prepubertal males generally are below 10 ng/dL (0.35 nmol/L), and in young adults they are above 300 ng/dL (10.4 nmol/L).

Control of Pubertal Timing

The mechanisms underlying regulation of pubertal timing have been partially elucidated in recent years, but much remains poorly understood. There is a strong influence of genetics on the timing of puberty, and it is estimated that 50–80% of the variance can be attributed to genetic factors [10, 11]. Genome-wide association studies have demonstrated associations between a number of loci and age at menarche. A recent study of 11 000 samples with associated clinical data identified loci relevant to timing of both female and male puberty [12]. Interestingly, loci associated with increased body mass index (BMI) were also associated with early puberty in females, while those loci were variably associated with either early or delayed puberty in boys, reflecting the conflicting clinical data of the impact of this parameter on male puberty [13]. Additional environmental inputs on pubertal timing include endocrine disrupting chemicals. Although exposure to these substances is often associated with early pubertal maturation, they may also lead to delayed maturation in both males and females. Investigation

of these effects is confounded by the multiplicity of exposures. Additionally, differing levels of exposure may lead to opposite effects on pubertal timing. Further, there appears to be a window of vulnerability, with effects being particularly apparent after exposure in early life [14]. The effects of these and other environmental factors may be through direct genomic mechanisms, but also likely involve non-genomic (epigenetic) mechanisms, such as effects on DNA and histone methylation, histone acetylation, and microRNA expression. This topic is discussed in more detail in Chapter 7.3.1., 'Recognizing Normal and Disordered Pubertal Development'.

Delayed Puberty and Hypogonadism

The causes of delayed puberty and hypogonadism can be divided into those involving delays or defects in hypothalamic regulation of the initiation of puberty and those involving primary defects of the gonads. These groups of conditions are best differentiated by the serum concentrations of gonadotropins after the age when puberty is expected. Hypothalamic and pituitary deficiencies are termed hypogonadotropic hypogonadism or central hypogonadism, and primary gonadal disorders are termed hypergonadotropic hypogonadism or primary hypogonadism.

Constitutional Delay of Growth and Puberty (Transient Hypogonadotropic Hypogonadism)

Constitutional delay of growth and puberty (CDGP) is a common variant of physiologic (normal) maturation. Its major outward characteristics include a slowing of linear growth during the late preteen and early teen years as well as a delay in the timing (and perhaps tempo) of puberty. In the typical scenario, a boy (or his parents) seeks endocrinological evaluation in the early teenage years because of significant concern for the discrepancy in growth and pubertal maturation between the patient and his/her age peers (Figure 7.3.2.1). The condition must be distinguished from hypogonadotropic hypogonadism [15, 16] (Table 7.3.2.1).

Clinically, the height age (the age for which the patient's height is the 50th centile) is delayed with respect to the calendar age but is concordant with the 'biological age' as indexed by the bone age. Sexual maturation is either prepubertal or lagging behind that of their peers, although it is often appropriate for the bone age. The height velocity is normal for a **prepubertal** child, although it may decline to subnormal values if the delay is more than 2–3 years (prepubertal 'dip'). When the height is plotted on the standard growth curve, the height gain appears to be 'falling off' the previously defined height centile, since the standard growth curve incorporates the pubertal growth spurt at an 'average' age. This discrepancy in growth between the normal adolescent age peers and the one with CDGP only accentuates the difference between them. The height discrepancy and the delay in pubertal maturation are often compelling concerns of the patient or the family, bringing the adolescent to medical consultation. Tanner has devised longitudinal growth curves that account for the later growth and adolescent maturation, because this pattern is so common. The large majority of those with CDGP follow this 'custom' curve, emphasizing that the standard growth curves based upon cross-sectional data, although adequate for the population 'averages', do not necessarily describe the growth of an **individual**.

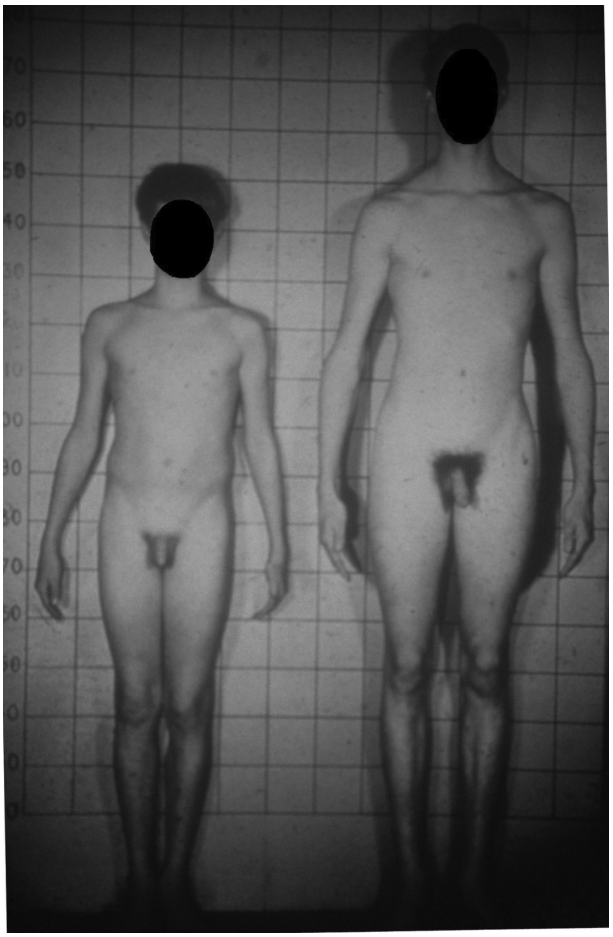


Figure 7.3.2.1 Healthy fraternal twin boys, age 15 years. The boy on the right has had normally timed pubertal development. His brother on the left has constitutional delay of growth and puberty.

Biochemically, adolescents with CDGP resemble their peers with comparable biological (bone) ages. Thus, the pubertal increases in haemoglobin, haematocrit, creatinine, and alkaline phosphatase will not be present. Serum levels of growth hormone (pulsatile pattern), IGF-I, IGFBP-3, LH, FSH, and the sex steroids may be diminished for chronological age, but normal when compared to younger adolescents of the same stage of pubertal maturation. The suppressed HPG axis found in adolescents with CDGP represents an extension of the physiologic hypogonadotropic hypogonadism (the 'juvenile pause') noted since infancy.

Without intervention with sex steroids, adolescents with CDGP will undergo spontaneous pubertal maturation and will reach their target height range as calculated from parental stature. Maturation occurs as much as several years after that of their peers. Failure to enter puberty spontaneously indicates long-term hypogonadism. Many adolescents with CDGP find the delay intolerable and suffer significant emotional distress because they differ in their appearance from their peers during these years. That is often the rationale for the **short-term** use of gonadal steroid therapy. Linear growth and a more mature 'appearance' are more objective outcomes of gonadal steroid administration.

Although some studies have shown little effect of CDGP on psychosocial well-being [17], others demonstrate an association with

Table 7.3.2.1 Comparison of CDGP and IHH

Clinical/lab test	CDGP	IHH
History of undescended testicles	–	±
Anosmia/hyposmia	–	+ (30–50% of patients)
Family history of delayed puberty	+	±
Presence of endogenous progressive pubertal maturation by age 18 years	+	–
Puberty pattern	Delayed adrenarche	More likely to have normal adrenarche
Height for age	Usually short	Usually normal
General physical exam	Normal	Usually normal but may have: Synkinesia Cleft lip and/or palate Sensorineural hearing loss Dental agenesis Skeletal malformations
Genital exam	Delayed pubertal maturation. Testicular and penile length appropriate for bone age	Small testes (1–2 ml) ± small phallus
Laboratory evaluation	Unreliable to distinguish between CDGP and IHH	

Table data obtained from Wehkalampi, *et al.* [16] and from Ambler [15].

poor psychological outcomes. Some of these negative outcomes may persist into adolescence and emerging adulthood. The Oakland Growth Study, for example, found that late maturing boys were not only perceived as less mature and less attractive than peers, but also had more feelings of inadequacy, which persisted into emerging adulthood [18]. Delayed pubertal timing has also been associated with higher rates of depression, anxiety, and substance use in adolescence and young adulthood. Academic achievement may be adversely affected as well [17].

Hypogonadotropic Hypogonadism

There are many causes of combined pituitary hormone deficiency, both congenital and acquired, that may include deficiency of GnRH or gonadotropins. These conditions are reviewed in subsections 2.3 and 2.4.

The neuroendocrine control of mammalian reproduction is governed by a single gene coding for gonadotropin-releasing hormone (GnRH). A neural network of approximately 1500 to 2000 neurons integrates various upstream genes that are responsive to environmental cues such as food (energy) availability, stress, and perhaps light-dark cycles (at least in seasonally breeding mammals).

A cascade of signalling molecules and transcription factors plays a crucial role in pituitary development, cell proliferation, patterning, and terminal differentiation [19]. Genes are expressed in an orderly sequence to activate or inhibit downstream processes (target genes) that have specific roles in the terminal differentiation of pluripotential precursor cells. Genes identified in association with isolated hypogonadotropic hypogonadism are listed in Table 7.3.2.2. It should be noted however, that there is increasing evidence that disorders of puberty may be caused by mutations in multiple genes in the same individual (oligogenicity), with some disorders presenting from the accumulative burden of mutations in genes such as FGFR1 and CHD7. In a recent large study of 116 subjects with Kallmann syndrome (KS) and normosmic isolated hypogonadotropic hypogonadism (nIHH), abnormalities in genes known to cause isolated hypogonadotropic hypogonadism (IHH) were identified in 51%. Fifteen per cent (15%) of subjects had abnormalities in two or more genes [20]. This oligogenicity, with synergism between gene defects, likely broadens the phenotypic spectrum and the endocrine profiles of the subjects.

Kallmann syndrome is the combination of hypogonadotropic hypogonadism (HH) and a diminished sense of smell—hyposmia or anosmia. It is mainly due to a failure of the GnRH-secreting neurons, which have an extra-CNS origin in the nasal placode, to leave their origin in the olfactory epithelium and migrate into the CNS. This is accomplished via the olfactory bulb and tract, finally ending at the arcuate nucleus of the hypothalamus. These neurons form a network among themselves and project dendrites toward the median eminence. Through an inadequately understood mechanism, the neural network generates pulses of GnRH, which travel to the median eminence, are secreted into the hypothalamic-pituitary portal system, and then cause the pituitary gonadotrophs to produce LH and FSH pulses in the general circulation. These activate the gonads to produce testosterone or oestrogen/progesterone. Genes involved in the migration of GnRH neurons may lead to KS, although not all patients with these genetic abnormalities will also have anosmia or hyposmia. Those with normosmic isolated hypogonadotropic hypogonadism (nIHH) have an intact sense of smell. They present with a lack of sexual maturation at the appropriate age associated with inappropriately low gonadotropin levels in the presence of prepubertal concentrations of sex steroid hormones, otherwise normal anterior pituitary function, findings on brain imaging, and appropriate response to exogenous pulsatile GnRH administration. Patients with nIHH may not have mutations in genes regulating GnRH neuron migration but may instead have defects in genes important in other aspects of gonadotropin secretion. Approximately 50% of patients presenting with IHH have KS, with the other half having nIHH [20]. Abnormalities of at least 30 genes have been shown to cause both Kallmann syndrome and/or nIHH, and gene mutations have been identified in 31–51% of affected patients [20–22]. Individuals with classic KS and nIHH may be found in the same kindred [23]. Many of the identified genes known to cause KS and nIHH encode receptor-ligand pairs or are involved in networks regulating developmental pathways.

Genetic Abnormalities Associated with Kallmann Syndrome and nIHH

Fibroblast growth factor-8 (FGF8), its receptor FGFR1, and related proteins have been identified as critical to the normal development and migration of GnRH precursor neurons from the fetal olfactory placode to the arcuate nucleus of the hypothalamus [21]. Mutations of some genes in this system (as listed in Table 7.3.2.2) have also been associated with defects in the ears and kidneys as well as additional areas of the CNS. Together, defects in genes involved in the FGF8 system account for approximately 23% of cases of congenital hypogonadotropic hypogonadism [24].

Table 7.3.2.2 Single gene defects leading to isolated hypogonadotropic hypogonadism

Genes		OMIM numbers
Genes important for GnRH neuron migration associated with anosmic/hyposmic IHH (Kallmann syndrome)		
FGF8 system genes	FGF8	600483
	FGFR1	136350
	ANOS1	300836
	HS6ST1	604846
	FGF17	603725
	IL17RD	606807
	DUSP6	602748
	SPRY4	607984
	KLB	611135
	FLRT3	604808
Other genes associated with Kallmann syndrome	PROK2	607002
	PROKR2	607123
	CHD7	608892
	SOX10	602229
	WDR11	606417
	SEMA3A	603961
	SEMA3E	608166
	CCDC141	616031
	FEZF1	613301
	IGSF10	617351
	AXL	109135
	NELF	608137
	SMCHD1	614982
Genes associated only with normosmic IHH		
	KISS1	603286
	KISS1R	604161
	GNRH1	152760
	GNRHR	138850
	TAC3	162330
	TACR3	162332
	LEP	164160
	LEPR	601007
	PCSK1	162150
	NR0B1	300473
	LHB	152780
	FSHB	136530

Mutations in the FGF8 gene in association with KS were first described in 2008 [24]. FGF8 is transcribed as four isoforms and is expressed in a variety of tissues, including the brain, limbs, heart, eye, and ear. Patients with loss of function mutations have hypogonadism, which may be anosmic or normosmic. Associated abnormalities include hearing loss, hyperextensibility of the digits, and cleft lip/palate [25]. Interestingly, one patient with an FGF8 mutation developed adult-onset hypogonadism, presenting with infertility and decreased libido. He responded normally to pulsatile GnRH therapy [24].

Fibroblast growth factor receptor 1 (FGFR1, aka KAL2) is one of four tyrosine kinase receptors for the much larger family of FGF ligands (at least 23 members) and binds FGF8. Although only approximately 10% of all persons with KS have mutations in this receptor, almost as many with nIHH have the same set of mutations. The inheritance is autosomal dominant. Mutations in FGFR1 may occur in association with mutations in CHD7 with increased frequency (oligogenicity) [20]. Interestingly, some who clearly met the criteria for nIHH subsequently had normal puberty, sexual maturation, and fertility **after** receiving sex-hormone replacement therapy. Ten per cent (5/50) of the subjects in one large series showed this phenotypic response, including increased testicular size suggesting sustained gonadotropin secretion, pulsatile LH secretion, adult levels of testosterone, and a normal ejaculate and sperm count [26, 27]. Isolated anosmia without hypogonadism has also been identified within families with FGFR1 mutations.

ANOS1 (also called KAL1) was the first gene discovered to cause KS. Its protein product, anosmin-1, has neural cell adhesion properties and is produced in a variety of neural tissues, including the developing olfactory bulb [26, 28]. Anosmin-1 modulates FGFR1 signalling on the cell surface and has roles in neurogenesis and the motility and migration of neural cells. It is an absolute requirement for the developing GnRH neurons to traverse the cribriform plate and take residence in the arcuate nucleus [29]. In addition to hypogonadism, individuals with ANOS1 deficiency lack olfactory epithelium and the olfactory bulb and tracts. Associated anomalies include synkinesia (mirror movements of the extremities), unilateral renal agenesis, oculomotor abnormalities, sensorineural hearing loss, and mid-line facial clefts. The mode of inheritance is X-linked recessive; as such it is approximately 10-fold more common in males. Approximately 10–20% of males with KS have mutations in the ANOS1 gene [30].

HS6ST1 is the gene encoding heparan sulphate 6-O-sulfotransferase 1 and is required for the modification of heparan sulphate. 6-O-sulfated heparan sulphate is part of a proteoglycan complex located on the cell surface that is required for normal FGF signalling. HS6ST1 mutations have been identified in kindreds with both KS and nIHH, some members of which have palatal abnormalities. Hypogonadism is inherited in a complex non-Mendelian pattern [31].

Other genes involved in the FGF8 system include FGF17, IL17RD, DUSP6, SPRY4, KLB, and FLRT3, each of which are found in relatively small percentages of patients with KS or nIHH [23].

The prokineticin system is composed of two very similar receptors (PROKR1 and PROKR2) within the rhodopsin receptor family, analogous to the kisspeptin receptor, GPR54. These receptors have two polypeptide ligands, PROK1 and PROK2. The former and its receptor, PROKR1, are primarily found in the gastrointestinal tract, but PROK2 and PROKR2 are located in neuroendocrine areas, including the arcuate nucleus, olfactory tract, and the suprachiasmatic (clock) nucleus. The phenotype includes abnormal development of the olfactory bulbs combined with hypogonadotropic hypogonadism. Humans with mutations in PROK2 or PROKR2 may have the Kallmann or the nIHH phenotype and endocrine profile [26].

Mutations in chromodomain helicase DNA-binding protein 7 (CHD7) are found in the majority of patients with CHARGE syndrome (Coloboma of the eye, Heart defects, Atresia choanae, Retarded growth and development, Genitourinary defects, and Ear abnormalities). Some affected patients have pituitary defects, including growth hormone deficiency and HH. Patients initially identified as having Kallmann syndrome have been subsequently reclassified as having CHARGE syndrome based on the presence of ear abnormalities and detection of mutations in CHD7 [32]. Mutations in CHD7 may occur in association with mutations in FGFR1 with increased frequency (oligogenicity) [20].

Mutations in the *SOX10* gene have been associated with Waardenburg syndrome, a constellation of deafness and pigmentation abnormalities of the eyes and hair. Investigation of the frequency of SOX10 mutations in patients with KS demonstrated gene defects in 38% of those with KS and hearing impairment. SOX10 mutations were found in only 2% of individuals with KS who had normal hearing [33].

Other genes implicated in KS include *WDR11, SEMA3A, SEMA3E, CCDC141, FEZF1, IGSF10, AXL, NELF*, and *SMCHD1* [23, 34, 35].

Genetic Abnormalities Associated Only with nIHH

The *KISS1* and *KISS1R* genes encode kisspeptin and the kisspeptin receptor. A few years after the discovery of the ANOS1 gene came the isolation of the kisspeptin receptor and its cognate ligand, a 54 amino acid peptide comprising residues 68–121 (also known as metastin) of the 145 amino acid precursor, kisspeptin-1 [26, 36]. *KISS1R* is a G-protein-coupled receptor gene that, when mutated, causes autosomal recessive nIHH in both mice and humans, suggesting that it is an obligatory upstream controlling mechanism for pulsatile GnRH secretion. A subset of kisspeptin-secreting neurons demonstrates paracrine or autocrine regulation by neurokinin B and dynorphin. Subjects with mutations in the *KISS1R* gene lack pubertal development at the appropriate time, but have an intact sense of smell. Male and female subjects with mutations in the *KISS1R* gene achieve fertility and normal pregnancy following either exogenous gonadotropin therapy or long-term, pulsatile GnRH administration. As more mutations have been found, the phenotype has expanded to delayed puberty rather than absent pubertal development. In humans, inactivating mutations of this system have been associated with delays in puberty, while activating mutations lead to precocious puberty [37]. These findings solidify the position of the kisspeptin/GPR54 system acting before (upstream of) GnRHR.

The GnRH receptor, GNRHR, is a G-protein coupled receptor expressed on the gonadotropes. Mutations result in impaired GnRH binding, intracellular trafficking, recycling, or signal transduction and cause a spectrum of defects from completely deficient to partial insufficiency of the HPG axis. The mode of inheritance is autosomal recessive. Reports of several series of subjects with nIHH have noted a frequency of GnRHR gene insufficiency ranging from 3.5% to 16.7% [38, 39]. Several patients with homozygous frame shift mutations in the *GNRH1* gene encoding GnRH have been reported [40, 41], but this appears to be a rare cause of nIHH. Affected males had cryptorchidism and microphallus, and both males and females exhibited a complete absence of pubertal development, low gonadotropin concentrations, and low serum levels of testosterone and oestradiol, respectively. The patients were normosmic, and there were no other associated abnormalities. In addition, Chan *et al.* identified several heterozygous variants in patients with nIHH which are of uncertain significance [41].

TAC3/TACR3 encode neurokinin B and its receptor. It has been estimated that mutations in this system may be responsible for over 5% of cases of nIHH [42, 43]. Neurokinin B is co-located in some neurons expressing kisspeptin, where it appears to be involved in autocrine/paracrine regulation of kisspeptin secretion in association with dynorphin. Neurokinin B and its receptor are also expressed in other reproductive tissues, including the uterus and ovary. The majority of the patients identified as having mutations in this system have had abnormalities in the receptor [43, 44]. Nearly all of the male patients had micropenis, but testicular volumes varied, suggesting some degree of testicular function. Interestingly, many affected individuals appear to have had partial or complete recovery of gonadal function in adulthood when observed following treatment [43]. A smaller number of patients with mutations in the *TAC3* gene have been described. However, the phenotype of these individuals is indistinguishable from those with mutations in the receptor, including the potential for gonadal recovery. Study of the TAC3/TACR3 genes in a group of Finnish patients with constitutional delay of growth and puberty did not reveal any abnormalities [45].

Deficiency of leptin leads to a phenotype of early and severe hyperphagia, accelerating weight gain, insulin resistance, impaired T-cell function and nIHH as an adolescent [46]. Circulating leptin concentrations are below the level of sensitivity of common leptin assays. Although thyroid, adrenal and somatotropic functions in adults are normal, the levels of gonadotropins and sex-steroids are within the prepubertal range as is the physical examination. Pulsatile LH secretion is absent; however, with administration of recombinant human leptin, marked changes in body composition (decreased fat mass) and adolescent maturation occurred [46]. Leptin administration to prepubertal children did not induce puberty, indicating that leptin is a permissive factor for puberty. Those with mutations of the leptin receptor have the same phenotype except that leptin levels are elevated and the response to exogenous leptin is absent or attenuated. Recombinant human leptin administration prevented the experimental disruption of LH pulsatility induced by fasting and restored menstrual cyclicity in **some** women with functional hypothalamic amenorrhea [47].

Mutation of the prohormone convertase 1/3 gene (PCSK1) leads to a monogenic obesity syndrome that may include nIHH. The gene is expressed primarily in endocrine glands and hypothalamus, and the protein product is needed to process a variety of prohormones into the mature hormones, including the cleavage of pro-opiomelanocortin (POMC) to α-melanocyte stimulating hormone (α-MSH). Additional conditions seen in patients with PCSK1 mutations include altered thyroid and adrenal function, abnormal glucose metabolism, and intestinal malabsorption. nIHH appears to occur in about 50% of affected individuals [38, 48].

NR0B1, located on the short arm of the X chromosome, encodes DAX1 (dosage-sensitive sex reversal-adrenal hypoplasia congenita) an orphan nuclear receptor predicted to be a transcription factor important in the development of the adrenal cortex and pituitary gonadotrophs. Loss of function of the DAX1 protein leads to an X-linked recessive disorder characterized by hypogonadotropic hypogonadism and adrenal failure without the hyperandrogenism

of virilizing congenital adrenal hyperplasia. Patients with this disorder, called adrenal hypoplasia congenita, have both glucocorticoid and mineralocorticoid deficiencies and hypogonadotropic hypogonadism that may not be manifest until the second decade of life. DAX1 is a negative regulator of SF1-mediated activity. Puberty is usually delayed, especially in boys, and a diagnosis of nIHH can be made at the appropriate age. Some may have a mixed picture of partial nIHH with an added defect at the gonad, illustrating the importance of DAX1 and SF1 for steroidogenesis [49]. Duplication of NR0B1 is associated with an XY disorder of sex development, either with a female phenotype or ambiguous genitalia.

Inactivating mutations of the human LHβ subunit (LHB) lead to nIHH in men and women [50]. All male subjects had normal genitalia at birth, but no pubertal development or fertility. The mode of inheritance is autosomal recessive. Circulating LH levels are undetectable, without an LH response to exogenous GnRH administration but with a normal response to hCG. Females have a milder presentation. One female subject with two brothers having nIHH due to LHβ deficiency presented with full sexual development and secondary amenorrhea and infertility. Menarche occurred at 13 years. She had premature menopause, but remained hypogonadotropic [50]. Thus there can be widely disparate phenotypes with the same mutation, especially as modified by sex. A specific explanation is lacking for the normal spontaneous pubertal development but subsequent premature menopause in the affected woman.

Females with isolated deficiency of the follicle stimulating hormone beta subunit (FSHB) present with delayed pubertal development and primary amenorrhea, normal or high levels of LH, and low or undetectable levels of FSH. Males have normal pubertal development but small testes and oligospermia. Exogenous GnRH administration raises LH, but not FSH levels [51].

Exercising Adolescent and Adult Women and Men

Exercise has an important impact on the HPG axis, and high levels of exercise can induce HH. The prevailing concept in this arena is that of energy conservation or the laws of thermodynamics. One must have 'enough' energy to support body growth and to store energy, predominantly fat, for longer-term energy requirements such as the menstrual cycle, pregnancy, and lactation [52]. The concept is that for current energy requirements:

$$\text{Energy retention} = [\text{energy intake } (\text{EI}) - \text{energy expenditure } (\text{EE})].$$

In experiments with female athletes, Loucks modified this to:

$$\text{Energy availability } (\text{EA}) = [\text{EI} - \text{exercise EE}]/\text{fat-free mass } (\text{FFM}).$$

The latter equation takes into account that it is the FFM that is the metabolically active (fuel burning) tissue and emphasizes how EA may be reduced by either restricting EI or increasing exercise EE (EEE) [52].

This latter equation may be rearranged to make the message clearer for experimental studies:

$$\text{EI} = \text{EEE} + [\text{EA x FFM}].$$

Randomized clinical trials controlling for both EI and EEE have shown that energy balance occurs at EA = 45 kcal/kg FFM *day in healthy young women and that there is damped pulsatile release of LH after 5 days with EA below 30 kcal/kg FFM*day, which roughly corresponds to the resting metabolic rate in healthy young adults. The susceptibility of the HPG axis to alterations in EA is strongly age dependent as might be hypothesized from the very high incidence of subclinical menstrual disorders shortly after menarche. This concept was experimentally tested by decreasing the energy availability to 25 kcal/kg FFM*day in gynecologically younger (5–8 years after menarche) and older (14–18 years after menarche) young women. After 5 days of caloric restriction and exercise, it was noted that only the gynecologically younger women had disrupted pulsatile release of LH. Thus, the gynecologically older women had a more 'robust' HPG axis, and the data likely explain the high incidence of 'athletic amenorrhea' in young women with the female athlete triad: eating disorder, osteopenia, and amenorrhoea [52].

This concept was updated for the International Olympic Committee in 2014 [53] which noted that compelling evidence existed that showed that the aetiologic factor underpinning the female athlete triad was an 'energy deficit relative to the intake and the energy expenditure required to support homeostasis, health and the activities of daily living, growth and sporting events'. They further noted dysregulation and dysfunction in many aspects of the body's physiologic functions. For these reasons they proposed changing the syndrome from Female Athlete Triad to Relative Energy Deficiency in Sport (RED-S) [53]. The authors neatly summarized the health consequences as shown in Figure 7.3.2.2. These include neuropsychiatric issues such as depression and irritability, difficulty concentrating, decreased coordination, and impaired judgment. Metabolic problems include decreases in glycogen stores, muscle strength, and endurance. Impacts on training lead to increased risk for injury and decreased training response.

The concept of high energy output and high psychological stress at low energy intake occurs in men as well, as noted in American Army Rangers during field training [54].

Treatment may be difficult and multiple options are outlined in the RED-S article [53]. The most straightforward approach would be to prescribe a greater caloric intake or to decrease exercise energy expenditure. This is a difficult treatment plan for a highly competitive athlete, likely a gymnast, dancer, or long-distance runner. The American Academy of Pediatrics Committee of Sports Medicine and Fitness, 1999–2000 has presented a series of recommendations [55]. Those relevant to the female athlete triad are shown in Box 7.3.2.1. A detailed update for the Female Athlete Triad (RED-S) taking into account the data shown earlier has been published [56]. Recommendations for screening, considerations for return to play and guidance for the clinician generally follow those in Box 7.3.2.1.

Hypergonadotropic Hypogonadism

The karyotype abnormality consisting of two or more X chromosomes and one or more Y chromosomes is known as Klinefelter syndrome. Klinefelter syndrome is the most common alteration of chromosome number, with a prevalence of 1:650 males in the general population [57]. It is thought that as many as 75% remain undiagnosed, even in adulthood. Infants and young children often have problems with expressive language development, and school aged children may have difficulty with reading. An assortment of

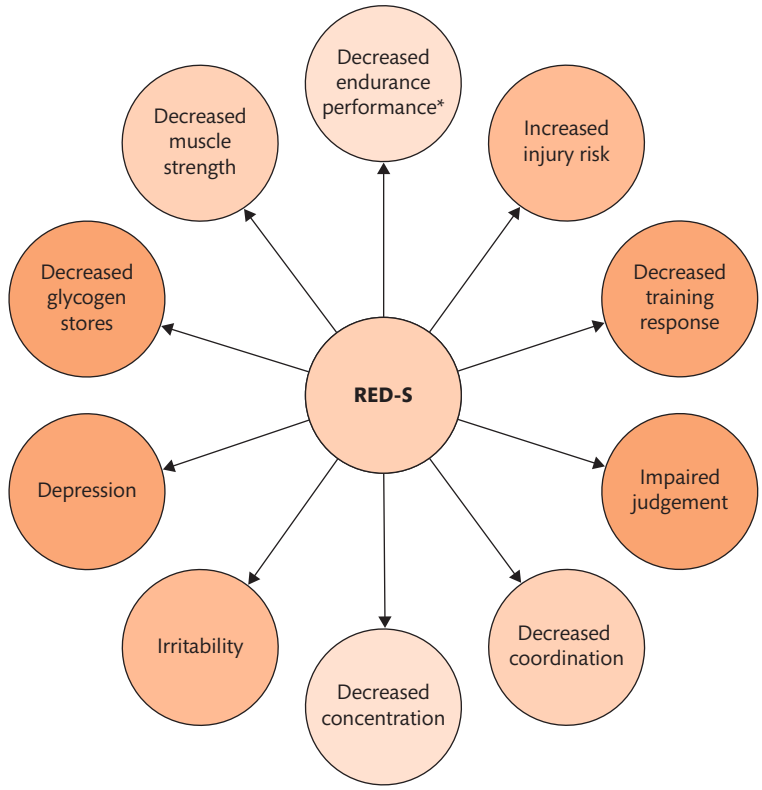


Figure 7.3.2.2 Health consequences of Relative Energy Deficiency in Sport (RED-S), including neuropsychiatric, metabolic problems, and impacts on training.

Reproduced with permission from Mountjoy M, SundgotBorgen J, Burke L, Carter S, Constantini N, Lebrun C, *et al*. The IOC consensus statement: beyond the Female Athlete Triad—Relative Energy Deficiency in Sport (REDS). *Br J Sports Med*. 2014 Apr;48(7):491–7. Copyright © 2014, BMJ Publishing Group Ltd and the British Association of Sport and Exercise Medicine.

Box 7.3.2.1 Guidelines for assuring safe and healthy sports participation for children and adolescents

- Dietary practices; exercise intensity, duration, and frequency; and menstrual history should be reviewed during evaluations that precede participation in sports
- Amenorrhea should not be considered a normal response to exercise. That is, the underlying cause should be sought before denoting it as "athletic amenorrhea" in these gynecologically young women
- Disordered eating should be considered in adolescents with amenorrhea
- Education and counselling should be provided to athletes, parents, and coaches regarding disordered eating, menstrual dysfunction, decreased bone mineralization, and adequate energy and nutrient intake to meet energy expenditure and maintain normal growth and development.
- An adolescent with menstrual dysfunction attributed to exercise should be encouraged to increase energy intake and to modify excessive exercise activity to return to energy balance
- Oestrogen-progesterone supplementation may be considered in mature amenorrhoeic (late adolescent) athletes
- Measurement of bone mineral density and level of 25-hydroxyvitamin D should be considered tools to help make treatment decisions. The adolescent athlete is unlikely to have reached peak bone mass and thus, it is appropriate to use age-adjusted Z-scores rather than T-scores

Data derived in part from the American Academy of Pediatrics, Committee on Sports Medicine and Fitness, 1999–2000, with relevance to the female athlete triad. *Pediatrics*. 2000 Sep;106(3):610-3. (Ref 55)

behavioural challenges is often present. Psychological testing often shows disorders of executive function as well. Physically, the testes appear normal during infancy and childhood, and levels of FSH and LH are normal before puberty. As pubertal maturation unfolds, the testes do not increase in size normally, and the seminiferous tubules gradually become hyalinized, with loss of germ cells and Sertoli cells. Clinically, the testes remain small and may become very firm to palpation. In one study, the mean bitesticular volume (sum of the volume of both testes) was 4.6 ml [58]. LH and FSH levels begin to rise into the upper portion of the normal adult range early in puberty. By mid-puberty, LH and FSH concentrations are often abnormal. Although the onset of puberty is typically normally timed, 80% of affected individuals do not achieve normal adult concentrations of testosterone. The abnormal testosterone secretion results in a slow tempo of physical changes and lack of attainment of normal pubic hair and other sexual hair growth, as well as small penis size and lack of muscular development. The relatively low levels of testosterone and high concentrations of oestradiol predispose adolescents and adults to gynecomastia, which occurs in about 40% of affected individuals.

Essentially all affected men with Klinefelter syndrome have azoospermia or severe oligospermia and are infertile. However, intracytoplasmic germ cell injection (ICSI) has proven to be a feasible approach for those who have viable spermatozoa isolated from ejaculates or after microdissection testicular sperm extraction (mTESE). Additional health problems occurring with higher frequency in men

Table 7.3.2.3 Historical and physical examination features important in the evaluation of delayed puberty and hypogonadism

Historical findings	Physical examination
• Partial pubertal development • Family history of delayed puberty • Sense of smell • Dental development • Chronic disease • Head trauma • Chemotherapy • Radiation therapy • Headache • Visual problems • Galactorrhoea • Delayed language development • Poor school performance • Otitis media in females • Lymphedema in females • Congenital heart disease • Trauma of testes or ovaries • Undescended testes • Viral orchitis • Congenital malformations	• Height percentile • Height velocity • Weight percentile • Stigmata of Turner syndrome in girls • Stigmata of Prader–Willi syndrome • Other dysmorphic features • Delayed dentition • Anosmia or hyposmia • Visual fields • Funduscopy • Presence of thyromegaly • Breast development Tanner stage • Axillary hair • Pubic hair Tanner stage • Presence of other sexual hair growth • Genital Tanner stage • Testicular volume • Synkinesia

with Klinefelter syndrome include metabolic syndrome, type 2 diabetes, osteopenia, venous thromboembolism, ischemic heart disease, and a variety of autoimmune conditions including hypothyroidism. The risk for breast cancer, mediastinal cancers, and non-Hodgkin lymphoma is also higher than in the general population [58]. Men with higher-order aneuploidies such as 48,XXXY or 48,XXYY tend to have more severe and varied manifestations.

Partial or complete loss of one of the two X chromosomes in phenotypic females is known as Turner syndrome. Turner syndrome has an incidence of 1:2000 live born female infants. Turner syndrome is the most common cause of first trimester spontaneous abortions, and only about 1% of 45,X conceptuses are live born. The most common karyotype is 45,X, comprising about 50% of affected girls. Most of the remainder have various forms of mosaicism or partial deletion, usually including loss of at least the short arm of the second X chromosome.

The ovaries of fetuses affected with Turner syndrome show accelerated loss of germ cells [59]. At birth, the oocyte number is reduced to far below normal. The high rate of oocyte loss continues, and the ovaries of affected girls are typically depleted of germ cells within a few years of birth. Classically, girls with Turner syndrome fail to enter puberty, with an absence of breast development. Pubic and axillary hair typically develop normally due to adrenal androgen production. However, approximately 20% of girls will have spontaneous breast development, more commonly those with mosaic or partial forms of Turner syndrome. Spontaneous menses can occur, again most often in girls with mosaic forms, although secondary amenorrhea nearly always develops. Pregnancy may very rarely occur [60].

Gonadotropin levels in the neonatal period and infancy may be normal. In early and mid-childhood, LH and FSH concentrations are also normal, due to the high degree of negative feedback from low levels of oestrogen on the hypothalamus and pituitary. However, by late childhood and early adolescence, gonadotropin levels are often elevated well above the adult range. A karyotype is necessary to confirm the diagnosis and should be obtained regardless of the presence of physical stigmata of Turner syndrome.

The 47,XXX karyotype is common, with a prevalence of 1:900–1000 in the general population. There are few or no recognizable phenotypic features of the condition, although reports indicate that affected individuals are taller than average. There is a higher than normal incidence of neurodevelopmental disorders, such as poor attention span, academic difficulties, decreased verbal fluency, and poor spatial cognition. Although ovarian function is usually normal, primary ovarian dysfunction occurs in a subset of individuals. This may present as delayed puberty or as premature ovarian insufficiency. In studies of adult women with premature ovarian insufficiency, the 47,XXX syndrome occurs in approximately 3% of patients. Gonadotropin levels are elevated, and a karyotype analysis is diagnostic [61]. Females with larger numbers of additional X chromosomes (48,XXXX or 49,XXXXX) are more likely to have phenotypic and developmental abnormalities.

A large number of genes important for normal ovarian function reside on the long arm of the X chromosome. Deletions of portions of Xq and balanced translocations with breakpoints on Xq are associated with ovarian dysfunction. The location of breakpoints associated with hypergonadotropic hypogonadism cluster in two regions, Xq13.3-21.1 and Xq26-qter. The ovarian dysfunction may take the form of either primary or secondary amenorrhea [62].

XY and XX gonadal dysgenesis are terms describing heterogeneous groups of disorders of gonadal differentiation. These conditions are also discussed in Chapter 7.2.2, 'Disorders of Sex Development (DSD in the Newborn', with a focus on the newborn. Their effects on pubertal maturation are reviewed here. Individuals with XY and XX gonadal dysgenesis typically have normal female genitalia and are often not recognized until they fail to enter puberty. Those with XY complete gonadal dysgenesis (Swyer syndrome) have failure of testis determination early in fetal life, with formation of streak gonads and subsequent failure to secrete testosterone and anti-Mullerian hormone (AMH). In the absence of testosterone and AMH, both internal and external genitalia develop along female lines. If partial testis determination occurs, leading to partial Leydig and Sertoli cell function, incomplete masculinization of the internal and external structures occurs, resulting in ambiguous genitalia. Abnormalities of several genes have been implicated as causes of XY complete gonadal dysgenesis, including defects in SRY accounting for 10–15% of children (OMIM #480000) [63], defects in WT-1 associated with Denys–Drash and Frasier syndromes (OMIM #607102), abnormalities of SOX9 associated with campomelic dysplasia (OMIM #608160), NR5A1 encoding SF-1 associated with adrenal hypoplasia (OMIM #184757), and duplication of NR0B1 encoding DAX1 (OMIM #300473). Recently, a number of individuals with 46,XY gonadal dysgenesis have been found to have autosomal dominantly inherited activating mutations in MAP3K1 (OMIM #600982), which activate the genetic cascade leading to ovarian differentiation and inhibiting testis formation [64]. In addition to absent puberty, affected individuals have a 30% incidence of gonadal tumours, most commonly gonadoblastoma and dysgerminoma [65]. Spontaneous pubertal changes in a patient known to have XY complete gonadal dysgenesis should prompt a search for a sex steroid-secreting gonadoblastoma.

Girls with XX gonadal dysgenesis do not have physical features of Turner syndrome, but they are typically somewhat shorter than

average. Similar to individuals with Swyer syndrome, they have normal female internal and external genitalia. Ovarian histology ranges from fibrous streaks to hypoplastic ovaries. Gonadal tumours are uncommon in this population. Identified causes of 46,XX gonadal dysgenesis include mutations in NR5A1 (SF-1, OMIM #184757) and NOBOX (OMIM #610934) [66, 67], and FSH resistance caused by a mutation in the FSH receptor (OMIM #136435).

Those affected by either XY or XX gonadal dysgenesis have elevations of gonadotropin levels by the time of expected puberty. Karyotype analysis will reveal the diagnosis in 46,XY patients, and imaging studies of the pelvis will demonstrate the absence of ovaries. 46,XX gonadal dysgenesis must be distinguished from premature ovarian insufficiency caused by a number of other conditions, including autoimmune oophoritis or exposure to radiation or chemotherapeutic agents.

Complete androgen insensitivity syndrome (CAIS, OMIM #300068, see also Chapter 7.2.2, 'Disorders of Sex Development (DSD) in the Newborn') is caused by mutations in the androgen receptor gene. The prevalence of CAIS has been reported to be between 1:20 400 and 1:99 000 genetic males [68]. The effects on pubertal maturation are reviewed here.

At the time of puberty, the testes secrete testosterone, and testosterone levels typically rise into the adult male range or higher. Because of the androgen resistance, there are few or no clinical signs of androgen activity, such as pubic or axillary hair. Breast development appears to proceed normally, caused by aromatization of circulating testosterone. Primary amenorrhea occurs because of the absence of the uterus.

Gonadotropin levels are often normal at birth and typically are in the normal prepubertal range during childhood. As the individual reaches the age of normal puberty, LH secretion increases because of the lack of negative feedback from testosterone via the androgen receptor. High testosterone and LH concentrations in a female with primary amenorrhea and clinical signs of androgen resistance is virtually diagnostic, and the diagnosis is then confirmed with a karyotype demonstrating a 46,XY composition and genetic studies of the AR gene.

Galactosemia (OMIM #230400) is an inborn error of metabolism most commonly caused by a mutation in the GALT gene encoding galactose-1-phosphate uridyltransferase. Premature ovarian failure occurs in 75–96% of female patients with galactosemia. The age at onset of ovarian dysfunction ranges from childhood to adulthood, and patients may present with absent puberty or may have normal pubertal maturation and menarche but develop secondary amenorrhea later. Ovarian function may wax and wane, with periods of amenorrhea alternating with spontaneous ovarian cycles and possible fertility. Those individuals harbouring more severe mutations are more likely to experience consistent and lifelong ovarian failure. Patients with milder forms of galactosemia caused by defects in genes other than GALT may not have overt features of galactosemia but may present with primary ovarian insufficiency [69].

The mechanism of ovarian damage in cases of galactosemia is unknown. Ovarian tissue has a high content of galactose and its metabolites and normally has high galactose-1-phosphate uridyltransferase activity. In contrast, the testis has low enzymatic activity and low galactose content, presumably accounting for the absence of testicular dysfunction in males with galactosemia. It is thought that accumulation of galactose and galactose-1-phosphate in ovarian cells has direct cytotoxic effects by decreasing the activity of a number of metabolic pathways.

Defects in the receptors for LH and FSH are very rare causes of abnormal pubertal maturation. The LH receptor in the male is critical for normal testosterone secretion *in utero*. Hence, partial loss of LH receptor function (OMIM #152970) causes inadequate testicular secretion of testosterone and ambiguous genitalia in the 46,XY fetus. Alternatively, complete loss of the LH receptor leads to an inability to secrete any testosterone and a subsequent lack of masculinization of the 46,XY fetus leading to normal female external genitalia. If the individual is assigned to the female sex and not diagnosed early in life, there is complete absence of pubertal development and primary amenorrhea, as the testicular tissue will not secrete testosterone and there will be no aromatization to oestrogens. 46,XX females with loss of LH receptor function usually have normally timed breast development but experience primary or secondary amenorrhea and hypoestrogenemia. This highlights the importance of normal FSH activity for females in early puberty and the importance of LH activity to establish normal menses and oestrogen levels in later puberty and adulthood [70].

Abnormalities of the FSH receptor have mainly been described in the Finnish population. Females carrying mutations in the FSH receptor gene (OMIM #136435) usually present with 46,XX gonadal dysgenesis, with absent puberty and primary amenorrhea. Some affected individuals, however, will have spontaneous pubertal development and even menarche, although those identified as having the disorder have all become amenorrhoeic. Males with defects of the FSH receptor have normal pubertal development, normal testosterone levels, and normal or near normal gonadotropin levels. However, they may have oligospermia [71].

In addition to those discussed earlier, a large number of other single gene defects and genetic syndromes are associated with hypergonadotropic hypogonadism. Defects of steroidogenesis may cause disorders of sex development that are recognized in the newborn period as ambiguous genitalia. Some of these disorders, however, will result in a phenotypic female who is unable to synthesize either androgens or oestrogens. These disorders may cause congenital adrenal hyperplasia and include steroidogenic acute regulatory protein (StAR) deficiency (OMIM #600617) and 17-hydroxylase deficiency (OMIM #609300). Patients with carbohydrate-deficient glycoprotein syndrome produce abnormally glycosylated gonadotropins that are biologically inactive. The disordered puberty is more severe in females than males. Noonan syndrome, caused by a defect of the *PTPN11* gene in 50% of individuals (OMIM #176876), is a constellation of features including short stature, characteristic facies, and right-sided cardiac defects, as well as undescended testes. Although females with Noonan syndrome have normal ovarian function, some males may have abnormal Leydig cell function. Some boys with Noonan syndrome have delayed puberty with low gonadotropins. Aside from any evident dysmorphic features, these boys would not be distinguishable from those with CDGP or IHH. Other well-recognized genetic syndromes associated with hypergonadotropic hypogonadism include the fragile X pre-mutation (OMIM #300624), pseudohypoparathyroidism

type 1a (Albright hereditary osteodystrophy, OMIM #103580), autoimmune polyglandular syndrome type 1 (OMIM #607358), progressive external ophthalmoplegia, blepharophimosis syndrome, myotonic dystrophy, and ataxia-telangiectasia syndrome (OMIM #607585) [72, 73].

Vanishing testis syndrome refers to the case of the phenotypically normal male born with bilaterally absent testes. Normally functioning testicular tissue is presumably present in early gestation, as the external and internal genitalia are normally formed and there are typically no Müllerian remnants, implying normal secretion of testosterone and AMH *in utero*. This condition is thought to be due to antenatal bilateral torsion of the testes or other vascular events. This condition is uncommon, occurring in approximately 1:20 000 males. Careful physical examinations at birth and during childhood will reveal apparent bilateral undescended testes, and further evaluation by measuring AMH, inhibin B, or human chorionic gonadotropin (hCG)-stimulated testosterone will demonstrate the absence of functioning testicular tissue. However, if a good physical examination is not performed in childhood, this condition may remain undetected and present with delayed pubertal development. In some cases, the vascular insult may occur near the time of delivery, and bilateral testicular necrosis may be identified [74].

Gonadal tissue is very radiosensitive, with germ cells being particularly prone to radiation injury. Although loss of germ cells in the male leads to infertility, Leydig cells are more resistant to radiation-induced damage. Hence, at lower doses of radiation, there may be loss of fertility with preservation of endocrine function, diagnosed by elevation of FSH with normal LH and testosterone levels. At higher doses of radiation, both fertility and hormone secretion are affected, with elevation of both FSH and LH and low testosterone concentrations. With any degree of radiation exposure in the male child or adolescent, germ cell loss can occur, while Leydig cell injury does not usually occur until doses exceed 20–30 Gy. This situation contrasts with females, in whom germ cell loss is closely tied to loss of endocrine function due to loss of follicle development. The number of oocytes in the female is limited, and exposure later in adolescence or in adulthood, when there are normally fewer oocytes present, is associated with worse endocrine and reproductive outcomes than exposure early in childhood, when the number of oocytes present is larger. Radiation exposure in doses above 10 Gy in pubertal girls is associated with adverse reproductive outcomes, while doses above 15 Gy place prepubertal girls at risk [75].

Chemotherapeutic medications, especially alkylating agents, commonly cause gonadal injury in both prepubertal and pubertal patients. Higher dose protocols are more likely to cause gonadal dysfunction. This group of medications includes cyclophosphamide, ifosfamide, cisplatin, procarbazine, busulfan, chlorambucil, and others. Similar to the case of radiation exposure, females are at higher risk for chemotherapy-induced fertility and hormonal sequelae, while defects of testosterone secretion in males exposed to alkylating agents are uncommon. Overall, males who have survived cancer in childhood have a 24% decrease in fertility, while females have a 10-fold increase in the incidence of premature ovarian insufficiency [76].

A variety of approaches have been used to preserve gonadal function in children and adolescents undergoing cancer treatments. Oophoropexy, which refers to surgical relocation of the ovaries, may move at-risk ovaries out of the field of radiation but results in loss of spontaneous fertility and may make assisted reproductive techniques more difficult. Although oocyte and embryo cryopreservation are accepted techniques for preserving fertility in adults, these are not options for the prepubertal patient, as they require ovarian hyperstimulation with or without a sperm donor. Although these techniques may be available for post-menarchal adolescents, they require a 2–4-week interval of ovarian hyperstimulation before chemotherapy can begin. This may result in an unacceptable delay in many cases [77]. Other techniques, such as ovarian tissue cryopreservation, are being studied but are currently not available outside of clinical trials. Treatment with long acting GnRH analogues has been proposed for suppression of ovarian function with the goal of preserving fertility after cancer treatment. Data have shown variable degrees of success, and this approach is not recommended as the sole fertility preservation technique [78]. For male adolescents undergoing chemotherapy or radiation treatment, semen samples obtained before treatment may be frozen, and this should be offered to all those at risk [76]. Testicular tissue cryopreservation may be performed as part of clinical trials and is considered experimental at this time [78].

Autoimmune oophoritis (AO) presents as premature ovarian insufficiency or less commonly as absent puberty, arrested puberty, or primary amenorrhea. It is estimated that 1–5% of women with premature ovarian insufficiency have ovarian autoimmunity. AO is commonly reported in type 1 autoimmune polyglandular syndrome (APS, OMIM #607358), and may be less often found in type 2 APS. It is nearly always associated with autoimmune adrenalitis, and approximately 20% of females with primary adrenal insufficiency, will have AO. In the setting of type 1 APS, 36% of females will have AO, and 4% of males will have autoimmune orchitis. Affected individuals have clinical ovarian insufficiency, low oestradiol concentrations, and elevation of FSH. Anti-Müllerian hormone levels are low. Other manifestations of autoimmunity may be present, including hypothyroidism, systemic lupus erythematosus, Sjögren syndrome, rheumatoid arthritis, haemolytic anaemia, pernicious anaemia, inflammatory bowel disease, and others [79]. Antibodies to several cytochrome P450 steroidogenic enzymes have been documented in patients with AO, but assays for these autoantibodies are not commonly available. However, because of the close association between autoimmune adrenalitis and AO, antiadrenal antibodies directed against the 21-hydroxylase enzyme may serve as a surrogate marker in the patient with clinical ovarian insufficiency. Because primordial follicles are preserved early in the course of AO, treatment with immunosuppressive agents such as glucocorticoids may be effective [73, 80].

Viral orchitis is an uncommon problem that usually affects adult men. Infection with the mumps virus causes orchitis in 15–30% of postpubertal males, although orchitis is rare in children. In 15–30% of cases, orchitis is bilateral. Symptoms include pain, oedema, and erythema of the scrotum. After resolution, approximately half of affected men have decreases in testicular volume. Some patients will have minor alterations in endocrine function, but sterility is rare [81]. In women, mumps oophoritis is less common, affecting 5% of infected adult women. Mumps oophoritis is very rare in childhood, and rarely causes alterations of endocrine function or fertility in any affected female [82].

Other causes of orchitis and oophoritis include bacterial infections with *Chlamydia trachomatis, Neisseria gonorrhoea*, and *Escherichia coli*; other viral pathogens such as Coxsackie virus and varicella; and non-infectious causes such as Henoch–Schönlein purpura and other vasculitides.

Diagnosis of Delayed Puberty and Hypogonadism

Diagnostic algorithms for the evaluation of delayed puberty and possible hypogonadism are presented in Figures 7.3.2.3 and 7.3.2.4. The evaluation starts with a careful history and physical examination (Table 7.3.2.3). Important historical features include the presence or absence of any signs of puberty, including the age at onset and the tempo of progression. Inquiry about the patient's sense of smell is important, because patients and families will not volunteer this information in this setting. The growth pattern of the patient must be assessed by examination of a standard growth chart. Finally, the timing of puberty in the parents, siblings, and other relatives is critical, as many of the possible conditions are heritable.

Important physical features include the patient's height and weight, the presence or absence of any signs of puberty, and the quantification of these signs if possible. Quantification of pubertal development includes assessment of Tanner stages, measurement of testicular volume and penile length in males, and assessment of breast size in females or gynecomastia in males.

The laboratory diagnosis begins with determinations of LH and FSH concentrations. Normal or low gonadotropin levels direct the evaluation along the hypogonadotropic hypogonadism pathway, while elevations of gonadotropins suggest a diagnosis involving primary testicular or ovarian abnormalities. Sex steroid levels should be measured in the morning because of the diurnal variation, particularly for testosterone. Many commercially available assays for oestradiol are insensitive, and biologic effects of oestradiol may be present at serum concentrations below the detection limit. Serial measurements of gonadotropins and sex steroids may be helpful to quantify the progression of puberty or lack thereof.

Treatment of Delayed Puberty and Hypogonadism

The principle goal of treatment of adolescents with delayed or absent puberty is the attainment of sex steroid levels and physical

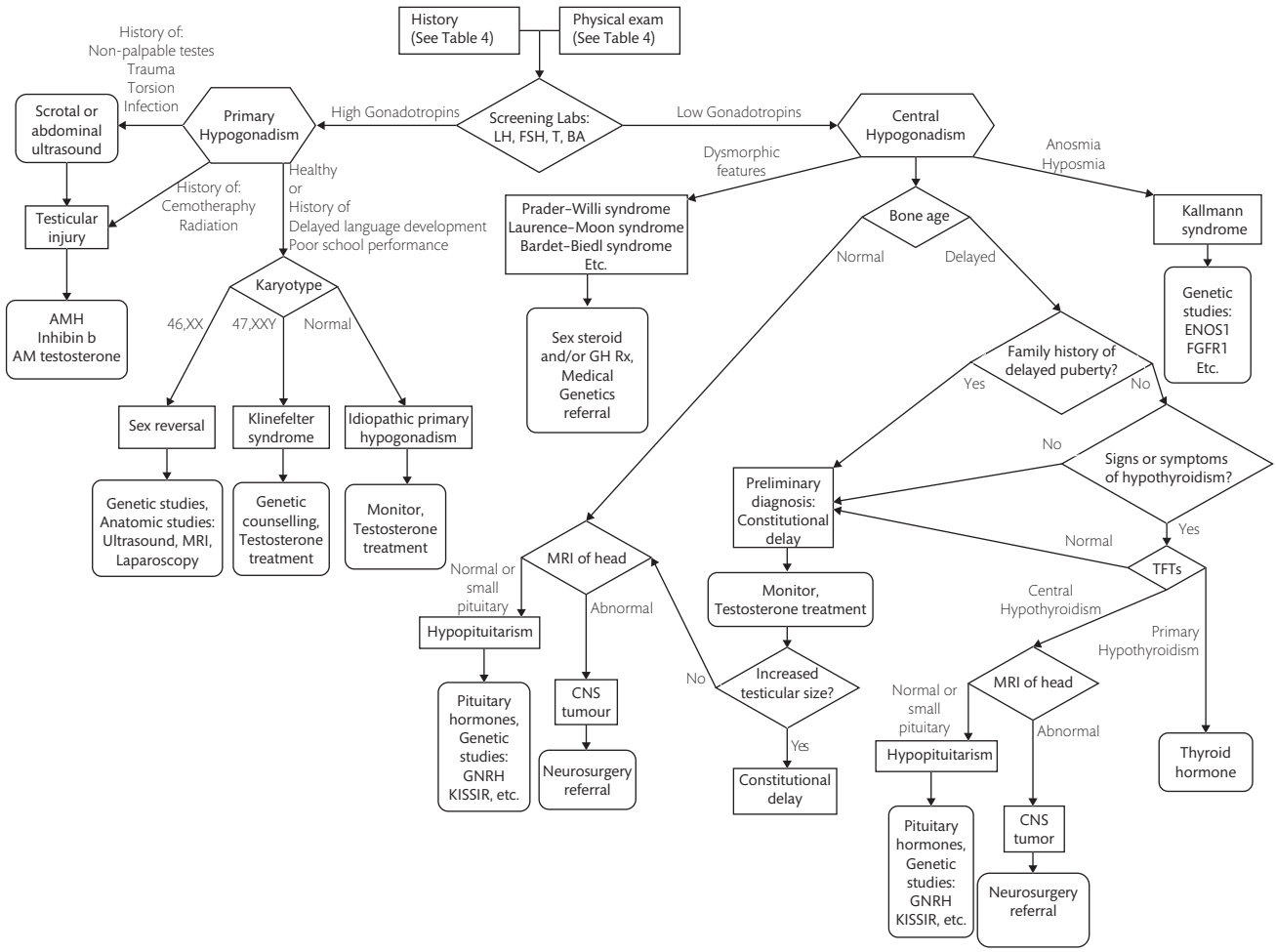


Figure 7.3.2.3 Algorithm for the evaluation and management of delayed puberty and hypogonadism in boys. AMH, Anti-Müllerian hormone; BA, bone age; CNS, central nervous system; FSH, follicle stimulating hormone; GH, growth hormone; LH, luteinizing hormone; MRI, magnetic resonance imaging; Rx, treatment; T, testosterone; TFTs, thyroid function studies.

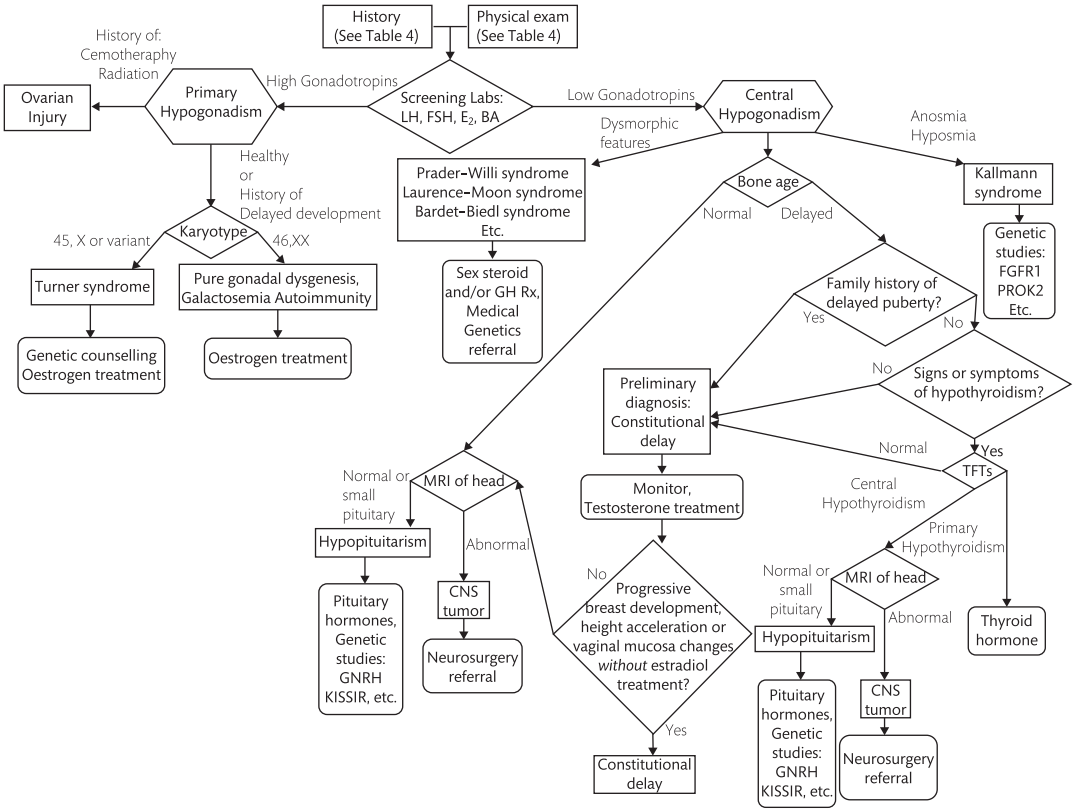


Figure 7.3.2.4 Algorithm for the evaluation and management of delayed puberty and hypogonadism in girls. BA, bone age; CNS, central nervous system; E2, oestradiol; FSH, follicle stimulating hormone; GH, growth hormone; LH, luteinizing hormone; MRI, magnetic resonance imaging; Rx, treatment; TFTs, thyroid function studies.

features that are appropriate for the stage of adolescent maturation. Replacement may be temporary in cases of transient delayed puberty, such as constitutional delay of growth and puberty (CDGP), or long-term in adolescents with permanent absence of pubertal maturation. Subsequent goals of sex steroid therapy in the adolescent are to promote physiological linear growth and development of secondary sexual characteristics and to permit the acquisition of normal body composition, including muscle mass and bone mineral content, to mimic the normal physiologic process. Regardless of whether the patient has hyper- or hypogonadotropic hypogonadism, long-term sex steroid replacement is accomplished similarly.

Androgen Preparations

Agents presently available for androgen replacement are listed in Table 7.3.2.4. Not all of the preparations are universally available, and treatment may need to be individualized using the locally available preparations. Few preparations are suggested for the induction of puberty, mainly because the dosage forms are metered to full androgen replacement for the adult. As most are drug delivery devices, they cannot be easily altered to deliver the small, and then increasing, doses of testosterone required to permit normal pubertal maturation in hypogonadal adolescents or in those with CDGP.

The oral 17α-hydroxylated preparations are virtually never used because of the concern of liver toxicity, and there is very little experience in adolescents with the buccal or nasal formulations. Testosterone undecanoate is not considered hepatotoxic. Unmodified testosterone, administered orally, is rapidly inactivated by first-pass hepatic metabolism.

Oxandrolone is a non-aromatizable, non-5-α reducible oral steroid hormone, which interacts directly with the androgen receptor. It augments height velocity in boys with CDGP without disproportionate advancement of skeletal maturation, which would theoretically decrease adult height [83]. In prepubertal boys, a marked increase in body mass index, a decrease in the triceps and subscapular skin folds, and an increase in the upper body muscle area have been noted following oxandrolone, 2.5 mg/day [84]. At that dose, oxandrolone was an anabolic steroid without significant virilizing action.

The primary clinical uses for androgen therapy in adolescent males are to induce pubertal maturation and as replacement therapy in those with permanent hypogonadism of either the hypogonadotropic or hypergonadotropic variety. The most common cause, although its precise incidence is unknown, is CDGP.

Without intervention, most adolescents with CDGP will undergo normal pubertal maturation spontaneously and most, but not all, will reach their genetically determined mid-parental height **range** [85]. Many adolescents suffer significant emotional distress because they differ in their appearance from their peers during these years. Androgen therapy was initially proposed for boys with CDGP to alleviate their psychological discomfort, in addition to the beneficial effects on bone mineral accrual, lean body mass (protein anabolism) and the regional distribution of body fat.

Table 7.3.2.4 Androgen preparations

Agent	Dosing route	Induction of puberty	Adult dose
Enanthate/cypionate	Intramuscular	50–100 mg IM every 4 weeks for 4–6 months 1. For CDGP: Dose can be increased after 3 months if no signs of pubertal maturation. Dose increase by 25–50 mg/month with evaluation at 3–4-month intervals until testes 8–10 ml 2. IHH: Increase dose by 25–50 mg/month at 4–6-month intervals until full adult replacement dose of ~200 mg/2 week or 100 mg/week	200 mg twice monthly
Mixture of esters (Sustanon)	Intramuscular	50–75 mg every 4 weeks for 4–6 months 1. For CDGP: Dose can be increased after 3 months if no sighs of pubertal maturation. Dose increased by 25 mg/month with evaluation at 3–4-month intervals until testicular volume reaches 6–8 mL 2. IHH: Dose may be escalated gradually to 100 to 150 mg/ month before changing to 250 mg/3 weeks, the approximate adult dose	250 mg every 2–3 weeks
Undecanoate	Intramuscular	Not recommended	750–1000 mg every 10–12 weeks (range 10–14 weeks)
Undecanoate	Oral	40 mg daily; titrate gradually to adult dose	80 mg twice daily
Testosterone	Transdermal patch	Begin with 2 mg/day patch for perhaps 12 h/day and gradually titrate to full adult dose	2.5–5 mg
Testosterone	Transdermal gel	0.5 gm/day increasing dose based on T levels (low level of evidence) Increase dose in stepwise fashion by 0.5 g increments to the full adult dose	5–10 g/day

The authors recognize that the majority of boys who have sought subspecialist evaluation are anxious to begin androgen therapy, and are generally pleased with the results, albeit subtle, even after 3 months of therapy with 50–75 mg long-acting testosterone esters per month. Their reasons to begin therapy fall into the appearance (too young), social (not considered a peer), and athletic (cannot compete because of size and lack of strength) spheres. The dose is increased by 25–50 mg per month every 3 months if spontaneous pubertal maturation has not occurred. This may be assessed by an increase in testicular size, indicating gonadotropin release despite the negative feedback effects of the exogenous testosterone, or by rising early morning levels of testosterone obtained at least 3 weeks following the previous testosterone injection. Therapy is discontinued when the testicular volume is approximately 6–8 mL. The longest acting ester available is testosterone undecanoate, and because of its ~3-month duration of action, it is not appropriate for adolescents with presumed CDGP.

For those with permanent hypogonadism, the escalation of the testosterone cypionate or enanthate ester continues until one reaches approximately 150 mg monthly, after which one might consider switching to twice monthly at 100 mg each administration, increasing to a maximum of 200 mg twice monthly, which is the average adult dose. At about the time of moving to twice monthly injections, one might consider the cutaneous gel which is available in sachet packages of 2.5 and 5 g or a metered pump dispensing 1.25 g, which we consider a mid-pubertal dose. The advantage of the gel is that the levels of testosterone, dihydrotestosterone (DHT), and oestradiol are all within the physiological range for the entire day. Subsequent alterations in dose can be made by measuring the circulating level of testosterone. Those receiving intramuscular testosterone have higher than normal levels of T, DHT, and oestradiol for the first part of the dosing interval and lower than normal for the latter part of the interval.

Oestrogen Preparations

Puberty can be induced using an oestrogen preparation started at approximately 12 years, an age appropriate to induce breast development without affecting the physiological rate of bone maturation or growth potential [86]. The initial dose should be low, one sixth to one quarter of the adult dose (Table 7.3.2.5), and increased gradually at intervals of 3–6 months. The administration of very low-dose depot oestradiol (initial dose of 0.2 mg/month, IM) permitted

Table 7.3.2.5 Oestrogen preparations

Agent	Dosing route	Induction of puberty	Adult dose
Oestradiol cypionate	Intramuscular	0.2 mg/month	~2.5 mg/month
Micronized oestradiol	Oral	0.25 mg/day	1–4 mg/day
Oestradiol	Transdermal patch*	6.25–12.5 µg twice weekly	100 µg twice weekly
Oestradiol	Transdermal gel	0.1 mg/day	1.5 mg/day

*The designated patch dosage is the amount delivered daily; how often one changes the patch depends on the manufacturer's specifications.

relatively age-appropriate (12–13 years of age) feminization without interfering with the effect of growth hormone on the enhancement of height potential [87].

In a study of girls with Turner syndrome, 56 subjects who were receiving rhGH therapy received low dose, oral micronized oestradiol (5 μg/kg*day) for 2 years followed by one year at 7.5 μg/kg*day and then 10 μg/kg*day [88]. The main purpose of the study was to induce feminization as close to physiologically as practicable without negatively affecting adult height. The majority had similar breast development and progression compared to a population of Dutch girls, but delayed by approximately 2 years. As previously reported, adult uterine size was not attained, likely due to the 45,X karyotype and not due to the protocol for the escalation of the oestradiol dose. No direct comparison with the transdermal application of oestradiol was made.

Transdermal oestradiol patches have been used with some advantages over the traditional oral administration of oestradiol or one of its synthetic analogues. Nocturnal application (3.1–12.5 μg/day of 17β-oestradiol) in girls with hyper- or hypogonadotropic hypogonadism produced levels of oestradiol that are similar to those measured in girls during spontaneous adolescent maturation [89]. Cutaneous administration of oestradiol in hydroalcoholic gel is another therapeutic possibility that can be used to induce puberty.

Detailed recommendations for low dose oestrogen treatment options for pubertal induction in adolescent girls with Turner syndrome and considerations for use may be found in [90]. The authors 'suggest the following practical approach to feminize girls with Turner syndrome: initiate puberty with low dose transdermal oestradiol, when available, starting with half of a 14 μg patch applied weekly or a whole 14 or 25 μg patch for one week per month at 11–12 years of age'. These investigators also recommend against the use of conjugated oestrogens unless other forms are unavailable. In general, the dose of oestrogen can be increased every 6–12 months to reach full replacement dose after two or three years of therapy.

Replacement therapy in most patients eventually involves cyclic oestrogen/progesterone therapy. Once full oestrogen replacement has been reached, cyclical progesterone (e.g. 5–10 mg of medroxyprogesterone acetate) can be added every month to induce monthly menstrual bleeding. Once full pubertal maturation has been reached, the oestrogen dosage should be the minimum that will maintain normal menstrual periods, prevent calcium loss from bone, and permit the accrual of peak bone mass early in the third decade [90]. At that time, low dose birth control pills are an alternative option; however, by definition the dose of oestrogen is greater than the physiologic dose for an adult woman.

REFERENCES

1. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*, 1970; **45**: 13–23.
2. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*, 1969; **44**: 291–303.
3. Sempe M. Croissance et developpement. In: Sizonenko P, Griscelli C, eds. *Precis de Pediatrie*. Lausanne: Editions Payot, 1996.
4. Sizonenko P, Limoli C. Height velocity (HV) during puberty in relation to bone age (BA). *Pediatr Res*, 1993; **22**: S89.
5. Rosenfield RL, Lipton RB, Drum ML. Thelarche, pubarche, and menarche attainment in children with normal and elevated body mass index. *Pediatrics*, 2009; **123**: 84–8.
6. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988–1994. *Arch Pediatr Adolesc Med*, 2001; **155**: 1022–8.
7. Marti-Henneberg C, Vizmanos B. The duration of puberty in girls is related to the timing of its onset. *J Pediatr*, 1997; **131**: 618–21.
8. Reiter EO, Lee PA. Have the onset and tempo of puberty changed? *Arch Pediatr Adolesc Med*, 2001; **155**: 988–9.
9. Styne D, Grumbach M. Physiology and disorders of puberty. In: Melmed S, Polonsky KS, Larsen P, Kronenberg H, eds. *Williams Textbook of Endocrinology*, 13th edn. Philadelphia, PA: W.B. Saunders Company, 2016: 1074–218.
10. Palmert MR, Boepple PA. Variation in the timing of puberty: clinical spectrum and genetic investigation. *J Clin Endocrinol Metab*, 2001; 86: 2364–8.
11. Wohlfahrt-Veje C, Mouritsen A, Hagen CP, *et al*. Pubertal onset in boys and girls is influenced by pubertal timing of both parents. *J Clin Endocrinol Metab*, 2016; 101: 2667–74.
12. Cousminer DL, Stergiakouli E, Berry DJ, *et al*. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum Mol Genet*, 2014; **23**: 4452–64.
13. Li W, Liu Q, Deng X, Chen Y, Liu S, Story M. Association between obesity and puberty timing: a systematic review and meta-analysis. *Int J Environ Res Public Health*, 2017; **14**: pii: E1266.
14. Kempinas Wde G. Environmental factors in dysregulation of puberty timing and progression. *Reprod Toxicol*, 2014; **44**: v–vi.
15. Ambler GR. Androgen therapy for delayed male puberty. *Curr Opin Endocrinol Diabetes Obes*, 2009; **16**: 232–9.
16. Wehkalampi K, Widen E, Laine T, Palotie A, Dunkel L. Patterns of inheritance of constitutional delay of growth and puberty in families of adolescent girls and boys referred to specialist pediatric care. *J Clin Endocrinol Metab*, 2008; **93**: 723–8.
17. Zhu J, Chan YM. Adult consequences of self-limited delayed puberty. *Pediatrics*, 2017; **139**: pii: e20163177.
18. Gross RT, Duke PM. The effect of early versus late physical maturation on adolescent behavior. *Pediatr Clin North Am*, 1980; **27**: 71–7.
19. Mehta A, Dattani MT. Developmental disorders of the hypothalamus and pituitary gland associated with congenital hypopituitarism. *Best Pract Res Clin Endocrinol Metab*, 2008; **22**: 191–206.
20. Cassatella D, Howard SR, Acierno JS, *et al*. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*, 2018; **178**: 377–88.
21. Miraoui H, Dwyer AA, Sykiotis GP, *et al*. Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 are identified in individuals with congenital hypogonadotropic hypogonadism. *Am J Hum Genet*, 2013; **92**: 725–43.
22. Sykiotis GP, Plummer L, Hughes VA, *et al*. Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. *Proc Natl Acad Sci U S A*, 2010; **107**: 15140–4.
23. Topaloglu AK. Update on the genetics of idiopathic hypogonadotropic hypogonadism. *J Clin Res Pediatr Endocrinol*, 2017; **9**(Suppl 2): 113–22.

24. Falardeau J, Chung WC, Beenken A, *et al*. Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *J Clin Invest*, 2008; 118: 2822–31.
25. Trarbach EB, Abreu AP, Silveira LF, *et al*. Nonsense mutations in FGF8 gene causing different degrees of human gonadotropin-releasing deficiency. *J Clin Endocrinol Metab*, 2010; **95**: 3491–6.
26. Crowley WF, Jr., Pitteloud N, Seminara S. New genes controlling human reproduction and how you find them. *Trans Am Clin Climatol Assoc*, 2008; **119**: 29–37; discussion –8.
27. Raivio T, Falardeau J, Dwyer A, *et al*. Reversal of idiopathic hypogonadotropic hypogonadism. *N Engl J Med*, 2007; **357**: 863–73.
28. Dode C, Hardelin JP. Kallmann syndrome: fibroblast growth factor signaling insufficiency? *J Mol Med (Berl)*, 2004; **82**: 725–34.
29. de Castro F, Seal R, Maggi R, Group of HcfKALn. ANOS1: a unified nomenclature for Kallmann syndrome 1 gene (KAL1) and anosmin-1. *Brief Funct Genomics*, 2017; **16**: 205–10.
30. Pedersen-White JR, Chorich LP, Bick DP, Sherins RJ, Layman LC. The prevalence of intragenic deletions in patients with idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. *Mol Hum Reprod*, 2008; **14**: 367–70.
31. Tornberg J, Sykiotis GP, Keefe K, *et al*. Heparan sulfate 6-O-sulfotransferase 1, a gene involved in extracellular sugar modifications, is mutated in patients with idiopathic hypogonadotrophic hypogonadism. *Proc Natl Acad Sci U S A*, 2011; **108**: 11524–9.
32. Jongmans MC, van Ravenswaaij-Arts CM, Pitteloud N, *et al*. CHD7 mutations in patients initially diagnosed with Kallmann syndrome—the clinical overlap with CHARGE syndrome. *Clin Genet*, 2009; **75**: 65–71.
33. Pingault V, Bodereau V, Baral V, *et al*. Loss-of-function mutations in SOX10 cause Kallmann syndrome with deafness. *Am J Hum Genet*, 2013; **92**: 707–24.
34. Salian-Mehta S, Xu M, Knox AJ, *et al*. Functional consequences of AXL sequence variants in hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2014; **99**: 1452–60.
35. Xu N, Kim HG, Bhagavath B, *et al*. Nasal embryonic LHRH factor (NELF) mutations in patients with normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Fertil Steril*, 2011; 95: 1613–20; e1–7.
36. Seminara SB, Messager S, Chatzidaki EE, *et al*. The GPR54 gene as a regulator of puberty. *N Engl J Med*, 2003; **349**: 1614–27.
37. Teles MG, Bianco SD, Brito VN, *et al*. A GPR54-activating mutation in a patient with central precocious puberty. *N Engl J Med*, 2008; **358**: 709–15.
38. Kaminski BA, Palmert MR. Genetic control of pubertal timing. *Curr Opin Pediatr*, 2008; **20**: 458–64.
39. Beranova M, Oliveira LM, Bedecarrats GY, *et al*. Prevalence, phenotypic spectrum, and modes of inheritance of gonadotropin-releasing hormone receptor mutations in idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2001; **86**: 1580–8.
40. Bouligand J, Ghervan C, Tello JA, *et al*. Isolated familial hypogonadotropic hypogonadism and a GNRH1 mutation. *N Engl J Med*, 2009; **360**: 2742–8.
41. Chan YM, de Guillebon A, Lang-Muritano M, *et al*. GNRH1 mutations in patients with idiopathic hypogonadotropic hypogonadism. *Proc Natl Acad Sci U S A*, 2009; **106**: 11703–8.
42. Francou B, Bouligand J, Voican A, *et al*. Normosmic congenital hypogonadotropic hypogonadism due to TAC3/TACR3 mutations: characterization of neuroendocrine phenotypes and novel mutations. *PLoS One*, 2011; **6**: e25614.
43. Gianetti E, Tusset C, Noel SD, *et al*. TAC3/TACR3 mutations reveal preferential activation of gonadotropin-releasing hormone release by neurokinin B in neonatal life followed by reversal in adulthood. *J Clin Endocrinol Metab*, 2010; **95**: 2857–67.
44. Topaloglu AK, Reimann F, Guclu M, *et al*. TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat Genet*, 2009; **41**: 354–8.
45. Vaaralahti K, Wehkalampi K, Tommiska J, Laitinen EM, Dunkel L, Raivio T. The role of gene defects underlying isolated hypogonadotropic hypogonadism in patients with constitutional delay of growth and puberty. *Fertil Steril*, 2011; **95**: 2756–8.
46. Farooqi IS, Wangensteen T, Collins S, *et al*. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*, 2007; **356**: 237–47.
47. Welt CK, Chan JL, Bullen J, *et al*. Recombinant human leptin in women with hypothalamic amenorrhea. N Engl J Med, 2004; **351**(10): 987–97.
48. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JW. PCSK1 mutations and human endocrinopathies: from obesity to gastrointestinal disorders. *Endocr Rev*, 2016; **37**: 347–71.
49. Habiby RL, Boepple P, Nachtigall L, Sluss PM, Crowley WF, Jr., Jameson JL. Adrenal hypoplasia congenita with hypogonadotropic hypogonadism: evidence that DAX-1 mutations lead to combined hypothalmic and pituitary defects in gonadotropin production. *J Clin Invest*, 1996; **98**: 1055–62.
50. Lofrano-Porto A, Barra GB, Giacomini LA, *et al*. Luteinizing hormone beta mutation and hypogonadism in men and women. *N Engl J Med*, 2007; **357**: 897–904.
51. Trarbach EB, Silveira LG, Latronico AC. Genetic insights into human isolated gonadotropin deficiency. Pituitary, 2007; **10**: 381–91.
52. Loucks AB. Energy availability and infertility. *Curr Opin Endocrinol Diabetes Obes*, 2007; **14**: 470–4.
53. Mountjoy M, Sundgot-Borgen J, Burke L, *et al*. The IOC consensus statement: beyond the Female Athlete Triad—Relative Energy Deficiency in Sport (RED-S). *Br J Sports Med*, 2014; **48**: 491–7.
54. Friedl KE, Moore RJ, Hoyt RW, Marchitelli LJ, Martinez-Lopez LE, Askew EW. Endocrine markers of semistarvation in healthy lean men in a multistressor environment. *J Appl Physiol (1985)*, 2000; **88**: 1820–30.
55. American Academy of Pediatrics. Committee on Sports Medicine and Fitness. Medical concerns in the female athlete. *Pediatrics*, 2000; **106**: 610–13.
56. Weiss Kelly AK, Hecht S, Council on Sports M, Fitness. The female athlete triad. *Pediatrics*, 2016; **138**: pii: e20160922.
57. Gravholt CH, Chang S, Wallentin M, Fedder J, Moore P, Skakkebaek A. Klinefelter syndrome: integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*, 2018; **39**: 389–423.
58. Chang S, Skakkebaek A, Trolle C, *et al*. Anthropometry in Klinefelter syndrome--multifactorial influences due to CAG length, testosterone treatment and possibly intrauterine hypogonadism. *J Clin Endocrinol Metab*, 2015; **100**: E508–17.
59. Reynaud K, Cortvrindt R, Verlinde F, De Schepper J, Bourgain C, Smitz J. Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil Steril*, 2004; **81**: 1112–19.
60. Sybert VP, McCauley E. Turner's syndrome. *N Engl J Med*, 2004; **351**: 1227–38.
61. Goswami R, Goswami D, Kabra M, Gupta N, Dubey S, Dadhwal V. Prevalence of the triple X syndrome in phenotypically normal women with premature ovarian failure and its association with autoimmune thyroid disorders. *Fertil Steril*, 2003; **80**: 1052–4.

62. Goswami D, Conway GS. Premature ovarian failure. *Hum Reprod Update*, 2005; **11**: 391–410.
63. Brennan J, Capel B. One tissue, two fates: molecular genetic events that underlie testis versus ovary development. *Nat Rev Genet*, 2004; **5**: 509–21.
64. Granados A, Alaniz VI, Mohnach L, *et al*. MAP3K1-related gonadal dysgenesis: six new cases and review of the literature. *Am J Med Genet C Semin Med Genet*, 2017; **175**: 253–9.
65. Sarafoglou K, Ostrer H. Clinical review 111: familial sex reversal: a review. *J Clin Endocrinol Metab*, 2000; **85**: 483–93.
66. Lourenco D, Brauner R, Lin L, *et al*. Mutations in NR5A1 associated with ovarian insufficiency. *N Engl J Med*, 2009; **360**: 1200–10.
67. Qin Y, Choi Y, Zhao H, Simpson JL, Chen ZJ, Rajkovic A. NOBOX homeobox mutation causes premature ovarian failure. *Am J Hum Genet*, 2007; **81**: 576–81.
68. Hughes IA, Deeb A. Androgen resistance. *Best Pract Res Clin Endocrinol Metab*, 2006; **20**: 577–98.
69. Flint AC, Hopwood NJ, Kasa-Vubu JZ. Management quandary. Diagnostic quandary: premature ovarian failure and galactosemia variants in adolescent girls with delayed puberty. *J Pediatr Adolesc Gynecol*, 2009; **22**: 330–5.
70. Laml T, Preyer O, Umek W, Hengstschlager M, Hanzal H. Genetic disorders in premature ovarian failure. *Hum Reprod Update*, 2002; **8**: 483–91.
71. Huhtaniemi I, Alevizaki M. Gonadotrophin resistance. *Best Pract Res Clin Endocrinol Metab*, 2006; **20**: 561–76.
72. Wittenberger MD, Hagerman RJ, Sherman SL, *et al*. The FMR1 premutation and reproduction. *Fertil Steril*, 2007; **87**: 456–65.
73. Nelson LM. Clinical practice. Primary ovarian insufficiency. *N Engl J Med*, 2009; **360**: 606–14.
74. Law H, Mushtaq I, Wingrove K, Malone M, Sebire NJ. Histopathological features of testicular regression syndrome: relation to patient age and implications for management. *Fetal Pediatr Pathol*, 2006; **25**: 119–29.
75. Rose SR, Horne VE, Howell J, *et al*. Late endocrine effects of childhood cancer. *Nat Rev Endocrinol*, 2016; **12**: 319–36.
76. Kurt BA, Armstrong GT, Cash DK, *et al*. Primary care management of the childhood cancer survivor. *J Pediatr*, 2008; **152**: 458–66.
77. Benard J, Sermondade N, Grynberg M. Fertility preservation in endocrine disorders during transition for girls. *Endocr Dev*, 2018; **33**: 128–48.
78. Oktay K, Harvey BE, Partridge AH, *et al*. Fertility preservation in patients with cancer: ASCO clinical practice guideline update. *J Clin Oncol*, 2018; **36**: 1994–2001.
79. Silva CA, Yamakami LY, Aikawa NE, Araujo DB, Carvalho JF, Bonfa E. Autoimmune primary ovarian insufficiency. *Autoimmun Rev*, 2014; **13**: 427–30.
80. Bakalov VK, Anasti JN, Calis KA, *et al*. Autoimmune oophoritis as a mechanism of follicular dysfunction in women with 46,XX spontaneous premature ovarian failure. *Fertil Steril*, 2005; **84**: 958–65.
81. Verajankorva E, Laato M, Pollanen P. Analysis of 508 infertile male patients in south-western Finland in 1980–2000: hormonal status and factors predisposing to immunological infertility. *Eur J Obstet Gynecol Reprod Biol*, 2003; **111**: 173–8.
82. Hviid A, Rubin S, Muhlemann K. Mumps. *Lancet*, 2008; **371**: 932–44.
83. Stanhope R, Brook CG. Oxandrolone in low dose for constitutional delay of growth and puberty in boys. *Arch Dis Child*, 1985; **60**: 379–81.
84. Papadimitriou A, Preece MA, Rolland-Cachera MF, Stanhope R. The anabolic steroid oxandrolone increases muscle mass in prepubertal boys with constitutional delay of growth. *J Pediatr Endocrinol Metab*, 2001; **14**: 725–7.
85. Crowne EC, Shalet SM, Wallace WH, Eminson DM, Price DA. Final height in boys with untreated constitutional delay in growth and puberty. *Arch Dis Child*, 1990; **65**: 1109–12.
86. Richmond E, Rogol A. Estrogen therapy to treat delayed puberty in adolescent girls. *Rev Endocrinol*, 2007; **1**: 32–4.
87. Rosenfield RL, Devine N, Hunold JJ, Mauras N, Moshang T, Jr., Root AW. Salutary effects of combining early very low-dose systemic estradiol with growth hormone therapy in girls with Turner syndrome. *J Clin Endocrinol Metab*, 2005; **90**: 6424–30.
88. Bannink EM, van Sassen C, van Buuren S, *et al*. Puberty induction in Turner syndrome: results of oestrogen treatment on development of secondary sexual characteristics, uterine dimensions and serum hormone levels. *Clin Endocrinol (Oxf)*, 2009; **70**: 265–73.
89. Piippo S, Lenko H, Kainulainen P, Sipila I. Use of percutaneous estrogen gel for induction of puberty in girls with Turner syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 3241–7.
90. Klein KO, Rosenfield RL, Santen RJ, *et al*. Estrogen replacement in turner syndrome: literature review and practical considerations. *J Clin Endocrinol Metab*, 2018; **103**: 1790–803.

7.3.3 Precocious Puberty

Diagnosis and Management

Juliane Léger and Jean-Claude Carel

Introduction

Precocious puberty (PP) may be caused by central or peripheral mechanisms. Precocious puberty leads to the progressive development of secondary sexual characteristics including breast development in girls and testicular enlargement in boys, together with the development of pubic hair, and an acceleration of growth velocity and bone maturation, resulting in premature fusion of the growth plates, potentially responsible for adult height deficit [1].

Premature sexual maturation is a frequent cause for referral to paediatric endocrinology. Clinical evaluation is generally sufficient to reassure the patient and family, but premature sexual maturation may reveal severe conditions and need a thorough evaluation to identify its cause and potential for progression, in order to propose an appropriate treatment [1]. Although the use of long-acting GnRH agonists has revolutionized the treatment of central precocious

puberty, questions remain regarding their optimal use [2]. One of the main ongoing controversial issues is the definition of normal pubertal development. The onset of puberty may also be subject to the effects of environmental (secular trends, adoption, absence of the father and possible exposure to estrogenic endocrine-disrupting chemicals), socioeconomic disadvantage, nutritional (body mass index), and constitutional (genetics, ethnicity) factors, with implications for the definition of precocious puberty [3, 4].

Normal and Premature Sexual Maturation

Normal pubertal development results from the activation of pulsatile hypothalamic GnRH secretion leading to activation of the pituitary-gonadal axis. The onset of puberty is marked clinically by breast development in girls and testicular enlargement in boys. Tanner stages (see Figure 7.3.1.1 in Chapter 7.3.1, 'Recognizing Normal and Disordered Pubertal Development') are used to evaluate pubertal development and the onset of puberty corresponds to Tanner 2 breast (B2) stage in girls (best assessed by both inspection and palpation) and Tanner 2 genital (G2) stage in boys (testicular volume greater than 4 ml or testicular length greater than 25 mm).

Defining the normal limits of pubertal development is difficult, given the paucity of truly normative data and the numbers of components to consider including not only pubertal onset, but also progression of puberty and onset of menarche. Although there have been discussions to decrease the cut-off defining early pubertal onset in girls, the traditional limits of 8 years in girls and 9.5 years in boys are still used by most paediatric endocrinologists [1]. Sexual hair development is a component of pubertal maturation that reflects the actions of androgen produced by the gonads or by the adrenals. Similarly, the traditional limits of pubic hair development have been set to 8 years in girls and 9.5 years in boys with wide ethnical variations.

There are several elements to remember when considering normal and abnormal pubertal development. First, the activation of the gonadotropic axis is not an all or nothing phenomenon, but evolves over several years, starting 2 to 3 years before the clinical onset of puberty. Second, the mean duration of the transition from one stage to the next is generally close to 6 months on average but varies among individuals. In slowly progressive puberty, pubertal development can remain at the B2 stage or revert to the B1 stage before resuming later. It is noteworthy that although the mean age at the B2 stage has decreased in the past decades, the age at menarche has been relatively stable, indicating a longer duration of puberty [5]. Third, the onset of puberty is affected by a number of factors in addition to ethnicity [6]. The prevalence of precocious puberty is about 10 times higher in girls than in boys and has been estimated at 0.2% of girls and less than 0.05% of boys. The number of cases increases with age for both sexes and onset of puberty below 6 years in girls and 7 years in boys accounts for only 15% of all cases [7, 8].

Aetiologies and Mechanisms Underlying Premature Sexual Development

Figure 7.3.3.1 summarizes the mechanisms underlying premature sexual development. Premature sexual development results from the action of sex steroids or compounds with sex steroid activity on target organs. The most common mechanism of progressive precocious puberty is the early activation of pulsatile GnRH secretion.

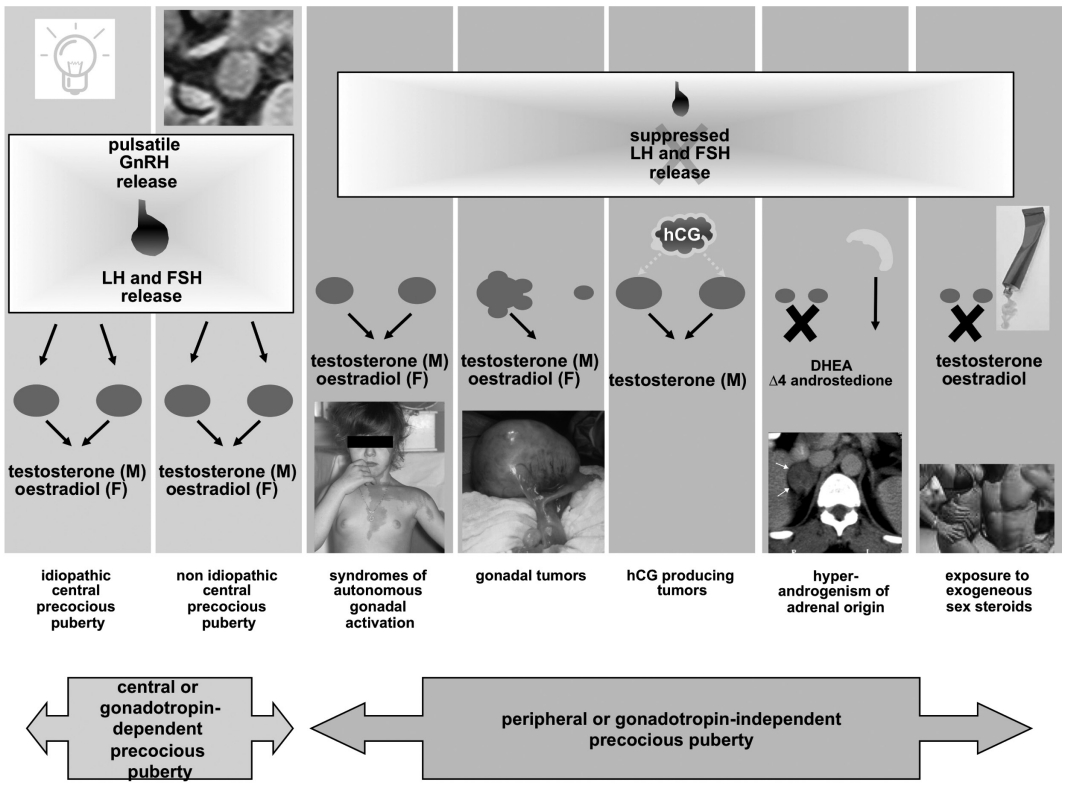


Figure 7.3.3.1 Principal mechanisms of premature sexual development.

Peripheral, or gonadotropin-independent precious puberty is due to the production of sex steroids by gonadal or adrenal tissue, independently of gonadotropins which are generally supressed. Exposure to exogenous sex steroids or to compounds with steroidal activity can also result in premature sexual development. It is also important to recognize variants of pubertal development that can mimic precocious puberty but do not lead to long-term consequences and are usually benign.

Central precocious puberty is due to the premature activation of GnRH secretion and results in a hormonal pattern that is similar to that of normal puberty, although early. Central precocious puberty can be due to hypothalamic tumours or lesions or be idiopathic in the majority of cases, in particular in girls (Table 7.3.3.1). Recent studies have implicated the inactivation of Makorin ring finger 3 (*MKRN3*) genes in 'idiopathic' CPP [9, 10]. *MKRN3* is an imprinted gene located on the long arm of chromosome 15, with a potentially inhibitory effect on GnRH secretion. *MKRN3* gene defects have been identified as a cause of paternally transmitted familial CPP, but such defects do not underlie maternally transmitted CPP and are rarely involved in sporadic forms [11, 12]. More recently, a paternally inherited DLK1 deletion has been reported in one family with isolated CPP [13]. Non-isolated idiopathic CPP has also been reported associated with complex and rare syndromic phenotypes such as Silver–Russell and Temple syndromes or encephalopathy disorders [14].

Peripheral precocious puberty can result from gonadal, adrenal or hCG-producing tumours, activating mutations in the gonadotropic pathway and exposure to exogenous sex steroids (Table 7.3.3.2). Peripheral precocious puberty can rarely lead to activation of pulsatile GnRH secretion and to central precocious puberty (Table 7.3.3.1).

It is essential to recognize that most cases of premature sexual maturation correspond to benign variants of normal development that can occur throughout childhood (Table 7.3.3.3). This is particularly true in girls below the age of 2 to 3 years where the condition is known as premature thelarche. Similarly in older girls, at least 50% of cases of premature sexual maturation will regress or stop progressing and no treatment is necessary [15]. Although the mechanism underlying these cases of non-progressive precocious puberty is unknown, the gonadotropic axis is not activated. Premature thelarche probably represents an exaggerated form of the physiological early gonadotropin surge that is delayed in girls relative to boys.

Consequences of Premature Sexual Maturation

Progressive premature sexual maturation can have consequences on growth and psycho-social development. Growth velocity is accelerated as compared to normal values for age and bone age is advanced in most cases. The acceleration of bone maturation can lead

Table 7.3.3.1 Clinical characteristics of the various forms of central precocious puberty

Cause	Symptoms and signs	Evaluation
Due to a CNS lesion		
Hypothalamic hamartoma	May be associated with gelastic (laughing attacks), focal or tonic-clonic seizures	MRI: Mass in the floor of the third ventricle iso-intense to normal tissue without contrast enhancement
Or other hypothalamic tumours: • Glioma involving the hypothalamus and/or the optic chiasm. • Astrocytoma, • Ependymoma, • Pinealoma, • Germ cell tumours.	May include headache, visual changes, cognitive changes, symptoms/signs of anterior or posterior pituitary deficiency (e.g. decreased growth velocity, polyuria/polydipsia), fatigue, visual field defects If CNS tumour (glioma) associated with neurofibromatosis, may have other features of neurofibromatosis (cutaneous neurofibromas, café au lait spots, Lisch nodules)	MRI: contrast-enhanced mass that may involve the optic pathways (chiasm, nerve, tract) or the hypothalamus (astrocytoma, glioma) or that may involve the hypothalamus and pituitary stalk (germ cell tumour), May have evidence of intracranial hypertension. May have signs of anterior or posterior pituitary deficiency (e.g. hypernatremia) If germ cell tumour: ßhCG detectable in blood or CSF
Cerebral malformations involving the hypothalamus: • Suprasellar arachnoid cyst, • Hydrocephalus, • Septo-optic dysplasia, • Myelomeningocele, • Ectopic neurohypophysis.	May have neurodevelopmental deficits, macrocrania, visual impairment, nystagmus, obesity, polyuria/polydipsia, decreased growth velocity.	May have signs of anterior or posterior pituitary deficiency (e.g. hypernatremia) or hyperprolactinemia
Acquired injury: • Cranial irradiation, • Head trauma, • Infections, • Perinatal insults.	Relevant history Symptoms and signs of anterior or posterior pituitary deficiency may be present	MRI may reveal condition-specific sequelae or may be normal
Idiopathic—no CNS lesion	≈ 92% of girls and ≈ 50% of boys History of familial precocious puberty or adoption may be present Syndromic/chromosomal disorders	No hypothalamic abnormality on the head MRI. The anterior pituitary may be enlarged *MKRN 3* gene evaluation if paternally transmitted
Secondary to early exposure to sex steroids		
Associated with any cause of gonadotropin-independent precocious puberty.	Relevant history	

Table 7.3.3.2 Clinical characteristics of the various forms of peripheral precocious puberty

Disorder	Characteristic symptoms and signs	Test results
Autonomous gonadal activation		
McCune–Albright syndrome and recurrent autonomous ovarian cysts due to somatic activating mutation of the GNAS gene resulting in increased signal transduction in the Gs pathway	Mostly in girls. Typically rapid progression of breast development and early occurrence of vaginal bleeding (before or within a few months of breast development) Precocious puberty may be isolated or associated with café-au-lait pigmented skin lesions (with irregular borders) or bone pain due to polyostotic fibrous dysplasia. More rarely other signs of endocrine hyperfunction (e.g. hypercortisolism, hyperthyroidism), liver cholestasis or cardiac rhythm abnormalities	Typically large ovarian cyst or cysts on pelvic ultrasound examination, Bone lesions of fibrous dysplasia May have laboratory evidence of hypercortisolism, hyperthyroidism, increased GH secretion, hypophosphatemia, liver cholestasis
Familial male-limited precocious puberty due to germinal activating mutations of the LH receptor gene	A familial history of dominant precocious puberty limited to boys (but transmitted by mothers) may be present but some cases are sporadic	Activating mutation of the LH receptor gene
Germline mutations of GNAS gene resulting in dual loss and gain of function. (rare)	Single case report of a boy with concomitant pseudohypoparathyroidism and gonadotropin-independent precocious puberty	
Tumours		
Granulosa cell tumours of the ovary.	Rapid progression of breast development, abdominal pain may occur. The tumour may be palpable on abdominal examination	Tumour detection on ultrasound or CT scan
Androgen producing ovarian tumours	Progressive virilization	Tumour detection on ultrasound or CT scan
Testicular Leydig cell tumours	Progressive virilization; testicular asymmetry (the tumour itself is rarely palpable)	Tumour detection on testicular ultrasound
hCG producing tumours	Tumours can originate in the liver or mediastinum. Pubertal symptoms in boys only. May be associated with Klinefelter syndrome	Elevated serum hCG
Adrenal disorders	Manifest with signs of androgen exposure	
Congenital adrenal hyperplasia	Increased androgen production leading to virilization in boys and girls	Increased adrenal steroid precursors in serum, mainly 17OH-progesterone (basal or after an ACTH stimulation test)
Adrenal tumour	Increased androgen production leading to virilization in boys and girls. Very rarely, oestrogen-producing adrenal tumour	Tumour on abdominal ultrasound or CT scan Elevated DHEAS, or adrenal steroid precursors
Generalized glucocorticoid resistance	Symptoms and signs of mineralocorticoid and adrenal androgen excess, such as hypertension and hypokalaemic alkalosis	Elevated free urinary cortisol and plasma cortisol
Environmental agents		
Exogenous sex steroids	Manifestations vary with the type of preparation (androgenic or estrogenic); most commonly described after topical exposure to androgens; tracing the source of exposure may be difficult	Endocrine evaluation can be misleading due to widely variable serum levels of sex steroids with time
Exposure to oestrogenic endocrine-disrupting chemicals	May play a role in precocious puberty in adopted children (by modulating the timing of pubertal gonadotropic axis activation) although this remains unproven	No validated biochemical test
Severe untreated primary hypothyroidism	Signs of hypothyroidism. No increase of growth velocity Manifest mostly with increased testicular volume in the absence of virilization. Due to a cross-reactivity of elevated TSH to the FSH receptor	Elevated serum TSH levels, low free T_4 level. No bone age advancement

to premature fusion of the growth plates and short stature. Several studies have assessed adult height in individuals with a history of precocious puberty. In older published series of untreated patients, mean heights ranged from 151 to 156 cm in boys and 150 to 154 cm in girls, corresponding to a loss of about 20 cm in boys and 12 cm in girls relative to normal adult height [16]. However, these numbers correspond to historical series of patients with severe early onset precocious puberty which are not representative of the majority of patients seen in the clinic today. Height loss due to precocious puberty is inversely correlated with the age at pubertal onset, and currently treated patients tend to have later onset of puberty than those in historical series [16].

Parents often seek treatment in girls because they fear early menarche [17]. However, there is little data to predict the age of menarche following early onset of puberty. In the general population, the time from breast development to menarche is longer for

Table 7.3.3.3 Benign variants of premature sexual maturation

Condition		
Non-progressive precocious puberty	See Table 7.3.3.5 for differential characteristics with progressive central precocious puberty	
Isolated precocious thelarche	Unilateral or bilateral breast development; particularly frequent before the age of 3 years	No further evaluation needed in most cases
Isolated precocious pubarche	Pubic hair development can be associated with adult body odour, axillary hair, or mild acne	Normal cortisol precursors in serum, including normal levels of 17OH-progesterone after ACTH stimulation
Isolated precocious menarche	Isolated vaginal bleeding without breast development or pubic hair, and no genital trauma. It is important to evaluate clinically for a vaginal lesion (sex abuse, foreign body, tumour)	

children with an earlier onset of puberty, ranging from a mean of 2.8 years when breast development begins at age 9 to 1.4 years when breast development begins at age 12 [18].

Adverse psychosocial outcomes are also a concern, but the available data specific to patients with precocious puberty have serious limitations [2]. In the general population, a higher proportion of early-maturing adolescents engage in exploratory behaviours (sexual intercourse, legal, and illegal substance use) and at an earlier age, than adolescents maturing within the normal age range or later [19]. In addition, the risk for sexual abuse seems to be higher in girls or women with early sexual maturation [20]. However, the relevance of these findings to precocious puberty is unclear, and they should not be used to justify intervention.

Evaluation of the Child with Premature Sexual Development

The evaluation of patients with premature sexual development should address several questions: (1) Is sexual development really occurring outside the normal temporal range? (2) What is the underlying mechanism and is it associated with a risk of a serious condition, such as an intracranial lesion? (3) Is pubertal development likely to progress? And (4) Would this impair the child's normal physical and psychosocial development?

Clinical Evaluation

A complete family history (age at onset of puberty in parents and siblings) and personal history including the age at onset and progression of sexual development should be taken. Any evidence suggesting possible central nervous system disorder, such as headache, increased head circumference, visual impairment or seizures should be sought. Growth should be evaluated by drawing a complete growth chart, because progressive precocious puberty is almost invariably associated with and sometimes preceded by an acceleration of growth velocity.

The stage of pubertal development should be classified according to Tanner (see Figure 7.3.1.1 in Chapter 7.3.1). Careful assessment is needed in obese girls to avoid overestimating breast development. The development of pubic hair results from the effects of androgens, which may be produced by testes or ovaries in central precocious puberty. Acne, oily skin, and hair may be present and result from the action of androgens. In girls, pubic hair in the absence of breast development is suggestive of adrenal disorders, premature pubarche or exposure to androgens. In boys, measurement of testicular volume may suggest the cause of puberty, as testicular volume increases in central precocious puberty as in normal puberty and in cases of peripheral precocious puberty due to testicular disorders (although generally less so), whereas it remains prepubertal in adrenal disorders, premature pubarche and other causes of peripheral precocious puberty. Physical examination should also assess for signs of specific causes of precocious puberty, such as hyperpigmented skin lesions suggesting neurofibromatosis or McCune–Albright syndrome. It is also important to recognize clinically the benign variants of precocious pubertal development with usually isolated and non-progressive secondary sexual characteristic (breast or pubic hair), normal or slightly increased growth velocity and no or slight bone age advancement, if performed (Table 7.3.3.3).

Premature sexual development can be associated with high levels of anxiety in girls and psychological evaluation of the child and of the familial environment may be useful.

Laboratory Evaluation and Imaging

Additional testing is generally recommended in all boys with precocious pubertal development, in girls who present with precocious Tanner 3 breast stage or higher or in girls with precocious B2 stage and additional criteria such as increased growth velocity, advanced bone age, symptoms or signs suggestive of central nervous system dysfunction or of peripheral precocious puberty.

Bone Age

Bone age measured using a reference atlas such as Greulich & Pyle evaluates the impact of sex steroids on epiphyseal maturation and is usually advanced in progressive precocious puberty. Caution should be taken in overinterpreting bone age, since there is a physiological scatter of approximately plus or minus one year of bone age vs chronological age in Caucasian children and a systematic advance of bone age in children of African descent when using references obtained in Caucasians. Bone age can also be used to predict adult height, although with a low precision (95% confidence interval of about ± 6 cm) and a tendency to overestimate adult height in precocious puberty.

Hormonal Measurements

Hormonal measurements that can be useful for the evaluation of premature sexual maturation are summarized in Table 7.3.3.4.

- Sex steroids should be determined in the morning, using assays with detection limits adapted to paediatric values. Most boys with precocious puberty have morning plasma testosterone values in the pubertal range. In girls, serum oestradiol levels are highly

Table 7.3.3.4 Hormonal evaluation for the evaluation of premature sexual maturation

	Technical requirements	Significance	Limitations	Usefulness
Serum oestradiol (girls)	Use morning values due to circadian variation. Use assay with a lower limit of detection of ≈ 5 pg/ml (18 pmol/L) or lower	Markedly elevated levels ≈ >100 pg/ml (367 pmol/L) suggest ovarian cyst or tumour	Levels can be normal in *bona fide* central precocious puberty. Difficulties in interpreting values measured with immuno-enzymatic methods (falsely high values close to the limit of detection of the assay)	First line test together with basal LH in girls. However, poor sensitivity to discriminate early pubertal from prepubertal levels
Serum testosterone	Use morning values due to circadian variation. Use assay with a lower limit of detection of ≈ 0.1 ng/ml (0.35 nmol/L)	Boys: reliable marker of testicular activation. Girls: use if signs of hyperandrogenism; elevated testosterone levels suggest adrenal disorders	Difficulties in interpreting values measured with immuno-enzymatic methods (falsely high values close to the limit of detection of the assay)	First line test with basal LH in boys. High sensitivity to confirm precocious puberty
Serum LH	Use morning values due to circadian rhythm Use ultrasensitive assays with a lower limit of detection of ≈ 0.1 UI/L or lower	Basal LH measurement poorly discriminates between prepubertal and early pubertal children Values >0.3 to 0.4 UI/L indicative of central precocious puberty with a high specificity and a low sensitivity in some series	Wide interassay variations; assay characteristics must be taken into account when interpreting the results	First line screening test in association with oestradiol or testosterone measurement. If clearly elevated can obviate the need for a stimulation test
Peak LH after stimulation with GnRH* or GnRH agonist	Can be performed at any time of the day Assay requirements similar to baseline measurements	Peak LH level above the pubertal cut off with elevated sex steroid levels indicate progressive central puberty Suppressed peak LH level with elevated sex steroid levels indicate peripheral precocious puberty	Wide interassay variations; assay characteristics must be taken into account when interpreting the results. Paucity of normative values to define cut-offs; values of5 to 8 UI/L are most often considered 'high' in children aged from 4 to 8 years. Higher cut-offs should be used in younger children due to transient activation of the gonadotropic axis. Peak values vary with the stimulating agent used (GnRH or GnRH agonist)	Gold standard for the diagnosis of central precocious puberty
Peak FSH after stimulation with GnRH* or GnRH agonist		Peak LH/FSH ratio typically increases during puberty; high ratios are used as a secondary criterion for progressive central puberty; this is less useful with more sensitive LH assays available	Poorly validated, in particular with sandwich-antibody assays for gonadotropin measurements	Can be useful as an additional criterion when a GnRH or GnRH agonist test is performed
Serum ßhCG		Produced by germ cell tumours. Can be detected in serum (peripheral tumours) or in CSF (intracranial tumour)	Peripheral production of ßhCG leads to pubertal development in boys and not in girls	Measurement warranted in boys with peripheral precocious puberty to identify a germ cell tumour and in the CSF when a lesion compatible with a germ cell tumour is detected by MRI.
Serum DHEAS		Produced by the adrenals, marker of androgen-producing adrenal tumours or of adrenal enzymatic defect	Also increased in precocious pubarche	Measure if androgenic signs (pubic hair) predominate
Serum 17OH-progesterone	Use morning (8 a.m.) values due to circadian rhythm or measure after ACTH stimulation.	Marker of adrenal enzymatic defects (congenital adrenal hyperplasia) Occasionally elevated with adrenal tumours	Borderline elevations are frequent in unaffected carriers of non-classical congenital adrenal hyperplasia	Measure if androgenic signs (pubic hair) predominate

variable and have a low sensitivity for the diagnosis of precocious puberty. Very high oestradiol levels are generally indicative of ovarian diseases (cysts or tumours).

- LH determinations are the key to diagnosis and should be based on ultrasensitive assays. Because prepubertal LH levels are <0.1 IU/L, LH assays used should have a detection limit near 0.1 IU/L. The measurement of gonadotropins following GnRH (or GnRH agonist) stimulation is considered the gold standard. However, normative values are scarce and cut-offs levels are not well validated. During normal puberty, the peak LH level increases progressively with a large overlap between successive pubertal stages resulting in an ability to fully discriminate only stage I and stage IV [16]. Peak LH levels of 5 to 8 IU/L suggest progressive central precocious puberty [17].
- Random LH measurements have been proposed as an alternative but variable cut-off values have been proposed. However, unless LH values are clearly elevated, we consider that it is preferable to confirm the diagnosis of progressive central precocious puberty by a stimulation test before initiating treatment. In girls below the age of 3 or 4 years, gonadotropin levels tend to be physiologically elevated and caution should be taken when interpreting the values to avoid over-diagnosing precocious puberty [21].
- FSH provides less information than LH measurements since FSH levels vary little through pubertal development. However, the stimulated LH/FSH ratio may help differentiate progressive precocious puberty (which tends to have higher LH/FSH ratios) from non-progressive variants that do not require GnRHa therapy.

Pelvic or Testicular Ultrasonography

In girls, pelvic ultrasonography can be used to detect ovarian cysts or tumours. Uterine changes due to oestrogen exposure can be used as an index of progressive puberty. A uterine volume greater than 2.0 ml and a uterine length of more than 34 mm have 89% and 80% sensitivity and 89% and 58% specificity respectively for precocious puberty in one series [22]. Testicular ultrasound scans should be performed if testicular volume is asymmetric or in peripheral precocious puberty, in order to detect Leydig cell tumours, which are generally not palpable.

Brain MRI

Neuroimaging is essential in the etiological evaluation of progressive central precocious puberty. Magnetic resonance imaging (MRI) is the examination of choice in the study of the brain and of the hypothalamic-pituitary region, for the detection of hypothalamic lesions. The prevalence of such lesions is higher in boys (30–80% of cases) than in girls (8–33%), and is much lower when puberty starts after the age of 6 years in girls, this population accounting for the majority of cases. It has been suggested that an algorithm based on age and oestradiol levels could replace MRI, but such an approach has not been clearly validated [23–25].

Differentiating Progressive and Non-Progressive Forms of Central Precocious Puberty

Clinical evaluation, hormonal measurements and imaging usually allow the identification of one of the following situations:

- **Progressive central or gonadotropin-dependent precocious puberty,** with high serum testosterone in boys, variable serum oestradiol in girls, peak serum LH after GnRH stimulation in the pubertal range, advanced bone age and oestrogenized uterus on ultrasound examination.
- **Peripheral or gonadotropin-independent precocious puberty,** with high serum testosterone in boys, generally high and occasionally markedly elevated serum oestradiol in girls, low (suppressed) peak serum LH after GnRH stimulation, advanced bone age and oestrogenized uterus on ultrasound examination.
- **Benign variants of precocious pubertal development**, with low serum sex-steroid levels, normal pelvic ultrasound examination and peak serum LH after GnRH stimulation in the prepubertal range (if done, not necessary in most cases).

Table 7.3.3.5 summarizes features reflecting the intensity and duration of the gonadotropic axis activation that are useful in distinguishing between progressive central precocious puberty and non-progressive forms of precocious puberty. Although these criteria are not fully evidence-based, and reflect personal experience

Table 7.3.3.5 Criteria to differentiate non-progressive forms and progressive central precocious puberty in girls

		Progressive central precocious puberty	Non progressive precocious puberty
Clinical	Pubertal stages	Progression from one stage to the next in 3 to 6 months	Stabilization or regression of pubertal signs
	Growth velocity	Accelerated (≈>6 cm/year)	Usually normal for age
	Bone age	Usually advanced by at least 1 year	Usually within one year of chronological age
	Predicted adult height	Below target height range or declining on serial determinations	Within target height range
Pelvic ultrasonography	Uterine development	Uterine volume >2.0 ml or length >34 mm Pearl-like shaped uterus Endometrial thickening (endometrial echo)	Uterine volume ≤2.0 ml or length ≤34 mm Prepubertal, tubular shaped uterus
Hormonal evaluation	Oestradiol	Usually measurable oestradiol level with advancing pubertal development	Oestradiol not detectable or close to the detection limit
	LH peak after GnRH or GnRH agonist	In the pubertal range	In the prepubertal range

as well as data obtained in cross-sectional and small-sized longitudinal studies, they can provide useful orientation. When discrepant results are obtained, it is recommended to wait a few months and reassess, to avoid unnecessary treatment [15].

The Normal Variants of Puberty

The distinction between early puberty and normal puberty is not clear-cut. There are several variants of normal puberty, which may pose problems for differential diagnosis, particularly as they have a high prevalence [26–28].

Isolated Premature Breast Development or Premature Thelarche

Premature thelarche is isolated breast development before the age of 8 years. There are two peaks in the frequency of premature thelarche: the neonatal period, which is marked by gonadotropin activation, this peak potentially lasting for 2 or 3 years, and the prepubertal period [29]. Premature thelarche differs from early puberty in the absence of any other aspect of sexual development, usually with a lack of scalability of breast development and no acceleration of height velocity or significant advance in bone maturation (≥2 years). Uterine ultrasound scans provide a simple means of checking that there is no change in the uterus. No further exploration or treatment is required and the outcome is the persistence of moderate breast development (in two-thirds of cases) or regression (one-third of cases). However, isolated premature breast development may precede the onset of central precocious puberty, which should not be ignored if patients develop other pubertal signs and an acceleration of height velocity.

Premature Development of Pubic Hair or Premature Pubarche

Premature pubarche is the appearance of pubic hair before the age of 8 years in girls and 9 years in boys. It may be accompanied by clinical signs of hyperandrogenism: acne, axillary hair, accelerated growth rate. It corresponds to adrenal maturation (adrenarche) and is not a differential diagnosis for central precocious puberty. Possible differential diagnoses to be systematically excluded include adrenal tumours and congenital adrenal hyperplasia [30, 31].

Slowly Progressive Forms of Precocious Puberty

Studies monitoring these benign variants of precocious puberty have shown that treatment with GnRH agonists is not appropriate because there tends to be either a total regression of pubertal signs or a slow progression towards puberty (Table 7.3.3.5)

Management

Central Precocious Puberty

GnRH Agonists

GnRH agonists are generally indicated in progressive central precocious puberty. GnRH agonists continuously stimulate the pituitary gonadotrophs, leading to desensitization and decreases in LH release and, to a lesser extent, FSH release [32]. Several GnRH agonists are available in various depot forms and their approval for use in precocious puberty varies between countries. Despite nearly 30 years of use of GnRH agonists in precocious puberty, there are still ongoing questions on their optimal use and an international consensus statement, has summarized the available information and the areas of uncertainty as of 2007 [2].

GnRH agonist treatments should be followed by experienced clinicians and result in the regression or stabilization of pubertal symptoms, decrease of growth velocity and bone age advancement [2]. GnRHa-injection dates should be recorded and adherence with the dosing interval monitored. A suppressed LH response to the stimulation by GnRH, GnRH agonist, or after an injection of the depot preparation (which contains a fraction of free GnRH agonist) is indicative of biochemical efficacy of the treatment but is not recommended routinely. Progression of breast or testicular development usually indicates poor compliance, treatment failure or incorrect diagnosis and requires further evaluation.

There are no randomized controlled trials assessing **long-term outcomes** of the treatment of central precocious puberty with GnRH agonists: at most, heightened outcomes have been evaluated. Among approximately 400 girls treated until a mean age of 11 years, the mean adult height was about 160 cm and mean gains over predicted height varied from 3 to 10 cm [16]. Individual height gains were very variable, but were calculated using predicted height, which is itself poorly reliable. Factors affecting height outcome include initial patient characteristics (lower height if bone age is markedly advanced and shorter predicted height at initiation of treatment) and, in some series, duration of treatment (higher height gains in patients starting treatment at a younger age and with longer durations of treatment).

Other outcomes to consider include bone mineral density, risk of obesity, and psychosocial outcomes. Bone mineral density may decrease during GnRH agonist therapy. However, subsequent bone mass accrual is preserved, and peak bone mass does not seem to be negatively affected by treatment [2]. There have been concerns that GnRH agonist use may affect BMI. However, childhood obesity is associated with earlier pubertal development in girls, and early sexual maturation is associated with increased prevalence of overweight and obesity. Altogether, the available data indicate that long-term GnRH agonist treatment does not seem to cause or aggravate obesity, as judged from BMI [2, 33]. However, the risk of obesity is a concern in girls with premature sexual maturation and BMI should be closely monitored. As discussed earlier, psychosocial evaluation data are scarce in patients with premature sexual maturation and there is little evidence to show whether treatment with GnRH agonists is associated with improved psychological outcome [2, 34, 35].

Although **tolerance** to GnRH agonist treatment is generally considered good, it may be associated with headaches and menopausal symptoms such as hot flushes. Local complications (3–13%) such as sterile abscesses may result in a loss of efficacy and anaphylaxis has been exceptionally described [36].

The **optimal time to stop treatment** has not been established and factors that could influence the decision to stop GnRH agonists include aiming at maximizing height, synchronizing puberty with peers, ameliorating psychological distress, or facilitating care of the developmentally delayed child. However, data only permit analysis of factors that affect adult height. Several variables can be used to decide on when to stop treatment including chronological age, duration of therapy, bone age, height, target height, growth velocity. However, these variables are closely interrelated and cannot be considered independently. In addition, retrospective analyses suggest

that continuing treatment beyond the age of 11 years is associated with no further gains [37]. Therefore, it is reasonable to consider these parameters as well as informed parent and patient preferences, with the goal of menarche occurring near the population norms [2]. Pubertal manifestations generally reappear within months of GnRH agonist treatment being stopped, with a mean time to menarche of 16 months [38]. Long-term fertility has not been fully evaluated, but preliminary observations are reassuring [38, 39].

The addition of growth hormone [40] or oxandrolone [41] when growth velocity decreases or if height prognosis appears to be unsatisfactory has been proposed. However, data are limited on the efficacy and safety of these drugs in children with precocious puberty.

Management of Causal Lesions

When precocious puberty is caused by a hypothalamic lesion (e.g. mass or malformation), management of the causal lesion has generally no effect on the course of pubertal development. Hypothalamic hamartomas should not be treated by surgery for the management of precocious puberty. Precocious puberty associated with the presence of a hypothalamic lesion may progress to gonadotropin deficiency.

Peripheral Precocious Puberty

Management of Causal Lesions

Surgery is indicated for gonadal tumours and postoperative chemo- or radiotherapy should be discussed as part of a multidisciplinary team including surgeons and oncologists.

Large ovarian cysts (greater than 20 ml or 3.4 cm in diameter and typically more than 75 ml or 5.2 cm) should be managed very carefully given the risk of adnexal torsion. In such cases, puncture possibly ultrasound-guided should be considered and allows molecular analysis of the cystic fluid for activating GNAS mutation.

Removal of exogenous exposure to sex steroids is obvious but the search for environmental exposure is often very difficult and requires careful investigation.

Benign Variants of Premature Sexual Maturation

Benign variants of premature sexual maturation should be followed clinically with reassurance to the parents. There is limited data on long term outcomes of individuals with these conditions and it has been suggested that premature pubarche is a risk factor for hyperandrogenism in adulthood.

Medications

There is no aetiological treatment for peripheral causes of precocious puberty and the rarity of the diseases renders evaluation of therapeutic strategies very difficult. In McCune-Albright syndrome and recurrent ovarian cysts, aromatase inhibitors and Selective oestrogen receptor modulators have been used to inhibit the production or action of oestrogens, respectively. These approaches are partly effective but no definitive strategy has emerged [42]. In familial male precocious puberty due to LH receptor activating mutations, ketoconazole, an inhibitor of androgen biosynthesis has been shown to be effective in the long term and the combination of anti-androgens and aromatase inhibitors has been proposed [43]. However, caution must be used with the use of ketoconazole given the risk of liver toxicity. Non classical and classical forms of congenital adrenal hyperplasia should be managed with glucocorticoids.

Conclusion

The main concern when examining a patient with premature sexual development should be the existence of a malignant or potentially threatening lesion, either intracranial, in the gonads, the adrenals, or elsewhere. However, these lesions are exceedingly rare and on a daily basis, the main difficulty is with the differentiation of progressive and non-progressive forms of precocious puberty and with the decision to treat, particularly for girls with an onset of puberty between the ages of 6 and 8 years. The psychological aspects of precocious puberty should also be evaluated during and after the treatment of these patients.

REFERENCES

1. Carel JC, Leger J. Clinical practice. Precocious puberty. *N Engl J Med*, 2008; **358**: 2366–77.
2. Carel JC, Eugster EA, Rogol A, *et al*. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics*, 2009; **123**: e752–62.
3. Sorensen K, Mouritsen A, Aksglaede L, Hagen CP, Mogensen SS, Juul A. Recent secular trends in pubertal timing: implications for evaluation and diagnosis of precocious puberty. *Hormone Res Paediatrics*, 2012; **77**: 137–45.
4. Sun Y, Mensah FK, Azzopardi P, Patton GC, Wake M. Childhood social disadvantage and pubertal timing: a national birth cohort from Australia. *Pediatrics*, 2017; **139**: pii: e20164099.
5. Aksglaede L, Sorensen K, Petersen JH, Skakkebaek NE, Juul A. Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*, 2009; **123**: e932–9.
6. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr Rev*, 2003; **24**: 668–93.
7. Teilmann G, Pedersen CB, Jensen TK, Skakkebaek NE, Juul A. Prevalence and incidence of precocious pubertal development in Denmark: an epidemiologic study based on national registries. *Pediatrics*, 2005; **116**: 1323–8.
8. Le Moal J, Rigou A, Le Tertre A, De Crouy-Channel P, Leger J, Carel JC. Marked geographic patterns in the incidence of idiopathic central precocious puberty: a nationwide study in France. *Eur J Endocrinol*, 2018; **178**: 33–41.
9. Abreu AP, Dauber A, Macedo DB, *et al*. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N Engl J Med*, 2013; **368**: 2467–75.
10. Bulcao Macedo D, Nahime Brito V, Latronico AC. New causes of central precocious puberty: the role of genetic factors. *Neuroendocrinology*, 2014; **100**: 1–8.
11. Macedo DB, Abreu AP, Reis AC, *et al*. Central precocious puberty that appears to be sporadic caused by paternally inherited mutations in the imprinted gene makorin ring finger 3. *J Clin Endocrinol Metab*, 2014; **99**: E1097–103.
12. Simon D, Ba I, Mekhail N, *et al*. Mutations in the maternally imprinted gene MKRN3 are common in familial central precocious puberty. *Eur J Endocrinol*, 2016; **174**: 1–8.

13. Dauber A, Cunha-Silva M, Macedo DB, *et al*. Paternally inherited DLK1 deletion associated with familial central precocious puberty. *J Clin Endocrinol Metab*, 2017; **102**: 1557–67.
14. Latronico AC, Brito VN, Carel JC. Causes, diagnosis, and treatment of central precocious puberty. *Lancet Diabetes Endocrinol*, 2016; **4**: 265–74.
15. Leger J, Reynaud R, Czernichow P. Do all girls with apparent idiopathic precocious puberty require gonadotropin-releasing hormone agonist treatment? *J Pediatr*, 2000; **137**: 819–25.
16. Carel JC, Lahlou N, Roger M, Chaussain JL. Precocious puberty and statural growth. *Hum Reprod Update*, 2004; **10**: 135–47.
17. Xhrouet-Heinrichs D, Lagrou K, Heinrichs C, *et al*. Longitudinal study of behavioral and affective patterns in girls with central precocious puberty during long-acting triptorelin therapy. *Acta Paediatr*, 1997; **86**: 808–15.
18. Marti-Henneberg C, Vizmanos B. The duration of puberty in girls is related to the timing of its onset. *J Pediatr*, 1997; **131**: 618–21.
19. Michaud PA, Suris JC, Deppen A. Gender-related psychological and behavioural correlates of pubertal timing in a national sample of Swiss adolescents. *Mol Cell Endocrinol*, 2006; **254**–5: 172–8.
20. Wise LA, Palmer JR, Rothman EF, Rosenberg L. Childhood abuse and early menarche: findings from the black women's health study. *Am J Public Health*, 2009; **99** Suppl 2: S460–6.
21. Vestergaard ET, Schjorring ME, Kamperis K, *et al*. The follicle-stimulating hormone (FSH) and luteinizing hormone (LH) response to a gonadotropin-releasing hormone analogue test in healthy prepubertal girls aged 10 months to 6 years. *Eur J Endocrinol*, 2017; **176**: 747–53.
22. de Vries L, Horev G, Schwartz M, Phillip M. Ultrasonographic and clinical parameters for early differentiation between precocious puberty and premature thelarche. *Eur J Endocrinol*, 2006; **154**: 891–8.
23. Stanhope R. Gonadotrophin-dependent [correction of dependant] precocious puberty and occult intracranial tumors: which girls should have neuro-imaging? *J Pediatr*, 2003; **143**: 426–7.
24. Chalumeau M, Hadjiathanasiou CG, Ng SM, *et al*. Selecting girls with precocious puberty for brain imaging: validation of European evidence-based diagnosis rule. *J Pediatr*, 2003; **143**: 445–50.
25. Pedicelli S, Alessio P, Scire G, Cappa M, Cianfarani S. Routine screening by brain magnetic resonance imaging is not indicated in every girl with onset of puberty between the ages of 6 and 8 years. *J Clin Endocrinol Metab*, 2014; **99**: 4455–61.
26. Atay Z, Turan S, Guran T, Furman A, Bereket A. The prevalence and risk factors of premature thelarche and pubarche in 4- to 8-year-old girls. *Acta Paediatr*, 2012; **101**: e71–5.
27. Mogensen SS, Aksglaede L, Mouritsen A, *et al*. Diagnostic work-up of 449 consecutive girls who were referred to be evaluated for precocious puberty. *J Clin Endocrinol Metab*, 2011; **96**: 1393–401.
28. Kaplowitz P. Clinical characteristics of 104 children referred for evaluation of precocious puberty. *J Clin Endocrinol Metab*, 2004; **89**: 3644–50.
29. de Vries L, Guz-Mark A, Lazar L, Reches A, Phillip M. Premature thelarche: age at presentation affects clinical course but not clinical characteristics or risk to progress to precocious puberty. *J Pediatr*, 2010; **156**: 466–71.
30. Oberfield SE, Sopher AB, Gerken AT. Approach to the girl with early onset of pubic hair. *J Clin Endocrinol Metab*, 2011; **96**: 1610–22.
31. Ghizzoni L, Gasco V. Premature pubarche. *Horm Res Paediatr*, 2010; **73**: 420–2.
32. Lahlou N, Carel JC, Chaussain JL, Roger M. Pharmacokinetics and pharmacodynamics of GnRH agonists: clinical implications in pediatrics. *J Pediatr Endocrinol Metab*, 2000; **13** Suppl 1: 723–37.
33. Lazar L, Lebenthal Y, Yackobovitch-Gavan M, *et al*. Treated and untreated women with idiopathic precocious puberty: BMI evolution, metabolic outcome, and general health between third and fifth decades. *J Clin Endocrinol Metab*, 2015; **100**: 1445–51.
34. Schoelwer MJ, Donahue KL, Bryk K, Didrick P, Berenbaum SA, Eugster EA. Psychological assessment of mothers and their daughters at the time of diagnosis of precocious puberty. *Int J Pediatr Endocrinol*, 2015; **2015**: 5.
35. Wojniusz S, Callens N, Sutterlin S, *et al*. Cognitive, emotional, and psychosocial functioning of girls treated with pharmacological puberty blockage for idiopathic central precocious puberty. *Front Psychol*, 2016; **7**: 1053.
36. Carel JC, Lahlou N, Jaramillo O, *et al*. Treatment of central precocious puberty by subcutaneous injections of leuprorelin 3-month depot (11.25 mg). *J Clin Endocrinol Metab*, 2002; **87**: 4111–16.
37. Carel JC, Roger M, Ispas S, et al. Final height after long-term treatment with triptorelin slow release for central precocious puberty: importance of statural growth after interruption of treatment. French study group of decapeptyl in precocious puberty. *J Clin Endocrinol Metab*, 1999; **84**: 1973–8.
38. Heger S, Muller M, Ranke M, *et al*. Long-term GnRH agonist treatment for female central precocious puberty does not impair reproductive function. *Mol Cell Endocrinol*, 2006; **25**4–5: 217–20.
39. Lazar L, Meyerovitch J, de Vries L, Phillip M, Lebenthal Y. Treated and untreated women with idiopathic precocious puberty: long-term follow-up and reproductive outcome between the third and fifth decades. *Clin Endocrinol*, 2014; **80**: 570–6.
40. Pasquino AM, Pucarelli I, Segni M, Matrunola M, Cerroni F. Adult height in girls with central precocious puberty treated with gonadotropin-releasing hormone analogues and growth hormone. *J Clin Endocrinol Metab*, 1999; **84**: 449–52.
41. Vottero A, Pedori S, Verna M, *et al*. Final height in girls with central idiopathic precocious puberty treated with gonadotropin-releasing hormone analog and oxandrolone. *J Clin Endocrinol Metab*, 2006; **91**: 1284–7.
42. Neyman A, Eugster EA. Treatment of girls and boys with McCune-Albright syndrome with precocious puberty—update 2017. *Pediatr Endocrinol Rev*, 2017; **15**: 136–41.
43. Leschek EW, Flor AC, Bryant JC, Jones JV, Barnes KM, Cutler GB, Jr. Effect of antiandrogen, aromatase inhibitor, and gonadotropin-releasing hormone analog on adult height in familial male precocious puberty. *J Pediatr*, 2017; **190**: 229–35.

7.4

Transition in Endocrinology

Helena K. Gleeson and Rohana J. Wright

Introduction

Healthcare professionals are increasingly aware that the transition from childhood to adulthood requires special consideration, in terms of meeting the needs of young people with any long term condition, both through healthcare design and delivery and the need for a specific skill and behaviour set in order to do this effectively. Young people can be a forgotten group, not adequately considered in paediatric or adult service development strategies resulting in disrupted care. This phase of life is key, as it is when a young person develops their ideas about, and relationships with the healthcare system and their own health condition, thus setting the scene for future interactions.

Young people with endocrine conditions, whether onset is in early childhood or during adolescence, often require lifelong care, and therefore both paediatric and adult healthcare professionals require knowledge about:

1. Biopsychosocial development and how it can both affect and be affected by having an endocrine condition
2. Key aspects of adolescent health
3. Developmentally appropriate healthcare
4. Effective transition between paediatric and adult services
5. Endocrine specific care during adolescence and young adulthood

Adolescence and Young Adulthood

The WHO have defined 'adolescence' as the period from age 10 to 19 years, but have also described a 'young person' as being from age 10 to 24 years [1]. This supports the move from developmental psychologists to consider biopsychosocial development not only in adolescence but extending into the third decade, embracing the concept of 'emerging adulthood' [2]. Further supporting evidence for the extension into young adulthood is provided from neuroscience research into brain development.

Differential pace of brain development during adolescence and young adulthood has been proposed as providing an explanation for 'adolescent' behaviour, for example, exploratory risk-taking behaviours. The limbic regions develop earlier, the areas associated with emotion and reward, followed much later by the pre-frontal cortex, which enables planning, considers inhibition, and allows future thinking. The pre-frontal cortex is thought to continue developing into the mid-20s [3].

Biopsychosocial development is an interdependent process (Figure 7.4.1), and there are a number of factors, including cultural and socioeconomic, that may alter the rate of progression through adolescence and into young adulthood. Understanding

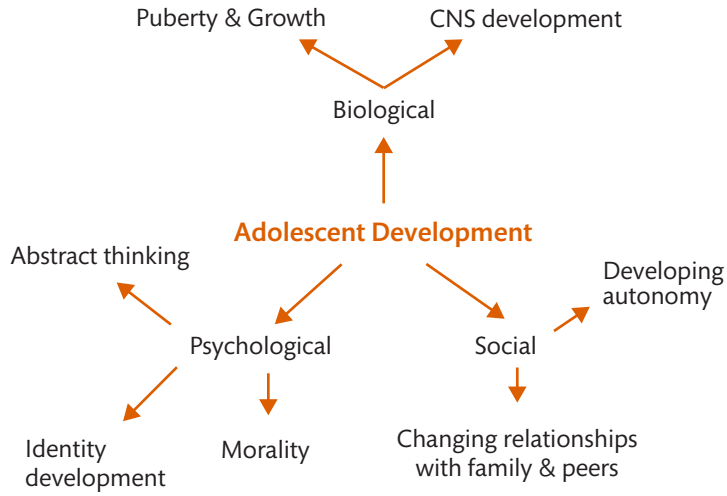


Figure 7.4.1 Biopsychosocial development of adolescence.

biopsychosocial development, (summarized in Table 7.4.1), and the potential complex interplay is invaluable for all clinical interactions between healthcare professionals and young people. The bidirectional relationship between 'normal' development and having an endocrine condition is significant.

Biological Development

Growth and puberty are covered in Chapter 7.1, but there are also changes in body composition beyond the achievement of adult height in late puberty with considerable sexual dimorphism [4–6].

Total body fat increases during puberty, with a steady increase in females and a slower increase in males due to the simultaneous rise in fat-free mass. This can be perceived as weight gain leading to impact on self-esteem and body image. Lean body mass (fat-free mass) increases to age 15 years in females but to age 18 years in males, with the maximal rate of change age between 12 and 15 years [7, 8]. The endocrine role of adipose tissue through release of adipocytokines during adolescence and young adulthood has potential implications for the development of future health problems such as cardiovascular disease, obesity, and diabetes; metabolic programming impacts on energy expenditure and insulin resistance [9–12].

Bone mineral content also increases, with 40% of bone density accrued during adolescence [13], with peak bone mass typically occurring in the early 20s in females and late 20s in males [14].

Muscle strength and mass increase through adolescence, most significantly in males due to the effect of androgens. This accrual peaks during the growth spurt in males and during menarche in females. By age 17 years, muscle mass is two-fold higher in males compared with females [15].

Other biological changes include:

- Blood pressure increases in puberty, more rapidly in males than females [16].
- Lipids -Total cholesterol and LDL cholesterol decrease in both males and females as puberty progresses; HDL is lower and triglycerides are higher in males by the end of puberty when compared to both males in earlier pubertal stages and females [17].
- Haemoglobin and haematocrit increase in males due to androgens, and can decrease in females due to menstruation [18].

Psychosocial Impact of Altered Biological Development

Pubertal Development

When compared to previous generations, girls now have menarche at a younger age [19]. Early pubertal maturation can have a detrimental impact on mental health [20]. Girls entering puberty at a younger age have been shown to have elevated rates of depression and these rates remain stable over the pubertal period. Earlier onset of puberty in girls is associated with earlier age of first sexual encounter, teenage pregnancy as well as the development of mood disorders [21–23].

Early-maturing males experience an initial benefit of entering puberty before their peers, conferring social benefits of increased muscle mass and older appearance. However over time rates of depression also increase in this group, potentially as a result of earlier introduction to risk taking behaviours, and association with older acquaintances [20].

Delayed or absent puberty can also have dramatic impact on psychosocial and emotional well-being [24–28]. Victimization and bullying have been reported. Young men diagnosed with congenital hypogonadotropic hypogonadism and Kallmann Syndrome have increased rates of anxiety and depression compared with their peer group [29–32]. Some young people become socially isolated as a result. Self-esteem and social outcomes do improve with the timely introduction of treatment at a physiologically appropriate age, as shown in both Kallmann syndrome and Turner syndrome [32, 33]. However, if the diagnosis is made late there can be ongoing sequelae for the young person, lasting into adult life. Psychological counselling is a key aspect of managing hypogonadism in adolescence.

Short Stature

Children and young people with short stature are reported to be at risk of social isolation, bullying and stigmatization, and tend to have low self-esteem and be underestimated by their parents, peers, and teachers [34, 35]. Parents also experience heightened anxiety and seem to have more awareness of potential impact on future functional limitations [36, 37], though levels of concern seem to be similar for families of children with other long-term conditions and may be more related to illness burden than height itself [38, 39].

Table 7.4.1 Biopsychosocial phases of adolescence

	Early 10–13 yrs	Mid 14–16 yrs	Late 17–19 yrs
Biological/ physical	**Girls** Breast bud and pubic hair development (Tanner stage II); initiation of growth spurt **Boys** Testicular enlargement; beginning of genital growth (Stage II)	**Girls** Mid to late puberty (Stage IV–V); completion of growth; menarche (Stage IV event); development of female body shape with fat deposition **Boys** Mid puberty (Stages III & IV); spermarche & nocturnal emissions; voice breaking; initiation of growth spurt (Stage III–IV)	**Boys** Completion of pubertal development (Stage V); continued androgenic effects on muscle bulk and body hair **Boys & Girls** Increase in bone mass
Psychological/ cognitive	Thinking remains concrete but with development of early moral concepts Progression of sexual identity development Sexual orientation Reassessment and restructuring of body image in face of rapid growth	Emergence of abstract thinking although ability to imagine future applies to others rather than self (self-seen as 'bullet-proof') Growing verbal abilities; adaptation to increasing educational demands conventional morality (identification of law with morality) Development of fervently held ideology (religious/political)	Complex abstract thinking Post-conventional morality (ability to recognise difference between law and morality) Increased impulse control Further completion of personal (and sexual) identity Further development or rejection of ideology and religion
Social	Realization of differences from parents Beginning of strong peer identification Early exploratory behaviours	Emotional separation from parents Strong peer group identification Exploratory/risk behaviours Heterosexual peer interests develop Early notions of vocational future	Development of social autonomy Development of intimate relationships Development of vocational capability

Adapted with permission from Pinquart M. Body image of children and adolescents with chronic illness: a meta-analytic comparison with healthy peers. *Body Image* 2013;10(2):141–8. Copyright © 2012 Elsevier Ltd. (Ref 40).

Psychosocial Development

Identity

The concept of 'self' evolves during adolescence, with the young person developing their own identity, and becoming autonomous in their thinking and acting. Part of this process is an evolving separation from the parents/main caregivers, with less reliance on their opinions and development of their own. Young people become more aware of how others see them, and more or less comfortable with who they are and what their life looks like from the outside. This includes the development of their sexual identity, along with moral/religious identity and academic/vocational identity. Self-esteem and self-efficacy form part of this process—the judgement or evaluation of one's ability and the belief in one's ability to achieve a particular task/goal. Where on the continuum of development an individual sits will hugely impact on their ability and confidence in managing their long-term condition, and this therefore needs to be taken into account during any clinical interactions. For example, the way a young person feels about their body image and how it is changing during adolescence will impact on their consultations; young people with long-term conditions have been shown to have lower self-esteem when it comes to body image than their counterparts without any underlying medical conditions [40, 41].

Thinking

The mode of thinking evolves during adolescence, with concrete thinking early in the process (rigid ideas with no sight of potential future effects of current actions), and the development of abstract thinking and logical thought process over time (the ability to link current actions with future consequences). It takes time for the 'bullet-proof' effect to progress to the complex abstract thinking needed to understand the long-term consequences on their future and their health of actions taken now. By late adolescence (age 17–20 years) young people are starting to understand these potential consequences, so the approach of healthcare professionals needs to adjust depending on the stage of development of a young person's thought processes.

Social Development

As young people develop biologically and psychologically, they are also adapting to a changing social environment. This involves changes in the way they interact with family, peers, and potential partners, with exposure to new behaviours and responsibilities. There can be significant cultural differences in how these changes evolve, depending on geography and ethnic background, with additional impact of socioeconomic status.

- **Relationships and support frameworks**

Throughout adolescence, there is an expected change in the interaction with parents/guardians, moving from a reliance on their support and guidance, to an active rejection of their input in favour of peer support. Ultimately there is re-acceptance of parental support in collaboration with peer support and the input of partners by late adolescence. With a long-term condition in the equation, the parental support relationship can be more complex: it is often relied upon, having been essential from childhood, potentially preventing young people from developing independence; for parents it can be hard to relinquish that control.

- **Exploratory and risk-taking behaviours**

During adolescence it is normal to explore behaviours that healthcare professionals consider risky. Substance use, including smoking, drugs, and alcohol, will commonly be tried for the first time during adolescence. Ninety-five percent of adult smokers have usually started by age 25 years [42], which indicates that health choices made during adolescence translate into longer term habits and adverse health outcomes [43]. The Key Data on Adolescence for 2017 [42] revealed that 5% of 15 year olds are regular smokers. Interestingly, the number of 15 year olds regularly using drugs has halved between 2001 and 2015, and only 9% of school pupils aged 11–15 years had consumed alcohol in the week preceding the survey, the lowest rate since the 1980s. This indicates that education and awareness raising in young people is having an impact. In consultations with young people with endocrine conditions, consideration should be given to whether these exploratory behaviours are occurring, potentially at a higher rate than in healthy peers [44], and that such behaviours may have more serious consequences for some young people, for example, childhood cancer survivors. There, therefore, needs to be a non-judgemental and open approach to discussing these issues. This is also a real opportunity to consider health promotion elements of consultations, as these behaviours often translate into adult life.

- Sexual and reproductive health

The average age of first sexual intercourse is 16 years of age, which falls in the middle of this adolescent development phase. There is therefore a duty to consider sexual health issues in consultations with young people in an endocrine clinic. Young people may have questions regarding sex and intimacy that they do not have the opportunity to discuss with other healthcare providers; indeed many young people will see the endocrine clinic as their main healthcare point of contact and will not attend their GP much, if at all. There is an opportunity to address safe sex, sexually transmitted infections (STIs) and contraception. Two thirds of new diagnoses with chlamydia happen in those under 25 years old. Teenage pregnancy remains a real issue, though rates have been consistently falling from the 1990s onwards, with the most recent reported under 18 pregnancy rate in 2015 in the UK being 20.8 per 1000 young women [42]. This remains higher than in other European countries. Some endocrine conditions may generate questions around ability to have sex, fertility, and whether there is a need for contraception—this needs to be openly addressed and questions should be welcomed.

- Education and vocation

Attainment of educational goals and progression to vocation are normal aspects of adolescent development. This can be affected by the presence of a long-term condition: attendance at school/college may be interrupted by health issues/hospital attendances. It has been shown that the presence of a long-term condition can lead to less educational attainment, less preparation for the move into the workplace, and less likelihood of achieving financial independence from parents [45, 46]. This is most marked in those with a significant disability, such as some survivors of childhood cancer [47].

- Concordance/adherence

Young people may not reliably take their medications, and if catastrophic consequences do not occur, it is difficult to convince young people to take their medications solely on the premise that bad things may happen. This relates most often to the concrete thinking of early adolescence, having yet to progress to abstract thought processes and an understanding of consequence. Other factors include lack of practice at looking after their own medications, differing priorities and psychological problems. This is a challenge in some endocrine disorders, the prime example being glucocorticoid insufficiency, where healthcare providers must ensure safety education highlights that glucocorticoid replacement is vital for a healthy life. Concordance, or the establishment of a therapeutic alliance between young person and healthcare provider, is the key to improving adherence with a long-term medication regimen for a long-term condition. The emphasis is on a shared approach to the condition and its treatment, rather than a judgemental, paternalistic approach, which is likely to be ineffective. Where there are cognitive impairments limiting adherence to treatment regimens (e.g. in some congenital/genetic disorders) or after brain radiotherapy, there needs to be collaboration with families and caregivers to develop practical solutions to the problem.

Common Physical Health Concerns in Young People

There are some common physical health concerns that present during adolescence. It is useful for healthcare professionals interacting with this age group to have an understanding of these, and where to signpost young people, as this may be the main focus for the young person in a consultation. Figure 7.4.2 illustrates common health concerns affecting young people [48].

- Skin

Acne is a problem for 85% of adolescents, with 10–20% affected by moderate to severe acne [49]. It usually begins with the onset of puberty. It can have significant psychosocial consequences for a young person, affecting their self-esteem and body image, and impacting on social interactions and interpersonal relationships. It can be treated either topically or systemically depending on severity. Endocrine disorders potentially aggravating acne during adolescence include PCOS and congenital adrenal hyperplasia due to increase in circulating androgens.

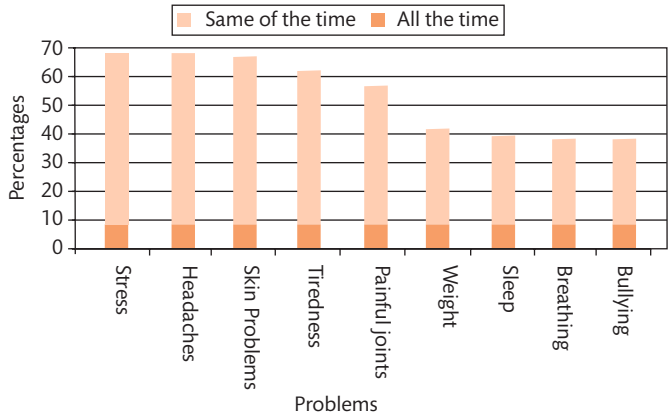


Figure 7.4.2 Self-reported health concerns among 13- to 15-year-olds.

- Sleep and fatigue

Sleep disorders can be both a cause of problems and the result of other issues in adolescence. The normal circadian rhythm slows during adolescence, meaning that young people are awake for longer, and it can be normal to be unable to sleep until after 11 p.m. [50]. Between 8 and 10 hours are needed per night for optimal functioning during adolescence, but many do not achieve this. It is also more likely that sleep patterns will vary day to day, which can disrupt sleep quality. Treatable sleep disorders can also occur, including insomnia, narcolepsy, sleep apnoea, and restless legs.

Fatigue is also a common problem in adolescence, and could be considered normal for many young people. It is defined as abnormal exhaustion after normal activities. There are increasing demands during adolescence as a result of physical changes, growth, increasing social and academic demands, all of which are compounded by reduced sleep time. Most young people will still manage to engage in usual activities despite experiencing fatigue. But for the minority this fatigue is significant and leads to reduced school attendance, impaired academic performance, reduced social interaction and impacts on relationships. When fatigue is persistent and impacting on function over a period of greater than 3 months, consideration should be given to investigation for underlying medical causes, psychological causes, and also the possibility of chronic fatigue syndrome. CFS is reported to occur in 0.2–0.6% of young people in the UK, a much higher rate than in the adult population [51]. CFS is the most common cause of school absence in the UK.

- Musculoskeletal issues and pain

A quarter of children are believed to suffer from recurrent or chronic pain, most commonly headache, back pain, or abdominal pain. The proportion affected is believed to increase during adolescence [52]. Sometimes these physical complaints can signify underlying psychological distress. Headache is reported in 20–25% of young people on at least a weekly basis [52]. Common causes are tension headache, migraine, ENT or dental causes, or dietary causes such as caffeine related. Back pain is common and nonspecific, often related to growth spurts and settling thereafter (e.g. Scheuermann's disease). Joint problems, sports injuries, and trauma are also common presentations.

- Breathing problems and asthma

A study of 12–14 year olds has shown that one third have experienced wheeze in the preceding year, with one-fifth having a diagnosis of asthma [53]. Most experience remission of symptoms by the end of puberty, but there is evidence that it is under-recognized during the teenage years.

- Overweight and underweight

One in five children aged 11–15 years is obese in England. It is now the commonest disorder in childhood and adolescence. Teenagers consume 8 times the recommended daily sugar allowance, and exercise levels are decreasing—by age 13–15 years, only 19% of boys and 7% of girls achieve 1 hour of exercise per day, and only 1 in 12 young people aged 11-18 years eat 5 portions of fruit and vegetables per day [42]. Some young endocrine clinic attenders have particular issues with appetite and weight gain, such as those with hypothalamic disorders. The sequelae of obesity in adolescence are multifactorial and include psychosocial issues, development of type 2 diabetes and increased cardiovascular risk, asthma, musculoskeletal problems, and liver disease.

At the opposite extreme, eating disorders including anorexia nervosa, bulimia nervosa, and binge eating can also present in adolescence. Factors influencing their development include biological, psychological, and cultural factors. Being underweight during adolescence can impact on normal endocrine development with poor pubertal progression or cessation of menses, poor growth, and poor bone development.

- Self-harm and mental health issues

75% of mental health issues start before 24 years of age, with 10% of young people experiencing a mental health disorder at any one time [54]. A quarter of young women aged 16–24 years exhibit symptoms of anxiety or depression [54]. There may be concurrent issues that point towards mental health problems, such as trouble at school, violence, and substance use. Long-term conditions increase the likelihood of mood disorder in young people as well as in adults, including endocrine conditions [55].

Consultations with Young People and Psychosocial Screening

To meet the needs of young people, as with patients of any age, the healthcare professional requires good communication skills. There are however some differences, which are important to emphasize. Confidentiality can be a concern for young people, particularly noted in older teenagers and young women [56]. It is important to explain the right to confidentiality including its limitations, rather than assuming it is understood. Encouragement of lone consulting for all or part of the consultation will also benefit the therapeutic alliance between healthcare professional and young person, hopefully increasing their confidence in negotiating consultations [57, 58]. To ensure that all aspects of a young person's life are covered, a psychosocial screen is useful, and one known as 'HEEADSSS' provides a useful framework (Table 7.4.2). It is important that the young person is the focus of the consultation, and while ideally this is done with the young person on their own, if accompanied the questions should be directed to them rather than their parents/caregivers [59, 60].

Developmentally Appropriate Healthcare

Due to the complexity of this period of life, health services must support the young person to engage with endocrine care; appropriate care in this age group is now called 'developmentally appropriate healthcare'. A recent review of the literature and a qualitative survey have identified the following themes that are required to deliver such care [61, 62]:

(i) Biopsychosocial development and holistic care
(ii) Acknowledgement of young people as a distinct group
(iii) Adjustment of care as the young person develops
(iv) Empowerment of the young person by prioritising health education and health promotion
(v) Interdisciplinary and interorganizational work.

Table 7.4.2 HEEADSSS communication framework

Theme	Relevance	Opportunity
Home	Identifying whether a young person feels happy and safe at home and has a supportive family who they can communicate with is a protective factor	Identifying safeguarding issues
Education & Employment	Young people with endocrine conditions may have lower educational and vocational attainment for any number of intrinsic and extrinsic reasons Identifying whether a young person is engaged with education or training and has a plan for employment when they leave is a protective factor	Assessing and addressing the impact of having an endocrine condition on education, training, and employment Assessing and addressing vocational readiness including aspiration of the young person; educational achievement; prior work experience; knowledge of resources, responsibilities, and rights; psychological aspects including self-esteem; expectations of the young person, their family, their teachers Addressing information needs of school, college, university, employment agencies and work to support the young person Discussing with the young person about disclosure to appropriate people about their condition to support education and vocation Offering neuropsychology for some young people, such as those treated for brain tumours, to guide them towards their strengths
Exercise & Eating	Young people with endocrine conditions are at risk of having or developing obesity which is associated with low self-esteem and potentially longer-term consequences such as diabetes, cardiovascular disease, reduced fertility, and sleep apnoea Young people who are childhood cancer survivors and have had treatment with the chemotherapy agent anthracyclines need advice about exercise such as weightlifting	Employing the 5As (Ask, Advise, Agree, Assist, Arrange) may support the young people becoming more active and eating more healthily
Activities (Including Peers)	Young people with an endocrine condition may have a disrupted peer network due to intrinsic or extrinsic factors Identifying young people with a supportive peer network and involvement in group activities is a protective factor	Enquiring about teasing and bullying may identify an area of risk and a young person in need of support Exploring potential for peer interaction through young person support groups, youth groups or with the assistance of a youth worker Enquiring about use of public transport and driving to support independence
Drugs (Smoking, Alcohol, Other Substances)	Young people with endocrine conditions are as likely as their healthy peers to explore the use of drugs Some young people with endocrine conditions are at risk of cardiovascular and/or respiratory disease, in particular childhood cancer survivors Young people on hydrocortisone or desmopressin who drink alcohol may increase their risk of adrenal crisis or hyponatraemia, respectively	Enquiring whether family or peers smoke, drink alcohol or use drugs is a recognized risk even if the young person denies use Employing the 5As (Ask, Advise, Agree, Assist, Arrange) may support the young people changing their drug use Educating young people on hydrocortisone or desmopressin about keeping themselves safe when drinking alcohol
Sexuality	Young people with endocrine conditions are as likely as their healthy peers to become sexually active in adolescence and be at risk of unwanted pregnancy or sexually transmitted infections Young people with endocrine conditions may have questions about sexuality, ability to have sex and fertility. Young people with PCOS on anti-androgens should be advised about avoiding pregnancy Young people with endocrine conditions caused by genetic disorders or associated with cardiac concerns should be advised about prepregnancy counselling	Identifying gender concerns Educating young people about sexual health and contraception and signposting to relevant services Discussing concerns about ability to have sex, in particular young people with disorders of sex development, and referring for further assessment if appropriate Discussing implications of their condition on fertility and referring for further assessment if appropriate Educating young people about the importance of pre-pregnancy genetic counselling and cardiac assessment and avoiding pregnancy in those on antiandrogens
Suicide, Depression, & Self-Image	Mental health issues are common in adolescence Protective factors include: having an adult to talk to when stressed, a peer support network and coping skills, as well as positive self-esteem and image	Asking them to rate their general happiness to encourage discussion Asking about self-harm and suicide Referring to appropriate services if required
Sleep	Sleep problems in adolescence are common and can be a sign of anxiety and depression	Enquiring about sleep and discussing sleep hygiene
Safety	Safety about driving alone or with friends Safety about internet use Safety about hydrocortisone and desmopressin in general and when away from home	Enquiring about safety with regards to driving and internet use Educating about medic alert jewellery, sick day rules, emergency hydrocortisone injection and keeping safe with diabetes insipidus Discussing with the young person about disclosure to appropriate people (friends, colleagues) about their condition and offering to provide them with education Empowering them to insist on optimal care in healthcare settings

This approach to care should be adopted by both children's and adult services. It is relevant not only to those undergoing transition and transfer to adult services but also for young people presenting for the first time to an endocrine service during adolescence and young adulthood. To assist healthcare organizations in provision of developmentally appropriate healthcare, a toolkit is available [63].

Training Healthcare Providers

Delivering developmentally appropriate healthcare requires trained, competent healthcare providers. Whilst this is logical, and is a feature of all of the national and international guidance available in this area, it is recognized that training in this area is suboptimal, and therefore presents a barrier to implementing optimal care. A survey of specialist trainees in Diabetes & Endocrinology in the UK revealed that they felt under-prepared to deal with young people in clinics, with 65% rating their training in adolescent and young adult health as minimal or non-existent (76% in endocrinology, 42% in diabetes) [64].

Transition and Transfer

Transition is defined as the purposeful, planned process that addresses the medical, psychosocial, and educational needs of adolescents and young adults with long term conditions, as they move from child-centred to adult-oriented healthcare systems [65]. This is not to be confused with transfer—which is the isolated event when the healthcare of a young person moves from services for children to services for adults.

The aims of transition are to:

- Provide high quality, coordinated, uninterrupted healthcare that is patient-centred, age and developmentally appropriate, culturally competent, flexible, responsive, and comprehensive with respect to all persons involved
- Promote skills in communication, decision-making, assertiveness and self-care, self-determination, and self-advocacy
- Enhance the young person's sense of control and interdependence
- Provide support for the parent(s)/guardian(s) of the young person during this process
- Maximize lifelong functioning and potential

There is a limited but emerging evidence base regarding transition of young people. A Cochrane review conducted in 2016 [66] revealed only four randomized controlled trials involving 238 participants. In the UK in 2016 NICE reviewed the evidence and published guidance on 'Transition from children's to adult services for young people using health or social care services' [67]; this makes recommendations on the overarching principles, how to plan and support transition and transfer, and on supporting infrastructure.

A recent trial [68] recruited 374 young people with long term conditions (diabetes, autistic spectrum disorder and cerebral palsy), retaining 274, with 149 crossing into adult services. The aims of the study were to address the needs and wishes of young people, to identify the features of care that were most effective and efficient, and to determine how transitional care should be designed, commissioned, and delivered. This trial found that, although the present guidance on the care of young people suggested a number of key requirements of a transition service, there were three factors that made a real impact on the young people: meeting the adult team before transfer, the development of self-efficacy/self-management skills, and being allowed to have parents/guardians involved if that suited them.

Due to the limited evidence base a Delphi study of international experts identified key elements that should be part of a transition program (**Box 7.4.1**) and what indicators could be used to assess its success (**Box 7.4.2**).

There is evidence, although limited, that transition in endocrinology could be improved. A single centre study identified that 25% of young people transferred from paediatric care were not seen within 12 months of transfer by adult services [69, 70]. In congenital adrenal hyperplasia, there is evidence that 50% of young people with CAH were lost to follow up during the transition period, and are not receiving specialist care despite the risks of long-term morbidity [61, 62]. One study found that patients with CAH who did transition successfully had a better quality of life [71]. In Turner syndrome, one survey of adult women with the condition found

Box 7.4.1 Key elements for successful transition: Delphi study

1. Co-ordination (timing of transfer, communication, and follow-up) between paediatric and adult healthcare professionals
2. Planning transition from an early age, possibly from the transition to high school aged approximately 12 years old (and at least 1 year before the planned date of transfer to adult services)
3. Setting the scene with young people and their families with regards to the aims for evolving self-management
4. Involving young people in the planning of transition—taking into account their thoughts and preferences
5. If developmentally appropriate, seeing the adolescent alone at least for part of the consultation (and talking about leading up to this in preparation)
6. Identifying an adult care provider willing to take on the young person before transfer
7. Tailoring the transition plan to the needs of the young person and family
8. Identifying someone within the team who will play the role of transition coordinator/named contact
9. Providing a written health summary and biopsychosocial profile summary to the young person and the adult care provider before transfer
10. Having a written transition protocol/plan that is available to young people, parents, and providers
11. Making sure that at least one appointment with the adult provider after transfer is fixed in anticipation
12. Ensuring healthcare professionals involved in these clinics have sufficient knowledge and skills in adolescent health
13. Parents and caregivers should be included in the process of transition
14. Keeping the primary care provider (GP, family doctor, nurse practitioner/advanced practice nurse) informed of the transition process
15. Putting in place mechanisms/resources to contact young people lost to follow-up
16. Discussing the differences between paediatric and adult care with the young person and their family
17. Discussing risk behaviours and healthy lifestyles and their influence on health

Box 7.4.2 Key indicators of successful transition: Delphi study

1. Young person not lost to follow-up
2. Attending scheduled visits in adult care (no missed consultations unless previously cancelled and rescheduled)
3. Young person building a trusting relationship with adult provider
4. Continuing attention to self-management
5. Young person's first visit in adult care no later than 3–6 months after transfer
6. Number of emergency department/unscheduled care visits for regular care in the past year (avoidable if routine medical care had been occurring)
7. Young person and family satisfaction with transfer of care
8. Maintenance/improvement of standard for disease control evaluation (such as glycosylated haemoglobin for diabetes).

that only 63.2% had regular follow-up and many were not on oestrogen replacement [72].

Young People with Endocrine Conditions Undergoing Transition

The approach to a young person with an endocrine condition depends on the condition and the timing of the presentation and includes the following aims:

- To ensure the young person who presented earlier in childhood has all the information previously shared with their parents about their diagnosis, previous investigations, and management
- To address the priorities of the young person presenting later in adolescence, particularly in respect of delayed puberty and short stature
- To revisit the diagnosis or degree of hormone deficiency/excess
- To discuss investigations and management with the young person regarding cardiometabolic, bone and reproductive health, as well as quality of life
- To raise awareness of maintaining safety for those on hydrocortisone and desmopressin and recognizing ill health and what to do and when to seek help
- To identify the level of endocrine care required in adult life and whether care can be provided by the family doctor, for example, in congenital hypothyroidism, or whether transfer to adult services are required

In addition, more generic approaches are also important:

- To encourage increasing independence in healthcare including organizing medication and prescriptions, arranging appointments at times suitable to them, coming in for part or all the consultation independently of their parents and contacting the hospital in between appointments to seek information and advice
- To address barriers to engagement with healthcare including adherence
- To discuss useful resources to get health information and support including patient support groups
- To address concerns about venepuncture or injections, particularly around transfer to adult services
- To highlight benefits and ways of disclosure about condition to friends, teachers, and employers
- Through psychosocial screening (see earlier) to address broader issues such as psychological, social, and educational/vocational issues, and health behaviours by identifying risk and protective factors
- To deliver health promotion, including discussion about sexual health and contraception

Ongoing Management of Congenital and Acquired Conditions Diagnosed in Childhood

Who Needs Re-Evaluation?

The completion of linear growth is an opportunity to review current endocrine status, and adequacy of endocrine replacement, and for some to revisit the diagnosis, and screen for health conditions more common in adult life. Agreement would be required between paediatric and adult endocrine teams and the young person and their carers where it is most appropriate for these evaluations to take place.

This is the case for:

- patients where the underlying diagnosis made in childhood was unclear, and establishing that the endocrine issues persist is important
- patients with hypopituitarism, particularly isolated hormone deficiencies which may recover, for example, growth hormone deficiency (see next) and hypogonadotropic hypogonadism (10–20% exhibit spontaneous recovery)
- patients at risk of evolving hypopituitarism: post radiotherapy and those with congenital hypopituitarism associated with structural abnormalities in the hypothalamic–pituitary area (e.g. ectopic posterior pituitary), midline brain, and optic nerve abnormalities and genetic defects (e.g. in the *GH-1* or *POU1F1* genes)
- patients with congenital adrenal hyperplasia; a short synacthen test to assess cortisol status may assist with those who are poorly compliant with glucocorticoids, by providing further information to support discussion with the young person, or in those with non-classical CAH as alternative management could be considered at this age
- patients who require assessment of reproductive health; for example, testicular ultrasound for testicular adrenal rest tumours in boys with CAH and assessment by gynaecology for some girls with CAH
- patients who require comprehensive health screening, for example, in young people with Turner syndrome (requirement for cardiovascular and hearing assessment) and in childhood cancer survivors (monitoring for non-endocrine late effects)

Who Needs to be Part of the Wider Multidisciplinary Team Involved in Transition in Endocrinology?

Involvement of the multidisciplinary team is important at this time; paediatric and adult endocrinologists and endocrine nurses need to liaise with each other. Young people should have access to the following people (dependent on the condition):

- Psychologists (although not always available within the service) can provide much needed support for all young people with endocrine conditions as they make the transition from childhood to adulthood
- Clinical geneticist should be a key member of the transition team from several perspectives; to provide information about those with a clear or potential genetic diagnosis that may have previously been shared with parents, to explain the potential implications of genetic mutations for future offspring (as in congenital adrenal hyperplasia)
- Urologist/gynaecologist is essential for many patients with a disorder of sex development (DSD) who may require a physical assessment and a discussion about sexual function and future pregnancy
- A fertility specialist is essential for those focussed on understanding more about options for fertility at this time, particularly if there may be opportunities for fertility preservation; for example in females with Turner syndrome with a menstrual cycle, female childhood cancer survivors at risk of premature ovarian insufficiency, males with Klinefelter syndrome, males with CAH who have testicular adrenal rest tumours
- A cardiologist is important for females with Turner syndrome and cancer survivors previously treated with anthracyclines or chest irradiation
- Neurorehabilitation specialists, neurologist and neurosurgeons for young people are required for those with a past history of a brain tumour who require ongoing surveillance or have neurological sequelae from the tumour or its treatment
- ENT/audiology are key for groups of young people with endocrine conditions associated with reduced hearing which in turn reduces quality of life; for example, Turner syndrome, Charge syndrome, and some childhood cancer survivors

The focus of care in late adolescence is related to quality of life, and cardiovascular, bone and reproductive health from the perspective of monitoring and optimizing hormone replacement. In adolescents after the achievement of adult height, there is a change in the management approach in some conditions from that in paediatrics to reflect adult practice, for example, the management of sex steroid deficiency or glucocorticoid replacement in CAH in adolescents who have completed puberty. However, adolescence and young adulthood is seen as a distinct treatment period for young people with growth hormone deficiency (GHD).

Childhood Onset Growth Hormone Deficiency in Adolescence and Young Adulthood

Following on from the established role of GH replacement in children with GHD starting in the 1960s and the availability of recombinant GH in the 1990s, the extension of indications for GH replacement in GHD were considered [73]. The term 'adult growth hormone deficiency syndrome' was used to describe the clinical entity that included decreased lean body mass and bone mineral density, increased visceral adiposity, abnormal lipid profile, decreased muscle strength and exercise endurance, diminished quality of life, and the potential for increased morbidity and mortality [74]. In addition, evidence from clinical trials showed that GH replacement improved aspects of the syndrome [75]. It was only logical that potentially a lifelong role for GH replacement should be explored in patients with childhood onset (CO) GHD on achieving adult height, to address developmental aspects described in CO GHD and/or metabolic aspects described in AO GHD [75].

One key study [76] compared body composition in 92 CO GHD patients with 35 age-matched untreated AO GHD patients (with the majority of both groups having other hormone deficits); the mean age of those with severe GHD was 21 years and they had been off GH replacement for a mean of 1.6 years. Compared with AO GHD, CO GHD patients were shorter and had a lower body mass index. Corrected for height, CO GHD compared with AO GHD patients had a lower bone mineral content (BMC), lean body mass (LBM) and fat mass (FM) [76]. The closer the CO GHD patients were to achieving their genetic target height, the higher the BMC and LBM (76). Therefore, there was a significant maturational deficit (16–20% less compared with AO GHD patients) in somatic development in CO GHD patients treated with GH replacement during childhood [76].

In 2003 there was a consensus meeting focussing on the management of patients treated with GH replacement for GHD during childhood [77]. Based on the available evidence at that time it was recommended that ongoing GH replacement should be offered to all adolescents at the end of growth who remain GHD. The main rationale for continuing GH replacement included optimizing body composition and peak bone mass with the potential outcome of reducing fractures in later life. With this 'new' indication came the challenge of defining the biochemical diagnosis of GHD in adolescence, reassessment strategy, the approach to initiation and monitoring of treatment.

Bone and Body Composition

Although there is a clear maturational deficit in both bone mass and lean body mass, there is little evidence that this low bone mass contributes to an increase in fractures. The evidence suggests an increased fracture risk in AO hypopituitarism, but is insufficient for isolated CO or AO GHD (reviewed in [78]). There is no evidence that isolated CO GHD or severe GH resistance (e.g. Laron syndrome) causes an increased fracture risk in children or adults. Only one study in GH-naive Russian adults with CO MPHD demonstrated an increase in non-traumatic fractures [79].

Overall, it is proposed that fracture risk is related to other hormone deficits with over/under replacement (e.g. sex steroid, glucocorticoids, or diabetes insipidus) or other non-endocrine factors rather than GHD.

One explanation for this is the skeleton has adapted both to relative short stature and low muscle mass. Using appropriate size-corrections, bone density is normal in children and adults with isolated GHD (reviewed in [78]). Cortical density, trabecular density and trabecular volume are normal when measured by peripheral quantitative computerized tomography (pQCT) and histomorphometry [80]. The only clear deficit affects cortical thickness (periosteal expansion). It has been proposed that untreated CO GHD in adolescence and young adulthood results in a reduction of muscle mass and force, which would ultimately have an impact on bone geometry, more specifically reducing cortical thickness, described as the mechanostat theory. In support of this theory, during

GH therapy muscle enlargement precedes and exceeds any gain in bone mass (reviewed in [78]).

After discontinuing GH replacement at adult height, several but not all studies have showed a significant reduced accrual of bone mass, decrease in lean mass (LM) (–8%) and increased fat mass (FM) (10–17%) in patients with CO GHD who were diagnosed with persistent GHD and had stopped GH replacement for at least 2 years (reviewed in [81]). In terms of muscle strength, previous studies have reported a lower maximum isotonic strength in young adults with CO GHD, as measured by hand grip force (reviewed in [81]).

Restarting GH replacement therapy resulted in a notable improvement in body composition, with LM increasing by 14% and FM being reduced by 7% over 2 years of therapy (reviewed in [81]). Some but not all studies also reported a positive effect on bone with increases in total body BMC and lumbar spine bone mineral density (BMD) after one or two years (reviewed in [81]).

Methods of assessing bone in these studies are key. A study using pQCT, which is uninfluenced by short stature, showed a reduction in cortical thickness after 6 months off GH replacement [82], while after 2 years on GH replacement a significant increase in cortical thickness was observed using digital X-ray radiogrammetry, another technique not influenced by short stature [83]. Assessment of bone quality, using more advanced non-invasive imaging tools, may provide a greater insight into the effects of GHD and GH replacement on bone [84].

In summary it is clear from the evidence that GH replacement in late adolescence and young adulthood increases LBM and reduces FM mass, and this in turn has a positive effect on bone geometry, although the benefit to long-term health and reduction in fractures has not been quantified.

Cardiovascular Risk

The potential that GH replacement may reduce cardiovascular morbidity and mortality in adults remains unproven and therefore an important area of research [85].

In adolescents and young adults with CO GHD some studies have reported an adverse lipid profile and also other indicators of increased cardiovascular risk at adult height despite conventional GH replacement [86–88], which deteriorates further on discontinuation [86–90] and improves on recommencement [87–89]. In contrast, some studies comparing discontinuation with continuation of GH replacement or recommencement have shown no difference in lipid profile changes or other indicators [86, 90–94]. In a database study, those patients with organic CO GHD that had been off GH for more than 2 years had a more adverse lipid profile with improvement in HDL cholesterol after 1 year on GH replacement [95].

Cardiac structure and function were assessed by echocardiography. At baseline, GHD subjects had a lower early-to-late mitral flow velocity ratio (E/A; marker of diastolic function) but a normal left ventricular (LV) mass index and ejection fraction compared with normal controls. Six months after GH replacement withdrawal, both LV mass index and E/A decreased, although remaining within the normal range. Six months after restarting GH replacement, cardiac parameters were brought back to the levels measured at study entry, with LV ejection fraction and E/A remaining lower than in normal controls [87]. Similar results were identified in older patients with CO GHD [96]. Two other studies have not identified any echocardiographic abnormalities [86, 92]. The clinical relevance of these cardiac changes is not clear. Reassuringly, continuation, discontinuation and recommencement of GH replacement have not been shown to have any effect on exercise capacity [92, 93].

Intima-media thickness (IMT) at the common carotid arteries has been reported to be both similar after 6 months off GH replacement [88] and increased after 3 years off GH replacement [94] in GHD patients compared with controls at baseline. Therefore, duration off GH replacement may be a factor.

Studies have demonstrated an increase in insulin sensitivity [88, 91, 97] with a reduction in fasting glucose [93] on discontinuation of GH replacement. Once GH replacement is restarted, insulin sensitivity decreases [88, 97], resulting in an increase in fasting glucose levels (93), but glucose tolerance was not impaired [86]. Importantly, there was no further deterioration in insulin sensitivity [86, 91] or fasting glucose with continued GH replacement [93].

In summary the results from studies of cardiovascular risk in adolescence are inconsistent; however there is some suggestion that with more prolonged discontinuation of GH replacement, there may be negative effects on cardiovascular risk, in particular lipid profile and intima media thickness, but only subtle effects on cardiac morphology and function Reassuringly, no clinically significant effects on insulin sensitivity or glucose homeostasis were observed.

Quality of Life

Quality of life is recognized as an area that improves with GH replacement in some adults with GHD, and is the main indication for treatment in some countries.

In the transition period, one study examined the effect of seamless continuation of GH replacement compared with discontinuation and found no effect on QoL at baseline or after 2 years off GH replacement [92]. In a study looking at recommencement of GH replacement, although QoL was significantly lower than in normal controls (females –0.4 vs. 1.2 SDS; males –0.7 vs. 1.1 SDS), no change in the overall score was identified with GH replacement despite significant improvement in individual parameters that were low at baseline, including sexual arousal and body shape [98]. No such effect on individual QoL parameters was confirmed by a similar study design [86]. However other studies have shown benefit. In one study of 22 subjects followed through 2 years discontinuation and 1 year recommencement of GH therapy, the number of psychological complaints and depression increased during the first 6 months of discontinuation, while anxiety decreased and QoL improved during the first 6 months that GH was recommenced [99]. Some differences were observed between those with MPHD and those with IGHD [99]. In a database study, a lower QoL was identified in those with organic GHD but not those with idiopathic GHD; the former group who had a higher BMI, had achieved less height gain and had been off GH for longer (more than 2 years). However, after 1 year on GH replacement, there was improvement in QoL in both organic and idiopathic GHD groups [95]. In a slightly older cohort of patients (25–29 year olds) with CO GHD having routine clinical care recorded in the KIMS database, QoL as measured by

adult growth hormone deficiency assessment (AGHDA) differed between different aetiologies but demonstrated similar levels of improvement with GH replacement [100]. It is possible that studies with less heterogeneous groups selected for poor QoL would show greater improvements in QoL with GH.

Using different measures of QoL in a study exploring discontinuation in 40 GHD adolescents and young adults, no changes were found in the Nottingham Health Profile and Psychological General Well-Being Index but differences did occur in the Mood Adjective Check List and visual analogue assessment [101]. QoL tools have not been specifically developed for this age group, and there is a pressing need to have an appropriate well-validated tool for these patients [102].

Retesting GH Status at Adult Height

Retesting is essential as patients diagnosed as GHD in childhood will not necessarily be GHD in late adolescence and young adulthood, in part because of the different GH stimulation cut offs applied in adolescence compared with paediatric practice and also with apparent recovery in significant numbers of GHD patients. This has been attributed to long-term exposure to sex steroids potentially augmenting pituitary size and GH secretion, transient GHD and incorrect diagnosis due to issues with biochemical testing [103].

Likelihood of Persistent Severe GHD

Information about the extent of hypopituitarism, structure of the hypothalamic pituitary area and the underlying aetiology allows retesting of GH status at end of growth to be stratified into those that do not need retesting (very high likelihood of persistent GHD) and those with a high and low likelihood of persistent GHD (Table 7.4.3).

Number of additional pituitary hormone deficits can predict likelihood of persistent GHD. Data from the adult population has identified that three or four additional pituitary hormone deficits has specificity of 96% and 99% respectively for diagnosing severe GHD [104], and this has been confirmed in the adolescent population [105]. In addition, evidence of severe GHD on testing in childhood increases the likelihood of persistent severe GHD in adolescence and young adulthood [106, 107].

The presence of structural abnormalities in the hypothalamic-pituitary area also predicts likelihood of ongoing GHD. The majority of patients diagnosed with GHD in childhood and with no structural pituitary abnormality on MRI will have adequate GH secretion when retested (66–85% depending on the test and cut-off used) [105, 108, 109–112]. The presence of structural abnormalities, such as an ectopic posterior pituitary (EPP), increases the likelihood of persistent GHD; 60% had persistent GHD out of a cohort of 18 with ectopic posterior pituitary [113]. The higher the location of the EPP in relation to the pituitary stalk, the more likely will be the persistence of GHD; in the majority of patients with persistent severe GHD (defined as GH peak <5 μg/L), the posterior pituitary was located at the median eminence (93%), whereas in around 80% of subjects with a GH peak of at least 5μg/L at reassessment, it was visible along the stalk [110].

The aetiology of CO GHD spans genetic and acquired causes and while proven genetic causes usually have a high likelihood of persistence [114], acquired causes are associated with more variation dependent on the insult. GHD is a frequent consequence in childhood cancer survivors treated for tumours in the hypothalamic–pituitary area [115] and in brain tumour and nasopharyngeal carcinoma survivors exposed to HP axis radiotherapy ≥30 Gy [116–118]. GHD can also be diagnosed in childhood following cranial radiotherapy for haematological malignancies exposed to lower doses of 18–24 Gy or doses used for haematopoietic stem cell transplantation. However, recovery is more likely in this group [119]. In an analysis of 73 patients who had received cranial irradiation and were diagnosed with GHD in childhood, those treated with higher doses, as used in the treatment of brain tumours, 61% had persistent severe GHD (peak GH <3 mcg/L) compared with 33% following cranial radiation or total body irradiation used in the treatment of haematological malignancies [106].

Traumatic brain injury may also result in GHD; the incidence of hypopituitarism is reported over the broad range of 10–60% [120–122]. Recovery is also variably reported. In one study, 23 patients aged 16 to 25 years were followed up at 3 and 12 months after the event; at 12 months hormone deficits persisted in 30%, consisting of mainly GH and gonadotropin deficiencies [120].

Retesting Strategy Based on Likelihood of Persistent Severe GHD

The first consensus on retesting for GHD [77] and subsequent guidelines [123, 124] have proposed strategies based on likelihood (Figure 7.4.3).

For those patients with panhypopituitarism (defined as three or more additional hormone deficiencies) retesting was optional and in these patients GH replacement can be continued seamlessly if the young person agrees. Those with known transcription factor or genetic mutations also did not require retesting [123].

Table 7.4.3 Factors increasing likelihood of persistent GHD (* indicates very high likelihood)

Higher likelihood	Lower likelihood
• Severe GHD (<3mcg/L) on testing in childhood • Three additional pituitary hormone deficits* • Abnormal pituitary MRI including ectopic posterior pituitary • Proven genetic cause for GHD* • Tumour in the hypothalamic-pituitary area • High-dose cranial irradiation (>30Gy)	• Peak GH during provocative testing during childhood >3 mcg/L • Isolated GHD with normal pituitary MRI • Low dose irradiation (<30 Gy) • Traumatic brain injury

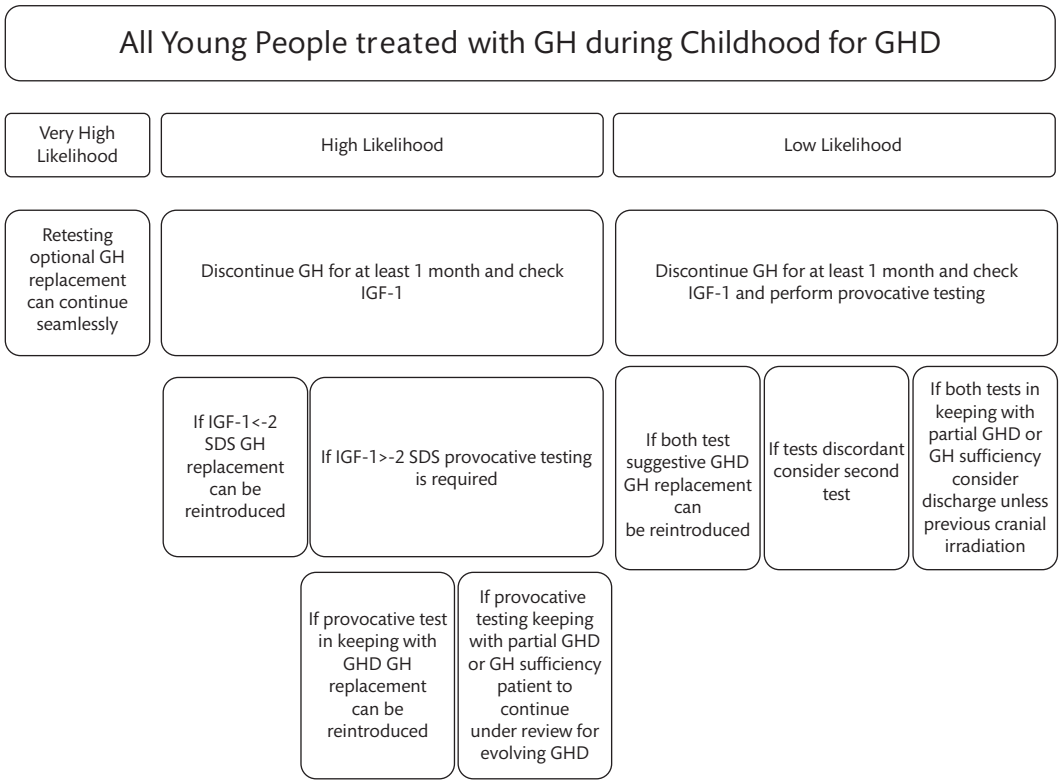


Figure 7.4.3 All young people treated with GH during childhood for GHD.

For those with a high likelihood of having persistent GHD, a measurement of serum IGF-1 concentration at least one month after finishing GH replacement should be done, progressing stimulation testing only if the IGF-1 is in the normal range. Patients with a low likelihood of persistent GHD, should have both serum IGF-1 estimation as well as a GH stimulation test.

Evidence Base for Retest Strategy

While an IGF-1level in general has low accuracy in diagnosing GHD, regardless of aetiology, an IGF-1 of less than –2SDS has a positive predictive value for severe GHD in adulthood of 96% [104]. This has been shown in studies in adolescents and young adults in the transition period with non-acquired GHD [125] and radiation-induced GHD [126].

If provocative testing is required only one provocative test is recommended by all the consensus statements in those with a high likelihood when IGF-1 is within the normal range, and in those with a low likelihood of persistent GHD [77, 123, 124]. However, guidelines from the Endocrine Society highlight that as idiopathic GHD in adults is very rare, more stringent criteria are proposed including the performance of two provocative tests, particularly if the IGF-1 is in the normal range [124].

The insulin tolerance test (ITT) has been validated for the diagnosis of persistent GHD in the transition period. There is a lack of agreement as to the peak GH cut-off in this age group, with all three consensus/guideline documents suggesting three different cut-off levels from the adult criterion of <3 mcg/L (124) to the evidence-based <6 mcg/L [123]. The latter was defined in a study of 26 patients (8 had isolated GHD and 18 had MPHD in childhood) which identified that a GH cut-off of 6.1 μg/L during an ITT had a sensitivity of 96% and specificity of 100% [127]. This was validated in a study of 79 patients divided by high and low likelihood of severe GHD, and found that a GH cut-off of 5.6 μg/L during an ITT had the highest accuracy for predicting persistent GHD with a sensitivity of 77%, specificity of 93%, and correct classification of 87% of patients [107]. More recently, a study looking at cut-offs in children and adolescents suggested that the optimal cut-off during an ITT was 5.1 μg/L, that during an arginine stimulation test was 6.5 μg/L, and that for clonidine was 6.8 μg/L [125]. The other test with supporting evidence in the transition period is the GH-releasing hormone (GHRH)—arginine test and a higher cut-off compared with adults (19 μg/L) had optimal sensitivity and specificity [128]. However, the GHRH-arginine test may be falsely normal in individuals with hypothalamic dysfunction. Although the ITT and GHRH-arginine test have been validated, GHRH is not always available and the ITT is contraindicated in some individuals and is not practical in many endocrine practices. The arginine test can be used but the response is dependent on BMI, so the arginine test should be limited to non-obese adolescents. Glucagon stimulation testing is a promising alternative for provocative testing in adults but has not been tested in the transition period.

Further Re-Evaluation of GH Status

Some groups may require further re-evaluation, regardless of the results of initial testing at the end of growth. Those at risk of evolving hypopituitarism, including GHD, are those treated with

radiotherapy [106] and patients with ectopic posterior pituitaries [110, 129]. In addition, patients with discordant results pose a particular challenge. For example, in one study 50% of patients (4 of 8) with IGHD and an ectopic posterior pituitary had discordant results with an adequate peak GH but with a low IGF-1 concentration [107]. Despite the normal GH peak, this cohort of patients should be monitored as they are at risk of development of severe GHD and evolving pituitary hormone deficiencies [110, 129].

Further evaluation using adult criteria around the time of assumed attainment of peak bone mass and attainment of full somatic development (~25 years) can be considered but is likely to be dependent on the severity of GHD and the local indication for ongoing GH replacement in adult life. Those with low likelihood of persistent GHD as described earlier, who retest as GHD at end of childhood, should be re-evaluated before a commitment is made to lifelong GH replacement as should those with discordant tests (normal IGF-1; GHD on provocative testing).

Growth Hormone Replacement in Adolescence and Young Adulthood

For those young people who test as being persistently GHD, there needs to be a discussion as to whether their preference is to continue GH replacement in the transition period with potential but unclear long-term benefits or have a holiday from daily injections, which may be preferable to young people for many reasons.

The optimal GH dose during the transition period is not clear. Studies have used GH doses that varied from 12.5 to 25 μg/kg/day (weight-based) to 200 μg/day (fixed dosing) [86, 98, 130, 131]. With respect to bone health and body composition, one study demonstrated no dose dependency [130, 131], whereas another found that a higher dose had a greater effect on spine BMD compared with a lower dose. Although similar effects were observed with LBM and FM at 6 months between the two doses, these were more likely to be sustained with the higher dose [86]. This study also demonstrated a greater improvement in LDL-C with the higher dose of GH replacement [86]. A dose effect in relation to quality of life has not been demonstrated [86, 98].

Some differences in dosing based on age have been recommended by The Endocrine Society Guidelines for Evaluation and Treatment of Adult Growth Hormone Deficiency [124]; patients <30 years of age may benefit from initial doses of 0.4–0.5 mg daily (as opposed to the initial doses of 0.2–0.3 mg daily for patients aged 30–60 years) and those transitioning from paediatric to adult replacement may need even higher doses. Females receiving oral (but not transdermal) oestrogen may need higher doses than other patients.

Doses should be titrated to normalize serum IGF-I concentrations for age and gender (0 to +2SDS). Despite these recommendations, a review of young people (aged 15–26 years) with childhood onset GHD on the KIMS database showed that their mean IGF-1 SD score was –1.36 (–1.83 SDS in females) after 1 year on a GH dose of 0.5 mg/day, suggesting inadequate dosing [132]. However, the relationship between IGF-1 level and biological endpoints is lacking [133]. As such, there is a lack of a biomarker for GH replacement in late adolescence and young adulthood [134].

Changes in body composition and bone mineral density are the most obvious candidates to be used as biomarkers and all the guideline documents recommend these parameters are assessed at baseline and during treatment. However, as previously described, there is little evidence of reduced bone mineral density, if correct methods are used such as volumetric BMD assessment or an adjustment to the areal BMD [78]. Finally, it is not clear how these measures would be used to alter management.

Other markers to assess could include cardiometabolic parameters including glucose homeostasis. These are most important to monitor in at risk groups such as adolescents and young adults with obesity or a family history of type 2 diabetes or at risk of diabetes or adverse lipid profile as a consequence of treatment, for example following total body irradiation or brain tumours. They are not required in all patients.

It is also recognized that adolescents may develop impaired quality of life and experience psychosocial problems related to GHD and/or the underlying chronic disease. There is a pressing need for validated instruments to assess QoL in this age group.

Holistic Approach to the Patient with Childhood Onset GHD

In keeping with the approach for all young people with long term conditions, it is important to focus on the wider aspects of growing up. Several studies in patients with childhood onset GHD compared with patients with AO GHD have identified reduced social outcomes including being less likely to live independently, living in partnership and having children, regardless of the underlying diagnosis [47]. Although some studies have shown similar educational and vocational outcomes to patients with AO GHD [47], others have shown lower educational level, lower income, higher risk of retirement compared with the general population and that this in turn impacts on mortality [135]. The endocrine team should routinely address these aspects with a view to offering support and encouraging independence as well as signposting.

Oestrogen Replacement in Adolescents

Girls with gonadotrophin deficiency or premature ovarian insufficiency (POI), including Turner Syndrome and some childhood cancer survivors, require oestrogen replacement throughout life. There are choices for the young person to consider, whether they are presenting in adolescence with delayed puberty or have completed pubertal induction and are considering their preferred oestrogen replacement in adult life. Guidelines for POI [136] and Turner syndrome [137] detailing the rationale for the most appropriate approach to oestrogen replacement have been published.

Girls with delayed puberty require an individualized approach to ensure that the priorities for achieving maturity or maximizing remaining growth are met. A shortened protocol for pubertal induction is generally successful, although available studies are too small to identify the reasons for unsatisfactory pubertal outcomes such as breast hypoplasia [138]. No data on long-term outcomes such as peak bone mass are available.

After pubertal induction, optimizing oestrogen replacement is important because oestrogen deficiency, particularly in the setting of POI, is associated with reduced life expectancy largely due to cardiovascular disease [136]. Oestrogen deficiency is also associated

with reduced bone mineral density (BMD) and an increased risk of fractures later in life [136, 137]. Despite lack of longitudinal outcome data, early initiation of oestrogen replacement is strongly recommended to control future risk of cardiovascular disease and optimize BMD; it should be continued at least until the average age of menopause. Lifestyle measures, such as regular weight-bearing exercise, avoiding smoking, maintaining a healthy body weight as well as adequate dietary calcium and vitamin D, are important.

Oestrogen prescribing for pubertal induction varies between countries and clinicians. After pubertal induction is complete, there is now a move towards offering young women with oestrogen deficiency physiological oestrogen replacement in the form of hormone replacement therapy (HRT) rather than the combined oral contraceptive pill, unless there are concerns about the possibility of spontaneous pregnancy. The reason for this move is based on HRT being more favourable for bone health and cardiovascular health with lower blood pressure, better renal function, and less activation of the renin angiotensin system [136].

Evidence for the optimal mode of oestrogen administration (oral or transdermal) is inconclusive. Transdermal oestradiol results in more physiological oestrogen concentrations and is preferred in young women with obesity, hypertension, migraines, and increased risk of venous thromboembolism [136]. Cyclical progestogen is required to maintain the health of the endometrium. Young women and their parents, particularly their mothers, should be reassured that concerns about the increased risk of breast cancer with HRT does not apply to this age group [136].

Altered sexual function and genitourinary symptoms may still occur despite adequate oestrogen replacement [136], and these should be discussed at their appointments. If these issues do occur, options such as increasing oestrogen dose as well as lubricants and local oestrogen can be considered. For some women with POI and reduced libido, the use of testosterone supplementation could be considered, although young women should be counselled about lack of data on long-term efficacy and safety. Duration of this therapy should potentially be restricted to 24 months.

Routinely annual evaluation should include smoking status, BMI and blood pressure; in some groups, such as those with Turner syndrome and some cohorts of survivors of childhood cancer, those with brain tumours and following bone marrow transplantation, this should also include a lipid profile and HbA1c. Assessment of bone mineral density (BMD) should be considered, particularly if other risk factors are present, and repeated at 5 years if evidence of osteoporosis has been found [136]. Appropriate adjustments to BMD should be made for short stature. A decrease in BMD should prompt review of oestrogen replacement and other potential factors.

Glucocorticoid Replacement in Adolescents with Congenital Adrenal Hyperplasia

Management of adolescents with congenital adrenal hyperplasia (CAH) presents challenges, but also offers opportunities to involve young people in making decisions about their management [139]. Control can deteriorate in adolescence due to physiological changes during puberty [140] and reduced adherence to medical therapy. Once linear growth is complete, there is a shift in focus of treatment to prevention of long-term adverse outcomes such as subfertility secondary to testicular adrenal rest tumours or menstrual irregularities, increased cardiovascular risk, and osteoporosis. These long-term adverse outcomes are in evidence during adolescence [141].

In adolescents with classical CAH, education about avoiding and recognizing signs and symptoms of adrenal crisis is key, as there is evidence that adrenal crises in CAH increase in frequency in young adulthood [142]. Although it is accepted that lifelong glucocorticoid therapy is required, the glucocorticoid regimen should be re-assessed after completion of growth and puberty, particularly in those struggling to achieve optimal control. While hydrocortisone is almost exclusively used in paediatrics for the management of CAH, studies from the UK and US demonstrate that two thirds of adults are on long-acting glucocorticoid preparations such as dexamethasone or prednisolone [143, 144]. There are no randomized controlled trials comparing these regimens, however a recent UK study did identify that increased adiposity, insulin resistance and use of prednisolone or dexamethasone are associated with impaired QoL in adults with CAH [145]. While hydrocortisone may be the optimal glucocorticoid, in some the priority will be simplifying the glucocorticoid regimen and changing to a long acting glucocorticoid once or twice daily as required. Long-acting hydrocortisone preparations are being trialled in patients with CAH [146]. Choice of long-acting glucocorticoid may depend on circumstances. In adolescent males with testicular adrenal rest tumours, short term dexamethasone may potentially preserve or restore fertility [147]. In females who are heterosexually active without adequate contraception, dexamethasone should be avoided because the placenta does not inactivate it. The exception to this would be when the young woman who is planning a pregnancy and has been assessed by a clinical geneticist as being at risk of having an affected child, when the pros and cons of dexamethasone should be discussed.

The overall aim is to use the lowest dose of glucocorticoid to normalize androstenedione concentrations (normalizing 17OHP concentrations will result in overtreatment). Long-acting glucocorticoid medications have reduced mineralocorticoid effects but monitoring plasma renin activity and blood pressure is always indicated. Fludrocortisone doses should be adjusted to maintain plasma renin activity within the normal range.

Patients with non-classic CAH represent a group less at risk of adrenal crisis and consequently may not need glucocorticoid therapy continuously lifelong. Guidelines suggest that patients should be given the option for discontinuing glucocorticoids [148]. Some authors have recommended weaning boys with non-classical CAH off hydrocortisone from mid puberty and girls 2–3 years after achieving menarche [149]. In patients diagnosed in late adolescence, glucocorticoid therapy might be indicated only at specific times. In females, glucocorticoids may be required to regularize periods to enhance fertility and reduce rate of miscarriage [149]; other options such as the use of the combined oral contraceptive pill and antiandrogens are available for other symptoms in those not seeking fertility [150]. Assessment of adrenal function is important as around a third will have subnormal cortisol levels during a short Synacthen test and, despite a lack of evidence in relation to the occurrence of adrenal crisis, the advice to cover sickness with hydrocortisone should be emphasized [151, 152]. Although data are lacking for long-term outcomes in non-classical CAH, follow up into adulthood is advised, particularly in females who may require management of hyperandrogenaemia or assistance with fertility.

Conclusion

Addressing the health needs of adolescents and young adults with long-term endocrine conditions requires developmentally appropriate healthcare with effective and safe transition and transfer to adult endocrine services. More evidence is required to inform how to deliver effective care to this age group. Paediatric and adult endocrinologists need to recognize the implications of having an endocrine condition on psychological and social development. They need the knowledge about common presentations of endocrine conditions and management during adolescence to allow appropriate counselling of patients about their choices and involve them in their care.

REFERENCES

1. World Health Organization. Adolescent Health [Internet]. Available at: http://www.who.int/maternal_child_adolescent/topics/adolescence/en/. (accessed: 15 August 2015).
2. Arnett JJ. Emerging adulthood: a theory of development from the late teens through the twenties. *Am Psychol*, 2000; **55**: 469–80.
3. Kilford EJ, Garrett E, Blakemore S-J. The development of social cognition in adolescence: an integrated perspective. *Neurosci Biobehav Rev*, 2016; **70**: 106–20.
4. Maynard LM, Wisemandle W, Roche AF, Chumlea WC, Guo SS, Siervogel RM. Childhood body composition in relation to body mass index. *Pediatrics*, 2001; 107: 344–350.
5. Guo SS, Chumlea WC, Roche AF, Siervogel RM. Age- and maturity-related changes in body composition during adolescence into adulthood: the Fels Longitudinal Study. *Int J Obes Relat Metab Disord*, 1997; **21**: 1167–75.
6. Nguyen TV, Maynard LM, Towne B, *et al*. Sex differences in bone mass acquisition during growth: the Fels Longitudinal Study. *J Clin Densitom*, 2001; **4**: 147–57.
7. Taylor RW, Grant AM, Williams SM, Goulding A. Sex differences in regional body fat distribution from pre- to postpuberty. *Obesity*, 2010; **18**: 1410–16.
8. Siervogel RM, Maynard LM, Wisemansdle WA, et al. Annual changes in total body fat (TBF) and fat free mass (FFM) in children from 8 to 18 years in relation to changes in body mass index (BMI). The Fels Longitudinal Study. *Ann NY Acad Sci*, 2000; **904**: 420–3.
9. Holst D, Grimaldi PA. New factors in the regulation of adipose differentiation and metabolism. *Curr Opin Lipidol*, 2002; **13**: 241–5.
10. Janeckova R. The role of leptin in human physiology and pathophysiology. *Physiol Res*, 2001; **50**: 443–59.
11. Remsberg KE, Talbott EO, Zborowski JV, Evans RW, McHugh-Pemu K. Evidence for competing effects of body mass, hyperinsulinemia, insulin resistance, and androgens on leptin levels among lean, overweight, and obese women with polycystic ovary syndrome. *Fertil Steril*, 2002; **78**: 479–86.
12. Ellis K, Nicolson M. Leptin levels and body fatness in children: effects of gender, ethnicity, and sexual development. *Pediatr Res*, 1997; **42**: 484–8.
13. Baxter-Jones A, Faulkner R, Forwood M, Mirwald R, Bailey D. Bone mineral accrual from 8 to 30 years of age: an estimation of peak bone mass. *J Bone Miner Res*, 2011; **26**: 1729–39.
14. Gordon CM, Zemel BS, Wren TA, *et al*. The determinants of peak bone mass. *J Pediatr*, 2017; **180**: 261–9.
15. Siervogel R M, Demerath E W, Schubert C, *et al*. Puberty and body composition. *Horm Res*, 2003; **60**: 36–45.
16. Cornoni-Huntley J, Harlan W, Leaverton P. Blood pressure in adolescence: the United States Health Examination Survey. *Hypertension*, 1979; **1**: 566–71.
17. Eissa M, Mihalopoulos N, Holubkov R, Dai S, Labarthe D. Changes in fasting lipids during puberty. *J Pediatr*, 2016; **170**: 199–205.
18. Daniel WA, Jr. Haematocrit: maturity relationship in adolescence. *Pediatrics*, 1973; **52**: 388–94.
19. Kaplowitz PB, Slora EJ, Wasserman RC, Pedlow SE, Herman-Giddens ME. Earlier onset of puberty in girls: relation to increased body mass index and race. *Pediatrics*, 2001; **108**: 347–53.
20. Rudolph KD, Troop-Gordon W, Lambert SF, Natsuaki MN. Long-term consequences of pubertal timing for youth depression: identifying personal and contextual pathways of risk. *Dev Psychopathol*, 2014; **26**: 1423–44
21. Deardorff J, Gonzales NA, Christopher FS, Roosa MW, Millsap RE. Early puberty and adolescent pregnancy: the influence of alcohol use. *Pediatrics*, 2005; **116**: 1451–6.
22. Downing J, Bellis MA. Early pubertal onset and its relationship with sexual risk taking, substance use and anti-social behaviour: a preliminary cross-sectional study. *BMC Public Health* 2009; **9**: 446.
23. Udry JR. Age at menarche, at first intercourse, and at first pregnancy. *J Biosoc Sci* 1979; **11**: 433–41.
24. Michaud PA, Suris JC, Deppen A. Gender-related psychological and behavioural correlates of pubertal timing in a national sample of Swiss adolescents. *Mol Cell Endocrinol*, 2006; 254–5: 172–8.
25. Golub MS, Collman GW, Foster PM, *et al*. Public health implications of altered puberty timing. *Pediatrics*, 2008; **121**(Suppl 3): S218–30.
26. Mrug S, Elliott MN, Davies S, Tortolero SR, Cuccaro P, Schuster MA. Early puberty, negative peer influence, and problem behaviors in adolescent girls. *Pediatrics*, 2014; **133**: 7–14.
27. Weingarden H, Renshaw KD. Early and late perceived pubertal timing as risk factors for anxiety disorders in adult women. *J Psychiatr Res*, 2012; **46**: 1524–9.
28. Benoit A, Lacourse E, Claes M. Pubertal timing and depressive symptoms in late adolescence: the moderating role of individual, peer, and parental factors. *Dev Psychopathol*, 2013: **25**: 455–71.
29. Aydogan U, Aydogdu A, Akbulut H, *et al*. Increased frequency of anxiety, depression, quality of life and sexual life in young hypogonadotropic hypogonadal males and impacts of testosterone replacement therapy on these conditions. *Endocr J*, 2012; **59**: 1099–105.
30. Lasaite L, Ceponis J, Preiksa RT, Zilaitiene B. Impaired emotional state, quality of life and cognitive functions in young hypogonadal men. *Andrologia*, 2014; **46**: 1107–12.
31. Turriff A, Levy HP, Biesecker B. Prevalence and psychosocial correlates of depressive symptoms among adolescents and adults with Klinefelter syndrome. *Genet Med*, 2011; **13**: 966–72.
32. Simm PJ, Zacharin MR. The psychosocial impact of Klinefelter syndrome—a 10 year review. *J Pediatr Endocrinol Metab* 2006; **19**: 499–505.
33. Carel JC, Elie C, Ecosse E, *et al*. Self-esteem and social adjustment in young women with Turner syndrome—influence of pubertal management and sexuality: population-based cohort study. *J Clin Endocrinol Metab*, 2006; **91**: 2972–9.

34. Bullinger M, Kołtowska-Häggström M, Sandberg D, *et al.* Health-related quality of life of children and adolescents with growth hormone deficiency or idiopathic short stature—part 2: available results and future directions. *Horm Res*, 2009; **72**: 74–81.
35. Keselman A, Martínez A, Pantano L, Bergadá C, Heinrich JJ. Psychosocial outcome in growth hormone deficient patients diagnosed during childhood. *J Pediatr Endocrinol Metab*, 2000; **13**: 409–16.
36. Becker A, Hagenberg N, Roessner V, Woerner W, Rothenberger A. Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *Eur Child Adolesc Psychiatry*, 2004; **13**: 17–24.
37. Silva N, Crespo C, Carona C, *et al.* Why the (dis) agreement? Family context and child—parent perspectives on health-related paediatric asthma. *Child Care Health Dev*, 2015; **41**: 112–21.
38. White-Koning M, Arnaud C, Dickinson HO, *et al.* Determinants of child-parent agreement in quality-of-life reports: a European study of children with cerebral palsy. *Pediatrics*, 2007; **120**: e804–14.
39. Annett R, Bender BG, DuHamel TR, Lapidus J. Factors influencing parent reports on quality of life for children with asthma. *J Asthma*, 2003; **40**: 577–87.
40. Cromer B. Adolescent development. Chapter 14. In: Kliegman R, Stanton B, St. Geme J, Scho N, eds. *Nelson Textbook of Pediatrics*, 20th edn. Philadelphia, PA: Elsevier, 2016.
41. Pinquart M. Body image of children and adolescents with chronic illness: a meta-analytic comparison with healthy peers. *Body Image*, 2013; **10**: 141–8.
42. Hagell A, Shah R, Coleman J. *Key Data on Adolescence 2017*. London: Association for Young People's Health, 2017.
43. Mokdad AH, Forouzanfar MH, Daoud F, *et al.* Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 2016; **387**: 2383–401.
44. Surís J-C, Michaud P-A, Akre C, Sawyer SM. Health risk behaviors in adolescents with chronic conditions. *Pediatrics*, 2008; **122**: e1113–18.
45. Pinquart M. Achievement of developmental milestones in emerging and young adults with and without pediatric chronic illness—a meta-analysis. *J Pediatr Psychol*, 2014; **39**: 577–87.
46. Maslow GR, Haydon A, McRee A-L, Ford CA, Halpern CT. Growing up with a chronic illness: social success, educational/vocational distress. *J Adolesc Health*, 2011; **49**: 206–12.
47. Mitra MT, Jönsson P, Åkerblad A-C, *et al.* Social, educational and vocational outcomes in young people with childhood-onset and young-adult-onset growth hormone deficiency. *Clin Endocrinol*, 2017; **86**: 526–33.
48. Churchill R, Allen J, Denman S, Williams D, Fielding K, Von Fragstein M. Do the attitudes and beliefs of young teenagers towards general practice influence actual consultation behavior? *Brit J Gen Pract*, 2000; **50**: 953–7.
49. Lynn D, Umari T, Dunnick C, Dellavalle R. The epidemiology of acne vulgaris in late adolescence. *Adolesc Health Med Ther*, 2016; **7**: 13–25.
50. Crowley S, Acebo C, Carskadon M. Sleep, circadian rhythms, and delayed phase in adolescence. *Sleep Med*, 2007; **8**: 602–12.
51. Turnbull N, Shaw EJ, Baker R, *et al. Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (or Encephalopathy): Diagnosis and Management of Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (or Encephalopathy) in Adults and Children*. London: Royal College of General Practitioners, 2007.
52. Currie C, Hurrelmann K, Settertobulte W, Smith R, Todd J, eds. *Health and Health Behaviour Among Young People (Health Policy for Children and Adolescents, No.1)*. Copenhagen: WHO Regional Office for Europe, 2000.
53. Austin JB, Kaur B, Anderson HR, *et al.* Hay fever, eczema and wheeze: a national UK study (ISAAC, international study of asthma and allergies in childhood). *Arch Dis Child*, 1999; **81**: 225–30.
54. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry*, 2005; **62**: 593–602.
55. Secinti E, Thompson EJ, Richards M, Gaysina D. Research review: childhood chronic physical illness and adult emotional health—a systematic review and meta-analysis. *J Child Psychol Psychiatry*, 2017; **58**: 753–69.
56. General Medical Council. *0–18 Years: Guidance for All Doctors*. London: GMC, 2007.
57. Gleeson H, Turner G. Transition to adult services. *Arch Dis Child Educ Pract Ed*, 2012; **97**: 86–92.
58. Allen D, Channon S, Lowes L, *et al.* Behind the scenes: the changing roles of parents in the transition from child to adult diabetes service. *Diabet Med*, 2011; **28**: 994–1000.
59. Goldenring J, Cohen E. Getting into adolescent heads. *Contemp Paediatr*, 1988; **5**: 75–90.
60. Wilson H, Bostock N, Phillip N, *et al.* Opportunistic adolescent health screening of surgical inpatients. *Arch Dis Child*, 2012; **97**: 919–21.
61. Gleeson H, Davis J, Jones J, O'Shea E, Clayton PE. The challenge of delivering endocrine care and successful transition to adult services in adolescents with congenital adrenal hyperplasia: experience in a single centre over 18 years. *Clin Endocrinol*, 2013; **78**: 23–8.
62. Arlt W, Willis DS, Wild SH, Krone N, *et al.* Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 young people. *J Clin Endocrinol Metab*, 2010; **95**: 5110–21.
63. Developmentally appropriate healthcare toolkit. Available at: https://www.northumbria.nhs.uk/wp-content/uploads/2017/04/nhs-making-healthcare-work-web-02.pdf (accessed: 2 April 2020).
64. Wright R J, Chapman S, Cheer K, *et al.* On behalf of the Adolescent and Young Adult Special Interest Group. Training needs in adolescent health and transition in paediatric and adult trainees in endocrinology and diabetes in the UK. *Br J Diabetes Vasc Dis*, 2015; **15**: 159–65.
65. Blum RW, Garell D, Hodgman CH, *et al.* Transition from child-centered to adult health-care systems for adolescents with chronic conditions. A position paper of the Society for Adolescent Medicine. *J Adolesc Health*, 1993; **14**: 570–6.
66. Campbell F, Biggs K, Aldiss SK, *et al.* Transition of care for adolescents from paediatric services to adult health services. *Cochrane Database Syst Rev*, 2016; **4**: CD009794.
67. Transition from children's to adults' services for young people using health or social care services. NICE guideline NG43. Available at: https://www.nice.org.uk/guidance/ng43 (accessed: 2 April 2020).
68. Colver A, McConachie H, Le Couteur A, *et al.* A longitudinal, observational study of the features of transitional healthcare associated with better outcomes for young people with long-term conditions. *BMC Med*, 2018; **16**: 111.

69. Suris J-C, Akre C. Key elements for, and indicators of, a successful transition: an international Delphi study. *J Adolesc Health*, 2015; **56**: 612–18.
70. Gabriel P, McManus M, Rogers K, White P. Outcome evidence for structured pediatric to adult health care transition interventions: a systematic review. *J Pediatr*, 2017; **188**: 263–9.e15.
71. Bachelot A, Vialon M, Baptiste A, *et al*. Impact of transition on quality of life in patients with congenital adrenal hyperplasia diagnosed during childhood. *Endocr Connect*, 2017; **6**: 422–9.
72. Pedreira CC, Hameed R, Kanumakala S, *et al*. Health-care problems of Turner syndrome in the adult woman: a cross-sectional study of a Victorian cohort and a case for transition. *Intern Med J*, 2006; **36**: 54–7.
73. Ranke MB, Wit JM. Growth hormone—past, present and future. *Nat Rev Endocrinol*, 2018; **14**: 285–300.
74. Cook DM. Adult growth hormone deficiency syndrome: a personal approach to diagnosis, treatment and monitoring. *Growth Horm IGF Res*, 1999; **9** Suppl A: 129–33.
75. Attanasio AF, Lamberts SW, Matranga AM, *et al*. Adult growth hormone (GH)-deficient patients demonstrate heterogeneity between childhood onset and adult onset before and during human GH treatment. Adult Growth Hormone Deficiency Study Group. *J Clin Endocrinol Metab*, 1997; **82**: 82–8.
76. Attanasio AF, Howell S, Bates PC, *et al*. Body composition, IGF-I and IGFBP-3 concentrations as outcome measures in severely GH-deficient (GHD) patients after childhood GH treatment: a comparison with adult onset GHD patients. *J Clin Endocrinol Metab*, 2002; **87**: 3368–72.
77. Clayton PE, Cuneo RC, Juul A, Monson JP, Shalet SM, Tauber M. Consensus statement on the management of the GH-treated adolescent in the transition to adult care. *Eur J Endocrinol*, 2005; **152**: 165–70.
78. Högler W, Shaw N. Childhood growth hormone deficiency, bone density, structures and fractures: scrutinizing the evidence. *Clin Endocrinol (Oxf)*, 2010; **72**: 281–9.
79. Bouillon R, Koledova E, Bezlepkina O, *et al*. Bone status and fracture prevalence in Russian adults with childhood-onset growth hormone deficiency. *J Clin Endocrinol Metab*, 2004; **89**: 4993–8.
80. Murray RD, Adams JE, Shalet SM. A densitometric and morphometric analysis of the skeleton in adults with varying degrees of growth hormone deficiency. *J Clin Endocrinol Metab*, 2006; **91**: 432–8.
81. Clayton P, Gleeson H, Monson J, Popovic V, Shalet SM, Christiansen JS. Growth hormone replacement throughout life: insights into age-related responses to treatment. *Growth Horm IGF Res*, 2007; **17**: 369–82.
82. Bechtold S, Bachmann S, Putzker S, Dalla Pozza R, Schwarz HP. Early changes in body composition after cessation of growth hormone therapy in childhood-onset growth hormone deficiency. *J Clin Densitom*, 2011; **14**: 471–7.
83. Hyldstrup L, Conway GS, Racz K, *et al*. Growth hormone effects on cortical bone dimensions in young adults with childhood-onset growth hormone deficiency. *Osteoporos Int*, 2012; **23**: 2219–26.
84. Ahmid M, Perry CG, Ahmed SF, Shaikh MG. Growth hormone deficiency during young adulthood and the benefits of growth hormone replacement. *Endocr Connect*, 2016; **5**: R1–11.
85. Jørgensen JOL, Juul A. Therapy of endocrine disease: growth hormone replacement therapy in adults: 30 years of personal clinical experience. *Eur J Endocrinol*, 2018; **179**: R47–R56.
86. Underwood LE, Attie KM, Baptista J. Growth hormone (GH) dose-response in young adults with childhood-onset GH deficiency: a two-year, multicenter, multiple-dose, placebo-controlled study. *J Clin Endocrinol Metab*, 2003; **88**: 5273–80.
87. Colao A, Di Somma C, Salerno M, Spinelli L, Orio F, Lombardi G. The cardiovascular risk of GH-deficient adolescents. *J Clin Endocrinol Metab*, 2002; **87**: 3650–5.
88. Colao A, Di Somma C, Rota F, *et al*. Common carotid intima-media thickness in growth hormone (GH)-deficient adolescents: a prospective study after GH withdrawal and restarting GH replacement. *J Clin Endocrinol Metab*, 2005; **90**: 2659–65.
89. Attanasio AF, Shavrikova E, Blum WF, *et al*. Continued growth hormone (GH) treatment after adult height is necessary to complete somatic development in childhood-onset GH-deficient patients. *J Clin Endocrinol Metab*, 2004; **89**: 4857–62.
90. Lanes R, Paoli M, Carrillo E, Villaroel O, Palacios A. Cardiovascular risk of young growth-hormone-deficient adolescents. Differences in growth-hormone-treated and untreated patients. *Horm Res*, 2003; **60**: 291–6.
91. Carroll PV, Drake WM, Maher KT, *et al*. Comparison of continuation or cessation of growth hormone (GH) therapy on body composition and metabolic status in adolescents with severe GH deficiency at completion of linear growth. *J Clin Endocrinol Metab*, 2004; **89**: 3890–5.
92. Mauras N, Pescovitz OH, Allada V, Messig M, Wajnrajch MP, Lippe B. Limited efficacy of growth hormone (GH) during transition of GH-deficient patients from adolescence to adulthood: a phase III multicenter, double-blind, randomized two-year trial. *J Clin Endocrinol Metab*, 2005; **90**: 3946–55.
93. Vahl N, Juul A, Jørgensen JO, Orskov H, Skakkebaek NE, Christiansen JS. Continuation of growth hormone (GH) replacement in GH-deficient patients during transition from childhood to adulthood: a two-year placebo-controlled study. *J Clin Endocrinol Metab*, 2000; **85**: 1874–81.
94. Capaldo B, Patti L, Oliviero U, *et al*. Increased arterial intima-media thickness in childhood-onset growth hormone deficiency. *J Clin Endocrinol Metab*, 1997; **82**: 1378–81.
95. Kołtowska-Häggström M, Geffner ME, Jönsson P, *et al*. Discontinuation of growth hormone (GH) treatment during the transition phase is an important factor determining the phenotype of young adults with nonidiopathic childhood-onset GH deficiency. *J Clin Endocrinol Metab*, 2010; **95**: 2646–54.
96. Feinberg MS, Scheinowitz M, Laron Z. Cardiac dimension and function in patients with childhood onset growth hormone deficiency, before and after growth hormone retreatment in adult age. *Am Heart J*, 2003; **145**: 549–53.
97. Nørrelund H, Vahl N, Juul A, *et al*. Continuation of growth hormone (GH) therapy in GH-deficient patients during transition from childhood to adulthood: impact on insulin sensitivity and substrate metabolism. *J Clin Endocrinol Metab*, 2000; **85**: 1912–17.
98. Attanasio AF, Shavrikova EP, Blum WF, Shalet SM. Quality of life in childhood onset growth hormone-deficient patients in the transition phase from childhood to adulthood. *J Clin Endocrinol Metab*, 2005; **90**: 4525–9.
99. Stouthart PJ, Deijen JB, Roffel M, Delemarre-van de Waal HA. Quality of life of growth hormone (GH) deficient young adults during discontinuation and restart of GH therapy. *Psychoneuroendocrinology*, 2003; **28**: 612–26.
100. Hoybye C, Jönsson P, Monson JP, *et al*. Impact of the primary aetiology upon the clinical outcome of adults with childhood-onset GH deficiency. *Eur J Endocrinol*, 2007; **157**: 589–96.

101. Wirén L, Johannsson G, Bengtsson BA. A prospective investigation of quality of life and psychological well-being after the discontinuation of GH treatment in adolescent patients who had GH deficiency during childhood. *J Clin Endocrinol Metab*, 2001; **86**: 3494–8.
102. Brütt AL, Sandberg DE, Chaplin J, *et al.* Assessment of health-related quality of life and patient satisfaction in children and adolescents with growth hormone deficiency or idiopathic short stature—part 1: a critical evaluation of available tools. *Horm Res*, 2009; **72**: 65–73.
103. Murray PG, Dattani MT, Clayton PE. Controversies in the diagnosis and management of growth hormone deficiency in childhood and adolescence. *Arch Dis Child*, 2016; **101**: 96–100.
104. Hartman ML, Crowe BJ, Biller BM, Ho KK, Clemmons DR, Chipman JJ; HyposCCS Advisory Board; U.S. HypoCCS Study Group. Which patients do not require a GH stimulation test for the diagnosis of adult GH deficiency? *J Clin Endocrinol Metab*, 2002; **87**: 477–85.
105. Quigley CA, Zagar AJ, Liu CC, *et al.* United States multicenter study of factors predicting the persistence of GH deficiency during the transition period between childhood and adulthood. *Int J Pediatr Endocrinol*, 2013; **2013**: 6.
106. Gleeson HK, Gattamaneni HR, Smethurst L, Brennan BM, Shalet SM. Reassessment of growth hormone status is required at adult height in children treated with growth hormone replacement after radiation therapy. *J Clin Endocrinol Metab*, 2004; **89**: 662–6.
107. Secco A, Di Iorgi N, Napoli F, *et al.* Reassessment of the growth hormone status in young adults with childhood-onset growth hormone deficiency: reappraisal of insulin tolerance testing. *J Clin Endocrinol Metab*, 2009; **94**: 4195–204.
108. Tauber M, Moulin P, Pienkowski C, Jouret B, Rochiccioli P. Growth hormone (GH) retesting and auxological data in 131 GH-deficient patients after completion of treatment. *J Clin Endocrinol Metab*, 1997; **82**: 352–6.
109. Maghnie M, Strigazzi C, Tinelli C, *et al.* Growth hormone (GH) deficiency (GH deficiency) of childhood onset: reassessment of GH status and evaluation of the predictive criteria for permanent GH deficiency in young adults. *J Clin Endocrinol Metab*, 1999; **84**: 1324–8.
110. Gelwane G, Garel C, Chevenne D, Armoogum P, Simon D, Czernichow P, Léger J. Subnormal serum insulin-like growth factor-I levels in young adults with childhood-onset nonacquired growth hormone (GH) deficiency who recover normal GH secretion may indicate less severe but persistent pituitary failure. *J Clin Endocrinol Metab*, 2007; **92**: 3788–95.
111. Maghnie M, Strigazzi C, Tinelli C, *et al.* Growth hormone (GH) deficiency (GHD) of childhood onset: reassessment of GH status and evaluation of the predictive criteria for permanent GHD in young adults. *J Clin Endocrinol Metab*, 1999; **84**: 1324–8.
112. Zucchini S, Pirazzoli P, Baronio F, *et al.* Effect on adult height of pubertal growth hormone retesting and withdrawal of therapy in patients with previously diagnosed growth hormone deficiency. *J Clin Endocrinol Metab*, 2006; **91**: 4271–6.
113. Léger J, Danner S, Simon D, Garel C, Czernichow P. Do all patients with childhood-onset growth hormone deficiency (GHD) and ectopic neurohypophysis have persistent GHD in adulthood? *J Clin Endocrinol Metab*, 2005; **90**: 650–6.
114. Mullis PE. Genetics of growth hormone deficiency. *Endocrinol Metab Clin North Am*, 2007; **36**: 17–36.
115. Karavitaki N, Brufani C, Warner JT, *et al.* Craniopharyngiomas in children and adults: systematic analysis of 121 cases with long-term follow-up. *Clin Endocrinol (Oxf)*, 2005; **62**: 397–409. Review.
116. Chemaitilly W, Li Z, Huang S, *et al.* Anterior hypopituitarism in adult survivors of childhood cancers treated with cranial radiotherapy: a report from the St Jude Lifetime Cohort study. *J Clin Oncol*, 2015; **33**: 492–500.
117. Merchant TE, Rose SR, Bosley C, Wu S, Xiong X, Lustig RH. Growth hormone secretion after conformal radiation therapy in pediatric patients with localized brain tumors. *J Clin Oncol*, 2011; **29**: 4776–80.
118. Schmiegelow M, Lassen S, Poulsen HS, *et al.* Growth hormone response to a growth hormone-releasing hormone stimulation test in a population-based study following cranial irradiation of childhood brain tumors. *Horm Res*, 2000; **54**: 53–9.
119. Holm K, Nysom K, Rasmussen MH, *et al.* Growth, growth hormone and final height after BMT. Possible recovery of irradiation-induced growth hormone insufficiency. *Bone Marrow Transplant*, 1996; **18**: 163–70.
120. Aimaretti G, Ambrosio MR, Di Somma C, *et al.* Hypopituitarism induced by traumatic brain injury in the transition phase. *J Endocrinol Invest*, 2005; **28**: 984–9.
121. Einaudi S, Matarazzo P, Peretta P, *et al.* Hypothalamo-hypophysial dysfunction after traumatic brain injury in children and adolescents: a preliminary retrospective and prospective study. *J Pediatr Endocrinol Metab*, 2006; 19: 691–703.
122. Niederland T, Makovi H, Gál V, Andréka B, Abrahám CS, Kovács J. Abnormalities of pituitary function after traumatic brain injury in children. *J Neurotrauma*, 2007; **24**: 119–27.
123. Ho K. Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, J. *Eur J Endocrinol*, 2007; **157**: 695–700.
124. Molitch ME, Clemmons DR, Malozowski S, Merriam GR, Vance ML. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 1587–609.
125. Guzzetti C, Ibba A, Pilia S, *et al.* Cut-off limits of the peak GH response to stimulation tests for the diagnosis of GH deficiency in children and adolescents: study in patients with organic GHD. *Eur J Endocrinol*, 2016; **175**: 41–7.
126. Cattoni A, Clarke E, Albanese A. The predictive value of insulin-like growth factor 1 in irradiation-dependent growth hormone deficiency in childhood cancer survivors. *Horm Res Paediatr*, 2018; **90**: 314–25.
127. Maghnie M, Aimaretti G, Bellone S, *et al.* Diagnosis of GH deficiency in the transition period: accuracy of insulin tolerance test and insulin-like growth factor-I measurement. *Eur J Endocrinol*, 2005; **152**: 589–96.
128. Corneli G, Di Somma C, Prodam F, *et al.* Cut-off limits of the GH response to GHrh plus arginine test and IGF-I levels for the diagnosis of GH deficiency in late adolescents and young adults. *Eur J Endocrinol*, 2007; **157**: 701–8.
129. di Iorgi N, Secco A, Napoli F, *et al.* Deterioration of growth hormone (GH) response and anterior pituitary function in young adults with childhood-onset GH deficiency and ectopic posterior pituitary: a two-year prospective follow-up study. *J Clin Endocrinol Metab*, 2007; **92**: 3875–84.
130. Attanasio AF, Shavrikova E, Blum WF, *et al.* Continued growth hormone (GH) treatment after adult height is necessary to

complete somatic development in childhood-onset GH-deficient patients. *J Clin Endocrinol Metab*, 2004; **89**: 4857–62.
131. Shalet SM, Shavrikova E, Cromer M, *et al*. Effect of growth hormone (GH) treatment on bone in postpubertal GH-deficient patients: a 2-year randomized, controlled, dose-ranging study. *J Clin Endocrinol Metab*, 2003; **88**: 4124–9.
132. Thankamony A, Capalbo D, Jonsson PJ, Simpson HL, Dunger DB. Predictors of insulin-like growth factor-I responses to growth hormone replacement in young adults with growth hormone deficiency. *Horm Res Paediatr*, 2016; **85**: 379–88.
133. Bengtsson BA, Abs R, Bennmarker H, *et al*. The effects of treatment and the individual responsiveness to growth hormone (GH) replacement therapy in 665 GH-deficient adults. KIMS Study Group and the KIMS International Board. *J Clin Endocrinol Metab*, 1999; **84**: 3929–35.
134. Johannsson G, Bidlingmaier M, Biller BMK, *et al*. Growth Hormone Research Society perspective on biomarkers of GH action in children and adults. *Endocr Connect*, 2018; **7**: R126–34.
135. Stochholm K, Juul S, Christiansen JS, Gravholt CH. Mortality and socioeconomic status in adults with childhood onset GH deficiency (GHD) is highly dependent on the primary cause of GHD. *Eur J Endocrinol*, 2012; **167**: 663–70.
136. European Society for Human Reproduction and Embryology (ESHRE) Guideline Group on POI, Webber L, Davies M, *et al*. ESHRE Guideline: management of women with premature ovarian insufficiency. *Hum Reprod*, 2016; **31**: 926–37.
137. Gravholt CH, Andersen NH, Conway GS, Dekkers OM, Geffner ME, Klein KO, *et al*. Clinical practice guidelines for the care of girls and women with Turner syndrome: proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. *Eur J Endocrinol*, 2017; **177**: G1–70.
138. Davenport ML. Moving toward an understanding of hormone replacement therapy in adolescent girls: looking through the lens of Turner syndrome. *Ann N Y Acad Sci*, 2008; **1135**: 126–37.
139. Merke DP, Poppas DP. Management of adolescents with congenital adrenal hyperplasia. *Lancet Diabetes Endocrinol*, 2013; **1**: 341–52.
140. Charmandari E, Brook CGD, Hindmarsh PC. Classic congenital adrenal hyperplasia and puberty. *Eur J Endocrinol*, 2004; **151** Suppl: U77–82.
141. Mooij CF, Webb EA, Claahsen van der Grinten HL, Krone N. Cardiovascular health, growth and gonadal function in children and adolescents with congenital adrenal hyperplasia. *Arch Dis Child*, 2017; **102**: 578–84.
142. Reisch N, Willige M, Kohn D, *et al*. Frequency and causes of adrenal crises over lifetime in patients with 21-hydroxylase deficiency. *Eur J Endocrinol*, 2012; **167**: 35–42.
143. Arlt W, Willis DS, Wild SH, *et al*. Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 patients. *J Clin Endocrinol Metab*, 2010; **95**: 5110–21.
144. Finkielstain GP, Kim MS, Sinaii N, *et al*. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2012; **97**: 4429–38.
145. Han TS, Krone N, Willis DS, *et al*. Quality of life in adults with congenital adrenal hyperplasia relates to glucocorticoid treatment, adiposity and insulin resistance: United Kingdom Congenital adrenal Hyperplasia Adult Study Executive (CaHASE). *Eur J Endocrinol*, 2013; **168**: 887–93.
146. Jones CM, Mallappa A, Reisch N, *et al*. Modified-release and conventional glucocorticoids and diurnal androgen excretion in congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2017; **102**: 1797–806.
147. King TFJ, Lee MC, Williamson EEJ, Conway GS. Experience in optimizing fertility outcomes in men with congenital adrenal hyperplasia due to 21 hydroxylase deficiency. *Clin Endocrinol (Oxf)*, 2016; **84**: 830–6.
148. Speiser PW, Azziz R, Baskin LS, *et al*. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2010; **95**: 4133–60.
149. Bidet M, Bellanné-Chantelot C, Galand-Portier M-B, *et al*. Fertility in women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2010; **95**: 1182–90.
150. Carmina E, Dewailly D, Escobar-Morreale HF, *et al*. Non-classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency revisited: an update with a special focus on adolescent and adult women. *Hum Reprod Update*, 2017; **23**: 580–99.
151. Nandagopal R, Sinaii N, Avila NA, *et al*. Phenotypic profiling of parents with cryptic nonclassic congenital adrenal hyperplasia: findings in 145 unrelated families. *Eur J Endocrinol*, 2011; **164**: 977–84.
152. Bidet M, Bellanné-Chantelot C, Galand-Portier M-B, *et al*. Clinical and molecular characterization of a cohort of 161 unrelated women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency and 330 family members. *J Clin Endocrinol Metab*, 2009; **94**: 1570–8.

SECTION 8

Female Reproductive Endocrine Disorders

8.1

Normal Female Endocrinology and Ovarian Disorders

8.1.1 Neuroendocrinology of Reproduction

The Role of Hypothalamus and Pituitary

Christopher R. McCartney and John C. Marshall

Introduction

Female reproductive physiology is characterized by highly complex hypothalamic–pituitary–ovarian interactions that, when functioning properly, achieve ovarian follicular development; episodic presentation of ova for possible fertilization (ovulation); and physiological preparation for possible pregnancy (e.g. endometrial decidualization). Reproductive neuroendocrine systems coordinate cyclic ovarian function to accomplish these goals, and gonadotropin-releasing hormone (GnRH)-secreting neurons represent the final common pathway for hypothalamic control of reproduction.

Anatomy of the Reproductive Hypothalamic–Pituitary Axis

The hypothalamus resides at the base of the brain, constituting the floor and inferior-lateral walls of the third ventricle. The hypothalamus is involved with a wide variety of biological processes including growth and development, energy homeostasis, weight maintenance, responses to stress, and reproduction. With regard to reproduction, the cell bodies of some 1000–1500 GnRH-secreting neurons are diffusely located throughout the preoptic area and the infundibular nucleus (the arcuate nucleus in lower mammalian species) (**Figure 8.1.1.1**) [1]. Selective destruction of the arcuate nucleus in mature female monkeys leads to severe hypogonadotropic hypogonadism [2], suggesting that normal gonadotropin secretion is particularly dependent on the population of GnRH neurons residing in the infundibular (arcuate) nucleus.

GnRH neurons extend processes through the tuberoinfundibular tract to the median eminence at the base of the third ventricle. GnRH neuron fibres have been called 'dendrons' because, in mice, they demonstrate characteristics of both axons and dendrites, including functional synaptic inputs throughout the length of the fibre [3]. Bursts of GnRH neuronal firing result in the release of GnRH into an extensive capillary network located within the median eminence—the proximal portion of the hypophyseal portal system. GnRH release may also be influenced by interactions with other cell populations in the median eminence. For example, projections from tanycytes—specialized ependymal cells lining the third ventricle—can reduce GnRH release by enveloping and physically isolating GnRH nerve terminals [4].

The hypophyseal portal circulation represents the anatomical and functional connection between the median eminence and cells of the anterior pituitary. The superior hypophyseal artery feeds the capillary network within the median eminence. This network drains into sinusoids that coalesce to form the hypophyseal portal veins, which traverse the pituitary stalk and constitute the chief blood supply for the anterior pituitary.

The pituitary gland is located within a saddle-like structure of the sphenoid bone (the sella turcica), and it is connected to the base of the hypothalamus by the pituitary stalk. The anterior lobe (pars distalis) of the pituitary contains specialized cells (gonadotropes)

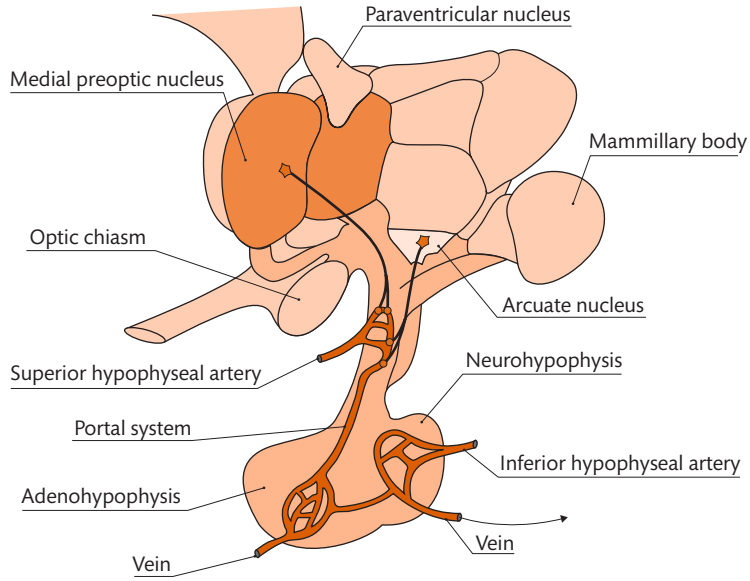


Figure 8.1.1.1 Anatomic foundations for GnRH neuron-gonadotrope interactions. GnRH neuron cell bodies reside in the infundibular nucleus (part of the mediobasal hypothalamus) and the preoptic area. GnRH neurons processes (dendrons) extend to the median eminence (also part of the mediobasal hypothalamus), where GnRH is released into the hypophyseal portal system. GnRH subsequently travels to the anterior pituitary where it interacts with gonadotrope cells.

Adapted with permission from Johnson and Everitt's *Essential reproduction*, ed 6. Blackwell Publishing, 2007, Fig. 6.4.

that produce luteinizing hormone (LH) and follicle-stimulating hormone (FSH). GnRH binds to GnRH receptors on the cell surface of gonadotropes, initiating an intracellular cascade that results in the synthesis and secretion of LH and FSH. These gonadotropins are released directly into the systemic circulation and ultimately direct ovarian follicular development, cyclic ovulation, and cyclic changes of ovarian sex steroid production.

Embryological Origins of GnRH Neurons

During embryologic development, nascent GnRH neurons are first identified in the nasal (olfactory) placode [5]. Initial GnRH cell migration toward the hypothalamus is guided by olfactory/vomeronasal neuron axons projecting toward the olfactory bulb, with subsequent migration guided by neurons that project caudally into the forebrain (Figure 8.1.1.2). Once reaching the hypothalamus, GnRH cells extend processes to the median eminence and gain access to the hypophyseal portal system.

As olfactory neurons form a pathway to guide GnRH neuronal migration, appropriate GnRH cell migration is dependent on normal olfactory system development. This is exemplified by Kallmann syndrome, in which abnormal olfactory system development renders an inadequate guidance infrastructure for GnRH neurons. Accordingly, individuals with Kallmann syndrome have congenital hypogonadotropic hypogonadism in addition to anosmia (absent

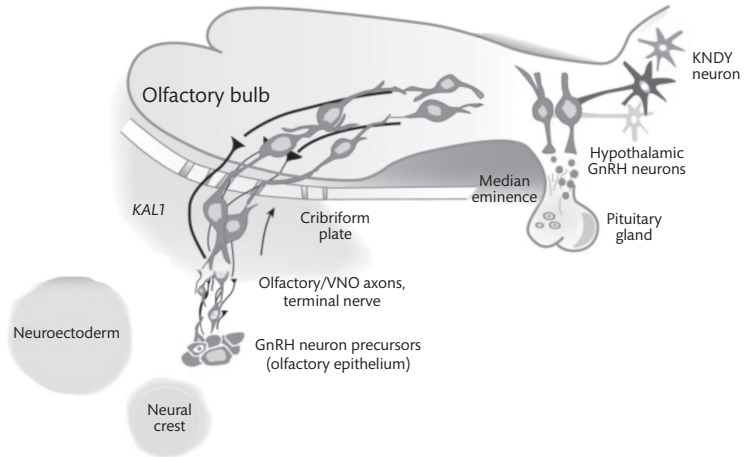


Figure 8.1.1.2 Schematic of embryological GnRH cell migration from the nasal epithelium to the hypothalamus, guided by olfactory/vomeronasal neurons.

Adapted with permission from Stamou MI, Cox KH, Crowley WF, Jr. Discovering Genes Essential to the Hypothalamic Regulation of Human Reproduction Using a Human Disease Model: Adjusting to Life in the '-Omics' Era. *Endocr Rev* 2015; 36:603–21. (Ref 5) Copyright 2015 Oxford University Press.

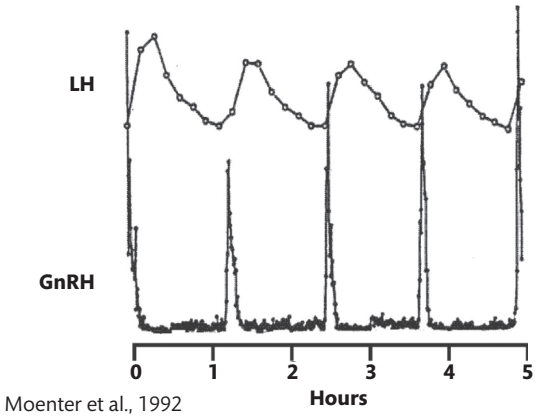


Figure 8.1.1.3 Temporal relationship between pulses of GnRH in the pituitary portal system and pulses of LH in the peripheral circulation (jugular vein). These data were obtained from a sheep model.

Adapted with permission from Moenter SM, Brand RM, Midgley AR, Karsch FJ. Dynamics of gonadotropin-releasing hormone release during a pulse. *Endocrinology* 1992; 130:503–10. (Ref 8) Copyright © 1992, Oxford University Press.

sense of smell). Kallmann syndrome was first described in patients with deletion of the *ANOS1* (formerly *KAL1*) gene, which encodes a matrix glycoprotein important for fetal olfactory development. Human fetuses with X-linked Kallmann syndrome exhibit arrested migration of GnRH-immunoreactive cells along terminal nerve fascicles in the nasal mucosa or at the cribriform plate [6]. Numerous other genes related to fetal GnRH cell migration underlie forms of congenital hypogonadotropic hypogonadism with and without anosmia [5].

GnRH Secretion

GnRH-1 is a decapeptide encoded by the *GNRH1* gene on human chromosome 8 (8p11.2-p21) [7]. *GNRH1* produces a 92-amino acid precursor peptide (prepro-GnRH) with the following segments: a 23-amino acid signal sequence, GnRH, a 3-amino acid proteolytic processing site, and a 56-amino acid GnRH-associated peptide. Essentially all mammalian species demonstrate the same amino acid structure of GnRH-1: (pyro) Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2.

Although loosely affiliated from an anatomical perspective, GnRH neurons are highly integrated from a functional perspective. GnRH neurons exhibit extensive connections to other GnRH neurons and to other neuronal populations. GnRH neuronal activity is characterized by episodic and coordinated bursts of action potential firing, which presumably correlate to changes in GnRH release at the median eminence. Episodic firing rate activity—with periods of high and low firing rates—appears to be an intrinsic property of GnRH neurons, but firing patterns and rates are substantially influenced by complex afferent neuronal networks (*vide infra*).

GnRH is released into the hypophyseal portal system in episodic bursts (Figure 8.1.1.3) [8]. Intermittent GnRH presentation to gonadotropes is an absolute requirement for maintenance of gonadotropin synthesis and secretion: there are no known parallel or backup systems. In particular, intermittent GnRH stimulation increases or maintains GnRH receptors on gonadotropes—the 'self-priming' or 'autopriming' effect, which is partly mediated via GnRH receptor density on gonadotrope cells. Absent or inappropriate GnRH secretion in humans leads to pubertal failure, hypogonadotropic hypogonadism, and infertility, all of which can be reversed with pulsatile exogenous GnRH administration [9].

Gonadotropin responses to GnRH can decrease with higher GnRH pulse frequencies [10]. Continuous GnRH receptor stimulation markedly reduces gonadotropin synthesis and secretion (desensitization) (Figure 8.1.1.4) [11], which is partly a reflection of reduced GnRH receptor expression on gonadotropes (i.e. receptor downregulation). Desensitization can be achieved for therapeutic purposes with long-acting GnRH receptor agonists, which transiently increase gonadotropin release, but subsequently cause desensitization of gonadotropin secretion. Competitive GnRH receptor antagonists are also available: these peptides reversibly bind to, but do not stimulate, the GnRH receptor, reducing gonadotropins within 24 to 72 hours.

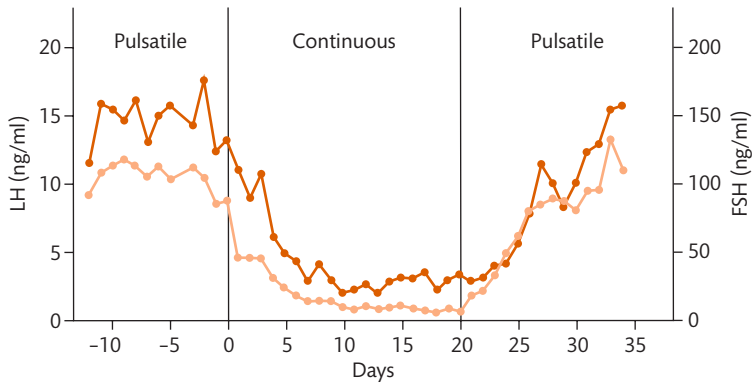


Figure 8.1.1.4 Impact of pulsatile vs. continuous GnRH administration in GnRH-deficient monkeys. These monkeys were gonadectomized (but sex steroid-replaced) and rendered GnRH deficient via radiofrequency ablation of the arcuate nucleus. Normal gonadotropin secretion is maintained with intermittent exogenous GnRH administration. However, continuous GnRH infusion markedly downregulates luteinizing hormone and follicle-stimulating hormone production.

Adapted with permission from Belchetz PE, Plant TM, Nakai Y, Keogh EJ, Knobil E. Hypophysial responses to continuous and intermittent delivery of hypothalamic gonadotropin-releasing hormone. *Science* 1978; 202:631–33. Copyright 1978 © The American Association for the Advancement of Science.

Upstream Regulation of GnRH Neurons

The regulation of pulsatile GnRH release is highly complex. A number of upstream neuronal populations—along with their associated neurotransmitters and neuropeptides—substantially impact GnRH neuron activity and largely mediate ovarian feedback—largely via circulating sex steroid concentrations—in addition to mediating the influence of energy availability and stress. Relevant signalling compounds include dopamine, norepinephrine, glutamate, γ-aminobutyric acid (GABA), nitric oxide, kisspeptin, neurokinin B, and endogenous opiates.

Kisspeptin is a critical requirement for normal GnRH secretion [12, 13]. The human kisspeptin gene (*KISS1*) encodes a 154-amino acid precursor protein (kisspeptin 1–145); and proteolytic modification yields a number of bioactive kisspeptin fragments: kisspeptin-54, -14, -13, and -10. Functional native kisspeptin species contain the carboxy-terminal amino acids 112 to 121, which are important for receptor binding and function. Kisspeptin neurons form synaptic contacts with GnRH neurons, including at GnRH neuron terminals in the median eminence. Kisspeptin stimulates GnRH neurons via the kisspeptin receptor (KISSR; also known as the G-protein coupled receptor 54, or GPR54)—a seven transmembrane domain, G-protein-coupled receptor expressed by a majority of GnRH neurons. Persons with inactivating *KISS1R* or *KISS1* mutations exhibit pubertal failure and hypogonadotropic hypogonadism. Kisspeptin-related stimulation of LH release is blocked by GnRH antagonists, and pulsatile GnRH administration can fully restore reproductive function in patients with *KISS1R* mutations. These findings suggest that the effects of kisspeptin on LH release is mediated by GnRH. Of interest, the effects of kisspeptin on LH release appear to be influenced by several factors including menstrual cycle phase and hormonal milieu.

Rodents exhibit two distinct populations of kisspeptin neurons. Kisspeptin neurons in the arcuate nucleus (mediobasal hypothalamus) mediate the negative feedback effects of oestrogen on GnRH release; while kisspeptin neurons in the anteroventral periventricular nucleus (AVPV) of the preoptic area mediate the positive feedback effects of oestrogen on GnRH release (i.e. GnRH surge generation). Kisspeptin neurons are primarily located in the infundibular (arcuate) nucleus in humans and monkeys. It remains uncertain whether unique populations of kisspeptin neurons play similar respective roles in primates [14].

Neurokinin B (NKB) is a decapeptide encoded by the tachykinin 3 gene (*TAC3*) [12, 15]. NKB preferentially binds to and acts via the neurokinin 3 receptor (*NK3R*), encoded by the *TACR3* gene. The role of NKB in central reproductive function is complex. Homozygous loss-of-function *TAC3* or *TACR3* mutations cause pubertal failure and hypogonadotropic hypogonadism. Senktide, a selective NK3R agonist, can provoke LH release, and this effect is abolished by GnRH receptor antagonism. Of interest, a number of studies suggest that NKB primarily influences GnRH release via its effects on kisspeptin release. For example, kisspeptin neurons express *NK3R;* senktide increases kisspeptin neuronal activity; LH responses to senktide are absent or impaired in Kiss1R knockout mice, with Kiss1R antagonism, or after Kiss1R desensitization; and continuous kisspeptin infusion enhances LH secretion in patients with loss-of-function *TAC3* or *TACR3* mutations. NKB analogues may have therapeutic use in the future.

Endogenous opioid peptides (EOP)—including endorphins, enkephalins, and dynorphins—participate in numerous neuroendocrine functions. Endorphins (e.g. β-endorphin) are derived from proopiomelanocortin (POMC) and act primarily via μ (micro)-opioid receptors; while dynorphins are products of the precursor prodynorphin and act chiefly at κ (kappa)-opioid receptors (KOR). Although derived from different precursor proteins, active endorphins, enkephalins, and dynorphins contain a common sequence at the amino-terminal (Tyr-Gly-Gly-Phe-[Met or Leu]). Ample evidence suggests that sex steroid negative feedback on GnRH release is at least partly mediated by hypothalamic opiates. Hypophyseal blood concentrations of β-endorphin increase during the luteal phase in monkeys [16]; opiate receptor antagonists increase LH pulse frequency when administered to women in the luteal phase [17], and exogenous opiates (e.g. morphine) can cause hypogonadotropic hypogonadism. Recent animal studies suggest that dynorphin is an important mediator of GnRH pulse frequency suppression by progesterone [12, 13]. For example, progesterone receptors colocalize with arcuate nucleus dynorphin neurons; dynorphin-containing varicosities associate with GnRH neurons; progesterone treatment increases dynorphin A concentrations in the third ventricle cerebrospinal fluid; and delivery of κ-opioid receptor antagonists to the mediobasal hypothalamus reverses progesterone inhibition of LH pulse secretion.

Animal studies suggest that kisspeptin, NKB, and dynorphin are often coexpressed in single neurons [12]. Such KNDy ('candy') neurons (Kisspeptin, Neurokinin B, Dynorphin) are extensively interconnected and project to GnRH neurons, including GnRH neuron fibres within the median eminence. Although corresponding data in humans are limited, autopsy studies generally support the notion that kisspeptin and NKB colocalize in a portion of neurons in the infundibular nucleus [18]. Regardless, kisspeptin, NKB, and dynorphin—released from neurons that do or do not colocalize with the other peptides—substantially influence GnRH neuronal function in humans, and such neurons are likely a fundamental component of the GnRH pulse generator (vide infra).

RFamide-related peptides (RFRPs)—the mammalian orthologues of gonadotropin-inhibitory hormone (GnIH)—also appear to exert an important influence on central reproductive function [19]. RFRP-immunoreactive cells have been identified in human hypothalami, and RFRP-immunoreactive fibres approximate GnRH neurons. Animal studies suggest that RFRP-3 reduces GnRH neuronal firing rates, but an understanding of the roles of RFRPs in humans requires further study.

The GnRH Pulse Generator

The fundamental nature of pulsatile GnRH secretion—including the neuroanatomical components of the GnRH pulse generator and mechanisms underlying episodic GnRH release—remains unclear. The mediobasal hypothalamus, which contains both the infundibular (arcuate) nucleus and the median eminence, appears to contain all requisite components of the GnRH pulse generator. In monkeys and sheep, LH pulse initiation is temporally associated with discrete episodes of neuronal firing (multiple unit electrical activity) in the area of the mediobasal hypothalamus (**Figure 8.1.1.5**) [20, 21]; and in monkeys, experimental electrical stimulation of the

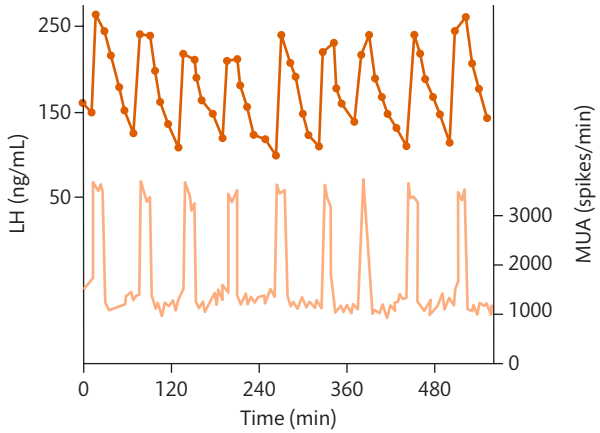


Figure 8.1.1.5 Temporal relationship between hypothalamic multiple unit activity (MUA) and pulses of luteinizing hormone (LH) in the peripheral circulation. These data were obtained from an ovariectomized monkey.

Adapted with permission from Knobil E. The electrophysiology of the GnRH pulse generator in the rhesus monkey. *J Steroid Biochem* 1989; 33:669–71. Copyright © 1989 Elsevier Ltd. (21).

mediobasal hypothalamus provokes GnRH release into the hypophyseal portal system [22]. Moreover, pulsatile LH secretion is maintained in monkeys after mediobasal hypothalamic isolation from the remainder of the brain; and isolated human mediobasal hypothalami release GnRH in a pulsatile fashion [23].

Numerous studies suggest that pulsatility is an intrinsic property of GnRH neurons; and coordination of GnRH release may reflect extensive interconnections among GnRH neurons. However, a number of observations suggest that kisspeptin, NKB, and dynorphin help to orchestrate coordinated GnRH neuronal activity [12, 24]. Although mechanisms remain unclear, some investigators have proposed a model whereby initial NKB release stimulates kisspeptin secretion, which initiates a GnRH pulse; and subsequent dynorphin release inhibits kisspeptin secretion, terminating the GnRH pulse. Although pulsatile GnRH release may occur in the absence of kisspeptin action (e.g. humans with KISSR mutations demonstrate low-amplitude LH pulses [25]), it seems clear that kisspeptin is mandatory for normal GnRH secretion.

Pituitary Gonadotropes

The anterior pituitary (adenohypophysis) is of ectodermal origin and forms during embryological development as an upward invagination of pharyngeal epithelium (Rathke pouch). Gonadotropes are specialized cells of the anterior pituitary that synthesize and secrete LH and FSH. GnRH binds to the GnRH type I receptor, a seven transmembrane G-protein-coupled receptor on the plasma membrane of gonadotropes: this activates a number of intracellular signalling cascades that enhance gonadotropin synthesis and secretion [26].

Each gonadotropin consists of α and β subunits, which dimerize. The biological specificity of LH and FSH relates to their unique β-subunits (the 121-amino acid LHβ and the 117-amino acid FSHβ, respectively). The α-subunit is common to LH, FSH, human chorionic gonadotropin (hCG), and thyrotropin. Gonadotropins also undergo variable post-translational modification—glycosylation in particular—which alters both bioactivity and elimination half-life.

Even though both LH and FSH are produced in a common cell type in response to GnRH stimulation, circulating LH and FSH concentrations vary differentially throughout ovulatory cycles. Numerous mechanisms underlie differential release of LH and FSH throughout the cycle [26]. For example, higher frequency GnRH pulses favour LH synthesis and secretion, while lower frequency GnRH pulses enhance FSH synthesis and secretion (**Figure 8.1.1.6**) [10]. In addition, inhibins and oestradiol selectively restrain FSH release during the mid- and late follicular phase and the luteal phase.

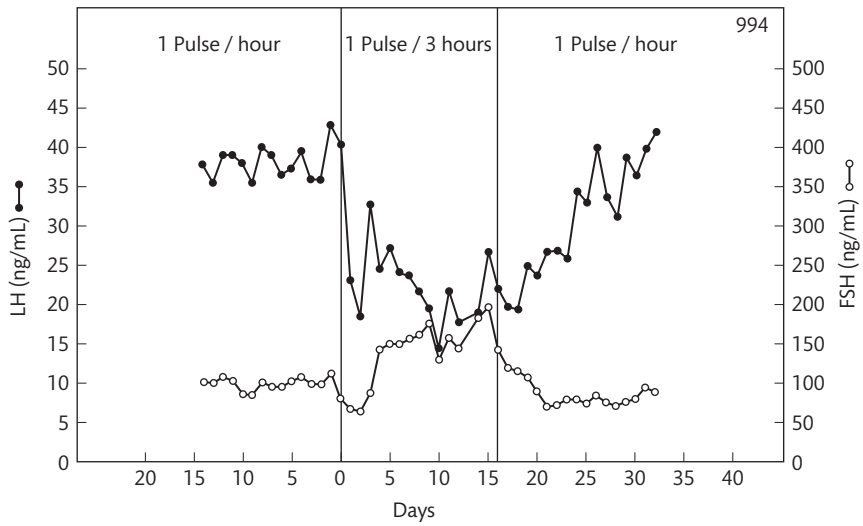


Figure 8.1.1.6 Impact of GnRH pulse frequency on LH and FSH secretion in GnRH-deficient monkeys. These monkeys were gonadectomized (but sex steroid-replaced) and rendered GnRH deficient via radiofrequency ablation of the arcuate nucleus. Normal gonadotropin secretion is maintained when exogenous GnRH was administered in a pulsatile fashion every hour. Decreasing the frequency of GnRH pulse administration to once every 3 hours was followed by a decrease in circulating LH but an increase in circulating FSH concentrations.

Adapted with permission from Wildt L, Hausler A, Marshall G, Hutchison JS, Plant TM, Belchetz PE, Knobil E. Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey. *Endocrinology* 1981; 109:376–85. Copyright © 1981 Oxford University Press (10).

Patterns of GnRH and Gonadotropin Secretion in Women

The hypophyseal portal system is inaccessible in humans, so direct *in vivo* measurements of pulsatile GnRH secretion are unavailable. However, LH pulses identified in the peripheral circulation correspond on a one-to-one basis with GnRH pulses identified in the hypophyseal portal system in animals (Figure 8.1.1.3) [8]; thus, LH (or α-subunit) pulse frequency—as determined by frequent peripheral blood sampling—accurately reflects GnRH pulse frequency. While LH (GnRH) pulse frequency approximates one pulse every 60–90 minutes in the follicular phase, LH pulse frequency decreases to approximately one pulse every 3 to 8 hours during the luteal phase [27, 28]. In open loop conditions (e.g. surgical or natural menopause), LH pulse frequency approximates one pulse per hour [29]. Similarly, isolated human mediobasal hypothalami generate a GnRH pulse every 60 to 100 minutes [23]. Thus, a 'circhoral' (once per hour) pulse frequency may be a fundamental characteristic of the adult GnRH pulse generator, with changes reflecting the application or withdrawal of sex steroid (primarily progesterone) negative feedback. Since high and low GnRH pulse frequency preferentially favours LH and FSH secretion, respectively, cyclic changes in GnRH pulse frequency contribute to differential changes in LH and FSH concentrations across ovulatory cycles [30].

LH pulse amplitude can be modulated by changes of GnRH released per pulse, by changes of gonadotrope responses to GnRH stimulation, or both. LH pulse amplitude also varies inversely with the preceding LH interpulse interval [31]. LH pulse amplitude tends to decrease across the follicular phase as a result of oestradiol negative feedback. However, the very high oestradiol concentrations at midcycle provoke increased gonadotropin release (the midcycle gonadotropin surge). During the luteal phase, LH pulse amplitudes are variable, but generally approximately double the amplitudes observed in the follicular phase.

In normally cycling women, circulating FSH concentrations are highest in the early part of follicular phase, wane across the follicular phase, but increase again as part of the midcycle gonadotropin surge. FSH concentrations remain relatively low during the luteal phase, but increase again at the luteal-follicular transition. The longer half-life of FSH tends to mask pulsatile release.

Negative Feedback Regulation of GnRH and Gonadotropin Secretion

In normally cycling women, changes in circulating oestradiol concentrations reflect follicular development during the follicular phase and corpus luteum function during the luteal phase. In general, oestrogens restrain gonadotropin secretion; and oestrogen deficiency (e.g. natural or surgical menopause) leads to a marked increase in gonadotropin secretion. Oestrogen restraint at the hypothalamus primarily reflects reductions in GnRH pulse amplitude rather than pulse frequency. GnRH pulse amplitude, but not GnRH pulse frequency, are increased in postmenopausal women and in ovariectomized post-pubertal female monkeys, and this is reversed with oestrogen replacement [32]. Low-dose oestradiol administration does not substantively alter LH release in GnRH-deficient women treated with fixed-dose exogenous GnRH pulses [33].

Day-to-day changes in GnRH pulse frequency primarily reflect the imposition and removal of progesterone negative feedback. In the early luteal phase, the increase in progesterone secretion from the corpus luteum slows LH pulse frequency, and exogenous progesterone slows follicular phase LH pulse frequency [34]. This effect of progesterone requires the permissive presence of oestradiol [35, 36], likely through increased hypothalamic progesterone receptor expression [37]. In contrast, androgens appear to reduce hypothalamic progesterone receptor expression and, thus, antagonize progesterone's ability to inhibit GnRH pulse frequency [38, 39]. At the gonadotrope, progesterone augments pituitary gonadotropin responses to GnRH [40, 41].

Sex steroid feedback on GnRH secretion appears to be largely mediated by kisspeptin (KNDy) neurons (Figure 8.1.1.7) [13]. While GnRH neurons express few sex steroid receptors, kisspeptin and KNDy neurons colocalize with oestrogen, progesterone, and androgen receptors. Oestrogen or oestrogen plus progesterone reduces kisspeptin expression in the arcuate nucleus in monkeys, while kisspeptin expression in the infundibular (arcuate) nucleus is increased in sex steroid deficient monkeys and women. Similarly, postmenopausal women demonstrate fewer infundibular

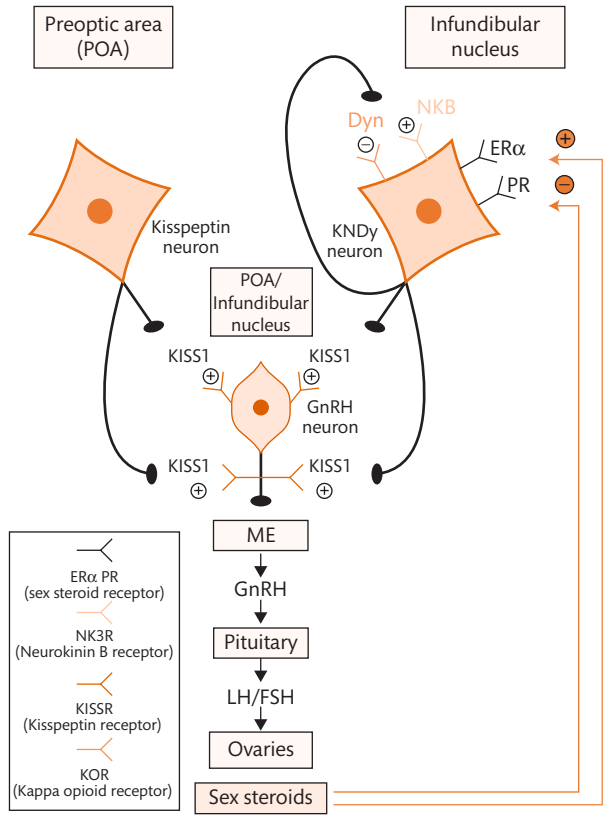


Figure 8.1.1.7 Model of kisspeptin and KNDy signalling to GnRH neurons. Kisspeptin (KISS1) stimulates GnRH neurons. KnDy neurons also release neurokinin B (NKB), which stimulates KISS1 release; and dynorphin (Dyn), which inhibits KISS1 release. ERα, oestrogen receptor alpha; KISS1, kisspeptin; ME, median eminence; POA, preoptic area; PR, progesterone receptor.

Adapted with permission from Skorupskaite K, George JT, Anderson RA. The kisspeptin-GnRH pathway in human reproductive health and disease. *Hum Reprod Update* 2014; 20:485–500. (13). Copyright © 2014, Oxford University Press.

prodynorphin-expressing neurons compared to premenopausal women.

FSH secretion is also selectively impacted by inhibins, activins, and follistatin [42]. Inhibins are heterodimer peptides that selectively inhibit FSH release from pituitary gonadotropes and contribute to differential LH and FSH release across the menstrual cycle. Two inhibin isoforms contain identical α-subunits but different β-subunits (βA for inhibin A, βB for inhibin B). Inhibin B is primarily secreted by ovarian granulosa cells in response to FSH stimulation during the early follicular phase; while inhibin A is primarily secreted by the corpus luteum in response to LH stimulation during the luteal phase. Activins and follistatin are pituitary gonadotrope products that influence FSH secretion via autocrine-paracrine signalling. Activins stimulate production of FSH, while follistatin reduces FSH production by binding activin. Follistatin production varies as a function of GnRH pulse frequency and partly mediates the reduction in FSH production in response to high GnRH pulse frequencies.

Sex Steroid Positive Feedback and the Midcycle Gonadotropin Surge

In contrast to the effects of relatively low oestradiol levels, high oestradiol concentrations at midcycle provoke the midcycle gonadotropin surge, characterized by an approximately tenfold increase in LH and an approximately fourfold increase in FSH levels. While some studies suggest that oestradiol alone can produce normal gonadotropin surges, the modest preovulatory increase in circulating progesterone may coordinate the timing of the surge, increase the duration of the surge, and/or contribute to the midcycle increase in FSH release [43, 44].

Midcycle levels of sex steroids markedly increase gonadotrope responsiveness to GnRH stimulation [45, 46]. Although midcycle surge generation in women is critically dependent on continued GnRH stimulation and can be prevented with GnRH receptor antagonists [47], it remains unclear whether midcycle GnRH surges occur in women. A study involving incomplete GnRH receptor antagonism suggested that GnRH secretion does not increase at midcycle in women [48]. Moreover, pulsatile administration of constant-dose exogenous GnRH—or even reduced dose—produces gonadotropin surges in GnRH-deficient women [46]. Taken together, these findings suggest that sex steroid augmentation of pituitary gonadotropin release is sufficient for gonadotropin surge generation in women.

Physiological Influences on GnRH Secretion

Energy Availability

Biological processes require metabolic energy. In settings of restricted energy availability, maintenance of life-sustaining functions is prioritized. Pregnancy and lactation are metabolically expensive; and since reproduction is not critical for individual survival, it is metabolically gated. Inadequate energy availability suppresses reproductive function: this forms the basis for both nutritional infertility and functional hypothalamic amenorrhea (discussed in Chapter 8.8.2). The neurobiological mechanisms underlying the influence of nutritional status on the reproductive system remain poorly understood. However, inadequate energy availability culminates in impaired GnRH and gonadotropin secretion. The reproductive effects of reduced energy availability appear to be mediated by circulating factors (e.g. leptin and insulin) in addition to complex central neural pathways involving myriad neuropeptides (e.g. kisspeptin) and neurotransmitters [49].

Stress

A 'stressor' is a potential threat to homeostasis (e.g. injury, illness, reduced energy availability, situations that provoke psychological distress, etc.). Both marked acute stress and chronic stress can reduce reproductive function, and in some cases this can be viewed as an appropriate adaptive response. As examples, critical illness is typically characterized by a reversible hypogonadotropic hypogonadism [50]; and psychological stress can play a role in functional hypothalamic amenorrhea [51]. Stress-related impairment of reproductive function primarily relates to suppression of GnRH secretion. The reproductive effects of stress appear to be mediated by corticotropin-releasing hormone (CRH), arginine vasopressin (AVP), and cortisol, in addition to complex central neural pathways involving numerous neuropeptides (e.g. kisspeptin) and neurotransmitters [52].

Lactation

Lactation is often accompanied by amenorrhea and subfertility. In the setting of limited resources, lactational amenorrhea can be viewed as an important adaptation since short interbirth intervals could jeopardize infant well-being. In the setting of lactation, pulsatile GnRH release is slowed, rendering hypogonadotropic hypogonadism. Mechanisms remain unclear, but this in part reflects the very high metabolic cost of milk production and, thus, reduced energy availability for reproductive functions. Reduced GnRH secretion during lactation may partly reflect hyperprolactinemia, and neural mechanisms may also be important since the frequency and duration of suckling appears to impact contraceptive effectiveness.

Circadian Changes and Sleep

Both basal gonadotropin secretion and LH surges may exhibit circadian patterns, but the degree to which this pertains is species dependent. For example, LH surges in female rodents are temporally constrained to the late afternoon hours; but in female monkeys and women, LH surges do not appear to be confined to a specific time of day. Similarly, although women exhibit diurnal changes in gonadotropin secretion, recent studies suggest that LH and FSH secretion does not change markedly when accounting for sleep status, body position, activity level, light exposure, and nutritional cues [53].

Sleep can have a profound influence on pulsatile LH (and by inference GnRH) secretion in humans, and such changes may contribute to normal gonadotropin production across puberty and during the early follicular phase (EFP). Early puberty is characterized by a sleep-associated increase in LH pulse amplitude and frequency: this begins shortly after sleep onset and also occurs during daytime sleep [54]. During puberty, LH pulse initiation appears to be encouraged by slow wave sleep but discouraged by rapid eye movement (REM) sleep [55]. Sleep-related slowing of LH pulse frequency occurs in late pubertal girls and in women, most prominently during the EFP, but also in the late follicular phase. In women

studied during the EFP, LH pulse initiation is encouraged by brief awakenings but discouraged by REM sleep [56]. Naloxone appeared to prevent the sleep-associated decrease in LH pulse frequency in the EFP, suggesting that sleep-related slowing may be mediated by hypothalamic opioids [57].

REFERENCES

1. Stopa EG, Koh ET, Svendsen CN, Rogers WT, Schwaber JS, King JC. Computer-assisted mapping of immunoreactive mammalian gonadotropin-releasing hormone in adult human basal forebrain and amygdala. *Endocrinology*, 1991; **128**: 3199–207.
2. Plant TM, Krey LC, Moossy J, McCormack JT, Hess DL, Knobil E. The arcuate nucleus and the control of gonadotropin and prolactin secretion in the female rhesus monkey (Macaca mulatta). *Endocrinology*, 1978; **102**: 52–62.
3. Herde MK, Iremonger KJ, Constantin S, Herbison AE. GnRH neurons elaborate a long-range projection with shared axonal and dendritic functions. *J Neurosci*, 2013; **33**: 12689–97.
4. Prevot V. Glial-neuronal-endothelial interactions are involved in the control of GnRH secretion. *J Neuroendocrinol*, 2002; **14**: 247–55.
5. Stamou MI, Cox KH, Crowley WF, Jr. Discovering genes essential to the hypothalamic regulation of human reproduction using a human disease model: adjusting to life in the "-omics" era. *Endocr Rev*, 2015; **36**: 603–21.
6. Quinton R, Hasan W, Grant W, *et al*. Gonadotropin-releasing hormone immunoreactivity in the nasal epithelia of adults with Kallmann's syndrome and isolated hypogonadotropic hypogonadism and in the early midtrimester human fetus. *J Clin Endocrinol Metab*, 1997; **82**: 309–14.
7. Millar RP, Lu ZL, Pawson AJ, Flanagan CA, Morgan K, Maudsley SR. Gonadotropin-releasing hormone receptors. *Endocr Rev*, 2004; **25**: 235–75.
8. Moenter SM, Brand RM, Midgley AR, Karsch FJ. Dynamics of gonadotropin-releasing hormone release during a pulse. *Endocrinology*, 1992; **130**: 503–10.
9. Santoro N, Filicori M, Crowley WF, Jr. Hypogonadotropic disorders in men and women: diagnosis and therapy with pulsatile gonadotropin-releasing hormone. *Endocr Rev*, 1986; **7**: 11–23.
10. Wildt L, Hausler A, Marshall G, *et al*. Frequency and amplitude of gonadotropin-releasing hormone stimulation and gonadotropin secretion in the rhesus monkey. *Endocrinology*, 1981; **109**: 376–85.
11. Belchetz PE, Plant TM, Nakai Y, Keogh EJ, Knobil E. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. *Science*, 1978; **202**: 631–3.
12. Lehman MN, Coolen LM, Goodman RL. Minireview: kisspeptin/neurokinin B/dynorphin (KNDy) cells of the arcuate nucleus: a central node in the control of gonadotropin-releasing hormone secretion. *Endocrinology*, 2010; **151**: 3479–89.
13. Skorupskaite K, George JT, Anderson RA. The kisspeptin-GnRH pathway in human reproductive health and disease. *Hum Reprod Update*, 2014; **20**: 485–500.
14. Plant TM. A comparison of the neuroendocrine mechanisms underlying the initiation of the preovulatory LH surge in the human, Old World monkey and rodent. *Front Neuroendocrinol*, 2012; **33**: 160–8.
15. Grachev P, Millar RP, O'Byrne KT. The role of neurokinin B signalling in reproductive neuroendocrinology. *Neuroendocrinology*, 2014; **99**: 7–17
16. Wehrenberg WB, Wardlaw SL, Frantz AG, Ferin M. Beta-endorphin in hypophyseal portal blood: variations throughout the menstrual cycle. *Endocrinology*, 1982; **111**: 879–81.
17. Gindoff PR, Jewelewicz R, Hembree W, Wardlaw S, Ferin M. Sustained effects of opioid antagonism during the normal human luteal phase. *J Clin Endocrinol Metab*, 1988; **66**: 1000–4.
18. Skrapits K, Borsay BA, Herczeg L, Ciofi P, Liposits Z, Hrabovszky E. Neuropeptide co-expression in hypothalamic kisspeptin neurons of laboratory animals and the human. *Front Neurosci*, 2015; **9**: 29.
19. Leon S, Tena-Sempere M. Dissecting the roles of gonadotropin-inhibitory hormone in mammals: studies using pharmacological tools and genetically modified mouse models. *Front Endocrinol (Lausanne)*, 2015; **6**: 189.
20. Thiery JC, Pelletier J. Multiunit activity in the anterior median eminence and adjacent areas of the hypothalamus of the ewe in relation to LH secretion. *Neuroendocrinology*, 1981; **32**: 217–24.
21. Knobil E. The electrophysiology of the GnRH pulse generator in the rhesus monkey. *J Steroid Biochem*, 1989; **33**: 669–71.
22. Claypool LE, Watanabe G, Terasawa E. Effects of electrical stimulation of the medial basal hypothalamus on the *in vivo* release of luteinizing hormone-releasing hormone in the prepubertal and peripubertal female monkey. *Endocrinology*, 1990; **127**: 3014–22.
23. Rasmussen DD, Gambacciani M, Swartz W, Tueros VS, Yen SS. Pulsatile gonadotropin-releasing hormone release from the human mediobasal hypothalamus *in vitro*: opiate receptor-mediated suppression. *Neuroendocrinology*, 1989; **49**: 150–6.
24. Maeda K, Ohkura S, Uenoyama Y, Wakabayashi Y, Oka Y, Tsukamura H, Okamura H. Neurobiological mechanisms underlying GnRH pulse generation by the hypothalamus. *Brain Res*, 2010; **1364**: 103–15.
25. Seminara SB, Messager S, Chatzidaki EE, *et al*. The GPR54 gene as a regulator of puberty. *N Engl J Med*, 2003; **349**: 1614–27.
26. Stamatiades GA, Kaiser UB. Gonadotropin regulation by pulsatile GnRH: signaling and gene expression. *Mol Cell Endocrinol*, 2018; **463**: 131–41.
27. Reame N, Sauder SE, Kelch RP, Marshall JC. Pulsatile gonadotropin secretion during the human menstrual cycle: evidence for altered frequency of gonadotropin-releasing hormone secretion. *J Clin Endocrinol Metab*, 1984; **59**: 328–37.
28. Filicori M, Santoro N, Merriam GR, Crowley WF, Jr. Characterization of the physiological pattern of episodic gonadotropin secretion throughout the human menstrual cycle. *J Clin Endocrinol Metab*, 1986; **62**: 1136–44.
29. Gill S, Lavoie HB, Bo-Abbas Y, Hall JE. Negative feedback effects of gonadal steroids are preserved with aging in postmenopausal women. *J Clin Endocrinol Metab*, 2002; **87**: 2297–302.
30. Marshall JC, Kelch RP. Gonadotropin-releasing hormone: role of pulsatile secretion in the regulation of reproduction. *N Engl J Med*, 1986; **315**: 1459–68.
31. O'Dea LS, Finkelstein JS, Schoenfeld DA, Butler JP, Crowley WF, Jr. Interpulse interval of GnRH stimulation independently modulates LH secretion. *Am J Physiol*, 1989; **256**: E510–15.
32. Chongthammakun S, Terasawa E. Negative feedback effects of estrogen on luteinizing hormone-releasing hormone release occur in pubertal, but not prepubertal, ovariectomized female rhesus monkeys. *Endocrinology*, 1993; **132**: 735–43.

33. Marshall JC, Case GD, Valk TW, Corley KP, Sauder SE, Kelch RP. Selective inhibition of follicle-stimulating hormone secretion by estradiol. Mechanism for modulation of gonadotropin responses to low dose pulses of gonadotropin-releasing hormone. *J Clin Invest*, 1983; **71**: 248–57.
34. Soules MR, Steiner RA, Clifton DK, Cohen NL, Aksel S, Bremner WJ. Progesterone modulation of pulsatile luteinizing hormone secretion in normal women. *J Clin Endocrinol Metab*, 1984; **58**: 378–83.
35. Karsch FJ, Weick RF, Hotchkiss J, Dierschke DJ, Knobil E. An analysis of the negative feedback control of gonadotropin secretion utilizing chronic implantation of ovarian steroids in ovariectomized rhesus monkeys. *Endocrinology*, 1973; **93**: 478–86.
36. Nippoldt TB, Reame NE, Kelch RP, Marshall JC. The roles of estradiol and progesterone in decreasing luteinizing hormone pulse frequency in the luteal phase of the menstrual cycle. *J Clin Endocrinol Metab*, 1989; **69**: 67–76.
37. Romano GJ, Krust A, Pfaff DW. Expression and estrogen regulation of progesterone receptor mRNA in neurons of the mediobasal hypothalamus: an in situ hybridization study. *Mol Endocrinol*, 1989; **3**: 1295–300.
38. Eagleson CA, Gingrich MB, Pastor CL, *et al.* Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab*, 2000; **85**: 4047–52.
39. Pielecka J, Quaynor SD, Moenter SM. Androgens increase gonadotropin-releasing hormone neuron firing activity in females and interfere with progesterone negative feedback. *Endocrinology*, 2006; **147**: 1474–9.
40. Nippoldt TB, Khoury S, Barkan A, Kelch RP, Marshall JC. Gonadotrophin responses to GnRH pulses in hypogonadotrophic hypogonadism: LH responsiveness is maintained in the presence of luteal phase concentrations of oestrogen and progesterone. *Clin Endocrinol (Oxf)*, 1987; **26**: 293–301.
41. Couzinet B, Brailly S, Bouchard P, Schaison G. Progesterone stimulates luteinizing hormone secretion by acting directly on the pituitary. *J Clin Endocrinol Metab*, 1992; **74**: 374–8.
42. Bilezikjian LM, Justice NJ, Blackler AN, Wiater E, Vale WW. Cell-type specific modulation of pituitary cells by activin, inhibin and follistatin. *Mol Cell Endocrinol*, 2012; **359**: 43–52.
43. Chang RJ, Jaffe RB. Progesterone effects on gonadotropin release in women pretreated with estradiol. *J Clin Endocrinol Metab*, 1978; **47**: 119–25.
44. Liu JH, Yen SS. Induction of midcycle gonadotropin surge by ovarian steroids in women: a critical evaluation. *J Clin Endocrinol Metab*, 1983; **57**: 797–802.
45. Wang CF, Lasley BL, Lein A, Yen SS. The functional changes of the pituitary gonadotrophs during the menstrual cycle. *J Clin Endocrinol Metab*, 1976; **42**: 718–28.
46. Hall JE, Martin KA, Whitney HA, Landy H, Crowley WF, Jr. Potential for fertility with replacement of hypothalamic gonadotropin-releasing hormone in long term female survivors of cranial tumors. *J Clin Endocrinol Metab*, 1994; **79**: 1166–72.
47. Kolp LA, Pavlou SN, Urban RJ, Rivier JC, Vale WW, Veldhuis JD. Abrogation by a potent gonadotropin-releasing hormone antagonist of the estrogen/progesterone-stimulated surge-like release of luteinizing hormone and follicle-stimulating hormone in postmenopausal women. *J Clin Endocrinol Metab*, 1992; **75**: 993–7.
48. Hall JE, Taylor AE, Martin KA, Rivier J, Schoenfeld DA, Crowley WF, Jr. Decreased release of gonadotropin-releasing hormone during the preovulatory midcycle luteinizing hormone surge in normal women. *Proc Natl Acad Sci U S A*, 1994; **91**: 6894–8.
49. Manfredi-Lozano M, Roa J, Tena-Sempere M. Connecting metabolism and gonadal function: novel central neuropeptide pathways involved in the metabolic control of puberty and fertility. *Front Neuroendocrinol*, 2018; **48**: 37–49.
50. Spratt DI, Cox P, Orav J, Moloney J, Bigos T. Reproductive axis suppression in acute illness is related to disease severity. *J Clin Endocrinol Metab*, 1993; **76**: 1548–54.
51. Pauli SA, Berga SL. Athletic amenorrhea: energy deficit or psychogenic challenge? *Ann N Y Acad Sci*, 2010; **1205**: 33–8.
52. Acevedo-Rodriguez A, Kauffman AS, Cherrington BD, Borges CS, Roepke TA, Laconi M. Emerging insights into hypothalamic-pituitary-gonadal (HPG) axis regulation and interaction with stress signaling. *J Neuroendocrinol*, 2018; **30**: e12590.
53. Klingman KM, Marsh EE, Klerman EB, Anderson EJ, Hall JE. Absence of circadian rhythms of gonadotropin secretion in women. *J Clin Endocrinol Metab*, 2011; **96**: 1456–61.
54. McCartney CR. Maturation of sleep-wake gonadotrophin-releasing hormone secretion across puberty in girls: potential mechanisms and relevance to the pathogenesis of polycystic ovary syndrome. *J Neuroendocrinol*, 2010; **22**: 701–9.
55. Shaw ND, Butler JP, McKinney SM, Nelson SA, Ellenbogen JM, Hall JE. Insights into puberty: the relationship between sleep stages and pulsatile LH secretion. *J Clin Endocrinol Metab*, 2012; **97**: E2055–62.
56. Hall JE, Sullivan JP, Richardson GS. Brief wake episodes modulate sleep-inhibited luteinizing hormone secretion in the early follicular phase. *J Clin Endocrinol Metab*, 2005; **90**: 2050–5.
57. Rossmanith WG, Yen SS. Sleep-associated decrease in luteinizing hormone pulse frequency during the early follicular phase of the menstrual cycle: evidence for an opioidergic mechanism. *J Clin Endocrinol Metab*, 1987; **65**: 715–18.

8.1.2 Ovarian and Uterine Development from Fetal Life to Puberty

Terhi Piltonen and Juha Tapanainen

Sex Organ Development

The development of embryonic gonad towards gender specific phenotype is a sequential and complex process. In humans, fetal

organogenesis starts 4 weeks post conception and is completed by the gestational week (GW) 9. After this, a rapid growth will follow and the fetus will maturate for organ functions. The primordial germ cells can be identified in the wall of the yolk sac at GW 3 onwards. At this bipotential state the germ cells migrate into the dorsal mesentery of the hindgut in the urogenital ridges of undifferentiated gonads (Figure 8.1.2.1). The karyotype of the embryo (46,XX or 46,XY) and the presence of the sex determining region on the Y-chromosome (SRY) determines whether the gonads will develop into an ovary or testis. Female or male gonads can be distinguished as early as by the GW 8 [1, 2]. Furthermore, several other genes like SRY box-related gene 9 (SOX9), Wilms tumour gene 1 (WT1), EMX2, LIM1, and steroidogenic factor 1 (SF-1) in males and WNT-4, RSPO-1, FOXL-2, and DAX-1 in females, have also been identified coordinating the gonadal development [3, 4]. After the differentiation process, the gonadal hormones, mainly dihydrotestosterone converted from testosterone and anti-Müllerian hormone (AMH, Müllerian inhibiting substance, MIF), have significant role in the development of genitalia (Figure 8.1.2.1). The Müllerian ducts will develop as a female tubule system, the uterus and fallopian tubes, whereas the Wolffian ducts will become epididymis and ductus dereference. In the presence or in absence of testosterone the urogenital sinus will form as male or female external genitalia.

Embryological Development of the Ovarian Follicle Pool and the Transient Activation of the Ovaries

In females, after migration to gonadal ridges, the primordial germ cells start proliferating through mitosis into oogonial stem cells that enter meiosis at GW 22, arresting at the diplotene stage to become primary oocytes. Only part of the oogonia will undergo meiosis having potential to ovulate eventually after puberty while the rest of them will undergo apoptosis. The oocytes that have reached the diplotene stage start associating with clusters of somatic cells that form pre-granulosa cells, forming primordial follicles that reach the highest number of 7 million around mid-gestation (Figure 8.1.2.2). After the saturation point, the number of primordial follicles starts to decrease by means of apoptosis and only 2 million primordial follicles are left at the time of birth.

The levels of follicle stimulating hormone (FSH) and luteinizing hormone (LH) in fetal circulation peak at mid-gestation and decrease by the end of the third trimester [5, 6]. The meaning of this rise in gonadotrophins is not well known, however, it is thought that this enhanced gonadotrophin secretion results into occasional follicle development in the fetal ovary, supported by the concomitant follicular expression of AMH that can also be measured from

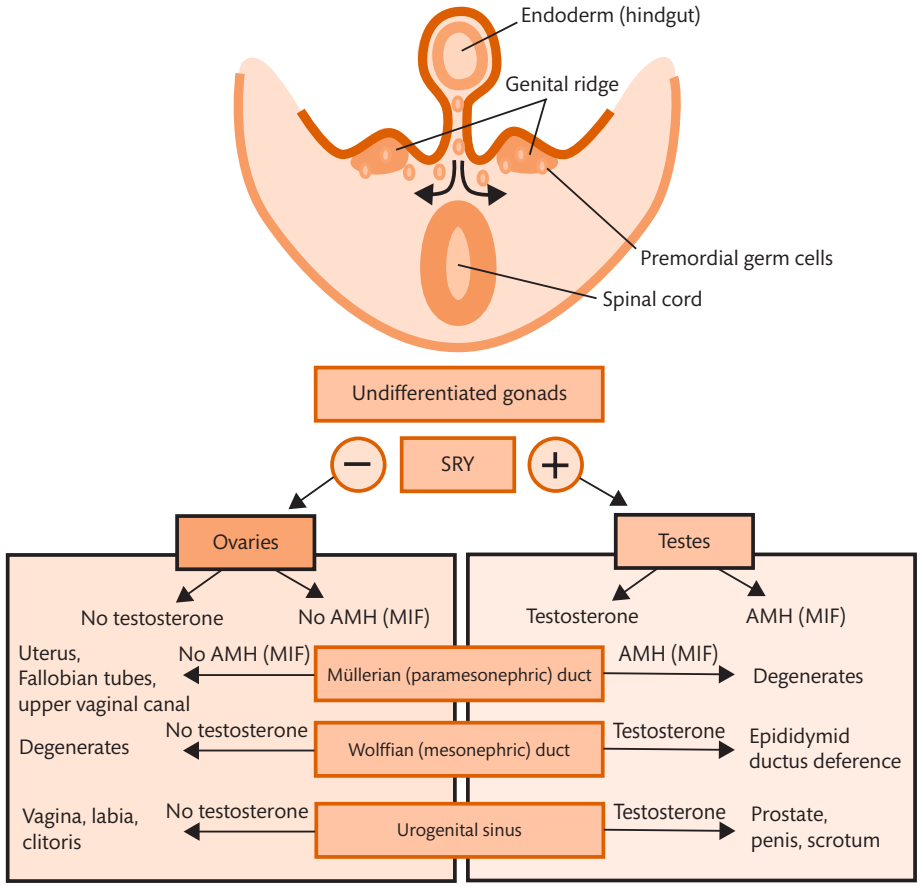


Figure 8.1.2.1 The sex organ development in human. At the bipotential state the germ cells migrate into the dorsal mesentery of the hindgut in the urogenital ridges. The karyotype of the embryo (46,XX or 46,XY) and the presence of the sex determining region on the Y-chromosome (SRY) determines whether the gonads will develop into an ovary or testis.

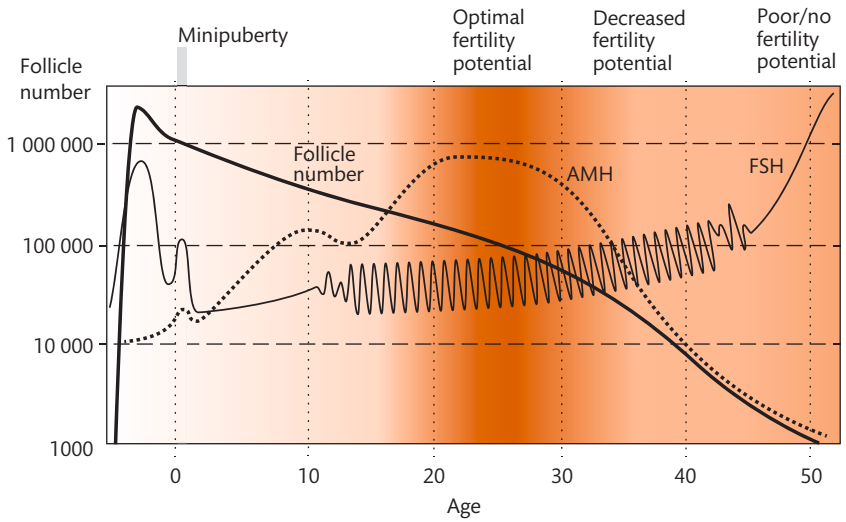


Figure 8.1.2.2 The number of ovarian follicles and the fluctuations in serum AMH and FSH levels throughout female reproductive life span. Primordial follicles reach the highest number of 7 million at gestational week 20. Through apoptosis the number of follicles decrease and only around 2 million are left at the time of birth. Thereafter, the follicle number decreases rapidly with 400 000 being present at the onset of the puberty. The follicle pool is diminished around age of 50 years leading into menopause.

the fetal circulation [7]. The levels of gonadotrophins decrease gradually towards the term of the pregnancy.

The withdrawal of placental steroids and thereby the lack of inhibitory feedback at birth induce a transient activation of hypothalamus–pituitary axis during the first months of infancy. This results into a short-term and transient release of gonadotrophins from the pituitary of the neonate, activating follicular development and proliferation of granulosa cells reflected by the rise in serum AMH levels. This phenomenon, known as minipuberty, is dependent on the fetal age at birth, prematurity delaying the onset [7–9]. After the postnatal ovarian activation, gonadotrophins and thereby sex steroid hormone and AMH levels decrease and remain low until puberty. The quiescent period ends before puberty as a consequence of maturation of the hypothalamus–pituitary–ovarian axis (Figure 8.1.2.2).

Initiation of the Puberty and Menstrual Cycles

Sensitization of the hypothalamus-pituitary-axis to ovarian steroid hormones and maturation of feedback mechanisms initiate the menstrual cycles (menarche) at puberty between the ages of 10 and 16, the mean age of the menarche being 12–13 years. During the uterine maturations, the tubal form of the uterus will change into a round adult shape under oestrogen influence and the endometrium will start proliferating. At this point, the ovarian follicle pool has further diminished beholding only around 400 000 follicles that will serve as oocyte reserve for ovulatory cycles during female reproductive life (Figure 8.1.2.2). After menarche, the menstrual cycles are irregular due to incomplete maturation of the hypothalamus–pituitary–ovarian axis, and it may take over 2 years for it to mature and the cycles to become regular. Eventually, approximately 400 follicles will ovulate during the reproductive life span. As the follicle pool diminishes, also the menstrual cycles become irregular. At menopause, around the age of 50 years, only a few hundred or possibly thousands of follicles are left as the follicle pool is exhausted.

REFERENCES

1. Sun YC, Sun XF, Dyce PW, Shen W, Chen H. The role of germ cell loss during primordial follicle assembly: a review of current advances. *Int J Biol Sci*, 2017; **13**: 449–57.
2. Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. *Endocr Rev*, 1996; **17**: 121–55.
3. Sinisi AA, Pasquali D, Notaro A, Bellastella A. Sexual differentiation. *J Endocrinol Investig*, 2003; **26**: 23–8.
4. Robboy SJ, Kurita T, Baskin L, Cunha GR. New insights into human female reproductive tract development. *Differentiation*, 2017; **97**: 9–22
5. Debieve F, Beerlandt S, Hubinont C, Thomas K. Gonadotropins, prolactin, inhibin A, inhibin B, and activin A in human fetal serum from midpregnancy and term pregnancy. *J Clin Endocrinol Metab*, 2000; **85**: 270–4.
6. Massa G, de Zegher F, Vanderschueren-Lodeweyckx M. Serum levels of immunoreactive inhibin, FSH, and LH in human infants at preterm and term birth. *Biol Neonate*, 1992; **61**: 150–5.
7. Kuiri-Hanninen T, Kallio S, Seuri R, *et al*. Postnatal developmental changes in the pituitary-ovarian axis in preterm and term infant girls. *J Clin Endocrinol Metab*, 2011; **96**: 3432–9.
8. Tapanainen J, Kellokumpu-Lehtinen P, Pelliniemi L, Huhtaniemi I. Age-related changes in endogenous steroids of human fetal testis during early and midpregnancy. *J Clin Endocrinol Metab*, 1981; **52**: 98–102
9. Tapanainen J, Koivisto M, Vihko R, Huhtaniemi I. Enhanced activity of the pituitary-gonadal axis in premature human infants. *J Clin Endocrinol Metab*, 1981; **52**: 235–8.

8.1.3 Menstrual Cycle and Ovulation

Gurkan Bozdag, Baris Ata, and Engin Türkgeldi

Introduction

The vast majority of our knowledge about follicle development had been retrieved from animal studies conducted before the 1950s [1]. Although initial attempts to understand human folliculogenesis has begun with histologic examinations of ovaries derived from post-mortem examination or oophorectomy, recent imaging studies and endocrinologic assessments have made significant improvements throughout the process of understanding the whole human menstrual cycle and related folliculogenesis. Nowadays, we can appreciate the fact that each follicle is distinct and controlled by endocrine and diverse paracrine factors that will determine its own fate [2].

According to studies performed on early human fetus, follicular development begins as early as the fourth month of fetal life [3]. In the early life of female, oogonia initiates the first meiotic division and arrests in the dictyate stage to generate the primary oocytes and then is surrounded by somatic cells. At the end of 20 weeks of gestation, number of those follicles containing oocytes constitutes true ovarian follicular reserve and reproductive potential with an estimated number of 7 million. Of them, only 400 become a Graafian follicle and rest deplete during reproductive ages (see Chapter 8.1.2). The current chapter aims to summarize the journey of primordial follicles while aiming to reach the stage of dominant follicle and its corresponding endometrial changes.

Follicle Activation and Growth

In the context of follicle activation, it is noteworthy to mention that medical or surgical hypophysectomy has been shown to be associated with decreased ovarian weight and a significant reduction in the total number of developing follicles in animal studies [4]. On the contrary, induction of those rats with an amount of follicle stimulating hormone (FSH) yields an improvement in ovarian weight and initiates follicle development from preantral stage up to the antral stage. Although development to the antral stage is not solely dependent on FSH as shown in FSH-null mice [5], preantral follicles are responsive to FSH treatment [2]. Those findings suggest that FSH has a critical role for the regulation of preantral follicle growth accompanied by oocyte and granulosa cell derived paracrine factors [2]. From the view of clinical practice, it might be postulated that a prolonged stimulation of FSH, longer than 2 weeks might eventually induce follicle development from preantral stages.

Oocyte–Granulosa Cell Interaction

Under the subheading of '*granulosa cell derived*' paracrine factors, one should mention about C-type natriuretic peptide (CNP) as an autocrine/paracrine follicle-stimulating agent secreted during secondary and antral follicle stages in response to FSH stimulation [6]. It activates its receptor natriuretic peptide receptor-B or 2 as expressed in granulosa cells of secondary follicles and signal via the production of cGMP to evoke desensitization reflected by dephosphorylation in the kinase-homology domain [7]. Whereas stimulation of cGMP production induces follicle development from primary and early secondary to the late secondary stages, it inhibits meiotic resumption of the oocyte [8].

Other than CNP, a large group of diverse peptide/protein ligands derived by granulosa cells have been explored to modulate follicle growth [2]. Of those activins, anti-Müllerian hormone (AMH) and bone morphogenetic protein-6 (BMP-6) act throughout RSK receptors. AMH is a glycoprotein from TGF-β family that is secreted by granulosa cells within 2–5 mm antral stage follicles and decreases when the follicles begin their FSH-dependent cyclic recruitment process [9, 10]. The main action of AMH is avoiding transition from primordial follicle to primary follicle and hence inhibiting the FSH-dependent cyclical recruitment [9].

It appears that not only granulosa cells but also oocyte itself is an active member of folliculogenesis. R-spondin2 [11], growth differentiation factor-9 (GDF9) and bone morphogenetic protein-15 (BMP15) [12] have been shown to promote granulosa cell growth as '*oocyte-derived*' structures. A diminished level of R-spondin2 in primary and larger follicles was noticed to be associated with a failure in follicle development during late reproductive life and premature ovarian insufficiency [13]. Although they are lacking in primordial follicles, *in-vivo* administration of R-spondin agonist might stimulate the development of primary follicles to the antral stage in animal models [11]. GDF-9 is also an important local factor for the development of follicles beyond the primary stage and promotes androgen production from theca cells [14]. Similarly, BMP15 is another potent stimulator of granulosa cell proliferation and yields different actions among species.

Steroids and Gonadotropin Receptors

In addition to local peptide hormones, oestrogens, androgens, and progesterone play important role for the regulation of folliculogenesis via specific receptors. Whereas oestrogen receptor-alpha variants and theca cell-specific deletion of oestrogen receptor-alpha leads to premature ovarian failure [15], deletion of androgen receptors may yield subfertility [16]. Notably, local increment in the concentration of oestrogens induces follicle growth and contributes to the selection of dominant follicle. Dihydrotestosterone, as an androgen, increases FSH receptor expression and provokes oestrogen and progesterone biosynthesis [2]. On the contrary, high levels are associated with follicular atresia.

As might be clearly understood from the aforementioned studies, the full process for detecting the dominant follicle is highly

complicated. The total duration of time to reach a preovulatory stage takes around 85 days. Although FSH cannot control every step of folliculogenesis, without any circulating FSH, particularly follicles 2–5 mm in diameter undergo atresia. During the lifespan of a Graafian follicle, with parallel to the molecular mechanisms given here, the initial change is morphological alteration, as granulosa cells timely become more cuboidal. Nevertheless, for the activation of a primordial follicle, BMP acts as a promoter. Nerve growth factor, brain derived neurotropic factor, neutrophin-3 and 4/5 are various neurotrophins that may alter the early follicular development. After the first awakening of the primordial follicles, FSH receptors also arise in number [17]. The decline in previous luteal phase both in the process of steroidogenesis and production of inhibin A from granulosa cells, a significant rise in concentration of FSH becomes evident particularly during the mid-late luteal phase.

Two-Cell, Two-Gonadotropin System

The next step after the awakening of a primordial follicle is formation of multilayer of granulosa cells surrounding the oocyte to form the preantral follicle stage (Figure 8.1.3.1). While zona pellucida appears around the oocyte, the outer stroma modifies to theca cells. FSH receptors not only induce proliferation of granulosa cells but also act to induce the function of aromatase enzyme that converts androgens to oestrogens. Hence, local increase of oestrogen concentration steadily increases the number of FSH receptors on granulosa cells. Although low dose of androgens are essential as a precursor of aromatase activity and induction of the enzyme, high doses are associated with overproduction of 5-alpha androgens that in turn inhibit the enzyme activity and causes atresia.

As the layers of granulosa cells increase and fluid accumulation begins within a preantral follicle, the later stage of those follicles is designated as antral follicle. During this process, granulosa cells around the oocyte differ from the ones that cover the whole follicle with its functionality and called as cumulus oophorus. The content of the fluid within an antral follicle essentially consists of oestrogen and hence concentration of the steroids reflects the functional capacity of granulosa and theca cells within the context of two-cell, two-gonadotropin system. Insulin like growth factor-II produced in theca and granulosa cells also stimulate aromatase activity. One should also address that LH receptors are present only theca cells, and solely granulosa cells have FSH receptors. LH secretion is responsible for the production of androgens, namely androstenedione and testosterone which in turn metabolize to estrone and oestradiol with aromatase activity, respectively. Notably, inhibin induces and activin stimulates action of FSH, LH, and IGF-I. These findings might suggest that LH is also important for the late development of folliculogenesis and maturation. Nevertheless, within the context of selection of dominant follicle, a fall in FSH levels due to increased concentration of oestrogen decreases the aromatase activity and associated oestrogen synthesis. Impairment in granulosa cell will lead to a significant reduction in aromatase activity and cause more androgenic microenvironment and finally atresia. Briefly, containing the highest potency of aromatase activity and LH receptors with response to FSH appears to be the crucial factor for a given follicle in order to be attained as a dominant follicle in the later phase of the menstrual cycle [17].

Novel Pathways of Follicle Activation

One of the novel discovered pathways in the process of folliculogenesis appears to be '*Hippo signalling*' that is also essential for organ size control. As an incision is made to the ovary, G-actin within granulosa cells polymerize to F-actin and inhibits several negative growth regulators acting in a serine/threonine kinase cascade that ultimately phosphorylates and inactivates key transcriptional coactivators, Yes-associated protein and transcriptional co-activator with PDZ-binding motif [2]. These alterations induce expression of downstream CCN growth factors and BIRC apoptosis inhibitors that in turn activates dormant follicles.

Another molecular mechanism to induce dormant follicles appears to be related with '*PTEN-PI3K-AKT pathway*'. Stimulation of phosphatidylinositol-3-kinase (PI3K) yields AKT stimulation that subsequently suppresses FOXO3 actions in nucleus to promote primordial follicle growth. Meanwhile, PTEN gene encodes an enzyme that converts PIP3 to PIP2, thus damping the actions of PI3K. Oocyte-specific deletion of the PTEN gene leads to global activation

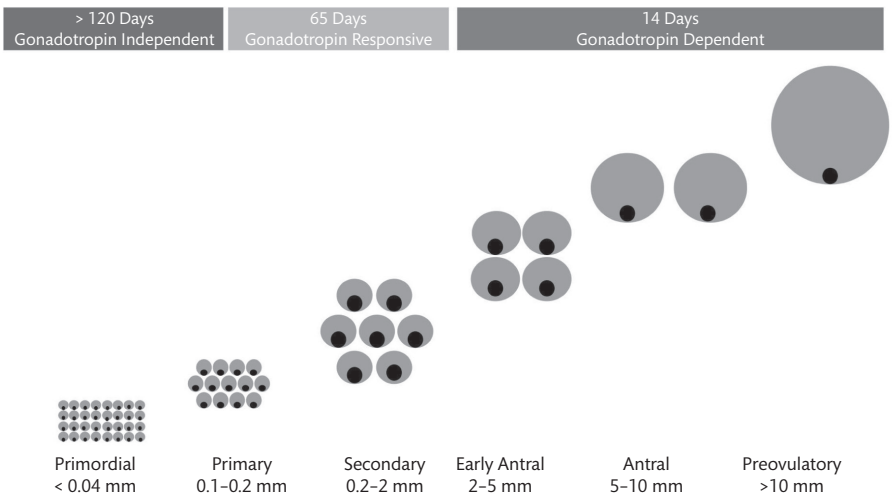


Figure 8.1.3.1 Phases and durations of follicular development.

of primordial follicles [18] whereas treatment with PTEN inhibitors promotes primordial follicle activation [19].

Early Follicular Phase

During early follicular phase, activin has a critical role with respect to an improvement in aromatase activity with parallel to FSH exposition and simultaneous inhibition of production of androgens. Activin has also a role in boosting the receptivity of FSH. According to imaging studies, recent theories of ovarian follicular recruitment proposed that there are two or more cohorts of antral follicles during the early phases of menstrual cycle (Figure 8.1.3.2). Of interest, whereas the dominant follicle develops from the final wave, other waves of cohorts do not produce an ovulatory follicle [1]. Those findings suggest that Graafian follicle selection might be observed more than once during natural menstrual cycles [20] rather than being once in the early follicular phase of the menstrual cycle, as mentioned earlier [21]. The competency of corpus luteum also seems to be similar in women with two versus three waves or in women with major versus minor waves preceding the ovulatory wave [22]. However, the appearance of second wave in the late luteal phase or early follicular phase, rather than in mid-luteal phase, facilitates the selection of a dominant follicle [20]. The functional status of dominant follicles that develop during anovulatory waves preceding the ovulatory wave in women is not fully understood; but available data suggest that dominant follicles of anovulatory waves may exhibit different physiologic characteristics than dominant follicles of ovulatory waves in women [1].

In the follicular phase, there are substantial changes in the endometrium with respect to increased levels of oestrogen and enhanced receptivity. All components of the endometrium including glandular structures, stromal cells, and vascular tissues present proliferation. Two-thirds of its upper part constitutes the functional layer, the remaining inner part makes up the basal layer. According to imaging studies, an average of 5 mm increase is observed in the thickness of endometrium. Ciliogenesis is also initiated 7/8th day of the cycle [17].

Mid-Late Follicular Phase

During mid-late follicular period, namely preovulatory phase, dominant follicle grows up with a rate of 1–4 mm per day and reaches to the maximal diameter of 30 mm. There is a steady increase in the aromatase activity and hence oestradiol levels but a contradictory decrease in fluidic AMH concentration. IGF-II continues to induce aromatase activity and oocyte-derived substances. GDF-9 and BMP-15 are involved in the process of follicle development and cumulus expansion [23]. There are also increased numbers of LH receptors, which will be used during ovulation, and time of triggering.

Ovulation

In a regular cycle of 28 days, ovulation occurs in mid-cycle, around the 14th day. In this short but defining period, two important events take place. First, the oocyte resumes meiosis and is released from the ovary along with a cumulus complex. Second, transition from follicular phase to luteal phase of the cycle is observed. The key step for both of these events is the mid-cycle LH surge.

Oestradiol rises constantly in the follicular period and reaches significant levels at the late follicular phase. Beyond a certain threshold, the inhibitory effect of oestradiol on GnRH is reversed and a positive feedback mechanism takes over. High oestradiol levels increase the sensitivity of the anterior hypophysis to GnRH, leading to increased secretion of LH and FSH, which in turn results in production of more oestradiol. An oestradiol level of around 200 pg/ml^{-1} for a minimum of 48 hours and some progesterone activity is required to induce this positive feedback mechanism [24]. Some researchers believe that increased inhibin A levels also play a role in the LH surge [25].

The rise in the serum progesterone levels contributes to the increased responsiveness of LH to GnRH. Progesterone is the dominant hormone in the luteal phase, however, studies show that progesterone starts to rise just prior to LH surge in ovulatory women. This elevation is believed to contribute to the increased sensitivity of LH to GnRH. While the rise in progesterone is relatively

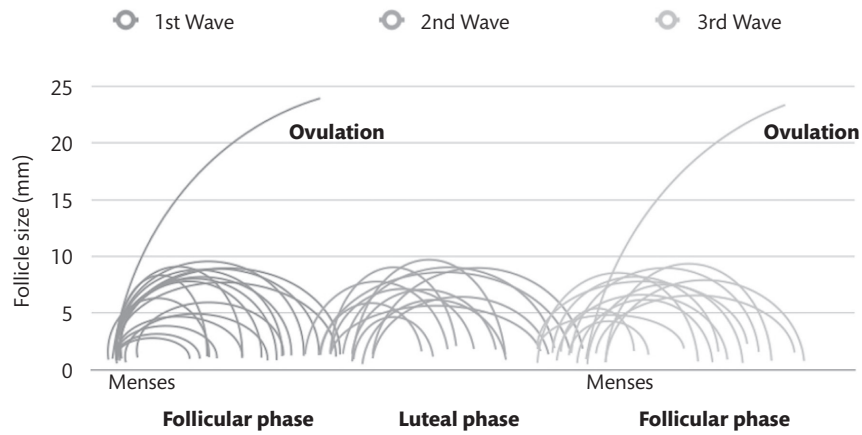


Figure 8.1.3.2 A schematic representation of a three-wave cycle. Each line represents growth and regression of a follicle.

Adapted with permission from Baerwald AR, Adams GP, Pierson RA. Ovarian antral folliculogenesis during the human menstrual cycle: a review. *Human reproduction update* 2012;18:73–91. Copyright © 2011, Oxford University Press.

small, it is essential for the LH surge. The surge is delayed or totally inhibited in patients receiving antiprogestin RU486 [26].

The positive feedback loop climaxes with the discharge of nearly all LH stored in the anterior hypophysis, causing 10- and 4-fold increase in LH and FSH levels, respectively [27]. The surge lasts about 40–48 hours, resulting in meiotic maturation and releasing of the cumulus-oocyte complex. During and after the LH surge, progesterone levels keep rising, whereas oestrogen levels show a significant and swift drop. This drop is what limits the LH surge, since progesterone alone cannot support the LH sensitivity to GnRH without oestradiol.

The LH surge blocks cGMP production in granulosa cells consequently decreases the cAMP levels in the oocyte and meiosis, which was arrested for years, resumes [28]. Then, second meiotic metaphase stage is completed, and first polar body is extruded. At this point, meiosis once more waits for completion, this time to be achieved with fertilization. One of the consequences of the LH surge is the initiation of a series of reactions that will lead to release of the cumulus-oocyte complex. The initial step is the expansion of the cumulus cells, also known as mucification. A group of epidermal growth factors (EGFs), namely amphiregulin, epiregulin, and betacellulin, cause secretion of glycoproteins from cumulus cells that result in the formation of an extracellular hyaluronan-rich matrix [29]. This results in the dispersion of the cumulus cells and significant expansion of the cumulus-oocyte complex [30]. Mucification is believed to be important in the release of the cumulus-oocyte complex from the ovary and its transport to the Fallopian tube.

Another event preceding the release of the oocyte-cumulus complex is the induction of prostaglandin production by the aforementioned progesterone rise prior to LH surge. The prostaglandins, produced by the follicle wall, stimulates the release of proteases that first thins and then perforates the apex of the follicle wall, an area known as macula pellucida. Thus, the cumulus-oocyte complex is released from the ovary along with 2–3 ml follicular fluid approximately 36 hours after the LH surge.

There appears to be no histologically significant change in the endometrium during ovulation; however, LH surge and ovulation are the key events that result in the transition from the proliferative phase to the secretory phase of the endometrium.

Luteal Phase and Menstruation

Unless pregnancy occurs, luteal phase begins with ovulation and lasts around 14 days. Mainly, it consists of two parts referred as luteinization and luteolysis.

Macroscopically, the ruptured follicle fills with fibrin clot and undergoes a series of changes. after ovulation. Theca and granulosa cells luteinize and form a cell complex with increased expression of LH receptors. Angiogenic and inflammatory changes contribute to this new formation as well [31]. The formation of corpus luteum now constitutes the principal unit in the ovary to support luteal phase.

Luteinization begins with ovulation and lasts about 7–9 days. During this period, corpus luteum is highly sensitive to LH. While LH pulses in this period are relatively less frequent compared to the preovulatory stage, they are still regular enough to induce significant production of progesterone and oestradiol [32]. Along with these two major hormones, 17-hydroxyprogesterone, estrone, inhibin A, and relaxin are also secreted from the corpus luteum and increase during luteinization.

If conception does not occur, luteolysis begins. This is marked with the regression of the corpus luteum and the gradual decline in progesterone and oestradiol levels. While a LH pulse occurs every 90–100 minutes in the follicular phase and almost every hour in the preovulatory period, their frequency drops to a pulse every 3–5 hours in the mid-luteal phase [33]. The most prominent mechanism behind the drop of pulse frequencies of LH and FSH in the luteal phase is the inhibitory effect of progesterone on the hypothalamus. It provokes production of an endogen opioid, β-endorphin, which reduces the secretion of GnRH, resulting in more sparse LH pulses as progesterone increases [34]. As these pulses get irregular and less frequent in the late luteal phase, the corpus luteum starts to degenerate and eventually undergoes apoptosis.

Two different pathways keep FSH levels low throughout the luteal phase. The first one is the inhibitory effect of increased oestradiol and inhibin A on FSH. The second is the reduction of GnRH pulses due to the increased production of β-endorphin by increased progesterone levels [35]. As the corpus luteum degenerates in the absence of human chorionic gonadotropin (hCG) or regular LH pulses, both of these pathways lose their inhibitory effect on FSH. Thus, towards the end of luteolysis, FSH starts to increase and the first step for initiating the recruitment of follicles in the next cycle is taken.

If conception occurs, luteolysis and apoptosis of the corpus luteum is avoided by hCG that is secreted by the fetal trophoblastic cells. LH and hCG share the same β-subunit, which allows hCG to act as an LH substitute by binding to LH receptors on the corpus luteum. hCG levels increase dramatically with the rapid growth of fetal trophoblastic cells as pregnancy advances, and this not only maintains but also increases the production of progesterone, oestradiol, and other ovarian steroids by the corpus luteum. This increased workload is visible as the size of the corpus luteum is doubled by the sixth gestational week due to hypertrophy of its cells [36]. Endometrial decidua is stabilized and supported by the progesterone secreted from the corpus luteum until the eighth gestational week. At this week, the placenta is developed enough to produce sufficient progesterone to support endometrium by itself, and gradually takes over this duty from the corpus luteum. This is known as the luteoplacental shift.

With regard to the luteal phase of the ovary, the secretory phase of the endometrium begins with the LH surge and ovulation. It can be divided into three periods, namely, early secretory, mid-secretory, and premenstrual phases.

In the early secretory phase, glycogen builds up in glandular epithelial cells, shifting the nuclei centrally. It is interesting to note that mitotic activity in the endometrium is observed almost exclusively in the first 3 days following ovulation [37].

Coiled arteries that can extend as endometrium thickens, known as spiral arteries, develop in the mid-secretory phase. Concomitantly, endometrial glands attain their peak activity and stroma takes an oedematous character. During this period, endometrium thickness may reach up to 8–10 mm as a maximum. In anticipation of a possible pregnancy, stromal cells proliferate around

blood vessels, natural killer cells accumulate around them, and a complex series of immunological, haematological, and metabolic changes take place in order to promote any possible implantation.

In the premenstrual phase, secretory activity is reduced due to decreased progesterone and oestrogen levels. The secretory endometrium cannot maintain itself and degrades. Inflammatory cells invade the endometrium as catabolic events that result in shrinkage of the endometrium follow. Finally, intermittent contraction of the arterioles in the endometrium results in local ischaemia and leads to shedding of the functional layer of the endometrium, manifesting itself as menstrual bleeding.

REFERENCES

1. Baerwald AR, Adams GP, Pierson RA. Ovarian antral folliculogenesis during the human menstrual cycle: a review. *Hum Reprod Update*, 2012; **18**: 73–91.
2. Hsueh AJ, Kawamura K, Cheng Y, Fauser BC. Intraovarian control of early folliculogenesis. *Endocr Rev*, 2015; **36**: 1–24.
3. Baker TG. A quantitative and cytological study of germ cells in human ovaries. *Proc R Soc Lond B Biol Sci*, 1963; **158**: 417–33.
4. McGee EA, Perlas E, LaPolt PS, Tsafriri A, Hsueh AJ. Follicle-stimulating hormone enhances the development of preantral follicles in juvenile rats. *Biol Reprod*, 1997; **57**: 990–8.
5. Abel MH, Wootton AN, Wilkins V, Huhtaniemi I, Knight PG, Charlton HM. The effect of a null mutation in the follicle-stimulating hormone receptor gene on mouse reproduction. *Endocrinology*, 2000; **141**: 1795–803.
6. Hobbs A, Foster P, Prescott C, Scotland R, Ahluwalia A. Natriuretic peptide receptor-C regulates coronary blood flow and prevents myocardial ischemia/reperfusion injury: novel cardioprotective role for endothelium-derived C-type natriuretic peptide. *Circulation*, 2004; **110**: 1231–5.
7. Schulz S. C-type natriuretic peptide and guanylyl cyclase B receptor. *Peptides*, 2005; **26**: 1024–34.
8. Norris RP, Ratzan WJ, Freudzon M, *et al.* Cyclic GMP from the surrounding somatic cells regulates cyclic AMP and meiosis in the mouse oocyte. *Development*, 2009; **136**: 1869–78.
9. Dewailly D, Robin G, Peigne M, Decanter C, Pigny P, Catteau-Jonard S. Interactions between androgens, FSH, anti-Mullerian hormone and estradiol during folliculogenesis in the human normal and polycystic ovary. *Hum Reprod Update*, 2016; **22**: 709–24.
10. Weenen C, Laven JS, Von Bergh AR, *et al.* Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum Reprod*, 2004; **10**: 77–83.
11. Cheng Y, Kawamura K, Takae S, *et al.* Oocyte-derived R-spondin2 promotes ovarian follicle development. *FASEB J*, 2013; **27**: 2175–84.
12. Juengel JL, Bodensteiner KJ, Heath DA, *et al.* Physiology of GDF9 and BMP15 signalling molecules. *Anim Reprod Sci*, 2004; 82–3: 447–60.
13. Bell SM, Schreiner CM, Hess KA, Anderson KP, Scott WJ. Asymmetric limb malformations in a new transgene insertional mutant, footless. *Mech Dev*, 2003; **120**: 597–605.
14. Vitt UA, McGee EA, Hayashi M, Hsueh AJ. *In vivo* treatment with GDF-9 stimulates primordial and primary follicle progression and theca cell marker CYP17 in ovaries of immature rats. *Endocrinology*, 2000; **141**: 3814–20.
15. Lee S, Kang DW, Hudgins-Spivey S, Krust A, Lee EY, Koo Y *et al.* Theca-specific estrogen receptor-alpha knockout mice lose fertility prematurely. *Endocrinology*, 2009; **150**: 3855–62.
16. Sen A, Hammes SR. Granulosa cell-specific androgen receptors are critical regulators of ovarian development and function. *Mol Endocrinol*, 2010; **24**: 1393–403.
17. Fritz MA, Speroff L. Regulation of the menstrual cycle. In: Fritz MA, Speroff L, eds. *Clinical Gynecologic Endocrinology and Infertility.* Philadelphia, PA: Lippincott Williams & Wilkins, Wolters Kluwer Business, 2011: 199–242.
18. Reddy P, Liu L, Adhikari D, *et al.* Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. *Science*, 2008; **319**: 611–13.
19. Li J, Kawamura K, Cheng Y, *et al.* Activation of dormant ovarian follicles to generate mature eggs. *Proc Natl Acad Sci U S A*, 2010; **107**: 10280–4.
20. Baerwald AR, Adams GP, Pierson RA. Characterization of ovarian follicular wave dynamics in women. *Biol Reprod*, 2003; **69**: 1023–31.
21. Gougeon A, Lefevre B. Evolution of the diameters of the largest healthy and atretic follicles during the human menstrual cycle. *J Reprod Fertil*, 1983; **69**: 497–502.
22. Baerwald AR, Adams GP, Pierson RA. Form and function of the corpus luteum during the human menstrual cycle. *Ultrasound Obstet Gynecol*, 2005; **25**: 498–507.
23. Gueripel X, Brun V, Gougeon A. Oocyte bone morphogenetic protein 15, but not growth differentiation factor 9, is increased during gonadotropin-induced follicular development in the immature mouse and is associated with cumulus oophorus expansion. *Biol Reprod*, 2006; **75**: 836–43.
24. Feher J. Female reproductive physiology. In: Feher J, ed. *Quantitative Human Physiology*. Boston, MA: Academic Press, 2012: 846–55.
25. Hall JE, Schoenfeld DA, Martin KA, Crowley WF, Jr. Hypothalamic gonadotropin-releasing hormone secretion and follicle-stimulating hormone dynamics during the luteal-follicular transition. *J Clin Endocrinol Metab*, 1992; **74**: 600–7.
26. Batista MC, Cartledge TP, Zellmer AW, Nieman LK, Loriaux DL, Merriam GR. The antiprogestin RU486 delays the midcycle gonadotropin surge and ovulation in gonadotropin-releasing hormone-induced cycles. *Fertil Steril*, 1994; **62**: 28–34.
27. Hall JE. Neuroendocrine control of the menstrual cycle. In: Barbieri RL, ed. *Yen and Jaffe's Reproductive Endocrinology*, 8th edn. Philadelphia, PA: Elsevier, 2019: 149–66.e5.
28. Conti M, Chang RJ. Folliculogenesis, ovulation, and luteogenesis. In: Groot LJD, Kretser DMd, Giudice LC, *et al.*, eds. *Endocrinology: Adult and Pediatric*, 7th edn. Philadelphia, PA: W.B. Saunders, 2016: 2179–91.e3.
29. Conti M, Hsieh M, Zamah AM, Oh JS. Novel signaling mechanisms in the ovary during oocyte maturation and ovulation. *Mol Cell Endocrinol*, 2012; **356**: 65–73.
30. Richards JS, Pangas SA. The ovary: basic biology and clinical implications. *J Clin Investig*, 2010; **120**: 963–72.
31. Brannstrom M, Norman RJ. Involvement of leukocytes and cytokines in the ovulatory process and corpus luteum function. *Hum Reprod*, 1993; **8**: 1762–75.
32. Nippoldt TB, Khoury S, Barkan A, Kelch RP, Marshall JC. Gonadotrophin responses to GnRH pulses in hypogonadotrophic hypogonadism: LH responsiveness is maintained in the presence of luteal phase concentrations

of oestrogen and progesterone. *Clin Endocrinol*, 1987; **26**: 293–301.
33. Djahanbakhch O, Warner P, McNeilly AS, Baird DT. Pulsatile release of LH and oestradiol during the periovulatory period in women. *Clin Endocrinol*, 1984; **20**: 579–89.
34. Ropert JF, Quigley ME, Yen SS. Endogenous opiates modulate pulsatile luteinizing hormone release in humans. *J Clin Endocrinol Metab*, 1981; **52**: 583–5.
35. Marshall JC. The menstrual cycle and disorders of ovulation. In: Groot LJD, Kretser DMD, Giudice LC, *et al*., eds. *Endocrinology: Adult and Pediatric*, 7th edn. Philadelphia, PA: W.B. Saunders, 2016: 2231–41.e3.
36. Davis JS, Rueda BR. The corpus luteum: an ovarian structure with maternal instincts and suicidal tendencies. *Front Biosci*, 2002; **7**: d1949–78.
37. Lessey BA, Young SL. Structure, function, and evaluation of the female reproductive tract. In: Strauss JF, Barbieri RL, eds. *Yen and Jaffe's Reproductive Endocrinology*, 8th edn. Philadelphia, PA: Elsevier, 2019: 206–47.e13.

8.2

Evaluation of the Female Patient with Suspected Reproductive Endocrine Disorders

8.2.1 Clinical Evaluation of Patients with Suspected Reproductive Endocrine Disorders

Rachel Roberts, Steve Franks, and Channa Jayasena

Introduction

Female reproductive endocrine disorders are common, and can result in significant morbidity for women due to their impact on fertility, in addition to physical, psychological, and sexual consequences. It is therefore important that women with menstrual disorders are recognized, receive thorough clinical evaluation, and have access to appropriate services, to facilitate prompt diagnosis so that hormone levels are maintained and the potential health consequences on fertility, bone, and cardiovascular health are minimized. Women with reproductive endocrine disorders are most likely to present with either menstrual irregularity or infertility, or, more rarely, with a physical, psychological, or sexual consequence of the underlying condition such as acne, depression, or dyspareunia. In women of reproductive age who present with difficulty in conceiving following at least 12 months of regular unprotected intercourse thorough clinical evaluation is warranted [1]. Earlier assessment after six months may be appropriate in women over 35 years, or in those with a known history of oligo- or amenorrhea, known or suspected uterine, tubal, or peritoneal disease, or moderate to severe endometriosis. In cases of infertility it is essential that both partners are evaluated as 35% of infertile couples have combined male and female factor infertility [2]. It is also important to remember that multiple factors may impair fertility in an individual couple, which may be both male and female factor. Furthermore, because both partners are affected by decisions surrounding fertility investigation and treatment, it is optimal to see couples together in clinic. Fertility evaluation may also benefit women planning to conceive using sperm from a known or anonymous donor. A thorough history and physical examination is essential in the assessment of reproductive disorders. Clinical assessment may help to suggest the cause of reproductive disorders, and allow for more focused investigations and diagnostic evaluation. Diagnosis enables appropriate treatment to be sought, which is particularly important to minimize long-term health consequences and given the time-sensitive treatment goal of pregnancy, if desired.

History

A detailed history is arguably the most powerful diagnostic tool in determining the underlying cause of female reproductive disorders. A good history will enable the clinician to perform a relevant examination and to target investigations to confirm a likely diagnosis. The history is also important to identify women who have modifiable risk factors, such as disordered eating or excessive exercise. In women with infertility, it is necessary to establish the duration that a couple has been trying to conceive and to review the results of any previous evaluation and treatments. It is also necessary to enquire about childhood development; sexual development during puberty; menstrual history; sexual history; contraceptive history; previous obstetric history; illnesses and infections; surgeries; medications used; exposure to certain environmental agents and social and personal history. Table 8.2.1.1 summarizes some of the key clinical findings for different causes of female reproductive endocrine disorders.

Childhood and pubertal development: Determining a woman's childhood and adolescent development may reveal abnormalities in growth and/or pubertal development, which in turn may help to identify the underlying cause of menstrual disturbance. Patients may have experienced primary amenorrhea, most frequently caused by

Table 8.2.1.1 Common symptoms and signs of conditions causing infertility

	Symptoms	Signs
Hypothalamic amenorrhea	Oligo/amenorrhea Weight fluctuations Eating disorders Strict eating habits and dieting Stress History of excessive exercise Previous fractures (e.g. stress fracture) Dyspareunia Poor libido Substance abuse Family history of eating and reproductive disorders	Low BMI Lanugo Mottled, cool extremities Hypercarotenaemia Signs of self-harm Hypotension Bradycardia Atrophic breasts Atrophic vaginitis
Hyperprolactinaemia	Oligo/amenorrhea Headaches Galactorrhoea Poor libido Visual disturbance History of relevant medications e.g. antidepressants	Scotoma or bitemporal hemianopia Breast secretions
Hyperthyroidism	Oligo/amenorrhea Heat intolerance Weight loss Increased appetite Irritability Mood changes Tremor Eye complaints	Tremor Hyperkinesis Goitre, bruit Warm vasodilated peripheries Tachycardia or atrial fibrillation Eye signs incl. exophthalmos, lid lag, 'stare', ophthalmoplegia
Hypothyroidism	Menorrhagia Lethargy Depression Weight gain Cold intolerance Dry skin and or hair History of thyroid disease or surgery	Mental slowness Dry, thin hair Goitre Neck scars Hypertension Bradycardia Slow relaxing reflexes
PCOS	Oligo/amenorrhea Weight gain Acne Hirsutism Family history of PCOS	Raised BMI with central adiposity Hirsutism Acne Acanthosis nigricans Skin tags Hypertension
Obesity	Oligo/amenorrhea Weight fluctuations Eating disorders Eating behaviours and diet	High BMI Central adiposity Acanthosis nigricans Hypertension
Primary ovarian insufficiency	Oligo/amenorrhea Poor libido Dyspareunia Previous fractures (e.g. stress fracture) Hot flushes Urogenital symptoms (e.g. frequency) Family history of POI/early menopause History of radiation therapy or chemotherapy	Often no signs Atrophic vaginitis Signs of other autoimmune disease (e.g. goitre, vitiligo)
Turner's syndrome	Primary or secondary amenorrhea Pubertal delay Hearing loss	Short stature Webbed neck Shield chest with widely spaced nipples Cubitus valgus Low posterior hairline Short fourth and fifth metacarpals Hypertension Abnormality of cardiac examination (e.g. murmur)
Late-onset congenital adrenal hyperplasia (21-hydroxylase deficiency)	Oligo/amenorrhea History of precocious puberty Acne Hirsutism	Short stature Hirsutism Acne Androgenic alopecia Hypertension Clitoral hypertrophy

Table 8.2.1.1 *Continued*

	Symptoms	Signs
Androgen-secreting tumour	Hirsutism Acne Rapid onset of symptoms	Hirsutism Acne Androgenic alopecia Hypertension Clitoral hypertrophy Deepening of voice
XY karyotype-Androgen Insensitivity Syndrome	Primary amenorrhea Lower abdominal pain History of gonadal tumours	Tall stature Thin Clitoral hypertrophy Underdeveloped labia Blind vaginal pouch Normal breast development Underdeveloped pale nipples Absent/sparse pubic and axillary hair
XY karyotype-gonadal dysgenesis	Primary amenorrhea Lower abdominal pain History of gonadal tumours	Tall stature Thin Female external genitalia Underdeveloped breasts Narrow pelvis and hips Sparse pubic hair
Cushing's syndrome	Oligo/amenorrhea Weight gain Change in appearance Depression Poor libido Thin skin/easy bruising Acne Muscular weakness Polyuria/polydipsia History of fractures History of steroid use	High BMI/central adiposity Moon face Plethora Acne Hirsutism Thin skin Bruising Striae Buffalo hump (dorsal fat pad) Proximal myopathy Visual fields Hypertension
Acromegaly	Oligo/amenorrhea Headaches Change in appearance Vision changes	Skin changes—thickening, skin tags, hirsutism, hyperhidrosis Peripheral vision loss Enlarged jaw Frontal bossing Macroglossia Visual field defects

primary ovarian insufficiency (e.g. Turner's syndrome), congenital deficiency of gonadotrophin releasing hormone (GnRH) (see next) or, much more rarely, anatomical abnormalities such as Müllerian agenesis [3]. Delayed or absent pubertal development may be a manifestation of congenital hypogonadotropic hypogonadism, a group of heterogeneous disorders typically caused by GnRH deficiency [4] or may be attributable to intersex disorders [5].

Menstrual history: Women with reproductive endocrine disorders predominantly present with menstrual disturbance, and the most common cause of female infertility is ovulatory dysfunction [2]. It is important to establish the regularity and duration of menses. A history of regular and consistent menses is highly suggestive of normal ovulatory function. In most ovulatory women, menstrual cycles are regular and predictable, occurring, on average, every 25–35 days, and should have consistent flow and characteristics. Some degree of variation in menses is normal, however; a study of more than 1000 menstrual cycles found that 56% of women experienced variations in intermenstrual interval exceeding 5 days within 6 months and in 75% of women within 1 year [6]. Assessing for regular symptoms of premenstrual molimina (breast tenderness, ovulatory pain, bloating) is also useful in suggesting a patient is likely to be ovulating regularly [7]. Irregular and infrequent menstrual periods (oligomenorrhea) usually signals irregular ovulation, and is defined as fewer than six to eight menstrual periods per year. Although it does not make pregnancy impossible, it can reduce the opportunities to conceive and clearly has an impact on fertility. Absent menstrual periods (amenorrhea) signals an absence of ovulation resulting in infertility, and can be classified as either primary or, more commonly, secondary amenorrhea. There are multiple causes of oligo- and amenorrhea, many of which are shared, which a full history and examination can help to determine. It is also important to ascertain the duration of menstrual disturbance, as this can influence what investigations and management is appropriate.

Identifying the flow of a woman's menstrual periods and the presence of abnormal uterine bleeding is another essential component of a menstrual history, and is of particular importance in women presenting with infertility, who may have a purely gynaecological condition, rather than a reproductive endocrine disorder. Abnormal uterine bleeding including menorrhagia can be a symptom of uterine polyps or fibroids, hypothyroidism, cancer, or perimenopause. The presence of painful menstrual periods (dysmenorrhoea) can also help to identify conditions that may affect fertility, such as endometriosis, pelvic inflammatory disease, fibroids, and

adenomyosis. If these conditions are suspected, further assessment with gynaecological review is indicated.

Many women record the timing of their menstrual cycle and characteristics of menstrual bleeding, especially in those having menstrual irregularity or attempting conception. Traditionally a paper record or diary may be kept; however, several electronic software applications have been developed to assist women in tracking their menstrual cycles, which may be accessible on computers and smartphones for minimal or no cost. It may be useful to look at these during a clinic appointment to give an accurate overview of a woman's menstrual patterns. Some women may chart serial basal body temperature (BBT) to help track ovulation and monitor their menstrual cycles. BBT is usually measured first thing in the morning before getting out of bed and in ovulatory cycles is meant to show a biphasic pattern, with a woman's temperature rising after ovulation; whereas anovulatory cycles typically result in monophasic patterns. However, BBT is an unreliable method of ovulation detection as some ovulatory women do not have clear biphasic BBT patterns [8] and interpretation of the charts can be difficult and subject to wide interobserver variation [9]. Furthermore, several cohort studies that evaluated the use of BBT to time intercourse reported no improvement in the chance of natural conception [10, 11]. Women should therefore be advised that charting BBT is generally not recommended.

Sexual history: The sexual history of the couple is a sensitive, yet important, element of clinical evaluation of women with infertility. Frequency and timing of intercourse as well as sexual dysfunction should be explored, since infrequent or ineffective intercourse can be an explanation for infertility. Spontaneous pregnancy rates are reported to be highest during the two-to-six days immediately prior to ovulation [12, 13]. However, there are limitations to timing intercourse in couples with infertility: ovulation can be difficult to predict in women with irregular menses; furthermore, timed intercourse has been reported to be emotionally stressful for couples [14]. Therefore, couples that are concerned about their fertility should be advised that vaginal sexual intercourse every 2 to 3 days optimizes the chance of pregnancy [15]. Women should be specifically asked about pain during sexual intercourse (dyspareunia), as it is rarely self-reported by patients and often only revealed after direct questioning [16]. Dyspareunia may contribute to infertility either by reducing the frequency of coitus, or because the underlying aetiology negatively affects fertility. For example, superficial dyspareunia may be caused by vaginal atrophy, as seen in hypo-oestrogenic conditions such as GnRH deficiency, hypothalamic amenorrhea, or premature ovarian insufficiency, whereas deep dyspareunia may be related to endometriosis or pelvic inflammatory disease. The use of sexual lubricants should also be assessed. *In vitro* studies have shown that vaginal lubricants can have a detrimental effect on sperm motility [17, 18], although the evidence for this is mixed [19]. It is possible that vaginal lubricant use during intercourse may negatively impact conception rates, particularly when the male partner has reduced sperm quality. For this reason, it is generally advised that common vaginal lubricants are not used during sexual intercourse, in couples trying to conceive [20]. It is also important to discuss previous sexually transmitted infections, and their treatment, as well as the number of sexual partners; the number of which increases the incidence of undiagnosed sexually transmitted diseases. *Chlamydia trachomatis* and *Neisseria gonorrhoea* are both important causes in the development of pelvic inflammatory disease, salpingitis, and subsequent tubal infertility.

Contraceptive history: It is important to ascertain the use of any previous contraceptive methods and duration since ceasing contraception. Progesterone-only injectable contraceptives have been reported to cause amenorrhea lasting up to 1 year following cessation of therapy [21, 22]. There is no clear evidence to suggest that oral contraceptives, progesterone-only subdermal implants, copper intrauterine devices or the levonorgestrel intrauterine system cause a delay in return of menses or fertility [23]. The use of a copper intrauterine device for contraception has historically been linked to increased rates of pelvic inflammatory disease and subsequent tubal infertility [24]. However this link is controversial, with recent evidence showing no increased risk of tubal infertility in copper intrauterine device users [25]. Women should therefore be counselled that the potential risks of pelvic inflammatory disease or infertility among intrauterine device users is likely to be very low.

Endocrine history: Symptoms of oestrogen deficiency such as hot flushes, vaginal dryness, dyspareunia, poor libido, depression, loss of concentration, and urogenital symptoms, may represent primary ovarian insufficiency or hypothalamic amenorrhea. Young women have the highest risk of developing thyrotoxic disorders such as Graves' disease, so it is important to screen for features such as hyperphagia, heat intolerance, palpitations, diarrhoea, tremor, mood disturbance, and insomnia. Symptoms of hirsutism and acne are associated with hyperandrogenic states most often caused by polycystic ovarian syndrome; however, other hyperandrogenic disorders include ovarian or adrenal tumours (often rapid onset of symptoms), congenital adrenal hyperplasia, and Cushing's syndrome. While Cushing's syndrome and acromegaly can commonly cause menstrual disturbance and infertility [26, 27], they usually present with other classical symptoms, which may include weight gain, changes in facial appearance, bruising, acne, weakness, polyuria and polydipsia, headaches, mood changes, and changes in vision. Pituitary tumours may disrupt gonadal function, either because of mass effect, or because of the effects of hormonal secretion from the tumour [28]. Symptoms may include changes in vision, headaches, or change in sense of smell. In addition, symptoms of hormone excess may be present if the tumour is functioning. Surgical treatment or irradiation of pituitary tumours may also result in infertility from disruption of gonadotrophic cells or resulting hypopituitarism [29, 30]. Radiation therapy for non-pituitary brain tumours is also associated with a significant risk of hypothalamic-pituitary dysfunction, and subsequent infertility [31]. The development of hypogonadism following cranial radiotherapy is dose-related and therefore a specific history of radiotherapy treatment is necessary [32]. Galactorrhoea and decreased libido are signs of hyperprolactinaemia. A history of traumatic brain injury can be associated with altered GnRH function and hypothalamic amenorrhea [33]. Infections and infiltrative disorders such as sarcoidosis, haemachromatosis or, tuberculosis can damage the hypothalamus or pituitary resulting in disrupted menstruation and infertility [34–37]. Diabetes mellitus may cause menstrual irregularities, oligomenorrhoea or secondary amenorrhoea [38].

Other medical history: Chronic illness or ill-health *per se* may disrupt menstrual cycles, which can be associated with impaired female fertility. Chronic kidney disease and renal replacement therapy

are associated with hyperprolactinaemia [39], ovulatory dysfunction, amenorrhoea, and reduced fertility [40]. Malabsorptive diseases, such as coeliac disease or inflammatory bowel disease, can cause reproductive changes including delayed onset of menarche, amenorrhea, impaired fertility, and adverse pregnancy outcomes, possibly related to autoimmune-mediated mechanisms or nutritional deficiency. Infertility may be the initial clinical feature of coeliac disease in women who are otherwise asymptomatic from a gastrointestinal perspective [41]. Women with Turner's syndrome may have symptoms of hearing loss, or valvular heart disease. It is widely recognized that cytotoxic chemotherapy drugs and pelvic irradiation can result in premature ovarian failure; it is therefore essential to elicit a past history of any cancer and cancer treatment, including childhood cancer [42, 43].

Surgical history: A surgical history is also of importance when evaluating a woman with a potential reproductive disorder. A history of total or partial thyroidectomy may be associated with ongoing thyroid hormone dysfunction that can adversely affect fertility [44]. A history of fractures, particularly stress fractures, may point to women whose subfertility is caused by a condition resulting in oestrogen deficiency, such as hypothalamic amenorrhea or primary ovarian insufficiency. It is also important to note previous abdominal and pelvic surgery such as appendicectomy or ovarian cystectomy, which predispose to adhesion formation and increase the likelihood of tubal dysfunction. Unilateral oophorectomy does not generally cause reduced fertility in young women, since they have many primordial follicles per ovary; however, prior unilateral oophorectomy may impact fertility in older women as they may develop diminished ovarian reserve sooner than women with two ovaries [45].

Obstetric and gynaecological history: In addition to a thorough menstrual history, a full obstetric and gynaecological history should be assessed for events potentially associated with subsequent infertility or with an adverse outcome in a future pregnancy. If necessary, further clinical evaluation and investigation by a gynaecologist may be appropriate. Important points to discuss are a history of abnormal smears, colposcopy, cervical surgery (including cone biopsies) or endometrial or cervical malignancies, which may cause cervical stenosis, a history of pelvic inflammatory disease or endometriosis and previous pregnancies and their outcome including ectopic pregnancies and their treatment, miscarriages and their management, molar pregnancies, and termination of pregnancies. A history of significant peripartum haemorrhage in previous pregnancies, or insufficient postpartum breast milk production, is classical for Sheehan's syndrome, which is caused by hypoperfusion and subsequent ischaemic necrosis of the pituitary gland [46]. Genital tuberculosis is rare in countries with high mean income, but it should be carefully considered in women from other countries, where it may be a more common cause of female infertility [47]. Clinical symptoms of genital tuberculosis are varied, but include chronic lower abdominal or pelvic pain, vaginal bleeding, menstrual irregularity, general malaise, and infertility.

Personal and social history: Several aspects of personal and lifestyle history may contribute to reproductive disorders and profoundly affect fertility, including age, weight, diet, exercise, smoking, alcohol, illicit drug use and occupation. Personal and lifestyle history can also help to determine whether any modifiable risk factors are present, and if so, to provide appropriate advice and support. It is important to specifically enquire whether trends in body weight and eating patterns mirror changes in reproductive function such as menstrual disturbance. Women with obesity have an increased risk of ovulatory dysfunction when compared with women with lower body mass index, even after adjusting for menstrual irregularity [48]. Participating in a group programme involving exercise, dietary advice and support improves pregnancy rates more than weight loss alone [49]. Low body weight or recent weight loss may signify hypothalamic amenorrhoea. Women with hypothalamic amenorrhoea have a higher than normal risk of having a previous eating disorder (whether or not formerly diagnosed). Restoration of body weight and reduction in exercise intensity may be sufficient to restore menses and improve rates of conception in some patients with hypothalamic amenorrhoea [50]. It is crucial that the subgroup of women with anorexia nervosa or bulimia are detected and referred for specialist care including psychiatric evaluation. High intensity strenuous exercise can also lead to disrupted menstrual cycles or amenorrhea, so this is another important lifestyle factor to explore while taking a history [51]. There is substantial evidence that female smokers have reduced fertility [52], and there is also an association between passive smoking in women and delayed conception [53]. Furthermore, it is widely known that maternal smoking in pregnancy is associated with increased risks for the infant. Women should therefore be strongly advised against smoking, and referred to appropriate smoking cessation intervention groups. Furthermore, it is important to realize that conception planning represents an effective window of opportunity to support and motivate smoking cessation in women [54]. There is inconsistent evidence about the impact of drinking alcohol on female fertility. However, it is known that excessive alcohol consumption is harmful to a developing fetus, and current advice from the UK Chief Medical Officer is to abstain from alcohol consumption while trying to conceive and when pregnant [55]. Drugs of abuse such as cocaine and marijuana can negatively impact ovulation and tubal function [56, 57]. A specific enquiry about occupation may also reveal potential aetiological factors in female infertility. Shift work, intense physical workload and long working hours have all been associated with reduced fecundability and prolonged time to conception [58, 59]. Furthermore, chemical substances used in lamp, wood, shoe, rubber, and textile manufacturing have been reported to negatively impact female fertility, although further detailed research is required [60].

Psychiatric/Psychological history: The belief that psychological factors play a role in menstrual disturbance and infertility is long-standing. In the seventeenth century, Richard Morten described that psychiatric illness and infertility were two cardinal features of anorexia nervosa [61]. It is important for a clinician to explore potential stressors in females presenting with menstrual disturbance or infertility, such as work- or study-related stress, personal stress, in addition to psychiatric disorders such as anxiety and depression. Psychological stress or psychiatric symptoms may contribute to the cause of infertility; be the consequence of infertility; or both. There is evidence that preconception stress, as measured by salivary alpha-amylase, is associated with a longer time-to-pregnancy and an increased risk of infertility [62]. Psychological stress is a common cause of hypothalamic amenorrhea, resulting in infertility [63]. Furthermore, the recognition, evaluation, and treatment of infertility are stressful for most couples, with one study finding 40% of infertility patients fulfilling the diagnosis of a psychiatric disorder,

most commonly anxiety disorders and depressive disorders [64]. Stress can negatively impact the success of fertility treatment [65]; as well as contribute to the discontinuation of fertility treatment before pregnancy is achieved [66]. It has been found that psychosocial interventions for couples during infertility treatment, in particular cognitive behavioural therapy, may be effective both in reducing psychological distress and in improving clinical pregnancy rates [67]. It is therefore essential that any clinician evaluating a woman for reproductive disorders pays attention to their emotional and psychological health when thinking about the cause, as well as when planning potential fertility treatment.

Family history: In women without consanguineous parents, reproductive disorders rarely have an obvious, Mendelian inheritance. However, several factors relevant to female reproductive disorders and infertility may affect multiple family members, which is likely to represent a combination of genetic and environmental factors, such as pubertal disorders, eating disorders, functional hypothalamic amenorrhea, polycystic ovary syndrome, premature ovarian insufficiency, diabetes mellitus, and autoimmune thyroid disease.

Drug history: Various prescribed and over-the-counter drugs can adversely affect reproductive function, and therefore a detailed drug history is an important component of clinical evaluation. Drugs which interfere with the female reproductive system typically do so by one of three mechanisms; inhibiting GnRH release, inducing hyperprolactinaemia or causing hypophysitis, resulting in hypopituitarism. Opioids profoundly inhibit hypothalamic kisspeptin and GnRH secretion, leading to hypogonadotropic hypogonadism [68], and are an under-recognized cause of secondary amenorrhea [69]. Several psychotropic medications including neuroleptics, selective serotonin reuptake inhibitors (SSRIs) and clomipramine cause hyperprolactinaemia by inhibiting the tonic inhibition of hypothalamic dopamine on prolactin secretion [70–72]. Drugs for gastrointestinal conditions, such as the dopamine antagonists metoclopramide and domperidone [73, 74], and certain antihypertensives [75, 76] have also been associated with hyperprolactinaemia. Cytotoxic agents Ipilimumab and tremelimumab are human monoclonal antibodies, which target CTLA-4, and promote an immune-mediated response against cancer cells. However, they can also cause autoimmune reactions in other areas of the body, including the pituitary; leading to hypophysitis and hypopituitarism in 2.3% of patients [77].

Examination

A physical examination should be performed to assess for potential causes of reproductive disorders. Examination should include basic anthropometric measurements, general inspection, examination of the neck, eyes, cardiovascular, and respiratory systems, abdominal and pelvic palpation, examination of secondary sexual characteristics, and gynaecological examination.

Anthropometric measurement: Anthropometric measurements are a simple but powerful tool in the assessment of female reproductive disorders. Measurement of height is useful to determine women with short stature, which may be associated with Turner's syndrome (often combined with a stocky or square appearance) or growth hormone deficiency caused by hypothalamic or pituitary disease; or women with unusually tall stature, which can be associated with XY karyotype. Height and weight can also be used to calculate body mass index (BMI). Underweight women with BMI <18.5 kg/m^2 are more likely to have ovulatory dysfunction, hypothalamic amenorrhea, eating disorders, and hyperthyroidism. Women with obesity are at increased risk of ovulatory dysfunction, as well as polycystic ovary syndrome, hypothyroidism, or rarer endocrine diseases such as Cushing's disease. Waist circumference or waist-to-hip ratio is also helpful as it can identify women with central obesity (defined by a waist circumference of >88 cm in adolescent and adult women [78]) and is associated with insulin resistance and polycystic ovarian syndrome. High waist circumference can also be a sign of Cushing's disease. If available in the clinic, it is useful to measure body fat percentage, using the bioelectrical impedance method, as some women who exercise strenuously or maintain a restrictive diet can have an abnormally low body fat percentage, in spite of a normal BMI, which is in itself associated with ovulatory dysfunction and amenorrhea [79].

Inspection of skin: Close inspection of the skin is invaluable for gaining diagnostic clues to the aetiology of female reproductive endocrine disorders. Women with hypothalamic amenorrhea caused by weight loss, restrictive eating, or an eating disorder may be characteristically thin, with lanugo, mottled, cool extremities or yellow pigmentation (hypercarotenaemia) of the skin. There may be evidence of non-suicidal self-injury (for example skin cutting, picking, or burning with a lit cigarette), which is common in anorexia nervosa [80]. Hyperthyroidism may be associated with tremor, hyperkinesis, and warm vasodilated peripheries; whereas patients with hypothyroidism may display characteristic 'peaches and cream' complexion, dry thin hair, loss of eyebrows, cold peripheries, dry skin, and oedema. Conditions causing androgen excess, most commonly polycystic ovarian syndrome, or more rarely congenital adrenal hyperplasia, 5-alpha reductase deficiency, or androgen-secreting ovarian or adrenal tumours, can cause hirsutism, acne, and androgenic alopecia. Hirsutism is defined clinically as an abnormal amount of terminal hair growth that occurs in a male pattern in women [81] and is commonly graded by the modified Ferriman–Gallwey system where hair growth is rated from 0 (no growth of terminal hair) to 4 (extensive hair growth) in nine areas of the body. A total score of $\geq$ 6–8 is considered as hirsutism. However, the scale is limited by its subjective nature and because it does not take into account all areas of the body that can be affected by androgenic hair growth, for example, buttocks and sideburns. Furthermore, it requires a full-body examination, which some patients may consider invasive. Given these limitations some clinicians prefer to use 'patient-important hirsutism', which takes into consideration all symptoms of hirsutism that cause the patient distress, irrespective of the degree of physical findings [82]. Hirsutism must be distinguished from hypertrichosis, the generalized growth of excessive vellus hair, which can be congenital, caused by metabolic disorders, such as thyroid dysfunction, or a side effect of some medications (see Chapter 8.4.4, for details). Excessive acne vulgaris is another important cutaneous manifestation of hyperandrogenaemia, although there is no universally accepted, well-validated scale for reporting acne severity. Androgenic alopecia, or male-pattern baldness, is a less common symptom of hyperandrogenaemia, usually affecting the fronto-temporo-occipital scalp. Insulin resistant states, which can be found in obesity or polycystic ovarian syndrome,

can also be associated with skin changes including skin tags and acanthosis nigricans (thickened, hyperpigmented skin, predominantly of the flexures). Turner's syndrome, a cause of primary ovarian insufficiency, is associated with a characteristic phenotype including a short-webbed neck, 'shield' chest with widely spaced nipples, cubitus valgus, and Madelung deformity of the forearm and wrist. Characteristic facial features may also be present including epicanthal folds, downslanting palpebral fissures, low set ears, and micrognathia. Women with Cushing's syndrome typically have a moon face, plethora, acne, hirsutism, thin skin, bruising, striae, dorsal fat pad, frontal balding, skin pigmentation, and oedema. Acromegaly can cause changes in appearance including frontal bossing, enlarged jaw, macroglossia, thick greasy skin, spade-like hands and feet, tight rings, and oedema.

Eye examination: Examination of the eyes is useful to determine patients with menstrual dysfunction caused by Graves' disease or a pituitary tumour. Hyperthyroidism caused by Graves' disease can be accompanied by thyroid eye disease, signs of which include exophthalmos, tearing, conjunctival oedema, and periorbital oedema. Examination of eye movements may also reveal ophthalmoplegia. Impairment of vision in one or both eyes is a common consequence of suprasellar extension of a pituitary tumour. Classically this may cause a bitemporal hemianopia, although initially more subtle visual loss can be elicited, such as a scotoma on the bitemporal fields [83].

Neck examination: Inspection of the neck may reveal a transverse scar from total- or subtotal thyroidectomy, and palpation may reveal a goitre, thyroid nodules, or thyroid tenderness in cases of infertility associated with thyroid disease. Parotid gland swelling, or sialadenosis, is a common sign of bulimia nervosa [84].

Breast examination: Inspection of the breasts may reveal atrophic breasts in cases of hypo-oestrogenemia such as GnRH deficiency, hypothalamic amenorrhea, or primary ovarian insufficiency. Applying pressure in a clockwise manner around the areola may elicit nipple discharge [85], which may be present in about 25% of cases of premenopausal women with hyperprolactinemia [86].

Cardiovascular examination: Blood pressure and radial pulse rate are important basic measurements that should be taken during clinical assessment. Bradycardia and hypotension are associated with anorexia nervosa; tachycardia or atrial fibrillation, a full pulse and/or systolic hypertension with hyperthyroidism; bradycardia and/or hypertension with hypothyroidism; and hypertension with obesity, or hyperandrogenic conditions, such as polycystic ovary syndrome, congenital adrenal hyperplasia, or an androgen-secreting tumour. Women with primary ovarian insufficiency caused by Turner's syndrome may have hypertension, and/or a number of cardiac malformations, including aortic valve abnormalities, aortic arch abnormalities, atrial and ventricular septal defects, and pulmonary venous abnormalities [87]. The risk of aortic dissection or rupture increases dramatically in pregnancy, and is associated with a 2% risk of mortality [88]. It is therefore critical that women with Turner's syndrome planning pregnancy (usually through *in vitro* fertilization), have a full medical evaluation, with particular attention paid to the cardiovascular system, before fertility treatment is considered [89].

Pelvic and abdominal examination: Abdominal and pelvic palpation should be performed on all women presenting with menstrual dysfunction or infertility. A palpable mass may suggest uterine fibroids or an ovarian mass. Scars indicate previous surgery, which may be a consequence of gynaecological disease (e.g. previous surgical treatment for endometriosis), or may be a cause of infertility due to subsequent adhesion formation. Tenderness on pelvic palpation may suggest pelvic inflammatory disease, ovarian cysts, or endometriosis.

Gynaecological examination: Gynaecological examination should be performed by a clinician with appropriate experience. Patient consent should be requested before conducting a gynaecological examination, and a chaperone offered, regardless of whether the clinician performing the examination is a male or female [90]. Some women may experience substantial fear, embarrassment, or anxiety regarding gynaecological examinations, which can be related to previous pelvic examinations, or rarely due to previous traumatic experiences such as sexual abuse or assault [91]. Providing an explanation of each step of the procedure, having a considerate and sensitive approach to examination and, have both been found to improve patient experience during a vaginal examination [92]. Patients should always be given the opportunity to stop the examination at any point, and clinicians should be sensitive to signs of patient discomfort. The external genitalia should be examined first. Vulval and vaginal atrophy, thin reddened skin, and lack of lubrication are signs of oestrogen deficiency. Clitoral hypertrophy is a sign of virilization and can be associated with congenital adrenal hyperplasia, androgen-secreting ovarian or adrenal masses, or androgen insensitivity syndrome [93]. External genitalia should be examined for female genital mutilation, also known as female circumcision or female genital cutting, which is a common cultural practice in some parts of Africa and Asia, affecting more than 125 million females worldwide [94], and increasingly seen by clinicians throughout the world as a result of immigration. Female genital mutilation should be classified according to the World Health Organization, and clearly documented. Infertility is more common in women with female genital mutilation, and there is a positive association between the anatomical extent of female mutilation and infertility [95]. Furthermore, women with female genital mutilation are significantly more likely than those without female genital mutilation to have adverse obstetric outcomes, with greater risks associated with more extensive female genital mutilation [96]. Anatomical abnormalities of the vagina and cervix should be assessed, including an intact hymen, transverse vaginal septum, or Müllerian agenesis, especially in cases of primary amenorrhea. Speculum examination is appropriate to examine the vagina and cervix. Abnormal mucopurulent discharge may signify infection and should be further investigated. A bimanual examination can be performed to assess the uterus for size, shape, position, and mobility, and the adnexae for tenderness or masses. Uterine enlargement, irregularity, or lack of mobility are signs of a uterine anomaly, including fibroids, endometriosis, or pelvic adhesions. Tenderness or masses in the adnexae or recto-uterine pouch is consistent with chronic pelvic inflammatory disease, ovarian cysts, or endometriosis. Endometriosis may also cause palpable tender nodules in the posterior fornix, recto-uterine pouch, uterosacral ligaments, or rectovaginal septum [97].

Secondary sexual characteristics: If indicated by the history, clinical examination of secondary sexual characteristics should be performed, including pubic hair and breast development, as assessed by Tanner stages (Figure 8.2.1.1) [98]. Tanner stages rate sexual maturity for pubic hair and breast development in five stages, with stage one representing pre-puberty and stage five

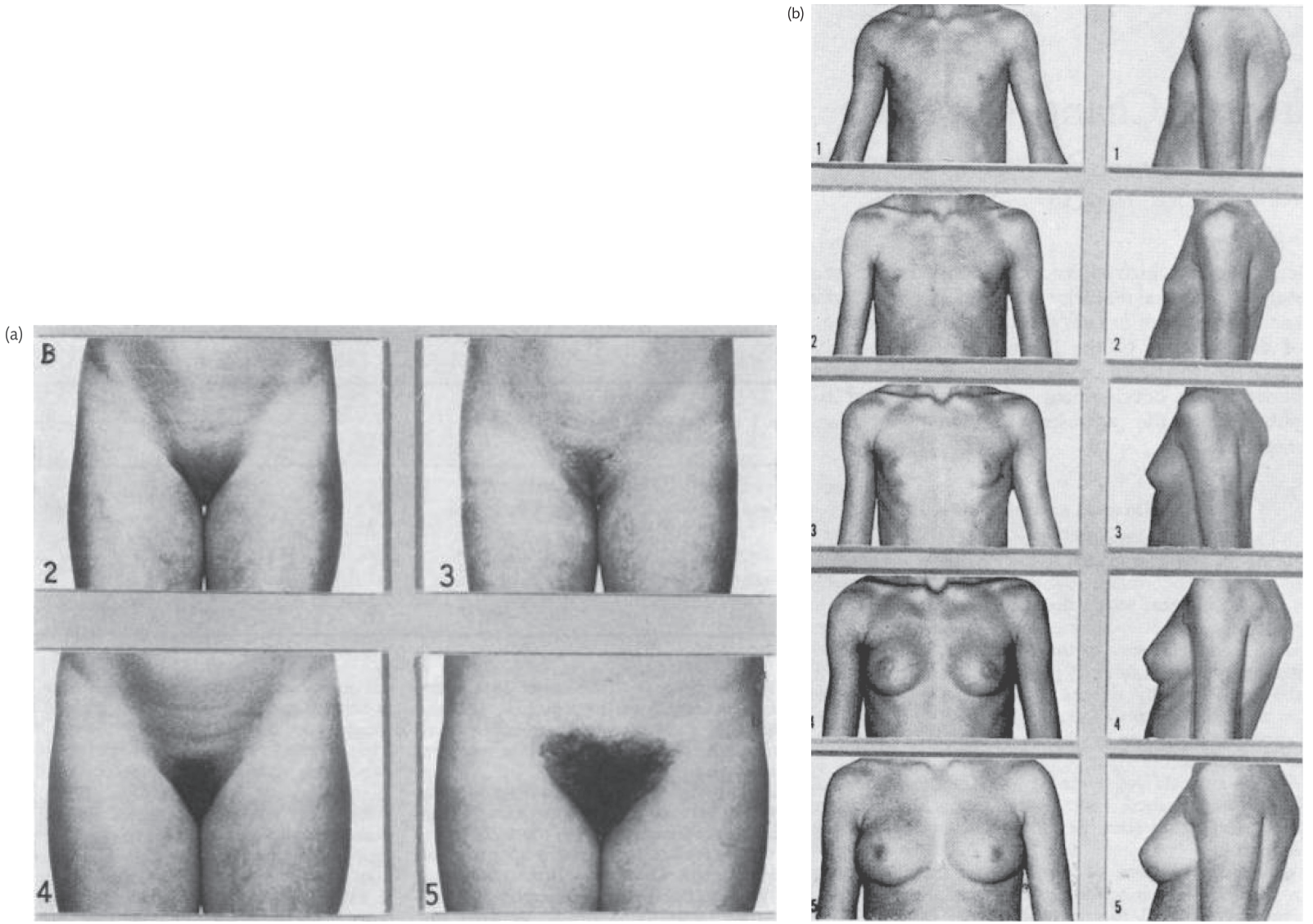


Figure 8.2.1.1 Tanner stages for rating sexual maturity. The secondary sexual developments of pubic hair and breast changes are assigned stages from 1 (pre-adolescent) to 5 (mature). Pubic hair development: Stage 1: Pre-adolescent; Stage 2: Sparse growth of long, slightly pigmented hair, straight, or curled along labia; Stage 3: Considerably darker, coarser, and more curled; Stage 4: Adult type but considerably smaller area; Stage 5: Adult in quantity and type. Breast development: Stage 1: Pre-adolescent; Stage 2: Breast bud stage; Stage 3: Further enlargement of breast and areola with no separation of their contours; Stage 4: Projection of the areola and papilla to form a secondary mound; Stage 5: Mature stage with projection of papilla only.

representing mature adult development. Incomplete pubertal development is usually associated with primary amenorrhea, and suggests oestrogen deficiency, which could be caused by hypothalamic or pituitary dysfunction, ovarian failure, and/or a chromosomal abnormality.

REFERENCES

1. Pfeifer S, Goldberg J, Lobo R, *et al*. Practice Committee of American Society for Reproductive Medicine. Definitions of infertility and recurrent pregnancy loss. *Fertil Steril*, 2008; **99**: 63.
2. World Health Organization (WHO) Technical Report Series. *Recent Advances in Medically Assisted Conception Number 820*. Geneva: WHO, 1992: 1–111.
3. Reindollar RH, Byrd JR, McDonough PG. Delayed sexual development: a study of 252 patients. *Am J Obstet Gynecol*, 1981; **140**: 371–80.
4. Boehm U, Bouloux PM, Dattani MT, *et al*. Expert consensus document: European Consensus Statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*, 2015; **11**: 547–64.
5. Lee PA, Houk CP, Ahmed SF, Hughes IA. Consensus statement on management of intersex disorders. International Consensus Conference on Intersex. *Pediatrics*, 2006; **118**: e488–500.
6. McCarthy JJ, Rockette, HE. Prediction of ovulation with basal body temperature. *J Reprod Med*, 1986; **31**: 742–7.
7. Magyar DM, Boyers SP, Marshall JR, Abraham GE. Regular menstrual cycles and premenstrual molimina as indicators of ovulation. *Obstet Gynecol*, 1979; **53**: 411–14.
8. Luciano AA, Peluso J, Koch EI, Maier D, Kuslis S, Davison E. Temporal relationship and reliability of the clinical, hormonal, and ultrasonographic indices of ovulation in infertile women. *Obstet Gynecol*, 1990; **75**: 412–16.
9. Moghissi KS. Prediction and detection of ovulation. *Fertil Steril*, 1981; **36**: 729–33.
10. Bauman JE. Basal body temperature: unreliable method of ovulation detection. *Fertil Steril*, 1981; **36**: 729–33.
11. Guermandi E, Vegetti W, Bianchi MM, Uglietti A, Ragni G, Crosignani P. Reliability of ovulation tests in infertile women. *Obstet Gynecol*, 2001; **97**: 92–6.
12. Wilcox AJ, Weinberg CR, Baird DD. Timing of sexual intercourse in relation to ovulation. Effects on the probability of conception, survival of the pregnancy, and sex of the baby. *N Engl J Med*, 1995; **333**: 1517–21.
13. Wilcox AJ, Dunson D, Baird DD. The timing of the 'fertile window' in the menstrual cycle: day specific estimates from a prospective study. *BMJ*, 2000; **321**: 1259–62.
14. Kopitzke EJ, Berg BJ, Wilson JF, Owens D. Physical and emotional stress associated with components of the infertility investigation: perspectives of professionals and patients. *Fertil Steril*, 1991; **55**: 1137–43.
15. Perloff WH, Steinberger E. In vivo survival of spermatozoa in cervical mucus. *Am J Obstet Gynecol*, 1964; **88**: 439–42.
16. Danielsson I, Sjöberg I, Stenlund H, Wikman M. Prevalence and incidence of prolonged and severe dyspareunia in women: results from a population study. *Scand J Public Health*, 2003; **31**: 113–18.
17. Anderson L, Lewis SE, McClure N. The effects of coital lubricants on sperm motility *in vitro*. *Hum Reprod*, 1998; **13**: 3351–6.
18. Goldenberg RL, White R. The effect of vaginal lubricants on sperm motility *in vitro*. *Fertil Steril*, 1975; **26**: 872–3.
19. Steiner AZ, Long DL, Tanner C, Herring AH. Effect of vaginal lubricants on natural fertility. *Obstet Gynecol*, 2012; **120**: 44–51.
20. Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Reproductive Endocrinology and Infertility. Optimizing natural fertility. *Fertil Steril*, 2008; **90**: S1–S6.
21. Garza-Flores J, Cardenas S, Rodriguez V, Cravioto MC, Diaz-Sanchez V, Perez-Palacios G. Return to ovulation following the use of long-acting injectable contraceptives: a comparative study. *Contraception*, 1985; **31**: 361–6.
22. Pardthaisong T, Gray RH, McDaniel EB. Return of fertility after discontinuation of depot medroxyprogesterone acetate and intra-uterine devices in Northern Thailand. *Lancet*, 1980; **1**: 509–12.
23. NICE Clinical Guidelines, No. 30. National Collaborating Centre for Women's and Children's Health (UK). *Long-Acting Reversible Contraception: The Effective and Appropriate Use of Long-Acting Reversible Contraception*. London: RCOG Press, 2005.
24. Tietze C, Lewit S. Amsterdam: Excerpta Medica International Congress Series No 54; 1962. Intra-uterine contraceptive devices. Proceedings of the Conference. New York City. 30 April–1 May 1962.
25. Hubacher D, Lara-Ricalde R, Taylor DJ, Guerra-Infante F, Guzmán-Rodríguez R. Use of copper intrauterine devices and the risk of tubal infertility among nulligravid women. *N Engl J Med*, 2001; **345**: 561–7.
26. Lado-Abeal J, Rodriguez-Arnao J, Newell-Price JD, Perry LA, Grossman AB, Besser GM, Trainer PJ. Menstrual abnormalities in women with Cushing's disease are correlated with hypercortisolemia rather than raised circulating androgen levels. *J Clin Endocrinol Metab*, 1998; **83**: 3083–8.
27. Kaltsas GA, Mukherjee JJ, Jenkins PJ, Satta MA, Islam N, Monson JP, Besser GM, Grossman AB. Menstrual irregularity in women with acromegaly. *J Clin Endocrinol Metab*, 1999; **84**: 2731–5.
28. Hirshfeld-Cytron J, Kim HH. Treatment of infertility in women with pituitary tumors. *Expert Rev Anticancer Ther*, 2006; **6**: S55–62.
29. Barker FG 2nd, Klibanski A, Swearingen B. Transsphenoidal surgery for pituitary tumors in the United States, 1996–2000: mortality, morbidity, and the effects of hospital and surgeon volume. *J Clin Endocrinol Metab*, 2003; **88**: 4709–19.
30. Snyder PJ, Fowble BF, Schatz NJ, Savino PJ, Gennarelli TA. Hypopituitarism following radiation therapy of pituitary adenomas. *Am J Med*, 1986; **81**: 457–62.
31. Constine LS, Woolf PD, Cann D, Mick G, McCormick K, Raubertas RF, Rubin P. Hypothalamic-pituitary dysfunction after radiation for brain tumors. *N Engl J Med*, 1993; **328**: 87–94.
32. Agha A, Sherlock M, Brennan S, *et al*. Hypothalamic-pituitary dysfunction after irradiation of nonpituitary brain tumors in adults. *J Clin Endocrinol Metab*, 2005; **90**: 6355–60.
33. Guerrero AF, Alfonso A. Traumatic brain injury-related hypopituitarism: a review and recommendations for screening combat veterans. *Mil Med*, 2010; **175**: 574–80.
34. Kelly TM, Edwards CQ, Meikle AW, Kushner JP. Hypogonadism in hemochromatosis: reversal with iron depletion. *Ann Intern Med*, 1984; **101**: 629–32.
35. Stuart CA, Neelon FA, Lebovitz HE. Hypothalamic insufficiency: the cause of hypopituitarism in sarcoidosis. *Ann Intern Med*, 1978; **88**: 589–94.

36. Summers VK, Hipkin LJ, Osborne Hughes R, Davis JC. Panhypopituitarism after cured tuberculous meningitis. *Br Med J*, 1968; **1**: 359.
37. Lam KS, Sham MM, Tam SC, Ng MM, Ma HT. Hypopituitarism after tuberculous meningitis in childhood. *Ann Intern Med*, 1993; **118**: 701–6.
38. Zarzycki W, Zieniewicz M. Reproductive disturbances in type 1 diabetic women. *Endocrinol Lett*, 2005; **26**: 733–8.
39. Sievertsen GD, Lim VS, Nakawatase C, Frohman LA. Metabolic clearance and secretion rates of human prolactin in normal subjects and in patients with chronic renal failure. *J Clin Endocrinol Metab*, 1980; **50**: 846–52.
40. Rush H, Neugarten J, Coco M. Women's health issues in a dialysis population. *Clin Nephrol*, 2000; **54**: 455–62.
41. Freeman HJ. Reproductive changed associated with celiac disease. *World J Gastroenterol*, 2010; **16**: 5810–14.
42. Bath LE, Hamish W, Wallace B, Critchley HO. Late effects of the treatment of childhood cancer on the female reproductive system and the potential for fertility preservation. *BJOG*, 2002; **109**: 107–14.
43. Meirow D, Nugent D. The effects of radiotherapy and chemotherapy on female reproduction. *Hum Reprod Update*, 2001; **7**: 535–43.
44. Vaiman M, Nagibin A, Hagag P, Kessler A, Gavriel H. Hypothyroidism following partial thyroidectomy. *Otolaryngol Head Neck Surg*, 2008; **138**: 98–100.
45. Lass A. The fertility potential of women with a single ovary. *Hum Reprod Update*, 1999; **5**: 5460.
46. Zargar AH, Singh B, Laway BA, Masoodi SR, Wani AI, Bashir MI. Epidemiologic aspects of postpartum pituitary hypofunction (Sheehan's syndrome). *Fertil Steril*, 2005; **84**: 523–8.
47. Shaheen R, Subhan F, Tahir F. Epidemiology of genital tuberculosis in infertile population. *J Pak Med Assoc*, 2006; **56**: 306–9.
48. Zaadstra BM, Seidell JC, Van Noord PA, *et al*. Fat and female fecundity: prospective study of effect of body fat distribution on conception rates. *BMJ*, 1993; **306**: 484–7.
49. Clark AM, Roberts BG. Maximizing weight loss in the overweight infertile patient: a prospective randomized controlled trial. 16th Annual Meeting of ESHRE, 2000, Bologna, Italy, 2000. Abstract No.O-162. *Hum Reprod*, 2000; **15**(Abstract Book 1): 65–6.
50. Bates GW, Bates SR, Whitworth NS. Reproductive failure in women who practice weight control. *Fertil Steril*, 1982; **37**: 373–8.
51. Gudmundsdottir SL, Flanders WD, Augestad LB. Physical activity and fertility in women: the North-Trøndelag Health Study. *Hum Reprod*, 2009; **24**: 3196–204.
52. Hughes EG, Brennan BG. Does cigarette smoking impair natural or assisted fecundity? *Fertil Steril*, 1996; **66**: 679–89.
53. Hull MG, North K, Taylor H, Farrow A, Ford WC. Delayed conception and active and passive smoking. The Avon Longitudinal Study of Pregnancy and Childhood Study Team. *Fertil Steril*, 2000; **74**: 725–33.
54. Kaya Y, Kizilkaya Beji N, Aydin Y, Hassa H. The effect of health-promoting lifestyle education on the treatment of unexplained female infertility. *Eur J Obstet Gynecol Reprod Biol*, 2016; **207**: 109–14.
55. Mamluk L, Edwards HB, Savović J, *et al*. Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently 'safe' levels of alcohol during pregnancy? A systematic review and meta-analyses. *BMJ Open*, 2017; **7**: e015410.
56. Mueller BA, Daling JR, Weiss NS, Moore DE. Recreational drug use and the risk of primary infertility. *Epidemiology*, 1990; **1**: 195–200.
57. Joesoef MR, Beral V, Aral SO, Rolfs RT, Cramer DW. Fertility and use of cigarettes, alcohol, marijuana, and cocaine. *Ann Epidemiol*, 1993; **3**: 592–4.
58. Bisanti L, Olsen J, Basso O, Thonneau P, Karmaus W. Shift work and subfecundity: a European multicenter study. *J Occup Environ Med*, 1996; **38**: 352–8.
59. Zielhuis G, Peelen SJM, Florack EIM, Roeleveld N. Hospital work and fecundability. *Scand J Work Environ Health*, 1999; **25**: 47–8.
60. Smith EM, Hammonds-Ehlers M, Clark MK, Kirchner HL, Fuortes L. Occupational exposures and risk of female infertility. *J Occup Environ Med*, 1997; **39**: 138–47.
61. Pearce JM. Richard Morton: origins of anorexia nervosa. *Eur Neurol*, 2004; **52**: 191–2.
62. Lynch CD, Sundaram R, Maisog JM, Sweeney AM, Buck Louis GM. Preconception stress increases the risk of infertility: results from a couple-based prospective cohort study—the LIFE study. *Hum Reprod*, 2014; **29**: 1067–75.
63. Berga SL, Girton LG. The psychoneuroendocrinology of functional hypothalamic amenorrhea. *Psychiatr Clin North Am*, 1989; **12**: 105–16.
64. Chen TH1, Chang SP, Tsai CF, Juang KD. Prevalence of depressive and anxiety disorders in an assisted reproductive technique clinic. *Hum Reprod*, 2004; **19**: 2313–8.
65. Boivin J, Schmidt L. Infertility-related stress in men and women predicts treatment outcome 1 year later. *Fertil Steril*, 2005; **83**: 1745–52.
66. Rajkhowa M, McConnell A, Thomas GE. Reasons for discontinuation of IVF treatment: a questionnaire study. *Hum Reprod*, 2006; **21**: 358–63.
67. Frederiksen Y, Farver-Vestergaard I, Skovgård NG, Ingerslev HJ, Zachariae R. Efficacy of psychosocial interventions for psychological and pregnancy outcomes in infertile women and men: a systematic review and meta-analysis. *BMJ Open*, 2015; **5**: e006592.
68. Narayanaswamy S, Prague JK, Jayasena CN, *et al*. Hypothesis in humans by coadministration of kisspeptin, neurokinin B, and naltrexone in Men. *J Clin Endocrinol Metab*, 2016; **101**: 3429–36.
69. Abs R, Verhelst J, Maeyaert J, *et al*. Endocrine consequences of long-term intrathecal administration of opioids. *J Clin Endocrinol Metab*, 2000; **85**: 2215–22.
70. Meltzer HY, Matsubara S, Lee JC. Classification of typical and atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin2 pKi values. *J Pharmacol Exp Ther*, 1989; **251**: 238–46.
71. Papakostas GI, Miller KK, Petersen T, *et al*. Serum prolactin levels among outpatients with major depressive disorder during the acute phase of treatment with fluoxetine. *J Clin Psychiatry*, 2006; **67**: 952–7.
72. Golden RN, Hsiao J, Lane E, Hicks R, Rogers S, Potter WZ. The effects of intravenous clomipramine on neurohormones in normal subjects. *J Clin Endocrinol Metab*, 1989; **68**: 632–7.
73. McCallum RW, Sowers JR, Hershman JM, Sturdevant RA. Metoclopramide stimulates prolactin secretion in man. *J Clin Endocrinol Metab*, 1976; **42**: 1148–52.
74. Sowers JR, Sharp B, McCallum RW. Effect of domperidone, an extracerebral inhibitor of dopamine receptors, on thyrotropin, prolactin, renin, aldosterone, and 18-hydroxycorticosterone secretion in man. *J Clin Endocrinol Metab*, 1982; **54**: 869–71.

75. Steiner J, Cassar J, Mashiter K, Dawes I, Fraser TR, Breckenridge A. Effects of methyldopa on prolactin and growth hormone. *Br Med J*, 1976; **1**: 1186–8.
76. Kelley SR, Kamal TJ, Molitch ME. Mechanism of verapamil calcium channel blockade-induced hyperprolactinemia. *Am J Physiol*, 1996; **270**: E96–E100.
77. Hodi FS, O'Day SJ, McDermott DF, *et al*. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*, 2010; **363**: 711–23.
78. Leibel NI, Baumann EE, Kocherginsky M, Rosenfield RL. Relationship of adolescent polycystic ovary syndrome to parental metabolic syndrome. *J Clin Endocrinol Metab*, 2006; **91**: 1275.
79. Stokić E1, Srdić B, Barak O. Body mass index, body fat mass and the occurrence of amenorrhea in ballet dancers. *Gynecol Endocrinol*, 2005; **20**: 195–9.
80. Cucchi A, Ryan D, Konstantakopoulos G, *et al*. Lifetime prevalence of non-suicidal self-injury in patients with eating disorders: a systematic review and meta-analysis. *Psychol Med*, 2016; **46**: 1345–58.
81. Martin KA, Anderson RR, Chang RJ, *et al*. Evaluation and treatment of hirsutism in premenopausal women: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1233–57.
82. Yildiz BO, Bolour S, Woods K, Moore A, Azziz R. Visually scoring hirsutism. *Hum Reprod Update*, 2010; **16**: 51–64.
83. Wilson P, Falconer M. Patterns of visual failure with pituitary tumours. Clinical and radiological correlations. *Brit J Ophthal*, 1968; **52**: 94.
84. Brown CA, Mehler PS. Medical complications of self-induced vomiting. *Eat Disord*, 2013; **21**: 287–94.
85. Klimberg VS. Nipple discharge: more than pathologic. *Ann Surg Oncol*, 2003; **10**: 98.
86. Bayrak A, Saadat P, Mor E, Chong L, Paulson RJ, Sokol RZ. Pituitary imaging is indicated for the evaluation of hyperprolactinemia. *Fertil Steril*, 2005; **84**: 181–5.
87. Mortensen KH, Andersen NH, Gravholt CH. Cardiovascular phenotype in Turner syndrome—integrating cardiology, genetics, and endocrinology. *Endocr Rev*, 2012; **33**: 677–714.
88. Karnis MF, Zimon AE, Lalwani SI, Timmreck LS, Klipstein S, Reindollar RH. Risk of death in pregnancy achieved through oocyte donation in patients with Turner syndrome: a national survey. *Fertil Steril*, 2003; **80**: 498.
89. Practice Committee of American Society For Reproductive Medicine. Increased maternal cardiovascular mortality associated with pregnancy in women with Turner syndrome. *Fertil Steril*, 2012; **97**: 282–4.
90. Committee on Ethics, American College of Obstetricians and Gynecologists. ACOG Committee Opinion No. 373: sexual misconduct. *Obstet Gynecol*, 2007; **110**: 441.
91. Petravage JB, Reynolds LJ, Gardner HJ, Reading JC. Attitudes of women toward the gynecologic examination. *J Fam Pract*, 1979; **9**: 1039.
92. Broadmore J, Carr-Gregg M, Hutton JD. Vaginal examinations: women's experiences and preferences. *N Z Med J*, 1986; **99**: 8.
93. Jorgensen PB, Kjartansdóttir KR, Fedder J. Care of women with XY karyotype: a clinical practice guideline. *Fertil Steril*, 2010; **94**: 105–13.
94. The Lancet. Changing culture to end FGM. *Lancet*, 2018; **391**: 401.
95. Almroth L, Elmusharaf S, El Hadi N, *et al*. Primary infertility after genital mutilation in girlhood in Sudan: a case-control study. *Lancet*, 2005; **366**: 385.
96. WHO study group on female genital mutilation and obstetric outcome, Banks E, Meirik O, Farley T, Akande O, Bathija H, Ali M. Female genital mutilation and obstetric outcome: WHO collaborative prospective study in six African countries. *Lancet*, 2006; **367**: 1835.
97. Hickey M, Ballard K, Farquhar C. Endometriosis. *BMJ*, 2014; **348**: 1752.
98. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child*, 1969; **44**: 291.

8.2.2 Laboratory Evaluation

Daniel Dumesic and Zain Al-Safi

Hormonal Evaluation

β-hCG

In cases of abnormal uterine bleeding (irregular, heavy or absent), it is important to rule out pregnancy. Measuring urine or serum β-hCG should be the initial laboratory evaluation in these women. Therefore, all patients with menstrual bleeding disorders should be tested for pregnancy.

Oestradiol, Progesterone, LH, and FSH

Measuring serum gonadotropins (FSH, LH) along with serum oestradiol and progesterone levels provides information to aid in the diagnosis of hypothalamic, pituitary, and ovarian causes of amenorrhea or oligomenorrhea. It can rule out causes such as ovarian insufficiency when elevated serum FSH and LH levels accompany low serum oestradiol (<20 pg/ml) and progesterone (<1 ng/ml) levels, indicating low or absent ovarian reserve.

Finding a transient elevation in serum FSH and LH levels with serum oestradiol levels <150 pg/ml and progesterone <2 ng/ml can indicate the midcycle ovulatory gonadotropin surge [1].

Compared to normal women, women with functional hypothalamic amenorrhea have a 53% reduction in LH pulse frequency but comparable LH pulse amplitude [2]. In these women, the diagnosis of functional hypothalamic amenorrhea is made after excluding other causes such as severe weight loss, excessive exercise resulting in low body fat, severe psychological stress, and physical wasting from severe systemic diseases (i.e. disseminated malignancies and pituitary or central nervous system lesions). Serum LH and FSH

levels are often normal; however, very low and often undetectable serum LH and FSH levels suggest organic hypothalamic amenorrhea due to genetic mutations affecting gonadotropin-releasing hormone (GnRH) production and function or central causes, such as pituitary, hypothalamic, or other brain tumours, and infiltrative lesions [1]. Patients with hypothalamic amenorrhea characteristically have a low level of serum oestradiol and low or low-to-normal levels of LH and FSH, whereas the gonadotropin response to GnRH stimulation can be preserved.

Serum oestradiol assessment can be limited by factors related to the assay used, such as poor sensitivity, variation among assays, and the fact that a measurement reflects a single time point [3].

Low serum oestradiol and progesterone (<1 ng/ml) levels are consistent with anovulation and may be seen in women with underlying polycystic ovary syndrome (PCOS). An elevated serum LH to FSH ratio previously has been reported in PCOS patients, but neither circulating levels of LH and FSH, nor the ratio of serum LH to FSH contribute significantly to the diagnosis of PCOS, particularly since LH secretion is pulsatile in nature and one-third of PCOS patients have normal circulating LH levels. Moreover, LH secretion and pulse amplitude are inversely related to body mass index (BMI) and percent body fat, suggesting that obesity in PCOS patients further modifies gonadotropin secretion [4].

Serum FSH and oestradiol checked on cycle days 2–4 can help to measure ovarian reserve. Serum FSH concentration starts to increase early in the menstrual cycle with diminished ovarian reserve and in reproductive ageing, but the assays have significant inter- and intracycle variability limiting their reliability [5].

Basal oestradiol alone should not be used to screen for diminished ovarian reserve. The test can aid to correctly interpret basal serum FSH value. With lower ovarian reserve as normally occurs with reproductive ageing, there is an early rise in serum oestradiol levels that can lower an otherwise elevated basal serum FSH concentration into the normal range

Anti-Müllerian hormone is another measure for ovarian reserve and is discussed later in this section.

Androgens (Testosterone, DHEAS)

Measuring serum testosterone is essential to identify hyperandrogenaemia as one of the criteria for diagnosing PCOS. Testosterone is produced either directly by the ovary or adrenal gland, or indirectly through metabolism of secreted androstenedione or DHEA (or its sulphate DHEAS) in peripheral tissues, such as fat and skin [6]. Approximately 50% of testosterone arises from peripheral conversion of androstenedione, whereas 25% is secreted by the ovaries and an additional 25% by the adrenal. LH is the stimulus for ovarian testosterone production.

High serum testosterone values in the adult-male range may indicate the presence of an androgenic tumour, while only a moderate increase of this androgen is a poor predictor of androgenic tumours. Androgen-producing ovarian tumours are uncommon, occurring only in about 1 in 500 hirsute women [7]. Ovarian tumours that produce androgens directly include Sertoli–Leydig cell tumours (androblastoma, arrhenoblastoma), granulosa-theca cell (stromal cell) tumours, and hilus cell tumours. Nevertheless, any large ovarian tumour (i.e. cystic teratomas, Brenner tumours, serous cystadenomas, and Krukenberg tumours) can produce androgens indirectly by causing hyperplasia of the surrounding normal stroma [8].

Several limitations exist, however, in measuring testosterone, which include lack of a standardized assay; poor sensitivity and accuracy in the female range; and variability based upon menstrual cycle phase, time of day, and feeding [9]. Avoiding interference by steroids of similar configuration requires extraction and separation of other circulating steroids chromatographically before subjecting the sample to immunoassay or mass spectrometry [9].

Serum total testosterone levels are above the upper normal limit in about 50% of women with PCOS by 1990 NIH criteria [10]. There is an overlap in total testosterone values between PCOS patients and normal women. When measured by liquid chromatography–mass spectrometry (LC-MS/MS), serum total testosterone levels that best distinguish PCOS from normal women are approximately 50 ng/dl in the general population (10% estimated PCOS prevalence) and 35 ng/dL in the clinical or referred, setting (70% estimated PCOS prevalence) [11].

Measuring free testosterone is recommended by the 2004 Rotterdam consensus to detect hyperandrogenism in PCOS women [12]. Free testosterone is more sensitive than total testosterone in assessing hyperandrogenaemia [13]. As a surrogate of free testosterone by equilibrium dialysis, the free androgen index (FAI) is calculated by measuring sex hormone binding globulin (SHBG) and total testosterone [14]. The FAI is elevated in about 54–89% of women with PCOS by 1990 NIH criteria [10, 15]. The FAI also is elevated in 30% of women with PCOS by Rotterdam criteria who have normal total testosterone levels, thereby identifying a subset of these PCOS women who should be considered as hyperandrogenic [15].

Serum levels of total testosterone in excess of 200 ng/dL are suspicious for an ovarian androgen-producing tumour. However, the best predictor of an androgen secreting neoplasm is virilization defined as male temporal balding, deepening of the voice, and enlargement of the clitoris, which occurs in 98% of such tumours, regardless of circulating testosterone levels [16].

The value of routinely measuring circulating androstenedione or DHEA as precursors of testosterone in hyperandrogenic women is more controversial. Serum androstenedione and total testosterone levels are elevated in 88% and 65% of PCOS women by Rotterdam criteria, respectively [17]. The combined use of total testosterone, androstenedione and FAI identifies hyperandrogenaemia in about 90% of PCOS patients [15].

Elevated serum DHEAS levels occur in 40–70% of hyperandrogenic women with PCOS [18]. Although increased in women with symptoms of androgen excess, only 16% of hyperandrogenic women have increased serum DHEAS in the presence of normal total and free testosterone levels [7]. Conversely, low serum DHEAS levels can occur in adolescents and young women with functional hypothalamic amenorrhea from anorexia nervosa, despite hypercortisolism and adequate adrenocorticotropic hormone (ACTH) [19, 20]. Measuring serum DHEAS, however, should be considered to rule out adrenal aetiologies [1]. In general, measurements of DHEAS are of lower yield unless indicated by clinical findings of severe hirsutism with high concentrations far exceeding the normal range (>6000 ng/dl), suggestive of an adrenal tumour [21].

17 α-Hydroxyprogesterone (17 OHP)

A basal serum 17-OHP level is useful to exclude late-onset congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency. It must be remembered, however, that an elevated basal serum 17-OHP level can also represent a normal finding in the luteal phase of the menstrual cycle [22], and can be repeated in the follicular phase of the menstrual cycle if necessary.

Random measurements of basal serum 17-OHP may be normal in mildly affected non-classic patients unless performed in the early morning (i.e. before 08.00 h) [23].

A morning serum 17 OHP level more than 200 ng/dL should prompt a measurement of serum cortisol and 17OHP at 30 to 60 minutes after ACTH administration (250 mcg intravenous or intramuscular). Post-ACTH 17OHP values in the 1000–10000 ng/dl range are typical of non-classic 21 hydroxylase enzyme deficiency [24–26]. If laboratory data are equivocal, genetic testing for mutations in the *CYP21A2* gene on DNA from peripheral blood cells is available.

Prolactin

A single measurement of serum prolactin level is usually adequate to document hyperprolactinemia. Prolactin secretion, however, is pulsatile and stress, as well as a number of physiological states such as sleep, and exercise can also mildly elevate prolactin levels. For these reasons, a test that shows a level of 25–40 ng/ml should be repeated before hyperprolactinemia is diagnosed [27]. When other causes of hyperprolactinemia have been ruled out (including medication intake) on the basis of history, physical examination, and assessment of thyroid and renal function, the diagnosis of a prolactinoma can be confirmed by gadolinium-enhanced magnetic resonance imaging (MRI) (see later in this section).

In general, serum prolactin levels are correlated with tumour size. Values between upper limits of normal and 100 ng/ml may be due to drugs or idiopathic causes, but can also be caused by microadenomas. Macroadenomas are typically associated with levels of over 250 ng/ml, and in some cases the level exceeds 1000 ng/ml [27, 28]. Care must be taken in interpreting a moderate elevation of the prolactin level (<100 ng/ml) in the presence of a macroadenoma, as this may be due to either compression of the pituitary stalk by the tumour or these values can be artificially low due to the 'hook effect'. The 'hook effect' occurs when a very high serum prolactin level saturates both the capture and signal antibodies used in immunoradiometric and chemiluminescent assays, preventing the binding of the two in a 'sandwich'. The artefact can be avoided by repeating the assay with serial serum dilution [29].

The degree of prolactin elevation correlates with the severity of hypogonadism. High prolactin levels greater than 100 ng/ml are associated with hypogonadism and low serum oestradiol levels, which result in amenorrhoea and vaginal dryness; while levels of 50 to 100 ng/ml can be associated with both oligomenorrhea and amenorrhea, and levels of 20 to 50 ng/ml can result in insufficient progesterone secretion and luteal phase shortening that may result in subfertility [30].

In PCOS, elevations in serum prolactin have been reported to occur and likely represent lactotrope stimulation by chronic oestrogen exposure [31, 32] rather than coexistence of a prolactinoma.

Growth Hormone

Growth hormone (GH) is secreted by the anterior pituitary somatotroph cells in a pulsatile fashion. Over 95% of patients with acromegaly harbour a GH secreting pituitary adenoma arising from somatotroph cells, leading to GH and insulin-like growth factor-1 (IGF-1) hypersecretion.

Patients with clinical features of acromegaly or with a pituitary mass require biochemical evaluation. Serum IGF-1 measurement is recommended as the initial screening as it is the best marker for integrated GH secretion [33]. A normal serum IGF-1 level effectively excludes the diagnosis of acromegaly. Hepatic and renal failure, hypothyroidism, malnutrition, severe infection, and poorly controlled diabetes mellitus may cause falsely elevated, normal, or low IGF-1 values [33]. Levels must be assessed in age-appropriate normal values with knowledge of the specific assay being used due to significant interassay variability [34].

Random measurements of GH are not useful due to the pulsatile nature of its secretion and the significant level variation throughout the day and in response to stimuli.

A dynamic test, such as an oral glucose tolerance test, is more specific and can be used to establish the diagnosis in those with elevated or equivocal IGF-1 levels. Normally GH should decrease to 1 ng/ml or less within 2 hours after ingestion of 75 g of glucose, and so a nadir serum GH level < 1 ng/ml excludes the diagnosis [35].

Anti-Müllerian Hormone (AMH)

AMH is a homodimeric disulphide-linked glycoprotein of the transforming growth factor-β (TGF-β) superfamily [36]. AMH level is lowest at birth and remains low during the prepubertal years; it then peaks during puberty and ultimately declines to undetectable levels at menopause [37, 38]. During ovarian follicle development, minimal AMH is produced by primordial follicles as was shown in immunohistochemistry studies in human ovarian tissue confirming the absence of AMH staining in primordial follicles [39]. The production of AMH increases in pre-antral and early antral follicles until a follicular diameter of 8 mm after which a sharp decline occurs [40]. It was demonstrated that 5–8 mm sized follicles contribute 60% of the circulating AMH. In contrast, granulosa cells of larger preovulatory follicles beyond 10 mm fail to produce AMH [40].

Anti-Müllerian hormone has been suggested as an ideal marker for assessing the remaining oocyte pool because it correlates with antral follicle counts (AFC), outcomes from ovarian stimulation, and the onset of menopause [41, 42]. AMH level can be an additional helpful assessment measure in women with PCOS [43], as it has been shown to correlate with polycystic ovaries. Serum AMH levels in women with PCOS are elevated 2- to 3-fold and are positively correlated with AFC and serum androgen levels [44, 45]. High serum AMH is specific but not sensitive to distinguish PCOS from asymptomatic women with polycystic ovaries [46].

Thyroid Hormone

The diagnosis of thyroid disease is easily made with modern immunoassays. Thyroid stimulating hormone (TSH) is the initial test to diagnose thyroid disease linked with reproductive dysfunction. An elevated serum TSH level greater than 10 mIU/L indicates hypothyroidism.

Clinical hypothyroidism is defined by both elevated serum TSH levels and reduced serum free T_4 concentrations, while subclinical hypothyroidism describes the mild or early changes when the TSH rises before the free T_4 falls below normal. The mild elevation in TSH compensates for the early decline in T_4 production by the thyroid.

During pregnancy, trimester-specific and lab-specific reference ranges should be used to determine the optimal serum TSH range and the diagnosis of clinical hypothyroidism [47].

Although not necessary for the diagnosis of hypothyroidism, additional tests of thyroperoxidase and antithyroglobulin antibodies may influence treatment decisions by establishing underlying Hashimoto thyroiditis.

Serum TSH measurement has the highest sensitivity and specificity of any single blood test used to evaluate suspected thyrotoxicosis and should be used as an initial screening test [48]. However, when thyrotoxicosis is strongly suspected, diagnostic accuracy improves when a serum TSH, free T_4, and total T_3 are assessed at the initial evaluation [49].

The diagnosis of thyrotoxicosis is clear with elevated free T_4 and low TSH levels. However, a low TSH level with a low or normal free T_4 level has several additional diagnoses to consider other than subclinical hyperthyroidism. Diagnosis of thyrotoxicosis during pregnancy should be made more cautiously than hypothyroidism, because the serum TSH level can be low in the first trimester due to hCG-mediated thyroid stimulation. Thyroid binding globulin is elevated during a normal pregnancy and so serum total T_4 would be increased accordingly. Serum free T_4 concentration can be measured instead and will be elevated in overt hyperthyroidism.

Cortisol

ACTH-dependent or independent endogenous Cushing syndrome may need to be ruled out in certain scenarios. Cushing's syndrome may arise from a cortisol-producing tumour of the adrenal gland or from an ACTH-producing pituitary adenoma (Cushing's disease). Cushing syndrome is a rare cause of androgen excess in women and may be mistaken for PCOS. Screening for Cushing syndrome is indicated if clinical signs of Cushing's syndrome are suspected, including obesity, increased fat over the face (moon faces), trunk, and cervico-dorsal as well as supraclavicular regions, hypertension, easy bruising from thinning of the skin, impaired glucose tolerance, muscle wasting of the upper legs and arms, osteoporosis, and purple abdominal striae. Other manifestations include hirsutism, acne, and irregular menses. Mental disturbances include excessive euphoria, irritability, insomnia, and depression. Depression may occur due to excess cortisol action on the CNS limbic system.

The initial screening tests for hypercortisolism (after ruling out exogenous glucocorticoid use) include: 24-hour urine free cortisol, overnight dexamethasone suppression test, or late-night salivary cortisol [50]. Both false-positive and false-negative results are common. Consequently, tests often must be repeated before a diagnosis can be made or excluded. Once hypercortisolism is confirmed, an ACTH measurement will determine if the disease is ACTH-dependent or ACTH-independent.

Patterns of hormonal changes with different endocrine disorders are summarized in Table 8.2.2.1.

Imaging Studies

Head Imaging

This can be performed with computed tomography (CT) or more commonly with brain MRI with gadolinium enhancement to provide better details. MRI is the single best imaging procedure for most sellar masses, and there is usually no need to perform any other imaging study. Specific appearance depends on the underlying aetiology:

Craniopharyngioma: Both CT and MRI can be helpful in diagnosing craniopharyngiomas. CT can show calcification that is characteristic of the adamantinomatous subtype, whereas MRI with gadolinium enhancement provides better structural analysis.

Rathke's cleft cyst: Most cysts are small and discovered incidentally by MRI. The MRI characteristics include a sellar and/or suprasellar, symmetrical, round, or ovoid mass that enhances on either T_1 or T_2-weighted images but does not concentrate gadolinium. MRI often shows a hypointense sellar lesion with rim enhancement and a hyperintense signal in the T_2-weighted images. Hyperintensity on T_1-weighted images and isointensity on T_2-weighted images

Table 8.2.2.1 Common causes of abnormal menstrual bleeding and accompanying laboratory patterns

	LH (IU/L)	FSH (IU/L)	LH/FSH	E2 (pg/mL)	P4 (ng/mL)	AMH (ng/mL)	PRL (ng/mL)	TSH (μU/mL)	T4 (μg/dL)	DHEA-S (μg/dL)	17 OHP (ng/dL)	T (ng/dL)
Functional hypothalamic anovulation	<10	<10	~1	<50	<1	>1	Low nl	Low nl	Low nl	nl	nl	Low nl
Ovarian insufficiency menopause	>15	>15	FSH>LH	<50	<1	<0.5	nl	nl or ↑	nl or ↓	nl	nl	Low nl
PCOS	<15	<10	LH>FSH	<50	<1	nl or ↑	High nl	nl	nl	High nl	nl	High nl or slight ↑
Non-classical CAH	<15	<10	LH>FSH	<50	≤1	nl	nl	nl	nl	High nl	↑	↑
Hyperprolactinemia	<10	<10	LH<FSH	<50	<1	nl	↑	nl or ↑	nl	nl or slight ↑	nl	nl

Abbreviations: 17OHP, 17-hydroxyprogesterone; nl, normal; P4, progesterone; PRL, prolactin; T, testosterone.

suggests that mucoid material is present in the cyst; hyperintensity in both T_1- and T_2-weighted images suggests the presence of blood.

Sarcoidosis: Lesions can appear as subdural plaques or infundibular plaques, infiltrating masses, or as multiple nodules on MRI studies.

Langerhans cell histiocytosis: Findings on MRI are abnormal in most patients with Langerhans cell histiocytosis and endocrinopathy. Most common abnormalities on MRI include the loss of the posterior pituitary bright spot on T1-weighted images and infundibular enlargement. Other less common radiological abnormalities include a thickened infundibular stalk, a partially or completely empty sella, and lesions in the hypothalamus [51].

Prolactinoma: Head MRI should be performed in a patient with any degree of hyperprolactinemia to identify a mass lesion in the hypothalamic–pituitary region, unless the patient is taking a medication known to cause hyperprolactinemia. MRI may show either a micro (<1 cm) or macro (>1 cm) adenoma. Most microadenomas have lower signal intensity than the normal bright pituitary gland on T1-weighted images, and they usually do not distort the architecture of the gland. A convex outline of the pituitary gland or deviation of the pituitary stalk can also be detected [52].

The enhancement of macroprolactinomas with gadolinium is quite variable. These macroprolactinomas may distort the architecture of the gland with the inferior portion, as they also tend to invade the cavernous sinus and erode the bony floor [52].

If a mass lesion is found in the sellar region, secretion of other pituitary hormones should also be evaluated. A pituitary adenoma can cause hypersecretion of other pituitary hormones, but any sellar mass can cause hyposecretion of one or more pituitary hormones.

Pituitary apoplexy: Pituitary apoplexy can be confirmed by MRI in essentially all cases, while CT may not show this disorder in 54% to 79% of patients [53]. However, because vascular emergencies are more common than pituitary apoplexy, head CT is often the first procedure performed to evaluate the symptoms that are common to both conditions.

Sheehan's syndrome: In the early stage of Sheehan syndrome, MRI shows an enlarged pituitary; in later stages, an empty sella can be present [54], with the typical finding of intrasellar cerebrospinal fluid (CSF) filling in continuity with overlying subarachnoid spaces, and residual pituitary gland with a semi-lunate morphology flattened against the sellar floor [55].

Lymphocytic hypophysitis: MRI often shows an enlarged pituitary mimicking an adenoma, sometimes with suprasellar extension and occasionally with thickening of the infundibulum. In rare cases, an empty sella is present [56].

Functional hypothalamic amenorrhoea: If there are no clear indications or other explanations for the amenorrhea (an eating disorder such as anorexia nervosa or history of excessive exercise, or weight loss), brain MRI should be considered [1]. Empty sella syndrome can also be present as an underlying diagnosis [57]. A history of significant head trauma should raise suspicion of pituitary stalk damage and associated pituitary hormone deficiencies [1].

Ovarian Imaging

In 2003, a polycystic ovary was defined at a consensus conference in Rotterdam, Netherlands as having 12 or more follicles per ovary, measuring 2–9 mm in diameter irrespective of location, or an ovarian volume of greater than 10 cm^3 [12]. With newer equipment and advanced technology, however, the precision of counting follicles has improved. A recent task force report recommends that the threshold for polycystic ovary morphology based on follicle count should be increased to greater than 25 follicles per ovary to avoid overinterpretation of polycystic ovary morphology and misdiagnosis of PCOS [58]. This new recommendation warrants further investigation and may need to be adjusted by patient age.

While the sonographic evidence of polycystic ovaries in the presence of clinical or biochemical hyperandrogenism and/or oligoanovulation is regarded as confirmatory for PCOS, polycystic ovary morphology can occur in normal ovulatory women without a history of hyperandrogenism [59]. Moreover, the polycystic ovary should not be confused with ultrasound appearance of an ovary containing multiple small antral follicles in women recovering from hypogonadotropic hypogonadism. Therefore, the implication of a multifollicular ovary or polycystic ovary should be considered within the clinical setting. Few cysts are seen in severe hyperthecosis and the ovaries appear more solid [60].

Pelvic ultrasound (US) may be helpful to rule out structural causes for abnormal uterine bleeding such as fibroids or polyps, but currently, there are not enough data to support measuring endometrial thickness to evaluate premenopausal women with abnormal uterine bleeding. On the other hand, thin endometrium (less than 4 mm) can be seen in women with hypoestrogenic causes of amenorrhea.

When clinically indicated and supported by severe androgen excess, e.g. testosterone levels exceeding 200 ng/dl, a pelvic ultrasound is warranted to rule out ovarian virilizing tumours. In addition, some granulosa cell tumours produce oestrogen and may present with symptoms of unopposed oestrogen, depending on their age of presentation.

Virilizing tumours of the ovary are most commonly solid, non-calcified, and unassociated with ascites [61]. Steroid cell tumours, Leydig tumours, and Sertoli–Leydig cell tumours are characteristic virilizing neoplasms. Leydig cell tumours are usually small tumours (mean size 2.4 cm) that may be difficult to depict via different imaging methods, and are reportedly isoechoic to the uterus by sonography and hypoattenuating by CT [62].

Sertoli–Leydig cell tumours (mixed sex cord-stromal tumours) have a non-specific appearance. On US, these tumours usually present either as a distinct hypoechoic mass or a heterogeneous mass that is primarily solid with multiple cystic spaces. Small virilizing Sertoli–Leydig cell tumours may be easily detected using colour Doppler US rather than transvaginal US alone [62, 63].

Granulosa cell tumours are typically unilateral masses (average size, 12 cm) [62]. Although granulosa cell tumours are the most common oestrogen-producing tumours, a small subset is androgenic [62, 63]. Women typically present with hyperoestrogenicity, including abnormal uterine bleeding [62, 63].

Variability in appearance depends in part on tumour type, with Sertoli–Leydig tumours mostly solid and virilizing granulosa cell tumours more cystic [61].

Ultrasound is typically sufficient for the evaluation of adnexal masses, but MRI may be used if further characterization of an adnexal mass is needed. If metastatic disease is expected and preoperative evaluation is attempted, CT, MRI, or positron emission tomography (PET) may be needed.

Adrenal Imaging

DHEAS levels exceeding 6000–7000 ng/dl (or testosterone levels >200 ng/dl) may be associated with an androgen-producing tumour of the adrenal gland [21]. Imaging studies of the adrenal glands may be considered to locate the lesion.

Although CT remains the primary adrenal imaging procedure [64], MRI has advantages in certain clinical situations to characterize local invasion from adrenal carcinoma without exposure to ionizing radiation.

The maximum diameter of the adrenal mass is predictive of malignancy. Most adrenal adenomas are less than 4 cm in diameter. In contrast, most adrenal carcinomas are greater than 4 cm in diameter when discovered [65].

Bone Density Imaging

After 6 months of amenorrhea, clinicians should consider a baseline bone density evaluation in any adolescent or woman with functional hypothalamic amenorrhea [1].

Dual energy X-ray absorptiometry (DXA) is the most commonly used densitometric technique for adolescents and adults throughout the world because of its speed, precision, safety, low cost, and widespread availability [1]. Z-score should be used in adolescents or premenopausal women (as opposed to T-score in postmenopausal women) as recommended by the International Society for Clinical Densitometry. A Z-score of –2.0 or lower is defined as 'below the expected range for age', and a Z-score above –2.0 is 'within the expected range for age' [66]. The Z-score compares the bone density measure to age-, sex-, and race-matched controls.

The goal of bone densitometry is to identify individuals with low bone mass who are at increased risk for skeletal fragility, and to guide and monitor treatment [67]. Baseline bone mineral density at –2.0 or less at any skeletal site has the greatest risk for fractures, and in these patients, clinicians should attentively monitor nutritional intake and a patient's skeletal status [67].

The spine (a trabecular-rich site) is the most common site of low bone density in adolescents and young women with amenorrhea and also predicts fracture risk; it is therefore an important site to monitor [68–71]. In older adolescents (above age 15 years) and women with functional hypothalamic amenorrhea, measuring hip bone density gives information about weight-bearing cortical bone and can be useful to monitor bone density longitudinally [67].

In one study of 130 young females with anorexia nervosa (mean age 24 years), more than 90% had osteopenia and nearly 40% met World Health Organization (WHO) criteria for osteoporosis (T-score ≤ 2.5) [72].

Patients with primary ovarian insufficiency on average have 2–3% lower bone mineral density at L1–L4, femoral neck, and total hip. More than 1-year delay in the diagnosis of oestrogen deficiency is a risk factor for reduced bone mineral density below the expected range for age (Z-score < -2) [73].

Thyroid Imaging

Radioactive iodine scans are not necessary to make the diagnosis of hyperthyroidism but are most useful for distinguishing Graves' disease from toxic adenoma, thyroiditis, and factitious disorder. Usually, clinical factors often obviate the need for diagnostic scans, especially since radioactive iodine scans are contraindicated during pregnancy.

In pregnancy, breastfeeding, or women trying to conceive, ultrasonography with Doppler blood flow may be useful for distinguishing Graves' disease from thyroiditis. The need for experienced technicians and radiologists, however, limits the ability to recommend the general use of ultrasonography for this purpose [47].

Other Tests

These tests may be indicated when encountering a diagnosis of hypergonadotropic hypogonadism related to spontaneous primary ovarian insufficiency (i.e. non-iatrogenic). Insulin resistance testing is also listed to be considered when a diagnosis of PCOS is encountered.

Karyotype

Turner syndrome occurs in 1 in 2500–3000 girls and young women [74]. Approximately 50% of patients have a monosomy X with complete loss of one X chromosome (45,X) [75], while the rest may have a mosaic chromosomal complement of 45, X. Therefore, once the diagnosis of primary ovarian insufficiency is established, a karyotype with 30 cells counted is recommended to identify women who have primary ovarian insufficiency related to abnormal karyotype [76].

Fragile X Testing

Premutation in the FMR1 gene (expansion of CGG trinucleotide repeats to 55–200 repeats) can lead to increased risk of primary ovarian insufficiency [77]. FMR1 premutation is found in 14% of patients with familial primary ovarian insufficiency and the prevalence is about 2% in sporadic cases [78]. Testing for FMR1 premutation is helpful to rule out fragile X associated primary ovarian failure and is indicated in these situations.

Autoimmune Testing

Autoimmune primary ovarian insufficiency is associated with multiple endocrine disorders (i.e. hypothyroidism, adrenal insufficiency, hypoparathyroidism, diabetes mellitus, hypophysitis) and non-endocrine abnormalities (chronic candidiasis, idiopathic thrombocytopaenic purpura, vitiligo, alopecia, autoimmune haemolytic anaemia, pernicious anaemia, systemic lupus erythematosus, rheumatoid arthritis, Crohn's disease, Sjögren's syndrome, primary biliary cirrhosis, and chronic active hepatitis) [79–81].

Testing for associated autoimmune conditions in women diagnosed with primary ovarian insufficiency is important to identify the mechanism of the disease (autoimmune lymphocytic oophoritis) [82], and to treat or monitor patients who have these other autoimmune disorders (i.e. polyglandular autoimmune endocrinopathies). These tests include checking for the presence of adrenal autoantibodies by either 21-hydroxylase (CYP21) immunoprecipitation or indirect immunofluorescence [76].

Measurement of serum thyroid autoantibodies is reasonable as thyroid autoimmune disease, most commonly Hashimoto's thyroiditis, is present in 14–27% of women at initial diagnosis [76, 83, 84].

Testing for diabetes with fasting/postprandial glucose or haemoglobin A1c appears justified in those patients with normal karyotype and spontaneous primary ovarian insufficiency [83].

Box 8.2.2.1 Tests for patients with confirmed ovarian insufficiency

Karyotype
Fragile X testing
Adrenal autoantibodies
Thyroid autoantibodies
Testing for diabetes
Testing for pernicious anaemia
Serum calcium
Vitamin D levels
Bone density scan

Other tests to rule out associated autoimmune disorders may be considered, including complete blood count, chemistry panel levels to rule out pernicious anaemia, and hypoparathyroidism, respectively.

These tests that can be ordered in patients with primary ovarian insufficiency are listed in Box 8.2.2.1.

Insulin Resistance Testing

About 60–95% of women with PCOS have insulin resistance [85], as do most women with simple obesity. Therefore, ascertainment of metabolic dysfunction is warranted, particularly in these high-risk individuals. Determining insulin resistance in the presence of euglycemia is limited by the lack of a highly sensitive test that is practical for population-based clinical use [86–88]. Nevertheless, an oral glucose tolerance test is simple and is the best screening measure for detecting glucose intolerance or diabetes, particularly in PCOS women who often have normal fasting glucose levels [89].

Haemoglobin A1C was a strong predictor of the development of glucose-defined diabetes during the Diabetes Prevention Program [90] and its follow-up [91]. Hence, it is reasonable to consider an A1C range of 5.7–6.4% as identifying individuals with prediabetes. Prediabetes should not be viewed as a clinical entity in its own right but rather as an increased risk for diabetes and cardiovascular disease [92]. Similar to those with impaired fasting glucose and/or impaired glucose tolerance, individuals with A1C of 5.7–6.4% should be informed of their increased risk for diabetes and cardiovascular risk and counselled about effective strategies to lower their risks [92]. Haemoglobin A1C of 6.5% or greater can be used as a diagnostic test for diabetes, as suggested by the current guidelines from the American Diabetes Association [92].

The efficacy of interventions for primary prevention of type 2 diabetes [93, 94] has primarily been demonstrated among individuals who have impaired glucose tolerance with or without elevated fasting glucose, not for individuals with isolated impaired fasting glucose or for those with prediabetes defined by A1C criteria.

From a cardiovascular risk perspective, a 2-hour 75-gram oral glucose tolerance test (Table 8.2.2.2), along with a fasting lipid panel (with target levels based on cardiovascular disease risk factors) [95, 96]) should be performed in obese individuals or in lean PCOS women with advanced age (40 years), personal history of gestational diabetes, family history of type 2 diabetes, or other cardiovascular disease risk factors (i.e. smoking, hypertension, etc) [97].

Women with normal glucose tolerance and lipid levels can be rescreened every 2 years or sooner if additional risk factors are identified, while those with impaired glucose tolerance should be screened annually to rule out developing type 2 diabetes [98].

Table 8.2.2.2 Two-hour glucose tolerance test values (values in mg/dl)

	Fasting	2 hour
Normal	<100	<140
Impaired fasting glucose	100–125*	
Impaired glucose tolerance		140–199
Diabetes	≥126	≥200

* It should be noted that the World Health Organization (WHO) and numerous other diabetes organizations define the impaired fasting hyperglycaemia (IFG) cut-off at 110 mg/dl.

REFERENCES

1. Gordon CM, Ackerman KE, Berga SL, *et al*. Functional hypothalamic amenorrhea: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 1413–39.
2. Berga SL, Mortola JF, Girton L, *et al*. Neuroendocrine aberrations in women with functional hypothalamic amenorrhea. *J Clin Endocrinol Metab*, 1989; **68**: 301–8.
3. Gordon CM. Clinical practice. Functional hypothalamic amenorrhea. *N Engl J Med*, 2010; **363**: 365–71.
4. Taylor AE, McCourt B, Martin KA, *et al*. Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 1997; **82**: 2248–56.
5. Kwee J, Schats R, McDonnell J, Lambalk CB, Schoemaker J. Intercycle variability of ovarian reserve tests: results of a prospective randomized study. *Hum Reprod*, 2004; **19**: 590–5.
6. Martin KA, Anderson RR, Chang RJ, *et al*. Evaluation and treatment of hirsutism in premenopausal women: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1233–57.
7. Azziz R, Sanchez LA, Knochenhauer ES, *et al*. Androgen excess in women: experience with over 1000 consecutive patients. *J Clin Endocrinol Metab*, 2004; **89**: 453–62.
8. Azziz R, Nestler JE, Dewailly D, ed. *Androgen Excess Disorders in Women*. Philadelphia, PA: Lippincott-Raven, 1997.
9. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*, 2007; **92**: 405–13.
10. Chang WY, Knochenhauer ES, Bartolucci AA, Azziz R. Phenotypic spectrum of polycystic ovary syndrome: clinical and biochemical characterization of the three major clinical subgroups. *Fertil Steril*, 2005; **83**: 1717–23.
11. Salameh WA, Redor-Goldman MM, Clarke NJ, Mathur R, Azziz R, Reitz RE. Specificity and predictive value of circulating testosterone assessed by tandem mass spectrometry for the diagnosis of polycystic ovary syndrome by the National Institutes of Health 1990 criteria. *Fertil Steril*, 2014; **101**: 1135–41 e1132.
12. Rotterdam. ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*, 2004; **81**: 19–25.
13. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*, 2007; **92**: 405–13.
14. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*, 1999; **84**: 3666–72.

15. Pasquali R, Zanotti L, Fanelli F, *et al.* Defining hyperandrogenism in women with polycystic ovary syndrome: a challenging perspective. *J Clin Endocrinol Metab*, 2016; **101**: 2013–22.
16. Dumesic DA. Amenorrhea, oligomenrrhea, and hyperandrogenic disorders. In: Hacker N, GJ, Hobel C, eds. *Hacker & Moore's Essentials of Obstetrics and Gynecology*, 5th edn. Philadelphia, PA: Elsevier Saunders, 2016: Chapter 33.
17. O'Reilly MW, Taylor AE, Crabtree NJ, *et al.* Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *J Clin Endocrinol Metab*, 2014; **99**: 1027–36.
18. Wild RA, Umstot ES, Andersen RN, Ranney GB, Givens JR. Androgen parameters and their correlation with body weight in one hundred thirty-eight women thought to have hyperandrogenism. *Am J Obstet Gynecol*, 1983; **146**: 602–6.
19. Zumoff B, Walsh BT, Katz JL, Levin J, Rosenfeld RS, Kream J, Weiner H. Subnormal plasma dehydroisoandrosterone to cortisol ratio in anorexia nervosa: a second hormonal parameter of ontogenic regression. *J Clin Endocrinol Metab*, 1983; **56**: 668–72.
20. Devesa J, Perez-Fernandez R, Bokser L, Gaudiero GJ, Lima L, Casanueva FF. Adrenal androgen secretion and dopaminergic activity in anorexia nervosa. *Horm Metab Res*, 1988; **20**: 57–60.
21. Terzolo M, Ali A, Osella G, Reimondo G, Pia A, Peretti P, Paccotti P, Angeli A. The value of dehydroepiandrosterone sulfate measurement in the differentiation between benign and malignant adrenal masses. *Eur J Endocrinol*, 2000; **142**: 611–17.
22. Abraham GE, Chakmakjian ZH. Serum steroid levels during the menstrual cycle in a bilaterally adrenalectomized woman. *J Clin Endocrinol Metab*, 1973; **37**: 581–7.
23. White PC, Speiser PW. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Endocr Rev*, 2000; **21**: 245–91.
24. Levine LS, Dupont B, Lorenzen F, *et al.* Cryptic 21-hydroxylase deficiency in families of patients with classical congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 1980; **51**: 1316–24.
25. Speiser PW. Congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Endocrinol Metab Clin N Am*, 2001; **30**: 31–59, vi.
26. Speiser PW, Knochenhauer ES, Dewailly D, Fruzzetti F, Marcondes JA, Azziz R. A multicenter study of women with nonclassical congenital adrenal hyperplasia: relationship between genotype and phenotype. *Mol Genet Metab*, 2000; **71**: 527–34.
27. Schlechte JA. Clinical practice. Prolactinoma. *N Engl J Med*, 2003; **349**: 2035–41.
28. Casanueva FF, Molitch ME, Schlechte JA, *et al.* Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. *Clin Endocrinol (Oxf)*, 2006; **65**: 265–73.
29. St-Jean E, Blain F, Comtois R. High prolactin levels may be missed by immunoradiometric assay in patients with macroprolactinomas. *Clin Endocrinol (Oxf)*, 1996; **44**: 305–9.
30. Seppala M, Ranta T, Hirvonen E. Hyperprolactinaemia and luteal insufficiency. *Lancet*, 1976; **1**: 229–30.
31. Corenblum B, Taylor PJ. The hyperprolactinemic polycystic ovary syndrome may not be an distinct entity. *Fertil Steril*, 1982; **38**: 549–52.
32. Falaschi P, del Pozo E, Rocco A, Toscano V, Petrangeli E, Pompei P, Frajese G. Prolactin release in polycystic ovary. *Obstet Gynecol*, 1980; **55**: 579–82.
33. Katznelson L, Laws ER, Jr., Melmed S, *et al.* Acromegaly: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 3933–51.
34. Frystyk J, Freda P, Clemmons DR. The current status of IGF-I assays—a 2009 update. *Growth Horm IGF Res*, 2010; **20**: 8–18.
35. Carmichael JD, Bonert VS, Mirocha JM, Melmed S. The utility of oral glucose tolerance testing for diagnosis and assessment of treatment outcomes in 166 patients with acromegaly. *J Clin Endocrinol Metab*, 2009; **94**: 523–27.
36. Visser JA, Themmen AP. Anti-Mullerian hormone and folliculogenesis. *Mol Cell Endocrinol*, 2005; **234**: 81–6.
37. Lee MM, Donahoe PK, Hasegawa T, *et al.* Mullerian inhibiting substance in humans: normal levels from infancy to adulthood. *J Clin Endocrinol Metab*, 1996; **81**: 571–6.
38. Hudson PL, Dougas I, Donahoe PK, *et al.* An immunoassay to detect human mullerian inhibiting substance in males and females during normal development. *J Clin Endocrinol Metab*, 1990; **70**: 16–22.
39. Weenen C, Laven JS, Von Bergh AR, *et al.* Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum Reprod*, 2004; **10**: 77–83.
40. Jeppesen JV, Anderson RA, Kelsey TW, *et al.* Which follicles make the most anti-Mullerian hormone in humans? Evidence for an abrupt decline in AMH production at the time of follicle selection. *Mol Hum Reprod*, 2013; **19**: 519–27.
41. Kwee J, Schats R, McDonnell J, Themmen A, de Jong F, Lambalk C. Evaluation of anti-Mullerian hormone as a test for the prediction of ovarian reserve. *Fertil Steril*, 2008; **90**: 737–43.
42. Ebner T, Sommergruber M, Moser M, Shebl O, Schreier-Lechner E, Tews G. Basal level of anti-Mullerian hormone is associated with oocyte quality in stimulated cycles. *Hum Reprod*, 2006; **21**: 2022–6.
43. Kushnir VA, Halevy N, Barad DH, Albertini DF, Gleicher N. Relative importance of AMH and androgens changes with aging among non-obese women with polycystic ovary syndrome. *J Ovarian Res*, 2015; **8**: 45.
44. Pigny P, Merlen E, Robert Y, *et al.* Elevated serum level of anti-mullerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. *J Clin Endocrinol Metab*, 2003; **88**: 5957–62.
45. Pigny P, Jonard S, Robert Y, Dewailly D. Serum anti-Mullerian hormone as a surrogate for antral follicle count for definition of the polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2006; **91**: 941–5.
46. Rosenfield RL, Wroblewski K, Padmanabhan V, Littlejohn E, Mortensen M, Ehrmann DA. Antimullerian hormone levels are independently related to ovarian hyperandrogenism and polycystic ovaries. *Fertil Steril*, 2012; **98**: 242–9.
47. Alexander EK, Pearce EN, Brent GA, *et al.* 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
48. de los Santos ET, Starich GH, Mazzaferri EL. Sensitivity, specificity, and cost-effectiveness of the sensitive thyrotropin assay in the diagnosis of thyroid disease in ambulatory patients. *Arch Intern Med*, 1989; **149**: 526–32.
49. Ross DS, Burch HB, Cooper DS, *et al.* 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. *Thyroid*, 2016; **26**: 1343–421.
50. Nieman LK, Biller BM, Findling JW, *et al.* The diagnosis of Cushing's syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2008; **93**: 1526–40.

51. Makras P, Samara C, Antoniou M, *et al*. Evolving radiological features of hypothalamo-pituitary lesions in adult patients with Langerhans cell histiocytosis (LCH). *Neuroradiology*, 2006; **48**: 37–44.
52. Imaoka I, Wada A, Matsuo M, Yoshida M, Kitagaki H, Sugimura K. MR imaging of disorders associated with female infertility: use in diagnosis, treatment, and management. *Radiographics*, 2003; **23**: 1401–21.
53. Verrees M, Arafah BM, Selman WR. Pituitary tumor apoplexy: characteristics, treatment, and outcomes. *Neurosurgical Focus*, 2004; **16**: E6.
54. Fleckman AM, Schubart UK, Danziger A, Fleischer N. Empty sella of normal size in Sheehan's syndrome. *Am J Med*, 1983; **75**: 585–91.
55. Chiloiro S, Giampietro A, Bianchi A, *et al*. Diagnosis of endocrine disease: primary empty sella: a comprehensive review. *Eur J Endocrinol*, 2017; **177**: R275–85.
56. Unluhizarci K, Bayram F, Colak R, *et al*. Distinct radiological and clinical appearance of lymphocytic hypophysitis. *J Clin Endocrinol Metab*, 2001; **86**: 1861–4.
57. Guitelman M, Garcia Basavilbaso N, Vitale M, *et al*. Primary empty sella (PES): a review of 175 cases. *Pituitary*, 2013; **16**: 270–4.
58. Dewailly D, Lujan ME, Carmina E, *et al*. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update*, 2014; **20**: 334–52.
59. Polson DW, Adams J, Wadsworth J, Franks S. Polycystic ovaries—a common finding in normal women. *Lancet*, 1988; **1**: 870–2.
60. Rousset P, Gompel A, Christin-Maitre S, Pugeat M, Hugol D, Ghossain MA, Buy JN. Ovarian hyperthecosis on grayscale and color Doppler ultrasound. *Ultrasound Obstet Gynecol*, 2008; **32**: 694–9.
61. Outwater EK, Marchetto B, Wagner BJ. Virilizing tumors of the ovary: imaging features. *Ultrasound Obstet Gynecol*, 2000; **15**: 365–71.
62. Horta M, Cunha TM. Sex cord-stromal tumors of the ovary: a comprehensive review and update for radiologists. *Diagn Interv Radiol*, 2015; **21**: 277–86.
63. Outwater EK, Wagner BJ, Mannion C, McLarney JK, Kim B. Sex cord-stromal and steroid cell tumors of the ovary. *Radiographics*, 1998; **18**: 1523–46.
64. Ilias I, Sahdev A, Reznek RH, Grossman AB, Pacak K. The optimal imaging of adrenal tumours: a comparison of different methods. *Endocr Relat Cancer*, 2007; **14**: 587–99.
65. Angeli A, Osella G, Ali A, Terzolo M. Adrenal incidentaloma: an overview of clinical and epidemiological data from the National Italian Study Group. *Horm Res*, 1997; **47**: 279–83.
66. Leslie WD, Adler RA, El-Hajj Fuleihan G, *et al*. Application of the 1994 WHO classification to populations other than postmenopausal Caucasian women: the 2005 ISCD Official Positions. *J Clin Densitom*, 2006; **9**: 22–30.
67. Crabtree NJ, Arabi A, Bachrach LK, *et al*. Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: the revised 2013 ISCD Pediatric Official Positions. *J Clin Densitom*, 2014; **17**: 225–42.
68. Gordon CM, Goodman E, Emans SJ, *et al*. Physiologic regulators of bone turnover in young women with anorexia nervosa. *J Pediatr*, 2002; **141**: 64–70.
69. Bachrach LK, Guido D, Katzman D, Litt IF, Marcus R. Decreased bone density in adolescent girls with anorexia nervosa. *Pediatrics*, 1990; **86**: 440–7.
70. Ackerman KE, Nazem T, Chapko D, *et al*. Bone microarchitecture is impaired in adolescent amenorrheic athletes compared with eumenorrheic athletes and nonathletic controls. *J Clin Endocrinol Metab*, 2011; **96**: 3123–33.
71. Mitchell DM, Tuck P, Ackerman KE, *et al*. Altered trabecular bone morphology in adolescent and young adult athletes with menstrual dysfunction. *Bone*, 2015; **81**: 24–30.
72. Grinspoon S, Thomas E, Pitts S, *et al*. Prevalence and predictive factors for regional osteopenia in women with anorexia nervosa. *Ann Intern Med*, 2000; **133**: 790–4.
73. Popat VB, Calis KA, Vanderhoof VH, *et al*. Bone mineral density in estrogen-deficient young women. *J Clin Endocrinol Metab*, 2009; **94**: 2277–83.
74. Bondy CA, Turner Syndrome Study G. Care of girls and women with Turner syndrome: a guideline of the Turner Syndrome Study Group. *J Clin Endocrinol Metab*, 2007; **92**: 10–25.
75. Wolff DJ, Van Dyke DL, Powell CM, Working Group of the ALQAC. Laboratory guideline for Turner syndrome. *Genet Med*, 2010; **12**: 52–5.
76. Nelson LM. Clinical practice. Primary ovarian insufficiency. *N Engl J Med*, 2009; **360**: 606–14.
77. Wittenberger MD, Hagerman RJ, Sherman SL, *et al*. The FMR1 premutation and reproduction. *Fertil Steril*, 2007; **87**: 456–65.
78. Sherman SL. Premature ovarian failure in the fragile X syndrome. *Am J Med Genet*, 2000; **97**: 189–94.
79. Betterle C, Dal Pra C, Mantero F, Zanchetta R. Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocr Rev*, 2002; **23**: 327–64.
80. Ebrahimi M, Akbari Asbagh F. Pathogenesis and causes of premature ovarian failure: an update. *Int J Fertil Steril*, 2011; **5**: 54–65.
81. Komorowska B. Autoimmune premature ovarian failure. *Przeglad Menopauzalny*, 2016; **15**: 210–14.
82. Bakalov VK, Anasti JN, Calis KA, *et al*. Autoimmune oophoritis as a mechanism of follicular dysfunction in women with 46,XX spontaneous premature ovarian failure. *Fertil Steril*, 2005; **84**: 958–65.
83. Kim TJ, Anasti JN, Flack MR, Kimzey LM, Defensor RA, Nelson LM. Routine endocrine screening for patients with karyotypically normal spontaneous premature ovarian failure. *Obstet Gynecol*, 1997; **89**: 777–9.
84. Hoek A, Schoemaker J, Drexhage HA. Premature ovarian failure and ovarian autoimmunity. *Endocr Rev*, 1997; **18**: 107–34.
85. Dumesic DA, Oberfield SE, Stener-Victorin E, Marshall JC, Laven JS, Legro RS. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocr Rev*, 2015; **36**: 487–525.
86. Legro RS, Finegood D, Dunaif A. A fasting glucose to insulin ratio is a useful measure of insulin sensitivity in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 1998; **83**: 2694–8.
87. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 1985; **28**: 412–19.
88. Katz A, Nambi SS, Mather K, *et al*. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J Clin Endocrinol Metab*, 2000; **85**: 2402–10.
89. Salley KE, Wickham EP, Cheang KI, Essah PA, Karjane NW, Nestler JE. Glucose intolerance in polycystic ovary syndrome—a position statement of the Androgen Excess Society. *J Clin Endocrinol Metab*, 2007; **92**: 4546–56.

90. Ackermann RT, Cheng YJ, Williamson DF, Gregg EW. Identifying adults at high risk for diabetes and cardiovascular disease using hemoglobin A1c National Health and Nutrition Examination Survey 2005–2006. *Am J Prev Med*, 2011; **40**: 11–17.
91. Diabetes Prevention Program Research G. HbA1c as a predictor of diabetes and as an outcome in the diabetes prevention program: a randomized clinical trial. *Diabetes Care*, 2015; **38**: 51–8.
92. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41**: S13–27.
93. Knowler WC, Barrett-Connor E, Fowler SE, *et al*. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*, 2002; **346**: 393–403.
94. Tuomilehto J, Lindstrom J, Eriksson JG, *et al*. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*, 2001; **344**: 1343–50.
95. Catapano AL, Graham I, De Backer G, *et al*. 2016 ESC/EAS Guidelines for the Management of Dyslipidaemias. *Eur Heart J*, 2016; **37**: 2999–3058.
96. Jellinger PS, Handelsman Y, Rosenblit PD, *et al*. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for management of dyslipidemia and prevention of cardiovascular disease. *Endocr Pract*, 2017; **23**: 1–87.
97. Wild RA, Carmina E, Diamanti-Kandarakis E, *et al*. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab*, 2010; **95**: 2038–49.
98. Genuth S, Alberti KG, Bennett P, *et al*. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*, 2003; **26**: 3160–7.

8.3

Female Reproductive Endocrinology

8.3.1 Disorders of Gonadotropin Secretion

Sarah L. Berga

Introduction

Folliculogenesis and ovulation require appropriate gonadotropin-releasing hormone (GnRH) input to drive appropriate pituitary secretion of LH and FSH. Ovarian dysfunction either causes or is caused by altered GnRH drive. The myriad of factors that modulate GnRH drive convey information about (1) developmental stage; (2) ovarian function and reserve, including stage of folliculogenesis; (3) energy balance, body composition, and metabolism; (4) circadian and circannual position; (5) stress and emotional states; and (6) cognitive assessments. When hypoestrogenism results from diminished or absent follicle pool, maximal GnRH-LH pulsatility ensues and reflects insensitivity to oestradiol negative feedback [1]. Polycystic ovary syndrome may be viewed as masculinization of the GnRH 'pulse generator' that results in an increase in LH pulse frequency and amplitude accompanied by a decrease in FSH that leads to chronic anovulation and thecal-stromal ovarian hyperstimulation [2]. GnRH-LH drive that is chronically too slow to support full folliculogenesis results in hypothalamic hypogonadism that reflects increased sensitivity to estradiol negative feedback.

Clinically, reduced GnRH drive results in a spectrum of ovarian compromise, ranging from luteal insufficiency to chronic anovulation and a variety of menstrual patterns such as amenorrhoea, polymenorrhoea, menorrhagia, and oligomenorrhoea (Figures 8.3.1.1 to 8.3.1.4). Rarely, there is an organic or congenital cause for reduced GnRH drive, such as brain tumours, coeliac disease, or migration of an insufficient number of GnRH neurons from the olfactory placode into the hypothalamus during fetal development [3].

Typically, the cause of hypothalamic hypogonadism is functional, that is, due to the endocrine consequences of psychological or behavioural variables. Anorexia nervosa provides the most dramatic example, but most women who develop functional hypothalamic hypogonadism (FHH) or functional hypothalamic anovulation and amenorrhea (FHA) do not meet diagnostic criteria for syndromal eating or psychiatric disorders. The nomenclature used to describe FHH reflects behavioural antecedents and also includes exercise amenorrhoea, stress-induced anovulation, functional hypothalamic chronic anovulation, and hypothalamic hypoestrogenism. Reifenstein [4] first introduced the term 'psychogenic amenorrhoea'. Despite the multiplicity of names, the pathogenesis of anovulation is similar. Typically FHH presents as secondary amenorrhoea. The diagnosis requires exclusion of organic causes. The term 'functional' also implies that amelioration of causal behavioural factors will restore ovulatory ovarian function. Support for the concept that the proximate cause of the anovulation in FHA is insufficient GnRH drive was initially provided by showing that administration of exogenous pulsatile GnRH restored folliculogenesis [5–7]. The reduction in GnRH pulsatility manifests in the circulation and can be detected as a reduction in LH pulsatility [8].

Appropriate treatment depends on understanding the psychoneuroendocrinology of FHH. The primary cause of anovulation in FHH is insufficient GnRH drive that reflects altered brain modulation of hypothalamic function in response to perceived or actual psychogenic and metabolic stressors. As shown in Figure 8.3.1.1, FHH is more than diminished GnRH drive and comprises a constellation of neuroendocrine secretory adaptations including obligatory activation of the hypothalamic–pituitary–adrenal (HPA) axis that aligns reproduction with circumstance [8]. Not surprisingly, reduced HPA activation precedes recovery of ovarian function [9, 10]. Appropriate treatment of FHA requires recognizing the critical role GnRH drive for ovarian function and the role of behavioural factors in activating the HPA axis.

Regulation of the GnRH Pulse Generator

The GnRH pulse generator refers to the collective functioning of GnRH neurons residing in the mediobasal hypothalamus that

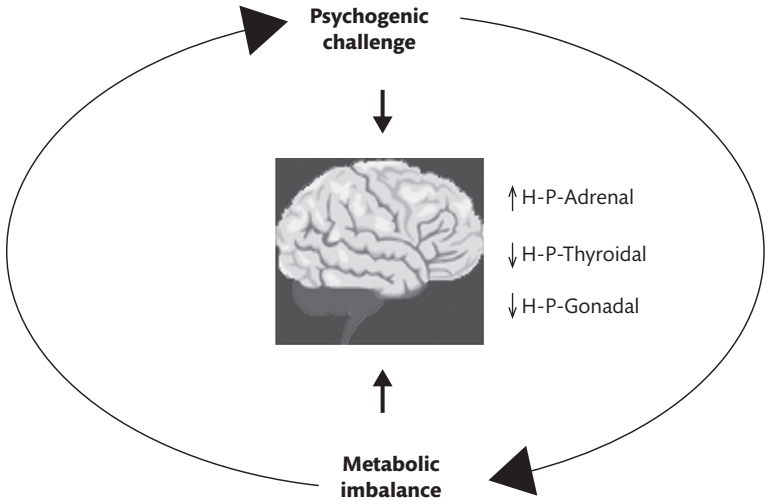


Figure 8.3.1.1 In stress-induced anovulation/functional hypothalamic amenorrhoea, energetic imbalance, and psychosocial challenge synergistically amplify the independent impact of the other upon neuroendocrine outputs to thereby suppress the hypothalamic–pituitary–ovarian (HPO) and hypothalamic–pituitary–thyroidal (HPT) axes, while concomitantly activating the hypothalamic–pituitary–adrenal (HPA) axis.

secrete GnRH into the portal vasculature and thereby regulate LH and FSH release from pituitary gonadotrophs (Figure 8.3.1.2). GnRH neurons are endogenously pulsatile. However, for the bolus of GnRH to be sufficient to trigger gonadotropins release, a group of GnRH neurons must secrete synchronously. Bolus size reflects the cohort size of GnRH-to-GnRH synapses or appositions. The GnRH pulse generator is active during fetal life and then is suppressed by central processes that cause enhanced negative feedback sensitivity to oestradiol until the onset of puberty. HPA activation increases the sensitivity of GnRH to oestradiol inhibition [11]. Figure 8.3.1.3 illustrates a hypothetical scheme to explain how different pulse frequencies might result from varying the size of the GnRH cohort that synchronously releases GnRH pulses into the portal circulation to drive the pituitary release of LH and FSH (Figures 8.3.1.5 and 8.3.1.6).

Figure 8.3.1.7 illustrates the proximate neuromodulatory cascade that alters GnRH function [12]. Behavioural activation of the HPA axis reduces GnRH drive and results in FHH. Hypercortisolaemia, as reflected by increased urinary free cortisol and elevated cerebrospinal fluid levels of cortisol [13], has been demonstrated in amenorrhoeic athletes and non-athletes [8, 10, 14, 15]. Acute nutritional deprivation activates the HPA axis and reduces LH pulsatility [16]. Given the energetic expense of reproduction, it is not surprising that metabolic factors gate reproductive function. Psychosocial influences, including external stressors and attitudes, heighten behavioural reactivity [17–19] and alter the

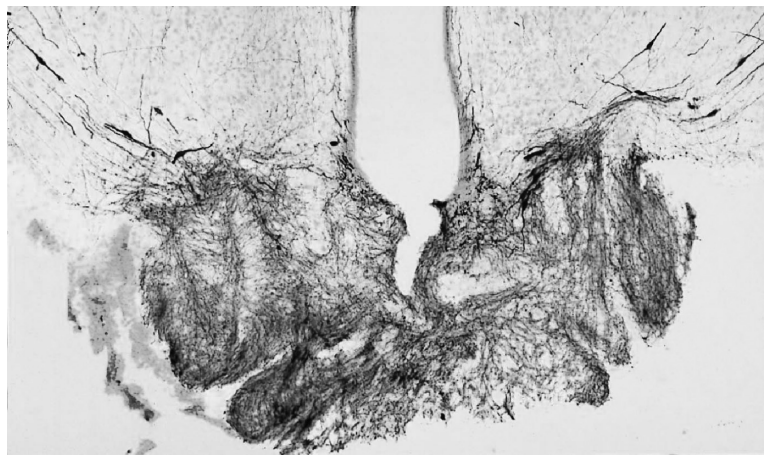


Figure 8.3.1.2 Neuroanatomical photomicrograph of gonadotrophin-releasing hormone pulse generator. Shown is the hypothalamic region from a rhesus monkey. Median eminence and adjacent basal hypothalamus are stained for gonadotrophin-releasing hormone in brown and counterstained with a methyl-grey Nissl stain. Gonadotrophin-releasing hormone neurons are visible at the border of the median eminence, within the median eminence, and within the hypothalamus. The dense accumulation of gonadotrophin-releasing hormone axons occurs where gonadotrophin-releasing hormone axons converge upon the portal loops that carry the gonadotrophin-releasing hormone to the pituitary.

Reproduced with permission from Berga SL *et al.* Secondary amenorrhoea, Chapter 2. In: *Atlas of Clinical Gynecology, Reproductive Endocrinology*, Vol III (Stenchever MA, Series Editor; Mishell DR Jr, Volume Editor). Current Medicine, Inc., Philadelphia, PA, USA, 1999, with special thanks to Gloria Hoffman, PhD who performed and contributed the work.

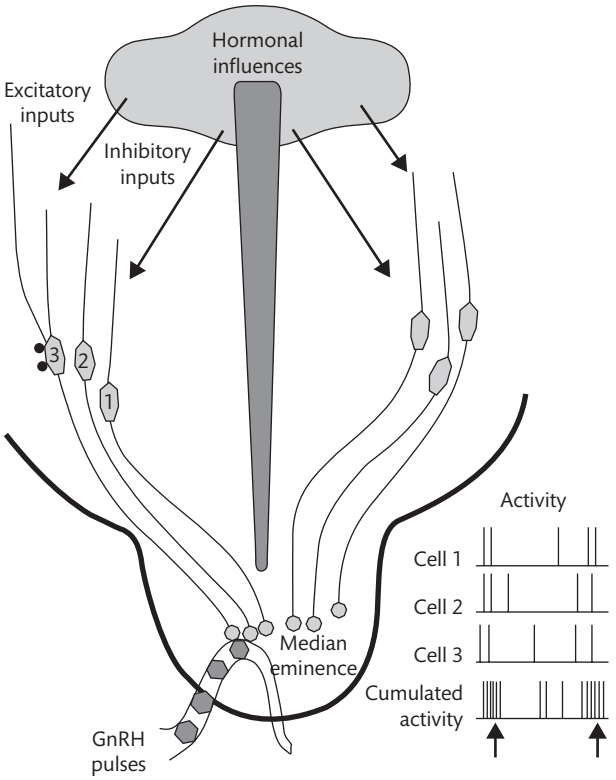


Figure 8.3.1.3 Diagram of the gonadotropin-releasing hormone pulse generator and its modulation. The gonadotrophin-releasing hormone pulse generator and mechanisms of its control are illustrated. The hypothetical activity of three gonadotrophin-releasing hormone neurons is shown in the lower right. When the activity of the neurons is summed at the level of the median eminence, bursts of activity that result in pulses of gonadotrophin-releasing hormone (short arrows) are produced. Some of the factors that influence the firing or release of gonadotrophin-releasing hormone are also illustrated. The gonadotrophin-releasing hormone neurons shown in this figure receive input from other gonadotrophin-releasing hormone neurons, both dendritic (in the form of bridges) and axonal (indicated by the terminal from the unlabelled neuron to cell 3). These cells also receive extrinsic excitatory and inhibitory inputs from neurons in various regions of the brain whose activity, in turn, is greatly modified by gonadal steroids and nutritional signals. In addition, pulse generator activity can be affected by changes in glial investment of gonadotrophin-releasing hormone soma and terminals that influence cohort size. Glia also can reversibly isolate gonadotrophin-releasing hormone cells from extrinsic influences or prevent gonadotrophin-releasing hormone from gaining easy access to the portal vasculature.

Reproduced with permission from Berga SL *et al.* Secondary amenorrhoea, Chapter 2. In: *Atlas of Clinical Gynecology, Reproductive Endocrinology*, Vol III (Stenchever MA, Series Editor; Mishell DR Jr, Volume Editor). Current Medicine, Inc., Philadelphia, PA, USA, 1999, with special thanks to Gloria Hoffman, PhD who contributed this diagram.

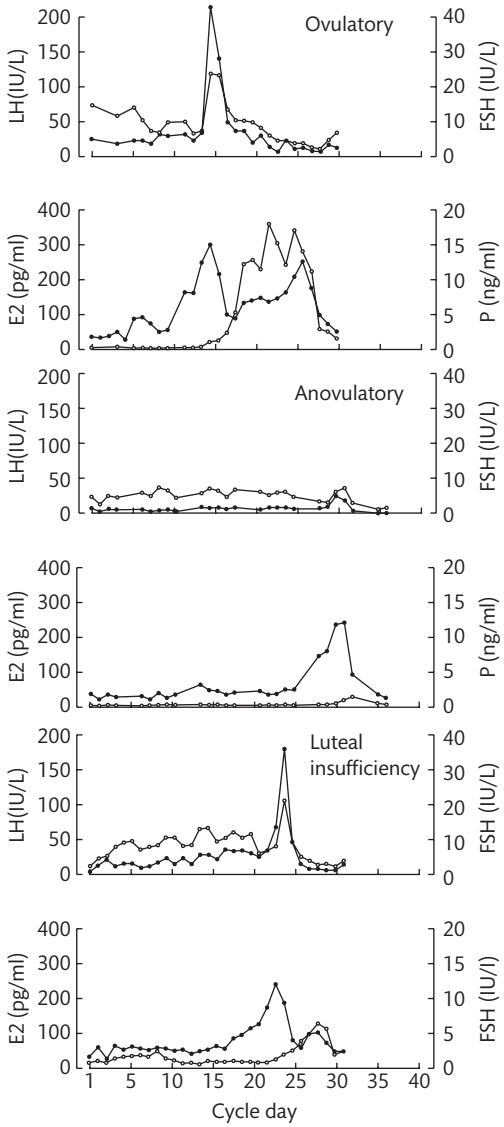


Figure 8.3.1.4 Graph of three types of menstrual cycles: ovulatory, anovulatory, and luteal insufficiency. Blood samples were taken daily or nearly daily across a menstrual interval (from the first day of vaginal bleeding of one cycle to the first day of bleeding of the next cycle) in three women between ages 20–30 years who reported regularly regular menstrual cycles. The top panels for each menstrual cycle type show the gonadotrophins, luteinizing hormone (closed circles) and FSH (open circles). The bottom panels show oestradiol (closed circles) and progesterone levels (open circles). In the ovulatory cycle (top two panels), FSH was highest in the early follicular phase and the luteinizing hormone surge occurred as expected at midcycle. There was the expected preovulatory exponential rise in oestradiol and roughly 14 days of progesterone secretion above 1 ng/ml after the midcycle surge. In the anovulatory cycle (middle two panels), gonadotrophin levels were relatively low, indicating that the cause of the anovulation was not follicular depletion. Late in the cycle, there was an elevation and decline in oestradiol unaccompanied by any significant progesterone secretion. The decline in oestradiol triggered withdrawal bleeding, leading to anovulatory cycling. In the cycle labelled luteal insufficiency (bottom two panels), there was a luteinizing hormone surge, but it occurred around day 22 and progesterone secretion occurred for only 6 days. None of these cycles were clinically distinguishable from the other and all women assumed that they were having normal menstrual cycles although the hormone levels reveal that they were not.

neuroendocrine cascade that modulates GnRH drive [20, 21]. Endocrine-disrupting chemicals, such as bisphenol A and some polychlorinated biphenyls, may sensitize GnRH and kisspeptin neuroregulatory cascades to stressors by modulating GnRH gene transcription and/or by acting as oestrogen agonists or antagonists [22]. Behavioural modifications that reduce cortisol levels reversed the associated neuroendocrine cascade and restored ovulation in women with FHA [23, 24].

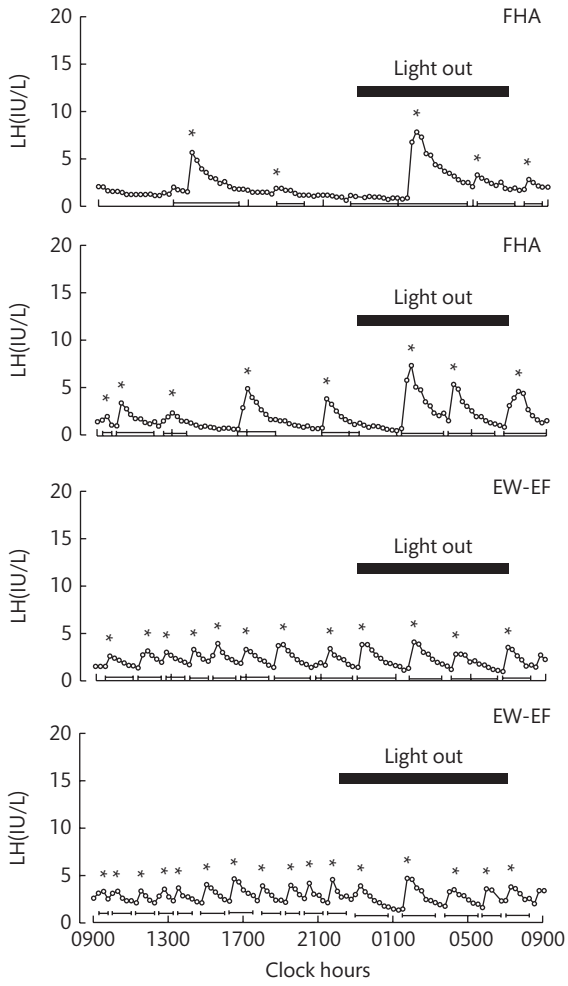


Figure 8.3.1.5 Luteinizing hormone pulse patterns in two women with FHA are shown in the top two panels and in two eumenorrheic, ovulatory women who were studied in the early follicular (EF) phase in the bottom two panels. Luteinizing hormone pulse patterns are the best *in vivo* surrogate for assessing gonadotrophin-releasing hormone activity in humans. To assess luteinizing hormone pulsatility in conditions in which slower frequencies are anticipated, one generally obtains blood samples from a forearm vein at 15-min intervals for 24 h. Computer-driven algorithms are then used to resolve the string of luteinizing hormone determinations into pulses so that frequency and amplitude can be computed. The eumenorrheic women regularly display regular luteinizing hormone pulses of even amplitude whereas the women with FHA display irregular luteinizing hormone pulses with an overall slower frequency and variable amplitude. In general, women with FHA have luteinizing hormone pulse frequency that is roughly half of that observed in eumenorrheic women. The expected nocturnal slowing is seen in the eumenorrheic women whereas some augmentation of luteinizing hormone is seen in women with FHA, a pattern that is somewhat similar to that seen in puberty.

As shown in Figure 8.6.1.7, kisspeptin is the G protein–coupled receptor ligand for its receptor, GPR54. Kisspeptin–GPR54 signalling plays a critical role in the initiation of GnRH secretion during puberty. Kisspeptin/neurokinin B/dynorphin (KNDy) neurons within the arcuate nucleus secrete kisspeptin, which stimulates GnRH neurons [25]. KNDy neurons integrate other neuromodulatory signalling systems linked to reduced GnRH pulsatility [26]. Stressors, regardless of type, activate the HPA axis and elicit a constellation of neuroendocrine alterations, including hypothalamic hypothyroidism that conserves and diverts energetic expenditure [8, 10]. Metabolic signals such as leptin regulate KNDy neurons via a GABAergic mechanism [12]. In monkey models, administration of a CRH antagonist, astressin reversed the impact of social stress upon central GABA-a binding in prefrontal cortex and altered responsivity to ghrelin [27–29].

Mechanisms implicated in the regulation of GnRH pulsatility include glial interposition into the synaptic cleft to reduce appositions and effective GnRH cohort size [30] and multiple neuromodulatory factors including corticotropin-releasing hormone, endogenous opioids, neuropeptide Y, leptin, γ-aminobutyric acid (GABA) and ghrelin [12, 29, 31]. In monkeys, challenge paradigms have included restraint stress, exercise, nutritional restriction, injection of insulin, endotoxin [32], and social instability; stressors are synergistic [20]. However, monkey studies do not reveal why humans initiate deleterious coping strategies or persistently engage in stressful behaviours. The response to some stressors, such as endotoxin, is blunted by sex steroids [32], while the response to restraint is heightened if the ovaries are present [15]. Further, some individuals are more sensitive to the same stressor than others [20, 21, 33, 34]. The neuroendocrine and endocrine responses to exercise were exaggerated in men whose cortisol levels did not suppress when given dexamethasone when compared with the responses in men whose cortisol levels did suppress [35]. In women with FHA, a 20-minute exercise challenge not only provoked a drop in glucose that was not seen in eumenorrheic women but also elicited a proportionally larger rise in cortisol [36]. In contrast, in lactational amenorrhoea, stress responses to challenge are blunted [37]. Reactivity to challenge may be mediated in part by the serotonergic system [33], particularly variants in the serotonin transporter [11, 34]. Overall, the impact of stressors is species-specific, context-dependent, and idiosyncratic.

Although the hypothalamus synthesizes and secretes specific releasing hormones that govern the synthesis and secretion of specific pituitary hormones, it nonetheless functions as an integrated unit to coordinate the endocrine responses to internal and external signals. Because of these complex interdependent relationships, it is simplistic to view FHA as an isolated disturbance in GnRH drive (Figure 8.3.1.1).

Recognition of Functional Hypothalamic Anovulation

The diagnosis of functional hypothalamic anovulation entails excluding all other causes of amenorrhoea and anovulation. Reduced GnRH drive manifests as FSH>LH with low oestradiol (<60 pg/ml) and progesterone concentrations (<1 ng/ml). A thorough search for organic causes must be conducted. While focal neurological signs would be certain to raise suspicion of a central organic cause, conditions such as adult-onset hydrocephalus [38] may be accompanied by dizziness or vague symptoms, rather than lateralizing symptoms. In most instances, if suspected, organic causes can be confirmed by magnetic resonance imaging (MRI). Conditions such as coeliac disease (gluten enteropathy) and irritable bowel syndrome that result in nutritional insufficiency should be excluded as well.

The epidemiology of FHA is not well described. In addition to individual factors, there appears to be a dose–response relationship

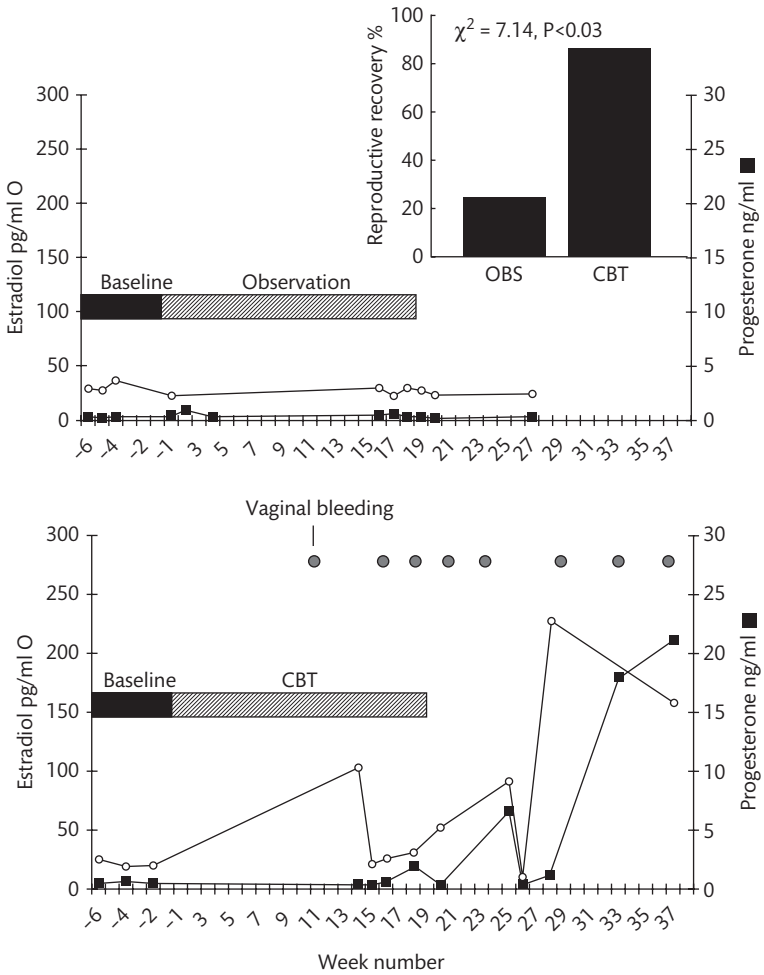


Figure 8.3.1.6 Serum oestradiol (o) and progesterone (■) levels and vaginal bleeding (•) in a woman with functional hypothalamic amenorrhoea (FHA) who was observed and did not recover (top panel) compared with a woman with FHA who was treated with cognitive behaviour therapy (CBT) and recovered (lower panel). The inset depicts the proportion of women with FHA who had evidence of ovarian recovery following randomization to observation (OBS, n = 8; o) or CBT (n = 8; ■).

Adapted with permission from Berga SL, Marcus MD, Loucks TL, Hlastala S, Ringham R, Krohn MA. Recovery of ovarian activity in women with functional hypothalamic amenorrhea who were treated with cognitive behavior therapy. *Fertil Steril*, 2003; 80: 976–81. Copyright © 2003 American Society for Reproductive Medicine.

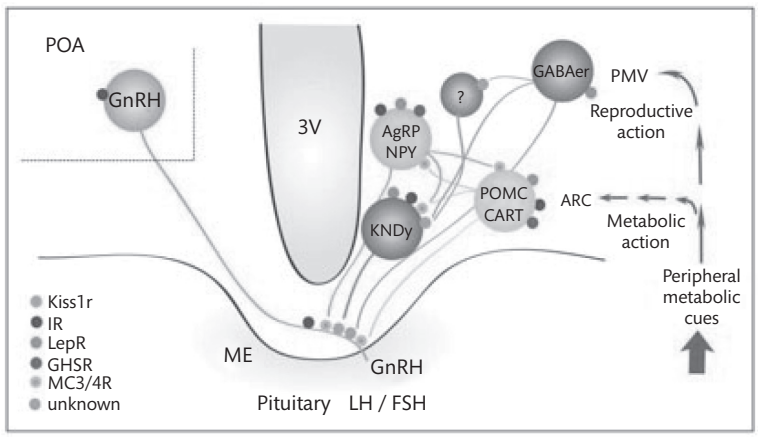


Figure 8.3.1.7 Schematic representation of neural interactions between metabolic and reproductive functions depicting likely sites of action of leptin, insulin, and ghrelin to control the release of gonadotropin-releasing hormone. Abbreviations: 3V, third ventricle; AgRP, agouti-related peptide; ARC, arcuate nucleus; CART, cocaine- and amphetamine-regulated transcript; GABA, gamma-aminobutyric acid; GHSR, growth hormone secretagogue receptor; IR, insulin receptor; Kiss1r, kisspeptin receptor; KNDy, kisspeptin/neurokinin B/dynorphin; LepR, leptin receptor; MC3r, melanocortin receptor 3; MC4r, melanocortin receptor 4; ME, median eminence; NPY, neuropeptide Y; PMV, ventral premammillary nucleus; POA, preoptic area; POMC, pro-opiomelanocortin.

Reproduced with permission from Navarro VM1, Kaiser UB. Metabolic influences on neuroendocrine regulation of reproduction. *Curr Opin Endocrinol Diabetes Obes*. 2013 Aug;20(4):335–41.

between the type and severity of stress and the proportion of individuals who develop amenorrhoea [39]. To what extent more subtle forms of anovulation and luteal insufficiency are due to intermittent or mild reductions in GnRH drive is difficult to ascertain in general clinical practice because the manifestations may be clinically subtle or occult [40]. Oligoasthenospermia may sometimes be the male analogue of functional hypothalamic anovulation. The behavioural causes of FHH may differ in men and women, but this is an area that has received minimal investigative attention [41]. A combined metabolic and psychogenic stressor, a war drill, acutely and reversibly raised cortisol levels while suppressing testosterone and gonadotropin levels in men [42]. Short-term fasting studies suggest that the GnRH pulse generator of men may be more sensitive to metabolic challenge than that of women [43]. Available evidence suggests that extreme forms of hypothalamic hypogonadism, such as anorexia nervosa or bulimia affect 3% of women [44], that FHA unrelated to an eating disorder affects roughly 5% of college-aged women [39], that recreational running is associated with a roughly 55% prevalence of cycles that are anovulatory or display luteal insufficiency [40], and that men are susceptible to functional reductions in GnRH drive without obvious clinical manifestations [41, 42].

Pathogenesis of Functional Hypothalamic Anovulation

The final common aetiological pathway to functional hypothalamic anovulation is reduced GnRH drive that manifests in the circulation as a reduction in LH pulse frequency [8]. Several studies have documented that the LH pulse frequency of amenorrhoeic women with FHA is about half that of women in the follicular phase [8]. (**Figure 8.3.1.5**). Pulsatile administration of GnRH resulted in ovulation and conception in women with FHA [5–7]. For folliculogenesis to result in ovulation, however, GnRH drive must remain above and below critical thresholds for approximately 12–14 days. Presumably, intermittent or lesser reductions in GnRH pulse frequency lead to poor follicular development with subsequent luteal insufficiency.

FHA is more than an isolated interruption of the hypothalamic–pituitary–ovarian axis. A constellation of neuroendocrine secretory aberrations has been demonstrated [8]. While cortisol levels are higher in FHA than in eumenorrheic women, the diurnal pattern remains intact [8, 10]. In both eumenorrheic and amenorrhoeic women, cortisol secretion peaks nocturnally. Among women with anovulation, increased cortisol secretion appears to be specific to FHA [10]. Ovarian function is not necessary for GnRH drive to be suppressed by stress. FSH levels in postmenopausal women admitted to the intensive care unit continue to decline after the peak in cortisol levels demonstrating a lag between recovery of the hypothalamic–pituitary–adrenal and resumption of GnRH drive [45]. Elevated cortisol levels have been observed in women with depression and anorexia nervosa but not bulimia [46]. In women athletes, only those with amenorrhoea displayed elevated cortisol levels and reduced luteinizing hormone pulse frequency [47]. Interestingly, women with FHA did not meet criteria for depression [17–19]. The development of anovulation sufficient to cause amenorrhoea may require greater or more sustained HPA axis activation than generally seen in depression and bulimia. Since energy deficits enhance HPA activation, the extent of energy insufficiency may explain differential rates of FHA in women with depression and bulimia [20].

Cortisol *per se* is not likely to be the cause of reduced GnRH pulse frequency in FHA. Acute, exogenous, high-dose administration of cortisol (over 300% of physiological) did not reduce GnRH drive [48]. Hydrocortisone infusions given to mimic moderate (136% increase) and severe (197% increase) stress had no acute impact on gonadotropin secretion [49]. In FHA, the increase in cortisol is in the range of 20–30%. In underweight anorectics, the increase was in the range of 150%, but after short-term weight recovery, the increase was about 70% of the control group [46]. Typically, weight restoration alone is insufficient to restore menses, despite reduced cortisol levels. Since the increased cortisol secretion observed in women with FHA and anorexia nervosa is lower than the levels needed to independently suppress gonadotropin secretion, other factors must play a causal role. Of note, the decline in LH caused by insulin-induced hypoglycaemia in rhesus monkeys was comparable in monkeys pretreated with metyrapone to block cortisol synthesis and those who were not treated [50]. Similarly, an intravenous infusion of corticotropin-releasing hormone caused a decrease in gonadotropin in monkeys with and without adrenalectomy [51]. While these observations do not obviate a role for cortisol as an indicator of stress, they do suggest that neural concomitants of the stress signal or concomitant metabolic signals, rather than cortisol *per se*, cause stress-induced suppression of GnRH pulsatility. While an increase in cerebrospinal corticotropin-releasing hormone has been detected in women with anorexia nervosa [46], women with FHA do not have a detectable increase in cerebrospinal corticotropin-releasing hormone [52] despite an increase in cerebrospinal fluid (CSF) cortisol concentrations [13]. The lack of suppression of CSF CRH in the presence of increased CSF cortisol indicated resistance to feedback inhibition of hypothalamic CRH release by cortisol and suggests that mechanisms other than CRH and cortisol mediate the stress-induced suppression of GnRH [13].

Cortisol is a hormone with profound metabolic actions and increased secretion produces metabolic mobilization, including inhibition of gluconeogenesis. Cortisol also dampens thyroid-stimulating hormone (TSH) release and causes 'sick euthyroid syndrome', presumably as a means of conserving or diverting energy expenditure. As noted earlier, studies have suggested that increased HPA drive predisposes to further endocrine reactivity in response to the subsequent metabolic challenge of exercise [36]. The converse may also hold. Prior metabolic alterations may predispose to neuroendocrine reactivity in the face of psychogenic challenge [20, 21]. We found that women with FHA showed reduced levels of thyronine and thyroxine despite similar TSH patterns [8]. During recovery from FHA, TSH increased dramatically following a reduction in cortisol levels were reduced [10, 24]. These data indicate that the recovery begins with a decrease in CRH followed by recovery of GnRH and TRH drive.

In a monkey model, metabolic challenge sensitized the hypothalamic–pituitary–ovarian (HPO) axis to the effects of subsequent social stress [20]. Monkeys with the greatest sensitivity to stress had differential serotonergic tone [33] and higher baseline heart rates. Administration of a CRH antagonist reversed the inhibition of LH pulsatility induced by the administration of a potent orexigenic signal, ghrelin [29]. In women, caloric deprivation sufficient to slow LH pulsatility in eumenorrhoeic women dropped circulating glucose

and increased circulating cortisol levels, indicating that metabolic stress activates the HPA axis [16]. Women with FHA responded to a graded exercise challenge with a small drop in circulating glucose while eumenorrheic women showed a rise in glucose during exercise [36]. Simultaneously, exercise provoked a larger rise in cortisol in amenorrhoeic as compared to eumenorrhoeic women [36]. These data indicate that the impact of exercise is state dependent. While exercise is often performed to reduce psychological stress, at an endocrine level, it may amplify the stress cascade and heighten stress reactivity. Thus, the management of stress needs to be contextual. Psychogenic solutions need to be found for psychosocial dilemmas and exercise is not a substitute for good mental hygiene.

Role of Behavioural Variables

Behaviours linked to the development of FHH include exercise [40, 47], weight loss [53, 54], affective [17–19, 54], and eating disorders [44], various personality characteristics [17–19, 55], drug use [54], and a number of external and intrapsychic stresses [17–19, 39, 54, 55]. Given individual variation in metabolism, autonomic tone, habitus, aptitudes, and psychological valences, what is stressful to one may be more or less so to another. It is likely that any given stressor, when the 'dose' is large enough, can activate the central neural pathways leading to reduced GnRH drive [39]. Typically, women with FHA report multiple, seemingly minor, stressors, such as a combination of job or school pressures, poor eating habits, and relatively increased energy expenditure through activity or exercise rather than a solitary stressor [17–19]. Others report overeating or 'emotional eating' and weight gain as indicators of stress but the correlation between overeating and FHH has not been well established. Our studies indicated that women with FHA differed from eumenorrheic women and women with other causes of anovulation and amenorrhea psychologically. In particular, women with FHA displayed unrealistic expectations and attitudes such as perfectionism and high need for social approval that are likely to impair coping responses [17–19]. Although women with FHA do not typically meet criteria for an eating disorder, they display attitudes and behaviours similar to women with eating disorders. One is a drive for thinness and disordered eating. The bottom line is that attitudes and expectations engender a panoply of behaviours, such as aberrant food intake and excessive exercise, that independently and synergistically activate the HPA axis and elicit a constellation of neuroendocrine adaptations including reduced GnRH drive.

To test the hypothesis that changing attitudes would reduce stress and restore ovarian function, we randomized women with FHA to cognitive behaviour therapy (CBT) or observation for 20 weeks. CBT consisted of 16 sessions that focused on attitudes rather than behaviours and provided general guidance but not explicit requirements about nutrition and exercise. The goal of CBT was to improve problem-solving strategies and coping mechanisms, to foster realistic expectations of self and others, and to help the individual develop strategies for dealing with specific and common stressors [23]. Of those randomized initially to CBT, 87% recovered ovarian function while only 25% of those randomized to observation recovered ovarian function (**Figure 8.3.1.6**). Recovery of ovarian function was accompanied by decreased nocturnal cortisol [24], increased leptin and TSH independent of weight gain, and an improvement in attitudes.

Taken together, available data indicate that behaviours or expectations that activate to the HPA axis suppress ovarian and thyroidal function. Generally, a mix of multiple, seemingly minor, psychogenic, and metabolic stressors appears to be more deleterious to reproductive function than a solitary stressor.

Treatment Considerations

Recovery from FHA is expedited by identifying attitudinal antecedents that fuel stress sensitivity. Society might admire the achievements made possible by unbridled ambition, but the endocrine system rewards moderation. Conflicted aspirations may seem mundane to the clinician, but to the individual facing such decisions, having to choose among goals may be perceived as life-defining and potentially life-threatening. Certain behaviours, such as exercise and dieting, may provide a sense of autonomy and control, but ironically, these behaviours also serve as metabolic challenges that sensitize the endocrine system to ongoing or subsequent psychogenic stress and vice versa. Thus, addressing behaviours and attitudes that fuel FHA is likely to foster neuroendocrine recovery and resumption of ovulation while institution of dieting and exercise regimens are likely to sustain FHA and hypercortisolism.

Long-term health complications of FHH/FHA include osteopenia/osteoporosis and infertility. A key goal is to restore menstrual cycles. Kisspeptin has been investigated as a possible modality [56]. Likewise, subcutaneous daily recombinant leptin has been investigated as a treatment option for FHA but its use only restored ovulatory function in less than half of the 8 women treated [57]. In contrast, CBT not only fostered recovery of ovarian function, but also reduced hypercortisolaemia and increased in leptin and TSH [24].

Ovulation induction is often undertaken for women with FHA who desire fertility. Unfortunately, use of clomiphene, a selective oestrogen receptor modulator, or letrozole, an aromatase inhibitor, may not increase FSH sufficiently to foster folliculogenesis to the point of ovulation because of increased sensitivity to oestrogen negative feedback [11]. Indeed, the absence of increased FSH levels in women with FHA despite low oestradiol levels indicates increased sensitivity to oestrogen negative feedback. Women with FHA rarely report hot flashes even in the face of profound hypoestrogenism; the presence of hot flashes in a woman with presumed FHH should engender a thorough search for an organic aetiology. Other options for ovulation induction include exogenous pulsatile administration of GnRH and exogenous administration of gonadotropins [5]. Should ovulation induction be undertaken? Women with FHA who are underweight during pregnancy have an increased risk of intrauterine growth restriction and preterm delivery [58]. Children born to mothers with subclinical hypothyroidism (30% lower maternal thyroxine) had a mean full-scale intelligence quotient 7 points lower than the control population [59]. Women with FHA display varying decrements in thyroxine compared to control populations [8, 44], but our data indicated a 25% decrement in women with FHA in the absence of low weight [8]. During the first trimester, the mother is the sole source of thyroxine for the fetus, and in the second trimester, she is the predominant source. Ovulation induction in the presence of metabolic mobilization, such as that which accompanies FHA, may carry as yet unspecified risks for the child's neuropsychological development [60]. Women with FHA may also display compromised

parenting skills. A woman who is stressed to the point of developing anovulation is unlikely to cope well with the additional stress of a newborn. Thus, ovulation induction in a woman with FHA carries risks such as intrauterine growth restriction, preterm delivery, and poor neuropsychological development in the offspring. The preferred course of action would be to foster recovery through behavioural and cognitive alterations that spur neuroendocrine recovery including resumption of ovulatory function before resorting to ovulation induction or assisted reproduction. Recognizing that stress has profound effects on reproductive physiology provides a lens for understanding the social determinants of health [61].

A common rationale for instituting hormone replacement or contraceptive therapies is prevention of osteoporosis. While hormonal contraception or exogenous hormone replacement is unlikely to cause harm, it may also provide few benefits [9]. Hormone replacement strategies are unlikely to promote bone accretion in the face of metabolic imbalance including hypercortisolism. In particular, excess cortisol reduces osteoblast activity and impairs bone formation. Sex steroid therapy will not ameliorate ongoing metabolic compromise. Indeed, women with FHH/FHA may be catabolic rather than anabolic. When eumetabolic, oestrogens inhibit bone resorption but androgens are needed to stimulate bone formation. Both androgens and oestrogens are low in FHA. Further, cortisol interferes with oestrogen action [62]. Women with anorexia nervosa have lower bone mass than women with FHA, but both have lower bone mass than age-matched eumenorrhoeic women [63]. Women with anorexia nervosa do not build bone in response to exogenous sex steroid administration because of concurrent nutritional deficits and ongoing metabolic compromise including hypercortisolaemia [63]. While sex steroids and bisphosphonates primarily retard bone resorption, women with anorexia nervosa, and FHA already have reduced bone formation, so it is not clear that the use of either will significantly improve bone accretion [9]. Further, bisphosphonates get incorporated into the maternal skeleton and are mobilized during pregnancy and thus carry the additional hazard of being potentially teratogenic. Resumption of menses is associated with bone accretion, but bone density remains below expected values, indicating that it may not be possible to fully recover 'lost bone'. Delayed menarche in young ballet dancers was associated with higher rates of scoliosis, as well as fracture [64]. In one small study of amenorrhoeic athletes, oral contraceptive use led to a small gain in bone over placebo, while cyclic progestin exposure led to greater bone loss than placebo [65].

The characteristic hypothalamic alterations induced by challenge only become problematic when challenge elicits a chronic, rather than acute, response. The long-term consequences of persistent HPA activation have been studied in animal models and hippocampal neuronal loss has been documented. The multiple health consequences of chronic stress in humans are impressive [66]. Persistent stress in elderly women (manifested as an elevated urinary free cortisol) was associated with cognitive decline across time, while stress reduction led to cognitive improvement [67]. Both perceived stress and chronicity of stress were significantly associated with advanced cellular ageing as evidenced by lower telomerase activity and shorter telomere length in healthy premenopausal women [68]. Hormone therapy *per se* is unlikely to be harmful, but more than hormone administration is needed. The stress process needs to be dampened.

Although psychopharmacological approaches have not been studied in FHA, they should be considered. Benzodiazepines are contraindicated during pregnancy, so their use would carry the greatest risk were conception to occur during pharmacological intervention. Antidepressants may reduce HPA activation, but it is not clear that their use will address associated metabolic imbalances that would be problematic were conception to occur. Naltrexone has been employed with variable success [69]. The optimal treatment is to reduce stress and restore metabolic balance so that the hypothalamus recovers and gonadal function resumes. An integral goal of treatment is to help women identify sources of psychogenic and metabolic stress, and to provide emotional support, while coping mechanisms other than dieting or exercise are learned. Non-pharmacological therapies, such as stress management, relaxation training, and psychoeducation empower individuals by fostering self-care and competency. In this regard, non-pharmacological therapies have the potential to produce long-term mental and physical health benefits extending beyond resumption of ovulatory function.

FHA is theoretically reversible, but few studies have documented the course of recovery or the likelihood. Hirvonen found that 72% recovered in 6 years [70]. One report showed that 9 of 16 women with FHA who underwent ovulation induction and conceived became eumenorrhoeic postpartum [71]. Roughly half of women with anorexia nervosa or bulimia have a full recovery [44]. To avoid disappointment and unrealistic expectations, patients should be counselled that reproductive recovery may not immediately ensue following lifestyle alterations.

Summary

The development of FHH/FHA demonstrates that attitudes, moods, and behaviours alter neuroendocrine function. Although a link between brain states and gonadal function has long been hypothesized, only recently have we been able to specify some of the mechanisms mediating this relationship. In the psychoneuroendocrinologic context, health depends on achieving psychological harmony through realistic expectations of self and others and metabolic harmony through balanced diet and energy expenditure. Although reduction in or avoidance of some stresses may be possible, the complete elimination of frustrations and demands is in itself unrealistic. Thus, the development of awareness and the institution of appropriate coping patterns are instrumental to maintaining health, including reproductive health. The burden of reducing and managing stress should not fall exclusively to the individual, but should also be a societal goal. While the medical profession participates in defining and understanding the role of stress in disease, it is by no means the only profession or social institution with an obligation to ameliorate this health burden.

ACKNOWLEDGEMENTS

Funding was provided by RO1MH-50748 and RR-00056.

REFERENCES

1. Weiss G, Skurnick JH, Goldsmith LT, Santoro NF, Park SJ. Menopause and hypothalamic-pituitary sensitivity to estrogen. *JAMA*, 2004; **292**: 2991–6.
2. Berga SL, Guzick DS, Winters SJ. Increased luteinizing hormone and alpha-subunit secretion in women with hyperandrogenic anovulation. *J Clin Endocrinol Metab*, 1993; **77**: 895–901.
3. Caronia LM, Martin C, Welt CK, *et al*. A genetic basis for functional hypothalamic amenorrhea. *N Engl J Med*, 2011; **364**: 215–25.
4. Reifenstein EC Jr. Psychogenic or hypothalamic amenorrhea. *Med Clin North Am*, 1946; **30**: 1103–14.
5. Martin KA, Hall JE, Adams JM, Crowley WF Jr. Comparison of exogenous gonadotropins and pulsatile gonadotropin-releasing hormone for induction of ovulation in hypogonadotropic amenorrhea. *J Clin Endocrinol Metab*, 1993; **77**: 125–9.
6. Miller DS, Reid RR, Cetel NS, Rebar RW, Yen SS. Pulsatile administration of low-dose gonadotropin-releasing hormone. Ovulation and pregnancy in women with hypothalamic amenorrhea. *JAMA*, 1983; **250**: 2937–41.
7. Hurley DM, Brian RJ, Burger HG. Ovulation induction with subcutaneous pulsatile gonadotropin-releasing hormone: singleton pregnancies in patients with previous multiple pregnancies after gonadotropin therapy. *Fertil Steril*, 1983; **40**: 575–9.
8. Berga SL, Mortola JF, Girton L, Suh B, Laughlin G, Pham P, Yen SS. Neuroendocrine aberrations in women with functional hypothalamic amenorrhea. *J Clin Endocrinol Metab*, 1989; **68**: 301–8.
9. Gordon CM, Ackerman KE, Berga SL, *et al*. Functional hypothalamic amenorrhea: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 1413–39.
10. Berga SL, Daniels TL, Giles DE. Women with functional hypothalamic amenorrhea but not other forms of anovulation display amplified cortisol concentrations. *Fertil Steril*, 1997; **67**: 1024–30.
11. Michopoulos V, Berga SL, Kaplan JR, Wilson ME. Social subordination and polymorphisms in the gene encoding the serotonin transporter enhance estradiol inhibition of luteinizing hormone secretion in female rhesus monkeys. *Biol Reprod*, 2009; **81**: 1154–63.
12. Navarro VM, Kaiser UB. Metabolic influences on neuroendocrine regulation of reproduction. *Curr Opin Endocrinol Diabetes Obes*, 2013; **20**: 335–41.
13. Brundu B, Loucks TL, Adler LJ, Cameron JL, Berga SL. Increased cortisol in the cerebrospinal fluid of women with functional hypothalamic amenorrhea. *J Clin Endocrinol Metab*, 2006; **91**: 1561–5.
14. Ackerman KE, Patel KT, Guereca G, Pierce L, Herzog DB, Misra M. Cortisol secretory parameters in young exercisers in relation to LH secretion and bone parameters. *Clin Endocrinol (Oxf)*, 2013; **78**: 114–19.
15. Villanueva AL, Schlosser C, Hopper B, Liu JH, Hoffman DI, Rebar RW. Increased cortisol production in women runners. *J Clin Endocrinol Metab*, 1986; **63**: 133–6.
16. Loucks AB, Thuma JR. Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. *J Clin Endocrinol Metab*, 2003; **88**: 297–311.
17. Giles DE, Berga SL. Cognitive and psychiatric correlates of functional hypothalamic amenorrhea: a controlled comparison. *Fertil Steril*, 1993; **60**: 486–92.
18. Marcus MD, Loucks TL, Berga SL. Psychological correlates of functional hypothalamic amenorrhea. *Fertil Steril*, 2001; **76**: 310–16.
19. Berga SL, Girton LG. The psychoneuroendocrinology of functional hypothalamic amenorrhea. *Psychiatr Clin North Am*, 1989; **12**: 105–16.
20. Williams NI, Berga SL, Cameron JL. Synergism between psychosocial and metabolic stressors: impact on reproductive function in cynomolgus monkeys. *Am J Physiol Endocrinol Metab*, 2007; **293**: E270–6.
21. Berga SL. Stress and reproduction: a tale of false dichotomy? *Endocrinology*, 2008; **149**: 867–8.
22. Gore AC, Chappell VA, Fenton SE, *et al*. EDC-2: The Endocrine Society's second scientific statement on endocrine-disrupting chemicals. *Endocr Rev*, 2015; **36**: E1–E150.
23. Berga SL, Marcus MD, Loucks TL, Hlastala S, Ringham R, Krohn MA. Recovery of ovarian activity in women with functional hypothalamic amenorrhea who were treated with cognitive behavior therapy. *Fertil Steril*, 2003; **80**: 976–81.
24. Michopoulos V, Mancini F, Loucks TL, Berga SL. Neuroendocrine recovery initiated by cognitive behavioral therapy in women with functional hypothalamic amenorrhea: a randomized, controlled trial. *Fertil Steril*, 2013; **99**: 2084–91.
25. Gottsch ML, Cunningham MJ, Smith JT, Popa SM, Acohido BV, Crowley WF, Seminara S, Clifton DK, Steiner RA. A role for kisspeptins in the regulation of gonadotropin secretion in the mouse. *Endocrinology*, 2004; **145**: 4073–7.
26. McCarthy MM. A piece in the puzzle of puberty. *Nat Neurosci*, 2013; **16**: 251–3.
27. Michopoulos V, Embree M, Reding K, *et al*. CRH receptor antagonism reverses the effect of social subordination upon central GABAA receptor binding in estradiol treated ovariectomized female rhesus monkeys. *Neuroscience*, 2013; **250**: 300–8.
28. Michopoulos V, Loucks T, Berga SL, Rivier J, Wilson ME. Increased ghrelin sensitivity and calorie consumption in subordinate monkeys is affected by short-term astressin B administration. *Endocrine*, 2010; **38**: 227–34.
29. Vulliemoz NR, Xiao E, Xia-Zhang L, Rivier J, Ferin M. Astressin B, a nonselective corticotropin-releasing hormone receptor antagonist, prevents the inhibitory effect of ghrelin on luteinizing hormone pulse frequency in the ovariectomized rhesus monkey. *Endocrinology*, 2008; **149**: 869–74.
30. Witkin JW, O'Sullivan H, Miller R, Ferin M. GnRH perikarya in medial basal hypothalamus of pubertal female rhesus macaque are ensheathed with glia. *J Neuroendocrinol*, 1997; **9**: 881–5.
31. Williams CL, Nishihara M, Thalabard JC, Grosser PM, Hotchkiss J, Knobil E. Corticotropin-releasing factor and gonadotropin-releasing hormone pulse generator activity in the rhesus monkey. Electrophysiological studies. *Neuroendocrinology*, 1990; **52**: 133–7.
32. Ferin M. Clinical review 105: stress and the reproductive cycle. *J Clin Endocrinol Metab*, 1999; **84**: 1768–74.
33. Bethea CL, Pau FK, Fox S, Hess DL, Berga SL, Cameron JL. Sensitivity to stress-induced reproductive dysfunction linked to activity of the serotonin system. *Fertil Steril*, 2005; **83**: 148–55.
34. Jarrell H, Hoffman JB, Kaplan JR, Berga S, Kinkead B, Wilson ME. Polymorphisms in the serotonin reuptake transporter gene modify the consequences of social status on metabolic health in female rhesus monkeys. *Physiol Behav*, 2008; **93**: 807–19.

35. Petrides JS, Mueller GP, Kalogeras KT, Chrousos GP, Gold PW, Deuster PA. Exercise-induced activation of the hypothalamic-pituitary-adrenal axis: marked differences in the sensitivity to glucocorticoid suppression. *J Clin Endocrinol Metab*, 1994; **79**: 377–83.
36. Sanders KM, Kawwass JF, Loucks T, Berga SL. Heightened cortisol response to exercise challenge in women with functional hypothalamic amenorrhea. *Am J Obstet Gynecol*, 2018; **218**: 230.e1–230.e6.
37. Altemus M, Deuster PA, Galliven E, Carter CS, Gold PW. Suppression of hypothalamic–pituitary–adrenal axis responses to stress in lactating women. *J Clin Endocrinol Metab*, 1995; **80**: 2954–9.
38. Lowry DW, Lowry DL, Berga SL, Adelson PD, Roberts MM. Secondary amenorrhea due to hydrocephalus treated with endoscopic ventriculocisternostomy. Case report. *J Neurosurg*, 1996; **85**: 1148–52.
39. Drew FL. The epidemiology of secondary amenorrhea. *J Chronic Dis*, 1961; **14**: 396–407.
40. De Souza MJ, Miller BE, Loucks AB, *et al*. High frequency of luteal phase deficiency and anovulation in recreational women runners: blunted elevation in follicle-stimulating hormone observed during luteal-follicular transition. *J Clin Endocrinol Metab*, 1998; **83**: 4220–32.
41. Nachtigall LB, Boepple PA, Pralong FP, Crowley WF, Jr. Adult-onset idiopathic hypogonadotropic hypogonadism-a treatable form of male infertility. *N Engl J Med*, 1997; **336**: 410–15.
42. Opstad PK. Androgenic hormones during prolonged physical stress, sleep, and energy deficiency. *J Clin Endocrinol Metab*, 1992; **74**: 1176–83.
43. Berga SL, Loucks TL, Cameron JL. Endocrine and chronobiological effects of fasting in women. *Fertil Steril*, 2001; **75**: 926–32.
44. Becker AE, Grinspoon SK, Klibanski A, Herzog DB. Eating disorders. *N Engl J Med*, 1999; **340**: 1092–8.
45. Spratt DI, Longcope C, Cox PM, Bigos ST, Wilbur-Welling C. Differential changes in serum concentrations of androgens and estrogens (in relation with cortisol) in postmenopausal women with acute illness. *J Clin Endocrinol Metab*, 1993; **76**: 1542–7.
46. Kaye WH, Gwirtsman HE, George DT, *et al*. Elevated cerebrospinal fluid levels of immunoreactive corticotropin-releasing hormone in anorexia nervosa: relation to state of nutrition, adrenal function, and intensity of depression. *J Clin Endocrinol Metab*, 1987; **64**: 203–8.
47. Loucks AB, Mortola JF, Girton L, Yen SS. Alterations in the hypothalamic–pituitary-ovarian and the hypothalamic-pituitary-adrenal axes in athletic women. *J Clin Endocrinol Metab*, 1989; **68**: 402–11.
48. Saketos M, Sharma N, Santoro NF. Suppression of the hypothalamic-pituitary-ovarian axis in normal women by glucocorticoids. *Biol Reprod*, 1993; **49**: 1270–6.
49. Samuels MH, Luther M, Henry P, Ridgway EC. Effects of hydrocortisone on pulsatile pituitary glycoprotein secretion. *J Clin Endocrinol Metab*, 1994; **78**: 211–15.
50. Van Vugt DA, Piercy J, Farley AE, Reid RL, Rivest S. Luteinizing hormone secretion and corticotropin-releasing factor gene expression in the paraventricular nucleus of rhesus monkeys following cortisol synthesis inhibition. *Endocrinology*, 1997; **138**: 2249–58.
51. Xiao E, Luckhaus J, Niemann W, Ferin M. Acute inhibition of gonadotropin secretion by corticotropin-releasing hormone in the primate: are the adrenal glands involved? *Endocrinology*, 1989; **124**: 1632–7.
52. Berga SL, Loucks-Daniels TL, Adler LJ, *et al*. Cerebrospinal fluid levels of corticotropin-releasing hormone in women with functional hypothalamic amenorrhea. *Am J Obstet Gynecol*, 2000; **182**: 776–81.
53. Vigersky RA, Andersen AE, Thompson RH, Loriaux DL. Hypothalamic dysfunction in secondary amenorrhea associated with simple weight loss. *N Engl J Med*, 1977; **297**: 1141–5.
54. Fries H, Nillius SJ, Pettersson F. Epidemiology of secondary amenorrhea. II. A retrospective evaluation of etiology with special regard to psychogenic factors and weight loss. *Am J Obstet Gynecol*, 1974; **118**: 473–9.
55. Shanan J, Brzezinski A, Sulman F, Sharon M. Active coping behavior, anxiety, and cortical steroid excretion in the prediction of transient amenorrhea. *Behav Sci*, 1965; **10**: 461–5.
56. Jayasena CN, Abbara A, Veldhuis JD, *et al*. Increasing LH pulsatility in women with hypothalamic amenorrhoea using intravenous infusion of Kisspeptin-54. *J Clin Endocrinol Metab*, 2014; **99**: E953–61.
57. Welt CK, Chan JL, Bullen J, *et al*. Recombinant human leptin in women with hypothalamic amenorrhea. *NEJM*, 2004; **351**: 987–97.
58. van der Spuy ZM, Steer PJ, McCusker M, Steele SJ, Jacobs HS. Outcome of pregnancy in underweight women after spontaneous and induced ovulation. *Br Med J (Clin Res Ed)*, 1988; **296**: 962–5.
59. Haddow JE, Palomaki GE, Allan WC, *et al*. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*, 1999; **341**: 549–55.
60. Morreale de Escobar G, Obregon MJ, Escobar del Rey F. Is neuropsychological development related to maternal hypothyroidism or to maternal hypothyroxinemia? *J Clin Endocrinol Metab*, 2000; **85**: 3975–87.
61. Marmot M. The health gap: the challenge of an unequal world. *Lancet*, 2015; **386**: 2442–4.
62. Whirledge S, Cidlowski JA. A role for glucocorticoids in stress-impaired reproduction: beyond the hypothalamus and pituitary. *Endocrinology*, 2013; **154**: 4450–68.
63. Miller KK, Klibanski A. Clinical review 106: amenorrheic bone loss. *J Clin Endocrinol Metab*, 1999; **84**: 1775–83.
64. Warren MP, Brooks-Gunn J, Hamilton LH, Warren LF, Hamilton WG. Scoliosis and fractures in young ballet dancers. Relation to delayed menarche and secondary amenorrhea. *N Engl J Med*, 1986; **314**: 1348–53.
65. Hergenroeder AC, Smith EO, Shypailo R, Jones LA, Klish WJ, Ellis K. Bone mineral changes in young women with hypothalamic amenorrhea treated with oral contraceptives, medroxyprogesterone, or placebo over 12 months. *Am J Obstet Gynecol*, 1997; **176**: 1017–25.
66. McEwen BS. Protective and damaging effects of stress mediators. *New Engl J Med*, 1998; **338**: 171–9.
67. Lupien SJ, Gaudreau S, Tchiteya BM, *et al*. Stress-induced declarative memory impairment in healthy elderly subjects: relationship to cortisol reactivity. *J Clin Endocrinol Metab*, 1997; **82**: 2070–5.
68. Epel ES, Blackburn EH, Lin J, *et al*. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci USA*, 2004; **101**: 17312–15.
69. Wildt L, Leyendecker G, Sir-Petermann T, Waibel-Treber S. Treatment with naltrexone in hypothalamic ovarian

failure: induction of ovulation and pregnancy. *Hum Reprod*, 1993; **8**: 350–8.
70. Hirvonen E. Etiology, clinical features and prognosis in secondary amenorrhea. *Int J Fertil*, 1977; **22**: 69–76.
71. Lewinthal D, Corenblum B, Brooks JH, Taylor PJ. Spontaneous return of menstruation in hypothalamic amenorrhea following gonadotropin-releasing hormone-induced pregnancy. *Fertil Steril*, 1987; **47**: 870–1.

8.3.2 Hyperprolactinaemia

Julian Davis and Agnieszka Święcicka

Prolactin Biology

Prolactin is a polypeptide hormone, named from its well-known effects to promote lactation. It is essential for successful reproduction in man and mammals, although it is known to have a wide variety of non-reproductive effects whose clinical significance remains uncertain. Prolactin is secreted by lactotrophic cells of anterior pituitary gland, but many other sites of production have been identified, including the endometrium, immune cells including T-lymphocytes, and skin [1–4]. Circulating prolactin is generally assumed to derive mainly from pituitary secretion, and extrapituitary production is thought to exert mainly local paracrine effects within tissues. The clinical condition of hyperprolactinaemia is almost always attributable to abnormal pituitary prolactin secretion and ectopic neoplastic causes of hyperprolactinaemia are extremely rare.

Lactotrophic cells constitute a large proportion (25–40%) of the endocrine cell types in the pituitary, though their numbers and size can vary during and after pregnancy [5]. This indicates a significant degree of plasticity in pituitary structure that may be important in the susceptibility of the pituitary to the frequent formation of microadenomas [6].

Unlike other pituitary hormones, pituitary prolactin secretion is under predominantly inhibitory control from the hypothalamus, and transection of the pituitary stalk results in sustained hyperprolactinaemia. The prolactin inhibitory factor was identified as dopamine, secreted by tubero-infundibular neurons, which acts on D2 receptors on lactotrophic cells to reduce prolactin synthesis and secretion [4, 7]. A series of other hypothalamic hormones may act as stimuli to prolactin secretion, including TRH, VIP, galanin, and fibroblast growth factor [4]. However the main physiological mechanism for pulses of prolactin secretion is thought to be the episodic interruption of dopamine inhibition.

The reproductive effects of prolactin are found at all levels of the hypothalamo-pituitary-gonadal axis, with effects on the frequency of gonadotropin-releasing hormone (GnRH) pulses, inhibition of the secretion of gonadotrophins, and suppression of luteolysis [8]. These multiple levels of suppression of gonadal function mean that the hyperprolactinaemia that normally occurs during breastfeeding effectively suppresses fertility [9].

A variety of non-reproductive effects of prolactin have been documented which are of less clear significance in man. The prolactin receptor (PRLR) is found in many peripheral tissues, and studies of prolactin and PRLR knockout mice have clearly confirmed not only the reproductive effects of prolactin, but also effects on maternal behaviour, adipose tissue depots, bone mineralization, and immune response [10, 11].

Hyperprolactinaemia

Hyperprolactinaemia, reflecting sustained overproduction from the pituitary, is relatively common in the population [12]. The causes are listed in **Box 8.3.2.1**. The commonest cause is the use of drugs that have dopamine D2 receptor antagonist activity, including antipsychotic agents such as phenothiazines, although some newer agents have little effect [13]. Pregnancy and lactation are the commonest

Box 8.3.2.1 Causes of hyperprolactinaemia

- Pregnancy and lactation
- Acute stress
- Medication
  - Metoclopramide
  - Domperidone
  - Sulpiride
  - Phenothiazines
  - Haloperidol
  - Selective serotonin reuptake inhibitors
  - Methyldopa
- Chest wall and nipple stimulation
  - Herpes zoster
  - Burns
  - Nipple stimulation, piercings
- Pituitary-hypothalamic disease
  - Non-tumoural (idiopathic) hyperprolactinaemia
  - Prolactinoma—microprolactinoma, macroprolactinoma
  - Disconnection hyperprolactinaemia
    - Hypothalamic tumour
    - Granuloma
    - Craniopharyngioma
    - Meningioma
    - Trauma
  - Acromegaly: mixed mammosomatotroph adenoma
- Associated endocrine conditions
  - Hypothyroidism
  - Polycystic ovary syndrome
- Miscellaneous conditions
  - Renal failure
  - Cirrhosis

physiological causes, and short-term acute stress, such as the anxiety provoked by blood sampling, is also a frequent cause of transient rises in serum prolactin that may be misinterpreted in clinical practice and necessitate a second confirmatory blood sample.

Pathological pituitary causes of hyperprolactinaemia may reflect a functioning pituitary prolactinoma, a prolactin-secreting microadenoma or macroadenoma, but in many cases no adenoma is detectable on scanning, in which case the condition is termed idiopathic or non-tumoural hyperprolactinaemia. Anatomical disturbance in the suprasellar region may cause effective functional disconnection of the anterior pituitary gland from the inhibitory effects of dopamine: the resulting 'stalk disconnection' hyperprolactinaemia may lead to diagnostic confusion between functioning prolactinomas and non-functioning pituitary adenomas or other non-pituitary lesions, which usually require different management.

Prolactin hypersecretion can coexist with various conditions. Prolactin is frequently cosecreted with growth hormone by mammosomatotroph cells, and mixed GH-prolactin-secreting tumours can present clinically as either acromegaly or hyperprolactinaemia. Thus the evaluation of hyperprolactinaemic patients should consider the possibility of acromegaly. Hyperprolactinaemia complicates the presentation of polycystic ovarian syndrome in up to 30% of patients [14], though the mechanism and nature of the association remain uncertain. Other miscellaneous recognized causes of hyperprolactinaemia include hypothyroidism, nipple, or chest wall stimulation (including herpes zoster, burns, nipple piercings), chronic renal failure, and cirrhosis.

Presentation

The typical clinical features that suggest hyperprolactinaemia are those of galactorrhoea and oligo-/amenorrhoea. Galactorrhoea is frequent though not universal, and may be unilateral or bilateral, and is explained by the direct effects of prolactin in the breast, to promote milk protein synthesis, and ductal epithelial cell proliferation. Oligomenorrhoea is due to suppression of the hypothalamic–pituitary–ovarian axis, though some women may develop irregular anovulatory menstrual bleeding. Prolonged hypogonadism due to hyperprolactinaemia may lead to osteopaenia [15].

Weight gain has been reported in hyperprolactinaemic women [16], which improves with dopamine agonist treatment, but whose mechanism is still not well understood. Insulin resistance has also been described in hyperprolactinaemic women [17].

Although prolactin receptors are widespread and prolactin has been described to have a series of effects on immune tissues, there appear to be no clinically significant abnormalities in immune function or autoimmune disease in patients with hyperprolactinaemia, although prolactin deficiency in severe hypopituitarism may have adverse effects [18, 19].

Investigation

Prolactin Measurement

Serum prolactin levels are readily measured by most clinical biochemistry laboratories, and most immunoassays generate results that are expressed in terms of a WHO international standard that comprises 23kD monomeric prolactin, and expressed in mU/L or ng/ml [20]. Prolactin levels should be measured on more than one occasion, and persistent unexplained hyperprolactinaemia requires evaluation. It should be remembered that pregnancy is a well-known physiological cause of hyperprolactinaemia, and the possibility should be considered before embarking on further investigation of a woman with amenorrhoea and hyperprolactinaemia.

Serum prolactin levels rise at night during sleep, but this is rarely a problem in routine clinical practice, and blood samples therefore do not need to be timed. The reference range for serum prolactin varies slightly according to age and gender, with slightly higher levels seen in premenopausal women. Dopamine antagonist drugs can markedly raise prolactin levels to several times the upper limit of the normal range, and in patients on prolactin-raising medication a judgement needs to be made as to whether further investigation is necessary or not.

In patients with pituitary adenomas the degree of prolactin elevation is generally roughly proportional to the size of the adenoma, with microprolactinomas (<10 mm tumour diameter) usually associated with prolactin levels below 8000 mU/L, and macroprolactinomas usually giving levels above 10 000mU/L. Patients with large pituitary adenomas and only slight prolactin elevation may have non-functioning pituitary tumours causing stalk disconnection hyperprolactinaemia.

Pitfalls in Prolactin Assays: Hook Effect and Macroprolactinaemia

Some prolactin assays are susceptible to a 'high dose hook effect', whereby extreme elevation of prolactin in the sample saturates both the capture and the detection antibodies in an immunometric sandwich assay resulting in gross under-reading of the sample concentration. If such an effect is suspected in a patient with clinical features of a large macroprolactinoma but only marginal hyperprolactinaemia, the blood sample should be re-analysed after dilution to overcome this hook effect.

The other important issue affecting serum prolactin assays is the frequent occurrence of macroprolactinaemia, in up to 10% of the population. Prolactin normally circulates as monomeric protein, but it is also found in a dimeric form and in high molecular weight complexes (>100kD), formed of prolactin-immunoglobulin complexes ('big-big prolactin', or 'macroprolactin') [21]. Some patient samples contain large amounts of macroprolactin, a condition termed macroprolactinaemia, though the complexes are thought to have limited bioactivity. Macroprolactinaemia is found in a proportion of patients with hyperprolactinaemia, so its presence cannot be taken to exclude genuine pituitary disease. On the other hand, the incidental finding of artefactually raised prolactin in a patient with no features of pituitary disease can lead to fruitless or misleading investigation, and therefore most laboratories now screen for the presence of macroprolactin complexes using polyethylene glycol precipitation [20].

Other Biochemical Investigation

Patients with genuine unexplained and persistent hyperprolactinaemia should be considered for investigation of pituitary function, including assessment of the pituitary gonadal axis (serum oestradiol, LH, and FSH measurement, and assessment of ovulation). Thyroid function should be tested as hyperprolactinaemia

can be a feature of hypothyroidism. The possibility of acromegaly should be considered, with assessment of GH and IGF-1 levels. Polycystic ovary syndrome should also be considered, and biochemical evaluation could include measurement of androgen levels. In patients with pituitary disease the adequacy of pituitary-adrenal function should be tested, by measurement of 9am plasma cortisol or a short tetracosactrin test (see Box 8.3.2.2).

Pituitary Imaging and Other Tests

Persistent hyperprolactinaemia may be due to pituitary or hypothalamic mass lesions, as outlined previously, and these should be evaluated by pituitary imaging (see online Figure 8.3.2.1). The main purpose of imaging is to exclude a large mass lesion that threatens adjacent structures: small pituitary microadenomas are common incidental findings in the normal population, and they may require no action at all. In most centres MR scanning is the investigation of choice. CT scanning is a good alternative but often less satisfactory for imaging of adjacent structures such as the optic chiasm and cavernous sinus. Although some patients find the machines less claustrophobic than MR scanners, CT involves a significant X-ray dose, and repeated scanning of the orbits should be avoided to minimize risk of cataract. In patients where a pituitary or hypothalamic mass lesion is detected, visual fields should be assessed formally by manual or computerized perimetry.

Treatment

Microprolactinomas and Non-Tumoural Hyperprolactinaemia

Patients with hyperprolactinaemia may require treatment for various reasons, including restoration of ovulatory function, maintenance of adequate oestrogenization, suppression of galactorrhoea, or reduction in size of a mass lesion. Some patients may not require treatment at all, for example patients who have no significant mass lesion and no galactorrhoea, and with adequate oestrogen status. The main current treatment option is dopamine agonist therapy, and this will be considered first, followed by the other available options. Treatment is mainly directed at the hormonal disturbance, and patients with small pituitary microprolactinomas are often treated in the same way as patients with non-tumoural hyperprolactinaemia.

Box 8.3.2.2 Investigation of hyperprolactinaemia

- Exclusion of pregnancy, review of medication
- Biochemical testing
  - Serum prolactin (two occasions)
    - Screen for macroprolactinaemia if appropriate
  - Thyroid function testing
  - Pituitary function testing
    - Gonadal function: oestradiol, LH, FSH
    - Screening for acromegaly: GH levels, IGF-1
    - Adrenal function: 9.00 a.m. cortisol, stimulation testing if appropriate
  - Screening for polycystic ovary syndrome if appropriate (serum androgens)
- Visual field testing
- Pituitary imaging: MR or CT scanning

The issues for patients with macroprolactinomas are slightly different, and these will be considered separately.

Dopamine Agonists

Pituitary prolactin synthesis and secretion are suppressed by dopamine, and ergot-derived dopamine D2 receptor agonists have become the standard treatment for hyperprolactinaemia since their introduction in the 1970s [22, 23]. The drugs in current use are all orally active and include bromocriptine, the first drug of this class to be used, and cabergoline; quinagolide is a non-ergot dopamine agonist, with similar effects at the D2 receptor, but slightly different side effects. A series of other ergoline derivatives have been used including pergolide and lisuride, but are not in routine use at present. Bromocriptine and quinagolide are usually given daily, whereas cabergoline is given once or twice weekly.

Prolactin levels fall rapidly, within hours of first administration, and the effect of treatment can be judged biochemically within days or weeks. Prolactin levels typically normalize within 12–24 months of dopamine agonist therapy in up to 95% of patients with microprolactinoma, whereas restoration of normal ovulatory function occurs in approximately 82% of females; recovery of gonadal function is less common in male patients [23, 24]. Resistance to dopamine agonists is uncommon, and treatment is more often constrained by side effects, which are outlined next. Thus women with anovulatory infertility due to hyperprolactinaemia can rapidly and effectively be rendered fertile with simple oral treatment, and ovulation may occur before the resumption of a first period after long spells or amenorrhoea. Patients should therefore be warned to take contraceptive precautions if pregnancy is not sought immediately.

Cabergoline is usually recommended as the drug of choice in hyperprolactinaemia because of its higher efficacy in normalizing prolactin levels, as well as a higher frequency of pituitary tumour shrinkage [25]. The greater efficacy of cabergoline may be explained by the higher affinity for dopamine receptor binding sites as well as potentially superior drug compliance, owing to lower incidence of unpleasant side effects compared to bromocriptine [25]. Although no clinical trials have directly compared the mass-reducing effects of different dopamine agonists, collective evidence suggests that bromocriptine decreases pituitary tumour size by approximately 50% in two-thirds of patients, compared with a 90% decrease with cabergoline [25].

All dopamine agonist drugs have potential side effects of nausea, postural hypotension, headache, and nasal stuffiness, but these are more pronounced with bromocriptine and these side effects are relatively unusual with quinagolide and cabergoline. Patients should be warned about the risk of developing mood disturbances or impulse control disorders such as pathological gambling, hypersexuality, and/or uncontrollable spending which have previously been reported in patients taking dopamine agonists. Psychotic depression is occasionally seen with all of these drugs, which should be avoided in patients with a psychotic predisposition.

Cabergoline and pergolide are used in higher doses in the treatment of Parkinson's disease, and high cumulative doses of ergot-derived drugs have been found to be associated with the development of a fibrotic cardiac valvulopathy in these patients, possibly due to their action at the 5-HT2B serotonin receptor [26]. The doses used for hyperprolactinaemia are generally 5–10 times lower than in Parkinson's disease, but caution

and echocardiographic monitoring have been advised for pituitary patients, although almost all studies so far reported have shown no excess of valvulopathy in pituitary patients treated with low cumulative doses [27, 28]. This caution does not apply to quinagolide, as a non-ergot derivative.

Oestrogen Replacement

In patients in whom galactorrhoea and tumour bulk are not a problem, and the only issue is oestrogen deficiency due to anovulation, it may be possible to avoid the use dopamine agonists by simply replacing oestrogen. This should be considered in premenopausal anovulatory women to avoid symptoms of oestrogen deficiency and bone mineral loss, if dopamine agonists are not considered necessary or cause excessive side effects. The effect of oral oestrogen therapy on the growth of microadenomas has not been examined in a randomized controlled fashion. However, patients treated with oral contraceptives and oestrogen/progesterone replacement for 2 years have not shown an increase in tumour size [29, 30].

Ovulation Induction Without PRL Suppression

In most patients dopamine agonists are highly effective in restoring normal ovulatory function and fertility. However, some patients prove to be genuinely resistant to high doses of dopamine agonists, or have intolerable side effects, and for these women it may be necessary to use exogenous gonadotrophin therapy to induce ovulation.

Observation Alone

In postmenopausal patients, oestrogen replacement therapy can be considered if necessary for control of symptoms and for maintenance of bone mineral density. As for premenopausal women, pituitary size and serum prolactin levels should be monitored, as pituitary adenoma growth may otherwise progress without symptoms.

In both pre and postmenopausal patients it may be possible to withhold any treatment. Thus, if cyclical ovarian function and endogenous oestrogen levels are preserved despite moderate hyperprolactinaemia, or in an asymptomatic postmenopausal woman, there may be no indication for any treatment at all. In such cases, annual checks of serum prolactin level should be continued, and pituitary MR scanning should be carried out if prolactin levels progressively rise.

Surgery and Radiotherapy

Surgery is rarely indicated for microprolactinomas, and success rates have been disappointingly low even in highly specialist centres [31]. This does however depend on the surgical expertise available, and selective microadenomectomy may be worth considering in patients who are resistant to or develop troublesome side effects on dopamine agonist treatment. Radiotherapy is almost never indicated—most of these patients have otherwise normal pituitary function, and stand to develop hypopituitarism as a long-term consequence of pituitary irradiation.

Long-Term Follow-Up

Patients with microadenomas or idiopathic hyperprolactinaemia who achieve normal serum prolactin concentrations on dopamine agonist therapy should continue the treatment for at least one year. After that, an attempt to reduce the dose of dopamine agonist can be made, ensuring prolactin remains within normal range. If the prolactin levels continue to remain normal for two or more years and no adenoma is seen on magnetic resonance imaging, discontinuation of the drug can be considered.

If the drug is discontinued, prolactin should be measured initially every three months and annually thereafter. The risk of recurrence of hyperprolactinaemia appears to be highest in the first 12 months after cessation of the therapy, and ranges from 26% to 69% [32, 33]. In the event of recurrence of hyperprolactinaemia, a magnetic resonance imaging of the pituitary should be performed and dopamine agonist therapy resumed at the dose which previously resulted in normoprolactinaemia and shrinkage of the pituitary lesion.

Macroprolactinomas

Pituitary prolactinomas larger than 10 mm in diameter may present not only with anovulation, but also with mass effects of headache and visual field loss due to optic nerve compression, and with features of hypopituitarism. Most of these patients can be successfully treated with dopamine agonists, but there are some specific issues to be considered. In such patients the diagnosis should be reviewed at the start of treatment, as it includes not only prolactinoma, but also acromegaly, and non-functioning adenoma.

Dopamine Agonists

Dopaminergic drugs have proved highly effective even with very large prolactinomas, and both reduce prolactin levels and shrink adenomas in 85–90% of cases [22]. Tumour shrinkage is generally rapid, and is usually sufficient to remove pressure on the optic chiasm and allow recovery visual field loss within days. Most tumour shrinkage occurs within the first few weeks of therapy, but it may continue for months or even years. The mechanism is not fully understood but is likely to involve apoptosis of tumorous lactotroph cells.

Prolactin levels should be measured regularly, and dopamine agonist dose increased, as necessary, until normoprolactinaemia is achieved.

A magnetic resonance imaging should be repeated in 6 to 12 months after commencing the therapy to determine if the size of the adenoma has decreased.

A small number of patients will demonstrate resistance to dopamine agonists, defined as either failure to achieve normoprolactinaemia on maximum tolerated dose

and/or absence of tumour shrinkage by at least 50%. Most patients will achieve normal prolactin level with 0.25–3 mg of cabergoline weekly. Treatment-resistant patients are often managed with much higher doses of cabergoline, which carries a higher risk of cardiac valvular abnormalities, as previously observed in patients with Parkinson's disease taking at least 3 mg of cabergoline daily. Interval echocardiography might be required in this group of patients to assess for the development of valvular abnormalities.

Surgery and Radiotherapy

Surgery is rarely used even for large prolactinomas, because the majority respond so well to dopamine agonist drugs, and because it is extremely difficult to achieve a long-term cure with larger tumours with extrasellar extension of tumour tissue. However, it may have a role in debulking large tumours that have failed to show useful

shrinkage with drug therapy, in order to decompress and protect the optic chiasm.

Radiotherapy is rarely used alone, and has long-term adverse effects of hypopituitarism and increased risk of cerebrovascular disease, and risk-benefit analysis needs to be undertaken for individual patients. It does have a role in selected patients with macroprolactinoma to reduce the risk of tumour regrowth after debulking surgery, although hyperprolactinaemia tends to persist for many years, possibly due to radiation damage to hypothalamic dopaminergic neurons.

Even large tumours may shrink to disappearance, as judged by MR scanning, and long-term remission is seen in up to 50–60% of cases, which should prompt trial withdrawal of dopamine agonists in most patients who achieve sustained tumour shrinkage. However, in these patients, cautious follow-up with interval MR scanning is mandatory, as these tumours have displayed their potential for substantial growth.

Patients with hyperprolactinaemia and pituitary macroadenomas may have either large prolactinomas or non-functioning macroadenomas that are causing stalk disconnection hyperprolactinaemia. In general, prolactin levels exceed 10 000 mU/L (500 ng/ml) in most patients with macroadenomas composed of lactotrophic cells, whereas non-functioning tumours cause disconnection hyperprolactinaemia that rarely exceeds 5000 mU/L (20 ng/ml) [22]. The response to dopamine agonists may be useful in discriminating the true diagnosis before resorting to pituitary surgery. Although serum prolactin levels are suppressed by dopamine agonist therapy regardless of the nature of the tumour, substantial shrinkage of a macroadenoma suggests that it is in fact a prolactinoma, whereas non-functioning adenomas usually show little or no change in size.

Withdrawal of Treatment

In the case of a stable clinical picture, normoprolactinaemia and marked decrease in the size of macroadenoma, the dopamine agonist therapy can be decreased gradually to lowest dose necessary to maintain normal prolactin levels.

Discontinuation can be considered in those patients whose serum prolactin concentrations have been normal for more than two years, and whose adenoma can no longer be visualized by MRI [25, 34]. If the drug is discontinued, prolactin concentration and the size of the adenoma by MRI must be monitored.

Management of Pregnancy

In anovulatory hyperprolactinaemic patients, an important objective of therapy is the resumption of ovulatory cycles and pregnancy. However, oestrogen is a potential stimulus to pituitary lactotroph proliferation, and the very high oestrogen levels seen during normal pregnancy can result in clinically significant enlargement of a prolactinoma. Pregnancy is associated with significant pituitary enlargement in normal subjects, attributed mainly to proliferation of lactotroph cells, but in patients with prolactinoma there is a risk of substantial increase in tumour volume, sufficient to threaten the optic nerve.

In most patients, pregnancy is achieved by treatment of hyperprolactinaemia with dopamine agonists. Bromocriptine has been used for this purpose for over 30 years, and the accumulated evidence suggests that is it safe to use in pregnancy [35]. Similarly, cabergoline use in this setting does not appear to be associated with excess risk of miscarriage or congenital malformation [36]. However, it is probably advisable to limit exposure of any of these agents as far as possible, and all dopamine agonists of course would impair or prevent lactation and breastfeeding postpartum. In general, therefore, it is prudent to stop dopamine agonist agents as soon as pregnancy is confirmed, but patients should be warned of the small risk of adenoma enlargement once the restraint of dopaminergic treatment has been removed.

The actual risk of clinically important pituitary enlargement has been addressed by important meta-analyses and guidelines that help advise patients about pregnancy management [37, 38]. Women with microprolactinomas have a low risk (1–2%) of tumour enlargement during pregnancy, and dopamine agonists should be stopped when pregnancy is confirmed. Patients should be warned to report symptoms of pituitary expansion, including headache or visual impairment. Prolactin measurements can be made during pregnancy, but are highly variable between individuals and of limited value in guiding management. Women with macroprolactinomas have a higher risk of clinically significant or symptomatic tumour enlargement, approximately 20–30%. These patients can therefore be offered a choice of continuation of dopamine agonists throughout pregnancy, or stopping the drug with close monitoring of symptoms and visual fields until delivery. If a patient develops severe headaches and/or visual field abnormalities, a pituitary MR scan can be performed to evaluate the tumour mass, and if there is a significant threat to the optic chiasm, a dopamine agonist can be restarted or pituitary surgery considered. Prolactinoma debulking before pregnancy may reduce the risk, but carries its own risks and is probably not justifiable for these patients.

Women whose micro- or macroadenoma remained stable during pregnancy should not be discouraged from breastfeeding, and re-initiation of dopamine agonist can be delayed until breastfeeding is complete. However, women who experienced symptoms and signs in keeping with enlargement of lactotroph adenoma during or after pregnancy should not breastfeed and dopamine agonist therapy should promptly be initiated. A substantial (40–60%) proportion of women display features of disease remission after pregnancy obviating the need for treatment with dopamine agonist [39, 40].

REFERENCES

1. Grattan DR. 60 years of neuroendocrinology: the hypothalamo-prolactin axis. *J. Endocrinol*, 2015; **226**: T101–22.
2. Gerlo S, Davis JRE, Mager DL, Kooijman R. Prolactin in man: a tale of two promoters. *Bioessays*, 2006; **28**: 1051–5.
3. Ben-Jonathan N, LaPensee CR, LaPensee EW. What can we learn from rodents about prolactin in humans? *Endocr Rev*, 2008; **29**: 1–41.
4. Freeman ME, Kanyicska B, Lerant A, Nagy G. Prolactin: structure, function, and regulation of secretion. *Physiol Rev*, 2000; **80**: 1523–631.
5. Castrique E, Fernandez-Fuente M, Le Tissier P, Herman A, Levy A. Use of a prolactin-Cre/ROSA-YFP transgenic mouse provides no evidence for lactotroph transdifferentiation after weaning, or increase in lactotroph/somatotroph proportion in lactation. *J Endocrinol*, 2010; **205**: 49–60.

6. Melmed S. Mechanisms for pituitary tumorigenesis: the plastic pituitary. *J Clin Invest*, 2003; **112**: 1603–18.
7. Grattan DR, Kokay IC. Prolactin: a pleiotropic neuroendocrine hormone. *J Neuroendocrinol*, 2008; **20**: 752–63.
8. McNeilly AS. Prolactin and the control of gonadotrophin secretion. *J Endocrinol*, 1987; **115**: 1–5.
9. McNeilly AS. Lactational control of reproduction. *Reprod Fertil Dev*, 2001; **13**: 583–90.
10. Bachelot A, Binart N. Reproductive role of prolactin. *Reproduction*, 2007; **133**: 361–9.
11. Horseman ND, Zhao W, Montecino-Rodriguez E, *et al*. Defective mammopoiesis, but normal hematopoiesis, in mice with a targeted disruption of the prolactin gene. *EMBO J*, 1997; **16**: 6926–35.
12. Soto-Pedre E, Newey PJ, Bevan JS, Greig N, Leese GP. The epidemiology of hyperprolactinaemia over 20 years in the Tayside region of Scotland: the Prolactin Epidemiology, Audit and Research Study (PROLEARS). *Clin Endocrinol (Oxf)*, 2017; **86**: 60–7.
13. Peuskens J, Pani L, Detraux J, De Hert M. The effects of novel and newly approved antipsychotics on serum prolactin levels: a comprehensive review. *CNS Drugs*, 2014; **28**: 421–53.
14. Franks S. Polycystic ovary syndrome. *N Engl J Med*, 1995; **333**: 853–61.
15. Klibanski A, Greenspan SL. Increase in bone mass after treatment of hyperprolactinemic amenorrhea. *N Engl J Med*, 1986; **315**: 542–6.
16. Greenman Y, Tordjman K, Stern N. Increased body weight associated with prolactin secreting pituitary adenomas: weight loss with normalization of prolactin levels. *Clin Endocrinol (Oxf)*, 1998; **48**: 547–53.
17. Shibli-Rahhal A, Schlechte J. The effects of hyperprolactinemia on bone and fat. *Pituitary*, 2009; **12**: 96–104.
18. De Bellis A, Bizzarro A, Pivonello R, Lombardi G, Bellastella A. Prolactin and autoimmunity. *Pituitary*, 2005; **8**: 25–30.
19. Mukherjee A, Helbert M, Ryder WDJ, Borrow R, Davis JRE, Shalet SM. Failure of antibody response to polysaccharide antigen in treated panhypopituitary adults. *Clin Exp Immunol*, 2009; **156**: 271–7.
20. Smith TP, Kavanagh L, Healy M-L, McKenna TJ. Technology insight: measuring prolactin in clinical samples. *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 279–89.
21. Hattori N, Inagaki C. Anti-prolactin (PRL) autoantibodies cause asymptomatic hyperprolactinemia: bioassay and clearance studies of PRL-immunoglobulin G complex. *J Clin Endocrinol Metab*, 1997; **82**: 3107–10.
22. Bevan JS, Webster J, Burke CW, Scanlon MF. Dopamine agonists and pituitary tumor shrinkage. *Endocr Rev*, 1992; **13**: 220–40.
23. Webster J, Piscitelli G, Polli A, Ferrari CI, Ismail I, Scanlon MF. A comparison of cabergoline and bromocriptine in the treatment of hyperprolactinemic amenorrhea. Cabergoline Comparative Study Group. *N Engl J Med*, 1994; **331**: 904–9.
24. Verhelst J, Abs R, Maiter D, *et al*. Cabergoline in the treatment of hyperprolactinemia: a study in 455 patients. *J Clin Endocrinol Metab*, 1999; **84**: 2518–22.
25. Melmed S, Casanueva FF, Hoffman AR, *et al*. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
26. Roth BL. Drugs and valvular heart disease. *N Engl J Med*, 2007; **356**: 6–9.
27. Molitch ME. The cabergoline-resistant prolactinoma patient: new challenges. *J Clin Endocrinol Metab*, 2008; **93**: 4643–5.
28. Drake WM, Stiles CE, Bevan JS, *et al*. A follow-up study of the prevalence of valvular heart abnormalities in hyperprolactinemic patients treated with cabergoline. *J Clin Endocrinol Metab*, 2016; **101**: 4189–94.
29. Testa G, Vegetti W, Motta T, *et al*. Two-year treatment with oral contraceptives in hyperprolactinemic patients. *Contraception*, 1998; **58**: 69–73.
30. Corenblum B, Donovan L. The safety of physiological estrogen plus progestin replacement therapy and with oral contraceptive therapy in women with pathological hyperprolactinemia. *Fertil Steril*, 1993; **59**: 671–3.
31. Nomikos P, Buchfelder M, Fahlbusch R. Current management of prolactinomas. *J Neurooncol*, 2001; **54**: 139–50.
32. Biswas M, Smith J, Jadon D, *et al*. Long-term remission following withdrawal of dopamine agonist therapy in subjects with microprolactinomas. *Clin Endocrinol (Oxf)*, 2005; **63**: 26–31.
33. Kharlip J, Salvatori R, Yenokyan G, Wand GS. Recurrence of hyperprolactinemia after withdrawal of long-term cabergoline therapy. *J Clin Endocrinol Metab*, 2009; **94**: 2428–36.
34. Colao A, Di Sarno A, Cappabianca P, Di Somma C, Pivonello R, Lombardi G. Withdrawal of long-term cabergoline therapy for tumoral and nontumoral hyperprolactinemia. *N Engl J Med*, 2003; **349**: 2023–33.
35. Molitch ME. Management of prolactinomas during pregnancy. *J Reprod Med*, 1999; **44**(12 Suppl): 1121–6.
36. Ono M, Miki N, Amano K, *et al*. Individualized high-dose cabergoline therapy for hyperprolactinemic infertility in women with micro- and macroprolactinomas. *J Clin Endocrinol Metab*, 2010; **95**: 2672–9.
37. Molitch ME. Endocrinology in pregnancy: management of the pregnant patient with a prolactinoma. *Eur J Endocrinol*, 2015; **172**: R205–13.
38. Casanueva FF, Molitch ME, Schlechte JA, *et al*. Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. *Clin Endocrinol (Oxf)*, 2006; **65**: 265–73.
39. Auriemma RS, Perone Y, Di Sarno A, *et al*. Results of a single-center observational 10-year survey study on recurrence of hyperprolactinemia after pregnancy and lactation. *J Clin Endocrinol Metab*, 2013; **98**: 372–9.
40. Domingue M-E, Devuyst F, Alexopoulou O, Corvilain B, Maiter D. Outcome of prolactinoma after pregnancy and lactation: a study on 73 patients. *Clin Endocrinol (Oxf)*, 2014; **80**: 642–8.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

8.3.3 Premenstrual Syndrome

Deepthi Lavu, Radha Indusekhar, and Shaughn O'Brien

Introduction

Premenstrual disorders (PMD) are psychoneuroendocrine disorders which affect millions of women and their families all over the world. These disorders are characterized by cyclical recurrence of psychological, physical, and/or behavioural symptoms experienced in the luteal phase of the menstrual cycle and relieved soon after the onset of menstruation. They have a negative effect on the personal, social, and occupational lives of affected women leading to a reduced quality of life during their reproductive years. Premenstrual syndrome (PMS) affects 30–40% of women worldwide and 3–8% experience a severe form of PMS called premenstrual dysphoric disorder (PMDD).

The absence of objective tests and poor reliability relating to retrospective symptoms complicate diagnosis and efficient treatment. Until the recent publication of the International Society for Premenstrual Disorders (ISPMD) classification, there was a lack of consensus internationally among authoritative bodies regarding the classification, diagnosis, and management of these disorders [1]. The ISPMD classification now enables diagnostic accuracy and guides management of PMD. This chapter will review the classification, aetiology, symptoms, diagnosis, and management options recommended for treating PMD (PMS and PMDD).

Prevalence and Morbidity

The difference in PMD classification and symptom interpretation, prior to the ISPMD consensus, resulted in several studies providing a wide estimate of the prevalence of PMD [2]. The extent of impact of PMD is under-recognized by both the medical and non-medical community. It is considered that as many as 80–89% of women suffer from physiological premenstrual symptoms. Approximately 40% suffer from PMS and 3–8% have extreme symptoms with a diagnosis of PMDD. Symptoms occur during reproductive life and are absent before menarche, during pregnancy, and post-menopause. PMD are experienced cyclically over most ovulatory cycles and spontaneous remission is rare. Withdrawal of effective treatment usually results in symptom resurgence after a few cycles.

Women with PMDD are estimated to be incapacitated for 3.8 years of their reproductive life [3]. Morbidity results from the symptom severity experienced, including cyclical symptom recurrence, symptom duration, and the interference caused relating to daily functioning and productivity. This includes interpersonal relationships, work, and social life. PMDD is a chronic medical condition with an economic impact relating to loss of work productivity, which exceeds the medical costs involved in management [4].

Classification of PMD

Although PMD are widely prevalent and have a major impact on millions of women worldwide, until recently there was no international agreement among authoritative bodies on their classification. The ISPMD classification of PMD was formulated based on existing systems in use for defining and categorizing the disorders. This included the World Health Organization's *International Classification of Diseases (ICD)* 10th edition, the American College of Obstetricians and Gynecologists (ACOG) definitions, the Royal College of Obstetricians and Gynaecologists (RCOG) definition and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria. According to the ISPMD, PMD can be classified into Core and Variant PMD based on strict clinical diagnostic criteria [1].

Core Premenstrual Disorders

The key characteristic is the timing of the symptom occurrence. It must involve all or part of the two premenstrual weeks and must resolve during or shortly after menstruation with a clear symptom free interval before ovulation. Although there is no need to establish ovulation for diagnosing core PMD, the elimination of ovulation results in a reduction or absence of symptoms. See **Figure 8.3.3.1** and **Box 8.3.3.1**.

Core PMD were defined as PMS and PMDD. A distinction can be made between the definitions of PMD suggested by authoritative bodies to determine the ones which meet core PMD criteria and those which do not. See **Table 8.3.3.1**.

Variant Premenstrual Disorders

Symptoms in this group do not meet the criteria for core PMD. They are further classified into four types:

Premenstrual Exacerbation

There is increased severity of symptoms of an underlying medical or psychiatric disorder in the luteal phase of the menstrual cycle. For example, diabetes, asthma, depression, migraine, etc. may exhibit premenstrual exacerbation. Symptoms reduce following menstruation, but are not absent. See **Figure 8.3.3.2**.

Non-Ovulatory Premenstrual Disorders

Unspecified non-ovulatory ovarian activity is incriminated in causing symptoms in a small proportion of women. This area is

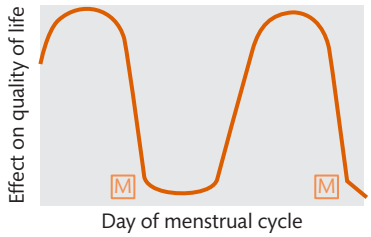


Figure 8.3.3.1 Core premenstrual disorder.
Adapted with permission from Management of premenstrual syndrome. *BJOG* 2016; DOI: 10.1111/1471-0528.14260. Copyright © 2016 Royal College of Obstetricians and Gynaecologists.

Box 8.3.3.1 Characteristics of symptoms in core PMD

1. Occur in ovulatory cycles
2. The type and number of symptoms are not specified—they may be somatic and/or psychological
3. Absent after menstruation and before ovulation (symptom free week in follicular phase)
4. Must recur in luteal phase
5. Must be prospectively rated for a minimum of two cycles
6. Must cause significant impairment to normal daily functioning, interfere with work, school, or interpersonal relationships or cause significant distress

poorly understood due to lack of robust evidence. Follicular activity in the ovary is hypothesized in symptom causation.

Progestogen-Induced Premenstrual Disorders

Symptoms are iatrogenic and arise due to use of exogenous progestogens such as cyclical hormone replacement therapy or combined oral contraception. Continuous administration of progestogens such as progestogen only contraception including a levonorgestrel intrauterine system (LNG-IUS) can cause PMD like side effects in women sensitive to progestogens but these would not be classified as core or variant PMD as there is no cyclicity. See Figure 8.3.3.3.

Premenstrual Disorders Without Menstruation

Symptoms are reported in some women who have medical or surgical amenorrhoea secondary to procedures such as hysterectomy with conservation of ovaries, endometrial ablation, or when using the LNG-IUS. These are due to an intact ovarian cycle. See Figure 8.3.3.4.

Symptoms

There are more than 200 reported symptoms. These encompass emotional, physical, behavioural, and cognitive domains (see Table 8.3.3.2). Probably, 80–95% of women experience some degree of premenstrual symptoms as a result of ovulation. These are called moliminal symptoms. The minimal severity of these

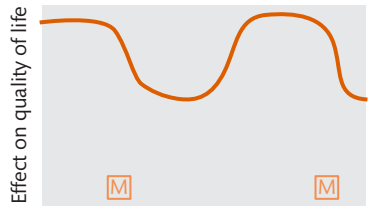


Figure 8.3.3.2 Premenstrual exacerbation.
Adapted with permission from Management of premenstrual syndrome. *BJOG* 2016; DOI: 10.1111/1471-0528.14260. Copyright © 2016 Royal College of Obstetricians and Gynaecologists.

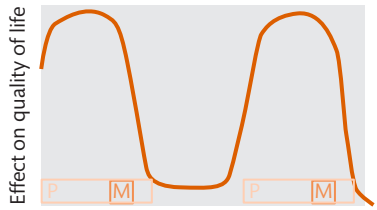


Figure 8.3.3.3 Progesterone-induced premenstrual disorder.
Adapted with permission from Management of premenstrual syndrome. *BJOG* 2016; DOI: 10.1111/1471-0528.14260. Copyright © 2016 Royal College of Obstetricians and Gynaecologists.

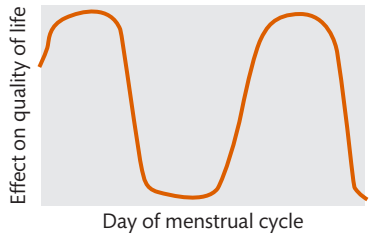


Figure 8.3.3.4 Premenstrual disorder with absent menstruation.
Adapted with permission from Management of premenstrual syndrome. *BJOG* 2016; DOI: 10.1111/1471-0528.14260. Copyright © 2016 Royal College of Obstetricians and Gynaecologists.

Table 8.3.3.1 Distinction between PMD definitions

PMS	PMDD
The negative impact of symptoms on daily functioning and severity of distress differentiates PMS from physiological premenstrual symptoms. Definitions which meet criteria for core PMD RCOG ACOG Definitions which do not meet core PMD criteria ICD-10	The diagnosis of PMDD is focused on psychological symptoms and lays emphasis on the number, character, and severity of symptoms. Definitions which meet criteria for core PMD DSM-5

Table 8.3.3.2 Characteristic symptoms of PMD

Physical	Emotional	Behavioural/ Cognitive
Breast tenderness or pain Abdominal swelling or bloating Headaches Acne Fatigue Weight gain Swelling of extremities Musculoskeletal pains or arthralgia	Mood swings Irritability Anger Loss of control Lack of interest in usual activities Loneliness Anxiety Depression Hopelessness Suicidal ideation or attempts	Change in appetite and food cravings Sleep disturbances Restlessness Social withdrawal Poor concentration Confusion

symptoms differentiates them from PMS, as they do not interfere with daily functioning of the individual. When symptoms are very severe, women might fulfil the criteria for premenstrual dysphoric disorder (PMDD). Often symptoms are similar to those experienced in other medical or psychiatric conditions. Their premenstrual timing and cyclical recurrence help differentiate PMD from other conditions. Complexity in diagnosis often arises when there is premenstrual exacerbation of an underlying medical or psychiatric condition, and one should be cautious when ruling out PMD in women suffering from chronic medical or psychiatric conditions.

Measurement of Symptoms

The ISPMD recommends two months of prospective rating of daily symptoms to enable accurate diagnosis. Studies have found retrospective recall of symptoms to be unreliable and bias in reporting can be due to the phenomenon of 'menstrual magnification' where any symptom exacerbated premenstrually is attributed to PMDs [5].

A variety of methods to quantify premenstrual symptoms is available with none universally preferable. Paper based methods include Calendar of Premenstrual Experiences, the Premenstrual Assessment Form, the Prospective Record of the Impact and Severity of Menstruation, Daily Record of Severity of Problems charts (DRSP) [6–9]. DRSP (Figure 8.3.3.5) is recommended by the RCOG as reliable and easy to use, but they are also noted to delay diagnosis and are reported to be inconvenient when completing and analysing symptoms hence are rarely used by clinicians [10]. Some clinicians use the premenstrual symptoms screening tool (PSST) to avoid unnecessary data collection when the diagnosis of PMS is unlikely [11]. Electronic methods of symptom quantification acceptable to women are being explored to overcome the limitations of using paper charts to aid diagnosis. One of the earlier methods was the Menstrual Symptometrics Device which quantified menstrual blood loss and premenstrual symptoms [12]. The Carolina Premenstrual Assessment Scoring System is a more recent computerized system developed to provide an electronic diagnosis of PMD [13, 14]. The latest advance in this area include an easy to use, commercially available mobile phone app, PreMentricS, which is currently being validated for use as a clinicians' aid to diagnosis (Figure 8.3.3.6).

Diagnosis

Establishing a diagnosis of PMD is complex as there are no objective tests available for this purpose. Clinical review should include a detailed review of medical, psychological, and psychiatric history, life stressors, substance abuse, and domestic violence. As a diagnosis of PMD is one of exclusion, physical examination is recommended including a breast and pelvic examination as dictated by the symptoms. Prospective symptom recording for 2 months prior to making a diagnosis is recommended [1, 10]. The benefits of prospective symptom charting include facilitation of diagnosis, providing a sense of symptom control for women by enabling a visualization of change in symptoms based on lifestyle modifications. and treatment tailoring. An accurate diagnosis supports the use of the most effective management strategies for PMD.

Differential Diagnosis

Due to the range of symptoms experienced, a number of medical, psychiatric, and psychological conditions can mimic PMD and vice versa (see Table 8.3.3.3).

Many of the aforementioned conditions exhibit premenstrual exacerbation of symptoms which can often complicate diagnosis and lead to ineffective treatments (see Table 8.3.3.4). Diligence in diagnosing and treating these conditions is essential for optimal symptom management.

Risk Factors and Comorbidity

Studies on twins have established a clear genetic influence in PMD [15]. Black women are noted to experience significantly less PMS and PMDD symptoms compared to Caucasian women [16]. Lifestyle risk factors for PMD include higher perceived stress levels, cigarette smoking, raised body mass index, earlier age of menarche, and greater alcohol or caffeine consumption. Recent evidence suggests that antiepileptic medication affects premenstrual symptoms experienced by women with epilepsy [17]. Studies have found an association between PMD and developing hypertension later in life [18].

Mood and anxiety disorders peak in the reproductive age group in women and are highly likely to be comorbidities with PMD [19]. Major depressive disorders are the most prevalent psychiatric disorder in women with PMD [20]. The relationship between bipolar disorder and PMD is less clear. Women with PMD have higher rates of obsessive-compulsive disorders and a quarter of women with PMD are affected by panic and anxiety disorders. Premenstrual exacerbation of schizophrenia has also been reported [21].

Pathophysiology

The definitive cause of PMD is unknown. There is no evidence to back hormonal imbalance as the cause. As there are no recorded differences in the hormonal levels, including thyroid stimulating hormone, oestrogen, progesterone, gonadotrophins, and adrenocorticotropic hormones, in women with PMD, and those without, there are no laboratory investigations to help diagnosis.

PMS appears to be triggered by ovulation as it is absent in anovulatory cycles. Progesterone released by the corpus luteum is implicated in symptom causation. It is hypothesized that the degree of sensitivity in the central nervous system (CNS) to the effects of hormones (especially progesterone) influences the severity of premenstrual symptoms experienced. A study where exogenous progestogen caused recurrence of symptoms in women with severe PMD post hysterectomy and bilateral salpingo-oophorectomy demonstrated the sensitivity of those women to the effects of progestogen. Women with PMD do not have the characteristic biological markers seen in depression [22]. They respond to antidepressant medication within days, while in women with depression and anxiety disorders, may require weeks of treatment before symptom

DAILY RECORD OF SEVERITY OF PROBLEMS

Please print and use as many sheets as you need for at least two FULL months of ratings.

Name or Initials _____

Month/Year _____

Each evening note the degree to which you experienced each of the problems listed below. Put an "x" in the box which corresponds to the severity: 1 - not at all, 2 - minimal, 3 - mild, 4 - moderate, 5 - severe, 6 - extreme.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Enter day (Monday="M", Thursday="R", etc)	>																															
Note spotting by entering "S"	>																															
Note menses by entering "M"	>																															
Begin rating on correct calendar day	>																															
1 Felt depressed, sad, "down,", or "blue" or felt hopeless; or felt worthless or guilty	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
2 Felt anxious, tense, "keyed up" or "on edge"	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
3 Had mood swings (i.e., suddenly feeling sad or tearful) or was sensitive to rejection or feelings were easily hurt	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
4 Felt angry, or irritable	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
5 Had less interest in usual activities (work, school, friends, hobbies)	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
6 **Had difficulty concentration**	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
7 Felt lethargic, tired, or fatigued; or had lack of energy	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
8 Had increased appetite or overate; **or had cravings for specific foods**	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
9 Slept more, took naps, found it hard to get up when intended; or had trouble getting to sleep or staying asleep	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
10 Felt overwhelmed or unable to cope; or felt out of control	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
11 Had breast tenderness, breast swelling, bloated sensation, weight gain, headache, joint or muscle pain, or other physical symptoms	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
At work, school, home, or in daily routine, at least one of the problems noted above caused reduction of productivity or **inefficiency**	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
At least one of the problems noted above caused avoidance of or less participation in hobbies or social activities	6																															
	5																															
	4																															
	3																															
	2																															
	1																															
At least one of the problems noted above interfered with relationships with others	6																															
	5																															
	4																															
	3																															
	2																															
	1																															

Figure 8.3.3.5 Daily Record of Severity of Problems chart for prospective recording of symptoms.

response. Several neurotransmitters have been implicated in the causation of progesterone hypersensitivity including GABA, serotonin, allopregnanolone, dopamine, endorphins, acetylcholine, cholecystokinin, and norepinephrine.

Serotonin

Serotonin has systemic hormone like effects influencing mood, behaviour, sexual function, appetite, and the immune system. Low serotonin levels cause panic, anxiety, irritability, and mood disturbances. Conditions characterized by depletion of serotonin such as a tryptophan poor diet result in symptoms similar to PMD. Antidepressants (selective serotonin reuptake inhibitors (SSRIs) and serotonin norepinephrine reuptake inhibitors (SNRIs)) which enhance synaptic serotonin levels, provide relief from symptoms in women with PMDs in contrast to medications which increase norepinephrine or dopamine.

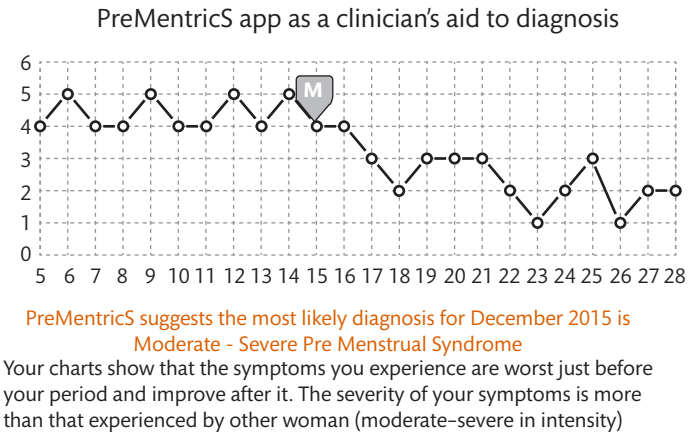


Figure 8.3.3.6 PreMentricS smartphone app allows daily recording of symptoms and graphically represents symptom scores to aid clinicians diagnosing premenstrual disorders.

Table 8.3.3.3 Differential diagnosis of PMD

Symptoms experienced	Disorders to exclude	Features differentiating symptoms from PMD
Headaches	Migraine	Timing of symptoms and absence of symptom free interval preovulation.
Abdominal bloating	Irritable bowel syndrome	Timing of symptoms and absence of symptom free interval preovulation.
Pelvic or abdominal pain	Chronic pelvic pain Endometriosis	Positive findings on clinical examination and investigations. Timing of symptoms helps exclude PMD diagnosis in women with chronic pelvic pain.
Fatigue	Chronic fatigue syndrome Anaemia Fibromyalgia Hypothyroidism	Absence of cyclical symptom recurrence. Clinical examination and laboratory investigations will yield positive results.
Breast symptoms	Fibrocystic breast disease Breast cancer	Clinical examination and radiological findings.
Emotional symptoms	Depression Anxiety disorders Substance abuse Post-traumatic stress disorder Psychosis	No pure cyclical symptoms Absence of asymptomatic phase pre ovulation. Past psychiatric history and failure to establish occurrence of symptoms in the luteal phase of the cycle should prompt psychiatric evaluation.
Combination of symptoms: Headaches, breast tenderness, sleep disturbances, hot flushes	Perimenopause	Lack of luteal timing. Elevated follicle stimulating hormone in perimenopause is of limited value.

Table 8.3.3.4 Disorders with premenstrual exacerbation

Somatic	Psychological
Endometriosis, adenomyosis, dysmenorrhoea Side effects of medications (especially hormonal contraception) Perimenopausal symptoms Migraines Anaemia Autoimmune disorders (e.g. systemic lupus erythematosis, multiple sclerosis) Hypothyroidism Fibromyalgia, chronic fatigue syndrome Breast disorders: fibrocystic, breast cancer, galactorrhoea Eating disorders	Major depressive disorder Bipolar disorder Anxiety disorders Seasonal affective disorder Personality disorders Substance abuse Stress Domestic violence

Allopregnanolone and GABA

Allopregnanolone, a neuroactive steroid, is a metabolite of progesterone which has been studied widely in relation to PMD. Allopregnanolone affects GABA, a major inhibitory neurotransmitter system in the CNS (see Figure 8.3.3.7). Of the different types of GABA receptors, GABA $_A$ receptor carries out the basic inhibitory control mechanism which is essential to the functioning of the CNS and regulation of mood and behaviour. This receptor is the site of action of allopregnanolone, alcohol, anticonvulsants, benzodiazepines, and barbiturates [23]. It is hypothesized that alteration of the GABA $_A$ receptor isoform is responsible for symptoms of PMD by causing decreased neurosteroid sensitivity and CNS excitability [23]. This is based on a study which used saccadic eye velocity as a proxy for central GABAergic tone and showed decreased responsiveness to pregnanolone and benzodiazepines in the luteal phase in women with PMS. SSRI treatments increase the GABA concentration in the occipital cortex of the brain hence providing evidence of a link between the serotonergic and GABAergic neurotransmitter systems.

Ovarian Steroids

The positive impact of ovulation suppressing treatments on premenstrual symptoms in women with PMD provides evidence of the involvement of ovarian steroids, oestrogen, and progesterone, in symptom development. Ovulation suppression using combined oral contraception has had varied results with symptoms being experienced all throughout the cycle in some women. Although still unclear, the mechanism of symptom development in relation to ovarian steroids is hypothesized on the fluctuations in the presence of the sex hormones and effects of the oestradiol on the renin angiotensin system and sensitivity of the CNS to the circulating (endogenous and exogenous) progestogens. Unlike other endocrine abnormalities where there are differences in the levels of the related hormones (such as thyroid hormones in thyroid disorders), studies have not found any difference in the levels of ovarian hormones in women with PMD compared to asymptomatic women, which strengthens the hypothesis suggesting that the differences lie in the sensitivity of women to these hormones.

Treatment of PMD

The key to effective management of PMD lies in accuracy of the diagnosis. A diagnosis based on two months of prospective symptom recording is recommended prior to initiating treatment for these disorders. In case of discrepancy in diagnosis between two months of symptom recording, further months recording is recommended. Once diagnosis is established, management should be individualized and the woman should be involved in the decision-making process as this leads to better understanding of the condition by the woman and greater treatment compliance. Contraceptive needs should be discussed in every case and a multidisciplinary team approach, involving the family physician, gynaecologist, psychiatrist/psychologist, should be the ideal although this is rarely possible. Severe PMD which involves suicidal ideation should involve urgent consultation with a psychiatrist.

The most logical way of treating PMD involves the elimination of ovulation and ovarian steroid activity. Alternatively, as serotonin depletion is hypothesized in symptom causation, medications which increase the serotonergic functioning can be used. Dietary and environmental interventions reduce sensitivity of the neuroendocrine system. Pilot studies investigating the use of the GABA$_A$ receptor modulating steroid, sepranolone, in treating symptoms have shown initial positive results [24].

During the time of completion of the prospective symptom charts it may be prudent to initiate treatment by giving advice regarding lifestyle changes. Although there is no symptom-based classification for PMD, in situations where there is one dominant symptom, symptom-based treatment is likely to be most effective. In general, treatment options for PMD can be divided into those which involve suppression of ovulation and those which do not. See Figure 8.3.3.8.

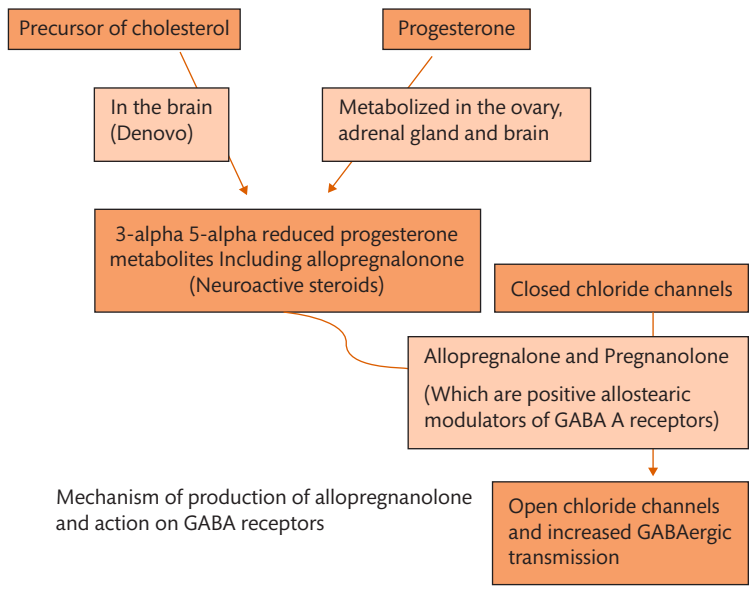


Figure 8.3.3.7 Mechanism of production of allopregnanolone and action on GABA receptors.

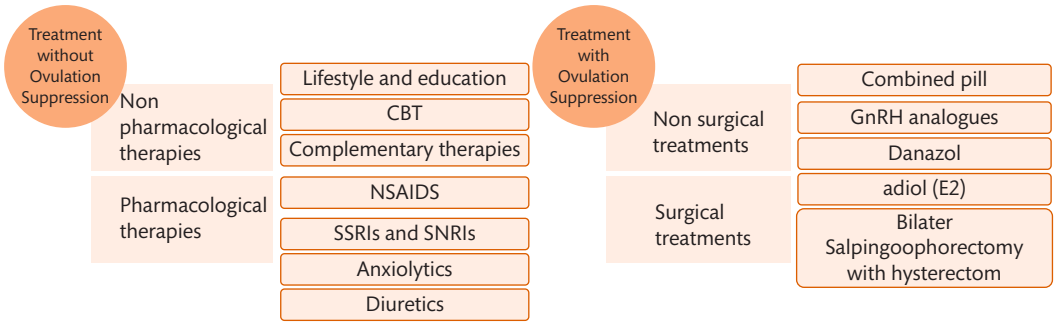


Figure 8.3.3.8 Management options for treating premenstrual disorders.

Treatments Without Ovulation Suppression

The strongest evidence for efficacy of non-pharmacological therapies include calcium, chasteberry, and cognitive behaviour therapy (CBT).

Lifestyle and Education

There are no randomized control trials (RCTs) to support the reported decrease in premenstrual symptoms with regular aerobic exercise. It is believed that relief in symptoms is due to the release of endorphins [25]. Education regarding the condition to allow satisfactory lifestyle modifications including reducing stress levels, incorporating exercise and weight loss have been suggested as an initial intervention in women presenting with PMD symptoms.

Dietary and Nutritional Supplements

Very few of the many dietary recommendations have been scientifically explored. Calcium supplementation has the strongest evidence of benefit among the nutritional supplements explored. Dosage of 1200 mg/day of calcium carbonate can be used for this purpose. 100 mg of vitamin B_6 daily has been shown to be slightly superior to placebo. Caution is recommended when increasing the dose of vitamin B_6 as it may precipitate peripheral neuropathy and it should be discontinued if any tingling or numbness is experienced in the extremities. Chasteberry (*Vitex agnus castus*) has shown encouraging results when dealing with emotional and physical PMD.

Cognitive Behavioural Therapy

This treatment can be initiated when prospective symptom recording is being carried out and its efficacy is thought to stem from improved coping strategies and modification of irrational thinking [25]. Recent studies have indicated some benefit of couples CBT compared to one-to-one therapy [26]. CBT in combination with SSRIs was found to have longer maintenance of treatment effect compared to SSRI alone.

NSAIDS

Useful for pain management, but not effective in the presence of emotional or behavioural symptoms.

Anxiolytics

Anxiolytics such as benzodiazepines may be of use in women with predominant anxiety/irritability symptoms. However, there is poor evidence to support their use in PMD and concerns regarding drug dependence have suggested usage restricted to the luteal phase of the menstrual cycle. Buspirone, a 5HT1A partial agonist, is an alternative as it has low dependence and has shown some efficacy in treating PMD.

Selective Serotonin Reuptake Inhibitors (SSRIs) and Serotonin Norepinephrine Reuptake Inhibitors (SNRIs)

There is strong evidence of symptomatic benefit with the use of SSRIs for PMD. It is important to rule out bipolar disorder prior to prescribing this class of medication as they may trigger a manic episode if used without a mood stabilizer. As the onset of action is short for SSRIs when used for PMD, their use can be limited to the luteal phase of the cycle, with efficacy similar to continuous daily dosing of SSRIs as demonstrated in recent Cochrane review [27]. Dosage of use is recommended to be similar to that used in the treatment of depression: fluoxetine 20–40 mg/day; sertraline 50–150 mg/day; paroxetine 10–20 mg or 12.5–25 mg of paroxetine continuous release (paroxetine CR) per day; and citalopram 10–20 mg/day. Most of the side effects of SSRIs are transient and include nausea, insomnia, headache, fatigue, diarrhoea, dizziness, and decreased libido or delayed orgasm. Although most of these disappear with continued use, the effects on sexual dysfunction reduce only after discontinuation of treatment, which can be a cause of discontinuation of treatment. The overall response to SSRIs is 50–60% and low response usually triggers the use of an alternative SSRI [25].

Diuretics

Spironolactone, an aldosterone receptor antagonist, may be useful in the management of women with mastalgia and severe bloating [28].

Complementary and Alternative Medicine

A number of other modalities which have had positive initial results but should only be offered in the context of a clinical trial are: acupuncture, Qi therapy, relaxation, reflexology, massage, krill oil, lavender oil, Chinese herbs, and transmagnetic stimulation.

Treatment by Suppression of Ovulation

These treatments are not recommended for women who wish to conceive as they involve the suppression of ovulation.

Combines Oral Contraception (COC)

The 'pill' is the most commonly used contraceptive method worldwide. As the COC prevents ovulation, it eliminates the luteal phase of the cycle. In spite of this, traditional COCs have not shown efficacy in the treatment of PMD. This is related to sensitivity of women

to the progestogen component of the pill leading to PMS-like side effects.

RCTs have shown that newer oral contraceptives, containing ethinyl oestradiol (20 μg) and drospirenone (3 mg) in a 24/4 regimen are effective in treating women with PMD. Drospirenone is an antiandrogenic progestogen. It is derived from 17-α-spironolactone and has antimineralocorticoid properties. A shorter hormone free interval in the 24/4 regimen maintains sufficient sex steroid hormone levels to prevent follicular development and supresses ovarian steroid synthesis. This regimen is approved by the United States Food and Drug Administration (FDA) for use in the treatment of PMDD. Most clinicians use this pill continuously (omitting the four placebo tablets), but there is no supportive research for this strategy.

Gonadotrophin-Releasing Hormone (GnRH) Analogues

GnRH agonists down-regulate gonadotrophin release, hence suppressing ovarian function and the production of ovarian steroids. This reduces oestrogen and progesterone to postmenopausal levels. This treatment provides relief to women with PMD symptoms, but use is restricted by the cost and hypo-oestrogenic side effects. Add-back therapy with oestradiol and progestogen, although reducing the short-term vasomotor side effects and long term adverse effects (vaginal atrophy, increased cardiovascular risk, bone demineralization and potentially osteoporosis), may cause resurgence of PMS-like symptoms. A recent review found that add-back therapy does not decrease the efficacy of GnRH analogue treatment [29]. Various types of add-back therapies, including tibolone, have been attempted, but none was found to be superior. Treatment with GnRH analogues is reserved for situations where other approaches have failed. It is a useful way of predicting efficacy of hysterectomy and bilateral oophorectomy in women contemplating surgery as a treatment option for severe PMD. It can also be used to determine whether severe symptoms are related to PMDD or another comorbidity. GnRH antagonists have the advantage of being orally active and avoid the agonist 'flare', but there are no published studies.

Danazol

Danazol is an androgen analogue which inhibits gonadotrophins at high levels. Danazol has, in the past, been effective in treating PMD, but is rarely used due to its masculinizing side effects. There is a risk of virilization of a female fetus should pregnancy occur while taking it, necessitating use of reliable long acting contraception.

Oestradiol (E2)

Transdermal oestradiol (200 μg) effectively inhibits ovulation. No published studies support the use of oral oestrogens for this purpose. To prevent endometrial hyperplasia and irregular bleeding secondary to unopposed oestrogen, addition of a progestogen is recommended. This has been given in the form of 2.5 mg of norethisterone or 100 mg of oral micronized progesterone for the first 7–10 days of each calendar month or as an intrauterine system with 52 mg levonorgestrel in its core. The short duration of progestogen is to prevent the recurrence of symptoms in women already sensitive to progestogens. An intrauterine system may be an effective alternative treatment option in women as there is (usually but not always) minimal systemic absorption after the initial few months; it also provides effective contraception.

Tibolone

This is a synthetic steroid with oestrogenic, progestogenic, and androgenic properties. There is some evidence that 2.5 mg tibolone may be effective in treating PMS [30]. Further trials are necessary before recommending it as a treatment option.

Surgery

This treatment option involves bilateral oophorectomy with hysterectomy and may be contemplated in women over 35 years of age, whose family is complete, suffering from severe symptoms or where hysterectomy is indicated for other comorbidities. Careful selection of cases is necessary as it involves major surgery. GnRH administration to confirm the efficacy of the planned surgery is advised [10]. Hormone replacement therapy is then recommended at least up to the age of menopause. Unopposed oestrogen can be used in women where hysterectomy has been performed. In women who have had bilateral oophorectomy with conservation of the uterus, a progestogen is required for endometrial protection when oestrogen replacement therapy is prescribed. This can reintroduce symptoms of PMS and therefore is not a recommended surgical option for treating PMD.

Summary

Premenstrual syndrome affects millions of women all over the world with up to 3–8% of women suffering from a very severe form called PMDD. PMD occur during a woman's reproductive years causing significant impairment in relationships, productivity, and quality of life. The aetiology is unknown. This in combination with the absence of objective diagnostic tests complicate diagnosis and treatment. Prospective symptom quantification is required to make a diagnosis. Lifestyle modifications, CBT, and calcium can be initiated in the diagnostic phase. Once diagnosed, a personalized case-based approach is recommended for managing this complex condition. Treatments involving COC, GnRH analogues with add-back hormone therapy, for example, oestradiol patches along with progestin can be attempted in young women where ovulation suppression is acceptable. Luteal phase SSRIs are recommended for those not wanting hormonal therapies. In very severe conditions surgical options involving bilateral oophorectomy and hysterectomy can be considered. The ICD-11 was released recently in 2018 and this, unlike the previous version, recognises the category of PMDD [31].

REFERENCES

1. O'Brien PM, Bäckström T, Brown C, *et al*. Towards a consensus on diagnostic criteria, measurement and trial design of the premenstrual disorders: the ISPMD Montreal consensus. *Arch Women's Ment Health*, 2011; **14**: 13–21.
2. Direkvand-Moghadam A, Sayehmiri K, Delpisheh A, Kaikhavandi S. Epidemiology of premenstrual syndrome—a systematic review and meta-analysis study. *JCDR*, 2014; **8**: 106–9.
3. Halbreich U, Borenstein J, Pearlstein T, Kahn LS. The prevalence, impairment, impact, and burden of premenstrual dysphoric disorder (PMS/PMDD). *Psychoneuroendocrinology*, 2003; **28**: 1–23.

4. Chawla A, Swindle R, Long S, Kennedy S, Sternfeld B. Premenstrual dysphoric disorder: is there an economic burden of illness? *Med Care*, 2002; **40**: 1101–12.
5. Rapkin AJ, Chang LC, Reading AE, *et al*. Comparison of retrospective and prospective assessment of premenstrual symptoms. *Psychol Rep*, 1988; **62**: 55–60.
6. Mortola JF, Girton L, Beck L, Yen SS. Diagnosis of premenstrual syndrome by a simple, prospective, and reliable instrument: the calendar of premenstrual experiences. *Obstet Gynecol*, 1990; **76**: 302–7.
7. Endicott J, Halbreich U. Retrospective report of premenstrual depressive changes. *Psychopharmacol Bull*, 1982; **18**: 109–12.
8. Reid R. Premenstrual syndrome. *Curr Probl Obstet Gynecol Fertil*, 1985; **8**: 1–57.
9. Endicott J, Nee J, Harrison W. Daily record of severity of problems: reliability and validity. *Arch Women's Ment Health*, 2006; **9**: 41–9.
10. Green L. Management of premenstrual syndrome. *BJOG*, 2017; **124**: e73–e105.
11. Steiner M, Macdougall M, Brown E. The premenstrual symptoms screening tool for clinicians. *Arch Womens Ment Health*, 2003; **6**: 203–9.
12. Wyatt KM, Dimmock PW, Hayes-Gill B, Crowe J, O'Brien PM. Menstrual symptometrics: a simple computer-aided method to quantify menstrual cycle disorders. *Fertil Steril*, 2002; **78**: 96–101.
13. Epperson CN, Hantsoo LV. Making strides to simplify diagnosis of premenstrual dysphoric disorder. *Am J Psychiatry*, 2017; **174**: 6–7.
14. Halbreich U. Diagnosis of PMS/PMDD—the current debate. O'Brien S, Scmidt PJ, Rapkin AJ, eds. *The Premenstrual Syndromes: PMS and PMDD*. London: Informa Health Care UK, 2007: 17.
15. Jahanfar S, Lye MS, Krishnarajah IS. The heritability of premenstrual syndrome. Twin Research & Human Genetics, 2011; **14**: 433–6.
16. Pilver CE, Kasl S, Desai R, Levy BR. Health advantage for black women: patterns in pre-menstrual dysphoric disorder. *Psychol Med*, 2011; **41**: 1741–50.
17. Herzog AG, Smithson SD, Fowler KM, *et al*. Premenstrual dysphoric disorder in women with epilepsy: relationships to potential epileptic, antiepileptic drug, and reproductive endocrine factors. *Epilepsy & Behavior*, 2011; **21**: 391–6.
18. Bertone-Johnson ER, Whitcomb BW, Rich-Edwards JW, Hankinson SE, Manson JE. Premenstrual syndrome and subsequent risk of hypertension in a prospective study. *Am J Epidemiol*, 2015; **182**: 1000–9.
19. Kessler R. Epidemiology of women and depression. *J Affect Disord*, 2003; **74**: 5–13.
20. Cohen LS, Soares CN, Otto MW, *et al*. Prevalence and predictors of premenstrual dysphoric disorder in older premenopausal women. The Harvard Study of Mood Cycles. *J Affect Disord*, 2002; **70**: 125–32.
21. Yonkers KA, McCunn KL. Comorbidity of premenstrual syndrome and premenstrual dysphoric disorder with other psychiatric conditions. In: O'Brien PM, Rapkin AJ, Schmidt PJ, eds. *The Premenstrual Syndromes: PMS and PMDD*. Informa Healthcare, 2007: 184.
22. Rapkin AJ. The role of serotonin in premenstrual syndrome. *Clin Obstet Gynecol*, 1992; **35**: 629–36.
23. Sundstrom Poromaa I, Smith S, Gulinello M, *et al*. GABA receptors, progesterone and premenstrual dysphoric disorder. *Arch Women's Ment Health*, 2003; **6**: 23–41.
24. Bixo M, Ekberg K, Poromaa IS, *et al*. Treatment of premenstrual dysphoric disorder with the GABAA receptor modulating steroid antagonist Sepranolone (UC1010)-A randomized controlled trial. *Psychoneuroendocrinology*, 2017; **80**: 46–55.
25. Nevatte T, O'Brien PM, Bäckström T, *et al*. ISPMD consensus on the management of premenstrual disorders. *Arch Women's Ment Health*, 2013; **16**: 279–91.
26. Ussher JM, Perz J. Evaluation of the relative efficacy of a couple cognitive-behaviour therapy (CBT) for premenstrual disorders, in comparison to one-to-one CBT and a wait list control: a randomized controlled trial. *PLOS One*, 2017; **12**: e0175068.
27. Marjoribanks J, Brown J, O'Brien PMS, Wyatt K. Selective serotonin reuptake inhibitors for premenstrual syndrome. *Cochrane Database Syst Rev,* 2013; **6**: CD001396.
28. Rapkin AJ, Winer SA. The pharmacologic management of premenstrual dysphoric disorder. *Expert Opin Pharmacother*, 2008; **9**: 429–45.
29. Naheed B, Herman Kuiper J, O'Brien S. A systematic review-ovarian suppression with GnRH analogues for management of PMS. *BJOG*, 2016; **123**: 186.
30. Taskin O, Gökdeniz R, Yalcinoglu A, *et al*. Placebo-controlled cross-over study of effects of tibolone on premenstrual symptoms and peripheral beta-endorphin concentrations in premenstrual syndrome. *Human Reprod*, 1998; **13**: 2402–5.
31. World Health Organization. International classification of diseases for mortality and morbidity statistics (11th Revision), 2018.

8.4

Polycystic Ovary Syndrome and Other Androgen Excess Disorders

8.4.1 Polycystic Ovary Syndrome

Definitions, Phenotypes, Prevalence, and Genetics

Sezcan Mumusoglu and Bulent Okan Yildiz

Introduction

Polycystic ovary syndrome was first reported in 1935 by Stein and Leventhal in seven cases, with large stiff ovaries, amenorrhoea, and infertility. Five of the patients were hirsute and one had acne. Wedge resection resulted in two pregnancies and regular cycles in remaining [1]. As an heterogenous disorder, the definition of PCOS remains still controversial. Nevertheless, available international guidelines emphasize that androgen excess, ovulatory dysfunction and polycystic ovaries are the main features of the syndrome [2].

Clinical Features Defining Polycystic Ovary Syndrome

PCOS is not a disease, but a syndrome; therefore, it reflects various potential aetiologies with diverse clinical presentations. The principal clinical features of PCOS are hyperandrogenism (HA), ovulatory dysfunction (OD) and polycystic ovarian morphology (PCOM). A recent meta-analysis of 24 studies on PCOS prevalence from all around the globe including Europe, United States (US), Asia, and Australia, revealed that overall reported prevalence (95% CI) of hirsutism, biochemical HA, OD, and PCOM in population were 13% (8–20%), 11% (8–15%), 15% (12–18%), and 28% (22–35%), respectively [3].

Androgen Excess (Hyperandrogenism)

Androgen excess is detectable either by clinical exam or biochemical analysis. Hirsutism is the main diagnostic criterion of clinical HA whereas acne and androgenic alopecia are among other skin manifestations. Androgen excess presented by elevated serum androgen levels is referred as hyperandrogenaemia or biochemical HA. Most of the women with PCOS, but not all, have both clinical and biochemical evidence of HA [4].

Hirsutism is defined as excess growth of body or facial terminal hair in females, in a male-like pattern and assessed by using a modified Ferriman–Gallwey (mFG) scoring system in which nine areas of the body (the upper lip, chin, chest, upper and lower abdomen, thighs, upper and lower back, and upper arms) are scored from 0 to 4 by the observer [5]. Hirsutism is usually defined by a total mFG score of ≥4–6 depending on ethnicity. Hirsutism is present in up to 80% of patients with HA. The rates of hirsutism would show variation depending on geographical region and ethnicity and it is important to use a population-specific cut-off level. Standardized visual scales [6] should be preferred when assessing hirsutism with acknowledging that self-treatment is common and can limit clinical assessment [2]. Approximately 70% of women with hirsutism have PCOS [7].

Other uncommon and less specific markers of clinical HA are acne and alopecia. A meta-analysis of pooled data on studies of PCOS prevalence reported that the prevalence of acne and alopecia are 16% (95% CI, 8–26%) and 2% (95% CI, 0–5%) in adult women, respectively [3]. Approximately 40% of women complaining of severe acne have a risk of having PCOS, whereas only 18% of women with androgenic alopecia were found to have PCOS [8]. On the other hand, acne or alopecia alone are not commonly associated with and should not be considered as a sign of clinical HA [7].

Assessment of biochemical HA in clinical practice is most useful in establishing the diagnosis of PCOS and/or phenotype where clinical signs of HA, particularly hirsutism is unclear or absent [2]. Biochemical HA could not be reliably assessed in women on hormonal contraception and drug withdrawal is required for at least three months. The use of calculated free or bioavailable testosterone or free androgen index (FAI) is recommended for evaluation of

biochemical HA. FAI is calculated as the ratio of total testosterone (tT) divided by sex hormone binding globulin (SHBG) and multiplied by hundred (FAI = TT × 100/SHBG). Direct measurement of tT by chemiluminescence or radioluminescence immunoassays without extraction/chromatography is not generally recommended for children or women including PCOS due to low precision, sensitivity, and specificity of these assays. Commercially available assays for direct measurement of free testosterone (fT) are not reliable and should not be used in assessment of biochemical HA [9]. Most reliable results could be obtained by properly validated extraction/chromatography immunoassays or liquid chromatography-mass spectrometry (LC-MS). Measurement of androstenedione (A4) and dehydroepiandrosterone sulphate (DHEAS) might provide limited information if total or free testosterone are not elevated. Normal values of androgens should be determined from healthy non-hirsute, eumenorrheic women, or by a cluster analysis of a large sample of general population. At least 70% of PCOS patients presenting to the clinics show biochemical HA when high-quality assays are used [10]. We have previously reported in a population study that 76% of unselected women with PCOS had biochemical and/or clinical HA [11].

Ovulatory Dysfunction

Ovulatory dysfunction presents with menstrual irregularities (<21 or >35 days) often characterized by oligomenorrhoea, due to infrequent or absent ovulation. Oligomenorrhea is defined as menstrual cycle length >35 days or having <8 menstrual cycles in a year whereas amenorrhea is the absence of menstruation for three or more consecutive months. PCOS is the most common cause of oligo-anovulation. Among women with oligo-anovulation, ~70% will have a diagnosis of PCOS [12], while ~80% of the women with PCOS have oligo-amenorrhea [10].

Most women with PCOS have a history of menstrual irregularities starting from puberty or primary amenorrhea, whereas, some may have regular cycles at the beginning and subsequently develop irregular menstruation in association with weight gain [13]. Weight loss in those obese PCOS patients usually leads to a recovery of menstrual cycles [14]. Additionally, many women with PCOS tend to have more regular cycles by ageing [15], probably due to follicle loss through the process of ovarian ageing or a decrease in androgen levels by ageing [16].

Regular menses indicate ovulatory cycles in almost every woman in the general population [17]. On the other hand, having regular cycles might not be confirm regular ovulation particularly in women with HA [18]. A study revealed that 40% of 132 hirsute women claiming 'regular cycles' had actually OD and this entity is known as 'subclinical-OD' [10]. Accordingly, in patients with hirsutism or PCOM who have apparently regular menstrual cycles, luteal phase (days 21–24 of the cycle) progesterone levels need to be measured and a progesterone level ≥5 ng/ml should be used as a proof of ovulation.

Polycystic Ovaries

Polycystic ovarian morphology on ultrasound is the most common feature of PCOS affecting up to 1 in 3 adult women in the population (22–35% in prevalence studies) [3]. Introduction of the high frequency ultrasound into clinical use, in1990s, enabled observation of the histologic findings of polycystic ovaries non-invasively. Initially, presence of ≥10 antral follicles in each ovary between 2 to 8 mm in diameter with peripheral location and relatively increased amount of stroma on ultrasound had been defined as PCOM, without any threshold for ovarian volume [19]. In 2003, presence of PCOM was included in the set of diagnostic criteria known as Rotterdam criteria and presence of ≥12 antral follicles in either ovary between 2 to 9 mm in diameter and/or increased ovarian volume (>10 ml; calculated by the formula 0.5× length × width × thickness) was described as ultrasound criteria for PCOM [20]. It was recommended to use transvaginal rather than transabdominal ultrasound [10].

A recent meta-analysis including 12 population-based PCOS prevalence studies [3] reported that the lowest prevalence of PCOM (11%) was obtained from a study conducted with 5 MHz probe [21], whereas, the highest prevalence (53%) was observed in a study with 5–9 MHz ultrasound probe [8]. It appears that the ultrasound technology and frequency of the transducer effect definition of PCOM, the selected threshold for antral follicle count (AFC), the application of ovarian volume and the precise diagnosis of PCOM. Accordingly, 2018 international evidence-based guideline for the assessment and management of PCOS recommends the use of ≥20 antral follicles and/or an ovarian volume ≥10 ml on either ovary as a threshold for PCOM when an endovaginal ultrasound transducer with a frequency bandwidth that includes 8 MHz is used [2]. If using older technology or transabdominal ultrasound, the threshold for PCOM could be an ovarian volume ≥10 ml on either ovary due to difficulty or reliability of assessing follicle number with these approaches [2]. Considering the decrease of ovarian volume and AFC with ageing, age-specific cut-off values are needed to define PCOM.

It is noteworthy to mention that PCOM has a low predictive value for the diagnosis of PCOS in that only one out of four adult women with PCOM would eventually be diagnosed with PCOS [3]. On the other hand, pooled analysis of 13 studies including 3361 women with a diagnosis of PCOS revealed that ~75% of the women with PCOS have PCOM on ultrasound with using both above-mentioned sets of ultrasound criteria [10].

Anti-Müllerian Hormone (AMH), a polypeptide secreted only by granulosa cells of the pre-antral and antral follicles, has been proposed as an alternative surrogate marker for the detection of PCOM due to limitations and challenges of the ultrasound in the diagnosis of PCOS. Even though AMH levels are higher in women with PCOS than healthy women and correlate with AFC on ultrasound, the available assays show significant heterogeneity and lack standardization. Therefore, routine use of AMH as part of laboratory evaluation in PCOS is not yet recommended by current guidelines [2, 22].

Diagnostic Criteria of Polycystic Ovary Syndrome

Diagnostic criteria of PCOS in adult women are not necessarily applicable in adolescents. Some normal features of puberty including irregular cycles, ovulatory dysfunction, multicystic ovaries, and acne could be identical with features of PCOS. In this context, sets of diagnostic criteria for adults and adolescents have been reviewed separately (Table 8.4.1.1).

Table 8.4.1.1 Sets of diagnostic criteria for polycystic ovary syndrome in adults and adolescents

Features of PCOS	In adults			In adolescents	
	NIH 1990	Rotterdam* 2004	Androgen Excess Society 2006	ICPE 2017#	International PCOS Network* 2018
Hyperandrogenism (HA; clinical and/or biochemical HA)	✓	±	✓	✓	✓
Ovulatory dysfunction (OD)	✓[a]	±	±	✓[b]	✓[d]
Polycystic ovarian morphology (PCOM)	–	±	±	±[c]	–[e]
Diagnosis statement	HA and OD are indispensable features	Any 2 out of 3 features (HA, OD, PCOM) is sufficient	HA is a must and it should be accompanied by any one out of OD and PCOM or both	HA and OD are indispensable features	HA and OD are indispensable

NIH, National Institutes of Health; ICPE, International Consortium of Pediatric Endocrinology.

* Recommended sets of diagnostic criteria by the "*The international evidence-based guideline for the assessment and management of PCOS*" in 2008. # Severe cystic acne is suggested as optional criteria used in concert with indispensable criteria.

✓ A must criteria for diagnosis; ± It may present or not (optional); – Not identified or suggested features in diagnostic set.

[a] detected by only menstrual irregularities

[b] should persist 2 years beyond menarche

[c] PCOM used in concert with the required criteria, and defined as ovarian volume>10 cm^3

[d] reassessment is advised 8 years postmenarche

[e] should not be used <8 years after menarche

Diagnostic Criteria of Polycystic Ovary Syndrome in Adults

For the diagnosis of PCOS in adults, different groups have proposed three different sets of diagnostic criteria in the last three decades (Table 8.4.1.1) [20, 23]. The common prerequisite in all three sets is that the mimicking disorders should be ruled out before making the diagnosis of PCOS.

In 1990, an expert meeting sponsored by National Institutes of Health (NIH) produced the first set of diagnostic criteria [24]. As a result of a survey of participants, after the exclusion of mimicking disorders including non-classical congenital adrenal hyperplasia (NCAH), thyroid dysfunction, and hyperprolactinemia, the presence of clinical and/or biochemical HA and presence of menstrual irregularities was proposed as indispensable features of PCOS for diagnosis (Table 8.4.1.1). Majority suggested that detection of PCOM by ultrasound was not essential for the diagnosis even though it was in the list of possible diagnostic criteria [24].

In 2003, European Society for Human Reproduction and Embryology/American Society for Reproductive Medicine (ESHRE/ASRM) sponsored consensus workshop was held in Rotterdam and the second set of diagnostic criteria was proposed [20]. In addition to NIH criteria, PCOM on ultrasound was accepted as a diagnostic feature. Presence of any 2 out of 3 following features was sufficient for the diagnosis of PCOS: (i) clinical and/or biochemical HA; (ii) OD; and (iii) PCOM (Table 8.4.1.1). Rotterdam criteria expanded the definition of PCOS by adding new phenotypes to the NIH criteria (Figure 8.4.1.1).

In 2006, the Androgen Excess & PCOS Society (AE-PCOS) published a task force report on diagnosis of PCOS [23]. The conclusion was that PCOS is a mainly hyperandrogenic disorder and the presence of clinical and/or biochemical HA is indispensable. The second criterion in addition to HA could be OD or PCOM (Table 8.4.1.1). Hence, the AE-PCOS criteria were a compromise between NIH and Rotterdam criteria excluding the new non-hyperandrogenic phenotypes proposed by Rotterdam [23].

There are many similarities between each proposed set of diagnostic criteria. However, the following differences and limitations are worth discussing: (i) the original NIH criteria defines menstrual dysfunction by <6 menses per year, it fails to include the women with regular menses and 'subclinical-OD', whereas the Rotterdam

NIH 1990

AE-PCOS 2006

Rotterdam 2003

	Phenotype A	Phenotype B	Phenotype C	Phenotype D
Androgen excess (Clinical and/or biochemical-HA)	✓	✓	✓	–
Ovulatory dysfunction (OD)	✓	✓	–	✓
Polycystic ovaries (PCO)	✓	–	✓	✓
Distributions of PCOS phenotypes (%)	40–45		~35	~20

Figure 8.4.1.1 Phenotypes of polycystic ovary syndrome according to diagnostic criteria sets.

and AE-PCOS criteria do not; (ii) in Rotterdam and AE-PCOS criteria the diagnosis depends on the use of a reliable method and cut-off to describe PCOM; (iii) neither Rotterdam nor AE-PCOS criteria described how hypothalamic amenorrhea with PCOM will be differentiated from PCOS; (iv) using different sets of diagnostic criteria decreases the comparability of PCOS research worldwide and causes confusion in clinical practice.

In 2012, NIH Evidence-based Methodology Workshop on PCOS was convened and announced the following recommendations [25]; (i) assigning a new name which reflects the complex metabolic, hypothalamic, pituitary, ovarian, and adrenal interactions that characterize the syndrome, however, no specific name was suggested; (ii) using the broad, inclusionary diagnostic criteria of Rotterdam which includes the 'classic NIH' and the Androgen Excess and PCOS Society criteria, however with specifically identifying each subphenotype in research and clinical initiatives: phenotype A: HA + OD + PCO; phenotype B: HA + OD; phenotype C: HA + PCO; and phenotype D: OD + PCO (Figure 8.4.1.1); and (iii) improving the methods and criteria used to assess androgen excess and OD.

The guidelines and position statement papers published after the NIH workshop all support the use of broader Rotterdam criteria for the diagnosis of PCOS with explicit reporting of the specific phenotypes [2, 26, 27].

Diagnostic Criteria of Polycystic Ovary Syndrome in Adolescents

In pubertal transition after menarche, 85% and 59% of menstrual cycles are anovulatory in the first and third years, respectively [26]. Therefore, OD has not been considered as a sign of PCOS until it persists 2 years beyond menarche [28]. Similarly, PCOM is a very common finding among adolescents even by abdominal ultrasound when the Rotterdam criteria are applied [26]. However, after 2 years from menarche the thresholds for increased ovarian size in adolescents becomes similar of those for adults [29]. Therefore, using ovarian volume is a better option than using AFC for defining PCOM in adolescents [29]. With regard to definition of HA in adolescents, mFG scores are not affected by age and tT and fT levels 1 to 2 years after menarche are generally comparable with those in adults [30]. Therefore, documented persistent HA could be diagnostic feature in adolescents. Overall, evidence suggests that assessment and diagnosis of PCOS in adolescents, particularly before the age of 18 years may cause overdiagnosis of the syndrome [31].

In the last few years, four sets of diagnostic criteria have been proposed for PCOS in adolescents: First, in 2012, the Amsterdam ESHRE/ASRM-sponsored 3rd PCOS Consensus Workshop Group recommended to observe all three features (HA, OD, and PCO) for the diagnosis of PCOS but only in adolescents who are at least 2 years postmenarche and/or those with primary amenorrhea by the age of 16 years before diagnosis, after exclusion of related disorders [7]. Second, in 2013, Endocrine Society suggested that diagnosis should be based on clinical and/or biochemical HA in the presence of persistent oligomenorrhea after exclusion of other pathologies. However, Endocrine Society recommended against the use of PCO appearance to make diagnosis in adolescent [26]. Third, in 2017, International Consortium of Pediatric Endocrinology (ICPE) recommended that for the appropriate diagnosis of adolescent PCOS clinical and/or biochemical HA and menstrual irregularities 2 years beyond menarche are required after exclusion of related disorders. According to ICPE guideline PCO on ultrasound and severe cystic acne are optional features of PCOS and recommended to be used in concert with HA and menstrual irregularities to diagnose PCOS in adolescents (Table 8.4.1.1) [32]. Lastly, in 2018, 'The international evidence-based guideline for the assessment and management of PCOS' recommended both oligo-anovulation and HA as required criteria, however, suggested against use of PCO on ultrasound <8 years after menarche [2]. The guidelines also suggest that in adolescents with a feature of PCOS who do not meet diagnostic criteria an increased risk could be considered planning reassessment at or before full reproductive maturity after 8 years postmenarche (Table 8.4.1.1) [2].

Differential Diagnosis of Polycystic Ovary Syndrome and Exclusion of Mimicking Disorders

All mimicking and similar disorders should be excluded before making the diagnosis of PCOS. Thyroid dysfunction and hyperprolactinemia could mimic PCOS causing OD and should be excluded by measuring thyroid stimulating hormone and prolactin, respectively. These two disorders are relatively rare in PCOS. Accompanying thyroid dysfunction in 7563 PCOS patients was reported to be 1.15% whereas prevalence of hyperprolactinemia was 2.8% in women with PCOS [10].

In patients with HA, non-classical congenital adrenal hyperplasia (NCAH) should be excluded. NCAH is an autosomal recessive disease characterized with the deficiency of 21-hydroxylase. Depending on ethnicity, the prevalence of NCAH ranges between 1% and 10% in patients with HA [33]. The clinical presentation of NCAH is identical with the classical features of PCOS; including OD, clinical- and biochemical HA, and PCOM ~40% of the patients [34]. Basal morning 17-hydroxyprogesterone (17-HP) is measured in the early follicular phase for the diagnosis of 21-OH-deficient NCAH. A cut-off value >2 ng/ml detects NCAH with 90% sensitivity. Patients with a screening level greater than 2 ng/ml should undergo an acute adrenocorticotropic hormone (ACTH) stimulation test.

Other rare disorders such as adrenal or ovarian androgen-secreting neoplasms, Cushing's syndrome, acromegaly, disorders of severe insulin resistance including hyperandrogenic insulin-resistant acanthosis nigricans (HAIR-AN) and lipodystrophy should be excluded by appropriate tests in case of clinical suspicion.

Phenotypes of Polycystic Ovary Syndrome

PCOS presents with various phenotypes. The NIH 2012 evidence-based methodology PCOS workshop recommends the use of broader ESHRE/ASRM diagnostic criteria, with explicit reporting of specific phenotypes in all research studies and clinical care [25]. In 2009; an AE-PCOS Task Force recommended using nine subphenotypes considering the four features of OD, clinical HA, biochemical HA, and PCOM [10]. Later, in 2012, NIH Evidence-based Methodology

Workshop on PCOS recommended to use phenotypic approach with four subphenotypes in order to maximize the homogeneity and comparability in research and clinical initiatives NIH as just stated [25] (Figure 8.4.1.1). This approach was also endorsed by the 2018 international evidence-based guideline for the assessment and management of PCOS [2].

Phenotype A and B (Classic PCOS)

About 15–20% of women with PCOS will have phenotype A (HA + OD + PCO) whereas 30–35% will show phenotype B (HA + OD) (Figure 8.4.1.1). Women with classic PCOS appear to be more hirsute and obese, show a more irregular menstrual pattern and more likely to have insulin resistance, dyslipidaemia, hepatic steatosis with an increased risk for metabolic syndrome compared to ovulatory or non-hyperandrogenic phenotypes (C and D) [31, 35].

Phenotype C (Ovulatory PCOS)

Phenotype C (HA + PCOM) or ovulatory PCOS seem to affect ~35% of the women with PCOS in unselected population studies of the syndrome (Figure 8.4.1.1) [31]. Hirsutism scores, androgen and lipid levels, and risk of metabolic syndrome of women with phenotype C are in between classic PCOS and phenotype D [36].

Phenotype D (Non-Hyperandrogenic PCOS)

Phenotype D or non-hyperandrogenic PCOS is the least prevalent phenotype comprising ~15–20% of the women with PCOS. These patients have normal androgen levels and the mildest degree of metabolic dysfunction including insulin resistance and (Figure 8.4.1.1) [31]. Patients with phenotype D who have increased body mass index (BMI) might be more insulin-resistant compared to healthy women even though at a level lower than classic PCOS [37].

Referral Bias in Phenotypes of PCOS

Most of the available data linking PCOS with endocrine and metabolic dysfunction and long-term comorbidities are derived from studies where included patients are those referred to the clinics for medical care. It should be kept in mind that there are significant differences between clinical versus unselected populations of PCOS patients. Patients presenting to the clinics have more severe phenotypes, more severe clinical and biochemical HA and higher BMI compared to patients identified in unselected populations reflecting significant referral bias [38].

Prevalence of Polycystic Ovary Syndrome

There are substantial discrepancies between the results of the epidemiological studies investigating the prevalence of PCOS. For example, reported prevalence with Rotterdam criteria is between the range of 2–21% [3]. The variation might be due to differences of selected study populations (referral bias) [38], limitations within the sampling and protocols applied, and a lack of standardized definitions for the features of PCOS. The effects of race and ethnicity, particularly, on the clinical presentation of androgen excess and enhancement in the ultrasound technology over time for visibility of AFC might also contribute to the inconsistencies between prevalence studies.

Recently, we have reported a systematic review and meta-analysis of 24 studies on prevalence of PCOS from all around the globe including Europe, US, Asia, and Australia [3]. Since there was significant heterogeneity among the studies, we have also generated an appraisal tool, which showed that available studies had low to moderate methodological quality. The PCOS prevalence (95% CI) according to the diagnostic criteria of NIH, Rotterdam and AE-PCOS Society were 6% (5–8%, n = 18 trials), 10% (8–13%, n = 15 trials) and 10% (7–13%, n = 10 trials), respectively. Higher prevalence rates with Rotterdam and AE-PCOS Society criteria compared to NIH are due to expansion of definition with inclusion of more phenotypes in addition to classic PCOS defined by NIH criteria (Figure 8.4.1.2) [3].

Genetics of Polycystic Ovary Syndrome

It has long been recognized that PCOS clusters in families and that both genetic and environmental factors contribute to the development of the syndrome [39]. A twin study on heritability of PCOS from The Dutch Twin Register database estimated about 70% influence of genetic factors on the pathogenesis of PCOS [40]. Recently, genome-wide association studies (GWAS) have supported the idea that PCOS has an oligogenic/polygenic inheritance similar to type 2 diabetes and cardiovascular disease [41]. However, clarifying the type of inheritance is difficult due to incomplete penetrance and epigenetic modifications with contributions of environmental effects [39].

Several studies have examined association of PCOS with more than 100 candidate genes involved in gonadotropin secretion and action, androgen biosynthesis, and action, folliculogenesis, insulin action, and diabetes, metabolism, and inflammation [42, 43]. These studies have limitations including small sample size, not well-matched cases and controls and lack of replication [39, 44]. In this context, GWAS studies provided a new insight and a more comprehensive approach to identify the genetic basis of PCOS [45–50]. The mapped genes, genotype-phenotype associated genes, pathway associated genes, and functional genes elucidated by GWAS studies in Han Chinese and European ancestry are shown in Table 8.4.1.2. Although there are substantial ethnic differences, cross-ethnic meta-analysis of genetic variants for PCOS revealed that 12 of 17 genetic variants mapping to the Chinese PCOS loci, including *YAP1, RAB5B/SUOX, LHCGR, THADA, DENND1A, FSHR, c9orf3, SUMO1P1*, similar effect size and identical direction in PCOS patients from Northern European ancestry, indicating a common genetic risk profile for PCOS across populations and supports the hypothesis that PCOS is an ancient disorder that may have persisted for more than 50 000 years [51, 52].

Despite major advances in the field of PCOS genetics, identified susceptibility loci in GWAS studies so far appear to contribute less than 10% of the heritability of the disorder. In the future, identification of rare variants and functional gene networks by next generation sequencing along with epigenetic studies might further increase our understanding of the pathophysiology of the syndrome [53].

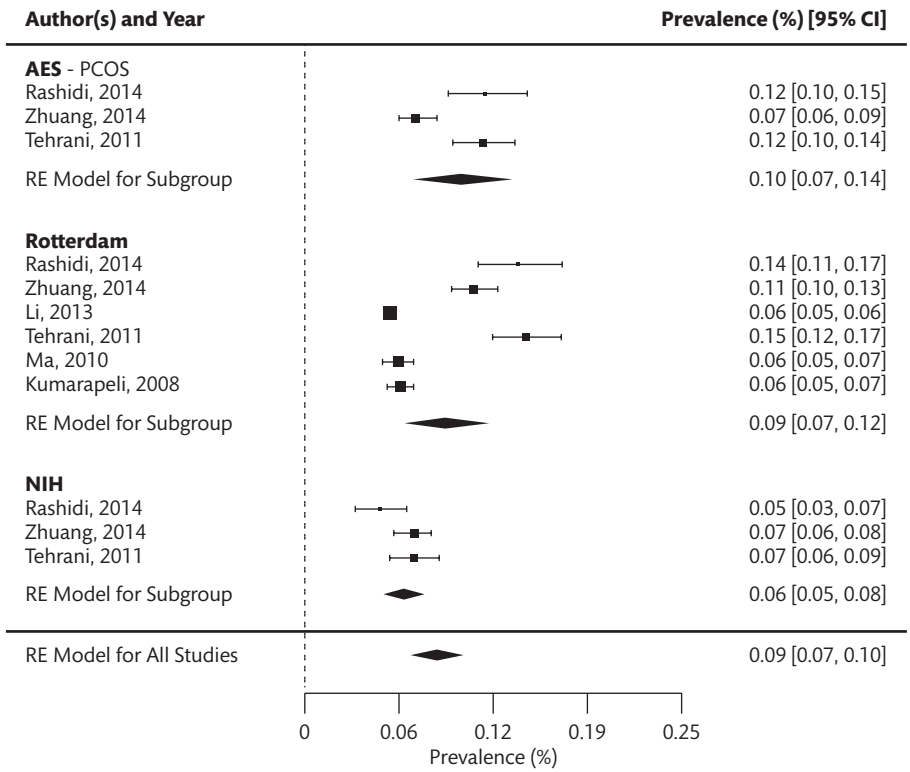


Figure 8.4.1.2 Prevalence (95% CI) of polycystic ovary syndrome according to unselected population studies.

Adapted with permission from Bozdag, G., Mumusoglu, S., Zengin, D., Karabulut, E., Yildiz, B.O., 2016. The prevalence and phenotypic features of polycystic ovary syndrome: A systematic review and meta-analysis. *Hum Reprod*. 31, 2841–55. Copyright © 2016, Oxford University Press.

Table 8.4.1.2 Genetic studies and identified genes in PCOS patients

Types of study	Asian ancestry	European ancestry
Associated gene polymorphisms with PCOS in case-control genetic studies [39, 44]	GnRHR, FSHR, FTO, IR, IRS	StAR, GnRHR, FTO, IR, IRS, VDR, PPAR-γ, IL-1β/ IL-1Ra
Mapped genes in GWAS studies [51]	*YAP1*, *RAB5B*, and *SUOX*, *LHCGR*, *THADA*, *DENND1A*, *FSHR*, *C9orf3*, *SUMO1P1/ZNF217*, *HMGA2*, *TOX3*, *INSR*	*YAP1*, *RAB5B*, *LHCGR*, *THADA*, *DENND1A*, *FSHR*, *c9orf3*, *SUMO1P1*, *GATA4–NEIL2*, *KCNA4–FSHB*, *ERBB4*, *ERBB2*, *ERBB2*, *RAD50*, *KRR1*
Genotype-phenotype associated studies [43]	*THADA*, *DENND1A*—associated with endocrine and metabolic disturbances and polycystic ovaries *LHCGR* and *INSR*—associated with anovulation *C9orf3*—associated with all classical PCOS features	*DENND1A*—associated with androgen excess and anovulation *FSHR*—associated with lower levels of FSH *RAB5B*—associated with glucose metabolism dysfunction
Pathway analysis studies using a data set obtained from GWAS studies [42, 45–52]	*INS*, *GNAQ*, *PLCB3*, *STXBP1*, *SMC3*, *PLCB2*, and *PLCZ1*—associated with oocyte meiosis and the regulation of insulin secretion	*LHCGR*, *FSHR*, *THADA*, *DENND1A*, *YAP1*, *RAB5B*, and *SUOX*—associated with alteration of gonadotropin secretion. *GATA4*—associated with gonadal development and transcription of steroidogenic genes. NEIL2—associated with DNA repair. FDFT1-associated with cholesterol-biosynthesis pathway
Functional studies elucidated by GWAS [45–52]	*LHCGR* and *INSR*- genetic variants of might have changed the expression level via modification on methylation	NA

GWAS, genome-wide association studies; FSH, follicle-stimulating hormone; GnRHR, gonadotropin-releasing hormone receptor; FSHR, follicle-stimulating hormone receptor; FTO, fat mass and obesity associated gene; IR, insulin receptor; IRS, insulin receptor substrate; VDR, vitamin D receptor; MTHFR, methylenetetrahydrofolate reductase; PPAR-γ, peroxisome proliferator-activated receptor gamma; IL-1β/ IL-1Ra, interleukin-1beta/IL-1 receptor antagonist; DENND1A, DENN/MADD domain-containing 1 A; NEIL2, endonuclease 8-like 2; GATA4, zinc-finger transcription factor; FDFT1, farnesyl-diphosphate farnesyltransferase; NA, not available.

Criteria	Phenotype A	Phenotype B	Phenotype C	Phenotype D
Hyperandrogenism	Present	Present	Present	Absent
Anovulation	Present	Present	Absent	Present
PCO	Present	Absent	Present	Present

According to the Rotterdam consensus criteria, the diagnosis of PCOS can be established by presence of 2 out of 3 criteria listed in the 1st column (1). Hence, there are 4 possible phenotypes of PCOS based on the various combinations of these criteria (1). Phenotypes A and B fulfil the NIH criteria and are associated with the greatest degree of insulin resistance and metabolic dysfunction (6).

REFERENCES

1. Stein IF, Leventhal ML. Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol*, 1935; **29**: 181–91.
2. Teede HJ, Misso ML, Costello MF, *et al*. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod*, 2018; **110**: 364–79.
3. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod*, 2016; **31**: 2841–55.
4. Azziz R, Sanchez LA, Knochenhauer ES, *et al*., Androgen excess in women: experience with over 1000 consecutive patients. *J Clin Endocrinol Metab*, 2004; **89**: 453–62.
5. Ferriman D, Gallwey JD. Clinical assessment of body hair growth in women. *J Clin Endocrinol Metab*, 1961; **21**: 1440–7.
6. Yildiz BO, Bolour S, Woods K, Moore A, Azziz R. Visually scoring hirsutism. *Hum Reprod Update*, 2010; **16**: 51–64.
7. Fauser BC, Tarlatzis BC, Rebar RW, *et al*. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. *Fertil Steril*, 2012; **97**: 28–38.e25.
8. Lauritsen MP, Bentzen JG, Pinborg A, *et al*., The prevalence of polycystic ovary syndrome in a normal population according to the Rotterdam criteria versus revised criteria including anti-Mullerian hormone. *Hum Reprod*, 2014; **29**: 791–801.
9. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*, 2007; **92**: 405–13.
10. Azziz R, Carmina E, Dewailly D, *et al*. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril*, 2009; **91**: 456–88.
11. Yildiz BO, Bozdag G, Yapici Z, Esinler I, Yarali H. Prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome under different diagnostic criteria. *Hum Reprod*, 2012; **27**: 3067–73.
12. Hamilton-Fairley D, Taylor A. Anovulation. *BMJ*, 2003; **327**: 546–9.
13. Laitinen J, Taponen S, Martikainen H, *et al*., Body size from birth to adulthood as a predictor of self-reported polycystic ovary syndrome symptoms. *Int J Obes Relat Metab Disord*, 2003; **27**: 710–15.
14. Yildiz BO, Knochenhauer ES, Azziz R. Impact of obesity on the risk for polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 162–8.
15. Elting MW, Korsen TJ, Rekers-Mombarg LT, Schoemaker J. Women with polycystic ovary syndrome gain regular menstrual cycles when ageing. *Hum Reprod*, 2000; **15**: 24–8.
16. Winters SJ, Talbott E, Guzick DS, Zborowski J, McHugh KP. Serum testosterone levels decrease in middle age in women with the polycystic ovary syndrome. *Fertil Steril*, 2000; **73**: 724–9.
17. Malcolm CE, Cumming DC. Does anovulation exist in eumenorrheic women? *Obstet Gynecol*, 2003; **102**: 317–18.
18. Chang WY, Knochenhauer ES, Bartolucci AA, Azziz R. Phenotypic spectrum of polycystic ovary syndrome: clinical and biochemical characterization of the three major clinical subgroups. *Fertil Steril*, 2005; **83**: 1717–23.
19. Adams J, Polson DW, Franks S. Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. *Br Med J (Clin Res Ed)*, 1986; **293**: 355–9.
20. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*, 2004; **19**: 41–7.
21. Moran C, Tena G, Moran S, Ruiz P, Reyna R, Duque X. Prevalence of polycystic ovary syndrome and related disorders in mexican women. *Gynecol Obstet Invest*, 2010; **69**: 274–80.
22. Dewailly D, Lujan ME, Carmina E, *et al*. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. *Hum Reprod Update*, 2014; **20**: 334–52.
23. Azziz R, Carmina E, Dewailly D, *et al*. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab*, 2006; **91**: 4237–45.
24. Zawadski JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif A, Givens JR, Haseltine FP, Merriam GR, eds. *Polycystic Ovary Syndrome*. Boston, MA: Blackwell Scientific Publications, 1992: 377–84.
25. National Institutes of Health. Evidence-based Methodology Workshop on Polycystic Ovary Syndrome, 3–5 December 2012, Executive Summary, Final Report, 2012.
26. Legro RS, Arslanian SA, Ehrmann DA, *et al*. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2013; **98**: 4565–92.
27. Conway G, Dewailly D, Diamanti-Kandarakis E, *et al*. European survey of diagnosis and management of the polycystic ovary syndrome: results of the ESE PCOS Special Interest Group's Questionnaire. *Eur J Endocrinol*, 2014; **171**: 489–98.
28. Homburg R, Lambalk CB. Polycystic ovary syndrome in adolescence—a therapeutic conundrum. *Hum Reprod*, 2004; **19**: 1039–42.
29. Fruzzetti F, Campagna AM, Perini D, Carmina E. Ovarian volume in normal and hyperandrogenic adolescent women. *Fertil Steril*, 2015; **104**: 196–9.
30. Rosenfield RL. The diagnosis of polycystic ovary syndrome in adolescents. *Pediatrics*, 2015; **136**: 1154–65.
31. Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril*, 2016; **106**: 6–15.
32. Ibanez L, Oberfield SE, Witchel S, *et al*. An international consortium update: pathophysiology, diagnosis, and treatment of polycystic ovarian syndrome in adolescence. *Horm Res Paediatr*, 2017; **88**: 371–95.
33. Azziz R, Dewailly D, Owerbach D. Clinical review 56: nonclassic adrenal hyperplasia: current concepts. *J Clin Endocrinol Metab*, 1994; **78**: 810–15.
34. Moran C, Azziz R. 21-hydroxylase-deficient nonclassic adrenal hyperplasia: the great pretender. *Semin Reprod Med*, 2003; **21**: 295–300.
35. Panidis D, Tziomalos K, Papadakis E, *et al*. Associations of menstrual cycle irregularities with age, obesity and phenotype in patients with polycystic ovary syndrome. *Hormones (Athens)*, 2015; **14**: 431–7.

36. Guastella E, Longo RA, Carmina E. Clinical and endocrine characteristics of the main polycystic ovary syndrome phenotypes. *Fertil Steril*, 2010; **94**: 2197–201.
37. Panidis D, Tziomalos K, Misichronis G, *et al*. Insulin resistance and endocrine characteristics of the different phenotypes of polycystic ovary syndrome: a prospective study. *Hum Reprod*, 2012; **27**: 541–9.
38. Ezeh U, Yildiz BO, Azziz R. Referral bias in defining the phenotype and prevalence of obesity in polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2013; **98**: E1088–96.
39. Azziz R, Carmina E, Chen Z, *et al*., Polycystic ovary syndrome. *Nat Rev Dis Primers*, 2016; **2**: 16057.
40. Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary syndrome in a Dutch twin-family study. *J Clin Endocrinol Metab*, 2006; **91**: 2100–4.
41. Strauss JF, 3rd, McAllister JM, Urbanek M. Persistence pays off for PCOS gene prospectors. *J Clin Endocrinol Metab*, 2012; **97**: 2286–8.
42. Shim U, Kim HN, Lee H, Oh JY, Sung YA, Kim HL. Pathway analysis based on a genome-wide association study of polycystic ovary syndrome. *PLoS One*, 2015; **10**: e0136609.
43. Cui L, Li G, Zhong W, *et al*., Polycystic ovary syndrome susceptibility single nucleotide polymorphisms in women with a single PCOS clinical feature. *Hum Reprod*, 2015; **30**: 732–6.
44. Chen Y, Fang SY. Potential genetic polymorphisms predicting polycystic ovary syndrome. *Endocr Connect*, 2018; **7**: R187–95.
45. Chen ZJ, Zhao H, He L, *et al*., Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat Genet*, 2011; **43**: 55–9.
46. Shi Y, Zhao H, Shi Y, *et al*., Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet*, 2012; **44**: 1020–5.
47. Hwang JY, Lee EJ, Jin Go M, *et al*. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J Hum Genet*, 2012; **57**: 660–4.
48. Lee H, Oh JY, Sung YA, *et al*., Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum Reprod*, 2015; **30**: 723–31.
49. Hayes MG, Urbanek M, Ehrmann DA, *et al*. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat Commun*, 2015; **6**: 7502.
50. Day FR, Hinds DA, Tung JY, *et al*. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat Commun*, 2015; **6**: 8464.
51. Louwers YV, Stolk L, Uitterlinden AG, Laven JS. Cross-ethnic meta-analysis of genetic variants for polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2013; **98**: E2006–12.
52. Unluturk U, Sezgin E, Yildiz BO. Evolutionary determinants of polycystic ovary syndrome: part 1. *Fertil Steril*, 2016; **106**: 33–41.
53. McAllister JM, Legro RS, Modi BP, Strauss JF, 3rd, Functional genomics of PCOS: from GWAS to molecular mechanisms. *Trends Endocrinol Metab*, 2015; **26**: 118–24.

8.4.2 Polycystic Ovary Syndrome

Reproductive Aspects

R. Jeffrey Chang

Introduction

The characteristic features of polycystic ovary syndrome (PCOS) are irregular menstrual bleeding, hirsutism, and distinctive ovarian morphology. Menstrual irregularity is due to chronic anovulation whereas excess hair growth results from androgen overproduction by the ovaries. The ovaries as demonstrated by ultrasonography are slightly enlarged and contain increased number of small follicles compared to that of normal ovaries. This appearance gives rise to the term polycystic ovary and the syndrome by which it is known. Despite this distinctive appearance, the morphogenesis of the polycystic ovary is unknown. While PCOS also is associated with metabolic dysfunction this section will focus on the reproductive aspects of the disorder.

Clinical Features

Menstrual Irregularity

Menstrual dysfunction characterized by irregular or absent bleeding is extremely common in PCOS. Generally, women with this disorder have never assumed regular menstrual cyclicity. At late puberty, the onset of menstrual function is commonly associated with irregular bleeding episodes that may persist for several years before assuming a cyclical monthly pattern. During this time, young adolescent women are anovulatory and bleeding may be infrequent or absent. The duration of menstrual irregularity may be variable. As a result, suspicion of PCOS in early adolescence based solely on menstrual bleeding is unreliable. However, clinical expediency may warrant an evaluation for menstrual irregularity after 2 years. Despite the preponderance of anovulation in this condition, a small percentage of hirsute women with PCOS may exhibit normal ovulatory function with monthly menstrual bleeding. The mechanism for maintaining regular ovulatory cycles is unknown.

As women with PCOS enter late reproductive life, a resumption of ovulation and normal menstrual cyclicity may ensue for unknown reasons. These individuals have a smaller cohort of follicles, elevated levels of follicle-stimulating hormone (FSH), and reduced serum

androgens compared to age-matched anovulatory PCOS women. The reasons for this transformation in some PCOS women are unclear.

Hirsutism

Excessive hair growth usually appears on the face, chin, and lower abdomen in a male pubic hair pattern. Hirsutism may also include hair on the extremities, abdominal flank, and back. More severe cases include the appearance of hair on the chest. A rapid onset of hair growth is important as this may reflect an androgen-producing neoplasm. Virilization is uncommon in women with PCOS. While the amount of hirsutism has been correlated to serum androgen concentrations, variation among individuals may exist due to ethnic differences. Thus, determination of androgen excess is based on clinical and biochemical evidence. The primary source of androgen overproduction is the ovary. In approximately 30–50% of individuals, serum adrenal androgens are elevated.

Excessive hair growth based on activity of the androgen receptor (AR) in PCOS women has been studied with inconsistent results. The AR gene encodes variable length CAG repeat polymorphism in the X chromosome; notably, length of the CAG repeat, and AR activity are inversely correlated. While shorter CAG repeat number has been associated with PCOS women compared to normal controls, it is unclear whether AR activity accounts for the inconsistent relationship between hirsutism and serum androgen levels in some ethnic populations of PCOS women.

Obesity, which is common in women with PCOS, may contribute to hirsutism by virtue of lowered sex-hormone-binding globulin (SHBG). Reduction in SHBG is directly correlated to obesity, thus giving rise to increased free testosterone levels. In particular, obese women with acanthosis nigricans are highly likely to have insulin resistance with compensatory hyperinsulinemia, which is known to suppress SHBG and contribute to excess ovarian androgen production.

The Polycystic Ovary

In women with PCOS the ovaries are slightly enlarged and contain a 2–3-fold increase in the numbers of growing pre-antral and small antral follicles compared with those of normal ovaries. The accumulation of small follicles of approximately 2–5 mm size appears to result from arrested growth at the mid-antral stage of development (Figure 8.4.2.1). Currently, the polycystic ovary (PCO) is defined as more than 12 follicles per ovary or an ovarian volume over 10 ml as demonstrated ultrasonographically [1]. Whether the greater follicle number is due to an increased rate of entry into the growing pool or a decreased rate of follicle atresia has not been established. It has been suggested that anti-Müllerian hormone (AMH) may be involved. AMH is produced by granulosa cells (GCs) of growing pre-antral and small antral follicles and negatively regulates follicle maturation. Its expression in growing pre-antral follicles of PCOS ovaries is decreased compared to that of normal ovaries, which may, in part, contribute to a greater number of small follicles [2]. While AMH levels in women with PCOS are elevated 2–3-fold compared to that of normal women, this increase probably reflects an increased number of growing pre-antral and small antral follicles in women with this disorder.

Another factor that may contribute to PCO morphology is androgen excess. It has been described that hyperandrogenic women with congenital adrenal hyperplasia or androgen-producing ovarian tumours display PCO. In addition, PCO are found in male-to-female transsexuals treated with long-term androgen. Notably, administration of high-dose testosterone to non-human primates increased ovarian size and follicle number, suggesting that PCO may be a consequence of increased local androgen production [3]. Recent studies have demonstrated that androgens decrease the expression of AMH and AMH receptor mRNA [4]. This suggests that an effect of androgen on the ovarian follicle population also may involve suppression of AMH.

Infertility

The leading cause of anovulatory infertility in women is PCOS as approximately 80% demonstrate irregular menstrual cycles. It

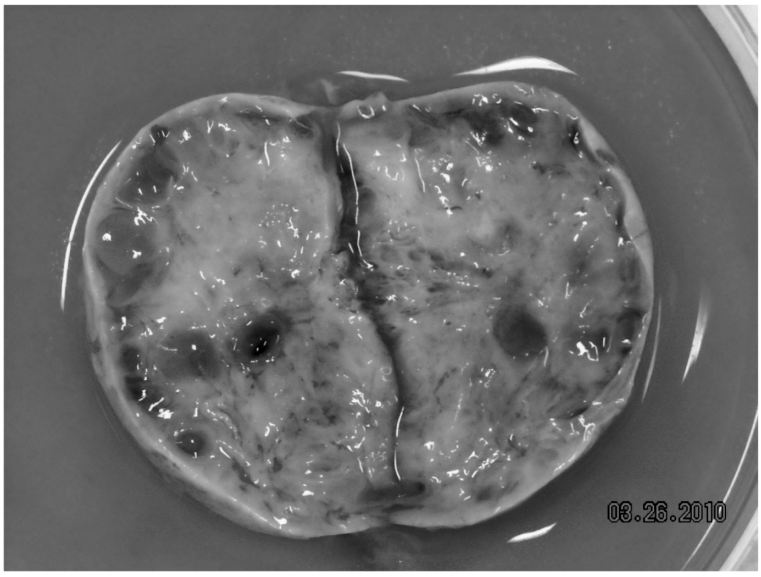


Figure 8.4.2.1 A cut section of a polycystic ovary. Note the peripheral distribution of small follicles within the ovarian cortex and the prominent central stroma. The ovary is encased by a thickened outer capsule.

is noteworthy that the presence of regular menstrual cycles does not preclude anovulation as demonstrated in 16% of women with PCOS as defined by hyperandrogenism and oligo-anovulation [5]. Obesity in a majority of PCOS women also may contribute to infertility. While the mechanism underlying ovulatory dysfunction in obese PCOS women is unclear, the interaction of PCOS and obesity likely serves to disrupt ovulation [6]. This may underlie restored ovulation in many obese PCOS women who sustain weight loss as opposed to PCOS individuals without obesity that fail to ovulation following weight loss.

In normal weight PCOS women, increased luteinizing hormone (LH) secretion and hyperandrogenaemia seem to impact follicle development, with insulin also enhancing gonadotropin actions on ovarian steroidogenesis and GC differentiation. By comparison, in obese women with PCOS the presence of metabolic disruption, including abnormal glucose-insulin homeostasis and adipogenic dysfunction, may be more important than LH in altering follicle development and oocyte maturation.

Altered Reproductive Physiology

The basis for the failure of ovulation in women with PCOS has not been clearly established. Despite evidence incriminating alteration of hypothalamic–pituitary–ovarian interaction, neither the inciting event nor dominant disruptor responsible for the reproductive phenotype has been defined. Inappropriate gonadotropin secretion, excess ovarian androgen production, arrested follicle development, and distinctive ovarian morphology have been well described and conceptualized into numerous pathophysiological schema (Figure 8.4.2.2). However, the underlying basis for PCOS remains to be established.

Inappropriate Gonadotropin Secretion

Women with PCOS exhibit increased LH secretion as shown by greater pulse frequency and amplitude, higher mean serum concentrations, and greater response to gonadotropin-releasing hormone (GnRH) compared to that of normal women [7]. As LH secretion reflects hypothalamic GnRH activity, it is assumed that activity of the hypothalamic GnRH pulse generator is increased. Conversely, increased activity of GnRH delivery favours the expression of LHβ gene expression [8]. In addition, excess androgen production may dictate LH pulse frequency in this disorder. Administration of progesterone and oestrogen to women with PCOS failed to slow LH pulse frequency whereas in normal women a significant reduction in the pulsatile LH secretion was noted. However, when women with PCOS were pretreated with an androgen-blocking agent before oestrogen and progesterone treatment, normal LH pulse release was subsequently restored [9].

In contrast, FSH secretion in women with PCOS is decreased compared to that observed in normal women. The mechanism for decreased FSH secretion in PCOS has not been fully determined, although negative feedback by chronic oestrogen secretion has been suggested. Alternatively, reduced serum FSH levels may result from increased activity of the GnRH pulse generator, which favours LHβ gene expression at the expense of the FSHβ gene [8].

Whether hyperinsulinemia influences inappropriate gonadotropin secretion in PCOS is not clear. In contrast to studies performed in rodents that suggest insulin may enhance LH response to GnRH, an effect of insulin in PCOS women has not been able to show consistent alterations in LH secretion or LH release following GnRH stimulation [10, 11]. Recently, it has been shown that AMH may influence GnRH secretion. *In vitro* studies of rat hypothalamic tissue have demonstrated that exposure to AMH was associated with an increase of GnRH neuronal activity [12]. This finding was supported *in vivo* by an injection of AMH into the lateral ventricle of female mice that produced a rapid increase of serum LH. Because serum AMH is elevated in women with PCOS, these results suggest that AMH may influence inappropriate gonadotropin secretion in this disorder.

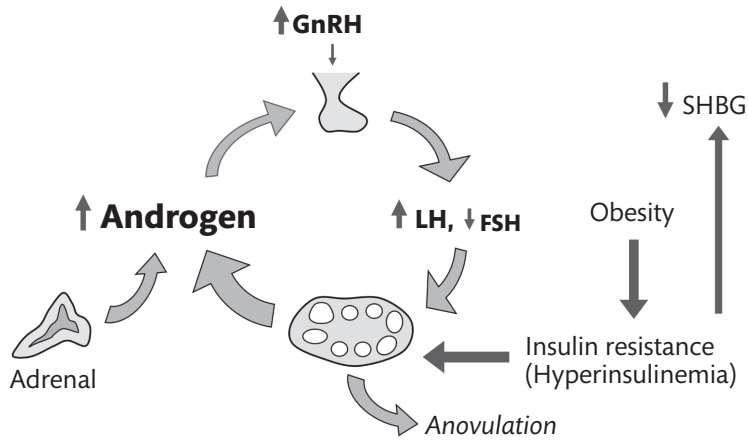


Figure 8.4.2.2 Proposed pathophysiology of polycystic ovary syndrome (PCOS). Increased hypothalamic GnRH activity drives pituitary LH hypersecretion that stimulates excess theca androgen production. Conversely, FSH secretion is decreased, which contributes, at least in part, to impaired follicle growth. Hyperandrogenaemia may inhibit steroid negative feedback on GnRH activity. Androgen excess in PCOS also is associated with increased visceral fat that contributes to insulin resistance, leading to hyperinsulinemia and dyslipidaemia. Altered metabolic function may disrupt ovarian folliculogenesis and increase theca cell androgen production. Hyperinsulinaemia lowers hepatic sex-hormone-binding globulin (SHBG) that results in an increase of bioactive testosterone in the circulation.

Disordered Follicle Function

Inadequate FSH Secretion

It is well established that women with PCOS exhibit reduced serum FSH levels. The relative lack of FSH has been incriminated as a cause for anovulation in PCOS women as ovulation may be achieved with the administration of FSH. The relative suppression of FSH secretion may be the result of increased activity of hypothalamic GnRH in women with PCOS as described previously. Alternatively, reduction of circulating FSH may also arise from the negative feedback of chronic oestrogen secretion characteristic of this disorder. However, GCs of PCOS women exhibit increased FSH binding and are readily responsive to FSH *in vitro* and *in vivo* [13]. This raises the issue of whether other factors within the microenvironment of the ovary exist to hinder follicle maturation.

Anti-Müllerian Hormone

Because AMH inhibits aromatase and circulating AMH levels are elevated in women with PCOS, excessive local production of this protein, combined with reduced FSH secretion, could set the stage for follicle arrest and decreased oestradiol (E_2) production by the ovary. This is consistent with *in vitro* studies showing that GCs incubated for 48 hours in the presence of FSH exhibited reduced AMH production [14]. In addition, during controlled ovarian hyperstimulation, serum AMH levels declined with length of treatment. However, these observations may also be explained by the decrease in AMH production seen in the final stages of follicle development in normal ovaries. The possible role of AMH in follicle arrest in PCOS is uncertain as endogenous factors that regulate this protein are currently unknown.

Premature Luteinization

It has been shown *in vitro* that GCs obtained from small PCOS follicles produced more progesterone following LH stimulation than that observed in GC from larger sized normal follicles [15]. These results suggested that PCOS GCs acquire LH receptors at an earlier stage of development and, subsequently, undergo luteinization prematurely. Consistent with this notion LH receptor mRNA expression has been shown to be greater in PCOS GCs compared to GCs from normal follicles [16]. Because LH secretion is increased in PCOS, this proposed mechanism may be a prelude to terminal differentiation and arrested follicle development.

It is unclear how small antral follicles in PCOS might gain early acquisition of LH receptors. It is known that FSH induces LH receptor content in cultured GCs. However, in women with PCOS, FSH secretion is reduced. Insulin may amplify FSH-induced LH receptor in GCs although inhibition of LH receptors by insulin has also been reported [17]. Moreover, androgen has been shown to inhibit FSH-induced LH receptor expression in cultured rat GCs [18]. Despite these findings, insulin and LH were shown to act synergistically in stimulating steroid production in GCs from PCOS follicles [19]. These results suggest that increased hyperinsulinemia in women with PCOS promotes LH action on GCs, and by extension, contributes to follicle atresia and eventual arrest.

Growth and Differentiation Factor-9

Studies have demonstrated that the most pronounced increase of growing follicles in PCOS occurs at the primary stage [20, 21]. This suggests that developmental transition from the primary to secondary stage may be slower in PCOS. A potential mechanism for restricted early follicle growth was reported by Filho *et al.* [22]. In women with PCOS the expression of growth differentiation factor-9 (GDF-9), an oocyte-specific member of the TGF-β family known to play a critical role in folliculogenesis, was decreased compared to that observed in normal oocytes at all developmental stages. Only 8–12% of oocytes of primary follicles in PCOS contained signal for GDF-9 compared to 96% noted in normal controls. In addition, the signal, when present in more advanced follicles, was weaker than corresponding normal follicles. In mice, adequate amounts of GDF-9 are required for growth beyond the primary follicle as female GDF-9 knock-out mice are infertile with a complete block in folliculogenesis at the primary stage [23]. Together, these results suggest that defects in follicle growth in PCOS may begin at the earliest stages of folliculogenesis.

Granulosa Cell Function

Despite chronic anovulation, isolated PCOS granulosa cells appear to be extremely sensitive to FSH compared to normal GCs. Clinically, FSH-stimulated E_2 responses have been shown to be greater in women with PCOS compared to those observed in normal women [24]. The heightened oestrogen response to FSH is consistent with ovarian hyper-responsiveness to gonadotropin stimulation commonly encountered in this disorder. In these clinical studies, FSH-induced rises of E_2 levels within three hours of administration, which suggests that any aromatase inhibition, previously suspected in PCOS, is relatively mild.

It is unknown whether PCOS granulosa cell are more sensitive to FSH primarily or secondarily due to intraovarian or extraovarian factors. Previously, it has been demonstrated that GCs obtained from unstimulated follicles of PCOS women have greater FSH binding of compared to GCs from normal ovaries [13]. Thus, increased FSH receptor expression in PCOS GCs may account for increased ovarian responsiveness to FSH.

Early *in vitro* studies have shown that androgens may enhance FSH-induced aromatase activity and thus, possible involvement of the FSH receptor. Subsequent studies in non-human primates revealed that testosterone treatment induced increases of FSH receptor mRNA expression, which likely contributed to greater ovarian size and increased follicle number observed in these animals [25]. Collectively, these findings suggest that androgen excess is responsible, at least in part, for enhanced GC responsiveness to FSH.

Insulin is another consideration that might impact GC function. *In vitro* studies have shown that PCOS GCs exhibit extreme sensitivity to insulin whether in the presence or absence of gonadotropin stimulation [19]. Thus, insulin may act as a 'co-gonadotropin' within the ovary. In PCOS women treated with an insulin-lowering drug, thiazolidinedione, GC responsiveness to FSH during insulin infusion was significantly increased compared with that observed prior to treatment [26]. These findings suggested that the GC in women with PCOS may be insulin resistant.

Theca Cell Function

Androgen excess appears to be the result of an inherent defect of theca cell (TC) steroidogenesis in women with PCOS. The enzyme cytochrome P450 c17A1 (CYP17) catalyses two distinct activities, 17α-hydroxylase and 17,20 lyase, that are essential to the production of ovarian androgens. Studies have indicated greater activities of 17α-hydroxylase, 17,20 lyase, and 3β-hydroxysteroid dehydrogenase in PCOS TCs than those observed in TCs from normal women, which likely form the basis for androgen overproduction in PCOS [27]. These results are commensurate with clinical studies, in which PCOS exhibited increased responsiveness of 17-hydroxyprogesterone to gonadotropin stimulation compared with normal women and has led to the concept of dysregulated expression of the CYP17A [28]. In addition, in PCOS women enhanced hCG-stimulated 17-hydroxyprogesterone production persisted despite suppression of endogenous gonadotropins by GnRH agonist. Together, these findings support the concept of a primary abnormality of TC steroid production.

Within the TC androgen production is driven by LH secretion, which activates CYP17A. It has been shown in TCs that LH stimulated androgen production is dose-related. These results are consistent with the positive correlation between elevated serum LH levels and circulating testosterone concentrations in women with PCOS. Moreover, in PCOS women treated with sex steroids or GnRH agonists, elevated LH levels are reduced or eliminated with a corresponding reduction in circulating androgen levels and hair growth [29]. Summarily, increased secretion of LH, combined with a highly responsive TC, is an essential component of androgen overproduction in PCOS.

In PCOS other factors may contribute to androgen excess, the most notable of which is insulin. Receptors for insulin, insulin-like growth factors I and II have been demonstrated in theca tissue from both normal and PCOS women. In addition, *in vitro* studies of TCs from normal women have shown that these growth factors may independently stimulate androgen production as well as enhance androgen responses to LH [30].

Consistent with *in vitro* studies, insulin-induced hyperandrogenaemia has been reported in women with PCOS following prolonged insulin infusion over 16 hours [31]. During the administration of insulin changes in gonadotropin levels were not observed. These findings coincide with studies that have shown reduced hyperinsulinemia and associated decreases of serum androgens without corresponding changes of LH in women with PCOS treated with insulin-lowering drugs.

Treatment

Hirsutism

The primary therapeutic approach to excessive hair growth is reduction of ovarian androgen production by suppression of ovarian steroidogenesis. Combination oral contraceptives containing combination oestrogen-progestin readily decrease circulating androgen levels and effectively slow the rate of hair growth. However, the range of response may be variable largely due to the severity of hirsutism at the time of treatment. Oral contraceptive therapy inhibits serum gonadotropin levels leading to a decrease of ovarian androgen production. In addition, the oestrogen component increases SHBG leading to a reduction of bioactive testosterone, whereas progestin serves to increase the metabolic clearance rate of testosterone as well as compete for 5α-reductase and the AR.

In adolescent PCOS with androgen excess, a proposed alternative to oral contraceptives includes a combination of low-dose pioglitazone–flutamide–metformin [32]. This therapy was as effective as ethinyl oestradiol-cyproterone acetate in reducing hyperandrogenaemia and also improved metabolic dysfunction and reduced markers of cardiovascular risk.

Antiandrogenic agents, either alone or more commonly in combination with oral contraceptives, may be considered in the treatment of hirsutism. Spironolactone is an aldosterone antagonist that competes for testosterone binding sites to lower the clinical effects of androgen excess. In addition, spironolactone appears to interfere with cytochrome P450 to inhibit steroid enzyme action and decrease androgen production. Considering its antiandrogenic activity, pregnancy should be avoided in individuals using spironolactone. Because this medication may decrease the action of aldosterone, serum potassium levels may increase and therefore should be monitored.

Other antiandrogens shown to have therapeutic benefit include flutamide and finasteride. Flutamide competes for the AR whereas finasteride inhibits 5α-reductase. Flutamide is not approved by the United States Food and Drug Administration for the treatment of hirsutism due to a reported increased risk for hepatic toxicity particularly in individuals that are obese. Because of its antiandrogenic effects, flutamide should be avoided during pregnancy.

Menstrual Irregularity

In PCOS women with irregular bleeding, the primary treatment option is combination oral contraceptives. In the majority of cases, regular withdrawal bleeding will result. In individuals who are unable to tolerate combination oral contraceptives, administration of progestin only containing oral contraception or intermittent progestin alone may decrease or eliminate bleeding. Long-acting progestational agents should be considered carefully due to the enduring effect after administration. Alternatively, an intrauterine device containing progestin also may be used. An additional benefit of progestin therapy is the reduced risk of endometrial cancer.

Insulin-Lowering Drugs

Insulin-lowering drugs improve insulin sensitivity in PCOS women with insulin resistance. Metformin, a biguanide, increases insulin sensitivity in the liver to reduce gluconeogenesis and hyperinsulinemia. Administration of metformin to women with PCOS decreases androgen levels, increases rates of spontaneous ovulation, and enhances ovulatory responses to clomiphene. However, in cases of severe obesity metformin has been shown to be considerably less effective. In addition, evidence for metformin ameliorating excessive hair growth is lacking. Side effects of metformin include gastrointestinal symptoms, which are dose-related and tend to resolve after several weeks. A rare adverse effect of metformin therapy is lactic acidosis, which may occur in individuals with systemic and debilitating diseases.

Infertility

Medical Treatment

In anovulatory women with PCOS that are overweight or obese, diet and lifestyle intervention are highly recommended as first-line therapy. Reduction in weight by 5–10% may successfully restore ovulation and fertility. When pharmacological therapy is required, clomiphene citrate has been considered first-line therapy for ovulation induction in PCOS women due to its ability to inhibit oestrogen negative feedback and induce an increase in FSH secretion. Recent studies have demonstrated that letrozole, an aromatase inhibitor, may also be considered as another first-line therapy for ovulation induction in PCOS women. Letrozole diminishes hypothalamic oestrogen negative feedback and is thought to transiently increase FSH release without depleting oestrogen receptors. Both clomiphene and letrozole appear to have equivalent efficacy with respect of successful ovulation induction and pregnancy rates. Another alternative therapy is metformin that reduces circulating insulin and androgen levels, inhibits hepatic glucose production, decreases intestinal glucose uptake, and increases peripheral insulin sensitivity. If monotherapy with these agents is not successful, then metformin and clomiphene citrate may be combined in an effort to improve ovulation and pregnancy rates before proceeding to low-dose gonadotropin therapy or laparoscopic ovarian drilling.

In the absence of ovulation, induction using the aforementioned drugs would warrant use of low-dose gonadotropin therapy. Because women with PCOS are prone to excessive follicle responsiveness to gonadotropin administration and potential ovarian hyperstimulation, caution must be exercised when using these preparations. The goal is to create a transient increase in FSH above a threshold dose for selecting a limited number of developing follicles. Use of exogenous gonadotropins is associated with increased risks for multiple pregnancy and severe ovarian hyperstimulation syndrome.

Surgical Approach

In the original description of PCOS in 1935, several women that had undergone wedge resection resumed ovulation. Over the past several decades therapeutic wedge resection has gained in popularity as electrocautery of the ovarian capsule during laparoscopy (i.e. laparoscopic ovarian drilling) induced high rates of ovulation and pregnancy [33]. Moreover, spontaneous ovulation appeared to persist following the procedure, although the duration was highly variable [34]. Laparoscopic ovarian drilling (LOD) is best suited for anovulatory women with PCOS who are resistant to clomiphene citrate or letrozole treatment. Disadvantages of this surgical method include surgical complications, adhesion formation, and premature menopause, particularly when large numbers of punctures are used, although, long-term surveillance of PCOS women undergoing LOD has been reassuring.

In vitro fertilization (IVF) is indicated for PCOS women who have associated infertility conditions, including tubal damage, severe endometriosis, and male factor infertility, or who require preimplantation genetic testing. In addition, IVF also is a reasonable option for PCOS women who fail clomiphene therapy, are older or at increased risk of multiple birth from gonadotropin therapy because the multiple birth rate can be reduced by transferring one embryo at a time.

REFERENCES

1. Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*, 2004; **19**: 41–7.
2. Stubbs SA, Hardy K, Da Silva-Buttkus P, *et al.* Anti-Mullerian hormone protein expression is reduced during the initial stages of follicle development in human polycystic ovaries. *J Clin Endocrinol Metab*, 2005; **90**: 5536–43.
3. Vendola KA, Zhou J, Adesanya OO, Weil SJ, Bondy CA. Androgens stimulate early stages of follicular growth in the primate ovary. *J Clin Invest*, 1998; **101**: 2622–9.
4. Laird M, Thomson K, Fenwick M, Mora J, Franks S, Hardy K. Androgen stimulates growth of mouse preantral follicles *in vitro*: interaction with follicle-stimulating hormone and with growth factors of the TGFβ superfamily. *Endocrinology*, 2017; **158**: 920–35.
5. Chang WY, Knochenhauer ES, Bartolucci AA, Azziz R. Phenotypic spectrum of polycystic ovary syndrome: clinical and biochemical characterization of the three major clinical subgroups. *Fertil Steril*, 2005; **83**: 1717–23.
6. Dumesic DA, Richards JS. Ontogeny of the ovary in polycystic ovary syndrome. *Fertil Steril*, 2013; **100**: 23–38.
7. Rebar R, Judd HL, Yen SS, Rakoff J, Vandenberg G, Naftolin F. Characterization of the inappropriate gonadotropin secretion in polycystic ovary syndrome. *J Clin Invest*, 1976; **57**: 1320–9.
8. Kaiser UB, Sabbagh E, Katzenellenbogen RA, Conn PM, Chin WW. A mechanism for the differential regulation of gonadotropin subunit gene expression by gonadotropin-releasing hormone. *Physiology*, 1995; **92**: 12280–4.
9. Eagleson CA, Gingrich MB, Pastor CL, *et al.* Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab*, 2000; **85**: 4047–52
10. Dunaif A, Graf M. Insulin administration alters gonadal steroid metabolism independent of changes in gonadotropin secretion in insulin-resistant women with the polycystic ovary syndrome. *J Clin Invest*, 1989; **83**: 23–9.
11. Mehta RV, Patel KS, Coffler MS, *et al.* Luteinizing hormone secretion is not influenced by insulin infusion in women with polycystic ovary syndrome despite improved insulin sensitivity during pioglitazone treatment. *J Clin Endocrinol Metab*, 2005; **90**: 2136–41.
12. Cimino I, Casoni F, Liu X, *et al.* Novel role for anti-Mullerian hormone in the regulation of GnRH neuron excitability and hormone secretion. *Nat Commun*, 2016; **7**: 1–12.
13. Almahbobi G, Anderiesz C, Hutchinson P, McFarlane JR, Wood C, Trounson AO. Functional integrity of granulosa cells from polycystic ovaries. *Clin Endocrinol*, 1996; **44**: 571–80.
14. Pellatt L, Hanna L, Brincat M, *et al.* Granulosa cell production of anti-Mullerian hormone is increased in polycystic ovaries. *J Clin Endocrinol Metab*, 2007; **92**: 240–5.
15. Willis DS, Watson H, Mason HD, Galea R, Brincat M, Franks S. Premature response to luteinizing hormone of granulosa cells from anovulatory women with polycystic ovary syndrome: relevance to mechanism of anovulation. *J Clin Endocrinol Metab*, 1998; **83**: 3984–91.

16. Jakimiuk AJ, Weitsman SR, Navab A, Magoffin DA. Luteinizing hormone receptor, steroidogenesis acute regulatory protein, and steroidogenic enzyme messenger ribonucleic acids are overexpressed in thecal and granulosa cells from polycystic ovaries. *J Clin Endocrinol Metab*, 2001; **86**: 1318–23.
17. May JV, McCarty K, Jr, Reichert LE, Jr, Schomberg DW. Follicle-stimulating hormone-mediated induction of functional luteinizing hormone/human chorionic gonadotropin receptors during monolayer culture of porcine granulosa cells. *Endocrinol*, 1980; **107**: 1041–9.
18. Jia XC, Kessel B, Welsh TH, Jr, Hsueh AJ. Androgen inhibition of follicle-stimulating hormone-stimulated luteinizing hormone receptor formation in cultured rat granulosa cells. *Endocrinol*, 1985; **117**: 13–22.
19. Willis D, Mason H, Gilling-Smith C, Franks S. Modulation by insulin of follicle-stimulating hormone and luteinizing hormone actions in human granulosa cells of normal and polycystic ovaries. *J Clin Endocrinol Metab*, 1996; **81**: 302–9.
20. Webber LJ, Stubbs SA, Stark J, *et al*. Prolonged survival in culture of preantral follicles from polycystic ovaries. *J Clin Endocrinol Metab*, 2007; **92**: 1975–8.
21. Maciel GA, Baracat EC, Benda JA, *et al*. Stockpiling of transitional and classic primary follicles in ovaries of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 5321–7.
22. Teixeira Filho FL, Baracat EC, Lee TH, *et al*. Aberrant expression of growth differentiation factor-9 in oocytes of women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2002; **87**: 1337–44.
23. Dong J, Albertini DF, Nishimori K, Kumar TR, Lu N, Matzuk MM. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature*, 1996; **383**: 531–5.
24. Coffler MS, Patel KS, Dahan MH, *et al*. Evidence for abnormal granulosa cell responsiveness to follicle stimulating hormone in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2003; **88**:1742–7.
25. Weil SJ, Vendola K, Zhou J, *et al*. Androgen receptor gene expression in the primate ovary: cellular localization, regulation, and functional correlations. *J Clin Endocrinol Metab*, 1998; **83**: 2479–85.
26. Coffler MS, Patel K, Dahan MH, Yoo RY, Malcom PJ, Chang RJ. Enhanced granulosa cell responsiveness to follicle stimulating hormone during insulin infusion in women with polycystic ovary syndrome treated with pioglitazone. *J Clin Endocrinol Metab*, 2003; **88**: 5624–31.
27. Nelson VL, Legro RS, Strauss JF 3rd, McAllister JM. Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. *Mol Endocrinol*, 1999; **13**: 946–57.
28. Barnes RB, Rosenfield RL, Burstein S, Ehrmann DA. Pituitary-ovarian responses to nafarelin testing in the polycystic ovary syndrome. *N Engl J Med*, 1989; **320**: 559–65.
29. Chang RJ, Laufer LR, Meldrum DR, *et al*. Steroid secretion in polycystic ovarian disease after ovarian suppression by a long-acting gonadotropin-releasing hormone agonist. *J Clin Endocrinol Metab*, 1983; **56**: 897–903.
30. Bergh C, Carlsson B, Olsson JH, Selleskog U, Hillensjö T. Regulation of androgen production in cultured human thecal cells by insulin-like growth factor I and insulin. *Fertil Steril*, 1993; **59**: 323–4.
31. Tosi F, Negri C, Perrone F, *et al*. Hyperinsulinemia amplifies GnRH agonist stimulated ovarian steroid secretion in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2012; **97**: 1712–19.
32. Ibanez L, Diaz M, Sebastiani G, *et al*. Treatment of androgen excess in adolescent girls: ethinylestradiol-cyproteroneacetate versus low-dose pioglitazone-flutamide-metformin. *J Clin Endocrinol Metab*, 2016; **96**: 3361–6.
33. Gjonnaess H. Ovarian electrocautery in the treatment of women with polycystic ovary syndrome (PCOS). Factors affecting the results. *Acta Obstet Gynecol Scand*, 1994; **73**: 407–12.
34. Lunde O, Djoseland O, Grottum P. Polycystic ovarian syndrome: a follow-up study on fertility and menstrual pattern in 149 patients 15–25 years after ovarian wedge resection. *Hum Reprod*, 2001; **16**: 1479.

8.4.3 Polycystic Ovary Syndrome

Metabolic Aspects

David A. Ehrmann and Susan Sam

Introduction

Polycystic ovary syndrome (PCOS) is a common endocrinopathy among reproductive age women associated with reproductive features related to disordered gonadotropin release and increased androgen production leading to anovulation, infertility, and clinical hyperandrogenism Figure 8.4.3.1. The most commonly used and

accepted criteria for diagnosis is based on the 2003 Rotterdam consensus that has been endorsed internationally [1]. The definitions, phenotypes, and prevalence of PCOS is covered in Chapter 8.4.1.

In the 1980s, a series of studies clearly established that PCOS is associated with a significant reduction in peripheral insulin sensitivity that is independent of adiposity [2–5]. Furthermore, obesity is common among women with PCOS especially in the United States and exacerbates insulin resistance associated with the disorder. The combination of both intrinsic and obesity-related insulin resistance places women with PCOS at high risk for other metabolic complications such as type 2 diabetes (DM2), metabolic syndrome, and possibly cardiovascular disease. Although the reproductive features of the disorder may improve with age, the associated metabolic abnormalities may actually worsen. Furthermore, the risk for insulin resistance varies according to the phenotype of PCOS identified by the Rotterdam criteria. Women who present with all three criteria for PCOS (phenotype A) or those who fulfil the NIH criteria (phenotype B) exhibit the greatest degree of insulin resistance [6]. This chapter will explore the pathophysiology of aberrant insulin action in women with PCOS and discuss the long-term metabolic complications and treatment strategies.

Pathophysiology of Insulin Resistance in PCOS

Insulin resistance is a common feature of PCOS that is independent of body weight and is the underlying cause for the heightened metabolic risk [2–5]. Non-obese women with PCOS have the same degree of insulin resistance as obese reproductively normal women and obese women with PCOS are more insulin resistant compared to reproductively normal women of the same body mass index (BMI) [3, 7]. Insulin resistance in PCOS is characterized by reduced sensitivity and responsiveness to insulin-mediated glucose utilization primarily in the skeletal muscle and adipose tissue even among non-obese women [3, 7, 8]. The cellular mechanisms for insulin resistance in PCOS have not been fully characterized. Studies on subcutaneous adipocytes of women with PCOS have revealed impaired insulin sensitivity [7, 9–11] and reduced glucose transport [7–9] related to lower expression of the insulin-sensitive glucose transporter, GLUT-4 [12, 13]. This defect is present even among non-obese women with this condition [7] although the finding has not been confirmed in all studies [8]. Additionally, subcutaneous adipocytes from women with PCOS display reduced insulin sensitivity to the antilipolytic effect of insulin [10]. These appear to be post-receptor defects since no reduction in insulin receptor number or alteration in insulin binding have been identified [7, 9, 11].

A predisposition to accumulate fat in central depots independent of BMI has been reported in PCOS especially in non-obese women [14–17] and predisposes to insulin resistance although this finding has not been demonstrated in overweight or obese women with this condition [18]. The predisposition to abdominal adiposity maybe related to chronic hyperandrogenism. Administration of a weak anabolic androgen to postmenopausal women over 9 months increases visceral fat [19]. Female to male transsexuals have an increase in visceral fat following administration of androgens [20]. Some, but not all, short-term interventions designed to lower androgen levels or block androgen action in women with PCOS demonstrate improved insulin sensitivity [21–25]. These findings have not been universal especially among obese women although inadequate sample size has been a limitation [26–31]. A reduction in adipose tissue expression of adiponectin associated with lower circulating levels of this adipokine has been demonstrated even among

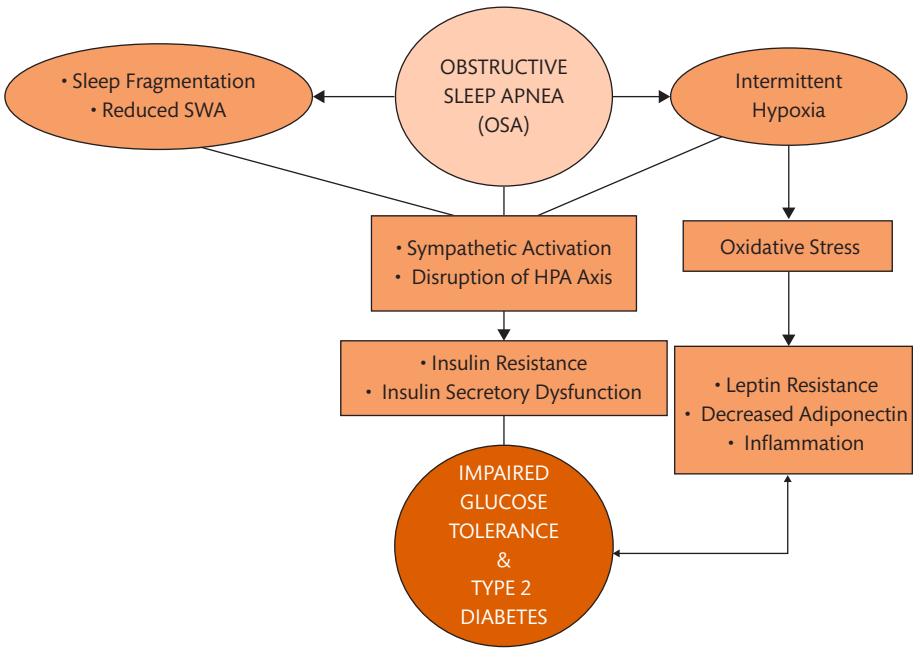


Figure 8.4.3.1 Women with PCOS display disordered gonadotropin release that favours increased luteinizing hormone (LH) compared to follicle-stimulating hormone (FSH) resulting in overproduction of testosterone with defects in follicular development. Insulin resistance is an intrinsic abnormality in PCOS and further contributes to ovarian androgen production as well as reduction in SHBG leading to higher levels of bioavailable testosterone.

non-obese women with PCOS [17, 32, 33]. In addition to reduced insulin action, women with PCOS have evidence for inadequate insulin secretion related to β-cell dysfunction placing them at heightened risk for DM2 [34].

Hyperinsulinemia resulting from insulin resistance promotes ovarian and adrenal androgen overproduction. *In vitro*, theca cells from women with PCOS overproduce androgens in response to insulin [4]. Treatment of women with PCOS with agents that lower insulin levels decreases both ovarian and adrenal androgens [35]. Hyperinsulinemia increases the availability of bioactive androgens such as bioavailable testosterone through its suppressive effect on circulating sex hormone-binding globulin (SHBG) [36]. Hence, insulin resistance and hyperinsulinemia adversely impact hyperandrogenism and reproductive complications of PCOS as well as contribute to development of metabolic consequences.

Additionally, there is heterogeneity in insulin resistance among the various phenotypes of PCOS according to the Rotterdam criteria. Insulin resistance and metabolic dysfunction is largely confined to women who present with the classical form of PCOS consisting of oligomenorrhea and hyperandrogenism especially in combination with polycystic ovaries [6].

Type 2 Diabetes Mellitus

Women with PCOS are at increased risk for glucose intolerance related to the intrinsic insulin resistance present in many women and worsened by the high prevalence of obesity. Furthermore, women with PCOS demonstrate pancreatic β-cell dysfunction and inadequate insulin secretion for the degree of insulin resistance [34]. About 30–40% of obese reproductive-aged PCOS women have been found to have impaired glucose tolerance (IGT), and about 10% have frank type 2 DM based on a 2-hour glucose level >200 mg/dl [37].

In a meta-analysis of 40 studies that examined the relationship between PCOS and glucose intolerance, a diagnosis of PCOS was associated with an odds ratio of 3.26 (95% CI: 2.17–4.90) for IGT and 2.87 (95% CI:1.445.72) for DM2 [38]. However, the risk varied according to ethnicity and geographic location. Asian women had a fivefold increase, followed by women in the Americas (fourfold increase) and women in Europe (threefold increase) [38]. Furthermore, women with PCOS are at increased risk of conversion from normal to IGT or from IGT to DM2 over time. The overall conversion rate from normal glucose tolerance to glucose intolerance over time is in the range of 2–5% per year [14, 39]. However, the risk is higher among obese women approaching 5–15% within 3 years [40].

Dyslipidaemia

Studies have demonstrated a number of lipid abnormalities in women with PCOS including elevated low-density lipoprotein (LDL) cholesterol and triglyceride levels compared to control women of the same age, BMI, and ethnicity [41, 42]. Smaller studies have also revealed lower high-density lipoprotein (HDL) cholesterol in PCOS compared to matched control women [43]. Women with PCOS have an atherogenic lipoprotein profile even if non-obese [44]. Current guidelines recommend a screening lipid profile on all women with PCOS [45].

Metabolic Syndrome

The diagnosis of metabolic syndrome is often based on ATP III criteria. These criteria include a waist circumference above 88 cm in women, HDL cholesterol less than 50 mg/dl, triglyceride above 150 mg/dl, hypertension defined as BP above 130/85 and impaired fasting glucose equal or greater than 100 mg/dl [46]. The presence of three of these criteria will establish the diagnosis of metabolic syndrome which is a strong predictor for cardiovascular disease.

The risk for metabolic syndrome is increased among women with PCOS. A meta-analysis of 16 studies comparing the prevalence of metabolic syndrome in women with and without PCOS, demonstrates an OR 3.01 (CI 2.06, 4.41); however, there is significant statistical heterogeneity [47]. After PCOS and control women are matched for BMI, the statistical heterogeneity was resolved, yet women with PCOS still had a greater prevalence of metabolic syndrome OR 2.2 (CI 1.36, 3.56). Additional subgroup analyses that included only lean women (BMI <25 kg/m^2) indicated a significant increase in risk for metabolic syndrome in PCOS with OR 3.00 (CI 1.30, 6.93) [47]. These data clearly indicate that the prevalence of metabolic syndrome is increased in PCOS women independent of obesity.

Non-Alcoholic Fatty Liver

Non-alcoholic fatty liver disease (NAFLD) is a hepatic complication that is strongly associated with insulin resistance and is the second most common indication for liver transplantation in the United States. A number of small cohort and cross-sectional studies indicate that the risk for NAFLD is increased among women with PCOS. in addition, a recent systematic review and meta-analysis [48] as well as a large retrospective analysis [49] show that the odds ratio for NAFLD is ~2.5 (95% CI 2.19–2.95) [48] and 2.23 (95% CI 1.86–2.66), respectively [49]. The increase in risk appears to be related to androgen excess in the current studies.

The current guidelines recommend that women with PCOS with metabolic risk factors may be screened using serum markers of liver dysfunction. If serum markers are elevated, non-invasive quantification of fibrosis by ultrasound and liver biopsy may be considered [45].

Cardiovascular Disease

Women with PCOS tend to have multiple risk factors for cardiovascular disease such as dyslipidaemia, DM2, abdominal adiposity, and metabolic syndrome. A number of studies reveal an increased prevalence of surrogates for cardiovascular disease such as increased carotid intima media thickness [50, 51] and coronary artery calcification [52, 53]. Yet at this time, conclusive evidence for increased cardiovascular disease morbidity and mortality among women with PCOS is lacking. A number of older studies did not find the prevalence of non-fatal or fatal cardiovascular disease to be increased although risk for non-fatal cerebrovascular disease was higher [54, 55].

In more recent studies, such as a substudy of the Women's Ischemia Evaluation Study (WISE), multivessel coronary disease was observed in 35% of women with PCOS compared to 25% of women without PCOS and the risk for stroke was also increased [43, 56]. Accordingly, a consensus statement by the Androgen

Excess and Polycystic Ovary Syndrome Society considers women with PCOS who are obese with a history of cigarette smoking, hypertension, glucose intolerance and subclinical vascular disease to be at risk and those with metabolic syndrome/DM2 to be at high risk for development of cardiovascular disease [43].

Sleep Apnoea

Women with PCOS have an increased prevalence of obstructive sleep apnoea (OSA) that may even exceed those in men [57, 58]. Both obesity and hyperandrogenism have been proposed to account for the high prevalence of OSA in PCOS [59, 60]. However, it appears that the increased risk may be independent of these two risk factors; the risk has been shown to be increased by nine times even after adjusting for BMI compared to control women [60, 61]. Furthermore, in women with PCOS the presence of OSA was associated with 2-fold higher fasting insulin levels and other surrogate measures of insulin resistance (Figure 8.4.3.2) [62]. The current guidelines for clinical management of women with PCOS suggest screening overweight/obese adolescents and women with PCOS for symptoms suggestive of OSA, and when identified, obtaining a definitive diagnosis using polysomnography. If OSA is diagnosed, patients should be referred for institution of appropriate treatment [45].

Gynaecological Cancers

Women with PCOS have many risk factors for endometrial cancer such as chronic anovulation with unopposed endometrial stimulation by oestrogen, central adiposity, and type 2 diabetes. A meta-analysis assessing the association between PCOS and endometrial cancer found an increased risk of developing endometrial cancer in PCOS (RR = 2.7; 95% CI, 1.0–7.29) [63]. Similar results were confirmed by a subsequent systematic review that revealed a threefold increased risk for this complication [64]. There are currently no data supporting routine screening with pelvic ultrasound or endometrial biopsy in asymptomatic women for endometrial cancer and such practices are not routinely recommended by the guidelines [45].

Evaluation of Women with PCOS for Metabolic Abnormalities

Evaluation should cover the multiple metabolic abnormalities discussed here which could be present in women with PCOS. The risk factors for metabolic complications are not uniform among women with PCOS [6]. Women who display all three diagnostic criteria for PCOS by Rotterdam consisting of menstrual irregularity, hyperandrogenism and polycystic ovaries or those who fulfil the NIH criteria consisting of menstrual irregularity and hyperandrogenism have been shown to manifest the greatest degree of insulin resistance and are at the highest risk for metabolic abnormalities [6]. Furthermore, the risk for metabolic abnormalities is modified by the presence of obesity, OSA, and family history of DM2 and cardiovascular disease. Hence, obtaining a comprehensive history that includes a detailed assessment of menstrual cycles, weight, and lifestyle factors such as smoking in addition to family history of PCOS, diabetes, and cardiovascular disease is essential. It is also important to assess for obesity including abdominal obesity and hypertension to further stratify metabolic risk.

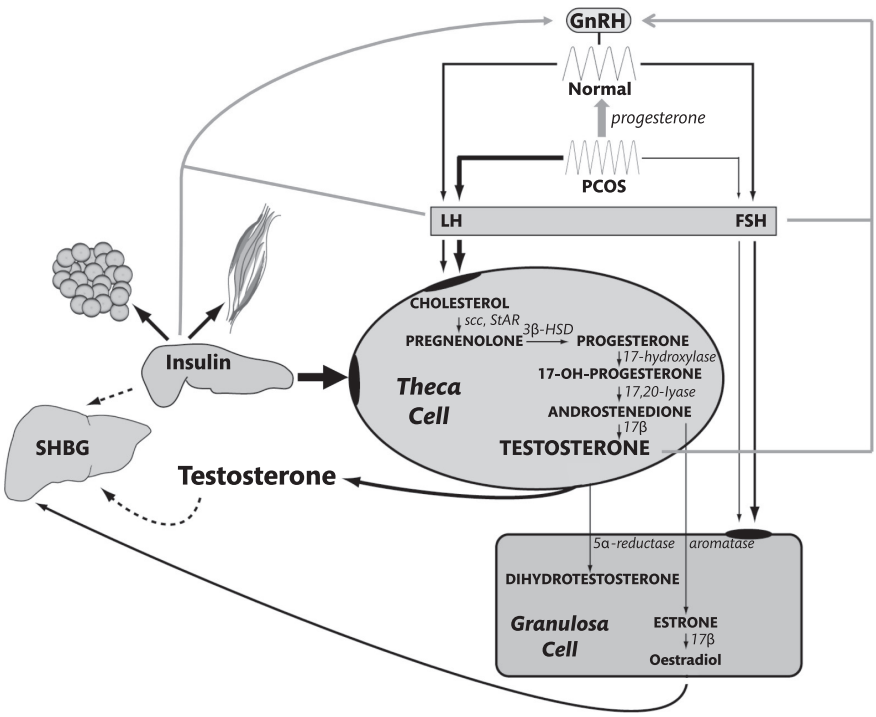


Figure 8.4.3.2 Obstructive sleep apnoea contributes to development of glucose intolerance and type 2 diabetes by activation of sympathetic pathways related to fragmented sleep and intermittent hypoxia. These defects predispose to insulin resistance, inflammation, and insulin secretory defects that eventually lead to development of glucose intolerance.

Due to the high prevalence of abnormal glucose tolerance in PCOS, screening for prediabetes and DM2 is indicated. The current guidelines recommend use of an oral glucose tolerance test with a 75-g oral glucose load to screen for IGT or DM2 [45, 65]. Haemoglobin A1c may have poor sensitivity in detecting IGT and women with PCOS have been shown to have 2-hour glucose values above 200 mg/dl on a 75-g oral glucose load consistent with DM2 despite normal fasting glucose [66]. Studies have also demonstrated increased risk of conversion to glucose intolerance over time in women with PCOS [67] although the rate of overall conversion among all women with PCOS may not be higher than 2% per year [39]. The risk for conversion from normal to IGT and from IGT to DM2 is high among obese women with PCOS, reportedly 5 to 15% within 3 years [40]. Because of the higher risk of conversion to IGT and DM2 in PCOS, most current guidelines recommend periodic screening although an interval screening period has not been clearly established [45, 65]. The risk for abnormal glucose tolerance is higher in women with PCOS with a family history of DM2 especially if obese so an individualized approach to screening for glucose tolerance is essential. Screening for metabolic abnormalities should not include fasting insulin levels since fasting insulin levels lack adequate accuracy and sensitivity [45] and are not useful in the diagnosis or management of PCOS or its metabolic sequelae. Specifically, there is no threshold value that defines insulin resistance that is independent of obesity. Any woman with PCOS who is overweight or obese should be presumed to be insulin resistant.

Women with PCOS should undergo screening with a fasting lipid profile due to a higher prevalence of dyslipidaemia [45, 65]. The guidelines do not provide specific recommendations regarding how often the screening should be repeated. Clinical judgement incorporating factors such as presence of dyslipidaemia at baseline, obesity, DM2, tobacco use, and personal or family history of cardiovascular disease are important considerations in this regard.

Due to a heightened risk for OSA especially among obese women with PCOS, screening for symptoms of this disorder is essential and supported by guidelines [45]. If women report symptoms suggestive of OSA, they should undergo a polysomnograph to obtain a definitive diagnosis followed by treatment [45].

Approach to Long-Term Treatment of Women with PCOS

The best long-term therapy for women with PCOS is a matter of debate, and often deduced from diabetes or cardiovascular prevention trials in similar populations, because such studies do not exist for women with PCOS. An individualized approach to each patient depending on symptoms and long-term risk is advised.

Lifestyle Modification

Lifestyle intervention resulting in weight loss is the principal treatment of metabolic dysfunction in obese women with PCOS and additionally often results in improvements in menstrual irregularity and even fertility. Even a small reduction in body weight by as little as 5% has been shown to improve ovulation in women with PCOS [68]. There is a paucity of data in regard to the 'optimal' diet for the management of obesity in women with PCOS. In a systematic review of dietary studies in PCOS, weight loss improved the presentation of PCOS regardless of dietary composition [69]. Subtle differences were detected among the diets with greater reduction in insulin resistance and improved quality of life with a low-glycaemic index diet, greater weight loss for a monounsaturated fat-enriched diet and greater improvements in depression and self-esteem for a high protein diet [69].

Exercise improves insulin resistance and reduces cardiovascular risk in women with PCOS if performed for at least 30 minutes per day for 5 days per week. However, exercise without concomitant caloric reduction is unlikely to induce weight loss in women with PCOS [70].

Bariatric Surgery

Bariatric surgery is increasingly utilized in morbidly obese patients as a first line obesity therapy [71]. The current recommendations are to utilize bariatric surgery in patients with a BMI greater than 40 kg/m^2 or with a BMI greater than 35 kg/m^2, and serious medical comorbidities [71]. Bariatric surgery should be considered in women with PCOS who fulfil the general criteria based on BMI and comorbidities. A meta-analysis of 13 studies of 2130 women of whom 45.6% had PCOS, demonstrated that bariatric surgery and subsequent weight loss results in significant improvements in menstrual irregularity from 56.2% to 7.7%, hirsutism from 67% to 32%, and infertility from 18.2% to 4.3% [72].

Metformin

In the 1990s, metformin was shown to ameliorate hyperandrogenism in both obese and non-obese women with PCOS [35]. Metformin lowers testosterone levels by approximately 20–25% in women with PCOS [35]. This effect may be more pronounced among non-obese women with PCOS. There is significant variability in the clinical response to metformin treatment in women with PCOS including in its ability to reduce testosterone and insulin levels, regulate menses and improve body weight and composition [73].

Current guidelines for management of PCOS recommend metformin treatment for women with IGT or type 2 diabetes who do not respond to lifestyle modification [45]. The impact of metformin on weight loss and body composition has been examined in a number of observational and randomized clinical trials but the results have been conflicting. A meta-analysis of 16 randomized clinical trials of 630 participants treated with metformin for 6 months at an average daily dose of 1500 mg reported no evidence of an effect on BMI and only a marginal reduction of waist:hip ratio [74]. However, there may be evidence for a beneficial effect of metformin on BMI and abdominal obesity when added to lifestyle modification. In a meta-analysis published in 2015, comparing nine randomized clinical trials (n = 493 participants) of metformin plus lifestyle vs. lifestyle alone, metformin plus lifestyle was more effective in reducing BMI [49]. Metformin may also have a favourable impact on body composition by reducing abdominal fat, although larger well-designed randomized studies are necessary to establish a clear benefit [75].

GLP-1 Agonists

Glucagon-like peptide-1 analogues have proven effective in reducing body weight among obese and overweight individuals without diabetes [76–78]. Small studies have also demonstrated effectiveness in this regard in women with PCOS [79]. Additional small studies have shown that the combination of these agents and metformin is more effective in inducing weight loss in women with PCOS [80–82]. In addition to lowering body weight, these agents on their own or in combination to metformin have been shown to reduce testosterone levels and may have additional metabolic benefits [82, 83]. Larger randomized clinical studies of longer duration are warranted to examine the potential benefits of these agents in women with PCOS.

Combination Oral Contraceptives

Combination oral contraceptives are commonly used as first-lined treatment for menstrual abnormalities, hirsutism, and acne in women with PCOS. Oral contraceptives suppress gonadotropin release with subsequent reduction in ovarian androgen secretion and stimulate SHBG production from the liver, thus reducing free testosterone level. Oral contraceptives are associated with a significant reduction in risk for endometrial cancer [84]. The choice for 'best' oral contraceptives for women with PCOS is unknown. A low dose formulation including ethinyl oestradiol and progestin is recommended. There are often adverse effects on insulin sensitivity that may be dose and drug dependent [85]. Oral contraceptives may also be associated with a significant elevation in circulating triglycerides, as well as in HDL levels, although these do not appear to progress over time. However, there is no evidence that risk for cardiovascular disease is increased with oral contraceptive use in PCOS although appropriate risk stratification is necessary in each individual case based on patient's personal cardiometabolic risk profile at baseline [86].

Conclusion

Insulin resistance, obesity, dyslipidaemia, and type 2 diabetes are common among women with PCOS. Management of these women should include an evaluation for metabolic risk at baseline and close follow-up especially for those who are identified at higher risk. Treatment tends to be symptom-based and should be individualized.

REFERENCES

1. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*, 2004; **19**: 41–7.
2. Burghen GA, Givens JR, Kitabchi AE. Correlation of hyperandrogenism with hyperinsulinism in polycystic ovarian disease. *J Clin Endocrinol Metab*, 1980; **50**: 113–16.
3. Dunaif A, Segal KR, Futterweit W, Dobrjansky A. Profound peripheral insulin resistance, independent of obesity, in polycystic ovary syndrome. *Diabetes*, 1989; **38**: 1165–74.
4. Chang RJ, Nakamura RM, Judd HL, Kaplan SA. Insulin resistance in nonobese patients with polycystic ovarian disease. *J Clin Endocrinol Metab*, 1983; **57**: 356–9.
5. Marsden PJ, Murdoch AP, Taylor R. Tissue insulin sensitivity and body weight in polycystic ovary syndrome. *Clin Endocrinol*, 2001; **55**: 191–9.
6. Barber TM, Dimitriadis GK, Andreou A, Franks S. Polycystic ovary syndrome: insight into pathogenesis and a common association with insulin resistance. *Clin Med (Lond)*, 2016; **16**: 262–6.
7. Dunaif A, Segal KR, Shelley DR, Green G, Dobrjansky A, Licholai T. Evidence for distinctive and intrinsic defects in insulin action in polycystic ovary syndrome. *Diabetes*, 1992; **41**: 1257–66.
8. Ciaraldi TP, Aroda V, Mudaliar S, Chang RJ, Henry RR. Polycystic ovary syndrome is associated with tissue-specific differences in insulin resistance. *J Clin Endocrinol Metab*, 2009; **94**: 157–63.
9. Ciaraldi TP, el-Roeiy A, Madar Z, Reichart D, Olefsky JM, Yen SS. Cellular mechanisms of insulin resistance in polycystic ovarian syndrome. *J Clin Endocrinol Metab*, 1992; **75**: 577–83.
10. Ciaraldi TP, Morales AJ, Hickman MG, Odom-Ford R, Olefsky JM, Yen SS. Cellular insulin resistance in adipocytes from obese polycystic ovary syndrome subjects involves adenosine modulation of insulin sensitivity. *J Clin Endocrinol Metab*, 1997; **82**: 1421–5.
11. Marsden PJ, Murdoch A, Taylor R. Severe impairment of insulin action in adipocytes from amenorrheic subjects with polycystic ovary syndrome. *Metabolism*, 1994; **43**: 1536–42.
12. Rosenbaum D, Haber RS, Dunaif A. Insulin resistance in polycystic ovary syndrome: decreased expression of GLUT-4 glucose transporters in adipocytes. *Am J Physiol*, 1993; **264**: E197–202.
13. Chen YH, Heneidi S, Lee JM, *et al*. miRNA-93 inhibits GLUT4 and is overexpressed in adipose tissue of polycystic ovary syndrome patients and women with insulin resistance. *Diabetes*, 2013; **62**: 2278–86.
14. Kirchengast S, Huber J. Body composition characteristics and body fat distribution in lean women with polycystic ovary syndrome. *Hum Reprod*, 2001; **16**: 1255–60.
15. Yildirim B, Sabir N, Kaleli B. Relation of intra-abdominal fat distribution to metabolic disorders in nonobese patients with polycystic ovary syndrome. *Fertil Steril*, 2003; **79**: 1358–64.
16. Carmina E, Bucchieri S, Esposito A, *et al*. Abdominal fat quantity and distribution in women with polycystic ovary syndrome and extent of its relation to insulin resistance. *J Clin Endocrinol Metab*, 2007; **92**: 2500–5.
17. Huang ZH, Manickam B, Ryvkin V, *et al*. PCOS is associated with increased CD11c expression and crown-like structures in adipose tissue and increased central abdominal fat depots independent of obesity. *J Clin Endocrinol Metab*, 2012; **98**: E17–24.
18. Barber TM, Golding SJ, Alvey C, *et al*. Global adiposity rather than abnormal regional fat distribution characterizes women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 999–1004.
19. Lovejoy JC, Bray GA, Bourgeois MO, *et al*. Exogenous androgens influence body composition and regional body fat distribution in obese postmenopausal women—a clinical research center study. *J Clin Endocrinol Metab*, 1996; **81**: 2198–203.
20. Elbers JM, Asscheman H, Seidell JC, Megens JA, Gooren LJ. Long-term testosterone administration increases visceral fat in female to male transsexuals. *J Clin Endocrinol Metab*, 1997; **82**: 2044–7.

21. Moghetti P, Tosi F, Castello R, *et al*. The insulin resistance in women with hyperandrogenism is partially reversed by antiandrogen treatment: evidence that androgens impair insulin action in women. *J Clin Endocrinol Metab*, 1996; **81**: 952–60.
22. Shoupe D, Lobo RA. The influence of androgens on insulin resistance. *Fertil Steril*, 1984; **41**: 385–8.
23. Elkind-Hirsch KE, Valdes CT, Malinak LR. Insulin resistance improves in hyperandrogenic women treated with Lupron. *Fertil Steril*, 1993; **60**: 634–41.
24. Dahlgren E, Landin K, Krotkiewski M, Holm G, Janson PO. Effects of two antiandrogen treatments on hirsutism and insulin sensitivity in women with polycystic ovary syndrome. *Hum Reprod*, 1998; **13**: 2706–11.
25. Zulian E, Sartorato P, Benedini S, *et al*. Spironolactone in the treatment of polycystic ovary syndrome: effects on clinical features, insulin sensitivity and lipid profile. *J Endocrinol Invest*, 2005; **28**: 49–53.
26. Geffner ME, Kaplan SA, Bersch N, Golde DW, Landaw EM, Chang RJ. Persistence of insulin resistance in polycystic ovarian disease after inhibition of ovarian steroid secretion. *Fertil Steril*,1986; **45**: 327–33.
27. Dunaif A, Green G, Futterweit W, Dobrjansky A. Suppression of hyperandrogenism does not improve peripheral or hepatic insulin resistance in the polycystic ovary syndrome. *J Clin Endocrinol Metab*, 1990; **70**: 699–704.
28. Dale PO, Tanbo T, Djoseland O, Jervell J, Abyholm T. Persistence of hyperinsulinemia in polycystic ovary syndrome after ovarian suppression by gonadotropin-releasing hormone agonist. *Acta Endocrinol (Copenh)*, 1992; **126**: 132–6.
29. Lasco A, Cucinotta D, Gigante A, *et al*. No changes of peripheral insulin resistance in polycystic ovary syndrome after long-term reduction of endogenous androgens with leuprolide. *Eur J Endocrinol*, 1995; **133**: 718–22.
30. Diamanti-Kandarakis E, Mitrakou A, Hennes MM, *et al*. Insulin sensitivity and antiandrogenic therapy in women with polycystic ovary syndrome. *Metabolism*, 1995; **44**: 525–31.
31. Nagamani M, Van Dinh T, Kelver ME. Hyperinsulinemia in hyperthecosis of the ovaries. *Am J Obstet Gynecol*, 1986; **154**: 384–9.
32. Aroda V, Ciaraldi TP, Chang SA, Dahan MH, Chang RJ, Henry RR. Circulating and cellular adiponectin in polycystic ovary syndrome: relationship to glucose tolerance and insulin action. *Fertil Steril*, 2008; **89**: 1200–8.
33. Chazenbalk G, Trivax BS, Yildiz BO, *et al*. Regulation of adiponectin secretion by adipocytes in the polycystic ovary syndrome: role of tumor necrosis factor-{alpha}. *J Clin Endocrinol Metab*, 2010; **95**: 935–42.
34. Dunaif A, Finegood DT. Beta-cell dysfunction independent of obesity and glucose intolerance in the polycystic ovary syndrome. *J Clin Endocrinol Metab*, 1996; **81**: 942–7.
35. Nestler JE, Jakubowicz DJ. Decreases in ovarian cytochrome P450c17 alpha activity and serum free testosterone after reduction of insulin secretion in polycystic ovary syndrome. *N Engl J Med*, 1996; **335**: 617–23.
36. Nestler JE, Powers LP, Matt DW, *et al*. A direct effect of hyperinsulinemia on serum sex hormone-binding globulin levels in obese women with the polycystic ovary syndrome. *J Clin Endocrinol Metab*, 1991; **72**: 83–9.
37. Legro RS, Kunselman AR, Dodson WC, Dunaif A. Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. *J Clin Endocrinol Metab*, 1999; **84**: 165–9.
38. Kakoly NS, Khomami MB, Joham AE, *et al*. Ethnicity, obesity and the prevalence of impaired glucose tolerance and type 2 diabetes in PCOS: a systematic review and meta-regression. *Hum Reprod Update*, 2018; **24**: 455–67.
39. Legro RS, Gnatuk CL, Kunselman AR, Dunaif A. Changes in glucose tolerance over time in women with polycystic ovary syndrome: a controlled study. *J Clin Endocrinol Metab*, **90**: 3236–42.
40. Goodman NF, Cobin RH, Futterweit W, *et al*. American Association of Clinical Endocrinologists, American College of Endocrinology, and Androgen Excess and PCOS Society Disease State Clinical Review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome—part 2. *Endocr Pract*, 2015; **21**: 1415–26.
41. Talbott E, Clerici A, Berga SL, *et al*. Adverse lipid and coronary heart disease risk profiles in young women with polycystic ovary syndrome: results of a case-control study. *J Clin Epidemiol*, 1998; **51**: 415–22.
42. Legro RS, Kunselman AR, Dunaif A. Prevalence and predictors of dyslipidemia in women with polycystic ovary syndrome. *Am J Med*, 2001; **111**: 607–13.
43. Wild RA, Carmina E, Diamanti-Kandarakis E, *et al*. Assessment of cardiovascular risk and prevention of cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab*, 2010; **95**: 2038–49.
44. Sidhwani S, Scoccia B, Sunghay S, Stephens-Archer CN, Mazzone T, Sam S. PCOS is associated with atherogenic changes in lipoprotein particle number and size independent of body weight. *Clin Endocrinol*, 2011; **75**: 76–82.
45. Legro RS, Arslanian SA, Ehrmann DA, *et al*. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2013; **98**: 4565–92.
46. Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III) final report. *Circulation*, 2002; **106**: 3143–421.
47. Moran LJ, Misso ML, Wild RA, Norman RJ. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update*. 2010; **16**: 347–63.
48. Rocha ALL, Faria LC, Guimaraes TCM, *et al*. Non-alcoholic fatty liver disease in women with polycystic ovary syndrome: systematic review and meta-analysis. *J Endocrinol Invest*, 2017; **40**: 1279–88.
49. Kumarendran B, O'Reilly MW, Manolopoulos KN, *et al*. Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: a longitudinal study based on a United Kingdom primary care database. *PLoS Med*, 2018; **15**: e1002542.
50. Talbott EO, Zborowski JV, Boudreaux MY, McHugh-Pemu KP, Sutton-Tyrrell K, Guzick DS. The relationship between C-reactive protein and carotid intima-media wall thickness in middle-aged women with polycystic ovary syndrome. *J Clin Endocrinol Metab*,.2004; **89**: 6061–7.
51. Carmina E, Orio F, Palomba S, *et al*. Endothelial dysfunction in PCOS: role of obesity and adipose hormones. *Am J Med*, 2006; **119**: 356.e1–6.

52. Talbott EO, Zborowski JV, Rager JR, Boudreaux MY, Edmundowicz DA, Guzick DS. Evidence for an association between metabolic cardiovascular syndrome and coronary and aortic calcification among women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 5454–61.
53. Christian RC, Dumesic DA, Behrenbeck T, Oberg AL, Sheedy PF, 2nd, Fitzpatrick LA. Prevalence and predictors of coronary artery calcification in women with polycystic ovary syndrome. *T J Clin Endocrinol Metab*, 2003; **88**: 2562–8.
54. Pierpoint T, McKeigue PM, Isaacs AJ, Wild SH, Jacobs HS. Mortality of women with polycystic ovary syndrome at long-term follow-up. *J Clin Epidemiol*, 1998; **51**: 581–6.
55. Wild S, Pierpoint T, Jacobs H, McKeigue P. Long-term consequences of polycystic ovary syndrome: results of a 31 year follow-up study. *Hum Fertil (Camb)*, 2000; **3**: 101–5.
56. Shaw LJ, Bairey Merz CN, Azziz R, *et al*. Postmenopausal women with a history of irregular menses and elevated androgen measurements at high risk for worsening cardiovascular event-free survival: results from the National Institutes of Health—National Heart, Lung, and Blood Institute sponsored Women's Ischemia Syndrome Evaluation. *J Clin Endocrinol Metab*, 2008; **93**: 1276–84.
57. Ehrmann DA. Metabolic dysfunction in PCOS: relationship to obstructive sleep apnea. *Steroids*, 2012; **77**: 290–4.
58. Helvaci N, Karabulut E, Demir AU, Yildiz BO. Polycystic ovary syndrome and the risk of obstructive sleep apnea: a meta-analysis and review of the literature. *Endocr Connect*, 2017; **6**: 437–45.
59. Fogel RB, Malhotra A, Pillar G, Pittman SD, Dunaif A, White DP. Increased prevalence of obstructive sleep apnea syndrome in obese women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2001; **86**: 1175–80.
60. Vgontzas AN, Legro RS, Bixler EO, Grayev A, Kales A, Chrousos GP. Polycystic ovary syndrome is associated with obstructive sleep apnea and daytime sleepiness: role of insulin resistance. *J Clin Endocrinol Metab*, 2001; **86**: 517–20.
61. Gopal M, Duntley S, Uhles M, Attarian H. The role of obesity in the increased prevalence of obstructive sleep apnea syndrome in patients with polycystic ovarian syndrome. *Sleep Med*, 2002; **3**: 401–4.
62. Tasali E, Van Cauter E, Hoffman L, Ehrmann DA. Impact of obstructive sleep apnea on insulin resistance and glucose tolerance in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 3878–84.
63. Chittenden BG, Fullerton G, Maheshwari A, Bhattacharya S. Polycystic ovary syndrome and the risk of gynaecological cancer: a systematic review. *Reprod Biomed Online*, 2009; **19**: 398–405.
64. Haoula Z, Salman M, Atiomo W. Evaluating the association between endometrial cancer and polycystic ovary syndrome. *Hum Reprod.* 2012; **27**: 1327–31.
65. Salley KE, Wickham EP, Cheang KI, Essah PA, Karjane NW, Nestler JE. Glucose intolerance in polycystic ovary syndrome—a position statement of the Androgen Excess Society. *J Clin Endocrinol Metab*, 2007; **92**: 4546–56.
66. Velling Magnussen L, Mumm H, Andersen M, Glintborg D. Hemoglobin A1c as a tool for the diagnosis of type 2 diabetes in 208 premenopausal women with polycystic ovary syndrome. *Fertil Steril*, 2011; **96**: 1275–80.
67. Norman RJ, Masters L, Milner CR, Wang JX, Davies MJ. Relative risk of conversion from normoglycaemia to impaired glucose tolerance or non-insulin dependent diabetes mellitus in polycystic ovarian syndrome. *Hum Reprod*, 2001; **16**: 1995–8.
68. Moran LJ, Noakes M, Clifton PM, Tomlinson L, Galletly C, Norman RJ. Dietary composition in restoring reproductive and metabolic physiology in overweight women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2003; **88**: 812–19.
69. Moran LJ, Ko H, Misso M, Marsh K, *et al*. Dietary composition in the treatment of polycystic ovary syndrome: a systematic review to inform evidence-based guidelines. *J Acad Nutr Diet*, 2013; **113**: 520–45.
70. Harrison CL, Lombard CB, Moran LJ, Teede HJ. Exercise therapy in polycystic ovary syndrome: a systematic review. *Hum Reprod Update*,2011; **17**: 171–83.
71. Jensen MD, Ryan DH, Apovian CM, *et al*. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll of Cardiol*, 2014; **63**(25 Pt B): 2985–3023.
72. Skubleny D, Switzer NJ, Gill RS, *et al*. The impact of bariatric surgery on polycystic ovary syndrome: a systematic review and meta-analysis. *Obes Surg*, 2016; **26**: 169–76.
73. Sam S, Ehrmann DA. Metformin therapy for the reproductive and metabolic consequences of polycystic ovary syndrome. *Diabetologia*, 2017; **60**: 1656–61.
74. Tang T, Lord JM, Norman RJ, Yasmin E, Balen AH. Insulin-sensitising drugs (metformin, rosiglitazone, pioglitazone, D-chiro-inositol) for women with polycystic ovary syndrome, oligo amenorrhoea and subfertility. *Cochrane Database Syst Rev*, 2012: **5**: CD003053.
75. Pasquali R, Gambineri A, Biscotti D, *et al*. Effect of long-term treatment with metformin added to hypocaloric diet on body composition, fat distribution, and androgen and insulin levels in abdominally obese women with and without the polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2000; **85**: 2767–74.
76. Su N, Li Y, Xu T, *et al*. Exenatide in obese or overweight patients without diabetes: a systematic review and meta-analyses of randomized controlled trials. *Int J Cardiol*, 2016; **219**: 293–300.
77. Zhang F, Tong Y, Su N, *et al*. Weight loss effect of glucagon-like peptide-1 mimetics on obese/overweight adults without diabetes: a systematic review and meta-analysis of randomized controlled trials. *J Diabetes*, 2015; **7**: 329–39.
78. Mehta A, Marso SP, Neeland IJ. Liraglutide for weight management: a critical review of the evidence. *Obes Sci Pract*, 2017; **3**: 3–14.
79. Niafar M, Pourafkari L, Porhomayon J, Nader N. A systematic review of GLP-1 agonists on the metabolic syndrome in women with polycystic ovaries. *Arch Gynecol Obstet*, 2016; **293**: 509–15.
80. Jensterle Sever M, Kocjan T, Pfeifer M, Kravos NA, Janez A. Short-term combined treatment with liraglutide and metformin leads to significant weight loss in obese women with polycystic ovary syndrome and previous poor response to metformin. *Eur J Endocrinol*, 2014; **170**: 451–9.
81. Jensterle M, Kravos NA, Goricar K, Janez A. Short-term effectiveness of low dose liraglutide in combination with metformin versus high dose liraglutide alone in treatment of obese PCOS: randomized trial. *BMC Endocr Disord*, 2017; **17**: 5.
82. Elkind-Hirsch K, Marrioneaux O, Bhushan M, Vernor D, Bhushan R. Comparison of single and combined treatment with exenatide and metformin on menstrual cyclicity in overweight women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2008; **93**: 2670–8.

83. Frossing S, Nylander M, Chabanova E, *et al.* Effect of liraglutide on ectopic fat in polycystic ovary syndrome: a randomized clinical trial. *Diabetes Obes Metab*, 2018; **20**: 215–18.
84. Dumesic DA, Lobo RA. Cancer risk and PCOS. *Steroids*, 2013; **78**: 782–5.
85. Meyer C, McGrath BP, Teede HJ. Effects of medical therapy on insulin resistance and the cardiovascular system in polycystic ovary syndrome. *Diabetes Care*, 2007; **30**: 471–8.
86. Yildiz BO. Approach to the patient: contraception in women with polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2015; **100**: 794–802.

8.4.4 Polycystic Ovary Syndrome

Hirsutism

Duarte Pignatelli, Ricardo Azziz, and Bulent Okan Yildiz

Physiology and Pathophysiology of Hair Growth

Androgen Metabolism in Women

Androgens are 19 carbon (C19) steroids, synthesized from the steroid substrate pregnenolone, which is itself derived from cholesterol. Androgens are produced by both the ovary and the adrenal cortex. They may also be derived from the conversion of precursor steroids by the liver and some peripheral tissues including the skin and adipose tissue. The main circulating androgens in women include testosterone, and its 5α -reduced metabolite dihydrotestosterone (DHT), androstenedione (A4), dehydroepiandrosterone (DHEA) and its metabolite dehydroepiandrosterone sulphate (DHEAS). Recently the importance of 11-oxygenated androgens such as 11-ketotestosterone (11-KT) and 11-ketoDHT (11-KDHT) has been recognized as these are also potent agonists of the human androgen receptor [1–3]. Further study is required, however, to determine the clinical utility of 11-KT and 11-KDHT in the evaluation of androgen excess disorders in women. See Figure 8.4.4.1.

Only free androgens are able to interact with androgen receptors on target tissues. Thus, the biological action of testosterone and DHT is significantly influenced by the circulating SHBG level. Even without a change in total hormone concentrations, a decrease in SHBG will result in increased free fractions of testosterone and DHT, increasing androgenic action. Conversely, higher SHBG levels will result in decreased free fractions, and a decreased androgenic action. SHBG levels are inversely correlated with androgen and insulin levels, whereas oestrogens increase circulating SHBG concentrations. SHBG also influences the clearance of androgens from the circulation.

Androgen production and clearance are influenced by various physiological states. Obesity, particularly the abdominal type, increases the formation of testosterone from A4, and decrease SHBG levels resulting in increased circulating free androgens. Additionally, because androgens are fat soluble, excess adipose tissue serves as an extravascular pool for androgens. The number of androgens metabolized to oestrogens by adipose tissue aromatase are also enhanced in obesity.

Normative ranges for androgens also differ depending on age. Circulating levels of A4, total and free testosterone, and, DHEA/DHEAS, decline with age, beginning in the mid/late 30s. It is well established that normal menopausal androgen levels are lower than those produced in the reproductive years.

Normal Hair Physiology

Hair covers the vast majority of the body, sparing only the lips, palms of the hands, and the soles of the feet. There are about 5 million hair follicles on a human, of which 1 million on the head. Almost all hair follicles are present at birth and no additional follicles arise thereafter, although the size of the follicles may change over time.

A hair follicle is present in conjunction with a sebaceous gland and an erector pili muscle, forming the pilosebaceous unit (PSU) (Figure 8.4.4.2). The physiology of hair growth is covered elsewhere [4].

Structurally, there are three types of hair. Lanugo is soft hair covering the surface of the fetus, which is shed sometime in late gestation or early postpartum. Vellus hair is soft, fine, non-pigmented, or containing little pigment, generally measuring less than 2 mm in length, and covering apparently hairless areas of the body. Terminal hair is long, coarse, thick, pigmented, and contains a central core of compacted keratinocytes (i.e. medullated), which vellus hairs do not have. Terminal hairs are found primarily on the upper and lower limbs, the scalp, the face, the upper and lower back, the chest, abdomen, and in the axillary and pubic areas. This type of hair shows significant regional morphological differences (i.e. longer in some sites, more pigmented in others, etc.) due to genetically determined differences in the follicles. Nonsexual terminal hair presents in the scalp, eyebrows, and eyelashes [5].

Race and ethnicity influence the body hair type and distribution. The number of hair follicles per unit skin area and the rate of hair growth vary among ethnic groups. For example, Asians have less dense hair than Blacks, who in turn have less dense hair than Whites. However, men and women within the same race or ethnic group have similar follicle numbers and the visible differences between them are related to the type of hair arising from these follicles (i.e. terminal versus vellus hairs).

Hair follicles undergo cyclic changes and there are three phases of the hair follicle growth cycle (Figure 8.4.4.3). Anagen is the active growing phase of hair. During this phase, keratinocytes are dividing extensively with downwards progression of the dermal papilla. Anagen is followed by the transitional catagen phase in which the hair stops growing and the hair bud shrinks forming a club end,

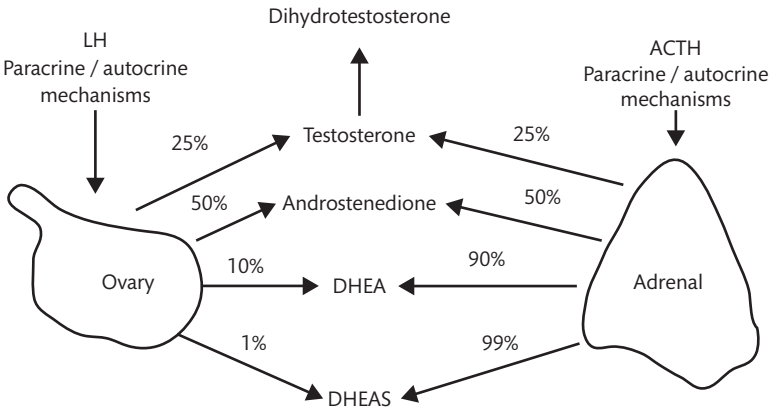


Figure 8.4.4.1 The regulation of androgen synthesis and secretion.
ACTH, adrenocorticotropic hormone; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulphate; LH, luteinizing hormone.

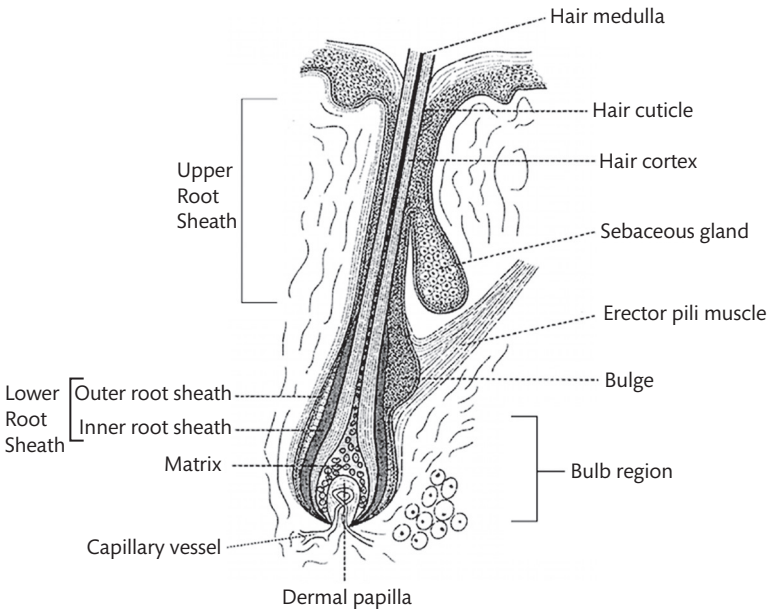


Figure 8.4.4.2 Anatomy of a pilosebaceous unit.
Reproduced with permission from Sanchez LA, Perez M, Azziz R. Laser hair reduction in the hirsute patient: a critical assessment. *Human reproduction update*. 2002;8(2):169–81. Copyright © 2002, Oxford University Press. (ref 5).

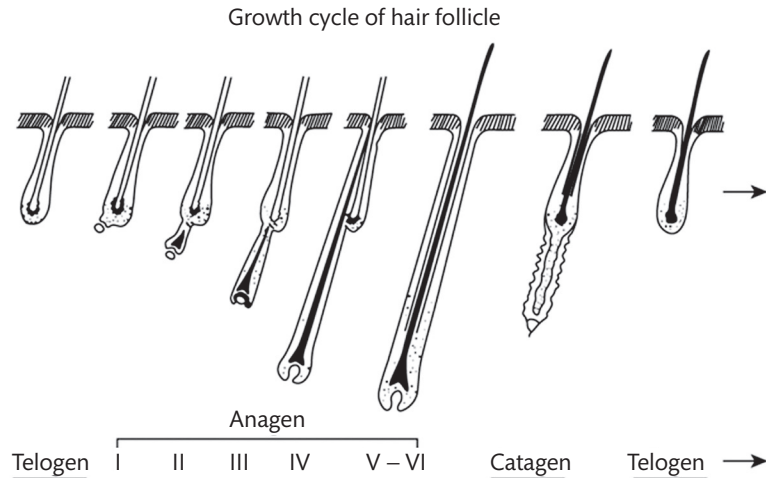


Figure 8.4.4.3 Growth cycle of a hair follicle.
Reproduced with permission from Uno H. Biology of hair growth. *Semin Reprod Endocrinol*. 1986; 4: 131–41. Copyright © 1986 by Thieme Medical Publishers, Inc.

and finally by a resting, or telogen phase, after which the hair sheds [5]. Although in many animals the growth cycles of all hair follicles are in synchrony, in humans, the growth phases of different hair follicles are not synchronous and, for that reason, hairs appear to be continuously growing. The length of hair cycle phases varies significantly in different parts of the human body. Scalp follicles have the longest anagen phase, which may last 2–6 years. They have a catagen phase of 1–3 weeks and a telogen phase of up to 3 months. Normally, 80–85% of scalp hairs are in anagen. The anagen phase of body hairs may as short as 3–6 months (e.g. forearms or legs) or as long as 2–3 years (e.g. on scalp).

Development and growth of hair follicles are regulated by hormonal factors. Growth and thyroid hormones stimulate a generalized increase in hair growth. Both hypo- and hyperthyroidism are, however, associated with hair loss. Pregnancy temporarily increases the number of hair follicles in anagen, of which many enter catagen or telogen postpartum resulting in diffuse hair loss. Oestrogens oppose the effects of androgens, by increasing SHBG levels and reducing free androgens rather than showing a direct effect on hair follicles [5].

Effects of Androgens on Hair Follicle

Androgens are the principal hormonal regulators of type and distribution of body and facial hairs. In hair follicles, circulating testosterone is metabolized by 5α-reductase to the more potent DHT. Both hormones (and to a limited extent, A4, and DHEA) bind to the androgen receptor. Androgen actions on the PSU, include: (1) increased sebum production; (2) the differentiation of the hair follicle from vellus to terminal hairs (in sensitive areas of the body), or vice-versa (in scalp); and (3) the prolongation of the anagen phase resulting in longer thicker hairs. See Figure 8.4.4.4.

The growth and differentiation of hair follicles in response to androgens varies greatly in different body areas, in association with the local content of AR, 5α-reductase, l-ornithine decarboxylase (ODC), 17β-hydroxysteroid dehydrogenase, and other enzymes.

Three types of skin areas concerning sensitivity to androgens are generally considered (Table 8.4.4.1), although we should note that there is wide variability in skin responsiveness across the body and face.

Some skin areas are relatively independent of the effect of androgens and are defined as *non-sexual skin areas*. Alternatively, other skin areas (e.g. lower pubic triangle and the axilla) are quite sensitive to androgens, and hair follicles are readily terminalized even in the presence of relatively low levels of circulating androgens. These skin areas begin to develop terminal hair in early puberty, when only minimal increases in adrenal androgens are observed, and are termed as *ambosexual skin areas*. Finally, other areas of the skin respond to androgens, but only to significantly higher levels. These areas are defined as *sexual skin areas* [6, 7]. In women, the presence of increased terminal hairs in sexual skin areas, in a male-like pattern, is considered pathological and is termed hirsutism.

Androgens, particularly in excess, may transform vellus into terminal hairs in androgen-sensitive areas in an irreversible manner (i.e. terminalization). Paradoxically, in certain areas androgens may transform terminal hairs into vellus hairs (i.e. miniaturization), as is observed in male-pattern balding. The process of transformation (i.e. terminalization or miniaturization), occurs progressively over many hair growth cycles, requiring months to years of androgen exposure. Interruption of the process sufficiently early (e.g. through use of antiandrogens in case of vellus hair terminalization) can reverse the observed effects.

Epidemiology of Hirsutism

The prevalence of hirsutism will depend on the method used to determine its presence, the definition used for hirsutism, and the population under investigation. Although objective methods are available for the assessment of hair growth including photographic evaluations and microscopic measurements, they are not suitable for widespread clinical use due to a significant degree of complexity and high cost. Alternatively, various methods, based on visual assessment of hair type and growth, have been proposed to evaluate patients suspected of having hirsutism [8].

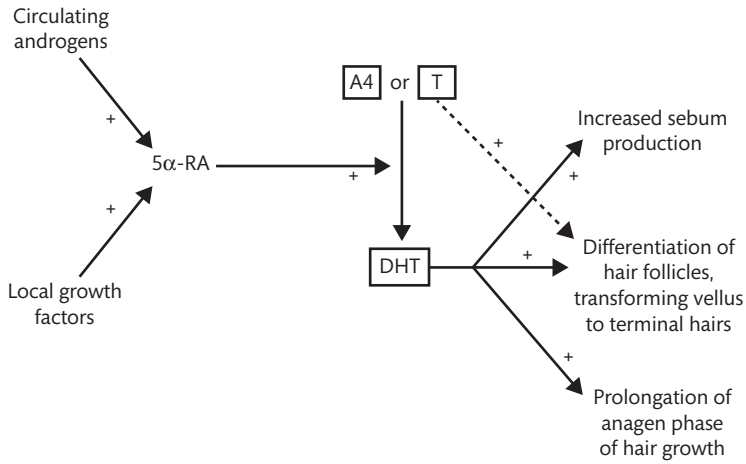


Figure 8.4.4.4 Effects of androgens at the pilosebaceous unit.

A4, Androstenedione; T, testosterone; DHT, Dihydrotestosterone. 5A-RA, 5a-reductase.

Table 8.4.4.1 Hair type and localization in relation to sensitivity to androgens

Hair type	Skin area	Androgen sensitivity
Nonsexual hair	Eyelashes, eyebrows, and lateral and occipital aspects of the scalp	Relatively independent of the effects of androgens
Ambosexual hair	Lower pubic triangle, the axilla, and the forearm and lower leg	Sensitive to low levels of androgens
Sexual hair	Upper lip, chin, and neck, cheek and sideburn areas, chest, upper and lower abdomen, upper and lower back, thighs, and upper arms	Sensitive to high levels of androgens

The most commonly used method of scoring body and facial terminal hair growth, for defining the presence of hirsutism, is based on a modification of the method originally described by Ferriman and Gallwey in 1961 [9, 10]. The modified Ferriman–Gallwey (mFG) visual assessment method scores the presence of hair growth between 0 (absence of terminal hairs) and 4 (extensive terminal hair growth) at nine different body sites (Figure 8.4.4.5) [8, 10]. However, we should note that this system is semi-quantitative at best, and subject to interobserver variability and there is also a lack of consensus on what score defines hirsutism [10]. See Figure 8.4.4.6.

To determine the prevalence of hirsutism in the general population some large studies have been done [11–14]. Overall, 7.5% of the population could be defined as being hirsute by an mFG scores ≥8. However, when using cluster analysis to define what the 'natural' cut-off should be, the investigators observed that a score of 3 or more, defined a population that complained of being hirsute more frequently (70% vs. 15% below this score) and who used some form of treatment for unwanted hair [11]. Consequently, it is likely that the mFG score defining 'hirsutism' should be lower than what is generally used clinically today. More studies on defining 'hirsutism' are needed.

The degree of body hair growth, may be affected by ethnicity and race. While the degree of facial and body terminal hair growth appears to be similar in Black and White women, at least in the United States [11], it is highly unlikely that Asian women have similar degrees of hair growth. Furthermore, while there may be significant differences in the degree of hirsutism between ethnic groups, it is less clear that there are significant differences in the cut-off values used to define hirsutism between populations.

Other Signs of Androgen Excess

Acne

Acne is a common disorder of the PSU. It occurs in adolescence, and may persist into adulthood. Acne presents most commonly on the face, neck, chest, shoulders, and back. Androgens stimulate sebocyte proliferation, cause the sebaceous glands to enlarge and produce more sebum. In women whose acne is severe, or associated with hirsutism or irregular menstrual periods, hyperandrogenism should be considered [15].

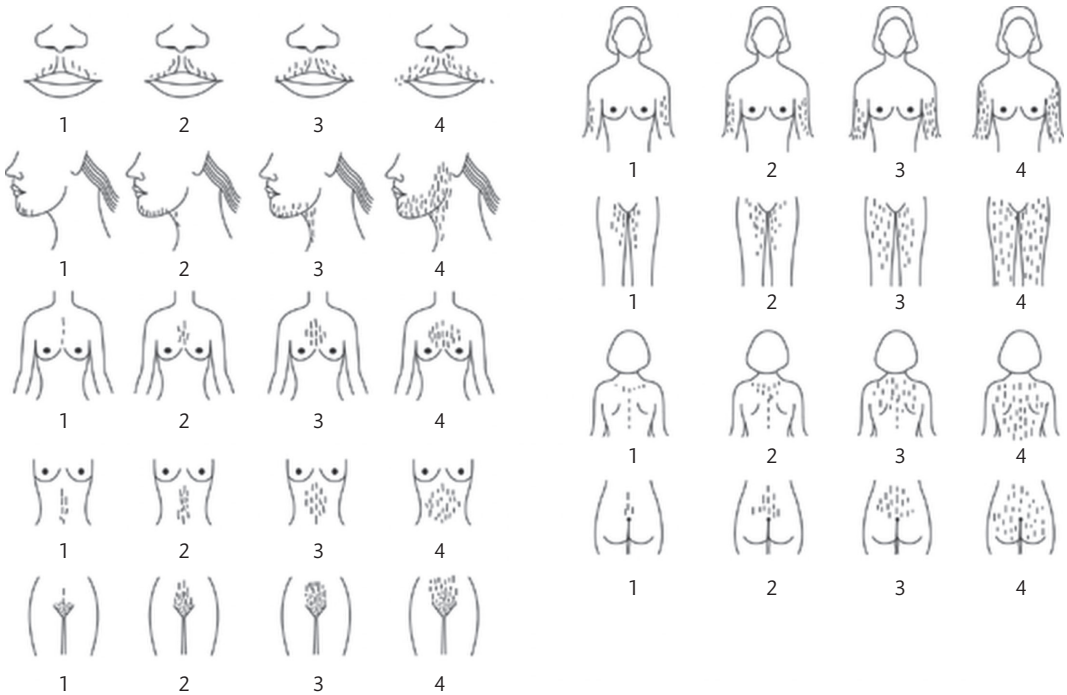


Figure 8.4.4.5 Modified Ferriman–Gallwey (mFG) hirsutism scoring system. Each of the nine body areas is rated from 0 (absence of terminal hairs) to 4 (extensive terminal hair growth), and the numbers in each area are added for a total score. A mFG score ≥6–8 generally defines hirsutism.

Adapted with permission from Hatch R, Rosenfield RL, Kim MH, Tredway D. Hirsutism: implications, etiology, and management. *American Journal of Obstetrics and Gynecology*. 1981;140(7):815–30. Copyright 1981 The C. V. Mosby Co. (ref 10).

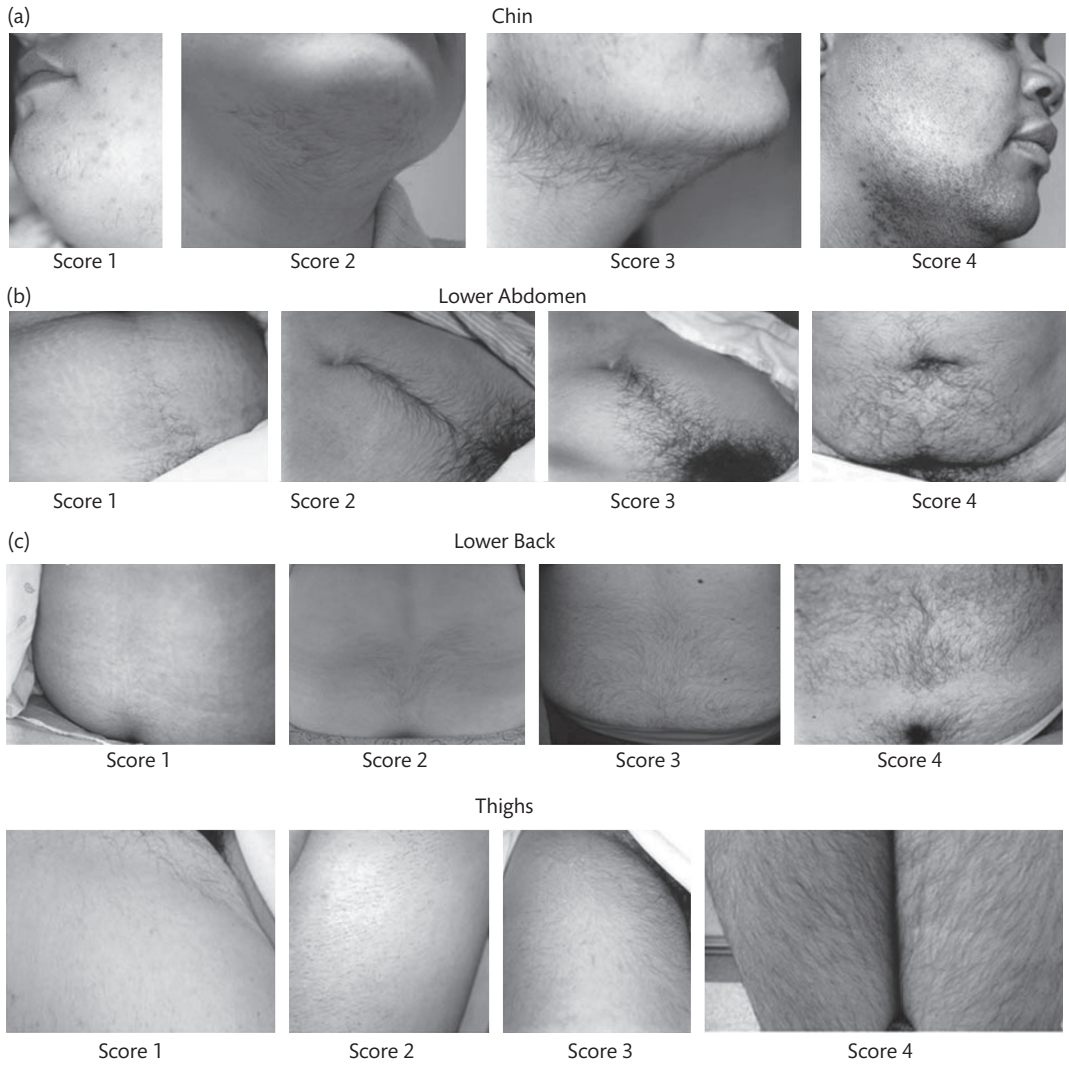


Figure 8.4.4.6 Photographic examples of facial and body terminal hair growth scored according to the modified FG method.

Adapted with permission from Yildiz BO, Bolour S, Woods K, Moore A, Azziz R. Visually scoring hirsutism. *Human reproduction update*. 2010;16(1):51–64. Copyright © 2009, Oxford University Press. (ref 8).

Androgenic Alopecia

The term 'alopecia' refers to loss of scalp hair. Androgenic alopecia is sometimes referred to as 'androgenetic' alopecia on the presumption of an underlying, yet to be determined genetic factor. In the presence of androgens, anagen phase of sensitive scalp hairs is shortened, and hair follicles shrink or become miniaturized. With successive anagen cycles, the follicles become smaller and shorter, and non-pigmented vellus hairs replace the thick pigmented terminal hairs.

Androgenic alopecia is a form of alopecia in women (female-pattern hair loss or FPHL), generally characterized by diffuse loss of hair in the sagittal (top) part of the scalp [16]. In severe forms of hyperandrogenism, hair loss in women may be similar to that of men ('male-pattern balding'), with frontal, sagittal and eventually occipital hair loss.

While most of the women with FPHL have normal endocrine function, and regular ovulatory cycles, it is not uncommon for androgenic alopecia to be accompanied by other androgenic skin manifestations, such as hirsutism and acne in the same patient. The prevalence of hyperandrogenaemia is higher when FPHL presents in association with other signs of androgen excess, such as hirsutism and menstrual dysfunction [17].

Virilization

Virilization is a relatively uncommon clinical finding, and its presence is usually associated with markedly elevated levels of circulating androgens and significant pathology. Depending on the duration of the underlying pathology, virilization is characterized by male-pattern balding, clitoromegaly, and severe hirsutism. It may also be accompanied by signs of masculinization including deepening of the voice, increased muscle mass, and loss of typical female fat deposition, including decreased breast size. Women with virilization are nearly always amenorrhoeic [18]. The presence of an androgen-secreting neoplasm (ovarian or adrenal) should always be suspected in any woman who develops signs of virilization, particularly if the onset is sudden with a rapid progression, although there may be other rare causes.

Differential Diagnosis of Hirsutism

The causes of hirsutism are summarized in Box 8.4.4.1. Over 80% of hirsute patients will have PCOS, with a small percentage of patients having idiopathic hirsutism, and other rare disorders including non-classic congenital adrenal hyperplasia (NCAH), the hyperandrogenism-insulin resistance-acanthosis nigricans (HAIRAN) syndrome, and androgen-secreting neoplasms [14, 19].

Cushing's syndrome, acromegaly, thyroid dysfunction, and hyperprolactinaemia may be associated with hirsutism, but patients usually present with other clinical features common to these disorders.

Polycystic Ovary Syndrome

PCOS is a common disorder characterized by androgen excess, ovulatory dysfunction, and polycystic ovaries [19, 20]. PCOS affects 5–15% of women of reproductive age and over 80% of hirsute women. There are currently three sets of criteria for diagnosing PCOS [21–24], although the 2003 Rotterdam classification system is the broadest and most widely used. This defines PCOS by the presence of as least two of the following three criteria: (1) oligo- and/or anovulation, (2) clinical/biochemical hyperandrogenism, (3) polycystic ovaries on ultrasonography, and exclusion of other related disorders [22, 23].

The aetiology(s) and genetic basis of the syndrome remain largely unknown. Patients with PCOS have several interrelated characteristics including dysregulated ovarian and adrenal steroidogenesis, altered gonadotropin dynamics, chronic anovulation, polycystic ovaries, and insulin resistance [20]. It is noteworthy that insulin resistance and hyperinsulinaemia are dominant features of PCOS both in obese and lean patients, and over 80% of patients with PCOS demonstrate varying degrees of insulin resistance [25]. Insulin directly enhances LH-stimulated androgen secretion from the ovarian theca cells. Increased insulin levels also decrease the synthesis of SHBG by the liver and reduce circulating SHBG levels, thus resulting in higher concentrations of free androgens.

Box 8.4.4.1 Differential diagnosis of hirsutism

- Functional androgen excess disorders
  - Polycystic ovary syndrome (PCOS)
  - Idiopathic hirsutism
- Specific identifiable disorders
  - Non-classic congenital adrenal hyperplasia (NCAH)
  - Hyperandrogenism-insulin resistance-acanthosis nigricans (HAIRAN) syndrome[b]
  - Androgen-secreting neoplasms[b]
  - Cushing's syndrome[b]
  - Acromegaly[b]
  - Thyroid dysfunction
  - Hyperprolactinemia
- Other causes
  - Drugs
  - Chronic skin irritation

[a] Hirsutism in a patient with normal ovarian function (normo-ovulation), no polycystic ovaries, and normal circulating androgen levels.
[b] If clinical findings are highly suggestive of these very rare disorders, further testing may be needed.

Additionally, PCOS is associated with increased risk of type 2 diabetes, dyslipidaemia, cardiovascular disease (CVD), and endometrial carcinoma [19].

Current treatment regimens are directed at reducing hirsutism, menstrual cycle regulation, and at times achieving pregnancy. In addition, improvement of insulin sensitivity, weight control, and prevention of long-term health consequences are also important therapeutic goals.

Idiopathic Hirsutism

Hirsute patients with normal ovarian function and morphology (i.e. regular ovulation, and no polycystic ovaries on ultrasound), along with normal androgen levels, in the absence of features that suggest other specific identifiable causes of hirsutism, are diagnosed as having idiopathic hirsutism (IH).

It must be said that routine androgen assays may not be suitable to detect mild to moderate hyperandrogenaemia in women (see next). Furthermore, available evidence suggest that up to 40% of hirsute women who claim to have 'regular menses' actually demonstrate oligo-ovulation and so, a percentage of these women should be diagnosed as having PCOS [26].

In some patients with idiopathic hirsutism the activity of 5α-reductase in the hair follicle, which converts testosterone to the more potent androgen DHT, appears to be increased. Increased numbers of ARs or the presence of polymorphisms in the AR gene that result in an increased receptor activity may also be the basis for the disorder in other patients.

Prevalences of idiopathic hirsutism of up to 17% of hirsute women have been reported [6, 27]. However, it is important to note that with the use of the more expansive Rotterdam criteria for PCOS, the prevalence of IH is nowadays significantly lower.

Overall, the aetiology and epidemiology of idiopathic hirsutism remains unclear, although the prevalence is decreasing as the definition of PCOS broadens and the quality of the clinical and laboratory evaluation improves.

Non-Classic Congenital Adrenal Hyperplasia (NCAH)

Between 1% and 10% of patients with hirsutism have NCAH. The most common form is adrenocortical 21-hydroxylase (21-OH) deficiency. In this autosomal recessive disorder, the steroid precursors for 21-OH, particularly 17α-hydroxyprogesterone (17-HP) and A4, accumulate and are secreted in excess. Hyperandrogenic symptoms most commonly appear in the peri- or postpubertal period, but may be absent (i.e. cryptic) even in the presence of elevated circulating androgen levels. In addition to hirsutism, other features may include oligo-ovulation, and polycystic ovaries. Some children may present with premature pubarche. Clinically, it is difficult to distinguish these patients from patients with PCOS [28]. Biochemically, the levels 17-OHP are elevated while the adrenal androgen DHEAS is not higher than that of other hyperandrogenic women [29].

Hyperandrogenic-Insulin Resistant-Acanthosis Nigricans (HAIRAN) Syndrome

Patients with the HAIRAN syndrome present with marked acanthosis nigricans (a velvety thickening and hyperpigmentation of the skin found on intertriginous areas, reflecting hyperinsulinemia) and significant degrees of hyperandrogenism [30]. Normal or low

luteinizing hormone levels accompany increased androgen levels. The clinical distinction between HAIRAN and PCOS is not always clear. Nevertheless, HAIRAN syndrome can be diagnosed in patients with hyperandrogenism and acanthosis, by the presence of severe insulin resistance and hyperinsulinemia, generally determined as fasting basal circulating insulin levels higher than 80 μU/ml and/or 500 μU/ml during an oral glucose challenge. The ovaries are enlarged and hyperthecotic with proliferating islands of luteinized theca cells in the ovarian stroma, and tend to be less cystic than the typical polycystic ovary.

Androgen-Secreting Tumours

Androgen-secreting tumours, whether ovarian or adrenal, are relatively rare. The clinical onset of these tumours is usually sudden and may rapidly lead to virilization and masculinization. Other systemic symptoms, such as weight loss and anorexia might also be observed in patients with carcinomas. Ovarian neoplasms are usually not malignant, and include Sertoli–Leydig cell tumours and lipoid cell tumours. They are usually palpable on pelvic exam and/or are associated with unilateral ovarian enlargement on imaging. Androgen-secreting tumours of the adrenal are often carcinomas. Androgen-secreting adrenocortical carcinomas are frequently associated with Cushingoid features.

The clinical presentation is generally the most sensitive indicator of an androgen-producing tumour. Biochemical suppression or stimulation tests are not recommended for the diagnosis of androgen-secreting tumours as these tests can be misleading. Likewise, basal levels of androgens vary widely in patients with these tumours. In patients suspected of suffering from an adrenal or ovarian androgen-producing tumour, imaging studies and ovarian and/or adrenal venous sampling may be of value in identifying the neoplasm [18].

Drugs

A number of non-androgenic drugs, such as phenytoin, cyclosporine, and diazoxide might result in generalized growth of body and facial hair, leading to vellus hypertrichosis. Alternatively, the use or abuse of androgenic/anabolic drugs, such as danazol and methyltestosterone, may produce hirsutism in addition to amenorrhoea and liver dysfunction.

Clinical Evaluation of Hirsutism

A thorough history and a focused physical examination are essential for the evaluation of the patient with hirsutism. It is firstly important to clinically distinguish between hirsutism and hypertrichosis in order to determine the necessary subsequent evaluation and management. Hypertrichosis is characterized by increased vellus hair growth and is not caused by androgen excess [31]. Nevertheless, some of the patients with hyperandrogenism will have excess growth of both terminal and vellus hairs.

History

In a patient with suspected hyperandrogenism, the use of androgenic/anabolic drugs or skin irritants use should be excluded. Onset and progression of hirsutism or signs of virilization should be determined. Peripubertal onset of hirsutism with slow progression is more consistent with non-neoplastic disorders, such as PCOS. Alternatively, rapid progression of hair growth with signs of virilization in a previously asymptomatic woman should raise suspicion of an androgen-secreting neoplasm.

A detailed history of the onset (menarche) and subsequent development of the patient's menstrual pattern should be obtained. Menstrual irregularities may accompany hirsutism, although some hirsute women can have regular cycles (eumenorrhea) or even normal ovulatory function. A history of galactorrhoea or symptoms of thyroid dysfunction should also be sought. Finally, a detailed family history for endocrine, metabolic, cardiovascular, neoplastic, and reproductive disorders should be obtained.

Physical Examination

Firstly, the physical exam should focus on determining whether features of hyperandrogenism are truly present. Hyperandrogenism in women may present as hirsutism, acne, androgenic alopecia, or virilization. The type, pattern, and extent of excessive hair growth should be established and preferably scored by using the mFG method.

Finally, in addition to evidence of hyperandrogenism, signs of other endocrine disorders should be sought for, including acanthosis nigricans, the presence and body distribution of obesity, Cushingoid features, and facial features suggestive of acromegaly.

Laboratory Evaluation

In patients with hirsutism, measurement of total and free testosterone, and possibly DHEAS and A4, may assist in detecting underlying hyperandrogenaemia. Patients with androgen-secreting neoplasms have very high levels of androgens.

Total testosterone should be measured with a high-quality immunoassay after sample extraction and chromatography, or preferably mass spectrometry. Currently available, direct assays for the measurement of free testosterone are not reliable to be used in hirsute women [32]. Free testosterone should be measured, based on the total testosterone level, using equilibrium dialysis, competitive binding, or ammonium sulphate precipitation, or estimated from the ratio of total testosterone and SHBG (i.e. the free androgen index or FAI). Overall, free testosterone is more sensitive for the detection of hyperandrogenaemia than is total testosterone alone. The measurement of A4 or DHEAS increases the proportion of women detected as hyperandrogenaemic modestly.

Recently a new class of androgens (11-oxy-c19 steroids) have been reported to potentially play a role in androgen excess, including 11-KT, and 11-KDHT, which may become clinically useful in the future [1–3].

The initial evaluation of a hirsute woman should also include the measurement of a morning follicular phase 17-HP level to exclude NCAH. If the 17-HP level is between 2 and 10 ng/ml an acute adrenocorticotropic hormone (ACTH) stimulation test should be performed for the diagnosis of NCAH. The diagnosis of 21-OH-deficient NCAH is made biochemically if the basal or the stimulated levels are greater than 10 ng/ml or 15 ng/ml.

Prolactin and thyroid stimulating hormone (TSH) levels should also be checked if menstrual dysfunction or galactorrhoea or signs of thyroid dysfunction are present.

In the evaluation for PCOS, it is important to determine whether they have ovulatory dysfunction and polycystic ovaries. Many

women with hirsutism present with ovulatory and/or menstrual dysfunction. However, in up to 40% of hirsute patients with apparent regular menstruation, anovulation may still be present [26]. Consequently, in hirsute women with apparently regular menses, ovulatory function should be confirmed by obtaining a luteal phase progesterone level, best between 22 to 24 days after the start of menstruation. Progesterone levels lower than 3–5 ng/ml in a eumenorrhoeic patient are consistent with anovulation. Additionally, all hirsute patients should undergo pelvic ultrasonography to determine whether they have polycystic ovaries. Patients diagnosed with PCOS should also assess glucose intolerance and cardiometabolic risk factors.

Treatment of Hirsutism

A combination of pharmacological therapies and cosmetic amelioration is recommended for the treatment of hirsutism. If the underlying cause is one of the rare disorders previously discussed, therapies for these disorders should be undertaken.

The primary aim of the treatment of hirsutism is to stop the development of new terminal hairs. Hormonal therapy may also decrease the growth rate, diameter, and pigmentation of terminal hairs that are already present. Patients should be informed that the effect of treatment will probably only be observed after at least 6 months and the achievement of optimal results will require 12–24 months of therapy.

Any terminal hairs remaining after medical therapy must be destroyed mechanically.

Current treatment of hirsutism may consist of: (1) suppression of androgen production; (2) blockade of peripheral androgen action; and (3) mechanical and cosmetic means of hair removal.

Suppression of Androgen Production

Suppression of ovarian androgen secretion can be accomplished using combination hormonal contraceptives. Long-acting gonadotropin-releasing hormone (GnRH) analogues and insulin sensitizers are currently not indicated in the routine treatment of hirsutism in most women [33]. Likewise, suppression of adrenal androgen secretion can be achieved using various glucocorticoids, although the effect on hair growth is minimal at best, not to mention the significant negative impact of these medications on metabolic status.

Combination (oestrogen-progestin) oral contraceptive pills (OCPs) have been a mainstay for the treatment of hirsutism [33]. Oral contraceptives suppress the pituitary secretion of luteinizing hormone and lead to a decrease in ovarian androgen production. The oestrogenic fraction in OCPs increases the levels of SHBG which, in turn, results in a decrease in the free testosterone fraction. The progestin in the pill can compete for 5α-reductase and the AR. Combined oral contraceptives have also been shown to modestly decrease adrenal androgen production by a yet unclear mechanism, possibly due to decreasing ACTH levels.

Most OCPs contain ethinyl oestradiol as the oestrogenic fraction. Progestins in the oral contraceptives vary in their androgenic potential and some may actually decrease SHBG levels. Norethindrone, norgestrel, and levonorgestrel are known to have some intrinsic androgenic activity. Alternatively, the third-generation progestins norgestimate and desogestrel are non-androgenic and have the advantage of less metabolic side effects, including minimal impact on glucose, insulin, and lipids. There are a number of combined oral contraceptives containing antiandrogenic progestins. Of those, the ethinyl oestradiol and cyproterone acetate combination has been the most widely used in hirsutism. See Box 8.4.4.2.

Blockade of Peripheral Androgen Action

Agents that suppress androgen production (see earlier) when used alone usually have modest effect on hair growth, and in most hirsute patients, peripheral androgen blockers need to be added for an adequate treatment response [34]. These include AR blockers (spironolactone, cyproterone acetate, or flutamide) and 5α-reductase inhibitors (finasteride or dutasteride). All these agents are similarly efficacious and the main difference rests in their side effects profile. All have teratogenic potential, inducing feminization of a male fetus, and therefore should be used with effective contraception. The addition of combination contraceptives to peripheral androgen blockers provides protection against the risk of unwanted pregnancy, reduces the risk of irregular menstrual bleeding, and suppresses androgen levels by a different mechanism, thus improving their efficacy.

Spironolactone

Spironolactone is a potent antimineralocorticoid and mild diuretic. It is an effective therapy for hirsutism by competing with androgens for the AR, 5α-reductase, and SHBG. It also inhibits at least partially the activity of some of the ovarian and adrenal enzymes involved in androgen biosynthesis. Doses of 50–200 mg/day are generally used for the treatment of hirsutism. Side effects include menstrual irregularity, dyspepsia, nausea, polyuria, nocturia, hypotension, and headaches. If the dose is increased from 25 mg/day in a progressive fashion to the dose desired, patients tend to develop less side effects. Menstrual irregularity may be avoided if spironolactone is given in conjunction with an OCP.

Cyproterone Acetate

Cyproterone acetate is an antiandrogenic progestin effective in treatment of hirsutism and acne. It acts mainly by competitively binding with the AR. In mild to moderate cases, cyproterone acetate in a dose of 2 mg/day combined with ethinyl oestradiol generally

Box 8.4.4.2 Pharmacological treatment of hirsutism[a]

- Suppression of ovarian or adrenal androgen production
  - Combined contraceptive pills
  - Long-acting gonadotropin-releasing hormone (GnRH) analogues
  - Insulin sensitizers
  - Glucocorticoids
- Blockade of peripheral androgen receptor action
  - Cyproterone acetate
  - Spironolactone
  - Flutamide
- Blockade of 5α-reductase activity[b]
  - Finasteride/Dutasteride
  - Combination therapy

[a] At least 6 months of treatment is needed for an observable clinical response.
[b] Minimizing the conversion of testosterone to dihydrotestosterone

improves the symptoms. However, in more significant hirsutism, higher doses of cyproterone acetate may be required (up to 100 mg/day) for significant improvement. Side effects include mood changes, loss of libido, and weight gain.

Flutamide

Flutamide is an AR blocker, initially used as adjuvant treatment for prostate cancer. It is as effective as spironolactone in the treatment of hirsutism in doses of 125–500 mg/day, but with significantly less side effects, with the exception of the rare occurrence (3/10 000 users) of dose-dependent severe hepatotoxicity including the possibility of death. Careful monitoring of liver function tests is required due to the potential hepatotoxicity. Some authors completely recommend against the use of flutamide for hirsutism, considering its risk of potentially fatal hepatotoxicity [33].

Finasteride/Dutasteride

Finasteride is a competitive inhibitor of type 2 5α-reductase and is used for the treatment of benign prostatic hyperplasia. Although type 1 is the predominant isoform of 5α-reductase found in the PSU, finasteride 5 mg daily has been found to be useful for the treatment of hirsutism. Alternatively, dutasteride targets the type 1 isoform of 5a-reductase and may prove to have even higher efficacy, but so far there were no randomized controlled clinical trials with this drug.

Mechanical Means of Hair Removal

Pharmacological agents should be combined with appropriate mechanical/cosmetic treatments for optimal results in hirsute patients. Shaving, bleaching, or chemical depilation may be useful to temporarily ameliorate unwanted hair. Shaving does not affect the rate or duration of the anagen phase or diameter of the hair. Thus, patients can be reassured that shaving does not lead to a worsening of hirsutism. However, shaving can lead to a blunt hair end, which would give the false impression of thicker hair.

Plucking or waxing are not very much recommended because they cause discomfort and may lead to folliculitis with the subsequent development of in-grown hair. Furthermore, excessive or indiscriminate use of any depilating agent can result in chronic skin irritation.

Efluornithine is a topical irreversible inhibitor of l-ornithine decarboxylase (ODC), an enzyme which catalyses follicular polyamine synthesis necessary for hair growth. Efluornithine hydrochloride, marketed in a 13.9% cream, has been found to reduce unwanted facial hair in women. Efluornithine does not remove the hair, but rather reduces the rate of hair growth. Adverse effects are usually mild and include dry skin and itching.

Electrolysis, recently most often replaced by laser epilation, can be used to achieve a more permanent destruction of unwanted hairs, although long-term efficacy of these therapies is not well established. Electrolysis may still have an important role in women with blond or white hair in which laser epilation is not so effective. Repeated sessions of electrolysis can result in 20–50% permanent hair loss, which may last for months to years [35], significantly longer even than laser epilation.

Laser epilation, a selective photothermolysis technique, is also available for the treatment of hirsutism. A recent meta-analysis of the available randomized controlled trials involving 444 subjects reported that laser epilation results in about a 50% hair reduction that is maintained at least for up to 6 months after the final treatment [36]. Laser epilation appears to be more effective for hirsute women with dark hair and light skin. Side effects, including scarring and discoloration, might be observed after either electrolysis or laser epilation.

Emotional Support/Therapy

Hirsute women usually have high levels of emotional distress and some patients will have significant psychological morbidity including anxiety and depression [19]. Thus, education and psychological support are key elements of the overall therapeutic approach. Diagnosis, treatment alternatives, and expectations should be discussed in detail. Patients need to participate in shared decision-making regarding the treatment choice that best addresses their concerns. Observable decreases in unwanted hair with therapy help reduce their emotional burden. Nevertheless, professional psychological counselling may be needed in some patients.

Key Points

- Hirsutism is a common and significant health problem in women, with a negative impact on their quality of life.
- Hirsutism often signals an underlying androgen excess disorder, most commonly PCOS.
- A thorough evaluation in a hirsute woman should include a detailed clinical history and physical examination, a diagnostic work-up comprising a focused hormonal profile and a pelvic ultrasound.
- The first-line pharmacological treatment is oral contraceptives and/or antiandrogens, accompanied by mechanical/cosmetic hair removal as needed.

REFERENCES

1. Rege J, Nakamura Y, Satoh F, *et al*. Liquid chromatography-tandem mass spectrometry analysis of human adrenal vein 19-carbon steroids before and after ACTH stimulation. *J Clin Endocrinol Metab*, 2013; **98**: 1182–8.
2. O'Reilly MW, Kempegowda P, Jenkinson C, *et al*. 11-oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. *J Clin Endocrinol Metab*, 2017; **102**: 840–8.
3. Turcu AF, Auchus RJ. Clinical significance of 11-oxygenated androgens. *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 252–9.
4. Sanchez LA, Perez M, Azziz R. Laser hair reduction in the hirsute patient: a critical assessment. *Hum Reprod Update*, 2002; **8**: 169–81.
5. Paus R, Cotsarelis G. The biology of hair follicles. *N Engl J Med*, 1999; **341**: 491–7.
6. Azziz R, Carmina E, Sawaya ME. Idiopathic hirsutism. *Endocr Rev*, 2000; **21**: 347–62.
7. Danforth CH. Studies on hair: with special reference to hypertrichosis. *Arch Dermatol Syphilol*, 1925; **11**: 637–53.
8. Yildiz BO, Bolour S, Woods K, Moore A, Azziz R. Visually scoring hirsutism. *Hum Reprod Update*, 2010; **16**: 51–64.
9. Ferriman D, Gallwey JD. Clinical assessment of body hair growth in women. *J Clin Endocrinol Metab*, 1961; **21**: 1440–7.

10. Hatch R, Rosenfield RL, Kim MH, Tredway D. Hirsutism: implications, etiology, and management. *Am J Obstet Gynecol*, 1981; **140**: 815–30.
11. DeUgarte CM, Woods KS, Bartolucci AA, Azziz R. Degree of facial and body terminal hair growth in unselected black and white women: toward a populational definition of hirsutism. *J Clin Endocrinol Metab*, 2006; **91**: 1345–50.
12. Carmina E, Rosato F, Janni A, Rizzo M, Longo RA. Extensive clinical experience: relative prevalence of different androgen excess disorders in 950 women referred because of clinical hyperandrogenism. *J Clin Endocrinol Metab*, 2006; **91**: 2–6.
13. Yildiz BO, Bozdag G, Yapici Z, Esinler I, Yarali H. Prevalence, phenotype and cardiometabolic risk of polycystic ovary syndrome under different diagnostic criteria. *Hum Reprod*, 2012; **27**: 3067–73.
14. Azziz R, Sanchez LA, Knochenhauer ES, *et al*. Androgen excess in women: experience with over 1000 consecutive patients. *J Clin Endocrinol Metab*, 2004; **89**: 453–62.
15. Yildiz BO. Diagnosis of hyperandrogenism: clinical criteria. Best practice & research. *Clin Endocrinol Metab*, 2006; **20**: 167–76.
16. Bergfeld WF, Redmond GP. Androgenic alopecia. *Dermatol Clin*, 1987; **5**: 491–500.
17. Vexiau P, Chaspoux C, Boudou P, *et al*. Role of androgens in female-pattern androgenetic alopecia, either alone or associated with other symptoms of hyperandrogenism. *Arch Dermatol Res*, 2000; **292**: 598–604.
18. Azziz R. The evaluation and management of hirsutism. *Obstet Gynecol*, 2003; **101**(5 Pt 1): 995–1007.
19. Conway G, Dewailly D, Diamanti-Kandarakis E, *et al*. The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. *Eur J Endocrinol*, 2014; **171**: P1–29.
20. Ehrmann DA. Polycystic ovary syndrome. *N Engl J Med*, 2005; **352**: 1223–36.
21. Zawadzki J. DunaifA. *Diagnostic Criteria for Polycystic Ovary Syndrome: Towards a Rational Approach*. Boston, MA: Blackwell Scientific Publications, 1992.
22. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*, 2004; **81**: 19–25.
23. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*, 2004; **19**: 41–7.
24. Azziz R, Carmina E, Dewailly D, *et al*. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab*, 2006; **91**: 4237–45.
25. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. *Endocr Rev*, 1997; **18**: 774–800.
26. Azziz R, Waggoner WT, Ochoa T, Knochenhauer ES, Boots LR. Idiopathic hirsutism: an uncommon cause of hirsutism in Alabama. *Fertil Steril*, 1998; **70**: 274–8.
27. Carmina E. Prevalence of idiopathic hirsutism. *Eur J Endocrinol*, 1998; **139**: 421–3.
28. Pall M, Azziz R, Beires J, Pignatelli D. The phenotype of hirsute women: a comparison of polycystic ovary syndrome and 21-hydroxylase-deficient nonclassic adrenal hyperplasia. *Fertil Steril*, 2010; **94**: 684–9.
29. Azziz R, Hincapie LA, Knochenhauer ES, Dewailly D, Fox L, Boots LR. Screening for 21-hydroxylase-deficient nonclassic adrenal hyperplasia among hyperandrogenic women: a prospective study. *Fertil Steril*, 1999; **72**: 915–25.
30. Barbieri RL, Ryan KJ. Hyperandrogenism, insulin resistance, and acanthosis nigricans syndrome: a common endocrinopathy with distinct pathophysiologic features. *Am J Obstet Gynecol*, 1983; **147**: 90–101.
31. Trueb RM. Causes and management of hypertrichosis. *Am J Clin Dermatol*, 2002; **3**: 617–27.
32. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*, 2007; **92**: 405–13.
33. Martin KA, Anderson RR, Chang RJ, *et al*. Evaluation and treatment of hirsutism in premenopausal women: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1233–57.
34. Swiglo BA, Cosma M, Flynn DN, *et al*. Clinical review: antiandrogens for the treatment of hirsutism: a systematic review and metaanalyses of randomized controlled trials. *J Clin Endocrinol Metab*, 2008; **93**: 1153–60.
35. Wagner RF, Jr. Physical methods for the management of hirsutism. *Cutis*, 1990; **45**: 319–21, 25–6.
36. Haedersdal M, Gotzsche PC. Laser and photoepilation for unwanted hair growth. *Cochrane Database Syst Rev*, 2006; **4**: CD004684.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

8.5

Female Hypogonadism in Pre- and Post-Menopause

8.5.1 Female Hypogonadism

Premature Ovarian Insufficiency

Ephia Yasmin and Gerard S. Conway

Introduction

The ovary is a unique organ in the body in that is has a defined lifespan of approximately 50 years. The life expectancy of the ovary is determined by germ-cell number which is a balance between germ-cell multiplication *in utero* and factors affecting programmed cell death throughout life. The peak germ-cell number occurs at six months gestation with >700 000 eggs. Current understanding of ovarian reserve suggests that once this maximum number of oocytes has been achieved, the ability to the make new oocytes is lost and the total number gradually declines. New research questions whether oogonial stem cells can be activated to make new oocytes [1]. It is estimated that nearly 50% of germ cells have been destroyed by the age of menarche, and by the onset of menopause, only a few thousand remain. Throughout life therefore, germ-cell apoptosis takes place at a rate of approximately 25–150 lost per day. The decline in oocytes is roughly parallel with a fall in the concentration of anti-Müllerian hormone (AMH) in the circulation from about the age of 30 [2]. Only a small fraction of oocytes are accounted for by ovulation; about 500 in all. As the number of germ cell falls, a critical point is reached when periods stop with subsequent reduction in circulating oestrogen, and to a lesser extent, testosterone concentrations.

The menopause is defined by the last menstrual period and occurs at an average age of 50.7 years in the Western world [3]. Premature ovarian insufficiency (POI) is defined as amenorrhoea with raised FSH (>25 IU) on two occasions and low oestradiol before the age of 40 years [4]. POI occurs in approximately 1% of females and becomes increasingly rare in younger age groups. For instance, in presence of normal karyotype, 1:1000 of women at 30 has POI, 1:10 000 at 20 and 1:100 000 of women will present with gonadal failure and primary amenorrhoea (**Figure 8.5.1.1**). The prevalence of POF varies by ethnicity with women of oriental origin having a lower risk and African Americans a higher risk compared to Caucasian Americans [5]. In terms of the mode of presentation, POI is the aetiology of 10–28% of cases with primary amenorrhoea and about 10% of those with secondary amenorrhoea [6, 7].

The term premature ovarian insufficiency is preferred to earlier alternative, premature ovarian failure, because it is an all-encompassing term that accounts for the variable course and occasional remission [7]. The term hypogonadotropic hypogonadism is also used to emphasize ovarian origin whereby the raised concentrations on luteinizing hormone (LH) and FSH contrast from low gonadotrophins in hypothalamic or pituitary causes of hypogonadism. Gonadal dysgenesis refers to hypogonadotropic hypogonadism with a known genetic cause such as an abnormal 46,XY or 45,X karyotype and implies that ovarian development was halted at an early stage of embryonic development. In the case of Turner syndrome it

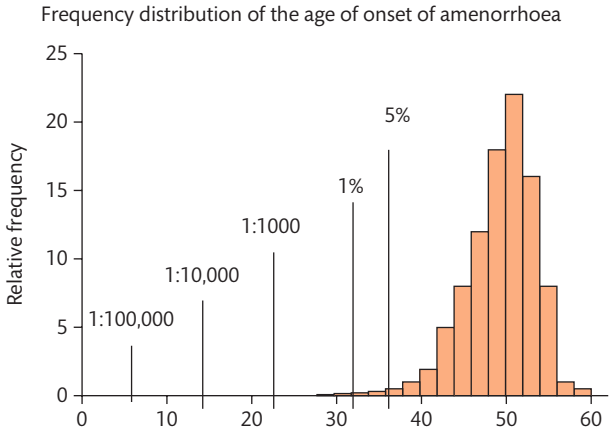


Figure 8.5.1.1 A schematic of the distribution of the age of onset of amenorrhoea according to the life span of the ovary. Data amalgamated from a variety of sources for the purposes of illustration.

is thought that early ovarian development is usually normal but that the chromosome anomaly leads to rapid germ-cell apoptosis [8].

The Menopause

Menopause describes the end of natural reproductive life when menstruation has been absent for at least 1 year. The process can be abrupt but in many women, symptoms precede amenorrhoea by months or even years. The age of menopause varies between 45 and 60 years and may have a familial link [9]. The interval between periods usually shortens as the menopause approaches from an average of 28 days often to 21 days. The diagnosis of menopause may not be obvious in women taking hormone treatments for period control or who have had a hysterectomy. In addition, non-specific symptoms such as joint and muscle stiffness, reduction of cognitive function and concentration, or simply low mood can occur in the absence of classical symptoms. Oestrogen deficiency should be considered as a possibility for anyone experiencing new onset of symptoms after the age of 40. The diagnosis of menopause is mainly based on history and, for the majority of women, no diagnostic test is required. A measurement of FSH can be unreliable as it may not correlate with symptoms of the premenopause.

The majority of women experience some symptoms of the menopause which often gradually diminish over several years [10]. In a proportion of women, disabling symptoms can be lifelong requiring indefinite treatment with hormone replacement therapy (HRT). Treatment of the menopause is individualized and includes information, counselling, lifestyle, hormone-based treatments, and treatments to improve sexual function. A discussion regarding the spectrum of symptoms that are associated with oestrogen deficiency can be reassuring and will ensure that no hidden symptom is left untreated. Symptoms relating to change of mood, libido, sexual function, and urogenital dysfunction should be elicited in history taking.

The risks and benefits of HRT require discussion as this treatment is not essential and non-pharmaceutical alternatives are available. Reports arising from the Women's Health Initiative (WHI) that highlighted the risks of HRT between 2002 and 2004 resulted in a decline in the use of HRT [11]. Subsequent assessments, however, have shown that the risks of HRT may have been overestimated. In addition, the risk of stroke can be reduced by the use of transdermal oestrogen instead of oral oestrogen that formed the basis of most of the early HRT studies.

Premature Ovarian Insufficiency

Clinical Evaluation and Diagnosis of POI

The clinical presentation of POI varies depending on the age of presentation [12, 13]. In very early onset in adolescence POI leads to pubertal delay and primary amenorrhoea. After menarche, the presentation is with symptoms of oestrogen deficiency and secondary amenorrhoea or as part of a work-up for irregular periods or infertility. In addition, POI can be part of clinical syndromes that can be genetic, such as Turner syndrome, or autoimmune endocrinopathies.

The formal diagnosis of POI is based on the presentation of amenorrhoea with finding of elevated serum FSH concentrations (FSH >25 IU/L) on at least two occasions separated by 4 weeks [4]. The requirement for two samples takes into account the fact that ovarian function can fluctuate and raised FSH can be a transient phenomenon. In addition, because the diagnosis has such severe implications, certainty of the diagnosis is required. While it is the usual expectation that the condition will be permanent, some women follow an unpredictable course of relapse and remission particularly in the first year after diagnosis [14]. There is a commonly quoted pregnancy rate of approximately 1–5% in women with POI and it is therefore important to inform women with POI of this phenomenon and advise use of contraception when appropriate. Because of this background fertility, case reports of effective treatment of POI must be viewed with caution. Fluctuating ovarian function probably accounts for many cases which are labelled 'resistant ovary syndrome'. It is now understood that ovarian biopsy is not predictive of remission and should no longer be included as part of the work up of POI [4]. Anti-Müllerian hormone has found a place as a diagnostic marker of low ovarian reserve but once the FSH concentration is raised, AMH offers little additional information [15].

A careful family history can identify other affected female members in as many as 30% of cases [16] and in this situation more genetic screening is indicated and genetic counselling for relatives may be appropriate. A karyotype should be performed in anyone with a positive family history because small chromosome deletions can be inherited. Fragile X (Fra-X) premutation analysis should be offered to all women with POI because of its value in detecting a new pedigree for the prevention of fragile X syndrome [17, 18].

Oestrogen receptors are widely distributed throughout the body and the effects of oestrogen deficiency are diverse. While the classical symptoms of hot flushes, sleep disturbance, and vagina dryness are obvious, subtle changes in mood and musculoskeletal pain may go unrecognized as part of this symptom spectrum. A careful history listing any change in wellbeing, no matter how minor, that coincides or precedes amenorrhoea will provide a useful guide to optimizing HRT.

The major medical issues for health surveillance in women with POI revolve around the quality of life and bone protection offered by HRT. Oestrogen deficiency is a risk for osteoporosis and DXA scans should be performed at intervals to assist in the correct dosing of HRT. Fertility is often a concern for women with POI. A discussion on options for reproduction include use of donated oocytes and adoption. Prepregnancy counselling should include health issues such as risk of aortic root dissection in Turner syndrome.

Psychological and emotional support to deal with impact of diagnosis on their health and relationships should form part of the multiprofessional care in POI. Long-term follow-up is essential to monitor HRT and to consider emerging associated autoimmune pathology.

Aetiology of POI

Ovarian insufficiency is the final outcome many pathogenic mechanisms. Table 8.5.1.1 shows the main categories of pathogenic mechanisms that lead to ovarian insufficiency. Despite diagnostic advances, particularly in the identification of genetic associations, the cause of POI remains unknown in the majority of cases [13].

Genetic Causes of POI

A genetic aetiology for POI is suggested not only by a positive family history but also if there are features of an associated syndrome. Cytogenetic chromosome anomalies occur in about 2% of cases with the majority involving the X chromosome.

Table 8.5.1.1 Summary of pathogenic mechanisms causing premature ovarian insufficiency

Chromosome anomalies	Chromosomal defects cause both abnormal ovarian development and accelerated apoptosis. Turner syndrome is the most common causing and females with Downs Syndrome have later onset POI.
Single gene defects	Many single gene defects have been described but in total they account for only a small minority of cases of POI. Mutations are commonly in genes vital for ovarian development or function. The most common gene defect is the Fra-X permutation.
Autoimmune	Autoimmune ovarian damage is usually presumed because of the association with other organ specific autoimmune conditions. Thyroid and adrenal autoimmunity is common. Ovarian autoantibodies are rarely positive.
Metabolic	In galactosaemia, abnormal metabolites are thought to mediate ovarian damage.
Iatrogenic	Ovarian damage following pelvic surgery and radiotherapy of chemotherapy during treatment of cancer.
Environmental factors	Viral infections and toxins such as pesticides are presumed. Smoking results in younger age of natural menopause.

X chromosome defects and Turner's syndrome: Defects of the X chromosome associated with POI include complete or partial deletion of monosomy X (Turner syndrome), trisomy X or X-autosome translocations. In the case of Turner syndrome variants, there is a great deal of variability in severity of the condition and this corresponds approximately with genotype and the amount of X chromosome disruption. Ovarian histology can vary from streak ovaries to those with well-preserved structure and follicles. Those with monosomy X (45X) tend to be the most severely affected while those with partial deletions of the X chromosome or mosaic 45X/46XX karyotype often lack the typical phenotypic features of the syndrome but may present with only with ovarian insufficiency and secondary amenorrhoea. Among women with Turner syndrome, the prevalence of ovarian failure is between 80% and 90% with most experiencing primary amenorrhoea and complete pubertal delay.

X chromosome deletions appear to segregate in two specific regions: POI1 at Xq26-qter and POI2 Xq13.3--Xq21.1 [19]. There are various proposed mechanisms by which X chromosome defects causes POI. Genes responsible for germ-cell development located along the critical region may be interrupted although break point genes that determine ovarian function have not been convincingly identified. It seems most likely that structural rearrangements of the X chromosome do not affect single genes but rather disrupt normal pairing at meiosis leading to meiotic arrest and subsequent atresia. It must also be noted however that some X chromosome break points are not associated with POI [19].

Single gene defects causing POI: A growing number of genes have been linked to ovarian insufficiency although the strength of evidence linking each anomaly with POI is variable [20, 21]. The most clinically useful genetic association is that between carriers of the Fra-X premutation and ovarian failure. Fra-X premutations occur in 3% of women with sporadic POI and 15% of those with the familial form and this genetic test is the only one advised in the United Kingdom as part of routine work up of POI [22]. In general, the phenotype of this subgroup of women with POI is indistinguishable from others although a minority are reported to a progressive intention tremor. The pathway by which Fra-X premutations damage ovarian function is unclear as this type of expansion of exon 1 of the gene is not thought to effect protein transcription even though the protein is expressed within the ovary.

Autoimmune Causes of POI

Autoimmune mechanisms may be involved in pathogenesis of up to 30% of cases of POI [23, 24]. POI has been reported to be associated with various endocrine (thyroid, adrenal, hypoparathyroidism, diabetes mellitus, and hypophysitis) and non-endocrine (chronic candidiasis, idiopathic thrombocytopenic purpura, vitiligo, alopecia, autoimmune haemolytic anaemia, pernicious anaemia, SLE, rheumatoid arthritis, Crohn's disease, Sjogren's syndrome, myasthenia gravis, primary biliary cirrhosis, and chronic active hepatitis) autoimmune disorders [25]. In many cases non-ovarian autoimmune involvement exists only at subclinical level [24]. POI may be part of the autoimmune polyglandular syndromes (APS) when accompanied by other autoimmune endocrinopathies. POI is more common with APS types I and III than with APS type II. The single most common autoimmune association is with hypothyroidism which occurs in 10% of women with POI.

Various pathways of autoimmune ovarian damage have been described but a reliable ovarian specific autoimmune marker is still lacking. Several putative pathogenic autoantibodies have been explored [26]. Antiovarian antibodies detected by routine *immunofluorescence* have been reported in several studies of women with POI but their specificity and pathogenic roles are questionable. The incidence of antiovarian antibodies in POI in different studies has been reported to vary from 4% to 69% [27]. Other candidate autoantibodies include those directed against steroidogenic enzymes (such as 3ß-hydroxysteroid dehydrogenase), gonadotropins, and their receptors, the corpus luteum, zona pellucida, and oocyte. None of these, however, have been validated as a diagnostic marker of autoimmune ovarian failure. Therefore, in the clinical work up of POI, screening for an autoimmune aetiology is usually only possible by seeking coexisting autoimmune diseases.

Women with idiopathic POI show an increased number of activated T cells in peripheral blood. Similar findings have been described in other autoimmune endocrinopathies, such as recent onset Graves' disease, type 1 diabetes mellitus, and Addison's disease. However, postmenopausal women also show raised numbers of activated peripheral T cells and oestrogen substitution has been shown to lower the number of activated peripheral T cells in women with POI. Therefore, it is difficult to be certain whether the raised numbers of activated blood T cells is the cause or the result of ovarian failure in this situation [25].

POI Resulting from Cancer Treatments

As the number of survivors of childhood cancer increases across the world, so iatrogenic POI becomes a greater proportion of women with this condition. Both chemotherapy, alkylating agents in particular, and radiotherapy used in the treatment of cancer or benign diseases are damaging to the ovary [28, 29]. The return of ovarian

function after cancer treatment is commonly observed although difficult to predict. For women estimated to be high risk of POI, oophorectomy or oocyte harvesting for cryopreservation may be considered [30].

Miscellaneous Causes of POI

Viral oophoritis is a possible occult aetiology that could theoretically account for the many cases of idiopathic POI. There is, however, little direct evidence of viral ovarian damage beyond case reports such as mumps oophoritis [31]. Cigarette smoking in known to be associated with earlier age of natural menopause but whether environmental effects are sufficiently strong to cause POI is unlikely. The available data regarding effects of endocrine disruptors, heavy metals, solvents, pesticides, plastics, and industrial chemicals on female reproduction is equivocal [32].

Hormone Replacement Therapy

Physiological replacement of ovarian steroid hormones until the age of normal menopause at 50, is generally accepted as routine for women with POI. However, there is little risk/benefit data for this young population. The principle of HRT use in young women differs only slightly from that in older women with the main treatment goal being optimal quality of life. Young women may require a higher oestrogen dose than that used in an older age group. Also, expectations for sexual function can be higher commonly requiring consideration for vaginal oestrogen and androgen replacement. An HRT regimen should be based on the individual preferences of each patient who should be encouraged to undertake a trial and error approach through the wide variety of product available.

Management of oestrogen replacement for young women presenting with primary amenorrhoea requires liaison with paediatric endocrinologists with experience in the induction of puberty in order to optimize breast and uterine development. For instance, a popular strategy is to maximize the time between the introduction of oestrogen and starting progesterone withdrawal bleeds is thought to benefit breast development. Conversely, the common practice of starting a low dose combined oral contraceptive in this circumstance may not offer the best outcome of uterine development.

Among **oral oestrogen** choices, oestradiol esters and 17β-oestradiol provide the mainstay of treatments [33]. Conjugated equine oestrogens are used less frequently than in the past because of a greater risk of hypertension and thrombosis compared to oestradiol [34, 35]. Some young women with POI find the **combined oral contraceptive pills** (COCP) pills a more acceptable option for oestrogen replacement but careful assessment of the pill free week is advised. The pill free week amounts to three months of oestrogen deficiency each year which may coincide with symptoms of oestrogen deficiency or bone loss. In choosing a type of COCP those with lower dose ethinylestradiol and second generation progestins have the lowest risk to thrombosis [36]. **Transdermal oestrogen** avoids first-pass liver metabolism involves non-invasive self-administration and attainment of therapeutic hormone levels with low daily doses [37]. This route of oestrogen administration also appears to be associated with a lower risk of thrombo-embolism. **Subcutaneous oestrogen** replacement involves placement of 25–50 mg oestradiol pellets usually in the lower abdomen or buttocks in a minor office procedure that is not currently available in some countries including the United Kingdom. **Topical vaginal oestrogen** may be used as an adjunct to systemic oestrogen. Creams, pessaries, tablets and vaginal ring appear to be equally effective for control of symptoms [38].

Once the choice of oestrogen has been made, separate consideration can be given to the **progestin** in women with an intact uterus. Progestins vary from the more potent such as norethisterone to the weaker such as dydrogesterone. Trial and error will allow the user to find the most suitable progesterone preparation. The route may be oral, transdermal, vaginal, or uterine. With the **oral** and **transdermal** routes there is a choice between continuous or sequential (for 10–14 days each month) delivery. Sequential regimen ensures monthly menstrual bleed. Continuous regimen avoids menstrual flow but break through bleeding may be more common in young women compared to an older age group in whom there is greater uterine atrophy. **Uterine delivery** with the levonorgestrel intrauterine device (Mirena) has the advantage of avoiding the adverse effects of oral progestins highlighted in the studies of older women [39].

Androgen replacement is useful in some instances when fatigue and loss of libido persist despite optimized oestrogen replacement [40] Transdermal testosterone administration and dehydroepiandrosterone treatment are two of the options for androgen replacement in these women. However, testosterone preparations are generally not licensed for use in females and most preparations available in the market are for use in men. Out of licence use in the female with involve using a significantly lower dose of testosterone.

Fertility Options for Women with POI

Different fertility options will be appropriate for each individual with POI. It is important that discussions on this topic from a professional with experience in the field, is made available not only soon after diagnosis but at intervals throughout follow-up as the technology and opportunities in this field are continually developing.

Spontaneous Fertility in POI

Women with POI have a 5–10% chance of spontaneous conception at some time after diagnosis as in some cases hormone levels and disease activity fluctuate and return to biochemical normality; this is often transient and the likelihood of recovery of ovulation is not possible to predict. Pregnancy loss in those who conceive is reported at 20%, which is similar to that of the normal population. Several medical therapies have been tried to induce ovulation in women with POI; however, in a systematic review all were reported to be equally ineffective [16]. In particular, glucocorticoids have been considered for those with autoimmune markers and although promising in early case reports, a controlled trial failed to show benefit [41].

Ovum Donation

Assisted conception with donated oocytes has been used to achieve pregnancy in women with POI for over 20 years and remains the only realistic fertility treatment for the majority of women with established POI [42]. The availability of donated oocytes varies from

country to country and this option is not acceptable in certain cultures. Embryos are developed with the using donated eggs fertilized with partner's sperm. The best embryo is chosen for transfer with the recipient having been prepared with oral or transdermal oestrogen. Because of the lack of a corpus luteum, progesterone support is required which can be administered orally, vaginally, or in injectable form.

Embryo Cryopreservation

Cryopreservation of embryos is a long-established technique as part of *in vitro* fertilization (IVF) but it will apply to only a small number of women with normal ovarian function who can anticipate imminent ovarian failure but have sufficient time to allow for controlled ovarian hyperstimulation(28). This circumstance might arise in some cases of malignancy or in someone who is known to be at risk of ovarian failure such as those with X chromosome anomalies or Fra-X premutation who may wish to delay starting a family for some years. Donor sperm could be used for those without a partner.

Oocyte Cryopreservation

Cryopreservation of oocytes has become popular since the advent of vitrification which improved the freezing process for such a large cell. Vitrification involves rapid cooling in high concentrations of penetrating cryoprotectants which avoids formation of intracellular ice and resulting damage during cooling and warming. Oocyte cryopreservation also requires controlled ovarian stimulation but in contrast to preservation of embryos, no sperm is required. In practice, this option is only applicable to women with normal ovarian function and old enough for oocyte harvesting. The success rate for achieving pregnancy is likely to be lower than for embryo cryopreservation because oocyte survival and subsequent fertilization is impaired. This option has been used for women with Turner syndrome who retain some ovarian function [43] but so far the overall fertility outcome is not known for this subgroup compared to the more common application in women diagnosed with cancer.

Ovarian Tissue Cryopreservation

The use of ovarian tissue cryopreservation for later use has been explored in young women undergoing anticancer treatment. Cryopreservation of ovarian cortex or strips of ovarian cortical tissue has resulted the birth of over 100 children [44]. This procedure has the advantage of not requiring sperm and being appropriate for prepubescent girls accepting that a laparoscopy is required. The cryopreserved tissue can be re-implanted into the pelvis with ovulation and pregnancy occurring spontaneously or after IVF. Re-implanted ovarian tissue has the added advantage of providing endogenous oestrogen but will be subject to the same mechanism causing POI primarily.

Psychology

The diagnosis of POI is distressing and a life experience with high levels of depression and low levels of self-esteem with negative effects on sexuality in affected women. Many women with POI report moderate to severe stress at the time of diagnosis and are unsatisfied with the amount of information provided by their clinician. Therefore, access to a clinical psychologist is recommended. Often several visits will be required for detailed information on aetiology and fertility options to be retained. Regular specialist follow-up is recommended so that women have access to accurate fertility information and so that the need for psychology support can be reassessed at intervals. Often crises arise some years after the original diagnosis, for instance, when a near relative or peer achieves a pregnancy.

Box 8.5.1.1 An overview of the management of POI

Education and counselling

- Return of ovarian function May occur in 5 – 10% of women with POI but it is not possible to predict when and if this may happen with precision.
- Adoption and oocyte donation are the only realistic options for fertility although cryopreservation ovarian tissue it is possible for those at risk of developing POI.
- Access to follow-up counselling is important as issues return with life events such as a close friend all family member becoming pregnant.

Investigations

- Thyroid function tests, vitamin B_{12}, and ferritin because of associated autoimmune conditions.
- Autoantibody screen including thyroid and adrenal antibodies.
- Karyotype for early onset POI and genetic screen for FRAXA premutation.
- Pelvic ultrasound and ovarian biopsy are not recommended.

Treatment

- Oestrogen and progesterone replacement is usually indicated an individualized.
- Education is required regarding risks and benefits of oestrogen preparations—oral, transdermal, and vaginal.
- Inform on media coverage of the use of HRT in postmenopausal women that are not relevant to young women.
- Consider vaginal oestrogen and testosterone supplements.
- Psychosexual and fertility counselling.

Conclusions

Premature ovarian failure is a complex condition that requires specialist services. The diagnostic work-up is aimed at determined the aetiology where possible and is followed by a screen for syndromic conditions. Oestrogen replacement and fertility options need to be reassessed at intervals and clinicians have to be vigilant for psychological sequelae (see Box 8.5.1.1).

REFERENCES

1. Clarkson YL, McLaughlin M, Waterfall, M, *et al*. Initial characterisation of adult human ovarian cell populations isolated by DDX4 expression and adelhyde dehydrogenase activity. *Sci Rep*, 2018; **8**: 6953.
2. Anderson RA, Nelson SM, Wallace WH. Measuring anti-Mullerian hormone for the assessment of ovarian reserve: when and for whom is it indicated? *Maturitas*, 2012; **71**: 28–33.
3. van Noord PA, Dubas JS, Dorland M, Boersma H, te Velde E. Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. *Fertil Steril*, 1997; **68**: 95–102.

4. Webber L, Davies M, Anderson R, *et al*. ESHRE Guideline: management of women with premature ovarian insufficiency. *Hum Reprod*, 2016; **31**: 926–37.
5. Luborsky JL MP, Sowers MF, Gold EB, Santoro N. Premature menopause in a multi-ethnic population study of the menopause transition. *Hum Reprod*, 2003; **18**: 199–206.
6. Master-Hunter T, Heiman DL. Amenorrhea: evaluation and treatment. *Am Fam Physician*, 2006; **73**: 1374–82.
7. Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. *Clin Endocrinol (Oxf)*, 2008; **68**: 499–509.
8. Reynaud K, Cortvrindt R, Verlinde F, De Schepper J, Bourgain C, Smitz J. Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil Steril*, 2004; **81**: 1112–19.
9. Davis SR, Lambrinoudaki I, Lumsden M, *et al*. Menopause. *Nat Rev Dis Primers*, 2015; **1**: 15004.
10. Monteleone P, Mascagni G, Giannini A, Genazzani AR, Simoncini T. Symptoms of menopause—global prevalence, physiology and implications. *Nat Rev Endocrinol*, 2018; **14**: 199–215.
11. Sassarini J, Lumsden MA. Oestrogen replacement in postmenopausal women. *Age Ageing*, 2015; **44**: 551–8.
12. Conway GS, Kaltsas G, Patel A, Davies MC, Jacobs HS. Characterization of idiopathic premature ovarian failure. *Fertil Steril*, 1996; **65**: 337–41.
13. Nelson LM. Clinical practice. Primary ovarian insufficiency. *N Engl J Med*, 2009; **360**: 606–14.
14. Bachelot A, Nicolas C, Bidet M, *et al*. Long-term outcome of ovarian function in women with intermittent premature ovarian insufficiency. *Clin Endocrinol (Oxf)*, 2017; **86**: 223–8.
15. Visser JA, Schipper I, Laven JS, Themmen AP. Anti-Mullerian hormone: an ovarian reserve marker in primary ovarian insufficiency. *Nat Rev Endocrinol*, 2012; **8**: 331–41.
16. van Kasteren YM, Hundscheid RD, Smits AP, Cremers FP, van Zonneveld P, Braat DD. Familial idiopathic premature ovarian failure: an overrated and underestimated genetic disease? *Hum Reprod*, 1999; **14**: 2455–9.
17. Conway GS. Premature ovarian failure and FMR1 gene mutations: an update. *Ann Endocrinol (Paris)*, 2010; **71**: 215–17.
18. Mila M, Alvarez-Mora MI, Madrigal I, Rodriguez-Revenga L. Fragile X syndrome: an overview and update of the FMR1 gene. *Clin Genet*, 2018; **93**: 197–205.
19. Baronchelli S, Villa N, Redaelli S, *et al*. Investigating the role of X chromosome breakpoints in premature ovarian failure. *Mol Cytogenet*, 2012; **5**: 32.
20. Ferrarini E, Russo L, Fruzzetti F, *et al*. Clinical characteristics and genetic analysis in women with premature ovarian insufficiency. *Maturitas*, 2013; **74**: 61–7.
21. Desai S, Rajkovic A. Genetics of reproductive aging from gonadal dysgenesis through menopause. *Semin Reprod Med*, 2017; **35**: 147–59.
22. Conway GS, Payne NN, Webb J, Murray A, Jacobs PA. Fragile X premutation screening in women with premature ovarian failure. *Hum Reprod*, 1998; **13**: 1184–7.
23. Welt CK. Autoimmune oophoritis in the adolescent. *Ann N Y Acad Sci*, 2008; **1135**: 118–22.
24. Wilson C. Autoimmunity: autoimmune Addison disease and premature ovarian failure. *Nat Rev Endocrinol*, 2011; **7**: 498.
25. La Marca A, Brozzetti A, Sighinolfi G, Marzotti S, Volpe A, Falorni A. Primary ovarian insufficiency: autoimmune causes. *Curr Opin Obstet Gynecol*, 2010; **22**: 277–82.
26. Melner MH, Feltus FA. Autoimmune premature ovarian failure—endocrine aspects of a T cell disease. *Endocrinology*, 1999; **140**: 3401–3.
27. Wheatcroft NJ, Salt C, Milford-Ward A, Cooke ID, Weetman AP. Identification of ovarian antibodies by immunofluorescence, enzyme-linked immunosorbent assay or immunoblotting in premature ovarian failure. *Hum Reprod*, 1997; **12**: 2617–22.
28. Wallace WH, Anderson RA, Irvine DS. Fertility preservation for young patients with cancer: who is at risk and what can be offered? *Lancet Oncol*, 2005; **6**: 209–18.
29. Morgan S, Anderson RA, Gourley C, Wallace WH, Spears N. How do chemotherapeutic agents damage the ovary? *Hum Reprod Update*, 2012; **18**: 525–35.
30. Wallace WH, Kelsey TW, Anderson RA. Ovarian cryopreservation: experimental or established and a cure for the menopause? *Reprod Biomed Online*, 2012; **25**: 93–5.
31. Morrison JC, Givens JR, Wiser WL, Fish SA. Mumps oophoritis: a cause of premature menopause. *Fertil Steril*, 1975; **26**: 655–9.
32. Sharara FI, Seifer DB, Flaws JA. Environmental toxicants and female reproduction. *Fertil Steril*, 1998; **70**: 613–22.
33. Lobo RA. Hormone-replacement therapy: current thinking. *Nat Rev Endocrinol*, 2017; **13**: 220–31.
34. Swica Y, Warren MP, Manson JE, *et al*. Effects of oral conjugated equine estrogens with or without medroxyprogesterone acetate on incident hypertension in the Women's Health Initiative hormone therapy trials. *Menopause*, 2018; **25**: 753–61.
35. Smith NL, Blondon M, Wiggins KL, *et al*. Lower risk of cardiovascular events in postmenopausal women taking oral estradiol compared with oral conjugated equine estrogens. *JAMA Intern Med*, 2014; **174**: 25–31.
36. Hugon-Rodin J, Gompel A, Plu-Bureau G. Epidemiology of hormonal contraceptives-related venous thromboembolism. *Eur J Endocrinol*, 2014; **171**: R221–30.
37. Henzl MR, Loomba PK. Transdermal delivery of sex steroids for hormone replacement therapy and contraception. A review of principles and practice. *J Reprod Med*, 2003; **48**: 525–40.
38. Faubion SS, Sood R, Kapoor E. Genitourinary syndrome of menopause: management strategies for the clinician. *Mayo Clin Proc*, 2017; **92**: 1842–9.
39. Marjoribanks J, Farquhar C, Roberts H, Lethaby A, Lee J. Long-term hormone therapy for perimenopausal and postmenopausal women. *Cochrane Database Syst Rev*, 2017; **1**: CD004143.
40. Wierman ME, Arlt W, Basson R, *et al*. Androgen therapy in women: a reappraisal: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 3489–510.
41. van Kasteren YM, Braat DD, Hemrika DJ, *et al*. Corticosteroids do not influence ovarian responsiveness to gonadotropins in patients with premature ovarian failure: a randomized, placebo-controlled trial. *Fertil Steril*, 1999; **71**: 90–5.
42. Luisi S, Orlandini C, Regini C, Pizzo A, Vellucci F, Petraglia F. Premature ovarian insufficiency: from pathogenesis to clinical management. *J Endocrinol Invest*, 2015; **38**: 597–603.
43. El-Shawarby SA, Sharif F, Conway G, Serhal P, Davies M. Oocyte cryopreservation after controlled ovarian hyperstimulation in mosaic Turner syndrome: another fertility preservation option in a dedicated UK clinic. *BJOG*, 2010; **117**: 234–7.
44. Richard A. Anderson WHBW, Evelyn E. Telfer. Ovarian tissue cryopreservation for clinical and research perspectives. *Hum Reprod Open*, 2017; **1**: hox001.

8.5.2 Female Hypogonadism

Endocrinology of the Menopause and Hormone Replacement Therapy

Stavroula A. Paschou, Panagiotis Anagnostis, and Dimitrios G. Goulis

Introduction

Menopause marks the permanent cessation of menstruation and the transition to the non-reproductive stage of a woman's life. The alteration in the hormonal milieu during menopause is the result of a decline in the ovarian reserve and varies significantly among individuals, affected by both genetic and environmental factors. A complex and challenging pattern of clinical and metabolic consequences is associated with the transition to menopause. The purpose of this chapter is to provide current knowledge regarding the definition of menopause and the different stages preceding and following the final menstrual period, the physiology of the endocrine changes occurring during the menopausal transition, the clinical manifestations, diagnosis, and differential diagnoses, as well as the pros and cons of its medical management.

Definitions

The term 'menopause' is used to describe the post-reproductive period in a woman's life, after completion of 12 months of amenorrhea or immediately after bilateral oophorectomy. The mean age of natural menopause is 50–52 years. In cases in which the menopause is recorded before the age of 45 years, the term 'early menopause' (EM) is used [1, 2]. EM has been recorded in 10% of postmenopausal women, while in 1% of cases, menopausal age has been reported before the age of 40 years. In the latter situation, the term 'premature ovarian insufficiency' (POI) is used. The prevalence of POI in women younger than 30 years of age is 0.1% [1–3]. 'Perimenopause' refers to the period during which a woman experiences variable menstrual cycle changes and symptoms attributed to different degrees of decline in ovarian reserve [4].

To describe the transition through the different stages of reproductive age during the years preceding or following the menopause, the Stages of Reproductive Aging Workshop (STRAW) system has been developed [4]. Its initial form was proposed in 2001 and its update in 2011, the latter known as STRAW + 10. It is considered as the standard staging system that describes the transition to menopause in a detailed way since it discriminates the onset of late reproductive life and early menopausal transition. It also provides a more detailed classification of postmenopausal status [4]. The STRAW system is based not only on the woman's symptomatology and menstrual pattern but also on the alterations in serum concentrations of specific endocrine markers of ovarian ageing, such as inhibin B and anti-Müllerian hormone (AMH) [4].

According to the STRAW classification system, the menopausal transition is divided into two stages, Stage –2 ('early menopausal transition'), which is designated by a persistent difference of seven or more days in the length of consecutive menstrual cycles. Stage –2 is also characterized by elevated follicular stimulating hormone (FSH) with low AMH and inhibin B concentrations, as well as, low antral follicle count (AFC). Decreased AMH, inhibin B and AFC with a shorter cycle duration (without any change in menstruation pattern and normal FSH concentrations) are the first signs observed in the Stage –3, which marks the late period of reproductive age [4]. Stage –1 ('late menopausal transition') is characterized by persistent elevation in FSH (more than 25 IU/L) concentrations, associated also with high oestradiol (E2) concentrations. AFC, AMH, and inhibin B concentrations are low. The first vasomotor symptoms may occur during this stage. The early postmenopausal Stage follows and is divided into Stages +1a (referring to the first 12 months and marking the end of perimenopause), +1b (covering the next 12 months) and Stage +1c, lasting for 3–6 years. During the Stage +1c, FSH concentrations stabilize, whereas AMH and inhibin B concentrations and AFC are extremely low [4]. In general, AMH and AFC seem to be superior to FSH as markers of ovarian ageing not only because they herald the late reproductive stages and transition to perimenopause, but also because they are not affected by menstrual cycle phase [4].

It must be underlined that the STRAW criteria are not applicable to women with polycystic ovarian syndrome (PCOS), hypothalamic amenorrhoea, and chronic illnesses, such as human immunodeficiency virus (HIV) infection. They cannot also be used in cases of chemotherapy-associated menopause (after administration of alkylating agents) or treatment with tamoxifen. Furthermore, in cases of hysterectomy or endometrial ablation, it is recommended to use the STRAW criteria only on the basis of endocrine biomarkers rather than on menstrual pattern, after at least three months postoperatively [4]. The lack of standardized assays and the variability for biomarkers, such as inhibin B and AMH, constitute another limitation of the STRAW classification system [4].

Endocrine Physiology

Transition to menopause is the result of the coalescence of an acceleration of follicle loss and a decline in the functional quality of oocytes. The reduction in ovarian reserve is indicated by the decrease in AFC, total non-growing follicle and primordial follicle pool [5]. The nerve growth factor (NGF) pool numbers, approximately seven million at birth, reaches the number of 300 000 at the onset of puberty and declines at 12% and 3% of the initial number at the age of 30 and 40 years, respectively. Non-growing follicles finally decrease

to less than 1000 at the age of natural menopause (50–52 years) [5]. During the late stages of reproductive life, a gradual decrease in the number of follicles entering the growth phase (from 45 during the age of 25–30 years to six during the age of 38–45 years). During perimenopause, 20% of cycles with a duration of less than 40 days and 80% of those with a duration of more than 40 days are anovulatory [5, 6].

The acceleration of follicle depletion is the result of an early rise in FSH concentrations, due to abnormal gonadotrophin-releasing hormone (GnRH) pattern. The latter is the consequence of dysregulation of the GnRH pulse generator in the hypothalamus, due to the progressive loss of control from brain centres [6]. FSH is also elevated due to the loss of negative feedback by inhibin B, whose concentrations decrease as the non-growing follicle pool declines with age. The sustained FSH elevation during late perimenopause and early postmenopausal period is responsible for the increase in cycle length variability [5, 6]. In parallel with FSH, E2 concentrations increase during the early follicular phase of the menstrual cycle, especially during early perimenopause, as a result of accelerated follicular growth. The primary circulating oestrogen during menopause is estrone, which is mainly produced through the peripheral conversion of the adrenal steroid hormone Δ_4-androstenedione by the enzyme aromatase, in adipose tissue. Only 5% of estrone is converted to E_2 via the enzyme 17β-hydroxysteroid dehydrogenase (17β-HSD). A decline in progesterone is observed during the mid-cycle and luteal phase with advancing reproductive age [5, 6].

Inhibins A and B are members of the transforming growth factor-β (TGF-β) superfamily and secreted by the granulosa cells of ovarian follicles. Their primary role is the inhibition of gonadotrophins, mainly FSH. Inhibin A is produced by the dominant follicle and the corpus luteum (during the luteal phase), whereas inhibin B is secreted by the granulosa cells of the antral follicles, being the principal regulator of the FSH concentrations during the follicular phase. During menopausal transition, inhibin B and AMH concentrations start to decrease along with AFC, as a result of the shrinkage of the follicle cohort [5, 6]. AMH is also a member of the TGF-β superfamily peptides, secreted by granulosa cells of late preantral and small antral follicles, and exerts an inhibitory effect on the primordial to primary follicle transition. AMH concentrations, reflect the reserve of follicle cohort and constitute a reliable tool for predicting the age of menopause [7]. Genetic predisposition and environmental factors, such as smoking, diet, physical activity, and socioeconomic status affect the time of transition to menopause, explaining in part the variability in the age of menopause [4–6].

Clinical Picture

The earliest and most prevalent symptom of a woman entering the menopause is a disordered pattern of menstruation, specifically an alteration in cycle length by more than seven to more than 60 days [1–4]. As just mentioned, the former is the characteristic pattern of the 'early menopausal transition' (Stage −2), whereas the latter is defined as the 'late menopausal transition' (Stage −1), lasting 1–3 years until the menopause [4]. Except for the menstrual irregularities, a variability in menopause-related symptoms exists during these stages, associated with the fluctuations in oestrogen concentrations [4].

Many tools have been developed to assess the severity of menopausal symptoms. One of the most widely used is the Greene Scale, which is a self-report questionnaire, consisting of 21 items, which measure physical and psychological symptoms (four options, from 0, corresponding to 'not at all', to 3, corresponding to 'extremely'). These items cover vasomotor symptoms (two items, range 0–6), anxiety (six items, range 0–18), depression (five items, range 0–15), somatic symptoms (seven items, range 0–21) and sexuality (one item, 0–3). The whole score ranges from 0 to 63 [8].

Vasomotor symptoms (VMS) are the predominant menopause-related clinical manifestations and include hot flushes, palpitations, night sweating, disturbed sleep pattern, and fatigue. VMS last 7.4 years on average, as shown by large epidemiological studies, such as the Study of Women's Health Across the Nation (SWAN). More specifically, VMS last for a median period of 4.5 years following the menopause. The total duration of VMS depends on the time of onset, being longer in cases of its initial appearance during the premenopausal or early perimenopausal period (11.8 years on average) and shorter in women who were postmenopausal at the onset of VMS (median duration 3.4 years). The race and the age at onset seem to affect the duration of VMS since it is the longest in African-American women (median duration 10.1 years) and women younger than 45 years. Low educational level and higher depressive symptomatology are also associated with a longer duration [9]. Hot flushes appear as a sudden sensation of heat located at the face and upper chest, which lasts for 2–4 min. They may be aggravated by stress and anxiety, spicy food, or alcohol. VMS can rarely occur in older ages (more than 65 years) and may be associated with increased risk of cardiovascular disease (CVD). The neuroregulators kisspeptin, neurokinin B, and dynorphin, along with disorders in the thermoregulatory system, are implicated in the pathogenesis of VMS [1].

Another category of menopausal symptomatology involves the neuromuscular system, including headaches, loss of memory, dizziness, breathing difficulties, diffuse muscle and joint pain, and numbness. Skin dryness may also be present. Transition to menopause is also associated with a variety of psychogenic symptomatology, including irritability, mood swings, anxiety, depression, poor concentration, general loss of interest and decreased libido [1, 2]. Finally, genitourinary symptomatology, mainly present during the late postmenopausal stage [4], includes vaginal dryness, dyspareunia, vaginitis, dysuria, and recurrent urinary tract infections [1].

The decline in oestrogen concentrations is associated with long-term consequences, such as dysregulation in glucose and lipid metabolism, creating an atherogenic lipid profile [increase in total cholesterol, low-density lipoprotein cholesterol (LDL-C) and triglycerides, and decrease in high-density lipoprotein cholesterol (HDL-C) concentrations] [10], body fat redistribution (central adiposity), increase in systolic and diastolic blood pressure, and loss of bone mass [11, 12]. These metabolic disturbances predispose to an increased risk of CVD, type 2 diabetes mellitus (T2DM), osteoporosis, and fractures, especially in the case of coexistence with environmental and genetic factors [12]. Postmenopausal women with EM or POI seem to be at a 1.5 to 2-fold increased risk of CVD and T2DM [13].

Diagnosis

In women with an intact uterus, the diagnosis of menopause is retrospective, based on clinical criteria and established after completion of 12 months of amenorrhea. The diagnosis is more straightforward when cessation of menses occurs at an average age of 50–52 years, in the absence of surgery or medication. Further investigation, including inhibin B, AMH, FSH, and E2, does not confer any benefit in the diagnosis and management, except for cases seeking fertility [1]. A hormonal investigation, including FSH and serum E2 assessment, is recommended in cases of hysterectomy without bilateral oophorectomy. In these cases, persistently elevated FSH with low E2 concentrations (less than 20 pg/ml) support the diagnosis [1].

A potentially postmenopausal woman should be asked about the presence and duration of the symptoms mentioned earlier and the extent to which these affect her quality of life. The pattern of menstruation (cyclicity and length of menses), the age of menarche and the number (if any) of established pregnancies (including both live births and abortions) should be recorded. The medical history should include investigation for chronic diseases (such as T2DM, hypertension, and CVD events), surgical operations, fractures, current and past medications, smoking habits, alcohol use, and allergies.

The physical examination should accompany the medical history and include body mass index (BMI) assessment (calculated in kg/m^2), waist and hip circumference, blood pressure, and heart rate. The clinician should also search for signs associated with specific medical conditions, such as polycystic ovary syndrome (PCOS). The latter may augment the risk for T2DM and, potentially, CVD. Signs, such as acanthosis nigricans, hirsutism, alopecia, and acne, should be searched for and recorded. A physical and sonographic gynaecological examination is needed to exclude anatomical causes of menstrual disturbances and uterine disorders.

A hormonal investigation is necessary only in cases in which a definite diagnosis cannot be made by the medical history and physical examination, especially in women younger than 45 years. In such cases, a combination of increased FSH with decreased E_2 concentrations is confirmatory. Oligomenorrhoea or amenorrhoea for at least four months and FSH concentrations >25 IU/l on two occasions, at least 4–6 weeks apart, establish the diagnosis of EM or POI [2, 14]. In all cases of non-iatrogenic POI, obtaining a karyotype and testing for Y-chromosomal material (if positive, a gonadectomy should be performed) and FMR1 (fragile X mental retardation 1) gene mutations are recommended [14]. Test for autosomal gene mutations should be considered only in cases of suspicion of a specific syndrome (e.g. epicanthus inversus syndrome). In case of POI of unknown aetiology or an indication of autoimmunity, further investigation should include assessment of antithyroid and antiadrenal autoantibodies followed by thyroid stimulating hormone (TSH) and Synacthen test in case of positivity, respectively [14], as well as, testing for autoimmune polyendocrine syndromes [3]. In cases of iatrogenic menopause due to chemotherapy or radiotherapy, clinicians should take into consideration that a resumption in gonadal function may occur even 12 months after intervention [1].

Other causes of menstrual disturbances, such as hyperprolactinaemia, thyroid dysfunction and PCOS should be excluded. A hormonal profile should include assessment of TSH, prolactin, luteinizing hormone (LH), total testosterone and sex hormone binding globulin (SHBG) concentrations [3]. In case that more rare situations are suspected (non-classical adrenal hyperplasia, ovarian hyperthecosis, androgen-secreting tumours, Cushing's syndrome, acromegaly), laboratory investigation should include baseline concentrations of the involved hormones or the respective dynamic tests, as suggested by international guidelines to confirm or exclude the diagnosis [15].

Further examination could include mammography and bone mineral density (BMD) assessment, with dual-energy X-ray absorptiometry (DXA), ideally in both lumbar spine and femoral neck. Repeated DXA scans at 1–2-year intervals in cases of osteoporosis or osteopenia with increased fracture risk should be obtained [1, 2].

Differential Diagnosis

The differential diagnosis should include causes that mimic menopausal (mainly vasomotor) symptomatology, such as thyrotoxicosis, carcinoid syndrome, pheochromocytoma, acromegaly, mastocytosis, renal or medullary thyroid carcinoma, pancreatic islet cell tumours or lymphomas, which may also manifest with flushing, palpitations, and/or sweating. Therefore, assessment of TSH, urinary or serum metanephrines, urinary 5-hydroxyindoleacetic acid (5-HIAA), serum tryptase, insulin-like growth factor-1 (IGF-1) and calcitonin concentrations, as well as, an abdominal ultrasound may be considered in cases of ambiguous diagnosis. Furthermore, medications [such as opioids and their withdrawal, selective serotonin reuptake inhibitors (SSRIs), nicotinic acid, calcium channel blockers (CCBs)], dietary factors (such as alcohol, spicy food, and food additives) and chronic infections should be excluded [1].

Special consideration should be given to women with hysterectomy without bilateral oophorectomy. In such cases and in the presence of vasomotor symptomatology, assessment of pituitary-gonadal axis, with measurement of FSH and serum E_2, as well as prolactin and TSH, will elucidate the aetiology of menstrual disturbances. This assessment is also suggested in cases of a pre-existing history of irregular cycles, in which the pattern of menstruation cannot contribute to the diagnosis of menopause. Women on oral contraceptive pills (OCPs) constitute a challenging case. In these, withdrawal of the OCP and hormonal assessment on day 3–5 of a spontaneous menstrual cycle may unmask potential menopausal transition.

It is of utmost importance to differentiate postmenopausal pattern of menstruation from other causes of abnormal uterine bleeding, such as coagulopathies, fibroids, polyps, adenomyosis, endometrial carcinoma, and endometrial intraepithelial neoplasia. In these cases, transvaginal ultrasonography is a useful diagnostic tool. When the diagnosis is obscure, further evaluation with hysteroscopy, plus dilation and curettage, is recommended [16]. In contrast, menopause-associated uterine bleeding is the result of the unopposed effect of E_2 due to anovulation, causing significant proliferation and thickening of the endometrium [17].

Medical Management

Hormone Replacement Therapy

Hormone replacement therapy (HRT) is the treatment of choice for women with menopausal symptoms and urogenital atrophy.

Additionally, HRT is indicated for the treatment of osteoporosis and prevention of fracture, but only for women who present climacteric symptoms. There has been some concern about possible adverse effects of HRT regarding increased risk of breast cancer, especially with the use of prolonged combined therapy with certain progestins, and venous thromboembolism (VTE), especially with the use of oral oestrogens. In general, however, the benefits of HRT outweigh these risks, particularly for women under 60 years of age or for those who are in menopause for less than ten years. HRT is the most effective treatment for menopausal symptoms, while many of the regimens given systematically, and all of the local applications are effective in treating vaginal atrophy, vaginal dryness, bruising, dyspareunia, and pruritus. HRT can also positively affect mood disorders and sleep disturbances. Furthermore, it exerts beneficial effects on bone metabolism, as standard dose regimens reduce the incidence of fragility fractures, while low or very low dose regimens improve bone density and bone turnover markers. HRT has been associated with improvement of cardiometabolic risk factors too and, when given in women younger than 60 years of age or women who are in menopause for less than ten years, protects them against CVD. The use of HRT has also been associated with a decrease in the risk of colon and rectal cancer, while there is some evidence for a reduction in risk of Alzheimer's disease [18–20].

HRT Regimens

Women after hysterectomy should receive formulations with oestrogens only, while for women with intact uterus progestogen needs to be added, to counteract the possible harmful effects of oestrogen on the endometrium. A combination of conjugated oestrogen and bazedoxifene has been introduced in Europe for use in women with an intact uterus that have contra-indications or are intolerant in progestogens. Alternatively, tibolone is a synthetic steroid with oestrogenic, androgenic and progestogenic activities and can be used in women with and without hysterectomy [21–24]. Table 8.5.2.1 presents the available types of oestrogen formulations according to the route of administration.

Table 8.5.2.1 Oestrogenic formulations of HRT and route of administration

Oral medications	
• 17β-E_2 • Steroidized oestrogens • Tibolone • Conjugated equine oestrogens • Oestrogen-progestogen combinations • Conjugated oestrogen/bazedoxifene	Once per day
Patches	
• 17β-E_2 • 17β-E_2/norethisterone	Twice per week
Gel	
• 17β-E_2	Once per day
Vaginal applications	
• 17β-E_2 • Estriol	Tablets or creams daily for 15–20 days, then 1–2 per week

17β-E_2:17β-oestradiol.

The dosage should be personalized according to the therapeutic goals, the age of woman, and the years after menopause. In Table 8.5.2.2, a classification of oestrogen dosages and indications is presented.

The progestogens used in HRT are distinguished in natural human progesterone and synthetic forms. Among the most common used progestogens are microsomal progesterone, norethindrone acetate (NETA), dydrogesterone, drospirenone, and medroxyprogesterone acetate. Concerning the mode of administration, oral progestogens are administered once or twice daily constantly, resulting in amenorrhea, or periodically, 12–14 days per month, resulting in regular monthly bleeding. Transdermal progestogens (norethisterone) are administered twice per week in a continuous manner, leading to amenorrhea, or in a cyclical manner, for 14 days per month, leading to regular monthly bleeding. Vaginal progesterone can be used once a day before bedtime in a continuous way leading to amenorrhea, or in a periodical manner, 12–14 days per month, leading to regular monthly bleeding. An intrauterine device that releases levonorgestrel can also be used. This device remains active for 3–5 years and provides contraception during perimenopause, while it can be used in combination with oestrogens as proper HRT. In the case that libido is not restored after adequate replacement, low-dose testosterone therapy may be effective. Even if virilizing effects with low-dose and proper follow-up are rare, the long-term effects of testosterone treatment in women, particularly regarding breast cancer risk, are not known [18, 25–27]. In Table 8.5.2.3, the recommended dosages of progestogens according to oestrogens doses are presented.

Follow-Up of Women on HRT

Initial monitoring of women receiving HRT should be performed 2–3 months after the introduction of treatment to evaluate the efficacy and possible adverse effects. Persistence of menopausal symptoms usually requires an increase of oestrogen dose. Women on HRT often complain of unexpected vaginal bleeding, breast tenderness, flatulence, fluid retention, headaches, and mood changes. Symptoms are more common in women with prolonged time in menopause but tend to decline over the time of treatment. In that case, treatment modifications can include a decrease in the dose of oestrogens, change of the type of progestogen or change of the route of administration. When the woman is satisfied by HRT, follow-up and monitoring should be performed on an annual basis. Then, the cost-to-benefit ratio should be assessed, while the dose of oestrogens should be redefined. The duration of treatment depends on the patient characteristics and preferences, the individualized cost-to-benefit analysis and the oestrogens dose used [27–29].

HRT and CVD

HRT has been associated with a favourable effect on the cardiometabolic profile by improving serum lipids concentrations, body fat composition, glucose metabolism, insulin resistance and vascular function [17, 30]. However, primary and secondary analyses of the Women's Health Initiative (WHI) study have shown that the exact effects of HRT on CVD risk largely depend on the age of the women. Younger menopausal women aged 50–59 years or women who are within the first ten years after menopause present, in general, lower risk of CVD when on HRT, while initiation of HRT at an older age (more than 60 years of age or more than

Table 8.5.2.2 Dosages and indications of oestrogens

Dosage	Indication	Oral oestrogens	Transdermal 17β-E_2	Tibolone
Standard dose	Premature ovarian insufficiency	17β-E_2 2–4 mg or CEE 0.625–1.25 mg	50–100 µg	-
	Perimenopause	17β-E_2 2 mg or CEE 0.625 mg	50 µg	-
	Menopause: poor response at a low dose	17β-E_2 2 mg or CEE 0.625 mg	50 µg	2.5 mg
Low dose	Perimenopause	17β-E_2 1 mg or CEE 0.300–0.450 mg	25–37 µg	-
	Menopause	17β-E_2 1 mg or CEE 0.300–0.450 mg	25–37 µg	2.5 mg
Ultra-low-dose	Menopause: starting dose or maintenance dose after initial control of symptoms at low dosage	17β-E_2 0.25–0.5 mg	14 µg	1.25 mg

17β-E_2, 17β-oestradiol; CEE, conjugated equine oestrogens.

10 years after the menopause) can increase the CVD risk. The findings of two additional landmark studies, Kronos Early Estrogen Prevention Study (KEEPS) and Early versus Late Intervention Trial with Estradiol (ELITE) trial, support the notion of a 'window of opportunity', demonstrating a cardioprotective activity of HRT, if they are initiated close to the time of menopause, when women still present good vessel health and low prevalence of atherosclerosis. Therefore, all postmenopausal women should be individually assessed for CVD, in order a decision for HRT initiation to be taken [18, 28–32]. Table 8.5.2.4 presents a template of individualization of HRT according to CVD risk.

HRT and Venous Thromboembolism

Treatment with standard oral oestrogen doses, either as monotherapy or in combination with progestogen, has been associated with the development of VTE [31–32]. Large epidemiological studies have shown that the risk of VTE does not increase with administration of transdermal oestrogens in contrast to the administration of oral regimens. Smaller doses of oral oestrogens may have less adverse effects. The addition of a progestogen increased the risk for VTE significantly, but this risk is highly associated with the type of progestogen. Natural progesterone appears to have a favourable profile, while derivatives of norpregnane and medroxyprogesterone acetate have been associated with an increased risk for VTE. The absolute risk for VTE is small in the age group of 5–60 years of age. However, the presence of one or more additional risk factors can increase it. Healthcare providers should clinically assess the risk for VTE in each patient individually, before initiation of HRT. Genetic screening for thrombogenic mutations is not recommended for all women, but individuals in high risk due to personal or family history of thromboembolic events may profit from such an approach. Among other thrombogenic mutations, prothrombin and Factor V Leiden mutations have been associated with a higher risk for VTE [27, 33–34]. Recommendations regarding the individualization of HRT according to the risk for VTE are presented in Box 8.5.2.1.

HRT and Breast Cancer

Combined HRT with oestrogen and progestogen has been associated with a small increase in breast cancer risk in middle-aged women. This risk is mainly attributed to the effects of progestogen, as oestrogen monotherapy for seven years in the WHI study did not increase the risk of breast cancer. Major epidemiological studies have provided evidence that natural progesterone and dydrogesterone present a more favourable safety profile concerning breast cancer. Furthermore, cyclical HRT may have a more favourable risk profile compared to continuous combined regimens, probably due to the lower cumulative exposure of the breast cells to the progestogen. From a clinical point of view, a detailed history can help in identifying women at high risk for breast cancer [17, 24, 26, 29, 31]. Nevertheless, these women represent the minority and, therefore, HRT can be safely administered in the majority of postmenopausal women after evaluation for breast cancer risk factors (Box 8.5.2.2).

Non-Hormonal Treatment of Menopausal Symptoms

Alternative Drug Therapies

Non-hormonal treatment options can be used for the treatment of hot flashes for women who are not good candidates for HRT. Clonidine, selective serotonin, and norepinephrine reuptake

Table 8.5.2.3 Dosages of progestogens according to oestrogens doses

Oestrogen dose	Natural progesterone	Dydrogesterone	Norethisterone	Drospirenone
Standard dose	200 mg	10–20 mg	*Per os* 1–5 mg *Transdermal* 0.25 mg	–
Low dose	100 mg	5–10 mg	*Per os* 0.5 mg *Transdermal* 0.125 mg	2 mg
Ultra-low-dose	50 mg	2.5–5 mg	*Per os* 0.25 mg *Transdermal* 0.063 mg	1 mg

Table 8.5.2.4 Individualization of HRT according to CVD risk

High CVD risk	
• >60 years old • >10 years in menopause • Long-term uncontrolled T2DM • Clinically evident CVD	No HRT
Intermediate CVD risk	
• 5–10 years in menopause • Dyslipidaemia • Central obesity • Smoking • Hypertension • T2DM	Transdermal HRT
Low CVD risk	
• Peri- or recently postmenopausal • None of the above factors present	Any HRT

HRT, hormone replacement treatment; CVD, cardiovascular disease, T2DM, type 2 diabetes mellitus.

inhibitors (SNRIs) and SSRIs can be effective for the treatment of vasomotor symptoms. Clonidine presents alpha-2 agonist properties and has been associated with a decrease in hot flushes after an average time of three months, obtaining approval for this indication in some countries. Among SSRIs, paroxetine has been shown to be most effective for the vasomotor symptoms. Venlafaxine and desvenlafaxine are the most commonly administered SNRIs in women who cannot take HRT. Gabapentin also has beneficial effects, but adverse effects, such as severe drowsiness, can limit its use in everyday clinical practice [26, 35].

Non-Pharmaceutical Alternative Therapies

Although topical low-dose oestrogen therapy is the treatment of choice for the symptoms of urogenital atrophy, an alternative approach is the use of non-hormonal products that hydrate the vaginal tissues. Long-acting vaginal moisturizers are a confirmed first-line treatment for women with complications of vaginal atrophy. The treatment is usually given as a daily application booster for 2 to 3 weeks and then as maintenance treatment with one to two applications per week. Dyspareunia usually requires additional application of water-based lubricants, silicone, or oils before sexual contact in combination with the regular application of long-acting vaginal moisturizers [36–37].

Phytoestrogens are substances produced by plants with a chemical structure similar to that of human E_2. The four main categories include coumestans, prenylflavonoids, isoflavones, and lignans. Phytoestrogens can interact with both types of oestrogen receptors, type α and β. The effectiveness of the individual substances depends on their intestinal absorption and metabolism, the affinity to oestrogen receptors and their oestrogenic activity. The evidence is indicative of beneficial properties resulting from the administration of phytoestrogens, but further research is needed. Studies evaluating the effect of phytoestrogens have shown some benefit concerning the frequency of hot flashes and the degree of vaginal dryness, but the effect was inferior to that resulting from HRT. Therefore, phytoestrogens represent an alternative suggestion for women, who do not wish or cannot take HRT for menopausal symptoms [35–37].

Meditation, cognitive behavioural therapy, relaxation techniques, and controlled breathing, acupuncture, and homoeopathy have also been used to treat postmenopausal symptoms. Data regarding the efficacy and safety of these approaches are sparse so that no evidence-based recommendations can be made [18, 35].

Lifestyle Changes

Many consequences of the menopause can be positively affected by lifestyle changes, such as optimal diet, increasing physical activity, giving up smoking, and decreasing alcohol consumption. Weight loss should be achieved through specific clinical nutrition recommendations, such as consumption of mono- and polyunsaturated fat rather than saturated forms, reduction in the total amount of carbohydrates and preference to those deriving from fruits and whole grains, as well as an increase of protein intake deriving from fish, poultry, or skimmed dairy products. Smoking represents an important risk factor for many diseases associated with menopause and ageing, such as CVD, osteoporosis, and cancer, while adversely affects the effectiveness of osteoporosis treatment. Therefore, smoking discontinuation should be an

Box 8.5.2.1 Individualization of HRT according to VTE risk

Risk factors
- Personal history of VTE
- Family history of VTE
- >60 years old
- >10 years in menopause
- Obesity
- T2DM
- Smoking
- Immobilization

Risk factors present

No HRT

or

Very-low or low-dose transdermal 17β-E_2 in combination with natural progesterone

No risk factors

Any HRT

17β-E_2: 17β-oestradiol, HRT: hormone replacement treatment; VTE: venous thromboembolism; T2DM: type 2 diabetes mellitus.

Box 8.5.2.2 Individualization of HRT according to breast cancer risk

Risk factors
- Personal history of breast cancer
- Family history of breast cancer
- Previous surgery in the breast for a benign reason
- Obesity
- Non-parous or increased age of parity
- Alcohol consumption
- High density in mammography

HRT individualization
- Use of oestrogen in dose inversely proportional to age
- Natural progesterone or dydrogesterone
- Cyclical mode of administration
- Annual cost-to-benefit analysis and dose adjustment

essential part of routine clinical consultation of women in menopause. Physical exercise can prevent weight gain and decreased muscle mass, improve bone quality, decrease T2DM, and CVD incidence and improve psychological status. Postmenopausal women should be encouraged for regular aerobic physical activity, spending at least 150 min per week of average exercise intensity or at least 75 min per week of high-intensity exercise. Anaerobic activities targeting and strengthening major muscle groups should also be performed at least twice per week. Menopausal women who cannot achieve the aforementioned recommendations for exercise should be advised to be as active as possible, according to their abilities [18, 38].

Postmenopausal Osteoporosis Treatment

The recommended calcium and vitamin D daily intake in women older than 50 years with risk of fracture is 1000–1200 mg and 600–800 IU, respectively. Calcium and vitamin D sufficiency improves bone quality and mineralization, prevents the development of secondary hyperparathyroidism, and regulates bone turnover. Furthermore, the maximum efficacy of antiosteoporotic treatment is only achieved in cases of adequate calcium and vitamin D intake. Therefore, women in menopause should be encouraged to consume dairy products or other calcium-rich foods. Alternatively, dietary supplements should be recommended. Vitamin D supplementation therapy should be adjusted at doses which will achieve 25-hydroxyvitamin D [25(OH)D)] concentrations of at least 30 ng/ml. If osteoporotic fracture or osteoporosis is present or the risk for fracture, as assessed by the FRAX score, is increased, antiosteoporotic treatment should be added according to international guidelines. Selective oestrogen receptor modulators (SERMs, raloxifene, and bazedoxifene), bisphosphonates, denosumab, and teriparatide are the currently approved agents for the treatment of postmenopausal osteoporosis [39].

Contraception in Perimenopause and Late Parenthood

Women experiencing perimenopausal menstrual cycle disorders tend to believe that they are no longer fertile and often stop the use of contraception. However, the annual pregnancy probability for women aged 40–44 years is approximately 30% and for women aged 45–49 years is approximately 10%. Indeed, sporadic ovulation may occur even in women with non-regular vaginal bleeding. Therefore, in women who are at risk of an unwanted pregnancy, contraception should stop only two years after the last menstrual episode, if they are less than 50 years, and one year after the last menstrual episode, if they are older than 50 years. The criteria for the appropriate use of contraceptives in women over 40 years of age are set according to their underlying health condition. All currently available contraceptive methods can be given as choices to women older than 45 years, but physicians should weigh the decision depending on the age-related CVD risk, smoking status, or other complications. Women older than 40 years without particular CVD risk factors can take safely hormonal contraception. With regard to oral contraceptives, formulations with second-generation progestogens should be preferred, as these have a better profile regarding VTE. Especially in the perimenopausal population, oral administration of contraceptives is associated with non-specific benefits, such as the restoration of menstrual cycle, the treatment of bleeding problems and menopausal symptoms, the reduction of risk of ovarian and endometrial cancer, as well as bone loss prevention. In women with CVD risk factors, such as obesity, dyslipidaemia, smoking, hypertension, or T2DM, the use of contraceptives should be weighed against the potential CVD risk; in most cases, they should be avoided [40].

In case of desire for late parenthood, the perimenopausal woman should be advised that advanced age is associated not only with infertility but also with pregnancy complications and may have an impact on the long-term health of the offspring. Appropriate hormonal assessment and therapeutic approach can assist in the management of infertility. An experienced multidisciplinary team is needed in the case of such a pregnancy [41].

Conclusions

Menopause represents the end of reproductive life in women and is associated with a substantial decrease in endogenous oestrogen production. These hormonal changes have various metabolic, bone, CVD, and psychological effects. HRT is the treatment of choice for women with menopausal symptoms and urogenital atrophy. In general, the benefits of HRT outweigh possible risks, such as breast cancer and VTE risk, particularly for women under 60 years of age or for those who are in menopause for up to 10 years. Postmenopausal women should also be managed with lifestyle interventions, including diet, exercise, smoking cessation, and decrease in alcohol consumption, with a particular focus on bone and cardiovascular health. Alternative drug or non-pharmaceutical therapies may contribute to the alleviation of postmenopausal symptoms. The ultimate goal of the healthcare providers should be an integrated plan for the postmenopausal woman, bearing in mind that most women will spend more than one-third of their lifespan in menopausal status.

REFERENCES

1. Stuenkel CA, Davis SR, Gompel A, *et al.* Treatment of symptoms of the menopause: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2015; **100**: 3975–4011.
2. Baber RJ, Panay N, Fenton A. 2016 IMS Recommendations on women's midlife health and menopause hormone therapy. *Climacteric*, 2016; **19**: 109–50.
3. Vujovic S, Brincat M, Erel T, *et al.* EMAS position statement: managing women with premature ovarian failure. *Maturitas*, 2010; **67**: 91–3.
4. Harlow SD, Gass M, Hall JE, *et al.*; STRAW +10 Collaborative Group. Executive summary of the Stages of Reproductive Aging Workshop +10: addressing the unfinished agenda of staging reproductive aging. *J Clin Endocrinol Metab*, 2012; **97**: 1159–68.
5. Santoro N, Randolph JF Jr. Reproductive hormones and the menopause transition. *Obstet Gynecol Clin North Am*, 2011; **38**: 455–66.
6. Te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update*, 2002; **8**: 141–54.
7. Ledger WL. Clinical utility of measurement of anti-Müllerian hormone in reproductive endocrinology. *J Clin Endocrinol Metab*, 2010; **95**: 5144–54.

8. Greene JG. Constructing a standard climacteric scale. *Maturitas*, 1998; **29**: 25–31.
9. Avis NE, Crawford SL, Greendale G, *et al.* Duration of menopausal vasomotor symptoms over the menopause transition. *JAMA Intern Med*, 2015; **175**: 531–9.
10. Anagnostis P, Stevenson JC, Crook D, Johnston DG, Godsland IF. Effects of menopause, gender and age on lipids and high-density lipoprotein cholesterol subfractions. *Maturitas*, 2015; **81**: 62–8.
11. Lovejoy JC, Champagne CM, de Jonge L, Xie H, Smith SR. Increased visceral fat and decreased energy expenditure during the menopausal transition. *Int J Obes (Lond)*, 2008; **32**: 949–58.
12. Qiu C, Chen H, Wen J, *et al.* Associations between age at menarche and menopause with cardiovascular disease, diabetes, and osteoporosis in Chinese Women. *J Clin Endocrinol Metab*, 2013; **98**: 1612–21.
13. Tao XY, Zuo AZ, Wang JQ, Tao FB. Effect of primary ovarian insufficiency and early natural menopause on mortality: a meta-analysis. *Climacteric*, 2016; **19**: 27–36.
14. European Society for Human Reproduction and Embryology (ESHRE) Guideline Group on POI, Webber L, Davies M, Anderson R, Bartlett J, Braat D, Cartwright B, Cifkova R, de Muinck Keizer-Schrama S, Hogervorst E, Janse F, Liao L, Vlaisavljevic V, Zillikens C, Vermeulen N. ESHRE Guideline: management of women with premature ovarian insufficiency. *Hum Reprod*, 2016; **31**: 926–37.
15. Paschou SA, Anagnostis P, Goulis DG, Lambrinoudaki I. Androgen excess and post-reproductive health. *Maturitas*, 2018; **115**: 115–16.
16. ACOG Committee Opinion No. 734: the role of transvaginal ultrasonography in evaluating the endometrium of women with postmenopausal bleeding. *Obstet Gynecol*, 2018; **131**: e124–9.
17. Whitaker L, Critchley HO. Abnormal uterine bleeding. *Best Pract Res Clin Obstet Gynaecol*, 2016; **34**: 4–65.
18. Armeni E, Lambrinoudaki I, Ceausu I, *et al.* Maintaining postreproductive health: a care pathway from the European Menopause and Andropause Society (EMAS). *Maturitas*, 2016; **89**: 63–72.
19. de Villiers TJ, Hall JE, Pinkerton JV, *et al.* Revised global consensus statement on menopausal hormone therapy. *Maturitas*, 2016; **91**: 153–5.
20. Mørch LS, Lidegaard Ø, Keiding N, Løkkegaard E, Kjær SK. The influence of hormone therapies on colon and rectal cancer. *Eur J Epidemiol*, 2016; **31**: 481–9.
21. Komm BS, Mirkin S, Jenkins SN. Development of conjugated estrogens/bazedoxifene, the first tissue selective estrogen complex (TSEC) for management of menopausal hot flashes and postmenopausal bone loss. *Steroids*, 2014; **90**: 71–81.
22. Valera MC, Gourdy P, Trémollières F, Arnal JF. From the Women's Health Initiative to the combination of estrogen and selective estrogen receptor modulators to avoid progestin addition. *Maturitas*, 2015; **82**: 274–7.
23. Biglia N, Maffei S, Lello S, Nappi RE. Tibolone in postmenopausal women: a review based on recent randomised controlled clinical trials. *Gynecol Endocrinol*, 2010; **26**: 804–14.
24. Lambrinoudaki I. Progestogens in postmenopausal hormone therapy and the risk of breast cancer. *Maturitas*, 2014; **77**: 311–17.
25. Utian WH, Archer DF, Bachmann GA, *et al.* Estrogen and progestogen use in postmenopausal women: July 2008 position statement of The North American Menopause Society. *Menopause*, 2008; **15**: 584–602.
26. Panay N, Hamoda H, Arya R, Savvas M; British Menopause Society and Women's Health Concern. The 2013 British Menopause Society & Women's Health Concern recommendations on hormone replacement therapy. *Menopause Int*, 2013; **19**: 59–68.
27. Baber RJ, Panay N, Fenton A; IMS Writing Group. 2016 IMS Recommendations on women's midlife health and menopause hormone therapy. *Climacteric*, 2016; **19**: 109–50.
28. Kling JM, Lahr BA, Bailey KR, Harman SM, Miller VM, Mulvagh SL. Endothelial function in women of the Kronos Early Estrogen Prevention Study. *Climacteric*, 2015; **18**: 187–97.
29. Manson JE, Chlebowski RT, Stefanick ML, *et al.* Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *JAMA*, 2013; **310**: 1353–68.
30. Schenck-Gustafsson K, Brincat M, Erel CT, *et al.* EMAS position statement: managing the menopause in the context of coronary heart disease. *Maturitas*, 2011; **68**: 94–7.
31. Anderson GL, Limacher M, Assaf AR, *et al.* Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*, 2004; **291**: 1701–12.
32. Rossouw JE, Anderson GL, Prentice RL, *et al.* Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, 2002; **288**: 321–33.
33. Canonico M, Fournier A, Carcaillon L, *et al.* Postmenopausal hormone therapy and risk of idiopathic venous thromboembolism: results from the E3N cohort study. *Arterioscler Thromb Vasc Biol*, 2010; **30**: 340–5.
34. Trémollières F, Brincat M, Erel CT, *et al.* EMAS position statement: managing menopausal women with a personal or family history of VTE. *Maturitas*, 2011; **69**: 195–8.
35. Mintziori G, Lambrinoudaki I, Goulis DG, *et al.* EMAS position statement: non-hormonal management of menopausal vasomotor symptoms. *Maturitas*, 2015; **81**: 410–13.
36. Rees M, Pérez-López FR, Ceasu I, *et al.* EMAS clinical guide: low-dose vaginal estrogens for postmenopausal vaginal atrophy. *Maturitas*, 2012; **73**: 171–4.
37. Moreira AC, Silva AM, Santos MS, Sardão VA. Phytoestrogens as alternative hormone replacement therapy in menopause: what is real, what is unknown. *J Steroid Biochem Mol Biol*, 2014; **143**: 61–71.
38. Lambrinoudaki I, Ceasu I, Depypere H, *et al.* EMAS position statement: diet and health in midlife and beyond. *Maturitas*, 2013; **74**: 99–104.
39. Kanis JA, McCloskey EV, Johansson H, *et al.* European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporos Int*, 2013; **24**: 23–57.
40. Ruan X, Mueck AO. Oral contraception for women of middle age. *Maturitas*, 2015; **82**: 266–70.
41. Mintziori G, Lambrinoudaki I, Kolibianakis EM, *et al.* EMAS position statement: late parenthood. *Maturitas*, 2013; **76**: 200–4.

8.6

Female Infertility

8.6.1 Female Infertility and Assisted Reproduction

Adam H. Balen and Susie Jacob

Introduction

Infertility is common. It has recently been suggested that approximately 9% of couples are involuntarily childless although the exact number inevitably depends on how the complaint is defined [1]. Medical definitions of infertility tend to emphasize the immediate problem brought to the consultation, reflecting the typically short-term interaction of many doctors, particularly specialists, with their patients. Most accepted definitions therefore involve the number of months prior to the consultation during which the couple has been exposed to the chance of a pregnancy. When the lifetime experience of a couple's attempt to raise a family is considered, a quite different picture emerges: at least a quarter of all couples experience unexpected delays in achieving their desired family size [2, 3], although only a half may seek treatment [3].

The single most important determinant of a couple's fertility is the age of the female partner. For women up to and including the age of 25 the cumulative conception rate is 60% at 6 months and 85% at a year. For couples where the female partner is 35 years of age or older, the conception rates are 60% at a year and 85% at 2 years. Women are born with a finite complement of eggs, which do not undergo further cell division until just after fertilization. Thus, an oocyte ovulated today is pretty well the same age as the woman from whose ovary it came. Even DNA, the most stable molecule in biology, is not completely invulnerable to the passage of years; this impact of age on oocytes is consistent with its effect on the risk of congenital abnormalities, well known in many cases to increase with maternal age.

When one refers to a patient as being 'infertile' one is referring to a slow rate of conception—infertility is rarely absolute; indeed some prefer the expression 'subfertility'. If, despite a regular menstrual cycle and a normal sex life, pregnancy has not occurred by 12 months, most authorities would accept that that couple has a fertility problem and would offer investigation and treatment. If there is a history of a menstrual disturbance, pelvic inflammatory disease, or in the male partner an attack of orchitis or a history of cryptorchidism, investigation should begin sooner rather than later.

A more difficult problem is defining infertility in the couple with an older female partner. In one way, one might consider delaying investigation because it takes longer for a woman of 35 years and older to achieve a particular conception rate. On the other hand, the slope of the line relating the risk of childlessness to age gets much steeper as one approaches the age of 40. Furthermore, the prospects of achieving a pregnancy with treatment is parallel to this curve. There is therefore little time to lose and we are more active in advising investigation and treatment. There seems little point in waiting beyond a year and in many women (particularly those with some diagnostic clue in their history) we recommend initiating investigation after 6 months of unprotected intercourse.

A review published in 2007 has examined the collective prevalence of infertility from 25 population surveys (of 172 413 women) from around the world [1] there was a wide range of infertility rates from 3.5% to 16.7% in developed countries and 6.9% to 9.3% in less developed countries, with a median overall prevalence of 9%, which equates to over 70 million women worldwide. Overall approximately 56% (range 27–76%), that is 40 million couples seek medical care although only an estimated 22% receive care [1]. In 1993 and 1995 two surveys in England of 2377 and 728 women reported prevalence rates of 26.4% and 17.3%, respectively, of whom 50–61% sought assistance [4, 5].

According to the UK Government Statistical Services there is a steadily rising proportion of women in the UK who have never had a child. The mean age of mothers at childbirth fell from 28.7 years for women born in 1920 to a low of 26.0 years for women born in the mid-1940s (**Figure 8.6.1.1**) [6]. Women born in the 1940s had the lowest average age at childbirth contributing to the 1960s 'baby boom', when family size was also larger [7].

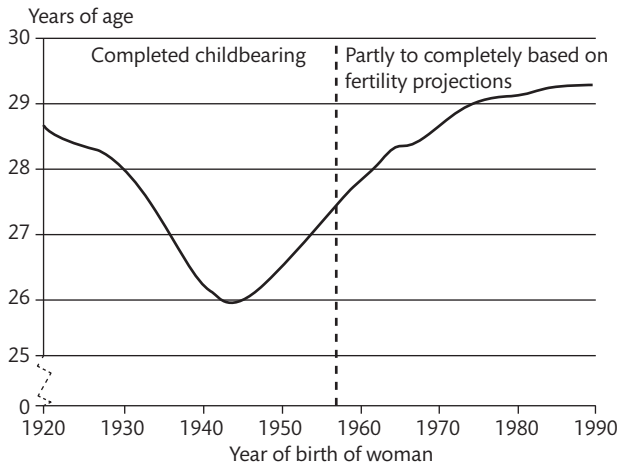


Figure 8.6.1.1 The mean age of UK mothers at first childbirth.

The proportion of childless women reaching the end of their childbearing years continues to rise. The most recent data [8] for those born in 1971, found that 18% are childless by 45 years compared with 11% for those born in 1944 who are presumed to be their mothers' cohort [8]. The average family size for the 1971 cohort is 1.9 children, the lowest level recorded. A trend to later childbearing continues, shown by those who are childless by the age of 30 years; 48% of those born in 1986 are childless at their 30th birthday compared with 44% born in 1971. The secular change in delayed childbirth relates to a number of factors, including contraceptive usage, the desire for a career, rising rates of sexually transmitted infections and the decline in fertility with age.

Obesity

Obesity is a common problem among women of reproductive years, with 56% of women in the UK being either overweight or obese. Obesity has a negative impact on spontaneous conception, miscarriage, pregnancy and the long-term health of both mother and child due to both an increased rate of congenital anomalies and the possibility of metabolic disease in later life. Obesity also has a negative impact on male fertility. Women who are obese respond less well to drugs that are used for ovarian stimulation for the treatment of both anovulation and assisted conception, although this does not always equate with a reduction in ongoing pregnancy rates. Furthermore, obesity may affect the safety of procedures, for example the ability to see ovaries on ultrasound scan or the provision of safe anaesthesia for laparoscopy or oocyte retrieval. Obesity also has a major impact during pregnancy and at delivery.

A normal body mass index (BMI) is considered to be 19–24.9 kg/m^2, although some would consider the lower limit of normal to be 20 kg/m^2. Being underweight leads to hypothalamic amenorrhoea and increases risk to pregnancy if conception does occur. Overweight is defined as a BMI of ≥25 kg/m^2 (World Health Organization (WHO) definitions) with 'pre-obese' being 25.0–29.9 kg/m^2, moderate obesity (class I) 30.0–34.9 kg/m^2, severe obesity (class II) 35.0–39.9 kg/m^2, and very severe ('morbid') obesity (class III) ≥ 40 kg/m^2.

Miscarriage rates appear to be increased with increasing maternal weight [9]. In those who conceive spontaneously there is an increased risk of miscarriage in those who are moderately overweight (BMI 25–27.9 kg/m^2). This has also been demonstrated in those who conceive by IVF [10] or who are recipients of donated oocytes. Pregnancy carries significant risks for those who are obese with increased rates of congenital anomalies (neural tube (OR 3.5), omphalocele (OR 3.3) and cardiac defects (OR 2.0)), miscarriage, gestational diabetes, hypertension, and problems during delivery [11]. The risks of congenital anomalies appear real, although there are also technical difficulties in assessing the fetus by ultrasound because adipose tissue attenuates the signal. Pregnancy itself exacerbates any underlying insulin resistance and as a result women with PCOS and/or obesity have an increased risk of gestational diabetes.

Obesity is associated with an increased risk to the mother during pregnancy. Risks include increased incidence of hypertension, pre-eclampsia, gestational diabetes, and thromboembolic disorders as well an increased caesarean section rate. Obesity contributes significantly to the risk of maternal mortality [9]. Macrosomia, admission to neonatal intensive care, birth defects, stillbirth, and perinatal death are all increased in the infants of women who are obese.

Polycystic ovary syndrome (PCOS) affects 20–25% of women and the prevalence appears to be rising because of the current epidemic of obesity [12]. PCOS accounts for 90–95% of women who attend infertility clinics with anovulation. At least 40% of women with PCOS are obese and they are more insulin resistant than weight-matched individuals with normal ovaries. Increasing abdominal obesity is correlated with reduced menstrual frequency and fertility together with greater insulin resistance. Several studies have shown that weight loss in women with PCOS improves the endocrine profile, menstrual cyclicity, rate of ovulation, and likelihood of a healthy pregnancy. Even a modest loss of 5–10% of total body weight can achieve a 30% reduction of central fat, an improvement in insulin sensitivity and restore ovulation. Lifestyle modification is clearly a key component for the improvement of reproductive function for overweight, anovulatory women with PCOS [9].

Weight loss should therefore be encouraged prior to ovulation induction treatments, such as clomiphene citrate or gonadotropin therapy, both to improve the likelihood of ovulation and enhance ovarian response. Monitoring treatment is also harder in the obese as visualization of the ovaries is more difficult which raises the risk of multiple ovulation and multiple pregnancy. National guidelines in the UK for the management of overweight women with PCOS advise weight loss, preferably to a BMI of <30 kg/m^2 prior to commencing drugs for ovarian stimulation [9]. The British Fertility Society suggests that ideally women should not commence assisted conception treatments until they have reduced their BMI to less than 35 kg/m^2, but if time is on their side (e.g. under the age of 35 years with normal ovarian reserve, they should aim for a BMI of less than 30 kg/m^2) [9].

Investigating Infertility

Fertility investigations should normally be instigated as soon as the couple seeks help. Even if they have been trying for less than a year, it is worthwhile asking some general questions to ensure that major problems, such as irregularities of the menstrual cycle, a history of pelvic surgery or orchidopexy have not been ignored. If the couple's medical history is normal the expected cumulative chance of conception over a period of time should be explained and investigations deferred until they have been trying for a year. When the

female partner is aged 35 years or older, monthly fecundity is significantly reduced but we do not believe that investigations should be delayed proportionately because of the concomitant age-related decline in the success of treatment.

Once the decision has been taken to investigate a couple it should be possible to perform the basic screening tests within 2–3 months and provide them with a management plan, which may involve reassurance, more detailed investigations, or treatment. A pragmatic approach should be taken. Infertility is rarely absolute and treatment options may be discussed to enhance a couple's fertility even in the absence of a clear diagnosis.

General Investigations

The fertility clinic should be used for general health screening and preconception counselling. Particular attention should be paid to body weight, blood pressure, urinalysis, cervical cytology, and rubella immunity. Some clinics ascertain hepatitis B, C, and HIV status before offering assisted conception—this has become routine practice in the UK because of the putative risk of viral contamination of cryopreserved embryos via liquid nitrogen.

It is important to perform a general physical examination and a pelvic examination should be performed. Endometriosis is suggested by the presence of nodules in the vagina, thickening of the posterior fornix, tenderness, and fixity of the pelvic organs. If the examination is painful one should be alerted to the possibility of pelvic pathology and include a laparoscopy early in the course of investigations. Adnexal masses should be investigated by ultrasound in the first instance.

Chlamydia Screening

A controversial subject is the routine swabbing of the cervix for *Chlamydia trachomatis*. Chlamydial DNA has been recovered from 50% of women with tubal infertility compared with approximately 12% in pregnant women or women with non-tubal infertility. Chlamydia infection is the commonest cause of tubal infertility in developed countries and is the commonest sexually transmitted pathogen in the UK. It is thought that at least 1 in 20 women in the UK between the ages of 18–25 years may have undiagnosed infection. *Chlamydia trachomatis* causes urethritis and epididymitis in men and cervicitis, salpingitis, and endometritis in women, although symptoms can be mild and non-specific. The high efficacy of 'self-swabbing' to test for chlamydia and gonorrhoea infections, has been proven when used in conjunction with reliable nucleic acid amplification techniques, which can negate the need for repeated vaginal examinations [13]. Chlamydia serology provides evidence of past infection and is a routine screening test in some clinics. The presence of chlamydial antibodies correctly predicts tubal damage in 90% of cases, of whom over half have no history of pelvic inflammatory disease. We advise the use of chlamydia screening to help identify patients whose tubal status should be tested early in the investigative process. There is evidence, however, that screening tests may be negative in the presence of infection in the upper genital tract and so there is rationale for prophylactic antibiotics prior to any procedure that involves instrumentation of the cervix (doxycycline or azithromycin).

Diagnosis of Anovulatory Infertility

Determining the cause of anovulatory infertility is the key to treatment as correction of the cause will result in cumulative conception rates that mimic those expected for normal women of the same age. It is first necessary to ascertain whether ovulation is occurring. Patients with anovulatory infertility will have oligomenorrhea or amenorrhea and a low luteal phase progesterone. A progesterone concentration of greater than 30 nmol/L suggests ovulation but it can be difficult to know when to take the blood if the patient has an erratic cycle—and impossible if she is amenorrhoeic. If the progesterone is 15–30 nmol/L the timing may have been incorrect. It is then necessary to check the timing of the blood test to subsequent menstruation and repeat the test in the following cycle (sometimes two progesterone measurements in the same cycle are helpful). The optimal way to assess ovulation in women with irregular cycles is by a combination of serial ultrasound scans and serum endocrine measurements (follicle stimulating hormone (FSH) and luteinizing hormone (LH) in the follicular phase and progesterone in the luteal phase).

The optimal frequency of intercourse is every 2–3 days in the follicular phase of the cycle and, if possible, daily for 2–3 days at the predicted time of ovulation. Abstinence until the 'day of ovulation' can be detrimental to sperm function. It is therefore important to advise couples about the frequency of intercourse and try to diffuse the tensions that often result from timed intercourse 'to order'.

The timing of sexual intercourse in relation to ovulation has a strong influence on the chance of conception. The precise number of fertile days in a woman's menstrual cycle is uncertain. A recent study [14] demonstrated that the probability of conception was 10% when intercourse occurred 5 days before ovulation and 33% when it took place on the day of ovulation. The fertile period appears to last 6 days and ends on the day of ovulation. The rapid decline in the probability of conception after this time is due either to a short survival time of the oocyte or a swift change in the nature of the cervical mucus. If commercially available kits for detecting the mid-cycle surge of LH in the urine are used to focus a couple to have intercourse on the day of the LH surge and the following day, they may be missing 3 or 4 fertile days prior to this and reducing their chance of conception. With respect to the precise timing of the 'fertile window' in the menstrual cycle, this occurs between days 10 and 17 in only about 30% of women [14].

The luteal phase of the cycle normally lasts for between 10 and 17 days and the concept of 'luteal phase deficiency' (LPD) is controversial. Probably the most convincing argument against the phenomenon of LPD is the failure of luteal support—with either progesterone or hCG—to improve pregnancy rates in spontaneous pregnancies. Endometrial biopsy has been used to assess the quality of ovulation further by equating histological changes with serum progesterone levels. Histological dating is, however, an unreliable indicator of the endometrial response to hormonal stimulation and is open to considerable biological variability and observer error. We do not recommend endometrial biopsy for determining whether the patient has ovulated.

Endocrine Profile

A baseline endocrine profile is optimally performed during the first 3 days of the cycle. It is essential to be aware of the normal reference range for the assay in the laboratory in which it is being performed. Reference ranges vary from laboratory to laboratory and can be quite different if different types of assay are used, for example, radioimmunoassays and immunoradiometric assays give

very different results for gonadotrophin measurements. There are a variety of recent advances in assay technology including chemiluminescence assays and mass spectrometry. It is therefore important to have knowledge of normal ranges for the assays used by your laboratory and also to ensure that they are appropriately calibrated for your 'normal' population.

Standard tests include a baseline measurement of FSH, LH, and oestradiol, thyroid function and in those with menstrual irregularity or amenorrhoea, prolactin, and testosterone. In women with evidence of hyperandrogenism or PCOS a more detailed assessment of androgen profile and metabolic screen may be indicated.

Ovarian Reserve Tests

It is natural for a woman to wish to have an idea of her potential fertility. A measurement of serum FSH concentration taken during days 1–3 of menstruation has been the traditionally used test of 'ovarian reserve'—a term that refers both to the number of oocytes within the ovary and their fertility potential (see Table 8.6.1.1). Ovarian reserve, or the number of releasable oocytes, declines with ovarian age, which does not always equate with the age of the woman. An elevated FSH level indicates reduced ovarian reserve and, generally, if greater than 10 IU/L on more than one occasion, the ovaries are unlikely to be ovulating regularly and will also be resistant to exogenous stimulation. When the serum concentration of FSH is above 15 IU/L the chance of ovarian activity is slim and levels greater than 25 IU/L are suggestive of the menopause or premature ovarian insufficiency. Even if ovulation is occurring in the presence of an elevated serum FSH concentration the fertility potential of the oocyte within the follicle is significantly impaired and in the unlikely event of fertilization taking place, there is an increased likelihood of a chromosomally abnormal embryo developing, and consequent risk of miscarriage and fetal chromosomal abnormality.

Additional measurements can be made in order to increase the positive predictive value of FSH, including an assessment of ovarian volume and the number of visible antral follicles on ultrasound scan, serum inhibin B and anti-Müllerian hormone (AMH) [15]. It has even been suggested that these tests may help determine a woman's future fertility over forthcoming years—although the evidence for longer term predictions has been disappointing [16]. There is debate about the widespread use of ovarian reserve testing outside of the context of planning infertility treatment. More clinics now offer the opportunity to have a fertility 'check' which includes a review of ovarian reserve. These may be counterproductive if the aim is to plan motherhood as may falsely reassure women about their own fertility leading to delays in attempting to conceive [17].

Table 8.6.1.1 Gonadotrophin hormonal profile and potential diagnosis

FSH	LH	Oestradiol	Diagnosis
Normal	Elevated	Usually normal	PCOS
Normal	Low	Low	Weight-related amenorrhoea
Low	Low	Low	Hypogonadotropic hypogonadism, functional or organic
Elevated	Elevated	Low	If oligo-/amenorrhoeic: ovarian failure
Elevated	Elevated	High	If mid-cycle, think of mid-cycle surge

The number of **antral follicles** in the ovary, as assessed by pelvic ultrasound (see on) has been reported as the best single predictor of poor ovarian response to stimulation for IVF [18]. Indeed it is the number of small antral follicles, 2–6 mm in diameter, that declines significantly with age, while there is little change in the larger follicles of 7–10 mm, which is still below the size at which growing follicles have been recruited.

Anti-Müllerian hormone (AMH) is a dimeric glycoprotein and member of the transforming growth factor-β (TGF-β) superfamily, which is best known as a product of the testes during fetal development that suppresses the development of Müllerian structures. AMH is also produced by the granulosa cells of preantral and antral follicles and appears to be a more stable predictor of the ovarian follicle pool than FSH, as it does not fluctuate through the menstrual cycle. For many fertility clinics, AMH has superseded the traditionally used FSH level in measuring ovarian reserve and planning ovarian stimulation [19]. Indeed it has been reported that higher AMH concentrations are associated with increased numbers of mature oocytes, embryos, and clinical pregnancies during IVF treatment [20]. Assays for AMH are now available for routine use despite a problematic history with assay stability [21]. AMH has also been proposed as a biochemical marker for the presence of polycystic ovaries, with high AMH concentrations being characteristic although a consensus has not been agreed. A recent series from our clinic found cut-off values of xxx and xxx for the presence of polycystic ovaries and the polycystic ovary syndrome, respectively.

Chromosomal Analysis

It is sensible to study the chromosomes of women with infertility and any dysmorphic features, and those with premature ovarian insufficiency (POI). Routine chromosomal analysis for those with recurrent miscarriage is not recommended as the chance of finding an abnormality is slim. If there is a congenital abnormality within the offspring, unbalanced chromosomal abnormalities in the family or a detected translocation in pregnancy tissue, chromosomal analysis may be prudent [22]. Men with severe oligospermia (<1 million/ml) should also have an endocrine profile and a chromosomal analysis.

Autoantibodies

Women with premature ovarian failure sometimes have ovarian autoantibodies or signs of other autoimmune disease (thyroid, pernicious anaemia, diabetes mellitus, SLE). The presence of autoantibodies alerts one to the risk that these conditions may become manifest in the future.

Antiphospholipid Syndrome

Women with recurrent miscarriage might have elevated levels of lupus anticoagulant, anticardiolipin antibodies and β2 glycoprotein I antibodies; and may benefit from an acquired thrombophilia screen [22].

Pelvic Ultrasound

An ultrasound assessment of ovarian volume and antral follicle count in the early follicular phase has been used as a predictor for ovarian response prior to IVF treatment, with small volume ovaries indicating

reduced ovarian reserve. We recognize in the ovary three distinct morphological appearances: normal, polycystic, and multicystic. Multicystic ovaries are characteristically observed in pubertal girls and women recovering from weight loss-related amenorrhea. These multicystic (or multifollicular) ovaries are normal in size or slightly enlarged and contain six or more cysts that are 4–10 mm in diameter; in contrast to women with polycystic ovaries (PCO), the stroma is not increased. The multicystic ovary appears to develop as a consequence of reduced hypothalamic secretion of gonadotrophin-releasing hormone (GnRH) which results in subnormal stimulation of the ovaries by the gonadotrophins. Polycystic ovaries are a separate entity and have a distinct response to induction of ovulation and ovarian stimulation for *in vitro* fertilization (IVF). Polycystic ovaries may be present in women who are non-hirsute and who have regular menstrual cycles. It is important to differentiate between PCO and PCOS. The former describes the morphological appearance of the ovary whereas the latter term is only appropriate when PCO are found in association with a menstrual disturbance (amenorrhea or, more commonly, oligomenorrhea) and/or the complications of hyperandrogenization (seborrhoea, acne and hirsutism) [23]. Polycystic Ovary Syndrome is also associated with endocrinological abnormalities and in particular with elevated serum concentrations of androgens (testosterone, androstenedione), LH, prolactin, and oestrogens. As with the clinical picture, these changes are variable and patients with PCOS may have normal endocrine concentrations.

With the advent of high-resolution ultrasound, identification of polycystic ovaries is simple and the polycystic ovary should have at least one of the following: either 12 or more follicles measuring 2–9 mm in diameter or increased ovarian volume (>10 cm^3) [24].

Besides making a careful assessment of ovarian morphology, it is necessary to perform a baseline ultrasound scan of the ovaries before commencing ovarian stimulation in order to detect the presence of ovarian cysts, which may be a physiological remnant of ovulation or representative of ovarian pathology. An endometrioma has the characteristic hazy, echodense appearance of blood in a cyst. Dermoid cysts are sometimes seen in women of reproductive age and may be difficult to distinguish from endometriomas, as both may be bilateral with a hazy, homogeneously echodense appearance of lipid matter in dermoids and blood in endometriomas. All but obviously simple cysts should be treated with caution as ovarian malignancy may occur in young women. Therapeutic stimulation of the ovaries should not be performed until complex ovarian cysts have resolved, either spontaneously or surgically.

The baseline ultrasound scan also permits inspection of the other pelvic structures and might reveal the presence of hydrosalpinges, fibroids, or congenital developmental anomalies of the uterus. Endometrial changes can be seen clearly using pelvic ultrasound. The endometrial thickness in the early follicular phase is 4–6 mm, by the time of ovulation it is about 8–10 mm and in the mid-luteal phase it reaches 14 mm. It has been suggested that there is a reduced chance of pregnancy if the triple line appearance is absent or if the preovulatory endometrial thickness is less than 7 mm [25].

Assessment of Tubal Patency and the Uterine Cavity

Hysterosalpingography

Tubal infertility is diagnosed in between 15% and 50% of couples presenting with subfertility. X-ray hysterosalpingography (HSG) provides a delineation of both the uterine cavity and the fallopian tubes. An HSG is the simplest preliminary test for the delineation of the uterine cavity and fallopian tubes and has few complications. An HSG is performed if there are no pointers in the history to an increased risk of tubal disease, for example pelvic infection, peritonitis, pelvic pain.

Hysterosalpingo-Contrast-Sonography (HyCoSy) and Saline Infusion Sonohysterography (SIS)

Similar to the HSG, a HyCoSy and SIS can be carried out relatively easily in an outpatient setting. Saline with/without the addition of a foam, can be used to outline both the uterine cavity and passage of fluid through the fallopian tubes during an ultrasound assessment. There is no risk of radiation exposure unlike with the HSG. This technique should be carried out when there are no tubal disease risk factors. An additional benefit is the concurrent assessment of uterine and ovarian morphology.

Laparoscopy and Hysteroscopy

Hysteroscopy permits assessment of the uterine cavity at the same time as laparoscopic assessment of the pelvic contents. Congenital anomalies of uterine development occur in about 4% of women; although rarely affecting fertility they may sometimes predispose to an increased risk of second trimester miscarriage. It is our practice to consent all patients undergoing diagnostic laparoscopy for treatment of mild endometriosis or adhesiolysis, which should not prolong the procedure by more than 15–20 minutes. There is evidence that even mild endometriosis may adversely affect fertility and so ablation, with diathermy or laser, can be performed during the initial diagnostic procedure. Fine periovarian and peritubular adhesions can often be broken down at the time of the initial laparoscopy. The presence of hydrosalpinges warrants salpingectomy to improve the chance of success in subsequent IVF treatment.

Investigating the Male

The general examination should include an assessment of BMI, blood pressure, secondary sexual characteristics, the abdomen, and genitalia. Some chest diseases are associated with infertility (congenital absence of the vas, spermatic duct obstruction) and might be elicited at the time of the examination. An absent or deficient sense of smell in patients with hypogonadotropic hypogonadism gives the diagnosis of Kallmann's syndrome.

Men with androgen deficiency of prepubertal origin will have a high-pitched voice, small soft testes, and a small penis, lack of adult hair, and decreased muscle mass. They are often tall with a large arm span that exceeds their height. If hypogonadism develops after puberty the skin becomes fine, body hair and beard growth diminish. There may be gynaecomastia, as in Klinefelter's syndrome. Gynaecomastia may also occur with hyperthyroidism, liver disease, oestrogen, or hCG-producing tumours or with some drugs (most notably antiandrogens such as cimetidine, spironolactone, digitalis). Transient gynaecomastia is normal during puberty. Other signs of endocrine disease (Cushing's syndrome, thyroid disease, pituitary tumour) should also become evident on the general examination. A full neurological examination is required when there are problems with sexual function.

Congenital deformities of the penis or hypo-/epispadias may cause problems with semen deposition. Testicular size should be

assessed using an orchidometer and is normal if over 15 ml. Small testes that are soft are usually associated with gonadotrophin deficiency, as in hypopituitarism or Kallmann's syndrome. Small testes that are firm (implying fibrosis) are usually associated with severe and permanent destruction of germinal epithelium (as in Klinefelter's syndrome) and androgenization may be normal. Testicular masses or asymmetry warrant further investigation by ultrasound in the first instance.

The Semen Analysis

The specimen of semen should be produced by masturbation into a clean, dry container and delivered to the laboratory within 30 minutes of its production. There should have been a period of abstinence of 3 days. A fixed period of abstinence not only improves the standardization of the test but more than 5 days abstinence is associated with a decrease in motility despite an increase in sperm number. There are large swings in semen parameters in healthy, fertile sperm donors and so the results of a single semen analysis should be viewed with caution and repeated on two or more occasions, 3 months apart. Sperm production by the testis takes 10–12 weeks and so an abnormal semen specimen is a reflection of testicular function 3 months previously. See **Table 8.6.1.2**.

Defining 'normal' sperm parameters has been a historically difficult subject due to the lack of consensus on who makes up a 'normal' population. Recognized globally, the WHO has provided standardized measurements adopted by many laboratories. The most recent edition [26], has attempted to address previous concerns regarding levels falling well within a fertile range; this has led to a reduction in the three main parameters (count, motility, morphology). The male populations used have proven fertility within a relationship with time to pregnancy ≤12 months [27]. The chance of natural conception falls significantly when the sperm concentration is less than 5×10^6/ml. When the total count is low there is often a corresponding reduction in motility. It has been suggested that sperm morphology is one of the better prognosticators for fertility [28]. Immotility can be caused by infection, superoxide production by leukocytes, antisperm antibodies, or defects in the microtubules and dynein arms of the sperm tail.

Sperm function can be impaired by lipid peroxidation in the sperm plasma membrane. Oxidative stress correlates with reduced motility and a decreased capacity for oocyte fusion. Reactive oxygen species (ROS) initiate lipid peroxidation and are produced either within the dysfunctional spermatozoa or by leukocytes. Seminal plasma contains a rich concentration of antioxidants and removal of sperm from seminal plasma during preparative procedures for assisted conception can expose the sperm to damaging ROS. A matter of some concern is the notion that ROS are well-known mutagens and sperm-derived genetic damage to embryos might occur through chromosomal breakage (while oocyte-derived damage occurs through chromosomal rearrangement) [29]. It has been suggested that spermatozoa that have been exposed to ROS are at increased risk of carrying chromosomal breakages. Measurements of DNA fragmentation, are one suggested method to assess spermatozoa integrity and have been related to recurrent pregnancy loss, although are not recommended for routine screening [30].

Table 8.6.1.2 Normal semen parameters (WHO, 2010)

Volume	≥ 1.5 ml
pH	≥ 7.2
Sperm concentration	≥ 15 x 10^6/ml
Total sperm count	≥ 39 x 10^6/ejaculate
Total Motility (progressive + non-progressive)	40%
Progressive motility	32%
Vitality (live spermatozoa)	58%
Morphology (normal forms)	4%
Peroxidase-positive leukocytes	<1.0 x 10^6/ml
MAR test (motile spermatozoa with bound particles)	<50%
Immunobead test (motile spermatozoa with bound beads)	<50%
Seminal zinc	≥ 2.4 µmol/ejaculate
Seminal fructose	≥ 13 µmol/ejaculate
Seminal neutral glucosidase	≥ 20 mU/ejaculate

Management of Tubal Infertility

In vitro fertilization has revolutionized many forms of fertility therapy, yet the question of IVF versus tubal surgery for mild to moderate tubal disease is still debated. Successful tubal surgery can provide a permanent cure, with the possibility of more than one pregnancy and can be performed laparoscopically

The techniques employed in tubal surgery are of paramount importance and require adequate training, whether performed at laparotomy or laparoscopy. Open tubal surgery is optimally performed using an operating microscope. While some surgeons advocate the continued use of open microsurgery, the laparoscopic approach has gained favour in recent years. One study compared the outcome of microsurgical and laparoscopic adhesiolysis and found no statistically significant difference in cumulative conception rates, which were a little over 40% after 12 months [31]. Peritubal adhesions interfere with ovum pick-up and tubal transport, while periovarian adhesions may inhibit ovulation. When the tubes are patent and the ovaries freely mobile, adhesiolysis will result in good cumulative conception rates (60% in 24 months), although at second-look laparoscopy there is often a recurrence of the adhesions to some degree. Dense adhesions carry a worse prognosis than fine, filmy adhesions.

The mainstay of salpingostomy is the fashioning of a small ostium at the tip of the tube, with eversion of the tubal mucosa so that the reconstructed fimbriae are positioned to allow the ostium free movement over the ovary. Raw areas and linear incisions in the tube will heal over and should be avoided. The best cases to treat are those in which the tubes have thin walls, normal mucosa, and no periovarian adhesions, although when the distal end of the tube is blocked there are usually periovarian and peritubular adhesions. Large hydrosalpinges, greater than 1.5 cm in diameter, carry a worse prognosis and are often excised.

Cornual occlusion due to infection (salpingitis isthmica nodosa, pelvic inflammatory disease, tuberculosis) is often associated with microscopic damage along the length of the tube and so there is a worse prognosis and greater risk of ectopic pregnancy than after reversal of sterilization. Reversal of sterilization leads to the best results, not only because the patient is of proven fertility, but also

because damage is to a very small portion of the tube. Pregnancy rates are between 60% and 80%, with ectopic pregnancy rates usually less than 5%.

Women with moderate to severe tubal infertility are optimally treated with IVF. If they have a history of repeated ectopic pregnancy there is a case for performing a sterilization prior to IVF, as there is nothing more traumatic than developing a further ectopic pregnancy after the stresses of an IVF treatment cycle. The overall rate of ectopic pregnancy after IVF is 5% (i.e. higher than normal) because uterine transfer of the pre-embryo(s) does not ensure that it will remain in the uterine cavity.

There is good evidence to suggest that the presence of hydrosalpinges affects the outcome of IVF by having an effect on the endometrial environment, possibly through the passage of toxic fluid into the uterine cavity, which disrupts implantation. If the tubes are completely blocked and there are large hydrosalpinges there is a case for surgical intervention. A large prospective randomized controlled trial evaluated salpingectomy versus no intervention. Despite no significant difference in the pregnancy rate between the salpingectomy group (36.6%) and the non-intervention group (23.9%), the live birth rates were increased (28.6% vs. 16.3%, $P = 0.045$) [32]. The differences were more significant in the presence of bilateral hydrosalpinges and particularly so with ultrasound visible hydrosalpinges (clinical pregnancy rate 45.7% vs. 22.5%, $p = 0.029$, live birth rate 40% vs. 17.5%, $P = 0.038$). More recently, it has been suggested that tubal occlusion may offer an alternative to salpingectomy with equal results. This is particularly useful for cases where significant adhesions or pelvic pathology may make salpingectomy a more difficult procedure. A recent Cochrane review evaluates any surgical intervention for those with hydrosalpinges prior to IVF and concludes that both occlusion and salpingectomy increase the clinical pregnancy rate, with no significant advantage over either one [33]. Salpingectomy can usually be performed laparoscopically and care should be taken not to compromise ovarian blood supply. Interest in hysteroscopic tubal occlusion prior to IVF has waned due to inferior success rates and increased risk of miscarriage [34].

Fibroids and Myomectomy

Fibroids are common with increasing incidence with age. Prevalence has been reported as low as 3% in Swedish Caucasian women aged 25–32 years and 8% in those aged 33–40 [35]. While rates have been reported as high as 70% in white Americans and 80% in African Americans age 50 years [36]. Imaging and initial assessment of fibroids is by ultrasonography but magnetic resonance imaging (MRI) can be extremely helpful in further delineating the position of multiple fibroids and distinguishing fibroids from adenomyoma. Classification of fibroids is by their position, with serosal fibroids being of least significance to fertility, there is then increasing significance of the presence of subserosal fibroids (in which >50% projects out of the serosal surface), intramural fibroids, and submucous fibroids, which in turn may be pedunculated into the cavity of the uterus (type 0), sessile with intramural extension of either ≤50% (type I) or ≥50% (type II). It is thought that fibroids are most likely to affect fertility if they either distort the endometrial cavity or have an intramural component of >4 cm.

Fibroids are often removed indiscriminately and myomectomy can result in extensive pelvic adhesion formation and damage to the integrity of the uterine cavity. The debate continues as to how many fibroids should be removed and from what location. There is increasing evidence that intramural fibroids affect implantation, even when there is no deformation of the uterine cavity. A meta-analysis of 17 studies concluded that all fibroids may affect fertility, with a greater influence on delivery than pregnancy rates [37]. The most recent review from the American Society for Reproductive Medicine, concludes that there is insufficient evidence that the presence of myomas reduce the likelihood of achieving pregnancy, however, myomectomy for cavity distorting myomas (open, laparoscopically or hysteroscopic) improves pregnancy rates and reduces the chance of miscarriage [38]. Myomectomy is a major procedure with potential risks to the integrity and viability of the uterus. Preoperative treatment with a gonadotropin-releasing hormone agonist for 6–8 weeks will cause significant shrinkage of the fibroids and reduce vascularity and blood loss during surgery

Less invasive procedures than operative myomectomy are being evaluated for the management of fibroids, including uterine artery embolization and MRI-guided laser coagulative necrosis or high intensity focused ultrasound for the destruction of fibroids. The place of these techniques in the management of infertility is still being evaluated. Furthermore, while uterine artery embolization has become popular in the management of fibroids it is not recommended for those who wish to preserve fertility because of the potential adverse effect on both uterine and ovarian blood supply.

Endometriosis

Endometriosis can cause pelvic pain and infertility. Treatment is best achieved with surgery without delaying the chance of conception by hormonal therapies that are contraceptive. Careful laparoscopic assessment of the pelvis reveals signs of endometriosis in up to 18% of women with proven fertility [39]. While a number of theories have been proposed for the pathogenesis of endometriosis, that of retrograde menstruation is the most popular and plausible. Retrograde menstruation is common, being seen in 75–90% of women who have had laparoscopies performed at the time of menstruation. Menstrual blood does not always contain endometrial cells and the factors that influence implantation of ectopic endometrium are uncertain, for the prevalence of endometriosis has been estimated as 1–20%, not 75–90%. Women with endometriosis appear to have altered immune function, which may permit implantation of regurgitated endometrium. Abnormalities of cellular adhesion molecules, including the integrins and extracellular matrix proteins, are also thought to play a role in pathogenesis. The detection of endometriosis in women being investigated for subfertility is thought to reflect their lack of conception and exposure to frequent menstruation rather than being a cause of the infertility. Indeed, the likelihood of finding evidence of endometriosis in women who attend for sterilization is increased in proportion to the interval since the birth of their last child.

Women with symptomatic endometriosis may have a genetic disposition to endometrial implantation on the peritoneum and a further disposition to an inflammatory response to the cyclical changes that occur in the ectopic endometrium. As is well known, the degree of endometriosis does not correlate with symptomatology: pelvic pain, dyspareunia, and dysmenorrhoea. It is not possible, moreover, to predict which patients will develop progressive disease with resultant pelvic adhesions and ovarian cysts.

It is easy to envisage how severe endometriosis can affect fertility by distorting pelvic anatomy, with adhesions that smother the ovaries and tubes and with endometriotic ovarian cysts. Furthermore, the prevalence of endometriosis in infertile women is as high as 20–68% [40]. There is debate about the extent to which endometriosis affects fertility in the absence of pelvic deformity. It has been suggested that the peritoneal environment is altered, with an increased concentration of macrophages which impede sperm motility, phagocytose spermatozoa, and interfere both with oocyte pick-up by the fallopian tube and with fertilization.

The management of endometriosis depends upon the wishes of the patient, specifically whether her predominant complaint is pain or infertility. If fertility is required but pain is also a problem then management has usually been with analgesics, either alone or combined with surgical treatment. Non-steroidal anti-inflammatory drugs (NSAIDs) are regularly used although their effectiveness is questionable. There is some evidence that NSAIDs inhibit the process of ovulation through their antiprostaglandin action but endometriotic pain usually occurs at the time of menstruation rather than mid-cycle and so these drugs should be safe in women wishing to conceive.

Laparoscopy is the mainstay of the classification of endometriosis and the best known system of classification is that of the American Fertility Society (AFS, now American Society of Reproductive Medicine, ASRM). in which the appearance of the disease, the degree of adhesions, and obliteration of the pouch of Douglas provide a score. It has been suggested that the AFS classification is limited by its inability to provide an indication of the activity of the disease and has no predictive value with respect to either pain or subfertility. Surgical therapy for the treatment of endometriosis can be performed at the time of the diagnostic laparoscopy, although only if the diagnosis has been suspected and the patient has been given appropriate information and consent.

There is little to choose between the medical therapies (e.g. progestogens, GnRH agonists, danazol or the combined oral contraceptive pill) with respect to subsequent fertility and a body of evidence that indicates no benefit when compared with expectant management. These have been collected together in a systematic review [41, 42] in which 23 trials involving 3043 women were included. The odds ratio for pregnancy following ovulation suppression versus placebo or no treatment for all women randomized was 0.79 (95% CI 0.54–1.14), $P = 0.21$ and 0.80 (95% CI 0.51–1.24), $P = 0.32$, respectively for subfertile couples only despite the use of a variety of suppression agents. This absence of demonstrable efficacy, together with the fact that the treatments are contraceptive, means that medical therapies are inadvisable for women who wish to conceive.

There is now convincing evidence that laparoscopic treatment of mild to moderate endometriosis is superior to diagnostic laparoscopy alone for both ongoing pain management and clinical pregnancy rate. Laparoscopic treatment provides improved pain scores at 6 and 12 months, respectively (odds ratio (OR) 6.58, 95% CI 3.31–13.10, 3 randomized controlled trials (RCTs), 171 participants, and OR 10.00, 95% CI 3.21–31.17, 1 RCT, 69 participants). Excision or ablation of endometriosis yields a higher live birth rate or ongoing pregnancy rate past 20 weeks when compared with diagnostic laparoscopy alone (OR 1.94, 95% CI 1.20–3.16, $P = 0.007$, 2 RCTs, 382 participants, moderate-quality evidence) [43].

In considering surgery for endometriosis a distinction should be made between ovarian endometriomata and deeply infiltrating endometriosis (i.e. endometriosis that penetrates more than 5 mm below the peritoneal surface). Cystic ovarian endometriosis tends to be associated with adhesions, while deep infiltrating endometriosis is not and is often found in the pouch of Douglas, on the uterosacral ligaments, and in the uterovesical fold. Sometimes the lesions can be very deep yet have only a small visible surface area. MRI can be helpful in localizing the lesions and guiding the surgery.

There is some evidence that excisional surgery for endometriomata of greater than 3 cm in size provides for a more favourable outcome than simple drainage and ablation [43, 44]. A systematic review reported that laparoscopic excision of the cyst wall of the endometrioma was associated with a reduced rate of recurrence of the endometrioma (OR 0.41 95% CI 0.18–0.93), reduced requirement for further surgery (OR 0.21 95% CI 0.05–0.79), reduced recurrence rate of the symptoms of dysmenorrhoea (OR 0.15 95% CI 0.06–0.38), dyspareunia OR 0.08 95% CI 0.01–0.51), and non-menstrual pelvic pain (OR 0.10 95% CI 0.02–0.56). It was also associated with a subsequent increased rate of spontaneous pregnancy in women who had documented prior subfertility (OR 5.21 CI 2.04–13.29) [45]. While surgery for endometriomata may improve outcomes, care must be taken to retain normal ovarian tissue to prevent further depletion of the ovarian reserve.

Laparoscopic surgery should only be performed by appropriately trained and skilled surgeons as endometriosis taxes the skill of the surgeon more than any other disease in the pelvis. It may be necessary to resect affected bowel or bladder and the help of a colorectal surgeon or a urologist may be required. Great care is required when operating near the ureter and ureteric stenting may be helpful. Large lesions often require laparotomy although such major surgery is usually reserved for patients with severe pain who have completed their family rather than for those with infertility, in whom GnRH agonist therapy combined with IVF is usually more appropriate.

Aggressive treatment of deeply infiltrating endometriosis and cystic ovarian endometriosis is associated with cumulative pregnancy rates of up to 60% over 12 months, after which IVF will probably provide a greater chance of conception than a second-look procedure. Severe endometriosis may reduce the success of IVF therapy, by impairing the rates of fertilization and implantation. It is commonplace to suppress active endometriosis with a GnRH agonist for 2–3 months prior to IVF, particularly if pituitary desensitization is part of the IVF treatment protocol [42].

Adenomyosis, a benign condition where endometrial glands and stroma are found within the myometrium, may impact on fertility. Theories to this effect include altered uterine contractility, impaired uterotubal transport and reduced sperm function secondary to high uterine cavity nitric oxide levels [46]. A detrimental effect on IVF outcomes is recognized with a reduction in implantation and increased early pregnancy loss. A 41% reduction in live birth rate for those with adenomyosis following an embryo transfer has been shown (OR 0.59, 95% CI 0.42–0.82) [46]. Diffuse adenomyosis appears to be a worse prognostic feature than focal adenomyosis, which is more amenable to surgical excision. Pre-IVF treatment with a GnRH agonist appears to be beneficial and is certainly less invasive then surgical options.

Anovulatory Infertility, PCOS, and Ovulation Induction

These subjects are dealt with in Section 8.4, 'Polycystic Ovary Syndrome and Other Androgen Excess Disorders'.

Unexplained Infertility

One can consider two approaches to the diagnosis and management of unexplained infertility. The first is strictly scientific, with a quest for and exclusion of each known cause of infertility before the label 'unexplained infertility' can be given. The second approach is a pragmatic one based upon a management-oriented policy, whereby treatment is commenced after the common obstacles to fertility have been excluded. The treatment of unexplained infertility essentially aims to boost fertility, usually by a combination of superovulation and close apposition of sperm and egg(s).

Studies of populations of patients with infertility indicate that approximately 10–25% have unexplained infertility, 20–30% ovulatory dysfunction, 20–35% tubal damage, 10–50% sperm dysfunction, 5–10% endometriosis, and 5% coital dysfunction. A degree of subfertility is found in both partners in 30–50% of couples, as usually a couple's subfertility is a relative rather than an absolute barrier to conception. Unexplained infertility has been defined as the inability to conceive after 1 year in the absence of any abnormalities. Between 40% and 65% of couples given this label will conceive spontaneously over the following 3 years and it has been suggested that treatment should be deferred until the couple has been trying to conceive for at least 3 years, as before this time therapy does not confer any benefit over the natural chance of conception [47, 48]. It appears that the most important prognostic factors are the duration of infertility and the age of the female partner.

A number of approaches have been employed in the management of unexplained infertility. Therapy should aim to boost the monthly pregnancy rate above the natural rate of 1.5–3% that is expected for couples who have been trying to conceive for over a year.

It used to be thought that clomiphene enhanced fertility by correcting a subtle defect in ovarian function—either follicular development or luteal phase defect. A recent three arm randomized controlled trial in couples with infertility for over 2 years, confirmed ovulation, patent fallopian tubes, and motile sperm randomized 580 women to expectant management (n = 193), oral clomiphene citrate (n = 194), or unstimulated intrauterine insemination (n = 193) for 6 months [49]. Live birth rates were 32/193 (17%), 26/192 (14%), and 43/191 (23%), respectively. Compared with expectant management, the odds ratio for a live birth was 0.79 (95% CI 0.45–1.38) after clomiphene citrate and 1.46 (0.88 to 2.43) after unstimulated intrauterine insemination. Despite no benefit with respect to live birth rates, more women randomized to clomiphene citrate (159/170, 94%) and unstimulated intrauterine insemination (155/162, 96%) found the process of treatment acceptable than those randomized to expectant management (123/153, 80%) ($P = 0.001$ and $P < 0.001$, respectively). A subsequent review also concluded that clomiphene citrate was no more effective than no treatment or placebo for live birth (odds ratio (OR) 0.79, 95% CI 0.45–1.38; $P = 0.41$) [50]. The addition of clomiphene also increased the risk of multiple pregnancy

Although it is reasonable to assume that the combination of gonadotropins to induce superovulation, with the release of two or three oocytes, with insemination of a prepared sample sperm into the uterine cavity should boost fertility, this appears to be unfounded. A meta-analysis of 14 RCTs including 1867 women with unexplained infertility found no conclusive evidence of benefit in cumulative live birth rate when compared with timed intercourse, both with and without ovarian stimulation [51]. A recent trial questions this indifference, evaluating stimulated intrauterine insemination (IUI) against expectant management. An increased live birth rate was seen with IUI (risk ratio (RR) 3.41, 95% CI 1.71–6.79; $P = 0.0003$) [52].

IVF confers the advantages of being able to study fertilization and the selection of good quality pre-embryos for transfer into the uterus. The Cochrane database [53] reports a higher live birth rate per woman with IVF compared with expectant management (OR 22 95% CI 2.56–189.4, 1 RCT, 51 women, very low-quality evidence) or unstimulated IUI (OR 2.47, (%% CI 1.19–5.12, 2 RCTs, 156 women, low-quality evidence). For those women who had no pretreatment with either IUI or ovarian stimulation, there was no conclusive evidence of a difference in live birth rates with IVF. The overall paucity of evidence means firm conclusions as to the benefit of IVF cannot yet be made in unexplained infertility. Overall it seems sensible to progress to IVF in couples with unexplained infertility after 2 years of regular unprotected intercourse.

In Vitro Fertilization (IVF)

Assisted conception techniques involve the laboratory preparation of gametes, artificially bringing them closer together and hence enhancing fertility by either bypassing an absolute obstruction to fertilization or boosting fecundity above that expected without treatment. Assisted conception is indicated if the prognosis for tubal surgery is considered too poor or if conception has failed to occur within 12 months of tubal surgery. IVF is indicated for moderate to severe endometriotic disease if conception has failed to occur within 12 months of ablative laparoscopic surgery. When there is severe sperm dysfunction and sperm preparation provides an inadequate specimen for superovulation intrauterine insemination or if conception has failed to occur after 3–4 cycles of superovulation/IUI, IVF should be offered. Micromanipulation techniques (i.e. intracytoplasmic sperm injection (ICSI)) may be required to achieve fertilization if there is severe male factor infertility.

Prior to assisted conception treatment, in addition to baseline infertility investigations, it is usual for most clinics to test couples for HIV, hepatitis B, and hepatitis C, in order to avoid iatrogenic transmission from one partner to the other and also to protect laboratory staff who are handling bodily fluids. Furthermore, cryopreserved gametes and embryos have the potential—albeit unproven—of cross-contamination through liquid nitrogen.

For a couple to undergo IVF, the female partner should have functioning ovaries and a normal uterus and the male partner at least one sperm per ejaculate. However, the lack of ovarian function can be bypassed with oocyte donation, the absence of sperm can be bypassed with sperm donation, and the absence of a uterus by IVF surrogacy. Sometimes both sperm and oocytes, or surplus embryos

from another couple, are donated so that the resultant child has inherited no genetic material from either parent. Such parents have in reality 'adopted' the embryos but do, of course, gain from the experience of pregnancy and childbirth.

IVF is sometimes embarked upon before all other treatment modalities have been exhausted and while we do not advocate unnecessary delay, particularly in older patients, the notion that IVF is the high-tech modern answer to every couple's subfertility is erroneous. The stresses placed upon a couple by IVF (and other assisted conception procedures) are immense and the treatment has risks and complications (e.g. ovarian hyperstimulation syndrome (OHSS) and multiple pregnancy).

Regimens for IVF

IVF therapy has become increasingly simplified in recent years. The use of gonadotropin-releasing hormone (GnRH) agonists and antagonists with gonadotropins has resulted in greater ease of planning the superovulation stimulation. When GnRH agonists or antagonists are used the oocyte retrieval can be precisely timed to occur 34–38 hours after the administration of hCG. The latter acts as a surrogate for the normal mid-cycle LH surge and causes resumption of meiosis within the oocytes and their preparation for fertilization. Furthermore, there is good evidence that the oocytes do not become over-mature within follicles that are considered to be ready for collection and so the administration of hCG can be delayed to avoid oocyte collection at weekends. Most large clinics, however, provide flexibility with a 6- or 7-day service.

A disadvantage of the use of GnRH agonists in a classic 'long protocol' is the 2 weeks or more lead-in to the therapy during which pituitary desensitization ('downregulation') is achieved before stimulation with gonadotropins can be commenced. Pituitary desensitization is assessed by a combination of endometrial shedding and low serum concentrations of oestradiol (ultrasound confirmation of a thin endometrium and quiescent ovaries is adequate without recourse to biochemistry).

The GnRH agonists can be administered intranasally, subcutaneously, or intramuscularly (by depot in some instances). The shorter acting preparations can be used to induce a flare response, being commenced on day 1 of the cycle, with gonadotropin stimulation starting the following day. The agonist is then either continued through to the day of hCG administration or given for 3 days only. The flare response can be utilized in those patients who have had a poor response in the past in order to try to maximize the response to stimulation—this it does to varying degrees.

The advent of the third-generation GnRH antagonists enables us to dispense with pituitary desensitization and commence ovarian stimulation on day 2, with the daily administration of an antagonist on day 6 of stimulation or once the leading follicle(s) has reached a diameter of 14 mm (usually day 6 or 7). Although it appears that success rates are better when commenced on day 6 rather than using a flexible protocol [54]. The GnRH antagonist acts immediately to inhibit pituitary secretion of FSH and LH, without the flare effect of agonists or the need for 10–14 days' desensitization. An endogenous LH surge can be prevented, thereby allowing oocyte retrieval at the desired time. GnRH antagonist cycles are certainly much shorter and more convenient for patients than the 'long protocol' and many clinics are now increasingly using them.

Initial studies found pregnancy rates were lower when using the antagonist protocol. A recent review, however concluded that there is a similar probability of livebirth when either GnRH agonists or antagonists are used (OR 1.02, 95% CI 0.85–1.23; 12 RCTs n = 2303, moderate-quality evidence) [55] with no difference in miscarriage rates (OR 1.03, 95% CI 0.82–1.29 34 RCTs, n = 7082, moderate-quality evidence). The advantage of the antagonist protocol is a reduction in incidence of OHSS (OR 0.61, 95% CI 0.51–0.72; 36 RCTs, n = 7944, moderate-quality evidence). The pharmacokinetics of the antagonist protocol allows use of a GnRH agonist to initiate oocyte maturation, rather than human chorionic gonadotrophin (hCG), further reducing the risk of OHSS due to the agonists' shorter half-life. Using a GnRH agonist as a trigger was initially found to be detrimental to the pregnancy rate [56] (OR 0.47, 95% CI 0.31–0.70; five RCTs, 532 women, moderate-quality evidence) but with intensive luteal phase support, using a combination of oestradiol, progestogen, and hCG, similar pregnancy rates to hCG are now reported [57].

Gonadotropin Therapy

Gonadotropin therapy for the stimulation of superovulation can be with either human menopausal gonadotropins (hMG), which contain urinary-derived FSH and LH in differing proportions depending on the preparation, or with urinary-derived FSH alone; or with recombinant-FSH (rFSH). The advent of the recombinantly derived gonadotropins has broadened the scope of therapeutic agents and resulted in a potentially unlimited supply.

Most RCTs comparing gonadotrophin preparations have evaluated efficacy in terms of number of oocytes retrieved, focusing on ovarian response and potency rather than on treatment outcome. Only very few RCTs have been powered to compare gonadotrophin preparations with respect to ongoing pregnancy rates, and none have been powered for live birth rate which is the outcome of interest to couples seeking infertility treatment. Live birth rate can however be appropriately addressed by meta-analysis. A recent meta-analysis of RCTs compared rFSH to all other gonadotrophins, irrespective of downregulation protocol used, and did not find any statistically significant difference in live birth rate (OR 0.97, 95% CI 0.87–1.08, 28 RCTs, 7339 couples) [58]. If urinary gonadotrophins were considered separately, there were significantly fewer live births after rFSH than HMG (OR 0.84, 95% CI 0.72–0.99), 11 RCTs, n = 3197). There were no significant differences between hMG and rFSH with respect to miscarriage rate or multiple pregnancy. The authors conclude that any differences between urinary gonadotrophins were unlikely to be clinically significant and therefore choice of gonadotrophin should depend on availability and cost. In ovulation induction, similar pregnancy outcomes between urinary-derived gonadotrophins and rFSH have also been documented [59].

Oocyte Retrieval

The preovulatory hCG 'trigger' is usually administered when the leading follicle is at least 17–18 mm in diameter and there are at least three follicles greater than 17 mm. Ultrasound-guided oocyte retrieval is usually performed under light sedation plus analgesia; combinations of benzodiazepines, midazolam, and opiates are given intravenously or intramuscularly, with appropriate monitoring during and after the procedure. Administration of a local anaesthetic (1% lidocaine (lignocaine)) into the vaginal fornices is of

additional benefit. The procedure should be pain free. The patient is awake or lightly sedated and may be shown the oocytes on a closed-circuit video monitor attached to the embryologist's microscope. Oocyte retrieval should take about 20 minutes.

After oocyte retrieval, the semen is washed and prepared. Insemination is usually performed 1–6 hours after oocyte retrieval with 50–200 000 motile spermatozoa being placed with each oocyte; 16–18 hours later the oocytes are examined to ensure that correct fertilization has occurred, as defined by the presence of two pronuclei. Multiple pronuclei indicate polyspermic fertilization or digyny (i.e. failure to extrude the second polar body) and are not suitable for transfer.

Embryo Transfer

Embryo transfer was traditionally performed 2–3 days after oocyte collection (at the 4–8 cell stage). There has been a recent shift in favour of extended culture to the blastocyst stage (day 5 of embryo development). Improved clinical pregnancy rates (OR 1.30, 95% CI 1.14–1.47; 27 RCTs, 4031 women, moderate-quality evidence) and live birth rates (OR 1.48, 95% CI 1.2–1.82; 13 RCTs, 1630 women, low-quality evidence) are seen when compared with fresh cleavage stage transfers [60]. This equates to 29% of women achieving a live birth after fresh cleavage stage but 32–42% following blastocyst transfer. Arguments favouring blastocyst transfer include: avoidance of premature exposure of embryos to the highly oestrogenized uterine environment; reduced uterine pulsatility; and improved implantation as the most viable embryos reach blastocyst stage. Lower embryo freezing rates (OR 0.48, 95% CI 0.4–0.57; 14 RCTs, 2292 women) and higher failure to transfer cycles (OR 2.50, 95% CI 1.76–3.55; 17 RCTs, 2577 women) are seen at the blastocyst stage. Limited evidence suggests no difference in cumulative pregnancy rates following fresh and frozen embryo transfer cycles after a single oocyte retrieval between blastocyst and cleavage stage embryos. An RCT to compare day 5 transfer of a single blastocyst with transfer of a single cleavage embryo on day 3 had to be terminated early after a prespecified interim analysis found a higher rate of pregnancy and delivery in the blastocyst group (32.0% vs. 21.6%, RR 1.48 95% CI 1.04–2.11) [61].

Number of Embryos for Transfer

One major problem that has arisen from the growth of assisted conception treatment in a competitive environment is the dramatic rise in multiple births. Triplets and greater have been prevented by legislation introduced by the Human Fertilisation and Embryology Authority (HFEA) in the UK in 2002 limiting the number of embryos transferred to two for women under 40, as there is no evidence that the transfer of three significantly increases the chance of pregnancy and recommends the transfer of no more than 3 in women over the age of 40. In the UK, over the last 10 years, there has been a strong drive towards elective single embryo transfer (eSET). This has led to a significant reduction in multiple birth from 28% of all IVF treatment cycles in 1991 to 11% in 2016 [62]. American data, supports elective single embryo transfer with no significant impact on cumulative live birth rate. Good prognostic features include maternal age <38 years, surplus good quality embryos for cryopreservation and a previous live birth from IVF [63]. Evidence suggests that by adopting a SET policy and cryopreserving the spare embryos for subsequent replacement if the initial cycle should fail, the live birth rate is not significantly different to that following a double embryo transfer and multiple pregnancy rates can be reduced to 5% [64, 65]. Furthermore, there is a significant cost benefit with respect to maternity and paediatric care [64, 65].

Time lapse imaging (TLI) is now regularly used to help select the best embryo to transfer. There is some evidence that TLI improves clinical outcomes due to increased knowledge about the embryos' development and the closed incubation system. A recent retrospective review of TLI reported an improved live birth rate with TLI (TLI 36.8 versus standard culture 33.9%, adjusted odds ratio 1.28, 95% CI 1.05–1.57). A reduction in preterm birth and low birth weight has also been reported [66].

Luteal Phase After IVF

The embryo transfer procedure usually takes 5–10 minutes. The procedure should be performed under ultrasound guidance rather than using the 'clinical touch' method, as this results in significant increase in ongoing pregnancy (OR 1.51 95% CI 1.31–1.74) and livebirth rates (OR 1.78 95% CI 1.19–2.67) [66]. After embryo transfer the patient can go about her normal daily activities. Indeed, inactivity is best avoided as the 2 weeks up to the pregnancy test are hard for couples to cope with as they are no longer attending the clinic for regular scans and monitoring. It is usual to provide luteal support until the results of the pregnancy test are known and this itself can delay the onset of menstruation and give the couple false hope. Luteal support can be provided by either hCG or parenteral or vaginal progesterone. The administration of hCG should be avoided if there is any risk of OHSS as it will continue to stimulate the ovaries, while exogenous progesterone will of course replace the secretion of the corpora lutea. Many clinics have now stopped giving hCG because OHSS is not always easy to predict.

Pregnancy Rates After IVF

A clinical pregnancy is defined as the ultrasound visualization of a gestational sac. Biochemical pregnancies are so named if hCG is present in the serum (in the absence of exogenously administered hCG for luteal support) yet bleeding occurs before a gestational sac is seen on ultrasound. It is a sensible convention not to include biochemical pregnancies in treatment results and care must be taken when comparing the results of different clinics or studies to ensure that the same definitions of pregnancy have been used.

Modifications to the treatment process, from superovulation strategies to create a larger cohort of mature oocytes, through to the advances in culture technology to allow embryos to thrive in the laboratory have led to a steady increase in live birth rates over the last 20 years with the overall live birth rate per cycle in the UK greater than 25% per cycle. Approximately 68 000 assisted conception treatments are performed annually in the UK, resulting in approximately 1% of all births. There are huge variations in both provision and outcomes of assisted conception treatments around Europe (and the globe).

The chance of a pregnancy following a single cycle of IVF is now approximately 30–40% in the larger units. With the eSET policy supported in the UK, the chance of multiple pregnancy is now 11%, with the majority being twin pregnancies. The miscarriage rate is about 20% and the chance of an ectopic pregnancy is approximately 5%.

The pregnancy rates achieved by IVF equate favourably with those expected for a couple without infertility when adjusted for the age of the female partner. Cumulative conception and live birth rates, calculated by life table analysis, provide the best form of comparison between treatments, although they do not take into consideration couples who drop out of treatment because they are perceived as having a poor chance or because they cannot cope with the stresses of the therapy. The major factors that determine the chance of an ongoing pregnancy are the age of the woman, with rates declining over the age of 35, increasing duration of infertility, parity, and the number of oocytes collected. Not surprisingly, couples who have achieved a pregnancy are more likely to do so if they try again.

Most IVF cycles fail after embryo transfer, and so research has focused on trying to identify which are the correct embryo/s to transfer. Non-invasive ways to assess embryo health have focused on embryo metabolism, and in particular their amino acid profile [67, 68]. Metabolically quiet embryos seem to have more developmental potential than those with a high amino acid turn over. Although these methods have appeared promising, there continues to be limited evidence of any benefit when metabolomic testing is implemented [69].

A high proportion of embryos are karyotypically abnormal, with this increasing substantially with progressing age. Preimplantation genetic screening (PGS) of those patients at highest risk of aneuploidy has been considered; initial outcomes were disappointing [70], but with changes in technique these are improving. Trophectoderm biopsy is currently favoured, for those embryos reaching the blastocyst stage. Use of next generation screening (NGS) rather than array comparative genomic hybridization (aCGH) is providing promise with improved implantation rates (NGS 71.6% vs. aCGH 64.6%) and live birth rates (NGS 62% vs. aCGH 54.4%). These improvements are in part due to the ability of NGS to detect mosaic embryos and those with partial aneuploidies or triploidy [71].

Micromanipulation of Gametes for Severe Male Factor Infertility

Standard IVF requires the presence of more than 500 000 motile sperm in the total ejaculate. In cases where the sperm count is lower, fertilization can be assisted by a variety of micromanipulation techniques, such as ICSI—the injection of a single spermatozoon directly into the cytoplasm (ooplasm) of the oocyte, which has revolutionized the management of male infertility and has provided the possibility of a pregnancy for men who previously would have required their partners to undergo donor insemination.

ICSI can be used not only for men with profound oligozoospermia or asthenoteratozoospermia but also for those with obstructive azoospermia, after microsurgical or direct aspiration of sperm from either the epididymis or the testis. Fertilization rates with ICSI are in the region of 60%, irrespective of the origin of the sperm, providing 90% of couples with an embryo transfer and chance of a pregnancy. Pregnancy rates after ICSI are the same as after IVF.

There is some evidence for an increased rate of strand breakages in the DNA of sperm from men with subfertility, some of whom have cystic fibrosis or are carriers of cystic fibrosis mutations or other recessive gene anomalies. Furthermore, the stage at which genomic imprinting takes place is not known and there is a suggestion that genes may be modified in the epididymis—in other words, distal to the site of aspirated testicular sperm. The data on children born to date as a result of micromanipulation techniques are reassuring with respect to major congenital abnormalities but there is an increased rate of sex chromosome anomalies.

Cryopreservation of Gametes and Embryos

Cryopreservation of sperm and oocytes may offer the preservation of fertility for those about to undergo potentially sterilizing therapy for cancer. The introduction of vitrification as a freezing technique has revolutionized the clinical outcome of oocyte freezing. Clinical pregnancy rates equivalent to fresh embryo transfers have been reported [72]. With this improvement in outcome, there has been a steady rise in those wishing to freeze their oocytes for 'social reasons', and in doing so delaying starting a family. This approach should be taken with caution as too many choose to freeze oocytes at an age when fertility is already in significant decline. It is also possible to cryopreserve ovarian tissue followed by reimplantation or *in vitro* culture of follicles to provide a chance of viable oocytes for women who are about to undergo chemo/radiotherapy or have an oophorectomy.

Embryo survival is in the region of 70% and if individual blastomeres are damaged, as each is pluripotent, there appears to be no harmful effect on the developing fetus. Thawed embryos are transferred time-accordingly to stage of freezing after ovulation in carefully monitored natural cycles or after the commencement of progesterone therapy in artificial cycles in which pretreatment has been performed, with a GnRH analogue and then with oral oestradiol which is administered until the endometrium has developed adequately. There is no apparent difference in clinical outcome between natural or hormonal replacement cycles [73]. With vitrification technology yielding superior clinical outcomes over older slow freeze techniques, there has been a move by some clinics to consider frozen embryo transfer cycles for all, foregoing any initial fresh embryo transfer. There appears to be little difference in clinical outcome between fresh or frozen transfer cycles when cumulative livebirth rates are considered [74]; although as would be expected there will be a delay to pregnancy for some women.

Cryopreservation has enabled sperm donation to be an achievable option for those men who are found to be azoospermic and for whom surgical sperm retrieval is unsuccessful. It is also a valuable necessity for those in same—sex female relationships or those wishing to embark on motherhood as a single parent. Similarly, oocyte donation has become more accessible with cryopreservation, although for many who require donor eggs, they are often donated 'fresh' allowing fertilization and embryo cryopreservation where possible.

Preimplantation Genetic Testing

The ability for couples with a known genetic condition or balanced translocation, to have unaffected children, or at most be carriers of the condition, has been made possible by the development of preimplantation genetic testing (PGT, formerly "diagnosis" (PGD)). Cystic fibrosis or Huntingdon's chorea are examples of conditions that can be prevented being passed on to offspring. When the genetic mutation and surrounding genetic region are known for the affected

person, individualized genetic probes can be developed to test material from developing embryos. Trophectoderm biopsy from a blastocyst is now favoured (as with PGS), with most clinics freezing embryos until the embryos' genetic identity is known. If there are unaffected or carrier embryos, these can then be transferred in a subsequent frozen transfer cycle. For this to have a chance of success, there should be a reasonable ovarian reserve and adequate sperm numbers to provide a group of potential viable embryos that can reach blastocyst and withstand a biopsy and freezing. Alternatively, women can choose to have antenatal screening or testing to establish if their pregnancy is affected.

Surrogacy

IVF surrogacy is an option for women with ovaries but without a uterus, either because of a congenital absence (e.g. Rokitansky syndrome) or after hysterectomy (e.g. after severe obstetric haemorrhage or cervical carcinoma), or for women for whom a pregnancy would be a medical risk (e.g. severe heart or lung disease). Sperm must be frozen and quarantined for 6 months to reduce the risk of infection with HIV. A standard IVF regimen is used and the surrogate host prepared for a frozen embryo replacement cycle. Egg collection can sometimes be difficult if the ovaries are situated high in the abdomen, in which case a transabdominal approach may be required.

Straight surrogacy is another option, less commonly performed within clinics, in which the surrogate host donates her own oocytes either to be inseminated *in vitro*, in a standard IVF protocol, or *in vivo*, in an IUI protocol. There are strict regulations concerning surrogacy arrangements and few clinics offer this treatment because of ethical concerns and the complexities of the arrangements. Key components of a successful programme are an experienced counsellor and the selection of properly motivated surrogates who are fully informed of all of the IVF processes, their risks, and complications.

Ovarian Hyperstimulation Syndrome

The OHSS is a consequence of superovulation therapy for assisted conception procedures. This potentially fatal condition is avoidable by the judicious use of gonadotropins and careful monitoring of stimulation regimens. Women who are at particular risk of developing the syndrome include those who have polycystic ovaries and those who are young (under 30 years).

The pathophysiological hallmark of the OHSS is a sudden increase of vascular permeability which results in the development of a massive extravascular exudate. This exudate accumulates primarily in the peritoneal cavity, causing a protein rich ascites. Loss of *fluid* into the 'third' space causes a profound fall in intravascular volume, haemoconcentration, and suppression of urine formation. Loss of *protein* into the third space causes a fall in plasma oncotic pressure which results in further loss of intravascular fluid. Secondary hyperaldosteronism occurs and may cause hyponatraemia.

The syndrome is graded according to severity. Mild ovarian hyperstimulation is characterized by fluid accumulation, as evidenced by weight gain, abdominal distension, and discomfort. Ultrasound examination shows enlarged ovaries with a mean diameter greater than 5 cm but less than 8 cm. Grade 2 (moderate) ovarian hyperstimulation is associated with the development of nausea and vomiting. The ovarian enlargement and abdominal distension are greater, with ovarian diameter 8–12 cm, and cause more discomfort and dyspnoea. Ascites can be detected by ultrasound.

Grade 3 (severe) OHSS is a life-threatening condition in which there is clinical evidence of contraction of the intravascular volume (subnormal central venous pressure with reduced cardiac output), severe expansion of the third space (tense ascites, pleural, and pericardial effusions, all of which compromise the circulation and breathing), severe haemoconcentration, and the development of hepatorenal failure. In addition to the circulatory crisis these patients are at risk from intravascular thrombosis. Deaths have been recorded in women with grade 3 OHSS, caused usually by cerebrovascular thrombosis, renal failure, or cardiac tamponade resulting from pericardial effusion. Some authors have classified the most severe group as 'Critical OHSS', when the haematocrit is >55%, oliguria/anuria, thromboembolism, and acute adult respiratory distress syndrome (ARDS).

Risk Factors for OHSS

OHSS generally only occurs after overstimulated ovaries have been exposed to human chorionic gonadotropin (hCG). The condition therefore results most commonly when sensitive ovaries are exposed to gonadotropin preparations that contain FSH and then to hCG. The finding that severe OHSS is often associated with pregnancy is probably related to the persistence of hCG in this situation. Even when the ovaries have been severely overstimulated, OHSS can usually be prevented by avoiding exposure of the ovaries to hCG and/or LH. Thus in the context of a woman undergoing a cycle of ovarian stimulation while donating oocytes, hyperstimulation is likely to be a self-limiting situation as a pregnancy, by definition, will not occur.

In IVF the rate of OHSS varies in published series from 1% to 10%, being highest in those combining gonadotropin stimulation with treatment with a GnRH analogue. Severe cases occur in 0.25–8% of IVF cycles with mild cases occurring in up to 33% of cases [75]. A distinction has been made between early and late OHSS, with those presenting early (that is, 3–7 days after hCG administration) having significantly higher serum oestradiol concentrations and more follicles than those presenting late (12–17 days after hCG). Those presenting early usually have a self-limiting condition of relatively short duration while those presenting late are more likely to be pregnant and have a severe and more prolonged form of the syndrome, due to persistent stimulation of the ovaries by hCG from the placenta.

The greatest cause of morbidity and potential mortality in OHSS is from thromboembolism. When considering the pathophysiology of the OHSS it is easy to appreciate the potential risk of deep venous thrombosis (DVT) and thromboembolic events. Not only is there a hypercoagulable state but also the combination of enlarged ovaries and ascites leads to reduced venous return from the lower limbs, which combined with immobility places the patient at risk of DVT. Venous thrombosis in the lower limb most often resolves without long-term sequelae, unless pulmonary embolism occurs, which may be fatal. Upper limb venous thrombosis may lead to disabling long-term disability, with persistent discomfort, cramp, weakness, and cold hands. Cerebral thrombosis may resolve completely

or lead to various forms of long-term disability. The prevalence of thrombophilia may be increased in women with severe OHSS and prophylactic screening for thrombophilia has been advocated in those who have experienced severe OHSS [76].

Management of the OHSS

It goes without saying that prevention is the key and this can be achieved by using mild stimulation regimens, particularly in those at increased risk (young women and those with polycystic ovaries). The antagonist protocol is recommended for those at high risk of OHSS due to the lower incidence of OHSS, in part due to a shorter duration of gonadotrophin stimulation. Use of the GnRH antagonist then allows the GnRH agonist to be used as a 'trigger' in place of hCG, utilizing the natural LH within the women's system and further reducing the risk of OHSS. Other options for those with a high follicle count (>20–25 follicles) include: withhold hCG and not perform the oocyte retrieval or collect the oocytes and cryopreserve all embryos, as by not transferring an embryo the potential for severe late OHSS is avoided. For those at high risk of OHSS, Dopamine agonists such as Cabergoline can be used shortly after hCG administration to reduce the likelihood of moderate-severe OHSS and ovarian volume with no detriment to clinical pregnancy rates. Mild ovarian hyperstimulation is common and is managed expectantly, its importance being that it should alert both patient and doctor to the risk of a more severe condition developing, particularly if a pregnancy occurs. The patient should be encouraged to weigh herself daily and take plenty of oral fluids. A marked increase in weight (more than 5 kg) with the development of abdominal distension, nausea, and vomiting indicate the onset of grade 2 hyperstimulation and the need for hospital review. In non-conception cycles, moderate ovarian hyperstimulation can be expected to resolve with the development of menstruation, although the ovarian cysts may persist for a month or more.

Patients with grade 2 hyperstimulation need reassurance and explanation, together with rest. Oral fluids are encouraged although vomiting may make an intravenous infusion necessary. If hospitalized, TED stockings are advised to reduce the risk of deep vein thrombosis. Adequate analgesia is required. Preferred drugs are paracetamol, with or without codeine, and pethidine for very severe pain.

The development of clinically detectable and usually painful ascites, together with a deterioration in respiration, circulation and renal function indicates the development of severe grade 3 hyperstimulation and may require admission to an intensive care unit. The intravascular volume should be monitored by measurements of central venous pressure, renal function by meticulous attention to input and urine output and haemoconcentration by measurement of haematocrit, whose level reflects intravascular volume depletion and blood viscosity. A haematocrit of over 45% is a serious warning sign and a measurement greater than 55% signals a life-threatening situation. There may be a striking leucocytosis, the WBC count rising up to 40 000/ml. Measurement of body weight, serum urea, creatinine, and electrolytes, together with serum albumin and liver function tests and periodic assessments of the coagulation profile are mandatory.

Infusion of colloid (e.g. human albumen or 6% hydoxyethyl starch (HES)) is required to maintain intravascular volume, as indicated by restoration of a normal central venous pressure and improved haematocrit. It can be implemented on a day case basis if the clinic has capacity for careful patient monitoring. Crystalloid (normal saline usually) is administered for rehydration, although with careful monitoring of fluid balance. Prophylactic low molecular weight heparin should be given to prevent thromboembolism and as the risk continues up to the end of the first trimester of pregnancy there is an argument to continue heparin until that time [77].

A further concern is the development of hyponatraemia, secondary to antidiuretic hormone hypersecretion. If urine output remains suppressed despite restoration of central venous pressure and rehydration, abdominal paracentesis, under ultrasound guidance, should be undertaken. The indications for this procedure are therefore the need for symptomatic relief of a tense ascites, oliguria, rising serum creatinine, falling creatinine clearance, and haemoconcentration unresponsive to medical therapy. Severe oliguria or renal failure persisting despite these measures usually necessitate dialysis.

Paracentesis of hydrothorax should be considered for relief of dyspnoea. Cardiac tamponade from pericardial effusion may prove fatal if not rapidly relieved. Careful cardiology assessment together with cardiac ultrasound should therefore feature in the management of these patients. One must be aware of the possibility of re-accumulation of fluid in any of these cavities.

OHSS is a condition that should be taken extremely seriously because of the physical and emotional distress that it can cause and the thromboembolic risks. The triennial confidential enquiry into maternal deaths and morbidity in the UK reports on the numbers and causes of maternal mortality [78]. The report for 2003–2005, entitled *Saving Mothers' Lives,* had for the first time recorded deaths related to OHSS. In the triennium 2003–2005 there were four deaths out of approximately 119 641 IVF stimulation cycles, a mortality rate of 1:30 000. While problems with appropriate follow-up and expert care were identified in the management of these patients, it goes without saying that there is no acceptable rate of mortality as a result of fertility treatment. Reassuringly, the most recent report entitled *Saving Lives, Improving Mothers' Care* does not attribute any deaths directly to OHSS although there were 6 women (3%) who died between 2013 and 2015 who had achieved a pregnancy through assisted reproductive techniques [79].

REFERENCES

1. Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment seeking: potential need and demand for infertility medical care. *Hum Reprod*, 2007; **22**: 1506–12.
2. Green E, Vessey M. The prevalence of subfertility: a review of the current confusion and a report of two new studies. *Fertil Steril*, 1990; **54**: 978–83
3. Datta J, Palmer MJ, Tanton C, *et al.* Prevalence of infertility and help seeking among 15000 women and men. *Hum Reprod*, 2016; **31**: 2108–18.
4. Gunnell DJ, Ewings P. Infertility prevalence, needs assessment and purchasing. *J Publ Health Med*, 1994; **16**: 29–35.
5. Buckett W, Bentick B. The epidemiology of infertility in rural population. *Acta Obstet Gynecol Scand*, 1997; **76**: 233–7.
6. Birth Registrations England, Wales, Scotland and Northern Ireland 1935 to 2002: Office for National Statistics, General

Register Office for Scotland, Northern Ireland Statistics and Research Agency. Birth order, England & Wales: Office for National Statistics, UK 2002-based national population projections, 2003 to 2035: Government Actuary's Department Completed family size 2007.
7. OPCS (1994) Fertility trends in England and Wales: 1984–94. London: Birth Statistics, OPCS, HMSO, 1994. Available at: https://www.statistics.gov.uk (accessed: 12 April 2020).
8. Childbearing for women born in different years, England and Wales: 2016. Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/childbearingforwomenbornindifferentyearsenglandandwales/2016 (accessed: 12 April 2020).
9. Balen AH, Anderson R. Impact of obesity on female reproductive health: British Fertility Society, Police and Practice Guidelines. *Hum Fertil (Camb)*, 2007; **10**: 195–206.
10. Provost MP, Acharya KS, Acharya CR, *et al*. Pregnancy outcomes decline with increasing body mass index: analysis of 239,127 fresh autologous *in vitro* fertilization cycles from the 2008–2010 Society for Assisted Reproductive Technology registry. *Fertil Steril*, 2016, **105**: 663–9.
11. Linné Y. Effects of obesity on women's reproduction and complications during pregnancy. *Obes Rev*, 2004; **5**: 137–43.
12. Balen AH, Michelmore K. What is polycystic ovary syndrome? Are national views important?. *Hum Reprod*, 2002; **17**: 2219–27.
13. Stewart CAW, Schoeman SA, Booth RA, Smith SD, Wilcox MH, Wilson JD. Assessment of self taken swabs versus clinician taken swab cultures for diagnosing gonorrhea in women: single centre, diagnostic accuracy study. *BMJ*, 2012; **345**: e8107.
14. Wilcox AJ, Dunson D, Baird DD. The timing of the 'fertile window' in the menstrual cycle: day specific estimates from a prospective study. *BMJ*, 2000; **321**: 1259–62.
15. Broekmans FJ, Kwee J, Hendricks DJ, Mol BW, Lambalk CB. A systematic review of test predicting ovarian reserve and IVF outcome. *Hum Reprod Update*, 2006; **12**: 685–718.
16. Broer SL, Eijkemans MJC, Scheffer GJ, *et al*. Anti-Mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. *J Clin Endocrinol Metab*, 2011, **96**: 2532–9.
17. Tremellen K, Savulescu J. Ovarian reserve screening: a scientific and ethical analysis. *Hum Reprod*, 2014. **29**: 2606–14.
18. The number of small antral follicles (2–6 mm) determines the outcome of endocrine ovarian reserve tests in a subfertile population. *Human Reprod*, 2007; **22**: 1925–31.
19. Nelson SM, Yates RW, Fleming R. Serum anti-Mullerian hormone and FSH: prediction of live birth and extremes of response in stimulated cycles—implications for individualization of therapy. *Hum Reprod*, 2007; **22**: 2414–21.
20. La Marca A, Volpe A. Antimullerian hormone (AMH) in female reproduction: is measurement of circulating AMH a useful tool? *Clin Endocrinol*, 2006; **64**: 603–10.
21. Fleming R, Fairbairn C, Blaney C, Lucas D, Gaudoin M. Stability of AMH measurement in blood and avoidance of proteolytic changes. *Reprod Biomed Online*, 2013; **26**: 130–2.
22. ESHRE guideline on the management of recurrent pregnancy loss. Available at: https://www.eshre.eu/Guidelines-and-Legal/Guidelines/Recurrent-pregnancy-loss.aspx (accessed: 12 April 2020).
23. The Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod*, 2004; **19**: 41–7.
24. Balen AH, Laven JSE, Tan SL, Dewailly D. Ultrasound assessment of the polycystic ovary: international consensus definitions. *Hum Reprod Update*, 2003; **9**: 505–14.
25. Zhao J, Zhang Q, Li Y. The effect of endometrial thickness and pattern measured by ultrasonography on pregnancy outcomes during IVF-ET cycles. *Reprod Biol Endocrinol*, 2012; **10**: 100.
26. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Cambridge: Cambridge University Press, 2010.
27. Cooper TG, Noonan E, von Eckardstein S, *et al*. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*, 2010; **16**: 231–45.
28. Kruger TF, Coetzee K. The role of sperm morphology in assisted reproduction. *Hum Reprod Update*, 1999; **5**: 172–8.
29. Aitken RJ, Clarkson JS, Fishel S. Generation of reactive oxygen species, lipid peroxidation and human sperm function. *Biol Reprod*, 1989; **40**: 183–97.
30. Bisht S, Faiq M, Tolahunase M, Dada R. Oxidative stress and male infertility. *Nat Rev Urol*, 2017; **14**: 470–85.
31. Saravelos HG, Li T-C, Cooke ID. An analysis of the outcome of microsurgical and laparoscopic adhesiolysis for infertility. *Hum Reprod*, 1995; **10**: 2887–94.
32. Strandell A, Lindhard A, Waldenstrom U, Thorburn J, Janson PO, Hamberger L. Hydrosalpinx and IVF outcome: a prospective, randomized multicentre trial in Scandinavia on salpingectomy prior to IVF. *Hum Reprod*, 1999; **14**: 2762–9.
33. Johnson N, van Voorst S, Sowter MC, Strandell A, Mol BWJ. Surgical treatment for tubal disease in women due to undergo *in vitro* fertilization. *Cochrane Database Syst Rev*, 2010; **1**: CD002125.
34. Wu B, Zhang Q, Zhao J, Wang Y, Xu D, Li Y. Pregnancy outcome of *in vitro* fertilization after Essure and laparoscopic management of hydrosalpinx: a systematic review and meta-analysis. *Fertil Steril*, 2017; **108**: 84–95.
35. Borgfeldt C, Andolf E. Transvaginal ultrasonographic findings in the uterus and the endometrium: low prevalence of leiomyoma in a random sample of women age 25–40 years. *Acta Obstet Gynecol Scand*, 2000; **79**: 202–7.
36. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. *Am J Obstet Gynecol*, 2003; **188**: 100–7.
37. Somigliana E, Vercellini P, Daguati R, Pasin R, De Giorgi O, Crosignani PG. Fibroids and female reproduction: a critical analysis of the evidence. *Hum Reprod Update*, 2007; **13**: 465–76.
38. Practice Committee of the American Society for Reproductive Medicine. Removal of myomas in asymptomatic patients to improve fertility and/or reduce miscarriage rate: a guideline. *Fertil Steril*, 2017; **108**: 416–25.
39. Vessey MP, Villard-Macintosh L, Painter R. Epidemiology of endometriosis in women attending family planning clinics. *Br Med J*, 1993; **306**: 182–4.
40. Bosteels J, Van Herendael B, Weyers S, D'Hooghe T. The position of diagnostic laparoscopy in current fertility practice. *Hum Reprod Update*, 2007; **13**: 477–85.
41. Hughes E, Brown J, Collins JJ, Farquhar C, Fedorkow DM, Vandekerckhove P. Ovulation suppression for endometriosis. *Cochrane Database Syst Rev*, 2007; **3**: CD000155.
42. Sallam HN, Garcia-Velasco JA, Dias S, Arici A. Long-term pituitary down-regulation before *in vitro* fertilization (IVF) for

women with endometriosis. *Cochrane Database Syst Rev*, 2006; **1**: CD004635.
43. Duffy JMN, Arambage K, Correa FJS, *et al*. Laparoscopic surgery for endometriosis. *Cochrane Database Syst Rev*, 2014.
44. Hart RJ, Hickey M, Maouris P, Buckett W. Excisional surgery versus ablative surgery for ovarian endometriomata. *Cochrane Database Syst Rev*, 2008; **2**: CD004992.
45. Brown J, Farquhar C. Endometriosis: an overview of Cochrane reviews. *Cochrane Database Syst Rev*, 2014: **3**: CD009590.
46. Younes G, Tulandi T. Effects of adenomysosis on *in vitro* fertilization treatment outcomes: a meta-analysis. *Fertil Steril*, 2017; **108**: 483–90.
47. Rousseau S, Lord J, Lepage Y, van Campenhout J. The expectancy of pregnancy for 'normal' infertile couples. *Fertil Steril*, 1983; **40**: 768.
48. Hull MGR. Infertility treatment: relative effectiveness of conventional and assisted conception methods. *Hum Reprod*, 1992; **7**: 785–96.
49. Bhattacharya S, Harrild K, Mollison J, *et al*. Clomifene citrate or unstimulated intrauterine insemination compared with expectant management for unexplained infertility: pragmatic randomised controlled trial. *BMJ*, 2008; **337**: 716–24.
50. Hughes E, Brown J, Collins JJ, Vanderkerchove P. Clomiphene citrate for unexplained subfertility in women. *Cochrane Database Syst Rev*, 2010; **1**: CD000057.
51. Veltman-Verhulst SM, Hughes E, Ayeleke R, Cohlen BJ. Intrauterine insemination for unexplained subfertility. *Cochrane Database Syst Rev*, 2016; **2**: CD001838.
52. Farquhar CM, Liu E, Armstrong S, Arroll N, Lensen S, Brown J. Intrauterine insemination with ovarian stimulation versus expectant management for unexplained infertility (TUI): pragmatic, open-label, randomized, controlled, two-centre trial. *Lancet*, 2018; **391**: 441–50.
53. Pandian Z, Gibreel A, Bhattacharya S. *In vitro* fertilisation for unexplained subfertility. *Cochrane Database Syst Rev*, 2012; **4**: CD003357.
54. Tarlatzis BC, Fauser BC, Kolibianikis EM, Diedrich K, Devroey P, on behalf of the Brussels GnRH Antagonist Consensus Workshop Group. GnRH antagonist in ovarian stimulation for IVF. *Hum Reprod Update*, 2006; **12**: 333–40.
55. Al-Inany HG, Youssef MA, Ayeleke RO, Brown J, Lam WS, Broekmans FJ. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. *Cochrane Database Syst Rev*, 2016; **4**: CD001750.
56. Youssef MAFM, Van der Veen F, Al-Inany HG, *et al*. Gonadotrophin releasing hormone agonist versus HCG for oocyte triggering in antagonist-assisted reproductive technology. *Cochrane Database Syst Rev*, 2014; **10**: CD008046.
57. Christopoulos G, Vlismas A, Carby A, Lavery S, Trew G. GnRH agonist trigger with intensive luteal phase support vs. human chorionic gonadotrophin trigger in high responders: an observational study reporting pregnancy outcomes and incidence of ovarian hyperstimulation syndrome. *Hum Fertil (Camb)*, 2016; **19**: 199–206.
58. van Wely M, Kwan I, Burt AL, *et al*. Recombinant versus urinary gonadotrophin for ovarian stimulation in reproductive technology cycles. *Cochrane Database Syst Rev*, 2011; **2**: CD005354.
59. Weiss NS, Nahuis M, Bayram N, Mol BJ, Van der Veen F, van Wely M. Gonadotrophins for ovulation induction in women with polycystic ovarian syndrome. *Cochrane Database Syst Rev*, 2015; **9**: CD010290.
60. Glujovsky D, Farquhar C, Retamar AMQ, Sedo CRA, Blake D. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. *Cochrane Database Syst Rev*, 2016: **6**: CD002118.
61. Papanikolaou EG, Camus M, Kolibianikis EM, van Landuyt L, van Steirteghem A, Devroey P. In vitro fertilization with single blastocycst-stage versus single cleavage stage embryos. *NEJM*, 2006; **354**: 1139–46.
62. Human Fertilisation & Embryology Authority. Fertility treatment 2014–2016: trends and figures. Available at: https://www.hfea.gov.uk/media/2544/hfea-fertility-treatment-2014-2016-trends-and-figures.pdf (accessed: 12 April 2020).
63. Practice Committee of the American Society for Reproductive Medicine. Guidance on the limits to the number of embryos to transfer: a committee opinion. *Fertil Steril*, 2017; **107**: 901–3
64. Kjellberg AT, Crlsson P, Bergh C. Randomized single versus double embryo transfer: obstetric and paediatric outcome and a cost-effectiveness analysis. *Hum Reprod*, 2006; **21**: 210–16.
65. Fiddelers AAA, van Montfoort APA, Dirksen CD, *et al*. Single versus double embryo transfer: cost-effectiveness analysis alongside a randomized, clinical trial. *Human Reprod*, 2006; **21**: 2090–7.
66. Mascarenhas M, Fox SJ, Thompson K, Balen AH. Cumulative live birth rates and perinatal outcomes with the use of time-lapse imaging incubators for embryo culture: a retrospective cohort study of 1882 ART cycles. *BJOG*, 2019; **126**: 280–6.
67. Abou-Setta AM, Mansour RT, Al-Inany HG, Aboulghar MM, Aboulghar MA, Serour GI. Among women undergoing embryo transfer, is the probablility of pregnancy and livebirth improved with ultrasound guidance over clinical touch alone? A systematic review and meta-analysis of prospective randomised trials. *Fertil Steril*, 2007; **88**: 333–41.
68. Brison DR, Houghton FD, Falconer D, *et al*. Identification of viable embryos in IVF by non-invasive measurement of amino acid turnover. *Hum Reprod*, 2004; **19**: 2319–24.
69. Siristatidis CS, Sertedaki E, Vaidakis D. Varounis C, Trivella M. Metabolomics for improving pregnancy outcomes in women undergoing assisted reproductive technologies. *Cochrane Database Syst Rev*, 2018; **3**: CD011872.
70. Twisk M, Mastenbroek S, van Wely M, Heineman MJ, Van der Veen F, Repping S. Preimplantation genetic screening for abnormal number of chromosomes (aneuploidies) in *in vitro* fertilization or intracytoplasmic sperm injection. *Cochrane Database Syst Rev*, 2006; **1**: CD005291.
71. Friedenthal J, Maxwell SM, Munne S, *et al*. Next generation sequencing for preimplantation genetic screening improves pregnancy outcomes compared with array comparative genomic hybridization in single thawed euploid embryo transfer cycles. *Fertil Steril*, 2018; **109**: 627–32.
72. Argyle CE, Harper JC, Davies MC. Oocyte cryopreservation: where are we now?. *Hum Reprod Update*, 2016; **22**: 440–9
73. Ghobara T, Gelbaya TA, Ayeleke R. Cycle regimens for frozen-thawed embryo transfer. *Cochrane Database Syst Rev*, 2017; **7**: CD003414.
74. Wong k, van Wely M, Mol F, Repping S, Mastenbroek S. Fresh versus frozen embryo transfers in assisted reproduction. *Cochrane Database Syst Rev*, 2017, **3**: CD011184.
75. Delvigne A, Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome. *Hum Reprod Update*, 2002; **8**: 559–77.
76. Dulitzky M, Cohen SB, Inbal A, Seidman DS, Soriano D, Lidor A, Mashiach S, Rabinovici J. Increased prevalence

of thrombophilia among women with severe ovarian hyperstimulation syndrome. *Fertil Steril*, 2002; 77: 463–7.
77. Rizk BRMB. *Ovarian Hyperstimulation Syndrome: Epidemiology, Pathophysiology, Prevention and Management.* Cambridge: Cambridge University Press, 2006: 234.
78. Lewis G, ed. *Saving Mothers' Lives—Reviewing Maternal Deaths to Make Motherhood Safer 2003–2005. The Seventh Report on the Confidential Enquiries into Maternal Deaths in the United Kingdom.* London: RCOG, 2007.
79. Knight M, Nair M, Tuffnell D, Shakespeare J, Kenyon S, Kurinczuk JJ, eds. on behalf of MBRRACE-UK. *Saving Lives, Improving Mothers' Care—Lessons Learned to Inform Maternity Care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity, 2013–15.* Oxford: National Perinatal Epidemiology Unit, University of Oxford, 2017.

8.6.2 Female Infertility

Fertility Preservation

Kutluk Oktay and Enes Taylan

Introduction

Treatments with antineoplastic agents and ionizing radiation have been successfully used in the management of a broad range of diseases including neoplastic diseases to various autoimmune and haematological disorders. With the remarkable advancements in modern treatment regiments, the survival rates have significantly improved in paediatric and reproductive age females. However, as the number of patients cured of the disease has increased the number of women facing infertility and early menopause after such treatments has also dramatically risen [1, 2].

Human ovarian reserve consists of a limited number of primordial germ cells numbers of which are determined *in utero*. Those germ cells possess the capacity to produce mature oocytes that can fertilize with sperm and provide offspring throughout the reproductive lifespan. However, treatment regimens, especially ones that include a high dose of alkylating agent and/or pelvic ionizing radiation, can irreversibly deplete ovarian reserve impairing reproductive capacity and shortening the fertile lifespan [3, 4]. In the last few decades, the number of females confronting these challenging issues has larger grown, hence, the necessity of fertility preservation has become more evident. Reproductive medicine specialists in collaboration with scientists from several other scientific disciplines have developed and introduced multiple options for preserving reproductive potential in women under the risk of gonadal failure or ovarian insufficiency (OI).

In this chapter, we will briefly describe the mechanisms of and risk factors for OI and then provide an overview of currently available options for fertility preservation in women.

Ovarian Reserve and Mechanisms of Ovarian Insufficiency

Contrary to continuous testicular germ cell regeneration in males, ovarian germ cell development, and proliferation process in females is limited to earlier intrauterine life. At birth, usually, the human ovary hosts approximately one million primordial germ cells surrounded by a single layer of squamous granulosa cells, and only 300 000–500 000 of those primordial follicles reach pubertal age. The large number of primordial follicles are lost due to an apoptotic process. This remaining pool of primordial follicles serves as the ovarian reserve throughout the reproductive years of a woman and nearly exhausts at menopause [5]. Primordial follicles initiate growth from intrauterine life, and their growth initiation is not dependent on pubertal status. By the age of 37 years, the number of primordial follicles in the ovarian reserve decreases to about 25 000 and the reserve depletes at the time of menopause. Recently, researchers reported the presence of mitotically active cells with stem cell characteristics in the adult ovary that can produce oocytes. However, their role in folliculogenesis and the follicle reserve is still unclear and needs to be further investigated [6, 7].

The Impact of Gonadotoxic Treatments on Ovarian Reserve

With the advancements in modern chemoradiation therapies and surgical practice, a growing group of patients with ovarian insufficiency has emerged. While modern cancer treatments can target cancer cells, in tandem high doses of chemotherapy and pelvic radiotherapy can irreversibly destruct ovarian follicles by causing DNA damage and induce massive germ cell apoptosis [3, 8].

Modern chemotherapeutic agents can have a spectrum of ovarian toxicity, depending on the class of the agent, patient age at the time of treatment, and the cumulative dose [4]. We have shown that the most gonadotoxic agents are those that mainly cause DNA double-strand breaks (DBSs) in oocytes [3]. DNA repair mechanisms can maintain genomic integrity in oocytes under normal circumstances. However, at the level of severe DNA damage due to genotoxic agents, those repair mechanisms remain insufficient and lead to apoptotic death of the oocyte [3]. Among all chemotherapeutic agents, the most gonadotoxic drugs are the members of the alkylating category such as cyclophosphamide [9]. Because these drugs are non-cell-cycle specific chemical compounds they can target and damage dormant primordial follicles that constitute ovarian reserve.

Although menstruation is a poor surrogate for ovarian reserve and fertility, using menstruation as an outcome in women with breast cancer who underwent chemotherapy researchers were able to demonstrate the negative impact of alkylating agents on ovarian function. For example, Gadduci *et al.* reported that after treatment with CMF protocol (cyclophosphamide + methotrexate + 5-fluorouracil) 20–70% of women younger than age 40 experienced

amenorrhea [10]. When CMF protocol was compared to the AC protocol (doxorubicin + cyclophosphamide), significantly lower rates of amenorrhea (69% vs. 34%, respectively) have been reported with the AC protocol [11]. This result is most likely associated with a lower cumulative dose of cyclophosphamide administered in the AC regimen. However, combination of a taxane agent (paclitaxel) with AC (AC-T) did not significantly increase the risk of amenorrhea compared with standard AC regimen [12].

Patient age at the time of chemotherapy is another critical factor that determines the nature of the ovarian response to the insult of the gonadotoxic agent. In women with breast cancer, the incidence of chemotherapy-induced amenorrhea was reported as 15–40% under the age of 30. However, this incidence dramatically increases to 49–100% for women older than 40 years of age [13]. This is because gonadotoxic chemotherapy regimens result in the loss of approximately 10 years' worth of ovarian reserve [14] and it is more likely to push older women towards the menopausal state as they have lower reserve compared to younger women. However, regardless of age, females of all ages including children are expected to experience early menopause after exposure to gonadotoxic chemotherapy agents. Therefore, fertility preservation and completion of the family building as early as possible is critical regardless of the age at chemotherapy exposure in most instances [15].

The Impact of BRCA Mutations on Ovarian Reserve

Mutations in *BRCA1* and *BRCA2* genes are well-known due to their association with the increased risk for multiple hereditary malignancies such as breast and ovarian cancers. *BRCA* genes are members of the ataxia-telangiectasia-mutated (ATM)-mediated DNA damage signalling pathway and are essential for DNA double-strand break (DSB) repair [16]. Moreover, in several studies, we followed by others showed that *BRCA*1 and *2* mutations are associated with diminished ovarian reserve [8, 17, 18]. In our recent report of controlled ovarian stimulation in women with breast cancer by using aromatase inhibitors for fertility preservation, we observed significantly lower ovarian response rates in *BRCA* mutation carries particularly, among those with *BRCA1* mutations [19]. In another study, *BRCA* mutation carrier women were found to experience menopause 3–4 years earlier than healthy controls [20]. Recently, we showed that in ovaries from both mouse and human with BRCA1 mutation the ovarian reserve was significantly reduced due to the increased age-related accumulation of DNA DBSs in primordial follicle oocytes. Additionally, we showed that affected women with *BRCA1* mutations had lower serum anti-Müllerian hormone (AMH) levels compared to controls [8]. In another very recent study, examining ovaries obtained from unaffected *BRCA* mutation carriers and age-matched organ donor cadavers, we provided direct evidence of diminished ovarian reserve and accelerated primordial follicle loss in both women with *BRCA1* and *2* mutations [21].

Based on the current evidence from animal and clinical studies, it appears that the ovarian reserve may be lower in women with BRCA mutations, and as we have shown recently, are more prone to chemotherapy-induced loss of ovarian reserve [22]. Hence, women with BRCA mutation should be informed about the possibility of higher risk of chemo-induced infertility and available fertility preservation options as well.

Table 8.6.2.1 Indications for fertility preservation in reproductive-age women

Benign conditions	Malignant conditions
• Benign ovarian diseases (endometrioma, benign ovarian tumours, etc.) • Haematological and autoimmune diseases that require chemotherapy or HSCT (SLE, sickle cell anaemia, rheumatoid arthritis, severe combined immunodeficiency syndrome, etc.) • Genetic conditions that can cause ovarian insufficiency *(BRCA 1* and *2* mutations, *FMR1* premutation, Mosaic Turner syndrome, etc.) • Conditions that can cause autoimmune oophoritis and ovarian insufficiency (polyglandular autoimmune syndrome type 1 and 2, SLE, pernicious anaemia, myasthenia gravis, viral infections, etc.) • Genetic conditions that can require prophylactic oophorectomy *(BRCA 1* and *2* mutations, *p53* mutation)	• Breast cancer • Hodgkin lymphoma • Non-Hodgkin lymphoma • Endometrial cancer • Uterine sarcomas • Cervical cancer • Ovarian cancer • Rhabdomyosarcoma • Tumours that require pelvic radiotherapy *(colorectal cancers, bladder cancer, cervical cancer, pelvic bone tumours, etc.)*

The Importance of Timely Referral and Intervention

As previously mentioned, OI can be caused by several intrinsic and external factors, and the rate of progression may vary from days to years depending on the underlying reason. Therefore, with early diagnosis patients can be given a chance to preserve their reproductive potential before complete loss of ovarian reserve. Accordingly, guidelines by the American Society of Clinical Oncology (ASCO) and the American Society of Reproductive Medicine (ASRM) for fertility preservation in cancer patients strongly recommend that physicians should inform their patients about the potentially harmful effects of the planned treatments on fertility before the initiation of it and promptly refer patients to reproductive specialist to discuss the risk of ovarian damage and currently available fertility preservation options [23, 24] (Table 8.6.2.1).

In a recent study, we have shown that in breast cancer patients, especially before breast surgery, early referral yields to retrieval of a more significant number of oocytes and embryos being cryopreserved for future fertility and reduced delay to the initiation of chemotherapy [25]. However, surveys evaluating the awareness of physicians regarding the fertility preservation reported that less than half of the oncologists in the United States always or often refer their cancer patients with fertility-related questions to fertility preservation specialist [26]. Fertility and ovarian endocrine health are important components of quality of life [27]. Hence, providing timely and accurate information about available fertility and ovarian preservation options for women of reproductive age with the risk of OI is critical before complete loss of the limited and irreplaceable ovarian reserve.

Current Options for Fertility Preservation

Embryo Cryopreservation

Embryo cryopreservation is a well-established fertility preservation option and currently offers the best chance of livebirth comparing to other fertility preservation options. However, embryo cryopreservation requires a partner or use of donor sperm. In one study, Frozen embryo transfer resulted in a clinical pregnancy rate of 60%, and around 34% live birth in infertility patients with mean age of 35.1 ± 4.03, which is comparable to that of fresh embryo transfer success rates [28]. In women with breast cancer with the mean age of 35.8 ± 4.1, we reported a live birth rate of 45%, which appeared to be superior to those undergoing frozen embryo transfer for infertility [29]. With the utilization of preimplantation genetic screening, the live birth rates can increase up to 77% per embryo transfer in women with average age of 36.7 [28–43] after transfer of euploid frozen-thawed embryos [30].

Oocyte Cryopreservation

For post-pubertal girls and women who do not have a partner or wish to use donor sperm due to legal, ethical or religious considerations, cryopreservation of mature oocytes is an established fertility preservation option. An adequate number of oocytes can be collected following approximately 10–12 days of controlled ovarian hyperstimulation treatment, and mature oocytes can be successfully cryopreserved using a vitrification method. With increasing experience and further advances in cryopreservation methods, fertilization, and pregnancy success rates for frozen-thawed oocytes have approached to those with fresh oocytes in young patients, though success rates with frozen embryos may still be better [31]. We have performed an individual patient data meta-analysis where we calculated age-based success rates for women undergoing egg freezing. Based on this publication we created an online success rate estimator [32]. This estimator can be found at this website link (http://www.fertilitypreservation.org/contents/probability-calculator) and may be useful for counselling patients on egg freezing.

Immature oocytes can also be obtained from patients who cannot undergo ovarian stimulation due to lack of time and even at the time of ovarian tissue harvesting for fertility preservation. Those oocytes can be successfully matured through *in vitro* maturation (IVM) procedure and subsequently cryopreserved. Performing IVM for immature oocytes before cryopreservation rather than post-thaw also results in significantly higher maturation and survival rates [33]. Recently, livebirths following the use of IVM in patients with infertility has been reported; however, it is still an experimental fertility preservation method limited to a number of fertility centres [34].

Controlled Ovarian Hyperstimulation Protocols for Women with Oestradiol-Sensitive Cancer

Controlled ovarian hyperstimulation (COH) can provide development and collection of a considerable number of oocytes in a single treatment cycle and yield multiple embryos for fertility preservation. On the other hand, the growth of numerous follicles at the same time results in elevated circulating oestradiol levels, which is considered unsafe in patients with oestradiol-sensitive cancers. For those patients, we have developed safer ovarian stimulation protocols using tamoxifen and aromatase inhibitors such as letrozole. These medications possess a dual pharmacological effect of stimulating ovarian follicle development and reducing serum oestradiol level concurrently. Use of tamoxifen alone for ovulation induction or in combination with low dose gonadotropins yielded significantly higher numbers of mature oocyte and embryo compared to natural cycle IVF [35, 36].

Likewise, letrozole in combination with gonadotropins can produce comparable outcomes to conventional COH while providing significantly lower oestradiol levels and decreased gonadotropin requirements [37]. We further showed that the letrozole-gonadotropin protocol in breast cancer patients was not associated with an increase in recurrence after short- and mid-term follow-up [38, 39]. Hence, these protocols can be safely utilized for fertility preservation in patients with oestradiol-sensitive tumours.

As COH typically requires 10–12 days of treatment, the delay in the initiation of adjuvant or neoadjuvant chemotherapy is may be of concern. However, studies in breast cancer have shown that initiation of chemotherapy can be delayed up to 12 weeks after breast surgery without any adverse effect on survival and recurrence rates [40, 41].

Random-Start Controlled Ovarian Hyperstimulation

Controlled ovarian stimulation for oocyte/embryo cryopreservation is traditionally initiated after menses at the beginning of the follicular phase. However, this approach may require 2–4 weeks to start ovarian stimulation depending on the patients' menstrual cycle phase at the time of the presentation and cause up to 6 weeks of delay in the initiation of cancer treatments. Recently, studies have shown that there are multiple major follicle recruitment waves during a normal menstrual cycle [42]. Based on this ovarian biological fact, we introduced an ovarian stimulation approach which we named 'random-start ovarian stimulation' [43]. Our and others' studies have shown that ovarian stimulation can be initiated any time during the menstrual cycle, without reducing the yield of oocytes and embryos [43, 44]. Random-start COH offers an effective strategy for emergency fertility preservation by eliminating the long delays for cancer treatments.

Ovarian Tissue Cryopreservation and Transplantation

Oocyte and embryo cryopreservation are currently considered as established methods of fertility preservation by the American Society of Clinical Oncology (ASCO) and the American Society of Reproductive Medicine (ASRM). However, these methods usually require on average 10–12 days of COH treatment to obtain mature oocytes. Women who do not have sufficient time for ovarian stimulation cannot benefit from these options. Moreover, controlled ovarian stimulation is not an appropriate strategy for prepubertal girls. In that instance, the only available approach other than immature oocyte retrieval and *in vitro* maturation is ovarian tissue harvesting and cryopreservation for future transplantation [45].

Cryopreservation of ovarian tissue before the initiation of gonadotoxic treatments enables preservation of a large number of primordial follicles embedded in ovarian cortex. This procedure can be done at any age and at any time of the menstrual cycle without the need for ovarian stimulation [46]. Typically, one ovary is harvested via a simple laparoscopic outpatient procedure. The cortical tissue is then separated from stroma and cut into small pieces and cryopreserved using cryoprotectants and slow freezing methods

[47]. Even though there are some encouraging data on ovarian vitrification, the clinical experience is very limited. When the patient is cured of her disease and desires having a child, the tissues are thawed and autotransplanted via minimally invasive surgical approaches including robotic technology [48, 49]. While the primary purpose of ovarian tissue cryopreservation and transplantation is restoring fertility, this approach has also been used for restoration of endocrine function, potentially resolving menopausal symptoms, and improving the quality of life [50].

Although ovarian tissue cryopreservation and subsequent autotransplantation has previously been considered an experimental fertility preservation method, there have been more than 130 live births with 37.7% of live birth rate after ovarian transplantation since the first case reported by our group [51–54]. As a result of improving success rates, a number of countries and the American Society of Reproductive Medicine in the United States have removed ovarian tissue freezing from experimental category [55].

Despite the concern of reintroducing malignant cells along with ovarian tissue in patients with a history of cancer, studies showed no evidence of cancerous cells in cryopreserved ovarian tissues from non-metastatic breast cancer patients and those with bone and soft tissue tumours [56]. However, in haematological malignancies there is the possibility of the presence of malignant cells in cryopreserved ovarian tissue. Even though clinical significance of this is unknown, patients should be informed about the potential risks, and the ovarian tissue should be carefully examined for malignant cells before proceeding with transplantation [57]. Furthermore, in the instance of leukaemias, the initial induction chemotherapy is generally non-gonadotoxic and does not require fertility preservation. However, fertility preservation is required before these patients receive high dose alkylating agents with or without total body irradiation to precondition before haematopoietic stem cell transplantation. At that point however, the patient is typically in remission and there are no circulating cancer cells. In fact, xenografting studies showed that, leukaemia cells are not present or transmitted via transplantation from ovarian tissue samples obtained from patients who are in remission [58]. Hence in general, the risk of reseeding of cancer cells and causing a relapse by ovarian autotransplantation is low, limited to few case scenarios, and currently theoretical.

The Use of Gonadotropin-Releasing Hormone (GnRH) Analogues for Ovarian Protection

Recently, few randomized studies in women with breast cancer suggested a beneficial effect of gonadotropin-releasing hormone analogues (GnRHa) against chemotherapy-induced damage [59, 60]. However, those studies were marred by numerous issues including the use of unreliable markers such as menstrual status, lack of placebo or blinding, and lack of correction for the difference in desire to conceive between study and control groups [61, 62]. Because irregular or even regular menstrual periods can be observed in patients with occult OI, menstrual status cannot be accepted as a surrogate for ovarian reserve. Currently, the most reliable ovarian reserve marker is AMH, and when it was used with serum follicle-stimulating hormone (FSH) level for defining ovarian failure, studies showed no benefit from GnRHa treatment in cancer patients [63, 64]. Among the several concerns raised against this hypothesis, the major concern is that primordial follicles that constitute the ovarian reserve are quiescent and do not express receptors for gonadotropins or GnRH analogues [65]. Therefore, manipulation of serum gonadotropin or GnRH levels has no plausible direct or indirect effect on primordial follicles [66]. Another critical concern is that gonadotoxic drugs such as cyclophosphamide induce primordial follicle death by causing DNA DBSs in oocytes in a non-cell cycle-dependent fashion; hence there is no mechanism for ovarian suppression by GnRHa to prevent chemotherapy-induced DNA damage in oocytes [2, 3, 8]. Moreover, gonadal suppression using GnRHa induces a hormonal state similar to the prepubertal stage, and if ovarian suppression were to be protective, children of prepubertal age would be resistant to gonadotoxic effects of chemotherapy. However, thousands of female children experience endocrine issues and gonadal failure following treatments with chemotherapeutic agents [2]. Based on these facts, the use of GnRHa for the prevention of ovaries from chemotherapy

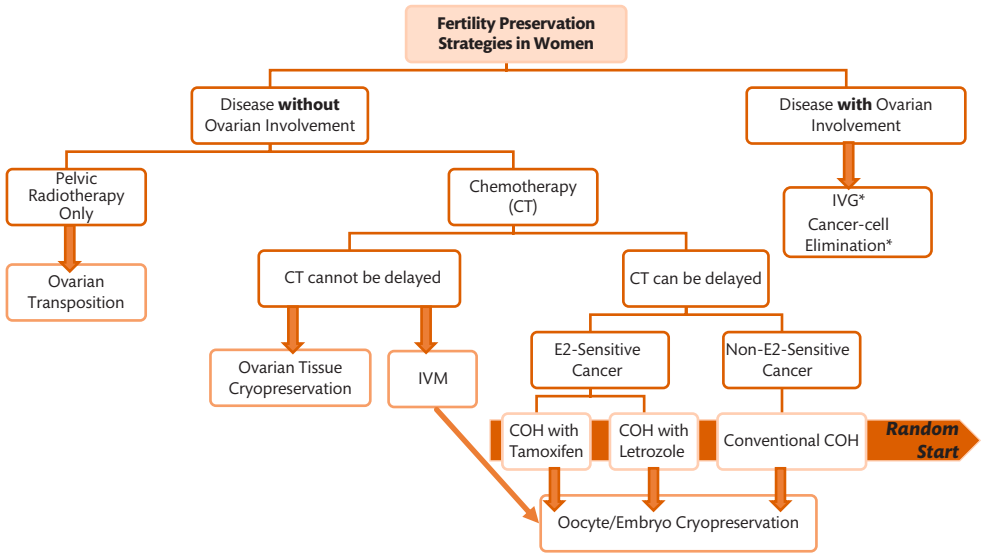


Figure 8.6.2.1 Fertility preservation strategies in women under the risk of ovarian insufficiency. Symbol (*) represents currently theoretical and experimental methods based on animal data. COH, controlled ovarian stimulation; E2, oestradiol; IVG, *in vitro* growth; IVM, *in vitro* maturation.

damage is still controversial and should not be recommended as an effective method of fertility preservation.

Conclusion

Fertility preservation is a rapidly developing field with an increasing pace of advancements in understanding ovarian physiology and cryobiology, and new surgical techniques using minimally invasive approaches. It is becoming a crucial part of cancer survivorship and an essential aspect of comprehensive women's healthcare [67]. Fortunately, there are several non-experimental treatment options including embryo, oocyte, and ovarian tissue cryopreservation, and other emerging ones such oocyte *in vitro* maturation [55, 68–70] (see Figure 8.6.2.1). In order to maximize the benefits from these available options and improve the quality of life, timely referral to fertility preservation counselling should be an integral part of the medical care of reproductive age women.

REFERENCES

1. Levine JM, Kelvin JF, Quinn GP, Gracia CR. Infertility in reproductive-age female cancer survivors. *Cancer*, 2015; **121**: 1532–9.
2. Sklar CA, Mertens AC, Mitby P, *et al*. Premature menopause in survivors of childhood cancer: a report from the childhood cancer survivor study. *J Natl Cancer Inst*, 2006; **98**: 890–6.
3. Soleimani R, Heytens E, Darzynkiewicz Z, Oktay K. Mechanisms of chemotherapy-induced human ovarian aging: double strand DNA breaks and microvascular compromise. *Aging (Albany NY)*, 2011; **3**: 782–93.
4. Bedoschi G, Navarro PA, Oktay K. Chemotherapy-induced damage to ovary: mechanisms and clinical impact. *Future Oncol*, 2016; **12**: 2333–44.
5. Oktem O, Oktay K. The ovary: anatomy and function throughout human life. *Ann N Y Acad Sci*, 2008; **1127**: 1–9.
6. Johnson J, Canning J, Kaneko T, Pru JK, Tilly JL. Germline stem cells and follicular renewal in the postnatal mammalian ovary. *Nature*, 2004; **428**: 145–50.
7. White YA, Woods DC, Takai Y, Ishihara O, Seki H, Tilly JL. Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. *Nat Med*, 2012; **18**: 413–21.
8. Titus S, Li F, Stobezki R, *et al*. Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci Transl Med*, 2013; **5**: 172ra121.
9. Rodriguez-Wallberg KA, Oktay K. Fertility preservation in women with breast cancer. *Clin Obstet Gynecol*, 2010; **53**: 753–62.
10. Gadducci A, Cosio S, Genazzani AR. Ovarian function and childbearing issues in breast cancer survivors. *Gynecol Endocrinol*, 2007; **23**: 625–31.
11. Bines J, Oleske DM, Cobleigh MA. Ovarian function in premenopausal women treated with adjuvant chemotherapy for breast cancer. *J Clin Oncol*, 1996; **14**: 1718–29.
12. Abusief ME, Missmer SA, Ginsburg ES, Weeks JC, Partridge AH. The effects of paclitaxel, dose density, and trastuzumab on treatment-related amenorrhea in premenopausal women with breast cancer. *Cancer*, 2010; **116**: 791–8.
13. Hortobagyi GN, Buzdar AU, Marcus CE, Smith TL. Immediate and long-term toxicity of adjuvant chemotherapy regimens containing doxorubicin in trials at M.D. Anderson Hospital and Tumor Institute. *NCI Monogr*, 1986; **1**: 105–9.
14. Oktem O, Oktay K. Quantitative assessment of the impact of chemotherapy on ovarian follicle reserve and stromal function. *Cancer*, 2007; **110**: 2222–9.
15. Rodriguez-Wallberg KA, Oktay K. Fertility preservation and pregnancy in women with and without *BRCA* mutation-positive breast cancer. *Oncologist*, 2012; **17**: 1409–17.
16. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol*, 2006; **7**: 739–50.
17. Wang ET, Pisarska MD, Bresee C, *et al*. BRCA1 germline mutations may be associated with reduced ovarian reserve. *Fertil Steril*, 2014; **102**: 1723–8.
18. Phillips KA, Collins IM, Milne RL, *et al*.; Kathleen Cuningham Consortium for Research into Familial Breast Cancer (kConFab). Anti-Müllerian hormone serum concentrations of women with germline BRCA1 or BRCA2 mutations. *Hum Reprod*, 2016; **31**: 1126–32.
19. Oktay K, Kim JY, Barad D, Babayev SN. Association of BRCA1 mutations with occult primary ovarian insufficiency: a possible explanation for the link between infertility and breast/ovarian cancer risks. *J Clin Oncol*, 2010: **28**: 240–4.
20. Lin WT, Beattie M, Chen LM, Oktay K, Crawford SL, Gold EB, Cedars M, Rosen M. Comparison of age at natural menopause in BRCA1/2 mutation carriers with a non-clinic-based sample of women in northern California. *Cancer*, 2013; **119**: 1652–9.
21. Lin W, Titus S, Moy F, Ginsburg ES, Oktay K. Ovarian aging in women with BRCA germline mutation. *J Clin Endocrinol Metab*, 2017; **102**: 3839–47.
22. Oktay KH, Bedoschi G, Goldfarb SB, *et al*. Increased chemotherapy-induced ovarian reserve loss in women with germline BRCA mutations due to oocyte deoxyribonucleic acid double strand break repair deficiency. Fertil Steril. 2020; **113**(6): 1251–60.e1. doi:10.1016/j.fertnstert.2020.01.033
23. Ethics Committee of American Society for Reproductive Medicine. Fertility preservation and reproduction in patients facing gonadotoxic therapies: a committee opinion. *Fertil Steril*, 2013; **100**: 1224–31.
24. Oktay K, Harvey BE, Partridge AH, *et al*. Fertility Preservation in Patients with Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol*, 2018; **36**:1994–2001.
25. Lee S, Ozkavukcu S, Heytens E, Moy F, Oktay K. Value of early referral to fertility preservation in young women with breast cancer. *J Clin Oncol*, 2010; **28**: 4683–6.
26. Quinn GP, Vadaparampil ST, Lee JH, *et al*. Physician referral for fertility preservation in oncology patients: a national study of practice behaviors. *J Clin Oncol*, 2009; **27**: 5952–7.
27. van der Schouw YT, van der Graaf Y, Steyerberg EW, Eijkemans JC, Banga JD. Age at menopause as a risk factor for cardiovascular mortality. *Lancet*, 1996; **347**: 714–18.
28. Check JH, Katsoff B, Wilson C, Choe JK, Brasile D. Pregnancy outcome following fresh vs. frozen embryo transfer into gestational carriers using a simplified slow freeze protocol. *Clin Exp Obstet Gynecol*, 2012; **39**: 23–4.
29. Oktay K, Turan V, Bedoschi G, Pacheco FS, Moy F. Fertility Preservation Success Subsequent to Concurrent Aromatase Inhibitor Treatment and Ovarian Stimulation in Women with Breast Cancer. *J Clin Oncol*, 2015; **33**: 2424–9.
30. Coates A, Kung A, Mounts E, *et al*. Optimal euploid embryo transfer strategy, fresh versus frozen, after preimplantation genetic screening with next generation sequencing: a randomized controlled trial. *Fertil Steril*, 2017; **107**: 723–30.e3.

31. Druckenmiller S, Goldman KN, Labella PA, Fino ME, Bazzocchi A, Noyes N. Successful oocyte cryopreservation in reproductive-aged cancer survivors. *Obstet Gynecol*, 2016; **127**: 474–80.
32. Cil AP, Bang H, Oktay K. Age-specific probability of live birth with oocyte cryopreservation: an individual patient data meta-analysis. *Fertil Steril*, 2013; **100**: 492–9.e3
33. Lee JA, Barritt J, Moschini RM, Slifkin RE, Copperman AB. Optimizing human oocyte cryopreservation for fertility preservation patients: should we mature then freeze or freeze then mature? *Fertil Steril*, 2013; **99**: 1356–62.
34. Chian RC, Huang JY, Gilbert L, *et al*. Obstetric outcomes following vitrification of *in vitro* and *in vivo* matured oocytes. *Fertil Steril*, 2009; **91**: 2391–8.
35. Oktay K, Buyuk E, Davis O, Yermakova I, Veeck L, Rosenwaks Z. Fertility preservation in breast cancer patients: IVF and embryo cryopreservation after ovarian stimulation with tamoxifen. *Hum Reprod*, 2003; **18**: 90–5.
36. Oktay K, Buyuk E, Libertella N, Akar M, Rosenwaks Z. Fertility preservation in breast cancer patients: a prospective controlled comparison of ovarian stimulation with tamoxifen and letrozole for embryo cryopreservation. *J Clin Oncol*, 2005; **23**: 4347–53.
37. Oktay K, Hourvitz A, Sahin G, Oktem O, Safro B, Cil A, Bang H. Letrozole reduces estrogen and gonadotropin exposure in women with breast cancer undergoing ovarian stimulation before chemotherapy. *J Clin Endocrinol Metab*, 2006; **91**: 3885–90.
38. Azim AA, Costantini-Ferrando M, Oktay K. Safety of fertility preservation by ovarian stimulation with letrozole and gonadotropins in patients with breast cancer: a prospective controlled study. *J Clin Oncol*, 2008; **26**: 2630–5.
39. Turan V, Bedoschi G, Moy F, Oktay K. Safety and feasibility of performing two consecutive ovarian stimulation cycles with the use of letrozole-gonadotropin protocol for fertility preservation in breast cancer patients. *Fertil Steril*, 2013; **100**: 1681–5.e1.
40. Cold S, Düring M, Ewertz M, Knoop A, Møller S. Does timing of adjuvant chemotherapy influence the prognosis after early breast cancer? Results of the Danish Breast Cancer Cooperative Group (DBCG). *Br J Cancer*, 2005; **93**: 627–32.
41. Lohrisch C, Paltiel C, Gelmon K, Speers C, Taylor S, Barnett J, Olivotto IA. Impact on survival of time from definitive surgery to initiation of adjuvant chemotherapy for early-stage breast cancer. *J Clin Oncol*, 2006; **24**: 4888–94.
42. Oktay K, Demirtas E, SonWY, Lostritto K, Chian RC, Tan SL. *In vitro* maturation of germinal vesicle oocytes recovered after premature luteinizing hormone surge: description of a novel approach to fertility preservation. *Fertil Steril*, 2008; **89**: 228.e19–22.
43. Sonmezer M, Turkcuoglu I, Coskun U, Oktay K. Random-start controlled ovarian hyperstimulation for emergency fertility preservation in letrozole cycles. *Fertil Steril*, 2011; **95**: 2125.e9–11.
44. Cakmak H, Katz A, Cedars MI, Rosen MP. Effective method for emergency fertility preservation: random-start controlled ovarian stimulation. *Fertil Steril*, 2013; **100**: 1673–80.
45. Oktay K, Buyuk E. Ovarian transplantation in humans: indications, techniques and the risk of reseeding cancer. *Eur J Obstet Gynecol Reprod Biol*, 2004; **113**: S45–7.
46. Taylan E, Oktay K. Autologous transplantation of human ovarian tissue. In: Leung P, Adashi E. *The Ovary*, 3rd edn. San Diego, CA: Elsevier, 2018: Chapter 30.
47. Taylan E, Oktay K. Application of decellularized tissue scaffolds in ovarian tissue transplantation. *Methods Mol Biol*, 2018; **1577**: 177–81.
48. Oktay K, Bedoschi G, Pacheco F, Turan V, Emirdar V. First pregnancies, live birth, and *in vitro* fertilization outcomes after transplantation of frozen-banked ovarian tissue with a human extracellular matrix scaffold using robot-assisted minimally invasive surgery. *Am J Obstet Gynecol*, 2016; **214**: 94.e1–9.
49. Oktay K, Taylan E, Sugishita Y, Goldberg GM. Robot-assisted Laparoscopic Transplantation of Frozen-thawed Ovarian Tissue. *J Minim Invasive Gynecol*, 2017; **24**: 897–8.
50. Taylan E, Oktay KH. Robotics in reproduction, fertility preservation, and ovarian transplantation. *Robotic Surg Res Rev*, 2017; **4**: 19–24.
51. Oktay K and Karlikaya G. Ovarian function after transplantation of frozen, banked autologous ovarian tissue. *N Engl J Med*, 2000; **342**: 1919.
52. Oktay K, Buyuk E, Veeck L, Zaninovic N, Xu K, Takeuchi T, Opsahl M, Rosenwaks Z. Embryo development after heterotopic transplantation of cryopreserved ovarian tissue. *Lancet*, 2004; **363**: 837–40.
53. Oktay K, Economos K, Kan M, Rucinski J, Veeck L, Rosenwaks Z. Endocrine function and oocyte retrieval after autologous transplantation of ovarian cortical strips to the forearm. *JAMA*, 2001; **286**: 1490–3.
54. Pacheco F, Oktay K. Current success and efficiency of autologous ovarian transplantation: a meta-analysis. *Reprod Sci*, 2017; **24**: 1111–20.
55. Practice_Committee_of_the_American_Society_for_Reproductive_Medicine. Fertility preservation in patients undergoing gonadotoxic therapy or gonadectomy: a committee opinion. Fertility and sterility 2019;112:1022-33.
56. Sánchez-Serrano M, Novella-Maestre E, Roselló-Sastre E, Camarasa N, Teruel J, Pellicer A. Malignant cells are not found in ovarian cortex from breast cancer patients undergoing ovarian cortex cryopreservation. *Hum Reprod*, 2009; **24**: 2238–43.
57. Azem F, Hasson J, Ben-Yosef D, *et al*. Histologic evaluation of fresh human ovarian tissue before cryopreservation. *Int J Gynecol Pathol*, 2010; **29**: 19–23.
58. Greve T, Clasen-Linde E, Andersen MT, *et al*. Cryopreserved ovarian cortex from patients with leukemia in complete remission contains no apparent viable malignant cells. *Blood*, 2012; **120**: 4311–16.
59. Lambertini M, Boni L, Michelotti A, *et al*.; GIM Study Group. Ovarian suppression with triptorelin during adjuvant breast cancer chemotherapy and long-term ovarian function, pregnancies, and disease-free survival: a randomized clinical trial. *JAMA*, 2015; **314**: 2632–40.
60. Leonard RCF, Adamson DJA, Bertelli G, *et al*.; Anglo Celtic Collaborative Oncology Group and National Cancer Research Institute Trialists. GnRH agonist for protection against ovarian toxicity during chemotherapy for early breast cancer: the Anglo Celtic Group OPTION trial. *Ann Oncol*, 2017; **28**: 1811–16.
61. Oktay K, Taylan E, Rodriguez-Wallberg KA, Bedoschi G, Turan V, Pacheco F. Goserelin does not preserve ovarian function against chemotherapy-induced damage. *Ann Oncol*, 2018; **29**: 512–13.
62. Oktay K, Turan V. Failure of ovarian suppression with gonadotropin releasing hormone analogs to preserve fertility: an assessment based on the quality of evidence. *JAMA Oncol*, 2016; **2**: 74–5.

63. Demeestere I, Brice P, Peccatori FA, *et al*. No evidence for the benefit of gonadotropin-releasing hormone agonist in preserving ovarian function and fertility in lymphoma survivors treated with chemotherapy: final long-term report of a prospective randomized trial. *J Clin Oncol*, 2016; **34**: 2568–74.
64. Elgindy E, Sibai H, Abdelghani A, Mostafa M. Protecting ovaries during chemotherapy through gonad suppression: a systematic review and meta-analysis. *Obstet Gynecol*, 2015; **126**: 187–95.
65. Oktay K, Bedoschi G. Appraising the biological evidence for and against the utility of GnRHa for preservation of fertility in patients with cancer. *J Clin Oncol*, 2016; **34**: 2563–5.
66. Horicks F, Van Den Steen G, Gervy C, Clarke HJ, Demeestere I. Both *in vivo* FSH depletion and follicular exposure to Gonadotrophin-releasing hormone analogs *in vitro* are not effective to prevent follicular depletion during chemotherapy in mice. *Mol Hum Reprod*, 2018; **24**: 221–32.
67. Taylan E, Oktay KH. Current state and controversies in fertility preservation in women with breast cancer. *World J Clin Oncol*, 2017; **8**: 241–8.
68. Sonmezer M, Oktay K. Orthotopic and heterotopic ovarian tissue transplantation. *Best Pract Res Clin Obstet Gynaecol*, 2010; **24**: 113–26.
69. Sugishita Y, Taylan E, Kawahara T, Suzuki N, Oktay K. Comparison of open and closed devices in human ovarian tissue vitrification. *Fertil Steril*, 2017; **108**: e172–3.
70. Kawahara T, Sugishita Y, Taylan E, Suzuki N, Moy F, Oktay K. Vitrification versus slow freezing of human ovarian tissue: a comparison of follicle survival and DNA damage. *Fertil Steril*, 2017; **108**: e56–7.

8.7

Hormonal Contraception

8.7.1 Hormonal Contraception

Jennifer Chin and Bliss Kaneshiro

Introduction

The first widely used hormonal contraceptive, the birth control pill, became available in the 1960s. Hormonal contraception revolutionized the way society views reproduction, sexuality, and the link between the two. For the first time in history, women with access to the birth control pill could determine when and if they had children. This ability to separate sex from childbearing led to the sexual revolution and brought reproductive rights into focus. The ability to plan pregnancy is now regarded as one of the most importance public health successes of the 20th century. Family planning services more than triple the savings of public funding spent on care for pregnant mothers and newborns [1].

Hormonal contraception has two main categories: combined hormonal contraceptive methods which contain both oestrogen and progestin, and progestin-only methods. With both categories, progestin creates the main contraceptive effect, while the oestrogen mainly helps with bleeding regulation through stabilization of the endometrial lining [2]. Combined hormonal contraceptives are available as an oral pill, an injectable, a vaginal ring, and a transdermal patch. Progestin-only methods are available as an oral pill, an injectable, a subdermal implant, and an intrauterine device (IUD).

All methods have a low risk profile, especially when compared to the health risks of pregnancy. The World Health Organization (WHO) developed criteria that clinicians can use to assess the appropriateness of different methods for patients with specific medical conditions. The WHO Medical Eligibility Criteria assigns contraceptive methods a score ranging from one to four depending on a patient's medical condition: 1 means no restrictions on use of the method; 2 means advantages generally outweigh the risks; 3 means the risks generally outweigh the benefits; and 4 means the risk is unacceptable and another method should be used. This system is based on an extensive review of the medical literature and, when evidence is lacking, expert opinion [3]. Throughout this chapter, efficacy will be explained as typical-use failure rate. Researchers have derived typical-use failure rates from data exploring how most women take the prescribed method and have calculated the risk of pregnancy based upon those patterns. Providers should use typical-use rates, rather than perfect-use rates, which assumes a woman takes a contraceptive method with perfect compliance, when counselling patients regarding efficacy.

Many methods have non-contraceptive benefits. These include the treatment of endometriosis, adenomyosis, dysmenorrhoea, abnormal uterine bleeding, and acne. Specific methods also can reduce the risk of certain cancers. These benefits will be described in detail throughout the chapter.

Progestin-Only Methods

Progestin-only methods do not contain oestrogen and thus do not incur the thromboembolic risks associated with oestrogen. These methods have few absolute contraindications and are generally safe for most women including those with comorbidities. Additionally, both the subdermal implant and the IUD are among the most effective types of hormonal contraception. Because little needs to be done beyond insertion for a woman to be compliant with an IUD or implant, the typical-use failure rate of 0.1% approaches the perfect-use failure rate [4].

Intrauterine Devices

IUDs are now the most widely used type of long acting reversible contraception worldwide [4]. Current IUDs are safe and have few contraindications as shown in **Box 8.7.1.1**. Though we focus our discussion on levonorgestrel IUDs, copper IUDs are also highly effective at pregnancy prevention and are an ideal option for women seeking emergency contraception and those unable to take exogenous oestrogen or progestin.

Levonorgestrel IUDs vary in the amount of levonorgestrel released daily and in size of the device. All have a T-frame though some of the newer levonorgestrel IUDs boast flexibility in the size of the arms of the IUD. Levonorgestrel IUDs work by thickening the cervical mucus making it impenetrable to sperm [5]. Approximately 50–75% of women continue to ovulate [6]. The primary mechanism of action of the levonorgestrel IUD does not involve interrupting implantation after fertilization has occurred [7].

The progestational effect results in less frequent and lighter menses. After the first three to six months of use, most women experience light periods or amenorrhea. Levonorgestrel IUDs are 99.9% effective at pregnancy prevention and have a 1-year continuation rate of 81% [8]. Return to fertility after removal is immediate [9]. The 52-mg levonorgestrel IUD is currently marketed for use up to 5 years, though recent studies suggest efficacy up to 7 years [10]. The 52-mg levonorgestrel IUD initially releases 19.5 mg of levonorgestrel which declines to approximately 10 mcg/day after 5 years [11, 12]. The 13.5-mg levonorgestrel IUD is smaller than the 52-mg levonorgestrel IUD (28 mm × 30 mm vs. 32 mm × 32 mm) and provides contraception for up to 3 years. It initially releases 14 mcg of levonorgestrel per day which declines to 5 mcg/day at 3 years [11]. The 19.5-mg levonorgestrel IUD is effective for up to five years and measures 28 mm × 30 mm. It initially releases 17.5 mcg/day which declines to 7.4 mcg/day at 3 years [11]. The devices with a lower levonorgestrel content do not result in amenorrhea as commonly as the 52-mg levonorgestrel IUD.

The IUDs have an excellent safety profile. At the time of insertion, the risk of perforation is less than 1% [13]. It is a common misconception that IUDs cause pelvic inflammatory disease (PID) and ectopic pregnancy. Patients are not at increased risk of PID outside of the first 20 days of insertion, and even within this time frame, the risk is less than 1% [14, 15]. Antibiotics are not indicated at the time of IUD insertion [15]. Screening for sexually transmitted infections (STI) prior to IUD placement has not been shown to be of benefit [16]. Even if a woman ultimately screens positive for chlamydia or gonorrhoea, the risk of upper genital infection with IUD insertion is low with absolute risk of 0–5% for those with an STI and 0–2% for those without [17]. In symptomatic women, testing and treatment should occur prior to IUD insertion.

If a woman is diagnosed with PID with an IUD in situ, patients may opt to start treatment with the IUD in place and proceed with removal only if they do not clinically improve after 48 to 72 hours [16]. In terms of ectopic pregnancy, IUDs decrease the risk of both intrauterine and ectopic pregnancy; however, if a pregnancy does occur with the IUD in place, it is more likely to be an ectopic pregnancy.

The postpartum period is an excellent time to address long-acting reversible contraceptives (LARC) methods. IUDs can be inserted anytime postpartum; however, certain times incur more risks of expulsion than others. Providers should discuss and offer to interested patients immediate post-placental placement of IUDs during antenatal visits [18]. Although the expulsion rate is higher with immediate post-placental placement compared to placement at six weeks postpartum (10–27% vs. 3%), studies have shown that the cost:benefit ratio favours post-placental placement, especially in instances where postpartum follow-up may be difficult. Continuation rates with post-placental placement are as high as 85% after 12 months [19]. Two separate studies found an increased risk of perforation with post-placental IUD placement in breastfeeding women and in women who delivered at or less than 36 weeks of gestation [3].

Historically, clinicians preferred to insert the IUD at the start of menses so they could be assured the patient was not pregnant. Insertion was thought to be easier at the time of menses, though no studies have demonstrated that this is the case. More recent studies show a higher rate of expulsion with insertion at the time of menses [20]. Current recommendations state IUDs can be inserted as long as a clinician is reasonably certain a woman is not pregnant [21].

After IUD insertion, women do not need a routine follow-up visit; however, they should be advised to return at any time to discuss side effects or different methods if they are dissatisfied. At routine gynaecologic visits, providers can check IUD strings and women can palpate their own IUD strings periodically if they desire. Failure to visualize or palpate strings could be a sign of expulsion or perforation, though most of the time, the IUD will be in the uterus with the strings pulled into the cervix. Pelvic ultrasound can determine intrauterine IUD placement when strings cannot be identified. If the IUD is not visualized on ultrasound, and a patient does not recall expulsion, an abdominal X-ray can be used to determine if a perforation has occurred.

Satisfaction with the IUD requires proper counselling about the bleeding profile after insertion. Irregular spotting and bleeding are common in the first 3 to 6 months after insertion. Approximately 40% of women become amenorrhoeic after placement of the 52-mg levonorgestrel IUD [22, 23] and providers should counsel women that amenorrhea while using a hormonal IUD is common and not harmful.

Box 8.7.1.1 WHO medical eligibility criteria contraindications to levonorgestrel IUD insertion (category 4 or risk is unacceptable)

- Pregnancy
- Puerperal sepsis
- Immediate post-septic abortion
- Gestational trophoblastic disease (persistently elevated beta hCG levels or malignant disease)
- Cervical cancer (do not place IUD but may continue use if already in)
- Endometrial cancer (do not place IUD but may continue use if already in)
- Current breast cancer
- Uterine fibroids with distortion of the uterine cavity
- Distorted uterine cavity that is incompatible with IUD insertion
- Current infection (once treated, no longer a contraindication)
- Pelvic tuberculosis

Data from World Health Organization. Medical Eligibility Criteria for Contraceptive Use. Fifth Edition, 2015. Available at http://www.who.int/reproductivehealth/publications/family_planning/MEC-5/en/ Retrieved 15 May 2018.

The IUD offers many non-contraceptive benefits in addition to being a highly effective contraceptive. Because most women will experience lighter bleeding and some will experience amenorrhea, it is an effective treatment for dysmenorrhoea due to endometriosis and heavy bleeding due to adenomyosis [24–27]. Due to its progestational effect on the endometrium, it has been described as a treatment for endometrial hyperplasia and for endometrial protection in the setting of tamoxifen use or postmenopausal oestrogen therapy [28–30].

Implants

The first widely utilized contraceptive implant was a levonorgestrel implant comprised of six separate silicone capsules, each containing 36 mg of levonorgestrel. It was effective for up to seven years in normal weight women. Removal could be difficult because of the number of capsules and it was phased out of production by 2008 and replaced by a two-rod levonorgestrel implant that is used in many parts of the world today. The two rods are 43 mm long, contain a total of 75 mg of levonorgestrel, and provide contraception for up to five years.

A single 4 cm rod containing 68 mg of etonogestrel, the active metabolite of desogestrel, is also available. The etonogestrel implant is marketed for use up to 3 years; however, more recent studies have shown efficacy for up to 5 years [31]. The contraceptive implant does not increase the risk of osteopenia, osteoporosis, or fracture [32, 33].

Contraceptive implants are highly effective with a 1-year typical-use failure rate of 0.09% for the etonogestrel implant and 0.1% for the levonorgestrel implant [34]. After 2 years, 82% of women will continue to use this method [35]. The implant effectively prevents ovulation, though it also thickens cervical mucus and thins the endometrium [4]. Return to fertility with discontinuation is rapid with 90% of women ovulating within 3 weeks after removal [34].

Providers should counsel patients that a common side effect is irregular, unpredictable bleeding, which, unlike the IUD, generally does not improve over time. Though bleeding is rarely heavy, it is unpredictable which can lead to dissatisfaction [36].

Older versions of this implant could not be visualized on X-ray though the most current version of the etonogestrel implant is impregnated with barium sulphate. Providers can now use computed tomography, ultrasound, X-ray, or magnetic resonance imaging to visualize non-palpable implants. When the implant is not easily palpated, ultrasound guidance can be used to assist with removal. Consult with a surgeon familiar with the anatomy of the arm is also helpful.

Injectable Contraception

Injectable medroxyprogesterone (DMPA) was first used in 1967 and consists of a 150 mg intramuscular dose of medroxyprogesterone that is administered every 12 weeks. A 104 mg subcutaneous dose is also available. DMPA provides contraception by inhibiting ovulation, while also thickening the cervical mucus and thinning the endometrium. After 1 year of use, 50% of women will develop amenorrhea and 80% develop amenorrhea within 5 years [35]. Though many women experience unscheduled bleeding with early injections, the frequency and duration of unscheduled bleeding decreases with continued use. Supplemental oestrogen, cyclooxygenase-2 inhibitors, or mifepristone have been used to treat unscheduled bleeding with DMPA though expectant management is most commonly used [37, 38]. DMPA has a 10-year continuation rate of 23% and typical-use failure rate of 6% in the first year [35]. Compared to other methods, return to fertility with DMPA is not immediate. The average time to the resumption of normal cycles is 6 to 8 months for patients. For the subcutaneous formulation, the return to fertility is a median of 7 months with 97% of patients returning to baseline fertility after 1 year [39].

Clinicians can start patients on DMPA at any time they are reasonably certain a patient is not pregnant. However, if a patient has a negative urine pregnancy test on the day of the visit but does not meet the criteria to ensure she is not pregnant, clinicians can administer DMPA with counselling regarding the possibility of a very early pregnancy and a delay in the diagnosis of pregnancy. No studies demonstrate teratogenicity of DMPA [40]. The only absolute contraindications to DMPA are known current pregnancy and breast cancer.

Concerns about temporary decreases in bone mineral density should not limit the use or duration of use of DMPA. Inhibition of gonadotropins causes a reversible decline in bone mineral density by 3–6% after 2 years of use. However, studies have shown that this loss does not continue after two years of use and is similar to the loss experienced during breastfeeding or the postpartum period. A year after discontinuation of DMPA, previous bone mineral densities are the same as non-users [41]. Providers should carefully weigh these potential risks against the negative consequences of an unintended pregnancy and should not limit use or duration of use [42]. Patients already prone to low bone density, including heavy smokers, patients with anorexia, amenorrhoeic athletes, or chronic steroid users, should use DMPA with caution. Providers should counsel all patients using DMPA on appropriate calcium intake, regular exercise, and smoking cessation. Routine monitoring of bone mineral density is not required.

Providers and patients also worry about the weight gain associated with DMPA use. Studies have had conflicting results as to whether DMPA truly causes weight gain. One randomized trial did not note significant weight gain with DMPA use [43]. Observational studies have reported overweight and obese adolescents gain more weight when using this method than when using no contraceptives or oral contraceptive pills [44–47].

DMPA has many unique non-contraceptive benefits. Unlike combined hormonal contraceptives, DMPA is unaffected by concomitant antiseizure medications and raises the seizure threshold [48]. In patients with sickle cell anaemia, DMPA stabilizes the red cell membrane and significantly decreases menstrual blood loss [49]. Due to the high rate of amenorrhea, DMPA is also effective for achieving amenorrhea and in treating conditions such as heavy menses, dysmenorrhoea, endometriosis, and ovarian cyst formation [50, 51].

Progestin-Only Pill

For patients who desire a birth control pill but are unable to take oestrogen, progestin-only pills, consisting of 0.35 mg of norethindrone

taken daily, are a good option. Typical-use failure rates are 9%, which is higher than other progestin-only methods [9]. The main mechanism of action is thickening of cervical mucus, so sperm do not pass into the uterus. Progestin-only pills also decrease tubal motility and thin the endometrial lining. Thickening of the cervical mucus happens approximately two to four hours after the pill is taken and lasts for about 22 hours. Patients need to take the pill at the same time every day. Taking a pill more than 3 hours late results in decreased efficacy and women should use a backup method for 48 hours [35]. Return to fertility is rapid due to the short-term effects of the progestin-only pill. Sixty per cent (60%) of women continue to ovulate while taking the progestin-only pill and 40–50% retain normal menstrual cycles. A new pill containing 4 mg of drospirenone works by inhibiting ovulation.

Absolute contraindications include breast cancer and known pregnancy though inadvertent administration during pregnancy does not result in teratogenicity. The progestin-only pill has diminished efficacy in those taking rifampin, phenytoin, carbamazepine, barbiturates, primidone, topiramate, oxcarbazepine, and St. John's Wort as these medications increase hormone metabolism [35].

Combined Hormonal Contraception

The first oral contraceptive pill was developed in 1957 for menstrual disorders and contained 150 mcg of mestranol and 9.85 mg of norethynodrel. In 1960, this hormonal combination became available to women for contraception. Over time, the oestrogen dose in pills has decreased and new progestin and oestrogen formulations have been developed. Combined hormonal contraception is now available as a pill, patch, ring, or injection.

All combined hormonal contraceptives work primarily by inhibiting ovulation. They also thicken the cervical mucus and thin the endometrium. The typical-use failure rate is approximately 9% [9]. Return to fertility is rapid with most women ovulating one to two weeks after discontinuing combined hormonal contraceptives. Combined hormonal contraceptives can provide multiple non-contraceptive benefits including reduced bleeding, decreased pain from dysmenorrhoea or endometriosis, regulating anovulatory cycles and improvement in acne.

The risks of combined hormonal contraceptives are associated with the thrombotic effects of the oestrogen component of the pill. The first reported thrombotic event associated with oral contraceptive pill use occurred in 1961 in a patient taking a 150 mcg mestranol pill for endometriosis. Newer formulations of oral contraceptive pills today have much lower oestrogen dose and thus the risk for thromboembolic events has greatly decreased. Overall, the risk is 2–3 per 10 000, lower than the risk of thromboembolism during pregnancy of 6 per 10 000. Absolute contraindications are listed in **Box 8.7.1.2**. Relative contraindications include women who smoke and are less than 35 years old, chronic hypertension that is adequately controlled with medication or with blood pressures<160/100, diabetes mellitus without vascular disease, connective tissue disorders, migraines without aura, and gallbladder disease [3].

Oral Contraceptive Pills

Combined hormonal oral contraceptive pills are easy to initiate, are safe and effective, and have many non-contraceptive benefits [52].

Box 8.7.1.2 WHO medical eligibility criteria contraindications to combined hormonal contraception (category 4)

- Less than 6 weeks postpartum in breastfeeding women
- Less than 21 days postpartum in women with risk factors for Venous Thromboembolism
- Smoking and age 35 or older
- Multiple risk factors for cardiovascular disease (e.g. older age, smoking, diabetes, hypertension, known dyslipidaemias)
- Systolic ≥160 or diastolic ≥100 mmHg
- Hypertension with vascular disease
- History of or acute deep venous thrombosis or pulmonary embolism
- Major surgery with prolonged immobilization
- Known thrombogenic mutations
- Current and history of ischaemic heart disease
- Stroke
- Complicated valvular heart disease
- Positive antiphospholipid antibodies
- Migraine with aura
- Migraine without aura, 35 years of age or older (may continue but should not initiate)
- Current breast cancer
- Diabetes with nephropathy, retinopathy, neuropathy, or any vascular disease or diabetes of 20 years' duration or more
- Liver disease including acute viral hepatitis, severe decompensated cirrhosis, hepatocellular adenoma, malignant hepatoma

Data from World Health Organization. Medical Eligibility Criteria for Contraceptive Use. Fifth Edition, 2015. Available at http://www.who.int/reproductivehealth/publications/family_planning/MECguidelinePart-2.pdf. Retrieved 15 May 2018.

Because of these qualities, they are one of the most commonly used contraceptive methods [4]. Traditionally, providers told women to start taking their pills on the next Sunday after their next menstrual period. Quick start, where a woman starts the pill the day she receives it and uses a backup method for the first seven days is an easy way to initiate the oral contraceptive pill. Quick start ensures that patients do not delay initiating a contraceptive method and does not increase the risk of breakthrough bleeding as compared to the other start methods [35]. Although oral contraceptive pills are easy to initiate and prescribe, women must remember to take the pill regularly [53]. The WHO has published guidelines on how to manage missed oral contraceptive pills as shown in **Box 8.7.1.3**.

Commonly prescribed low dose pills contain 20–35 mcg of ethinyl oestradiol and a progestin. Studies have shown similar efficacy, side effects, and safety profiles between these oestrogen doses [35]. Epidemiologic studies have classified generations of oral contraceptive pills based on their amount of oestrogen and the type of progestin used. First generation pills contain 50 mcg or more of ethinyl oestradiol and are not commonly used anymore. Second generation pills contain either levonorgestrel, norgestimate, or other types of norethindrone. Third generation pills contain desogestrel or gestodene. Fourth generation pills contain drospirenone. When the third and fourth generation pills were introduced, there was initially concern for a higher risk of thromboembolic events. However, prospective cohort studies reveal a similar risk for thrombotic events with different pills suggesting the findings in case–control studies, which suggested an elevated risk, were likely due to prescribing biases [54–56].

Most oral contraceptive pills contain 21 days of hormone and seven days of placebo; women have a week of bleeding related to hormonal withdrawal. Newer pills shorten this placebo time to

Box 8.7.1.3 Missed combined oral contraceptive pills

30–35 μg ethinyl oestradiol pills: Missed 1–2 hormonal pills or starts a pack 1–2 days late

20 μg ethinyl oestradiol pills: Missed 1 active pill or starts a pack 1 day late

- Take hormonal pill as soon as possible an continue taking one pill a day (will take 2 pills on the day she realizes she has missed a pill)
- Continue taking one pill a day
- No need for back up contraception

30–35 μg ethinyl oestradiol pills: Missed 3 or more hormonal pills or starts a pack 3 or more days late

20 μg ethinyl oestradiol pills: Missed 2 active pills or starts a pack 2 or more days late

- Take active pill as soon as possible
- Continue taking one pill a day
- Use back up contraception until she has taken hormonal pills for 7 days in a row
- If pills are missed in the 1st week, consider emergency contraception
- If pills are missed in the 3rd week, take active pills in current pack and start the next pill pack instead of taking inactive pills

Data from World Health Organization. Selected Practice Recommendations for Contraceptive Use. Second edition, 2015. Available at http://apps.who.int/iris/bitstream/handle/10665/252267/9789241565400-eng.pdf?sequence=1. Retrieved 15 May 2018.

three to four days resulting in a shorter, lighter period. A shorter placebo week results in lower contraceptive failures due to a lower risk of follicular development and subsequent ovulation [53]. Formulations have varied the amount of hormone throughout the cycle to mimic the fluctuations in hormone levels during a regular menstrual cycle. Both biphasic and triphasic patterns exist, but neither have shown any difference in efficacy or side effect profile.

Having a monthly withdrawal bleed while on hormonal contraceptives does not have a biological benefit. For patients who prefer not to bleed monthly or are taking oral contraceptives for heavy menses, anaemia, dysmenorrhoea, premenstrual syndrome, endometriosis, or menstrual migraines, continuous or extended dosing is a good option [35]. In extended dosing, women take hormonally active pills for several months without taking the placebo pills thereby postponing a hormonal withdrawal bleed. With continuous dosing, women take hormonally active pills indefinitely. With continuous and extended administration, women may experience more breakthrough bleeding and spotting which will improve over time. After discontinuation, women will return to the same pattern of menses they experienced prior to initiating the contraception.

Contraceptive Patch

The contraceptive patch is formulated as a transdermal patch that is worn for one week at a time for three weeks and then removed for a hormone free week to permit a withdrawal bleed. Each 4.5 cm square patch delivers 20 mg of ethinyl oestradiol and 150 mcg of norelgestromin daily. The patch can be applied on the lower abdomen, buttocks, upper outer arm, or upper torso (except for the breasts). The patch provides contraceptive effect for up to nine days though women are instructed to change the patch every seven days. Efficacy is similar to oral contraceptive pills, though studies suggest a higher failure rate in women weighing over 90 kg. Side effects unique to the patch include partial detachment of the patch, complete detachment, skin irritation, and pigment changes under the site of the patch [57]. When compared to oral contraceptive pills, women on the patch are more likely to experience nausea, breast discomfort, and headaches [35]. These side effects are more pronounced when women use the patch in a continuous fashion [58–61].

Contraceptive Ring

Contraceptive rings are placed in the vagina where oestrogen and progestin are absorbed through the vaginal mucosa. Different contraceptive rings provide anywhere from 1 week to 1 year of contraception [4]. The etonogestrel vaginal ring is a flexible ethylene vinyl acetate copolymer that releases 15 mcg of ethinyl oestradiol and 120 mcg of etonogestrel daily. It is 5.4 cm in diameter and 4 mm in thickness. The ring provides adequate contraceptive hormone for up to 5 weeks though women are instructed to leave the ring in place for 3 weeks and then remove it for 1 week to allow for a withdrawal bleed. While it is not recommended that patients remove the ring regularly, they can do so for up to 3 hours with no decrease in efficacy.

In contrast to the pill and the patch, the ring releases relatively low constant levels of hormone and thus results in fewer oestrogen-related side effects and breakthrough bleeding. While the ring may cause more discharge, vaginal flora and cervical cytology are not affected [62]. Women should dispose of their ring in a sealed plastic bag in order to minimize hormone leakage into the environment.

Combined Hormonal Contraceptive Injection

Although not widely utilized outside of Asia and Latin America, a monthly combined hormonal contraceptive injection is available. The injection contains 25 mg of medroxyprogesterone acetate and 5 mg of oestradiol or 150 mg of dihydroxyprogesterone acetophenide and 10 mg of oestradiol enanthate. In comparison to DMPA, patients have a quicker return to fertility and have less irregular bleeding. An automatic device for self-administration is available which may improve compliance by eliminating the need for monthly follow-up [63].

Emergency Contraception

Providers should counsel all patients about emergency contraception, when it is needed, and how it can be obtained. Although not a hormonal contraceptive, the most effective emergency contraceptive is the copper IUD. Trained providers can insert the copper IUD for up to five days after unprotected intercourse and it has the added benefit of providing long-term contraception.

Levonorgestrel emergency contraception consists of one dose of 1.5 mg dose of levonorgestrel which delays or inhibits ovulation [64, 65]. This method can be taken within 120 hours of unprotected intercourse, though efficacy is higher the sooner it is taken. Ulipristal acetate, a selective progestin receptor modulator, has higher efficacy in pregnancy prevention upto 120 hours after unprotected

intercourse. Ulipristal acetate is also more effective in overweight and obese women and is thus the preferred method over levonorgestrel emergency contraception for women with a body mass index (BMI) over 25. Because ulipristal acetate is a progestin receptor modulator, it may reduce the efficacy of other hormonal contraceptives and thus women should use either abstinence or a non-hormonal method of contraception for the first seven days after using ulipristal acetate [35].

Cancer Risk Reduction

One of the unanticipated results of long-term hormonal contraceptive use has been its role in cancer risk reduction. The type of long-term studies that can demonstrate a reduction in cancer have only been done with the combined oral contraceptive pills though similar effects are suspected with the contraceptive patch and the ring.

Combined hormonal contraceptives significantly decrease the risk of ovarian cancer. This is most likely due to a reduction in ovulation which decreases the inflammatory and reparative process that normally occurs during each cycle. DMPA has not been shown to provide the same effect although it also prevents ovulation [66]. Even among women who use combined oral contraceptives for as little as 3 to 6 months, a reduction in ovarian cancer protection was noted up to 20 years after discontinuation [67]. In patients who use oral contraceptive pills for less than 10 years, their risk of ovarian cancer is reduced by 40% and this reduction is doubled if used for more than 10 years [68–71]. This is particularly helpful for women with a strong family history of ovarian cancer, and studies have shown that for these women, the use of oral contraceptive pills can reduce their risk to that of the general population [72].

Because all hormonal contraceptives thin the endometrial lining and prevent cyclical growth and shedding, they all decrease the risk of endometrial cancer [73, 74]. After using combined hormonal contraceptives for four years, the incidence of endometrial cancer is reduced by 56% and further drops by 72% after 12 years [75–77]. For DMPA, the risk is reduced by 80% and extends for 8 years after discontinuing DMPA [78].

Facilitating Contraceptive Use

The goal of contraceptive counselling is to provide women with education about contraceptive options so that they can select a method that best fits their individual needs and desires for childbearing. Though effectiveness is an important aspect of each method, it may not be the most important contraceptive characteristic for a particular woman. Providers should respect a woman's reproductive desires which include the option of not using or discontinuing a contraceptive method.

Providers should ensure that all women have access to safe, appropriate, hormonal contraception if they desire it. Removing barriers to contraception includes using quick start and providing methods, including IUDs and implants, on the same days as contraceptive counselling. Overcoming systemic barriers including increasing health insurance coverage for all women and training an adequate number of clinicians skilled in contraceptive counselling is also important.

The WHO has created guidelines as to what is required of a clinician before initiating different methods of contraception. For the hormonal contraceptives, with the exception of IUDs, a pelvic exam is not necessary [3]. While it is advised that all women receive proper health maintenance and cancer screening, contraceptives should not be withheld or delayed if women do not wish to participate in cancer screening or other testing [79–85].

Providing women with a complete picture of contraceptive benefits, side effects, and risks provides them with realistic expectations and increases satisfaction. In particular, women should understand anticipated changes to the menstrual cycle with hormonal contraceptives. Patient-centred care allows a woman to be fully invested in her contraceptive method.

Summary

Since the advent of the first oral contraceptive pill, science, healthcare professionals, and society have made great strides in expanding the hormonal contraceptives available to women. Women can now choose between many different contraceptive formulations and routes of administration. Hormone dosages have greatly decreased. Side effect profiles have diminished. Women have the ability to choose the contraceptive method that fits their needs. Hormonal contraception has helped women and families safely prevent unwanted pregnancies. An unintentional advantage has been to offer many non-contraceptive benefits.

Hormonal contraception has proven to be one of the greatest inventions of this century. However, this advancement is often limited to women living in areas with more healthcare resources. Women around the world deserve the right to choose if and when they wish to become pregnant. Fertility has always been associated with the female body. All currently available methods of hormonal contraception require that women incur the risks and side effects of birth control. Development of a reversible male contraceptive would expand the options available to couples wishing to plan their families. We need to continue to strive, expand, and develop until every woman and man have the ability to determine their reproductive future.

REFERENCES

1. Frost JJ, Sonfield A, Gold RB, Ahmed F. Estimating the impact of serving new clients by expanding funding for Title X: Guttmacher Institute Occasional Report, 2006.
2. Blumenthal PD, Edelman A. Hormonal contraception. *Obstet Gynecol*, 2008; **112**: 670–84.
3. Medical Eligibility Criteria for Contraceptive Use. Available at: http://www.who.int/reproductive-health/publications/mec/ (accessed 12 April 2020).
4. Speroff L, Darney PD. *A Clinical Guide for Contraception*. Philadelphia, PA: Lippincott Williams & Wilkins, 2005.
5. Critchley HO, Wang H, Jones RL, *et al*. Morphological and functional features of endometrial decidualization following long-term intrauterine levonorgestrel delivery. *Hum Reprod*, 1998; **13**: 1218–24.

6. Barbosa I, Bakos O, Olsson SE, Odlind V, Johansson ED. Ovarian function during use of a levonorgestrel-releasing IUD. *Contraception*, 1990; **42**: 51–66.
7. Videla-Rivero L, Etchepareborda JJ, Kesseru E. Early chorionic activity in women bearing inert IUD, copper IUD and levonorgestrel-releasing IUD. *Contraception*, 1987; **36**: 217–26.
8. Trussell J, Vaughan B. Contraceptive failure, method-related discontinuation and resumption of use: results from the 1995 National Survey of Family Growth. *Fam Plann Perspect*, 1999; **31**: 64–72, 93.
9. Trussell J. Contraceptive failure in the United States. *Contraception*, 2004; **70**: 89–9.
10. Rowe P, Farley T, Peregoudov A, Piaggio G, Boccard S, Landoulsi S, Meirik O. Safety and efficacy in parous women of a 52-mg levonorgestrel-medicated intrauterine device: a 7-year randomized comparative study with the TCu380A. *Contraception*, 2016; **94**: 288.
11. Highlights of Prescribing Information. Mirena.com. Bayer HealthCare Pharmaceuticals Inc, 2000. Available at: https://www.mirena-us.com (accessed : 12 April 2020).
12. Highlights of Prescribing Information. Liletta.com. Allergan USA, Inc, 2017. Available at: https://www.liletta.com (accessed: 12 April 2020).
13. ACOG practice bulletin. Clinical management guidelines for obstetrician-gynecologists. Number 59, January 2005. Intrauterine device. *Obstet Gynecol*, 2005; **105**: 223–32.
14. Farley TM, Rosenberg MJ, Rowe PJ, Chen JH, Meirik O. Intrauterine devices and pelvic inflammatory disease: an international perspective. *Lancet*, 1992; **339**: 785–8.
15. Grimes DA, Schulz KF. Antibiotic prophylaxis for intrauterine contraceptive device insertion. *Cochrane Database Syst Rev*, 2000; **2**: CD001327.
16. Pelvic Inflammatory Disease. Centers for Disease Control and Prevention. US Department of Health and Human Services. 4 June 2015. Available at: https://www.cdc.gov (accessed: 12 April 2020).
17. Mohllajee AP, Curtis KM, Peterson HB. Does insertion and use of an intrauterine device increase the risk of pelvic inflammatory disease among women with sexually transmitted infection? A systematic review. *Contraception*, 2006; **73**: 145–53.
18. ACOG committee opinion. Clinical guidance and publications. Number 670, August 2016. Postpartum long acting reversible contraception. *Obstet Gynecol*, 2016; **128**: e32–7.
19. White MK, Ory HW, Cwiak C Cordes S. Postpartum intrauterine device placement: a patient-friendly option. *Contraception and Reproductive Medicine*, 2018; **3**: 3.
20. Rooks JB, Rochat RW. Intrauterine device termination rates and the menstrual cycle day of insertion. *Obstet Gynecol*, 1980; **55**: 220–4.
21. US Selected Practice Recommendations for Contraceptive Use. Centers for Disease Control and Prevention. US Department of Health and Human Services. 27 September 2017. Available at: https://www.cdc.gov (accessed: 12 April 2020.
22. Hidalgo M, Bahamoxndes L, Perrotti M, Diaz J, Dantas-Monteiro C, Petta C. Bleeding patterns and clinical performance of the levonorgestrel-releasing intrauterine system (Mirena) up to two years. *Contraception*, 2002; **65**: 129–32.
23. Backman T, Huhtala S, Blom T, Luoto R, Rauramo I, Koskenvuo M. Length of use and symptoms associated with premature removal of the levonorgestrel intrauterine system: a nation-wide study of 17,360 users. *BJOG*, 2000; **107**: 335–9.
24. Lockhat FB, Emembolu JO, Konje JC. The evaluation of the effectiveness of an intrauterine-administered progestogen (levonorgestrel) in the symptomatic treatment of endometriosis and in the staging of the disease. *Hum Reprod*, 2004; **19**: 179–84.
25. Bragheto AM, Caserta N, Bahamondes L, Petta CA. Effectiveness of the levonorgestrel-releasing intrauterine system in the treatment of adenomyosis diagnosed and monitored by magnetic resonance imaging. Contraception, 2007; **76**: 195–9.
26. Fedele L, Bianchi S, Raffaelli R, Portuese A, and Dorta M. Treatment of adenomyosis-associated menorrhagia with a levonorgestrel-releasing intrauterine device. *Fertil Steril*, 1997; **68**: 426–9.
27. Nilsson CG. Comparative quantitation of menstrual blood loss with a d-norgestrel-releasing iud and a Nova-T-copper device. *Contraception*, 1977; **15**: 379–87.
28. Orbo A, Arnes M, Hancke C, Vereide AB, Pettersen I, Larsen K. Treatment results of endometrial hyperplasia after prospective D-score classification: a follow-up study comparing effect of LNG-IUD and oral progestins versus observation only. *Gynecol Oncol*, 2008; **111**: 68–73.
29. Gardner FJ, Konje JC, Abrams KR, *et al*. Endometrial protection from tamoxifen-stimulated changes by a levonorgestrel-releasing intrauterine system: a randomised controlled trial. *Lancet*, 2000; **356**: 1711–17.
30. Suhonen S, Holmstrom T, Lahteenmaki P. Three-year follow-up of the use of a levonorgestrel-releasing intrauterine system in hormone replacement therapy. *Acta Obstet Gynecol Scand*, 1997; **76**: 145–50.
31. McNicholas C, Maddipati R, Zhao Q, Swor E, Peipert JF. Use of the Etonogestrel Implant and Levonorgestrel Intrauterine Device Beyond the U.S. Food and Drug Administration–Approved Duration. *Obstet Gynecol*, 2015; **125**: 599–604.
32. Bahamondes L, Monteiro-Dantas C, Espejo-Arce X, Dos Santos Fernandes AM, Lui-Filho JF, Perrotti M, Petta CA. A prospective study of the forearm bone density of users of etonorgestrel- and levonorgestrel-releasing contraceptive implants. *Hum Reprod*, 2006; **21**: 466–70.
33. Beerthuizen R, van Beek A, Massai R, Makarainen L, Hout J, Bennink HC. Bone mineral density during long-term use of the progestagen contraceptive implant Implanon compared to a non-hormonal method of contraception. *Hum Reprod*, 2000; **15**: 118–22.
34. Funk S, Miller M, Mishell DR, Archer D, Poindexter A, Schmidt J, Zampaglione E. Safety and efficacy of Implanon, a single-rod implantable contraceptive containing etonogestrel. *Contraception*, 2005; **71**: 319–26.
35. Zieman M, Hatcher RA, Allen AZ, Lathrop E, Haddad L. *Managing Contraception 2016.* Tiger, GA: Bridging the Gap Foundation, 2016.
36. Affandi B. An integrated analysis of vaginal bleeding patterns in clinical trials of Implanon. *Contraception*, 1998; **58**(6 Suppl): 99 S–107 S.
37. Nathirojanakun P, Taneepanichskul S, Sappakitkumjorn N. Efficacy of a selective COX-2 inhibitor for controlling irregular uterine bleeding in DMPA users. *Contraception*, 2006; **73**: 584–7.
38. Tantiwattanakul P, Taneepanichskul S. Effect of mefenamic acid on controlling irregular uterine bleeding in DMPA users. *Contraception*, 2004; **70**: 277–9.
39. Jain J, Dutton C, Nicosia A, Wajszczuk C, Bode FR, Mishell DR. Jr, Pharmacokinetics, ovulation suppression and return to ovulation following a lower dose subcutaneous formulation of DMPA-Provera. *Contraception*, 2004; **70**: 11–18.

40. Prahalada S, Carroad E, Hendrickx AG. Embryotoxicity and maternal serum concentrations of medroxyprogesterone acetate (MPA) in baboons (Papio cynocephalus). *Contraception*, 1985; **32**: 497–515.
41. Scholes D, LaCroix AZ, Ichikawa LE, Barlow WE, Ott SM. Injectable hormone contraception and bone density: results from a prospective study. *Epidemiology*, 2002; **13**: 581–7.
42. ACOG committee opinion. Clinical guidance and publications. Number 602, June 2014. Depot medroxyprogesterone and acetate and bone effects. *Obstet Gynecol*, 2014; **123**: 1398–402.
43. Pelkman CL, Chow M, Heinbach RA, Rolls BJ. Short-term effects of a progestational contraceptive drug on food intake, resting energy expenditure, and body weight in young women. *Am J Clin Nutr*, 2001; **73**: 19–26.
44. Bonny AE, Ziegler J, Harvey R, Debanne SM, Secic M, Cromer BA. Weight gain in obese and nonobese adolescent girls initiating depot medroxyprogesterone, oral contraceptive pills, or no hormonal contraceptive method. *Arch Pediatr Adolesc Med*, 2006; **160**: 40–5.
45. Mangan SA, Larsen PG, Hudson S. Overweight teens at increased risk for weight gain while using depot medroxyprogesterone acetate. *J Pediatr Adolesc Gynecol*, 2002; **15**: 79–82.
46. Bonny AE, Britto MT, Huang B, Succop P, Slap GB. Weight gain, adiposity, and eating behaviors among adolescent females on depot medroxyprogesterone acetate (DMPA). *J Pediatr Adolesc Gynecol*, 2004; **17**: 109–15.
47. Clark MK, Dillon JS, Sowers M, Nichols S. Weight, fat mass, and central distribution of fat increase when women use depot-medroxyprogesterone acetate for contraception. *Int J Obes (Lond)*, 2005; **29**: 1252–8.
48. Mattson RH, Cramer JA, Caldwell BV, Siconolfi BC. Treatment of seizures with medroxyprogesterone acetate: preliminary report. *Neurology*, 1984; **34**: 1255–8.
49. Legardy JK, Curtis KM. Progestogen-only contraceptive use among women with sickle cell anemia: a systematic review. *Contraception*, 2006; **73**: 195–204.
50. Cullins VE. Noncontraceptive benefits and therapeutic uses of DMPAt medroxyprogesterone acetate. *J Reprod Med*, 1996; **41**(5 Suppl): 428–33.
51. Lackie E, Fairchild A. The birth control pill, thromboembolic disease, science and the media: a historical review of the relationship. *Contraception*, 2016; **94**: 295–302.
52. Burkman RT, Collins JA, Shulman LP, Williams JK. Current perspectives on oral contraceptive use. *Am J Obstet Gynecol*, 2001; **185**(2 Suppl): S4–12.
53. Chabbert-Buffet N, Jamin C, Lete I, *et al*. Missed pills: frequency, reasons, consequences and solutions. *Eur J Contracept Reprod Health Care*, 2017; **22**: 165–9.
54. Jick SS, Hernandez RK. Risk of non-fatal venous thromboembolism in women using oral contraceptives containing drospirenone compared with women using oral contraceptives containing levonorgestrel: case-control study using United States claims data. *BMJ*, 2011; **342**: d2151.
55. van Hylckama Vlieg A, Helmerhorst FM, Vandenbroucke JP, Doggen CJ, Rosendaal FR. The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. BMJ, 2009; **339**: b2921.
56. Rosenberg M, Waugh MS. Causes and consequences of oral contraceptive noncompliance. *Am J Obstet Gynecol*, 1999; **180**(2 Pt 2): 276–9.
57. Abrams LS, Skee D, Natarajan J, Wong FA. Pharmacokinetic overview of Ortho Evra/Evra. *Fertil Steril*, 2002; **77**(2 Suppl 2): S3–12.
58. Stewart FH, Kaunitz AM, Laguardia KD, Karvois DL, Fisher AC, Friedman AJ. Extended use of transdermal norelgestromin/ethinyl estradiol: a randomized trial. *Obstet Gynecol*, 2005; **105**: 1389–96.
59. Jick SS, Kaye JA, Russmann S, Jick H. Risk of nonfatal venous thromboembolism in women using a contraceptive transdermal patch and oral contraceptives containing norgestimate and 35 microg of ethinyl estradiol. *Contraception*, 2006; **73**: 223–8.
60. Jick S, Kaye JA, Li L, Jick H. Further results on the risk of nonfatal venous thromboembolism in users of the contraceptive transdermal patch compared to users of oral contraceptives containing norgestimate and 35 microg of ethinyl estradiol. *Contraception*, 2007; **76**: 4–7.
61. Cole JA, Norman H, Doherty M, Walker AM. Venous thromboembolism, myocardial infarction, and stroke among transdermal contraceptive system users. *Obstet Gynecol*, 2007; **109**(2 Pt 1): 339–46.
62. Roumen FJ, Boon ME, van Velzen D, Dieben TO, Coelingh Bennink HJ. The cervico-vaginal epithelium during 20 cycles' use of a combined contraceptive vaginal ring. *Hum Reprod*, 1996; **11**: 2443–8.
63. Bahamondes L, Marchi NM, de Lourdes Cristofoletti M, Nakagava HM, Pellini E, Araujo F, and Rubin J. Uniject as a delivery system for the once-a-month injectable contraceptive Cyclofem in Brazil. *Contraception*, 1996; **53**: 115–19.
64. Randomised controlled trial of levonorgestrel versus the Yuzpe regimen of combined oral contraceptives for emergency contraception. Task Force on Postovulatory Methods of Fertility Regulation. *Lancet*, 1998; **352**: 428–33.
65. von Hertzen H, Piaggio G, Ding J, *et al*. Low dose mifepristone and two regimens of levonorgestrel for emergency contraception: a WHO multicentre randomised trial. *Lancet*, 2002; **360**: 1803–10.
66. DMPAt-medroxyprogesterone acetate (DMPA) and risk of epithelial ovarian cancer. The WHO Collaborative Study of Neoplasia and Steroid Contraceptives. *Int J Cancer*, 1991; **49**: 191–5.
67. Oral contraceptive use and the risk of ovarian cancer. The Centers for Disease Control cancer and steroid hormone study. *JAMA*, 1983; **249**: 1596–9.
68. Vessey MP, Painter R. Endometrial and ovarian cancer and oral contraceptives—findings in a large cohort study. *Br J Cancer*, 1995; **71**: 1340–2.
69. Hankinson SE, Colditz GA, Hunter DJ, Spencer TL, Rosner B, Stampfer MJ. A quantitative assessment of oral contraceptive use and risk of ovarian cancer. *Obstet Gynecol*, 1992; **80**: 708–14.
70. Ness RB, Grisso JA, Klapper J, *et al*. Risk of ovarian cancer in relation to estrogen and progestin dose and use characteristics of oral contraceptives. SHARE Study Group. Steroid Hormones and Reproductions. *Am J Epidemiol*, 2000; **152**: 233–41.
71. Royar J, Becher H, Chang-Claude J. Low-dose oral contraceptives: protective effect on ovarian cancer risk. *Int J Cancer*, 2001; **95**: 370–4.
72. Gross TP, Schlesselman JJ. The estimated effect of oral contraceptive use on the cumulative risk of epithelial ovarian cancer. *Obstet Gynecol*, 1994; **83**: 419–24.
73. Combination oral contraceptive use and the risk of endometrial cancer. The Cancer and Steroid Hormone Study of the Centers

for Disease Control and the National Institute of Child Health and Human Development. *JAMA*, 1987; **257**: 796–800.
74. Schlesselman JJ. Net effect of oral contraceptive use on the risk of cancer in women in the United States. *Obstet Gynecol*, 1995; **85**(5 Pt 1): 793–801.
75. Hormonal contraception: recent advances and controversies. *Fertil Steril*, 2008; **90**(5 Suppl): S103–13.
76. Mulders TM, Dieben TO. Use of the novel combined contraceptive vaginal ring NuvaRing for ovulation inhibition. *Fertil Steril*, 2001; **75**: 865–70.
77. Roumen FJ, Apter D, Mulders TM, Dieben TO. Efficacy, tolerability and acceptability of a novel contraceptive vaginal ring releasing etonogestrel and ethinyl oestradiol. *Hum Reprod*, 2001; **16**: 469–75.
78. DMPAt-medroxyprogesterone acetate (DMPA) and risk of endometrial cancer. The WHO Collaborative Study of Neoplasia and Steroid Contraceptives. *Int J Cancer*, 1991; **49**: 186–90.
79. Ye Z, Thomas DB, Ray RM. Combined oral contraceptives and risk of cervical carcinoma in situ. WHO Collaborative Study of Neoplasia and Steroid Contraceptives. *Int J Epidemiol*, 1995; **24**: 19–26.
80. Gram IT, Macaluso M, Stalsberg H. Oral contraceptive use and the incidence of cervical intraepithelial neoplasia. *Am J Obstet Gynecol*, 1992; **167**: 40–4.
81. Moreno V, Bosch FX, Munoz N, *et al*. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. *Lancet*, 2002; **359**: 1085–92.
82. Lin J, Zhang SM, Cook NR, Manson JE, Buring JE, Lee IM. Oral contraceptives, reproductive factors, and risk of colorectal cancer among women in a prospective cohort study. *Am J Epidemiol*, 2007; **165**: 794–801.
83. Fernandez E, La Vecchia C, Balducci A, Chatenoud L, Franceschi S, Negri E. Oral contraceptives and colorectal cancer risk: a meta-analysis. *Br J Cancer*, 2001; **84**: 722–7.
84. Beral V, Hermon C, Kay C, Hannaford P, Darby S, Reeves G. Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46 000 women from Royal College of General Practitioners' oral contraception study. *BMJ*, 1999; **318**: 96–100.
85. Stewart FH, Harper CC, Ellertson CE, Grimes DA, Sawaya GF, Trussell J. Clinical breast and pelvic examination requirements for hormonal contraception: current practice vs evidence. JAMA, 2001; **285**: 2232–9.

8.8

Exogenous Factors and Female Reproductive Health

8.8.1 Exogenous Factors and Female Reproductive Health

Common Extragonadal Endocrinopathies Affecting Reproduction

Alessandra Gambineri and Daniela Ibarra-Gasparini

Introduction

Female infertility is estimated to affect between 8 and 15% of reproductive-aged couples worldwide. It is defined as the failure to establish a clinical pregnancy after 12 months of regular and unprotected sexual intercourse in women younger than 35 and after 6 months in women over the age of 35 [1]. In a World Health Organization (WHO) study of 8500 infertile couples, female infertility was reported in 37% of cases, male infertility in 8% of cases, and both male and female infertilities in 35% of cases [2]. Different causes may be behind female fertility, but ovulatory disorders account for more than a quarter of cases. Other causes of female infertility are endometriosis (15%), pelvic adhesions (11%), tubal blockage (11%), other tubal abnormalities (11%), and hyperprolactinemia (7%). Ovulatory disorders are more frequent in women with absent or irregular menses, whereas they are unusual in women who report monthly menses. Given that many conditions are potentially associated with ovulatory dysfunctions, the WHO classified anovulation into three main groups summarized in Box 8.8.1.1; hyperprolactinemia is recognized as an additional aetiology. Advanced age plays a triggering role in aggravating female infertility [3]. One study performed in 2004 on European women found that fertility peaks in the early and mid 20s; thereafter, there is a gradual fertility decline, and at age 40, a woman has only 3% of her ovarian reserve [4].

Box 8.8.1.1 World Health Organization (WHO) classification of anovulation

WHO class 1: Hypogonadotropic hypogonadal anovulation
These women have low or low-normal circulating levels of gonadotropins and low serum oestradiol concentrations due to decreased hypothalamic secretion of the gonadotropin-releasing hormone (GnRH), or pituitary unresponsiveness to GnRH, or decreased pituitary secretion of gonadotropins.

WHO class 2: Normogonadotropic normo-oestrogenic anovulation
These women have normal circulating levels of gonadotropins and of oestradiol. However, FSH secretion during the follicular phase of the menstrual cycle is subnormal. This group includes women with polycystic ovary syndrome (PCOS).

WHO class 3: Hypergonadotropic hypoestrogenic anovulation
These women have high circulating levels of gonadotropins and low serum oestradiol concentrations. The primary causes are premature ovarian failure and ovarian resistance to gonadotropins.

Hyperprolactinaemic anovulation
The anovulation is due to the inhibitory effect of prolactin on gonadotropin secretion. These women usually have normal circulating levels of gonadotropins and low serum oestradiol concentrations.

The Complex Hormonal Regulation of Folliculogenesis and Ovarian Steroidogenesis

The female reproductive function is influenced by the ovarian life span. The ovary is a metabolically active organ and acts as a germ cell reservoir. It consists of approximately 0.3 million primordial follicles that contain diplotene-arrested oocytes which represent the ovarian basic functional units. Folliculogenesis is a complex mechanism leading to the development of a mature follicle and ovulation. Each follicle has its own unique fate which is controlled by many endocrine and paracrine factors that support the passage from the first stage of follicle development, including the activation

of a primordial follicle that develops into primary and secondary follicles (preantral follicles and small antral follicles), through to the final stage of follicle development. This is where the small antral follicle develops into the antral follicle or Graafian follicle under the stimulatory control of pituitary gonadotropins.

In women of reproductive age, Graafian follicles are the major source of the cyclic secretion of ovarian oestrogens. In response to the preovulatory gonadotropin surge during each reproductive cycle, the dominant Graafian follicle ovulates to release a mature oocyte for fertilization, whereas the remaining theca cells (TCs) and granulosa cells (GCs) undergo transformation to become the corpus luteum that contributes to the circulating progesterone source. The production of steroid hormones by the ovarian cells is essential for supporting each stage of folliculogenesis; steroidogenesis is therefore a prerequisite for successful reproduction. Although the follicle-stimulating hormone (FSH) and luteinizing hormone (LH) are the primary regulators of folliculogenesis and of ovarian steroidogenesis, other hormones such as the thyroid hormones (THs), growth hormone (GH), prolactin (PRL), glucocorticoids and androgens participate in the regulation of follicular development and modulate the production of ovarian steroids (**Figure 8.8.1.1 and 8.8.1.2**).

THs receptors, triiodothyronine (T_3) and thyroxine (T_4) are present in the follicular fluid, whereas thyroid-stimulating hormone (TSH) receptors, THs transporters, and THs receptors are expressed in the ovary at the various developmental stages of the follicle. Studies in rats have demonstrated that T_3 amplifies proliferative FSH action at the GCs of the preantral follicle, stimulating its transition to the small antral follicle. In addition, T3 inhibits GCs apoptosis, through the activation of the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt) pathway, which is the pathway involved in the FSH-induced regulation of GCs survival [5]. Accordingly, THs administration in hypothyroid rats has been demonstrated to improve the maturation of ovarian follicles, increasing the number of healthy large antral follicles and ovulatory oocytes synergistically with gonadotropins [6]. Studies in humans are scarce. One *in vitro* study performed on fresh ovarian tissues form fertile women did not find an effect of T_4 or TSH administration on the follicle development [7], however other studies have demonstrated that THs induce oestradiol and progesterone ovarian synthesis acting directly at GCs. In particular, T_4 amplifies gonadotropins action on luteinization and progesterone secretion [8], whereas T_3 amplifies FSH stimulatory action on cytochrome P450 lanosterol 14α-demethylase (CYP51) expression. The latter is a key enzyme involved in sterol and steroid biosynthesis, expressed in oocytes and GCs of primordial follicles and of growing follicles [5].

Other studies have supported the role of GH in modulating folliculogenesis as well as ovarian steroidogenesis in humans. These actions may reflect the direct autocrine and paracrine actions of GH in the ovary or may be mediated by the Insulin Growth Factor (IGF)-1, whose hepatic production is induced by GH, since both GH- and IGF1-receptors have been found in the human ovaries. In addition, the GH releasing hormone (GHRH) and GHRH-receptors have been found in the ovary, thus supporting the presence of an ovarian regulation of GH synthesis by GH secretagogues in addition to the traditional hypothalamic–pituitary–regulatory axis. There is increasing evidence that GH plays an important role in controlling both follicular development and atresia. In particular, GH directly stimulates both the gonadotropin independent and the gonadotropin dependent phases of the follicle growth, acting at the level of small, preantral, and antral follicles. It has also been suggested that GH enhances follicular survival and cell proliferation by potentiating LH action at the ovary. Accordingly, GH deficiency is associated with a decrease in LH-receptors in the ovaries. GH also plays a role in follicle selection and in stimulating the development of the dominant follicle by enhancing nuclear and cytoplasmic oocyte maturation, by increasing the sensitivity of the follicle to gonadotrophins and by reducing the incidence of apoptosis of preovulatory ovarian follicles. In addition, GH facilitates ovulation by increasing tissue plasminogen activator synthesis, which plays an important role in the rupture of the ovarian capsule. GH also has

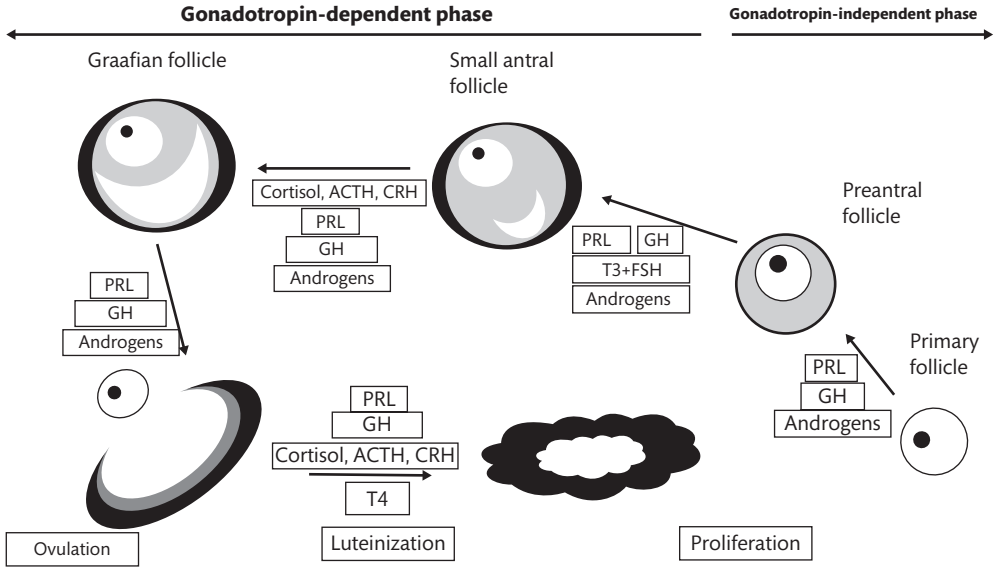


Figure 8.8.1.1 The complex hormonal regulation of folliculogenesis. Hormones of pituitary and/or ovarian origin bind their receptors on thecal cells (black), granulosa cells (grey) and luteal cells, and promote gametogenesis, thus inducing the development of primary follicles into preantral, antral, and mature antral (Graafian) follicles and taking part in ovulation and luteinization. ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; FSH, follicular stimulating hormone; GH, growth hormone; PRL, prolactin; T_3, triiodothyronine; T_4, thyroxine.

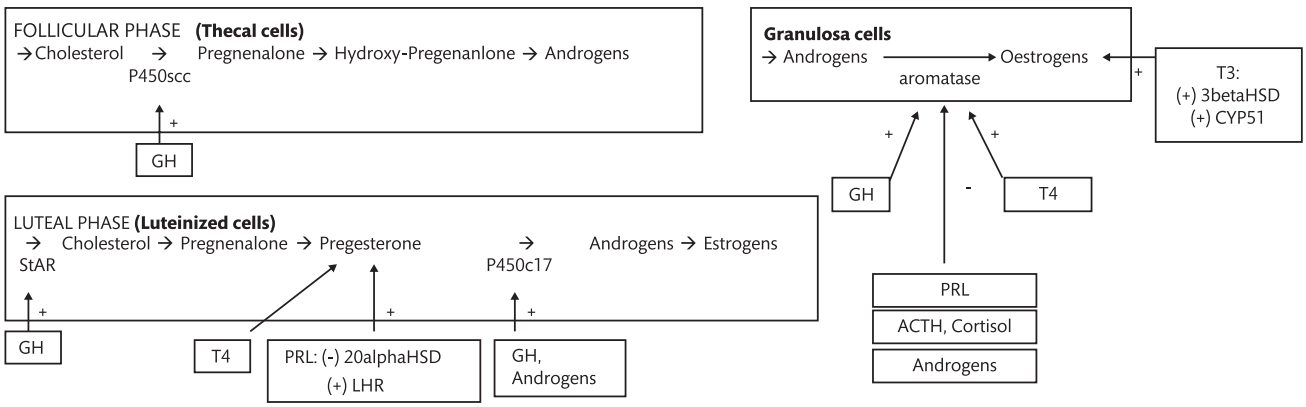


Figure 8.8.1.2 The complex hormonal regulation of ovarian steroidogenesis. Both thecal and granulosa cells are required for the production of oestrogens during the follicular phase. Following ovulation, granulosa cells become luteinized and acquire the ability to synthesize and secrete both oestrogen and progesterone. (+) indicates a stimulatory role; (–) indicates an inhibitory role. GH, Growth hormone; T_3, triiodothyronine; T_4, thyroxine; PRL, prolactin; ACTH, adrenocorticotropic hormone; FSH, follicular stimulating hormone; LHR, luteinizing hormone receptor; 20alphaHSD, 20 alpha hydroxyl-steroid dehydrogenase; StAR, steroidogenic acute regulatory protein; P450scc, cytochrome P450 side-chain cleavage; P450c17, 17alpha-hydroxylase; 3betaHSD, 3beta-Hydroxysteroid dehydrogenase; CYP51, cytochrome P450 lanosterol 14α-demethylase.

a stimulatory impact on local Aquaporin 1 (AQP1) synthesis. The amount of AQP1 at the level of the GCs and TCs membrane regulates cellular permeability to water, whereas its amount at the level of the cytoplasm regulates the intracellular action of steroid hormones, thus influencing the formation of human follicular antrum, the growth of follicles, and finally folliculogenesis [9]. GH regulates steroidogenesis in the GCs and TCs of the follicle through the stimulation of the activity of several steroidogenic enzymes, such as steroidogenic acute regulatory protein (StAR), cytochrome P450 side-chain cleavage (P450scc), 17alpha-hydroxylase (P450c17), aromatase, and 3beta-hydroxysteroid dehydrogenase (3betaHSD). This thus increases both oestrogen and progesterone synthesis and secretion during the follicular or luteal phase of the menstrual cycle [10].

Prolactin (PRL) is primarily known for its role in breast development and lactation, however it is also involved in the regulation of a wide variety of other physiological functions, including reproduction. In the ovary, PRL acts synergistically with gonadotropins to stimulate follicular development through interaction with the PRL receptor (R), a member of the cytokine superfamily. There are several isoforms of PRL-R, above all the long (long-R) and the short (short-R) isoforms, which are identical in their extracellular domains but differ in the lengths and sequences of their cytoplasmic tail [11]. This difference leads to different signalling properties and, therefore, to different biological actions. Although the long-R is the most expressed isoform of PRL-R in the ovary, coexpression of both long-R and short-R isoforms is necessary for the regular development of ovarian follicles [12]. Long-R shows an intrinsic thyrosine kinase activity. Acting through the long-R, PRL activates many kinases associated with the cytoplasmic domain, including Janus kinase 2 (Jak2), MAP kinase, and PI3K [11]. JaK2 phosphorylates several cellular proteins such as a family of transcription factors termed signal transducers and activators of transcription (Stats) which translocate into the nucleus. Here they regulate gene expression by inducing several PRL-responsive genes involved in cell proliferation and differentiation. Short-R is not thyrosine phosphorylated [13], thus showing different actions from long-R. The overactivation of short-R, which follows an excess in PRL circulating levels, leads to the dephosphorylation of a dual specific phosphatase-uridine diphosphate (D-UDP1), which in turn deactivates MAP kinase leading to an ovarian defect [14]. Short-R also down-regulates forkhead box O-3 (FOXO3) and galactose-1-phosphate uridyltransferase (GALT), two genes involved in normal ovarian development [11]. Therefore, to guarantee a normal ovulatory process, short-R should not be overactivated. Accordingly, female mice with only the short-R display an important ovarian defect and infertility. PRL also influences follicle development through a stimulatory impact on AQP1 [9]. In rodents, PRL also shows a luteotropic effect due to the ovarian upregulation of LH receptor expression [15] and repression of 20 alpha hydroxyl-steroid-dehydrogenase (20-alpha-HSD), both resulting in the upregulation of ovarian progesterone secretion which is fundamental for a correct implantation [16]. When PRL is in excess, however, the inhibitory effect of PRL prevails over aromatase activity, thus leading to low ovarian oestrogen production [11]. In addition, PRL inhibits gonadotropin-releasing hormone (GnRH) secretion and the LH pituitary response to GnRH stimulation, thus reducing LH pulse amplitude and frequency with a consequent inhibition in ovarian function. The PRL inhibitory effect on GnRH release is not direct but is mediated through the inhibitory effect of PRL on kisspeptin [17], which is the most important stimulator of GnRH secretion. Accordingly, kisspeptin administration in hyperprolactinaemic anovulation restores gonadotropin secretion and ovulation [18].

Glucocorticoids are also significantly involved in ovulation. Glucocorticoids, in fact, act as anti-inflammatory agents in the ovary to promote healing and repair after ovulation and restore ovarian function under stress conditions by inhibiting oxidative stress [19]. In addition, both glucocorticoids as well as the adrenocorticotropic hormone (ACTH) and corticotropin-releasing hormone (CRH) stimulate follicular development [20], but inhibit oestradiol and progesterone production in the ovary [19–22]. In addition, endogenous glucocorticoids released in response to perceived stress are protective against gonadotropin secretion, by inhibiting prostaglandin synthesis in the brain, thereby maintaining LH secretion regardless of the nature of the stressor [23]. However, an excess of glucocorticoid secretion or exogenous glucocorticoid administration may lead to an inhibition of GnRH pulsatility, and a decrease in gonadotropin responsiveness and release from the pituitary [24], as

demonstrated by a reduction of about 60% in the peak of LH in response to intravenous GnRH after prednisolone administration [22]. Glucocorticoids in excess also impair LH action at the ovary thereby reducing the ovarian LH-receptor level [19]. Similar results have been observed in the presence of ACTH excess [25]. Studies in mice have demonstrated an increased number of atretic antral follicles with a decrease in corpora lutea (CLs) and a consequent impairment in follicular development in stressed mice which show high cortisol levels [19]. In human studies, stressed-induced hypercortisolism increased the level of reactive oxygen species (ROS) beyond the physiological range, resulting in cell cycle arrest and apoptosis in follicular oocytes, reducing fertilization and pregnancy rates [26].

Finally, androgens also play an important role in regulating folliculogenesis. Androgen receptors (ARs) are expressed in GCs, TCs, and stromal cells of follicles at all the maturity levels. The highest expression level of ARs is in the GCs of preantral and small antral follicles, while their expression gradually decreases when follicles enter the preovulation stage. Androgens, particularly testosterone, promote preantral follicle growth, through the activation of the PI3K/Akt/FOXO3 pathway [27] and the modulation of the expression of key ovarian growth factors, such as growth differentiation factor-9 (GDF9). They also accelerate the development of primary into antral follicles and inhibit follicular atresia in mammals, by enhancing the expression of an antiapoptotic microRNA (miR-125b) [28]. The mitogenic properties of androgens may be due in part to the activation of glucose transporter type 4 (GLUT4) signalling and the consequent increased ovarian glucose metabolism [29]. In addition, androgens may up-regulate FSH-Receptor levels in a transcription independent manner, thus sensitizing preantral follicles towards FSH actions [30]. The activity of some steroidogenic enzymes located in the TCs has also been demonstrated to be increased by androgens. In particular, GDF-9, which is induced by androgens, stimulates P450c17 expression and inhibits aromatase activity leading to an androgen-rich environment [30]. Androgens also induce the expression of the vascular endothelial growth factor (VEGF), which is important for vascularization of the ovarian follicle and the corpus luteum [27].

The Hormonal Regulation of the Uterus Decidualization and of Implantation

THs, GH, PRL, glucocorticoids, and androgens act also at the uterine level, stimulating decidualization, and favouring implantation.

THs stimulate the production of progesterone from GCs, which is responsible for building up the endometrial lining for optimal implantation. They also up-regulate the leukaemia inhibitory factor (LIF) which is involved in embryo implantation and increase the expression of various adhesion molecules in early placental extravillous trophoblasts such as metalloproteinases-2 and -3 (MMP-2, MMP-3), fetal fibronectin, and integrin, thus promoting the invasion process.

Also, PRL and GH exert paracrine and autocrine actions at the level of the uterus finally regulating reproduction and are also synthesized locally. PRL is synthesized by the human endometrium around day 23 of the normal menstrual cycle primarily under the control of progesterone and oestradiol [11]. PRL decidual production rapidly increases soon after implantation in order to maintain pregnancy, since it inhibits detrimental factors such as interleukin-6 (IL-6) and 20-alpha-HSD, by linking to both long and short isoforms of PRL-R [31].

GH stimulates mitogenesis and promotes uterine proliferation and endometrial cell growth through the upregulation of oestradiol receptors and the stimulation of the uterine production of IGF-binding proteins (IGFBPs) and prostaglandins [10]. Through its mitogenic activity, GH increases the placental size, thus stimulating fetal growth and inducing the production of placental lactogens, oestradiol, and progesterone, thus supporting pregnancy and the maternal metabolism to induce nutrient repartitioning for fetal development. GH has a stimulatory effect on AQP1 which is localized in the uterus, thus stimulating uterine imbibition and facilitating embryo implantation [9].

Glucocorticoids inhibit oestradiol-stimulated uterine growth. In fact, the administration of dexamethasone in rats has been found to attenuate the uterine weight gain induced by oestradiol, through a reduction in intracellular oestrogen receptor concentrations or by glucocorticoid receptor-mediated inhibition of the c-fos/c-jun transcription factor. The latter is used in the signal transduction pathways of many growth factors and is directly or indirectly stimulated by oestrogens. CRH, expressed in human decidualized endometrial stroma, seems to stimulate the process of decidualization synergistically with progesterone [22]. In addition, during the luteal phase of the menstrual cycle, it is secreted into the lumen of the uterus, where it participates in the inflammatory phenomena of blastocyst implantation and, later in the menstruation cycle [32]. CRH also seems to stimulate the production of the blood supply to the fetus by activating the nitric oxide synthase of placental vessels [22]. Finally, placental CRH stimulates cortisol secretion during the latter part of pregnancy, and thus is responsible for the maternal hypercortisolism of pregnancy [22].

In the endometrium, testosterone induces collagen type 1 expression and VEGF, thus finally regulating neoangiogenesis which is important for implantation. However, high levels of progesterone and/or 17-OH progesterone during the preovulatory phase of the menstrual cycle have been associated with inadequate endometrial maturation and impaired embryo implantation [33].

In summary, the ovaries and the uterus are under the control of many hormones and an excess or defect of these hormones may account for infertility, thus complicating the diagnosis of reproductive disorders.

Extragonadal Endocrinopathies Affecting Reproduction

Thyroid Diseases

Among the extraovarian causes of infertility, the most frequent endocrine diseases are those affecting the thyroid gland.

THs disorders in terms of both an excess or defect and thyroid-peroxidase autoantibody (TPO-Ab) or thyroglobulin autoantibody (Tg-Ab) production alterations are associated with a disturbance in folliculogenesis, fertilization, and embryogenesis, as well as with some pregnancy complications.

Hypothyroidism

Hypothyroidism has mainly been associated with menstrual irregularities and anovulation, although there are some data that describe a link between hypothyroidism and many pregnancy complications.

The prevalence of menstrual irregularities is 23% among hypothyroid females, compared to 8% in euthyroid females [34]. More recent reports describe a lower prevalence of menstrual abnormalities among hypothyroid females, which may be attributed to earlier detection and treatment of hypothyroidism compared to the past [2].

Few studies describe the association between hypothyroidism and low fertility rates both in animals [35] and in humans where high serum TSH levels would seem to be a predictor of fertilization failure in women undergoing *in vitro* fertilization [36]. Hypothyroidism is also associated with early and late obstetric complications, namely miscarriage, preterm birth, placental abruption, fetal death, postpartum haemorrhage, respiratory distress syndrome, and impaired neurological development during childhood [37].

There may also be a link between low fertility rate and subclinical hypothyroidism which is defined as serum TSH levels over 4.5 mIU/L with normal T_4 and T_3 circulating levels [37]. In addition, some studies have described an association between subclinical hypothyroidism and reduced pregnancy outcome [37, 38] which improves after T_4 supplementation [39]. Accordingly, the 2017 American Thyroid Association (ATA) guidelines suggest considering treatment with T_4 with TSH levels above 4 mIU/L [40].

The link between hypothyroidism and infertility is related to three main factors: (i) a decrease in folliculogenesis (see previous section); (ii) a decrease in the sex hormone binding globulin (SHBG) synthesis and secretion by the liver, with a consequent increased peripheral bioavailability of oestrogens and, particularly, of androgens; and (iii) to an increased pituitary PRL secretion stimulated by TRH [2]. Reduced fertility in hypothyroid patients is also due to the peripheral insulin resistance and to the consequent hyperinsulinaemia that frequently follows hypothyroidism [37].

Altered Thyroid Autoimmunity with or Without Hypothyroidism

The most frequent form of autoimmune thyroid disorder is that sustained by TPO-Ab and/or Tg-Ab which may or may not be associated with hypothyroidism. The impact on fertility seems to be different among the forms with or without hypothyroidism [41]. In fact, although many studies have reported an increased prevalence of thyroid autoimmunity in women attending infertility centres [42], most studies suggest that pregnancy rates and pregnancy outcomes are negatively influenced only if thyroid antibodies are associated with hypothyroidism [38, 43]. However, some studies have reported a decreased pregnancy rate [44] and an increased miscarriage rate and obstetric complications also with normal TSH levels [45].

Although TPO-Ab has been found in the follicular fluid of women with autoimmune thyroid disorders [44], there are no data demonstrating a direct negative impact of TPO-Ab on folliculogenesis, nor on fertilization, implantation, or embryogenesis, although some authors have suggested the role of inflammation induced by autoimmunity in reducing oocyte quality. In any case, the 2017 ATA guidelines suggest considering treatment with T_4 with TSH levels above 2.5 mIU/L if there is a positivity for TPO-Ab [40].

Hyperthyroidism

Hyperthyroidism has been associated with a 5.8% infertility rate [46], which seems to be higher in the presence of autoimmune hyperthyroidism [41].

Menstrual irregularities are present in 8–60% of patients affected by hyperthyroidism and include oligomenorrhea or amenorrhea, hypermenorrhoea, and even anovulation [34, 37]. Hyperthyroidism in pregnancy is associated with an elevated prevalence of miscarriage, stillbirth, preterm birth, and intrauterine growth restriction. In addition, hyperthyroid pregnant females may develop preeclampsia or placental abruption [37].

Whether the negative impact of hyperthyroidism on fertility is entirely attributable to the direct action of THs on the ovaries and uterus (see previous section) or whether other mechanisms are involved is unknown. Indeed, thyrotoxicosis is associated with increased circulating levels of SHBG [47], with a consequent decreased peripheral bioavailability of oestrogens and, particularly, of androgens. An increased production rate of testosterone and androstenedione in hyperthyroid females has also been described [48], as well as an altered GnRH-induced LH secretion with a consequent significantly higher LH level in both the follicular and luteal phase of the menstrual cycle, thus contributing to anovulation [41].

Diseases of the Hypothalamic–Pituitary Axis

Several diseases affecting the hypothalamic–pituitary axis may impact on reproductive function, through a reduction in the GnRH drive to the pituitary and/or a reduction of FSH and LH input to the ovary with a consequent insufficient folliculogenesis and ovarian function.

Clinically variable menstrual cycle alterations, including oligomenorrhoea, amenorrhoea, or polymenorrhoea may be detected.

In rare cases, the reduction in GnRH drive may be due to gene mutations such as the fibroblast growth factor receptor 1 (FGFR1), prokineticin receptor 2 (PROKR2) and GnRH-receptor, to congenital GnRH deficiency due to the migration of an insufficient number of GnRH neurons from the olfactory placode into the hypothalamus during fetal development (Kallmann's syndrome) or to hypothalamic lesions [49]. More frequently it may be functional. Similarly, pituitary causes can be congenital or acquired. Congenital causes are alterations in genes involved in pituitary embryogenesis such as Homeobox protein prophet of PIT-1 (PROP-1) and Homeobox expressed in embryonic stem (ES) cells 1 (HESX1) [50]. Acquired causes are pituitary adenomas or other pituitary tumours, empty sella syndrome, trauma, or radiation of the pituitary area, sarcoidosis, lymphocytic hypophysitis (autoimmune diseases), and ischaemic alterations such as Sheehan's syndrome [50]. Hyperprolactinemia is the most frequent cause of infertility due to hypothalamic–pituitary axis dysfunction and may be due to organic or functional causes.

Pituitary Adenomas

Pituitary adenomas may be associated with infertility due to hypogonadotropic hypogonadism. There are three main underlying mechanisms: (i) hormonal hypersecretion (PRL, GH or cortisol); (ii) pathological interruption of hypothalamic–pituitary dopaminergic pathways (i.e. pituitary stalk compression), with consequent hyperprolactinemia; and (iii) the destruction or compression of

gonadotroph cells as a mass effect. Pituitary hormonal hyper- and hyposecretion also influence pregnancy outcomes [51]. However, with the improvement in clinical and surgical treatment, women harbouring pituitary adenomas can still become pregnant and give birth.

Hyperprolactinaemia

A number of physiological states including pregnancy, breast-feeding, stress, lack of exercise, and sleep can cause an increase in PRL. Non-physiological hyperprolactinemia may be due to pituitary lactotroph adenomas (prolactinomas), which account for approximately 40% of all pituitary tumours, pharmacological stimulation of PRL pituitary secretion, such as D2 antagonists administration, pathological interruption of hypothalamic–pituitary dopaminergic pathways (pituitary stalk compression), hypothyroidism or renal disorders, and, sometimes, due to unknown reasons (the idiopathic form). Regardless of aetiology, hyperprolactinemia may clinically manifest with galactorrhoea, menstrual abnormalities, hypogonadism, and infertility, which is found in approximately 10% of the patients affected, or it may remain asymptomatic [52]. In addition, women often note a decrease in libido and dyspareunia due to oestrogen deficiency [53].

Infertility in hyperprolactinemia is due both to the direct impact of PRL on folliculogenesis and to a PRL-induced hypogonadotropic hypogonadism, since PRL reduces GnRH secretion (see previous section). When pulsatile GnRH is administered in hyperprolactinaemic patients, ovulatory cycles and fertility may be restored [2].

Acromegaly

Acromegaly is a chronic disorder caused by GH hypersecretion which affects fertility in approximately 50% of females. Infertility in acromegalic patients is principally due to anovulation mainly for pituitary reasons, such as the destruction or compression of gonadotroph cells by the GH-secreting adenoma or hyperprolactinemia derived from tumour cosecretion or the stalk effect (pituitary stalk compression). Anovulation in acromegaly may also be due to polycystic ovary syndrome (PCOS) which derives from the direct effect of excessive GH/IGF-I secretion on the ovaries and/or from insulin resistance secondary to GH excess [53]. In fact, hyperinsulinemia, which is compensatory to insulin resistance, stimulates androgen production by TCs and reduces SHBG hepatic synthesis thus increasing the peripheral bioavailability of androgens. In acromegaly, however, infertility is also due to the increased prevalence of spontaneous abortion [54] and to pregnancy complications, such as gestational diabetes and hypertension [51, 55].

Cushing's Disease

The term Cushing's syndrome (CS) identifies a group of diseases caused by a chronic excess of glucocorticoids, and the term Cushing's disease identifies the forms of CS that are ACTH dependent. Menstrual irregularities and chronic anovulation are common in women with CS, affecting up to 80% of these patients and who clinically manifest with oligomenorrhea, amenorrhea, or polymenorrhoea. As a consequence, infertility is frequent in females affected by CS.

The mechanisms responsible for chronic anovulation in CS are still not completely clear, although high circulating levels of cortisol inhibits GnRH and gonadotropin release, thus resulting in hypogonadotropic hypogonadism [24]. In addition, high cortisol levels reduce the ovarian LH-receptor expression thus impairing LH action at the ovary, (19) thus contributing to anovulatory infertility [19, 26] (for more details, see previous section). The adrenal hyperandrogenism that accompanies all the forms of CS also plays a role in infertility [25].

Hypopituitarism

Hypopituitarism is defined by the decreased secretion of one or more of the hormones normally produced by the pituitary gland. The signs and symptoms of hypopituitarism vary depending on which hormone is lacking, with infertility always present when FSH and LH secretion are affected. A low serum oestradiol concentration with equally low LH and FSH circulating levels is the diagnostic key of gonadotropin deficiency. Hypoandrogenism is also a typical feature of the pituitary diseases affecting ACTH and gonadotropin secretion, with a consequent defect in the follicle maturation which stops growing at the small follicle stage. The consequence is a phenotype of low functional ovarian reserve, characterized by low oestradiol and anti-Müllerian hormone (AMH) levels due to declining GCs mass. The deficit in TSH and GH production contributes to hypogonadism and to anovulation through their impact on folliculogenesis and steroidogenesis as described in the previous section, although fertility is always not compromised in the isolated deficiency of the GH axis. In fact, a high proportion of GH-deficient women have normal menstrual cycles and conceive normally [10]. However, a deficiency in GH is frequently associated with intrauterine growth retardation. In addition, GH administration in hypopituitaric patients with both gonadotropin and GH deficiency significantly reduces the dosage and duration of the gonadotropin treatment required for ovulation induction and for a positive fertility outcome [50].

Diseases of the Adrenal Gland

Infertility in diseases of the adrenal gland may arise from hypercortisolism as well as from a glucocorticoid deficiency [19, 26]. There are two main adrenal causes of glucocorticoid deficiency; Addison's disease and congenital adrenal hyperplasia.

Addison's Disease

Addison's disease or primary adrenal insufficiency is characterized by a deficiency in cortisol, aldosterone, and androgen adrenal synthesis and secretion, usually caused by an autoimmune reaction towards the adrenal cortex. Data on the impact of Addison's disease on fertility and pregnancy outcomes are scarce. Recently, however, many studies have demonstrated reduced fertility in these patients [27] with a concomitant reduction in parity and an increased risk of unfavourable pregnancy outcomes such as caesarean section and preterm delivery, although an adequate glucocorticoid treatment generally improves pregnancy outcomes [56].

The underlying causes of reduced fertility in Addison's disease still need to be clearly defined, however many other factors are potentially involved than just a deficit in glucocorticoids. Excess glucocorticoid therapy has been suggested, together with the possible coexistence of other autoimmune diseases affecting fertility, such as autoimmune thyroid disease or premature ovarian failure

(POF), which has been found in 7% of women with Addison's disease [57], and also, interestingly, adrenal hypoandrogenism. A defect in adrenal androgens, in particular, seems to play a central role in fertility impairment and in the increase in spontaneous abortions observed in females affected by Addison's disease.

In fact, low testosterone results in poor follicle maturation and in the arrest of follicles at the small follicle growth stages. The consequence is a typical phenotype characterized by low oestradiol and high FSH levels, low testosterone and dehydroepiandrosterone (DHEA) and DHEA sulphate (DHEAS) levels, but high AMH levels, which are sometimes erroneously diagnosed as a premature ovarian insufficiency (POI). These cases, however, are not POI and should be diagnosed and specifically treated.

Since DHEA and DHEAS are exclusively produced by the adrenal and, therefore, reflect adrenal androgen production in the zona reticularis, the measurement of DHEA and DHEAS circulating levels could help clinicians in diagnosing these forms [58] which are surprisingly common in tertiary fertility centres but often go unrecognized [59]. A correct diagnosis is detrimental for the prognosis of these forms since androgen supplementation can improve ovarian function [60]. DHEA supplementation in such patients in fact improves both spontaneous fertility and *in vitro* fertilization (IVF) cycle outcomes [59].

Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency is a common autosomal recessive disorder due to mutations in the *CYP21A2* gene. Irregular menses are common in females with CAH, with amenorrhea being more frequent in patients with salt wasting (SW) and simple virilizing (SV) forms compared to non-classical (NCAH) forms [61]. In a recent report, pregnancy and delivery rates were significantly lower in 62 women with CAH (aged 18–63) with respect to controls, despite fertility treatments. These cases were also influenced by the severity of the genotype with a more severe negative impact on fertility outcomes observed in SW and the less severe impact observed in the NCAH forms [62].

The reduction in pregnancy and birth rates described in CAH can be attributed to several factors. It may be due to decreased sexual activity due to genital anomalies and to the consequences of genital surgery; psychological factors (i.e. homosexuality) and single status; together with the secondary PCOS phenotype, as a consequence of the hyperandrogenic state. Inappropriate therapy is also implicated in the infertility of CAH which has been observed both if corticosteroid supplementation is insufficient and if there is a glucocorticoid overtreatment for the mechanisms described in previous section.

When glucocorticoid supplementation is insufficient, androgen overproduction from the adrenal suppresses normal follicular development inducing follicular atresia and suppresses gonadotropin secretion from the pituitary, thus leading to anovulation and infertility. In addition, elevated progesterone concentrations that characterize CAH due to 21-hydroxylase deficiency interfere with the quality of the cervical mucus [2, 33]. Conversely, with glucocorticoid overtreatment, the development of obesity and insulin resistance can contribute to decrease fertility in females affected by CAH. Accordingly, when CAH patients are adequately treated, they have good fertility outcomes and pregnancies are usually spontaneous. Thus, an optimized glucocorticoid therapy regimen is important in CAH women also to guarantee adequate fertility [62].

REFERENCES

1. Practice Committee of the American Society for Reproductive Medicine. Definitions of infertility and recurrent pregnancy loss: a committee opinion. *Fertil Steril*, 2013; **99**: 63.
2. Unuane D, Tournaye H, Velkeniers B, Poppe K. Endocrine disorders and female infertility. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 861–73.
3. Wallace WHB, Kelsey TW. Human ovarian reserve from conception to the menopause. *PLoS One*, 2010; **5**: e8772.
4. Dunson DB, Baird DD, Colombo B. Increased infertility with age in men and women. *Obstet Gynecol*, 2004; **103**: 51–6.
5. Liu J, Tian Y, Yu D, *et al*. Role of CYP51 in the regulation of T3 and FSH-induced steroidogenesis in female mice. *Endocrinology*, 2017; **158**: 3974–87.
6. Jiang JY, Umezu M, Sato E. Improvement of follicular development rather than gonadotrophin secretion by thyroxine treatment in infertile immature hypothyroid rdw rats. *J Reprod Fertil*, 2000; **119**: 193–9.
7. Aghajanova L, Lindeberg M, Carlsson IB, *et al*. Receptors for thyroid-stimulating hormone and thyroid hormones in human ovarian tissue. *Reprod Biomed Online*, 2009; **18**: 337–47.
8. Wakim AN, Polizotto SL, Burholt DR. Influence of thyroxine on human granulosa cell steroidogenesis *in vitro*. *J Assist Reprod Genet*, 1995; **12**: 274–7.
9. Skowronski MT, Mlotkowska P, Tanski D, *et al*. Pituitary gonadotropins, prolactin and growth hormone differentially regulate AQP1 expression in the porcine ovarian follicular cells. *Int J Mol Sci*, 2018; **19**: 5.
10. Hull KL, Harvey S. Growth hormone: roles in female reproduction. *J Endocrinol*, 2001; **168**: 1–23.
11. Devi YS, Halperin J. Reproductive actions of prolactin mediated through short and long receptor isoforms. *Mol Cell Endocrinol*, 2014; **382**: 400–10.
12. Russell DL, Richards JS. Differentiation-dependent prolactin responsiveness and STAT (signal transducers and activators of transcription) signaling in rat ovarian cells. *Mol Endocrinol*, 1999; **13**: 2049–64.
13. Devi YS, Shehu A, Halperin J, *et al*. Prolactin signaling through the short isoform of the mouse prolactin receptor regulates DNA binding of specific transcription factors, often with opposite effects in different reproductive issues. *Reprod Biol Endocrinol*, 2009; **7**: 87.
14. Devi YS, Seibold AM, Shehu A, *et al*. Inhibition of MAPK by prolactin signaling through the short form of its receptor in the ovary and decidua: involvement of a novel phosphatase. *J Biol Chem*, 2011; **286**: 7609–18.
15. Bachelot A, Carré N, Mialon O, *et al*. The permissive role of prolactin as a regulator of luteinizing hormone action in the female mouse ovary and extragonadal tumorigenesis. *Am J Physiol Endocrinol Metab*, 2013; **305**: E845–E852.
16. Bouilly J, Sonigo C, Auffret J, Gibori G, Binart N. Prolactin signaling mechanisms in ovary. *Mol Cell Endocrinol*, 2012; **356**: 80–7.
17. Sonigo C, Bouilly J, Carré N, *et al*. Hyperprolactinemia-induced ovarian acyclicity is reversed by kisspeptin administration. *J Clin Invest*, 2012; **122**: 3791–5.

18. Gaytán F, Gaytán M, Castellano JM, *et al*. KiSS-1 in the mammalian ovary: distribution of kisspeptin in human and marmoset and alterations in KiSS-1 mRNA levels in a rat model of ovulatory dysfunction. *Am J Physiol Endocrinol Metab*, 2009; **296**: E520–31.
19. Kala M, Nivsarkar M. Role of cortisol and superoxide dismutase in psychological stress induced anovulation. *Gen Comp Endocrinol*, 2016; **225**: 117–24.
20. Rae MT, Hillier SG. Steroid signaling in the ovarian surface epithelium. *Trends Endocrinol Metab*, 2005; **16**: 327–33.
21. Su AI, Cooke MP, Ching KA, *et al*. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A*, 2002; **99**: 4465–70.
22. Chrousos GP, Torpy DJ, and Gold PW. Interactions between the hypothalamic-pituitary-adrenal axis and the female reproductive system: clinical implications. *Ann Intern Med*, 1998; **129**: 229–40.
23. Matsuwaki T, Kayasuga Y, Yamanouchi K, Nishihara M. Maintenance of gonadotropin secretion by glucocorticoids under stress conditions through the inhibition of prostaglandin synthesis in the brain. *Endocrinology*, 2006; **147**: 1087–93.
24. Oakley AE, Breen KM, Clarke IJ, *et al*. Cortisol reduces gonadotropin-releasing hormone pulse frequency in follicular phase ewes: influence of ovarian steroids. *Endocrinology*, 2009; **150**: 341–9.
25. Kaltsas GA, Korbonits M, Isidori AM, *et al*. How common are polycystic ovaries and the polycystic ovarian syndrome in women with Cushing's syndrome? *Clin Endocrinol (Oxf)*, 2000; **53**: 493–500.
26. Prasad S, Tiwari M, Pandey AN, Shrivastav TG, Chaube SK. Impact of stress on oocyte quality and reproductive outcome. *J Biomed Sci*, 2016; **23**: 36.
27. Lebbe M and Woodruff TK. Involvement of androgens in ovarian health and disease. *Mol Hum Reprod*, 2013; **19**: 828–37.
28. Sen A, Prizant H, Light A, *et al*. Androgens regulate ovarian follicular development by increasing follicle stimulating hormone receptor and microRNA-125b expression. *Proc Natl Acad Sci U S A*, 2014; **111**: 3008–13.
29. Sato K, Iemitsu M, Aizawa K, Ajisaka R. Testosterone and DHEA activate the glucose metabolism-related signaling pathway in skeletal muscle. *Am J Physiol Endocrinol Metab*, 2008; **294**: E961–8.
30. Jie-Xue P, Jun-Yu Z, Zhang-Hong K, *et al*. Androgens as double-edged swords: induction and suppression of follicular development. *Hormones (Athens)*, 2015; **14**: 190–200.
31. Bachelot A, Binart N. Reproductive role of prolactin. *Reproduction*, 2007; **133**: 361–9.
32. Makrigiannakis A, Margioris AN, Le Goascogne C, *et al*. Corticotropin-releasing hormone (CRH) is expressed at the implantation sites of early pregnant rat uterus. *Life Sci*, 1995; **57**: 1869–75.
33. Hagenfeldt K, Janson PO, Holmdahl G, *et al*. Fertility and pregnancy outcome in womem with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Hum Reprod*, 2008; **23**: 1607–13.
34. Krassas G. E., Poppe K., and Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
35. Bernal A, DeMoraes GV, Thrift TA, Willard CC, Randel RD. Effects of induced hypothyroidism on ovarian response to superovulation in Brahman (Bos indicus) cows. *J Anim Sci*, 1999; **77**: 2749–56.
36. Cramer DW, Sluss PM, Powers RD, *et al*. Serum prolactin and TSH in an *in vitro* fertilization population: is there a link between fertilization and thyroid function? *J Assist Reprod Genet*, 2003; **20**: 210–5.
37. Dittrich R, Beckmann MW, Oppelt PG, *et al*. Thyroid hormone receptors and reproduction. *J Reprod Immunol*, 2011; **90**: 58–66.
38. Benhadi N, Wiersinga WM, Reitsma JB, Vrijkotte TG, Bonsel GJ. Higher maternal TSH levels in pregnancy are associated with increased risk for miscarriage, fetal or neonatal death. *Eur J Endocrinol*, 2009; **160**: 985–91.
39. Kim CH, Ahn JW, Kang SP, Kim SH, Chae HD, Kang BM. Effect of levothyroxine treatment on *in vitro* fertilization and pregnancy outcome in infertile women with subclinical hypothyroidism undergoing *in vitro* fertilization/intracytoplasmic sperm injection. *Fertil Steril*, 2011; **95**: 1650–4.
40. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
41. Mansourian AR. Female reproduction physiology adversely manipulated by thyroid disorders: a review of literature. *Pak J Biol Sci*, 2013; **16**: 112–20.
42. Poppe K, Velkeniers B, Glinoer D Thyroid autoimmunity and female infertility. *Thyroid Int*, 2008; **4**: 1–11.
43. Kilic S, Tasdemir N, Yilmaz N, Yuksel B, Gul A, Batioglu S. The effect of anti-thyroid antibodies on endometrial volume, embryo grade and IVF outcome. *Gynecol Endocrinol*, 2008; **24**: 649–55.
44. Monteleone P, Parrini D, Faviana P, *et al*. Female infertility related to thyroid autoimmunity: the ovarian follicle hypothesis. *Am J Reprod Immunol*, 2011; **66**: 108–14.
45. Poppe K, Glinoer D, Van Steirteghem A, *et al*. Thyroid dysfunction and autoimmunity in infertile women. *Thyroid*, 2002; **12**: 997–1001.
46. Joshi JV, Bhandarkar SD, Chadha M, Balaiah D, Shah R. Menstrual irregularities and lactation failure may precede thyroid dysfunction or goitre. *J Postgrad Med*, 1993; **39**: 137–41.
47. Akande EO, Hockaday TD. Plasma luteinizing hormone levels in women with thyrotoxicosis. *J Endocrinol*, 1972; **53**: 173–4.
48. Doufas AG, Mastorakos G. The hypothalamic–pituitary–thyroid axis and the female reproductive system. *Ann N Y Acad Sci*, 2000; **900**: 65–76.
49. Fraietta R1, Zylberstejn DS, Esteves SC. Hypogonadotropic hypogonadism revisited. *Clinics (Sao Paulo)*, 2013; **68** Suppl 1: 81–8.
50. Higham CE, Johannsson G, Shalet SM. Hypopituitarism. *Lancet*, 2016; **388**: 2403–15.
51. Glezer A, Jallad RS, Machado MC, Fragoso MC, Bronstein MD. Pregnancy and pituitary adenomas. *Minerva Endocrinol*, 2016; **41**: 341–50.
52. Melmed S, Casanueva FF, Hoffman AR, *et al*. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
53. Kaltsas GA, Androulakis II, Tzieveriotis K, *et al*. Polycystic ovary phenotype in women with active acromegaly. *Clin Endocrinol (Oxf)*, 2007; **67**: 917–22.
54. Bétéa D, Valdes Socin H, Hansen I, Stevenaert A, Beckers A. Acromegaly and pregnancy. *Ann Endocrinol*, 2002; **63**: 457–63.
55. Katznelson L, Laws ER Jr, Melmed S, *et al*. Acromegaly clinical practice guidelines. *J Clin Endocrinol Metab*, 2014; **99**: 3933–51.
56. Björnsdottir S, Cnattingius S, Brandt L, *et al*. Addison's disease in women is a risk factor for an Adverse pregnancy outcome. *J Clin Endocrinol Metab*, 2010; **95**: 5249–57.

57. Ambrosi B, Barbetta L, Morricone L. Diagnosis and management of Addison's disease during pregnancy. *J Endocrinol Invest*, 2003; **26**: 698–702.
58. Gleicher N, Seier K, Kushnir VA, *et al.* Associations between peripheral androgens and cortisol in infertile women. *J Steroid Biochem Mol Biol*, 2016; **158**: 82–9.
59. Gleicher N, Kushnir VA, Darmon SK, *et al.* New PCOS-like phenotype in older infertile women of likely autoimmune adrenal etiology with high AMH but low androgens. *J Steroid Biochem Mol Biol*, 2017; **167**: 144–52.
60. Gleicher N, Kushnir VA, Weghofer A, Barad DH. The importance of adrenal hypoandrogenism in infertile women with low functional ovarian reserve: a case study of associated adrenal insufficiency. *Reprod Biol Endocrinol*, 2016; **14**: 23.
61. Speiser PW, Azziz R, Baskin LS, *et al.* Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2010; **95**: 4133–60.
62. Bachelot A, Grouthier V, Courtillot C, Dulon J, Touraine P. Management of endocrine disease. Congenital adrenal hyperplasia due to 21-hydroxylase deficiency: update on the management of adult patients and prenatal treatment. *Eur J Endocrinol*, 2017; **176**: r167–81.

8.8.2 Exogenous Factors and Female Reproductive Health

Nutrition and Reproduction

Siew Lim, Aya Mousa, Soulmaz Shorakae, and Lisa Moran

Introduction

Reproductive function is closely related to nutritional status. Undernutrition and overnutrition as reflected in body weight and body composition significantly impact fertility and reproductive outcomes. In addition, maternal micronutrient status in the periconception, pregnancy, and lactation period also influences the development of the offspring.

Undernutrition

The proportion of women of reproductive age that are underweight has decreased globally in the last few decades. Currently, the prevalence of maternal underweight is under 10% in all developed and developing regions except in South Asia and Africa.

Undernutrition and Ovulation

Undernutrition reduces fertility. A European multicentre study found that having BMI under 20 kg/m² is associated with delayed conception, defined as time to pregnancy exceeding 9.5 months of unprotected intercourse. Similarly, the Nurses' Health Study II found that 12% of ovulatory infertility in the US may be due to underweight (BMI<20 kg/m²) [1]. A 10–15% reduction in body weight from ideal body weight could result in amenorrhea. This may explain the higher prevalence of amenorrhea among anorexia nervosa patients, athletes, or dancers with very low body mass.

Undernutrition and Pregnancy Outcomes

In addition to reducing fecundity, undernutrition may also adversely affect pregnancy outcomes. A recent meta-analysis of 265 760 women reported that low prepregnancy BMI (<18.5) is associated with increased risk of miscarriage (relative risk 1.08, 95% CI 1.05–1.11) [2]. Maternal underweight is also associated with preterm births (BMI< 16: relative risk 1.54, 95% CI 1.40–1.68; BMI 17–18.49: relative risk 1.17, 95% CI 1.14–1.21) and small-for-gestational-age (SGA) birth (adjusted odds ratio 1.36, 95% CI 1.25–1.49). Achieving less than the recommended gestational weight gain further exacerbates these risks.

Undernutrition and Fertility Treatment Outcomes

A recent study in the US (n = 402 742 transfers) reported that underweight women had a significant decreased chance of intrauterine pregnancy (adjusted risk ratio [aRR] 0.97; 95% CI, confidence interval 0.96–0.99) and live birth (aRR 0.95; 95% CI, 0.93–0.98) per transfer [3]. As with naturally conceived pregnancies, maternal underweight is also associated with increased risk of low birthweight and preterm birth among cycles resulting in singleton pregnancies [3].

Overnutrition

More than half of all adults are overweight or obese in Western countries such as Australia and US. Overweight is defined as having BMI between 25 and 29.9 kg/m², while obese is defined as having BMI of 30 kg/m² and above. The prevalence of obesity has also increased dramatically in developing countries. The global epidemic of obesity is likely to be due to changes in dietary patterns towards greater consumption of energy-dense food and decreased physical activities associated with work, home, transport, or leisure. Obesity contributes significantly to the development of chronic diseases

such as diabetes, cardiovascular diseases, certain types of cancers (e.g. endometrial, ovary, cervix, and postmenopausal breast cancer) and reproductive disorders.

Obesity and Fertility

Overweight or obesity reduces fecundity in men and women. Twenty-five percent of ovulatory infertility in the US could be attributable to overweight or obesity (BMI $\geq$ 25 kg/m^2) [1]. Maternal overweight or obesity increases time to pregnancy regardless of age, cycle regularity, partners BMI, physical activity, smoking, parity, alcohol, and intercourse frequency. Body weight or body composition changes during the pubertal period may have an important impact on the development of the reproductive system. The onset of obesity during adolescence is associated with increased menstrual irregularities and ovulatory disorders. Weight gain or obesity in young adulthood also increases the risk of infertility in men and women. Obesity in men also increases the risk of infertility (OR 1.66, 95% CI 1.53–1.79) [4].

Obesity and Fertility Treatment Outcomes

Not only does obesity affect natural fertility, it may also compromise the success of assisted reproductive technology. Women who were overweight or obese had lower clinical pregnancy (RR 0.87, 95% CI 0.80–0.96), live birth (RR 0.80, 95% CI 0.71–0.90) and higher miscarriage (RR 1.43, 95% CI 1.22–1.67) following IVF treatment compared to normal weight women [5] Male obesity also contribute to decreased chances of live birth from assisted reproductive technologies (OR 0.65, 95% CI 0.44–0.97) [4].

Obesity and Pregnancy Outcomes

There is an increasing number of pregnant women who are obese. Obesity increases the risk of complications during pregnancy and delivery. Obese women are at higher risk of early miscarriage, pre-eclampsia, pregnancy-induced hypertension, gestational diabetes, caesarean delivery, and postpartum haemorrhage [6]. Higher maternal body weight also confers additional risks to the infants, including increased risk of being large for gestational age, having macrosomia, being stillborn, or being admitted to a neonatal intensive care unit [7]. These risks are further exacerbated by excessive gestational weight gain [8].

Polycystic Ovary Syndrome (PCOS)

PCOS is one of the most common endocrine conditions affecting up to 20% of women of reproductive age (see Chapter 8.4.1 for definition, diagnosis, and prevalence). Women with PCOS are more likely to be overweight (RR 1.95, 95% CI 1.52–2.50), obese (RR 2.77, 95% CI 1.88–4.10) or centrally obese (1.73, 95% CI 1.31–2.30) compared to women without PCOS [9].

Women with PCOS may be at greater risk of adverse pregnancy outcomes, including increased risk of gestational diabetes, pregnancy-induced hypertension, pre-eclampsia, preterm birth, and admission of infants to a neonatal intensive unit [10]. When fertility treatment is sought, women with PCOS may have a higher rate of cycle cancellation and longer duration of stimulation but the pregnancy and live birth rates are comparable to normal healthy women [11]. See online only Tables 8.8.2.1 and 8.8.2.2.

Box 8.8.2.1 The effect of insulin on androgen production

- Pituitary
  - Increases sensitivity to GnRH
- Ovaries
  - Increases the activity of 17,20 lyase
  - Increases the activity of 3β-hydroxysteroid dehydrogenase and aromatase in the granulosa cells
  - Increases LH receptors
  - Promotes ovarian growth and cyst formation
- Liver
  - Inhibits sex hormone binding globulin (SHBG) production

Obesity, Hyperinsulinaemia, Hyperandrogenaemia

Obesity is likely to promote hyperandrogenism and PCOS via hyperinsulinaemia, a compensatory response to insulin resistance. An increase in insulin level usually causes a concomitant increase in testosterone or androstenedione levels. The mechanisms underlying the relationship between hyperinsulinaemia and hyperandrogenism are summarized in Box 8.8.2.1 (Also see online only Table 8.8.2.3).

Treatment and Prognosis

Lifestyle Modification and Fertility

Weight loss through energy restriction, with or without exercise, improves reproductive function in overweight or obese women. In women with PCOS, about half of them had significant improvements in menstrual cyclicity or ovulation following weight loss through lifestyle modification. Among those who have lost at least 5% of their body weight, up to 80% experienced improvements in reproductive function. Improvements in reproductive hormones have been observed shortly after energy restriction (2 weeks) before weight loss occurs. This is consistent with the current understanding of reproductive function being influenced by energy balance instead of adiposity. In a Cochrane review, lifestyle modification reduced body weight, body mass index and free androgen index [12].

Diet and Fertility

A number of dietary factors have been associated with ovulatory infertility. In a prospective cohort study involving 18 555 pre-menopausal women, increasing consumption of vegetable as opposed to animal protein, choosing low glycaemic index carbohydrates and choosing unsaturated fatty acids (monounsaturated or polyunsaturated as opposed to transfats) were associated with reduced ovulatory infertility [13]. Research on dietary patterns found that greater adherence to a Mediterranean diet is associated with a higher likelihood of achieving pregnancy, clinical pregnancy, and live birth. Further research is warranted examining the effects of these dietary factors in intervention trials.

Evidence relating to supplements is more mixed. Consumption of iron supplements, other sources of non-heme iron, or multivitamins (particularly B vitamins and folic acid intake) is associated with decreased risk of ovulatory infertility in a prospective cohort study of 18 555 premenopausal women [14]. However, a Cochrane review of randomized controlled trials of antioxidant supplementation in women reported no benefits for increasing pregnancy or live birth rate. The high variability of interventions assessed makes it difficult to draw strong conclusions from the current evidence and further studies are needed.

Metformin and Fertility

Metformin has been used in women with PCOS, with effects on reducing the pulse amplitude of luteinizing hormone (LH) secretion, and on granulosa cell steroidogenesis and oocyte maturation. Metformin is also considered a safe option, with similar efficacy to insulin, for management of gestational diabetes during pregnancy with no increased risk of major birth defects [15]. It is given orally in doses of 1500 to 2000 mg in divided daily doses. A recent Cochrane review demonstrated that metformin use in women with PCOS compared to placebo may improve live birth rates, rates of clinical pregnancy, ovulation, and menstrual frequency although with increased gastrointestinal side effects. There was no clear evidence of an increase in miscarriage rates with metformin use. While metformin combined with clomiphene citrate had no clear benefit over clomiphene alone for live birth rates, it however resulted in higher rates of clinical pregnancy and ovulation. It is currently recommended that metformin could be used in women with PCOS at high metabolic risk including those with impaired glucose tolerance and in addition to clomiphene citrate for induction of ovulation in women with PCOS who are resistant to clomiphene citrate.

Bariatric Surgery and Fertility

Bariatric surgery is increasingly recognized as a treatment for morbid obesity (BMI >35 kg/m^2) in women of reproductive age who have failed other treatments including lifestyle modification. Current guidelines suggest that bariatric surgery be considered for patients with a BMI above 40 kg/m^2 or BMI above 35 kg/m^2 in the presence of obesity-related comorbidities, with PCOS considered to be one of those by some experts.

Bariatric surgery may improve menstrual cyclicity and ovulation through weight loss. A recent systematic review and meta-analysis reported 58% (22–92%) of infertile women become pregnant after bariatric interventions (weighted mean difference for successful pregnancy 0.580 (95%CI 0.539, 0.621)). A lower likelihood of gestational diabetes, hypertension and macrosomia are reported in pregnancies following bariatric surgery [16], however, it is notable that those pregnancies pose surgical, medical, and obstetric challenges including nutritional deficiencies, potential surgical complications, and difficulties with the diagnosis of gestational diabetes. It is generally recommended to avoid pregnancy within the first 12–24 months post bariatric surgery to allow for achievement of the full therapeutic benefit of the procedure and stabilization of the nutritional state and to reduce the risk of intrauterine growth retardation. To avoid nutritional deficiencies, it is advisable to optimize the nutritional status of women prior to and following bariatric surgery. Extra care must be taken to appropriately distinguish and treat other potential surgical and non-surgical complications of previous bariatric surgery during pregnancy.

Dietary Intakes and Pregnancy Outcome

Besides body weight and energy status, maternal dietary intakes before conception and during pregnancy can affect pregnancy outcomes. The following section summarizes the effect of various macronutrients and micronutrients on pregnancy outcomes.

Energy

During pregnancy, energy intakes are reported to range from 7710 to 9260 kJ/day, with relatively higher levels in the Americas and the Eastern Mediterranean than in Southeast Asia, the Western Pacific, and Africa. In a meta-analysis of three trials (n = 384), energy restriction during pregnancy in women with high BMI or with excess gestational weight gain resulted in reduced gestational weight gain [17]. However, it also reduced birthweight, with two of the three trials reporting adverse effects on birthweight. No improvements in pregnancy-induced hypertension or preeclampsia (gestational hypertension with proteinuria) were found. There is a lack of evidence to advise on energy restriction during pregnancy, however, prevention of excessive gestational weight gain in high-risk women (e.g. overweight and obese women, women at high risk of gestational diabetes) through diet and exercise is associated with reduced risk of infant macrosomia and infant respiratory distress syndrome.

Protein

In developed countries, pregnant women report consuming 14.7% to 16.1% of total energy from protein, which is adequate based on current recommendations. In a Cochrane review of 12 trials (n = 6705), balanced energy/protein supplementation (<25% energy from protein) increased birthweight and reduced the risk of stillbirth and SGA infants, but had no effect on gestational weight gain [18]. On the other hand, high protein supplementation (≥25% energy from protein) resulted in a significantly increased risk of SGA, with no effect on stillbirth, neonatal death, preterm birth, birthweight, or gestational weight gain (one trial, n = 505 African American women with a history of low birthweight infants) [18, 19]. Isocaloric protein supplementation (replacing other macronutrients with protein without changing the overall energy intake) had no effect on birthweight or gestational weight gain (two trials, n = 184) [18]. It is unclear whether high protein intake results in a subconscious reduction in overall energy intake, as is usually seen in high protein *ad libitum* weight loss trials. Current evidence supports balanced energy-protein supplementation (<25% energy from protein) to increase birth weight in those at risk of having SGA infants but further evidence is needed on the safety of high protein diet or supplement use during pregnancy.

Glycaemic Index, Glycaemic Load, and Fibre

The glycaemic index (GI) is a measure of the effect of dietary carbohydrate intake on blood glucose levels. High GI foods

cause a greater increase in blood glucose level within 2 hours of consumption. In a 2010 systematic review, two trials (n = 74) reported that low GI diets in healthy pregnancies reduced the risk of large-for-gestational-age (LGA) infants; however, an increased risk of SGA was reported in one epidemiological study of 1082 pregnancies [20]. In pregnancies complicated by gestational diabetes mellitus, benefits of a low GI diet have been consistently reported, including a reduced need for insulin to maintain optimal glucose management [20].

Prepregnancy dietary glycaemic load (GL) and fibre intake may also affect the risk of gestational diabetes as well as preeclampsia. In the Nurses Health Study II (n = 13 110), higher dietary GL was associated with greater risk of gestational diabetes. Conversely, a 10 g/day increment in total dietary fibre in prepregnancy was associated with a 26% reduced risk of developing gestational diabetes, with each 5 g/day increment in fibre from cereals or fruit reducing the risk by 23% or 26%, respectively [21]. Total dietary fibre has also been associated with improved lipid profiles and reduced risk of preeclampsia. Based on current evidence, low GI, low GL, and/or high fibre diets could be recommended preconceptionally to women at risk of gestational diabetes, pre-eclampsia, or having LGA infants.

Fatty Acids

Marine foods and fish oil supplements (mainly omega-3 supplements) are rich sources of *n*-3 long-chain polyunsaturated fatty acids such as docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA). DHA is proposed to influence the neural and visual development of infants, while EPA may reduce the synthesis of thromboxane A2 from arachidonic acid, which in turn may influence the risk of preeclampsia and timing to parturition. A recent review (54 trials, n >15 000) found that omega-3 supplementation during pregnancy reduced early preterm birth (<34 weeks' gestation) but increased the rate of prolonged pregnancies (>42 weeks' gestation) [22]. Other meta-analyses found no effect of omega-3 supplements on recurrent preterm birth or recurrent intrauterine growth restriction [23, 24]. In summary, evidence regarding the benefits of supplementation with *n*-3 fatty acids has been mixed, and further studies are needed before these supplements can be recommended for use in pregnancy to prevent hypertensive disorders or preterm birth, or to improve fetal outcomes.

Micronutrients

The requirements for most micronutrients rise during pregnancy. While some of these could be met by the increased absorption which occurs during pregnancy, certain populations may be at risk of developing micronutrient deficiency. Deficiency in one or more micronutrients could have long-term consequences for both the mother and the offspring.

Folate

Folate is present in various fruits and green leafy vegetables. Folate is essential for DNA and RNA synthesis, amino acid metabolism, and formate oxidation, and is considered an antioxidant as it removes oxidizing free radicals [25]. Folate is particularly important during phases of rapid cell division and tissue growth such as during embryonic and fetal periods. Circulating folate concentrations usually decrease during pregnancy in women not supplemented with folic acid (the synthetic form of folate). This is likely due to increased folate demand resulting from fetal and uteroplacental organ development, increased blood volume, increased folate catabolism or clearance, decreased folate absorption, or hormonal influences during pregnancy. Women from developing countries, particularly in areas with a high prevalence of malaria or sickle cell disease, are at high risk of folate inadequacy during pregnancy. Low levels of maternal folic acid can elevate serum homocysteine concentrations and increase the risk of fetal abnormalities.

Periconceptional supplementation with folic acid three months before and early in pregnancy can reduce the risk of neural tube defects by almost 75%. This protective effect is smaller in obese women compared to non-obese women. Evidence regarding whether folic acid prevents other fetal complications is less clear. In a 2015 Cochrane review (five trials, n = 7391), folic acid supplements prevented neural tube defects but had no effects on miscarriage, or other birth defects including cleft palate, cleft lip, or congenital heart defects [25].

The recommended daily folate requirement is 400 µg for women of reproductive age, 600 µg for pregnant women, and 500 µg for lactating women. From 12 weeks prior to conception until at least 12 weeks' gestation, all women should be supplemented with 400 µg/day of folic acid to reduce the risk of having a baby with a neural tube defect. A dose of 2.8 mg/week (or 5 mg intermittently) is or poor lifestyle habits including inadequate diet, smoking and alcohol consumption. To achieve optimal outcomes, women supplementing with folic acid should be advised to also maintain a folate-rich diet.

Vitamin A

Vitamin A is a fat-soluble vitamin present in carrots, sweet potatoes, eggs and dairy, and green leafy vegetables such as kale. The main physiological functions of vitamin A include vision, reproduction, growth, immunity, and cell proliferation. Deficiency is assessed by either a history of night blindness or serum or plasma retinol concentrations less than 0.70 µmol/L (subclinical vitamin A deficiency). Globally, the estimated prevalence of night blindness in pregnant women is 7.8% (9.8 million women), while 15.3% (19.1 million women) have deficient serum retinol concentrations. Maternal night blindness has been associated with low birthweight infants and infant mortality. However, a meta-analysis of 19 trials and over 310 000 women found that taking vitamin A supplements during pregnancy reduced risk of maternal anaemia, maternal infection, and maternal night blindness, but did not prevent maternal, perinatal, or newborn deaths, stillbirth, preterm birth, low birthweight, or anaemia in the newborn [26]. Excess retinol, the active form of vitamin A, could have teratogenic effects, hence the upper limit for vitamin A supplements is 10 000 IU per day (3000 µg retinol) and the non-toxic form (B-carotene) is preferred during pregnancy. Overall, due to the potential risk of toxicity, vitamin A supplementation should only be initiated after careful assessment of a woman's current intake. In most developed countries where hypovitaminosis A is rare, there is no need for supplementation.

Vitamin B_1 (Thiamine), Vitamin B_2 (Riboflavin), Vitamin B_3 (Niacin), Vitamin B_6 (Pyridoxine)

Vitamin B_1 (thiamine), vitamin B_2 (riboflavin), vitamin B_3 (niacin), and vitamin B_6 (pyridoxine) are involved in energy production and the metabolism of protein, fat, and carbohydrates. Greater energy and protein requirements during pregnancy increase the requirement for these vitamins. Thiamine deficiency could impair fetal brain development due to the role of thiamine-dependent enzymes in lipid and nucleotide synthesis in the brain. Deficiency in riboflavin has been associated with low birth weight, preeclampsia, and congenital heart defects; however, evidence regarding the use of riboflavin supplements to prevent these outcomes has been inconsistent. Periconceptional intakes of thiamine, niacin, and pyridoxine have been associated with lower risk of orofacial cleft, but causality has not been established due to the paucity of randomized trials testing these vitamins. In a systematic review of four trials (n = 1646), supplementation with pyridoxine did not prevent eclampsia or preeclampsia, and there was insufficient data to assess its effects on other maternal or neonatal outcomes. In summary, the effects of supplementing these vitamins during pregnancy remain unknown.

Vitamin B_{12} (Cyanocobalamin)

Vitamin B_{12} is vital for cellular growth, differentiation, and development [25]. Together with folate, vitamin B_{12} is responsible for the conversion of homocysteine to methionine, a process which is important for the methylation of DNA, proteins, neurotransmitters, and phospholipids. Because vitamin B_{12} is mainly derived from animal sources, those with restricted meat intake are at higher risk of deficiency. Global estimates suggest that vitamin B_{12} insufficiency is prevalent in 25% of pregnancies, ranging between 21%, 19%, and 29% in the first, second, and third trimesters, respectively. Women in the Indian subcontinent and the Eastern Mediterranean have high rates of vitamin B_{12} insufficiency.

Low vitamin B_{12} concentrations are associated with high plasma homocysteine levels, which have been implicated in various adverse pregnancy outcomes including placental abruption, stillbirths, low birthweight, and preterm deliveries. A meta-analysis of 18 longitudinal studies including individual patient data (11 216 observations) found that vitamin B_{12} deficiency (<148 pmol/L) was associated with an increased risk of low birthweight infants and preterm birth [27]. Similar to folate, low maternal vitamin B_{12} levels have also been associated with increased risk of neural tube defects and spina bifida. The effects of vitamin B_{12} supplementation during pregnancy are currently unknown and given the risks associated with B_{12} deficiency, further investigation by means of randomized controlled trials is warranted.

Vitamin C and E

Vitamin C (ascorbic acid) and vitamin E (tocopherols/tocotrienols) are important antioxidants which inhibit free radical formation, and vitamin C is additionally involved in collagen synthesis and iron and folate metabolism. Increased oxidative stress by free radicals has been implicated in the pathogenesis of preeclampsia and premature rupture of membranes (PROM), prompting interest into whether antioxidants such as vitamins C and E may protect against these conditions. In two separate Cochrane reviews, vitamin C (n = 29 trials, n = 24 300) and vitamin E (21 trials, n = 22 129), alone, or in combination with other supplements, were associated with a decreased risk of placental abruption, but an increased risk of self-reported abdominal pain [26, 27]. Vitamin C also resulted in a small increase in gestational age at birth and, when supplemented alone, reduced the risk of preterm and term PROM. Conversely, supplementation of vitamins C and E together, or vitamin E alone, increased the risk of term PROM [28, 29]. No other benefits or harms were reported for either vitamin. Overall, current evidence does not support the use of vitamins C or E for improving maternal or perinatal outcomes, and further research is needed to elucidate their roles in placental abruption and PROM.

Vitamin D

Vitamin D is a fat-soluble hormone with a well-established role in calcium metabolism and bone homeostasis. Vitamin D is synthesized by the skin upon exposure to sunlight and it can also be obtained from dietary sources such as oily fish or fortified dairy products, or from supplements in the form of ergocalciferol (vitamin D_2) or cholecalciferol (vitamin D_3). Vitamin D deficiency is determined by measuring its major circulating form, 25-hydroxyvitamin D (25(OH)D), where levels less than 75, 50, and 25 nmol/L are considered insufficient, deficient, and severely deficient, respectively. In developed countries, poor vitamin D status may result from low intake of fortified foods, highly pigmented skin, and minimal sunlight exposure due to long hours spent indoors and the use of sunscreen and protective clothing to prevent skin cancer. Global reports suggest that 40–98% of pregnant women are vitamin D-deficient, of which 15–84% are severely deficient.

Vitamin D deficiency is a known risk factor for rickets in infants and osteomalacia in adults. A meta-analysis of 15 trials (n = 2833) found that vitamin D supplements during pregnancy may reduce the risk of preeclampsia, low birthweight, and preterm birth. However, when vitamin D and calcium were combined, the risk of preterm birth was increased [30]. Limitations in the available studies and lack of reporting on adverse events were also noted. Hence, while preventing and treating vitamin D deficiency is important to maintain optimal maternal and fetal bone health, there is limited evidence to support the use of vitamin D supplements for other pregnancy outcomes.

Calcium

Calcium is an essential nutrient for various biological functions including skeletal development, muscle contraction, and enzyme and hormone homeostasis. Calcium absorption increases in pregnancy and a dietary intake of 1200 mg daily is recommended. Inadequate calcium intake during pregnancy can have adverse effects on both mother (osteopenia, paraesthesia, tremor, tetanus, muscle cramping) and fetus (delayed growth, low birthweight, poor fetal mineralization). However, trials evaluating the effect of calcium supplementation on maternal bone mineral density and fetal bone mineralization have been inconclusive.

Reduced calcium intake has also been implicated in the development of hypertension in pregnancy. A 2013 report by the World

Health Organization (WHO) combined two Cochrane reviews (21 trials; n >90 000) and found that calcium supplementation reduced the risk of preeclampsia by 52%, irrespective of baseline risk of hypertension or calcium intake status. The protective effect was greater in women at higher risk of hypertension and/or with lower calcium intakes (less than 900 mg daily). Modest reductions in preterm birth were also reported in women who received 1500–2000 g of elemental calcium daily compared to lower intakes; however, the quality of evidence was poor for this outcome [31]. Supplementation with 1.5–2.0 g/day of elemental calcium for prevention of preeclampsia has been recommended by the WHO for all pregnant women where dietary calcium intake is low, but particularly among those at higher risk of hypertension [31].

Iodine

Iodine is a non-metallic trace element essential for the synthesis of thyroid hormones. Dietary sources of iodine include kelp, seafood, and plants from iodine-rich soil, although the main source of iodine is via fortified salt. Nearly two-thirds of the populations in Western and Central Europe live in regions of mild to severe iodine deficiency, with those in mountainous and flooded areas being at increased risk. Iodine deficiency is the most common cause of preventable mental deficits globally, manifesting in a spectrum of iodine deficiency disorders (IDDs). These IDDs range from mild intellectual blunting to endemic cretinism (permanently and severely stunted physical and mental development). Other IDDs include maternal and fetal goitre and hypothyroidism, increased pregnancy loss, and infant mortality.

Given that the fetus depends on maternal iodine to synthesize its own thyroid hormones, even marginal iodine deficiency and hypothyroidism in pregnancy can influence the physical and mental development of the fetus. In a systematic review of 11 trials (n >2900), women receiving iodine supplements were 68% less likely to experience the adverse effect of postpartum hyperthyroidism and 15 times more likely to experience nausea and vomiting during pregnancy. A non-significant 34% reduction in perinatal mortality was also observed [32]. WHO recommends a daily intake of 250 μg during pregnancy and breastfeeding and periconceptional iodine supplementation should be considered for women at high risk of iodine deficiency. With the currently available evidence, the benefits and harms of iodine supplementation before, during, or after pregnancy remain unclear.

Iron

Meat, fish, legumes, green leafy vegetables, and iron-fortified cereals are the main dietary sources of iron. The most important role of iron is oxygen transport, although it is also involved in growth, reproduction, and healing. Pregnancy increases the risk of maternal anaemia, specifically iron-deficiency anaemia, due to an increased demand for maternal iron to support maternal and fetal needs. Global estimates suggest that 38.2% of pregnant women are anaemic, with the highest prevalence in South Asia (48.7%) and Africa (46.3%). Iron-deficiency anaemia contributes to 20% of post-birth maternal deaths in Asia and Africa.

Meeting the increased iron requirements during pregnancy is difficult from food sources alone, even with the increased iron absorption which occurs in the second and third trimesters. In a meta-analysis of 44 trials (n = 43 274), iron supplementation reduced the risk of maternal anaemia and iron deficiency in pregnancy. There was also some indication for improved birthweight and risk of preterm birth; however, the evidence for these outcomes was less clear [33]. As iron stores at conception are strong predictors of maternal iron status and risk of anaemia in later pregnancy, iron supplementation may have its greatest benefit if started at conception or in early pregnancy. Adequate iron status during pregnancy can also prevent postpartum anaemia, which is associated with postpartum depression. However, these benefits are weighed against a potential dose–response relationship with adverse events such as constipation, nausea, vomiting, and diarrhoea. Routine iron supplementation of pregnant women is recommended in some countries such as the US and France, but not in others such as the UK and Australia, which instead recommend screening pregnant women for anaemia and treating those with iron-deficiency anaemia (IDA). In countries where there is a high risk of IDA and screening is inadequate, daily supplementation of 60 mg during pregnancy is recommended by the WHO. Intermittent supplementation regimens should also be considered as they are as effective as daily supplementation but with fewer side effects.

Zinc

Zinc is present in meat and seafood. It is a cofactor in over 80 metalloenzymes with a ubiquitous role in biological functions including DNA transcription, protein synthesis, cellular division, wound healing, neurological function, immunity, folate utilization, and vision. Zinc also has antioxidant properties. It binds the sulphydryl groups in proteins and displaces iron or copper while preventing them from binding to lipids, proteins, and DNA. Zinc status is determined by measuring plasma or serum zinc levels, zinc-dependent enzyme levels, or 24-hour urinary zinc excretion. Cut-offs for defining zinc deficiency differ by age, sex, and time of day of sampling, making it difficult to establish global zinc deficiency status. Based on dietary data, however, it is estimated that 17% of the global population is at risk of zinc deficiency. Strict vegetarians and those with chronic diseases or infections are at increased risk. In two meta-analyses, zinc supplementation during pregnancy reduced the incidence of preterm birth by 14% but had no effects on neonatal mortality, birthweight, or preeclampsia [34, 35]. The effect of zinc on preterm birth was observed primarily in women from low-income settings, and thus may have been amplified by the poor overall nutritional status of these women. Altogether, the evidence shows limited benefits of zinc supplementation during pregnancy. Improving overall nutrition, particularly in women from low-income settings, is likely a more prudent approach for improving pregnancy outcomes than with zinc supplements alone.

Alcohol and Caffeine

Globally, it is estimated that 5–10% of pregnancies are at risk of alcohol-related birth defects. In addition to the known impact of alcohol abuse on fetal alcohol syndrome, regular alcohol intake (at least once a week) or high alcohol consumption (more than 14 units a week or more than 3 units a day) increase the risk of

first-trimester miscarriage and SGA infants. The effect of moderate or occasional alcohol consumption on pregnancy outcomes is less clear. Two systematic reviews reported that prenatal alcohol exposure was associated with poorer child cognition, behaviour, and mental development, irrespective of the level of alcohol consumed [36, 37]. In contrast, two other systematic reviews concluded that there was no convincing evidence of adverse effects from low to moderate prenatal alcohol exposure. Inconsistencies and weaknesses in the evidence, including lack of adjustment for sociodemographic confounders, preclude the ability to determine safe levels of drinking during pregnancy. As such, it is advisable that women avoid alcohol intake during pregnancy.

Evidence regarding the effects of caffeine on pregnancy outcomes has been inconsistent. Maternal caffeine intake appears to have no effect on cardiovascular malformations, oral clefts, or early miscarriage. Some prospective cohort studies have reported that high caffeine intake (>200 mg/day) increases the risk of miscarriage after adjusting for several confounders, while others found no relationship between caffeine consumption and miscarriage, growth restriction, low birthweight, or preterm delivery. Current evidence is insufficient to confirm or refute the benefits of caffeine avoidance in pregnancy. Until safety limits are established, women with high daily caffeine intake (>200 mg) should be advised to minimize caffeine intake during pregnancy.

Nutrition and Reproduction in Men

Compared to the literature in women, there is relatively limited information on nutrition and reproduction in men. Obesity is a risk factor for infertility in males, as it is in females. A dose–response relationship between BMI and infertility in men has been reported, and a high BMI has been associated with lower sperm count, density, and motility; reduced testosterone and testosterone/oestrogen ratio; and a greater number of sperm cells with chromatic damage measured using DNA fragmentation index. Paternal obesity may also

Table 8.8.2.4 Dietary recommendations for optimal reproductive outcomes

Energy	• There is lack of evidence to advise on energy restriction during pregnancy. Prevention of excessive gestational weight gain in high-risk women (e.g. overweight and obese women, women at high risk of gestational diabetes) is associated with reduced risk of infant macrosomia and infant respiratory distress syndrome.
Protein	• Balanced energy/protein intake (<25% energy from protein) may increase birthweight and reduce the risk of SGA infants and stillbirth, particularly in women at high risk of nutritional inadequacy • Isocaloric or high protein intake (>25% energy from protein) may increase the risk of SGA infants.
Glycaemic index (GI), glycaemic load (GL) and dietary fibre	• Low GI diets during pregnancy can reduce the risk of LGA infants and improve glycaemic control in gestational diabetes • Low GL and high fibre diets before pregnancy may prevent the development of gestational diabetes, and high fibre diets may also reduce the risk of preeclampsia
N-3 polyunsaturated fatty acids	• Insufficient evidence to support the use of fatty acid supplements to improve maternal or fetal outcomes.
Folate	• Supplementation is recommended for women from 12 weeks prior to conception until at least 12 weeks' gestation during preconception (400 µg) to prevent neural tube defects • Women supplementing with folic acid should be advised to also maintain a folate-rich diet.
Vitamin A	• Supplementation, preferably with β-carotene, should only be initiated in women with deficiency after careful assessment of current vitamin A intake. Excess retinol intakes may have teratogenic effects.
B vitamins (B_1, B_2, B_3, B_6, and B_{12})	• Deficiency may increase the risk of adverse pregnancy outcomes, but effects of supplementation are unknown.
Vitamin C and E	• Evidence for supplementation in pregnancy is inconclusive due to reports of both benefits and harms in relation to placental abruption and premature rupture of membranes.
Vitamin D	• Supplementation in women with inadequate vitamin D status may help maintain optimal maternal bone health and prevent poor bone mineralization in the fetus.
Calcium	• Supplementation of 1.5–2.0 g elemental calcium/ day is recommended by the WHO in women with low calcium intakes (<900 mg/day) and/or women at high risk of gestational hypertension to prevent preeclampsia and promote fetal bone mineralization.
Iodine	• The WHO recommends a daily intake of 250 µg during pregnancy and breastfeeding and periconceptional iodine supplementation should only be considered for women at high risk of iodine deficiency • The current evidence on the benefits and harms of iodine supplementation before, during, or after pregnancy is unclear.
Iron	• Pregnant women should be screened and treated for iron-deficiency anaemia • Daily supplementation of 30–60 mg from early pregnancy is recommended where screening is inadequate.
Zinc	• Some evidence of benefit in preventing preterm birth in low-income settings; however, the evidence is insufficient to support routine supplementation.
Alcohol and Caffeine	• Safety of low to moderate intake cannot be established from current evidence. Women are advised to limit or avoid alcohol and minimize caffeine consumption during pregnancy.

Abbreviations: GI/GL, glycaemic index/load; SGA/LGA, small/large for gestational age.

have adverse effects on the reproductive health of the offspring, leading to further amplification of subfertility.

Despite limited studies in this field, recent systematic reviews have revealed a number of trends linking nutrition with male reproductive health. First, higher sperm quality has been associated with diets which are low in saturated and transfatty acids and high in omega-3 fatty acids, folate, zinc, β-carotene, and vitamins D, E, and C. Second, healthy dietary patterns characterized by higher intakes of seafood, poultry, whole grains, fruits and vegetables, and low-fat dairy have been associated with higher sperm quality, whereas diets rich in processed meat, soy, potatoes, full-fat dairy, coffee, alcohol, sweets, and sugary drinks have shown the opposite relationship. Third, a Cochrane review of 34 trials (n = 2786 couples) reported that semen quality and clinical pregnancy rates were improved by antioxidant supplementation [36]. However, heterogeneity in study populations and regimens, and the expansive definition of antioxidants makes it difficult to ascertain which specific antioxidants, and at which doses, were responsible for these effects. Last, moderate intake of caffeine and alcohol do not appear to have a meaningful impact on semen quality or fecundability. However, further well-designed studies are needed to confirm these effects and to establish which patterns, level, and duration of alcohol or caffeine consumption may affect male reproductive health. See Table 8.8.2.4 and [39].

REFERENCES

1. Rich-Edwards JW, Spiegelman D, Garland M, *et al.* Physical activity, body mass index, and ovulatory disorder infertility. *Epidemiology*, 2002; **13**: 184–90.
2. Balsells M, Garcia-Patterson A, Corcoy R. Systematic review and meta-analysis on the association of prepregnancy underweight and miscarriage. *Eur J Obstet Gynecol Reprod Biol*, 2016; **207**: 73–9.
3. Kawwass JF, Kulkarni AD, Hipp HS, Crawford S, Kissin DM, Jamieson DJ. Extremities of body mass index and their association with pregnancy outcomes in women undergoing *in vitro* fertilization in the United States. *Fertil Steril*, 2016; **106**: 1742–50.
4. Campbell JM, Lane M, Owens JA, Bakos HW. Paternal obesity negatively affects male fertility and assisted reproduction outcomes: a systematic review and meta-analysis. *Reprod Biomed Online*, 2015; **31**: 593–604.
5. Rittenberg V, Seshadri S, Sunkara SK, Sobaleva S, Oteng-Ntim E, El-Toukhy T. Effect of body mass index on IVF treatment outcome: an updated systematic review and meta-analysis. *Reprod Biomed Online*, 2011; **23**: 421–39.
6. Rahman MM, Abe SK, Kanda M, *et al.* Maternal body mass index and risk of birth and maternal health outcomes in low- and middle-income countries: a systematic review and meta-analysis. *Obes Rev*, 2015; **16**: 758–70.
7. Liu P, Xu L, Wang Y, *et al.* Association between perinatal outcomes and maternal pre-pregnancy body mass index. *Obes Rev*, 2016; **17**: 1091–102.
8. Goldstein RF, Abell SK, Ranasinha S, *et al.* Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. *JAMA*, 2017; **317**: 2207–25.
9. Lim SS, Davies MJ, Norman RJ, Moran LJ. Overweight, obesity and central obesity in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update*, 2012; **18**: 618–37.
10. Yu HF, Chen HS, Rao DP, Gong J. Association between polycystic ovary syndrome and the risk of pregnancy complications: a PRISMA-compliant systematic review and meta-analysis. *Medicine (Baltimore)*, 2016; **95**: e4863.
11. Heijnen EM, Eijkemans MJ, Hughes EG, Laven JS, Macklon NS, Fauser BC. A meta-analysis of outcomes of conventional IVF in women with polycystic ovary syndrome. *Hum Reprod Update*, 2006; **12**: 13–21.
12. Lim SS, Hutchison SK, Van Ryswyk E, Norman RJ, Teede HJ, Moran LJ. Lifestyle changes in women with polycystic ovary syndrome. Cochrane Database Syst Rev, 2019, 3: CD007506.
13. Chavarro JE, Rich-Edwards JW, Rosner BA, Willett WC. A prospective study of dietary carbohydrate quantity and quality in relation to risk of ovulatory infertility. *Eur J Clin Nutr*, 2009; **63**: 78–86.
14. Chavarro JE, Rich-Edwards JW, Rosner BA, Willett WC. Use of multivitamins, intake of B vitamins, and risk of ovulatory infertility. *Fertil Steril*, 2008; **89**: 668–76.
15. Sinai Talaulikar V, Tang T, Yasmin E. Role of metformin in women's health: review of its current place in clinical practice and emerging indications for future. *Obstet Gynecol Surv*, 2016; **71**: 307–17.
16. Yi Xy, Li QF, Zhang J, Wang ZH. A meta-analysis of maternal and fetal outcomes of pregnancy after bariatric surgery. *Int J Gynaecol Obstet*, 2015; **130**: 3–9.
17. Kramer MS, Kakuma R. Energy and protein intake in pregnancy. *Cochrane Database Syst Rev*, 2003: **4**: CD000032.
18. Ota E, Hori H, Mori R, Tobe-Gai R, Farrar D. Antenatal dietary education and supplementation to increase energy and protein intake. *Cochrane Database Syst Rev*, 2012; **9**: CD000032.
19. Rush D, Stein Z, Susser M. A randomized controlled trial of prenatal nutritional supplementation in New York City. *Pediatrics*, 1980; **65**: 683–97.
20. Louie JCY, Brand-Miller JC, Markovic TP, Ross GP, Moses RG. Glycemic index and pregnancy: a systematic literature review. *J Nutr Metab*, 2010; **2010**: 282464.
21. Zhang C, Liu S, Solomon CG, Hu FB. Dietary fiber intake, dietary glycemic load, and the risk for gestational diabetes mellitus. *Diabetes Care*, 2006; **29**: 2223–30.
22. Middleton P, Gould J, Shepherd E, Makrides M. Omega-3 supplementation during pregnancy: an updated Cochrane review. *J Paediatr Child Health*, 2017; **53**: 68–9.
23. Saccone G, Berghella V. Omega-3 supplementation to prevent recurrent preterm birth: A systematic review and metaanalysis of randomized controlled trials. *Am J Obs Gynecol,* 2015; 213: 135–140.
24. Saccone G, Berghella V, Maruotti GM, Sarno L, Martinelli P. Omega-3 supplementation during pregnancy to prevent recurrent intrauterine growth restriction: Systematic review and meta-analysis of randomized controlled trials. *Ultrasound Obstet Gynecol,* 2015; 46: 659–64.
25. De-Regil LM, Peña-Rosas JP, Fernández-Gaxiola AC, Rayco-Solon P. Effects and safety of periconceptional oral folate supplementation for preventing birth defects. *Cochrane Database Syst Rev*, 2015; **12**: CD007950.
26. McCauley ME, van den Broek N, Dou L, Othman M. Vitamin A supplementation during pregnancy for maternal and newborn outcomes. *Cochrane Database Syst Rev*, 2015; **10**: CD008666.
27. Rogne T, Tielemans MJ, Chong MF-F, *et al.* Maternal vitamin B12 in pregnancy and risk of preterm birth and low birth weight: a systematic review and individual participant data meta-analysis. *Am J Epidemiol*, 2017; **185**: 212–23.
28. Rumbold A, Ota E, Hori H, Miyazaki C, Crowther CA. Vitamin E supplementation in pregnancy. *Cochrane Database Syst Rev*, 2015; **2**: CD004069.

29. Rumbold A, Ota E, Nagata C, Shahrook S, Crowther CA. Vitamin C supplementation in pregnancy. *Cochrane Database Syst Rev*, 2015; **9**: CD004072.
30. De-Regil LM, Palacios C, Ansary A, Kulier R, Peña-Rosas JP. Vitamin D supplementation for women during pregnancy. *Cochrane Database Syst Rev*, 2012; **2**: CD008873.
31. WHO. Guideline: *Calcium Supplementation in Pregnant Women*. Geneva: World Health Organization, 2013.
32. Harding KB, Peña-Rosas JP, Webster AC, *et al*. Iodine supplementation for women during the preconception, pregnancy and postpartum period. *Cochrane Database Syst Rev*, 2017; **3**: CD011761.
33. Peña-Rosas JP, De-Regil LM, Garcia-Casal MN, Dowswell T. Daily oral iron supplementation during pregnancy. Cochrane Database Syst Rev, 2015; **7**: CD004736.
34. Chaffee BW, King JC. Effect of zinc supplementation on pregnancy and infant outcomes: a systematic review. *Paediatr Perinat Epidemiol*, 2012; **26** Suppl 1: 118–37.
35. Ota E, Mori R, Middleton P, *et al*. Zinc supplementation for improving pregnancy and infant outcome. *Cochrane Database Syst Rev*, 2015; **7**: CD000230.
36. Flak AL, Su S, Bertrand J, Denny CH, Kesmodel US, Cogswell ME. The association of mild, moderate, and binge prenatal alcohol exposure and child neuropsychological outcomes: a meta-analysis. *Alcohol Clin Exp Res*, 2014; **38**: 214–26.
37. Testa M, Quigley BM, Das Eiden R. The effects of prenatal alcohol exposure on infant mental development: a meta-analytical review. *Alcohol and Alcoholism*, 2003; **38**: 295–304.
38. Showell MG, Brown J, Yazdani A, Stankiewicz MT, Hart RJ. Antioxidants for male subfertility. *Cochrane Database Syst Rev*, 2011; **1**: CD007411.
39. Metwally M, Ong KJ, Ledger WL, Li TC. Does high body mass index increase the risk of miscarriage after spontaneous and assisted conception? A meta-analysis of the evidence. *Fertil Steril*, 2008; **90**: 714–26.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

8.8.3 Exogenous Factors and Female Reproductive Health

Environment and Reproduction

Evanthia Diamanti-Kandarakis and Eleni A. Kandaraki

Overview of Female Reproduction and Environment

Female reproductive system is one of the most delicate systems of the human body. Under the orchestrating signals of hypothalamus, each component of the activated axis of hypothalamus–pituitary–ovary has to cointeract harmonically, in order to create the optimal hormonal milieu for its proper function.

Increased global industrial activity has exposed humans to a wide variety of modern chemical substances. In fact, we live in a world in which chemicals are an integral part of our everyday life. These exogenous insults can perturb hormonal balance of female body and threaten reproductive health and ultimately reproductive capacity.

Endocrine-Disrupting Chemicals (EDCs)

Definition and Mechanism of Action of EDCs

EDCs or endocrine disruptors are 'exogenous chemicals, or mixture of chemicals that interfere with any aspect of hormone action, leading to adverse effects in an intact organism or its progeny or (sub)populations'. Humans are continuously exposed to EDCs in daily life via multiple ways of contact [1].

Currently, nearly 800 chemicals are known or suspected to play a negative role in hormonal disruption, and this list is rapidly growing. EDCs are a heterogeneous group of molecules with the high potential to interact directly with hormone receptors (e.g. oestrogen, androgen, thyroid) as agonists or antagonists, triggering different molecular pathways. Some EDCs can also interfere with proteins involved in the transport of hormones, such as sex hormone–binding globulin (SHBG), and thus disrupt the delivery of endogenous hormones to target cells. Their molecular structure usually consists of a phenol group and may possess halogen group substitution by chlorine and bromine. Therefore, they can easily mimic steroid hormones and 'interfere with the synthesis, secretion, transport, metabolism, binding action, or elimination of natural blood-borne hormones that are present in the body and are responsible for homeostasis, reproduction and developmental process' (definition of the US Environmental Protection Agency, EPA). These effects, depending on exposure dose and timing, give rise to alternative phenotypes, which may lead to increased disease susceptibility (Table 8.8.3.1).

The molecular structure and lipophilic properties of some EDCs may promote their bioaccumulation in the adipose tissue of animals and humans. EDCs have been detected also in biological fluids such as sera, urine, amniotic fluid, and breast milk. Some EDCs persist in the body for a long time, such as persistent organic pollutants (POPs), and some (e.g. pesticides) are quickly metabolized and cleared from the organism, although chronic exposure can lead to bioaccumulation [1].

The affinity of the hormone receptors for the EDCs is usually much lower than for natural ligands. For example, the affinity of the classical oestrogen receptors (ER), ER-a, and ER-b, for one of the most commonly used EDCs, bisphenol A (BPA) is 1000–10 000-fold lower than for 17b-estradiol (E2). However, EDCs can act in several tissues, even in low doses, also below the level

Table 8.8.3.1 Classification and physiological effects of common EDCs

Substance categories	Name	Source/Uses	Effects
Bisphenols	Bisphenol A, Bisphenol F	Polycarbonate plastics, epoxy resins, plastic toys, and bottles, lining of food cans	Oestrogenic, obesogenic, Neurological effects, adverse thyroid hormone action, reproductive and developmental effects
Non-steroidal synthetic oestrogen	Diethylstilboestrol (DES)	Pharmaceutical	Transplacental carcinogen, teratogen
Plasticizers	Phthalates (DEHP, BBP, DBP, DPP, DPrP)	Contaminated food, PVC plastics and flooring, personal care products, medical devices and tubing	Carcinogen, liver damage, reproductive and developmental effects, asthma, obesogen
Organochloride	Polychlorinated biphenyl (PCBs)	Contaminated air and food, skin contact with old electrical equipment	Carcinogen, chloracne, stomach and liver damage, reproductive and nervous system effects and thyroid injury
Dioxins and dibenzofurans	2,3,7,8-TCDD, 2,3,7,8-p-TCDD, 2,3,7,8-PCDF, 2,3,4,7,8-PeCDF, 1,2,3,7,8 (9)-PeCDF	Arise as unwanted by-products from certain incineration and industrial chemical process	Antioestrogenic properties
Phyto-oestrogens	Isoflavones, coumestrol, Lignans, and stilbens	Synthesized in plants	Oestrogenic properties

of exposure at which no adverse effects in organism have been observed by toxicologists.

Timing of EDC exposure during lifespan is another crucial parameter for the severity of their biological effects. Firstly, embryonic development, organogenesis, and tissue differentiation *in utero* is a critical developmental window, during which any environmental perturbation can lead to structural and functional changes that can profoundly deflect the developmental trajectory, often leading to lifelong phenotypic changes. Childhood and puberty are also considered periods of rapid change in an endocrine-dependent manner and are recognized as sensitive periods. In other words, the younger organisms are more susceptible to EDCs' action.

Timing of exposure to EDCs is highlighted as such an important factor, due to the unique ability of EDCs to cause epigenetic changes. Epigenetics are defined as changes in gene expression that occur without changes in DNA sequence. Several possible mechanisms of epigenetics exist, including methylation of cytosine residues on DNA, post-translational modification of histones, and altered microRNA expression. These changes alter gene expression and lead to transient or persistent changes in gene transcription [2].

Epigenetic changes can be transmitted through mitosis and/or meiosis. Thus, an altered epigenome, after EDC exposure, will transfer to the subsequent generations by germline in the process called 'epigenetic transgenerational inheritance' [3]. When the F_o generation female is exposed to EDCs during her pregnancy, the fetus and the germline of the fetus of F_1 generation including the cells in the gonads of the F_1 fetus, which will become F_2 generation, are directly exposed to these chemicals too [3, 4].

This transgenerational component of EDC's effects has become really obvious in 'the diethylstilboestrol catastrophe'. Diethylstilboestrol (DES), a potent synthetic non-steroidal oestrogen, is one of the first studied endocrine disruptors. It was prescribed to pregnant women from the 1940s to 1975 to prevent miscarriage and other complications. Its use was discontinued when a subset of exposed daughters presented with early-onset vaginal clear-cell adenocarcinoma, with a 40-fold increase in risk compared to unexposed individuals. The prevalence of neoplasia caused by DES in exposed population is estimated to be 0.1%, but this is much lower than the prevalence of subfertility and infertility occurred in DES intrauterine exposed population which is estimated to be >90%, occurring in both sons and daughters. What's more, new data suggest that increased disease risk persists even in their grandchildren a potential epigenetic effect [1, 5, 6].

There has been extensive evidence from experimental studies showing that exposure to endocrine disruptors has a negative impact on ovarian physiology leading to morphological and functional alterations of the female reproductive system [7, 8]. In fact, environmental pollutants are shown to interfere with several aspects of follicle growth and ovarian steroid hormone production by modifying the expression and/or activity of regulatory enzymes or by altering hormone receptor binding and action. Furthermore, early exposure to EDCs has also been experimentally linked to reproductive tract malformations, earlier pubertal timing, alterations of the hypothalamic–pituitary–ovarian axis, disorders of ovulation, fertility, and fecundity [7].

EDCs Affect Early Gonadal Development

Early gonadal development begins with the primordial germ cells originating in the proximal epiblast, far from the testes or ovaries. At this stage of development, the gonadal ridges are still bipotential, and they remain so until gestational week 7 in humans [1]. The expression of the Y chromosome-specific gene (*Sry)* in XY gonadal ridges triggers the differentiation of the gonadal ridges into testes, whereas, in the absence of a Y chromosome, the XX gonadal ridges differentiate into ovaries. The differentiation of testes and ovaries is chiefly under genetic control and involves several morphogenetic pathways.

Recent data suggest that these early stages of folliculogenesis are vulnerable to exogenous perturbations [7]. For example, intrauterine exposure to the analgesic and antipyretic drug, acetaminophen, reduced the number of germ cells in female mouse fetuses, due to disrupted proliferation of the primordial germ cells.

Furthermore, continuous exposure of female mice to paracetamol until birth resulted in a reduction in ovarian follicle reserves and total number of follicles, which led to subfertility.

The industrial chemical BPA, which is widely reported to have oestrogenic and other endocrine-disrupting properties, also interferes with gonadal development. In mouse embryonic stem cells, exposure to BPA results in the downregulation of the testis-promoting genes *Sox9* and *Fgf9* and in the upregulation of the ovary-promoting genes *Wnt4* and *Foxl2* during differentiation, suggesting that exposure to BPA favours ovary differentiation4,. Acting in a similar way, tamoxifen, an oestrogen receptor (ER)-modulator used in breast cancer therapy, promotes ectopic upregulation of Sox9 and downregulation of Foxl2 in the XX gonads. Ultimately, disrupting the balance between these opposing factors can lead to sex-reversal phenotypes and cause reprogramming of the cell types in adulthood, which, in turn, will adversely affect the germ cell population, future oocyte reservoir, and reproductive lifespan, leading to infertility. Notably, *FOXL2* mutations in humans cause premature ovarian insufficiency [9, 10].

Another process of early gonadal development that is susceptible to exogenous effects is meiotic division. In females, meiosis begins during fetal life and arrests in prophase I until puberty. Disruption of germ cells at this stage can have long-lasting consequences for reproductive health parameters due to reduction in the number of oocytes. For example, gestational exposure to BPA [7], through its estrogenic effects, can disrupt meiotic division in mice and monkeys, via altered the expression of genes that control meiosis (*Stra8, Dazl, Nobox*). Furthermore, *in vitro* studies indicate that BPA impaired meiotic progression in cultured human fetal oocytes, increased levels of recombination (MLH1 foci), and induced epigenetic changes that may contribute to chromosome congression failure. Analogous data exist in literature for endocrine disruptor di(2-ethylhexyl) phthalate (DEHP) and paracetamol [1].

EDCs Affect Follicular Assembly and Folliculogenesis

Follicle assembly is crucial for establishing the primordial follicle pool. The assembly or formation of the primordial follicles requires individual oocytes to segregate and associate with squamous (i.e. precursor) granulosa cells. Nests of associated oocytes undergo random apoptosis of individual oocytes to derive isolated oocytes that then associate with precursor squamous granulosa cells [1]. In humans, the timing of primordial follicle assembly is very different from that of rodents, with follicle assembly occurring during mid-gestation stage. The mechanisms that control follicular assembly remain unclear. Evidence suggests that, in mouse fetuses, high levels of maternal oestradiol and progesterone inhibit follicle assembly and that a reduction in the levels of steroid hormone around the time of birth initiates the process. Furthermore, the human fetal ovary expresses steroidogenesis enzymes and steroid hormone receptors, particularly ERβ, during the second trimester, which enables oestrogenic, progestogenic, and androgenic signalling to occur.

These data indicate that oestrogen has an important role in follicle assembly but also highlights the potential susceptibility to disruption by oestrogenic compounds. Thus, several chemicals including BPA and DEHP can interfere with this process, possibly by disrupting apoptotic-signalling pathways. For example, oral exposure of pregnant mice to BPA results in an increased percentage of oocytes in germ cell nests, an increase in the total number of oocytes per ovary section and a decrease in the number of primordial follicles in offspring. Similarly, subcutaneous exposure of neonatal mice to the phyto-oestrogen, genistein, inhibits the breakdown of the germ cell nest and disrupts primordial follicle assembly. Taken together, these studies support the view that many EDCs can have capacity to disrupt follicle assembly, potentially affecting both the quantity and quality of follicles later in life [4, 5, 9].

After follicle assembly and the formation of primordial follicle pool, the first wave of follicular recruitment and folliculogenesis is initiated. Following exposure to BPA from 1 to 14 days postpartum, 30-day-old lambs had reduced ovary weight, fewer primordial follicles, more transitional and primary follicles, but a stable total number of oocytes compared with lambs that were not exposed to BPA. In proportion a decrease in the number of primordial follicles, concomitant with an increase in the number of secondary and antral follicles, has been observed in 15-day-old mice after exposure to DEHP from 7 to 14 days postpartum. Thus, EDCs can accelerate the rate of follicular recruitment, leading to adverse consequences later in life, such as a shorter reproductive lifespan due to a reduced number of primordial follicles [4].

EDCs Affect Steroidogenesis

EDCs also have the potential to interfere with ovarian steroidogenesis. In rat theca-interstitial and granulosa cell cultures, BPA significantly increases testosterone and progesterone levels by increasing the expression of several key cytochrome p450 steroidogenic enzymes, such as 17-a hydroxylase and cholesterol side chain cleavage enzyme, respectively. There is a hypothesis that BPA stimulates ovarian theca-interstitial cells to elevated androgens synthesis and also inhibits testosterone catabolism, thereby leading to elevated androgen levels. In fact, human studies have confirmed this positive correlation between serum BPA level and elevated androgen levels, such as total and free testosterone, androstenedione, and dehydroepiandrosterone (DHEA), in women with polycystic ovary syndrome (PCOS) [10].

Similar data exist for various EDCs. For example, exposure to the organochlorine pesticides were associated with a slower drop in the oestradiol ratio and progesterone metabolites after ovulation in women, with decreased progesterone levels and a shorter luteal phase in women, findings that were in accordance with inhibited production of oestradiol, testosterone, androstenedione, and progesterone in isolated mouse antral follicles [6].

EDCs Metabolic Disruption Combined with Female Reproductive Deregulation

After the initial observation that obesity affects unfavourably fertility and fecundability, it is now widely accepted that the metabolic profile of a woman stigmatizes her reproductive function, starting from *in utero* fetal development and offspring programming. In addition to that, there is considerable experimental evidence linking EDCs with metabolic alterations through deregulation of normal lipid and glucose homeostasis. Importantly, the world's modern epidemics—obesity and diabetes—are believed to be partly related

to the increasing presence of such substances including the intra and extrauterine environment of developing organisms. The so-called **obesogens** are environmental chemicals presumed to favour obesity development through inappropriate regulation of energy and lipid metabolism. Many of their effects are mediated through the peroxisome proliferator-activated receptor-γ (PPAR-γ) who acts as a regulator of adipocyte differentiation [1].

Interestingly, an environmental chemical may alter concomitantly reproductive as well as metabolic pathways. BPA is a characteristic example of such interaction as it has been incriminated for multiple reproductive abnormalities as well as metabolic ones. Remarkably, epidemiological evidence link BPA detected in urine with cardiovascular disease, type 2 diabetes, and liver enzyme abnormalities in a representative sample of adult US population, while data are suggestive that BPA is also elevated in obese women with fertility problems. Animal models of prenatal and perinatal BPA exposure demonstrate a disruption of weight control mechanisms leading to increased adipose storage in early postnatal and adult life in addition to impair follicular maturation process, ovarian steroidogenesis, oestrous cyclicity, ovulation, and ultimately fertility [4, 9].

EDCs Exposure and Reproductive Health

Taking all these aforementioned together, experimental data suggest that EDCs have the potency/potential to disrupt the reproductive function, targeting at multiple stages directly and indirectly at the level of the ovary, affecting follicle maturation and ovarian steroid hormone production [4]. Scientific community has introduced the term '**ovarian dysgenesis syndrome**' to describe schematically this disruption. Ovarian dysgenesis syndrome is defined as early alterations in ovarian structure or function that cause an impairment of reproductive parameters in adulthood. This compromised reproductive development or function constitutes an underlying pathophysiological mechanism for a group of related female reproductive diseases and disorders, such as reduced fertility, PCOS, premature ovarian insufficiency, and reproductive-site cancers (Figure 8.8.3.1).

Firstly, there is no doubt that a significant decrease in total fertility rates, measured as live births per woman, has been observed during the past few years in modern, industrialized countries all over the world [8]. This decrement has a multifactorial aetiology and can be partly explained by changing attitudes towards family planning, as well as by socioeconomic changes that took place in contemporary societies. However, we cannot argue the fact that our chemically burdened environment has contributed to the increase in women reporting difficulties with conceiving and maintaining a pregnancy [6].

Infertility: Although any causal relationship between EDC exposure and infertility has not been established yet, indirect evidence supporting this interaction are ample. The widespread medical use of assisted reproductive techniques in the modern era has contributed to this direction. Various studies of women undergoing *in vitro* fertilization (IVF) have documented the close association between EDC exposure levels and increased odds of implantation failure, higher risk of miscarriage, poor embryo quality, and decreased pregnancy and live birth rates. Furthermore, even when pregnancy is achieved, preterm birth, decreased gestational age at birth, decreased birth weight, and decreased body length can be observed in EDC exposed women [6, 11].

Polycystic ovary syndrome: PCOS is a multifactorial endocrinopathy of unknown aetiology affecting approximately 5–10% of women in the reproductive age spectrum. Cardinal features of the syndrome are reproductive and metabolic aberrations central to which is abnormal gonadotropin secretion, disrupted ovarian follicular maturation, and steroid hormones secretion as well as disturbed insulin action leading to insulin resistance and compensatory hyperinsulinemia. The role of the environment in

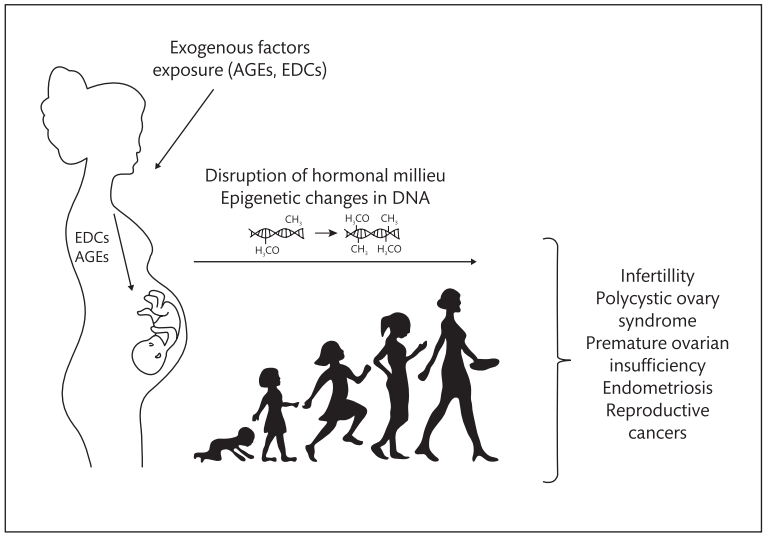


Figure 8.8.3.1 *In utero* exposure to exogenous factors (EDCs, AGEs) significantly affects multiple stages of early gonadal development. Via altering hormonal milieu and inducing epigenetic changes in DNA, endocrine disruptors have a negative impact on ovarian physiology leading to morphological and functional alterations of the female reproductive system. Ultimately, this compromised reproductive development/function constitutes an underlying pathophysiological mechanism for a group of related female reproductive diseases and disorders during adulthood, such as reduced fertility, polycystic ovary syndrome, premature ovarian insufficiency, and reproductive-site cancers.

PCOS pathogenesis has recently expanded to include EDCs, due to their potency to interfere with all hormone-sensitivity systems [11, 12].

The first study supporting the hypothesis that EDCs are actively implicated in the aetiology of PCOS was made by Fernandez *et al.* [7]. In this study, the effects of early life BPA exposure on reproductive parameters in a rat animal model, were investigated. When female rats were neonatally exposed to BPA, they exhibited a PCOS-like syndrome characterized by biochemical hyperandrogenaemia, anovulation, infertility, polycystic ovarian morphology, and increased gonadotropin-releasing hormone (GnRH) pulse frequency when they reached adulthood [7]. Human data have come to support this hypothesis. In a large study of 71 PCOS and 100 healthy women stratified by body mass index (BMI), it was showed that BPA levels are higher in PCOS women compared to controls, independently of body weight as both lean and overweight PCOS individuals had elevated BPA levels compared to normal ovulating non-hyperandrogenaemic women of matched body weight. Except for BPA, several studies have explored the higher serum concentration of other EDCs in PCOS patients: perfluorooctanoate, perfluorooctane sulfonate, polychlorinated biphenyls (PCBs), organochlorine pesticides, and polycyclic aromatic hydrocarbons [9].

Premature ovarian insufficiency: Premature ovarian insufficiency (POI) is a medical and biological diagnosis that affects young women, altering their quality of life and their fertility. In most cases, its aetiology remains undetermined, reflecting the fact that ovarian reserve can be perturbed by multifactorial, combining factors that are genetic, infectious, metabolic, as well as environmental. Given the multiple effects of EDCs in gonadal development and folliculogenesis, it is reasonable to assume that EDCs can be catalytic in POI pathogenesis. Human data to support this hypothesis are limited but illuminating. Epidemiological studies have shown that women with higher levels of phthalates and pesticides had an earlier mean age at menopause. Furthermore, DES *in utero* exposure was associated with an increased lifetime risk of early menopause in women, while increased urinary concentrations of BPA was positively associated with markers of oxidative stress and inflammation in women, suggesting BPA exposure promotes oxidative stress and inflammation, two key mechanisms of cellular senescence and ageing [1].

Endometriosis: Endometriosis affects up to 10% of women of childbearing age, causing infertility in about half of those women. EDCs have recently been implicated in the development or exacerbation of endometriosis. Women exposed to DES *in utero* may have an 80% higher risk of endometriosis than unexposed women. Furthermore, significantly higher levels of BPA, phthalates and PCB were found both in serum and peritoneal fluid of women with endometriosis, promoting chronic inflammation and triggering the stimulation of endometrial cells derived from retrograde menstruation [4, 8].

Reproductive cancers: Incidences of breast, endometrial, and ovarian cancers are increasing, and it is suspected that EDCs and other environmental factors may contribute to this increase. EDCs may act at many levels to disrupt female reproductive tissue development and to make tissue more vulnerable to a subsequent hit of an environmental chemical or hormonal insult. Several epidemiological studies have revealed an increased risk of reproductive cancer in association with EDC exposure. Occupational exposure to dioxins increased endometrial cancer risk in female workers, while women previously exposed to chlorotriazine herbicides showed a significant 2.7-fold increased risk for ovarian neoplasms [8, 11, 13].

Human Exposure Levels

Although the data about direct adverse health effects after human exposure to EDCs are still limited, abundant evidence from *in vitro* and animal studies, as well as from wildlife observations, are sufficiently alarming that several professional organizations have issued statements about a call to action. However, there is still a crucial question that has to be addressed; how experimental data can be applied in humans and whether the same chemicals can cause similar effects in humans at doses that are relevant to real-life exposure.

Several biomonitoring studies, undertaken by International Authorities are trying to estimate daily human EDCs exposure levels and establish safety margins, below which human exposure can be considered safe. Although there is chronic exposure to EDCs through inhalation and skin contact, the major route of human exposure is ingestion of food and water (e.g. meat, fish, dairy products, and vegetables), which can be contaminated with various pollutants. Indeed, plastic packaging of food, which commonly contains BPA, is now established as one of the most important daily sources of EDCs. The European Food Safety Authority has estimated that mean and high dietary exposure levels are 0.1 μg/kg and 0.4 μg/kg of bodyweight per day of BPA, respectively. Since animal studies have indicated that BPA affects ovary development at doses of ~20 μg/kg of bodyweight per day, there is a no-effect level below ~20 μg/kg of bodyweight per day in animals. When these data are applied to human risk assessment, a 100-fold difference (safety margin) between no-effect levels in animal studies human exposure levels is generally required, in order for human exposure to be considered safe. Thus, persons with high exposure levels to BPA can really be at risk for its adverse reproductive effects.

Finally, another parameter that needs to be highlighted is that humans at the same time are continuously exposed to numerous chemicals, including EDCs. Given the fact that most experimental studies typically study the effect of a single chemical under a certain period of time, it may be possible that synergistic action of various EDCs, for a longer period of time, can lead to even more deleterious consequences. Although some studies focusing on chemical mixtures have been carried out so far, there is still paucity of knowledge on how to interpret EDCs complex dose–response curves due to their overlapping additive/synergistic effects.

Nutrition: Oxidative Stress and Heat-Processed Food, Acting as Environmental Factors, Disrupting Female Ovarian Function

The function of female reproductive axis is closely linked to type and amount of nutritional intake. The impact of the quality of nutrition on female reproduction is expressed via the type of food

toxins and the oxidative stress produced from nutritional status. Westernized type of diets based on animal-derived food products are an important source of exogenous food toxins, glycotoxins (advanced glycosylated end products—AGEs) [4, 13, 14], which have been shown to interfere with ovarian function in experimental animals and in humans. Therefore, a diet containing high amounts of heat-processed food, especially with high sugar content may significantly increase the amount of these glycotoxins. Additionally, nutritional production of oxidative molecules is a critical regulator of intraovarian environment and subsequently folliculogenesis. A plethora of studies have been conducted so far in order to explore the putative role of oxidative stress in fertility. The majority of them involve women undergoing assisted reproductive techniques (ART), where various systemic and increased levels follicular fluid markers of oxidative stress were correlated with lower ovarian response, poorer embryo quality, decreased fertilization and pregnancy rates [11, 14]. Therefore, oxidative stress postprandially could be a potent, underlying mechanism of various reproductive abnormalities, including ovarian ageing, endometriosis, PCOS, and unexplained infertility. AGEs are important regulators of postprandial oxidative stress, initiating a clustering of molecular changes in several signalling pathways, including ovarian tissues. AGEs are known to induce effects within the body by two separate routes: receptor-independent or receptor—dependent [11, 12]. They can act intracellularly, by crosslinking with proteins, directly altering their structure and therefore their properties and function, or circulate and act on cell surface receptors such as the receptor for AGEs (RAGEs), which act as signal transduction receptors triggering activation of proinflammatory conditions and oxidative stress. AGEs once bonded to their multiligand receptor for advanced glycation end products (RAGE), initiate multiple downstream signalling cascades, such as nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), protein kinase C (PKC) and mitogen-activated protein kinase (MAPK) pathways. NF-κB activates nicotinamide adenine dinucleotide phosphate (NADPH) oxidase and promotes reactive oxygen species (ROS) production and a sustained inflammatory response that exacerbates further oxidative stress [12, 14]. In fact, AGE-RAGE system seems also to be actively involved in ovarian function. Furthermore, AGEs can also interfere in the glucose homeostasis of granulosa cells, lead to reduced glucose uptake, and disrupt normal follicular growth. Human data, regarding the action of dietary AGEs in female reproductive system, mainly originate from women with PCOS. Specifically, AGEs were found to be distinctively elevated in the serum of women with PCOS, independent of the presence of obesity and insulin resistance. Experimental animal and human studies have underpinned the detrimental effects of dietary AGEs in reproductive and metabolic function of PCOS [14, 15]. In female high-AGE fed mice increased deposition of AGEs and RAGE in theca and granulosa cells was observed, altering the ovarian milieu. In addition, high-AGE diet, in synergy with androgen excess, was shown to downregulate glyoxalase-1 in the ovary, a protective enzyme against glycation, promoting AGE accumulation and ovarian dysfunction. Subsequently, AGE accumulation in the microenvironment of ovary has detrimental effects in the reproductive function [11, 12]. Through downregulating signalling cascades and enzymatic activity, there is no doubt that AGEs contribute significantly to follicular arrest, anovulation, and ovarian dysfunction in PCOS [16]. Analogous findings exist for AGEs, with increased concentration of them both in the follicular fluid and in serum of women undergoing IVF to be negatively correlated with follicular growth, fertilization, and embryonic development. On the other hand, sRAGE, the decoy scavenger of AGEs, via binding and inactivating circulating and follicular fluid AGEs, might be able to serve as a useful biological marker of the follicular environment. In women undergoing IVF it was shown that higher follicular fluid sRAGE levels were accompanied by increased number of oocytes retrieved. Altogether, the literature to date indicates that AGEs have a negative impact on reproductive outcome in women undergoing ART [16].

Summary

In conclusion, acute or chronic exposure to EDCs and AGEs through each stage of life cycle has the potential to alter the structure as well as the functional or hormonal homeostasis in female reproductive system. Extensive data from different scientific models, collected in recent years, have confirmed the increased concern on their negative role. Concomitantly, potentiating further the scientific efforts to explore efficient ways of avoiding their harmful effects on every system and especially on the female reproduction.

REFERENCES

1. Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, *et al.* Endocrine-disrupting chemicals: an Endocrine Society scientific statement. *Endocr Rev*, 2009; **30**: 293–342.
2. Rattan S, Zhou C, Chiang C, Mahalingam S, Brehm E, Flaws JA. Exposure to endocrine disruptors during adulthood: consequences for female fertility. *J Endocrinol*, 2017; **233**: R109–29.
3. Johansson HKL, Svingen T, Fowler PA, Vinggaard AM, Boberg J. Environmental influences on ovarian dysgenesis—developmental windows sensitive to chemical exposures. *Nat Rev Endocrinol*, 2017; **13**: 400–14.
4. Rutkowska AZ, Diamanti-Kandarakis E. Polycystic ovary syndrome and environmental toxins. *Fertil Steril*, 2016; **106**: 948–58.
5. Chatzigeorgiou A, Kandaraki E, Piperi C, *et al.* Dietary glycotoxins affect scavenger receptor expression and the hormonal profile of female rats. *J Endocrinol*, 2013; **218**: 331–7.
6. Marques-Pinto A, Carvalho D. Human infertility: are endocrine disruptors to blame? *Endocr Connect*, 2013; **2**: R15–29.
7. Fernandez M, Bianchi M, Lux-Lantos V, Libertun C. Neonatal exposure to bisphenol A alters reproductive parameters and gonadotropin releasing hormone signalling in female rats. *Environ Health Perspect*, 2009; **117**: 757–62.
8. Karwacka A, Zamkowska D, Radwan M, Jurewicz J. Exposure to modern, widespread environmental endocrine disrupting chemicals and their effect on the reproductive potential of women: an overview of current epidemiological evidence. *Hum Fertil (Camb)*, 2017; **22**: 2–25.
9. Kandaraki E, Chatzigeorgiou A, Livadas S, *et al.* Endocrine disruptors and polycystic ovary syndrome (PCOS): elevated serum levels of bisphenol A in women with PCOS. *J Clin Endocrinol Metab*, 2011; **96**: E480–4.
10. Palioura E, Diamanti-Kandarakis E. Polycystic ovary syndrome (PCOS) and endocrine disrupting chemicals (EDCs). *Rev Endocr Metab Disord*, 2015; **16**: 365–71.

11. Diamanti-Kandarakis E, Papalou O, Kandaraki EA, Kassi G. Mechanisms in endocrinology: nutrition as a mediator of oxidative stress in metabolic and reproductive disorders in women. *Eur J Endocrinol*, 2017; **176**: R79–99.
12. Kandaraki E, Chatzigeorgiou A, Piperi C, Palioura E, Palimeri S, Korkolopoulou P, Koutsilieris M, Papavassiliou AG. Reduced ovarian glyoxalase-I activity by dietary glycotoxins and androgen excess: a causative link to polycystic ovarian syndrome. *Mol Med*, 2012; **18**: 1183–9.
13. Grieger JA, Grzeskowiak LE, Bianco-Miotto T, *et al*. Pre-pregnancy fast food and fruit intake is associated with time to pregnancy. *Hum Reprod*, 2018; **33**: 1063–70.
14. Kandaraki EA, Chatzigeorgiou A, Papageorgiou E, *et al*. Advanced glycation end products interfere in luteinizing hormone and follicle stimulating hormone signaling in human granulosa KGN cells. *Exp Biol Med (Maywood)*, 2018; **243**: 29–33.
15. Tantalaki E, Piperi C, Livadas S, *et al*. Impact of dietary modification of advanced glycation end products (AGEs) on the hormonal and metabolic profile of women with polycystic ovary syndrome (PCOS). *Hormones (Athens)*, 2014; **13**: 65–73.
16. Merhi Z, Kandaraki EA, Diamanti-Kandarakis E. implications and future perspectives of AGEs in PCOS pathophysiology. *Trends Endocrinol Metab*, 2019; **30**: 150–62.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

SECTION 9

Endocrine Disorders of Pregnancy

9.1

General Considerations Relating to Thyroid Disease in Pregnancy

Peter N. Taylor, L.D.K.E. Premawardhana, and John H. Lazarus

Introduction

Thyroid hormone is essential for maintaining a pregnancy and ensuring fetal development [1]. Thyroid disorders are common in women of childbearing age and as pregnancy has a substantial impact on the hypothalamic–pituitary–thyroid axis, abnormal thyroid function is frequently encountered in antenatal clinics. It is also well established that overt thyroid disease is associated with adverse obstetric and offspring neuro-developmental outcomes [1, 2]. There is now growing concern that more marginal degrees of thyroid dysfunction particularly subclinical hypothyroidism (elevated TSH and normal fT_4 concentration) and isolated hypothyroxinaemia (normal TSH and low fT_4) are also associated with fetal loss, prematurity, and impaired offspring cognitive function [3–5]. In some studies, maternal thyroid autoimmunity has also been identified as a potential risk for fetal loss [1] even in euthyroid women. Correction of overt hypothyroidism and hyperthyroidism dramatically reduces the risk of major adverse obstetric outcomes including fetal loss and premature birth.

There has been a dramatic increase in our understanding of both thyroid physiology and immunology in pregnancy as well as the importance of more personalized thyroid hormone reference-ranges and the potential for adverse outcomes with even borderline thyroid function abnormalities. Equally, we now realize other factors including thyroid autoimmunity, hCG levels and iodine status play a role. Iodine is essential for thyroid hormone production. Iodine deficiency is also common and pregnant women and children at greatest risk of being iodine deficient. Iodine deficiency is associated with adverse profound pregnancy and obstetric outcomes. Endocrine disruptors may also interact with iodine status thereby influencing thyroid status and pregnancy outcomes.

Interpretation of Thyroid Function Tests During Pregnancy

The profound physiological changes that occur during pregnancy have substantial consequences for the thyroid axis and in the assessment of thyroid function. It is clear there is a significant overlap between the signs and symptoms of the hypermetabolic state typical of normal euthyroid pregnant women, and those due to thyroid dysfunction. Taken together, the diagnosis of thyroid disease during pregnancy can be challenging and the availability of reliable accurate tests for gestational thyroid function is critical. Unfortunately, the notable underlying physiological changes which occur during pregnancy also create challenges in laboratory assessments of thyroid function. The substantial but reversible changes in thyroid function which occur in pregnancy are summarized in **Table 9.1.1** and **Figure 9.1.1**.

Abnormal thyroid function is best defined using pregnancy-specific reference ranges. calculated in a population of pregnant women free of major factors that interfere with thyroid function. These factors include pre-existing thyroid disease and autoimmunity, iodine deficiency, or high hCG states (twin pregnancy or following IVF) [1]. Excluding thyroid autoimmunity can substantially lower the upper limit of the TSH threshold. The most recent American Thyroid Association guidelines advocate the use of pregnancy specific and population-based reference ranges [6]. Additionally, calls have been made for more personalized reference ranges based on ethnicity [7] BMI parity, and fetal sex [8].

Traditionally, TSH was the mainstay of assessing thyroid status in pregnancy, however there is increasing evidence of important obstetric and offspring outcomes with fT_4 status [1, 9]. Furthermore, fT_4 levels are needed to distinguish between subclinical and overt thyroid disease. fT_4 immunoassays which are commonly used may

Table 9.1.1 Physiological changes that influence thyroid function in pregnancy

Physiological change	Effect on thyroid function test	Key point for clinicians
↑ Thyroxine-binding globulin	↑ Serum total T_3 and T_4 concentration	Total thyroid hormone concentrations may be misleading; rely on free thyroid hormone concentrations
Secretion of human chorionic gonadotrophin	↑ Free T_4 and ↓ TSH	High human chorionic gonadotrophin levels may result in gestational thyrotoxicosis. This usually only requires symptomatic treatment but needs to be distinguished from pathological thyroid disease
↑ Iodine clearance	↓ Hormone production in iodine-deficient areas	Need to be mindful of iodine deficiency and ensure optimal intake ideally from before conception
↑ Plasma volume	↑ T_3 and T_4 pool size	
Increased type 3 5-deiodinase (inner ring deiodination) activity from the placental	↑ T_3 and T_4 degradation	Increasing thyroid demand in pregnancy
Thyroid enlargement (in some women)	Increased thyroglobulin	Be aware that a small goitre is common in pregnancy, but may be a sign of low thyroid function so merits thyroid function testing

be less accurate in the third trimester due to alterations in thyroid hormone binding proteins. Fortunately, results are more accurate in the first trimester which has greater influence on obstetric and offspring outcomes.

Overall, there is a transient rise in free T_4 in the first trimester due to the relatively high circulating human chorionic gonadotropin (hCG) concentration which stimulates the thyroid to produce thyroid hormone. This is followed by a decrease of free T_4 in the second and third trimester. In iodine-deficient areas (including marginal iodine deficiency seen in many European countries), pregnant woman may become significantly hypothyroxinaemic with preferential triiodothyronine (T_3) secretion. The thyroidal 'stress' is also evidenced by a rise in the median thyroid-stimulating hormone (TSH) and serum thyroglobulin levels. A summary of TSH fT_4 and hCG concentration during pregnancy are shown in Figure 9.1.1.

Thyroid Hormone Supply to the Fetus

The fetal thyroid begins concentrating iodine at 10–12 weeks' gestation and is under the control of fetal pituitary TSH at approximately 20 weeks' gestation. Maternal thyroid function is therefore critically important in the first trimester as thyroid hormone is crucial for the development of many organs including the brain. Maternal circulating T_4 crosses the placenta to the fetus at all stages of pregnancy by incompletely understood mechanisms but involving both the type 2 and type 3 deiodinase enzymes, which are both significantly expressed in the placenta. Type 3 deiodinase (D_3), which degrades thyroid hormones [10], is also expressed in pregnant uterus, placenta, and fetal and neonatal tissues, and may act as a 'gatekeeper' to prevent too much thyroid hormone transport. Type 2 deiodinase, also located in the uterus and other parts of the genital tract, degrades T_4 in the fetus to provide T_3 for tissue growth and differentiation, and may have a role in fetal implantation. Genetic association studies have also indicated that variation in the deiodinases are associated with maternal pregnancy outcomes, particularly pre-eclampsia [11].

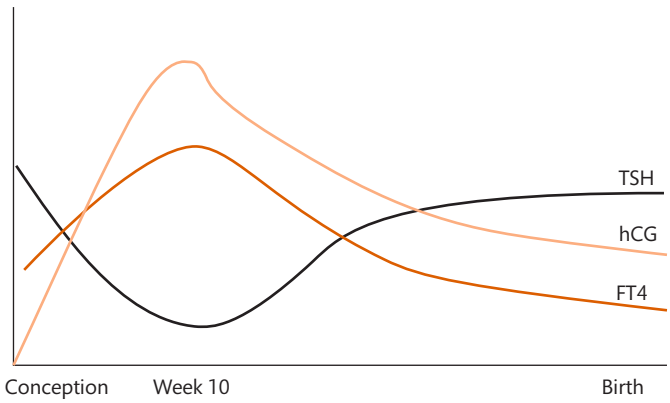


Figure 9.1.1 TSH, fT_4, and hCG levels over pregnancy.

Pregnancy and the Immune System

Human immune regulation involves homeostasis between T helper 1 (Th1) and T helper 2 (Th2) activity, with Th1 cells driving cellular immunity and Th2 cells humoral immunity [2]. In pregnancy, there is a bias towards a Th2 lymphocyte response evidenced by the fetal/placental unit producing Th2 cytokines, which inhibit Th1. Th1 cytokines are potentially harmful to the fetus, for example interferon-α, which is a known abortifacient.

Pregnancy also has a significant effect on the immune system in order to maintain the fetal–maternal allograft and prevent rejection. The trophoblast does not express the classic major histocompatibility complex (MHC) class Ia or II which are needed to present antigenic peptides to cytotoxic cells and T helper cells, respectively. Instead HLA-G, a non-classic MHC Ib molecule, is expressed which may be a ligand for the natural killer (NK) cell receptor thereby protecting the fetus from NK cell damage; it may also activate CD8 T cells that may have a suppressor function. Human trophoblasts also express abundant Fas ligand, thereby contributing to the immune privilege by mediating apoptosis of activated Fas-expressing lymphocytes of maternal origin.

Pregnancy and Thyroid Peroxidase Antibodies

Anti-thyroid peroxidase antibodies (TPOAbs) are found in around 10% of otherwise normal pregnant women when measured at the end of the first trimester. They are a marker of thyroid autoimmunity and the main risk factor for thyroid dysfunction in pregnancy and

in the postpartum period. Women who are TPOAb positive have higher TSH concentrations, lower T4 concentrations and a higher risk of thyroid dysfunction during pregnancy than women who are TPOAb negative [1]. Even in euthyroid women TPOAb positivity is associated with premature delivery and fetal loss [4, 12]. One randomized controlled trial has shown that levothyroxine administration reduced the risk of fetal loss and premature delivery in euthyroid TPOAb-positive women [13]. A later study failed to replicate the benefits for fetal loss but did find a reduction in pre-term delivery [14].

Worse pregnancy outcomes are seen in TPO antibody-positive women, especially those with higher TSH [15]. Recent data suggests that TPOAb positive women do not mount a thyroidal response to hCG [8]; this may, in part, explain the observed worse obstetric outcomes in TPOAb positive women as well as the interactive effect of TPO antibodies with higher levels of TSH. However, the benefits of treating with levothyroxine are unclear. An open label randomized trial in Chinese women with normal thyroid function and TPOAb positivity who were undergoing IVF found that levothyroxine did not reduce miscarriage rates or increase live birth rates [16]. The TABLET trial [17], a large double-blind, placebo-controlled trial also found no benefit from levothyroxine treatment on live birth rates in women with TPOAb positivity and a history of previous miscarriage or infertility. Furthermore, this trial [17] found no benefit of levothyroxine treatment on pregnancy loss, preterm birth, or neonatal outcomes. Nevertheless, it is still important to consider whether there is a synergistic effect of thyroid autoimmunity with borderline low thyroid dysfunction and whether results might have been different in the general population. Therefore, these trials, although elegant, do not address whether universal thyroid screening in pregnancy should be undertaken and trials here are still needed.

Effects of Iodine Status and Nutrition

It has long been established that iodine is essential for thyroid hormone synthesis [18]. Iodine deficiency is also common throughout the world with pregnant women and children at greatest risk of deficiency. Iodine deficiency during pregnancy is associated with maternal goitre due to the imbalance between the intake and increased requirements for iodine during gestation and results eventually in a reduced circulating maternal thyroxine (T_4) concentration. This gestational goitrogenesis is preventable by iodine supplementation [19], not only in areas of severe iodine deficiency, but also has been observed in areas of mild to moderate deficiency [20].

In addition to effects on thyroid growth, iodine deficiency is also associated with adverse obstetric outcomes including fetal loss, profound intellectual disability (cretinism), and goitre. The effects of iodine deficiency are most profound in areas of severe deficiency, but mild to moderate iodine deficiency may have adverse consequences at the population level. Even mild to moderate iodine deficiency in pregnant women is associated with adverse maternal effects including goitre [21] and lower IQ in offspring [22]. Although a recent trial did not show benefit of iodine supplementation on child neurocognition during pregnancy in areas of mild iodine deficiency [23] further studies are required, especially in areas of moderate iodine deficiency.

Table 9.1.2 Iodine status in pregnancy by urine iodine concentration

Median urinary iodine concentration (mcg/L)	Iodine status
Pregnant women	
<150	Insufficient iodine status
150–249	Satisfactory iodine status
250–499	Above required intake
≥500	Excessive iodine status

However, assessment of iodine status, particularly at the individual level is challenging and at present median urinary iodine concentration is used to determine the iodine status of a population. Alternative methods for assessing iodine status such as thyroglobulin levels are also being evaluated. Often there is an increase in urinary iodine in the first trimester compared to non-pregnant women, but where the population has a high median iodine concentration this difference may not occur. Urinary iodine excretion in pregnancy is maximal in the first trimester followed by a decline in the second and third trimesters.

The recommended daily iodine intake in pregnancy has been increased to 250 mcg/day which implies a urinary iodine excretion of 150–250 mcg/day as being adequate [24, 25] (Table 9.1.2). However, despite decades of advocacy, this is not always achieved even in developed parts of the world.

Iodine supplementation is essential to provide iodine sufficiency. This is usually undertaken using iodized salt. This vehicle has been chosen as salt is regularly added to meals during cooking and is used in staple foods such as bread and commercial ready meals. Universal salt iodization has been shown to be extremely effective in reducing iodine deficiency. In keeping with other food fortification initiatives, universal salt iodization programs require national legislation as regional or voluntary programs usually fail to succeed. In the last 30 years, there has been mandatory salt iodization in many countries particularly in the developing world. Iodine fortification of all food-grade salt is now mandated in ~120 countries [26]. While effective iodine supplementation can increase the risk of thyrotoxicosis in older individuals and thyroid autoimmunity in younger people, however, this is far less than the benefits of ensuring iodine sufficiency in pregnant women and children [18].

Screening for Thyroid Dysfunction in Pregnancy

Whether or not to initiate universal thyroid screening is a key debate in thyroidology [3]. Thyroid disorders represent a key easily modifiable risk factor for adverse maternal and offspring outcomes [1] and are common in women of childbearing age [27]. These factors have raised the possibility that universal thyroid screening may be desirable in pregnancy [28]. Some countries including Spain China and Poland are undertaking universal screening although the latest American Thyroid Association guidance discusses screening, but makes no decision for or against universal screening [6]. It is however accepted that targeted screening should be performed in those women at high risk for thyroid disease [6, 29] (Box 9.1.1).

Box 9.1.1 Women at high risk of thyroid disease during pregnancy

- Women with a history of thyroid disease (including hyperthyroidism, hypothyroidism, and post-partum thyroiditis) or thyroid surgery
- Women with a goitre
- Women with symptoms or signs suggestive of hypothyroidism or hyperthyroidism
- Women with a family history of thyroid disease
- Women known to be positive for thyroid antibodies
- Women with type 1 diabetes or other autoimmune disorders
- Women with a history of infertility, miscarriage, or preterm delivery
- Women with a history of head or neck irradiation

However, this targeted approach would miss approximately one third of women with significant thyroid dysfunction [30]. This approach has been further called into question as universal thyroid screening is pregnancy appears to very cost-effective; screening solely for overt hypothyroidism also had a cost-effectiveness ratio of $6776/QALY (quality adjusted life-year) [31] which is favourable compared to gestational diabetes mellitus screening ($12 078/QALY) and is well below the $50 000/QALY figure used in the United States as a criterion for screening decisions. This work also indicated universal screening appears to be more cost-effective than targeted screening.

Universal screening for thyroid dysfunction in pregnancy also meets the majority of the essential ten criteria for screening as recommended by Wilson and Junger [32]. Even if only overt thyroid disease is considered, it is still an important health problem (criteria 1). It is a common problem in pregnancy, overt hypothyroidism occurs in approximately 0.2–0.6% of pregnancies [33, 34] while overt hyperthyroidism, occurs with a frequency of about 0.2% [2]. Subclinical hypothyroidism occurs in 2–3% of pregnancies [1, 6] and isolated hypothyroxinaemia, may vary between 1% and 10% in iodine-replete populations [35] depending on the cut-off for the reference ranges used

Treatments for thyroid dysfunction are well established and acceptable to patients (criteria 2) with facilities for diagnosis readily available (criteria 3). In particular, levothyroxine is a safe treatment for hypothyroidism and we have over 50 years' experience with it [36]. The diagnosis of thyroid disease and treatment provision can also be easily be managed in outpatient clinics. Thyroid disease, particularly hypothyroidism, is frequently asymptomatic (criteria 4). In particular, subclinical hypothyroidism may represent an early disease state of hypothyroidism. In pregnancy, about 25% of women with subclinical hypothyroidism will have persistent TSH elevation following delivery [37].

Thyroid function testing relies on a simple blood test which is both suitable and acceptable (criteria 5 and 6) and these simple non-invasive tests are relatively inexpensive and available in modern laboratories. These blood tests can also be readily incorporated into routine blood tests which need to be done in early pregnancy. However, although pre-pregnancy screening would be likely to give the best outcomes, this is usually not practical. The natural history of thyroid disease is well understood (criteria 7) it is also cost effective (see earlier) (criteria 9) and the nature of screening in pregnancy means it will be a continuing process (criteria 10). Taken together these factors make a compelling case for universal thyroid in screening in pregnancy.

Only criteria 8, 'there should be an agreed policy on whom to treat as patients', is not satisfied. Whilst all relevant endocrine and obstetric societal guidelines endorse treatment of overt thyroid disease, it is the management of subclinical hypothyroidism and isolated hypothyroxinaemia which is less clear. It is also likely that there is an interaction with thyroid autoimmunity. Here the ATA guidance [6] is helpful which recommends to consider treatment with levothyroxine at TSH levels 4.0–10 mU/L if antibody negative and consider treatment at TSH levels 2.5–4 mU/L if antibody positive. Clinicians should definitely treat if TSH > 4.0 mU/L and antibody positive or TSH >10 mU/L [6] this provides much needed flexibility for clinicians although guidance is generally less clear.

An additional concern is that treatment with levothyroxine may not be free from consequences although widely regarded to be safe. Data from the Generation-R project [9] did show an inverse U shape with maternal fT_4 levels and offspring IQ. Thus, overtreatment may result in negating the benefits of supplementation with levothyroxine with the potential for a deleterious effect on offspring IQ. Preliminary data from the CATS II study also showed adverse outcomes in women who were over-treated with levothyroxine on offspring behaviour. While observational data [38] identified that SCH during pregnancy is associated with multiple adverse maternal and neonatal outcomes, the value of levothyroxine therapy in preventing these adverse outcomes remains uncertain. It is also worth highlighting that current management of thyroid hormone replacement in women established on levothyroxine prior to pregnancy is suboptimal with poorer control associated with increased odds of fetal loss [39].

The ultimate aim of screening for maternal thyroid disease in pregnancy is to optimize feto-maternal outcomes. Thus, therapeutic intervention when indicated should be implemented as early as possible in the course of fetal development. This will still need addressing with trials of universal screening and treating for low thyroid function in **early** (<10 weeks) gestation with key obstetric outcomes such as fetal loss, birthweight and gestational age evaluated as short term outcomes and with longer follow-up required for childhood neurological development outcomes. Until clearer evidence of benefit from treating marginal thyroid dysfunction is available from further randomized controlled trials in early pregnancy this controversy will continue.

A pragmatic approach may be to measure TSH and then reflex FT4 and TPOAb if TSH is outside of the relevant reference range. This will be challenging, but can with careful planning can be integrated into routine community health services. Preconception health examinations including thyroid hormone measurement have been managed successfully in rural parts of China [40]. A recent analysis of 184 611 pregnant participants in this programme has demonstrated an association between higher preconception TSH concentrations and risks of miscarriages and pre-term loss, highlighting key opportunities exist for modifying pregnancy outcomes with effective and timely preconception care [41].

A number of practical considerations will also need to be addressed before a universal thyroid screening program could be considered. There is a clear need for population specific and gestational age specific reference ranges in addition to a consensus regarding treatment thresholds and treatment targets. Furthermore, there will need to be enhancement in the skills in interpreting thyroid function tests in pregnancy by both endocrinologists and obstetricians.

REFERENCES

1. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. *Nat Rev Endocrinol*, 2017; **13**: 610–22.
2. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*, 2013; **1**: 238–49.
3. Velasco I, Taylor P. Identifying and treating subclinical thyroid dysfunction in pregnancy: emerging controversies. *Eur J Endocrinol*, 2018**; 178**: D1–12.
4. Korevaar TI, Schalekamp-Timmermans S, de Rijke YB, *et al.* Hypothyroxinemia and TPO-antibody positivity are risk factors for premature delivery: the generation R study. *J Clin Endocrinol Metab* 2013; **98**: 4382–90.
5. Chan S, Boelaert, K. Optimal management of hypothyroidism, hypothyroxinaemia and euthyroid TPO antibody positivity preconception and in pregnancy. *Clin Endocrinol (Oxf)*, 2015; **82**, 313–26.
6. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
7. Korevaar TI, Medici M, de Rijke YB, *et al*. Ethnic differences in maternal thyroid parameters during pregnancy: the Generation R study. *J Clin Endocrinol Metab*, 2013; **98**: 3678–86.
8. Korevaar TI, de Rijke YB, Chaker L, *et al*. Stimulation of thyroid function by human chorionic gonadotropin during pregnancy: a risk factor for thyroid disease and a mechanism for known risk factors. *Thyroid*, 2017; **27**: 440–50.
9. Korevaar TI, Muetzel R, Medici M, *et al*. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*, 2015; **4**: 35–43.
10. Hernandez A, Martinez ME, Fiering S, Galton VA, St Germain D. Type 3 deiodinase is critical for the maturation and function of the thyroid axis. *J Clin Invest*, 2006; **116**: 476–84.
11. Verloop H, Dekkers OM, Peeters RP, Schoones JW, Smit JW. Genetics in endocrinology: genetic variation in deiodinases: a systematic review of potential clinical effects in humans. *Eur J Endocrinol*, 2014; **171**: R123–35.
12. Thangaratinam S, Tan A, Knox E, *et al*. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. *BMJ*, 2011; **342**: d2616.
13. Negro R, Formoso G, Mangieri T, Pezzarossa A, Dazzi D, Hassan H. Levothyroxine treatment in euthyroid pregnant women with austoimmune thyroid disease: effects on obstetrical complications. *J Clin Endocrinol Metab*, 2006; **91**: 2587–91.
14. Nazarpour S, Ramezani Tehrani F, *et al*. Effects of levothyroxine treatment on pregnancy outcomes in pregnant women with autoimmune thyroid disease. *Eur J Endocrinol*, 2017; **176**: 253–65.
15. Liu H, Shan Z, Li C, Mao J, *et al*. Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid*, 2014; **24**: 1642–9.
16. Wang H, Gao H, Chi H, *et al*. Effect of levothyroxine on miscarriage among women with normal thyroid function and thyroid autoimmunity undergoing in vitro fertilization and embryo transfer: a randomized clinical trial. *JAMA*, 2017; **318**: 2190–8.
17. Dhillon-Smith RK, Middleton LJ, Sunner KK, *et al*. Levothyroxine in women with thyroid peroxidase antibodies before conception. *N Engl J Med*, 2019; **380**: 1316–25.
18. Zimmermann MB. Iodine deficiency. *Endocr Rev*, 2009; **30**: 376–408.
19. Zimmermann MB. Iodine deficiency in pregnancy and the effects of maternal iodine supplementation on the offspring: a review. *Am J Clin Nutr*, 2009; **89**: 668S–72S.
20. Taylor PN, Okosieme OE, Dayan CM, Lazarus JH. Therapy of endocrine disease: Impact of iodine supplementation in mild-to-moderate iodine deficiency: systematic review and meta-analysis. *Eur J Endocrinol*, 2014; **170**: R1–R15.
21. Taylor PN, Razvi S, Pearce SH, Dayan CM. A review of the clinical consequences of variation in thyroid function within the reference range. *J Clin Endocrinol Metab*, 2013; **98**: 3562–71.
22. Bath SC, Steer CD, Golding J, Emmett P, Rayman MP. Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: results from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Lancet*, 2013; **382**, 331–7.
23. Gowachirapant S, Jaiswal N, Melse-Boonstra A, *et al*. Effect of iodine supplementation in pregnant women on child neurodevelopment: a randomised, double-blind, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 853–63.
24. Zimmermann MB, Gizak M, Abbott K, Andersson M, Lazarus JH. Iodine deficiency in pregnant women in Europe. *Lancet Diabetes Endocrinol*, 2015; **3**, 672–4.
25. Taylor P, Vaidya B, Taylor P, Vaidya B. Iodine supplementation in pregnancy—is it time? *Clin Endocrinol (Oxf)*, 2016; **85**: 10–14.
26. Dasgupta PK, Liu Y, Dyke JV. Iodine nutrition: iodine content of iodized salt in the United States. *Environ Sci Technol*, 2008; **42**: 1315–23.
27. Taylor PN, Albrecht D, Scholz A, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*, 2018; **14**: 301–16.
28. Taylor PN, Okosieme OE, Premawardhana L, Lazarus JH. Should all women be screened for thyroid dysfunction in pregnancy? *Women's Health*, 2015; **11**: 295–307.
29. De Groot L, Abalovich M, Alexander EK, *et al*. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 2543–65.
30. Vaidya B, Anthony S, Bilous M, *et al*. Detection of thyroid dysfunction in early pregnancy: universal screening or targeted high-risk case finding? *J Clin Endocrinol Metab*, 2007; **92**: 203–7.
31. Dosiou C, Barnes J, Schwartz A, Negro R, Crapo L, Stagnaro-Green A. Cost-effectiveness of universal and risk-based screening for autoimmune thyroid disease in pregnant women. *J Clin Endocrinol Metab*, 2012; **97**: 1536–46.
32. Wilson J, Jungner G. *Principles and Practice of Screening for Disease*. Geneva: World Health Organization, 1968. Public Health Papers, 2011; **34**.
33. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
34. Medici M, Korevaar TI, Visser WE, Visser TJ, Peeters RP. Thyroid function in pregnancy: what is normal? *Clin Chem*, 2015; **61**: 704–13.
35. Negro R, Soldin OP, Obregon MJ, Stagnaro-Green A. Hypothyroxinemia and pregnancy. *Endocr Pract*, 2011; **17**: 422–9.
36. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. *Lancet*, 2017; **390**: 1550–62.
37. Shields BM, Knight BA, Hill AV, Hattersley AT, Vaidya, B. Five-year follow-up for women with subclinical hypothyroidism in pregnancy. *J Clin Endocrinol Metab*, 2013; **98**: E1941–5.

38. Maraka S, Ospina NM, O'Keeffe DT, *et al*. Subclinical hypothyroidism in pregnancy: a systematic review and meta-analysis. *Thyroid*, 2016; **26**: 580–90.
39. Taylor PN, Minassian C, Rehman A, *et al*. TSH levels and risk of miscarriage in women on long-term levothyroxine: a community-based study. *J Clin Endocrinol Metab*, 2014; **99**: 3895–902.
40. Zhou Q, Acharya G, Zhang S, Wang Q, Shen H, Li X. A new perspective on universal preconception care in China. *Acta Obstet Gynecol Scand*, 2016; **95**: 377–81.
41. Chen S, Zhou X, Zhu H, *et al*. Preconception TSH and pregnancy outcomes: a population-based cohort study in 184 611 women. *Clin Endocrinol (Oxf)*, 2017; **86**: 816–24.

9.2

Management of Thyroid Disorders Before Assisted and Spontaneous Pregnancies

Kris Poppe, Flora Veltri, and David Unuane

Introduction

Thyroid disorders are frequent in women of reproductive age, as shown in a recent survey in Spain in an area without iodine deficiency. Here, the prevalence of subclinical hypothyroidism (SCH) was 4–6%, hyperthyroidism 1% and thyroid autoimmunity (TAI) 10% [1]. Furthermore, the prevalence of TAI seems to be higher in women with particular causes of infertility such as polycystic ovary syndrome (PCOS) and idiopathic infertility compared to that in fertile women or with other causes of infertility [2]. Severe thyroid dysfunction may lead to irregular periods and infertility through direct and indirect interactions with the reproductive system. Although restoring normal thyroid function often results in a regular menstrual pattern and/or a normalized reproductive hormonal profile, it is not always followed by pregnancy. Indeed, in case of male infertility, endometriosis, tubal obstruction or PCOS, surgery and/or assisted reproductive technology (ART) may be necessary. Prior to the in-vitro part of ART treatment, an ovarian stimulation procedure (OH) is necessary. The latter is characterized by high oestradiol (E2) levels, adding additional strain on the thyroid gland during pregnancy. Especially in women with TAI this may be a supplementary reason to optimize thyroid function before pregnancy in order to obtain good quality embryos, and to avoid the progression from subclinical to overt hypothyroidism during pregnancy [3].

Before pregnancy, hypothyroid patients treated with levothyroxine (LT4) are advised to have a serum TSH level of less than 2.5 mIU/L, both in the case of assisted and spontaneous pregnancies. Women with Graves' disease desiring pregnancy should be advised of the increased risk of maternal and fetal complications as well as of possible side effects associated with antithyroid drugs (ATD). If necessary, contraception must be proposed and pregnancy postponed until euthyroidism is reached and confirmed [3].

With the exception of women planning ART or those known to have TAI, at present there are no recommendations regarding universal screening for thyroid function in the preconception phase. However, many physicians determine thyroid function parameters and antibody levels in general check-ups before pregnancy (sometimes upon request of the women/couple). Against this background, recommendations on how to manage fortuitous findings are useful [3, 4].

Thyroid (Dys)function and Female (In)fertility

Interaction Between Thyroid and Gonadal Function

The prevalence of hypothyroidism in women of reproductive age varies between 0.2% and 0.4% and is predominantly related to TAI [1, 2]. Moreover, the prevalence of TAI seems to be elevated in young women facing fertility problems, with thyroglobulin autoantibodies (Tg-abs) and thyroid peroxidase autoantibodies (TPO-abs) being reported in up to 25% of women affected by recurrent miscarriages [5, 6]. Thyroid hormones (TH) may influence fertility aspects in a direct or indirect manner. TSH and its receptor, as well as thyroid hormone receptors (TR-α1 and TR-β1) have been located in ovarian surface epithelium and in oocytes of primordial, primary, and secondary follicles. Therefore, they appear to play a role in the regulation of ovarian function [7]. Indirect mechanisms linking thyroid hormone to fertility issues may involve altering gonadotropin releasing hormone (GnRH) and prolactin secretion as well as changes in sex hormone binding globulin (SHBG) levels and coagulation factors [2]. Since TH influence the hormonal environment of the ovarian follicle and are essential in the signalling during implantation and early pregnancy, (mild) thyroid failure is plausible cause of infertility.

Thyroid Dysfunction and Reproduction

Subclinical and Overt Hyperthyroidism

Due to limited data the precise impact of hyperthyroidism on fertility is not well established.

However, in overtly hyperthyroid women, menstrual irregularities haven been reported in up to 65% compared to 17 % in healthy controls [8]. More recent reports have pointed towards a lower prevalence of menstrual abnormalities of about 22% compared to 8 % in healthy controls [2]. In hyperthyroidism, the most common abnormality was hypomenorrhea followed by polymenorrhagia, oligomenorrhea and hypermenorrhoea. Nevertheless, despite thyrotoxicosis being linked to reduced fertility, studies including endometrial biopsies have demonstrated that most hyperthyroid women seem to remain ovulatory [9].

Subclinical and Overt Hypothyroidism

There are inconsistent reports on the association between SCH and infertility, largely caused by different cut-offs used in various studies to define the condition. In general, despite a slight increase in the mean TSH concentration in infertile women, the incidence of SCH is comparable to that in the general population [5, 10]. Among these women affected by infertility, TSH levels were highest in those with ovulatory dysfunction.

As for hyperthyroidism, limited data are available on the prevalence of infertility in women with overt hypothyroidism. Nonetheless, some evidence suggests an association of overt hypothyroidism and an increased risk of infertility. Indeed, hypothyroidism may lead to a number of menstrual disturbances in women of fertile age. There may be changes in cycle length as well as the volume of menstrual bleeding secondary to breakthrough bleeding following anovulation and/or disturbances in haemostatic factors. The prevalence of menstrual abnormalities (especially oligomenorrhea) has been reported in 25–60% of hypothyroid women compared to 10% in euthyroid women [2].

Thyroid Autoimmunity

The prevalence of TAI varies considerably with ethnicity but is generally estimated around 10% [1, 11]. A higher prevalence of TAI has been reported in women affected by fertility problems [2, 5]. The underlying causal mechanisms that may account for any possible negative impact of TAI on fertility also remain unclear. On the one hand it has been proposed that the presence of TAI reflects a general immune imbalance possibly leading to implantation failure. Another explanation could be that thyroid antibodies are directly causative whereby the presence of thyroid antibodies could have an unfavourable effect on oocyte and embryo quality. Indeed, it has been hypothesized that thyroid antibodies may pass the 'follicle-blood' barrier during maturation of a secondary follicle. Cytotoxicity of these antibodies could then damage the maturing oocyte and eventually reduce oocyte quality and fertilization potential [12]. On the other hand, a positive TAI status increases the risk of developing (subclinical) hypothyroidism. Women with TAI are at increased risk of developing (sub)clinical hypothyroidism after ovarian hyperstimulation and during pregnancy [1, 2, 11].

General Considerations Relating to Fertility

Infertility is defined as the inability to conceive by a sexually active, couple not using contraception over a 12-month period for women <35 years and after 6 months for women >35 years. Infertility is a common condition and about 10–15% of all couples will experience difficulties to conceive (primary infertility) or to conceive the number of children they want (secondary infertility) [10, 13]. Subfertility is sometimes distinguished from infertility, where the latter is a synonym of sterility, defined as the absolute inability to conceive (absence of sperm, premature menopause, or complete tubal obstruction). According to the World Health Organization (WHO) task force on diagnosis and treatment of infertility, medical conditions contributing to infertility in developed countries were associated with female related infertility in 37 %, male infertility in 8% and both male and female infertility in 35% of the cases. Here, the most frequent causes of female infertility were: ovulatory disorders (25%), endometriosis (15%), pelvic adhesions (12%), tubal blockage (11%), other tubal abnormalities (11%), and hyperprolactinemia (7%) [13].

Infertility Causes Associated with Thyroid Disorders

In the case of idiopathic infertility, a higher prevalence of TAI has been reported when compared with fertile women [14, 15]. Additionally, an increased prevalence of TAI in women with polycystic ovary syndrome (PCOS) has been described. Possible underlying mechanisms may be polymorphisms of the PCOS-related gene for fibrillin 3 influencing the activity of TGF-β, a key regulator of immune tolerance. Together with lower TGF-β, low vitamin D levels and the high oestrogen-to-progesterone ratio these factors may contribute to autoimmunity [2, 16]. There are inconsistent reports with respect to women with endometriosis (by some considered as an autoimmune disease as such) having an increased prevalence of TAI, although this is not confirmed in other studies [5, 17].

Thyroid Disorders and Assisted Reproduction

Assisted Reproductive Technology (ART)/Ovarian Hyperstimulation (OH)

Once thyroid function is normal(ized) and lifestyle modifications and fertility counselling have not resulted in pregnancy, ART treatment can be proposed. Based on the fertility work up, different treatment strategies can be put forward. A first line treatment is intrauterine insemination (IUI), often used in couples with unexplained subfertility, cervical factor subfertility and male subfertility. If not successful, *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI) represents alternatives, especially in cases of male infertility. IUI is sometimes preceded by a stimulation procedure with oral clomiphene citrate (an antioestrogenic drug), or injections of gonadotrophins in order to optimize egg maturation and ovulation prediction. IVF/ICSI is always preceded by an intense two-step hormonal treatment. In a first step the natural cycle is suppressed to prevent premature luteinization. To achieve this, two different protocols can be used either with prolonged administration of gonadotropin-releasing hormone (GnRH) agonists (analogues) desensitizing the pituitary gland, or GnRH antagonists instantly blocking the pituitary gland. In a second step, OH takes place with hormones that mimic effects of FSH (and LH). Gonadotropin preparations (human menopausal

gonadotropin (hMG) or recombinant FSH trigger the development of several follicles simultaneously, and the administration is done under hormonal and ultrasound control, to avoid overstimulation. Finally, the ovulation is triggered with a hCG-injection [18]. During natural cycles, E2 levels rise up to 121 pg/ml in the periovulatory phase (compared with 55 pg/ml in the early follicular phase). This, however, does not usually lead to changes in serum TSH levels [19]. Likewise, treatment with clomiphene citrate (CC) leads to increases of E2 levels of up to 1200 pg/ml, without impact on thyroid function [20]. After OH with gonadotropins however, E2 levels increase to supraphysiologic levels of 2700 pg/ml usually only seen in the late second or third trimester of pregnancy. These high E2 levels may lead to a strain on the thyroid gland through several mechanisms. First, by an E2-mediated increase in thyroxin-binding globulin (TBG) causing a drop in free thyroxine (fT_4) concentration. Second, by a direct central action of E2 on the release of thyrotropin-releasing hormone (TRH). A third mechanism maybe an E2-related immunomodulation favouring the occurrence of TAI. Finally, recombinant FSH may act directly as one of the triggers responsible for the serum TSH increase [21].

The impact of OH on thyroid function may be diverse, ranging from no impact to an isolated increase in serum TSH or even to decreased serum fT_4 levels [3, 21]. This variability can be explained by the type of the pituitary downregulating drugs (GnRH agonists versus antagonists) used, different stimulation protocols implemented, the duration of TSH measurement follow-up after OH, and whether or not the presence of TAI was determined. Significantly increased serum TSH levels are more commonly described in women treated with GnRH antagonists compared to women treated with GnRH agonists. Overall, TSH values were >4 mIU/L in 41% of the women, of whom 89% were treated with an antagonist and 11% with an agonist. Increased TSH levels were independent from E2 levels and may be due to the presence of GnRH-receptors in thyrotrope cells of the adenohypophysis [22]. Although based on limited evidence, the latest guidelines propose to consider a serum TSH target of <2.5 mIU/L in the case of TAI positivity and/or in women already being treated with LT4 prior to OH /ART [3]. When TAI is present, thyroid function adaptation to OH is attenuated compared with that in women without TAI. Changes in thyroid function usually last until the first trimester of pregnancy, compared to an only temporary impact of several weeks in women without TAI [23].

A common complication of OH is the ovarian hyperstimulation syndrome (OHSS) which occurs in 20–30% of ART cycles. It is characterized by very high E2 levels (>5000 pg/ml) and has been associated with the development of severe hypothyroidism in TAI positive women [24, 25]. Women with hypothyroidism treated with LT4 (for pre-existing hypothyroidism) should increase their dosage before OH to anticipate the increased demands. Indeed, in a study by Busnelli *et al.*, the authors showed that 80% of hypothyroid women pregnant after OH and with TAI had to increase their dosage of LT4, within the first 7 weeks of pregnancy compared to 40% in those without TAI [26]. The degree of increase in serum TSH after OH during early pregnancy also depends on the prestimulation TSH level [27]. Another explanation for the different results relating to the impact of OH on thyroid function is the follow-up period after OH, as changes in TSH usually take at least 2 weeks to occur [28]. Therefore, it is recommended to determine thyroid function before and 1–2 weeks after OH [3].

Pregnancy Outcome Data

At present, the exact impact of preconceptional thyroid (dys)function and/or TAI on outcomes after ART is still subject of debate. In general, there seems to be no difference in terms of implantation, pregnancy, and delivery rates between women with preconception serum TSH levels between 2.5–5.0 mIU/L and <2.5 mIU/L [2, 10]. Little evidence exists on pregnancy outcomes in women with a preconception TSH level >4.0 mIU/L and without TAI [29]. More coherent study results are available on the impact of TAI, showing similar implantation rates, but more first trimester miscarriages compared with women without TAI [2, 3, 14]. A cause-effect relationship remains to be proven, and therefore, several hypotheses might explain this finding of which the most important ones are the higher mean ages in women with TAI, the higher risk to develop SCH during pregnancy, the higher prevalence of concomitant antibodies (e.g. anticardiolipin), the presence of impaired immune mechanisms (e.g. in the Treg cells), and an attenuated TSH response to high hCG levels during the first trimester of pregnancy [2, 30–32]. Finally, the concomitant presence of both a higher serum TSH and TAI seems to lead to lower live birth rates compared with the presence of either factor alone [33]. In a recent meta-analysis, including infertile women with and without TAI that were treated only with ICSI and did not receive LT4, miscarriage rates were comparable between both study groups. The systematic use of ICSI in case of TAI may therefore become a standard/additional treatment if these findings are confirmed in future studies [34].

Evidence and Recommendations on the Treatment of Thyroid Disorders Before ART

In a meta-analysis in TPO-ab positive women (in a large majority with a serum TSH >4.0 mIU/L) and undergoing ART, the effects of pre-ART LT4 treatment showed that it did not ameliorate clinical pregnancy rates, but increased live birth rates (RR 2.76) [35]. Evidence relating to the impact of LT4 therapy in euthyroid women with TPO-abs is limited to two retrospective and one prospective study [36]. In none of them, LT4 treatment improved pregnancy outcomes, but several criticisms to these studies are summarized in a recent paper [37]. One of the weak points in these studies is that the diagnosis of TAI is based often on the presence of TPO-abs only [34, 37]. However, positivity for Tg-abs alone was more frequent in infertile women compared with that of TPO-abs. The presence of Tg abs was associated with higher serum TSH levels and with higher miscarriage rates [38, 39]. Therefore, not testing for Tg-abs may lead to misdiagnosis of AIT, resulting in misinterpretation of findings and mismanagement of women before and during pregnancy [3, 40].

Insufficient evidence exists thus to determine whether LT4 therapy improves pregnancy outcomes following ART in euthyroid women with TPO-abs (TSH <4.0 mIU/L). However, administration of 25–50 mcg of LT4 may be considered given its potential benefits in comparison to its minimal risk, and the prevention to progressing to hypothyroidism once pregnancy is achieved. Alternatively, a 'wait and see' option may be preferred, and LT4 treatment can be started

when TSH >4 mIU/ L after OH or during early pregnancy [3, 10]. Women with SCH (typically TSH >4.0 mIU/L or above the upper limit of the local reference range) and overt hypothyroidism undergoing an ART treatment should be treated with LT4, and aiming a TSH level <2.5 mU/L before starting an OH procedure.

In women who achieve pregnancy following OH, TSH elevations should be treated according to the recommendations for pregnant women, and when no pregnancy is achieved, but mild TSH elevations were present, serum TSH measurements should be repeated after 2–4 weeks because they will probably normalize, especially when no TAI is present [3].

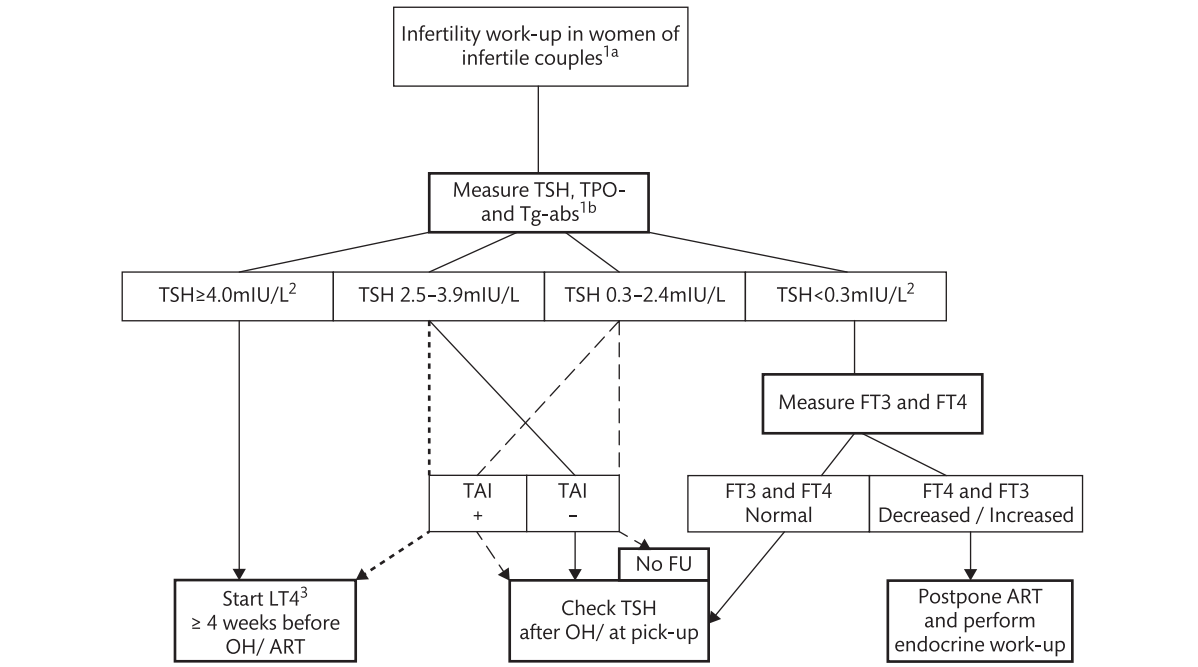
Of note, guidelines also point out that steroid treatment before ART is not beneficial and is not recommended in women with TPO-abs [3].

In Figure 9.2.1, an algorithm is proposed to screen and manage women of infertile couples.

Management of Thyroid Disorders Before Spontaneous Pregnancy

Hypothyroidism Before Pregnancy

The major physiologic thyroid change during pregnancy is that the total body T_4 pool must increase by approximately 40–50% to maintain a euthyroid state [2, 3, 30]. In healthy pregnant women, placental hCG is the major stimulus of maternal TH production, especially during the first trimester of pregnancy. However, in women with a history of thyroid disorders (post radioiodine, after thyroidectomy and/or inadequately treated with LT4), serum hCG will not stimulate the thyroid enough and (subclinical) hypothyroidism might occur during pregnancy. Increased demand in TH and/or LT4 takes place as early as 4–6 weeks of pregnancy. These requirements continue through 16–20 weeks of pregnancy, to stabilize thereafter until delivery. Treated women with hypothyroidism should be advised to contact their physician immediately upon a confirmed or suspected pregnancy, increase their dose of LT4 by 20–40% (by taking two additional tablets of their weekly LT4 dose) and undergo prompt testing and evaluation [41]. In case of de-novo overt hypothyroidism, women should receive full replacement doses of LT4 (1.6 mcg/kg/day). The importance and safety of LT4 treatment should be discussed with the patient to assure compliance [42]. Since T3 is degraded in the placenta, women treated with liothyronine or desiccated thyroid preparations should be switched to LT4 monotherapy [3, 4]. Because most women will be taking antenatal vitamin supplements (containing iron, calcium, and magnesium), they should be advised to leave a 4-hour gap between ingesting pregnancy supplements and T_4 replacement, to avoid LT4 malabsorption. Although based on limited evidence, traditionally, it is still recommended that hypothyroid women treated with LT4 should aim for a preconception serum TSH levels <2.5 mU/L, which is also the proposed threshold during pregnancy [3, 4, 42].



[1]when a choice has to be made for economical reasons [1a]measure only in case of ovarian causes or idiopathic infertility [1b]measure only TPO-abs
[2]or above / below the institutional cut-off
[3]dose depending on baseline TSH level and woman's body weight when ≥4.0mIU/L, and start 25 mcg when TSH 2.5–3.9 mIU/L; target a serum TSH level <2.5 mIU/L
TAI: thyroid autoimmunity / LT4: thyroid hormone / OH: ovarian hyperstimulation / ART: assisted reproductive technology / FU: Follow-up
Figure reproduced from Poppe et al. (ref 34) with permission from Oxford press

Figure 9.2.1 Algorithm to screen/manage thyroid disorders in women of infertile couples.

Reproduced with permission from Poppe K, Autin C, Veltri F, Kleynen P, Grabczan L, Rozenberg S, *et al.* Thyroid autoimmunity and intracytoplasmic sperm injection outcome: a systematic review and meta-analysis. J *Clin Endocrinol Metab* 2018.

Hyperthyroidism (Graves' Disease) Before Pregnancy

Graves' disease (GD) should be distinguished from gestational transient thyrotoxicosis (GTT) which is self-limiting requiring supportive treatment only but not antithyroid drugs (ATD) [3]. In women with GTT the typical features of GD (raised TSH-receptor antibodies (TRAbs), concentration goitre, family history and exophthalmos) are absent. In women with de-novo GD seeking pregnancy, three therapeutic options should be discussed: ATD, surgery and radioiodine therapy (RAI). The choice will depend on the preference of the patient, history of the disease, presence of a goitre, presence of high levels of TRAbs, and the timescale in which the couple hopes to conceive [43].

With regard to ATD, patients should be informed about the increased risk of birth defects associated with the use of either propylthiouracil (PTU) such as renal tract and head and neck anomalies or methimazole (MMI), including aplasia cutis, choanal atresia, trachea-oesophageal fistula, and dysmorphic facial features [44, 45]. However, the possibility of stopping ATD during the critical period of pregnancy (weeks 6–10) should be discussed. Finally, the preference for PTU before and during the first trimester of pregnancy, ATD titration (avoiding block-replace therapy), and the use of the lowest dose of ATD as possible should be highlighted. Concerning RAI therapy, the possible TRAbs increase after RAI therapy which may remain elevated for several years and the potentially long latency period before reaching euthyroidism should be mentioned [46]. The latter often leads to a delay in pregnancy (a minimum of 6 months and possibly longer) until a stable euthyroid state has been reached (with or without LT4 treatment). A pregnancy test should precede RAI therapy to exclude pregnancy. Thyroidectomy, may be indicated in case of contraindications/refusal of the two other therapeutic options, in case of relapse of GD and/or the presence of a large goitre or persistence of high maternal TRAbs (>3× upper reference for the assay) [43, 47]. After surgery, biochemical euthyroidism together with decreasing TRAbs should be confirmed before efforts to conceive commence.

Women already treated with ATD for GD should postpone pregnancy until a euthyroid state is reached, which should be verified by thyroid function in the normal reference range on two occasions over a period of at least 2 months and on stable therapeutic regimen. The use of contraception should be discussed, and pregnancy should be postponed if TRABs remain high and/or there is inadequate control of the hyperthyroidism. According to recent data, MMI should be switched to an equivalent PTU dose, at least 3 months before pregnancy to decrease the risk of congenital malformations due to MMI [45]. Women should be instructed to confirm pregnancy as soon as possible and contact their treating physician.

Screening in Daily Practice

There is insufficient evidence to recommend for or against universal screening for abnormal thyroid function preconception, with the exception of women with infertility, planning ART, a history of pregnancy loss, preterm delivery, or those known to have raised TPO-abs. Furthermore, it is unclear if only screening for thyroid dysfunction is recommended [3, 10]. Screening for TPO-abs may be of interest and indicated for several reasons. The prevalence of TPO-abs is around 10–15% in the general population and may identify women at risk for developing hypothyroidism during gestation or after OH, and be a predictive marker for the development of post-partum thyroiditis [3]. Furthermore, women with PCOS and idiopathic infertility are known to have a high prevalence of raised TPO-abs [5]. However, despite this rational basis, there are currently no data in terms of pregnancy outcomes that support such an approach. Nonetheless it is crucial that potential pregnancy wish is discussed with women undergoing treatment for hypo- or hyperthyroidism,

DISCLOSURE

No funding was received to support this work.

The authors declare no support from any organization in relation to the current work.

REFERENCES

1. Valdes S, Maldonado-Araque C, Lago-Sampedro A, *et al.* Population-based national prevalence of thyroid dysfunction in spain and associated factors: Di@bet.es study. *Thyroid*, 2017; **27**: 156–66.
2. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
3. Alexander EK, Pearce EN, Brent GA, *et al.* 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
4. Okosieme OE, Khan I, Taylor PN. Preconception management of thyroid dysfunction. *Clin Endocrinol (Oxf)*, 2018; **89**: 269–79.
5. Poppe K, Velkeniers B, Glinoer D. The role of thyroid autoimmunity in fertility and pregnancy. *Nat Clin Pract Endocrinol Metab*, 2008; **4**: 394–405.
6. Iravani AT, Saeedi MM, Pakravesh J, Hamidi S, Abbasi M. Thyroid autoimmunity and recurrent spontaneous abortion in Iran: a case-control study. *Endocr Pract*, 2008; **14**: 458–64.
7. Colicchia M, Campagnolo L, Baldini E, Ulisse S, Valensise H, Moretti C. Molecular basis of thyrotropin and thyroid hormone action during implantation and early development. *Hum Reprod Update*, 2014; **20**: 884–904.
8. Joshi JV, Bhandarkar SD, Chadha M, Balaiah D, Shah R. Menstrual irregularities and lactation failure may precede thyroid dysfunction or goitre. *J Postgrad Med*, 1993; **39**: 137–41.
9. Sturgis SH, Lerman J, Stanbury JB. The menstrual pattern in thyroid disease. *J Clin Endocrinol Metab*, 1952; **12**: 846–55.
10. Practice Committee of American Society for Reproductive M. Definitions of infertility and recurrent pregnancy loss. *Fertil Steril*, 2008; **90**: S60.
11. Veltri F, Belhomme J, Kleynen P, *et al.* Maternal thyroid parameters in pregnant women with different ethnic backgrounds: do ethnicity-specific reference ranges improve the diagnosis of subclinical hypothyroidism? *Clin Endocrinol (Oxf)*, 2017; **86**: 830–6.
12. Monteleone P, Parrini D, Faviana P, *et al.* Female infertility related to thyroid autoimmunity: the ovarian follicle hypothesis. *Am J Reprod Immunol*, 2011; **66**: 108–14.
13. Recent advances in medically assisted conception. Report of a WHO Scientific Group. *World Health Organ Tech Rep Ser*, 1992; **820**: 1–111.

14. van den Boogaard E, Vissenberg R, Land JA, *et al*. Significance of (sub)clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. *Hum Reprod Update*, 2011; **17**: 605–19.
15. Orouji Jokar T, Fourman LT, Lee H, Mentzinger K, Fazeli PK. Higher TSH levels within the normal range are associated with unexplained infertility. *J Clin Endocrinol Metab*, 2018; **103**: 632–39.
16. Gaberscek S, Zaletel K, Schwetz V, Pieber T, Obermayer-Pietsch B, Lerchbaum E. Mechanisms in endocrinology: thyroid and polycystic ovary syndrome. *Eur J Endocrinol*, 2015; **172**: R9–21.
17. Eisenberg VH, Zolti M, Soriano D. Is there an association between autoimmunity and endometriosis? *Autoimmun Rev*, 2012; **11**: 806–14.
18. Blockeel C, Engels S, De Vos M, *et al*. Oestradiol valerate pretreatment in GnRH-antagonist cycles: a randomized controlled trial. *Reprod Biomed Online*, 2012; **24**: 272–80.
19. Sawin CT, Hershman JM, Boyd AE, 3rd, Longcope C, Bacharach P. The relationship of changes in serum estradiol and progesterone during the menstrual cycle to the thyrotropin and prolactin responses to thyrotropin-releasing hormone. *J Clin Endocrinol Metab*, 1978; **47**: 1296–302.
20. Davis LB, Lathi RB, Dahan MH. The effect of infertility medication on thyroid function in hypothyroid women who conceive. *Thyroid*, 2007; **17**: 773–7.
21. Mintziori G, Goulis DG, Toulis KA, Venetis CA, Kolibianakis EM, Tarlatzis BC. Thyroid function during ovarian stimulation: a systematic review. *Fertil Steril*, 2011; **96**: 780–5.
22. Gizzo S, Noventa M, Quaranta M, *et al*. The potential role of GnRH agonists and antagonists in inducing thyroid physiopathological changes during IVF. *Reprod Sci*, 2016; **23**: 515–23.
23. Poppe K, Glinoer D, Tournaye H, *et al*. Impact of ovarian hyperstimulation on thyroid function in women with and without thyroid autoimmunity. *J Clin Endocrinol Metab*, 2004; **89**: 3808–12.
24. Poppe K, Unuane D, D'Haeseleer M, *et al*. Thyroid function after controlled ovarian hyperstimulation in women with and without the hyperstimulation syndrome. *Fertil Steril*, 2011; **96**: 241–5.
25. Poppe K, Glinoer D, Tournaye H, Devroey P, Velkeniers B. Impact of the ovarian hyperstimulation syndrome on thyroid function. *Thyroid*, 2008; **18**: 801–2.
26. Busnelli A, Somigliana E, Ferrari S, *et al*. The long-term impact of controlled ovarian hyperstimulation on thyroid function. *Endocr Pract*, 2016; **22**: 389–95.
27. Hammond KR, Cataldo NA, Hubbard JA, Malizia BA, Steinkampf MP. Gestational hypothyroidism: development of mild hypothyroidism in early pregnancy in previously euthyroid women. *Fertil Steril*, 2015; **103**: 1532–6 e1.
28. Reinblatt S, Herrero B, Correa JA, *et al*. Thyroid stimulating hormone levels rise after assisted reproductive technology. *J Assist Reprod Genet*, 2013; **30**: 1347–52.
29. Chai J, Yeung WY, Lee CY, Li HW, Ho PC, Ng HY. Live birth rates following in vitro fertilization in women with thyroid autoimmunity and/or subclinical hypothyroidism. *Clin Endocrinol (Oxf)*, 2014; **80**: 122–7.
30. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. *Nat Rev Endocrinol*, 2017; **13**: 610–22.
31. Thangaratinam S, Tan A, Knox E, Kilby MD, Franklyn J, Coomarasamy A. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. *BMJ*, 2011; **342**: d2616.
32. Busnelli A, Paffoni A, Fedele L, Somigliana E. The impact of thyroid autoimmunity on IVF/ICSI outcome: a systematic review and meta-analysis. *Hum Reprod Update*, 2016; **22**: 793–4.
33. Seungdamrong A, Steiner AZ, Gracia CR, *et al*. Eunice Kennedy Shriver National Institute of Child H, Human Development Reproductive Medicine N. Preconceptional antithyroid peroxidase antibodies, but not thyroid-stimulating hormone, are associated with decreased live birth rates in infertile women. *Fertil Steril*, 2017; pii: S0015-0282(17)31748-X.
34. Poppe K, Autin C, Veltri F, Kleynen P, *et al*. Thyroid autoimmunity and intracytoplasmic sperm injection outcome: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2018; doi: 10.1210/jc.2017-02633.
35. Velkeniers B, Van Meerhaeghe A, Poppe K, Unuane D, Tournaye H, Haentjens P. Levothyroxine treatment and pregnancy outcome in women with subclinical hypothyroidism undergoing assisted reproduction technologies: systematic review and meta-analysis of RCTs. *Hum Reprod Update*, 2013; **19**: 251–8.
36. Wang H, Gao H, Chi H, *et al*. Effect of levothyroxine on miscarriage among women with normal thyroid function and thyroid autoimmunity undergoing in vitro fertilization and embryo transfer: a randomized clinical trial. *JAMA*, 2017; **318**: 2190–98.
37. Poppe K, Veltri F, Autin C. Does levothyroxine improve pregnancy outcomes in euthyroid women with thyroid autoimmunity undergoing assisted reproductive technology? *Thyroid Res*, 2018; 11: 7.
38. Unuane D, Velkeniers B, Anckaert E, *et al*. Thyroglobulin autoantibodies: is there any added value in the detection of thyroid autoimmunity in women consulting for fertility treatment? *Thyroid*, 2013; **23**: 1022–8.
39. Liu H, Shan Z, Li C, et al. Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid*, 2014; **24**: 1642–9.
40. Lazarus J, Brown RS, Daumerie C, Hubalewska-Dydejczyk A, Negro R, Vaidya B. 2014 European Thyroid Association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J*, 2014; **3**: 76–94.
41. Yassa L, Marqusee E, Fawcett R, Alexander EK. Thyroid hormone early adjustment in pregnancy (the THERAPY) trial. *J Clin Endocrinol Metab*, 2010; **95**: 3234–41.
42. Taylor PN, Minassian C, Rehman A, *et al*. TSH levels and risk of miscarriage in women on long-term levothyroxine: a community-based study. *J Clin Endocrinol Metab*, 2014; **99**: 3895–902.
43. Kahaly GJ, Bartalena L, Hegedüs L, Leenhardt L, Poppe K, Pearce SH. 2018 European Thyroid Association guideline for the management of Graves' hyperthyroidism. *Eur Thyroid J*, 2018; **7**: 167–86.
44. Andersen SL, Olsen J, Laurberg P. Antithyroid drug side effects in the population and in pregnancy. *J Clin Endocrinol Metab*, 2016; **101**: 1606–14.
45. Seo GH, Kim TH, Chung JH. Antithyroid drugs and congenital malformations: a nationwide Korean cohort study. *Ann Intern Med*, 2018; **168**: 405–13.
46. Laurberg P, Wallin G, Tallstedt L, Abraham-Nordling M, Lundell G, Tørring O. TSH-receptor autoimmunity in Graves' disease after therapy with anti-thyroid drugs, surgery, or radioiodine: a 5-year prospective randomized study. *Eur J Endocrinol*, 2008; **158**: 69–75.
47. van Dijk MM, Smits IH, Fliers E, Bisschop PH. Maternal thyrotropin receptor antibody concentration and the risk of fetal and neonatal thyrotoxicosis: a systematic review. *Thyroid*, 2018; **28**: 257–64.

9.3

Thyroid Disease During Pregnancy

Tim I.M. Korevaar and Robin P. Peeters

Introduction

Adequate thyroid hormone (TH) availability is essential for a normal pregnancy and fetal development. During early pregnancy, an increase in maternal TH regulates the increased metabolic demand associated with pregnancy and ensures TH availability for the fetus which predominantly depends on the supply of maternal TH. As a consequence, untreated maternal hypothyroidism during pregnancy is associated with an increased risk of adverse pregnancy as well as child outcomes [1]. Although thyroid function is frequently assessed before and during pregnancy, accurate clinical assessment, and interpretation of maternal (and fetal) thyroid function during pregnancy is complicated by the maternal-placental-fetal interaction. Given this complexity, it remains difficult how and when to evaluate for thyroid dysfunction, as well as if and how to treat (mild) thyroid disease during pregnancy. This chapter discusses current insights on physiology, diagnosis, risk of adverse outcomes, and treatment options.

Thyroid Physiology

Pregnancy has clear effects on thyroid physiology. The maternal supply of TH to the fetus, an increase in thyroxine-binding globulin (TBG) concentrations under the influence of high maternal oestrogens, an increase in iodine clearance, and the degradation of TH by placental type 3 deiodinase (D3) all require increased TH production to ensure adequate TH availability for mother and fetus (Figure 9.3.1) [1]. In normal physiology, this increased demand for TH production is mediated by high concentrations of the pregnancy

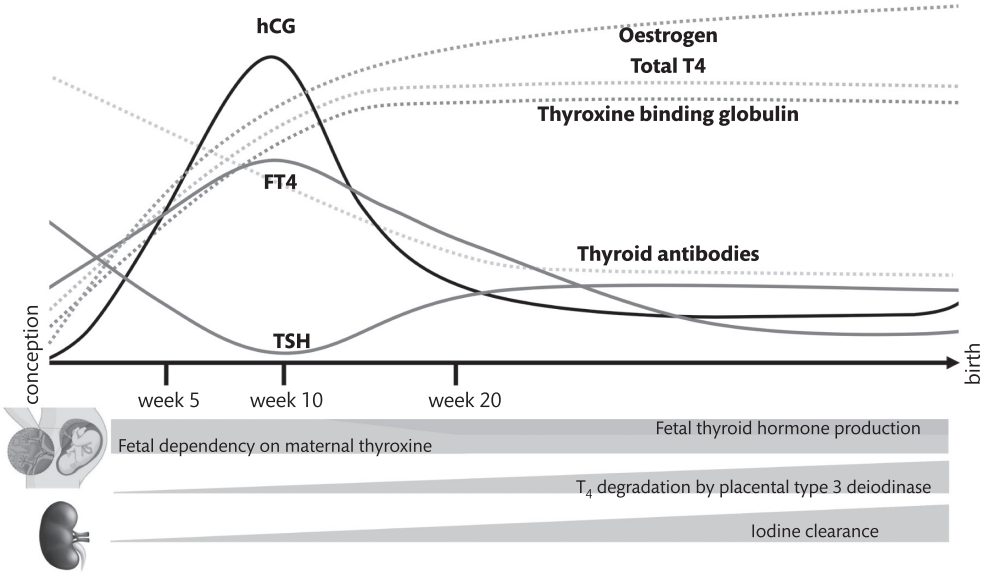


Figure 9.3.1 Gestational changes in thyroid function and related parameters.

hormone human chorionic gonadotrophin (hCG), a weak agonist of the TSH receptor, which stimulates the thyroid and thereby increases TH production (Figure 9.3.1) [2]. These pregnancy-specific changes and increased demand may expose pre-existing mild thyroid dysfunction. We further highlight the role of major gestational thyroid physiology concepts including iodine and thyroid autoimmunity next.

Iodine is an essential trace element that is a major component of TH and essential for its production. During pregnancy, an increase in demand for TH and an increase in renal iodine excretion necessitates a higher iodine intake to maintain adequate iodine availability. Severe maternal iodine deficiency leads to overt hypothyroidism, goitre, and offspring mental retardation (i.e. cretinism) [3]. Because of the crucial role of iodine and thyroid hormone in brain development, the World Health Organization recognizes iodine deficiency as 'the single most important preventable cause of brain damage'. The consequences of severe iodine deficiency are well established and include hypothyroidism, goitre, and child cretinism.

While prevention programs have eradicated severe iodine deficiency in most countries, many pregnant women still reside on areas of mild to moderate iodine deficiency (i.e. pregnant women in approximately 30% of all European countries still have suboptimal iodine status [4–7]). The effects of mild-to-moderate maternal iodine deficiency during pregnancy are less clear. Studies on thyroid function and iodine status have predominantly shown that a low iodine status is associated with a lower thyroid function and mild-to-moderate maternal iodine deficiency seems to impair neurocognitive function of children in some but not all studies. Although some international guidelines recommend the use of iodine supplements, such treatment decisions can only be made in light of local iodine status. Moreover, because the thyroid can store up to 3 months' worth of iodine, iodine status optimization should preferably be achieved before pregnancy. In line with this, a recent study showed that pre-pregnancy iodine deficiency is associated with a roughly 7-point lower child IQ [4]. Notably, also too much iodine intake is associated with lower thyroid function [5]; however, the effects of excessive iodine on offspring neurocognitive development remain largely unknown.

Thyroid hormone is crucial for normal fetal brain development because it regulates migration, proliferation, and differentiation of fetal neuronal cells, as well as synaptogenesis and myelination. In human fetuses, early neurogenesis starts from approximately 5 weeks post-conception and thyroid hormone receptors have been detected in the fetal brain from as early as 8 weeks gestation. Because the fetal thyroid gland is not functionally mature before week 18–20 of pregnancy, the fetus largely depends on the supply of maternal thyroxine during the early stages of intrauterine brain development. Also later in pregnancy, maternal thyroxine still contributes to fetal TH availability (Figure 9.3.1).

Thyroid autoimmunity, generally reflected by thyroperoxidase antibody (TPOAb) positivity, is the most important risk factor for thyroid dysfunction during pregnancy. During pregnancy, TPOAb positivity occurs in 5.6% to 22.1% of all women and is consistently associated with a higher risk of pregnancy complications such as miscarriage and premature delivery in different studies [6]. Another marker of thyroid autoimmunity, thyroglobulin antibody (TgAb) positivity, can occur independent of the presence of TPOAbs [7]. However, studies on isolated TGAb positivity and clinical outcomes are sparse. The pathophysiological mechanism through which thyroid autoimmunity is associated with a higher risk of adverse outcomes remains to be elucidated. Recent studies from our group have shown that the thyroidal response to hCG stimulation is severely impaired in TPOAb positive women (Figure 9.3.2). In addition, we showed that particularly those TPOAb positive women that had a low fT_4 despite a high hCG (i.e. an impaired thyroid functional capacity) had a higher risk of premature delivery suggesting that TPOAb positivity results in adverse pregnancy outcomes via alterations in gestational thyroid function (also see section on thyroid autoimmunity, later on).

Diagnosis of Thyroid Disease

The major changes in thyroid physiology just described necessitate the use of pregnancy-specific reference ranges to diagnose gestational thyroid disease entities. Although many centres in the past used a fixed upper limit for TSH of 2.5 mU/L in the first trimester and 3.0 mU/L in the second or third trimester, data from the last ten years have shown that these cut-offs are too low and may cause considerable overdiagnosis [8]. Therefore, a stepwise approach should be used to define the best possible reference range for each specific institution. Lab-specific TSH and fT_4 reference ranges are considered the gold standard and, when available, are the reference range of choice within a hospital infrastructure. A lab-specific reference range for TSH and fT_4 can be calculated as the 2.5th to 97.5th percentiles for a group of at least 400 women that are free of thyroid function interfering factors (major disease, TPOAb positivity, thyroid (interfering) drug use, higher physiological hCG (e.g. twin pregnancies and/or IVF treatment)) [8]. If it is not possible to obtain local lab-specific reference ranges, lab-specific reference ranges can be adopted from a centre that uses the same assay and has a patient population with similar ethnic background and iodine status. Finally, if appropriate reference ranges cannot be obtained via either of the previous steps, a fixed upper limit for TSH of 4.0 mU/L could be used. Notably, these recommendations are predominantly relevant for thyroid function tests obtained during the first trimester (from 6 weeks onwards) or second trimester. Data on the third trimester remain sparse.

To be able to distinguish between overt and subclinical thyroid disease, which has direct clinical consequences, measurement of fT_4 is required. With current knowledge, it is however not possible to identify an evidence-based fixed lower limit for fT_4 because published reference ranges differ widely and are very assay-specific. Therefore, the lower range for fT_4 is best defined using a lab-specific or 'adopted' reference range approach. Commonly used fT_4 immunoassays have been reported to be less accurate due to a shift in TH binding proteins during pregnancy [9–12]. This shift mostly affects fT_4 measurements during the third trimester and therefore is not likely to be relevant for the majority of thyroid function tests, which typically are taken in the first half of pregnancy. Furthermore, the correlation of fT_4 concentrations between assays is very high, which means that every assay can be used to adequately identify women with an abnormal fT_4 if pregnancy-specific and assay-specific reference ranges are used [13]. Previously, it was suggested that total T_4 concentrations may be used to assess thyroid dysfunction during

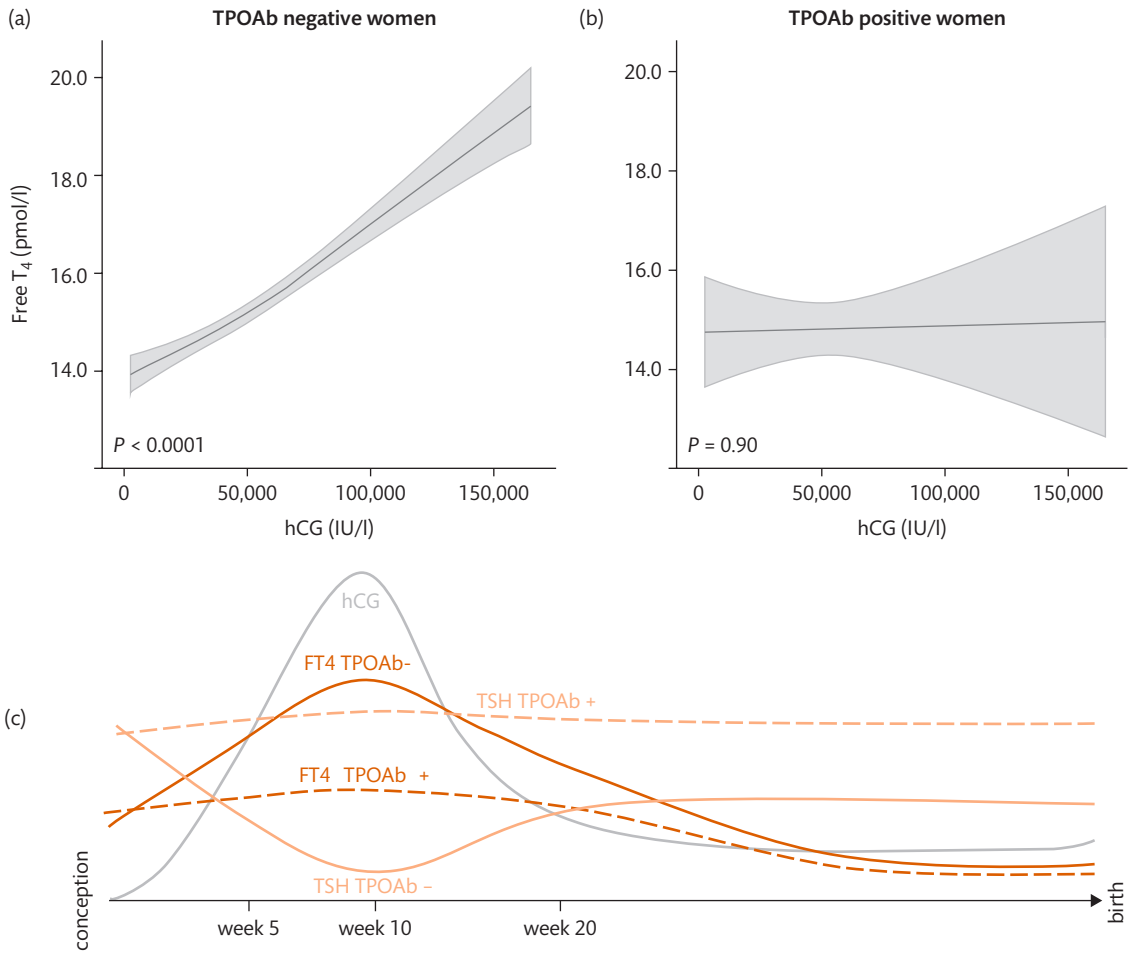


Figure shows association of hCG with FT4 during the first half of pregnancy inTPOAb negative women (A) and TPOAb positive women (B) using cross sectional data from 5,924 women participating in the Generation R study, and a translation of these data through out pregnancy (C).

Figure 9.3.2 The thyroidal response to hCG stimulation in TPOAb negative and TPOAb positive women.

pregnancy by raising the lower limit of the non-pregnancy reference range to 150% [11]. However, total T_4 is only a very crude estimate of the fraction of available TH because >99% of total T_4 is bound to TH binding proteins. This makes total T_4 concentrations highly dependent on changes in TBG [14, 15]. Furthermore, because early pregnancy total T_4 concentrations are not associated with adverse pregnancy or child outcomes [15, 16] fT_4 seems a more clinically relevant measurement of thyroid function [15].

The identification of women with gestational thyroid disease is complicated by the fact that classical clinical symptoms of hypo- and hyperthyroidism may also occur during normal pregnancy. Moreover, the majority of women with biochemical thyroid dysfunction, particularly subclinical forms, are asymptomatic [17]. Studies show that clinical characteristics such as maternal obesity, smoking, ethnicity, and gestational age at presentation are associated with the risk of gestational thyroid disease. However, these clinical characteristics altogether do not accurately predict the risk of thyroid dysfunction in the general population [18].

Screening for thyroid disease during pregnancy has been a topic of debate. On one hand, universal screening would identify a small group of women that would certainly benefit from treatment, e.g. overt hypothyroidism [prevalence ~0.3%] or Graves' disease [prevalence ~0.05%] (see section on overt thyroid disease). However, it remains unknown how many (asymptomatic) women with such disease are currently missed. In addition, screening can also identify women for which there is some evidence that treatment may be beneficial based on small randomized controlled trials, e.g. TPOAb positive women with a TSH above 4.0 mU/L [prevalence ~1.1%] (see part on thyroid antibodies) [19, 20]. On the other hand, there is no good evidence that treatment has beneficial effects for the majority of women that have an abnormal screening result (subclinical hypothyroidism [~3.5%], hypothyroxinaemia [~2.3%], subclinical hyperthyroidism [~2.5%] or overt hyperthyroidism [~1.5%]) (see respective subheadings as follows) which may lead to overtreatment in a substantial proportion of women. Experts agree that targeted screening for thyroid disease during pregnancy seems warranted for women at a high-risk of adverse pregnancy outcomes (for example women with a history of gestational thyroid disease, those with repeated miscarriages or adverse pregnancy outcomes such as pre-eclampsia or preterm delivery). However, the decision to universally screen for thyroid disease during pregnancy currently remains at the discretion of the physician's interpretation of current evidence.

Hypothyroidism

Overt Hypothyroidism

Overt maternal hypothyroidism (elevated TSH with low fT_4) during pregnancy occurs in roughly 0.2 to 0.6% of pregnant women [1, 8, 21]. It has consistently been associated with a higher risk of adverse pregnancy complications including premature delivery, low birth weight, miscarriage, pre-eclampsia [22], as well as negative effects on fetal neurodevelopmental outcomes [23]. A large case–control study demonstrated a seven-point lower child IQ in children born to untreated hypothyroid women compared to euthyroid controls, as well as a delay in motor skill development, language development and attention at 7–9 years of age [23].

Subclinical Hypothyroidism

Subclinical maternal hypothyroidism (elevated TSH with normal fT_4) is much more prevalent than overt hypothyroidism, occurring in roughly 3.5% of all pregnancies. While some studies have reported prevalences up to 18%, it has become apparent that such numbers represent an overestimation due to the use of an inappropriate TSH cut-off (see section on reference ranges). Autoimmune thyroid disease is the main risk factor for subclinical hypothyroidism, and TPOAb positivity occurs in roughly one third of pregnant women with subclinical hypothyroidism [24]. Multiple studies have demonstrated the negative consequences of maternal subclinical hypothyroidism on pregnancy outcomes, as it has been associated with a higher risk of pregnancy loss, placental abruption, premature delivery, and neonatal death [25–27]. Although the risks of these adverse outcomes are generally much lower than in overt disease, the high prevalence makes it a relevant public health issue. In contrast to its association with various adverse pregnancy outcomes, there is virtually no evidence that subclinical hypothyroidism is associated with adverse neurobehavioral outcomes in the offspring. Recent studies have shown that particularly the combination of subclinical hypothyroidism with TPOAb positivity is associated with a higher risk of adverse outcomes such as miscarriage, gestational diabetes and premature delivery, and that this higher risk is also present when TSH concentrations are in the high-normal range (i.e. for TPOAb positive women with TSH >2.5 mU/L) [28–31].

Hypothyroxinaemia

Hypothyroxinaemia (low fT_4 with normal TSH) has been associated with adverse neurobehavioral outcomes in the offspring in multiple studies from different cohorts [30, 32–34].

Initially, hypothyroxinaemia was considered as a pregnancy-specific disease entity that reflects a state of mild iodine deficiency. However, as hypothyroxinaemia also occurs in iodine sufficient areas and since (f)T_4 concentrations typically do not increase following iodine supplementation, it is likely that a much more multifactorial and pregnancy-specific pathophysiology underlies hypothyroxinaemia.

While subclinical hypothyroidism is associated with various adverse pregnancy outcomes, hypothyroxinaemia has predominantly been associated with adverse neurobehavioral outcomes in the offspring [30, 32, 34]. In 3659 mother–child pairs from a prospective birth cohort, hypothyroxinaemia was associated with a 1.8-fold increased risk of expressive language delay at both 18 and 30 months and a 2-fold increased risk of non-verbal cognitive delay [35]. Follow-up data subsequently revealed that maternal hypothyroxinaemia was associated with a 4.3 point lower non-verbal IQ at 6 years of age [36]. When various fT_4 cut-offs were assessed, low maternal fT_4 up to the tenth percentile was associated with a 1.5–3.8-point lower IQ [37]. Similar findings were also reported by Julvez *et al.* who reported a dose-response relationship of low maternal fT_4 with child mental score [34]. Hypothyroxinaemia has also been associated with other postnatal markers of intrauterine brain development, including a 2.6-fold higher risk of clinical autistic symptoms [38], a 1.7-fold higher risk of schizophrenia [39], as well as a higher risk of offspring ADHD, a slower reaction time, suboptimal school performance and lower grey matter and, cortex volume [37, 40–42]. As discussed already, the value of a fT_4 measurement during pregnancy has been questioned. However, a recent individual participant based meta-analysis of three different cohorts with three different fT_4 assays, showed a consistent association of hypothyroxinaemia during early pregnancy with offspring IQ [43]. This suggests that fT_4 is a sensitive marker of thyroid function during early pregnancy independent of the assay used [15].

Should (Subclinical) Hypothyroidism During Pregnancy Be Treated?

When considering treatment with levothyroxine during pregnancy it is important to consider the potential harms, particularly of overtreatment. Although the principle is similar to the potential harms of antithyroid drug treatment in the setting of Graves' hyperthyroidism during pregnancy (see section on hyperthyroidism), the potential for levothyroxine treatment-related harm is much less recognized. Various studies show that high maternal thyroid function is associated with a higher risk of low birth weight, pre-eclampsia and lower child IQ which indicates that levothyroxine treatment aiming for a high-normal thyroid function may carry the risk of overtreatment. While a full dose of levothyroxine based on patient weight in the range of 1.3–1.8 μg/kg is a good starting dosage for women with overt hypothyroidism, for women with subclinical disease a fixed starting dosage of 25–50 μg and titration after approximately 2 weeks is recommended [44].

Overt Hypothyroidism

There are no randomized controlled trials of levothyroxine treatment for overt hypothyroidism during pregnancy. However, given the well-known detrimental effects of untreated overt hypothyroidism on various adverse outcomes and the fact that there is no evidence of a higher risk of pregnancy complications in women with adequately treated hypothyroidism, there is general consensus is that overt hypothyroidism during pregnancy should be treated as early as possible and that it is unethical to perform a placebo controlled trial [44]. This is supported by the fact that in the case–control study showing a seven-point reduction in IQ among children born to untreated hypothyroid women, there were no deficits in children born from mothers who received levothyroxine treatment later in pregnancy [23]. Levothyroxine is the treatment of

choice, considering the important role of maternal T_4 for fetal brain development.

Subclinical Hypothyroidism

Similarly to overt hypothyroidism, there is a lack of data on the effects of treatment for subclinical hypothyroidism. A recent observational study by Maraka *et al.* found that treatment of subclinical hypothyroidism (defined by a TSH between 2.5 and 10 mU/L, fT_4 >0.8 ng/dl and/or total thyroxine >7.5 mcg/dl) with a median dosage of 50 μg (IQR 25–62.5) was associated with lower pregnancy loss (10.6% vs. 13.5%; OR 0.62 [0.48–0.82]) but a higher rate of premature delivery (7.1% vs. 5.2%; OR 1.60 [1.14–2.24]). Interestingly, subsequent stratification revealed that the beneficial effect on pregnancy loss was dependent on the TSH concentrations at baseline and predominantly present in women with TSH concentrations above 4.0 mU/L. No data on TPOAbs concentrations were available, whereas it has been shown that particularly the combination of subclinical hypothyroidism with TPOAb positivity is associated with a higher risk of adverse outcomes (see below in subsection on thyroid autoimmunity). A recent randomized controlled trial from Iran showed that treatment of subclinical hypothyroidism (defined using a TSH cut-off of <4.0 mU/L) with levothyroxine was associated with a 61% lower risk of premature delivery (RR 0.39 [0.15–0.98]) [45]. However, secondary analyses from an American randomized controlled trial could not identify any beneficial effects of levothyroxine treatment in women with subclinical hypothyroidism (LT4 group 9% versus 11% in placebo group, $P = 0.44$) [46]. The main difference between these two studies was the timing of treatment, with the Iranian study starting treatment during week 12 and the American study starting treatment at week 18. This difference suggests that levothyroxine treatment after the critical timeframe of early pregnancy may have no benefit, but further studies are needed to substantiate this. Studies investigating the effects of levothyroxine treatment in women with subclinical hypothyroidism on offspring neurocognition are discussed in the section on hypothyroxinaemia (see next).

Hypothyroxinaemia

Two randomized controlled trials have so far assessed the effects of treatment of mild thyroid dysfunction on offspring IQ. In the CATS trial, women were randomized to no screening or screening and subsequently treated with 150 μg levothyroxine if the TSH concentration was above the 97.5th percentile and/or the fT_4 concentration was below the <2.5th percentile [47]. There was no difference between the 390 children of treated mothers and 404 children of untreated mothers in mean IQ or the proportion of children with an IQ below 85 (also not for the low fT_4 or high TSH groups separately) [47]. Various arguments have been proposed for this negative finding, namely that treatment started too late (median 13 weeks), that IQ cannot be assessed reliably at age 3 and that the loss to follow-up was too high (24%). Alternatively, the relatively high levothyroxine dosage of 150 μg may have contributed to the lack of a net beneficial effect since recent evidence shows that high fT_4 concentrations are associated with suboptimal offspring neurodevelopment similar to the extent of low fT_4 concentrations [37].

The second RCT is a recent multicentre study from the US, in which 526 women with hypothyroxinaemia and 677 with subclinical hypothyroidism were randomized to receive either placebo or levothyroxine treatment (50 or 75 μg) and children underwent IQ testing at the age of 3 or 5 years (92.3% assessed at 5 years). The sample size of this study was defined based on the number of randomized women needed to detect a statistically significant difference of 5 IQ points (see trial registration NCT00388297). However, since the start of this trial, prospective cohort studies have shown that mild maternal thyroid dysfunction is associated with a 3 to 4 IQ point difference [48]. This trial showed that levothyroxine treatment is associated with a median increase of 3 IQ points in the offspring of women with subclinical hypothyroidism and hypothyroxinaemia, a result that failed to reach statistical significance based on the power calculations [49]. This increase in IQ was seen despite a relatively late start of treatment and is in line with what is expected based on observational data [46]. For this reason, further studies are still required to elucidate the effects of levothyroxine treatment in women with mild thyroid dysfunction on offspring neurodevelopment in an adequately powered RCT. Based on current knowledge, it is not possible to advise for or against levothyroxine treatment for women with subclinical disease. Individual risk factors such as a history of pregnancy complications or other comorbidities as well as the potential risk for overtreatment should be taken into account during individualized clinical decision-making.

Overt Hyperthyroidism

During pregnancy, the underlying cause can roughly distinguish overt hyperthyroidism in two major subtypes. On one hand, hyperthyroidism can be transient when it is the consequence of very high hCG concentrations (classically referred to as gestational hyperthyroidism or gestational transient thyrotoxicosis). On the other hand, overt hyperthyroidism can be related to an underlying pathological process as in Graves' disease (newly diagnosed or pre-existing but suboptimally controlled) or due to autonomous TH production (e.g. multinodular toxic goitre or toxic adenoma(s)). The incidence, risk of adverse pregnancy outcomes and the expected treatment differ widely between these two subtypes. Gestational hyperthyroidism is caused by high hCG concentrations, which typically peak at the tenth week of pregnancy, but can be high enough anywhere between the eighth to fifteenth week of pregnancy. Gestational hyperthyroidism has an estimated prevalence of1.5% but it is seen in up to 50% of women with twin pregnancies, IVF pregnancies and/or women with hyperemesis gravidarum. Large observational studies show that gestational hyperthyroidism is associated with a lower birth weight and an up to 2-fold higher risk of pre-eclampsia [50]. Nevertheless, only supportive treatment is advised because the potential harmful side-effects of antithyroid drugs (ATD; see also next) do not weigh-up against any hypothetical treatment benefit. Follow-up assessments could include measurements of TSHRAb (to exclude Graves' disease) and thyroid function test to monitor the transient nature of the disease, additional measurement of total T_4 or T_3 is not beneficial given the changes in TH binding globulins described above. However, there is some evidence that an additional hCG measurement may help to distinguish gestational hyperthyroidism from pathological forms [51]. Pathological forms of hyperthyroidism

during pregnancy are more rare, with estimated prevalence rates of 0.5–1.3% for pre-existing Graves' disease, 0.05% for new onset Graves' disease, and 0.1% for autonomous TH production [52, 53]. These types of hyperthyroidism are associated with a higher risk of pre-eclampsia, preterm birth, low birth weight and cases of maternal heart failure have been reported [52, 54–60]. One case–control study on 208 hyperthyroid women (89.4% diagnosed before pregnancy, 95% received treatment) and 403 matched controls [60] showed that hyperthyroid women had a 3.9-fold higher risk of pre-eclampsia, a 2.2-fold higher risk of fetal growth restriction, a 1.7-fold higher risk of preterm birth and a 3.6-fold increased risk of induction of labour [60]. Observational studies show that the risk of such adverse outcomes is lower when pathological hyperthyroidism is biochemically well-controlled, while the onset of disease (before or during pregnancy) is of less importance.

Specifically for Graves' disease, the different abilities of thyroid-related factors and drugs to cross the placenta (Figure 9.3.3) form the cornerstone for clinical assessment and treatment decision-making. When considering treatment, the potential harms associated with ATD are important to take into account particularly during early pregnancy. The use of propylthiouracil (PTU) during early pregnancy is associated with a higher risk of birth defects such as preauricular sinus/cysts and hydronephrosis and cases of maternal liver injury have been reported [61, 62]. The use of methimazole or carbimazole during early pregnancy has been associated with a higher risk of birth defects such as choanal atresia, aplasia cutis and omphalocele and cases of agranulocytosis are reported [42, 43]. To limit the potential adverse outcomes associated with ATD, several pregnancy-specific ATD strategies can be applied.

First of all, in women of childbearing age with hyperthyroidism, the option of definitive treatment before pregnancy should be discussed. When radioactive iodine is chosen as a definitive treatment, the fact that with TSHRAbs can flare-up and remain high up to 1.5 years after treatment with radioactive iodine but not with other treatment modalities should be taken into account.

Secondly, if the patient prefers ATDs or is already pregnant while on ATDs, therapy should consist of monotherapy with the lowest possible dose, as opposed to a block and replace therapy. As pregnancy progresses, the maternal immune response becomes less active due to a more liberal maternal immune tolerance, this generally leads to a reduction in maternal TSHrAbs (Figure 9.3.1) and thus often necessitates a reduction of the ATD dosage during the (late) second trimester [63, 64]. Large population studies have indicated that the use of PTU during early pregnancy is associated with a slightly lower risk of adverse outcomes, but also with less severe fetal anomalies, compared to methimazole [61]. Therefore, women receiving methimazole who are in need of continuing therapy during pregnancy should be switched to PTU as early as possible. This switch may be beneficial during the critical window of fetal organogenesis lasting up until the sixteenth week of pregnancy.

Third, when Graves' hyperthyroidism is biochemically controlled with a stable low dose of methimazole (5–10 mg) or PTU (100–200 mg) before pregnancy, the risk of relapse is relatively low and will most likely occur after a few months. Therefore, it can be considered

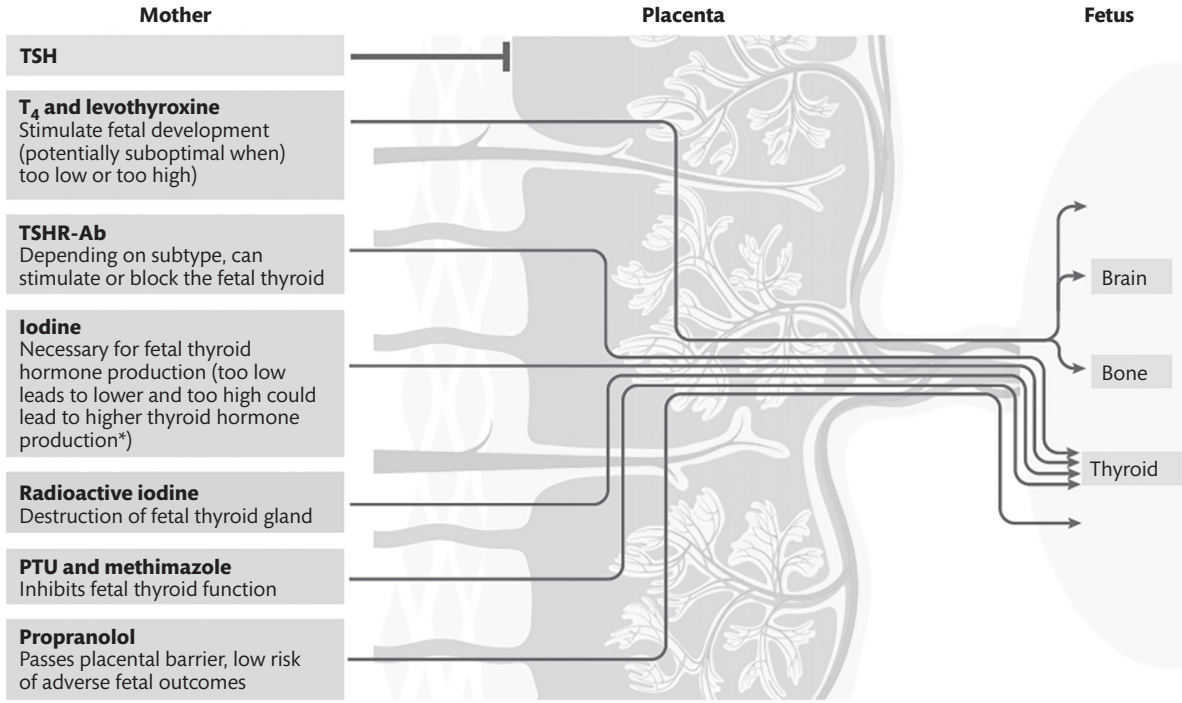


Figure 9.3.3 Thyroid-related substances and crossing of the placental barrier. The figure graphically depicts thyroid-related drugs and substances capable of affecting the function of the fetal thyroid gland as well as other tissues crossing the placenta. Of note, the fetus cannot escape the Wolff–Chaikow effect, that is, a phenomenon in which low iodine concentrations lead to decreased thyroid hormone production and high thyroid hormone concentrations lead to increased thyroid hormone production, until after roughly the 36th week of pregnancy, at which time high iodine concentrations are likely to cause low thyroid hormone production. TSHR-Ab, TSH receptor antibody; PTU, propylthiouracil.

to stop the treatment upon the wish to become pregnant or at the time of the first positive pregnancy test (in case of non-planned pregnancy), and monitor thyroid function closely (i.e. every 1–2 weeks) [65].

Finally, to assess disease severity but also the risk of neonatal Graves' hyperthyroidism, an early pregnancy measurement of TSHRAbs should be performed. If TSHRAbs are elevated, also the fetus should be monitored from mid-pregnancy onwards via fetal heart rate auscultation, fetal growth, and/or fetal thyroid ultrasound. Together with the maternal thyroid function, these measures are a reflection of the fetal TH status [66, 67].

Thyroid Autoimmunity

As a reflection of thyroid autoimmunity, thyroid antibodies (thyroglobulin antibodies TgAbs and thyroperoxidase antibodies (TPOAbs)) are an important risk factor for thyroid dysfunction during pregnancy. While both antibodies are a reflection of thyroid autoimmunity and their occurrence often overlaps, only TPOAbs seem to have clinical value because of a more prominent association with thyroid function and adverse pregnancy outcomes. It has consistently been shown that TPOAb positive women have higher TSH concentrations, lower fT_4 concentrations and also a higher risk of thyroid dysfunction during pregnancy [68, 69]. In general, TPOAbs are measured and then classified into positive or negative based on cut-offs provided by the manufacturer of the different assays. However, an individual participant data meta-analysis of three Dutch birth cohorts recently showed that there is a continuous association of TPOAb concentrations with TSH and fT_4 and that currently used cut-offs may fail to identify relevant thyroid autoimmunity [70]. It is well-recognized that thyroid autoimmunity affects the functional capacity of the thyroid. In line with this, we recently showed that TPOAb positive women have an impaired thyroidal response to hCG stimulation (**Figure 9.3.2**). These and other data indicate that TPOAb positive women lack the hCG-mediated increase in thyroid function during early pregnancy which may lead to a lower TH availability (area under the curve) during pregnancy. This may be the mechanism through which thyroid autoimmunity is associated with adverse outcomes since TPOAb positivity is associated with a higher risk of miscarriage and premature delivery, especially in combination with a higher TSH [70]. Two clinical trials have shown that LT4 treatment in TPOAb positive women markedly decreases the risk of miscarriage and/or premature delivery [52, 53]. Negro *et al.* randomized 115 TPOAb positive women to receive LT4 at a mean gestational age of 10 weeks with a dosage based on their TSH (<1 mU/L: 0.5 μg/kg·d; 1–2 mU/L: 0.75 μg/kg·d; >2 mU/L or TPOAbs >1500 kIU/L: 1 μg/kg·d) [20] In this study, LT4 treatment markedly reduced the rate of late miscarriage (from 13.8% versus 3.4%) and of premature delivery (22.4% vs. 7.0%) [20]. Using the same protocol, Nazarpour *et al.* randomized 131 TPOAb positive women to receive LT4 at a mean gestational age of 11 weeks [71]. This study also found that LT4 reduced the rate of premature delivery (23.7% versus 7.1%) but there was no beneficial effect of LT4 on the rate of miscarriage (3.6% vs. 3.4%) [71]. In addition, Nazarpour *et al.* stratified their analyses based on the TSH concentration at presentation and found that the benefit of LT4 treatment in reducing preterm delivery was mainly present in the group with TSH concentrations >4.0 mU/L (29.4% vs. 5.3%, P <0.01) compared with women with TSH levels <4.0 mU/L (16.7% vs. 11.1%, P = 0.69). These findings fit with the synergistically higher risk of adverse outcomes in women with both TPOAb positivity and a high TSH [29–31, 72]. Taken together, TPOAb positivity is the most important risk factor for thyroid dysfunction during pregnancy at least partly through impairing the thyroidal response to hCG stimulation. Therefore, TPOAbs are an important tool to improve the interpretation of TSH and, to a lesser extent fT_4 concentrations, during pregnancy.

Conclusions

A detailed understanding of pregnancy-specific thyroid physiology can help to interpret gestational thyroid function measurements. While thyroid disease should optimally be defined according to lab-specific reference ranges, the clinical assessment of whether to start treatment or not remains complicated. In part, this is because the risk of adverse outcomes seems to be related to other factors that influence thyroid function (i.e. TPOAb positivity, hCG). Another equally important reason is the lack of adequate randomized controlled trials that investigate the effects of levothyroxine treatment, preferably during early pregnancy. The potential risks associated with LT4 (over)treatment during pregnancy should be taken into account during clinical decision-making.

REFERENCES

1. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
2. Hershman JM. The role of human chorionic gonadotropin as a thyroid stimulator in normal pregnancy. *J Clin Endocrinol Metab*, 2008; **93**: 3305–6.
3. Zimmermann MB. The effects of iodine deficiency in pregnancy and infancy. *Paediatr Perinat Epidemiol*, 2012; **26** Suppl 1: 108–17.
4. Robinson SM, Crozier SR, Miles EA, *et al.* Preconception maternal iodine status is positively associated with IQ but not with measures of executive function in childhood. *J Nutr*, 2018; **148**: 959–66.
5. Shi X, Han C, Li C, Mao J, *et al.* Optimal and safe upper limits of iodine intake for early pregnancy in iodine-sufficient regions: a cross-sectional study of 7190 pregnant women in China. *J Clin Endocrinol Metab*, 2015; **100**: 1630–8.
6. De Leo S, Pearce EN. Autoimmune thyroid disease during pregnancy. *Lancet Diabetes Endocrinol*, 2018; **6**: 575–86.
7. Unuane D, Velkeniers B, Anckaert E, *et al.* Thyroglobulin autoantibodies: is there any added value in the detection of thyroid autoimmunity in women consulting for fertility treatment? *Thyroid*, 2013; **23**: 1022–8.
8. Medici M, Korevaar TI, Visser WE, Visser TJ, Peeters RP. Thyroid function in pregnancy: what is normal? *Clin Chem*, 2015; **61**: 704–13.
9. Anckaert E, Poppe K, Van Uytfanghe K, Schiettecatte J, Foulon W, Thienpont LM. FT4 immunoassays may display a pattern during pregnancy similar to the equilibrium dialysis ID-LC/tandem MS candidate reference measurement procedure in spite of susceptibility towards binding protein alterations. *Clinica Chimica Acta*, 2010; **411**: 1348–53.

10. Berta E, Samson L, Lenkey A, *et al*. Evaluation of the thyroid function of healthy pregnant women by five different hormone assays. *Pharmazie*, 2010; **65**: 436–9.
11. Lee RH, Spencer CA, Mestman JH, *et al*. Free T4 immunoassays are flawed during pregnancy. *Am J Obstet Gynecol*, 2009; **200**: 260.e1–6.
12. Kahric-Janicic N, Soldin SJ, Soldin OP, West T, Gu J, Jonklaas J. Tandem mass spectrometry improves the accuracy of free thyroxine measurements during pregnancy. *Thyroid*, 2007; **17**: 303–11.
13. Bliddal S, Feldt-Rasmussen U, Boas M, *et al*. Gestational age-specific reference ranges from different laboratories misclassify pregnant women's thyroid status: comparison of two longitudinal prospective cohort studies. *Eur J Endocrinol*, 2014; **170**: 329–39.
14. Glinoer D, de Nayer P, Bourdoux P, *et al*. Regulation of maternal thyroid during pregnancy. *J Clin Endocrinol Metab*, 1990; **71**: 276–87.
15. Korevaar TI, Chaker L, Medici M, *et al*. Maternal total T4 during the first half of pregnancy: physiologic aspects and the risk of adverse outcomes in comparison with free T4. *Clin Endocrinol (Oxf)*, 2016; **85**: 757–63.
16. Oken E, Braverman LE, Platek D, Mitchell ML, Lee SL, Pearce EN. Neonatal thyroxine, maternal thyroid function, and child cognition. *J Clin Endocrinol Metab*, 2009; **94**: 497–503.
17. Pop VJ, Broeren MA, Wiersinga WM, Stagnaro-Green A. Thyroid disease symptoms during early pregnancy do not identify women with thyroid hypofunction that should be treated. *Clin Endocrinol (Oxf)*, 2017; **87**: 838–43.
18. Korevaar TI, Nieboer D, Bisschop PH, *et al*. Risk factors and a clinical prediction model for low maternal thyroid function during early pregnancy: two population-based prospective cohort studies. *Clin Endocrinol (Oxf)*, 2016; **85**: 902–9.
19. Nazarpour S, Ramezani Tehrani F, *et al*. Effects of levothyroxine treatment on pregnancy outcomes in pregnant women with autoimmune thyroid disease. *Eur J Endocrinol*, 2017; **176**: 253–65.
20. Negro R, Formoso G, Mangieri T, Pezzarossa A, Dazzi D, Hassan H. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. *J Clin Endocrinol Metab*, 2006; **91**: 2587–91.
21. Pop V, Broeren M, Wiersinga W. The attitude toward hypothyroidism during early gestation: time for a change of mind? *Thyroid*, 2014; **24**: 1541–6.
22. van den Boogaard E, Vissenberg R, Land JA, *et al*. Significance of (sub)clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. *Hum Reprod Update*, 2011; **17**: 605–19.
23. Haddow JE, Palomaki GE, Allan WC, *et al*. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*, 1999; **341**: 549–55.
24. Casey BM, Dashe JS, Wells CE, McIntire DD, Leveno KJ, Cunningham FG. Subclinical hyperthyroidism and pregnancy outcomes. *Obstet Gynecol*, 2006; **107**(2 Pt 1): 337–41.
25. Sheehan PM, Nankervis A, Araujo Junior E, Da Silva Costa F. Maternal thyroid disease and preterm birth: systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2015; **100**: 4325–31.
26. Maraka S, Ospina NM, O'Keeffe DT, *et al*. Subclinical hypothyroidism in pregnancy: a systematic review and meta-analysis. *Thyroid*, 2016; **26**: 580–90.
27. Negro R, Stagnaro-Green A. Diagnosis and management of subclinical hypothyroidism in pregnancy. *BMJ*, 2014; **349**: g4929.
28. Karakosta P, Alegakis D, Georgiou V, *et al*. Thyroid dysfunction and autoantibodies in early pregnancy are associated with increased risk of gestational diabetes and adverse birth outcomes. *J Clin Endocrinol Metab*, 2012; **97**: 4464–72.
29. Liu H, Shan Z, Li C, *et al*. Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid*, 2014; **24**: 1642–9.
30. Korevaar TI, Schalekamp-Timmermans S, de Rijke YB, *et al*. Hypothyroxinemia and TPO-antibody positivity are risk factors for premature delivery: the generation R study. *J Clin Endocrinol Metab*, 2013; **98**: 4382–90.
31. Ying H, Tang YP, Bao YR, *et al*. Maternal TSH level and TPOAb status in early pregnancy and their relationship to the risk of gestational diabetes mellitus. *Endocrine*, 2016; **54**:742–50.
32. Henrichs J, Ghassabian A, Peeters RP, Tiemeier H. Maternal hypothyroxinemia and effects on cognitive functioning in childhood: how and why? *Clin Endocrinol (Oxf)*, 2013; **79**: 152–62.
33. Medici M, Timmermans S, Visser W, *et al*. Maternal thyroid hormone parameters during early pregnancy and birth weight: the Generation R Study. *J Clin Endocrinol Metab*, 2013; **98**: 59–66.
34. Julvez J, Alvarez-Pedrerol M, Rebagliato M, *et al*. Thyroxine levels during pregnancy in healthy women and early child neurodevelopment. *Epidemiology*, 2013; **24**: 150–7.
35. Henrichs J, Bongers-Schokking JJ, Schenk JJ, *et al*. Maternal thyroid function during early pregnancy and cognitive functioning in early childhood: the generation R study. *J Clin Endocrinol Metab*, 2010; **95**: 4227–34.
36. Ghassabian A, El Marroun H, Peeters RP, *et al*. Downstream effects of maternal hypothyroxinemia in early pregnancy: nonverbal IQ and brain morphology in school-age children. *J Clin Endocrinol Metab*, 2014; **99**: 2383–90.
37. Korevaar TI, Muetzel R, Medici M, *et al*. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*, 2016; **4**: 35–43.
38. Roman GC, Ghassabian A, Bongers-Schokking JJ, *et al*. Association of gestational maternal hypothyroxinemia and increased autism risk. *Ann Neurol*, 2013; **74**: 733–42.
39. Gyllenberg D, Sourander A, Surcel HM, Hinkka-Yli-Salomaki S, McKeague IW, Brown AS. Hypothyroxinemia during gestation and offspring schizophrenia in a national birth cohort. *Biol Psychiatry*, 2016; **79**: 962–70.
40. Finken MJ, van Eijsden M, Loomans EM, Vrijkotte TG, Rotteveel J. Maternal hypothyroxinemia in early pregnancy predicts reduced performance in reaction time tests in 5- to 6-year-old offspring. *J Clin Endocrinol Metab*, 2013; **98**: 1417–26.
41. Modesto T, Tiemeier H, Peeters RP, *et al*. Maternal mild thyroid hormone insufficiency in early pregnancy and attention-deficit/hyperactivity disorder symptoms in children. *JAMA Pediatr*, 2015; **169**: 838–45.
42. Pakkila F, Mannisto T, Hartikainen AL, *et al*. Maternal and child's thyroid function and child's intellect and scholastic performance. *Thyroid*, 2015; **25**: 1363–74.
43. Levie D, Korevaar TIM, Bath SC, *et al*. Thyroid function in early pregnancy, child iq, and autistic traits: a meta-analysis of individual participant data. *J Clin Endocrinol Metab*, 2018; **103**: 2967–79.
44. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.

45. Nazarpour S, Ramezani Tehrani F, Simbar M, *et al*. Effects of levothyroxine on pregnant women with subclinical hypothyroidism, negative for thyroid peroxidase antibodies. *J Clin Endocrinol Metab*, 2018; **103**: 926–35.
46. Casey BM, Thom EA, Peaceman AM, *et al*. Treatment of subclinical hypothyroidism or hypothyroxinemia in pregnancy. *N Engl J Med*, 2017; **376**: 815–25.
47. Lazarus JH, Bestwick JP, Channon S, *et al*. Antenatal thyroid screening and childhood cognitive function. *N Engl J Med*, 2012; **366**: 493–501.
48. Korevaar TIM, Tiemeier H, Peeters RP. Clinical associations of maternal thyroid function with foetal brain development: epidemiological interpretation and overview of available evidence. *Clin Endocrinol (Oxf)*, 2018; **89**: 129–38.
49. Casey B. Effect of treatment of maternal subclinical hypothyroidism or hypothyroxinemia on IQ in offspring. *AJOG*, 2016; **214**: S2.
50. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. *Nat Rev Endocrinol*, 2017; **13**: 610–22.
51. Korevaar TI, Steegers EA, Chaker L, *et al*. The risk of preeclampsia according to high thyroid function in pregnancy differs by hCG concentration. *J Clin Endocrinol Metab*, 2016; **101**: 5037–43.
52. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*, 2013; **1**: 238–49.
53. Carle A, Pedersen IB, Knudsen N, *et al*. Epidemiology of subtypes of hyperthyroidism in Denmark: a population-based study. *Eur J Endocrinol*, 2011; **164**: 801–9.
54. Sheffield JS, Cunningham FG. Thyrotoxicosis and heart failure that complicate pregnancy. *Am J Obstet Gynecol*, 2004; **190**: 211–17.
55. Sahu MT, Das V, Mittal S, Agarwal A, Sahu M. Overt and subclinical thyroid dysfunction among Indian pregnant women and its effect on maternal and fetal outcome. *Arch Gynecol Obstet*, 2010; **281**: 215–20.
56. Luewan S, Chakkabut P, Tongsong T. Outcomes of pregnancy complicated with hyperthyroidism: a cohort study. *Arch Gynecol Obstet*, 2011; **283**: 243–7.
57. Millar LK, Wing DA, Leung AS, Koonings PP, Montoro MN, Mestman JH. Low birth weight and preeclampsia in pregnancies complicated by hyperthyroidism. *Obstet Gynecol*, 1994; **84**: 946–9.
58. Davis LE, Lucas MJ, Hankins GD, Roark ML, Cunningham FG. Thyrotoxicosis complicating pregnancy. *Am J Obstet Gynecol*, 1989; **160**: 63–70.
59. Pillar N, Levy A, Holcberg G, Sheiner E. Pregnancy and perinatal outcome in women with hyperthyroidism. *Int J Gynaecol Obstet*, 2010; **108**: 61–4.
60. Aggarawal N, Suri V, Singla R, *et al*. Pregnancy outcome in hyperthyroidism: a case control study. *Gynecol Obstet Invest*, 2014; **77**: 94–9.
61. Andersen SL, Olsen J, Wu CS, Laurberg P. Birth defects after early pregnancy use of antithyroid drugs: a Danish nationwide study. *J Clin Endocrinol Metab*, 2013; **98**: 4373–81.
62. Seo GH, Kim TH, Chung JH. antithyroid drugs and congenital malformations: a nationwide Korean cohort study. *Ann Intern Med*, 2018; **168**: 405–13.
63. Zakarija M, McKenzie JM. Pregnancy-associated changes in the thyroid-stimulating antibody of Graves' disease and the relationship to neonatal hyperthyroidism. *J Clin Endocrinol Metab*, 1983; **57**: 1036–40.
64. Barbesino G, Tomer Y. Clinical review: clinical utility of TSH receptor antibodies. *J Clin Endocrinol Metab*, 2013; **98**: 2247–55.
65. Alexander EK, Pearce EN, Brent GA, *et al*. 2016 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
66. Momotani N, Noh J, Oyanagi H, Ishikawa N, Ito K. Antithyroid drug therapy for Graves' disease during pregnancy. Optimal regimen for fetal thyroid status. *N Engl J Med*, 1986; **315**: 24–8.
67. Korevaar TI, Chaker L, Jaddoe VW, Visser TJ, Medici M, Peeters RP. Maternal and birth characteristics are determinants of offspring thyroid function. *J Clin Endocrinol Metab*, 2016; **101**: 206–13.
68. Thangaratinam S, Tan A, Knox E, Kilby MD, Franklyn J, Coomarasamy A. Association between thyroid autoantibodies and miscarriage and preterm birth: meta-analysis of evidence. *BMJ*, 2011; **342**: d2616.
69. Medici M, de Rijke YB, Peeters RP, *et al*. Maternal early pregnancy and newborn thyroid hormone parameters: the Generation R study. *J Clin Endocrinol Metab*, 2012; **97**: 646–52.
70. Korevaar TIM, Pop VJ, Chaker L, *et al*. Dose dependency and a functional cutoff for TPO-antibody positivity during pregnancy. *J Clin Endocrinol Metab*, 2018; **103**: 778–89.
71. Nazarpour S, Ramezani Tehrani F, Simbar M, *et al*. Effects of levothyroxine treatment on pregnancy outcomes in pregnant women with autoimmune thyroid disease. *Eur J Endocrinol*, 2016; **176**: 253–65.
72. Karakosta P, Alegakis D, Georgiou V, *et al*. Thyroid dysfunction and autoantibodies in early pregnancy are associated with increased risk of gestational diabetes and adverse birth outcomes. *J Clin Endocrinol Metab*, 2012; **97**: 4464–72.

9.4

Management of Thyroid Disorders After Pregnancy

Nobuyuki Amino and Naoko Arata

Postpartum Thyroid Dysfunction

Postpartum thyroid dysfunction is defined as an exacerbation of subclinical autoimmune thyroiditis during the postpartum period [1–3]. Patients do not develop thyroid autoimmunity at the onset of postpartum thyroid dysfunction, but had 'subclinical autoimmune thyroiditis' [4] beforehand and this is exacerbated after delivery. Typically, an exacerbation induces destructive thyrotoxicosis followed by transient hypothyroidism. However, various types of thyroid dysfunction may occur, including Graves' disease. Therefore, any kind of thyroid dysfunction observed during postpartum period, is referred to as 'postpartum thyroid dysfunction'.

Pathogenesis

The pathogenesis of postpartum thyroid dysfunction is similar to that of exacerbation of Graves' disease or Hashimoto's thyroiditis, which occurs by the enhancement of immune activities after parturition. The difference is only their stage of autoimmune disease. In postpartum thyroiditis, immune activation causes the transition of subclinical into overt autoimmune thyroid disease; whereas in previously manifest Graves' or Hashimoto's disease, immune activation results in exacerbation or relapse after parturition. During pregnancy, maternal immune activities are suppressed in order to prevent rejection of the fetus. Sudden release from the immune suppression at the time of delivery intensifies immune activities above the normal level, similar to the sudden cessation of immunosuppressive drugs giving rise to the exacerbation of autoimmune diseases [3]. The serial changes in titres of microsomal (thyroid peroxidase, TPO) antibodies in pregnant women with Graves' disease and Hashimoto's disease (**Figure 9.4.1**) support this view. This immune rebound appears to be a general phenomenon observed in the postpartum period, since serum levels of immunoglobulins, and counts of lymphocytes and NK/K cell activity decrease in late pregnancy and increase after delivery even in pregnant women without thyroid dysfunction [5, 6]. Since the immunological changes following termination of pregnancy are similar to those during the postpartum period, thyroid dysfunction may occur following spontaneous of elective abortion [7, 8].

The postpartum rebound of immune activities comprises two phases. Cytotoxic T cells and natural killer (NK) cells increase from 1 to 4 months postpartum (**Figure 9.4.2**) [9, 10]. The enhancement of cellular immunity may exacerbate tissue injury in Hashimoto's thyroiditis. During the second phase, CD5+B cells, which produce autoantibodies, increase typically at 7 to 10 months postpartum (**Figure 9.4.2**) [9]. The enhancement of humoral immunity may cause postpartum Graves' disease, by an increase of anti-TSH receptor autoantibodies. Hashimoto's thyroiditis is commonly aggravated from 1 to 4 months postpartum and Graves' disease may develop or relapse from 3 to 10 months postpartum (**Figure 9.4.3**).

Studies on the production of cytokines have shown that Th1-type and Th2-type cytokines decrease during pregnancy, Th1-type cytokines increase during early postpartum period and Th2-type cytokines increase during later postpartum period [11], supporting the immune rebound hypothesis [2, 3]. A possible role for fetal microchimerism has been proposed as the mechanism of postpartum exacerbation of autoimmune thyroid disease [12] but insufficient evidence to support this hypothesis is available and further studies are needed.

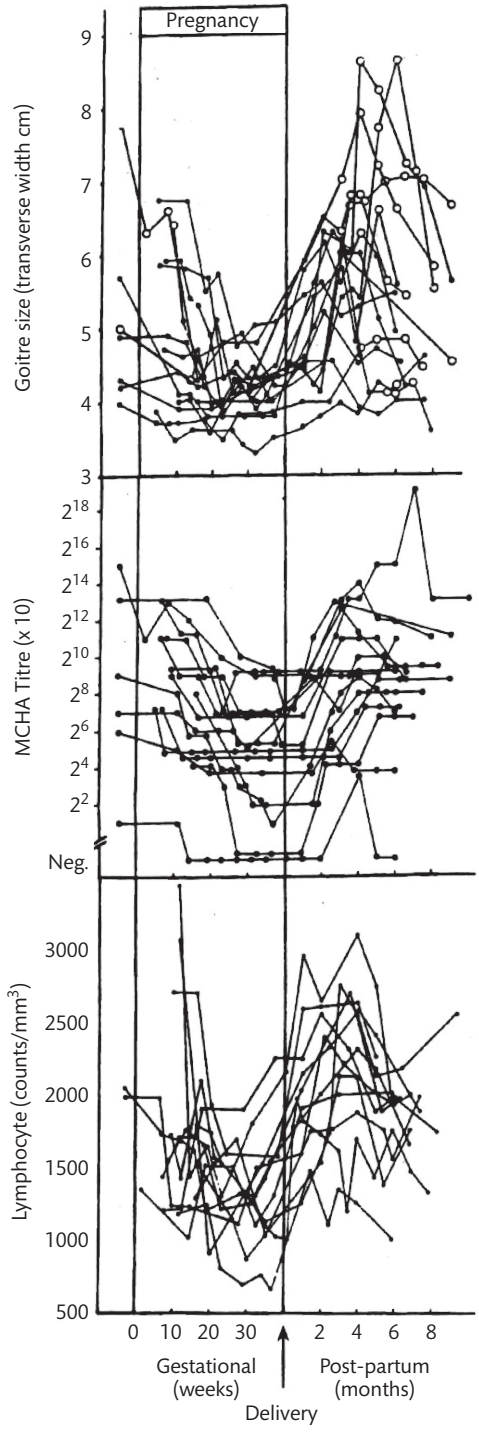


Figure 9.4.1 Serial changes in goitre size, titres of antithyroid microsomal antibody, and the counts of peripheral lymphocytes during pregnancy and postpartum period in patients with Hashimoto's thyroiditis. Open circles denote that TSH was more than 10 mU/l at time of measurement. MCHA: microsomal haemagglutination antibody.

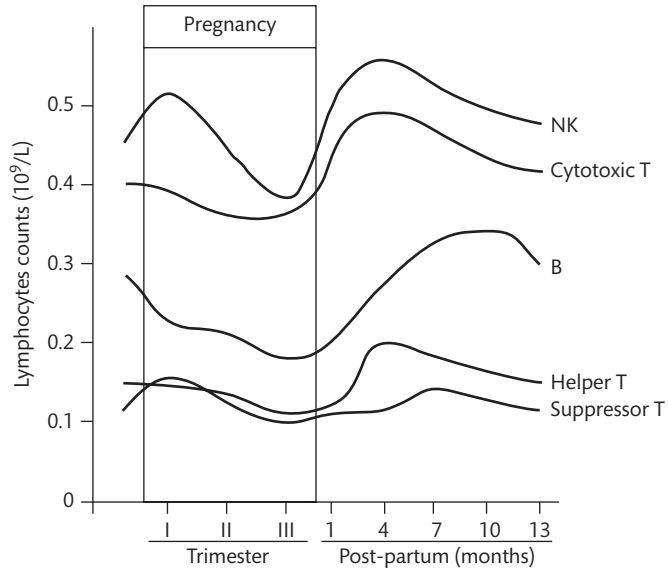


Figure 9.4.2 Changes in peripheral T cells, B cells, and NK cells during pregnancy and the postpartum period in the normal women.

Prevalence

Postpartum thyroid dysfunction is very common [3, 13] with an incidence of about 5% (1.1–16.7) (Table 9.4.1) in mothers in the general population, that is one in 20 pregnant women develop postpartum thyroid dysfunction [13]. Postpartum thyroid dysfunction can be classified into five groups by the clinical features – hyperthyroidism and/or hypothyroidism, transient or persistent (Figure 9.4.4) [2]:

- **I.** Persistent thyrotoxicosis
- **II.** Transient thyrotoxicosis
- **III.** Destructive thyrotoxicosis followed by transient hypothyroidism
- **IV.** Transient hypothyroidism
- **V.** Persistent hypothyroidism

Persistent thyrotoxicosis (group I) and a proportion of transient thyrotoxicosis (group II) reveal a high radioiodine uptake. Transient thyrotoxicosis with a high radioiodine uptake is common in postpartum Graves' disease; the overproduction of thyroid hormones ceases spontaneously within a year. The prevalence of postpartum Graves' disease (both persistent and transient) is estimated at 11% of those with postpartum thyroid dysfunction and 0.54% of the general population [14]. Thyrotoxicosis due to postpartum Graves' disease occurs between 3- and 10-months postpartum.

The other three types of postpartum thyroid dysfunction are associated with thyroid tissue damage due to an exacerbation of autoimmune thyroiditis. They often manifest themselves as transient thyrotoxicosis (destructive thyrotoxicosis) developing at 1 to 3 months postpartum. Depending on the extent of the destruction, transient hypothyroidism may follow (group III) or not (group II, with a low radioiodine uptake).

Occasionally, Graves' disease occurs closely following, or concomitantly with destructive thyrotoxicosis [15]. When cellular damage occurs slowly, hypothyroidism alone, rather than destructive thyrotoxicosis, may be observed after delivery. In many cases, it is transient (group IV). However, it may be persistent in a few cases (group V). Destructive transient thyrotoxicosis (group II, with a low radioiodine uptake, and group III) is the most common form of postpartum thyroid dysfunction, accounting for 50–60% of all postpartum thyroid dysfunction. The rest (group IV, V) only have a hypothyroid phase, however, persistent hypothyroidism is very rare (less than 0.1%).

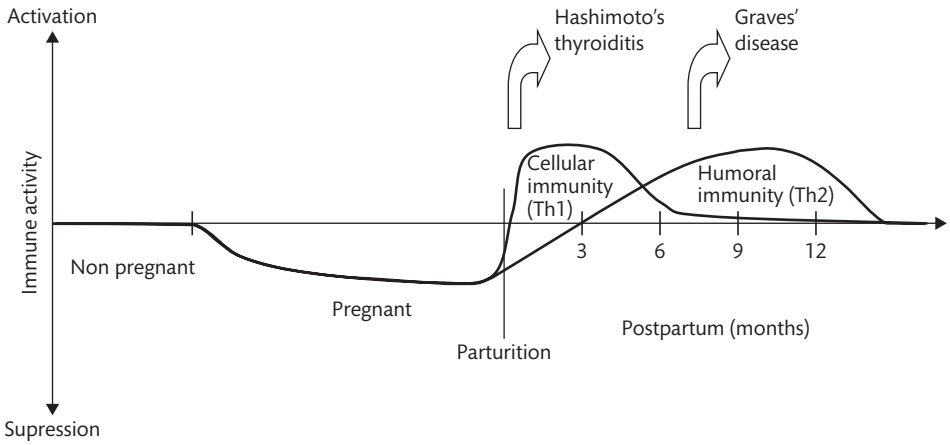


Figure 9.4.3 Immune rebound hypothesis of postpartum autoimmune thyroid diseases. One to 3 months after delivery, cellular immunity dominates, and development or exacerbation of autoimmune thyroiditis is observed. Three to six months after delivery, humoral immunity dominates, and development or exacerbation of Graves' disease is observed.

Symptoms and Signs

Thyroid dysfunction is most often subclinical: the patient usually has no features of hyper- or hypothyroidism, and thyroid function tests reveal only mild changes in serum TSH and thyroid hormones. Symptoms and signs in overt postpartum thyroid dysfunction are the same as in the non-postpartum state. Hypermetabolic and hyperdynamic symptoms, such as palpitation, sweating and tremor can be observed in any type of postpartum thyrotoxicosis.

In postpartum Graves' disease, eye signs and/or pretibial myxoedema may be present. In postpartum hypothyroidism, symptoms such as weakness, fatigue, dry skin, constipation, and cold intolerance, and signs such as cold skin, bradycardia, and thyroid enlargement are common. Since the hypothyroidism is of short duration, there is little risk of myxoedema. Postpartum depression, sometimes found with postpartum thyroid dysfunction [16], may be a significant consequence and is the presence of elevated antithyroid autoantibodies rather than with hypothyroidism [17].

Table 9.4.1 Incidence of postpartum thyroid dysfunction

Year	Author	Country	Prevalence(%)
1982	Amino	Japan	5.5
1984	Jansson	Sweden	6.5
1985	Walfish	Canada	7.1
1986	Freeman	USA	1.9
1987	Nikolai	USA	6.7
1987	Lervang	Denmark	3.9
1988	Fung	UK	16.7
1990	Rasmussen	Denmark	3.3
1990	Rajatanavin	Thailand	1.1
1991	Roti	Italy	8.7
1991	Lobig	Germany	2.0
1992	Walfish	Canada	6.0
1992	Stagnaro-Green	USA	8.8
1998	Kuijpens	Netherlands	12.4
2000	Lucas	Spain	7.8
2000	Barca	Brazil	13.3
2001	Shahbazian	Iran	11.4
2001	Bagis	Turkey	5.5

Diagnosis

The diagnosis of postpartum thyroid dysfunction is straightforward when the patient shows abnormal thyroid function tests and positive thyroid autoantibodies during the postpartum period. Thyroid dysfunction, however, is most often subclinical. For the cases with overt thyrotoxicosis, it is essential to differentiate between postpartum Graves' disease and destructive thyrotoxicosis. Usually an educated guess can be made from the time of onset (3–12 months after parturition in Graves' disease versus 1–6 months in destructive thyrotoxicosis). Figure 9.4.5 summarizes the time of onset of four different types of pregnancy-associated thyrotoxicosis [18]. Anti-TSH receptor antibodies and other markers helpful for the differential diagnosis are summarized in Table 9.4.2 [19–22]. Figure 9.4.6 shows measurements of anti-TSH receptor antibody concentrations and thyroid blood flow in postpartum destructive thyrotoxicosis (n = 42) and postpartum Graves' disease (n = 102) [22]. Blood tests, however, are not conclusive since anti-TSH-receptor autoantibodies are sometimes found in Hashimoto's thyroiditis and other tests do not have distinct cut-off values. The quantitative measurement of thyroid blood flow using Doppler ultrasonography is useful for differentiation and typically is high in Graves' disease and low in destructive thyrotoxicosis [23]. The measurement of radioactive iodine uptake provides definitive information in the differential diagnosis between Graves' disease (high uptake) and destructive thyrotoxicosis (low uptake). However, this investigation should not be performed in breastfeeding mothers. Notably, there is no reliable way to differentiate between transient and persistent Graves' disease.

For hypothyroidism, most cases are transient and due to an exacerbation of autoimmune thyroiditis. The finding of positive antithyroid microsomal (peroxidase) antibodies and/or antithyroglobulin antibody supports the existence of autoimmune

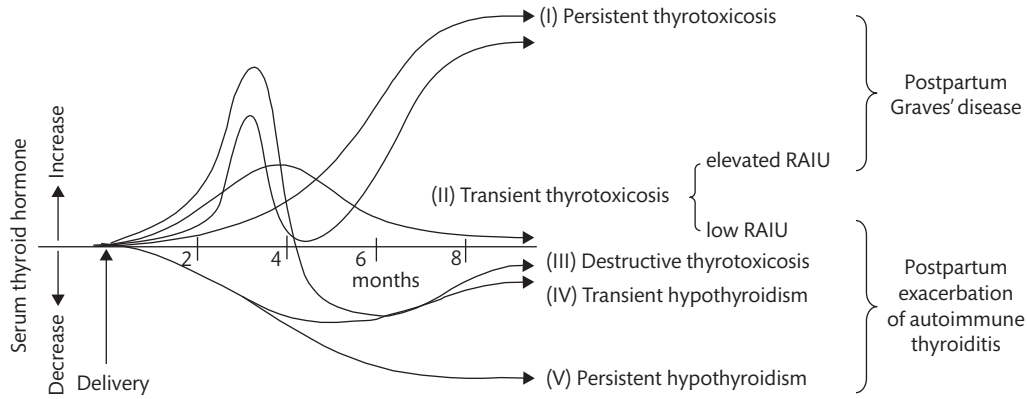


Figure 9.4.4 Time course of five different types of postpartum thyroid dysfunction.

thyroiditis, but negative results are obtained in 5–30%. Cases of iodine deficient hypothyroidism may occur in areas where iodine intake is marginal or mildly deficient, such as some parts of Europe [24], although it is likely that mothers already had hypothyroidism during pregnancy.

Once the diagnosis is established, patients should be followed up for 1 year since Graves' disease may occur shortly after destructive thyrotoxicosis.

In postpartum thyroid dysfunction some immunological abnormalities are observed before the onset of thyroid dysfunction (and therefore, before and during pregnancy). Among these, the finding of raised TPO or microsomal antibodies is the most useful marker for the prediction of the occurrence of postpartum thyroid dysfunction [25]. If TPO antibodies are present, there is always lymphocytic infiltration into the thyroid, and therefore 'subclinical autoimmune thyroiditis' [4] which may be exacerbated after delivery. Sixty to seventy per cent of women with a positive measurement of thyroid peroxidase antibodies in early pregnancy develop postpartum thyroid dysfunction [3], whereas the risk to develop postpartum thyroid dysfunction in TPO antibody negative women is estimated to be 0.6%. Mothers with high TPO titres (> 5000–10 000 reciprocal dilution) almost always develop postpartum thyroid dysfunction. However, the measurement of thyroid peroxidase antibodies does not provide any information on the type of dysfunction that will occur. Although the measurement of thyroid peroxidase antibodies with semiquantitative antimicrosomal particle agglutination (MCPA) tests is simple and cheap, the value of screening for postpartum autoimmune thyroid dysfunction remains unclear [26], and is in part dependent of different healthcare systems [27]. Antithyroid peroxidase antibody measurement is as useful as semiquantitative particle agglutination tests for predicting postpartum thyroid dysfunction [28].

A recent study, using highly sensitive immunoassays, has shown that measurement of antithyroglobulin antibody is more sensitive than that of antithyroid peroxidase antibody to detect thyroid

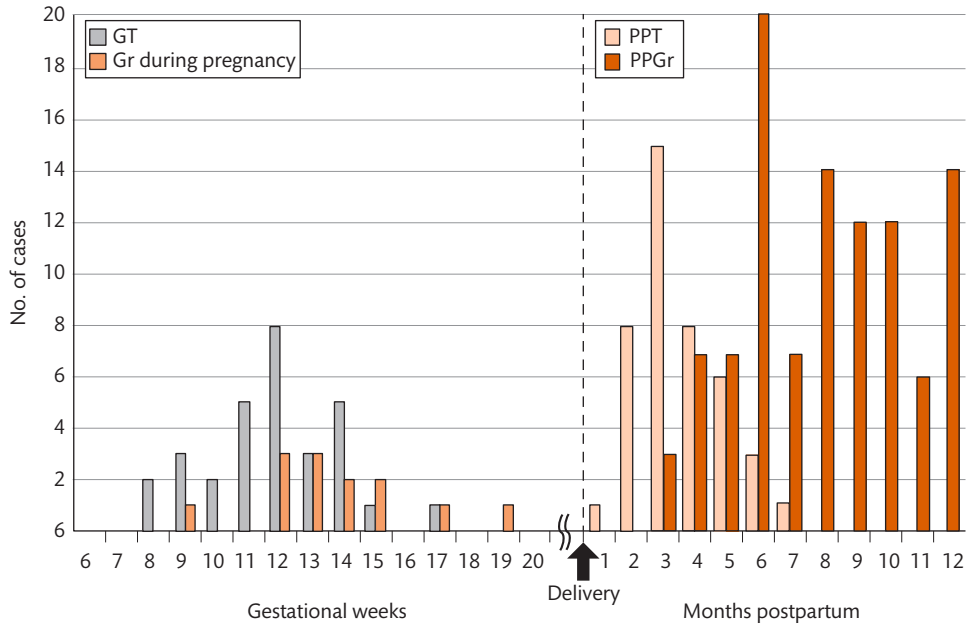


Figure 9.4.5 Time of onset and frequency of four different types of pregnancy-associated thyrotoxicosis. GT, gestational thyrotoxicosis; Gr, Graves' disease; PPT, postpartum thyroiditis; PPGr, postpartum onset Graves' disease.

Table 9.4.2 Differential diagnosis between postpartum Graves' disease and postpartum destructive thyrotoxicosis

	Graves' disease	Destructive thyrotoxicosis
Onset	3–12 months postpartum	1–6 months postpartum
Anti-TSH receptor antibody	Positive	Negative
Eye signs	Yes	No
Total T_3/total T_4 ratio (ng/μg)	>20 in 80% of cases	< 20
FT3/FT4 ratio	>2.8 in 77% of cases	< 2.8
Thyroid blood flow	High	Low
Radioactive iodine uptake*	High	Low
Serial change in serum thyroglobulin**	<50% increase from a month before the onset	>50% increase from a month before the onset

* Contraindicated in lactating mother; ** Thyroglobulin should be measured with a method not influenced by antithyroglobulin autoantibody(TgAb)

autoimmune abnormalities [29]. Further studies are required to confirm the value of sensitive antithyroglobulin antibody measurements in the diagnosis of postpartum thyroiditis.

Graves' disease is aggravated in early pregnancy, ameliorates in the latter half of pregnancy, but often relapses postpartum [30]. Human chorionic gonadotrophin (hCG) plays a crucial role in the exacerbation of Graves' thyrotoxicosis in early pregnancy [31]. Relapse of Graves' thyrotoxicosis after parturition may occur even if mothers are in remission before pregnancy.

The new onset of Graves' disease in the postpartum period is of great interest, since early diagnosis and treatment early during the disease period can result in rapid remission [14]. Overall, 40% of patients with Graves' disease who have had one or more deliveries, develop their disease postpartum [32]. TSH-receptor-stimulating antibodies are a good indicator of postpartum development of Graves' disease, since they are often positive before onset of the disease [33] when measured with a sensitive bioassay. Pregnant women with positive TSH-receptor-stimulating antibodies in early pregnancy have a high-risk of developing postpartum Graves' disease. Seventy-one pregnant women with positive TPO antibodies in early pregnancy were prospectively observed from early pregnancy to the postpartum period [14]. Among them, seven showed positive TSH-receptor-stimulating antibodies, five (70%) of whom developed postpartum Graves' disease. Thyrotoxicosis in three of those five was transient and spontaneously improved within a year. Graves' disease did not occur in the TSH-receptor-stimulating antibody-negative patients. A conventional radioreceptor assay for anti-TSH-receptor antibodies was not able to discriminate postpartum Graves' disease.

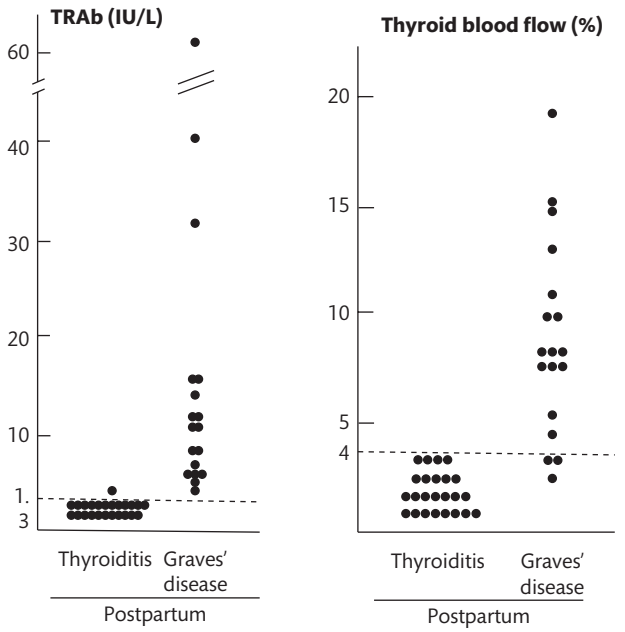


Figure 9.4.6 Differentiation of postpartum Graves' thyrotoxicosis from postpartum destructive thyrotoxicosis (postpartum thyroiditis) using antithyrotropin receptor antibodies(TRAb) and thyroid blood flow.

Modified with permission from Ide A *et al*. Differentiation of postpartum Graves' thyrotoxicosis from postpartum destructive thyrotoxicosis using antithyrotropin receptor antibodies and thyroid blood flow. *Thyroid*, 2014;24:1027–31. Copyright © 2014, Mary Ann Liebert, Inc. (ref 22).

Management of Postpartum Thyroid Dysfunction

In postpartum Graves' disease, treatment options are antithyroid drugs, radioactive iodine, or subtotal thyroidectomy similar to Graves' disease outside this setting. Antithyroid drug treatment is a good initial choice because (1) postpartum Graves' hyperthyroidism is often transient; (2) even in persistent Graves' disease, those patients diagnosed early are easily controlled with antithyroid drugs [14]; and (3) mothers may not want breastfeeding to be interrupted by radioiodine therapy. Radioactive iodine treatment can still be applied when hyperthyroidism persists after 1 year and has not gone into remission during antithyroid drug treatment.

In destructive thyrotoxicosis, the thyrotoxic phase is always transient and spontaneously ceases in 1–3 months. Treatment should be symptomatic, mainly with ß-adrenergic antagonists for cardiovascular hyperdynamic symptoms. Antithyroid therapy is not indicated. The treatment of hypothyroidism is required only when the patient has symptoms of hypothyroidism. Usually levothyroxine therapy with a gradual reduction in dose may be indicated, but recovery of patient's thyroid function may be difficult to demonstrate. Alternatively, replacement with a submaximal dose of liothyronine (15–50 μg 3 daily by mouth in 1–3 divided doses) is useful in most transient cases, since spontaneous recovery of thyroid function can be monitored by an increase of serum T_4 or free T_4. In permanent hypothyroidism, when serum T_4 or free T_4 does not recover after several months of T_3 treatment, levothyroxine replacement is indicated.

Recently successful prevention of postpartum thyroid dysfunction was achieved by short-term immunosuppressive therapy in patients who were predicted to develop postpartum hypothyroidism [34]. It is also reported that selenium supplementation during pregnancy and in the postpartum period reduces thyroid inflammatory activity and the incidence of hypothyroidism [35].

Prognosis

Little is known about the long-term prognosis of postpartum Graves' hyperthyroidism, although a better outcome than in 'ordinary' Graves' disease might be expected in view of early diagnosis just after the onset of disease. In destructive thyrotoxicosis and/or hypothyroidism due to exacerbation of autoimmune thyroiditis, thyroid dysfunction is transient, and most patients recover spontaneously to euthyroidism. Only in a few cases, hypothyroidism may persist. High titres of thyroglobulin antibodies and/or thyroid peroxidase antibodies are risk factors of persistent hypothyroidism. Even after recovery from hypothyroidism, abnormalities in ultrasonography and/or iodide perchlorate discharge tests may persist for a long time [36, 37], reflecting underlying chronic autoimmune thyroiditis. The patients almost certainly will develop postpartum thyroid dysfunction after the next parturition, with similar time of onset, type of thyroid dysfunction, and duration of dysfunction as in the previous episode.

Late development (after 5 years or more) of permanent hypothyroidism is found in 25–60% of the patients with postpartum thyroiditis [38–40]. Therefore, these patients should be followed up at appropriate intervals (once every 1–2 years) [26, 41]. Othman *et al.* reported that high titres of antimicrosomal antibodies and the severity of the hypothyroid phase of postpartum thyroiditis are risk factors for the late development of permanent hypothyroidism, but there was no association with HLA haplotype or family history of thyroid disease [38]. In contrast, in a Japanese population, high titres of thyroglobulin antibodies and HLA-DRw9 and/or B51 genotype were risk factors of permanent hypothyroidism [39].

Thyroid Disorders and Breastfeeding

The Influence of Thyroid Disease on Breastfeeding

Because hypothyroidism suppresses milk production and may restore with levothyroxine replacement [42, 43], women who have difficulties with lactation should be evaluated thyroid function and started on levothyroxine when hypothyroidism is revealed [41]. Data on the effects of hyperthyroidism on milk production are limited in humans.

Medications in Breastfeeding Women with Graves' Disease

Methimazole (MMI) concentration in breast milk is higher than propylthiouracil (PTU) concentration, with a mean milk/serum ratio of 0.98 in MMI [44] and 0.1 in PTU [45]. Regarding the influence on infants, drug exposure to infants through breast milk is important. When MMI 10 mg/day is administered to a mother, the amount of MMI exposed to infant through breast milk corresponds to 0.6 to 1.6 mg/day in terms of maternal body weight [46], and when 400 mg/day of PTU is administered to a mother, the exposure to infant of 0.8–9.2 mg/day is equivalent [45]. The peak of MMI concentration in blood and milk of 5 lactating women who received a single dose of carbimazole 40 mg (corresponding to 30 mg of MMI) was 1 hour after administration, and the mean blood half-life of MMI was 4.9 hours [44]. When women taking 10 mg of MMI 2 to 3 times per day, breastfeeding was performed 2 hours after taking MMI, and the blood MMI concentration of the infant was measured 2 hours after lactation, the MMI concentration of the infant was negligible [47].

According to several reports by Azizi *et al.*, infant hypothyroidism within one year due to the drug exposure by breast milk following administration of MMI of 10–30 mg/day to mothers was not observed. Growth (monitored postnatally for up to 60 months, somatic and neurodevelopment were also normal in the same series [47–49]. It should be noted that MMI 20–30 mg/day was administered only for the first month, and thereafter the dose was reduced to 10 mg/day or less in these studies. Similarly, hypothyroidism of the offspring following exposure to breastmilk of a mother taking PTU 750 mg daily has not been reported, although there are only a few case reports [50].

Based on these factors, both MMI (up to maximal dose of 20 mg/d) and PTU (up to maximal dose of 450 mg/d) can be administered, when antithyroid drugs is indicated for women with breastfeeding [41]. Monitoring for appropriate growth and development of their children by routine paediatric evaluations should be at least done [41]. It may be necessary to check the thyroid function of the infant when ATD is administered at moderate or higher doses. A mother should take her ATD dose just after breastfeeding, which should provide a 3–4 hour interval before she lactates again [51]. Considering the possibility of side effects of severe hepatic injury of PTU in mothers and infants [52], and high incidence of general side effects with PTU [53], MMI is the preferred drug in hyperthyroid women who is breastfeeding.

The influence on thyroid function of breastfed infants whose mothers were treated with inorganic iodine for postpartum relapsing Graves' disease has been reported from Japan, which is one of country with iodine excess intake [54]. Mothers were taking 37.5 mg per day of the median inorganic iodine, the median iodine concentration in breast milk was 15.1 mg/L, and the median urine iodine concentration of infants was 15.6 mg/L. In spite of extremely high iodine exposure to infants via breast milk, hypothyroidism in breastfed infants was found only in 3.8% (1 out of 26 children) [54]. Due to the presence of the sodium iodide symporter in mammary epithelial cells, inorganic iodine is concentrated and secreted in the milk. Moreover, suppression of thyroid function during infancy caused by excess iodine exposure may be a great negative influence on the growth and development of children. Therefore, inorganic iodine administration to breast-feeding mothers should be avoided, regardless of the report from Japan.

Transfer of Maternal Thyroid Hormone to Newborn via Breast Milk

A small amount of maternal thyroid hormone is present in the breast milk of lactating women. Mean thyroxine concentration in breast milk was reported as 0.83 μg/L [55], which showed maternal

transfer of thyroid hormone does not have a meaningful impact on the infants' thyroid hormone status.

Administration of Diagnostic or Therapeutic Radiopharmaceuticals to Breastfeeding Women

The use of ^{131}I is contraindicated during lactation. If required, ^{123}I can be used if breast milk is pumped and discarded for 3–4 days before breastfeeding is resumed. Similarly, Tc-^{99m} pertechnetate administration requires breast milk to be pumped and discarded during the day of testing [41].

REFERENCES

1. Amino N, Mori H, Iwatani Y, *et al*. High prevalence of transient post-partum thyrotoxicosis and hypothyroidism. *N Engl J Med*, 1982; **306**: 849–52.
2. Amino N, Tada H, Hidaka Y. The spectrum of postpartum thyroid dysfunction: diagnosis, management, and long-term prognosis. *Endocr Pract*, 1996; **2**: 406–10.
3. Amino N, Tada Hs, Hidaka Y. Postpartum autoimmune thyroid syndrome: a model of exacerbation of autoimmune disease. *Thyroid*, 1999**9**: 705–13.
4. Yoshida H, Amino N, Yagawa K, *et al*. Association of serum antithyroid antibodies with lymphocytic infiltration of the thyroid gland: studies of seventy autopsied cases. *J Clin Endocrinol Metab*, 1978; **46**: 859–62.
5. Amino N, Kuro R, Tanizawa O, *et al*. Changes of serum antithyroid antibodies during and after pregnancy in autoimmune thyroid disease. *Clin Exp Immunol*, 1978; **31**: 30–7.
6. Hidaka Y, Amino N, Iwatani Y, *et al*. Increase in peripheral natural killer cell activity in patients with autoimmune thyroid disease. *Autoimmunity*, 1992; **11**: 239–46.
7. Amino N, Miyai K, Kuro P, *et al*. Transient post-partum hypothyroidism: fourteen cases with autoimmune thyroiditis. *Ann Int Med*, 1977; **87**: 155–9.
8. Stagnaro-Green A. Post-miscarriage thyroid dysfunction. *Obstet Gynaecol*, 1992; 80: 490–2.
9. Watanabe M, Iwatani Y, Kaneda T, *et al*. Changes in T, B, and NK lymphocyte subsets during and after normal pregnancy. *Am J Reprod Immunol*, 1997; **37**: 368–77.
10. Stagnaro-Green A, Roman S, Cobin R, El-Harazy H, Wallenstein S, Davies T. A prospective study of lymphocyte-initiated immunosuppression in normal pregnancy: evidence of a T-cell etiology for post-partum thyroid dysfunction. *J Clin Endocrinol Metab*, 1992; **74**: 645–53.
11. Shimaoka Y, Hidaka Y, Tada H, *et al*. Changes in cytokine production during and after normal pregnancy. *Am J Reprod Immunol*, 2000; **44**: 143–7.
12. Ando T, Davies TF. Postpartum autoimmune thyroid disease: the potential role of fetal microchimerism. *J Clin Endocrinol Metab*, 2003; **88**: 2965–71.
13. Nicholson WK, Robinson KA, Smallridge RC, Ladenson PW, Powe NR. Prevalence of postpartum thyroid dysfunction: a quantitative review. *Thyroid*, 2006; **16**: 573–82.
14. Hidaka Y, Tamaki H, Iwatani Y, Tada H, Mitsuda N, Amino N. Prediction of post-partum onset of Graves' thyrotoxicosis by measurement of thyroid stimulating antibody in early pregnancy. *Clin Endocrinol* (*Oxf*), 1994; **41**: 15–20.
15. Momotani N, Noh J, Ishikawa N, Ito K. Relationship between silent thyroiditis and recurrent Graves' disease in the post-partum period. *J Clin Endocrinol Metab*, 1994; **79**: 285–9.
16. Bergink V, Pop VJM, Nielsen PR, Agerbo E, Munk-Olsen T, Liu X. Comorbidity of autoimmune thyroid disorders and psychiatric disorders during the postpartum period: a Danish nationwide register-based cohort study. *Psychol Med*, 2017; **20**: 1–9.
17. Kuijpens JL, Vader HL, Drexhage HA, Wiersinga WM, van Son MJ, Pop VJ. Thyroid peroxidase antibodies during gestation are a marker for subsequent depression postpartum. *Eur J Endocrinol*, 2001; **145**: 579–84.
18. Ide A, Amino N, Kudo T, *et al*. Comparative frequency of four different types of pregnancy-associated thyrotoxicosis in a single thyroid centre. *Thyroid Res*, 2017; **10**: 4.
19. Amino N, Yabu Y, Miyai K, *et al*. Differentiation of thyrotoxicosis induced by thyroid destruction from Graves' disease. *Lancet*, 1978; **2**: 344–6.
20. Hidaka Y, Nishi I, Tamaki H, *et al*. Differentiation of the postpartum thyrotoxicosis by serum thyroglobulin; Usefulness of the new multi-site immunoradiometric assay. *Thyroid*, 1994; **4**: 275–8.
21. Noh JY, Momotani N, Fukada S, Ito K, Miyauchi A, Amino N. Ratio of serum free triiodothyronine to free thyroxine in Graves' hyperthyroidism and thyrotoxicosis caused by painless thyroiditis *Endocr J*, 2005; **52**: 537–2
22. Ide A, Amino N, Kang S, *et al*. Differentiation of postpartum Graves' thyrotoxicosis from postpartum destructive thyrotoxicosis using antithyrotropin receptor antibodies and thyroid blood flow. *Thyroid*, 2014; **24**: 1027–31.
23. Ota H, Amino N, Morita S, *et al*. Quantitative measurement of thyroid blood flow for differentiation of painless thyroiditis from Graves' disease. *Clin Endocrinol (Oxf)*, 2007; **67**: 41–5.
24. Glinoer D, Delange F, Laboureur I, *et al*. Maternal and neonatal thyroid function at birth in an area of marginally low iodine intake. *J Clin Endocrinol Metab*, 1992; **75**: 800–5.
25. Amino N, Tada H, Hidaka Y, Izumi Y. Postpartum autoimmune thyroid syndrome. *Endocr J*, 2000; 47: 645–55.
26. De Groot LJ, Abalovich M, Alexander EK, *et al*. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 2543–65.
27. Hayslip CC, Fein HG, O'Donnell VM, Friedman DS, Klein TA, Smallridge RC. The value of serum antimicrosomal antibody testing in screening for symptomatic post-partum thyroid dysfunction. *Am J Obstet Gynaecol*, 1988; **159**: 203–9.
28. Feldt-Rasmussen U, Hoier MM, Rasmussen NG, Hegedus L, Hornnes P. Anti-thyroid peroxidase antibodies during pregnancy and postpartum. Relation to post-partum thyroiditis. *Autoimmunity*, 1990; **6**: 211–14.
29. Nishihara E, Amino N, Kudo T, *et al*. Comparison of thyroglobulin and thyroid peroxidase antibodies measured by five different kits in autoimmune thyroid diseases. *Endocr J*, 2017; **64**: 955–61.
30. Amino N, Amino N, Kudo T, *et al*. Exacerbation of thyrotoxicosis in early pregnancy and after delivery in Graves' disease. *J Clin Endocrinol Metab*, 1982; **55**: 108–12.
31. Tamaki H, Itoh E, Kaneda T, *et al*. Crucial role of serum human chorionic gonadotropin for the exacerbation of thyrotoxicosis in early pregancy in Graves' disease. *Thyroid*, 1993; **3**: 189–93.
32. Tada H, Hidaka Y, Tsuruta E, *et al*. Prevalence of post-partum onset of disease within patients with Graves' disease of childbearing age. *Endocr J*, 1994; **41**: 325–7.
33. Kasagi K, Hatabu H, Tokuda Y, Iida Y, Endo K, Konishi J. Studies on thyrotrophin receptor antibodies in patients with euthyroid Graves' disease. *Clin Endocrinol (Oxf)*, 1988; **29**: 357–66.

34. Tada H, Hidaka Y, Izumi Y, *et al*. A preventive trial of short-term immunosuppressive therapy in postpartum thyroid dysfunction. *Int J Endcrinol Metab*, 2003; **2**: 48–54.
35. Negro R, Greco G, Mangieri T, Pezzarossa A, Dazzi D, Hassan H. The influence of selenium supplementation on postpartum thyroid status in pregnant women with thyroid peroxidase autoantibodies. *J Clin Endocrinol Metab*, 2007; **92**: 1263–8.
36. Adams H, Jones MC, Othman S, *et al*. The sonographic appearances in post-partum thyroiditis. *Clin Radiol*, 1992; **45**: 311–15.
37. Creagh FM, Parkes AB, Lee A, *et al*. The iodide perchlorate discharge test in women with previous post-partum thyroiditis: relationship to sonographic appearance and thyroid function. *Clin Endocrinol (Oxf)*, 1994; **40**: 765–8.
38. Othman S, Phillips DIW, Parkes AB, *et al*. A long-term follow-up of post-partum thyroiditis. *Clin Endocrinol (Oxf)*, 1990; **32**: 559–64.
39. Tachi J, Amino N, Tamaki H, Aozasa M, Iwatani Y, Miyai K. Long term follow-up and HLA association in patients with post-partum hypothyroidism. *J Clin Endocrinol Metab*, 1988; **66**: 480–4.
40. Azizi F. The occurrence of permanent thyroid failure in patients with subclinical postpartum thyroiditis. *Eur J Endocrinol*, 2005; **153**: 367–71.
41. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
42. Buckshee K, Kriplani A, Kapil A, Bhargava VL, Takkar D. Hypothyroidism complicating pregnancy. *Aust N Z J Obstet Gynaecol*, 1992; **32**: 240–2.
43. Stein MT, Kessler DB, Hubbard E. Failure to thrive in a four-month-old nursing infant. *J Dev Behav Pediatr*, 2004; **25**: S69–73.
44. Johansen K, Andersen AN, Kampmann JP, Mølholm Hansen JM, Mortensen HB. Excretion of methimazole in human milk. *Eur J Clin Pharmacol*, 1982; **23**: 339–41.
45. Kampmann JP, Johansen K, Hansen JM, Helweg J. Propylthiouracil in human milk. Revision of a dogma. *Lancet*, 1980; **1**: 736–7.
46. Tegler L, Lindstrom B. Antithyroid drugs in milk. *Lancet*, 1980; **2**: 591.
47. Azizi F. Thyroid function in breast-fed infants is not affected by methimazole-induced maternal hypothyroidism: results of a retrospective study. *J Endocrinol Invest*, 2003; **26**: 301–4.
48. Azizi F, Afkhami M, Sarshar A, Nafarabadi M. Thyroid function and intellectual development of infants nursed by mothers taking methimazole. *J Clin Endocrinol Metab*, 2000; **85**: 3233–8.
49. Azizi F, Khoshniat M, Bahrainian M, Hedayati M. Intellectual development and thyroid function in children who were breast-fed by thyrotoxic mothers taking methimazole. *J Pediatr Endocrinol Metab*, 2003; **16**: 1239–43.
50. Momotani N, Yamashita R, Makino F, Noh JY, Ishikawa N, Ito K. Thyroid function in wholly breast-feeding infants whose mothers take high doses of propylthiouracil. *Clin Endocrinol (Oxf)*, 2000; **53**: 177–81.
51. Mandel SJ, Cooper DS. The use of antithyroid drugs in pregnancy and lactation. *J Clin Endocrinol Metab*, 2001; **86**: 2354–9.
52. Bahn RS, Burch HS, Cooper DS, *et al*. The role of propylthiouracil in the management of Graves' disease in adults: report of a meeting jointly sponsored by the American Thyroid Association and the Food and Drug Administration. *Thyroid*, 2009; **19**: 673–4.
53. Nakamura H, Noh JY, Itoh K, Fukata S, Miyauchi A, Hamada N. Comparison of methimazole and propylthiouracil in patients with hyperthyroidism caused by Graves' disease. *J Clin Endocrinol Metab*, 2007; **92**: 2157–62.
54. Hamada K, Mizokami T, Maruta T, *et al*. Effects of inorganic iodine therapy administered to lactating mothers with Graves' disease on infant thyroid function. *J Endocr Soc*, 2017; **1**: 1293–300.
55. van Wassenaer AG, Stulp MR, Valianpour F, *et al*. The quantity of thyroid hormone in human milk is too low to influence plasma thyroid hormone levels in the very preterm infant. *Clin Endocrinol (Oxf)*, 2002; **56**: 621–7.

9.5

Thyroid Disorders in Newborns

A.S. Paul van Trotsenburg and Nitash Zwaveling-Soonawala

Introduction

Thyroid hormone is essential for normal brain development starting in the early embryonic period and continuing throughout the first years of life [1]. While in adult onset hypothyroidism clinical manifestations are usually reversed with appropriate treatment, untreated congenital hypothyroidism has severe neurological consequences leading to permanent intellectual and motor disabilities. The important role of the thyroid in brain development was already recognized by 1850 when the British surgeon Curlings reported two mentally impaired children with large tongues, who appeared to have no thyroid gland at obduction [2]. In 1871, the British internist Fagge described four other mentally impaired and often deaf children and adolescents who were extremely short (adult height less than 120 cm), and who had a broad face with a flat root of the nose, thick nostrils, a large open mouth and thick lips, short broad hands and feet, and a swollen skin [3]. Although at that time a relationship was suggested between the striking disease and the absence of the thyroid, the function of this organ was still completely unknown. In the 1890s the first treatment trials in adults with acquired hypothyroidism, using thyroid extract from animals, led to improvement of clinical symptoms [4]. Treatment of children with clinically diagnosed congenital hypothyroidism in the 1950's restored growth and many other signs and symptoms [5]. These initial treatment protocols, however, were not successful in preventing mental retardation. In 1972 it was demonstrated that the key to preventing brain damage is treatment started within the first three months of life [6]. Since only one third of patients were clinically recognized as early as that, newborn screening programmes for congenital hypothyroidism were introduced in the late 1970s. These screening programmes have proven to be very effective in preventing brain damage by early detection and treatment [7, 8]. Children with congenital hypothyroidism treated within the first two to three weeks of life have near normal development, although mild neurological deficits and health impairments are observed in the most severely affected [9]. Congenital hypothyroidism has a reported incidence of one in 2000–4000 live births and may be of thyroidal (primary) or of hypothalamic–pituitary (central) origin. Primary congenital hypothyroidism accounts for approximately 90–95% of all cases and may be due to abnormal thyroid gland formation (dysgenesis) or defective thyroid hormone syntheses by a structurally normal gland (dyshormonogenesis). Central congenital hypothyroidism has a reported incidence of up to one in 16 000 live births. Central congenital hypothyroidism may occur in isolation but in the majority of cases (around 75%) is accompanied by additional pituitary hormone deficiencies.

Thyroid Development and Thyroid Hormone Biosynthesis

Thyroid development and thyroid hormone biosynthesis are reviewed in detail in Section 3. In short, the developing thyroid is first visible at the end of the third embryonal week as an endodermal thickening in the floor of the pharynx at the most anterior part of the foregut [10]. This median thyroid anlage enlarges and migrates downward. Initially the thyroid anlage stays connected to the pharyngeal endoderm via the thyroglossal duct but detaches from it by the end of the fourth embryonal week. It migrates further along the neck, forms two lobes and fuses with the lateral thyroid anlagen, the ultimobranchial bodies. The thyroid reaches its final, pretracheal position in the seventh embryonic week. After this, cellular differentiation and folliculogenesis continue. In the tenth embryonal

week, the thyroid is completely formed, and at the end of the first trimester thyroid hormone synthesis is achieved. Various transcription factors are involved in these stages of thyroid development. The four transcription factors NKX2-1, FOXE1, PAX8, and HHEX are expressed from the first development, while the TSH-receptor (TSHR), thyroglobulin (Tg), sodium-iodide symporter (NIS/SLC5A5) and thyroid peroxidase (TPO) are expressed in the stage of folliculogenesis [10]. Mutations in encoding genes may result in thyroid dysgenesis or thyroid dyshormonogenesis.

Development of Thyroid Hormone Synthesis During Gestation

Thyroid hormones are detectable in the fetal circulation at the end of the first trimester, although it is likely that a fraction of these hormones is of maternal origin [11]. Initially the concentrations of thyroxine (T_4) and thyroxine-binding globulin (TBG) are very low but subsequently increase to reach adult values at about 36 weeks. Fetal serum triiodothyronine (T_3) concentrations are low due to relatively low placental deiodinase type D1 activity, while reverse T_3 levels are high due to high placental deiodinase type D3. At term, the T_3 concentration are still low compared to the normal values for infants and adults. Birth induces a number of changes in thyroid hormone production and metabolism within a short period. This adaptation process starts with an acute surge of TSH into the circulation, reaching a maximum level around 30 min after birth. Thereafter, it gradually decreases and stabilizes within 1–2 days at slightly higher values than those in adults. The TSH surge significantly increases the thyroid hormone production. Serum T_4 and T_3 reach maximum levels approximately 24 hours after birth. Serum reverse T_3 concentration decreases rapidly in the first week of life caused by the loss of placental deiodinase D3 activity. In preterm infants, the functional maturation of the thyroid at birth is incomplete. Timing of the TSH surge is similar to that of term neonates, but quantitatively lower, especially in preterm infants with respiratory distress syndrome. During the first day following the TSH surge, increasing T_4 and T_3 concentrations can be observed, but the T_4 and T_3 peak levels are lower as the pregnancy is shorter and in the case of complications, such as intrauterine growth retardation and respiratory distress syndrome, a nadir is observed at about one week after birth, followed by a second TSH increase. This decrease in T_4 concentrations in premature infants, is referred to as transient hypothyroxinaemia of prematurity. Although premature neonates have lower postpartum free T_4 (fT_4) levels than would be normal for intrauterine life at the same age, administration of T_4 immediately after birth has no significant influence on mortality and morbidity, except for extremely premature infants (less than 27 weeks gestation) in whom such a temporary T_4 supplement may be beneficial for brain development [12].

Primary Congenital Hypothyroidism

Thyroid dysgenesis has long been reported to account for 85% of cases of primary congenital hypothyroidism, while dyshormonogenesis only accounts for 15%. However, in the last decade the contribution of thyroid dyshormonogenesis has increased and is reported to be 30–40%. Especially milder and often transient cases are increasingly identified, most likely resulting from the lowering of TSH cut-offs in neonatal screening [13].

Thyroid Dysgenesis

Thyroid dysgenesis may result in mild to severe hypothyroidism. The gland may be completely absent (agenesis) or remnants of variable size may be present along the tract of the thyroglossal duct (see Figure 9.5.1). These structures, called dystopic (synonym: ectopic) remnants are often localized in the sublingual area. Agenesis is characterized by complete absence of any thyroid tissue (indicated by technetium-99m or 123-I scintigraphy and ultrasound imaging), and complete inability to produce thyroid hormone and thyroglobulin. However, patients with negative 123-I scintigram and (almost) complete absence of circulating thyroid hormone, but clearly measurable serum thyroglobulin levels, have been described [14]. As the thyrocytes are the only cells able to produce thyroglobulin, this cell type has to be present, although it cannot be localized. We

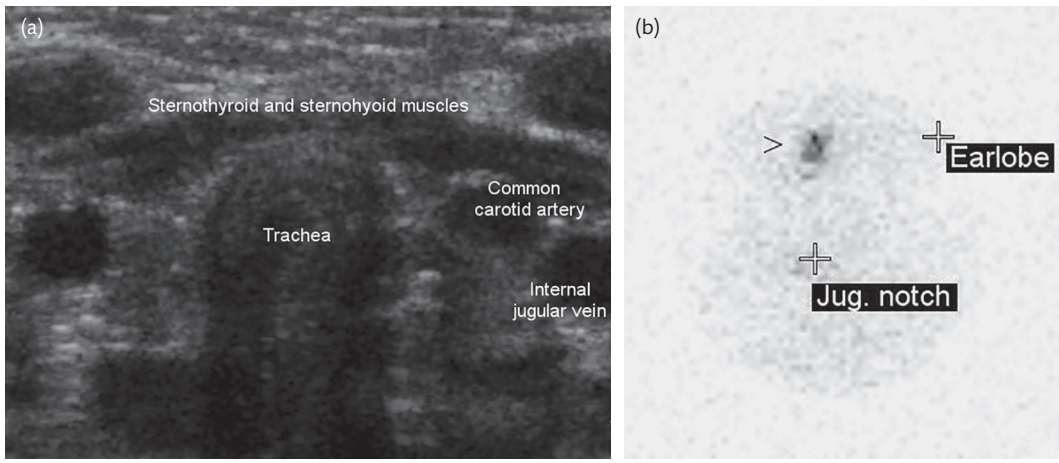


Figure 9.5.1 Ultrasound (a) and thyroid $^{123\text{-}I}$ scintigraphy (b) of the neck and head/neck, respectively, of a 14-day-old boy with primary congenital hypothyroidism due to thyroid dysgenesis. Ultrasonography showed no thyroid tissue in the trachea-'muscles'-carotid artery triangle. Scintigraphy showed a dystopic thyroid remnant (>).

introduced the term 'cryptic thyroid remnant' to describe this type of disorder.

In most cases it remains unclear why the migration and development of the thyroid becomes disturbed. Approximately 2% of cases are familial [15–17]. Although this familial occurrence strongly suggests the presence of genetic factors, mutations in genes known to be involved in thyroid development still explain not more than ten percent of all cases of dysgenesis. The high discordance between monozygotic twins (92%) suggests the involvement of other mechanisms than mutations in these genes [18]. Thyroid dysgenesis usually occurs in isolation but in very rare cases may be associated with other abnormalities. The most frequently encountered genetic form of isolated thyroid dysgenesis is due to TSH-receptor defects. The phenotype varies from severe congenital hypothyroidism due to a hypoplastic or even absent thyroid gland to mild TSH elevation only. It is an autosomal recessive disorder although a few monoallelic mutations have been shown to cause mild TSH elevation in a dominant way, with intracellular entrapment of wild-type TSH receptors by mutant receptors as the underlying mechanism is [19–21]. Other rare genetic forms of isolated thyroid dysgenesis are due to defects in the *PAX8* and *CDCA8 (or BOREALIN*) genes. Rare syndromic forms of thyroid dysgenesis include Bamforth–Lazarus syndrome (*FOXE1* gene defects), brain–lung–thyroid syndrome (*NKX2-1* gene defects), 22q11.2 deletion syndrome (*TBX1* gene defects), *GLIS3* gene defects (associated with permanent neonatal diabetes) and Alagille syndrome (*JAG1* gene defects) (see Table 9.5.1). The frequency of congenital heart defects is increased in patients with thyroid dysgenesis. This is probably explained by the close relation of thyroid development to the adjacent mesenchyme and the developing heart. Genes associated with both congenital heart defects and thyroid dysgenesis are *NKX2-5* and *NTN1*.

Thyroid Dyshormonogenesis

Dyshormonogenesis is characterized by a defect in one of the molecular steps involved in thyroid hormone synthesis within, and at the apical side of the thyroid follicular cells. In dyshormonogenesis, the thyroid gland has a normal position and structure. However, decreased thyroid hormone levels result in increased pituitary TSH secretion and ongoing thyroid gland stimulation, leading to thyroid gland enlargement (goitre); at birth or developing later on, especially when thyroid hormone treatment is delayed or with low patient compliance. Biochemically, patients present with varying degrees of hypothyroidism (low or normal fT_4, with high TSH). Many dyshormonogenesis cases are detected by neonatal screening but mild cases may be missed. Maternal iodide intake may influence the postnatal thyroid hormone levels. In general, thyroglobulin levels are increased in thyroid dyshormonogenesis, except in thyroglobulin synthesis defects. Thyroid imaging by ultrasound or scintigraphy using 123-I or technetium-99m may distinguish thyroid dysgenesis from dyshormonogenesis. However, radionuclide uptake may be low or absent in sodium-iodide symporter defect, emphasizing the importance of combining ultrasound and scintigraphy in the diagnostic work-up. Administration of perchlorate after 123-I scintigraphy may identify iodide organification defects; perchlorate blocks the sodium-iodide symporter, resulting in discharge of unbound 123-I in case of an organification defect. Normally, discharge is less than 10%. Discharge >90% indicates total iodide organification defect, while values between 10% and 90% indicate partial iodide organification defect (see Figure 9.5.2). Abnormal discharge is expected in *TPO, DUOX2, DUOXA2, SLC26A4*, and *SLC26A7* gene defects, while normal discharge is expected in *SLC5A5, TG*, and *IYD* defects. In contrast to thyroid dysgenesis, which is usually sporadic, most cases of dyshormonogenesis are hereditary in an autosomal recessive manner, with the exception of a few cases of dominant inheritance [22, 23]. Another difference between thyroid dysgenesis and dyshormonogenesis is the large phenotypic variability in thyroid dyshormonogenesis, with variable expression and penetrance of genetic defects. Lack of clear genotype-phenotype correlation may point to a role for other genetic or environmental influences, for instance multigenic involvement or dietary iodine intake. Increasing iodine intake may improve certain forms of thyroid dyshormogenesis (*DUOX2, DUOXA2, SLC26A4, SLC26A7, IYD*, and partial *SLC5A5*).

Defects in Iodine Transport

The first step in thyroid hormone synthesis is the active transport of iodide into the thyrocytes. Iodide transport across the basal membrane is mediated by the sodium-iodide symporter (NIS). NIS is also present in a number of non-thyroidal tissues, like the lactating mammary gland, stomach, salivary and lacrimal glands. NIS deficiency results in insufficient uptake of iodide as substrate for thyroid hormone synthesis. It is an autosomal recessive disorder characterized by goitre and hypothyroidism of variable severity. Diagnostic characteristics are (1) limited or absent radioiodide uptake in a normally positioned thyroid gland; (2) elevated blood thyroglobulin levels; and (3) a low saliva (radio-) iodide concentration resulting in low (radio-)iodide saliva/plasma ratio. Clinically there is marked heterogeneity which may be related to differences in iodide intake. Although high iodide intake may compensate for impaired NIS function, treatment with levothyroxine (LT4) is preferred, especially in young children.

Defects in the Synthesis of Thyroglobulin

Thyroglobulin is a very large glycoprotein homodimer, that serves as a macromolecular precursor for thyroid hormone synthesis and storage. Thyroglobulin deficiency is an autosomal recessive disorder and causes goitre with mild to severe hypothyroidism in combination with typically low or undetectable serum thyroglobulin levels. Abnormal iodoproteins/-peptides may be found in blood and urine. Upon scintigraphy radionuclide uptake (123-I) is high with a normal perchlorate discharge. The estimated prevalence is 1 in 100 000 newborns with a probably somewhat higher prevalence in Japan (1 in 67 000) [24]. *TG* defects show a large clinical heterogeneity, ranging from simple goitre to a very large thyroid with severe hypothyroidism. Molecular studies of different mutations have provided more insight into thyroglobulin transport/trafficking, which explains some of the clinical heterogeneity. For example, mutations in the acetylcholinesterase-like domain and mutations replacing cysteine residues inhibit thyroglobulin from being transported out of the endoplasmic reticulum into the Golgi apparatus, causing premature degradation as seen in other endoplasmic reticulum storage diseases [25, 26].

Table 9.5.1 Genetic causes of primary congenital hypothyroidism

Gene name	Inheritance	Thyroid location and size	Associated characteristics
Thyroid dysgenesis			
CDCA8 (BOREALIN)	AD/AR	Athyreosis Thyroid ectopy Hemiagenesis Asymmetric thyroid size	
FOXE1	AR	Athyreosis Severe hypoplasia/normal position	Cleft palate, choanal atresia, spikey hair, bifid epiglottis (Bamforth-Lazarus syndrome)
GLIS3	AR	Athyreosis Normal size and position	Neonatal diabetes mellitus, renal cystic dysplasia, hepatic fibrosis, congenital glaucoma, skeletal abnormalities
JAG1	AD	Athyreosis Thyroid ectopy Normal size and position	Alagille syndrome
NKX2-1	AD	Athyreosis hypoplasia, normal position Normal size and position	Choreoathetosis, respiratory problems (brain-Lung-Thyroid syndrome)
NKX2-5	unclear	Athyreosis Thyroid ectopy	Congenital heart disease
NTN1	unknown	Thyroid ectopy	Congenital heart disease
PAX8	AD	Athyreosis Thyroid ectopy Hypoplasia, normal position Normal size and position	Renal abnormalities
TBX1	AD	Hypoplasia, normal position Hemiagenesis Left lobe hypoplasia	22q11.2 deletion syndrome (DiGeorge syndrome, velocardiofacial syndrome)
TSHR	AR/AD	Athyreosis Hypoplasia, normal position Normal size and position	

Gene name	Inheritance	Thyroid size	Serum thyroglobulin	Functional study Radioiodine uptake + perchlorate discharge	Associated characteristics
Thyroid dyshormonogenesis					
DUOX2	AR/AD	normal to enlarged	elevated	high uptake, abnormal discharge	
DUOXA2	AR	normal to enlarged	elevated	high uptake, abnormal discharge	
IYD (DEHAL1)	AR	normal to enlarged	elevated	high uptake, normal discharge	
SLC26A7	AR	normal to enlarged	elevated	high uptake, abnormal discharge	
SLC5A5 (NIS)	AR	normal to enlarged	elevated	absent to low uptake	
SLC26A4 (PENDRIN)	AR	normal to enlarged	elevated	high uptake, abnormal discharge	sensorineural deafness
TG	AR	normal to enlarged	decreased/ undetectable	high uptake, normal discharge	
TPO	AR	normal to enlarged	elevated	high uptake, abnormal discharge	

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

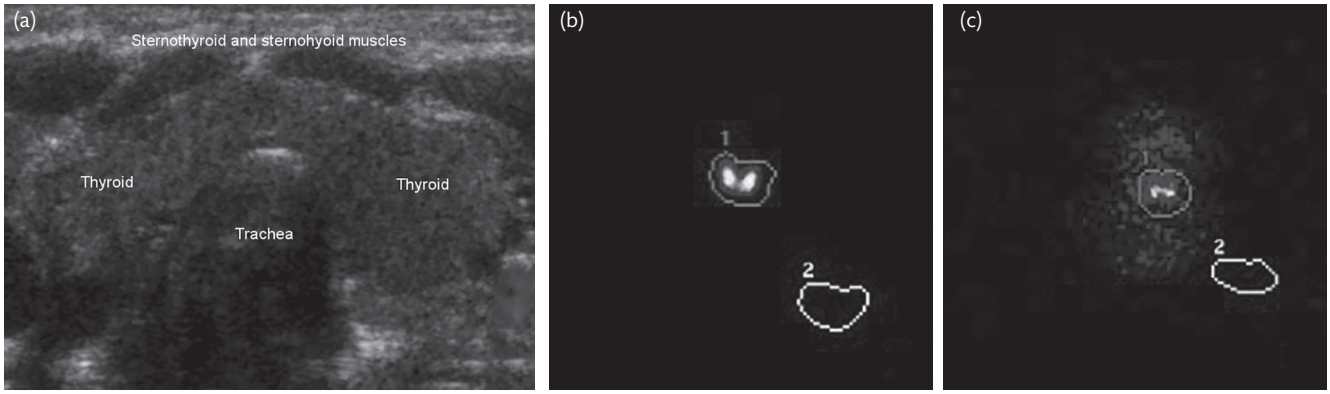


Figure 9.5.2 Ultrasound (a) and thyroid $^{123\text{-}I}$ scintigraphy before (b) and after (c) perchlorate administration in a 14-day-old boy with primary congenital hypothyroidism due to thyroid dyshormonogenesis. Ultrasonography and scintigraphy showed a normally localized thyroid. After the administration of perchlorate, $^{123\text{-}I}$ uptake decreased from 20.7% to 4.3%, confirming the diagnosis iodide organification defect.

Defects in Iodination of Thyroglobulin

Thyroperoxidase (TPO) is an enzyme located at the apical membrane of the follicular cell, and responsible for iodide oxidation, tyrosyl iodination, and iodotyrosine coupling. *TPO* defects are a common cause of permanent congenital hypothyroidism due to dyshormonogenesis. The estimated prevalence in a Dutch cohort was 1 in 66 000 newborns [27]. It is an autosomal recessive disorder usually leading to a total iodide organification defect, but also to partial defects. Hypothyroidism, especially in total iodide organification defects, is severe and associated with goitre. Thyroglobulin levels are elevated. Thyroid scintigraphy shows high radionuclide uptake and an abnormal perchlorate discharge test indicating defective organification. In case of total organification defect (discharge >90%) *TPO* gene defects are the most likely candidate, with *DUOX2* as an alternative. In case of partial organification defects (discharge 10–90%) other causes should also be considered, such as *DUOXA2* and *SLC26A4* mutations [28].

The enzymes dual oxidase 1 and 2 (DUOX1 and DUOX2), are responsible for thyroidal H_2O_2 generation at the apical membrane of the thyroid follicular cell. DUOX2 seems to be the dominant factor since it is more highly expressed in the thyroid than DUOX1. In 2002, the first four cases of *DUOX2* mutation causing congenital hypothyroidism were described [29]. *DUOX2* variants are found in high frequency in cohorts of congenital hypothyroidism patients with partial iodide organification defects and is one of the most frequent causes thyroid dyshormonogenesis. In a Chinese study of patients with overt and subclinical congenital hypothyroidism the *DUOX2* mutation rate was as high as 29% [30]. Recently, *DUOX2* mutations were found in patients with thyroid dysgenesis as well, suggesting a role of *DUOX2* also in thyroid dysgenesis [31]. The phenotypic variability of *DUOX2* mutations may in part be explained by other H_2O_2 generating systems compensating for the defective DUOX2 function, such as DUOX1. The finding of digenic *DUOX1* and *DUOX2* mutations emphasizes that congenital hypothyroidism may be oligogenic and that differences in severity may be explained by the presence of other genetic defects [32]. Other factors possibly explaining the phenotypic variability are thought to be age-related differences in thyroid hormone requirement, and differences in iodide intake [33].

DUOXA1 and DUOXA2 are endoplasmic reticulum proteins that are required for the maturation and plasma membrane localization of the DUOX enzymes [34]. In 2008 the first case of a *DUOXA2* mutation was reported in a case of mild permanent congenital hypothyroidism due to a partial iodide organification defect [35]. Since then only a few cases have been reported. *DUOXA2* mutations also seem to be associated with a variable phenotype, ranging from permanent to transient congenital hypothyroidism in the neonatal period. Just like for DUOX there seems to be an age-dependent decrease in severity and compensating DUOX1/DUOXA1 H_2O_2 generation is thought to influence severity. A high incidence of a specific *DUOXA2* mutation (Y138X) was found in a Korean population, emphasizing ethnic differences in prevalence [36].

SLC26A4 (or pendrin) is a member of the solute carrier family26A. It is localized to the apical membrane of the thyroid follicular cell and mediates iodide efflux from the cellular side to the follicular lumen, making it available for oxidation and tyrosyl iodination. In the inner ear SLC26A4 maintains acid-base homeostasis of the endolymphatic fluid [37]. Pendred syndrome is an autosomal recessive disorder characterized by sensorineural deafness, goitre, and a partial iodide organification defect. Patients present with hearing loss from infancy or later in childhood. Impaired hearing is associated with inner ear malformation (enlarged vestibular aquaduct) which can be visualized by MR or CT imaging. Development of goitre and hypothyroidism is influenced by iodide intake, and usually develops later in childhood or adolescence. Therefore, patients are mostly missed by neonatal screening. The incidence of Pendred syndrome is estimated at 7.5 to 10 in 100 000 individuals and accounts for approximately 10% of all hereditary forms of hearing loss. SLC26A7 is an anion transporter that was recently found to be localized at the apical membrane of the thyroid follicular cells. Although its precise function remains to be elucidated it seems to be involved in iodide transport. *SLC26A7* gene defects have been described in patients with goitrous congenital hypothyroidism exhibiting a partial iodide organification defect.

Defects in the Recycling of Iodide

The enzyme iodotyrosine deiodinase (IYD), also known as iodotyrosine dehalogenase 1 (DEHAL1) degrades monoiodotyrosines (MIT) and diiodotyrosines (DIT) into iodide and tyrosine,

recycling them for hormonogenesis [38]. IYD deficiency leads to increased blood iodotyrosines (MIT and DIT) levels and increased urinary excretion. The resulting urinary iodide loss leads to iodide deficient hypothyroidism and goitre. IYD deficiency is a very rare autosomal recessive disorder with *IYD* defects only reported in seven patients so far (three missense mutations and one deletion). Since the onset of the hypothyroidism and goitre is variable and not always present at birth, IYD deficiency is probably missed in neonatal screening programmes. This genetic form of hypothyroidism should be considered in case of nonautoimmune hypothyroidism and goitre occurring in childhood/adolescence and perhaps even later on in life.

Central Congenital Hypothyroidism

In central congenital hypothyroidism deficient thyroid hormone production is due to insufficient TSH stimulation of an otherwise normal thyroid gland [39]. If the defect is limited to pituitary thyrotrope cells, central hypothyroidism may be isolated. In most cases it is associated with other pituitary hormone deficiencies. In combined pituitary hormone deficiency, growth hormone deficiency and adrenocorticotrophic hormone (ACTH) deficiency are potentially brain damaging and life-threatening, due to hypoglycaemia and circulatory insufficiency. Isolated central congenital hypothyroidism may be due to defects in thyrotropin releasing hormone receptor gene (*TRHR*), thyroid stimulating hormone β-subunit gene (*TSHB*), and the more recently described genes *IGSF1, TBL1X*, and *IRS4* [40–46]. *TRHR* and *TSHB* gene mutations are found only rarely. Since the first cases of *IGSF1* gene mutations were reported in 2012, many new cases have been described making *IGSF1* mutations a frequent cause of isolated central hypothyroidism [47]. *TBL1X* and *IRS4* have been described only very recently and their prevalence still needs to be investigated in large cohorts [45, 46]. The majority of central congenital hypothyroidism cases are not isolated but occur within the framework of combined pituitary hormone deficiencies. Mutations in transcription factors involved in pituitary formation/differentiation are found in less than 5% of patients (see **Table 9.5.2**). Most patients with congenital combined pituitary hormone deficiency exhibit a pituitary malformation characterized by an absent or thin pituitary stalk, a hypoplastic anterior pituitary lobe and an ectopic posterior pituitary lobe (see **Figure 9.5.3**). This malformation is known as pituitary stalk interruption syndrome (PSIS) [48]. PSIS often occurs in isolation but may be accompanied by additional midline brain abnormalities. Since monogenic causes of isolated PSIS are only rarely detected, a polygenic and multifactorial aetiology is more likely [49].

Table 9.5.2 Genetic causes of congenital central hypothyroidism

Gene name	Inheritance	Associated characteristics
Isolated central hypothyroidism		
IGSF1	XL	Prolactin deficiency[1], growth hormone deficiency[2], delayed maturation, macro-orchidism in males, ovarian cysts in females
IRS4	XL	
TBL1X	XL	Hearing deficits
TRHR	AR	
TSHB	AR	
Central hypothyroidism within combined pituitary hormone deficiencies[3]		
CHD7	AD	CHARGE syndrome, PSIS
FGF8	AR	Kallmann syndrome, holoprosencephaly, corpus callosum agenesia
FGFR1	AD	Kallmann syndrome, septo-optic dysplasia, PSIS
FOXA2	AD	Craniofacial and endoderm-derived organ abnormalities, hyperinsulinism
HESX1	AR/AD	Septo-optic dysplasia
LEPR	AR	Severe obesity and hyperphagia
LHX3	AR	Short, rigid cervical spine, variable deafness, small to large pituitary
LHX4	AR/AD	Cerebellar anomalies, PSIS, Arnold-Chiari malformation, corpus callosum hypoplasia
NFKB2	AD	Variable immune deficiency (DAVID syndrome)
OTX2	AD	Ocular anomalies, PSIS
POU1F1	AR/AD	Midface hypoplasia
PROKR2	AR/AD	Septo-optic dysplasia, PSIS
PROP1	AR	Small to large pituitary
SOX2	AD	Microphthalmia, variable learning difficulties, pituitary hypoplasia
SOX3	XL	Variable mental retardation, PSIS

[1] 61% of hemizygous males, and 22% of heterozygous females; [2] 16% of hemizygous males in childhood [47]; [3] combination of hormone deficiencies is variable [39].
Abbreviations: AD, autosomal dominant; AR, autosomal recessive; PSIS, pituitary stalk interruption syndrome; XL, X-linked.

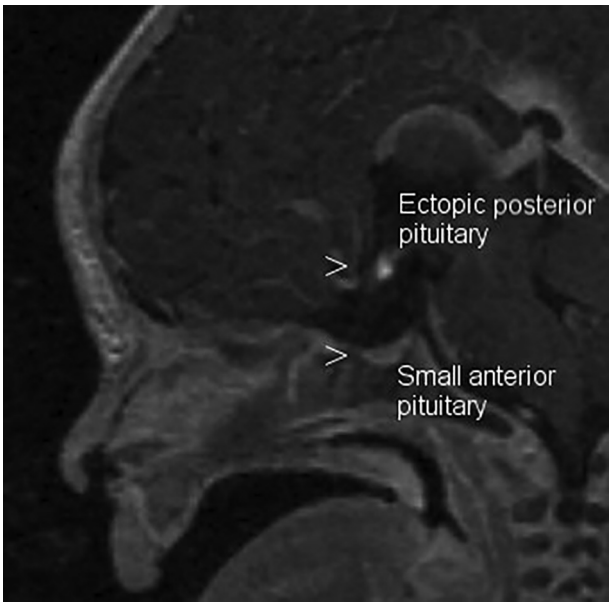


Figure 9.5.3 MRI of the hypothalamic-pituitary region of a 4-week-old girl with pituitary stalk interruption syndrome consisting of the classic triad of a small anterior pituitary lobe, an ectopic posterior pituitary lobe, and an absent pituitary stalk.

Albright's Hereditary Osteodystrophy and Down's Syndrome

Albright's hereditary osteodystrophy (AHO) is characterized by short stature, obesity, round face, subcutaneous ossifications and brachydactyly. In the late 1980s AHO was linked to mutations in the *GNAS* gene, explaining the resistance to parathyroid hormone and TSH often seen in these patients [50]. Nowadays, four types pseudohypoparathyroidism (PHP) are recognized, of which two are associated with hormonal resistance: PHP1a (AHO phenotype plus hormonal resistance to PTH, TSH, GnRH, LH and FSH, and PHP1b (hormonal resistance only). In PHP1a the severity of hypothyroidism ranges from subclinical (elevated TSH with normal fT_4) to manifest (elevated TSH and low fT_4). In PHP1a, congenital hypothyroidism is the earliest clue to this diagnosis in up to 30% of cases, but not in PHP1b which is associated with milder hypothyroidism [51]. Down's syndrome, caused by the presence of (part of) an extra chromosome 21, has been associated with congenital hypothyroidism for many decades [52]. In the last few years, it has become clear that, as a group, Down's syndrome children have a mild form of primary congenital hypothyroidism, probably caused by subtle thyroid dysgenesis which might be caused by overexpression of the chromosome 21 gene DYRK1A [53–55]. Some children have manifest hypothyroidism which is an undisputed treatment indication. However, most children have mild subclinical disease. LT4 treatment of young Down's syndrome children with subclinical disease during their first 2 years of life seemed to improve cognitive and motor development, but in the long term may only result in somewhat better growth [56, 57].

Transient Congenital Hypothyroidism

Exposure of the neonate or fetus to excessive quantities of iodine (e.g. iodine-containing radiographic contrast agents or disinfectants) may lead to transient primary congenital hypothyroidism. Incidentally, maternal TSH receptor blocking antibodies may cause transient primary congenital hypothyroidism for several weeks or months, depending on the initial concentration of circulating antibodies [58]. In both situations affected neonates may be detected by neonatal screening and selected for LT4 treatment during the first months of life. Maternal antithyroid drug use for hyperthyroidism is a third cause transient congenital hypothyroidism.

Transient central congenital hypothyroidism has been reported in newborns of women with uncontrolled Graves' hyperthyroidism during pregnancy. It may be caused by exposure of the fetal hypothalamic–pituitary–thyroid system to higher than normal thyroid hormone concentrations, impairing its physiological maturation during intrauterine life. Although transient, the central hypothyroidism may persist for several months to over a year, necessitating treatment [59].

Signs and Symptoms of Congenital Hypothyroidism

Due to the protective effects of maternal thyroid hormone placental transfer, newborns with congenital hypothyroidism are typically asymptomatic at birth. Classic signs and symptoms of untreated congenital hypothyroidism include large fontanels, macroglossia, abdominal distention, umbilical hernia, lethargy, hoarse cry, poor feeding, hypothermia, constipation, prolonged jaundice, hypotonia with delayed reflexes, bradycardia, and a mottled dry skin. In cases of thyroid dyshormonogenesis, goitre may be present but hardly ever large enough to obstruct neck flexion during delivery or cause airway obstruction. In syndromic forms of primary congenital hypothyroidism certain specific features may be present; cleft palate, choanal atresia, spikey hair and bifid epiglottis (Bamforth–Lazarus syndrome due to *FOXE1* gene defects), neurological and respiratory problems (brain–lung–thyroid syndrome due to *NKX2-1* gene defects), congenital heart disease (*NKX2-5* and *NTN1* gene defects), neonatal diabetes mellitus (*GLIS3* gene defects), renal abnormalities (*PAX8* gene defects), and sensorineural hearing loss (Pendred syndrome due to *SLC26A4* gene defects) (see Table 9.5.1).

In central congenital hypothyroidism signs of additional pituitary hormone deficiencies may be hypoglycaemia due to growth hormone and/or adrenocorticotropic hormone deficiency, or micropenis and/or undescended tests due to gonadotropin deficiency. In rare instance, diabetes insipidus may be present due to vasopressin deficiency (usually in case of septo-optic dysplasia). Other midline abnormalities may be present such as cleft palate, ocular abnormalities, and solitary central incisor.

Neonatal Screening

In the 1970s it became possible to determine T_4 and TSH concentrations in just a few drops of blood, obtained by a heel puncture, and absorbed in filter paper. Since then many countries have implemented neonatal mass-screening programmes for congenital hypothyroidism. Worldwide, the majority of these screening programmes are TSH-based and effectively detect cases of primary congenital hypothyroidism. Central congenital hypothyroidism is only detected

in programmes that also measure T_4 or fT_4 [60]. The Dutch neonatal screening programme consists of a three-step approach with a primary T_4 measurement followed by additional TSH measurement in the lowest 20% T_4 concentrations, and an additional measurement of T_4-binding globulin (TBG) in the lowest 5%. This screening strategy has proven to be effective in detecting both primary and central forms of congenital hypothyroidism [61, 62]. The incidence of permanent central congenital hypothyroidism in the Netherlands is one of the highest in the world, around 1 in 16 000 neonates [63].

Diagnosis and Treatment of Congenital Hypothyroidism

In a newborn with an abnormal neonatal screening result a venous blood sample should be taken for TSH and fT_4 measurement in order to confirm the diagnosis. A low fT_4 in combination with a high TSH level indicates primary hypothyroidism, while a low fT_4 together with a low, normal, or mildly elevated TSH points to central hypothyroidism. A normal fT_4 together with an elevated TSH is referred to as subclinical hypothyroidism. In central hypothyroidism TSH may be mildly elevated due to the presence of TSH with reduced bioactivity. This may complicate distinguishing mild primary hypothyroidism from central hypothyroidism. In a large group of neonates with central congenital hypothyroidism it was found that their TSH levels at diagnosis were never higher than 12.9 mIU/L [64]. Current guidelines advise starting levothyroxine (LT4) treatment immediately without waiting for the venous blood test result unless these results are available the same day [65]. In case of borderline screening results the decision to treat may be delayed until after the venous blood results are known. In these cases, further diagnostic studies may aid in obtaining a definitive diagnosis. Diagnostic work-up includes imaging studies (thyroid ultrasound, or scintigraphy using 123-I followed by perchlorate administration or technetium-99m), measurement of serum thyroglobulin and serum/salivary/urine iodine measurement (see sections on thyroid dysgenesis and thyroid dyshormonogenesis) and measurement of maternal and neonatal anti-TSH receptor antibodies. In case of suspected central hypothyroidism additional pituitary function tests are needed to search for possible accompanying pituitary hormone deficiencies. PSIS can be shown or ruled out by magnetic resonance imaging of the brain.

Treatment for primary congenital hypothyroidism should be started with an initial LT4 dose of 10–15 μg/kg once a day [65]. LT4 tablets should be crushed and mixed with a bit of water, breast milk, or formula and administered on a small spoon. In case of a severely hypothyroid neonate, the body's T_4 deficit can be corrected by one additional LT4 dose, 12 hours after the initial dose. There is no evidence of benefit of adding LT3 (liothyronine) to the LT4 treatment. T_4 absorption may be compromised by certain substances such as soya formulas, iron, and calcium preparations. To prevent under- or overtreatment, the fT_4 and TSH concentrations should be monitored at regular intervals. Recommended follow-up is every 2 weeks until serum TSH is normalized, every 1 to 3 months during the first year of life, every 2 to 4 months between 1 and 3 years of age and every 6 to 12 months thereafter [65]. Due to decreasing T_4 turn over with age, the required dose of 10–15 μg/kg in the first months of life decreases to around 4–5 μg/kg at the age of 5 years and to around 1.6 μg/kg in adults.

The short-term goals of LT4 treatment in the neonatal period are to normalize the fT_4 concentration within 2 weeks and TSH concentration within 4 weeks. When treating children with severe congenital hypothyroidism, normalization of TSH concentrations can sometimes only be achieved by establishing supra-physiological fT_4 concentrations. This seems to be due to an altered TSH-fT_4 set point possibly due to the effect of intrauterine hypothyroidism on the maturing hypothalamic–pituitary–thyroid axis [66]. While this set point change seems to normalize with age there is evidence of persistence [67]. In these cases, we advise keeping fT_4 concentrations around the upper limit of the reference interval and accepting slightly elevated TSH concentrations, although this may lead to goitre in cases of dyshormonogenesis.

In case of primary congenital hypothyroidism without definitive diagnosis, treatment may be safely discontinued at the age of 3 years to evaluate the permanent or transient character of the hypothyroidism.

For the treatment of central congenital hypothyroidism, a starting dose of approximately 6–8 μg/kg per day is suitable. In central hypothyroidism TSH levels are useless for therapy control and one has to rely on fT_4 concentrations. The aim is to keep fT_4 concentrations in the upper half of the age specific reference range. It is important to treat accompanying ACTH deficiency before starting T_4 supplementation. Genetic testing for mutations in *TRHR, TSHB, IGSF1, TBL1X*, and *IRS4* genes will substantiate a diagnosis of isolated central hypothyroidism.

The main treatment goal in congenital hypothyroidism is preventing brain damage, especially in the first 3 years of life when brain development is critically dependent on thyroid hormone. Long-term effect evaluation of the cognitive and motor development of patients with congenital hypothyroidism has demonstrated that timely and adequate LT4 treatment results in neuropsychological test results within the normal range [8, 68]. However, especially in severe cases of congenital hypothyroidism, mild learning disabilities, fine motor impairment and attention problems are still reported [69, 70]. Also, a higher prevalence of sensorineural hearing loss is reported emphasizing the role of thyroid hormone in cochlear development [71].

Congenital Hyperthyroidism

Congenital or neonatal hyperthyroidism is usually caused by maternal TSH receptor stimulating antibodies crossing the placenta and stimulating the fetal thyroid from mid-gestation onwards. Much rarer causes are activating mutations in the genes encoding the TSH receptor or stimulatory G protein (McCune–Albright syndrome) [72]. Although neonatal hyperthyroidism caused by stimulating antibodies affects only 1–2% of the children of mothers with past or active Graves' disease and is a self-limiting disease, it is associated with considerable morbidity and even mortality if not recognized and left untreated [72, 73]. Prenatal signs may be intra-uterine growth retardation, microcephaly, tachycardia, and goitre. After birth, neonates may be extremely restless and irritable, with an exophthalmos-like appearance, and may display signs of hypermetabolism, for instance tachycardia and sweating, and even multiple organ failure.

In the past decades, it has been shown that the key to preventing morbidity and mortality due to fetal and neonatal hyperthyroidism

caused by anti-TSH receptor antibodies is measurement of these antibodies in pregnant women with past or active Graves' disease: early in, and—if detectable—at mid gestation, and (late) in the third trimester. If antibodies are present at mid gestation, the fetus should be closely monitored. When there are signs of fetal hyperthyroidism, antithyroid drug treatment should be started via the mother. At birth, anti-TSH receptor antibodies, and fT_4 and TSH should be measured in cord blood, and thyroid function should be assessed every 2–3 days thereafter. Predictors of neonatal hyperthyroidism are anti-TSH receptor antibodies concentrations higher than three times the upper limit of the reference interval and higher than normal fT_4 levels [74].

If a neonate has clinical or biochemical signs of hyperthyroidism, treatment should be started consisting of methimazole (0.5–1.0 mg/kg per day, orally in three divided doses) and, depending on the severity of the condition, propranolol (2 mg/kg per day, orally in two divided doses) [72]. In severe cases, treatment with iodine (1 drop of Lugol's solution every 8 h after the start of antithyroid drug therapy; Lugol's solution contains 126 mg iodine/ml) and corticosteroids may be necessary. If heart failure is imminent, digitalization is indicated. When euthyroidism is reached, methimazole treatment should be continued for several months, and LT4 should be added to prevent hypothyroidism ('block and replace'). An alternative approach is to taper the methimazole dose to maintain euthyroidism ('titration'). In our experience, this leads to unwanted fluctuating fT_4 concentrations. Therefore, we prefer the block and replace strategy. In most infants the anti-TSH receptor antibodies will have disappeared by the age of 3–4 months, when treatment can be stopped. In case of congenital hyperthyroidism due to gain-of-function mutation of the TSH-receptor gene, remission will not occur. After initial thyreostatic drug treatment, the only reliable long-term treatment is surgical removal of the thyroid gland, sometimes followed by 131I treatment. Mental impairment, microcephaly, and growth problems may occur when treatment is delayed.

REFERENCES

1. Rovet JF. The role of thyroid hormones for brain development and cognitive function. *Endocr Dev*, 2014; **26**: 26–43.
2. Curling TB. Two Cases of absence of the thyroid body, and symmetrical swellings of fat tissue at the sides of the neck, connected with defective cerebral development. *Med Chir Trans*, 1850; **33**: 303–6.
3. Fagge CH. On sporadic cretinism, occurring in England. *Med Chir Trans*, 1871; **54**: 155–70.
4. Murray GR. Note on the treatment of myxoedema by hypodermic injections of an extract of the thyroid gland of a sheep. *Br Med J*, 1891; **2**: 796–97.
5. Hurxthal LM, Musulin N. Cretinism. *Am J Med*, 1946; **1**: 56–82.
6. Klein AH, Meltzer S, Kenny FM. Improved prognosis in congenital hypothyroidism treated before age three months. *J Pediatr*, 1972; **81**: 912–15.
7. Gruters A, Krude H. Detection and treatment of congenital hypothyroidism. *Nat Rev Endocrinol*, 2011; **8**: 104–13.
8. Krude H, Kuhnen P, Biebermann H. Treatment of congenital thyroid dysfunction: achievements and challenges. *Best Pract Res Clin Endocrinol Metab*, 2015; **29**: 399–413.
9. Leger J, Ecosse E, Roussey M, Lanoe JL, Larroque B, French Congenital Hypothyroidism Study Group. Subtle health impairment and socioeducational attainment in young adult patients with congenital hypothyroidism diagnosed by neonatal screening: a longitudinal population-based cohort study. *J Clin Endocrinol Metab*, 2011; **96**: 1771–82.
10. Nilsson M, Fagman H. Development of the thyroid gland. *Development*, 2017; **144**: 2123–40.
11. Thorpe-Beeston JG, Nicolaides KH, Felton CV, Butler J, McGregor AM. Maturation of the secretion of thyroid hormone and thyroid-stimulating hormone in the fetus. *N Engl J Med*, 1991; **324**: 532–36.
12. van Wassenaer AG, Westera J, Houtzager BA, Kok JH. Ten-year follow-up of children born at <30 weeks' gestational age supplemented with thyroxine in the neonatal period in a randomized, controlled trial. *Pediatrics*, 2005; **116**: e613–18.
13. Deladoey J, Ruel J, Giguere Y, Van Vliet G. Is the incidence of congenital hypothyroidism really increasing? A 20-year retrospective population-based study in Quebec. *J Clin Endocrinol Metab*, 2011; **96**: 2422–9.
14. Djemli A, Fillion M, Belgoudi J, Lambert R, Delvin EE, Schneider W, Van Vliet G. Twenty years later: a reevaluation of the contribution of plasma thyroglobulin to the diagnosis of thyroid dysgenesis in infants with congenital hypothyroidism. *Clin Biochem*, 2004; **37**: 818–22.
15. De Felice M, Di Lauro R. Thyroid development and its disorders: genetics and molecular mechanisms. *Endocr Rev*, 2004; **25**: 722–46.
16. Castanet M, Lyonnet S, Bonaiti-Pellie C, Polak M, Czernichow P, Leger J. Familial forms of thyroid dysgenesis among infants with congenital hypothyroidism. *N Engl J Med*, 2000; **343**: 441–2.
17. Fernandez LP, Lopez-Marquez A, Santisteban P. Thyroid transcription factors in development, differentiation and disease. *Nat Rev Endocrinol*, 2015; **11**: 29–42.
18. Perry R, Heinrichs C, Bourdoux P, *et al.* Discordance of monozygotic twins for thyroid dysgenesis: implications for screening and for molecular pathophysiology. *J Clin Endocrinol Metab*, 2002; **87**: 4072–7.
19. Calebiro D, de Filippis T, Lucchi S, *et al.* Intracellular entrapment of wild-type TSH receptor by oligomerization with mutants linked to dominant TSH resistance. *Hum Mol Genet*, 2005; **14**: 2991–3002.
20. Kleinau G, Neumann S, Gruters A, Krude H, Biebermann H. Novel insights on thyroid-stimulating hormone receptor signal transduction. *Endocr Rev*, 2013; **34**: 691–724.
21. Grasberger H, Refetoff S. Resistance to thyrotropin. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 183–94.
22. Park SM, Chatterjee VK. Genetics of congenital hypothyroidism. *J Med Genet*, 2005; **42**: 379–89.
23. Grasberger H, Refetoff S. Genetic causes of congenital hypothyroidism due to dyshormonogenesis. *Curr Opin Pediatr*, 2011; **23**: 421–8.
24. Hishinuma A, Fukata S, Nishiyama S, *et al.* Haplotype analysis reveals founder effects of thyroglobulin gene mutations C1058R and C1977S in Japan. *J Clin Endocrinol Metab*, 2006; **91**: 3100–4.
25. Medeiros-Neto G, Kim PS, Yoo SE, *et al.* Congenital hypothyroid goitre with deficient thyroglobulin. Identification of an endoplasmic reticulum storage disease with induction of molecular chaperones. *J Clin Invest*, 1996; **98**: 2838–44.
26. Citterio CE, Machiavelli GA, Miras MB, *et al.* New insights into thyroglobulin gene: molecular analysis of seven novel mutations

associated with goitre and hypothyroidism. *Mol Cell Endocrinol*, 2013; **365**: 277–91.
27. Bakker B, Bikker H, Vulsma T, de Randamie JS, Wiedijk BM, De Vijlder JJ. Two decades of screening for congenital hypothyroidism in the Netherlands: TPO gene mutations in total iodide organification defects (an update). *J Clin Endocrinol Metab*, 2000; **85**: 3708–12.
28. Ris-Stalpers C, Bikker H. Genetics and phenomics of hypothyroidism and goitre due to TPO mutations. *Mol Cell Endocrinol*, 2010; **322**: 38–43.
29. Moreno JC, Bikker H, Kempers MJ, *et al*. Inactivating mutations in the gene for thyroid oxidase 2 (THOX2) and congenital hypothyroidism. *N Engl J Med*, 2002; **347**: 95–102.
30. Fu C, Luo S, Zhang S, *et al*. Next-generation sequencing analysis of DUOX2 in 192 Chinese subclinical congenital hypothyroidism (SCH) and CH patients. *Clin Chim Acta*, 2016; **458**: 30–4.
31. Kizys MML, Louzada RA, Mitne-Neto M, *et al*. DUOX2 mutations are associated with congenital hypothyroidism with ectopic thyroid gland. *J Clin Endocrinol Metab*, 2017; **102**: 4060–71.
32. Aycan Z, Cangul H, Muzza M, *et al*. Digenic DUOX1 and DUOX2 mutations in cases with congenital hypothyroidism. *J Clin Endocrinol Metab*, 2017; **102**: 3085–90.
33. Muzza M, Fugazzola L. Disorders of H2O2 generation. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 225–40.
34. Grasberger H, Refetoff S. Identification of the maturation factor for dual oxidase. Evolution of an eukaryotic operon equivalent. *J Biol Chem*, 2006; **281**: 18269–72.
35. Zamproni I, Grasberger H, Cortinovis F, *et al*. Biallelic inactivation of the dual oxidase maturation factor 2 (*DUOXA2*) gene as a novel cause of congenital hypothyroidism. *J Clin Endocrinol Metab*, 2008; **93**: 605–10.
36. Park KJ, Park HK, Kim YJ, *et al*. DUOX2 mutations are frequently associated with congenital hypothyroidism in the Korean Population. *Ann Lab Med*, 2016; **36**: 145–53.
37. Wemeau JL, Kopp P. Pendred syndrome. *Best Pract Res Clin Endocrinol Metab*, 2017; **31**: 213–24.
38. Moreno JC, Visser TJ. Genetics and phenomics of hypothyroidism and goitre due to iodotyrosine deiodinase (DEHAL1) gene mutations. *Mol Cell Endocrinol*, 2010; **322**: 91–8.
39. Persani L, Brabant G, Dattani M, *et al*. 2018 European Thyroid Association guidelines on the diagnosis and management of central hypothyroidism. *Eur Thyroid J*, 2018; **7**: 225–37.
40. Bonomi M, Busnelli M, Beck-Peccoz P, *et al*. A family with complete resistance to thyrotropin-releasing hormone. *N Engl J Med*, 2009; **360**: 731–4.
41. Collu R, Tang J, Castagne J, *et al*. A novel mechanism for isolated central hypothyroidism: inactivating mutations in the thyrotropin-releasing hormone receptor gene. *J Clin Endocrinol Metab*, 1997; **82**: 1561–5.
42. Medeiros-Neto G, Herodotou DT, Rajan S, *et al* A circulating, biologically inactive thyrotropin caused by a mutation in the beta subunit gene. *J Clin Invest*, 1996; **97**: 1250–6.
43. Dacou-Voutetakis C, Feltquate DM, Drakopoulou M, Kourides IA, Dracopoli NC. Familial hypothyroidism caused by a nonsense mutation in the thyroid-stimulating hormone beta-subunit gene. *Am J Hum Genet*, 1990; **46**: 988–93.
44. Sun Y, Bak B, Schoenmakers N, *et al*. Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. *Nat Genet*, 2012; **44**: 1375–81.
45. Heinen CA, Losekoot M, Sun Y, *et al*. Mutations in TBL1X are associated with central hypothyroidism. *J Clin Endocrinol Metab*, 2016; **101**: 4564–73.
46. Heinen CA, de Vries EM, van den Akker ELT, *et al*. Mutations in IRS4 are associated with central hypothyroidism. Washington DC: 10th International Meeting of Pediatric Endocrinology, 2017.
47. Joustra SD, Heinen CA, Schoenmakers N, *et al*. IGSF1 deficiency: lessons from an extensive case series and recommendations for clinical management. *J Clin Endocrinol Metab*, 2016; **101**: 1627–36.
48. Bar C, Zadro C, Diene G, *et al*. Pituitary stalk interruption syndrome from infancy to adulthood: clinical, hormonal, and radiological assessment according to the initial presentation. *PLoS One*, 2015; **10**: e0142354.
49. Zwaveling-Soonawala N, Alders M, Jongejan A, *et al*. Clues for polygenic inheritance of pituitary stalk interruption syndrome from exome sequencing in 20 patients. *J Clin Endocrinol Metab*, 2018; **103**: 415–28.
50. Levine MA, Ahn TG, Klupt SF, *et al*. Genetic deficiency of the alpha subunit of the guanine nucleotide-binding protein Gs as the molecular basis for Albright hereditary osteodystrophy. *Proc Natl Acad Sci U S A*, 1988; **85**: 617–21.
51. Picard C, Decrequy A, Guenet D, *et al*. Diagnosis and management of congenital hypothyroidism associated with pseudohypoparathyroidism. *Horm Res Paediatr*, 2015; **83**: 111–17.
52. Fort P, Lifshitz F, Bellisario R, *et al*. Abnormalities of thyroid function in infants with Down syndrome. *J Pediatr*, 1984; **104**: 545–9.
53. van Trotsenburg AS, Kempers MJ, Endert E, Tijssen JG, de Vijlder JJ, Vulsma T. Trisomy 21 causes persistent congenital hypothyroidism presumably of thyroidal origin. *Thyroid*, 2006; **16**: 671–80.
54. Zwaveling-Soonawala N, Witteveen ME, Marchal JP, *et al*. Early thyroxine treatment in Down syndrome and thyroid function later in life. *Eur J Endocrinol*, 2017; **176**: 505–13.
55. Kariyawasam D, Rachdi L, Carre A, *et al*. DYRK1A BAC transgenic mouse: a new model of thyroid dysgenesis in Down syndrome. *Endocrinology*, 2015; **156**: 1171–80.
56. van Trotsenburg AS, Vulsma T, van Rozenburg-Marres SL, *et al*. The effect of thyroxine treatment started in the neonatal period on development and growth of two-year-old Down syndrome children: a randomized clinical trial. *J Clin Endocrinol Metab*, 2005; **90**: 3304–11.
57. Marchal JP, Maurice-Stam H, Ikelaar NA, *et al*. Effects of early thyroxine treatment on development and growth at age 10.7 years: follow-up of a randomized placebo-controlled trial in children with Down's syndrome. *J Clin Endocrinol Metab*, 2014; **99**: E2722–9.
58. Brown RS, Bellisario RL, Botero D, *et al*. Incidence of transient congenital hypothyroidism due to maternal thyrotropin receptor-blocking antibodies in over one million babies. *J Clin Endocrinol Metab*, 1996; **81**: 1147–51.
59. Kempers MJ, van Tijn DA, van Trotsenburg AS, de Vijlder JJ, Wiedijk BM, Vulsma T. Central congenital hypothyroidism due to gestational hyperthyroidism: detection where prevention failed. *J Clin Endocrinol Metab*, 2003; **88**: 5851–7.
60. Ford G, LaFranchi SH. Screening for congenital hypothyroidism: a worldwide view of strategies. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 175–87.

61. Lanting CI, van Tijn DA, Loeber JG, Vulsma T, de Vijlder JJ, Verkerk PH. Clinical effectiveness and cost-effectiveness of the use of the thyroxine/thyroxine-binding globulin ratio to detect congenital hypothyroidism of thyroidal and central origin in a neonatal screening program. *Pediatrics*, 2005; **116**: 168–73.
62. van Tijn DA, de Vijlder JJ, Verbeeten B, Jr., Verkerk PH, Vulsma T. Neonatal detection of congenital hypothyroidism of central origin. *J Clin Endocrinol Metab*, 2005; **90**: 3350–9.
63. Kempers MJ, Lanting CI, van Heijst AF, *et al*. Neonatal screening for congenital hypothyroidism based on thyroxine, thyrotropin, and thyroxine-binding globulin measurement: potentials and pitfalls. *J Clin Endocrinol Metab*, 2006; **91**: 3370–6.
64. Zwaveling-Soonawala N, van Trotsenburg ASP, Verkerk PH. TSH and FT4 concentrations in congenital central hypothyroidism and mild congenital thyroidal hypothyroidism. *J Clin Endocrinol Metab*, 2018; **103**: 1342–8.
65. Leger J, Olivieri A, Donaldson M, *et al*. European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis, and management of congenital hypothyroidism. *J Clin Endocrinol Metab*, 2014; **99**: 363–84.
66. Fisher DA, Schoen EJ, La Franchi S, *et al*. The hypothalamic-pituitary-thyroid negative feedback control axis in children with treated congenital hypothyroidism. *J Clin Endocrinol Metab*, 2000; **85**: 2722–7.
67. Kempers MJ, van Trotsenburg AS, van Tijn DA, *et al*. Disturbance of the fetal thyroid hormone state has long-term consequences for treatment of thyroidal and central congenital hypothyroidism. *J Clin Endocrinol Metab*, 2005; **90**: 4094–100.
68. Grosse SD, Van Vliet G. Prevention of intellectual disability through screening for congenital hypothyroidism: how much and at what level? *Arch Dis Child*, 2011; **96**: 374–9.
69. Kempers MJ, van der Sluijs Veer L, Nijhuis-van der Sanden RW, *et al*. Neonatal screening for congenital hypothyroidism in the Netherlands: cognitive and motor outcome at 10 years of age. *J Clin Endocrinol Metab*, 2007; **92**: 919–24.
70. Rovet JF. Children with congenital hypothyroidism and their siblings: do they really differ? *Pediatrics*, 2005; **115**: e52–7.
71. Bruno R, Aversa T, Catena M, *et al*. Even in the era of congenital hypothyroidism screening mild and subclinical sensorineural hearing loss remains a relatively common complication of severe congenital hypothyroidism. *Hear Res*, 2015; **327**: 43–7.
72. Leger J. Management of fetal and neonatal Graves' disease. *Horm Res Paediatr*, 2017; **87**: 1–6.
73. Alexander EK, Pearce EN, Brent GA, *et al*. 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **27**: 315–89.
74. van Dijk MM, Smits IH, Fliers E, Bisschop PH. Maternal thyrotropin receptor antibody concentration and the risk of fetal and neonatal thyrotoxicosis: a systematic review. *Thyroid*, 2018; **28**: 257–64.

9.6

Pituitary Tumours in Pregnancy

Wenyu Huang and Mark E. Molitch

Introduction

Pituitary adenomas may cause problems because of over-secretion of hormones by the tumour as well as by causing hypopituitarism, thereby affecting fertility and pregnancy outcome if pregnancy does ensue. In addition, the pregnancy itself alters hormone secretion and pituitary tumour size, complicating the evaluation and management of patients with pituitary neoplasms. The influence on various types of therapy on the developing fetus also affects therapeutic decision-making.

Prolactinoma and Pregnancy

Hyperprolactinemia is responsible for about one third of all cases of female infertility, owing to its inhibition of ovulation [1]. When women with prolactinomas get pregnant, two important issues arise: (1) the effects of the dopamine agonist on early fetal development; and (2) the effect of pregnancy on prolactinoma size [2].

Effects of Dopamine Agonists on the Developing Fetus

For patients with prolactinomas, dopamine agonists are the treatment of choice, as they are effective in correcting the hyperprolactinemia and restoring ovulation in over 90% of women with amenorrhea and anovulation [3–5]. Generally, dopamine agonists should be stopped upon confirmation of pregnancy. Bromocriptine has been shown to cross the placenta in human studies [6]; while cabergoline has been shown to do so in animal studies but such data are lacking in humans. The long half-life of cabergoline means several days of fetal exposure after the drug has been stopped.

With such short-term exposure of generally less than 6 weeks in over 6000 pregnancies, bromocriptine has not been found to cause any increase in spontaneous abortions, ectopic pregnancies, trophoblastic disease, multiple pregnancies, or congenital malformations (Table 1) [7]. A long-term follow-up study of 64 children between the ages of 6 months and 9 years who were exposed to this drug during early gestation showed no ill effects on their development. Bromocriptine has been used throughout gestation in only a little over 100 women, with no abnormalities noted in the infants except for one with an undescended testicle and another with a talipes deformity [8–10].

Experience with the use of cabergoline in early pregnancy is more limited, with just over 1000 cases being reported (Table 1). As with bromocriptine, such use has similarly not shown an increase in adverse pregnancy outcomes (Table 1), including only 2.6% major malformations found in 908 pregnancies [11–20]. Follow-up studies of up to 12 years following exposure to cabergoline during gestation in 232 children showed a slight retardation in verbal fluency in two (0.8%), difficulty in achieving complete continence in one by age 4 (0.4%), seizures in two (0.8%) and 'pervasive developmental disorder', an autism spectrum disorder, in two (0.8%) [19, 21, 22]. A summary of cabergoline use throughout gestation indicated that healthy infants were delivered at term in thirteen and at 36 weeks in one; but one had an intrauterine death at 34 weeks when the mother had severe preeclampsia [20, 23]. Compared to what is expected in the general population in the United States [24, 25], there do not appear to be any increases in adverse pregnancy outcomes with either drug (Table 9.6.1).

In a review of 176 pregnancies, in which quinagolide (not available in the United States) was maintained for a median duration of 37 days into the gestations, Webster reported 24 spontaneous abortions, one ectopic pregnancy, and one stillbirth at 31 weeks of gestation [26]. Nine fetal malformations were reported [26]. Thus, quinagolide does not appear to be safe for the fetus if used when pregnancy is desired.

In contrast to the majority of studies that provide reassuring information regarding cabergoline and bromocriptine, Hurault-Delaure *et al.* reported some adverse outcomes of dopamine agonist use from a French database [27]. Of the 57 408 mother–baby outcome pairs, 183 (0.3%) had received dopamine agonists (bromocriptine 64.5%, cabergoline 20.2%, quinagolide 9.8%) at some time

Table 9.6.1 Pregnancy outcomes summarized for women who became pregnant while taking bromocriptine or cabergoline, compared to what is expected in the normal population

	Bromocriptine (N)	Bromocriptine (%)	Cabergoline (N)	Cabergoline (%)	Normal (%)
Pregnancies	6272	100	1061	100	100
• Spontaneous abortions	620	9.9	77	7.6	10–15
• terminations	75	1.2	66*	6.5	20
• Ectopic	31	0.5	3	0.3	1.0–1.5
• Hydatidiform moles	11	0.2	1	0.1	0.1–0.15
Deliveries (known duration)	4139	100	791	100	100
• at term (>37 weeks)	3620	87.5	715**	90.4	87.3
• preterm (<37 weeks)	519	12.5	76	9.6	12.7
Deliveries (known outcome)	5120	100	670	100	100
• single births	5031	98.3	655	97.8	96.8
• multiple births	89	1.7	15	2.2	3.2
Babies (known details)	5246	100	908	100	100
• normal	5061	98.2	884	97.4	97
• with malformations	95	1.8	24	2.6	3.0

* 14 of these terminations were for malformations; ** 7 of these births were stillbirths.

during their pregnancy (75% in the first trimester) [27]. Compared to a control group of 345 unexposed women matched for age and the month-and-year of the beginning of their pregnancy, dopamine agonist exposure was associated with an increased frequency of preterm birth and an increased rate of early pregnancy loss [27]. There were no differences between the dopamine agonists with respect to these outcomes. No difference in psychomotor development at ages 9 and 24 months was shown between the two groups [27]. The authors mentioned limitations of this study, including the fact that the indication for prescribing dopamine agonists was not known and data about compliance with therapy was not available. Furthermore, exposed and unexposed subjects were not matched for multifetal pregnancy and details of co-existing pathology were not given.

Bromocriptine clearly has the largest safety database and has a proven safety record for pregnancy, the recent French data notwithstanding. The database for the use of cabergoline in pregnancy is smaller, but there is no evidence at present indicating that it exerts deleterious effects on the developing fetus. The risk of malformations with either drug is not greater than that what is found in the general population.

Effect of Pregnancy on Prolactinoma Size

The increasing oestrogen from the placenta stimulates lactotroph hyperplasia and a progressive increase in PRL levels over the course of pregnancy [28, 29]. In normal women, MRI scans show a gradual increase in pituitary volume, beginning by the second month and peaking the first week postpartum with a final height reaching 12 mm [30].

Prolactinomas can enlarge during pregnancy as a result of both the high oestrogen levels and the discontinuation of the dopamine agonist. The risk of symptomatic tumour enlargement in pregnant women with microprolactinomas or macroprolactinomas is presented in Table 2 [2, 31]. In these series, clinically significant tumour enlargement criteria generally consisted of progressive, severe headaches and/or visual field defects. Asymptomatic increases in tumour size without visual defects found on scans were not counted. Furthermore, in some series some pregnancies did not go to term and how tumour progression in such patients was counted was not clear. Table 9.6.2 is a summary of these data but the limitations of these numbers based on the information outlined earlier should be recognized [19–22, 31–42]. The risk of symptomatic tumour enlargement for microadenomas was 2.5% (20/800), for macroadenomas that had not had prior surgery or irradiation was 18.1% (52/288), and for macroadenomas with prior surgery/irradiation was 4.7% (7/148). Another caveat regarding these numbers is that in a small proportion of those cases the tumour enlargement reflected tumour apoplexy. There are a number of additional case reports of pituitary apoplexy occurring in women with both micro- and macroprolactinomas during pregnancy [43, 44], however the true incidence of apoplexy is unknown.

In one study, 34 women who stopped cabergoline shortly after conception had MRI scans carried out between weeks 24 and 32. Of the 12 with macroadenomas, 5 had no change in tumour size, none had a decrease in tumour size, 3 had an increase of <5 mm and 4 had an increase of >5 mm in size; of the 22 with microadenomas, 9 had no change in tumour size, 3 had a decrease in tumour size, 8 had an increase of <5 mm in size and 2 had an increase of >5 mm in size [21]. In most cases with tumour enlargement, reintroduction of the dopamine agonist was successful in reversing the problem and surgery was very rarely required. Cabergoline was restarted in 6 of the patients because of symptomatic headaches and/or vision-threatening increases in size of the adenomas [21, 42].

Postpartum PRL levels and tumour sizes are often reduced as compared with values before pregnancy [45], but this has not been observed in all series [46]. In one series of 56 hyperprolactinaemic women, 23 had normal PRL levels postpartum and of the 33 with persistent hyperprolactinemia 31% had levels that had decreased by more than 50% [42]. In another study, no patients with

Table 9.6.2 Enlargement of prolactinomas during pregnancy

Series	Ref. no.	Year	Total	Microadenomas			Macroadenomas				
							No prior treatment			Prior treatment	
				# Enlarged		Total	# Enlarged		Total	# Enlarged	
Gemzell & Wang	32	1979	85	5		46	20		70	5	
Molitch	33	1985	246	4		45	7		46	2	
Holmgren *et al.*	34	1986	26	3		4	2		5	0	
Ampudia *et al.*	35	1992	8	1		1	0		4	0	
Kuppersmith *et al.*	36	1994	54	0		4	4		0	0	
Rossi *et al.*	37	1995	22	2		3	1		2	0	
Badawy *et al.*	38	1997	16	0		0	0		0	0	
Mallmann *et al.*	39	2002	5	0		3	1		0	0	
Bronstein *et al.*	40	2002	48	1		30	11		21	0	
Ono *et al.*	22	2010	56	0		29	0		0	0	
Lebbe *et al.*	21	2010	45	2		15	3		0	0	
Stalldecker *et al.*	19	2010	47	0		34	0		0	0	
Auriemma *et al.*	41	2013	76	0		10	0		0	0	
Domingue *et al.**	42	2014	30	0		14	1		0	0	
Rastogi *et al.*	31	2016	0	0		33	0		0	0	
Karaca *et al.*	20	2018	36	2		17	2		0	0	
TOTAL			**800**	**20**	**2.5%**	**288**	**52**	**18.1%**	**148**	**7**	**(4.7%)**

* Analysis of tumour growth from this series—personal communication from D. Maiter.

prolactinoma showed a sharp increase of PRL levels or complained of symptoms suggestive of tumour enlargement when breast feeding postpartum [47]. Thus, women with prolactinomas can breast-feed. Reinstitution of dopamine agonist has to await the cessation of breast-feeding and should only be done if the woman remains hyperprolactinaemic and anovulatory.

Management of Prolactinoma During Pregnancy

In women with prolactinoma who are anovulatory, the treatment choice is dopamine agonists or transsphenoidal selective adenoma resection. Dopamine agonists seem to be the best primary treatment because of their efficacy in restoring ovulation and very low (2.5%) risk of clinically serious tumour enlargement. The accumulating data for cabergoline suggest that its safety record is equal to that of bromocriptine and generally it is better tolerated and more efficacious. Transsphenoidal surgery causes a permanent reduction of PRL levels in only 60–70% of cases and entails morbidity and mortality, albeit at low rates [4, 48, 49]. Pregnancy can generally be achieved in over 85% of patients with dopamine agonists or surgery [16, 33, 46, 49]. Radiotherapy is not warranted for patients with microadenomas due to the risk of long-term sequelae, especially hypopituitarism [50].

A patient with a microadenoma who was treated only with a dopamine agonist and had medication withdrawal after conception needs only to be followed clinically throughout gestation. PRL levels do not always rise during pregnancy in women with prolactinomas [51]. Usually PRL levels rise over the first 6–10 weeks after stopping the dopamine agonist and then plateau [52]. PRL levels may also not rise with tumour enlargement [51]. Therefore, periodic checking of PRL levels is not recommended [5]. Because of the low incidence of tumour enlargement, routine visual field testing is not cost effective. Visual field testing and MRI scanning are performed only in patients who become symptomatic and when intervention is being considered. It is important to carry out MRI scans this way, as Karaca *et al*. showed MRI confirmed tumour enlargement only in 4 out of 10 patients suspected clinically to have tumour enlargement [20]. There are no data documenting harm to the developing fetus from MRI scans [53–56]. Recently, slight increases in risks of stillbirths, neonatal deaths, and the broad outcome of any rheumatological, inflammatory, or infiltrative skin condition were shown with first trimester gadolinium exposure [57]. No adverse effects from MRIs with or without gadolinium have been shown in the late second or third trimesters. In clinical practice, it is important to document a significant increase in tumour size before instituting an intervention, such as restarting a dopamine agonist or surgery. If the headache is sudden, it may be caused by pituitary apoplexy, which may require an entirely different management course, including hormone replacement if there is sudden onset of hypopituitarism [43, 44]. MRI may be very helpful in distinguishing between haemorrhage into a tumour vs. tumour enlargement [43].

The patient with a small intrasellar or inferiorly extending macroadenoma can probably be managed as those with microadenomas (i.e. with dopamine agonists) before conception, followed by medication withdrawal upon confirmation of pregnancy. The risk that such a tumour will enlarge sufficiently to cause clinically serious complications is probably only marginally higher than the risk in patients with microadenomas.

A woman with a larger macroadenoma, especially one with suprasellar extension, has a 18.1% risk of clinically significant tumour enlargement during pregnancy. It is helpful to have a baseline MRI scan just prior to pregnancy. There is no best therapeutic approach in such a patient and there has to be a highly individualized shared decision-making with the patient with a clear, documented discussion of the various therapeutic alternatives. The most common approach is to stop the dopamine agonist after pregnancy is diagnosed, as in the patient with a microadenoma. Another approach is to debulk the tumour via transsphenoidal surgery. This reduces the risk of serious tumour enlargement, but cases of tumour expansion during pregnancy after such surgery have been reported [58]. After surgical debulking, a dopamine agonist will generally be required to normalize PRL levels and allow ovulation. A third approach, namely giving the dopamine agonist continuously throughout gestation, has been used but data regarding effects on the fetus are quite meagre; therefore, such treatment cannot be recommended without reservation. On the other hand, should pregnancy at an advanced stage be discovered in a woman taking bromocriptine or cabergoline, the data that exist are reassuring and would not justify therapeutic abortion. A special case might be a patient with a very large tumour in whom the growth rate of that tumour was slow and any effects of pressure on surrounding brain structures or the optic chiasm were very gradual and of no consequence. If there was substantial tumour shrinkage with the dopamine agonist, then stopping the drug abruptly might cause a sudden enlargement of the tumour with potential pressure on surrounding structures. In such a case, the wisest course could be to continue the dopamine agonist.

For patients with macroadenomas treated with a dopamine agonist alone or after surgery, careful follow-up with 1–3 monthly formal visual field testing is warranted. Repeat MRI scanning is reserved for patients with symptoms of tumour enlargement and/or visual field defect or both, as outlined earlier for patients with microadenomas. Repeat MRI scanning after delivery to detect asymptomatic tumour enlargement may be useful as well.

Should symptomatic tumour enlargement occur with any of these approaches, reinstitution of the dopamine agonist is probably less harmful to the mother and child than surgery. Any type of surgery during pregnancy results in a 1.5–fold and a 5-fold increase in fetal loss in the first and second trimester respectively, but without risk of congenital malformations [58, 59]. In addition, pregnant (compared to nonpregnant) women have significantly increased risks of postoperative complications, including infections and even mortality [60]. Thus, dopamine agonist reinstitution would appear to be preferable to surgery. However, such medical therapy must be very closely monitored, and transsphenoidal surgery or delivery (if the pregnancy is far enough advanced) should be performed if there is no response to the dopamine agonist and vision is progressively worsening.

Although suckling stimulates PRL secretion in normal women for the first few weeks to months postpartum, there are no data to suggest that breast-feeding can cause tumour growth. Thus, there seems to be no reason to discourage nursing in women with prolactinomas.

Cushing's Disease and Pregnancy

Changes of Hypothalamic–Pituitary–Adrenal (HPA) Axis During Pregnancy

Normal pregnancy is associated with profound changes of the hypothalamic–pituitary–adrenal axis [61], as listed next.

1. Serum free cortisol levels increase by 2–4-fold during pregnancy. Urine free cortisol (UFC) is normal during first trimester but increases by ~ 3-fold by the third trimester [61].
2. Total cortisol levels rise by 2–3-fold towards the end of pregnancy, mainly due to an increase in cortisol binding globulin and free cortisol.
3. The circadian rhythm of cortisol is preserved but blunted.
4. The cortisol response to ACTH stimulation is enhanced.
5. Both CRH and ACTH levels increase during pregnancy, partially owing to placental production of both hormones.
6. CRH and ACTH secretion is desensitized to the negative feedback of cortisol, resulting in failure of suppression by dexamethasone.

Cushing's Disease in Pregnancy

Cushing's syndrome is very rare in pregnancy. Its true prevalence is unknown. Interestingly, unlike in normal population wherein Cushing's disease comprises over 70% of the cases of Cushing's syndrome, Cushing's disease only accounts for 30–40% of Cushing's syndrome during pregnancy [61]. Adrenal causes of Cushing's syndrome, mostly adenoma and hyperplasia but even carcinoma, make up about 50% of the cases first found in pregnancy [62].

Effect of Cushing's Disease on Pregnancy

Untreated Cushing's syndrome exerts a significant negative impact on pregnancy outcomes. Maternal morbidities caused by hypercortisolism occur in 70% of the cases and consist of hypertension, pre-eclampsia, eclampsia, hyperglycaemia or diabetes, and hyperlipidaemia [62]. Other complications include osteoporosis, fracture, impaired wound healing, and psychiatric disorders [61]. Fortunately, maternal mortality is rare. Fetal complications from Cushing's syndrome during gestation include early spontaneous abortion, prematurity, and intrauterine growth restriction [63], with fetal mortality occurring in up to 20% of cases [61].

Diagnosis of Cushing's Disease During Pregnancy

The complex changes of the HPA axis during pregnancy make the diagnosis of Cushing's disease a major challenge. Many of the abnormalities of the HPA axis found in non-gravid patients with Cushing's syndrome can be seen in normal pregnancy, including increased cortisol levels, reduced inhibition of the HPA axis by exogenous glucocorticoid and a blunted circadian rhythm of cortisol. In addition, the symptoms and signs of Cushing's syndrome also overlap with those seen in normal pregnancy (e.g. weight gain, mood changes, fatigue, and hirsutism) [64].

Even though the 24-hour UFC level increases during pregnancy, a level higher than 3-fold of the upper limit of normal has been proposed to suggest Cushing's syndrome during pregnancy [61, 64]. Late night salivary cortisol levels also undergo a gradual increase during normal pregnancy. However, using the cut-off value of 0.255 μg/dl (7.0 nmol/L), 0.260 μg/dL (7.2 nmol/L), and 0.285 μg/dl (7.9 nmol/L) for first, second, and third trimesters respectively yield high sensitivity (>80%) and specificity (>90%) in separating Cushing's disease from normal gestational changes [65]. The 1 mg dexamethasone overnight suppression test is not reliable for the diagnosis of Cushing's syndrome, since more than 60% of normal women may have abnormal cortisol levels after 1mg dexamethasone during pregnancy [61].

Investigations of the differential diagnosis of ACTH-independent and ACTH-dependent Cushing's syndrome can usually start with measurement of the plasma ACTH concentration, performing an adrenal ultrasound scan (due to the high prevalence of adrenal Cushing's syndrome in pregnancy) and an 8 mg dexamethasone suppression test. It should be noted that up to 50% of women with adrenal Cushing's syndrome may have non-suppressed ACTH levels during pregnancy due to placental CRH stimulation and ACTH production [61]. A borderline low or suppressed ACTH, no response to the suppression by 8mg dexamethasone and an adrenal mass on the ultrasound all suggest ACTH-independent Cushing's syndrome [61]. When adrenal ultrasound identifies an adrenal lesion, a non-contrast MRI can also be used to further characterize the lesion. In patients suspected to have Cushing's disease, in addition to the suppressed cortisol level by 8mg dexamethasone, an increase in ACTH after CRH injection usually suggests Cushing's disease. Cushing's syndrome due to ectopic ACTH or an adrenal aetiology usually fail to response to exogenous dexamethasone and CRH. Pituitary MRI without contrast is necessary to evaluate a pituitary lesion during pregnancy but many patients with Cushing's disease may have very small microadenomas [61, 64]. Due to concern for radiation exposure and technical challenges in performing the test in most centres, inferior petrosal sinus sampling (IPSS) should be reserved for the patients who are difficult to diagnose even after the non-invasive tests [61].

Management of Cushing's Disease During Pregnancy

No systematic evaluation of treatments for Cushing's disease is available for pregnant patients. Reported management options include watchful monitoring, surgical, and medical treatments. In a review of 136 pregnancies collected from the literature, Lindsay *et al* found that the frequency of live births increased from 76% to 89% when active treatment was instituted by gestational week 20 [66]. Therefore, treatment during pregnancy has been advocated [66–68]. For patients with mild disease, it may be reasonable to wait until delivery to treat Cushing's disease [69]. However, comorbidities including hyperglycaemia and hypertension from hypercortisolism should be well controlled during pregnancy.

For those with Cushing's disease requiring treatment during pregnancy, transsphenoidal surgery in the second trimester appears to be the treatment of choice [64]. Bilateral adrenalectomy can be considered in those in whom transsphenoidal surgery and medical treatment are not feasible and the patient needs rapid resolution of the hypercortisolism [64]. It should be emphasized that after either pituitary or adrenal surgery, patients may develop adrenal insufficiency and replacement with glucocorticoid and mineralocorticoid (only after adrenal surgery) is required throughout pregnancy [70].

If patient is not eligible for or refuses surgery or has failure of surgery, then medical treatments can be considered for those with severe Cushing's disease in order to achieve improvement of symptoms, control of tumour, or complications and to improve fetal outcomes. Cabergoline, has been used in uncontrolled

Cushing's disease during pregnancy and appears to confer good control of the disease [71–73]. Although ketoconazole can inhibit steroidogenesis, the existing limited data from a few cases suggest that it is safe and not associated with congenital malformation, including sexual development [71, 74, 75]. Metyrapone crosses placenta and has the potential to affect fetal steroidogenesis [64]. Its use has been reported during the second and third trimesters in a patient with Cushing's syndrome and was not associated with fetal complications [74]. However, due to the accumulation of deoxycorticosterone caused by metyrapone, maternal hypertension and pre-eclampsia may be exacerbated [76]. Pasireotide is a newly approved somatostatin analogue by FDA to treat Cushing's disease [50, 77] but its use in during pregnancy has not been reported. Mifepristone has been used in treating refractory Cushing syndrome [78]. However, due to its antagonistic action on the progesterone receptor, it is an abortifacient.

Acromegaly and Pregnancy

Changes in GH, GH-Variant, and IGF-1 During Normal Pregnancy and Pregnancy Complicated by Acromegaly

Placenta-derived GH-variant starts to appear in maternal circulation around gestational weeks 15–17 and is released in a continuous and non-pulsatile manner. It increases throughout pregnancy and becomes the predominant circulating form of GH in the second half of pregnancy. The placental GH-variant is biologically active, stimulating IGF-1 production [79], and cross-reacts with pituitary GH in most assays. In contrast, pituitary GH declines to very low levels in the latter part of pregnancy presumably due to the negative feedback action of increasing levels of IGF-1 [76, 80].

Compared to the decline of pituitary GH in normal pregnancy, autonomous GH secretion from a pituitary adenoma in acromegaly is not suppressed by the rising IGF-1. Placental GH-variant secretion though is similar in both acromegalic and normal pregnancies. Therefore, pituitary GH and placental GH-variant are both present at high concentrations in maternal circulation in late gestation [76]. Interestingly, IGF-1 levels in acromegalic patients may decrease during early pregnancy, even in those without medical treatment [81]. This initial decrease in IGF-1 level is postulated to be caused by stimulation of GH-binding protein production and inhibition of GH signalling pathways by oestrogen [80].

Interaction Between Pregnancy and Acromegaly

Complications associated with acromegaly, including hypertension, hyperglycaemia, and cardiac disease, potentially could worsen during pregnancy [76, 81, 82]. These complications should be closely monitored during pregnancy but, in practice, such exacerbations are rarely seen [81, 82]. Tumour enlargement in patients with acromegaly is uncommon, with only 5 cases having been reported [20, 36, 83–85]. Most studies have shown both biochemical and clinical stability or even improvement in acromegaly during pregnancy [86, 87].

Diagnosis of Acromegaly During Pregnancy

Pregnancy-related changes in GH, GH-variants, and IGF-1 have made the diagnosis of acromegaly very challenging. First, there is steady increase in IGF-1 levels during the second half of normal pregnancies. Second, in patients with acromegaly, both pituitary GH and placental GH-variant are present at high concentrations, especially in late pregnancy. However, the commonly used GH assays do not distinguish between pituitary GH and the placental GH-variant. While GH-specific assays are available, these assays are not readily available. Lastly, both placental and pituitary GH forms have been shown to not be suppressed by oral glucose [80], so an oral glucose tolerance test should not be used to diagnose acromegaly in pregnancy. Therefore, some authors recommend waiting to diagnose acromegaly until after delivery, when the placental GH levels fall precipitously. Pituitary MRI is usually reserved for those with symptoms suggesting acute intracranial processes (e.g. concern for pituitary apoplexy or those with known acromegaly reporting severe headache and newly developed vision defects).

Management of Acromegaly During Pregnancy

Several management options exist for women with acromegaly during pregnancy: expectant approach, transsphenoidal surgery, and medical treatment including somatostatin analogues (SSAs), dopamine agonists and pegvisomant.

An expectant approach is appropriate for patients with mild disease, since spontaneous improvement of acromegaly has been reported during pregnancy [86, 87]. Close monitoring of potential complications from acromegaly, including hypertension, hyperglycaemia, pre-eclampsia and eclampsia, is warranted. In a retrospective, multi-centre study, most of the patients who were treated with dopamine agonists and/or SSA before conception and had medication withdrawal upon confirmation of pregnancy had uneventful pregnancies [81]. Close monitoring of visual fields is necessary for patients with known macroadenomas. In patients with suspected tumour enlargement causing a visual defect or severe headache, an MRI may be necessary to confirm whether an intervention is necessary.

In most patients, medical treatment should be stopped either before planned conception or immediately after confirming pregnancy [70, 88, 89], which is consistent with the Endocrine Society Guideline [90]. Fortunately, most acromegaly patients who had such drug withdrawal did well during pregnancy [81]. Therefore, medical treatment during pregnancy should be reserved for those with active acromegaly requiring treatment due to severe symptoms from large tumours (e.g. severe headache, vision defects, or tumour enlargement) [88]. It is important to note that SSAs, dopamine agonists and pegvisomant are not FDA approved to be used in pregnancy.

SSAs are effective in controlling both tumour growth and IGF-1 levels in non-pregnant acromegalic patients [90]. Since SSAs cross the placenta, potentially affecting fetal brain somatostatin receptors, and decrease uterine artery blood flow, their use during pregnancy is not recommended [91, 92]. However, retrospective studies and small case series have shown that SSAs are generally effective and safe during pregnancy [92, 93], though cases of small for gestational age [81] and shorter birth length [92] have been reported.

Dopamine agonists have been used in the treatment of pregnant women with acromegaly. In the reported case series, the use of dopamine agonists appears to be safe [80, 81, 91]. One retrospective study showed small for gestational age in women taking both SSAs and dopamine agonists. More safety data about the use of DA agonists during pregnancy are from large series involving the treatment of prolactinoma [1, 2] (see previous section).

Pegvisomant acts as a GH receptor antagonist [82, 94, 95] and has shown both efficacy and safety in controlling acromegalic symptoms and IGF-1 levels in non-pregnant patients. Use of pegvisomant during pregnancy was recently assessed in the largest reported case series of 35 pregnancies (27 maternal and 8 paternal exposure), including 3 cases in which pegvisomant was continued throughout pregnancy [95]. This small observational study demonstrated no apparent adverse effect of pegvisomant on pregnancy outcome. However, children with GH-insensitivity syndrome (Laron syndrome) due to GH receptor abnormalities have multiple abnormalities so pegvisomant use cannot be recommended during pregnancy [96].

Considering the prolonged nature of the course of most patients with acromegaly, interruption of medical therapy for 9–12 months should not have a particularly adverse effect on the long-term outcome. On the other hand, these drugs can control tumour growth and for enlarging tumours, their reintroduction during pregnancy may be warranted as a preferable management strategy to surgery. In patients with severe acromegalic symptoms not controlled by, or intolerant to, medical treatments, or with acute symptoms due to tumour enlargement (e.g. vision loss), or concern for pituitary apoplexy, emergent transsphenoidal surgery can be considered, preferably during the second trimester.

Functioning Gonadotropin-Secreting Adenoma in Pregnancy

Most gonadotroph tumours are clinically silent and present as non-functioning adenomas. Functioning gonadotropin-secreting pituitary adenomas are actually quite rare [97]. Due to oversecretion of bioactive gonadotropins, pre-menopausal women with functioning gonadotropin-secreting adenomas commonly have excessive production of oestrogen, leading to secondary amenorrhea, oligomenorrhea, galactorrhoea, infertility, and rarely ovarian hyperstimulation syndrome. In most of these patients, oestrogen levels are elevated, while FSH levels are either normal or mildly raised, and LH levels usually suppressed. In many cases of gonadotropin-secreting tumours, prolactin levels are elevated, likely due to stalk effect or stimulation by high oestrogen [97, 98].

Only two cases of functioning gonadotropin-secreting adenomas recognized during a possible pregnancy have been reported. In one case, the patient conceived naturally but developed ovarian hyperstimulation syndrome at gestation week 5 and had termination of pregnancy at the nineth week due to development of deep vein thrombosis [99, 100]. In another case, the patient presented with positive pregnancy tests and markedly enlarged ovaries. The pregnancy was terminated due to abnormal hCG levels and lack of an intrauterine pregnancy on ultrasound [101].

TSH-Secreting Adenoma in Pregnancy

TSH-secreting pituitary adenomas are rare and account for 0.5–3% of all pituitary adenomas [102, 103]. Clinically, they are characterized by hyperthyroidism with elevated levels of free T_4 and T_3, inappropriately normal or elevated level of TSH, and predominantly pituitary macroadenomas [102–104]. The majority of these tumours only secrete TSH, though 20–25% may cosecrete growth hormone or prolactin [105].

Menstrual irregularity, oligomenorrhea and amenorrhea often can be seen in women with TSH-secreting tumours, owing to the large size of most of the tumours and hyperprolactinemia as a result of stalk effect or cosecretion from the tumour [105]. However, spontaneous pregnancy has been reported in women with TSH-secreting tumour [106].

To the authors' knowledge, only 6 cases of TSH-secreting pituitary adenomas have been reported during pregnancy [106–111]. Pituitary macroadenomas were found in all but one case. In two cases, the macroadenomas were noted to have significant enlargement during pregnancy, which necessitated treatment [107, 108]. In these reported cases, treatment of TSH-secreting adenomas during pregnancy were targeted to the tumour (surgery, radiation, or octreotide) and/or the hyperthyroidism *per se* with thionamides.

Clinically Non-Functioning Pituitary Adenoma (CNFA) in Pregnancy

A CNFA, by definition, does not cause a clinical syndrome due to overproduction of a pituitary hormone. In the non-pregnant population, CNFAs represent about 30%-40% of all pituitary tumours [112, 113] and most stain positively for gonadotropins [113]. In a recent UK Surveillance study over a 3-year period of 71 pregnant women with pituitary tumours, CNFAs made up 23% of cases, whereas prolactinomas made up 69%, and acromegaly and Cushing's disease each comprised 4% [109].

Most CNFAs remain stable in size during pregnancy [76, 109, 112], especially those that were treated with surgery or radiotherapy before conception [109]. In another recent series of 8 cases, none had tumour enlargement during surgery [20]. However, in the UK Surveillance study, tumour enlargement was reported in one of seven non-functioning adenomas diagnosed before conception and in three of five non-functioning adenomas diagnosed in pregnancy [109]. Additionally, significant lactotroph hyperplasia during pregnancy may push the CNFAs up to impact the optic chiasm [114]. Pregnancy outcomes do not seem to be affected by CNFAs [109].

CNFAs usually do not need treatment during pregnancy. However, attention should be paid to symptoms suggestive of tumour enlargement, including severe headaches or vision defects. Pituitary apoplexy has been reported in CNFAs, especially macroadenomas [109].

REFERENCES

1. Molitch ME. Prolactinoma in pregnancy. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 885–96.
2. Molitch ME. Endocrinology in pregnancy: management of the pregnant patient with a prolactinoma. *Eur J Endocrinol*, 2015; **172**: R205–13.
3. Casanueva FF, Molitch ME, Schlechte JA, *et al*. Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. *Clin Endocrinol (Oxf)*, 2006; **65**: 265–73.
4. Gillam MP, Molitch ME, Lombardi G, Colao A. Advances in the treatment of prolactinomas. *Endocr Rev*, 2006; **27**: 485–534.

5. Melmed S, Casanueva FF, Hoffman AR, *et al*. Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
6. Bigazzi M, Ronga R, Lancranjan I, *et al*. A pregnancy in an acromegalic woman during bromocriptine treatment: effects on growth hormone and prolactin in the maternal, fetal, and amniotic compartments. *J Clin Endocrinol Metab*, 1979; **48**: 9–12.
7. Krupp P, Monka C. Bromocriptine in pregnancy: safety aspects. *Klin Wochenschr*, 1987; **65**: 823–7.
8. Canales ES, García IC, Ruíz JE, Zárate A. Bromocriptine as prophylactic therapy in prolactinoma during pregnancy. *Fertil Steril*, 1981; **36**: 524–6.
9. Konopka P, Raymond JP, Merceron RE, Seneze J. Continuous administration of bromocriptine in the prevention of neurological complications in pregnant women with prolactinomas. *Am J Obstet Gynecol*, 1983; **146**: 935–8.
10. Ruiz-Velasco V, Tolis G. Pregnancy in hyperprolactinemic women. *Fertil Steril*, 1984; **41**: 793–805.
11. Abu-Fadil S, DeVane G, Siler TM, Yen SS. Effects of oral contraceptive steroids on pituitary prolactin secretion. *Contraception*, 1976; **13**: 79–85.
12. Gold EB. Epidemiology of pituitary adenomas. *Epidemiol Rev*, 1981; **3**: 163–83.
13. Maheux R, Jenicek M, Cleroux R, *et al*. Oral contraceptives and prolactinomas: a case-control study. *Am J Obstet Gynecol*, 1982; **143**: 134–8.
14. Marin F, Kovacs KT, Scheithauer BW, Young WF Jr. The pituitary gland in patients with breast carcinoma: a histologic and immunocytochemical study of 125 cases. *Mayo Clin Proc*, 1992; **67**: 949–56.
15. Molitch ME. Clinical features and epidemiology of prolactinomas in women. In: Olefsky JM, Robbins RJ, eds. *Prolactinomas: Practical Diagnosis and Management*. New York: Churchill Livingstone, 1986: 67–95.
16. Ono M, Miki N, Amano K, *et al*. Individualized high-dose cabergoline therapy for hyperprolactinemic infertility in women with micro- and macroprolactinomas. *J Clin Endocrinol Meta*b, 2010; **95**: 2672–9.
17. Sherman BM, Harris CE, Schlechte J, *et al*. Pathogenesis of prolactin-secreting pituitary adenomas. *Lancet*, 1978; **2**: 1019–21.
18. Teperman L, Futterweit W, Zappulla R, Malis LI. Oral contraceptive history as a risk indicator in patients with pituitary tumors with hyperprolactinemia: a case comparison study of twenty patients. *Neurosurgery*, 1980; **7**: 571–3.
19. Stalldecker G, Mallea-Gil MS, Guitelman M, *et al*. Effects of cabergoline on pregnancy and embryo-fetal development: retrospective study on 103 pregnancies and a review of the literature. *Pituitary*, 2010; **13**: 345–50.
20. Karaca Z, Yarman S, Ozbas I, *et al*. How does pregnancy affect the patients with pituitary adenomas: a study on 113 pregnancies from Turkey. *J Endocrinol Invest*, 2018; **41**: 129–41.
21. Lebbe M, Hubinont C, Bernard P, Maiter D. Outcome of 100 pregnancies initiated under treatment with cabergoline in hyperprolactinaemic women. *Clin Endocrinol (Oxf)*, 2010; **73**: 236–42.
22. Ono M, Miki N, Amano K, *et al*. Individualized high-dose cabergoline therapy for hyperprolactinemic infertility in women with micro- and macroprolactinomas. *J Clin Endocrinol Metab*, 2010; **95**: 2672–9.
23. Glezer A, Bronstein MD. Prolactinomas, cabergoline, and pregnancy. *Endocrine*, 2014; **47**: 64–9.
24. Canfield MA, Honein MA, Yuskiv N, *et al*. National estimates and race/ethnic-specific variation of selected birth defects in the United States, 1999–2001. *Birth Defects Res A Clin Mol Teratol*, 2006; **76**: 747–56.
25. Martin JA, Hamilton BE, Sutton PD, *et al*. Births: final data for 2007. *Natl Vital Stat Rep*, 2010; **58**: 1–85.
26. Webster J. A comparative review of the tolerability profiles of dopamine agonists in the treatment of hyperprolactinaemia and inhibition of lactation. *Drug Saf*, 1996; **14**: 228–38.
27. Hurault-Delarue C, Montastruc JL, Beau AB, Lacroix I, Damase-Michel C. Pregnancy outcome in women exposed to dopamine agonists during pregnancy: a pharmacoepidemiology study in EFEMERIS database. *Arch Gynecol Obstet*, 2014; **290**: 263–70.
28. Rigg LA, Lein A, Yen SS. Pattern of increase in circulating prolactin levels during human gestation. *Am J Obstet Gynecol*, 1977; **129**: 454–6.
29. Tyson JE, Friesen HG. Factors influencing the secretion of human prolactin and growth hormone in menstrual and gestational women. *Am J Obstet Gynecol*, 1973; **116**: 377–87.
30. Dinc H, Esen F, Demirci A, Sari A, Resit Gümele H. Pituitary dimensions and volume measurements in pregnancy and post partum. MR assessment. *Acta Radiol*, 1998; **39**: 64–9.
31. Rastogi A, Bhadada SK, Bhansali A. Pregnancy and tumor outcomes in infertile women with macroprolactinoma on cabergoline therapy. *Gynecol Endocrinol*, 2017; **33**: 270–3.
32. Gemzell C, Wang CF. Outcome of pregnancy in women with pituitary adenoma. *Fertil Steril*, 1979; **31**: 363–72.
33. Molitch ME. Pregnancy and the hyperprolactinemic woman. *N Engl J Med*, 1985; **312**: 1364–70.
34. Holmgren U, Bergstrand G, Hagenfeldt K, Werner S. Women with prolactinoma—effect of pregnancy and lactation on serum prolactin and on tumour growth. *Acta Endocrinol (Copenh)*, 1986; **111**: 452–9.
35. Ampudia X, Puig-Domingo M, Schwarzstein D, *et al*. Outcome and long-term effects of pregnancy in women with hyperprolactinaemia. *Eur J Obstet Gynecol Reprod Biol*, 1992; **46**: 101–7.
36. Kupersmith MJ, Rosenberg C, Kleinberg D. Visual loss in pregnant women with pituitary adenomas. *Ann Intern Med*, 1994; **121**: 473–7.
37. Rossi AM, Vilska S, Heinonen PK. Outcome of pregnancies in women with treated or untreated hyperprolactinemia. *Eur J Obstet Gynecol Reprod Biol*, 1995; **63**: 143–6.
38. Badawy SZ, Marziale JC, Rosenbaum AE, Chang JK, Joy SE. The long-term effects of pregnancy and bromocriptine treatment on prolactinomas—the value of radiologic studies. *Early Pregnancy*, 1997; **3**: 306–11.
39. Mallmann ES, Nacul A, Spritzer PM. Pregnancy in hyperprolactinemic women. *Acta Obstet Gynecol Scand*, 2002; **81**: 265–7.
40. Bronstein MD, Salgado LR, de Castro Musolino NR. Medical management of pituitary adenomas: the special case of management of the pregnant woman. *Pituitary*, 2002; **5**: 99–107.
41. Auriemma RS, Perone Y, Di Sarno A, *et al*. Results of a single-center observational 10-year survey study on recurrence of hyperprolactinemia after pregnancy and lactation. *J Clin Endocrinol Metab*, 2013; **98**: 372–9.
42. Domingue ME, Devuyst F, Alexopoulou O, Corvilain B, Maiter D. Outcome of prolactinoma after pregnancy and lactation: a study on 73 patients. *Clin Endocrinol (Oxf)*, 2014; **80**: 642–8.
43. Piantanida E, Gallo D, Lombardi V, *et al*. Pituitary apoplexy during pregnancy: a rare, but dangerous headache. *J Endocrinol Invest*, 2014; **37**: 789–97.

44. Grand'Maison S, Weber F, Bédard M-J, *et al*. Pituitary apoplexy in pregnancy: a case series and literature review. *Obstet Med*, 2015; **8**: 177–83.
45. Crosignani PG, Mattei AM, Scarduelli C, Cavioni V, Boracchi P. Is pregnancy the best treatment for hyperprolactinaemia? *Hum Reprod*, 1989; **4**: 910–12.
46. Samaan NA, Schultz PN, Leavens MD, *et al*. Pregnancy after treatment in patients with prolactinoma: operation versus bromocriptine. *Am J Obstet Gynecol*, 1986; **155**: 1300–5.
47. Ikegami H, Aono T, Koizumi K, Koike K, Fukui H, Tanizawa O. Relationship between the methods of treatment for prolactinomas and the puerperal lactation. *Fertil Steril*, 1987; **47**: 867–9.
48. Salvatori R. Surgical treatment of microprolactinomas: pros. *Endocrine*, 2014; **47**: 725–9.
49. Laws ER, Jr., Fode NC, Randall RV, Abboud CF, Coulam CB. Pregnancy following transsphenoidal resection of prolactin-secreting pituitary tumors. *J Neurosurg*, 1983; **58**: 685–8.
50. Molitch ME. Diagnosis and treatment of pituitary adenomas: a review. *JAMA*, 2017; **317**: 516–24.
51. Divers WA, Jr., Yen SS. Prolactin-producing microadenomas in pregnancy. *Obstet Gynecol*, 1983; **62**: 425–9.
52. Narita O, Kimura T, Suganuma N, *et al*. Relationship between maternal prolactin levels during pregnancy and lactation in women with pituitary adenoma. *Nippon Sanka Fujinka Gakkai Zasshi*, 1985; **37**: 758–62.
53. Kok RD, de Vries MM, Heerschap A, van den Berg PP. Absence of harmful effects of magnetic resonance exposure at 1.5 T in utero during the third trimester of pregnancy: a follow-up study. *Magn Reson Imaging*, 2004; **22**: 851–4.
54. De Wilde JP, Rivers AW, Price DL. A review of the current use of magnetic resonance imaging in pregnancy and safety implications for the fetus. *Prog Biophys Mol Biol*, 2005; **87**: 335–53.
55. Webb JA, Thomsen HS, Morcos SK, *et al*. The use of iodinated and gadolinium contrast media during pregnancy and lactation. *Eur Radiol*, 2005; **15**: 1234–40.
56. Garcia-Bournissen F, Shrim A, Koren G. Safety of gadolinium during pregnancy. *Can Fam Physician*, 2006; **52**: 309–10.
57. Ray JG, Vermeulen MJ, Bharatha A, Montanera WJ, Park AL. Association between MRI exposure during pregnancy and fetal and childhood outcomes. *JAMA*, 2016; **316**: 952–61.
58. Belchetz PE, Carty A, Clearkin LG, Davis JC, Jeffreys RV, Rae PG. Failure of prophylactic surgery to avert massive pituitary expansion in pregnancy. *Clin Endocrinol (Oxf)*, 1986; **25**: 325–30.
59. Brodsky JB, Cohen EN, Brown BW Jr, Wu ML, Whitcher C. Surgery during pregnancy and fetal outcome. *Am J Obstet Gynecol*, 1980; **138**: 1165–7.
60. Cohen-Kerem R, Railton C, Oren D, Lishner M, Koren G. Pregnancy outcome following non-obstetric surgical intervention. *Am J Surg*, 2005; **190**: 467–73.
61. Lindsay JR, Nieman LK. The hypothalamic-pituitary-adrenal axis in pregnancy: challenges in disease detection and treatment. *Endocr Rev*, 2005; **26**: 775–99.
62. Polli N, Pecori Giraldi F, Cavagnini F. Cushing's disease and pregnancy. *Pituitary*, 2004; **7**: 237–41.
63. Yawar A, Zuberi LM, Haque N. Cushing's disease and pregnancy: case report and literature review. *Endocr Pract*, 2007; **13**: 296–9.
64. Brue T, Amodru V, Castinetti F. Management of endocrine disease: management of Cushing's syndrome during pregnancy: solved and unsolved questions. *Eur J Endocrinol*, 2018; **178**: R259–66.
65. Lopes LM, Francisco RP, Galletta MA, Bronstein MD. Determination of nighttime salivary cortisol during pregnancy: comparison with values in non-pregnancy and Cushing's disease. *Pituitary*, 2016; **19**: 30–8.
66. Lindsay JR, Jonklaas J, Oldfield EH, Nieman LK. Cushing's syndrome during pregnancy: personal experience and review of the literature. *J Clin Endocrinol Metab*, 2005; **90**: 3077–83.
67. Bevan JS, Gough MH, Gillmer MD, Burke CW. Cushing's syndrome in pregnancy: the timing of definitive treatment. *Clin Endocrinol (Oxf)*, 1987; **27**: 225–33.
68. Vilar L, Freitas Mda C, Lima LH, Lyra R, Kater CE. Cushing's syndrome in pregnancy: an overview. *Arq Bras Endocrinol Metabol*, 2007; **51**: 1293–302.
69. Gopal RA, Acharya SV, Bandgar TR, *et al*. Cushing disease with pregnancy. *Gynecol Endocrinol*, 2012; **28**: 533–5.
70. Bornstein SR, Allolio B, Arlt W, *et al*. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 364–89.
71. Berwaerts J, Verhelst J, Mahler C, Abs R. Cushing's syndrome in pregnancy treated by ketoconazole: case report and review of the literature. *Gynecol Endocrinol*, 1999; **13**: 175–82.
72. Nakhleh A, Saiegh L, Reut M, *et al*. Cabergoline treatment for recurrent Cushing's disease during pregnancy. *Hormones (Athens)*, 2016; **15**: 453–8.
73. Woo I, Ehsanipoor RM. Cabergoline therapy for Cushing disease throughout pregnancy. *Obstet Gynecol*, 2013; **122**(2 Pt 2): 485–7.
74. Boronat M, Marrero D, López-Plasencia Y, Barber M, Schamann Y, Nóvoa FJ. Successful outcome of pregnancy in a patient with Cushing's disease under treatment with ketoconazole during the first trimester of gestation. *Gynecol Endocrinol*, 2011; **27**: 675–7.
75. Costenaro F, Rodrigues TC, de Lima PB, Ruszczyk J, Rollin G, Czepielewski MA. A successful case of Cushing's disease pregnancy treated with ketoconazole. *Gynecol Endocrinol*, 2015; **31**: 176–8.
76. Bronstein MD, Paraiba DB, Jallad RS. Management of pituitary tumors in pregnancy. *Nat Rev Endocrinol*, 2011; **7**: 301–10.
77. Fleseriu M. Medical treatment of Cushing disease: new targets, new hope. *Endocrinol Metab Clin North Am*, 2015; **44**: 51–70.
78. Biller BM, Grossman AB, Stewart PM, *et al*. Treatment of adrenocorticotropin-dependent Cushing's syndrome: a consensus statement. *J Clin Endocrinol Metab*, 2008; **93**: 2454–62.
79. Verhaeghe J. Does the physiological acromegaly of pregnancy benefit the fetus? *Gynecol Obstet Invest*, 2008; **66**: 217–26.
80. Cheng V, Faiman C, Kennedy L, *et al*. Pregnancy and acromegaly: a review. *Pituitary*, 2012; **15**: 59–63.
81. Caron P, Broussaud S, Bertherat J, *et al*. Acromegaly and pregnancy: a retrospective multicenter study of 59 pregnancies in 46 women. *J Clin Endocrinol Metab*, 2010; **95**: 4680–7.
82. Cheng S, Grasso L, Martinez-Orozco JA, *et al*. Pregnancy in acromegaly: experience from two referral centers and systematic review of the literature. *Clin Endocrinol (Oxf)*, 2012; **76**: 264–71.
83. Cozzi R, Attanasio R, Barausse M. Pregnancy in acromegaly: a one-center experience. *Eur J Endocrinol*, 2006; **155**: 279–84.
84. Kasuki L, Neto LV, TAkiya CM, *et al*. Growth of an aggressive tumor during pregnancy in an acromegalic patient. *Endocr J*, 2012; **59**: 313–19.

85. Okada Y, Morimoto I, Ejima K, *et al*. A case of active acromegalic woman with a marked increase in serum insulin-like growth factor-1 levels after delivery. *Endocr J*, 1997; **44**: 117–20.
86. Lau SL, McGrath S, Evain-Brion D, *et al*. Clinical and biochemical improvement in acromegaly during pregnancy. *J Endocrinol Invest*, 2008; **31**: 255–61.
87. Dias M, Boguszewski C, Gadelha M, *et al*. Acromegaly and pregnancy: a prospective study. *Eur J Endocrinol*, 2014; **170**: 301–10.
88. Fleseriu M. Medical treatment of acromegaly in pregnancy, highlights on new reports. *Endocrine*, 2015; **49**: 577–9.
89. Abucham J, Bronstein MD, Dias ML. Management of endocrine disease: acromegaly and pregnancy: a contemporary review. *Eur J Endocrinol*, 2017; **177**: R1–12.
90. Katznelson L, Laws ER, Melmed S, *et al*. Acromegaly: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 3933–51.
91. Muhammad A, Neggers SJ, van der Lely AJ. Pregnancy and acromegaly. *Pituitary*, 2017; **20**: 179–84.
92. Maffei P, Tamagno G, Nardelli GB, *et al*. Effects of octreotide exposure during pregnancy in acromegaly. *Clin Endocrinol (Oxf)*, 2010; **72**: 668–77.
93. Mikhail N. Octreotide treatment of acromegaly during pregnancy. *Mayo Clin Proc*, 2002; **77**: 297–8.
94. Brian SR, Bidlingmaier M, Wajnrajch MP, Weinzimer SA, Inzucchi SE. Treatment of acromegaly with pegvisomant during pregnancy: maternal and fetal effects. *J Clin Endocrinol Metab*, 2007; **92**: 3374–7.
95. van der Lely AJ, Gomez R, Heissler JF, *et al*. Pregnancy in acromegaly patients treated with pegvisomant. *Endocrine*, 2015; **49**: 769–73.
96. Laron Z. Lessons from 50 years of study of Laron syndrome. *Endocr Pract*, 2015; **21**: 1395–402.
97. Ntali G, Capatina C, Grossman A, Karavitaki N. Clinical review: functioning gonadotroph adenomas. *J Clin Endocrinol Metab*, 2014; **99**: 4423–33.
98. Macchia E, Simoncini T, Raffaelli V, Lombardi M, Iannelli A, Martino E. A functioning FSH-secreting pituitary macroadenoma causing an ovarian hyperstimulation syndrome with multiple cysts resected and relapsed after leuprolide in a reproductive-aged woman. *Gynecol Endocrinol*, 2012; **28**: 56–9.
99. Baba T, Endo T, Kitajima Y, Kamiya H, Moriwaka O, Saito T. Spontaneous ovarian hyperstimulation syndrome and pituitary adenoma: incidental pregnancy triggers a catastrophic event. *Fertil Steril*, 2009; **92**: 390.e1–3.
100. Kanaya M, Baba T, Kitajima Y, *et al*. Continsuous follicle-stimulating hormone exposure from pituitary adenoma causes periodic follicle recruitment and atresia, which mimics ovarian hyperstimulation syndrome. *Int J Womens Health*, 2012; **4**: 427–31.
101. Ghayuri M, Liu JH. Ovarian hyperstimulation syndrome caused by pituitary gonadotroph adenoma secreting follicle-stimulating hormone. *Obstet Gynecol*, 2007; **109**(2 Pt2): 547–9.
102. Amlashi FG, Tritos NA. Thyrotropin-secreting pituitary adenomas: epidemiology, diagnosis, and management. *Endocrine*, 2016; **52**: 427–40.
103. Tjornstrand A, Nystrom HF. Diagnosis of endocrine disease: diagnostic approach to TSH-producing pituitary adenoma. *Eur J Endocrinol*, 2017; **177**: R183–r197.
104. Beck-Peccoz P, Persani L, Mannavola D, Campi I. Pituitary tumours: TSH-secreting adenomas. *Best Pract Res Clin Endocrinol Metab*, 2009; **23**: 597–606.
105. Beck-Peccoz P, LAnia A, Beckers A, Chatterjee K, Wemeau J-L. 2013 European thyroid association guidelines for the diagnosis and treatment of thyrotropin-secreting pituitary tumors. *Eur Thyroid J*, 2013; **2**: 76–82.
106. Bolz M, Korber S, Schober HC. [TSH secreting adenoma of pituitary gland (TSHom)—rare cause of hyperthyroidism in pregnancy]. *Dtsch Med Wochenschr*, 2013; **138**: 362–6.
107. Caron P, Gerbeau C, Pradayrol L, Simonetta C, Bayard F. Successful pregnancy in an infertile woman with a thyrotropin-secreting macroadenoma treated with somatostatin analog (octreotide). *J Clin Endocrinol Metab*, 1996; **81**: 1164–8.
108. Chaiamnuay S, Moster M, Katz MR, Kim YN. Successful management of a pregnant woman with a TSH secreting pituitary adenoma with surgical and medical therapy. *Pituitary*, 2003; **6**: 109–13.
109. Lambert K, Rees K, Seed PT, *et al*. Macroprolactinomas and nonfunctioning pituitary adenomas and pregnancy outcomes. *Obstet Gynecol*, 2017; **129**: 185–94.
110. Okuyucu K, Alagoz E, Arslan N, *et al*. Thyrotropinoma with Graves' disease detected by the fusion of indium-111 octreotide scintigraphy and pituitary magnetic resonance imaging. *Indian J Nucl Med*, 2016; **31**: 141–3.
111. Perdomo CM, Árabe JA, Idoate MÁ, Galofré JC. Management of a pregnant woman with thyrotropinoma: a case report and review of the literature. *Gynecol Endocrinol*, 2017; **33**: 188–92.
112. Araujo PB, Vieira Neto L, Gadelha MR. Pituitary tumor management in pregnancy. *Endocrinol Metab Clin North Am*, 2015; **44**: 181–97.
113. Huang W, Molitch ME. Management of nonfunctioning pituitary adenomas (NFAs): observation. *Pituitary*, 2018; **21**: 162–7.
114. Molitch ME. Pituitary tumors and pregnancy. *Growth Horm IGF Res*, 2003; 13 Suppl A: S38–44.

9.7

Other Disorders of the Pituitary and Hypothalamus in Pregnancy

Paul V. Carroll, Niki Karavitaki, and Kirstie Lithgow

Introduction

Pregnancy results in significant alteration of pituitary anatomy and hormone physiology. Prolactin-producing lactotroph cells undergo progressive and significant hyperplasia throughout pregnancy contributing to increase in size of the anterior pituitary. Changes in pituitary hormone production, binding protein levels and target hormone action occur during pregnancy. As a result of these alterations, the assessment of pituitary status in the pregnant woman is complex and differs from the non-pregnant state. Disorders of hormone function may be well established prior to pregnancy or less frequently develop during pregnancy. These conditions may impact on maternal and fetal outcomes and influence pregnancy management. The physiology of pregnancy may also influence behaviour of endocrine conditions. Several disorders are specifically associated with pregnancy, lymphocytic hypophysitis, and Sheehan's syndrome being particular examples.

Pituitary Gland Anatomy and Imaging in Pregnancy

The anterior pituitary increases in size and volume during pregnancy. The upper contour typically becomes more convex in shape. Change in size develops in the first trimester and up to a doubling of size has been reported by the end of pregnancy [1]. Post-delivery normal size and configuration has been demonstrated at 6 months postpartum [2]. Magnetic resonance imaging (MRI) is considered safe during pregnancy but administration of IV enhancing-agents (e.g. gadolinium) should be avoided in most situations (see Chapter 9.10, 'Imaging of Endocrine Disorders in Pregnancy') [3].

Physiological Changes in the Anterior Pituitary Hormone Axes as a Result of Pregnancy

Prolactin levels increase early in pregnancy, and by term are 10-fold greater than in the non-pregnant state, in preparation for lactation [4]. LH and FSH are suppressed by the high levels of placental oestrogen and progesterone. Later in gestation oxytocin levels increase in preparation for labour and lactation [5].

Pregnancy results in increased hypothalamo–pituitary adrenal (HPA) axis activity characterized by increased production of cortisol, cortisol-binding-globulin (CBG), and urinary free cortisol secretion [6, 7]. Oestrogen produced by the placenta is responsible for the increase in CBG, resulting in a transient fall in free cortisol with a resultant rise in ACTH production. The placenta produces CRH and ACTH stimulating the maternal pituitary and adrenal glands. Free cortisol values rise throughout the trimesters reaching a plateau in the third trimester [8]. Similarly, the activity of the renin angiotensin aldosterone system is increased during the first trimester. Systemic vascular resistance falls and these changes result in volume expansion during pregnancy [9]. The circadian production of cortisol is maintained in healthy pregnancy [10–12]. Up to the 33rd week of gestation, 90–95% of fetal cortisol is derived from maternal adrenal secretion; thereafter, fetal adrenal cortisol production is more active. See **Figure 9.7.1**.

Placental growth hormone (GH) differs from pituitary GH by 13 amino acids and is manufactured and secreted by the synctiotrophoblastic epithelium of the placenta [13]. The regulation of placental GH secretion is not understood but this isoform increases throughout gestation [14, 15]. As a consequence, serum insulin-like growth factor 1 (IGF-1) levels increase in the second half of pregnancy [16], and via negative feedback, pituitary GH levels are reduced in later pregnancy and the early postpartum phase [14, 15].

There is a transient fall in TSH in the first trimester as a result of human chorionic gonadotropin (hCG) stimulation of the TSH

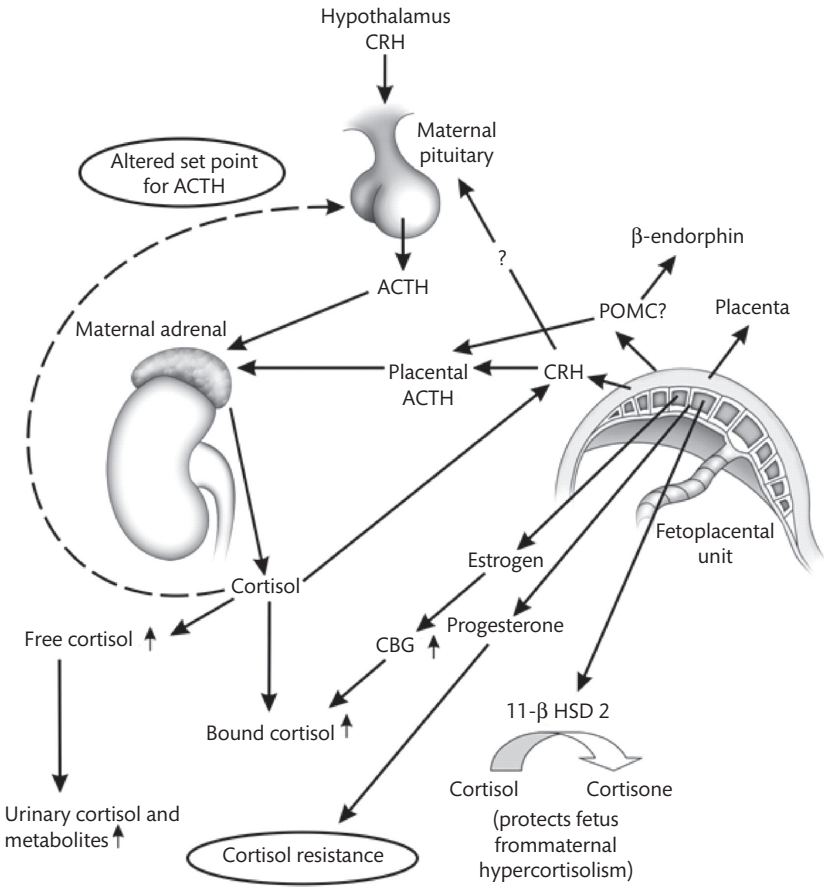


Figure 9.7.1 HPA axis during pregnancy. CRH stimulates ACTH secretion which results in adrenal cortisol production and secretion. Placental CRH stimulates fetoplacental ACTH which further increases cortisol production. Oestrogen increases CBG, with a rise in total cortisol. 11β-HSD 2 protects the fetus from maternal hypercortisolism.

Reproduced with permission from Karaca *et al*. Pregnancy and Pituitary Disorders. *EJE*, (2010) 162; 453–75.

receptor because of the structural homology between TSH and hCG molecules and their receptors [17]. The nadir level of TSH coincides with the peak in hCG [18]. There is a linear relationship between hCG and free T4 concentrations early in pregnancy [18]. For most women this stimulatory effect is transient and not clinically significant but in a significant number the TSH is suppressed at the end of the first trimester, especially in those with the highest hCG [19]. Multiple pregnancy results in higher hCG levels and more marked lowering of TSH [20]. The increase in oestrogens produced by the feto-placental unit stimulates hepatic production of thyroxine-binding globulin (TBG) and increases its sialylation thereby prolonging its half-life [21, 22]. This increase in TBG results in higher levels of total T_4 and T_3, starting at 4–6 weeks' gestation. Free thyroid hormone levels generally remain within the normal range throughout gestation, but pregnancy and trimester-specific references ranges have been developed [23]. A recent study described the methodology for validating local reference ranges for pregnancy [24].

Lymphocytic Hypophysitis

Lymphocytic hypophysitis (autoimmune hypophysitis) is a rare inflammatory/autoimmune disease that involves the pituitary gland and infundibulum resulting in enlargement of the gland and usually hypopituitarism [25]. The clinical features include headache, variable degrees of hormone insufficiency and in some cases diabetes insipidus. Mass effect with visual consequence may be also present. Both genders and all ages may be affected but the condition is classically identified during pregnancy or the early postpartum period [26]. The condition is more common in females. The diagnosis is confirmed histologically and the features include normal pituitary tissue with lymphocytic infiltration, plasma cells, histiocytes, and fibrosis [27]. Magnetic Resonance Imaging features indicative of lymphocytic hypophysitis include a symmetric enlargement of the pituitary gland, a thickened non-tapering pituitary stalk, and an intact sellar floor. There is prompt homogenous enhancement following gadolinium administration [28, 29]. Not all patients have histological diagnosis, as especially in the context of pregnancy a confident clinical diagnosis can be reached, facilitating clinical management. Treatment includes replacement of hormone deficiency (including ADH) and decision-making regarding conservative, medical, and surgical therapies. High-dose suppressive glucocorticoid remains the cornerstone of medical therapy but a variety of immunosuppressive treatments have been used. Surgery is indicated for non-responders, or if there is mass effect, headache, visual failure or when a tissue diagnosis is considered important. Radiotherapy may be useful when there is

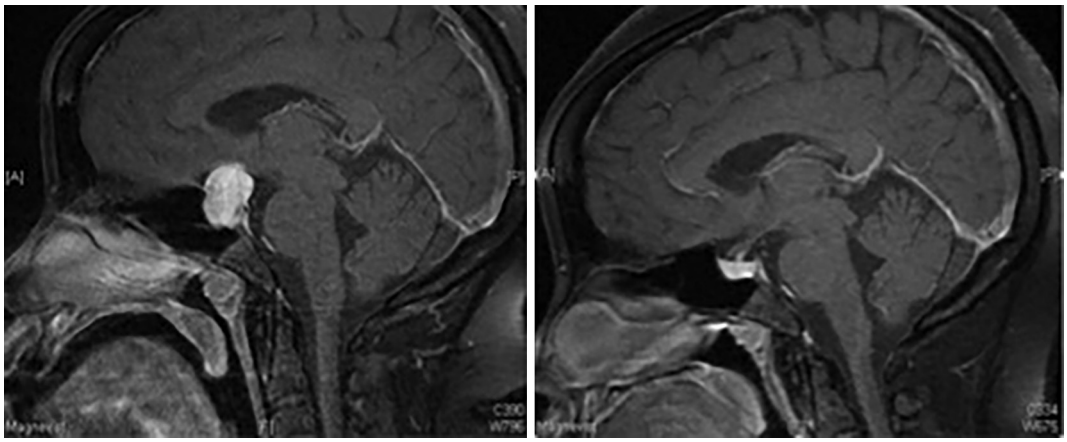


Figure 9.7.2 Sagittal MRI (T_1-weighted, post-contrast) demonstrating sellar and suprasellar enhancing mass, characteristic of lymphocytic hypophysitis in a woman presenting at 40 weeks gestation (left) and 3 months later after treatment with prednisolone 30 mg OD (right). Partial hypopituitarism (ACTH, GH and TSH deficiencies were permanent) but the patient went on to have a further spontaneous pregnancy with recurrence of hypophysitis requiring prednisolone.

relapse of disease and some patients require multimodal treatment [30, 31]. See Figure 9.7.2.

Lymphocytic hypophysitis is associated with other autoimmune disorders [32]. The pattern of involvement of the anterior pituitary, infundibulum and posterior pituitary determine the hormone deficits that are seen in this condition. Lymphocytic adenohypophysitis involves the anterior pituitary leading to loss of anterior pituitary hormone production and secretion. Lymphocytic neurohypophysitis has characteristic imaging findings and results in ADH deficiency with diabetes insipidus (DI). Panhypophysitis involves both anterior and posterior pituitary with panhypopituitarism and DI [33]. See Table 9.7.1.

In pregnancy associated lymphocytic hypophysitis, typically there is adenohypophysitis with characteristic presentation in the last month of pregnancy or within 2 months of delivery [34–36]. The fetus is unaffected by the process. It is not clear why pregnancy is associated with lymphocytic hypophysitis but a number of pathophysiological mechanisms have been proposed including; the physiological increase in size of the gland may result in increased pituitary antigenicity, there may be increased accessibility to immune reaction during pregnancy as a result of increased perfusion, alternatively placental tissue may invoke immune reactivity which cross reacts with pituitary antigenic targets [26, 28, 37, 38].

The pattern of pituitary hormone loss in lymphocytic hypophysitis differs from that seen in pituitary adenoma and other sellar/parasellar tumours. Often ACTH is the first hormone to become deficient and hyperprolactinaemia occurs in approximately 30% of non-pregnant cases [39, 40]. Hypoprolactinaemia with inability to lactate post-delivery is also seen. Lymphocytic hypophysitis may re-occur in subsequent pregnancies (either spontaneous or assisted conception [41]). See Figure 9.7.3.

Table 9.7.1 Classification of 379 patients with primary lymphocytic hypophysitis based on the anatomical location

	n	%
Lymphocytic adenohypophysitis	245	65
Lymphocytic infundibuloneurohypophysitis	39	10
Lymphocytic panhypophysitis	95	25

Derived from Caturegli P, Newschaffer C, *et al*. Autoimmune Hypophysitis. *Endocrine Reviews*, Volume 26, Issue 5, 1 August 2005, Pages 599–614. Copyright © 2005, Oxford University Press.

Management of Lymphocytic Hypophysitis

The treatment of lymphocytic hypophysitis involves initial resuscitation and supportive care with simultaneous replacement of hormone deficiency (especially hydrocortisone). Dopamine agonists can reduce elevated prolactin. High-dose glucocorticoids (typically prednisolone, by mouth or IV) are the mainstay of acute immunosuppressive treatment but there is no consensus on dose or duration. Surgery should be limited to those patients with deteriorating visual function, and /or lack of response to glucocorticoids [42, 43]. Patients are usually left with permanent hormone deficiencies, but

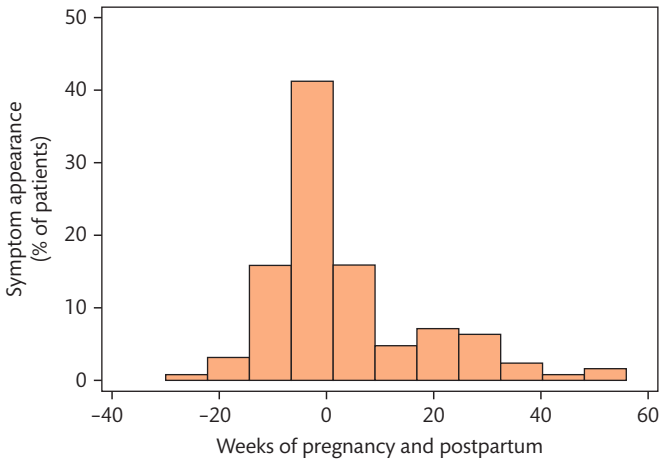


Figure 9.7.3 Typical timing of presentation/ diagnosis of lymphocytic hypophysitis in pregnancy.

Reproduced with permission from Caturegli P, Newschaffer C, Olivi A, Pomper MG, Burger PC, Rose NR. Autoimmune hypophysitis. *Endocrine Reviews*, 2005; 26: 599–614. Copyright © 2005 by the Endocrine Society.

follow-up evaluation of pituitary function is necessary as recovery occurs in some cases [44]. The long-term prognosis is good provided pituitary replacement is adequate.

Anterior Hypopituitarism in Pregnancy

Significant hypopituitarism with gonadotrophin deficiency results in infertility, and the majority of women with hypopituitarism in pregnancy have been identified and treated prior to conception. Rarely hypopituitarism develops during pregnancy or in the early postpartum period. Specific conditions resulting in loss of pituitary function include lymphocytic hypophysitis and postpartum pituitary necrosis, known as Sheehan's syndrome. This occurs as the anterior pituitary is vulnerable to abrupt hypotension due to the increased perfusion and expanded size seen in pregnancy. Sheehan's syndrome usually presents in the early postpartum period, as a result of volume loss due to postpartum haemorrhage (PPH, [45]). The patient may be acutely unwell with features of adrenal insufficiency or have a delayed presentation with fatigue, failure of lactation, and persistent amenorrhoea. Sheehan's syndrome is less common in developed countries where there is access to rapid and effective management of postpartum haemorrhage.

Whatever the aetiology of hypopituitarism, treatment with hormone replacement, should be continued or initiated in the pregnancy, depending on the hormone deficits. For most women this includes hydrocortisone and thyroxine. Internationally prednisolone and cortisone acetate are also used as replacement glucocorticoids. Most women with pre-existing hypopituitarism are also gonadotropin (follicle-stimulating hormone (FSH) and luteinizing hormone (LH)) deficient. Outside of assisted conception the majority of younger women receive oestrogen replacement, using oral or transdermal products. Most need assisted conception in the form of gonadotrophin injections, often with *in-vitro* fertilization. Therefore, if a woman with hypopituitarism wishes to conceive then she may require ovulation induction with gonadotropins, and successful pregnancies have been reported after both Sheehan's and lymphocytic hypophysitis [46].

It is thought that GH deficiency leads to difficulty with conception [47] and that GH administration increases effectiveness of gonadotrophins in ovulation induction [48]. The placenta synthesizes and secretes GH throughout pregnancy and in standard clinical practice GH replacement is discontinued once pregnancy is achieved [47].

Prior to the availability of glucocorticoid replacement therapy Addison's disease was associated with a high maternal mortality (~40%) and pregnancy was discouraged [49, 50]. Untreated adrenal insufficiency (AI) contributes to increased morbidity and mortality in pregnancy, [51]). Recognizing AI is not straightforward as the symptoms and signs (fatigue, nausea, reduction in serum sodium) are common in pregnancy. In general use of the synacthen (cosyntropin) test is avoided in pregnancy and most clinicians rely on measurement of early morning cortisol and ACTH to assess adrenal function. However, it can be used in the context of managing a woman with high index of suspicion for AI where the benefit of making a maternal diagnosis of this condition outweigh the theoretical risks of the test. Salivary cortisol may be a useful test in assessing the HPA axis [52]. The glucocorticoid requirement increases by the third trimester, typically with a 50% increase in dose. Parturition and labour should be covered with additional hydrocortisone in women with adrenal insufficiency (see Chapter 9.8, 'Adrenal Disease in Pregnancy' for precise details of recommended doses of glucocorticoid replacement) [53, 54]. The caesarean section rate is significantly higher in hypopituitary patients than in uncomplicated pregnancy. Exogenous glucocorticoid is excreted in breast milk, but at amounts that do not cause neonatal adrenal suppression [55].

Hypopituitary women usually have secondary hypothyroidism, characterized by low, or low 'normal' range serum TSH with low free thyroid hormones. The TSH level cannot be used to guide thyroxine dose adjustment and prior to conception most clinicians aim to maintain the serum free T_4 in the upper part of the reference range [56]. In early pregnancy the fetus is dependent on placental transfer of maternal thyroxine for neurodevelopment and often an early increase in thyroxine dose is required [57, 58].

Breastfeeding in the postpartum period depends to a great extent on the prolactin level of the mother. Although failure of milk production is a typical symptom of Sheehan's syndrome, hyperprolactinemia has also been reported in patients with hypopituitarism. Prolactin deficiency is common in hypopituitary patients and reflects the severity of hypopituitarism [59]. Deficiency of prolactin can lead to failure of lactation, alternatively hypersecretion can lead to galactorrhoea and nipple tenderness, neither of which is suitable for breastfeeding. Hypopituitary patients generally do not have sufficient milk production to breastfeed effectively.

Management of Pituitary Replacement in Pregnancy

The Endocrine Society has produced guidelines for the management of hypopituitarism and these include guidance on the management of hormone deficiencies in women with hypopituitarism [60]. The guidance recommends that hydrocortisone is the preferred glucocorticoid for use in pregnancy and that the dose required is individual. Higher doses than outside of pregnancy may be required, especially in the third trimester. The guidance recommends that close monitoring for clinical features of over- and under- replacement is performed. Dexamethasone is not inactivated by placental tissue action and is not recommended for routine use in pregnancy. Stress doses of hydrocortisone are recommended for labour/ delivery, similar to those used in major surgery (see Chapter 9.8). The Endocrine Society guidelines recommend that thyroid status is assessed 4–6 weekly when there is pituitary insufficiency with central hypothyroidism. Increased levothyroxine doses may be required to achieve target levels appropriate for pregnancy. The guidance recommends discontinuation of GH during pregnancy as there is a lack of information regarding efficacy and safety and the placenta produces GH throughout pregnancy [60].

Diabetes Insipidus in Pregnancy

Physiology of ADH Secretion and Action, Sodium Levels, and Resetting of the Osmostat in Pregnancy

Antidiuretic hormone (ADH, also known as vasopressin) is a nine amino acid peptide, secreted by the posterior pituitary and acting on the vasopressin (G protein-coupled) receptors. Release is regulated by hypothalamic osmoreceptors, arterial baroreceptors and renal blood supply [61, 62]. ADH stimulates V2 receptors resulting in insertion of aquaporin-2 in the cells of collecting ducts. This facilitates resorption of water and concentration of urine, maintaining electrolyte concentration and circulating volume [62]. Oxytocin and vasopressin are similar in structure and ADH can activate the oxytocin receptor [63].

During pregnancy there is a significant increase in circulating (plasma) volume with a decrease in plasma osmolality (approximately 10 mOsm/kg [64]). Total body water is increased by approximately 8 L in pregnancy, in part due to hormone mediated systemic circulation vasodilation [65]. The reduction in osmolality, and lower blood pressure result in a resetting of the osmostat and thirst perception, with desire to drink and secretion of vasopressin occurring at lower osmolality than occurs in the non-gravid state [65]. Changes in osmotic threshold occurs early (6–8 weeks gestation) and persists throughout the duration of pregnancy. Oxytocinase (cysteine aminopeptidase) is produced by the placenta and degrades vasopressin rendering it inactive [66]. Placental vasopressinase is most active from the middle of the second trimester to the end of pregnancy, normalizing quickly post-delivery [67]. ADH secretion therefore increases during pregnancy and by the third trimester there is a four-fold increase in ADH secretion. Vasopressinase is metabolically degraded by liver action and diseases that interfere with liver function result in increased vasopressinase and can contribute to DI or worsen symptoms in the patient with established DI [68, 69].

Diabetes Insipidus in the Pregnant Woman

The aetiology and incidence of diabetes insipidus (DI) in pregnancy are similar to those of the non-pregnant population. A number of pregnancy-specific conditions can result in transient DI; including pre-eclampsia, HELLP (haemolysis, elevated liver enzymes, and low platelets) syndrome and acute fatty liver of pregnancy (AFLP [69]). Gestational diabetes insipidus (GDI) is a pregnancy specific condition where metabolism of ADH by the placental enzyme vasopressinase results in reduced ADH action with polyuria. In this condition production of and sensitivity to ADH are unimpaired. Vasopressinase levels are usually even higher in multiple pregnancies [70]).

Established DI may become uncontrolled during pregnancy, due to physiological changes affecting vasopressin (anti-diuretic hormone (ADH)), including increased glomerular filtration rate, placental production of vasopressinase (responsible for ADH breakdown), and increased renal resistance to ADH [71]. Affected women present with worsening of previously controlled thirst and polyuria.

Outside of pregnancy the diagnosis of DI is usually confirmed by fluid deprivation testing, but clinicians typically avoid this test in pregnancy. Measurement of copeptin may become a further clinical tool to aid in the diagnosis [72]. Measurement of serum and urine osmolality and sodium is usually sufficient to allow sufficient judgement on the presence of DI and need for treatment.

Adequately treated DI should not have adverse effects on the pregnancy. Women diagnosed with diabetes insipidus should have regular consultant review in clinic with monitoring of serum electrolytes and hydration status. Antenatal anaesthetic review is important and labour should be managed on the delivery suite with adequately experienced obstetric and anaesthetic oversight. Conversely, if DI is undiagnosed or inadequately treated then there are the risks of severe dehydration and electrolyte disturbances, which may lead to maternal seizures or oligohydramnios [61].

Management of Diabetes Insipidus in Pregnancy

Desmopressin (DDAVP) is the most appropriate treatment of DI during pregnancy. DDAVP is a synthetic form of ADH with a different N-terminal that renders it less susceptible to degradation by vasopressinase [73, 74]. DDAVP can be administered orally, sublingually, intranasally or by subcutaneous, intramuscular, or intravenous injection. It is safe throughout pregnancy and breastfeeding, with no known adverse effects upon delivery of the neonate [75].

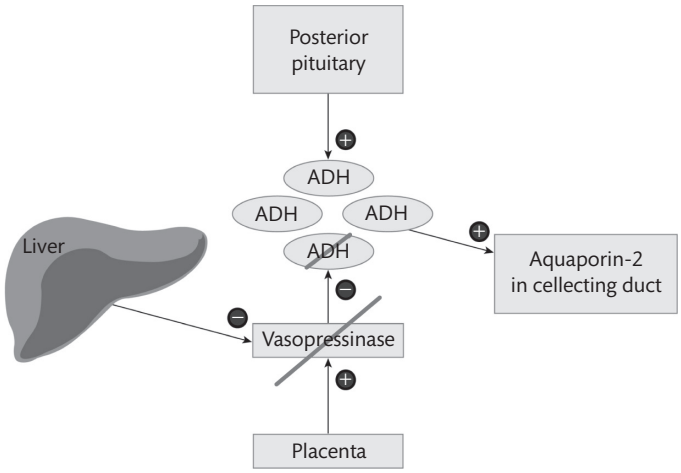


Figure 9.7.4 ADH physiology in normal pregnancy. In healthy pregnancy, the posterior pituitary produces ADH that acts to prevent the loss of water via aquaporins in the kidney. The placenta produces vasopressinase, which degrades ADH. In normal pregnancy, the liver degrades vasopressinase, and the posterior pituitary produces significant amounts of ADH, so that the effect of vasopressinase is not clinically significant. In conditions associated with liver dysfunction increased vasopressinase activity occurs, which can result in new-onset DI or an increased requirement for DDAVP.

Desmopressin has less affinity for the uterine oxytocin receptors then lysine or arginine-vasopressin [76]. As the circulation is expanded in pregnancy and the osmostat reset the desmopressin dose should be titrated to facilitate thirst management and maintain appropriate electrolyte levels. Desmopressin is not heavily metabolized by placental oxytocinase [77]. If desmopressin is being used care must be taken when IV fluid is administered as water intoxication can occur. In patients with new-onset DI the on-going need for desmopressin should be reviewed in the clinical context, and with appropriate investigations post-delivery. In general, the DDAVP requirement is unchanged during pregnancy, but a modest increase in dose may be required. If the dose of DDAVP is increased in pregnancy, it is usually possible to reduce back to the pre-pregnancy dose soon after delivery. DDAVP is safe for breastfeeding the newborn in lactating mothers [78]. See Figure 9.7.4.

REFERENCES

1. Gonzalez JG, Elizondo G, Saldivar D, Nanez H, Todd LE, Villarreal JZ. Pituitary gland growth during normal pregnancy: an in vivo study using magnetic resonance imaging. *Am J Med*, 1988; **85**: 217–20.
2. Dinc H, Esen F, Demirci A, Sari A, Resit GH. Pituitary dimensions and volume measurements in pregnancy and post-partum. MR assessment. *Acta Radiologica*, 1998; **39**: 64–9.
3. Kanal E, Barkovich AJ, Bell C, *et al*. ACR guidance document for safe MR practices. *Am J Roentgenol*, 2007; **188**: 1447–74.

4. Tyson JE, Hwang P, Guyda H, Friesen HG. Studies of prolactin secretion in human pregnancy. *Am J Obstet Gynecol*, 1972; **113**: 14–20.
5. Leake RD, Weitzman RE, Glatz TH, Fisher DA. Plasma oxytocin concentrations in men, nonpregnant women, and pregnant women before and during spontaneous labor. *J Clin Endocrinol Metab*, 1981; **53**: 730–3.
6. Carr BR, Parker CR Jr, Madden JD, MacDonald PC, Porter JC. Maternal plasma adrenocorticotropin and cortisol relationships throughout human pregnancy. *Am J Obstet Gynecol*, 1981; **139**: 416–22.
7. Demey-Ponsart E, Foidart JM, Sulon J, Sodoyez JC. Serum CBG, free and total cortisol and circadian patterns of adrenal function in normal pregnancy. *J Steroid Biochem*, 1982; **16**: 165–9.
8. Scott EM, McGarrigle HH, Lachelin GC. The increase in plasma and saliva cortisol levels in pregnancy is not due to the increase in corticosteroid-binding globulin levels. *J Clin Endocrinol Metab*, 1990: **71**: 639–44.
9. Elsheikh A, Creatsas G, Mastorakos G, Milingos S, Loutradis D, Michalas S. The renin-aldosterone system during normal and hypertensive pregnancy. *Arch Gynecol Obstet*, 2001; **264**: 182–5.
10. Nolten WE, Lindheimer MD, Rueckert PA, Oparil S, Ehrlich EN. Diurnal patterns and regulation of cortisol secretion in pregnancy. *J Clin Endocrinol Metab*, 1980: **51**: 466–72.
11. Allolio B, Hoffmann J, Linton EA, Winkelmann W, Kusche M, Schulte HM. Diurnal salivary cortisol patterns during pregnancy and after delivery relationship to plasma corticotrophin-releasing-hormone. *Clin Endocrinol*, 1990: **33**: 279–89.
12. Magiakou MA, Mastorakos G, Rabin D, *et al.* The maternal hypothalamic–pituitary–adrenal axis in the third trimester of human pregnancy. *Clin Endocrinol*, 1996; **44**: 419–28.
13. Karaca Z, Tanriverdi F, Unluhizarci K, Kelestimur F. Pregnancy and pituitary disorders. *Eur J Endocrinol*, 2010; **162**: 453–75.
14. Frankenne F, Closset J, Gomez F, Scippo ML, Smal J, Hennen G. The physiology of growth hormones (GHs) in pregnant women and partial characterization of the placental GH variant. *J Clin Endocrinol Metab*, 1988: **66**: 1171–80.
15. Eriksson L, Frankenne F, Eden S, Hennen G, Von Schoultz B. Growth hormone 24-h serum profiles during pregnancy—lack of pulsatility for the secretion of the placental variant. *Br J Obstet Gynaecol*, 1989; **106**: 949–53.
16. Caufriez A, Frankenne F, Englert Y, Golstein J, Cantraine F, Hennen G, Copinschi G. Placental growth hormone as a potential regulator of maternal IGF-1 during human pregnancy. *Am J Physiol*, 1990; **258**: E1014–19.
17. Yoshimura M, Hershman JM. Thyrotropic action of human chorionic gonadotropin. *Thyroid*, 1995; **5**: 425–34.
18. Pekonen F, Alfthan H, Stenman U-H, Ylikorkala O. Human chorionic gonadotropin (hCG) and thyroid function in early human pregnancy: circadian variation and evidence for intrinsic thyrotropic activity of hCG. *J Clin Endocrinol Metab*, 1988; **66**: 853–6.
19. Goodwin TM, Hershman JM. Hyperthyroidism due to inappropriate production of human chorionic gonadotropin. *Clin Obstet Gynecol*, 1997; **40**: 32–44.
20. Gruen JP, Meuris S, De Nayer P, Glinoer D. The thyrotropic role of human chorionic gonadotropin (hCG) in the early stages of twin (vs. single) pregnancy. *Clin Endocrinol*, 1997; **46**: 719–25.
21. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocr Rev*, 1997; **18**: 404.
22. Glinoer D, de Nayer P, Bourdoux P, *et al.* Regulation of maternal thyroid during pregnancy. *J Clin Endocrinol Metab*, 1990; **71**: 276–87.
23. Cotzias C, Wong SJ, Taylor E, Seed P, Girling J. A study to establish gestation-specific reference intervals for thyroid function tests in normal singleton pregnancy. *Eur J Obstet Gynecol Reprod Biol*, 2008; **137**: 61–6.
24. Donovan LE, Metcalfe A, Chin A, *et al.* A practical approach for the verification and determination of site- and trimester-specific reference intervals for thyroid function tests in pregnancy. *Thyroid*, 2019; **29**: 412–20.
25. Faje A. Hypophysitis: evaluation and management. *Clin Diabetes Endocrinol*, 2016; **2**: 15.
26. Caturegli P, Newschaffer C, Olivi A, Pomper MG, Burger PC, Rose NR. Autoimmune hypophysitis. *Endocr Rev*, 2005; **26**: 599–614.
27. Thodou E, Asa SL, Kontogeorgos G, *et al.* Lymphocytic hypophysitis: clinicopathological findings. *J Clin Endocrinol Metab*, 1995; **80**: 2302.
28. Gutenberg A, Hans V, Puchner MJ, Kreutzer J, Bruck W, Caturegli P, Buchfelder M. Primary hypophysitis: clinical–pathological correlations. *Eur J Endocrinol*, 2006; **155**: 101–7.
29. Lury KM. Inflammatory and infectious processes involving the pituitary gland. *Top Magn Reson Imaging*, 2005; **16**: 301–6.
30. Honegger J, Buchfelder M, Schlaffer S, *et al.* Treatment of primary hypophysitis in Germany. *J Clin Endocrinol Metab*, 2015; **100**: 3460–9.
31. Lupi I, Manetti L, Raffaelli V, *et al.* Diagnosis and treatment of autoimmune hypophysitis: a short review. *J Endocrinol Invest*, 2011; **34**: e245–52.
32. Bellastella G, Maiorino MI, Bizzarro A, *et al.* Revisitation of autoimmune hypophysitis: knowledge and uncertainties on pathophysiological and clinical aspects. *Pituitary*, 2016; **19**: 625–42.
33. Rivera JA. Lymphocytic hypophysitis: disease spectrum and approach to diagnosis and therapy. *Pituitary*, 2006; **9**: 35–45.
34. Beressi N, Beressi JP, Cohen R, Modigliani E. Lymphocytic hypophysitis. A review of 145 cases. *Annales de Medecine Interne*, 1999; **150**: 327–41.
35. Hashimoto K, Takao T, Makino S. Lymphocytic adenohypophysitis and lymphocytic infundibuloneurohypophysitis. *Endocr J*, 1997; **44**; 1–10.
36. Kidd D, Wilson P, Unwin B, Dorward N. Lymphocytic hypophysitis presenting early in pregnancy. *J Neurol*, 2003; **250**: 1385–7.
37. O'Dwyer DT, Smith AI, Matthew ML, *et al.* Identification of the 49-kDa autoantigen associated with lymphocytic hypophysitis as alpha-enolase. *J Clin Endocrinol Metab*, 2002: **87**: 752–7.
38. O'Dwyer DT, Clifton V, Hall A, Smith R, Robinson PJ, Crock PA. Pituitary autoantibodies in lymphocytic hypophysitis target both gamma- and alpha-enolase – a link with pregnancy? *Arch Physiol Biochem*, 2002; **110**: 94–8.
39. McCutcheon IE, Oldfield EH. Lymphocytic adenohypophysitis presenting as infertility. Case report. *J Neurosurg*, 1991; **74**: 821–6.
40. Thodou E, Asa SL, Kontogeorgos G, Kovacs K, Horvath E, Ezzat S. Clinical case seminar: lymphocytic hypophysitis: clinicopathological findings. *J Clin Endocrinol Metab*, 1995; **80**: 2302–11.
41. Biswas M, Thackare H, Jones MK, Bowen-Simpkins P. Lymphocytic hypophysitis and headache in pregnancy. *Br J Obstet Gynaecol*, 2002; **109**: 1184–6.

42. Leung GK, Lopes MB, Thorner MO, Vance ML, Laws ER Jr. Primary hypophysitis: a single-center experience in 16 cases. *J Neurosurg*, 2004; **101**: 262–71.
43. Tsametis C, Karavitaki N, Ansorge O, Turner HE, Wass JAH. Lymphocytic hypophysitis: what is the long-term outcome? *Endocrine Abstracts*, 2005; **9**: P122.
44. Sinha D, Sinha A, Pirie AM. A case of recurrent lymphocytic hypophysitis in pregnancy. *J Obstet Gynaecol*, 2006; **26**: 255–6.
45. Molitch ME. Pituitary disorders during pregnancy. *Endocrinol Metab Clin N Am*, 2006; **35**: 99–116.
46. Funazaki S, Yamada H, Hara K, Ishikawa SE. Spontaneous pregnancy after full recovery from hypopituitarism caused by lymphocytic hypophysitis. *Endocrinol Diabetes Metab Case Rep*, 2018: pii: EDM180081.
47. de Boer JA, Schoemaker J, van der Veen EA. Impaired reproductive function in women treated for growth hormone deficiency during childhood. *Clin Endocrinol*, 1997; **46**: 681–9.
48. Homburg R, Levy T, Ben Rafael Z. Adjuvant growth hormone for induction of ovulation with gonadotrophin-releasing hormone agonist and gonadotrophins in polycystic ovary syndrome: a randomized, double-blind, placebo-controlled trial. *Hum Reprod*, 1995; **10**: 2550–3.
49. Lekarev O, New MI. Adrenal disease in pregnancy. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 959–73.
50. Schneiderman M, Czuzoj-Shulman N, Spence AR, Abenhaim HA. Maternal and neonatal outcomes of pregnancies in women with Addison's disease: a population-based cohort study on 7.7 million births. *BJOG*, 2017: **124**: 1772–9.
51. Minneci PC, Deans KJ, Banks SM, Eichacker PQ, Natanson C. Meta-analysis: the effect of steroids on survival and shock during sepsis depends on the dose. *Ann Int Med*, 2004; **14**: 147–56.
52. Suri D, Moran J, Hibbard JU, Kasza K, Weiss RE. Assessment of adrenal reserve in pregnancy: defining the normal response to the adrenocorticotropin stimulation test. *J Clin Endocrinol Metab*, 2006; **91**: 3866–72.
53. Arlt W, Allolio B. Adrenal insufficiency. *Lancet*, 2003; **361**: 1881–93.
54. Langlois F, Lim DST, Fleseriu M. Update on adrenal insufficiency: diagnosis and management in pregnancy. *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 184–92.
55. McKenzie SA, Selley JA, Agnew JE. Secretion of prednisolone into breast milk. *Arch Dis Childhood*, 1975; **50**: 894–6.
56. LaFranchi S. Thyroid hormone in hypopituitarism, Graves' disease, congenital hypothyroidism, and maternal thyroid disease during pregnancy. *Growth Horm IGF Res*, 2006; **16**: 20–4.
57. LaFranchi SH, Haddow JE, Hollowell JG. Is thyroid inadequacy during gestation a risk factor for adverse pregnancy and developmental outcomes? *Thyroid*, 2005; **15**: 60–71.
58. DeGroot L, Abalovich M, Alexander EK, *et al.* Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 2543–65.
59. Teng-Teng C, Koch CA, Monson JP. Hypopituitarism. In: Feingold KR, Anawalt B, Boyce A, *et al.*, eds. www.Endotext.org. South Dartmouth (MA): MDText.com, 2000.
60. Fleseriu M, Hashim IA, Karavitaki N, Melmed S, Murad MH, Salvatori R, Samuels MH. Hormonal replacement in hypopituitarism in adults: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 3888–921.
61. Quigley J, Shelton C, Issa B, Sripada S. Diabetes insipidus in pregnancy. *Obstet Gynaecol*, 2018; **20**: 41–8.
62. Knepper MA, Kwon T-H, Nielsen S. Physiology of water balance. *N Engl J Med*, 2015; **372**: 1349–58.
63. Peter J, Burbach H, Adan RA, Lolait SJ, van Leeuwen FW, Mezey E, Palkovits M, Barberis C. Molecular neurobiology and pharmacology of the vasopressin/oxytocin receptor family. *Cell Mol Neurobiol*, 1995; **15**: 573–95.
64. Brunton PJ, Arunachalam S, Russel JA. Control of neurohypophysial hormone secretion, blood osmolality and volume in pregnancy. *J Physiol Pharmacol*, 2008; **59**(Suppl 8): 27–45.
65. Schrier RW. Systemic arterial vasodilation, vasopressin, and vasopressinase in pregnancy. *J Am Soc Nephrol*, 2010; **21**: 570–2.
66. Davison JM, Sheills EA, Philips PR, Barron WM, Lindheimer MD. Metabolic clearance of vasopressin and an analogue resistant to vasopressinase in human pregnancy. *Am J Physiol*, 1993; **264**(2 Pt 2): F348–53.
67. Refardt J, Christ-Crain M. Diabetes insipidus in pregnancy: how to advice the patient? *Minerva Endocrinol*, 2018; **43**: 458–64.
68. Mor A, Fuchs Y, Zafra K, Haberman S, Tal R. Acute presentation of gestational diabetes insipidus with pre-eclampsia complicated by cerebral vasoconstriction: a case report and review of the published work. *J Obstet Gynaecol Res*, 2015; **41**: 1269–72.
69. Gambito R, Chan M, Sheta M, *et al.* Gestational diabetes insipidus associated with HELLP syndrome: a case report. *Case Rep Nephrol*, 2012; **2012**: 640365.
70. Marques P, Gunawardana K, Grossman A. Transient diabetes insipidus in pregnancy. *Endocrinol Diabetes Metab Case Rep*, 2015; **2015**; 150078.
71. Durr JA, Lindheimer MD. Diagnosis and management of diabetes insipidus during pregnancy. *Endocr Pract*, 1996; **2**: 353–61.
72. Fenske W, Refardt J, Chifu I, *et al.* A copeptin-based approach in the diagnosis of diabetes insipidus. *N Engl J Med*, 2018; **379**: 428–39.
73. Ananthakrishnan S. Diabetes insipidus in pregnancy: etiology, evaluation, and management. *Endocr Pract*, 2009; **15**: 377–82.
74. Kalelioglu I, Kubat Uzum A, Yildirim A, Ozkan T, Gungor F, Has R. Transient gestational diabetes insipidus diagnosed in successive pregnancies: review of pathophysiology, diagnosis, treatment, and management of delivery. *Pituitary*, 2007; **10**: 87–93.
75. Ray JG. DDAVP use during pregnancy: an analysis of its safety for mother and child. *Obstet Gynecol Surv*, 1998; **53**: 450–5.
76. Edwards CR, Kitau MJ, Chart T, Besser GM. Vasopressin analogue DDAVP in diabetes insipidus: clinical and laboratory studies. *BMJ*, 1973; **3**: 375–8.
77. Kallen BA, Carlsson SS, Bengtsson BK. Diabetes insipidus and use of desmopressin (minirin) during pregnancy. *Eur J Endocrinol*, 1995; **132**: 144–6.
78. British National Formulary. *6.5.2 Posterior Pituitary Hormones and Antagonists: Desmopressin.* London: National Institute for Health and Care Excellence; 2016

9.8

Adrenal Disease in Pregnancy

David J. Torpy, Michael W. O'Reilly, and Sunita M.C. De Sousa

Introduction

Adrenal gland dysfunction in pregnancy is rare, but fetal and maternal complications are significantly increased, and prompt diagnosis and treatment are therefore critical to prevent morbidity and mortality. Adrenal disease may present for the first time during pregnancy, or may be diagnosed before conception and require careful monitoring and management throughout gestation [1]. Diagnosis of adrenal disorders during pregnancy is frequently confounded by physiological changes in maternal adrenal hormone metabolism as pregnancy progresses [2], which complicates interpretation of biochemical investigations. The hypothalamic–pituitary–adrenal (HPA) axis and renin-angiotensin-aldosterone system (RAAS) are significantly upregulated in pregnancy, although adrenomedullary function is not dramatically altered. Furthermore, dynamic endocrine testing is generally contraindicated in pregnancy. Symptoms of adrenal gland dysfunction may also mimic typical symptoms of pregnancy. The aim of this chapter is to review clinical aspects of the most common adrenal disorders in pregnancy, and to discuss approaches to diagnosis and management.

Addison's Disease—Primary Adrenal Insufficiency

Primary adrenal insufficiency (PAI) is glucocorticoid deficiency due to direct adrenal disease. Prevalence is 90–140 per million in Western countries. Predominant aetiologies are autoimmunity in adults and congenital adrenal hyperplasia in children; tuberculosis is important in prevalent areas [3–9]. Autoimmune PAI is associated with antibodies to the adrenal steroidogenic enzyme, 21-hydroxylase, and 50% of cases have autoimmune comorbidities [7]. Less frequent causes of PAI include adrenalectomy, adrenal haemorrhage, infiltration/infections, drugs, and genetic disorders. Characteristic presenting symptoms of adrenal insufficiency include fatigue, postural dizziness, nausea, abdominal pain, and weight loss, which may culminate in shock, known as an adrenal crisis [10]. Cortisol deficiency produces hypersecretion of ACTH/POMC peptides with consequent hyperpigmentation, and is often associated with deficient aldosterone which may cause hyperkalaemia and exacerbate hyponatraemia. Treatment, available since the 1950s, comprises glucocorticoid and mineralocorticoid (fludrocortisone) replacement. Fludrocortisone is monitored using electrolytes, BP, and plasma renin [11]. Glucocorticoid replacement is monitored clinically to avoid overdose (Cushing's syndrome) and underdosing (recrudescence of adrenal insufficiency). There has been a trend towards use of short-acting natural glucocorticoids in lower dose to avoid osteoporosis and cardiovascular effects, in the context of previous studies that overestimated natural cortisol (hydrocortisone) production rates [12–15]. Prognosis of PAI is generally good but all-cause mortality is slightly higher than the population; of that increase, adrenal crisis, which has an annual risk per PAI patient of approximately 8%, contributes to 15% of all deaths in autoimmune AI [3, 16–18], and 42% in CAH [19]. Despite standard hormone replacement, a substantial minority of PAI patients experience reduced quality of life, the basis of which is uncertain [20–22].

Untreated PAI may impair ovulation [23], folliculogenesis [24], and libido/sexual activity [25]. These factors together with the effects of associated autoimmunity, with or without ovarian immunological damage, result in reduced fecundity [23–25]. Untreated Addison's disease was previously associated with 35% maternal mortality, but fatalities are now rare [23]. Obstetric outcomes in treated PAI are now similar to normal pregnancies, apart from increased rates of preterm birth and Caesarean section [24, 25]. Speculatively, preterm birth may relate to disturbances of placental CRH secretion [25]. In a study of 517 women who gave birth after PAI diagnosis or within 3 years prior to diagnosis, maternal, and neonatal mortality, Apgar score and congenital malformation rates were similar to controls. Birth weight was also similar in women diagnosed with PAI prior to pregnancy but it was reduced in women

diagnosed after pregnancy, indicating that treatment mitigates the risk of hypocortisolism-induced low birth weight [25].

Pregnancy is a state of physiological hypercortisolism, with ACTH, total and free plasma cortisol, and corticosteroid-binding globulin (CBG), rising progressively during gestation, reaching 2- to 3-fold by parturition [26, 27]. Pregnant women with PAI should undergo endocrine review each trimester with titration of steroid hormone replacement [11]. Hydrocortisone is the preferred glucocorticoid because inactivation by placental 11β-hydroxysteroid dehydrogenase 2 (11βHSD2) protects the fetus from glucocorticoid excess, it has a shorter half-life allowing diurnal variation in dosing, and modification is not required for the drug to be active which may be important in adrenal crises [11, 24, 28]. Prednisolone, which has a longer plasma half-life than hydrocortisone (150 vs. 90 mins) is an acceptable alternative in pregnancy as it has comparable metabolism. Dexamethasone is avoided as it is not inactivated by 11βHSD2 and may have adverse fetal effects. As in non-pregnant individuals, glucocorticoid dose titration is individualized, based on avoidance of the clinical features of adrenal insufficiency or Cushing's. In pregnancy, this requires particular judgment as features of pregnancy such as postural hypotension and fatigue, or striae, oedema, and hyperglycaemia, may suggest under- or over-replacement, respectively. As a rule of thumb, the consensus-based increment of glucocorticoid is a 20–40% increase in total daily dose, roughly equating to 5–10 mg hydrocortisone, from 24 weeks onwards [11, 24]. Pregnancy is a state of secondary hyperaldosteronism. Elevated renin and aldosterone levels are seen in normal pregnancy; hence, renin cannot be used in dose titration, but urinary sodium and potassium levels may be informative. Generally, fludrocortisone dose does not need to be altered in pregnancy [11].

Labour or Caesarean section is a physiological stressor and glucocorticoid dosing should be increased as per major surgery guidelines and the NIHR guideline on the management of medical disorders in labour (https://www.nice.org.uk/guidance/ng121), incorporating a 100 mg IV hydrocortisone injection followed by 200 mg/24 h (50 mg q6hr or infusion 8.3 mg/h) for 24 hours, then double usual dose for 48 hours, then return to usual dose [11]. 'Sick day' management should be refreshed in pregnancy, incorporating increased glucocorticoid dose with fever (2-fold if temp >38°C, 3-fold if >39°C) for 2–3 days until recovery, use of home parenteral hydrocortisone IM/SC 100 mg when oral medication is not possible, a Medic Alert® bracelet, and a card/letter informing healthcare providers of the need to administer hydrocortisone with illness [11].

Adrenal crisis treatment in pregnancy does not differ from usual measures, but shock has fetal risks, which may be greater early in the first trimester when the fetal HPA axis is not yet producing substantial quantities of cortisol. A reliable treatment protocol follows: hydrocortisone 100 mg IV stat followed by up to 200 mg/24 hours by continuous infusion or IV/IM 50 mg boluses q6 hr, tapering rapidly depending on patient response with BP and symptoms followed closely. Intravenous glucose may be needed in rare cases with concomitant hypoglycaemia. Transfer to oral hydrocortisone within 24 hours is advisable unless there is an ongoing underlying inflammatory precipitant. Intravenous fluids, generally normal saline, are given: 1000 ml within the first hour, with further crystalloid fluid being administered according to standard resuscitation guidelines, taking into account the patient's circulatory status, body size, and relevant comorbidities. Treatment of the precipitating cause is essential [11].

PAI is rarely diagnosed in pregnancy. If suspected, blood should be drawn for simultaneous cortisol and ACTH measurements and the woman empirically treated with physiological or stress dose steroids, depending on clinical severity [11]. If the woman is haemodynamically stable, basal morning cortisol and ACTH should be measured, followed by a 250 mcg short synacthen (cosyntropin) test if results are equivocal [11, 28]. Because of total cortisol excess due to CBG and free cortisol excess due to placental CRH and ACTH, higher cut-off levels are required to diagnose PAI in pregnancy with basal plasma cortisol cut-offs of <300, <450, and <600 nmol/L in the first, second, and third trimesters, respectively [24, 27]. Normal pregnancy is associated with progressively rising ACTH [27]; however, plasma ACTH greater than 2-fold normal at any gestation is consistent with PAI [24, 28]. ACTH samples should be sent on ice and repeat measurement reduces the impact of physiological fluctuations [28]. Post-cosyntropin cortisol cut-offs of 700, 800 and 900 nmol/L should be used in the first, second and third trimesters, respectively, rather than typical non-pregnant thresholds around 500 nmol/L [24]. A single dose of cosyntropin appears to be safe in pregnancy [24]. Aldosterone and renin values need to be interpreted with caution, as they are elevated in normal pregnancy and reference ranges have not been established [28]. Women diagnosed with PAI should be screened for relevant autoimmune comorbidities, noting that untreated hypocortisolism may transiently increase TSH and that thyroxine commencement may precipitate an adrenal crisis due to increased cortisol clearance [23, 24].

Other Causes of Hypoadrenalism

Secondary hypoadrenalism (SAI) is 2-fold more common than PAI, but perhaps not in women in their childbearing years. SAI refers to hypocortisolism due to ACTH deficiency from a wide range of pituitary/hypothalamic disorders, most often pituitary adenomas and their treatment with surgery/radiotherapy, craniopharyngiomas, granulomatous/infiltrative disorders, autoimmunity (which may cause isolated ACTH deficiency) and mutations of pituitary developmental genes [29]. In contrast to PAI: SAI does not result in mineralocorticoid deficiency hence fludrocortisone treatment is not required; there is no ACTH/POMC hypersecretion so hyperpigmentation does not develop; and other pituitary hormones, particularly TSH, gonadotropins and growth hormone may necessitate hormonal replacement. Generally, SAI has a lower risk of adrenal crisis than PAI, although the quantitative risk of adrenal crisis specifically in pregnancy is not well described.

Glucocorticoid-induced adrenal insufficiency (GC-AI) refers to the prolonged ACTH deficiency and adrenal atrophy that follows long-term supraphysiological glucocorticoid therapy [30]. The risk of adrenal crisis among these patients appears to be very low, reflecting often incomplete cortisol suppression [30]. Use of systemic glucocorticoids (oral, inhaled) for inflammatory disorders is frequent (3% population), making GC-AI common, although data in pregnancy are limited. In potential GC-AI, the measurement of morning cortisol, taking into account the effect of pregnancy on cortisol and CBG, may often be reassuring about the presence of endogenous cortisol, perhaps assisted by placental CRH acting on the

pituitary as well as putative placental ACTH [31]. The ongoing use of systemic glucocorticoids is guided by the activity of the underlying inflammatory disease, which often improves in pregnancy.

Phaeochromocytoma

Phaeochromocytomas are catecholamine-producing tumours which originate histologically from chromaffin cells of the adrenal medulla. Extra-adrenal catecholamine-producing lesions typically originate from the sympathetic paravertebral ganglia, and are termed paragangliomas. Collectively, phaeochromocytomas and paragangliomas are referred to as PPGLs [32]. Phaeochromocytomas have an incidence of 2–8 million/year in the background population. In pregnancy, phaeochromocytoma has been estimated to occur in approximately one in 54 000 pregnancies [33]. Hereditary PPGLs are observed in up to 40% of cases, with up to 15 well-recognized PPGL driver genes and 12–15 genetic syndromes. Although rare, expedient diagnosis of an underlying phaeochromocytoma in a pregnant woman is critical, as maternal and fetal mortality may be as high as 50% in untreated cases [34]. Maternal mortality improved from 48% before 1969 to 4% in the 1990s; fetal outcomes remain poor, despite significant advances in diagnosis and treatment [35].

Physiological Changes in Adrenomedullary Function in Pregnancy

Active catecholamines such as adrenaline and noradrenaline play a major role in cardiovascular homeostasis, and are a reflection of global sympathetic and adrenal-medullary function. Pregnancy is not associated with dramatic alterations in catecholamine metabolism. In healthy pregnant women, plasma and urinary catecholamine levels approximate those of non-pregnant women [36]. The placenta is also an effective barrier to maternal catecholamines [37]. The placenta expresses the catecholamine-metabolizing enzymes catechol-O-methyltransferase (COMT) and monoamine oxidase. Intriguingly, maternal COMT deficiency is associated with pre-eclampsia (PE), and levels of COMT are significantly lower in women with severe PE variants [38]. This may be due to catecholamine-induced vasoconstriction in the uteroplacental circulation.

Clinical Manifestations and Associated Risks of Phaeochromocytoma in Pregnancy

Clinical signs and symptoms of phaeochromocytoma are driven by paroxysms of tumoral catecholamine production, and include headaches, hypertension, sweating, and palpitations. One small study has suggested that the prevalence of classic symptoms in pregnancy is lower than in non-pregnant patients [35]. However, as pregnancy progresses the incidence of symptoms increases; in a series of 35 pregnant women with phaeochromocytoma, 66% reported symptoms in the third trimester, compared to 11% and 20% in the first and second trimester, respectively [39]. Alongside tumour expansion, catecholamine surges may be driven by uterine growth, fetal movement and abdominal palpation as pregnancy progresses. In undiagnosed women, differentiating signs of catecholamine excess from PE and pregnancy-induced hypertension (PIH) may pose a challenge, resulting in a delayed diagnosis. Onset of hypertension before 20 weeks' gestation is not consistent with PE or PIH. A number of genetic susceptibility syndromes associated with significant familial risk of PPGL have been identified. These include multiple endocrine neoplasia type 2 (MEN2) and neurofibromatosis type 1 (NF1); also, mutations in mitochondrial succinate dehydrogenase (SDH) genes (A, B, D, and AF2), which encode subunits of the mitochondrial succinate dehydrogenase complex, confer a significant lifetime risk of PPGL. Evidence to support a structured surveillance strategy in pregnant mutation carriers is lacking, but clinical intuition would suggest close clinical and biochemical monitoring is warranted.

Adverse fetal and maternal outcomes in pregnant women with phaeochromocytoma are mediated by excessive tumoral catecholamine adrenoceptor stimulation. Similar to non-pregnant patients, cardiovascular complications in pregnant women include hypertensive crises, tachyarrhythmias, and left ventricular failure due to catecholamine-induced cardiomyopathy. The maternal mortality rate in undiagnosed or untreated phaeochromocytoma in pregnancy is as high as 50% in historical cases, with the greatest risk in the peripartum period [34]. Early antepartum diagnosis, with appropriate α-adrenoceptor blockade, has resulted in significantly improved maternal outcomes in the last 50 years. Fetal mortality remains high, even in the context of appropriate alpha blockade and specialist management.

Diagnosis

Phaeochromocytoma should be considered in pregnancy in the context of classic signs and symptoms of catecholamine excess. Biochemical testing for catecholamine excess should involve measurement of plasma metanephrines and normetanephrines, which are O-methylated metabolites of adrenaline and noradrenaline, respectively; these have the highest sensitivity (95–100%) and therefore best negative predictive value for diagnosis of phaeochromocytoma [40]. Urinary fractionated metanephrines are also highly sensitive. Specificity for both tests ranges between 85% and 90%; false positive results may be observed in the context of medications such as sympathomimetics, paracetamol, selective serotonin reuptake inhibitors, and methyldopa.

In pregnant women, due to concerns around radiation exposure, the radiological investigation of choice for tumour localization should be T_2-weighted magnetic resonance imaging (MRI) with gadolinium enhancement. Ultrasonography may be useful for localization of larger tumours, but has poor sensitivity for small lesions. Scintigraphy with ^{123}I-metaiodobenzylguanidine (MIBG) is contraindicated in pregnancy [41].

Management

Pharmacological α-adrenoceptor blockade is the cornerstone of preoperative phaeochromocytoma management in pregnant and non-pregnant patients, and is associated with significant reductions in fetal and maternal mortality compared to untreated patients [42]. There are no studies in pregnant women that can help to inform on the optimal duration of α-blockade before surgery. Endocrine Society guidelines recommend treatment for a minimum of 10 days before surgery in non-pregnant patients [43]. Preoperative α-blockade with either phenoxybenzamine or doxazosin appears safe in pregnancy. Phenoxybenzamine can cross the placenta, and therefore poses a theoretical risk to fetal cardiovascular stability [44]. Phenoxybenzamine is a non-competitive α_1- and α_2-adrenoceptor antagonist, with a starting

dose of 10 mg twice daily, increasing upwards to 1 mg/kg/day until blood pressure is controlled. There are no studies that inform on the optimal blood pressure target in either pregnant or non-pregnant patients. Phenoxybenzamine has a long duration of action due to the formation of a permanent covalent bond with α-adrenoceptors, which increases the likelihood of significant intra- or postoperative hypotension after tumour removal. Doxazosin is a selective α_1-adrenoceptor antagonist with a significantly shorter half-life, which results in significantly less reflex tachycardia and a shorter duration of postoperative hypotension. There are no head-to-head randomized studies comparing the efficacy of phenoxybenzamine with doxazosin in either pregnant or non-pregnant patients.

ß-blockade is contraindicated in the absence of haemodynamically effective α_1-receptor antagonism. This is due to the theoretical risk of unopposed α-adrenoceptor stimulation and consequent hypertensive crisis induced by tumoral catecholamines after ß-adrenoceptor blockade. Where appropriate, and in the context of pretreatment with α-blockade, ß-blockers are effective for treatment of catecholamine-induced arrhythmias and reflex tachycardia. Propranolol and the combined α- and ß-blocker labetalol may be used safely in pregnancy. For emergency perioperative management in untreated pregnant and non-pregnant patients with phaeochromocytoma, intravenous phenoxybenzamine, or phentolamine may be used.

Surgical resection is the only definitive treatment for phaeochromocytoma. The optimal timing of adrenalectomy in a patient diagnosed during pregnancy is the second trimester, before 24 weeks [35]; this reduces the risk of miscarriage associated with abdominal surgery in the first trimester, and avoids the anatomical complications associated with an enlarged uterus in the third trimester. In tumours diagnosed after 24 weeks, surgery should be delayed until after pregnancy where possible, and a prolonged period of treatment with α-blockade can be safely instituted until the fetus is viable. Laparoscopic surgery in a centre of expertise should be the gold standard intervention, although open surgery may be required in larger tumours. Elective caesarean section should be offered as the preferred mode of delivery in patients who have not undergone adrenalectomy. Although large data are lacking, a number of small previous studies identified increased maternal mortality with vaginal delivery [34]. Labour itself may result in significant catecholamine excursion due to pain and uterine contractions, and should be avoided in patients with tumours in situ.

Primary Aldosteronism

Fewer than 50 cases of primary aldosteronism (PA) in pregnancy have been reported in the literature since the 1960s [45]. While hypertension affects 6–8% of pregnant women, and the prevalence of PA in young hypertensive patients is estimated at up to 10%, published data do not currently support the inference that 0.6–0.8% of all pregnant women harbour an underlying diagnosis of PA. However, the likelihood is that a large proportion of cases of PA remain undiagnosed prior to and during pregnancy. The majority of published cases of PA in pregnancy are due to unilateral aldosterone-producing adenomas (APAs, or Conn's syndrome) rather than idiopathic bilateral adrenal hyperplasia, but prevalence data in large cohorts are lacking [46]. Furthermore, 65 pregnancies have been reported in women with familial hyperaldosteronism type 1 [47].

Patient cohorts with glucocorticoid-remediable aldosteronism (GRA) have provided further insights into the consequences of PA in pregnancy. In this condition, fusion of the 5' regulatory region of adrenal CYP11B1, which is a critical regulator of glucocorticoid production, with the coding region of CYP11B2 (aldosterone synthase) results in expression and activity of the latter falling under the control of pituitary adrenocorticotropic hormone (ACTH). The consequence is a subtype of PA with hypokalaemic hypertension, which is treatable with low-dose dexamethasone. In a case series of 29 pregnancies in a total of 8 women with GRA, one study observed deterioration of hypertension in 39%, with 23% requiring the addition of antihypertensive medication in tandem with conventional dexamethasone treatment [48].

Diagnosis

The renin-angiotensin-aldosterone system (RAAS) is upregulated in pregnancy. Plasma renin levels increase significantly from the first trimester onwards, due to direct extra-renal oestrogen-stimulated renin secretion in the ovaries and decidua. Plasma aldosterone levels are increased 3–8-fold during gestation, which is largely a compensatory response due to competitive inhibition of the mineralocorticoid receptor (MR) by progesterone in the renal tubular apparatus. Hepatic angiotensinogen production is also increased in pregnancy, resulting in elevations of both angiotensin II and aldosterone. These physiological biochemical changes mean that securing the diagnosis of PA in pregnancy is challenging.

PA in pregnancy should be considered in women with onset of hypertension before the 20th gestational week. Intuitively, the coexistence of hypokalaemia in pregnancy may increase the pretest probability of PA. A gestational biochemical diagnosis of PA is dependent on the presence of a low or suppressed plasma renin on repeated samples, as this should be physiologically elevated in healthy pregnant women. The aldosterone-to-renin ratio (ARR) should be theoretically normal or low in normal pregnancy, and elevated in those women with PA. However, caution must be exercised interpreting the ARR, and this parameter must be assessed in tandem with a suppressed renin. Although the Endocrine Society advises confirmatory testing in non-pregnant patients with saline suppression testing or a captopril challenge [49], these investigations are relatively contraindicated in pregnancy due to concerns about plasma expansion and teratogenicity, respectively. Imaging strategies for lateralization are limited to MRI or ultrasonography due to the risk of excessive radiation exposure with CT. Excessive radiation exposure associated with adrenal vein sampling (AVS), which is the gold standard technique for lateralization in non-pregnant patients, means that this technique is contraindicated in pregnancy. However, on the proviso that blood pressure is optimized and serum potassium levels are stable, investigations to lateralize a source of aldosterone hypersecretion, and, where appropriate, adrenalectomy *per se*, can often be deferred until after delivery.

Treatment

Therapeutic management of PA in pregnancy applies to women with a new diagnosis of PA during their pregnancy, as well as to

those with a pre-existing diagnosis who have either not had an adrenalectomy for unilateral disease, or who are medically treated for bilateral disease. In the majority of situations, intensive blood pressure control and tailored oral potassium replacement are the cornerstone of treatment for PA in pregnancy. MR antagonists are the drugs of choice for medical management of non-pregnant patients with PA. However, spironolactone has significant antiandrogenic effects, and is classified as FDA pregnancy category C. Spironolactone crosses the placenta [50], and therefore is not recommended for use in pregnancy due to concerns about undervirilization of the male fetus. Data to support this are not entirely consistent. Feminization of male fetuses has been reported in studies of female rats treated with high doses of spironolactone [51], but this has not been replicated in mice or rabbits. In humans, spironolactone therapy at doses below 400 mg/day did not result in male fetus feminization in six pregnancies from three patients with Bartter syndrome, or in two pregnancies from two patients with Gitelman syndrome. However, there is currently insufficient data to support safe use of spironolactone in pregnancy. Eplerenone does not have antiandrogenic action, and is entirely selective for the MR. It is classified as pregnancy category B by the FDA, but once again controlled safety studies in humans are lacking.

In women with a pre-existing diagnosis of PA, pregnancy should ideally not be contemplated until lateralization has been assessed by AVS. Those with unilateral disease should undergo adrenalectomy before conception in order to normalize blood pressure and reduce fetal and maternal risks. Women with bilateral adrenal hyperplasia should have optimization of their blood pressure before trying to conceive, at which point substitution of MR antagonists with antihypertensives known to be safe during pregnancy can be considered after discussion with the patient [45]. In those with poorly controlled hypertension or resistant hypokalaemia, conception should be discouraged until optimization has been achieved. Where this is not possible, or where the patient is unwilling to wait to conceive, complete discontinuation of MR antagonists is not advisable; eplerenone should be substituted for spironolactone. Similarly, eplerenone can be considered for use in pregnant women with pre-existing PA, or in women with a new pregnancy diagnosis of PA, if blood pressure cannot be controlled with conventional antihypertensive medications known to be safe in pregnancy, or if normokalaemia cannot be maintained. Laparoscopic adrenalectomy in the second trimester for patients with unilateral disease is a feasible option for patients with refractory hypertension and hypokalaemia [52]. A number of cases of laparoscopic adrenalectomy between 14- and 24-weeks' gestation for unilateral aldosterone-producing adenoma have been reported in the literature, with adverse fetal outcomes reported in four out of nine cases [46].

Adrenal Cushing's Syndrome

Cushing's syndrome has a predilection for women of childbearing age; however, it is a rare disorder and fertility is often impaired by hypercortisolism-induced hypogonadotropic hypogonadism and impaired folliculogenesis [28]. As of 2017, 263 pregnancies had been reported in 220 women with active or cured Cushing's syndrome [53]. Maternal complications from retrospective data include hypertension (41–68%), dysglycaemia (25–37%), pre-eclampsia (14–26%), osteoporosis (5%), psychiatric illness (4%), cardiac failure (3%), wound infection (2%), and maternal death (1–2%). Fetal complications include prematurity (43–66%), intrauterine growth restriction (15–21%), miscarriage/stillbirth (11–24%), and fetal mortality (up to 20%) [28, 31, 53]. Neonatal complications include hypoglycaemia, respiratory distress, jaundice, sepsis, intraventricular haemorrhage, and congenital malformations including cleft lip [31, 53]. It may be that milder degrees of hypercortisolism are less harmful to the fetus than expected due to placental 11βHSD2 inactivation of cortisol, perhaps explaining the low rates of hypocortisolism in neonates [28]. Long-term health outcomes in infants born to mothers with Cushing's syndrome are unclear [54]. Onset in pregnancy predicts worse outcomes [53], possibly because women with Cushing's syndrome prior to pregnancy have milder hypercortisolism that permits ovulation.

Cushing's syndrome in pregnancy has several features of special note. Firstly, clinical diagnosis of hypercortisolism is fraught due to pregnancy effects such as centripetal weight gain, striae, oedema, and fatigue as well as pregnancy complications such as diabetes mellitus and hypertension [20, 28]. Catabolic features of Cushing's such as proximal myopathy and fractures, often involving the feet, are more discriminatory [20]. Secondly, biochemical diagnosis of Cushing's must take into account the hypercortisolism of pregnancy [27]. Thirdly, approximately 50% of Cushing's in pregnancy is of primary adrenal cause, usually an adrenal adenoma, compared to 15% of non-pregnant cases and the attendant low ACTH levels are usually absent in pregnancy. Finally, treatment must take into account the special risks of surgery and medications in pregnancy [28].

Aetiologically, Cushing's in pregnancy is most commonly due to an adrenal (40–60%) or pituitary adenoma (33%). Rare causes include adrenocortical carcinoma where there may be evidence of prominent bi- or multihormonal adrenal hormone secretion as well as characteristic findings on imaging and perhaps worse prognosis than in non-pregnant individuals [54]. Bilateral macronodular adrenal hyperplasia is an exceptional cause of adrenal Cushing's in pregnancy as it may very rarely arise due to LH/HCG-stimulation of presumed aberrant receptors that may also be activated postmenopausally [53]. Ectopic ACTH secretion from neuroendocrine tumours is exceedingly rare as the hypercortisolism is often severe, inhibiting pregnancy.

Pregnancy-induced alterations in the HPA axis include increased CBG due to placental oestradiol, in turn elevating plasma total cortisol [54]. By mass spectrometry, plasma total cortisol rises 1.6-, 2.4- and 2.9-fold above non-pregnant controls during the first, second and third trimesters, respectively, commensurate with rising CBG [27]. Later in pregnancy, free cortisol rises due to placental CRH and ACTH [27, 28]. There is evidence suggesting that placental CRH and ACTH are stimulated by cortisol, resulting in a positive feedback loop, and the maternal adrenals have heightened sensitivity to ACTH [27, 28]. Compared to controls, 24-hour urinary free cortisol excretion is 1.7-, 2.4- and 3.1-fold higher during the first, second and third trimesters, respectively [27].

To avoid false-positive results for Cushing's syndrome due to increased CBG, HPA activation, and blunting of normal feedback loops and diurnal variation, urinary free cortisol and late-night salivary cortisol should be interpreted with trimester-specific

normative data [28, 55]. The primary screening test for Cushing's syndrome in pregnancy is urinary free cortisol [27]. Normative data should be derived from the same platform as urinary results by immunoassay are 50–100% higher than mass spectrometry due to cortisol metabolite cross-reactivity [27]. Though urinary free cortisol may overlap between pregnant women with and without Cushing's syndrome, the mean elevation in urinary free cortisol in affected women is higher, at 8-fold normal [31]. Diurnal variation persists in normal pregnancies, albeit with a higher nadir, but it is abolished in pregnancies complicated by Cushing's syndrome, making late night salivary cortisol an adjunct test [31]. Though late night salivary cortisol is highly assay dependent and there was previously insufficient evidence to guide decision-making thresholds [55], recent data suggest late night salivary cortisol normal ranges of <6.9 nmol/L, <7.2 nmol/L and <9.1 nmol/L in the first, second and third trimesters, respectively, to be 80–92% specific for Cushing's syndrome [56]. Dexamethasone suppression of cortisol secretion is blunted in pregnancy with post-DEX cortisol levels >50 nmol/L in over 60% of healthy pregnant women [54], hence the low-dose test is not useful for Cushing's diagnosis in pregnancy [31]. Moreover, this test usually relies on plasma total cortisol, which is misleadingly elevated by CBG in pregnancy [31].

Upon biochemical confirmation of hypercortisolism, plasma ACTH should be measured, noting that the physiological ACTH elevation of pregnancy may hinder identification of ACTH independence with non-suppressed ACTH levels in 50% of pregnant women with adrenal Cushing's [31]. Whilst the high-dose dexamethasone suppression test is typically used to distinguish pituitary versus ectopic causes of ACTH-dependent Cushing's syndrome, it may be used in pregnancy to differentiate adrenal from pituitary Cushing's as pregnant women with Cushing's disease typically suppress whereas pregnant women with ACTH-independent Cushing's syndrome typically show <80% cortisol suppression [31]. Hypercortisolaemic women with low or normal plasma ACTH, and non-suppressible cortisol following high-dose dexamethasone, should be investigated for adrenal Cushing's syndrome [31]. Adrenal ultrasound is safe but only 73% sensitive for adrenal lesions and non-contrast MRI may be required [31]. MRI appears safe in pregnancy but because of potential teratogenicity its timing should be considered [28]. Gadolinium is avoided as it crosses the placenta and pregnancy safety data are limited [31]. Pregnancy-related adrenal hypertrophy and background rates of adrenal incidentalomas should be taken into account [20]. Pregnant women with Cushing's syndrome and high plasma ACTH should be further investigated for ACTH-dependent causes [31], as discussed in other chapters.

Pregnant women with Cushing's syndrome may be treated medically or surgically, or conservatively managed with treatment of complications such as diabetes and hypertension [20]. There are no prospective studies to guide management but there are trends towards higher live birth rates following cortisol-lowering treatment compared to conservative management [31, 53]. Women with relatively mild hypercortisolism in late gestation may nonetheless be better suited to conservative management [54]. Surgery is preferred to medical therapy in women with an identifiable, irreversible and resectable cause of Cushing's syndrome who are in their second trimester of pregnancy, when maternal and fetal surgical complication rates are lower [20, 31]. Unilateral or bilateral adrenalectomy for adrenal Cushing's yields high cure rates and live birth rates of 87% [31]. Though some data show excess prematurity rates after third trimester surgery, laparoscopic adrenalectomy has been performed successfully and safely up to 32 weeks with no change in rates of prematurity or intrauterine growth restriction [54]. In the limited data on adrenocortical carcinoma in pregnancy, resection is often incomplete and surgical risks are greater [54]. Anaesthetic risks pertaining to pregnancy include aspiration pneumonitis due to decreased lower oesophageal sphincter activity and increased intra-abdominal pressure from the gravid uterus, and challenging airway management because of oropharyngeal oedema. No medication has been approved for the treatment of Cushing's syndrome in pregnancy [55], but several agents have been utilized. Metyrapone use in pregnancy is well described in case reports and series; it is often well tolerated and effective [53, 57]. However, metyrapone risks hypertension and pre-eclampsia due to accumulating 11-deoxycorticosterone with mineralocorticoid effects [31]. Metyrapone also can be transferred across the placenta and fetal hypocortisolism has been reported [28, 54]. Ketoconazole crosses the placenta and is teratogenic and abortifacient in rat studies but it has occasionally been used in pregnancy without adverse events [31, 54]. In addition, ketoconazole has the potential to feminize the male fetus due to inhibition of the adrenal 17,20 lyase enzyme [20, 28]. Cabergoline has been used in pregnancy to treat Cushing's disease [20]. Contraindicated approaches in pregnancy include radiotherapy and mitotane due to teratogenicity, and mifepristone as it is a direct abortifacient from progesterone receptor antagonism [20]. Spontaneous remission of Cushing's syndrome has been reported in pregnancy [31]. Remission of Cushing's syndrome is associated with trends towards improved obstetric outcomes with reduced fetal loss, preterm birth, and low birth weight [53]. Remission may lead to hypocortisolism, requiring hydrocortisone and sometimes fludrocortisone therapy with adjustments during pregnancy [54].

Congenital Adrenal Hyperplasia

More than 95% of congenital adrenal hyperplasia (CAH) is due to 21-hydroxylase deficiency, thus studies in pregnancy focus on this aetiology [58]. 21-hydroxylase deficiency is due to mutations of both *CYP21A2* alleles. Such mutations are common (1–10% population, with ethnic variation), but CAH incidence is approximately 1/10 000 births [59]. Genotype/phenotype variation relates, to an extent, to the more functional *CYP21A2* allele [59]. The phenotype involves: PAI due to impaired cortisol and aldosterone synthesis; hyperandrogenism due to accumulation of androstenedione, then converted by 17-ketosteroid reductase to testosterone (or via direct multistep conversion of 17OH progesterone to testosterone), leading to masculinization of the female infant genitalia (commonly clitoromegaly, labial fusion), then later hirsutism and in severe cases virilism; precocious puberty; adrenal enlargement and hyperpigmentation from excess ACTH; and ultimately short stature in males. Classical CAH (childhood features) can be categorized into simple (isolated) virilizing (25%) and more severe salt-wasting (75%) forms. Females may be detected at birth from genital inspection but males may not present for several days until an adrenal crisis develops, which may be fatal. Treatment involves glucocorticoid and mineralocorticoid

replacement, with careful dosing and timing to reduce ACTH and therefore adrenal androgen production. Non-classical CAH refers to adult-onset disease, generally milder than classical CAH, where a decision to pursue adrenal replacement/glucocorticoid suppression treatment should take into account the individualized risk and benefits.

Women with CAH face potential challenges in reproduction [60, 61]. Lower fecundity in CAH may result from a range of reasons including: possible androgenization of the brain affecting psychosexual development with reduced partnering; difficulties or perceived limitations arising from vulvovaginal function despite surgery, usually performed in childhood; anovulation due to androgen excess or exogenous Cushing's, with polycystic ovaries in some cases; ovarian adrenal rest tissue; and progesterone excess increasing cervical mucus viscosity. The greatest impediment to fecundity may be intercourse frequency—only 56% of women with simple virilizing CAH and 11% with salt-losing CAH ever attempt to conceive [62]. Women receiving adequate adrenal hormone treatment and having regular intercourse have pregnancy rates >90% [62], although miscarriage rates are higher than the general population (23% vs. 10.9%, P <0.001) [60]. In a study of 62 CAH women and 62 age-matched controls, CAH women were less likely to have been pregnant and they had fewer total pregnancies and fewer term pregnancies. Reproductive outcomes were worse in salt-wasting CAH than simple virilizing or non-classical CAH. Some women conceived spontaneously while others required increased glucocorticoids, concomitant fludrocortisone or assisted reproductive technologies. CAH pregnancies were characterized by more frequent Caesarean section, mostly due to past genital surgery, and gestational diabetes [61]. Contrary to earlier data, low birth weight was no more common in CAH, and there was no difference in birth weight between CAH subtypes. Postnatal development appeared normal on long-term follow-up [61]. A predominance of female offspring has been observed, which may reflect referral bias as girls face the risk of virilization but uterine changes favouring implantation of female embryos have been proposed [58, 61].

It is often necessary to increase glucocorticoid treatment prepregnancy to better suppress androgens and maximize ovulation frequency. Increasing glucocorticoid dose to achieve follicular phase progesterone concentrations <2 nmol/L is recommended [63]. Once pregnant, women with classical CAH and non-classical cases receiving adrenal hormone replacement are managed along similar lines to women with PAI, including personalized glucocorticoid dosing, increasing in the latter half of gestation [63]. Testosterone levels may be followed in pregnancy but not 17-hydroxyprogestrone as levels are elevated normally in pregnancy [59]. Delivery is generally by Caesarean section due to vaginal stenosis. Maternal androgen excess does not affect the female fetus due to placental aromatization.

Attempts to prevent masculinization of CAH female children have included trials of prenatal dexamethasone, commenced in early gestation prior to potential diagnosis of the fetus as a female *CYP21A2* mutation homozygote, with some, but incomplete benefit [58, 59]. This strategy exposes unaffected female and male fetuses to adrenal suppression and excess glucocorticoids with uncertain long-term consequences, and subjects pregnant women to glucocorticoid excess side effects. This approach is experimental only. Prevention may also be pursued through preimplantation or prenatal genetic diagnosis, in cases where the father has a mutant *CYP21A2* allele.

Miscellaneous Adrenal Disorders in Pregnancy

Adrenal Adenomas and Incidentalomas

The prevalence of adrenal adenomas in the background population is between 3% and 5%, increasing to 10% above the age of 70 years [64]. Detection of adrenal incidentalomas in pregnancy is rare, almost certainly due to the limited number of abdominal scans performed in the gestational period outside of obstetric ultrasound. The objectives of investigations of an incidental adrenal lesion diagnosed in pregnancy are to differentiate benign from malignant lesions, and to exclude adrenal hormone hypersecretion. In general, current imaging strategies are sensitive but poorly specific for malignancy in an adrenal mass [65], and several modalities such as FDG-PET and CT are relatively contraindicated in pregnancy. Radiological observation for stability of any detected lesion, such as MRI or ultrasound every 3 months, may be the most sensible management strategy for lesions not definitively proven to be benign on initial imaging [66]. Biochemical investigations for hormone excess should in general mirror those performed in the non-pregnant patient [67], with the caveat that the physiological perturbations in adrenal hormone metabolism described earlier may confound interpretation of results. Dexamethasone suppression testing for adrenal hypercortisolism should be avoided due to the confounding effect of increased oestrogen levels on hepatic CBG levels. Lesions with a biochemical signature suspicious for adrenocortical carcinoma may warrant closer observation or more definitive intervention such as adrenalectomy in the second trimester or the early postpartum period.

Adrenocortical Carcinoma

Adrenocortical carcinoma (ACC) is a rare malignancy, with an annual incidence of approximately one per million. Incidence is, however, higher in women than men, and is frequently diagnosed in women of reproductive age. Detection of ACC in pregnancy is extremely rare, and associated with a poor prognosis. Raffin-Sanson *et al.* reported outcomes in 12 female patients who were diagnosed either during pregnancy or in the first 6 months after delivery, and compared them to age-matched non-pregnant females [68]. Overall survival was worse, with a more advanced tumour stage in the pregnancy group. Unsurprisingly fetal outcomes were also poor. Some studies have speculated that ACC in pregnancy may display more aggressive properties due to oestrogen stimulation, but there are no conclusive data to support this. The clinical suspicion of ACC should be high in any patient with androgen excess associated with an adrenal mass, by the coexistence of cortisol and adrenal androgen hypersecretion, or in any adrenal mass greater than 4cm in transverse diameter.

Despite a recent report of normal pregnancy in a patient with metastatic ACC taking adjuvant mitotane up until the sixth gestational week, the latter is not recommended in pregnancy due to concerns about placental transfer [69]. In 17 women previously treated for ACC before conceiving, pregnancy did appear to be associated with worse clinical outcomes, and overall survival was not significantly different between the pregnancy group and matched

controls [70]. However, the authors caution about the potential bias introduced by a 'healthy mother effect', and recommend that women previously treated for ACC wait for a remission period of 5 years before trying to conceive.

REFERENCES

1. Lekarev O, New MI. Adrenal disease in pregnancy. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 959–73.
2. Monticone S, Auchus RJ, Rainey WE. Adrenal disorders in pregnancy. *Nat Rev Endocrinol*, 2012; **8**: 668–78.
3. Bergthorsdottir R, Leonsson-Zachrisson M, Oden A, Johannsson G. Premature mortality in patients with Addison's disease: a population-based study. *J Clin Endocrinol Metab*, 2006; **91**: 4849–53.
4. Bjornsdottir S, Sundstrom A, Ludvigsson JF, Blomqvist P, Kampe O, Bensing S. Drug prescription patterns in patients with Addison's disease: a Swedish population-based cohort study. *J Clin Endocrinol Metab*, 2013; **98**: 2009–18.
5. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. *Clin Immunol Immunopathol*, 1997; **84**: 223–43.
6. Laureti S, Vecchi L, Santeusanio F, Falorni A. Is the prevalence of Addison's disease underestimated? *J Clin Endocrinol Metab*, 1999; **84**: 1762.
7. Lovas K, Husebye ES. High prevalence and increasing incidence of Addison's disease in western Norway. *Clin Endocrinol (Oxf)*, 2002; **56**: 787–91.
8. Mason AS, Meade TW, Lee JA, Morris JN. Epidemiological and clinical picture of Addison's disease. *Lancet*, 1968; **2**: 744–7.
9. Willis AC, Vince FP. The prevalence of Addison's disease in Coventry, UK. *Postgrad Med J*, 1997; **73**: 286–8.
10. Rushworth RL, Torpy DJ, Falhammar H. Adrenal crises: perspectives and research directions. *Endocrine*, 2017; **55**: 336–45.
11. Bornstein SR, Allolio B, Arlt W, *et al*. Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 364–89.
12. Di Dalmazi G, Vicennati V, Rinaldi E, *et al*. Progressively increased patterns of subclinical cortisol hypersecretion in adrenal incidentalomas differently predict major metabolic and cardiovascular outcomes: a large cross-sectional study. *Eur J Endocrinol*, 2012; **166**: 669–77.
13. Esteban NV, Loughlin T, Yergey AL, *et al*. Daily cortisol production rate in man determined by stable isotope dilution/mass spectrometry. *J Clin Endocrinol Metab*, 1991; **72**: 39–45.
14. Rushworth RL, Torpy DJ. Adrenal Insufficiency in Australia: is it possible that the use of lower dose, short-acting glucocorticoids has increased the risk of adrenal crises? *Horm Metab Res*, 2015; **47**: 427–32.
15. Rushworth RL, Torpy DJ. Modern hydrocortisone replacement regimens in adrenal insufficiency patients and the risk of adrenal crisis. *Horm Metab Res*, 2015; **47**: 637–42.
16. Allolio B. Extensive expertise in endocrinology. Adrenal crisis. *Eur J Endocrinol*, 2015; **172**: R115–24.
17. Bensing S, Brandt L, Tabaroj F, *et al*. Increased death risk and altered cancer incidence pattern in patients with isolated or combined autoimmune primary adrenocortical insufficiency. *Clin Endocrinol (Oxf)*, 2008; **69**: 697–704.
18. Erichsen MM, Lovas K, Fougner KJ, *et al*. Normal overall mortality rate in Addison's disease, but young patients are at risk of premature death. *Eur J Endocrinol*, 2009; **160**: 233–7.
19. Falhammar H, Frisen L, Norrby C, *et al*. Increased mortality in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab*, 2014; **99**: E2715–21.
20. Bronstein MD, Machado MC, Fragoso MC. Management of endocrine disease: management of pregnant patients with Cushing's syndrome. *Eur J Endocrinol*, 2015; **173**: R85–91.
21. Ho W, Druce M. Quality of life in patients with adrenal disease: a systematic review. *Clin Endocrinol (Oxf)*, 2018; **89**: 119–28.
22. Torpy DJ. Glucocorticoid replacement in chronic adrenal insufficiency: evaluating the evidence for optimal therapy. *Endocr Pract*, 2017; **23**: 113–15.
23. Ambrosi B, Barbetta L, Morricone L. Diagnosis and management of Addison's disease during pregnancy. *J Endocrinol Invest*, 2003; **26**: 698–702.
24. Lebbe M, Arlt W. What is the best diagnostic and therapeutic management strategy for an Addison patient during pregnancy? *Clin Endocrinol (Oxf)*, 2013; **78**: 497–502.
25. Bjornsdottir S, Cnattingius S, Brandt L, *et al*. Addison's disease in women is a risk factor for an adverse pregnancy outcome. *J Clin Endocrinol Metab*, 2010; **95**: 5249–57.
26. Ho JT, Lewis JG, O'Loughlin P, *et al*. Reduced maternal corticosteroid-binding globulin and cortisol levels in pre-eclampsia and gamete recipient pregnancies. *Clin Endocrinol (Oxf)*, 2007; **66**: 869–77.
27. Jung C, Ho JT, Torpy DJ, *et al*. A longitudinal study of plasma and urinary cortisol in pregnancy and postpartum. *J Clin Endocrinol Metab*, 2011; **96**: 1533–40.
28. Lindsay JR, Nieman LK. The hypothalamic-pituitary-adrenal axis in pregnancy: challenges in disease detection and treatment. *Endocr Rev*, 2005; **26**: 775–99.
29. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. *Lancet (London, England)*, 2014; **383**: 2152–67.
30. Rushworth RL, Chrisp GL, Torpy DJ. Glucocorticoid induced adrenal insufficiency: a study of the incidence in hospital patients and a review of perioperative management. *Endocr Pract*, 2018; **24**: 437–45.
31. Lindsay JR, Jonklaas J, Oldfield EH, Nieman LK. Cushing's syndrome during pregnancy: personal experience and review of the literature. *J Clin Endocrinol Metab*, 2005; **90**: 3077–83.
32. Crona J, Taieb D, Pacak K. New perspectives on pheochromocytoma and paraganglioma: toward a molecular classification. *Endocr Rev*, 2017; **38**: 489–515.
33. Harper MA, Murnaghan GA, Kennedy L, Hadden DR, Atkinson AB. Phaeochromocytoma in pregnancy. Five cases and a review of the literature. *Br J Obstet Gynaecol*, 1989; **96**: 594–606.
34. Schenker JG, Chowers I. Pheochromocytoma and pregnancy. Review of 89 cases. *Obstet Gynecol Surv*, 1971; **26**: 739–47.
35. Ahlawat SK, Jain S, Kumari S, Varma S, Sharma BK. Pheochromocytoma associated with pregnancy: case report and review of the literature. *Obstet Gynecol Surv*, 1999; **54**: 728–37.
36. Natrajan PG, McGarrigle HH, Lawrence DM, Lachelin GC. Plasma noradrenaline and adrenaline levels in normal pregnancy and in pregnancy-induced hypertension. *Br J Obstet Gynaecol*, 1982; **89**: 1041–5.
37. Dahia PL, Hayashida CY, Strunz C, Abelin N, Toledo SP. Low cord blood levels of catecholamine from a newborn of a pheochromocytoma patient. *Eur J Endocrinol*, 1994; **130**: 217–19.

38. Kanasaki K, Palmsten K, Sugimoto H, *et al*. Deficiency in catechol-O-methyltransferase and 2-methoxyoestradiol is associated with pre-eclampsia. *Nature*, 2008; **453**: 1117–21.
39. Oishi S, Sato T. Pheochromocytoma in pregnancy: a review of the Japanese literature. *Endocr J*, 1994; **41**: 219–25.
40. Lenders JW, Pacak K, Walther MM, *et al*. Biochemical diagnosis of pheochromocytoma: which test is best? *JAMA*, 2002; **287**: 1427–34.
41. Freier DT, Thompson NW. Pheochromocytoma and pregnancy: the epitome of high risk. *Surgery*, 1993; **114**: 1148–52.
42. Burgess GE, 3rd. Alpha blockade and surgical intervention of pheochromocytoma in pregnancy. *Obstet Gynecol*, 1979; **53**: 266–70.
43. Lenders JW, Duh QY, Eisenhofer G, *et al*. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2014; **99**: 1915–42.
44. Santeiro ML, Stromquist C, Wyble L. Phenoxybenzamine placental transfer during the third trimester. *Ann Pharmacother*, 1996; **30**: 1249–51.
45. Riester A, Reincke M. Progress in primary aldosteronism: mineralocorticoid receptor antagonists and management of primary aldosteronism in pregnancy. *Eur J Endocrinol*, 2015; **172**: R23–30.
46. Morton A. Primary aldosteronism and pregnancy. *Pregnancy Hypertens*, 2015; **5**: 259–62.
47. Campino C, Trejo P, Carvajal CA, *et al*. Pregnancy normalized familial hyperaldosteronism type I: a novel role for progesterone? *J Hum Hypertens*, 2015; **29**: 138–9.
48. Mulatero P, di Cella SM, Williams TA, *et al*. Glucocorticoid remediable aldosteronism: low morbidity and mortality in a four-generation italian pedigree. *J Clin Endocrinol Metab*, 2002; **87**: 3187–91.
49. Funder JW, Carey RM, Mantero F, *et al*. The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 1889–916.
50. Kamoun M, Mnif MF, Charfi N, *et al*. Adrenal diseases during pregnancy: pathophysiology, diagnosis and management strategies. *Am J Med Sci*, 2014; **347**: 64–73.
51. Hecker A, Hasan SH, Neumann F. Disturbances in sexual differentiation of rat foetuses following spironolactone treatment. *Acta Endocrinol*, 1980; **95**: 540–5.
52. Kosaka K, Onoda N, Ishikawa T, *et al*. Laparoscopic adrenalectomy on a patient with primary aldosteronism during pregnancy. *Endocr J*, 2006; **53**: 461–6.
53. Caimari F, Valassi E, Garbayo P, *et al*. Cushing's syndrome and pregnancy outcomes: a systematic review of published cases. *Endocrine*, 2017; **55**: 555–63.
54. Brue T, Amodru V, Castinetti F. Management of endocrine disease: management of Cushing's syndrome during pregnancy: solved and unsolved questions. *Eur J Endocrinol*, 2018; **178**: R259–66.
55. Nieman LK, Biller BM, Findling JW, *et al*. The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2008; **93**: 1526–40.
56. Lopes LM, Francisco RP, Galletta MA, Bronstein MD. Determination of nighttime salivary cortisol during pregnancy: comparison with values in non-pregnancy and Cushing's disease. *Pituitary*, 2016; **19**: 30–8.
57. Lim WH, Torpy DJ, Jeffries WS. The medical management of Cushing's syndrome during pregnancy. *Eur J Obstet Gynecol Reprod Biol*, 2013; **168**: 1–6.
58. New MI, Carlson A, Obeid J, *et al*. Prenatal diagnosis for congenital adrenal hyperplasia in 532 pregnancies. *J Clin Endocrinol Metab*, 2001; **86**: 5651–7.
59. Speiser PW, Azziz R, Baskin LS, *et al*. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2010; **95**: 4133–60.
60. Eyal O, Ayalon-Dangur I, Segev–Becker A, Schachter-Davidov A, Israel S, Weintrob N. Pregnancy in women with nonclassic congenital adrenal hyperplasia: time to conceive and outcome. *Clin Endocrinol (Oxf)*, 2017; **87**: 552–6.
61. Hagenfeldt K, Janson PO, Holmdahl G, *et al*. Fertility and pregnancy outcome in women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Hum Reprod*, 2008; **23**: 1607–13.
62. Casteras A, De Silva P, Rumsby G, Conway GS. Reassessing fecundity in women with classical congenital adrenal hyperplasia (CAH): normal pregnancy rate but reduced fertility rate. *Clin Endocrinol (Oxf)*, 2009; **70**: 833–7.
63. Auchus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*, 2013; **98**: 2645–55.
64. Barzon L, Sonino N, Fallo F, Palu G, Boscaro M. Prevalence and natural history of adrenal incidentalomas. *Eur J Endocrinol*, 2003; **149**: 273–85.
65. Dinnes J, Bancos I, Ferrante di Ruffano L, *et al*. Management of endocrine disease: imaging for the diagnosis of malignancy in incidentally discovered adrenal masses: a systematic review and meta-analysis. *Eur J Endocrinol*, 2016; **175**: R51–64.
66. Fallo F, Pezzi V, Sonino N, Altavilla G, Barzon L. Adrenal incidentaloma in pregnancy: clinical, molecular and immunohistochemical findings. *J Endocrinol Invest*, 2005; **28**: 459–63.
67. Fassnacht M, Arlt W, Bancos I, *et al*. Management of adrenal incidentalomas: European Society of Endocrinology clinical practice guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2016; **175**: G1–34.
68. Abiven-Lepage G, Coste J, Tissier F, *et al*. Adrenocortical carcinoma and pregnancy: clinical and biological features and prognosis. *Eur J Endocrinol*, 2010; **163**: 793–800.
69. Tripto-Shkolnik L, Blumenfeld Z, Bronshtein M, Salmon A, Jaffe A. Pregnancy in a patient with adrenal carcinoma treated with mitotane: a case report and review of literature. *J Clin Endocrinol Metab*, 2013; **98**: 443–7.
70. de Corbiere P, Ritzel K, Cazabat L, *et al*. Pregnancy in women previously treated for an adrenocortical carcinoma. *J Clin Endocrinol Metab*, 2015; **100**: 4604–11.

9.9

Endocrine Bone Disease in Pregnancy

Jeremy Cox and Stephen Robinson

Vitamin D and Pregnancy

Introduction

Vitamin D and calcium metabolism is important both in health and disease, within normal and complicated pregnancy. There is an association between 25-hydroxyvitamin D (25-OHD) levels and intrauterine growth restriction, vitamin D is crucial to the fetus for growth and calcium accretion to bone. Vitamin D has effects beyond calcium homeostasis, maternal vitamin D concentrations are associated with increased risk of adverse pregnancy outcome including gestational diabetes, pre-eclampsia, and infections *in utero* and childhood. The concentrations of vitamin D need to be reviewed in terms of their appropriateness to ambient calcium concentrations and the homeostatic mechanisms involved. Therefore the changes that pregnancy brings to calcium physiology are reviewed before discussing the impact of maternal vitamin D status on the fetus, the mother, obstetric outcomes and the long-term health of the mother and child.

Epidemiology of Maternal Vitamin D

Vitamin D insufficiency and deficiency are common in Northern Europe, especially in women who have coloured skin or cover their skin. In Dutch women studied at 12 weeks of gestation the 25(OH)D was 15 ± 12 nmol/L in Turkish women, 20 ± 14 nmol/L in Moroccan women, and 53 ± 22 nmol/L in Western women [1]. There was an association with deprivation as well as reduced exposure to sunlight. In 970 Australian women 15% had a 25(OH)D less than 25nmol/L, whilst 11% of the neonates had a 25(OH)D less than 25nmol/L [2]. Vitamin D deficiency again was related to skin phototype, wearing a veil and young age. In 180 women studied in inner city London (n = 45 each group), the median 25(OH)D was 14 in Arab-Mediterranean, 21 in south Asian, 14 in African, and 48 nmol/L in Caucasian women (25(OH)D<25 nmol/L deficient, 25–75 nmol/L insufficient) [3]. Untreated, these are levels that are associated with a high parathyroid hormone (PTH) level in the mother (Figure 9.9.3).

Vitamin D and Calcium Physiology in Pregnancy

Calcium and bone metabolism in humans rely on adequate concentrations of vitamin D and its activated metabolite $1,25(OH)_2D$. This allows active calcium absorption from the gastrointestinal tract. Deficiency of $1,25(OH)_2D$ leads to secondary hyperparathyroidism in order to maintain plasma calcium concentration at a safe level for neuromuscular function, at the expense of bone mineral loss. The fetus at term will have accreted 30 g of calcium, 99% in the skeleton. Women transfer 300 mg daily to their fetus at term, and 210 mg/day in lactation [4]. Given that intake of calcium ranges from 800 mg to 1000mg in USA/Canada, and less elsewhere, calcium dynamics need to change in pregnancy and lactation to allow this transfer (Figure 9.9.1).

Therefore, during pregnancy adaptations of maternal physiology allow for enhanced uptake of calcium from the gut to support the fetus [4], and during lactation calcium mobilization from the bone to support breastfeeding the neonate. Normal calcium physiology with the changes of pregnancy and lactation are shown in Figure 9.9.2.

The normal range for vitamin D can be defined by hard outcomes, but this is difficult in humans with such long-term outcomes and the evidence base is suboptimal. It can also be defined by adaptive response, e.g. the rise in PTH concentrations that occurs when 25(OH)D falls below 30ng/ml (75nmol/L). In pregnancy there is an increase in $1,25(OH)_2D$ production from the maternal kidney, rather than decreased clearance, with a doubling of concentrations by 12 weeks, although the mechanisms that underlie control of this process are unknown [4]. Maternal PTH concentrations fall in early pregnancy, returning to non-pregnant values at term. Second trimester PTH suppression may not occur with low maternal calcium intake. Maternal parathyroid hormone related peptide (PTHrP) concentrations rise threefold by the third trimester, probably from placenta and breast, and may help maintain ionized calcium and stimulate calcitriol production [4].

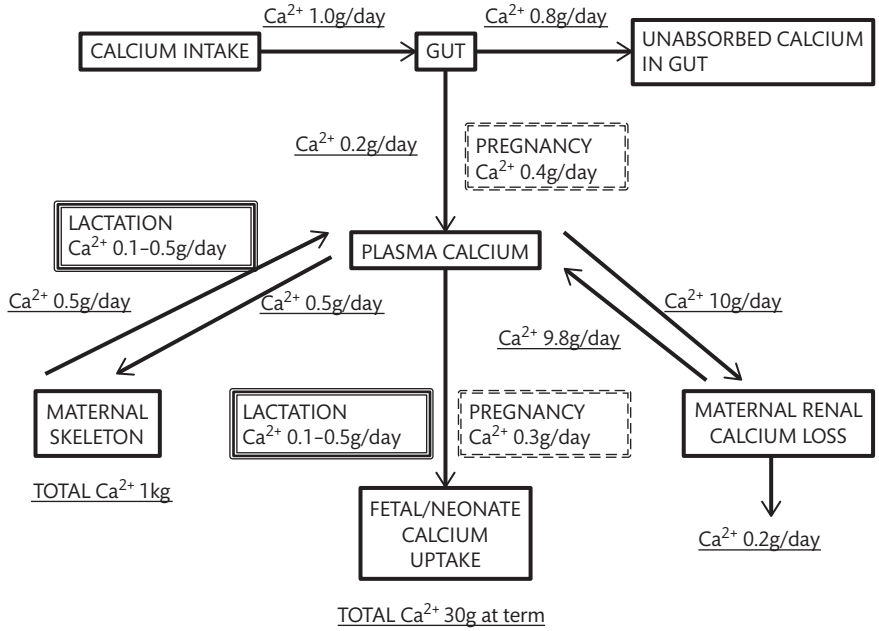


Figure 9.9.1 Normal calcium dynamics with the changes of pregnancy and lactation. Calcium is essential in bone and to neuromuscular function. Humans consume about 1.0 g calcium/day absorbing up to 20% of this. In steady state about 0.2 g calcium are lost in the urine each day. Normally the calcium taken up and released by the skeleton is in balance, but in the event of hypocalcaemia PTH causes increased osteoclast calcium release. During pregnancy higher levels of active vitamin D increase the fractional absorption of calcium from the gut, allowing for fetal calcium uptake. During lactation PTHrP increases osteoclast calcium release to the circulation allowing for the neonatal calcium uptake.

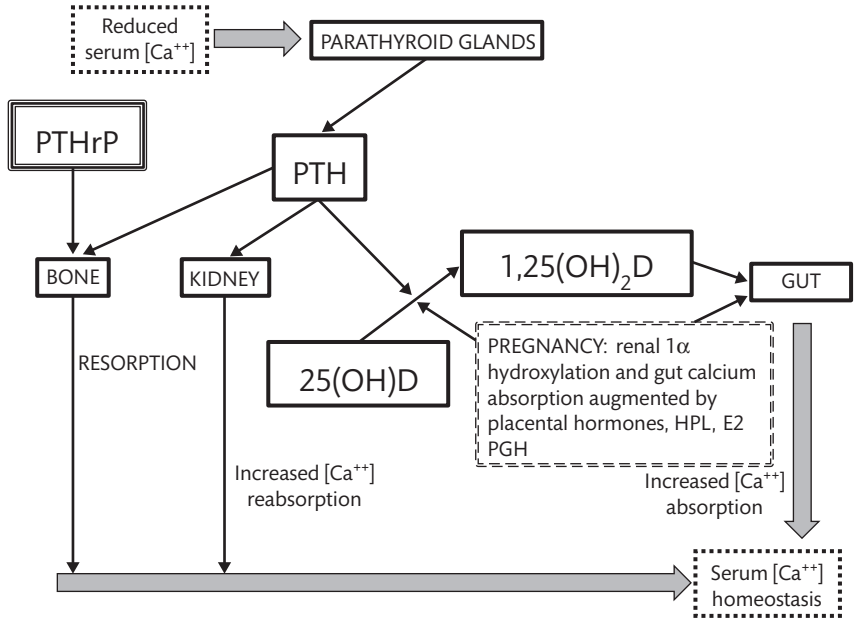


Figure 9.9.2 Normal calcium physiology with the changes of pregnancy and lactation. Normally reduced serum ionized calcium is detected by the calcium sensing receptor causing PTH release. This directly has effects on bone causing osteoclast activity with calcium release. PTH also increases renal calcium reabsorption from filtrate, thereby maintaining plasma calcium. In addition, PTH acts within the proximal tubule of the kidney to increase hydroxylation of 25(OH)D (calcifediol) in the 1 alpha position to produce $1,25(OH)_2D$ (calcitriol). Calcitriol increases fractional calcium absorption from the gut. The fundamental process in maternal physiology that drives this process in pregnancy is increased $1,25(OH)_2D$ action in the gut facilitating a twofold increase in calcium absorption. Placental hormones (HPL human placental lactogen, E2 oestradiol, PGH placental growth hormone) acting in the maternal kidney may increase vitamin D activation and also directly increase gut calcium absorption. When a mother is vitamin D replete and has good calcium intake this process can more than adequately allow for the fetal demands. Indeed in pregnancy the mother herself can drive a positive calcium balance. Parathormone related peptide (PTHrP) produced from the placenta in late pregnancy and then breasts in pregnancy and lactation increases osteoclast resorption and possibly osteocytic osteolysis, freeing calcium from maternal bone for the suckling neonate.

Fetal Outcomes

Birth weight can be defined by absolute weight or that corrected for gestational age; low birth weight and small for gestational age (SGA) have interconnected implications. Two meta analyses have shown an association between low 25(OH)D concentrations and low birth weight; RR 0.40 (95% CI 0.23–0.71) and RR 0.40 (95% CI 0.24–0.67), but one did not, RR 0.72 (95% CI 0.44–1.16) [5]. Possibly differing background risks and vitamin D ranges lead to these discrepancies.

Data regarding vitamin D concentration and small for gestational age (SGA) infants are similarly contradictory. One meta-analysis has shown an association between low 25(OH)D concentrations and SGA, RR 0.67 (95% CI 0.40–1.11) but vitamin D still thought to be protective; another did not, RR 0.78 (95% CI 0.50–1.21) [5]. In a study of nearly 4000 mother and 10-year-old offspring pairs there was no association between maternal vitamin D status and offspring total body or spinal bone mineral content [6].

Low maternal concentrations of 25(OH)D are associated with low neonatal concentrations of 25(OH)D [3] and neonatal hypocalcaemia. In a prospective study of severe deficiency, 59 Asian women were given 1000 IU vitamin D daily and compared to 67 untreated controls [7]. Vitamin D deficiency progressed in those who were untreated compared to an excellent response in mothers and neonates in terms of vitamin concentrations; none of the neonates of treated mothers developed symptomatic hypocalcaemia compared to five neonatal controls. Neonatal calcium in controls was <2.0 mmol/L in 24.6% compared to 6.3% in the treatment group.

Low maternal and neonatal vitamin D levels may be associated with increased childhood respiratory infections, through altered immunomodulation or bronchial development. Maternal intake of vitamin D was associated with reduced risk of recurrent wheeze at 3 years of age, OR 0.39 (95% CI 0.25–0.62) although there were multiple potential confounders [8]. However, in 436 Dutch children there was no association between vitamin D levels and lung function aged 6 years [9]. In a study of 3000 women the risk of later pre-eclampsia was reduced in those supplemented with vitamin D in the first year of life, OR 0.49 (95% CI 0.26–0.92) [10].

Maternal Outcomes

Low maternal 25(OH)D concentrations are associated with pre-eclampsia. Fifty-five singleton mothers with pre-eclampsia had 25(OH)D levels of 45nmol/L compared to 53 nmol/L in 219 controls [11]. Maternal 25(OH)D concentration was also associated with early onset severe pre-eclampsia; 25(OH)D <20 ng/ml was found in 54% of women with pre-eclampsia and 27% of controls [12]. Two trials have shown a lower risk of pre-eclampsia with vitamin D supplementation in pregnancy (8.9% vs. 15.5%; RR 0.52, 95% CI 0.25–1.05) [5]. Three trials have shown a lower risk of pre-eclampsia with vitamin D and calcium supplementation in pregnancy (5% vs. 9%; RR 0.51, 95% CI 0.32–0.80) [5].

When 23 423 women had their vitamin D intake estimated, there appeared to be a protective effect on hypertension of consuming 15–20 μg/d compared to 5 μg/day, OR 0.76 (95% CI 0.60–0.95) [13].

Some studies in the general population have shown an association between vitamin D concentrations and insulin resistance, T2DM and prediction of future T2DM [14]. However, Vitamin D supplementation has no effect on insulin sensitivity, insulin secretion or glycaemic control in subjects with existing type 2 diabetes. Studies have shown an inverse association between vitamin D concentrations and gestational diabetes and studies are investigating the therapeutic possibilities within these observations. Two have shown no benefit from maternal vitamin D supplementation and risk of gestational diabetes [5].

Obstetric Outcomes

Hypovitaminosis D has been investigated as an associate of preterm delivery. One metanalysis showed a clear significant association between 25(OH)D <37.5nmol/L and preterm delivery, RR 0.36 (95% CI 0.14–0.93); however, another did not show this, RR 1.26 (95% CI 0.60–2.63) [5].

Vitamin D deficiency was associated with primary caesarean section (CS) in 17% of 253 women having the procedure. Twenty eight percent had CS when 25(OH)D<37.5nmol/L but 14% when 25(OH)D>37.5nmol/L [15].

Outcome for Vitamin D in Mother

Both pregnancy and especially lactation are states that deprive a woman of vitamin D resources. Vitamin D insufficiency is particularly common in woman with multiple pregnancies or prolonged lactation. Calcium intake appears to have little effect on breast milk calcium output.

Supplement and Treatment Guidance

Vitamin D supplementation during pregnancy is associated with a significant increase in maternal and neonatal 25(OH)D levels (Figure 9.9.3). This in turn reduces the risk of neonatal hypocalcaemia and maternal pre-eclampsia. However, there are no clear data to show that vitamin D supplementation is associated with reduced risk of SGA, preterm delivery, or glucose intolerance. It is not known whether screening pregnant women has a clear utility, nor the method and dosing of a safe supplementation strategy.

In the absence of clear information there are sensible guidance documents.

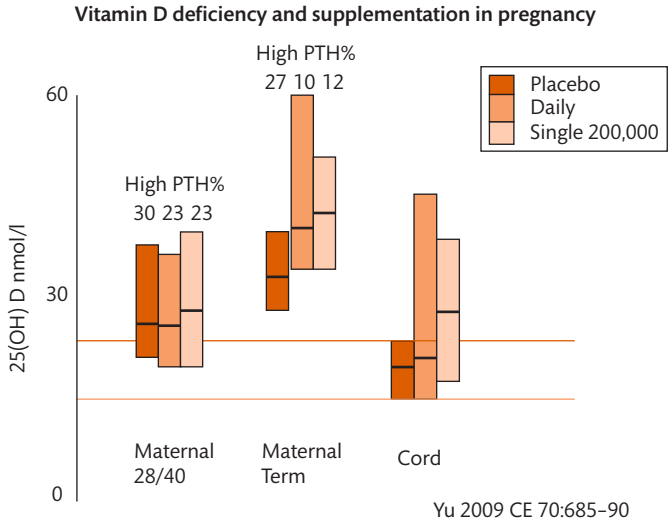


Figure 9.9.3 Vitamin D supplementation and PTH. 180 women were randomized to three groups. Daily ergocalciferol 800 IU, a single calciferol 200 000IU dose at 28 weeks and a no treatment group (routine advice and care). Either supplementation group was associated with a significant increase in term maternal and cord blood 25(OH)D levels. This was associated with a significant reduction in elevated PTH concentrations in the mother.

The current recommendation is for all pregnant and breastfeeding women to take 400 IU (10 μg) vitamin D daily. (NICE guidance 2017: https://www.nice.org.uk/guidance/ph56.) There is particular risk in some family origins, BMI>30kg/m², indoor lifestyle, covered skin, and certain diets. Dosing of up to 6000 IU/day appears safe in pregnancy and lactation [16]. Pregnant women with Vitamin D deficiency and increased one alpha hydroxylation are susceptible to sudden changes in calcium intake causing temporary hypercalcaemia, presumably related to high PTH concentrations removing ability to achieve homeostasis.

There is no recommendation for universal screening in pregnancy, so it is not recommended that vitamin D status is routinely assessed in pregnancy unless symptomatic (e.g. with bone pain) or particularly high risk (e.g. greatly reduced skin sunlight exposure). If a woman is found to be vitamin D deficient or insufficient she can be treated with 20 000 IU a week for 4 weeks, then continuing with a normal supplemental dose.

Primary Hyperparathyroidism

Introduction

Primary hyperparathyroidism (PHPT) is the third commonest endocrine disorder, however incidence rates increase markedly with age and so it is infrequent in pregnancy. Women of childbearing age have an annual incidence of about 6 per 100 000, compared to 99 per 100 000 in the age group 65–75 [16], and represent only 11% of the total incidence. The prevalence on routine biochemical screening of nearly 300 000 young Israeli women was 0.05% [17]. Its importance lies in the fact that about 50% of cases go undiagnosed, yet fetal and maternal complications are common [18].

Effect of Pregnancy on Primary Hyperparathyroidism

PHPT is under diagnosed, partly because of the dilutional effect of increased circulating volume and consequent hypoalbuminaemia from early pregnancy. Corrected or ionized calcium levels are unaffected by this. Pregnancy has variable effects on the course of the disease; although overall the peak recorded corrected calcium levels tend to be lower [17], critical elevations are often reported. In the first 12 weeks, there is a doubling of both calcitriol levels and gut calcium absorption, independent of PTH, resulting in a maternal positive calcium balance. The excess is not required by the fetus, so excretion increases, causing a state of absorptive hypercalciuria [4] and increasing renal stone risk. The high oestrogen levels may be thought to reduce excess bone resorption, but bone resorption markers increase steadily from the first trimester in normal pregnancy, so it is likely there is some increase in calcium flux out of bone [19]. In the third trimester calcium shunting across the placenta to the fetus increases markedly, but 300 mg per day is not sufficient to cause a substantial drop in maternal serum calcium, as is observed in rodents [4]. Simultaneously, increasing parathyroid hormone-related protein (PTHrP) levels from breast and placenta may reduce renal calcium loss and mobilize bone; these processes have been reported to cause hypercalcaemic crises in their own right [4]. This is particularly acute at delivery, when elevated PTHrP levels, the loss of the calcium shunt across the placenta and the sudden drop in oestrogen levels may result in an abrupt surge in calcium levels. Also during breast feeding, the average loss of 200 mg calcium daily into breastmilk may not offset the bone mobilizing effect of high PTHrP levels, resulting in worsening hypercalcaemia.

Maternal Complications

Most cases are asymptomatic to the mother or at least the symptoms are not distinguished from those of pregnancy; non-specific fatigue, nausea, vomiting, constipation, emotional lability, muscle aches, and urinary frequency [20]. In 124 affected pregnancies in 74 Israeli women, with a mean corrected calcium of 2.67 mmol/L, no increase in maternal complications was seen when compared to 431 pregnancies in 174 age matched normocalcaemic women [17]. However, the serum calcium concentration in this cohort was generally lower than in other reports of primary hyperparathyroidism in pregnancy. This may be because historical case reports and small series have a bias to report the more severely hypercalcaemic cases and their symptoms, and maternal complication rates have been reported to be high. There are increased rates of nephrolithiasis, acute pancreatitis, severe hyperemesis gravidarum and hypercalcaemic crises [20] whilst treatment of these poses significant challenges. Retrospective case control studies from Sweden show a marked increase in pre-eclampsia risks with PHPT diagnoses both up to 5 years before or after the affected pregnancies [21] and historical series suggest a 25% incidence [22].

Pregnancy and Fetal Complications

The risk to the pregnancy itself is the primary concern. In a large, single-centre series from Florida, in 62 affected pregnancies in which there was no operative intervention, 30 were lost, 3.5 times the rate in the general population [18]. Nine of these occurred later than is expected, during the second trimester. Importantly, the study linked the rates of loss to the calcium levels, with the percentage of pregnancies ending in miscarriage rising progressively from 12.5% at a serum calcium of 2.59–2.67 mmol/L to 80% at 3.04–3.24 mmol/L. The normal rate of fetal loss in mild hypercalcaemia that they demonstrated is consistent with the large series from Israel, where no increase was seen at a mean corrected calcium of 2.67mmo/L [17]. Other well-documented risks to the pregnancy include stillbirth, intrauterine growth retardation, preterm delivery, and low birth weight [20].

Neonatal Complications

The early neonatal time period has a high risk of hypocalcaemia in the baby, with symptoms such as opisthotonos, tetany, or seizures occurring in up to 50% in early series [20]. Fetuses run an elevated serum calcium and ionized calcium, typically 0.30–0.50 mmol/L above the maternal level, which is maintained by active calcium transport across the placenta. There is not a set gradient of serum calcium relative to the maternal serum calcium, but a fixed level independent of the maternal calcium value, which is maintained stable despite maternal hypercalcaemia [4]. The calcium-sensing receptor (CaSR) is not responsible for this and the higher fetal calcium levels stimulate the fetal CaSR, causing PTH secretion to be suppressed into the low range. Continued maternal hypercalcemia seems to enhance the transplacental flow of calcium to the fetus, further suppressing the fetal PTH level, even if the fetal serum calcium does not change. Post-delivery the neonatal parathyroid function is

impaired and hypocalcaemia ensues. This typically presents in the first two weeks, but may be delayed up to 2 months. It may occur even in asymptomatic PHPT or mild cases. Treatment is fortunately highly successful and spontaneous recovery normally occurs by 3 to 5 months, although damage can occasionally be permanent [4, 18]. Timely operative intervention in the mother does prevent this.

Diagnosis and Investigations

There are no specific guidelines on the diagnosis and treatment of primary hyperparathyroidism in pregnancy. The standard guidelines from the 4th international Workshop on Asymptomatic Hyperparathyroidism should therefore be followed [23]. Diagnosis is based on an elevated ionized or corrected serum calcium with PTH levels inappropriately elevated or in the upper half of the normal range. There is no specific PTH range for pregnancy and it should be noted that normally PTH levels are in the low normal range [4]. The differential diagnosis from familial hypocalciuric hypercalcaemia (FHH) and of other genetic conditions is thorny; in FHH the expected hypocalciuria is not present during pregnancy, with an increased fractional excretion of calcium by the kidneys [24]. Overall, one would expect an increased incidence of genetic conditions given the young age of diagnosis, perhaps as high as 20% [25] but this has not been borne out by the incidence of multigland disease on operation [18]. A proper family history is therefore critical and, if suspicious, urgent genetic screening is now possible, so avoiding unnecessary surgery.

Localization studies are normally limited to an ultrasound of the neck. Sensitivities are extremely user dependent, so a dedicated neck ultrasonographer is required; sensitivities are then above 60% [18]. There are case reports of modified dose ^{99M}Tc-Sestamibi scanning being used successfully, with estimated fetal exposure of < 5mGy, and 4D-CT at 10 mGy, much lower than known fetal toxicity cut-offs. These cannot yet be standardly recommended, but may have their place in complex cases [26].

Management

Noting that all pregnant women are operative candidates due to their young age and the risks to the pregnancy, a conservative approach may be followed if the mother is asymptomatic and calcium levels remain <2.75 mmol/L or <0.2 mmol/L above the upper limit of normal. Oral fluid intake should be encouraged and a low calcium diet followed. The mother should be counselled, meeting the surgeon and perinatologist. Regular maternal and fetal surveillance should be in place in a joint medical and obstetric antenatal clinic.

Medical Intervention

There are no pharmacological interventions proven safe in pregnancy, so the need for their introduction should be the signal for arranging a parathyroidectomy. Oral phosphate supplements of up to 9g per day have been used as a conservative option, but are poorly tolerated, cause diarrhoea and may precipitate hypokalaemia [20]. Cinacalcet, a calcimimetic suppressing PTH production, has been used in a number of case reports. It is poorly tolerated at the higher doses often needed, due to nausea and gastro-intestinal side-effects. Its efficacy has been limited in cases reports. It crosses the placenta and will stimulate the CaSR, which has a wide distribution in the fetus; parathyroid function will be suppressed, placental active calcium transport may be reduced and fetal bone accretion may theoretically be affected. In case reports neonatal hypocalcaemia has not been prevented [26]. Bisphosphonates are absolutely contraindicated in pregnancy, as they cross the placenta and have been shown to decrease fetal bone growth in animal studies, despite re-assuring data in accidental human usage [27].

In the acute scenario, whilst controlling severe hypercalcaemia before surgery, typical supportive management is indicated. Inpatient monitoring is used, with high volumes of intravenous normal saline, and once volume replete simultaneous IV furosemide boluses to promote calcium excretion. Subcutaneous calcitonin, at doses of up to 8 units 6 hourly, may be used, but tachyphylaxis occurs rapidly [20].

Surgical Intervention

Parathyroidectomy is generally carried out in the second trimester [22]. This can be individualized. Given the high rate of fetal loss in the late first trimester, there is an argument to intervene early in severe cases, with successful outcomes reported as early as at 11 weeks [18, 28]. Delayed operations in the third trimester have been felt to be associated with higher rates of complications, particularly pre-term labour, but these may be more related to the PHPT than the surgery. In an analysis of 16 reported cases with third trimester operations, the major complications were from pre-eclampsia (25%), pancreatitis and renal failure, which are likely to relate to the hypercalcaemia. There were no cases of anaesthesia induced premature labour and the main surgically related complication was transient maternal hypocalcaemia [22]. If hypercalcaemia therefore becomes problematic late in the pregnancy, operation is still indicated.

Pseudohyperparathyroidism

Pseudohyperparathyroidism is a very rare cause of hypercalcaemia in pregnancy and during lactation. In normal pregnancy PTHrP levels rise threefold from the first trimester to term, mainly coming from the breast and the placenta [4]. Breast hyperplasia can cause further elevation of levels and there have been rare cases reports of pregnancy hypercalcaemia being driven by markedly elevated PTHrP levels, whilst PTH is suppressed. Mastectomy has been required [29]. More cases have been reported during lactation, mostly presenting as osteoporosis, with the hypercalcaemia resolving on cessation of breastfeeding. Lactational PTHrP induced hypercalcaemia may be occurring more frequently than is appreciated [4, 30].

Familial Hypocalciuric Hypercalcaemia

Introduction

Familial hypocalciuric hypercalcaemia (FHH) is an autosomal dominant disorder most commonly resulting from heterozygous germline inactivating mutations in the CASR gene, and less frequently from mutations of GNA11 or AP2S1 [31]. It is responsible for up to 3% of cases of PTH-related hypercalcaemia in the under 45 age group [25] and needs to be borne in mind, so as to avoid an unnecessary operation.

Effect of Pregnancy on FHH

Outside of pregnancy, most cases are asymptomatic and mild biochemically. Two-thirds of cases will have serum corrected calcium

levels <10% above the upper limit of normal, and in almost 80% PTH levels will be in the normal range [32]. There is a theoretical risk of worsening hypercalcaemia in pregnancy, for similar reasons to those in PHPT. Interestingly, absorptive hypercalciuria still develops, despite patients normally having a low urine fractional calcium excretion. The 24-hour urine calcium/creatinine clearance ratio rises well above 0.01 [24], so that urine testing cannot be used to differentiate from PHPT, and genetic testing should be used if there is a suspicious family or clinical history.

Fetal Complications

Although the pregnancy is not affected, the maternal hypercalcaemia can cause suppression of the fetal parathyroid glands and temporary neonatal hypocalcaemia, dependent on the fetal genotype. Unaffected and heterozygote fetuses are at risk of hypocalcaemia, which in heterozygotes will turn to hypercalcaemia in the neonatal time period. Homozygote or compound heterozygote fetuses, inheriting a second mutation from the father or de novo, are at risk of neonatal severe hyperparathyroidism, which may require urgent parathyroidectomy. Rarely, heterozygotes may also present in this way [31]. Affected pregnant women should therefore be offered genetic counselling and the neonates have close calcium monitoring.

Effect of Lactation on FHH

During breastfeeding, PTHrP levels are normally under negative feedback control by the CaSR in mammary tissue. Inactivating mutations will lead to increased PTHrP levels and could worsen hypercalcaemia, but this has not been reported [4].

Hypoparathyroidism

Introduction

Hypoparathyroidism rarely presents during pregnancy [33], although it may be picked up in the mother when neonatal complications arise, including neonatal hypercalcaemia and fractures. It is normally known about before pregnancy, with the mother established on treatment, and the challenge is to maintain normocalcaemia despite the major alterations in calcium homeostasis that occur during pregnancy and breastfeeding. Failure to achieve this can have serious consequences.

Animal Hypoparathyroid Models

Data from parathyroidectomized rat and mouse PTH null models, show that PTH is not required for the marked increase in calcitriol levels and gut calcium absorption that occur in pregnancy. Pregnancies proceed to term without needing intervention, with no fetal or maternal adverse consequences. There is a dependence on the calcium content of the diet in rats, and if this is limited, maternal hypocalcaemia may occur [4].

Effect of Pregnancy on Hypoparathyroidism

In affected human mothers, case reports and series show that maternal symptoms tend to improve in pregnancy, with fewer episodes of tetany, cramps or paraesthesiae. Although animal data suggests activated Vitamin D supplements can be discontinued, and series prior to the introduction of Vitamin D analogues demonstrated successful outcomes [34], this approach is not used, as maternal dosing requirements seem to be extremely variable. Even within the same patient, different pregnancies may require dose titration in opposite directions [35]. This may partly be because dietary calcium intake is not controlled.

Pregnancy Complications

The consequences of unintended hypocalcaemia can be serious for the pregnancy. Spontaneous midterm abortion, premature labour and stillbirths have been reported in severe cases [33]. These may occur because hypocalcaemia increases uterine muscle excitability. A higher incidence of small for gestational age fetuses has also been suggested [35].

Fetal Complications

The fetus can maintain its own calcium levels despite maternal hypocalcaemia. The fetoplacental unit upregulates mechanisms to maintain the fetal blood calcium and increase placental calcium transport, with animal models showing increased PTHrp and PTH levels [36]. This marked increase in fetal PTH levels though is accompanied by parathyroid gland hyperplasia, and increased bone resorption, so that the skeleton is undermineralized. The skeletal consequences include reduced bone mass, neonatal rickets, and fractures in utero and on delivery [37]. Chest wall deformities may lead to respiratory distress. Neonatal hypercalcaemia secondary to parathyroid hyperplasia may occur.

The consequences of unintended hypercalcaemia have already been discussed under primary hyperparathyroidism.

Effect of Lactation on Hypoparathyroidism

Lactation results in an improvement in calcium homeostasis in hypoparathyroid mothers. Breast PTHrP production rises with breastfeeding, causing increased bone turnover, particularly in the trabecular compartment, and releasing calcium. PTHrP acts on the kidney to increase calcitriol production to normal levels and increase renal tubular calcium reabsorption. Patients can therefore go into positive calcium balance and lactation can result in hypercalcaemia if pregnancy doses of calcitriol or 1α calcidiol are continued. The rapidity of the decline of this state on weaning is variable. It may occur whilst breastfeeding frequency is reducing or be delayed post-weaning by weeks, but usually occurs within a few days [4, 38].

In the transition between delivery and initiation of breastfeeding, PTHrP levels from the placenta drop, before rising again from the breast. Temporary maternal hypocalcaemia has been reported at this time [4].

Treatment

There are no management guidelines published for hypoparathyroidism in pregnancy. Management is usually conservative, continuing on pre-pregnancy doses of 1α calcidiol or calcitriol and calcium supplementation to start with. Total calcium intake should be at least 1.2 g per day. Corrected or ionized calcium levels should be followed, as total calcium levels will drop with dilutional volume expansion in the first trimester and dose adjustments made on this basis will be erroneous. Change should be watched for with monthly calcium monitoring from 12 weeks. Dose titration is often required during the second trimester. Changes in urine calcium excretion

should not be monitored. Maintenance of normocalcaemia is the aim, but avoidance of hypocalcaemia is particularly important, so calcium levels should be run within the low-mid normal range (2.2–2.4 mmol/L) [4], rather than borderline low as is frequently practised outside of pregnancy. Mothers should be aware of the symptoms, so be able to seek help, and followed in a joint obstetric clinic with endocrine input. At delivery, it is advisable to at least return to prepregnancy doses if they have been increased, and then to watch closely with the establishment of breastfeeding, when dose reductions should be expected, moving to monthly monitoring again through lactation.

Pseudohypoparathyroidism

Introduction

Pseudohypoparathyroidism (PHP) refers to the metabolic picture of hypocalcaemia with hyperphosphataemia but unexpected high PTH levels, due to resistance to the action of PTH in target tissues, in the context of normal magnesium levels.

PHP type 1a patients have heterozygous mutations affecting the maternally inherited allele of GNAS, encoding the $G\alpha_s$ subunit, hence show resistance not only to PTH, but also to other hormones in tissues where the maternal allele of GNAS is preferentially expressed. Typically, this would include the pituitary, with resistance to gonadotropins and GHRH, and the thyroid with resistance to TSH, so that pregnancies are rare. These patients also show the somatic features of Albright's hereditary osteodystrophy (AHO). In PHP type 1b, patients have mainly PTH resistance due to methylation loss in the maternally inherited GNAS allele, resulting in decreased transcription levels. The AHO phenotype is not seen. Partial TSH resistance is commonly seen [39].

Effect of Pregnancy on Pseudohypoparathyroidism

Case reports in pregnancy are few, but suggest that the expected PTH independent increase in calcitriol levels does still occur, and therefore patients may remain normocalcaemic despite discontinuation of their normal calcitriol [40]. As with hypoparathyroidism, though, cases reports show that maternal dosing requirements are very variable and calcium content of the diet is probably critical [4].

Treatment

The basic management should be the same as for hypoparathyroidism. Additional factors though need to be considered in pregnancies affected by PHP. Thyroid status should be tracked and adequate replacement given. Counselling and prenatal diagnosis of the fetal genetic status are required, not just to predict neonatal issues such as hypocalcaemia, but also to predict changes in intrauterine growth [41].

Osteoporosis in Pregnancy and Lactation

Introduction

Osteoporosis is uncommon in women of child-bearing age and infrequently thought of at the time of pregnancy. Rarely, though, fractures may occur, typically during the third trimester or early in lactation, presenting with back pain and single or multiple vertebral compression fractures. It was recognized to be associated with the temporary changes in bone metabolism in pregnancy and lactation, and first proposed as a distinct syndrome in 1955.

Animal Bone Metabolism in Pregnancy

In rodent studies it is estimated that up to 92% of fetal skeletal calcium content comes from the maternal diet during pregnancy, with just a small amount near term from skeleton. When bone mineral density (BMD) is measured sequentially through pregnancy in a mouse model there is a continuous gain in total body mineral content from early pregnancy to up to 20% by birth. In rats measured at the start and end of pregnancy, there is no change in whole body BMD, a good marker for overall mineral balance, but small significant declines in lumbar spine BMD. These vary by study, with the calcium content of the diet being a significant factor. This remarkable conservation of the skeleton is achieved by a large increase in gut absorption of calcium, which doubles from mid-pregnancy, allowing the pregnant mother to accrete extra bone mineral in preparation for the higher fetal demands in late pregnancy and lactation [4].

Maternal Bone Metabolism and Mineral Density in Pregnancy

The physiology in human mothers is similar (Figures 9.9.1 and 2). The marked increase in gut calcium absorption during the first 12 weeks, allows the mothers to go into positive calcium balance by mid-pregnancy. Whether mothers accrete new bone mineral in preparation for late pregnancy and lactation is not clear [4]. Studies of bone markers show that resorption markers increase steadily from early pregnancy, being significantly elevated by 10 weeks [19, 42]. Bone formation markers, such as carboxy-terminal propeptide of type 1 procollagen (P1CP), are reduced in the first two trimesters, then climb rapidly in late pregnancy [19, 42]. There is therefore some uncoupling of bone resorption and formation in the first two trimesters, but overall this is consistent with a more marked increase in bone turnover in the last trimester.

Longitudinal bone mineral density measurements during pregnancy are limited, with variable results between and within studies [43]. One Sheffield study, looking at 16 Caucasian women with BMDs performed prior to and 2 weeks post-pregnancy, showed that overall BMD of total body changes little, but there does seem to be mineral redistribution occurring, with the BMD at sites rich in cortical bone increasing, 2.8% at the arm and 1.8% at the leg, and the BMD at sites rich in trabecular bone decreasing, with a 4.5% decrease at spine and 3.2% at pelvis [44]. The two largest controlled studies show differing results. One from Nottingham looked at 46 women, comparing them with 30 controls [45]. There was a small non-significant decline in BMD in the pregnant group at all sites, performed prior to pregnancy and within 2 weeks of delivery. A second larger study from Denmark also looked at forearm BMD in each trimester [42]. Of the 92 women recruited, 73 completed and were compared to 57 controls. DXA measurements at 2 weeks post-delivery, with all women having started breast-feeding, showed a 4% decrease at the ultradistal radius, a mainly trabecular site, and 0.5% gain at the distal proximal 1/3 forearm, a mainly cortical site. The rate of loss at the ultradistal radius was increasing through the trimesters and the magnitude significant by the 10-week visit. This mirrored changes seen at the total hip and spine, done only prior to

and post-pregnancy, where a 3.2% and 1.8% loss were seen respectively. Whole body BMD decreased by 2.4%.

These data suggest that there is normally only a small loss of maternal bone to help obtain the mineral for the formation of the fetal skeleton, and that this loss is mainly in the third trimester, as in animal models. Likely risk factors for increased loss include a low calcium diet and secondary hyperparathyroidism. Normally, due to the marked increase in gut calcium absorption and calcitriol levels, PTH levels are in the low normal range. However, in populations with low calcium intake or severe vitamin D deficiency secondary hyperparathyroidism still occurs [3]. Twin pregnancies, that would require a greater calcium shunt, have been shown to have elevated bone turnover markers [43].

Maternal Calcium and Bone Metabolism in Lactation

Lactation poses a greater burden on the mother. By 9 months a neonate will require another 30 g of calcium for skeletal development, which is provided by an average maternal shunt of 210 mg per day in breast milk. This essentially comes from her bone. The picture of absorptive hypercalciuria during pregnancy switches to a bone resorptive picture during breast feeding. Prolactin secretion in response to suckling suppresses the GnRH generator, dropping LH, FSH, and E2 levels, so that E2 related restraint on bone turnover is removed. PTHrP production from the breast, stimulated by both suckling and prolactin, enters the systemic circulation and causes bone mobilization, particularly in trabecular bone. This provides calcium for secretion into breast milk, under feedback control through the CaSR in mammary tissue. Stimulation of this CaSR by mobilized calcium causes both a reduction of PTHrP release into the systemic circulation and breast milk with increased secretion of calcium into the breast milk [4].

Other calcium regulatory systems play little role. The increased gut calcium absorption characteristic of pregnancy rapidly disappears, as calcitriol levels drop to the middle of the non-pregnant range. With the high systemic calcium load coming from bone, PTH is suppressed into the low normal range, but PTHrP does act on the kidney, increasing calcium conservation, so that hypercalciuria resolves.

Maternal Bone Mineral Density in Lactation

Women showed a marked loss in BMD with this resorptive phase, typically between 5% and 10% at trabecular sites [43]. This is most pronounced in the first 3–6 months, and with prolonged breastfeeding some regain actually starts to be seen at the lumbar spine, although loss continues at the hip [42]. Whereas in animal studies cortical bone loss is marked, with osteocytic osteolysis allowing resorption of pericanicular and lacuna spaces around the embedded osteocyte, in women this does not seem to be occurring and cortical bone is relatively spared, with cortical forearm sites showing little change [42]. Risk factors for greater bone loss include adolescence, continued amenorrhoea, prolonged breastfeeding and high PTHrP levels [43]. Calcium supplementation does not seem to alter the loss [46, 47].

BMD recovers by 12 months after weaning [47], with studies showing different rates at different sites. In the largest controlled cohort study, Moller showed that if breastfeeding stopped by 4 months, there was full recovery of whole body and lumbar spine densities to pre-pregnancy levels by 9 months, but that at this time it was less than 50% at the hip and minimal at the ultradistal radius [42]. Overall recovery is also not affected by a repeat pregnancy within 18 months. In rodent studies, there is never full recovery in some trabecular sites in the long bones, but bone dimensions change, with an increased cross-sectional diameter, such that bending and torsional strength are maintained [43]. Recent high resolution peripheral quantitative CT confirms that there is microarchitectural deterioration post-lactation in women, with trabecular loss and thinning as well as increased cortical porosity at the distal radius and distal tibia [48]. Strength analyses are not available from this study, but other reports show that high parity is associated with increased radial torsional bending strength, increased femoral neck size, and increased tibial cross-sectional area, paralleling the rodent data [49, 50]. A number of epidemiological studies show that increased parity is not associated with low BMD or fracture risk, with a recent meta-analysis suggesting no effect on femoral neck or lumbar spine density, and a small positive effect at total hip BMD [43, 51]. Calcium supplementation of 1000 mg daily post-weaning has been shown to speed up regain of BMD at lumbar spine, but it is not known if this will translate into a long-term benefit, especially at other sites [46].

Investigation of Pregnancy-Associated Osteoporosis

On discovering a case of postpartum osteoporosis, it should not be assumed that this is pregnancy or lactation related. This may be true, as there is wide variation in the rate of bone loss between individuals, prolonged breastfeeding or twins may increase loss, and certain medications used in pregnancy may enhance loss, such as LMWH or prednisolone, but other factors need to be considered. These would include typical risk factors for osteoporosis such as prior low BMI or anorexia, severe childhood illnesses meaning that normal peak bone mass was never achieved, systemic inflammatory conditions such as rheumatoid arthritis, malabsorptive disorders such as coeliac disease, drugs such as glucocorticoids, antiepileptics and depot progesterone, HIV, and a maternal history of fractures. More specific congenital or metabolic bone disorders may also first present this way, and should be considered, including primary HPT, idiopathic hypercalciuria, juvenile idiopathic osteoporosis, LRP5 mutations, and osteogenesis imperfecta [43].

Treatment of Pregnancy-Associated Osteoporosis

The management should ideally be conservative, as full recovery of bone density should occur rapidly on recommencement of menstruation. Although a number of pharmacological interventions have been reported, it is difficult to know if they were of benefit as regain occurs spontaneously. Indeed, they may impair bone regain and be contra-indicated in the context of further expected pregnancies, as bisphosphonates. Return of oestrogen effect is what is required. Breastfeeding should be discontinued with cabergoline being given if required. Calcium and vitamin D supplementation should be given; the trial evidence available is with 1000 mg calcium and 400 IU vitamin D_3 daily [46], but the latter should be titrated to bring 25(OH)vitamin D levels at least above 50 nmol/L. Mobilization should be encouraged. Depot progesterone contraception should be avoided. Mothers can be reassured that this will be unlikely to recur in future pregnancies [43].

On tracking BMD, full recovery should have occurred by 12 months. If osteoporosis still persists, further investigation for rarer conditions is indicated and a full family history desirable. A plan for future pregnancies would then be indicated.

Transient Osteoporosis of the Hip

Introduction

Transient osteoporosis of the hip was first described in three cases during pregnancy in1959, but has since been found to occur outside of pregnancy, at other weight-bearing joints and in older men [52]. It has an incidence of 6 per 10 000 pregnancies, but this is likely to be an underestimate, as joint pains are common in pregnancy [53].

Presentation

Typically it presents with spontaneous pain in the hip in the third trimester or early puerperium. This may come on suddenly or gradually, be felt in the groin or anterior-laterally in the thigh and worsens on weight-bearing or exercise. The patient may limp and on examination there may be soft tissue swelling, joint effusions and occasionally muscle wasting. Both hips may be affected and in rare cases hip fracture may occur [52].

Diagnosis

Diagnosis is based on clinical suspicion, followed by appropriate radiological investigation. X-rays show osteopaenia or radiolucency, and DEXA will show markedly reduced hip BMD on the affected side, especially when compared to the lumbar spine BMD. The magnitude of loss is such that from 20% to 40% regain of BMD has been seen on recovery [43]. MRI is diagnostic, showing epiphysial bone marrow oedema extending into the subchondral bone, and can be used to differentiate it from osteonecrosis of the hip, which can also present in pregnancy [52].

Aetiology

Although pregnancy may be a precipitating factor, transient regional osteoporosis (TRO) is not necessarily associated with the bone changes seen in pregnancy and lactation. The aetiology is poorly understood and histological studies have been conflicting. Venous stasis may be a critical component and it has been suggested that it is a variant of reflex sympathetic dystrophy [43].

Treatment

Treatment is supportive, as the changes typically resolve in a number of months, with MRI showing resolution of the bone marrow oedema. Reduction of weight-bearing activities and bedrest may be necessary. Adequate opioid analgesia is given. Pharmacological treatment is not indicated given the natural history of resolution. Women should be warned that recurrence can occur in later pregnancies.

REFERENCES

1. Van de Meer IM, Karamali NS, Boeke AJ, *et al*. High prevalence of vitamin D deficiency in pregnant non-Western women in The Hague, Netherlands. *Am J Clin Nutr*, 2006; **84**: 350–3.
2. Bowyer L, Catling-Paull C, Diamond T, Horner C, Davis G, Craig ME. Vitamin D, PTH and calcium levels in pregnant women and their neonates. *Clin Endocrinol (Oxf)*, 2009; **70**: 372–7.
3. Yu CK, Sykes L, Sethi M, Teoh TG, Robinson S. Vitamin D deficiency and supplementation during pregnancy. *Clin Endocrinol (Oxf)*, 2009; **70**: 685–90.
4. Kovacs CS. Maternal mineral and bone metabolism during pregnancy, lactation, and post weaning recovery. *Physiol Rev*, 2016; **96**: 449–547.
5. Palacios C, De-Regil LM, Lombardo LK, Pena-Rosas JP. Vitamin D supplementation during pregnancy: updated meta-analysis on maternal outcomes. J *Ster Biochem Biol*, 2016 **164**: 148–55.
6. Lawlor DA, Wills AK, Fraser A, Sayers A, Fraser WD, Tobias JH. Association of maternal vitamin D status during pregnancy with bone-mineral content in offspring: a prospective cohort study. *Lancet*, 2013; **381**: 2176–83.
7. Brooke OG, Brown IR, Bone CD, *et al*. Vitamin D supplements in pregnant Asian women: effects on calcium status and fetal growth. *Br Med J*, 1980; **280**: 751–4
8. Camargo CA, Rifas-Shiman SL, Litonjua AA, *et al*. Maternal intake of vitamin D during pregnancy and risk of recurrent wheeze in children at 3 y of age. *Am J Clin Nutr*, 2007; **85**: 788–95
9. Cremers E, Tijis C, Penders J, Jansen E, Mommers M. Maternal and child's vitamin D supplement use and vitamin D level in relation to childhood lung function: the KOALA Birth Cohort Study. *Thorax*, 2011; **66**: 474–80.
10. Hyppönen E, Hartikainen AL, Sovio U, Järvelin MR, Pouta A. Does vitamin D supplementation in infancy reduce the risk of pre-eclampsia? *Eur J Clin Nutr*, 2007; **61**: 1136–9.
11. Bodnar LM, Catov JM, Simhan HN, Holick MF, Powers RW, Roberts JM. Maternal vitamin D deficiency increases the risk of preeclampsia. *J Clin Endocrinol Metab* 2007; **92**: 3517–22.
12. Robinson CJ, Alanis MC, Wagner CL, Hollis BW, Johnson DD. Plasma 25-hydroxyvitamin D levels in early-onset severe preeclampsia. *Am J Obstet Gynecol* 2010; **203**: e1–6.
13. Haugen M, Brantsaeter AL, Trogstad L, *et al*. Vitamin D supplementation and reduced risk of pre-eclampsia in Nulliparous women. *Epidemiol*, 2009; **20**: 720–6.
14. Song Y, Wang L, Pittas AG, *et al*. Blood 25-hydroxy vitamin D levels and incident Type 2 diabetes: a meta analysis of prospective studies. *Diab Care*, 2013; **36**: 1422–8.
15. Merewood A, Mehta SD, Chen TC, Bauchner H, Holick MF. Association between vitamin D deficiency and primary Cesarean section. *J Clin Endo Metab*, 2009; **94**: 940–5.
16. Wermers RA, Khosla S, Atkinson EJ, *et al*. Incidence of primary hyperparathyroidism in Rochester, Minnesota, 1993–2001: an update on the changing epidemiology of the disease. *J Bone Miner Res*, 2006; **21**: 171–7.
17. Hirsch D, Kopel V, Nadler V, Levy S, Toledano Y, Tsvetov G. Pregnancy outcomes in women with primary hyperparathyroidism. *J Clin Endocrinol Metab*, 2015; **100**: 2115–22.
18. Norman J, Politz D, Politz P. Hyperparathyroidism during pregnancy and the effect of rising calcium on pregnancy loss: a call for earlier intervention. *Clin Endocrinol*, 2009; **71**: 104–9.
19. Black AJ, Topping J, Durham B, Farquharson RG, Fraser WD. A detailed assessment of alterations in bone turnover, calcium homeostasis, and bone density in normal pregnancy. *J Bone Miner Res*, 2000; **15**: 557–63.
20. Schnatz PF, Curry SL. Primary hyperparathyroidism in pregnancy: evidence-based management. *Obstet Gynecol Surv*, 2002; **57**: 365–76.
21. Hultin H, Hellman P, Lundgren E, *et al*. Association of parathyroid adenoma and pregnancy with preeclampsia. *J Clin Endocrinol Metab*, 2009; **94**: 3394–9.
22. Schnatz PF, Thaxton S. Parathyroidectomy in the third trimester of pregnancy. *Obstet Gynaeocol Surv*, 2005; **60**: 672–82.

23. Bilezikian JP, Brandi ML, Eastell R, *et al*. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab*, 2014; **99**: 3561–9.
24. Ghaznavi SA, Saad NM, Donovan LE. The biochemical profile of familial hypocalciuric hypercalcemia and primary hyperparathyroidism during pregnancy and lactation: two case reports and review of the literature. *Case Rep Endocrinol*, 2016; **2016**: 2725486.
25. Starker LF, Akerström T, Long WD, *et al*. Frequent germ-line mutations of the MEN1, CASR, and HRPT2/CDC73 genes in young patients with clinically non-familial primary hyperparathyroidism. *Horm Cancer*, 2012; **3**: 44–51.
26. Rubin MR, Silverberg SJ. Use of cinacalcet and 99mTc-sestamibi imaging during pregnancy. *J Endocr Soc*, 2017; **1**: 1156–9.
27. Djokanovic N, Klieger-Grossmann C, Koren G. Does treatment with bisphosphonates endanger the human pregnancy? *J Obstet Gynaecol Canada*, 2008; **30**: 1146–8.
28. Hui E, Osakwe O, Teoh TG, Tolley N, Robinson S. Three case reports of maternal primary hyperparathyroidism in each trimester and a review of optimal management in pregnancy. *Obstet Med*, 2010; **3**: 33–7.
29. Jackson IT, Saleh J, van Heerden JA. Gigantic mammary hyperplasia in pregnancy associated with pseudohyperparathyroidism. *Plast Reconstr Surg*, 1989; **84**: 806–10.
30. Reid IR, Wattie DJ, Evans MC, Budayr AA. Post-pregnancy osteoporosis associated with hypercalcaemia. *Clin Endocrinol*, 1992; **37**: 298–303.
31. Hannan FM, Babinsky VN, Thakker RV. Disorders of the calcium-sensing receptor and partner proteins: insights into the molecular basis of calcium homeostasis. *J Mol Endocrinol*, 2016; **57**: R127–42.
32. Nissen PH, Christensen SE, Heickendorff L, Brixen K, Mosekilde L. *CASR* Gene Mutations in Danish FHH Subjects. *J Clin Endocrinol Metab*, 2007; **92**: 4373–9.
33. Eastell R, Edmonds CJ, de Chayal RC, McFadyen IR. Prolonged hypoparathyroidism presenting eventually as second trimester abortion. *Br Med J (Clin Res Ed)*, 1985; 5; **291**: 955–6.
34. Graham WP, Gordan GS, Loken HF, Blum A, Halden A. Effect of pregnancy and of the menstrual cycle on hypoparathyroidism. *J Clin Endocrinol Metab*, 1964; **24**: 512–16.
35. Hatswell BL, Allan CA, Teng J, *et al*. Management of hypoparathyroidism in pregnancy and lactation: a report of 10 cases. *Bone Reports*, 2015; **3**: 15–19.
36. Kovacs CS. Bone development and mineral homeostasis in the fetus and neonate: roles of the calciotropic and phosphotropic hormones. *Physiol Rev*, 2014; **94**: 1143–218.
37. Loughead JL, Mughal Z, Mimouni F, Tsang RC, Oestreich AE. Spectrum and natural history of congenital hyperparathyroidism secondary to maternal hypocalcemia. *Am J Perinatol*, 1990; **7**: 350–5.
38. Cundy T, Haining SA, Guilland-Cumming DF, Butler J, Kanis JA. Remission of the hypoparathyroidism during lactation: evidence for a physiological role for prolactin in regulation of vitamin D metabolism. *Clin Endocrinol*, 1987; **26**: 667–74.
39. Levine MA. An update on the clinical and molecular characteristics of pseudohypoparathyroidism. *Curr Opin Endocrinol Diabetes Obes*, 2012; **19**: 443–51.
40. Breslau NA, Zerwekh JE. Relationship of estrogen and pregnancy to calcium homeostasis in pseudohypoparathyroidism. *J Clin Endocrinol Metab*,1986; **62**: 45–51.
41. Bréhin A, Cindy Colson C, Maupetit-Méhouas S, *et al*. Intrauterine growth and *GNAS* methylation. *J Clin Endocrinol Metab*, 2015; **100**: E623–31
42. Møller UK, Streym S, Mosekilde L, *et al*. Changes in calcitropic hormones, bone markers and insulin-like growth factor I (IGF-I) during pregnancy and postpartum: a controlled cohort study. *Osteoporos Int*, 2013; **24**: 1307–20.
43. Kovacs CS, Ralston SH. Presentation and management of osteoporosis presenting in association with pregnancy or lactation. *Osteoporos Int*, 2015; **26**: 2223–41.
44. Naylor KE, Iqbal P, Fledelius C, Fraser RB, Eastell R. The effect of pregnancy on bone density and bone turnover. *J Bone Miner Res*, 2000; **15**: 129–37.
45. Kaur M, Pearson D, Godber I, Lawson N, Baker P, Hosking D. Longitudinal changes in bone mineral density during normal pregnancy. *Bone*, 2003; **32**: 449–54.
46. Kalkwarf HJ, Specker BL, Bianchi DC, Ranz J, Ho M. The effect of calcium supplementation on bone density during lactation and after weaning. *N Engl J Med*, 1997; **337**: 523–8.
47. Polatti F, Capuzzo E, Viazzo F, Colleoni R, Klersy C. Bone mineral changes during and after lactation. *Obstet Gynecol*, 1999; **94**: 52–6.
48. Bjørnerem Å, Ghasem-Zadeh A, Wang X, *et al*. Irreversible deterioration of cortical and trabecular microstructure associated with breastfeeding. *J Bone Miner Res*, 2017; **32**: 681–7.
49. Specker B, Binkley T. High parity is associated with increased bone size and strength. *Osteoporos Int*, 2005; **16**: 1969–74.
50. Wiklund PK, Xu L, Wang Q, *et al*. Lactation is associated with greater maternal bone size and bone strength later in life. *Osteoporos Int*, 2012; **23**: 1939–45.
51. Song SY, Kim Y, Park H, Kim YJ, Kang W, Kim EY. Effect of parity on bone mineral density: a systematic review and meta-analysis. *Bone*, 2017; **101**: 70–6.
52. Cano-Marquina A, Tarín JJ, García-Pérez MÁ, Cano A. Transient regional osteoporosis. *Maturitas*, 2014; **77**: 324–9.
53. Steib-Furno S, Luc M, Pham T, *et al*. Pregnancy-related hip diseases: incidence and diagnoses. *Joint Bone Spine*, 2007; **74**:373–8.

9.10

Imaging for Endocrine Diseases in Pregnancy

Sandra Lowe

Introduction

The pregnant woman is exposed constantly to a range of ionizing, non-ionizing, and electromagnetic radiation in both her own environment and as a result of medical interventions [1]. Imaging techniques that do not involve electromagnetic radiation such as ultrasound and magnetic resonance imaging (MRI) are considered safe in pregnancy.

X-rays, computerized tomography (CT) with and without iodinated contrast, nuclear scanning, and positron emission tomography (PET) scanning all involve potential exposure of both mother and fetus to ionizing radiation. Such imaging may occur inadvertently or by intention during pregnancy, necessitating appropriate counselling. In addition, CT, MRI, or even ultrasound may require administration of a contrast agent to improve image definition or other additional information

For particular endocrine diseases, imaging involving radiation may be the gold standard (e.g. for pituitary or adrenal tumours), although non-ionizing options such as MRI or ultrasound may be an alternative in certain situations. Similarly, nuclear scanning is a standard modality for thyroid and parathyroid assessment. In more serious conditions such as phaeochromocytoma, a number of imaging modalities may be required including specialized nuclear and positron emission tomography (PET) scans, with or without CT.

The following chapter provides an overview of the principles and practical application of imaging modalities for endocrine disorders in pregnancy.

Medical Imaging Without Radiation

Ultrasound and magnetic resonance imaging (MRI) are considered safe modalities for imaging in pregnancy because of the absence of ionizing radiation. With MRI, heating of the fetal tissues by radiofrequency fields might theoretically lead to impairment of cell migration as demonstrated in animal models. Even with whole body imaging at the recommended allowable maximum specific absorption rate, the mother's body temperature might increase 0.6°C for 20–30 minutes but the maternal tissues dissipate this heat and the fetus is exposed to a much lower temperature and no adverse fetal effects have been reported [2]. As a precaution, the recommendation is to use 1.5-Tesla (T) scanners compared with 3.0-T scanners because the specific absorption rate quadruples when the magnetic field doubles [3].

Various guidelines have been published expressing confidence in the use of MRI throughout pregnancy, including the first trimester, with no evidence of adverse outcomes on pregnancy or the fetus/neonate [3–5].

Ionizing Radiation During Pregnancy

Box 9.10.1 describes the ways radiation exposure can be defined in pregnancy.

Recent data has estimated that total radiation exposure for an individual is made up of background exposure at 50%, medical exposure at 48% and radiation from consumer products and industrial, research, and occupational exposures, amounting to a

Box 9.10.1 Definitions

Absorbed dose:
- mean energy imparted per unit mass

Units of radiation:
- 1 gray (Gy) = 1000 milligray (mGy) = 100 rad

Radiation weighting factor:
- Equal doses of different types of energizing radiation do not necessarily lead to equal detrimental effects. These differences relate to microscopic energy dissipation and are corrected for by a radiation weighting factor (w_R) which when multiplied by the dose produces a quantity expressed as the sievert (Sv)
- Most diagnostic imaging the weighting factor is 1 (from photons and electrons)

Nuclear scanning with radiopharmaceuticals:
- fetal dose is expressed as mGy per megabecquerel (mGy/MBq), i.e. absorbed dose to the fetus per unit activity administered to the mother

2% [6]. The major radiation source for most women during pregnancy is environmental, arising from a number of sources including cosmic, gamma and radon. The magnitude of the exposure varies with a global average estimated radiation dose over the 9 months of pregnancy being 2.3 mSv (Table 9.10.1) [1, 7]. For a fetus, the dose is much smaller, 0.5–1 mSv, because of attenuation through the mother's tissue [8]. When a pregnancy is declared, by a female working in the presence of radiation, the embryo or fetus is afforded the same level of protection as the general public (i.e. 1mSv per year).

Fetal and Maternal Radiation Doses During Diagnostic Imaging

The calculation of radiation dose, both maternal whole body and organ specific as well as the estimated fetal absorbed dose may be derived from established dosimetry tables and/or by specific calculations/measurements made by a medical physicist on a case by case basis. All radiation does are potentially additive and where multiple procedures are performed, cumulative dosage should be calculated.

Table 9.10.1 Environmental radiation sources

Environmental exposures	mSv/exposure or per year
Airport scanner	0.005
Air travel: • at higher latitudes multiply by a factor of 2 or 3.	0.005 short flight 0.03 long flight
Aircrew/frequent flyers	1–5
Background radiation • UK or Australia • USA • Kerala, India • Some areas of Brazil and Sudan • Ramsar in Iran	 1.8 6.2 30 40 Up to 260
Uranium miners/nuclear industry workers	Additional to background: 1.5–2.5

Much research has been undertaken on the effects of low-level radiation which is the level most relevant when discussing risks of diagnostic imaging. It has generally been assumed that the relationships observed between radiation dose and adverse effects at high levels of radiation exposure also apply to low levels: the linear no-threshold (LNT) hypothesis. This concept remains a source of controversy among experts although the International Commission on Radiological Protection recommends that the LNT model should be assumed for the purpose of optimizing radiation protection practices, including protection during pregnancy [7, 9].

A number of national and international bodies have issued guidelines or policy documents which indicate a negligible risk to the fetus for radiation doses below 50 mGy. This is well above the dose from almost any single diagnostic imaging event although cumulative studies could lead to such a dose [10–12].

Risks of Ionizing Radiation in Pregnancy

In the assessment of any adverse effects of radiation upon the fetus it is necessary to consider:

- Absorbed dose
- Timing of the exposure relative to conception and
- Form of administration

Appropriate adjustments to standard X-ray and CT procedures can be made to lower the fetal absorbed dose without impairing image quality [8]. For most standard radiography, the fetal dose is very low (Table 9.10.2) [13].

In addition to the radiation technique, site and dose administered; maternal factors will influence the fetal absorbed dose. During X-rays and CT scanning the 'thickness' of the mother, which alters as pregnancy progresses, will influence the penetration of the dose and hence the fetal absorbed dose. The use of appropriate shielding will significantly reduce the fetal absorbed dose as well as the dose to adjacent maternal tissues.

In the case of nuclear medicine studies, fetal absorbed dose will represent the cumulative effect of external irradiation from the maternal tissues as well as placental transfer and fetal uptake of radiopharmaceuticals (Table 9.10.3) [14–16]. For radioisotopes

Table 9.10.2 Examples of fetal radiation doses for imaging of endocrine disease in pregnancy

Examination	Clinical scenario	Typical fetal dose range (mGy) in early pregnancy
CT head and/or neck	Thyroid or parathyroid	0.001–0.01
Xray hip	Avascular necrosis of hip	0.1–1.0
^{99}Tc thyroid scan	Thyroid toxic nodule	0.1–1.0
CT abdomen	Adrenal mass	1.0–10
^{18}F-FDG PET/CT	Malignancy Phaeochromocytoma and paragangliomas	10–50
Dual X-ray absorptiometry (DEXA): Posterior-anterior spine Proximal femur	Osteoporosis	1.7–4.9 1.0–2.7

Table 9.10.3 Fetal whole body dose from common nuclear medicine examinations for endocrine disorders comparing early pregnancy and at term. Dose includes maternal and fetal self-dose contributions

		Administered activity (MBq)	Early pregnancy (mGy)	Late pregnancy (mGy)
^{99m}Tc	Bone scan (phosphate)	750	4.58	1.8
^{99m}Tc	Thyroid scan (pertechnate)	250	3.45	0.41
^{99m}Tc	Sestamibi –parathyroid		6.0	10–50
^{123}I	Thyroid uptake[a]	30	0.4–0.6	0.3
^{131}I	Thyroid uptake[a]	0.55	0.03–0.04	0.15
^{131}I	Metastases imaging[a]	40	2.0–2.9	11.0

[a] Fetal thyroid doses are much higher than fetal whole body dose.

that are excreted in urine, radiation from bladder contents or the placenta will have a greater impact in early pregnancy compared with later pregnancy. This leads to a decline in fetal absorbed dose as pregnancy progresses for most nuclear imaging [15]. By using smaller administered doses and longer imaging times, the fetal absorbed dose may be further reduced, but there is concern that this may compromise the value of the study. The estimated fetal radiation doses quoted in Table 9.10.3 need to be adjusted for administered dose which may vary significantly between institutions. Although a reduction of the injected dose can be compensated for by increasing the scanning time, if the acquisition time is longer patients are more likely to move or ask to interrupt the scan to void the bladder (not a rare occurrence in pregnant women) [17]. Strategies to aid excretion of radiopharmaceuticals such as ensuring prompt and complete bladder emptying will reduce the fetal absorbed dose without compromising image quality and should be encouraged.

Older techniques for estimating fetal absorbed dose based on phantoms and models of the pregnant woman are being replaced by combinations of geometric constructs based on medical image data that are far more realistic [18]. However, estimating the absorbed dose to individual fetal organs remains difficult [19].

Nuclear scanning has been enhanced by the addition of hybrid SPECT/CT camera technology that allows simultaneous acquisition of combined multimodality imaging, with seamless fusion of three-dimensional volume datasets. The combination provides enhanced information with imaging of functional information and anatomy/localization. In endocrine disease, this may be particularly helpful for parathyroid and thyroid imaging as well as imaging for rare neuroendocrine tumours, adrenal cortical hyperfunction, phaeochromocytoma, and paraganglioma [20]. The fetal radiation dose from such procedures will be greater than from any single modality but the additional information may be critical.

Specific Risks of Diagnostic Radiation in Pregnancy

Experimental assessment of high dose radiation has identified seven specific areas of potential concern to the pregnant woman and her fetus.

- Lethality
- Genetic damage/epigenetic change
- Teratogenicity
- Growth impairment
- Sterility
- Oncogenicity

The effects of radiation exposure in pregnancy depend on the time of exposure as well as the fetal absorbed dose. Until the placenta implants, the cells of the conceptus are hypoxic and therefore less radiosensitive. In the very early embryo, the effect of radiation is more likely to be failure to implant or undetectable death of the embryo (i.e. an 'all-or-none' phenomenon) [21].

With regard to potential fetal dosage from diagnostic radiation, there is no evidence for any increase in lethality (miscarriage or stillbirth), teratogenicity, genetic damage/epigenetic changes, growth impairment, mental retardation, or sterility [9, 22, 23].

A large number of epidemiological studies have been performed to assess the possible effects of prenatal radiation on the incidence of malignant disease. All are flawed by problems such as retrospectivity, small study size, inadequate or inappropriate case and control selection, and variability in the determination of radiation exposure and measurement of outcome parameters. The studies in which prenatal radiation exposure has been associated with an increased incidence of malignancy have found a relative risk ratio (RR) of all cancer of between 1.5 and 2.4 [24]. However, a number of other studies have failed to establish any statistically significant association between prenatal exposure to radiation and childhood malignancy [25, 26]. At least part of the increased risk associated with irradiation could be accounted for by the fact that mothers with a higher incidence of illness during pregnancy (a susceptibility which might be associated with an increased risk of tumour in their offspring) had a greater incidence of exposure to diagnostic radiation. Court Brown noted that amongst the offspring of the 750 pregnant women identified as having been irradiated in the first trimester, there were no cases of leukaemia with an average follow-up period of 6 years [27].

There continues to be debate about whether a radiation dose estimated to be approximately 10 mGy could give rise to cancer as suggested by the widely quoted Oxford Survey of Childhood Cancer based on a Linear No Threshold assumption [28]. More recent assessment based on biologic and epidemiologic data suggests this model may be incorrect and there is no carcinogenic risk below 20 mGy [29]. Despite this, the National Radiological Protection Board has adopted an estimated additional risk (EAR) coefficient for cancer incidence under 15 years of age following low dose irradiation *in utero* of 0.006% per mGy compared with a risk of 0.0018% per mGy for a dose received just after birth. Given a baseline risk of childhood cancer of between 1.0 to 2.5 per 1000, the following

Table 9.10.4 Excess risk of cancer from fetal radiation, based on a lowest risk versus highest risk estimation

Fetal dose (mGy)	Low risk model	High risk model
10	1 in 4545	1 in 1667
20	1 in 2272	1 in 834
30	1 in 1515	1 in 556
40	1 in 1136	1 in 417
50	1 in 909	1 in 334

estimates of risk reflect a low risk versus the highest risk modelling based on available data (see Table 9.10.4) [30].

Perhaps, a simpler statistic is that quoted by Tirada *et al.* of an EAR of cancer following a **50 mGy** *in utero* exposure of 1.1–3.0 cases per 1000 patient exposures [8]. The data regarding the risk at different gestations and on various systems and organs is imprecise with wide confidence intervals owing to the low doses and the limited sample sizes in available studies [31].

Contrast Agents

A number of X-ray and CT procedures require the administration of iodine containing contrast agents. There are **no** reports of teratogenesis from iodinated contrast agents but the amount of inorganic iodine available to interfere with thyroid metabolism is about 0.1% of the dose administered. Non-ionic contrast agents have been shown to cross the placenta and inhibit type II and III deiodinases which can reduce intracellular triiodothyronine. In addition, depending on the dose of iodine, there is theoretical risk of fetal thyroid blockade, although this has not been reported to date. Breastfeeding may continue after administration of iodinated contrast as <0.01% of the maternal dose is absorbed by the feeding infant [4]. The use of gadolinium, although not teratogenic, has generated concern as gadolinium has been shown to cross the placental barrier and will remain in the amniotic fluid indefinitely with a potential for the toxic gadolinium ion to dissociate from its chelate molecule raising concerns of possible secondary adverse effects [5]. Expert bodies have concluded that the effects of MRI contrast agents remains unknown and may be harmful. There is no requirement to cease breastfeeding following the administration of gadolinium to a lactating woman as <0.0004% is transferred through breast milk [4].

Contrast enhance ultrasound is a technique that has been used in specific conditions to improve diagnostic information. Although the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) have recommended against its use because of limited experience, small studies in human pregnancy have demonstrated no harm [32, 33].

Stabin has published comprehensive data regarding interruption or cessation of breastfeeding following administration of nuclear pharmaceuticals [16].

Imaging for Specific Endocrine Conditions

Thyroid and Parathyroid

Thyroid ultrasound is the major modality used for thyroid imaging in pregnancy. It can allow detection of structures as small as 2 mm in diameter, in the thyroid or parathyroid glands. It can also be used to assess gland texture and blood flow.

In pregnant women with goitre, ultrasound may identify diffuse enlargement, non-palpable thyroid nodules, or much more rarely, features of lymphoma. It may also be used to detect or monitor lymph node status in patients with thyroid nodules, or suspected or known thyroid cancer.

In chronic autoimmune thyroid disease including Hashimoto's thyroiditis, Graves' disease, and subacute thyroiditis, the thyroid may be diffusely enlarged and have low echogenicity. In a study of subjects referred for ultrasound with goitre, thyroid dysfunction, neck discomfort, and/or difficulty in swallowing, hypoechogenicity was associated with a positive predictive value for autoimmune thyroid disease of 88.3% (85–91%) and a negative predictive value of 93% (88–98%) [34].

Thyroid ultrasound is the most accurate tool for detecting thyroid nodules, determining their sonographic features and pattern, monitoring growth, and evaluating cervical lymph nodes [35]. Certain sonographic patterns can be used to help predict the risk of malignancy [36] but with variable accuracy [37]. The American College of Radiology has introduced an imaging staging system called the Thyroid Imaging Reporting and Data System (TIRADS) [38]. TIRADS is based on ultrasonographic characteristics in the nodule to predict the risk of malignancy

Suspicious features include solid hypoechoic nodules, microcalcifications, irregular margins, a taller than wide shape, and absence of elasticity. For example, a sonographic pattern that includes solid hypoechoic nodules with irregular borders and microcalcifications correlates with a >70% risk of malignancy compared to a very low suspicion pattern of a noncalcified mixed cystic solid or spongiform nodule which has <3% cancer risk. Although some experts have recommended triaging the need for FNA based on ultrasound features and nodule size [35], the results of thyroid ultrasonography do not correlate perfectly with histopathologic findings [39] and are not adequate to replace cytologic examination for nodules >1–1.5 cm [40].

Thyroid ultrasound may be useful to assess for the presence of a toxic nodules or autoimmune thyroid disease but will not differentiate the various autoimmune causes of hyperthyroidism. Colour flow doppler imaging (CFD) with measurement of peak systolic, end diastolic, and mean velocities of inferior thyroid artery may offer additional clinical information. In one study CD was found to have a sensitivity of 88.9% (67.2–96.9%) and specificity 87.5% (52.9–97.8%) to differentiate between Graves' disease (increased vascularity) and thyroiditis (low vascularity) when compared with thyroid scanning by technetium-99m pertechnetate (^{99m}Tc) [41]. In pregnancy, CFD is a cheap simple technique with no ionizing radiation exposure and is cost-effective in the diagnosis of thyrotoxicosis.

In pregnant women with Graves' disease, ultrasonography of the fetal thyroid gland is an excellent diagnostic tool that can facilitate assessment of fetal thyrotoxicosis or goitre [42, 43]. Ultrasound may also be used to monitor fetal growth and well-being as well as features of fetal thyroid hormone dysfunction. Different expert bodies have suggested fetal monitoring in all cases of Graves' disease [44] while the ATA guideline only recommends monitoring in women with uncontrolled hyperthyroidism in the second half of pregnancy, and in women with high thyrotropin receptor antibody (TRAb) levels detected at any time during pregnancy (greater than 3 × the upper limit of normal) [36].

Nuclear Scanning for Thyroid Disease in Pregnancy

Thyroid nuclear scans are useful in the assessment of thyrotoxicosis, thyroiditis, the assessment of thyroid nodules, and goitre. In view of the potential radiation risks to the fetus, and the availability of other diagnostic approaches, nuclear scanning of the thyroid is generally avoided in pregnancy [36]. However, inadvertent scans do occur in early pregnancy and the following data is helpful when counselling such women.

Scintigraphy utilizes one of the radioisotopes of iodine (usually ^{123}I) or ^{99m}Tc. Fetal dose estimates from maternal thyroid nuclear scanning vary depending on the technique used. Motavalli has calculated fetal estimated doses of 5.52 mGy in early pregnancy and 0.66 mGy in later pregnancy from ^{99m}Tc thyroid scanning and 0.31 mGy (early) and 0.51 mGy (late pregnancy) with ^{123}I thyroid scanning [15]. In all cases, ^{123}I will produce a significantly greater fetal thyroid dose if administered after 12 weeks and therefore ^{99m}Tc is the preferred technique if thyroid scanning is required. Imaging with ^{99m}Tc should be delayed until after the first trimester because of the significant decrease in fetal dose.

CT, MRI, and Positron Emission Tomography (PET) and Thyroid Cancer

Although ultrasound is the main imaging modality used for thyroid assessment, in certain circumstances further diagnostic imaging is required. Preoperative ultrasonography of the central and lateral neck identifies abnormal lymph nodes in as many as 20–30% of patients and this may alter the planned surgical procedure [45, 46]. However, ultrasonography can miss as many as 50% of the involved lymph nodes in the central neck because the overlying thyroid gland hinders adequate visualization. In addition, patients with lymphadenopathy or other evidence of extensive lymph node involvement may have nodal involvement in regions inadequately evaluated with ultrasound (e.g. mediastinal, infraclavicular, retropharyngeal, or parapharyngeal lymph node chains). In addition, if there are signs or symptoms of locally invasive disease or if ultrasound suggests extrathyroidal extension, further perioperative imaging is indicated.

In these circumstances, CT or MRI imaging with intravenous contrast of the central and lateral neck is recommended [35]. In pregnancy, this will generally be performed with MRI although the fetal radiation dose from CT in this region is very low at all stages of pregnancy. Although positron emission tomography (PET) with the radioisotope fludeoxyglucose (^{18}FDG) is sometimes used for the assessment of thyroid nodules outside of pregnancy it is not recommended during pregnancy [35].

Parathyroid Disease

If hyperparathyroidism is suspected in pregnancy, imaging for localization and surgical planning may be required. The safety and success of minimally invasive and selective techniques for parathyroid exploration is dependent on accurate preoperative localization of the affected gland. There is no universally accepted algorithm for imaging of the parathyroid glands, and the choice of imaging approach is largely based on surgeon preference and local expertise.

In pregnancy, neck ultrasonography is preferred for parathyroid localization. The ultrasound characteristics of parathyroid adenomas include homogeneous hypoechogenicity and an extrathyroidal feeding vessel with peripheral vascularity seen on CFD. Ultrasound is operator dependant but in experienced hands, the sensitivity is excellent with a reported ability to localize the adenoma to a specific quadrant of the neck in 87% of cases and to a specific side of the neck in 94%. Sensitivity was not reduced by the presence of nodular disease of the thyroid gland [47]. Disadvantages to the use of ultrasound alone include decreased accuracy in patients with smaller parathyroid gland size, obesity, or mediastinal glands located behind the clavicles [48].

In the non-pregnant population technetium-99m scintigraphy (^{99m}Tc sestamibi) scanning combined with single photon emission computed tomography (SPECT) has the highest positive predictive value for localizing parathyroid tumours **Table 9.10.5** [49]. The estimated of fetal radiation dose from this procedure can be calculated based on the administered dose and the estimated fetal dosimetry at that stage of pregnancy [50]. If nuclear imaging of the thyroid is the only appropriate technique, a medical physicist should be consulted to calculate the potential fetal dose and allow appropriate counselling. In addition to SPECT, a subtraction thyroid scan, or fusion with computed tomography images (MIBI-SPECT-CT) may be useful, especially if there is concurrent nodular thyroid disease [51]. Other modalities such as four-dimensional computed tomography, MRI, or PET-CT may be required but have not been reported in pregnancy [49].

Adrenal Disease/Masses

In the course of obstetric ultrasound, occasionally adrenal masses will be detected. In most cases, adrenal 'incidentalomas' are non-functioning adrenocortical adenomas, but may also represent conditions requiring therapeutic intervention (e.g. adrenocortical carcinoma, phaeochromocytoma, hormone-producing adenomas, or metastases). The European Society for Endocrinology (ESE) along with the European Network for the Study of Adrenal Tumours have recently published guidelines which recommended aiming to establish if an adrenal mass is benign or malignant at the time of initial detection [52].

Although the visualization of retroperitoneal structures is difficult on US, it still has a role in incidentally detecting adrenal masses and in monitoring changes during treatment of already established

Table 9.10.5 Diagnostic accuracy of imaging modalities for preoperative parathyroid localization. SPECT: sestamibi-single photon emission computed tomography; CT: computed tomography; MRI: magnetic resonance imaging; N/A: not applicable; MET-PET-CT scan: 11C-methionine positron emission tomography and computed tomography

Imaging modality	Sensitivity (%)	Positive predictive value (%)
^{99} Tc Sestamibi	71–79	72–95
^{99} Tc Sestamibi-SPECT	70–81	91–95
Ultrasound	64–91	83–96
4D-CT	83–95	88–99
MRI	40–85	N/A
11C-methionine PET/CT	79–90	93–94

pathology. Contrast-enhanced ultrasonography using phospholipid stabilized microbubbles filled with sulphur hexafluoride is comparable to CT and MRI for detection of adrenal malignancy with sensitivity of 100% and specificity of 82% [53]. Outside of pregnancy, the recommended imaging is with non-contrast CT assessing for tissue density. Studies have shown the optimal sensitivity (71%) and excellent specificity (98%) for differentiating benign from malignant adrenal masses is achieved using a threshold attenuation value of 10 Hounsfield units (HU) on non-contrast CT [54]. This implies that some 30% of benign adenomas have an attenuation value of >10 HU and will not be distinguishable from malignant lesions or phaeochromocytomas. In pregnancy, non-contrast CT of the adrenals would potentially result in a fetal dose of up to 10 mGy [16].

In pregnancy, assuming a low likelihood of malignancy, the ESE recommend MRI with chemical shift imaging, a non-contrast technique to assess lipid content. The normal adrenal has low to intermediate signal intensity on both T_1- and T_2-weighted sequence. During chemical shift imaging, a drop in signal intensity has sensitivity (81–100%) and specificity (94–100%) similar to unenhanced CT densitometry for the differentiation of incidental adrenal lesions (**Figure 9.10.1**) [55]. For lipid-rich adenomas, there is effectively no difference between CT and MRI. When evaluating lipid-poor adenomas, CSI might be superior although quantitative assessment of loss in signal intensity is not well standardized meaning the interpretation of the images might be more dependent on the experience of the radiologist. Accuracy may be improved by sampling the in-phase and opposed-phase echoes in the correct order and during the same breath-hold and using the first echo pair [56]. If malignancy is suspected, imaging should be optimized to enhance diagnosis including CT and or ^{18}FDG PET (estimated fetal dose 7 mSv) combined with CT (estimated fetal dose up to 10 mSv).

In the special case of phaeochromocytoma, imaging should only be performed where the clinical picture and biochemical findings are suggestive of this rare diagnosis. If suspected, localization is critical. About 95% are within the abdomen and pelvis, with only 10% being extra-adrenal, usually in the superior and inferior abdominal para-aortic areas. In pregnancy, ultrasound and MRI are again the preferred modalities although if not definitive, then multidetector CT or nuclear scanning may be required because of the critical nature of the diagnosis.

On MRI most phaeochromocytomas demonstrate high signal intensity on T_2-weighted images, described as a 'light bulb' and regarded as characteristic for phaeochromocytoma. (**Figure 9.10.2**) [57]. However, recent studies found that 30% of phaeochromocytomas demonstrate intermediate to low signal on T_2-weighted images or are inhomogeneous secondary to haemorrhagic [58]. Ultrasound has proved its accuracy in detecting phaeochromocytoma confined to the adrenal, as these are usually large and well-marginated masses. However, ultrasound has a limited role in detecting small adrenal tumours and extra-adrenal phaeochromocytomas such as those found in the retroperitoneum [53].

Nuclear medicine imaging can be used when CT or MRI does not demonstrate a mass in patients with high index of suspicion. Meta-iodo-benzylguanidine (MIBG) labelled with ^{131}I has sensitivity of 80–90% for detection of a phaeochromocytoma with a specificity of 90–100%. A minority (25%) of phaeochromocytomas are not detected by MIBG scanning but may be seen using ^{111}In octreotide.

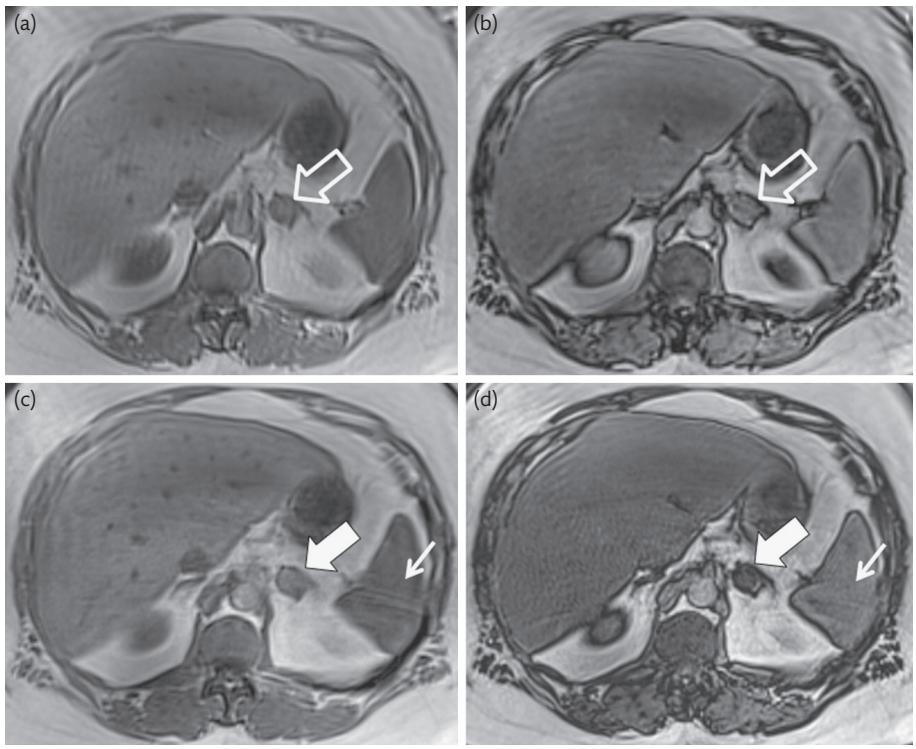


Figure 9.10.1 MRI in phase and out/opposed phase demonstrating a significant drop of signal intensity characteristic of an adrenal adenoma (solid arrow).

Reproduced with permission from Elsayes KM, Emad-Eldin S, Morani AC *et al*. Practical Approach to Adrenal Imaging. *Radiologic Clinics of North America*, 2017;55(2):279–301.

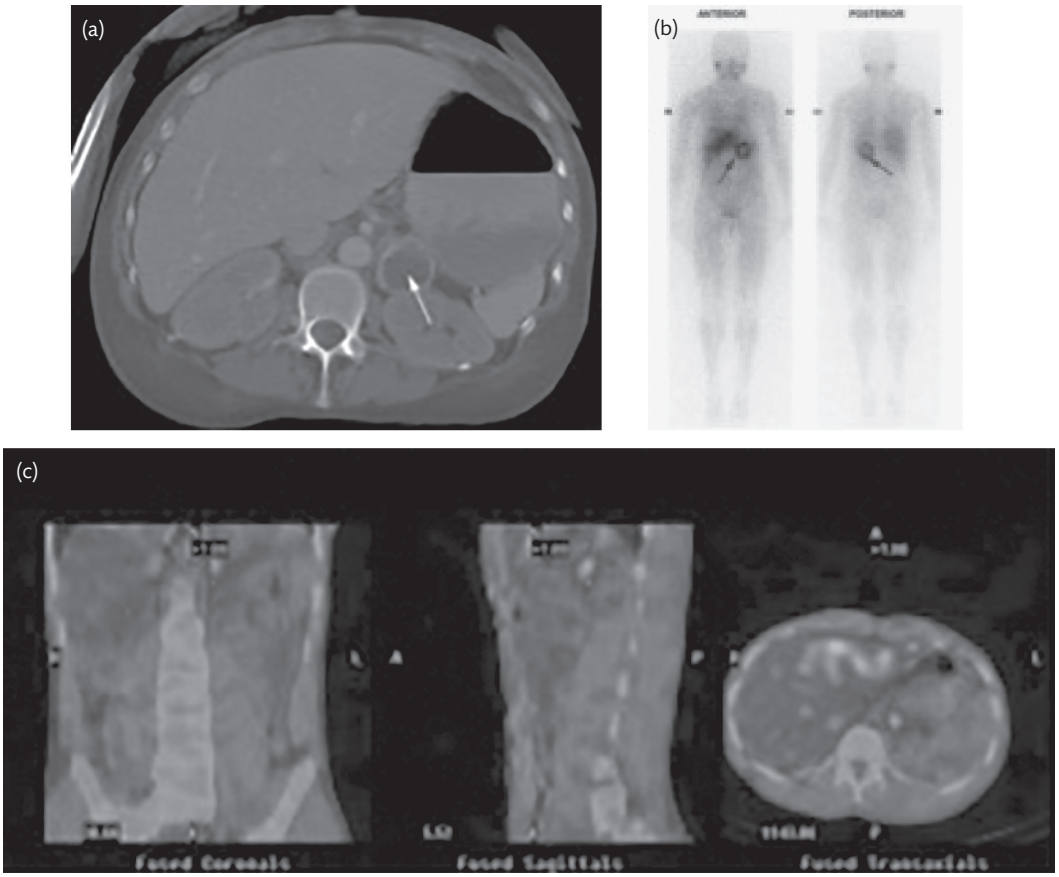


Figure 9.10.2 Various modalities demonstrating the same phaeochromocytoma (non-pregnant) (a) CT, (b) ^{123}I-MIBG, (c) SPECT
Reproduced with permission from Diagnostic Imaging Pathways-Phaeochromocytoma Government of Western Australia, 2012 [Available from: http://www.imagingpathways.health.wa.gov.au/index.php/component/content/article?id=201&tab=ct#images. (Accessed 28 April 2018). Copyright © 2019 Government of Western Australia. (ref 57).

For malignant phaeochromocytomas, ^{18}F-fluoro-dopamine positron emission tomography appears to be very helpful [53]. More recently, positron-emitting radiolabelled somatostatin analogues have emerged which, in combination with high-resolution PET integrated with CT (PET/CT), may improve the detection and staging of a variety of neuroendocrine tumours, including paraganglioma. Estimates of fetal radiation doses from these procedures have not been published but can be calculated if the procedures were considered essential.

In a systematic review of published cases of phaeochromocytoma in pregnancy, imaging modalities used during pregnancy included MRI 36%, ultrasonography 26%, MRI and ultrasonography 30% and CT plus one of the above 9% [59].

Pituitary Disease

The investigation of suspected pituitary masses, hypophysitis, tumours, or monitoring of existing pituitary tumours requires an appreciation of the normal anatomical changes of the pituitary throughout pregnancy. The size of the pituitary gland increases in three dimensions with a maximum volume in the first three postpartum days, returning to normal within 6 months [60]. Asymmetrical growth, deviation of the stalk, and height greater than 9–10 mm during pregnancy are not normal and should be further investigated [10]. The enlargement of the pituitary makes assessment of microadenomas more difficult. Monitoring of existing macroadenomas for growth, infarction, or haemorrhage is generally only performed if symptoms consistent with visual field impairment or, rarely, pituitary apoplexy occur.

The mainstay of imaging of the pituitary in pregnancy is non-contrast MRI. Normal pituitary tissue and most sella lesions, pituitary adenomas, and other tumours, have a signal that is similar to or slightly greater in intensity than that of central nervous system tissue. Cystic lesions, such as Rathke's cleft cysts, often have a low-intensity signal on T_1-weighted images and high intensity on T_2-weighted images. Haemorrhage into the pituitary gland results in a high-intensity signal on both T_1- and T_2-weighted images.

Imaging for Sheehan's syndrome will usually reveal an empty or partially empty sella with a normal sella size. Pituitary MRI is the most sensitive method of imaging, but CT may also be helpful. In the early stages, the pituitary gland shows non-haemorrhagic enlargement. Subsequently, the gland atrophies, and an empty sella develops.

Lymphocytic hypophysitis is characterized by diffuse, symmetrical pituitary enlargement, thickening of the pituitary stalk without deviation, and uniformly flat sella floor. Without contrast, the characteristic enhancement of the gland, and loss of the neurohypophyseal 'bright spot' will not be evident. MRI is also able to detect blood if pituitary haemorrhage is suspected. CT of the pituitary may be useful to assess calcification (e.g. with suspected craniopharyngiomas). Although MRI is the gold standard imaging

modality for the diagnosis of sella/suprasella lesions, some surgeons recommend additional imaging with thin-cut CT through the sella/sphenoid region and/or CT angiogram for preoperative planning and intraoperative navigation. Pituitary surgery is rarely required in pregnancy, but CT can safely be performed in the head as fetal radiation dose is minimal and in these circumstances, the use of iodinated contrast would be appropriate to optimize surgical planning [61, 62].

Multiple Endocrine Neoplasia and Pregnancy

Imaging for multiple endocrine neoplasia (MEN) syndromes may be required during pregnancy because of elevated biochemical markers or symptoms/signs of tumour (Figure 9.10.3). A number of these tumours may significantly impact pregnancy and urgent investigation is required if this is suspected. In the rare case when a MEN syndrome is suspected in pregnancy, multiple imaging may be required and should be performed in a timely manner. The specific imaging required will depend on the results of symptoms, signs, and biochemical testing, and multidisciplinary input is required.

Osteoporosis and Associated Syndromes

The standard screening test for osteoporosis is dual X-ray absorptiometry (DEXA). Studies (using phantoms) of DEXA of the posteroanterior lumbar spine/proximal femur during pregnancy generated estimated radiation doses of 1.7/2.7 mGy in the first trimester, 1/4/2.7 mGy in the second trimester, and 1.0/4.9 mGy in the third trimester. These were all well below the 10 mGy counselling threshold [63].

Investigation of suspected osteoporotic or stress fractures in pregnancy may require a range of imaging options. Reluctance to image pregnant patients due to concerns over radiation exposure is a major factor contributing to the delayed diagnosis of osteoporotic related fractures in pregnancy. In transient osteoporosis of pregnancy, plain X-rays may be normal initially, with a lag of weeks to months before osteopenia can be demonstrated. Osteopenia differentiates transient osteoporosis and other bone marrow oedema syndromes. Plain X-ray provides an overview of the region of concern and assessment of alignment, and allows screening for other causes of pain such as bone tumour.

MRI can be used for safe and reliable imaging in situations where multiple x-rays are required (e.g. for oedema), or if the initial X-rays are normal but there is suspected non-displaced fracture [64–65]. Although altered gadolinium enhancement may be seen in osteonecrosis of the femoral head (OFNH), the use of gadolinium is not recommended in pregnancy.

A bone scan with ^{99m}Tc m-methylene diphosphonate (MDP) may give a fetal absorbed dose up to 4.7 mSv in early pregnancy but much less in late pregnancy (1.8 mSv) and if clinically indicated, should be performed [16]. One of the main indications for further imaging in the setting of pregnancy is to differentiate hip pain due to transient osteoporosis of pregnancy (TOP) from ONFH. In early stages of the disease, X-ray findings will be indistinguishable [66]. ONFH is an idiopathic form of avascular necrosis which exhibits lack of uptake on bone scan and may show a characteristic 'double line sign' on MRI. In contrast, MRI of TOP shows diffuse homogenous marrow oedema of the involved segment. The appearance is not specific with radiological differentials including transient bone marrow oedema (without osteoporosis), osteonecrosis, infection, and tumour infiltration. If diagnosed early, ONFH may be amenable to a core decompression or revascularization procedures to avoid progression to arthritis.

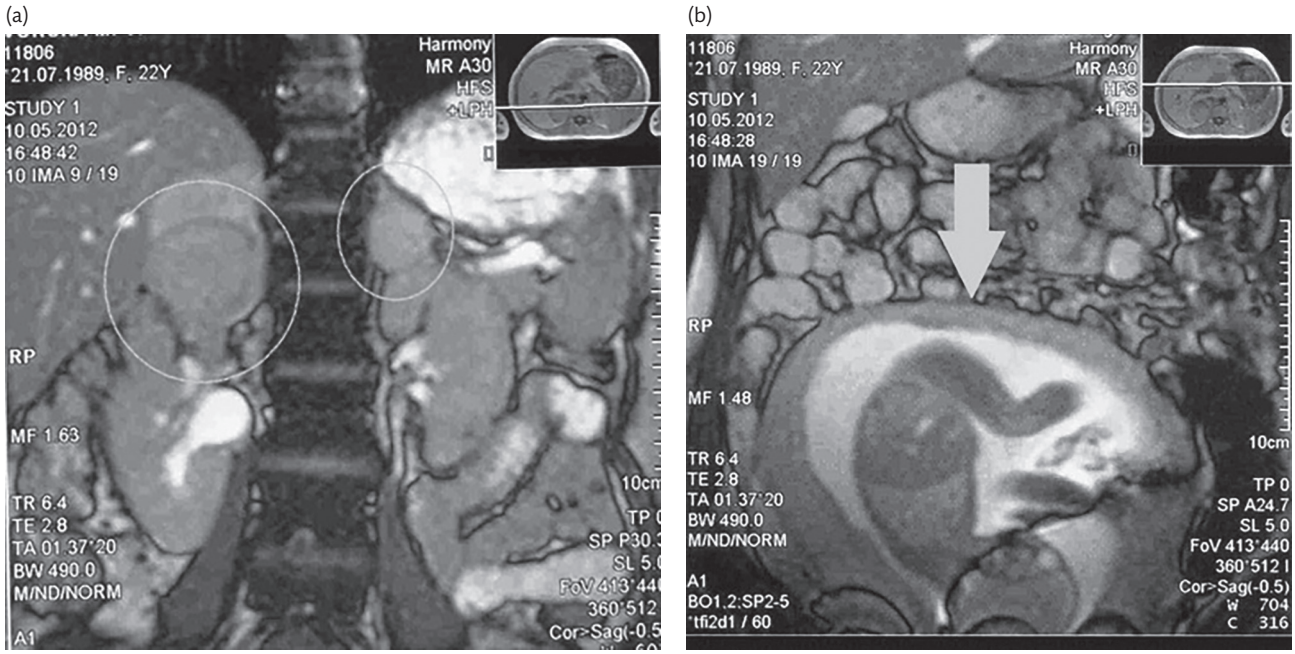


Figure 9.10.3 MRI of a women with MEN 2A at 27 weeks' gestation demonstrating bilateral adrenal tumours. These were proven to be phaeochromocytomas measuring 4 × 2.3cm on the left and 6.4 x 4.4 cm on the right. Below is an MRI of her fetus at the same examination.

Counselling for Diagnostic Imaging for Endocrine Conditions in Pregnancy

Informed consent for X-ray, CT, MRI, or nuclear scanning will be required if these investigations are ordered during pregnancy or breastfeeding. Counselling should be provided initially by the clinician ordering the investigation and then by the radiologist/nuclear physician performing the procedure. All involved allied staff must also be well informed to ensure the patient receives a consistent message about the risks and benefits of the proposed test(s).

Whether particular institutions use written consent forms or verbal consent, this interaction should be documented in the patient's medical record and in compliance with local law. The American College of Radiology provides a sample consent from which may be useful to your Institution [21].

For practical purposes, no specific counselling is required for women undergoing diagnostic imaging with a predicted fetal absorbed dose of less than 10 mGy. This includes all X-ray and CT scanning not involving the abdomen and most nuclear scans. For direct exposures or nuclear scanning with a potential exposure >10 mGy, the women should be counselled on a risk/benefit basis. The specific risk appears to be childhood malignancy, but, as described earlier, for each 10 mGy exposure, theoretical projections based on a LNT assumption indicate a likely **maximum risk** of 1 additional case of childhood cancer (not death) per 1700 exposures of 10 mGy [23]. This must be balanced against the benefit of the imaging or treatment in terms of management of the maternal condition.

There are no diagnostic imaging procedures that can be considered a risk factor for genetic damage, malformation, or neurodevelopmental effects based on current knowledge. It is particularly important to liaise with the radiologist or nuclear physician to ensure the most appropriate imaging is performed to obtain maximal information with minimal fetal absorbed dose. If possible, diagnostic imaging should be delayed until after delivery if the information is not likely to alter immediate management.

Summary

The detailed review of medical imaging in pregnancy given in this chapter should prevent the reflex 'no radiation' approach to selection of diagnostic imaging modalities in pregnancy. As in the non-pregnant subject, the principle of 'as low as reasonably achievable' (ALARA) should be applied in the pregnant subject, ensuring that non-ionizing radiation options are considered first. In almost all cases, radiation and contrast risks are negligible and therefore the most sensitive and specific imaging should be performed at the most appropriate time, whether during or after pregnancy where applicable. However, when X-ray, CT, nuclear scan, or ET are the most suitable modality; these should be applied with appropriate care and counselling. As pointed out by Zanotti-Fregonara, 'pregnant women should be imaged using the standard imaging protocols used for any other patient' [17]. The American College of Radiology particularly note that 'It is best to create written protocols for all imaging of pregnant patients to avoid reactive, non-optimal protocol adjustments by physicians attempting to reduce radiation exposures' [21]. If dose reduction strategies that do not compromise image quality are available for pregnant women, then these strategies should be used for all patients [17].

Endocrine conditions during pregnancy may require a range of imaging techniques but with care and consideration, an appropriate diagnostic and management plan can be achieved if required.

REFERENCES

1. USEPA. Radiation Sources and Doses. Available at: https://www.epa.gov/radiation/radiation-sources-and-doses (accessed: 28 April 2018).
2. Ciet P, Litmanovich, DE. MR safety issues particular to women. *Magn Res Imaging Clin*, 2015; **23**: 59–67.
3. Ray J, Vermeulen, MJ, Bharatha, A, *et al*. Association between MRI exposure during pregnancy and fetal and childhood outcomes. *JAMA*, 2016; **316**: 952–61.
4. Webb J, Thomsen, HS, Morcos, SK. The use of iodinated and gadolinium contrast media during pregnancy and lactation. *Eur Radiol*, 2005; **15**: 1234–40.
5. Kanal E, Barkovich AJ, Bell C, *et al*. ACR guidance document for safe MR practices: 2007. *AJR Am J Roentgenol*, 2007; **188**: 1447–74.
6. Schauer DA, Linton OW. National Council on Radiation Protection and Measurements report shows substantial medical exposure increase. *Radiology*, 2009; **253**: 293–6.
7. World Nuclear Association. Nuclear radiation and health effects, 2016. Available at: http://www.world-nuclear.org/information-library/safety-and-security/radiation-and-health/nuclear-radiation-and-health-effects.aspx (accessed: 28 April 2018).
8. Tirada N, Dreizin D, Khati NJ, *et al*. Imaging pregnant and lactating patients. *Radiographics*, 2015; **35**: 1751–65.
9. IRCP. The 2007 Recommendations of the International Commission on Radiological Protection, 2007. Available at: https://www.icrp.org/publication.asp?id=ICRP%20Publication%20103 (accessed: 5 April 2020).
10. National Council on Radiation Protection and Measurements: risk estimates for radiation protection. Bethesda, 1993 [Report Number 115]. Available at: https://ncrponline.org/shop/reports/report-no-115-risk-estimates-for-radiation-protection-1993/ (accessed: 5 April 2020).
11. ICRP. Pregnancy and Medical Radiation. *Ann ICRP*, 2000; **84**: 1–43.
12. McCollough CH, Schueller BA, Atwell TD, *et al*. Radiation exposure and pregnancy: when should we be concerned? *Radiographics*, 2007; **27**: 909–17.
13. Wall B, Meara J, Muirhead C, *et al*. Protection of pregnant patients during diagnostic medical exposures to ionising radiation. RCE-9. Health Protection Agency, 2009. Available at: https://www.rcr.ac.uk/publication/protection-pregnant-patients-during-diagnostic-medical-exposures-ionising-radiation (accessed: 28 April 2018).
14. Bural GG, Laymon CM, Mountz JM. Nuclear imaging of a pregnant patient: should we perform nuclear medicine procedures during pregnancy? *Mol Imaging Radionucl Ther*, 2012; **21**: 1–5.
15. Motavalli LR, Hakimabad HM, Azghadi EH. Fetal and maternal dose assessment for diagnostic scans during pregnancy. *Phys Med Biol*, 2016; **61**: 3596–608.
16. Stabin M. New generation fetal dose estimates for radiopharmaceuticals. *J Nucl Med*, 2017; **59**: 1005–6.
17. Zanotti-Fregonara P, Hindie E. Performing nuclear medicine examinations in pregnant women. *Phys Med*, 2017; **43**: 159–64.

18. Stabin MG, Xu XG, Emmons MA, *et al.* RADAR reference adult, pediatric, and pregnant female phantom series for internal and external dosimetry. *J Nucl Med*, 2012; **53**: 1807–13.
19. Stabin MG, Tagesson M, Thomas SR, *et al.* Radiation dosimetry in nuclear medicine. *Appl Radiat Isot*, 1999; **50**: 73–87.
20. Wong KK, Gandhi A, Viglianti BL, *et al.* Endocrine radionuclide scintigraphy with fusion single photon emission computed tomography/computed tomography. *World J Radiol*, 2016; **8**: 635–55.
21. Wagner LK, Applegate K. ACR-SPR practice parameter for imaging pregnant or potentially pregnant adolescents and women with ionizing radiation, 2013. Available at: https://www.acr.org/-/media/ACR/Files/Practice-Parameters/pregnant-pts.pdf (accessed: 28 April 2018).
22. Tharmalingam S, Sreetharan S, Kulesza AV, *et al.* Low-dose ionizing radiation exposure, oxidative stress and epigenetic programing of health and disease. *Radiat Res*, 2017; **188**: 525–38.
23. Valentin J. Biological effects after prenatal irradiation (embryo and fetus): ICRP Publication 90 Approved by the Commission in October 2002. *Ann ICRP*, 2003; **33**: 1–206.
24. Doll R, Wakeford R. Risk of childhood cancer from fetal irradiation. *Br J Radiol*, 1997; **70**: 130–9.
25. Naumburg E, Bellocco R, Cnattingius S, Hall P, Boice Jr JD, Ekbom A. Intrauterine exposure to diagnostic X-rays and risk of childhood leukemia subtypes. *Radiat Res*, 2001; **156**: 718–23.
26. Meinert R, Kaletsch U, Kaatsch P, Schüz J, Michaelis J. Associations between childhood cancer and ionizing radiation: results of a population-based case-control study in Germany. *Cancer Epidemiol Biomarkers Prev*, 1999; **8**: 793–9.
27. Brown WC, Doll R, Hill AB. Incidence of leukaemia after exposure to diagnostic radiation in utero. *BMJ*, 1960; **2**: 1539.
28. Wakeford R. Childhood leukaemia following medical diagnostic exposure to ionizing radiation in utero or after birth. *Radiat Prot Dosim*, 2008; **132**: 166–74.
29. Tubiana M, Feinendegen LE, Yang C, Kaminski JM. The linear no-threshold relationship is inconsistent with radiation biologic and experimental data. *Radiology*, 2009; **251**: 13–22.
30. Coakley FV, Cody DD, Mahesh M. The pregnant patient: alternatives to CT and dose-saving modifications to CT technique, 2010. Available at: https://www.imagewisely.org/imaging-modalities/computed-tomography/imaging-physicians/articles/the-pregnant-patient (accessed: 28 April 2018).
31. Streffer C, Shore R, Konermann G, *et al.* Biological effects after prenatal irradiation (embryo and fetus). A report of the International Commission on Radiological Protection. *Ann ICRP*, 2003; **33**: 5–206.
32. Piscaglia F, Nolsøe C, Dietrich CA, *et al.* The EFSUMB guidelines and recommendations on the clinical practice of contrast enhanced ultrasound (CEUS): update 2011 on non-hepatic applications. *Ultraschall Med*, 2012; **33**: 33–59.
33. Murotsuki J. Contrast-enhanced ultrasound in obstetrics and gynecology. *Donald School J Ultrasound Obstet Gynecol*, 2007; **1**: 16–19.
34. Pedersen OM, Aardal NP, Larssen TB, *et al.* The value of ultrasonography in predicting autoimmune thyroid disease. *Thyroid*, 2000; **10**: 251–9.
35. Haugen BR, Alexander EK, Bible KC, *et al.* 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*, 2016; **26**: 1–133.
36. Alexander EK, Pearce EN, Brent GA, *et al.* 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid*, 2017; **3**: 315–89.
37. Remonti LR, Kramer CK, Leitao CB, *et al.* Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid*, 2015; **25**: 538–50.
38. Tessler FN, Middleton WD, Grant EG, *et al.* ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol*, 2017; **14**: 587–95.
39. Singaporewalla RM, Hwee J, Lang TU, *et al.* Clinico-pathological correlation of thyroid nodule ultrasound and cytology using the TIRADS and Bethesda classifications. *World J Surg*, 2017; **41**: 1807–11.
40. Abalovich M, Amino N, Barbour LA, *et al.* Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2007; **92**(8 Suppl): S1–7.
41. Donkol RH, Nada AM, Boughattas S. Role of color Doppler in differentiation of Graves' disease and thyroiditis in thyrotoxicosis. *World J Radiol*, 2013; **5**: 178–83.
42. Luton D, Le Gac I, Vuillard E, *et al.* Management of Graves' disease during pregnancy: the key role of fetal thyroid gland monitoring. *J Clin Endocrinol Metab*, 2005; **90**: 6093–8.
43. Cohen O, Pinhas-Hamiel O, Sivan E, *et al.* Serial in utero ultrasonographic measurements of the fetal thyroid: a new complementary tool in the management of maternal hyperthyroidism in pregnancy. *Prenat Diagn*, 2003; **23**: 740–2.
44. De Groot L, Abalovich M, Alexander EK, *et al.* Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 2543–65.
45. Stulak JM, Grant CS, Farley DR, *et al.* Value of preoperative ultrasonography in the surgical management of initial and reoperative papillary thyroid cancer. *Arch Surg (Chicago, Ill: 1960)*, 2006; **141**: 489–94.
46. Kouvaraki MA, Shapiro SE, Fornage BD, *et al.* Role of preoperative ultrasonography in the surgical management of patients with thyroid cancer. *Surgery*, 2003; **134**: 946–54.
47. Gilat H, Cohen M, Feinmesser R, *et al.* Minimally invasive procedure for resection of a parathyroid adenoma: the role of preoperative high-resolution ultrasonography. *J Clin Ultrasound*, 2005; **33**: 283–7.
48. Berber E, Parikh RT, Ballem N, *et al.* Factors contributing to negative parathyroid localization: an analysis of 1000 patients. *Surgery*, 2008; **144**: 74–9.
49. Yip L SM, El-Hajj Fuleihan G. Preoperative localization for parathyroid surgery in patients with primary hyperparathyroidism. UpTo Date. Available at: https://www.uptodate.com/contents/preoperative-localization-for-parathyroid-surgery-in-patients-with-primary-hyperparathyroidism (accessed: 28 April 2018).
50. Greenspan BS, Dillehay G, Intenzo C, *et al.* SNM practice guideline for parathyroid scintigraphy 4.0. *J Nucl Med Technol*, 2012; **40**: 111–18.
51. Shafiei B, Hoseinzadeh S, Fotouhi F, *et al.* Preoperative 99mTc-sestamibi scintigraphy in patients with primary hyperparathyroidism and concomitant nodular goiter: comparison of SPECT-CT, SPECT, and planar imaging. *Nucl Med Commun*, 2012; **33**: 1070–6.

52. Fassnacht M, Arlt W, Bancos I, *et al*. Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*, 2016; **175**: G1–G34.
53. Das CJ, Baruah MP, Baruah UM. Radiological imaging in endocrine hypertension. *Indian J Endocrinol Metab*, 2011; **15**(Suppl 4): S383.
54. Boland GW, Lee MJ, Gazelle GS, Halpern EF, McNicholas MM, Mueller PR. Characterization of adrenal masses using unenhanced CT: an analysis of the CT literature. *Am J Roentgenol*, 1998; **171**: 201–4.
55. Schieda N, Al Dandan O, Kielar AZ, *et al*. Pitfalls of adrenal imaging with chemical shift MRI. *Clin Radiol*, 2014; **69**: 1186–97.
56. Elsayes KM, Emad-Eldin S, Morani AC, *et al*. Practical approach to adrenal imaging. *Radiol Clin N Am*, 2017; **55**: 279–301.
57. Diagnostic Imaging Pathways—phaeochromocytoma. Government of Western Australia, 2012. Available at: http://www.imagingpathways.health.wa.gov.au/index.php/component/content/article?id=201&tab=ct#images (accessed: 28 April 2018).
58. Blake MA, Cronin CG, Boland GW. Adrenal imaging. *Am J Roentgenol*, 2010; **194**: 1450–60.
59. Biggar MA, Lennard TW. Systematic review of phaeochromocytoma in pregnancy. *Br J Surg*, 2013; **100**: 182–90.
60. Karaca Z, Tanriverdi F, Unluhizarci K, *et al*. Pregnancy and pituitary disorders. *Eur J Endocrinol*, 2010; **162**: 453–75.
61. Chen CC, Carter BS, Wang R, *et al*. Congress of neurological surgeons systematic review and evidence-based guideline on preoperative imaging assessment of patients with suspected nonfunctioning pituitary adenomas. *Neurosurgery*, 2016; **79**: E524–6.
62. Donatini G, Kraimps JL, Caillard C, *et al*. Pheochromocytoma diagnosed during pregnancy: lessons learned from a series of ten patients. *Surg Endos*, 2018; **32**: 3890–900.
63. Damilakis J, Perisinakis K, Vrahoriti H, *et al*. Embryo/fetus radiation dose and risk from dual X-ray absorptiometry examinations. *Osteoporos Int*, 2002; **13**: 716–22.
64. ACOG. ACOG Committee Opinion No. 723: guidelines for diagnostic imaging during pregnancy and lactation. *Obstet Gynecol*, 2017; **130**: E210–16.
65. Lowe SA. Diagnostic radiography in pregnancy: risks and reality. *Aust N Z J Obstet Gynaecol*, 2004; **44**: 191–6.
66. Guerra JJ, Steinberg ME. Distinguishing transient osteoporosis from avascular necrosis of the hip. *J Bone Joint Surg*, 1995; **77**: 616–24.

SECTION 10

Male Reproductive Endocrine Disorders

10.1

Normal Male Reproductive Endocrinology

10.1.1 Endocrine and Local Regulation of Testicular Hormone and Sperm Production

Ilpo Huhtaniemi and Jorma Toppari

Introduction

The two main functions of the testis are androgen production and spermatogenesis, and the key role in their **endocrine** regulation is played by the two pituitary gonadotrophins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Other hormones and growth factors also influence testicular function, often by modulating gonadotrophin action. In addition, a plethora of other bioactive molecules mediate regulatory signals within and between the testicular cell compartments. They can be **paracrine** (between two dissimilar neighbouring cells), **autocrine** (the same or a similar cell is the origin and target of the regulating factor) or **intracrine** (the factor functions within the cell of its synthesis without being secreted). The main testicular hormone is testosterone (T), a product of Leydig cells located in the testicular interstitial cell compartment. T regulates spermatogenesis that takes place in testicular seminiferous tubules in paracrine and indirect fashion, by stimulating the functions of Sertoli cells in the adjacent seminiferous tubules. Following its secretion into the peripheral circulation, T exerts its endocrine actions, androgenic and anabolic, in various extratesticular tissues.

Endocrine Regulation of Testicular Function

The Hypothalamic–Pituitary–Testicular (HPT) Axis

The HPT axis is a typical endocrine regulatory circuit with hierarchical cascades of forward and feedback regulatory events (Figure 10.1.1.1). According to the classical concept, specific hypothalamic nuclei synthesize the decapeptide gonadotrophin-releasing hormone (GnRH). Axon terminals of the GnRH neurons make contact in the median eminence with hypophyseal portal vessels, which transport GnRH to the anterior pituitary gland where it stimulates the synthesis and secretion of LH and FSH [1, 2]. These two gonadotrophins are the key endocrine regulators of testicular function.

Several other hypothalamic hormones regulate the maturation and fine-tuning of GnRH neurons [3], including glutamate and γ-aminobutyric acid (GABA), and a plethora of neuropeptides such as neuropeptide Y, galanin-like peptide, opioid peptides, and orexins. Most importantly, however, it has been recently discovered that the GnRH neurons are under regulation of an upstream neuronal system where the gatekeeper role is played by the neuromodulatory peptide kisspeptin, encoded by the *KISS1* gene (Figure 10.1.1.1) [4]. Some kisspeptin neurons coexpress neurokinin B and dynorphin, and are hence collectively termed kisspeptin-neurokinin B-dynorphin (KNDy) neurons [5]. Kisspeptin functions through a G-protein-coupled receptor (GPCR) originally called GPR54, but now termed

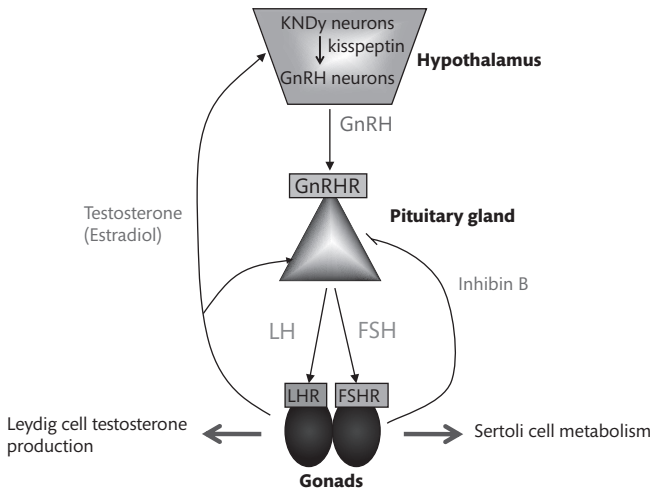


Figure 10.1.1.1 The hypothalamic–pituitary–testicular (HPT) axis. The main hormones functioning in the axis are depicted, as well as the functions of LH and FSH in the testis. GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; FSH, follicle-stimulating hormone; KNDy, kisspeptin/neurokinin B/dynorphin; inhibitory signal, ⟍, stimulatory signal ⟶.

KISS1R, expressed in GnRH neurons, where its activation stimulates *GnRH* expression.

The *GnRH* gene encodes a propeptide that is cleaved into a 24-amino acid signal peptide, the GnRH decapeptide, and a 56 amino acid GnRH-associated peptide. GnRH is secreted in short pulses of varying amplitude, in adult men with 1–2-hour interval. A recent study has demonstrated that kisspeptin neurons in the hypothalamic arcuate nucleus may be the anatomic site of the GnRH pulse generator [6].

GnRH interacts in the anterior pituitary gonadotroph cells with a high-affinity receptor (R) belonging to the GPCR gene family. Pulsatile GnRH secretion is vital for GnRHR activation, since tonic GnRH stimulation (e.g. during GnRH agonist treatment) downregulates the GnRHR signalling, thus suppressing gonadotrophin synthesis and release. The secretory peaks of LH are more distinct, due to their shorter half-life than those of FSH in the systemic circulation. Pulsatile release, however, is not important for gonadotrophin action at the gonadal level.

Both LH and FSH are glycoproteins with molecular mass of 30–40 kDa. They are synthesized in pituitary gonadotroph cells as heterodimers of the common α-subunit and a hormone-specific β-subunit (**Figure 10.1.1.2a**). More extensive glycosylation of FSH makes its half-life longer vis-à-vis LH (3–4 h vs. 20 min). In addition, the carbohydrate side chains in LH are heavily sulphated (50%), and there is a specific hepatic receptor for sulphated glycoproteins accelerating their elimination [7].

Like their ligands, the receptors (R) for LH and FSH are structurally related glycoproteins with a molecular mass of about 80 kDa, and belong to class A rhodopsin-like GPCRs (**Figure 10.1.1.2b**) [8, 9]. The molecular events mediating the actions of LH and FSH on Leydig and Sertoli cells, respectively, are in principle similar [8, 9]. The main, and most extensively studied signalling mechanism uses the classical adenylyl cyclase/cAMP/G-protein cascade, but other signalling mechanisms are also employed in gonadotrophin action; further details are presented in **Figure 10.1.1.3a**.

In Leydig cells, cell differentiation, growth, and steroidogenesis entail the important responses to LH stimulation. FSH function in the testis is age-dependent, and in prenatal and prepubertal life, it stimulates Sertoli cell proliferation. This influences fertility, since the Sertoli cell number correlates with total length of the seminiferous epithelium and testicular sperm production capacity. In adulthood FSH maintains the functions of Sertoli cells thus indirectly supporting spermatogenesis.

The reverse side of the HPT function is the negative (inhibitory) feedback of gonadal steroid and peptide hormones at the hypothalamic and pituitary levels, to maintain a physiological balance of gonadotrophin secretion (**Figure 10.1.1.1**) [1, 2]. T, partly after conversion to oestradiol, exerts its negative feedback action on GnRH secretion indirectly through KNDy neurons, by suppressing kisspeptin and neurokinin B release, since GnRH neurons do not express sex steroid receptors [4]. Sex steroids also directly suppress gonadotrophin synthesis in the pituitary gland [10]. Although testicular steroids also regulate FSH, this hormone is mainly under negative control at the pituitary level by Sertoli cell-derived inhibin B, a glycopeptide hormone of the TGF-β family (see next) (**Figure 10.1.1.1**).

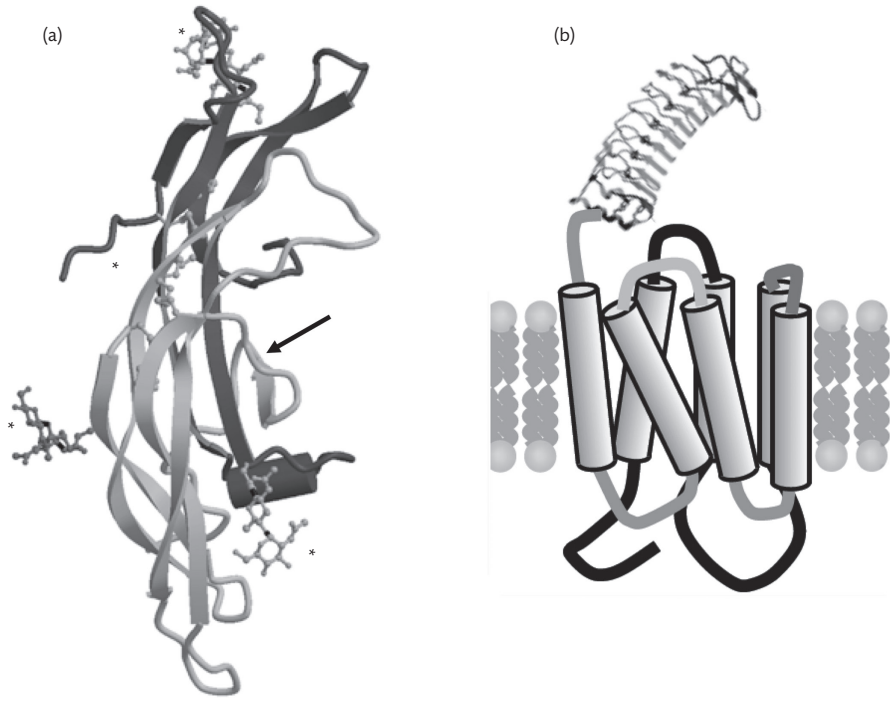


Figure 10.1.1.2 Schematic 3D structures of LH (a) and LHR (b). In panel A black colour depicts the α-subunit (CGA) and dark grey colour the β-subunit (LHB). The arrow points to the 'seat belt' structure of the β-subunit, which stabilizes the dimeric structure of the hormone. The asterisks depict the carbohydrate moieties attached to the hormone molecule. In panel B, the structure in the top part depicts the leucine rich repeats of the extracellular receptor domain, which forms the primary binding site of the hormone. The cylindrical structures in the middle of the picture depict the seven transmembrane domains of the receptor molecule. *For a colour version of this figure, please see colour plate section.*

Figures are courtesy of Drs. J. Dias and E. Reiter.

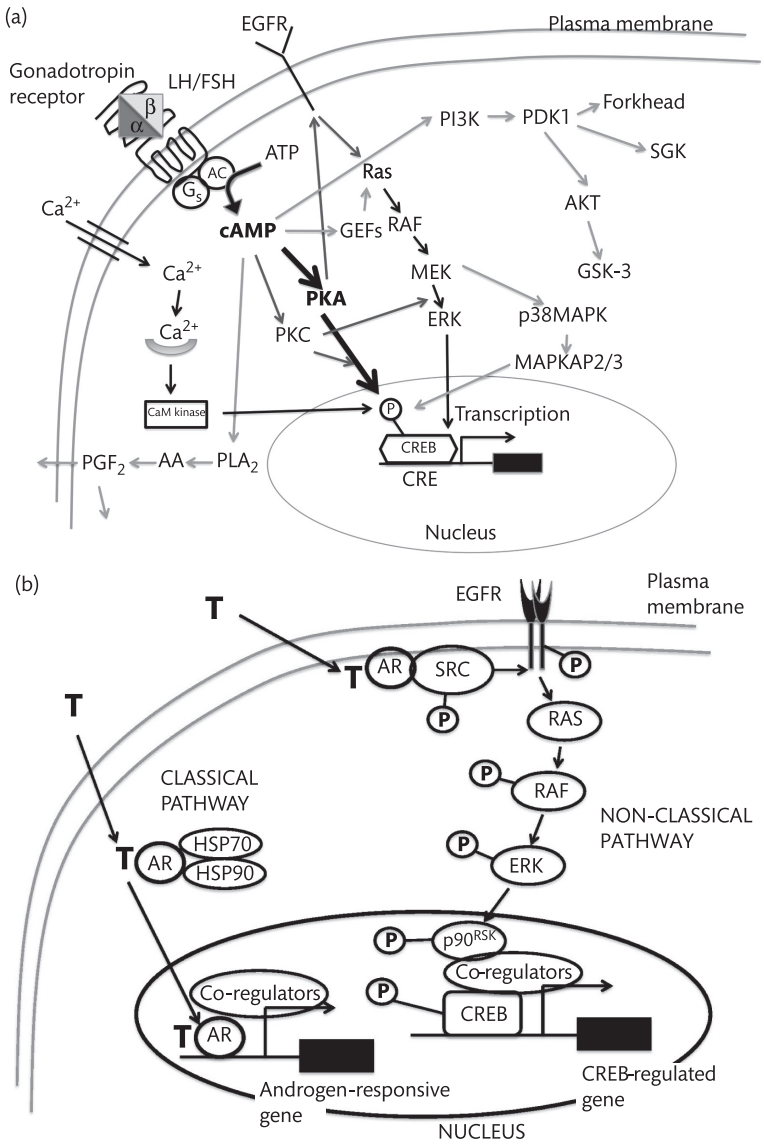


Figure 10.1.1.3 The main signalling cascades employed in LH and FSH action. Black arrows depict pathways common for both hormones, red arrows those specific for LH and light grey arrows those specific for FSH. AA, arachidonic acid; AC, adenylyl cyclase; ATP, adenosine trisphosphate; CaM, calmodulin; cAMP, cyclic adenosine-3':5'-monophosphate; CRE, cAMP response element; CREB, cAMP response element binding protein; EGFR, epidermal growth factor receptor; ERK, extracellular signal-regulated kinase; Forkhead, forkhead transcription factor; GEF, guanine nucleotide exchange factor; G_s, stimulatory guanine nucleotide binding protein; GSK-3, glycogen synthase kinase-3; MEK, mitogen/extracellular signal-regulated kinase; PDK1, phosphoinositide-dependent protein kinase; MAPK, mitogen activated protein kinase; MAPKAP, MAPK activated protein kinase; PGF_2, prostaglandin F_2; PI3K, phosphoinositide 3 kinase; PKA, protein kinase A; PKC, protein kinase C; PLA_2, phospholipase A_2; RAF, rapidly accelerated fibrosarcoma—Ser/Thre specific protein kinase; Ras, small GTPase family; SGK, serum glucocorticoid-regulated kinase. *For a colour version of this figure, please see colour plate section.*

Effects of Other Hormones on Testicular Function

Besides LH and FSH, several other circulating hormones affect testicular function. Since many 'extragonadal' hormones are also synthesized within the testis, it is difficult to delineate whether their testicular actions, mostly documented *in vitro*, are endocrine or para/autocrine *in vivo*.

Prolactin regulates Leydig cell function in rodents, but whether it has direct effects on the human testis remains unclear. Hyperprolactinaemia in man impairs testicular function, most likely indirectly through inhibitory action of prolactin on gonadotrophin secretion (see chapters in Section 10.3.2, 'Adult Hypogonadism: Aetiology').

Growth hormone (GH) stimulates the formation of insulin-like growth factor-I in the testis. Evidence for testicular effects of GH/IGF-1 comes from observations in rodents that GH deficiency or resistance are associated with delayed puberty and poor Leydig cell function.

Insulin receptors are found in Leydig cells, insulin and LH reciprocally upregulate the receptor of each other, and insulin augments basal and LH-stimulated Leydig cell steroidogenesis [11].

Thyroid hormones have an important role in the maintenance of Leydig cell steroidogenesis [12]. They increase, in additive fashion with LH, the level of steroidogenic acute regulatory protein (StAR) expression in Leydig cells. Without optimal thyroid hormone action, the steroidogenic capacity of Leydig cells is severely compromised.

Two hormones important in bone metabolism, osteocalcin, and vitamin D, may also have direct testicular effects. The former stimulates Leydig cell T production in rodents [13], and the latter may improve semen quality in humans [14].

Steroid hormones also modulate Leydig cell functions. Androgen and oestrogen receptors are expressed in Leydig cells, and thus these hormones can have para/autocrine regulatory effects in their site of synthesis [15, 16]. Glucocorticoids of adrenal origin also influence the endocrine regulation of the testis [17]. High systemic glucocorticoid levels, for example, in Cushing's syndrome and during physical and mental stress, suppress Leydig cell androgen production by inhibiting the conversion of cholesterol to steroid hormones.

Despite direct testicular actions of numerous blood-borne non-gonadotrophic hormones, their physiological role remains obscure and the dogma still prevails that gonadotrophins provide the main driving force for testicular function. The role of the other regulators is likely to fine-tune gonadotrophin actions.

Testicular Hormone Production

Testicular Steroidogenesis

Testicular steroidogenesis takes place in Leydig cells, and the most important hormone produced is T. LH is the key regulator of testicular steroidogenesis, but its action can be modulated by numerous other endocrine, paracrine, and autocrine factors. Most details of the synthesis, secretion, and metabolism of testicular steroid hormones were unravelled in the 1950s and 1960s. More recent studies on androgens have concerned identification and characterization of the steroidogenic enzymes at genome and protein level, mechanisms of cholesterol supply for steroid biosynthesis, and mechanisms of androgen action (see Chapter 10.1.2, 'Sex Steroid Actions in the Male').

Androgens are essential for all masculine functions of the body, including sexual differentiation, development of secondary sex characteristics, growth of sex organs, spermatogenesis, the masculine features of the musculo–skeletal apparatus, and sexual behaviour (**Box 10.1.1.1**) [18]. Essential for all of them is a single regulatory step, the binding of androgen to its receptor, which then evokes the variety of structural and functional responses specific to each target tissue.

The synthesis of all steroid hormones starts from cholesterol, for which the Leydig cells have multiple sources: (a) *de novo* synthesis from acetyl coenzyme A; (b) stored cholesteryl esters; (c) exogenous lipoprotein-supplied cholesterol; and (d) plasma membrane-derived cholesterol following hormonal stimulation [19]. The most utilized source are the plasma low- and/or high-density lipoprotein–cholesterol complexes, endocytosed by receptor-mediated mechanisms. Cholesterol is thereafter esterified and stored in intracellular lipid droplets.

The next step is the transcytoplasmic transport of cholesterol from lipid droplets to the outer mitochondrial membrane by still poorly understood mechanisms. More is known about the subsequent transfer of cholesterol from the outer to the inner mitochondrial membrane, a rapid step (within minutes), which in the testis is critically dependent on gonadotrophin stimulation. A crucial role therein is played by the functional interplay of at least two recently discovered molecules, the steroidogenic acute regulator protein (stAR) and the translocator protein (TSPO) [19] (see also Chapter 5.8.1, 'Genetics of Adrenal Insufficiency').

Box 10.1.1.1 The physiological actions of androgens

- Feedback regulation of gonadotrophin synthesis and secretion
- Androgenic actions
  - Differentiation of the male sexual organs
  - Secondary sex characteristics
    - Growth of male sex organs
      - Testis
      - Epididymis
      - Seminal vesicle
      - Prostate
      - Penis
      - Scrotum
    - Pubic hair
    - Axillary hair
    - Beard
    - Male-type hair distribution and balding
  - Regulation of spermatogenesis
- Psychological actions
  - Cognitive functions
  - Libido and potency
  - Sexual behaviour
  - Aggression
- Anabolic actions
  - Growth spurt at puberty
  - Epiphyseal closure
  - Growth of larynx
  - Thickening of vocal cords
  - Effects on blood lipids
  - Muscle mass
  - Distribution of adipose tissue
  - Haematopoiesis
  - Thickening of skin
  - Function of sebaceous gland

The first and rate-limiting step of steroid biosynthesis [19] at the mitochondrial inner membrane is the conversion of cholesterol to pregnenolone, catalysed by the cytochrome P450 cholesterol side chain cleavage enzyme (P450scc, CYP11A1) and auxiliary electron transferring proteins (**Figure 10.1.1.4**). For the following steps, from pregnenolone onwards, the steroid molecule is translocated to the smooth endoplasmic reticulum. Depending on the order of enzymatic reactions, two alternative pathways, Δ^5 or Δ^4, are employed (**Figure 10.1.1.4**), the former being more important in the human testis. Hence, pregnenolone is first converted by the bifunctional CYP17A1 enzyme to 17-hydroxypregnenolone (17-hydroxylation step) and further to dehydroepiandrosterone (17, 20 lyase step). The following steps, in alternative order, are *either* the conversion of the 3β-hydroxy-5-ene structure of dehydroepiandrosterone by 3β-hydroxysteroid dehydrogenase/$\Delta^{5\text{-}4}$ isomerase (3β-HSD) to the 3-keto-4-ene structure of androstenedione, *or* the reduction of dehydroepiandrosterone by 17β-hydroxysteroid dehydrogenase (17β-HSD; mainly type III) to 5-androstene-3β,17β-diol, with final formation of T after the two reactions taking place in alternate order. In the alternative Δ^4 pathway, the first metabolic step after pregnenolone is its 3β-HSD-catalysed conversion to progesterone, after which the

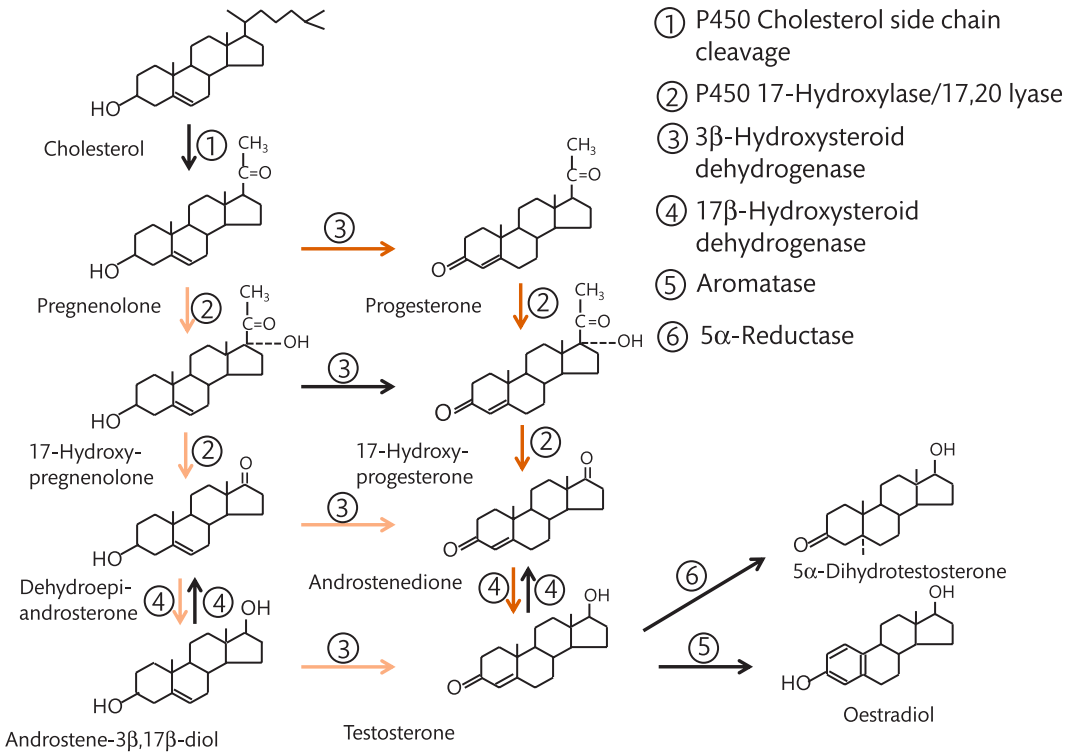


Figure 10.1.1.4 The key steroid metabolic steps in the testis leading to formation of T, 5α-DHT and oestradiol. The Δ5 pathway, used by the human testis, is marked with light orange arrows, and the Δ4 pathway, more important in rodents, with dark orange arrows. The same enzyme with dual function, P450 17-hydroxylase/17,20-lyase (enzyme 2) catalyses both 17-hydroxylation and ring D side chain cleavage in pregnenolone and progesterone. Apart from the interconversion of 17-keto and 17-hydroxy steroids all other reaction in the steroid metabolic pathway are irreversible. Multiple isoform of 17β-hydroxysteroid dehydrogenase (enzyme 4), with differing substrate specificities, catalyse these reactions. *For a colour version of this figure, please see colour plate section.*

metabolism proceeds further through 3-keto-4-ene intermediates to T (Figure 10.1.1.4).

The daily T production in a male is 6–7 mg; about 95% of it originates from the testes and the rest through peripheral metabolism of adrenal androgenic precursors (mainly dehydroepiandrosterone (DHEA) and its sulphate conjugate). A number of other steroid hormones are also secreted, of which many are Δ^4 and Δ^5 intermediates of T synthesis (Table 10.1.1.1). Some of them are weak androgens themselves (e.g. androsterone) or are metabolized further to more active androgens or oestrogens in peripheral tissues (androstenedione). Some steroids are stored in the testis and secreted as sulphate conjugates (Table 10.1.1.1) [20]. Steroid sulphates have no known hormonal function, and apparently represent storage and secretory forms.

T serves as a precursor for two important steroid hormones, 5α-DHT and oestradiol [21, 22] (Figure 10.1.1.4). Although these two T metabolites are secreted by the testis (Table 10.1.1.1), the major proportion of them is formed in peripheral tissues. 5α-DHT is 5–10-fold more potent than T due to its higher affinity for androgen receptor (AR), and it is also the main active molecule in some androgen target tissues, such as hair follicle and prostate, where a large proportion of circulating 5α-DHT is formed [23]. 5α-Reductase, converting T to 5α-DHT, exists as two isoenzymes [22]. Type 1 is expressed in sebaceous glands and liver, and type 2 in male urogenital tract, genital skin, and liver. In prostate, 5α-reductase is upregulated by androgen, in liver by thyroid hormone, and in skin fibroblasts by insulin-like growth factor (IGF) -I. Disturbances of male-type differentiation and sexual functions in connection with *5α-reductase* mutations demonstrate the physiological significance of this metabolic step in androgen physiology (see also Chapter 10.1.2, 'Sex Steroid Actions in the Male').

Oestradiol, either alone or in combination with androgens, participates in some important physiological functions of testicular steroids. About 25% of circulating oestradiol originates from the testes,

Table 10.1.1.1 The mean testicular, spermatic vein and peripheral vein concentrations (in nmol/L) of the key testicular steroids in man (Leinonen *et al.*, 1981).

Steroid	Testis	Spermatic vein	Peripheral vein
Pregnenolone sulphate	2600	430	90
Progesterone	130	23	0.8
17-Hydroxyprogesterone	690	45	3.2
Dehydroepiandrosterone	680	35	8.2
Dehydroepiandrosterone sulphate	2000	1400	1000
5-Androstene-3β,17β-diol	820	590	500
Androstenedione	740	45	2.5
Testosterone	2600	720	20
Testosterone sulphate	1400	150	13
5α-Dihydrotestosterone	50	14	1.5
Oestradiol	15	0.4	0.1

and the rest is formed through peripheral aromatization of androgenic precursors, mainly in adipose tissue and liver [21]. Human males with inactivating mutations of the *aromatase* (*CYP19*) or *oestrogen receptor-α* (*ESR1*) genes, as well as the corresponding knockout mouse models [24], have emphasized the physiological significance of oestrogen action in the male. Deficient oestrogen action in the male is associated with incomplete long bone epiphyseal closure, osteoporosis, insulin resistance, abnormalities in plasma lipids, and body fat deposition (see also Chapter 10.1.2, 'Sex Steroid Actions in the Male').

In the catabolism of T [25] the first step is its conversion to androstenedione through an interconvertible reaction catalysed by a specific oxidative isoenzyme of 17β-HSD (type 2) present in the liver. Androstenedione is then the preferential substrate for further reduction and hydroxylation reactions. The major metabolism involves 5α- and 5β-reduction of the Δ^4 double bond in ring A of the steroid nucleus, which is followed mainly by 3α- and 3β-hydroxylation of the 3-keto group. Thereafter, the 3α-metabolites are mainly conjugated with glucuronic acid, the 3β-metabolites with sulphate. About 90% of the androgen metabolites are excreted into urine and 10% into faeces. Of the urinary steroid metabolites, 20–40% occur as glucuronides, 40% as sulphates and the rest in free form.

Steroid Hormone Secretion and Transport

The secretion of steroid hormones into the circulation is assumed to be a passive diffusion process due to their lipid solubility and ease of transit through cell membranes. The main steroids produced and secreted by the human testis are listed in Table 10.1.1.1. T and the sulphate conjugates of pregnenolone, dehydroepiandrosterone, and 5-androstene-3β,17β-diol are quantitatively most abundant. There is no intratesticular storage of bioactive steroid hormones, with the exception of their sulphate conjugates, and the regulation of circulating steroid hormone concentrations occurs mainly at the level of their biosynthesis and variation of testicular blood flow. Since only about 30 μg of T is stored in the normal testes, the total content must turn over 200 times per day to achieve the daily production of 6 mg. Testicular steroid secretion has diurnal variation due to the night-time accentuation of LH secretion. Secretion pulses of serum T after the LH secretion peaks (average interval of 2 h) are inconsistent, apparently due to the relatively slow response of human testicular steroidogenesis to gonadotrophin stimulation, and to the buffering effect of steroid binding to plasma transport proteins (see next). Whereas animal experiments show dramatic (over 10-fold) responses of serum T to LH/human chorionic gonadotrophin (hCG) stimulation, the acute response of human testicular T release, within 1–2 hours, is only in the order of 30–50%.

In circulation, only 1–2% of T appears in free (i.e. non-protein-bound) form, 44% is bound to sex-hormone-binding globulin (SHBG), and 54% to albumin and other proteins [26]. SHBG is a dimeric β-globulin with a molecular weight of about 95 kDa, its carbohydrate content is 30%, and each of its monomers binds one androgen or oestrogen molecule. The affinity of albumin for T is only about 0.1% of that of SHBG, but its high concentration in circulation explains its overall importance in androgen transport.

Due to the avid binding of T to SHBG, their complexes do not enter androgen target cells. However, they may have some, as yet incompletely characterized, functions at the cell membrane [26]. T dissociates from the binding proteins in capillaries, where their interaction with endothelial glycocalyx reduces the affinity for T, allowing free T to diffuse into target cells. In contrast, the binding of T to albumin is easily dissociable, and this fraction, together with free T, forming about half of total plasma T, is considered to form the so-called bioavailable fraction of androgen.

The plasma level of SHBG is under endocrine and metabolic regulation [26], increasing clearly by oestrogen and ageing, and decreasing somewhat by androgen, and clearly by obesity. Consequently, the SHBG level in men is about half of that of women, and it increases in hypogonadism. If the HPT axis functions normally, changes in SHBG levels do not affect the balance of bioactive androgens, since they are quickly compensated for by changes in the feedback regulation.

For reading on mechanism of androgen action, see Chapter 10.1.2 'Sex Steroid Actions in the Male').

Physiological Effects of Androgens

The physiological effects of androgens are listed in Box 10.1.1.1. The major androgenic functions include the feedback regulation of gonadotrophin secretion by the hypothalamic–pituitary system, regulation of initiation and maintenance of spermatogenesis, male-type differentiation of the sexual organs during embryogenesis, stimulation of secondary sexual differentiation during puberty and control of male sexual behaviour. An alternative way of grouping the androgen effects is to define them as androgenic, psychological, and anabolic (Box 10.1.1.1). Besides the classical androgen target organs, low levels of AR are present in almost every tissue. T is the prohormone for some androgen actions, which, at target tissue level, are exerted by 5α-DHT or oestradiol. The anabolic androgen actions on muscle are predominantly due to T action, whereas its effects on bone, at least partly, require aromatization to oestrogens. The androgen effects on the central nervous system are mediated both through oestrogen (following aromatization of T) and ARs.

Testicular Protein and Peptide Hormones: Inhibin, Activin, and Follistatin

The testis produces a number of bioactive proteins and peptides (Table 10.1.1.1) and various autocrine and paracrine functions have been proposed for most of them [27–29]. Some of these hormones are functional in fetal life: the Leydig cell product insulin-like 3 (INSL3) regulates testicular descent, and the Sertoli cell product anti-Müllerian hormone (AMH) induces the regression of Müllerian ducts in the male fetus. Endocrine effects outside the adult testes have clearly been demonstrated only for inhibins, whereas the other members of the same family, activins, and their binding protein, follistatin, mainly exert their actions as paracrine or autocrine growth factors within the testis and several other tissues. The endocrine role of inhibins is to mediate the negative feedback regulation of pituitary FSH secretion.

Inhibin and activin are dimeric gonadal proteins, members of the transforming growth factor (TGF)–β family [1, 29]. In addition, there are specific transport proteins for activin, follistatin and α_2-macroglobulin. A plethora of studies have described the regulatory functions of these molecules in different tissues and their potential clinical significance.

Inhibin is a dimeric glycoprotein hormone of 32 kDa size, consisting of two dissimilar subunits, connected by disulphide bonds: a 18 kDa α-subunit and a 14 kDa β-subunit of type βA or βB, forming

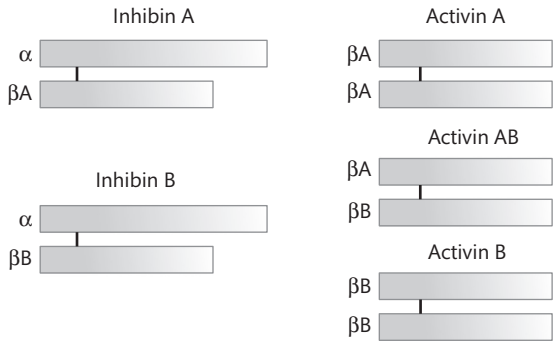


Figure 10.1.1.5 Schematic representation of inhibins and activins. The different subunits (α, βA and βB) can heterodimerize to form inhibins A (α/βA) and B (α/βB) or homodimerize to form activins A (βA/βA), AB (βA/βB) or B (βB/βB).

inhibin A and B (Figure 10.1.1.5). Homodimers of two β-subunits (βA/βA, βA/βB, βB/βB), have opposite biological action to inhibin, that is, para/autocrine stimulation of pituitary FSH secretion, hence the name activin. Activins bind to two types of receptors, I and II, which are single transmembrane-domain serine/threonine kinase molecules. Inhibin together with its coreceptor betaglycal blocks activin binding to its receptors. Follistatin, a glycoprotein of 31–42-kDa molecular size, bioneutralizes peptides of the TGFβ family, including activin.

In the human adult testis Sertoli cells strongly express inhibin α and βB mRNA, but βA mRNA very weakly. Hence, Sertoli cell-derived inhibin B is the main circulating inhibin with endocrine action [30, 31]. The Leydig cells express inhibin-α subunit and βB. βA is also expressed by peritubular myoid cells, and βB by spermatogenic cells. Follistatin is expressed by Sertoli and germ cells.

FSH stimulates inhibin α subunit expression in Sertoli cells, whereas the β subunits are under paracrine regulation. Serum FSH and inhibin B levels are inversely correlated, in support of the physiological role of inhibin (in concert with T) as a negative feedback regulator of FSH synthesis and release at the pituitary level. The expression of βA- or βB-subunits (that is, activin synthesis) is largely independent of FSH. The role of activin in FSH regulation is auto/paracrine locally in pituitary gland, and it is modulated by neutralizing action of follistatin. The inhibin peptides participate within the testis in the regulation of spermatogenesis.

Circulating inhibin B levels closely reflect the number and function of Sertoli cells. Inhibin B measurements offer a tool for monitoring Sertoli cell function in individual patients with infertility, in clinical and toxicological studies on male fertility, as well as in studies of developmental deficiencies of testicular function [31]. Low levels of inhibin B occur in men with non-obstructive azoospermia, in untreated men with hypo- or hypergonadotropic hypogonadism, in infertile men with elevated FSH, in men with Klinefelter syndrome and in orchidectomized men [32, 33]. Men with inactivating mutation of the *FSH receptor* gene also have very low level of this peptide [34]. Inhibin B is an early marker of male puberty, increasing about threefold between stages I and II [35]. The negative correlation between inhibin B and FSH is attained in late puberty. The levels of activin A are similar in men and women, and follistatin levels are somewhat lower in men.

Local Regulation of Testicular Function

Besides gonadotrophins and other hormones reaching the testis from circulation, there is a complex network of local paracrine, autocrine, and intracrine regulatory interactions within the testis (Box 10.1.1.2). This type of regulation is easy to demonstrate *in vitro* in cell coculture experiments, and numerous such interactions

Box 10.1.1.2 The different types of paracrine and autocrine signals detected in the testis tissue. The list is based on references cited in Saez 1994 [27] and Smith & Walker 2015 [1], and additional references found in PubMed (not cited)

- Neurohormones and neuropeptides
  - Growth hormone-releasing hormone
  - Pituitary adenylate cyclase-activating peptide
  - Gonadotrophin-releasing hormone
  - Corticotrophin-releasing hormone
  - Oxytocin
  - Arginine vasopressin
  - Thyrotropin-releasing hormone
  - Somatostatin
  - Opioids
  - Substance P
  - Galanin
  - γ-Aminobutyric acid
  - Catecholamines
  - Histamine
  - Neuropeptides B and W
  - Kisspeptin
  - Nociceptin
- Peptides originally identified in the testis
  - Inhibin
  - Activin
  - Follistatin
  - Anti-Müllerian hormone
  - Peritubular Sertoli cell regulating substance (PmodS)
- Growth factors
  - Insulin-like growth factors and their binding proteins
  - INSL3 (and other relaxin-peptides)
  - Insulin
  - Transforming growth factors α and β
  - Fibroblast growth factors
  - Platelet-derived growth factor
  - Nerve growth factor
  - Kit ligand
  - Gastrin-releasing peptide
  - Glial cell-derived neurotrophic factor
  - Platelet-derived growth factor
  - Bone morphogenetic proteins
  - Epidermal growth factor
  - Growth differentiation factor-9
  - Wilms tumour 1
  - Osteocalcin
- Immune derived cytokines
  - Interleukins
  - Interferons
  - Tumour necrosis factor-α
  - Oncostatin M
  - Leukaemia inhibitory factor
- Vasoactive/cardiovascular peptides
  - Endothelin

(*continued*)

Box 10.1.1.2 *Continued*

- Angiotensin II
- Atrial natriuretic peptide
- Vasoactive intestinal peptide
- Vascular endothelial growth factor
- Adrenomedullin
- Prostaglandins
- Natriuretic peptides

- Orexigenic hormones
  - Leptin
  - Ghrelin
  - Orexin
  - Neuropeptide Y
  - Adiponectin
  - Endocannabinoids
- Non-peptide hormones
  - Androgens
  - Oestrogens
  - Glucocorticoids
  - Progesterone
  - Vitamin D
  - Retinoic acids
  - Prostaglandins

are known today [27, 28]. However, their physiological significance remains uncertain in the absence of conclusive *in vivo* data.

One explanation for the local regulatory network is biological redundancy. To assure a vital physiological process, such as spermatogenesis, partly overlapping mechanisms are needed; if one fails the others take over. The other explanation relates to the complex testicular anatomy where different cell types may have dynamic interactions through paracrine coordination between cells at different functional and developmental stages. A perplexing feature is the wide expression, without apparent function, of a multitude of genes in germ cells during their maturation process. The main types of testicular local regulation are described next, and those most likely of physiological significance are elaborated.

Besides being the main testicular hormone, T is also the best example of a physiologically significant paracrine factor within the testis [27, 28]. Being formed by Leydig cells, it plays a key role in the local regulation of spermatogenesis. The Sertoli cells produce, under the influence of T, nutrients and other paracrine signals for the maintenance of spermatogenesis; hence the role of T is essential but indirect. AR are also present in the peritubular myoid cells, and T also has an autocrine or intracrine role in regulating Leydig cell function.

Almost all classes of regulatory peptides have been shown *in vitro* either to be produced in the testis or to have receptors in this organ, and a concept has been put forward that these peptides form an intratesticular network of paracrine and autocrine signalling (Box 10.1.1.2). Testis-specific knockout animal models can inform the possible physiological relevance of the paracrine or autocrine regulation. However, it may not apply to the human testis, since vast species differences exist in testicular cell-to-cell interactions. Conspicuously, experimental disruption of function of many peptides with putative intratesticular function has turned out to have only marginal effects on testicular function in rodents. There are a few exceptions; the disruption of the inhibin-α subunit has demonstrated that inhibin is an intragonadal tumour suppressor molecule, and general or testicular overexpression of enkephalin, interleukin-2 and TGF-β genes impair spermatogenesis and fertility [28]. In addition, impaired spermatogenesis is a non-specific finding in a number of genetically modified animal models with no apparent relationship to testicular function.

The best example of a physiologically significant paracrine effect within the testis, apart from T, is provided by the transgenic experiments with the insulin-like growth factor genes [36]. Overexpression of insulin-like growth factor-II increases testicular size. Mice with disrupted insulin-like growth factor-I gene have impaired Leydig cell maturation and suppressed spermatogenesis. However, since both the endocrine and paracrine component of testicular insulin-like growth factor-I actions are eliminated in this model, the final evidence for the significance of intratesticular insulin-like growth factor-I awaits testis-specific disruption of its gene. Altogether, it remains a conundrum why almost every biologically active regulatory molecule is expressed in the testis and can be shown to have direct effects (usually *in vitro*) on some aspects of testicular function.

Spermatogenesis

Testis Development and Germ Cells

Differentiation of germ cells from spermatogonia to spermatozoa is called spermatogenesis. See Figure 10.1.1.6. It takes place in the seminiferous tubules in the testis. It includes three phases: mitotic proliferation of spermatogonia, meiosis of spermatocytes, and morphogenesis of haploid spermatids to mature spermatozoa (spermiogenesis). Kinetics of germ cell differentiation is tightly regulated, and complete spermatogenesis takes 74 days in humans. Quantity of sperm production is highly variable, and World Health Organization (WHO) has lowered the reference limits for normal semen quality over decades. The current lower reference limit for total sperm number per ejaculate is 39 million and for sperm concentration 15 million per millilitre (WHO 2010). The median sperm concentration in young men of the general population varies between 40 and 60 million per millilitre, whereas men with proven fertility have much higher sperm concentrations [37]. Even in healthy, fertile men, most spermatozoa are structurally abnormal, and according to WHO reference limits, semen is considered normal if more than 4% of spermatozoa are morphologically normal [38]. See Chapter 10.2.3 Semen Analysis.

Germ cell specification occurs early in embryonic development. Primordial germ cells migrate from the yolk sac via hind gut to the bipotential gonadal primordium which differentiates to a testis in the presence of the sex-determining region of Y (*SRY*) gene. Several other genes are necessary for a male specific gonadal differentiation, such as *SOX9*, but they are not discussed further here (for review, see [39] for further details). Germ cells are called gonocytes in the fetal testis. They become surrounded by Sertoli cells and these form the seminiferous cords. The gonocytes undergo epigenetic reprogramming when the whole genome is demethylated and then methylated again, except for few gene areas that remain outside reprogramming. When the gonocytes migrate from the centre of the seminiferous cord to the

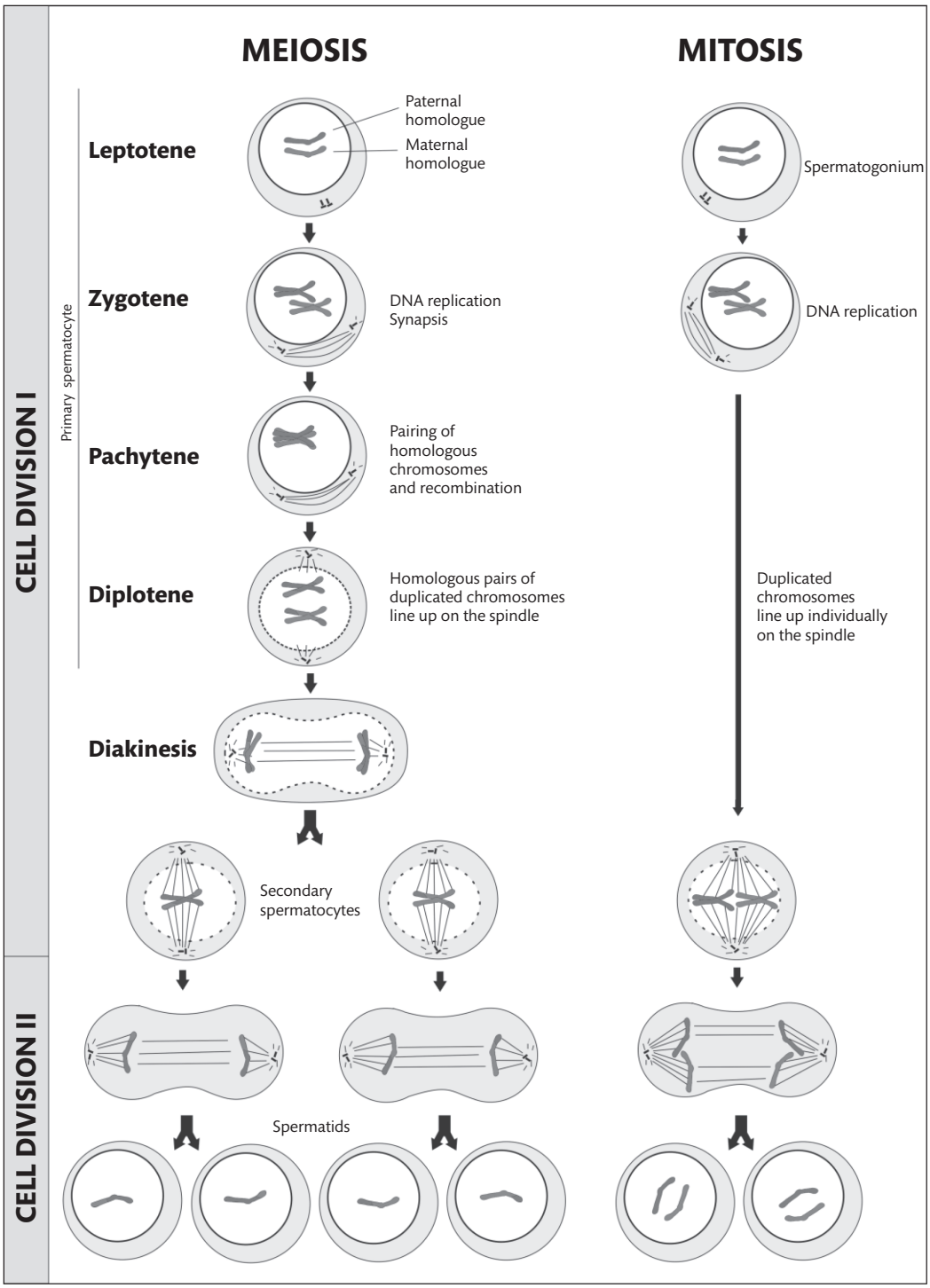


Figure 10.1.1.6 Schematic presentation of mitosis and meiosis. Spermatogonia multiply by mitosis like somatic cells, whereas spermatocytes undergo meiosis where genetic material is exchanged between chromosomes by homologous recombination. Chromosomes separate from each other in the first meiotic division when haploid secondary spermatocytes are formed. These divide rapidly again when chromatids separate and haploid spermatids are formed.

Sheyla Cisneros Montalvo is acknowledged for drawing the figure.

basement membrane in the periphery they form the spermatogonial stem cell pool. In rodents these cells proceed immediately to the first spermatogenic cycle, whereas the human spermatogonial stem cell pool is held in an undifferentiated stage for more than ten years until puberty. Survival of spermatogonia is critically dependent on optimal temperatures lower than core body temperature. Testes have to descend into the scrotum to be able to create a countercurrent heat exchange system between the testicular veins and arteries. Spermatogonia undergo apoptosis in an undescended testis (cryptorchidism) causing infertility in bilateral cryptorchidism, unless the testes are brought to the scrotum early in childhood [40].

Spermatogonia

Before puberty only the spermatogonial stem cell pool is maintained and spermatogonia do not enter meiosis. Spermatogenesis beyond meiosis starts in puberty in response to increasing FSH stimulation and continues through adult life. The balance between renewal of the spermatogonial stem cell pool and commitment to differentiation is critical for maintenance of lifelong fertility. If the spermatogonial stem cell pool is depleted, sperm production will cease when all differentiating spermatogonia have developed into spermatozoa. On the other hand, accumulation of stem cells leads to tumour-like development and reduction of sperm production. Glial cell-derived neurotrophic factor (GDNF) promotes the self-renewal of stem cell spermatogonia [41], and when their retinoic acid receptor gamma (RARγ) expression increases, retinoic acid drives cohorts of these cells to differentiation [42].

Mitoses of spermatogonial stem cells increase with ageing, which is considered as a possible reason for mutations to accumulate, particularly in some hot spots, such as genes for fibroblast growth factor receptors. Approximately two novel point mutations per year appear leading to doubling of mutations every 20 years. As a consequence, several genetic diseases occur more often in children of old men than in those of young fathers. Achondroplasia, Apert syndrome, and Costello syndrome are examples of such diseases [43]. Other diseases, such as autism spectrum disorders, the genetics of which is not known, are also more common in children of old men than in young men's offspring.

Spermatogonial proliferation differs between rodents and men. Mice and rats have several generations of spermatogonia due to many rounds of overlapping mitotic proliferation, whereas men have only three types of spermatogonia: A_{dark}, A_{pale}, and B. They are located on the basement membrane of the seminiferous tubules surrounded by Sertoli cells. After mitotic division of type B spermatogonia, the first meiotic cells, preleptotene spermatocytes appear. DNA is replicated and the cells are moved through the blood–testis barrier to the adluminal side of the seminiferous epithelium where the extracellular milieu is very different from normal, mimicking intracellular space with high potassium and low sodium concentration. The adluminal space is also immune-privileged protecting unique germ cells from immune surveillance.

Testes start to grow at the onset of puberty when the Sertoli cells proliferate first, and then spermatogonia are committed to differentiation and enter meiosis. The rapid multiplication of spermatogenic cells makes the testes grow. This first visible sign of puberty in boys occurs at the median age of 11.5 years, while the age range spans from 9.5 to 13.5 years [44].

A_{dark} spermatogonia divide actively and maintain the spermatogonial pool before puberty, whereas in adult testes these cells serve as reserve stem cells and stay mostly quiescent [45, 46]. In adult testes A_{pale} spermatogonia proliferate either to renew themselves (self-renewal) or give rise to type B spermatogonia (differentiation commitment). A_{dark} spermatogonia are needed to replenish the pool of A_{pale} spermatogonia when these become depleted. Thus, A_{pale} spermatogonia are transit-amplifying progenitor cells rather than true stem cells. Difficulty in studies of human spermatogonia is the subtlety of morphological differences between A_{dark} and A_{pale} spermatogonia. Rarefaction in the nucleus of A_{dark} spermatogonia is used to recognize this cell type.

Proliferation and differentiation of rodent spermatogonia have been well-characterized, and despite apparent differences between rodent and human spermatogenesis, regulatory mechanisms appear to be similar. Retinoic acid signalling commits the spermatogonia from self-renewal to differentiation and meiosis [47]. Quiescent stem cell spermatogonia do not express KIT, the receptor for stem cell factor (KIT ligand), that appears in differentiating spermatogonia [48]. Stem cell factor is important for germ cell survival and KIT is expressed also in fetal gonocytes, spermatocytes, and in a truncated form in spermatids. Differentiation of spermatogonia is accompanied with an epigenetic change, since these cells express DNA methyltransferases (DNMT) and possess altered histone modifications compared to undifferentiated spermatogonia [49]. This may prevent the differentiating spermatogonia to switch back to stem cells.

Meiosis

Meiosis is a long process taking 2–3 weeks depending on species. It starts with DNA replication in preleptotene spermatocytes that transit from the basal compartment via a transitory space to the adluminal compartment of the seminiferous tubules [50]. Meiosis proceeds through leptonema, zygonema, pachynema and diplonema (one can also use terms leptotene, zygotene, pachytene, and diplotene phases) to diakinesis. In leptotene spermatocytes chromosomes that consist of two sister chromatids start to condense and the synaptonemal complex starts to form between DNA strands. In zygonema, pairing of the homologous chromosomes takes place. Crossing-over occurs in pachytene spermatocytes when non-sister chromatids of homologous chromosomes switch DNA with each other. The sites of crossing-over, chiasmata, become apparent in diplotene spermatocytes when the synaptonemal complex disintegrates and the chromosomes separate from each other. In diakinesis, meiotic spindle starts to form and nuclear membrane disintegrates. In the first meiotic division chromosomes are divided into two haploid secondary spermatocytes (with 2C amount of DNA). They do not replicate DNA but rapidly undergo another division giving rise to haploid spermatids (with 1C amount of DNA).

Spermiogenesis

Spermatids undergo a remarkable metamorphosis from a conventional round cell to a highly differentiated carrier of the genome that can move rapidly and fertilize an egg. Spermatids develop an acrosomal structure to cover nucleus that condenses and polarizes to form the head of the sperm. Condensation of the nucleus requires the changing of histones to protamines in the chromosomes, which allows a toroid structure, packing the chromosomes very tightly. Transition proteins appear in between the exchange process. Some histone proteins are also left and their methylation status is considered important. A flagellum is built behind the head, and the midpiece at the proximal end of the tail is loaded with mitochondria that are like rocket motors for the space craft. The whole process takes 2–3 weeks, like meiosis, depending on species. Spermiogenesis

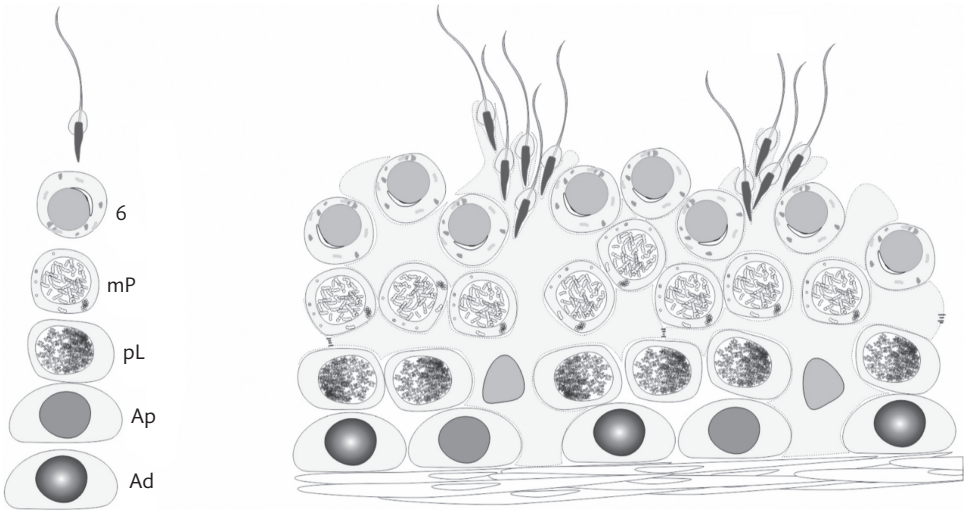


Figure 10.1.1.7 Schematic presentation of a section of human seminiferous epithelium at stage VI of the cycle. Sertoli cells envelope all spermatogenic cells. The blood–testis barrier is formed by Sertoli cells between spermatogonia and spermatocytes, isolating spermatogonia to basal compartment and spermatocytes together with spermatids to the adluminal compartment of the seminiferous epithelium. Preleptotene spermatocytes reside briefly in the intermediate compartment when transferred from the basal lamina to the adluminal area.
Sheyla Cisneros Montalvo is acknowledged for drawing the figure.

ends at spermiation when mature spermatids are released from the seminiferous epithelium to the lumen of the seminiferous tubules.

Transcription stops when the nuclear condensation occurs. However, many new proteins are needed after that, and the cells have to store RNA for later translation. Spermatids have a special ribonucleoprotein structure, called chromatoid body that is a platform for RNA storage and post-transcriptional modification [51, 52]. Components of the chromatoid body can be seen already in intermitochondrial cement of pachytene spermatocytes, but it becomes conspicuous in round spermatids where its movements around the nucleus can be followed in live cell microscopy [53].

Stages of the Seminiferous Epithelial Cycle

Different phases of spermiogenesis are easy to recognize, because the morphology of the cells change rapidly and the cellular features are very distinct. This has helped to understand also the whole organization of the spermatogenic cells in the testis. Yves Clermont and Charles Leblond introduced a staging system on the basis of spermatid morphology in 1952 [54]. In the seminiferous epithelium, spermatogonia, spermatocytes, and spermatids form layers where distinct cell types are always found together and they differentiate synchronously to the next developmental phases. These distinct cell associations characterize different stages of the seminiferous epithelium. They proceed in the seminiferous tubules in a wave-like fashion, and in time, in a cyclic fashion. In rodent seminiferous tubules these are easy to follow, because the stages occupy whole segments of the tubules and they can be microdissected separately from each other. In human seminiferous tubules the structure is more complex, since one cross-sectional segment of the tubule can contain pieces of several stages. This can be explained by the helical arrangement of succeeding stages. Clermont (1963) originally described six stages of the human seminiferous epithelium [55], whereas more recently, 12 stages were characterized [56], which makes comparison to e.g. the mouse easier. Sertoli cells nurture the spermatogenic cells and modify their functions according to the needs of different stages. This makes spermatogenic function cyclical, and the length of the cycle reflects the species-specific characteristics of spermatogenesis. The length of the cycle in human is 16 days [57], and the whole spermatogenesis takes around 4.5 epithelial cycles (see Figures 10.1.1.7 and 10.1.1.8).

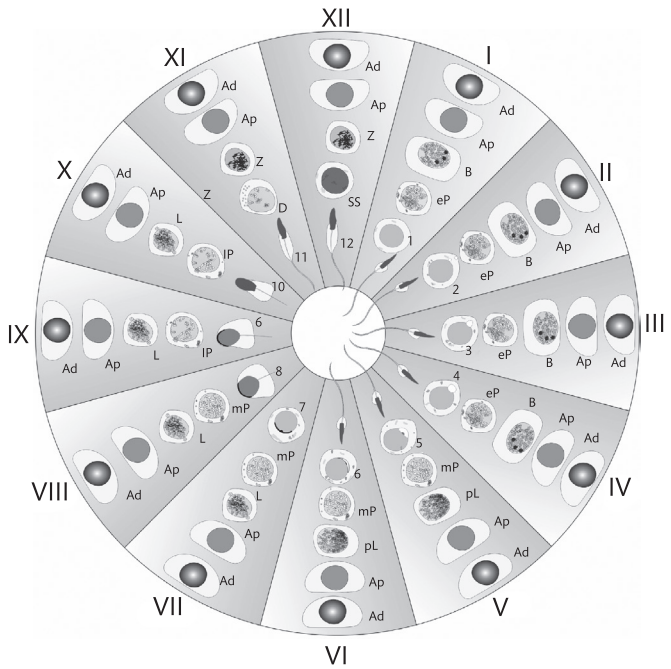


Figure 10.1.1.8 Spermatogenic cells form distinct associations, called stages of the seminiferous epithelial cycle, that follow each other in a cyclic fashion and germ cells differentiate through the cycle up to mature spermatid that is released to the lumen after stage VI in human seminiferous epithelium.
Sheyla Cisneros Montalvo is acknowledged for drawing the figure.

LH/T and FSH Actions in the Regulation of Spermatogenesis

It is generally held that FSH action is needed in the early stages of spermatogenesis, whereas the progression of postmeiotic round spermatids to elongated spermatids is the quintessential step needing androgen action [58, 59]. To what extent the actions of LH-stimulated T and FSH are needed for completion of spermatogenesis remains a contentious topic. The hormonal requirements for initiation of the first spermatogenic wave at puberty, for its maintenance in adult age, and for the re-initiation after transient suppression are apparently different. Androgen alone may not be able to drive spermatogenesis to completion beyond the early spermatid stage in the immature testis, and if a prepubertal animal is hypophysectomized, LH only partially prevents germ cell loss. On the other hand, full spermatogenesis can be initiated in various gonadotrophin-deficient rodent models by T treatment alone, and in man, prolonged hCG treatment alone, through stimulation of T production, is able to initiate spermatogenesis [60]. However, these patients may not have been totally deficient in FSH.

Treatment of healthy men with T enanthate (200 mg/week i.m.) suppresses both gonadotrophins and intratesticular T while maintaining peripheral T action. As a result, spermatogenesis is severely suppressed—to the extent that this provides a successful means of male contraception (see Chapter 10.3.6, 'Hormonal Male Contraception'). If these men receive injections of LH or hCG, their spermatogenesis recovers **qualitatively**, due to the restoration of high intratesticular T [61], although their FSH level remains suppressed. However, the sperm counts remain at only 50% of the pretreatment levels. This has been considered evidence that the re-initiation of spermatogenesis is possible with T alone, but its **quantitative** recovery also needs FSH. When FSH was added to the above treatment regimen, spermatogenesis was fully restored. The T-suppressed men were also treated with purified FSH alone [61], which was able to stimulate spermatogenesis, though not quantitatively. Hence, it appears that neither LH/T nor FSH is absolutely required for qualitatively normal spermatogenesis if either of the other hormone is available at sufficient levels.

Human inactivating *FSHR* mutations, as well as animal models with disrupted *Fshb* or *Fshr* gene function, have shed more light on the role of FSH in spermatogenesis [62]. Men with inactivating mutation in *FSHR* are normally masculinized, but subfertile with reduced testicular size and poor sperm quality. However, sporadic men with inactivating *FSHB* mutation have been found to be azoospermic. The reason for the discrepancy between phenotypes of the hormone and receptor inactivation remains unclear. Both the *Fshb* and *Fshr* knockout mouse data are in agreement with phenotype of the FSHR deficient men, indicating that spermatogenesis, though quantitatively and qualitatively compromised, may proceed without FSH action. Apparently, T is the 'master switch' of spermatogenesis and other factors, including FSH, are needed to maintain it qualitatively and quantitatively normal.

One of the remaining enigmas of androgen action is why the regulation of spermatogenesis seems to require about 100-fold higher concentrations of T than is needed for extragonadal androgen actions, despite the fact that apparently the same AR is mediating these effects. A simple explanation may be that such high intratesticular levels of T are in fact not needed—the T levels are high only because this organ is the site of T production. Recent animal data are in support of this explanation [63].

REFERENCES

1. Smith LB, Walker WH. Hormone signaling in the testis. In: Plant TM, Zeleznik AJ, eds. *Knobil and Neil's Physiology of Reproduction*, Vol. 1, 4th edn. Waltham: Academic Press, 2015: 637–90.
2. Huhtaniemi IT, Howard S, Dunkel L, Anderson RA. The gonadal axis: a life perspective. In: Pfaff DW, Joels M, eds. *Hormones, Brain, and Behavior*, 3rd edn. London: Elsevier, 2017: 3–58.
3. Herbison AE. Physiology of the adult gonadotropin-releasing hormone neuronal network. In: Plant TM, Zeleznik AJ, eds. *Knobil and Neil's Physiology of Reproduction*, Vol. 1, 4th edn. Waltham: Academic Press, , 2015: 399–467.
4. Skorupskaite K, George JT, Anderson RA. The kisspeptin-GnRH pathway in human reproductive health and disease. *Hum Reprod Update*, 2014; **20**: 485.
5. Goodman RL, Lehman MN, Smith JT, *et al*. Kisspeptin neurons in the arcuate nucleus of the ewe express both dynorphin A and neurokinin B. *Endocrinology*, 2007; **148**: 5752.
6. Clarkson, J, Han, SY, Piet, R, *et al*. Definition of the hypothalamic GnRH pulse generator in mice. *Proc Natl Acad Sci U S A*, 2017; **114**: E10216.
7. Fiete D, Srivastava V, Hindsgal O, Baenziger JU. A hepatic reticuloendothelial cell receptor specific for SO4-4GalNAc beta 1,4GlcNAc beta 2,2Man alpha that mediates rapid clearance of lutropin. *Cell*, 1991; **67**: 1103.
8. Ascoli M, Fanelli F, Segaloff DL. The lutropin/choriogonadotropin receptor, a 2002 perspective. *Endocr Rev*, 2002; **23**: 141.
9. Ulloa-Aguirre A, Zariñán T. The follitropin receptor: matching structure and function. *Mol Pharmacol*, 2016; **90**: 596.
10. Hammes A, Andreassen TK, Spoelgen R, *et al*. Role of endocytosis in cellular uptake of sex steroids. *Cell*, 2005; **122**: 751.
11. Neirijnck Y, Calvel P, Kilcoyne KR, *et al*. Insulin and IGF1 receptors are essential for the development of steroidogenic function of adult Leydig cells. *FASEB J*, 2018; fj201700769RR.
12. Wagner MS, Wajner SM, Maia AL. The role of thyroid hormone in testicular development and function. *J Endocrinol*, 2008; **199**: 351.
13. De Toni L, Di Niso A, Rocca MS, De Rocca Ponce M, Ferlin A, Foresta C. Osteocalcin, a bone-derived hormone with important andrological implications. *Andrology*, 2017; **5**: 664.
14. Boisen IM, Bøllehuus Hansen L, Mortensen LJ, Lanske B, Juul A, Blomberg Jensen M. Possible influence of vitamin D on male reproduction. *J Ster Biochem Mol Biol*, 2017; **173**: 215.
15. Akingbemi BT, Ge R, Rosenfeld CS, *et al*. Estrogen receptor-alpha gene deficiency enhances androgen biosynthesis in the mouse Leydig cell. *Endocrinology*, 2003; **144**: 84.
16. O'Hara L, McInnes K, Simitsidellis I, *et al*. Autocrine androgen action is essential for Leydig cell maturation and function, and protects against late-onset Leydig cell apoptosis in both mice and men. *FASEB J*, 2015; **29**: 894.
17. Hardy MP, Gao HB, Dong Q, *et al*. Stress hormone and male reproductive function. *Cell Tissue Res*, 2005; **322**: 147.
18. Nieschlag E, Behre HM, eds. *Testosterone: Action, Deficiency Substitution*, 4th edn. Cambridge: Cambridge University Press, 2012: 1–569.

19. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*, 2011; **32**: 81.
20. Leinonen P, Ruokonen A, Kontturi M, Vihko R. Effects of estrogen treatment on human testicular unconjugated steroid and steroid sulfate production *in vivo*. *J Clin Endocrinol Metab*, 1981; **53**: 569.
21. Cooke PS, Nanjappa MK, Ko CM, Prins GS, Hess RA. Estrogens in male physiology. *Physiol Rev*, 2017; **97**: 995.
22. Swerdloff RS, Dudley RE, Page ST, Wang C, Slamew WA. Dihydrotestosterone: biochemistry, physiology, and clinical implications of elevated blood levels. *Endocr Rev*, 2017; **38**: 220.
23. Davey RA, Grossmann M. Androgen receptor structure, function and biology: from branch to bedside. *Clin Biochem Rev*, 2016; **37**: 3.
24. Toocheck C, Clister T, Shupe J, *et al*. Mouse spermatogenesis requires classical and nonclassical testosterone signaling. *Biol Reprod*, 2016; **94**: 1.
25. Sundaram K, Kumar N. Metabolism of testosterone in Leydig cells and peripheral tissues. In: Payne AH, Hardy MP, Russell LD, eds. *The Leydig Cell*. Vienna, IL: Cache River Press, 1996: 287–305.
26. Hammond GL. Plasma steroid-binding proteins: gatekeepers of steroid hormone action. *J Endocrinol*, 2016; **230**: R13.
27. Saez JM. Leydig cells: endocrine, paracrine, and autocrine regulation. *Endocr Rev*, 1994; **15**: 574.
28. Gnessi L, Fabri A, Spera G. Gonadal peptides as mediators of development and functional control of the testis: an integrated system with hormones and local environment. *Endocr Rev*, 1997; **18**: 541.
29. Barakat B, Itman C, Mendis SH, Loveland KL. Activins and inhibins in mammalian testis development: new models, new insights. *Mol Cell Endocrinol*, 2012; **359**: 66.
30. de Kretser, DM, Buzzard, JJ, Okuma, Y, *et al*. The role of activin, follistatin and inhibin in testicular physiology. *Mol Cell Endocrinol*, 2004; **225**: 57.
31. Luisi S, Florio P, Reis FM, Petraglia F. Inhibins in female and male reproductive physiology: role in gametogenesis, conception, implantation and early pregnancy. *Hum Reprod Update*, 2005; **11**: 123.
32. Anawalt BD, Bebb RA, Matsumoto AM, *et al*. Serum inhibin-B levels reflect Sertoli cell function in normal men and in men with testicular dysfunction. *J Clin Endocrinol Metab*, 1996; **81**: 3341.
33. Illingworth PJ, Groome NP, Byrd W, *et al*. Inhibin-B: a likely candidate for the physiologically important form of inhibin in men. *J Clin Endocrinol Metab*, 1996; **81**: 1321.
34. Tapanainen JT, Aittomäki K, Jiang M, Vaskivuo T, Huhtaniemi IT. Men homozygous for an inactivating mutation of the follicle-stimulating hormone (FSH) receptor gene present variable suppression of spermatogenesis and fertility. *Nat Genet*, 1997; **15**: 205.
35. Andersson AM, Juul A, Petersen JH, Müller J, Groome NP, Skakkebaek NE. Serum inhibin B in healthy pubertal and adolescent boys: relation to age, stage of puberty, and follicle-stimulating hormone, luteinizing hormone, testosterone, and estradiol levels. *J Clin Endocrinol Metab*, 1997; **82**: 3976.
36. Hardy MP, Ganjam VK. Stress, 11beta-HSD, and Leydig cell function. *J Androl*, 1997; **18**: 475.
37. Virtanen HE, Jørgensen N, Toppari J. Semen quality in the 21(st) century. *Nat Rev Urol*, 2017; **14**: 120.
38. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva, WHO, 2010.
39. Mäkelä JA, Koskenniemi JJ, Virtanen HE, Toppari J. Testis development. *Endocr Rev*, 2018; **40**: 857–905.
40. Virtanen HE, Toppari J. Cryptorchidism and fertility. *Endocrinol Metab Clin N Am*, 2015; **44**: 751.
41. Meng X, Lindahl M, Hyvönen ME, *et al*. Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science*, 2000; **287**: 489.
42. Ikami K, Tokue M, Sugimoto R, *et al*. Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis. *Development*, 2015; **142**: 1582.
43. Maher GJ, McGowan SJ, Giannoulatou E, Verrill C, Goriely A, Wilkie AO. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. *Proc Natl Acad Sci U S A*, 2016; **113**: 2454.
44. Sadov S, Koskenniemi JJ, Virtanen HE, *et al*. Testicular growth during puberty in boys with and without a history of congenital cryptorchidism. *J Clin Endocrinol Metab*, 2016; **101**: 2570.
45. Simorangkir DR, Marshall GR, Ehmcke J, Schlatt S, Plant TM. Prepubertal expansion of dark and pale type A spermatogonia in the rhesus monkey (Macaca mulatta) results from proliferation during infantile and juvenile development in a relatively gonadotropin independent manner. *Biol Reprod*, 2005; **73**: 1109.
46. Ehmcke J, Simorangkir DR, Schlatt S. Identification of the starting point for spermatogenesis and characterization of the testicular stem cell in adult male rhesus monkeys. *Hum Reprod*, 2005; **20**: 1185.
47. Sugimoto R, Nabeshima Y, Yoshida S. Retinoic acid metabolism links the periodical differentiation of germ cells with the cycle of Sertoli cells in mouse seminiferous epithelium. *Mech Dev*, 2012; **128**: 610.
48. Yan W, Suominen J, Toppari J. Stem cell factor protects germ cells from apoptosis *in vitro*. *J Cell Sci*, 2000; **113**: 161.
49. Shirakawa T, Yaman-Deveci R, Tomizawa S, *et al*. An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a Kit-positive identity. *Development*, 2013; **140**: 3565.
50. Smith BE, Braun RE. Germ cell migration across Sertoli cell tight junctions. *Science*, 2012; **338**: 798.
51. Meikar O, Da Ros M, Liljenbäck H, Toppari J, Kotaja N. Accumulation of piRNAs in the chromatoid bodies purified by a novel isolation protocol. *Exp Cell Res*, 2010; **316**: 1567.
52. Meikar O, Vagin VV, Chalmel F, *et al*. An atlas of chromatoid body components. *RNA*, 2014; **20**: 483.
53. Parvinen M. The chromatoid body in spermatogenesis. *Int J Androl*, 2005; **28**: 189.
54. Leblond CP, Clermont Y. Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann NY Acad Sci*, 1952; **55**: 548.
55. Clermont Y. The cycle of the seminiferous epithelium in man. *Am J Anat*, 1963; **112**: 35.
56. Muciaccia B, Boitani C, Berloco BP, *et al*. Novel stage classification of human spermatogenesis based on acrosome development. *Biol Reprod*, 2013; **89**: 60.
57. Heller CG, Clermont Y. Spermatogenesis in man: an estimate of its duration. *Science*, 1963; **140**: 184.
58. McLachlan RI, O'Donnell L, Meachem SJ, de Kretser DM, Pratis K, Robertson DM. Identification of specific sites of hormonal

regulation in spermatogenesis in rats, monkeys, and man. *Recent Prog Horm Res*, 2002; **57**: 149.
59. Huhtaniemi I. A short evolutionary history of FSH-stimulated spermatogenesis. *Hormones*, 2015; **14**: 368.
60. Vicari E, Mongioi A, Calogero AE, *et al*. Therapy with human chorionic gonadotrophin alone induces spermatogenesis in men with isolated hypogonadotrophic hypogonadism—long-term follow-up. *Int J Androl*, 1992; **15**: 320.
61. Matsumoto AM. Hormonal control of human spermatogenesis. In: Burger H, de Kretser D, eds. *The Tests*, 2nd edn. New York: Raven Press, 1989: 181–96.
62. Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev*, 2000; **21**: 551.
63. Oduwole OO, Vydra N, Wood NEM, *et al*. The overlapping dose-responses of spermatogenic and extragonadal testosterone actions jeopardise the principle of hormonal male contraception. *FASEB J*, 2014; **28**: 2566–76.

10.1.2 Sex Steroid Actions in the Male

Dirk Vanderschueren, Leen Antonio, Na Ri Kim, and Frank Claessens

Introduction

Sex hormones are steroids synthesized from cholesterol, predominantly in the gonads and adrenal glands. Testosterone (T) represents the main circulating androgen and oestradiol (E2) the main circulating oestrogen. Leydig cells in the testis of an adult man produce 5–7 mg testosterone daily (see Figure 10.1.2.1), regulated by luteinizing hormone (LH) from the anterior pituitary gland. Around 5–10% of circulating testosterone is converted to the more potent androgen dihydrotestosterone (DHT) by the 5-alpha-reductase enzymes, thereby amplifying specific androgen actions in specific androgen responsive tissues such as prostate and skin (Figure 10.1.2.1). E2 is converted from androgens (T and androstenedione) via aromatization by the aromatase enzyme; 15–25% of the aromatization occurs in, and is secreted into the circulation, by the testis, and the rest in extragonadal tissues such as adipose tissue. Conversion to E2 accounts for the diversified 'dual mode' of action of androgens. Although only 0.1% of testosterone is converted into E2 and circulating concentrations of E2 are about 100 times lower than those of T, this diversification pathway of androgens is of major physiological importance in males. Sex steroids exert their effect by binding to intracellular nuclear receptors that are present in many tissues and act as ligand-dependent transcription factors. Androgens and oestrogens bind to the ligand-binding domain of the androgen receptor (AR) and oestrogen receptor α or β (ERα/ERβ), respectively. In this chapter, we discuss the physiological role of AR and ER mediated androgen actions throughout life.

Sex Steroid Metabolism and Transport

Conversion of Testosterone to Dihydrotestosterone and Oestradiol

Testosterone has a direct androgenic effect following activation of the AR. This is the most important pathway in muscle (Figure 10.1.2.1) [1]. In other androgen-sensitive tissues such as the prostate, intracellular androgen action is amplified by local enzymatic conversion of T to DHT by the 5-alpha-reductase enzymes (mainly 5ARD 1,2) [2–4]. The higher potency of DHT is believed to be due to higher binding affinity for the AR and more efficient coactivator recruitment. It is likely that this amplification system has evolved to selectively expose tissues requiring high local AR activity (such as prostate or hair follicle) while sparing other tissues in whom such high activities are not necessary or could be harmful. Furthermore, T also functions as a prohormone for E2 in reproductive organs, bone, brain, and adipose tissue, where the aromatase enzyme (CYP19A1) is highly expressed [3, 5].

Other enzymes are also involved in sex steroid synthesis and metabolism in peripheral target tissues. In this tissue-specific sex steroid synthesis, also called intracrinology, adrenal sex steroid precursors with weak androgenic potency, such as dehydroepiandrosterone (DHEA), are converted to more potent androgens as well as oestrogens. While some tissues depend on androgens delivered from the bloodstream, others are more reliant on this intracrine hormone metabolism [4]. In addition, the enzyme activities that inactivate sex steroids through catabolism or conjugation vary considerably even between cell types within the same tissue. Taken together, this implies that sex steroid concentrations measured in serum may not always represent the full gamut of androgen and oestrogen activities at the tissue/cellular level due to the differential local synthesis and metabolism of sex steroids.

Transport of Sex Steroids in Circulation

In the circulation, sex steroids are mainly bound to sex hormone-binding globulin (SHBG) and albumin. Albumin, the most abundant protein in human serum, binds sex steroids with a relatively low affinity (10^{-4}M range), whereas SHBG specifically binds sex steroids, mainly T and E2, with a high affinity (10^{-9}M range), with the affinity of SHBG for T being three times higher than for E2. In normal men, about 55% of T and 30% of E2 is bound to SHBG. Only a small fraction (1–3%) of T and E2 circulates in a non-protein bound or free form [6, 7].

According to the 'free hormone hypothesis', the free fraction is responsible for the biological activities of sex steroids, as only the free unbound hormone can enter the cell and activate its nuclear receptor. However, because sex steroids are only loosely bound to albumin, it is hypothesized that the bound hormone can dissociate from albumin, enter the cell, and may also activate the receptor. Together with the free fraction, the albumin-bound hormone

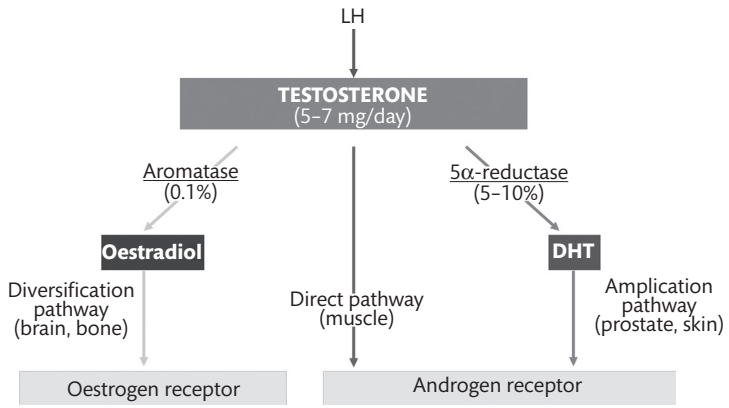


Figure 10.1.2.1 Pathways of testosterone action. The physiological responses to testosterone are either direct via activation of the AR or are modulated via reduction into the more potent DHT (amplification pathway) or via aromatization into oestrogens that act via ERs (diversification pathway).

Adapted and reproduced from PY Liu *et al*: Androgens and Cardiovascular Disease. *Endocr. Rev 2003*. 24: 313–40.

fraction is therefore considered the biologically available (bioavailable) fraction, while the fraction that actually binds to the receptor is called the biologically active (bioactive) fraction (Figure 10.1.2.2) [5]. However, it remains unclear to what extent the bioavailable fraction equates or equilibrates with the bioactive fraction, as intracellular metabolism of the steroid hormone and rate of blood flow in the target tissue also have to be taken into account.

Molecular Mechanisms of Androgen and Oestrogen Actions

Androgens (mainly T and DHT) are highly selective steroid ligands for the AR. The AR (also known as NR3C4) is a member of the nuclear receptor superfamily. The AR is composed of three different functional domains (Figure 10.1.2.3): the central DNA-binding domain (DBD) recognizes androgen response elements (ARE) via two zinc finger modules, the N-terminal domain (NTD) has transactivating properties (AF-1) and the C-terminal ligand-binding domain (LBD) has a ligand-binding pocket and a second activation function (AF-2) [8, 9].

Unliganded AR is primarily located in the cytoplasm. Once ligands, androgens, bind to the AR, the LBD undergoes a conformational change resulting in the dissociation of heat shock proteins (HSPs) from the cytoplasmic AR. This triggers dimerization of the receptor, the exposure of a nuclear localization signal inducing translocation to the nucleus and the induction of the AF-2 surface that is recognized by a number of coactivators (Figure 10.1.2.4). The androgen/AR complex will then recruit these coactivators to AREs located near the androgen target genes in the genome.

The coactivators of the AR function as pioneering factors, histone modifiers, chromatin modifiers, or general transcription factors. Along with AR, actions of coactivators will lead to the upregulation of the expression levels of the target genes, whereas corepressors suppress AR transactivation, resulting in downregulation of certain target genes [10–12]. How the same transcription factor can trigger diverse responses in different tissues can at least in part be explained by the tissue-specific concentration of the ligands and the expression levels of the receptors themselves, but also by the chromatin architecture and the cooperativity of the AR with tissue-specific transcription factors [13].

Apart from the classical, sex steroid-induced gene transcription, androgen and oestrogen actions can also be mediated through non-genomic mechanisms. In such non-genomic actions, the ligand binds to the cytoplasmic receptor, which is then translocated to the plasma membrane, where it can activate other signal transduction pathways (Figure 10.1.2.4) [14]. In contrast to the slow classical nuclear receptor actions, requiring hours to days, these non-genomic actions result in rapid effects, occurring within seconds or minutes of ligand binding. The best-documented event is the triggering of the release of intracellular calcium, which can lead to activation of kinases such as mitogen-activated protein kinase (MAPK), protein kinase A (PKA), Akt and protein kinase C (PKC). Ultimately, this process indirectly affects gene transcription through phosphorylation and subsequent protein synthesis. The exact role of the membrane-initiated events in androgen-related physiology and pathologies still needs to be established.

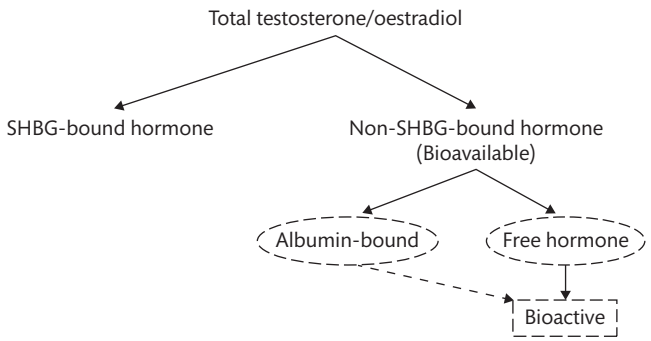


Figure 10.1.2.2 Sex steroid fractions in human serum.

Ontogeny of Androgen Action

Testosterone levels change considerably during the male life span (Figure 10.1.2.5).

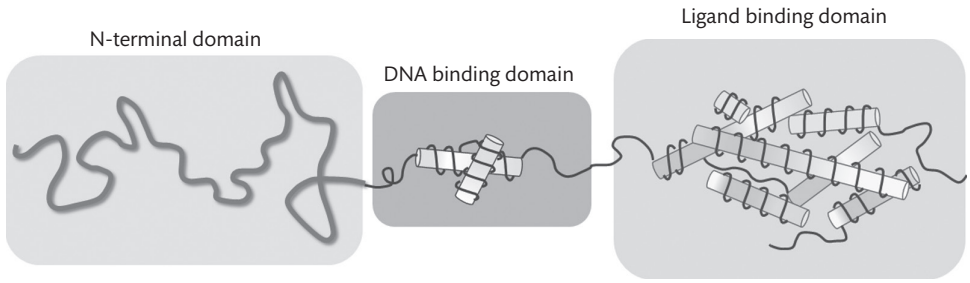


Figure 10.1.2.3 Schematic model of the modular structure of the AR, with an unstructured NTD, a central DBD, and a carboxyterminal ligand-binding domain.

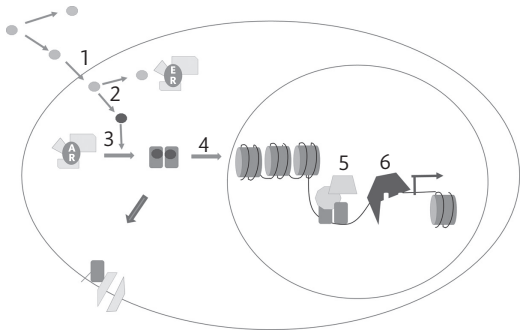


Figure 10.1.2.4 Multiple steps in the nuclear and membrane-initiated actions of androgen receptor: (1) Testosterone (orange) enters the cell where it can be converted to DHT (2) (red) or oestrogen (green). When T or DHT bind to AR, HSPs (light blue) dissociate, the receptor is dimerizing (3) and enters the nucleus (4), where it will bind AREs (5), recruiting coactivators (grey) ultimately leading to increased transcription of androgen target genes by RNA polymerase II (dark blue). *For a colour version of this figure, please see colour plate section.*

Antenatal

Around the sixth week of embryogenesis, the undifferentiated gonads develop into testes (see also Chapter 7.2.1 and 10.1.1). This is regulated by testis-determining factor, a transcription factor encoded by the sex-determining region of the Y chromosome. Around week 8, the fetal Leydig cells start producing T. This prenatal T exposure determines further differential sex development in the fetus. T induces differentiation of the Wolffian ducts into male reproductive organs such as epididymis, vas deferens, and seminal vesicles. Additionally, immature Sertoli cells produce anti-Müllerian hormone (AMH). This triggers the regression of Müllerian ducts, eventually inhibiting the formation of uterus and fallopian tubes in the male fetus [15–17].

DHT is important for differentiation and development of male external genitalia during embryogenesis [18, 19]. This is illustrated by individuals with mutations in the gene encoding 5α-reductase, resulting in absent or very low levels of DHT. Although these patients have a 46, XY-karyotype, they are born with female external genitalia. As the testes can produce normal levels of T, internal male urogenital structures develop normally in these individuals. Moreover, increase of T at the onset of puberty leads to development of typical male features, such as an increase in muscle mass and voice deepening [20].

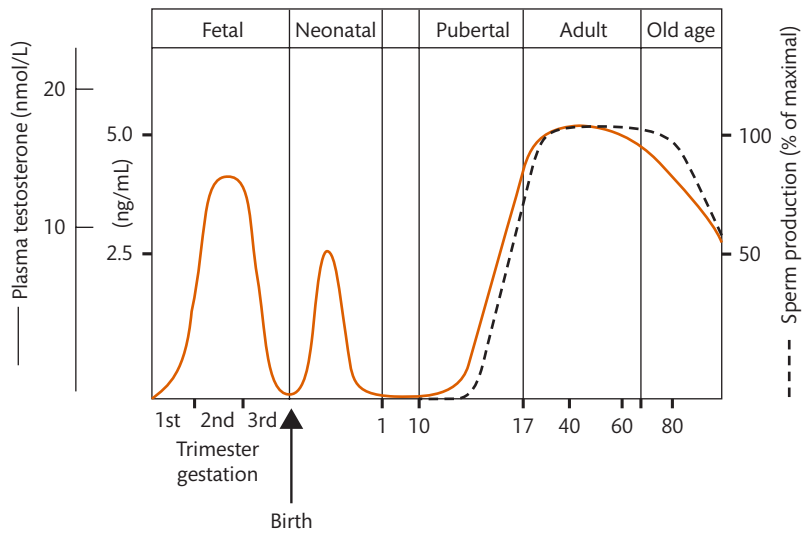


Figure 10.1.2.5 Concentrations of testosterone and appearance of spermatogenesis across the male lifespan.
Reproduced from JE Griffin *et al*: The Testis. In: Bondy PK, Rosenberg LE (eds). *Metabolic control and disease*, 8th edition. Philadelphia, Pa: WB Saunders, 1980:1535–78.

Individuals with inactivating mutations in the AR also have a 46, XY-karyotype, but are born with a female external phenotype due to abrogation of both DHT and T mediated effects antenatally. In these patients with the complete androgen insensitivity syndrome (CAIS), T levels are normal or even high, but the Wolffian ducts do not differentiate into epididymis and vas deferens. As AMH action is normal, a uterus and fallopian tubes are absent. For a detailed discussion on disorders of sex development, see Chapter 7.2.2.

Neonatal and Childhood

At birth, hypothalamic gonadotropin-releasing hormone (GnRH) neurons are transiently activated, probably due to cessation of the negative feedback by placental oestrogens. This triggers transient activation of the hypothalamic–pituitary–gonadal (HPG) axis, resulting in a second peak of T levels (mini-puberty). Androgen exposure in the neonatal period is important for growth in childhood and reproductive function in later life [21, 22].

After these two peaks in early development, circulating T remains low until puberty. Moreover, SHBG levels are very high in childhood, resulting in even lower free sex steroid concentrations in prepubertal children.

Puberty

Upon entry into puberty, T secretion resumes and SHBG levels gradually decrease, contributing to a progressive increase in free sex steroids. High androgen levels induce development of secondary sexual characteristics and axial skeleton growth. Increased testicular size, thinning of the scrotum, and growth of pubic hair are the most obvious physical signs of pubertal sexual development. In addition, increasing levels of T lead to an increase in muscle mass and deepening of voice. T deficiency during puberty leads to absence of or insufficient secondary sexual development as well as infertility (i.e. hypogonadism in adulthood if persistent). This is covered in more detail in Chapter 7.3.2.

Puberty is a critical period for peak bone mass acquisition and establishment of sexual dimorphism in bone mass and geometry and overall physique. Together with T, oestrogens play an important role in bone development during male puberty. This became evident from case reports of men with aromatase deficiency. Aromatase deficiency is caused by a loss-of-function mutation in the *CYP19A1* gene, resulting in impaired oestrogen synthesis and very low or undetectable E2 levels [23]. Without oestrogen action, the long bone epiphyses do not close and the growth termination from epiphyseal fusion expected at the end of puberty is lacking. This results in continuous linear bone growth and men with aromatase deficiency therefore present with tall stature and eunuchoid body proportions. Furthermore, these men fail to achieve adult peak bone mass leading to osteopenia or osteoporosis and skeletal deformities such as scoliosis and genu valgum. Oestrogen treatment results in a rapid completion of bone maturation, epiphyseal closure, and growth cessation. Furthermore, oestrogens can increase bone mineral density. Other features of the syndrome include impaired spermatogenesis as well as the metabolic syndrome with abdominal obesity, insulin resistance, and dyslipidaemia [23].

One case report of a man with a mutation in ERα reports a similar phenotype [24], which confirms animal data that the major receptor for oestrogen action in the male is ERα. Thus far, no men with ERβ mutations have been described.

Adulthood

Circulating T levels reach their peak in early adulthood. This is then followed by a gradual decline in T levels in subsequent decades, with a concomitant increase of SHBG concentrations [25–27]. This age-related decline in T is gradual and lacks a sharp inflection point akin to the menopause in females. The extent to which the age-related decline in T is due to chronological ageing alone or to the accumulation of age-related comorbidities remains however incompletely understood (see Chapter 10.3.5).

Obesity is an important contributor to low T levels in adult men. Severely obese men (body mass index (BMI) > 40 kg/m^2) have lower serum total and free T levels compared to men with normal BMI [28] and their T concentrations drop more quickly [29]. Also chronic illnesses, such as cardiovascular disease and type 2 diabetes, are associated with lower testosterone levels in adult men. Similarly, SHBG levels can be affected by multiple hormonal and metabolic factors (see Chapter 10.2.2).

Androgens play a key role in the physiology of the adult male. However, recent studies with experimentally induced androgen and oestrogen deficiency in healthy young men made it clear that oestrogens also play a fundamental role in several target organs, responsible for effects that were traditionally attributed to androgens [30]. Androgen and oestrogen mediated effects are summarized in **Table 10.1.2.1**.

High intratesticular testosterone levels are essential for normal spermatogenesis [31]. On the other hand, sexual function is mediated by both androgens and oestrogens. Experimentally induced short-term sex steroid deficiency in healthy young men resulted in a decrease in sexual desire and erectile function. When these men were treated with T replacement therapy together with an aromatase inhibitor, thereby normalizing T levels but maintaining E2 deficiency, sexual desire and erectile function remained decreased compared to men treated with T only [30]. Furthermore, androgen deficiency also induced a decrease in lean mass and muscle strength [30], whereas E2 deficiency induced a decrease in bone mineral density [32].

E2 is not only important for sexual/reproductive function and skeletal development, but it also regulates hypothalamic–pituitary–thyroid (HPT) axis activity, both at hypothalamic and pituitary level. In humans, the relevance of androgen aromatization to E2 for normal gonadotropin feedback has been studied in men with idiopathic hypothalamic hypogonadism (IHH) who were treated with pulsatile GnRH to normalize gonadotropin levels [33]. Not only T but also E2 infusion resulted in decreased levels of LH and follicle-stimulating hormone (FSH), which indicates that E2 plays a role in mediating negative feedback actions on gonadotrophin secretion. It is known that E2 exerts feedback regulation via ERα expressing KNDy (Kisspeptin/Neurokinin B/Dynorphin) neurons in arcuate nucleus of hypothalamus [34]. Furthermore, transdermal E2 treatment decreased LH pulse frequency, LH pulse amplitude, and GnRH-stimulated LH secretion in men suffering from aromatase deficiency, again illustrating the importance of oestrogens for hypothalamic-pituitary feedback in men [35].

Table 10.1.2.1 The differential role of androgens and oestrogens in male physiology

Sex steroid target organ	Androgens		Oestrogens E2	Remarks	Ref
	T	DHT			
Reproduction					
Spermatogenesis	+		+	Intratesticular T levels are crucial for normal spermatogenesis. Impaired spermatogenesis is also observed in men with aromatase deficiency.	[23, 31]
Sexual desire Erection	+		+	Experimentally induced short-term sex steroid deficiency in healthy young men resulted in a decrease in sexual desire and erectile function. When these men were treated with T replacement therapy together with an aromatase inhibitor, thereby normalizing T levels but maintaining E2 deficiency, sexual desire, and erectile function remained decreased compared to men treated with T only.	[30]
Prostate		+		DHT mediates prostate growth and the development of benign prostate hyperplasia. Treatment with a 5α-reductase inhibitor results in a decrease of prostate volume.	[36]
Bone and body composition					
Maintenance of bone mass	+		++	Severe oestrogen deficiency leads to a decline in bone mineral density. In patients with aromatase deficiency, oestrogen treatment restores bone mineral density. Murine models with osteoblasts and osteocytes specific knockouts of AR also revealed a direct role for androgens on trabecular bone, independent of aromatization.	[23, 32, 37] [38, 39]
Maintenance of muscle mass and strength	+			Androgen deficiency leads to a decrease in lean mass, muscle strength, and muscle size.	[30]
Body fat			+	Oestrogen deficiency results in an increase in body fat.	[30]
Hypothalamic-pituitary feedback	+		+	Both T and E2 provide hypothalamic-pituitary feedback.	[40]
Central nervous system					
Gender-related behaviour	+		+	Early prenatal androgen exposure contributes to the development of gender-related behaviour. Murine models suggest neonatal oestrogen exposure is important for sexually dimorphic behaviour in adult life.	[41–43]
Aggression	+			Serum T levels showed a low, but positive correlation with aggressive behaviour and higher correlation with dominance.	[44]
Skin and hair					
Sebum production	+	+		Androgens stimulate sebum production and are a contributing factor in the development of acne vulgaris.	[45]
Pubic and axillary hair growth	+			Due to the rise of androgen levels in puberty, axillary, and pubic hair starts growing. Patients with CAIS have no pubic and axillary hair.	[45]
Facial hair growth	+	++		Due to the rise of androgen levels in puberty, facial hair starts growing. Patients with CAIS or 5α-reductase deficiency have absent or sparse facial hair.	[45]
Male balding		+		Androgens suppress hair growth on the frontal and vertex scalp, leading to male balding. Patients with CAIS or 5α-reductase deficiency do not develop male balding.	[45]
Erythropoiesis	+		?	Androgens stimulate haemoglobin synthesis and erythropoietin synthesis. Exogenous testosterone suppresses hepcidin, resulting in a higher haematocrit.	[46] [47]

REFERENCES

1. Yeap BB. Androgens and cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes*, 2010; **17**: 269–76.
2. Callewaert F, Boonen S, Vanderschueren D. Sex steroids and the male skeleton: a tale of two hormones. *Trends Endocrinol Metab*, 2010; **21**: 89–95.
3. Labrie F. All sex steroids are made intracellularly in peripheral tissues by the mechanisms of intracrinology after menopause. *J Steroid Biochem Mol Biol*, 2015; **145**: 133–8.
4. Swerdloff RS, Dudley RE, Page ST, Wang C, Salameh WA. Dihydrotestosterone: biochemistry, physiology, and clinical implications of elevated blood levels. *Endocr Rev*, 2017; **38**: 220–54.
5. Hammond GL. Diverse roles for sex hormone-binding globulin in reproduction. *Biol Reprod*, 2011; **85**: 431–41.
6. Mendel CM. The free hormone hypothesis: a physiologically based mathematical model. *Endocr Rev*, 2015; **10**: 232–74.
7. Hammond GL, Wu TS, Simard M. Evolving utility of sex hormone-binding globulin measurements in clinical medicine. *Curr Opin Endocrinol Diabetes Obes*, 2012; **19**: 183–9.

8. Claessens F, Denayer S, Van Tilborgh N, Kerkhofs S, Helsen C, Haelens A. Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nucl Recept Signal*, 2008; **6**: e008.
9. Helsen C, Kerkhofs S, Clinckemalie L, *et al.* Structural basis for nuclear hormone receptor DNA binding. *Mol Cell Endocrinol*, 2012; **348**: 411–17.
10. Moras D, Gronemeyer H. The nuclear receptor ligand-binding domain: structure and function. *Curr Opin Cell Biol*, 1998; **10**: 384–91.
11. Stephens RG, Kogon SL. New U.S. guidelines for prescribing dental radiographs—a critical review. *J Can Dent Assoc*, 1990, **56**: 1019–24.
12. Heemers HV, Tindall DJ. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev*, 2007; **28**: 778–808.
13. Pihlajamaa P, Sahu B, Lyly L, Aittomäki V, Hautaniemi S, Jänne OA. Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. *EMBO J*, 2014; **33**: 312–26.
14. Hammes SR, Levin ER. Minireview: recent advances in extranuclear steroid receptor actions. *Endocrinology*, 2011; **152**: 4489–95.
15. Waites G. The testis. In: Bondy PK, Rosenberg LE, eds. *Metabolic Control and Disease*. Philadelphia, PA: WB Saunders, 1970: 241–79.
16. Griffin JE, Wilson JD. Disorders of androgen receptor function. *Ann NY Acad Sci*, 1984; **438**: 61–71.
17. Josso N, Racine C, Di Clemente N, Rey R, Xavier F. The role of anti-Müllerian hormone in gonadal development. *Mol Cell Endocrinol*, 1998; **145**: 3–7.
18. Wachtel SS, Wachtel G, Nakamura D. Sexual Differentiation. *Vertebr Endocrinol Fundam Biomed Implic*, 1991; **26**: 149–80.
19. Nef S, Parada LF. Hormones in male sexual development. *Genes Dev*, 2000; **14**: 3075–86.
20. Okeigwe I, Kuohung W. 5-Alpha reductase deficiency: a 40-year retrospective review. *Curr Opin Endocrinol Diabetes Obes*, 2014; **21**: 483–7.
21. Kuiri-Hänninen T, Seuri R, Tyrväinen E, *et al.* Increased activity of the hypothalamic-pituitary-testicular axis in infancy results in increased androgen action in premature boys. *J Clin Endocrinol Metab*, 2011; **96**: 98–105.
22. Kiviranta P, Kuiri-Hänninen T, Saari A, Lamidi ML, Dunkel L, Sankilampi U. Transient postnatal gonadal activation and growth velocity in infancy. *Pediatrics*, 2016; **138**: e20153561.
23. Rochira V, Carani C. Aromatase deficiency in men: a clinical perspective. *Nat Rev Endocrinol*, 2009; **5**: 559–68.
24. Smith EP, Boyd J, Frank GR, *et al.* Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med*, 1994; **331**: 1056–61.
25. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocr Rev*, 2005; **26**: 833–76.
26. Camacho EM, Huhtaniemi IT, O'Neill TW, *et al.* Age-associated changes in hypothalamic-pituitary-testicular function in middle-aged and older men are modified by weight change and lifestyle factors: longitudinal results from the European Male Ageing Study. *Eur J Endocrinol*, 2013; **168**: 445–55.
27. Feldman HA, Longcope C, Derby CA, *et al.* Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *J Clin Endocrinol Metab*, 2002; **87**: 589–98.
28. Wu FCW, Tajar A, Pye SR, *et al.* Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metab*, 2008; **93**: 2737–45.
29. Bohne H, Wanjiku J, Rathke AK, Humpert S, Gerhard D. Influence of nutrient loading on growth, N-, P- and K-concentrations and outplanting performance of Rosa majalis. *Eur J Hortic Sci*, 2014; **79**: 29–35.
30. Finkelstein JS, Lee H, Burnett-Bowie SM, *et al.* Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*, 2013; **369**: 1011–22.
31. Dohle GR, Smit M, Weber RFA. Androgens and male fertility. *World J Urol*, 2003; **21**: 341–5.
32. Finkelstein JS, Lee H, Leder BZ, *et al.* Gonadal steroid-dependent effects on bone turnover and bone mineral density in men. *J Clin Invest*, 2016; **126**: 1114–25.
33. Bagatell CJ, Dahl KD, Bremner WJ. The direct pituitary effect of testosterone to inhibit gonadotropin secretion in men is partially mediated by aromatization to estradiol. *J Androl*, 1994; **15**: 15–21.
34. Lehman MN, Coolen LM, Goodman RL. Minireview: kisspeptin/neurokinin B/dynorphin (KNDy) cells of the arcuate nucleus: a central node in the control of gonadotropin-releasing hormone secretion. *Endocrinology*, 2010; **151**: 3479–89.
35. Rochira V, Zirilli L, Genazzani AD, *et al.* Hypothalamic–pituitary–gonadal axis in two men with aromatase deficiency: evidence that circulating estrogens are required at the hypothalamic level for the integrity of gonadotropin negative feedback. *Eur J Endocrinol*, 2006; **155**: 513–22.
36. Wilson JD. The critical role of androgens in prostate development. *Endocrinol Metab Clin N Am*, 2011; **40**: 577–90.
37. Vanderschueren D, Laurent MR, Claessens F, *et al.* Sex steroid actions in male bone. *Endocr Rev*, 2014; **35**: 906–60.
38. Notini AJ, McManus JF, Moore A, *et al.* Osteoblast deletion of exon 3 of the androgen receptor gene results in trabecular bone loss in adult male mice. *J Bone Miner Res*, 2007; **22**: 347–56.
39. Sinnesael M, Claessens F, Laurent M, *et al.* Androgen receptor (AR) in osteocytes is important for the maintenance of male skeletal integrity: evidence from targeted AR disruption in mouse osteocytes. *J Bone Miner Res*, 2012; **27**: 2535–43.
40. Pitteloud N, Dwyer AA, DeCruz S, *et al.* Inhibition of luteinizing hormone secretion by testosterone in men requires aromatization for its pituitary but not its hypothalamic effects: evidence from the tandem study of normal and gonadotropin-releasing hormone-deficient men. *J Clin Endocrinol Metab*, 2008; **93**: 784–91.
41. Hines M. Gender development and the human brain. *Annu Rev Neurosci*, 2011; **34**: 69–88.
42. Wu MV, Manoli DS, Fraser EJ, *et al.* Estrogen masculinizes neural pathways and sex-specific behaviors. *Cell*, 2009; **139**: 61–72.
43. Honda SI, Harada N, Ito S, Takagi Y, Maeda S. Disruption of sexual behavior in male aromatase-deficient mice lacking exons 1 and 2 of the cyp19 gene. *Biochem Biophys Res Commun*, 1998; **252**: 445–9.
44. Archer J. Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci Biobehav Rev*, 2006; **30**: 319–45.
45. Ceruti JM, Leirós GJ, Balañá ME. Androgens and androgen receptor action in skin and hair follicles. *Mol Cell Endocrinol*, 2018; **465**: 122–33.
46. Shahani S, Braga-Basaria M, Maggio M, Basaria S. Androgens and erythropoiesis: past and present. *J Endocrinol Investig*, 2009; **32**: 704–16.
47. Bachman E, Feng R, Travison T, *et al.* Testosterone suppresses hepcidin in men: a potential mechanism for testosterone-induced erythrocytosis. *J Clin Endocrinol Metab*, 2010; **95**: 4743–7.

10.2

Evaluation of the Male Patient with Suspected Hypogonadism and/or Infertility

10.2.1 Clinical Evaluation

Bradley D. Anawalt

Clinical Evaluation

Although male hypogonadism (androgen deficiency and/or infertility due to a defect in spermatogenesis) are common in men, there is no value in screening for these disorders [1]. There are a number of questionnaires available for screening or case-finding, but these instruments should be reserved for research purposes [1]. The clinician should evaluate men for androgen deficiency when they present with suggestive symptoms such as low or decreased libido, new onset of erectile dysfunction or gynaecomastia, infertility (attempting to conceive with at least weekly vaginal intercourse without contraception for at least 12 months), or unexplained muscular weakness (**Box 10.2.1.1**). Male androgen deficiency is associated with many vague and non-specific symptoms that overlap with depression, anxiety, and psychosomatic disorders; sexual symptoms (e.g. decreased libido, decreased morning erections, and decreased penile tumescence) appear to be most specific for androgen deficiency [2]. Men with signs of androgen deficiency such as lack of full pubertal development (e.g. small testes), osteoporosis, or unexplained anaemia should be evaluated for hypogonadism even in the absence of suggestive symptoms [3–5].

When androgen deficiency or infertility is suspected or confirmed in a man, the remainder of the history and the physical examination should be focused on determination of the following: (1) the onset and duration of androgen deficiency or infertility; (2) the cause of androgen deficiency or infertility; (3) and determination of the likely benefit and risk of hormonal therapy (testosterone or gonadotropin replacement therapy).

The timing of onset of male hypogonadism has important diagnostic and therapeutic implications. For adult men who present with prepubertal or congenital onset of hypogonadism, hormonal therapy may cause dramatic changes in facial appearance, musculature, and bone density, but it is unlikely to fully restore all end-organ function to normal compared to age-matched peers. For example, testosterone replacement therapy will not increase bone mineral density in men who are diagnosed as adults with congenital hypogonadism to that of age-matched peers who are eugonadal [6, 7]. Likewise, gonadotropin therapy is more likely to be rapidly effective for men with acquired, postpubertal hypogonadotropic hypogonadism than for men with congenital hypogonadotropic hypogonadism (e.g. Kallmann syndrome) [8]. In addition, men with congenital hypergonadotropic or hypogonadotropic hypogonadism often have syndromes with extragonadal sequelae (e.g. increased risk of diabetes mellitus type 1 and 2 and autoimmune thyroid disease in Klinefelter syndrome or anosmia and single 'horseshoe' kidney in Kallmann syndrome associated with *ANOS1* [*KAL1*] mutation) [9, 10].

The cause of hypogonadism is either primary (testicular) or secondary (dysfunction or disease of the hypothalamus and pituitary). The distinction between primary and secondary hypogonadism

Box 10.2.1.1 Symptoms and signs of adult male hypogonadism

Specific, but uncommon, symptoms and signs
- Incomplete or delayed sexual development
- Inability to conceive after ≥12 months of frequent intercourse without contraception
- Loss or decreased growth of androgen-dependent body hair
  Small testes (<10 ml)

Suggestive symptoms and signs
- Decreased libido, spontaneous erections, and sexual activity
- New onset of breast growth and tenderness
- Low trauma fracture or significant height loss
- Hot flashes

is important diagnostically; men with primary hypogonadism might benefit from karyotyping for Klinefelter syndrome whereas men with secondary hypogonadism must have an evaluation for diseases and medications that affect hypothalamic and pituitary function [1]. Determination of the cause of hypogonadism is also important therapeutically; infertility due to primary hypogonadism cannot be treated medically whereas the majority of men with hypogonadotropic hypogonadism can have fertility restored or improved with gonadotropin therapy [8].

The history and physical examination can be useful in determining the onset and cause of hypogonadism. The history and examination are also indispensable in determining the potential benefits and risks of testosterone or gonadotropin replacement therapy. For example, a man who presents during adulthood with congenital hypogonadism is likely to have decreased spermatogenesis that will increase but not normalize with gonadotropin replacement therapy. On the other hand, a man who presents with recent onset of postpubertal hypogonadism is less likely to have evidence of end-organ harm from androgen deficiency; for this man, testosterone replacement might not cause measurable changes in androgen-dependent tissues, but it will maintain androgen-dependent organ function within the normal range. The history and physical examination might also identify specific risks of androgen therapy. For example, testosterone therapy might cause accelerated male pattern balding in men with a family history of androgenic alopecia or increased risk of prostate cancer in men with risk factors for prostate cancer.

Specific Elements of the History and Physical Examination for All Men Suspected of Having Androgen Deficiency or Infertility

All adult men with suspected androgen deficiency or male infertility should be queried about the following:

To determine the onset of hypogonadism: The man should be asked about the onset and achievement of pubertal milestones—were they are at normal age milestones compared to peers? How long have symptoms and signs of hypogonadism been present? Did the man have a decline in libido? (Men with untreated prepubertal hypogonadism have never had a normal libido and will not report a decline.) Did the man ever father a child? (It is true that such history would not prove previous normal fertility as the man might not be the biological father, but the information is still useful.)

To determine the cause of hypogonadism or male infertility: The man should be asked about exposure to testicular toxins (e.g. alkylating chemotherapeutics), testicular trauma or infection (e.g. mumps orchitis). The man should also be asked about symptoms or signs of sellar masses (new onset of blurry vision or headaches; symptoms of other pituitary hormone hypo- or hyperfunction) and risk factors for iron overload syndromes (family history of hemochromatosis or personal history of blood transfusions). Finally, the man should be queried about exogenous prescription or nonprescription use of corticosteroids, opioids, anabolic androgenic steroids, or medications that cause hyperprolactinemia and about symptoms suggestive of corticosteroid excess (easy bruisability or muscle weakness).

Men with suspected male hypogonadism should have a physical examination that includes the following:

Genital examination: The most important part of the physical examination in man with suspected male hypogonadism is the genitourinary examination. The penis should be examined for normal location of the urethra opening. The vasa deferens should be palpated. Absence of one or both of the vasa suggests cystic fibrosis, a cause of male infertility [4]. The testes should be measured with a Prader orchidometer; Prader orchidometry correlates well with ultrasonographic measurements although ultrasonographic measurements are usually slightly smaller [11]. Testes that are each 3 ml or smaller (by Prader orchidometry) indicate that the onset of hypogonadism occurred before the onset of puberty and usually indicates primary hypogonadism due to Klinefelter syndrome or congenital hypogonadotropic hypogonadism due to a genetic mutation.

The remainder of the exam should be focused on evidence of androgen effects. The skin should be examined for acne and hair in androgen-dependent areas (e.g. face, chest, and below the umbilicus). Acne suggests that the skin has recently been exposed to androgens. Hair in androgen-dependent areas indicates that the skin has had some lifetime exposure to androgens. It takes several months of androgen exposure to cause hair growth, but several years of hypogonadism to decrease hair in androgen-dependent areas. The chest should be examined for breast tissue (see Chapter 10.3.7, 'Gynaecomastia'). Breast tissue is firm and rubbery and discoid around the areolae; it can be discriminated by the firm edge of the breast tissue that can flipped up from surrounding fat that is soft. The muscles should be examined for strength and mass; profound, longstanding hypogonadism may cause decreased muscle mass and strength. A man who is infertile and has well-developed musculature but low serum testosterone and gonadotropin concentration might be using androgenic anabolic steroids.

In specific settings, the clinician should examine for unusual physical findings. In men with small testes and suspected congenital hypogonadotropic hypogonadism, the man should be examined for eunuchoidal proportions that is most easily detected as arm span that is at least 5 cm greater than the height of the man. In addition, men with congenital hypogonadotropic hypogonadism should be queried or tested for decreased sense of smell, coloboma, or dysdiadokinesis; these findings suggest specific heritable genetic mutations (e.g. Kallmann syndrome due to *ANOS1* mutation or CHARGE syndrome). Men with postpubertal onset of hypogonadotropic hypogonadism should be examined for signs of Cushing syndrome including thin skin with diffuse ecchymoses or broad purple striae and proximal muscle weakness (e.g., inability to rise from the seated postion without using arms to push or pull up).

Conclusions

A history and physical examination is useful for helping determine the onset of hypogonadism and may be useful in determining specific causes of hypogonadism or infertility. For example, a man who presents with small testes, decreased or absent sense of smell, and a coloboma likely has a genetic mutation causing congenital hypogonadotropic hypogonadism. A man with infertility and absence of one or both vasa deferentia has cystic fibrosis. Finally, the history and exam provide important clues about the likelihood of therapeutic benefit and adverse effects of testosterone or gonadotropin therapy. For example, an infertile man with hypogonadotropic

hypogonadism, testes that each measure 6 ml, and a family history of androgenic balding is likely to have improved fertility with gonadotropin therapy but might experience male pattern baldness with any form of androgen replacement therapy. Thus, a careful history and physical examination is an important component of the evaluation of a man suspected of having hypogonadism or infertility.

REFERENCES

1. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
2. Wu FC, Tajar A, Beynon JM, *et al*. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*, 2010; **363**: 123–35.
3. Watts NB, Adler RA, Bilezikian JP, *et al*. Osteoporosis in men: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 1802–22.
4. Snyder PJ, Kopperdahl DL, Stephens-Shields AJ, *et al*. Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone. *JAMA Int Med*, 2017; **177**: 471–9.
5. Roy CN, Snyder PJ, Stephens-Shields AJ, *et al*. Association of testosterone levels with anemia in older men. *JAMA Int Med*, 2017; **177**: 480–90.
6. Finkelstein JS, Klibanski A, Neer RM, *et al*. Increases in bone density during treatment of men with idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 1989; **69**: 776–83.
7. Tahani N, Nieddu L, Prossomariti G, *et al*. Long-term effect of testosterone replacement therapy on bone in hypogonadal men with Klinefelter syndrome. *Endocrine*, 2018; **61**: 327–35.
8. Anawalt BD. Approach to male infertility and induction of spermatogenesis. *J Clin Endocrinol Metab*, 2013; **98**: 3532–42.
9. Groth KA, Skakkebæk A, Høst C, Gravholt CH, Bojesen A. Clinical review: Klinefelter syndrome—a clinical update. *J Clin Endocrinol Metab*, 2013; **98**: 20–30.
10. Boehm U, Bouloux PM, Dattani MT, *et al*. Expert consensus document: European consensus statement on congenital hypogonadotropic hypogonadism—pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol*, 2015; **11**: 547–64.
11. Sakamoto H, Saito K, Ogawa Y, Yoshida H. Testicular volume measurements using Prader orchidometer versus ultrasonography in patients with infertility. *Urology*, 2007; **69**: 158–62.

10.2.2 Endocrine Evaluation

Jean-Marc Kaufman

Men Evaluated for Suspected of Hypogonadism

Making the diagnosis of hypogonadism requires the documentation of deficient testosterone (T) production as reflected by low T blood concentrations [1]. After having established that the patient is T deficient, further hormonal evaluation is needed to determine the underlying cause of hypogonadism with measurement of serum gonadotropins to differentiate secondary hypogonadism of hypothalamic-pituitary origin from primary testicular failure as the next step. Additional hormonal measurements may be of help to refine the diagnosis. Optimally, planning of the endocrine investigations is guided by the findings of the clinical evaluation. The clinical and hormonal findings jointly provide guidance as to the need of additional investigations such as imaging and genetic testing. In general, there is a rather straight forward logic sequence of hormonal investigations in men with suspected hypogonadism. The approach is summarized in Figure 10.2.2.1.

Hormonal Diagnosis of Hypogonadism

The diagnosis of hypogonadism is based on the presence of symptoms and signs of hypogonadism and a consistently and unequivocally low serum total T and/or free T [1]. Some conditions to be fulfilled to reliably demonstrate a low serum T concentration and thus deficient T production, are summarized in Box 10.2.2.1.

Measurement of Serum T Level

In order to account for the physiologic circadian variation of serum T levels in healthy men, with usually ≥ 40% difference between peak T levels in the morning compared to nadir T levels in late afternoon to early evening [2], blood sampling to confirm a clinical suspicion of hypogonadism should be performed in the morning, ideally between 7:00 and 10:00 a.m. Although the circadian variations of serum T concentrations tend to be blunted in ageing men, this effect nevertheless persists partially and variably so that sampling in older men should also be performed in the morning. Sampling should be performed in fasting state because of reported decreases in blood T concentrations of around 25–30% after food intake [3, 4]. Because T is secreted episodically with resulting within and between-day variations in T levels, the finding of a low T should be confirmed by demonstration of low T in a second blood sample. Acute stress (e.g. surgery) and disease can result in transient marked suppression of T levels, so that blood sampling for diagnosis of hypogonadism should not be performed during or soon after such events (i.e. preferably not earlier than 2–4 weeks following full recovery). Moreover, to account for possible transient T decrease of any cause, the second blood sample for confirmation of low T should be obtained on a separate occasion, ideally with an interval of at least 2 weeks after the initial sampling. Conversely, a single clearly normal serum T generally excludes significant hypogonadism. Only occasionally, e.g. in symptomatic men with borderline low-normal serum T and evidence for a disorder of the gonadal axis (e.g. elevated gonadotropins), one may want to confirm the normal T in a second sample.

To measure T levels in serum, mass-spectrometry-based methods which include serum extraction and a chromatography step, i.e. gas chromatography-mass spectrometry (GC-MS) and mainly liquid

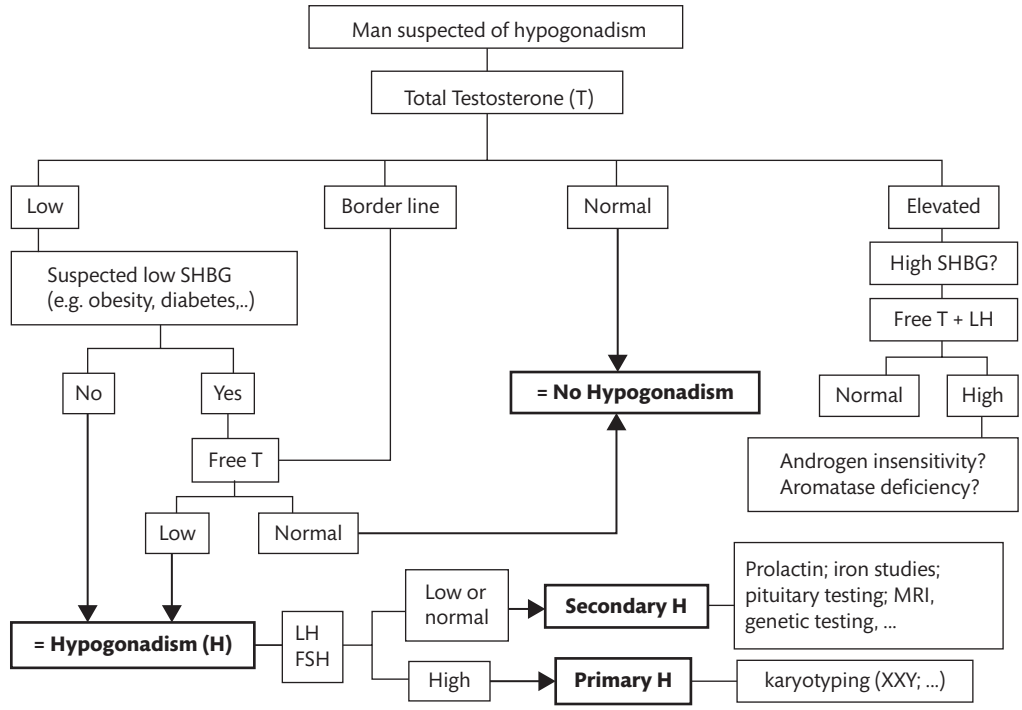


Figure 10.2.2.1 Flow-chart for hormonal evaluation of men suspected of hypogonadism.

chromatography-tandem mass spectrometry (LC-MS/MS), are now considered the preferred state-of-the-art methods, whereas automated immunoassay platforms are the most commonly used methods by clinical chemistry laboratories. The radioimmunoassay methods for T introduced in the early 1970s included serum or plasma extraction and chromatographic purification steps before immunoassay, which improved specificity by removing potentially cross-reacting steroids and allowed increased sensitivity by extracting larger volumes of serum. Evolutions in the past decades have seen radio-immunoassays mostly replaced by alternative immunoassay techniques (e.g. chemiluminescence-immunoassays), and the extraction and chromatography steps have been omitted with resultant diminished specificity as a trade-off for gains in time, costs, and suitability for automation in multianalysers. These automated platform immunoassays have been criticized for poor assay performance [5, 6]. Clinically, this is particularly an issue in children and women, where these immunoassays cannot reliably quantify the low serum T concentrations. By contrast, most T immunoassays in commonly used automated platforms perform reasonably well for routine diagnosis of hypogonadism in men [7, 8], even if there are some issues of calibration and accuracy. Irrespective of whether immunoassay or LC-MS/MS is used, proper assay validation and ongoing quality control in individual clinical laboratories is essential and should include the use of a reference range validated for the particular laboratory conditions.

Poor standardization of T assays limiting comparison of results across assay platforms has been an issue of concern. In response the USA Centers for Disease Control and Prevention (CDC) Clinical Reference Laboratory developed a reference method calibrated with higher order reference material and implemented a standardization and certification programme for T assays open to subscription by external laboratories.

There is no consensus on what is considered a subnormal T level, with proposed cut-off levels varying from 8 to 12 nmol/L (230–350 ng/dl) according to different recommendations and guidelines from scientific societies [1, 9, 10]. Moreover, reference ranges for serum T levels vary between laboratories due to methodological issues related to both T assay calibration and how the reference data

Box 10.2.2.1 Points to consider when assessing serum testosterone (T) for diagnosis of hypogonadism

Pre-analytical conditions

- Blood sampling performed in the morning (preferably between 7 and 10 a.m.)
- Blood sampling after overnight fast
- No sampling during- and earlier than 2–4 weeks after full recovery from acute physically stressful event or disease
- No sampling after night work shift

Analysis and interpretation

- Measurement of serum total T with a well validated assay: LC-MS/MS (ideal) or immunoassay
- Use of a valid reference range (laboratory-validated or through cross-calibration with external assay)
- Assessment of serum-free T if total T borderline low/normal (e.g. 200–400 ng/dl or 7–14 nmol/L) in particular in clinical situations with altered SHBG (see Box 10.2.2.2)
- Free T ideally measured by a validated equilibrium dialysis- or ultrafiltration-based technique, but limited access
- Free T estimated by calculation from total T, SHBG (and albumin) using a validated equation (see text): acceptable practical alternative but with limitations
- Low (free) T to be confirmed in repeat sample, preferably after an interval (weeks) allowing for recovery of possible transient cause of low (free)T
- Single clearly normal (free) T generally excludes significant hypogonadism
- Occasionally, borderline low-normal (free)T may deserve confirmation in repeat sample if clinic suggestive of deficiency and other indications of disorder of gonadal axis (e.g. elevated gonadotropins)

was generated. In the 2018 update of the US Endocrine Society clinical practice guideline on T therapy in men with hypogonadism (1), the only lower T cut-off level put forward is 9.2 nmol/L (264 ng/dl) for T assays calibrated against the CDC reference method. This 9.2 nmol/L cut-off corresponds to the 2.5th percentile for T in non-obese men 19–39 years of age, according to harmonized T reference ranges generated from cross-calibrated data from US- and Europe-based cohorts of community-dwelling men [11]. Although there is no universally accepted T cut-off for diagnosis of hypogonadism, there is broad agreement that men with T below 8 nmol/L are likely to be hypogonadal, whereas men with T above 12 nmol/L are not [9].

When serum T is measured with the purpose to monitor T treatment in a hypogonadal men, interpretation of T levels can be difficult as there may be very large fluctuations in measured levels depending on the timing of sampling relative to administration of the last T dose and on the route of administration (e.g. transdermal versus intramuscular) and the pharmacokinetics of the T preparation being used. Blood sampling is often performed shortly before administration of the next dose, in order to capture the trough T level. However, for most preparations there can be significant within- and/or between-day variations of T levels such that single time point measurement may not be fully representative. These issues are important for clinicians to bear in mind when considering T dose adjustments in patients being started or maintained on T therapy and when assessing efficacy or compliance to treatment.

Assessment of Serum-Free T Level

Testosterone in the circulation is mostly bound to plasma proteins, i.e. in part specifically and with high affinity to SHBG (about 40–60%; variable according to SHBG concentration), in part more loosely bound to albumin (about 40–50%), and minor fractions bound to cortisol-binding globulin and orosomucoid. In men, only 1 to 3% (median 1.5%) of T circulates as free hormone [12]. According to the free hormone hypothesis, only this small fraction can freely access the target tissues and is directly available for biological action [13]. At least part of T loosely bound to albumin might dissociate in the capillary circulation and thus also become available for diffusion into the tissues. The combined free and albumin-bound fractions of T are often referred to as 'bioavailable T'. The relation between free T and bioavailable T serum concentration is fairly constant, unless there are large deviations of serum albumin levels from normal [14].

In conditions which markedly affect SHBG levels (**Box 10.2.2.2**) free T is a more accurate index of testicular function than total T. As an example, obesity results in decreased SHBG and consequently also decreased total T levels, whereas free T is still maintained in many obese men with low total T, who are not in fact hypogonadal. Thus, to rely solely on total T will lead to overdiagnosis of hypogonadism in obese men. However, if free T is also decreased in addition to total T, usually encountered in severe obesity, these men are likely to be truly and symptomatically hypogonadal. On the other hand, SHBG increases with age so that gonadal function in older men may be lower than suggested by the total T level and more accurately reflected by the free T level. Moreover, men with markedly increased SHBG concentrations (e.g. due to HIV/AIDS, antiepileptic medication, or liver disease) may have a low free T and be hypogonadal despite a total T in the normal range. All current guidelines recommend total T as the first-line diagnostic test for diagnosis of hypogonadism [1, 9, 10, 15–17]. Generally they recommend assessment of free T only as a supplementary diagnostic test in those symptomatic men with a condition known to affect SHBG concentrations, where total T is a less reliable index of gonadal status, and in those with borderline total T result at the lower limit of the reference range. However, not all recommendations support the use of free T assessment [10, 17]. Nevertheless, several studies have illustrated the usefulness of free T assessment (e.g. [18, 19]).

Box 10.2.2.2 Clinical situations associated with altered serum SHBG levels

Low SHBG
- Obesity
- Type 2 diabetes mellitus
- Hypothyroidism
- Cushing syndrome
- Glucocorticoid treatment
- Acromegaly
- Growth hormone therapy
- Nephrotic syndrome

High SHBG
- Ageing
- Low BMI
- Thyrotoxicosis
- Chronic liver disease (cirrhosis; hepatitis)
- HIV infection
- Growth hormone deficiency
- Anticonvulsant drugs (some, e.g. phenytoin)
- Oestrogen, tamoxifen, clomiphene citrate use

There are methodological issues as to the clinical implementation of free T assessment. The reference range for free T is even less well established than that for total T and there is again no generally accepted lower cut-off level indicating T deficiency. A level around 225 pmol/L (6.5 ng/dl), but ranging from 170 to 240 pmol/L (5.0–7.0 ng/dl) has been suggested [5, 9, 16, 20–22]. Measurement of free T is based on dialysis or ultrafiltration techniques. In the indirect dialysis and ultrafiltration methods a tracer is added (usually tritium-labelled T), percentage free T is measured after dialysis or ultrafiltration and free T concentration is calculated from percentage free T and separately determined total T level. In state-of-the-art direct methods the actual concentration free T is measured directly in the dialysate or ultrafiltrate using a highly sensitive LC-MS/MS assay [12, 14]. Implementation of these reference methods are technically challenging, cumbersome, expensive, and thus not well suited for routine clinical use. Consequently, free T measured by a dialysis (or ultrafiltration) method is not available in most clinical chemistry laboratories and thus not accessible to clinicians. As an alternative for routine clinical purposes, an estimate of free T can be calculated from total T and SHBG (and albumin) concentrations. Different equations for calculation of free T have been proposed: they all have limitations and do not perform equally well when validated against actually measured free T [12]. Best established are calculations based on the law of mass action with the version proposed by Vermeulen *et al.* [14] being the most widely used in clinical practice and research. Even this equation has limitations: although the calculated estimate correlates strongly with the

actual measured free T over the full range of clinically encountered SHBG levels and performs reasonably well for routine clinical purpose, it tends to systematically overestimate the actual measured free T by about 15–20% [12]. Importantly, the calculated estimates of free T are critically dependent on the reliability of the assays used for determination of total T and SHBG concentrations. There are possible pitfalls which can occasionally affect the reliability of the calculated free T estimates: e.g. very rare SHBG variants with altered affinity for T; use of drugs competing with the binding of T to SHBG (e.g. methyltestosterone, fluoxymesterone); states of disequilibrium between free and bound T (e.g. sampling shortly following parenteral T administration). Moreover, calculations should not be used in case of serum concentrations of SHBG far outside the reference range at the extremes of highs and lows, which is not rare in women (e.g. very high SHBG under use of oral contraceptives) but is uncommon in men, or in case of unusually high T concentrations clearly above the upper normal range (e.g. shortly following parenteral T administration).

The non-SHBG –bound or 'bioavailable' T is measured by an ammonium sulphate precipitation technique, which involves addition of tritium-labelled T tracer and precipitation of SHBG-bound T with ammonium sulphate. The unprecipitated tracer remaining in the supernatant is albumin-bound or free and represents the bioavailable fraction. Bioavailable T can then be calculated from separately measured total T and the percentage tracer in the supernatant. An estimate of bioavailable T can also be calculated from total T, SHBG, and albumin (e.g. using the equation based on the law of mass action according to Vermeulen *et al.*) [14]. The ammonium sulphate precipitation method is technically challenging, with results from different labs difficult to compare and there is no evidence that determination of bioavailable T offers added value compared to free T for diagnosis of male hypogonadism [1, 5, 14]. Finally, T concentration in saliva reflects closely serum-free T concentration if measured reliably, which requires the use of an LC-MS/MS-based assay [23]. Although this approach has potential, there are pitfalls (e.g. contamination of the saliva with blood), reliable assay of saliva T is not generally available in clinical chemistry labs and experience with use of saliva T as diagnostic tool is presently lacking.

Differentiation Between Causes of Hypogonadism

Gonadotropins

Luteinizing hormone (LH) and follicle-stimulating hormone (FSH) secretion increases when negative feedback exerted at the hypothalamic-pituitary level is diminished in the presence of low circulating levels of sex steroids (T and oestradiol) and inhibin B, respectively, unless this expected response is prevented by hypothalamic-pituitary malfunction. In a hypogonadal man, low or inappropriately normal gonadotropin levels therefore indicate secondary hypogonadism due to functional or organic hypothalamic-pituitary dysfunction with the low T secondary to inadequate LH stimulation of the testicular Leydig cells. Conversely, in a hypogonadal man elevated gonadotropin levels indicates primary hypogonadism with diminished negative feedback on gonadotropin secretion resulting from primary testicular failure. Measurement of serum levels of LH and FSH is the primary step in the search for the cause of hypogonadism and orientates further investigations. As an example, elevated gonadotropins in absence of another manifest cause of primary hypogonadism will direct towards performing karyotyping for possible Klinefelter's syndrome. On the other hand, low gonadotropins in absence of a manifest functional cause of hypogonadism will direct towards investigating for possible organic hypothalamic-pituitary disease (e.g. determination of serum prolactin and iron load, MRI imaging, possibly testing of other pituitary hormonal axes, or search for genetic defects affecting gonadotropin-releasing hormone (GnRH) secretion or action or directly the LH gene).

Serum LH and FSH are measured by immunoassay. Even if there are differences in sensitivity and accuracy between assays, most are sufficiently robust to perform satisfactorily for clinical purposes. The secretion of gonadotropins is pulsatile and this is reflected in moment to moment variations and a pulsatile pattern of serum LH levels, whereas the pulsatile pattern is less prominent and serum levels are less variable for FSH. In most instances this variability of LH levels does not cause diagnostic problems because in primary hypogonadism LH and FSH levels are usually substantially elevated and clearly distinct from the low to normal levels characteristic of secondary hypogonadism. Only in rare instances with borderline findings is repeated measurement of LH in a separate sample needed to account for variability of LH levels. At the lower end of gonadotropin levels, there is overlap between low and normal levels and assays differ in their ability to distinguish between both. Again, this is rarely a problem in clinical practice as in men with secondary hypogonadism the prevailing low, low-normal, or normal gonadotropin levels are generally clearly distinct from the elevated levels seen in primary hypogonadism.

In the GnRH-stimulation test, the pituitary response of LH and FSH to stimulation with GnRH is assessed as a measure of the pituitary gonadotropin secretory capacity. Protocol details may vary, but commonly LH and FSH levels are measured before and 30 to 45 minutes following intravenous injection of 100 μg GnRH. In healthy men, the increase of LH in response to GnRH tends to be proportional to the baseline LH level and the response of FSH, although much lower, is proportional to that of LH. The GnRH test has been used mainly to differentiate between normal and decreased gonadotropin secretion. However, there is no general agreement on criteria for normal response and although the discrimination between the response in healthy men and men suspected of disorders of the hypothalamic-pituitary-gonadal axis may be better for absolute than for relative LH increase, overall neither can reliably discriminate [24]. With the currently available improved immunoassays for gonadotropins, the GnRH test offers little added value above assessment of basal levels in the evaluation of hypogonadal men and indications for performing a GnRH test are now limited (e.g. in some instances for differential diagnosis of precocious puberty). In particular, the GnRH test does not reliably differentiate between delayed puberty and hypogonadotropic hypogonadism (HH) or in men with HH between a hypothalamic and pituitary defect.

A major component of the negative feedback inhibition of LH secretion by sex steroids is exerted by circulating oestradiol, which is in part of testicular origin but originates mainly from aromatization of T in peripheral tissues (e.g. in fat tissues) [25]. Patients with a functional hypothalamic disorder of GnRH secretion as cause of HH will usually respond with augmentation of both gonadotropin

and T levels in response to reduced hypothalamic oestradiol negative feedback by treatment either with a compound with (partial) antagonistic action on hypothalamic oestrogen receptors (e.g. clomiphene citrate or tamoxifen) or with an aromatase inhibitor (e.g. letrozole). However, there are no formally established or validated protocols for diagnostic purposes.

Prolactin

In men with secondary hypogonadism serum prolactin level should be measured. Prolactin is measured in serum by immunoassay. A prolactin concentration above the reference range in a single serum sample obtained randomly during the day infers the presence of hyperprolactinemia if excessive venipuncture stress has been avoided. High levels, more than tenfold the upper reference range (>250 μg/L or 5300 mIU/L), usually indicate the presence of a prolactinoma. However, more modest elevations of prolactin levels may also indicate the presence of a prolactinoma even though other possible causes of hyperprolactinemia should be considered [26]. Modest hyperprolactinemia may result from lesions of hypothalamus or pituitary stalk (e.g. hypothalamic or pituitary tumour, trauma, surgery, lymphocytic hypophysitis, infiltrative disease such as sarcoidosis) which interfere with the hypothalamic tonic inhibitory control of prolactin secretion by dopamine. Further, modest elevations of prolactin can be due to (psychological and physical) stress at time of blood sampling. Several medications stimulate prolactin secretion, sometimes to markedly elevated serum levels (up to >200μg/L). This is in particular the case for psychotropic drugs such as risperidone and gastroprokinetic drugs such as metoclopramide and domperidone with antidopaminergic activity. Several other drugs can cause hyperprolactinemia, among which verapamil, morphine, and other opiates, and amoxapine. Prolactin levels can be increased in hypothyroidism, renal failure, cirrhosis of the liver, and by chest wall stimulation. Finding of an elevated serum prolactin in asymptomatic men may result from presence of circulating macro-prolactin with absent or only low bioactivity, which can be tested by assay of the serum for prolactin after precipitation of possible macroprolactin by sample pretreatment with polyethylene glycol (PEG). Some clinical laboratories check all serum samples with markedly high prolactin concentration for presence of macro-prolactin.

Insulin-Like Peptide 3

Insulin-like peptide 3 (INSL3), a member of the relaxin-insulin hormone family, is produced by testicular Leydig cells. It acts via the Relaxin Family Peptide Receptor2 (RXFP2), which is expressed on testis germ cells and in lower concentrations on the Leydig cells, suggesting a paracrine and autocrine role. It indirectly facilitates testicular descend by promoting the development of the gubernaculum. In adults it may have a role in germ cells survival. Its concentration in serum increases at puberty in response to increased LH levels. In adults, INSL3 concentrations reflect Leydig cells maturation and number [27]. INSL3 was reported not to differentiate between boys with delayed puberty and HH [28]. In HH INSL3 are low, remain low under treatment with T or FSH monotherapy, but the INSL3 levels do increase with hCG or hCG-FSH treatment. However, the INSL3 rise in response to hCG (or hCG-FSH) was reported to be less pronounced in HH patients with history of cryptorchidism [29]. There is presently no established clinical role for INSL3 measurements in the work-up of adult men suspected of hypogonadism.

Sex Steroid Hormones Other Than T

Through action of 5α-reductase (type 1 and 2) T can be metabolized to its 5α-reduced, more potent androgenic metabolite 5α-dihydrotestosterone (DHT). Only small amounts of DHT are secreted by the testes, most DHT in the circulation (≥80%) derives from 5α-reduction of T by 5α-reductase type2 in peripheral target tissues. Therefore, serum DHT is not a representative measure of testicular function and its measurement, which is of interest in rare cases of suspected 5α-reductase deficiency, does not contribute to the work-up of adult men with suspected hypogonadism [30].

Androstenedione is secreted by both the testes and adrenal cortex. Its androgenic action is dependent on its biotransformation to T (blood conversion rate around 15%). This contribution is marginal relative to the testicular T production and measurement of serum androstenedione levels does not contribute to the work-up of men suspected of hypogonadism. This is also the case for measurement of serum levels of the androgen precursor dehydroepiandrosterone (DHEA), which originate mainly from secretion by the adrenal cortex and peripheral desulphatation of DHEA-sulphate (DHEAS), and for no more than 10% from testicular secretion. Androstanediol glucuronide, which has been proposed as an important marker of androgenic activity in women, is derived in men from T by up to 70% and the remaining from DHEAS, thus offering no useful information in addition to the assessment of serum T in the work-up of male hypogonadism [30].

Oestradiol is a physiologically important sex steroid in men [31, 32]. Most circulating oestradiol is derived from aromatization of T in peripheral tissues and only ≤ 20% is directly secreted by the testes. Measurement of oestradiol may potentially be informative in specific clinical situations such as in patients with gynaecomastia (possible relative or absolute oestrogen excess), morbid obesity (inconsistent reports of oestrogen excess), osteoporosis (increased bone loss and fracture risk associated with low oestradiol), or rare cases of severely delayed skeletal maturation (possible oestradiol deficiency in aromatase deficiency). The oestradiol serum levels in healthy men before age 65 years lie around 10–53 pg/mL (38–196pmol/L) [33]. To reliably measure the low oestradiol levels in men may require a well validated LC-MS/MS-based method as many immunoassays lack the required sensitivity. Moreover, there are no established consensus cut-off levels for either oestradiol excess or deficiency.

Inhibin B

Inhibin B is a testicular glycoprotein hormone which belongs to the TGF-β superfamily. It is a heterodimer composed of disulphide-bound alpha and beta B subunits. Inhibin B is apparently secreted by the Sertoli cells which, however, requires the presence of germ cells in the adult testis. Following transiently elevated serum levels during the neonatal activation of the gonadal axis ('mini-puberty'), its concentration in serum is low in prepubertal boys to rise again during puberty under influence of increased FSH secretion and initiation of spermatogenesis. In early pubertal stages there is a positive association between serum levels of FSH and inhibin B. In later puberty (from Tanner stages G3 and G4 on) and in adult men, there is a close inverse relationship between serum levels of FSH

and inhibin B. This reflects that in adult men inhibin B production is primarily regulated by germ cells with only a secondary role for FSH and that serum inhibin B is the principal negative feedback regulator of FSH secretion as long as spermatogenesis is not severely deteriorated with ensuing markedly low serum inhibin B. Serum levels are measured by immunoassay, which requires high specificity for inhibin B and low cross-reactivity with other inhibin subtypes, activins, and their respective subunits. Inhibin B levels display a significant circadian rhythm, which closely parallels that of serum T [34, 35]. In adult men, serum inhibin B levels reflect global testicular spermatogenesis and are strongly positively associated with both testis volume and sperm count [36]. In subfertile men, there is a generally close relationship between degree of impairment of spermatogenesis and inhibin B secretion, with arrest of spermatogenesis at earlier stages accompanied by the lowest inhibin B levels [32]. The basal serum inhibin B level has been shown to offer excellent discrimination (better than LH) between HH (idiopathic HH or combined pituitary hormone deficiency) and constitutional delay of puberty in boys with Tanner genital stage G1 (i.e. testis volume <3 ml) with 100% sensitivity and specificity for an inhibin B of ≤35 pg/ml [37]. In boys with early genital stage G2 (testis volume 3–6 ml) discrimination was still fair and better than serum LH with sensitivities of 86% and 80% and specificities of 92% and 88% in boys with idiopathic HH and combined pituitary hormone deficiencies, respectively, for a serum inhibin B of ≤65 pg/ml [37]. In men with HH, inhibin B increases under FSH or pulsatile GnRH therapy with a lesser response in men with history of cryptorchidism and a more rapid response in men with acquired HH compared to men with congenital HH. A higher baseline inhibin B level, greater than 60pg/ml has been reported to be a favourable predictor for achieving adult testicular size and optimizing spermatogenesis during long-term GnRH therapy in men with idiopathic HH [38]. Besides in the differentiation between HH and constitutional delayed puberty, measurement of inhibin B does not play a contributory role in the work-up of adult men suspected of hypogonadism.

Anti-Müllerian Hormone

Anti-Müllerian hormone (AMH), a testicular homodimeric glycoprotein hormone belonging to the TGF-β superfamily, is a marker of prepubertal Sertoli cell function. Its production is stimulated by FSH and downregulated by T. Serum levels are measured by immunoassay. In boys before puberty, serum levels reflect the number, integrity, and immature state of Sertoli cells. The AMH levels are high in prepubertal boys (55–250 ng/ml; 400–1800 pmol/L) and decline during puberty (Tanner stage 2: 70–1000 pmol/L: Tanner stage 4: 30–400 pmol/L), when Sertoli cells express the androgen receptor (AR) and are exposed to increasing intratesticular T concentrations, to reach low levels (3–18 ng/ml; 25–130 pmol/L) in healthy adult men [39, 40]. In untreated men with congenital HH, serum AMH levels are generally high compared to eugonadal men, reflecting the immature state of their Sertoli cells. Treatment with FSH results in further increase of AMH levels, whereas the levels decrease markedly towards normal low adult levels during combined FSH—hCG treatment in response to exposure of the Sertoli cells to hCG-induced rise of intratesticular T concentrations [41].

AMH has no prominent role in the diagnostic hormonal work-up of adult men suspected of hypogonadism. However, AMH is more contributory as diagnostic tool in prepubertal boys [38, 39]. Low serum AMH can be an early marker of (central or gonadotropin-independent) precocious puberty. Conversely, AMH levels remain elevated in boys with delayed puberty. In boys with prepubertal testes (stage 1), serum AMH can help differentiate between adolescents with constitutional delayed puberty and adolescents with HH as the latter have lower AMH levels due to the lower FSH exposure and/or a smaller pool of Sertoli cells. However, for this indication inhibin B levels are more discriminative [37]. AMH levels are also lower in some children with cryptorchidism, monorchidism, and partial gonadal dysgenesis [39, 40]. In boys with non-palpable gonads, serum AMH is highly predictive for presence of (intra-abdominal) testis tissue; undetectable AMH is suggestive of congenital or acquired anorchidism or complete gonadal dysgenesis [39, 40]. In this context, the information obtained by dosage of serum AMH largely limits the indications for performing an hCG-stimulation test (measurement of serum T before and following intramuscular hCG injection(s); applied protocols for hCG-stimulation test vary by duration, hCG dosage, single or repeated injections).

Genetic Markers

Polymorphisms of the promotor of the *SHBG* gene can affect SHBG concentration and thus serum total T levels. However, this should not raise major diagnostic problems as the magnitude of these effects is not large relative to the considerable native between subject variations of SHBG concentrations seen in community-dwelling men. Moreover, these polymorphisms are not expected to affect serum-free T levels. Polymorphisms of SHBG affecting the affinity for T have also been reported. These also should not affect measured free T levels, although they may invalidate calculated estimates of free T in carriers of these very rare polymorphisms [42–45].

Polymorphisms of the androgen receptor can affect androgen sensitivity. This has in particular been documented for the polyglutamine tract polymorphism encoded by a highly polymorphic CAG repeat sequence in exon 1 of the androgen receptor gene, with longer repeat associated with lower receptor transactivation upon binding with its ligand. In theory, taking into account such genetic polymorphisms can refine the interpretation of T levels. However, the effect of this and other reported polymorphisms that can affect T levels is quantitatively too marginal to be clinically relevant in the context of the diagnostic work-up of men suspected of hypogonadism [45–47].

Rare mutations in the androgen receptor gene can cause partial androgen insensitivity and rare mutations in the aromatase gene can cause oestrogen deficiency. Because of diminished negative feedback exerted by T and oestradiol, respectively, hormonally both conditions are characterized by elevated LH and T levels, whereas oestradiol levels tend to be elevated in androgen insensitivity and low in aromatase deficiency.

Men Evaluated for Subfertility

Semen analysis according to World Health Organization (WHO) criteria is the gold standard for evaluation of men with suspected subfertility. Hormonal investigations provide mainly corroborating evidence. The clinical context will usually indicate whether or not the fertility problem may be part of a broader problem of

hypogonadism affecting also testicular hormonal function, which can be ascertained by measurement of serum T and gonadotropin levels.

Elevated serum FSH levels indicate testicular dysfunction with deficient spermatogenesis and consequent diminished inhibin B secretion. Serum FSH and inhibin B levels are strongly inversely correlated and both rise of FSH and decrease of inhibin B levels are to some extent proportional to the severity of impaired spermatogenesis. However, in individual cases findings for hormonal levels and spermatogenesis may be discordant. Both FSH and inhibin B are expected to be within normal range in men with obstructive azoospermia, which can be further documented by analysis of the seminal plasma. However, both FSH and inhibin B may also be within normal range in non-obstructive azoospermia with arrest of spermatogenesis in a late stage, in some men with azoospermia and focal Sertoli-cell-only, and in some men with hypospermatogenesis [34–36]. Findings of decreased serum AMH in men with impaired spermatogenesis have been inconsistent and serum AMH has not been found to be a clinically useful marker of spermatogenesis [35, 48]. Furthermore, measurement of inhibin B and AMH after stimulation with recombinant FSH does not appear to offer added diagnostic value compared to basal levels of these hormones and, similarly, limited available data on measurement of inhibin B and AMH in seminal plasma does not indicate added value compared to serum levels [35].

Although in some studies diagnostic accuracy of inhibin B and of inhibin B and FSH combined was slightly greater than that of FSH, this is not of practical significance for evaluation of individual patients with subfertility, for which measurement of inhibin B offers no clear added value above FSH dosage [34–36, 49]. Importantly, in men with non-obstructive azoospermia neither the levels of FSH, inhibin B, or AMH, nor their combination can reliably predict the outcome of procedures to retrieve sperm from the testis either by fine-needle aspiration (FNA) or by open biopsy (testicular sperm extraction—TESE) [34, 35].

REFERENCES

1. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
2. Diver MJ, Imtiaz KE, Ahmad AM, Vora JP, Fraser WD. Diurnal rhythms of serum total, free and bioavailable testosterone and of SHBG in middle-aged men compared with those in young men. *Clin Endocrinol*, 2003; **58**: 710–17.
3. Caronia LM, Dwyer AA, Hayden D, Amati F, Pitteloud N, Hayes FJ. Abrupt decrease in serum testosterone levels after an oral glucose load in men: implications for screening for hypogonadism. *Clin Endocrinol (Oxf)*, 2013; **78**: 291–6.
4. Lehtihet M, Arver S, Bartuseviciene I, Pousette A. S-testosterone decrease after a mixed meal in healthy men independent of SHBG and gonadotrophin levels. *Andrologia*, 2012; **44**: 405–10.
5. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*, 2007; **92**: 405–13.
6. Handelsman DJ. Mass spectrometry, immunoassay and valid steroid measurements in reproductive medicine and science. *Hum Reprod*, 2017; **32**: 1147–50.
7. Huhtaniemi IT, Tajar A, Lee DM, *et al*. Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men. *Eur J Endocrinol*, 2012; **166**: 983–91.
8. Haring R, Baumeister SE, Nauck M, *et al*. Testosterone and cardiometabolic risk in the general population—the impact of measurement method on risk associations: a comparative study between immunoassay and mass spectrometry. *Eur J Endocrinol*, 2013; **169**: 463–70.
9. Wang C, Nieschlag E, Swerdloff R, *et al*. Investigation, treatment, and monitoring of late-onset hypogonadism in males: ISA, ISSAM, EAU, EAA, and ASA recommendations. *Eur J Endocrinol*, 2008; **159**: 507–14.
10. Khera M, Adaikan G, Buvat J, *et al*. Diagnosis and treatment of testosterone deficiency: recommendations from the Fourth International Consultation for Sexual Medicine. *J Sex Med*, 2016; **13**: 1787–804.
11. Travison TG, Vesper HW, Orwoll E, *et al*. Harmonized reference ranges for circuating testosterone levels in men of four cohorts studies in the United States and Europe. *J Clin Endocrinol Metab*, 2017; **102**: 1161–73.
12. Fiers T, Wu F, Moghetti P, Vanderschueren D, Lapauw B, Kaufman JM. Reassessing free testosterone calculation by liquid chromatography-tandem-mass-spectrometry direct equilibrium dialysis. *J Clin Endocrinol Metab*, 2018; **103**: 2167–74.
13. Hammond GL. Plasma steroid-binding proteins: primary gatekeepers of steroid hormone action. *J Endocrinol*, 2016; **230**: R13–R25.
14. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab*, 1999; **84**: 3666–72.
15. Morales A, Bebb RA, Manjoo P, et al. Diagnosis and management of testosterone deficiency syndrome in men: clinical practice guideline. *Can Med Assoc J*, 2015; **187**: 1369–77.
16. Buvat J, Maggi M, Guay A, Torres LO. Testosterone deficiency in men: systematic review and standard operating procedures for diagnosis and treatment. *J Sex Med*, 2013; **10**: 245–84.
17. Yeap BB, Grossmann M, McLachlan RI, *et al*. Endocrine Society of Australia position statement on male hypogonadism (part 1): assessment and indications for testosterone therapy. *Med J Aus*, 2016; **205**: 173–8.
18. Antonio L, Wu FC, O'Neill TW, *et al*. Low free testosterone is associated with hypogonadal signs and symptoms in men with normal total testosterone. *J Clin Endocrinol Metab*, 2016; **101**: 2647–57.
19. Rastrelli G, O'Neill TW, Ahern T, *et al*. Symptomatic androgen deficiency develops only when both total and free testosterone decline in obese men who may have incident biochemical secondary hypogonadism: prospective results from the EMAS. *Clin Endocrinol (Oxf)*, 2018; **89**: 459–69.
20. Wu FC, Tajar A, Beynon JM, *et al*. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*, 2010; **363**: 123–35.
21. Araujo AB, Esche GR, Kupelian V, *et al*. Prevalence of symptomatic androgen deficiency in men. *J Clin Endocrinol Metab*, 2007; **92**: 4241–7.
22. Bhasin S, Pencina M, Jasuja GK, *et al*. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. *J Clin Endocrinol Metab*, 2011; **96**: 2430–9.
23. Fiers T, Delanghe J, T'Sjoen G, Van Caenegem E, Wierckx K, Kaufman JM. A critical evaluation of salivary testosterone as a

method for the assessment of serum testosterone. *Steroids*, 2014; **86**: 5–9.
24. Bang AK, Nordkap L, Almstrup K, *et al.* Dynamic GnRH and hCG testing: establishment of new diagnostic reference levels. *Eur J Endocrinol*, 2017; **176**: 379–91.
25. Raven G, de Jong FH, Kaufman JM, de Ronde W. In men, peripheral estradiol levels directly reflect the action of estrogens at the hypothalamo-pituitary evel to inhibit gonadotropin secretion. *J Clin Endocrinol Metab*, 2006; **91**: 334–28.
26. Melmed S, Casanueva FF, Hoffman AR, *et al.* Diagnosis and treatment of hyperprolactinemia: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2011; **96**: 273–88.
27. Ivell R, Wade JD, Anand-Ivell R. INSL3 as a biomarker of Leydig cell functionality. *Biol Reprod*, 2013; **88**: 147, 1–8.
28. Rohayem J, Nieschlag E, Kliesch S, Zitzmann M. Inhibin B, AMH, but not INSL3, IGF1 or DHEAS support differentiation between constitutional delay of growth and puberty and hypogonadotropic hypogonadism. *Andrology*, 2015; **3**: 882–7.
29. Trabado S, Maione L, Bry-Gauillard H, *et al.* Insulin-like peptide 3 (INSL3) in men with congenital hypogonadotropic hypogonadism/Kallmann syndrome and effects of different modalities of hormonal treatment: a single-center study in 281 patients. *J Clin Endocrinol Metab*, 2014; **99**: E268–E275.
30. Kaufman JM, Vermeulen A. The decline of androgen levels in elderly men and its clinical and therapeutic implications. *Endocr Rev*, 2005; **26**: 833–76.
31. Vanderschueren D, Laurent MR, Claessens F, *et al.* Sex steroid actions in male bone. *Endocr Rev*, 2014; **35**: 906–60.
32. Finkelstein JS, Lee H, Burnett-Bowie S-AM, *et al.* Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*, 2013; **369**: 1011–22.
33. Handelsman DJ, Yeap B, Flicker L, Martin S, Wittert GA, Ly LP. Age-specific population centiles for androgen status in men. *Eur J Endocrinol*, 2015; **173**: 809–17.
34. Meachem SJ, Nieschlag E, Simoni M. Inhibin B in male reproduction: pathophysiology and clinical relevance. *Eur J Endocrinol*, 2001; **145**: 561–71.
35. Iliadou PK, Tsametis C, Kaprara A, Papadimas I, Goulis DG. The Sertoli cell: novel clinical potentiality. *Hormones*, 2015; **14**: 504–14.
36. Pierik FH, Vreeburg JTM, Stijnen T, De Jong FH, Weber RFA. Serum inhibin B as a marker of spermatogenesis. *J Clin Endocrinol Metab*, 1998; **83**: 3110–14.
37. Coutant R, Biette-Demeneix E, Bouvattier C, *et al.* Baseline inhibin B and anti-Müllerian hormone measurements for diagnosis of hypogonadotropic hypogonadism (HH) in boys with delayed puberty. *J Clin Endocrinol Metab*, 2010; **95**: 5225–32.
38. Pitteloud N, Hayes FJ, Dwyer A, Boepple PA, Lee H, Corwley Jr WF. Predictors of outcome of long-term GnRH therapy in men with idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2002; **87**: 4128–36.
39. Condorelli RA, Cannarella R, Calogero AE, La Vignera S. Evaluation of testicular function in prepubertal children. *Endocrine*, 2018; **62**: 274–80.
40. Edelsztein NY, Grinspon RP, Schteingart HF, Rey RA. Anti-Müllerian hormone as a marker of steroid and gonadotropin action in the testis of children and adolescents with disorders of the gonadal axis. *Int J Pediatr Endocrinol*, 2016; **2016**: 20.
41. Young J, Rey R, Couzinet B, Chanson P, Josso N, Chaison G. Antimüllerian hormone in patients with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 1999; **84**: 2696–9.
42. Eriksson AL, Lorentzon M, Mellström D, *et al.* SHBG gene promoter polymorphisms in men are associated with serum sex hormone-binding globulin, androgen and androgen metabolite levels, and hip bone mineral density. *J Clin Endocrinol Metab*, 2006; **91**: 5029–37.
43. Vanbillemont G, Bogaert V, De Bacquer D, *et al.* Polymorphism of the SHBG gene contribute to the interindividual variation of sex steroid hormone blood levels in young, middle-aged and elderly men. *Clin Endocrinol (Oxf)*, 2009; **70**: 303–10.
44. Coviello D, Haring R, Wellons M, *et al.* A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple loci implicated in sex steroid hormone regulation. *PLoS Genet*, 2012; **8**: e1002805.
45. Ohlsson C, Wallaschofski H, Lunetta KL, *et al.* Genetic determinants of serum testosterone concentrations in men. *PLoS Genet*, 2011; **7**: e1002313.
46. Crabbe P, Bogaert V, De Bacquer D, Goemaere S, Zmierczak H, Kaufman JM. Part of the interindividual variation in serum testosterone levels in healthy men reflects differences in androgen sensitivity and feedback set point: contribution of the androgen receptor polyglutamine tract polymorphism. *J Clin Endocrinol Metab*, 2007; **92**: 3604–10.
47. Huhtaniemi IT, Pye SR, Holliday KL, *et al.* Effect of polymorphisms in selected genes involved in pituitary-testicular function on reproductive hormones and phenotype in aging men. *J Clin Endocrinol Metab*, 2010; **95**: 1898–908.
48. Aksglaede L, Olesen IA, Carlsen E, Petersen JH, Juul A, JØrgensen N. Serum concentration of anti-Müllerian hormone is not associated with semen quality. *Andrology*, 2018; **6**: 286–92.
49. Kelsey TW, McConville L, Edgar AB, *et al.* Follicle stimulating hormone is an accurate predictor of azoospermia in childhood cancer survivors. *PLoS One*, 2017; **12**: e0181377.

10.2.3 Diagnostic Semen Analysis

Jackson C. Kirkman-Brown and Sarah J. Conner

Introduction

Human semen is comprised of spermatozoa (commonly referred to as sperm) produced during spermatogenesis in the testes, matured in the epididymis, and seminal plasma produced by the accessory sex glands: the prostate and the seminal vesicles. Prior to ejaculation, during the arousal phase, the alkaline pre-ejaculate, a clear and colourless viscous fluid, is released from the bulbourethral gland. This neutralizes any urine left in the urethra prior to the passage of semen. Pre-ejaculate fluid emission is very variable in amounts between men, from not noticeable through to several millilitres; it may contain sperm from prior ejaculations and occasionally patients may erroneously believe that this is, in fact, their ejaculate. Although the focus is often on the sperm, the epididymal environment and seminal plasma are also important for reproductive function providing protection from the acidic environment of the vagina and from oxidative damage as well as supplying compounds, essential for optimal sperm function. Secretions into the epididymal lumen include L-carnitine, α-glucosidase, and glycerylphosphorylcholine, a pump mechanism selectively moving L-carnitine from circulating blood at the epididymal level. The fluid secreted by the seminal vesicles has an alkaline pH and makes up more than 60% of the ejaculate volume. Fructose, produced in the seminal vesicles under the stimulation of androgens, plays an important role in energy metabolism and motility of spermatozoa. Prostatic secretion, with an acidic pH between 6.4 and 6.8, comprises about 15–20% of the ejaculate volume; it is rich in citric acid, zinc, acid phosphatase, magnesium, and polyamines. Citric acid is important in the maintenance of osmotic equilibrium of the seminal plasma and coagulation/liquefaction of semen after ejaculation.

During ejaculation, the secretions of various accessory glands and the testicular fraction are released in a precise sequence. The first fraction is predominantly prostatic with an epididymal and testicular fraction (rich in sperm), and is followed by fractions that derive from the seminal vesicles.

Mixing of the fluids from the prostate and seminal vesicles leads to coagulation of the ejaculate which increases retention of the semen within the upper vagina to facilitate sperm passage into the cervix. In humans, the ejaculate liquefies after approximately 30 minutes under the action of prostatic proteolytic enzymes.

Guidelines for Initial Assessment

Basic semen analysis is the primary diagnostic tool in the male fertility assessment pathway. The complete pathway, as recommended by a recent World Health Organization (WHO) working group should initially include '*a physical examination performed by an examiner with appropriate training and expertise, a reproductive history and at least one properly performed (high quality) semen analyses. A full evaluation by a urologist or other specialist in male reproduction should be done if the initial screening evaluation demonstrates an abnormal physical examination, an abnormal male reproductive or sexual history, or an abnormal semen analysis is found*' [1] (see Chapter 10.3.3, 'Management of Idiopathic Male Infertility').

During semen analysis an assessment is made of the number, viability, motility, and morphology of the sperm cells as well as the quality of the seminal plasma and the presence of non-sperm cells. It is very important to note that the diagnostic reliability of semen analysis depends crucially on the quality control of the entire system, alongside experience and ability of the laboratory staff. Therefore laboratories participating in an external quality assurance (EQA) scheme should always be used, in countries where this is possible they should also be accredited to the relevant local standards (see next). A serious confounder of any results is the application of different criteria to the evaluation of sperm parameters, making it difficult to compare tests carried out in different laboratories [2].

To ensure consistent and comparable results, since 1980, the WHO has produced a regularly updated manual of guidelines for semen analysis. The current version, published in 2010 [3], is the 5th edition and, in addition to detailed laboratory protocols, provides a set of reference ranges and limits based on analysis of semen samples from several thousand recent fathers in eight countries on three continents. Other semen analysis protocols, designed to meet or exceed requirements of the WHO recommendations, such as those from The European Society for Human Reproduction and Embryology—Special Interest Group in Andrology (ESHRE-SIGA) recommendations [4] and the Practical Guide to Basic Laboratory Andrology [5], have been adapted to suit the needs of modern Assisted Reproductive Treatment (ART) laboratories. Results received from ISO15189 accredited laboratories working to either the WHO or ESHRE-SIGA standards can be considered reliable as long as the laboratory is participating in an EQA scheme.

Guidelines for Collecting Semen Samples

It is well recognized that there can be a high level of variability in the parameters of semen samples from a single individual due to biological and analytical reasons [6] and caution should therefore be exercised in drawing firm conclusions based on the analysis of a single sample, a single abnormal result does not necessarily mean any problem is present. The process of spermatogenesis takes 64 days in humans. Therefore, a second test after 2–3 months should be considered when abnormal results are returned but there is no need to repeat a normal first analysis [1]. Febrile illnesses cause significant sperm abnormalities which become apparent within 7–10 days, and can persist for up to 10–12 weeks.

Some of the variations in semen analysis may be minimized by following certain rules prior to sample collection. The period of sexual abstinence before taking a sample should be between 2 and 7 days, which serves to match the criteria used by WHO to set the reference ranges [3, 7]. This period includes any ejaculation, not only sexual intercourse, a detail often overlooked by patients and staff. Various studies have shown a positive correlation between semen volume, sperm concentration, and length of abstinence. Effects on sperm quality are less clear for abstinences under 2 days, longer abstinences correlate with decreased sperm quality due to prolonged storage in the epididymis/vas deferens [7]. It is worth discussion with patients that regular ejaculation is likely to provide optimal sperm quality.

Masturbation with ejaculation of the sample into a sterile container, such as those used for urine, is the recommended procedure for collection of semen samples. Where masturbation proves difficult, coitus, using specialized 'sperm safe' condoms developed and marketed specifically for this purpose, is recommended. Coitus

interruptus should not be considered, as this method tends to lose part of the ejaculate and makes it difficult to distinguish between the man and his partner's epithelial cells, white cells, and red blood cells; and can moreover cause bacterial contamination. It is notable that some patients will expect or require using a lubricant for masturbation, generally these are toxic to sperm, even those marketed as 'sperm friendly' intimate lubricants may have adverse effects. Good diagnostic laboratories will usually provide the same mineral oil used in *in vitro* fertilization (IVF) processes for masturbation as it is assured at a relevant medical grade and will have passed sperm toxicity testing. It is similarly worth discussion with patients to check that these lubricants are not being used at home during intercourse attempting conception [8, 9].

The sample should by default be obtained at the site of the laboratory for a number of reasons: to examine the ejaculate within the first 30 minutes; to control the environment and handling of the ejaculate; and finally, if subsequent samples are to be produced for treatment or storage, they will usually have to be produced on site for legal traceability.

Semen samples may sometimes have to be collected offsite for psychological, or practical reasons. These should still be collected in a designated provided container, with accurate record the time of ejaculation, the sample kept at 20–37°C and delivered to the laboratory within 1 hour of collection. The patient should also be asked to provide information regarding any pathologies from which he may have suffered during the three preceding months including the use of medication, fevers, viral or bacterial infections, antibiotic therapy, and local or general anaesthetic. Any of these factors may influence semen characteristics.

Macroscopic and Microscopic Evaluation of the Ejaculate

The semen sample should be processed within 1 hour of ejaculation in order to evaluate the time and nature of liquefaction. Microscopic evaluation must be carried out after a complete liquefaction of the sample. A normal sample usually liquefies within 15–30 minutes at room temperature, so processing should occur within 60 minutes after ejaculation. Failure to liquefy, either partially or completely, within 1 hour should be noted.

Volume

The normal semen volume is ≥1.5 ml [3]. Reduced semen volume (<1.0 ml) may indicate obstructive pathologies of the ejaculatory ducts or of a secretory defect of the seminal vesicles for functional or anatomical reasons. Rarely, and usually accompanied by other clinical evidence, it may suggest reduced production of testosterone [10]. A total absence of ejaculated semen is termed aspermia (as opposed to azoospermia—which is absence of spermatozoa in an ejaculate). In cases of low volume or aspermia, postejaculatory urine analysis should always be performed to consider for the possibility of retrograde ejaculation, prevalent in diabetics with autonomic neuropathy and men using certain medications (e.g. alpha-adrenergic antagonists). It is worth noting that retrograde ejaculation in some patients has been treated successfully by a number of medications, including alpha agonists [11, 12] without need for recourse to ART treatment.

Hyperspermia (elevated ejaculate volume) (>6.0 ml, represents the upper 90 Centile and >6.8 ml the 95 Centile of 'normal' [3]) can be associated with inflammatory pathologies and/or infections of the seminal vesicle and/or the prostate.

pH and Appearance

Seminal plasma has an alkaline pH (normal variation 7.2–8.0) reflecting the predominant contribution of the seminal vesicles. pH (≥8.0) may indicate inflammatory pathologies and dysfunction of the prostate. Acidic pH (<7.0) is more informative, and is often associated with obstructive pathologies of the ejaculatory ducts, or with congenital (agenesis) or acquired hypotrophy or atrophy of the seminal vesicles and vas deferens (if associated with azoospermia).

The physiological appearance of semen is opalescent ivory. This becomes milky when the ejaculate derives exclusively from the prostate, as occurs in some cases of genital tract obstruction. A yellowish appearance can indicate a high number of white blood cells (pyospermia) [13], and is often associated with acute or subacute infections of the male genital tract. Pink, intense red, or red-brown colours indicate the presence of blood (haematospermia). Haematospermia can be caused by trauma (such as following prostate biopsy), inflammations, and infections of the male genital tract, duct obstructions, cysts (bladder or accessory sex glands), and neoplastic pathologies [14].

Liquefaction and Viscosity

Immediately after ejaculation, human seminal fluid undergoes a process of coagulation which transforms the liquid into a gelatinous coagulate. Immediately after coagulation, the process of liquefaction begins and should be completed within 60 minutes. The assessment process ascertains whether liquefaction is complete or not and whether it occurs within an appropriate physiological time span. Liquefaction is the process of the ejaculate becoming homogenous liquid and not having 'jelly-like' parts, it must not be confused with viscosity.

When incomplete liquefaction occurs, it becomes more difficult to analyse sperm parameters (concentration and motility) accurately. Since coagulation and liquefaction depend on factors of seminal vesicular and prostatic origin, an alteration of these processes can indicate pathologies of these structures. Perturbed liquefaction is often found in infection or inflammation of the accessory glands, and may explain infertility, if only in part.

The evaluation of seminal viscosity provides information on rheological characteristics common to all biological fluids. The viscosity is assessed by measuring the elasticity of the semen post-liquefaction with a normal semen sample forming threads of less than 2 cm [3]. An increase in viscosity makes microscopic analysis difficult and hinders the assessment of sperm parameters. Hyperviscosity can indicate pathologies of the accessory glands, while reduced viscosity often occurs in cases of serious oligozoospermia or in azoospermia.

Sperm Concentration

An accurate sperm count concentration is essential for the evaluation of fertility and determination of treatment options. WHO or ESHRE methods should be followed including: initial assessment of a 10 µl aliquot from well mixed semen on a slide; subsequent dilution to immobilize all cells according to protocols using a positive displacement pipette for accurate volume of the viscous semen;

a duplicate count on an improved Neubauer Haemocytometer chamber counting >200 sperm per side; and assessment of acceptability of the result from the variance of the two counts as per described methods [3, 5].

Fertile men between the ages of 20 and 40, who are normal from anatomical, urological, andrological, and endocrinological standpoints, may show great variation in the number of spermatozoa between ejaculates. However, this intraindividual variation can be minimized provided that sample collection is carried out properly, and the physical and psychological conditions of the subject do not vary. The lowest 'normal' threshold concentration for potential natural fertilization is 15 million/ml (with >39 million sperm in the entire ejaculate). This figure is based on the fifth centile for fertile men who get partners pregnant within 1 year [3]. Concentrations below this number are indicative of oligozoospermia. The absence of spermatozoa in the ejaculate is termed azoospermia. Semen samples without spermatozoa in a first examination should be centrifuged at 200 g for 15 minutes and the supernatant recentrifuged at 3000 g for 10 minutes; both pellets should be evaluated to confirm the absence of sperm.

Sperm Motility

Sperm motility should be measured after liquefaction or at fixed time points after ejaculation (1–2 h), again according to set WHO/ESHRE protocols [3, 5, 15]. It is important to consider not only the percentage of motile cells, but also the grades of motility [5]. Sperm motilities should be classified as: rapid progressive, with a velocity of at least 25 μm per second (grade A); slow or sluggish progressive (grade B); non-progressive (grade C); or absent (grade D). In normal subjects, 1 hour after ejaculation, the percentage of sperm with motility of grades A + B should be at least 32%, with at least 25% grade A. Values below these are indicative of asthenozoospermia. The importance of characterizing grade A from B sperm motility is because rapid sperm are recognized as those capable of fertilization naturally and in IVF.

Sperm Morphology

Human sperm morphology is classified according to the Tygerberg Strict Criteria approach [3, 5, 16]. Although commonly referred to as '% normal sperm', they are better thought of as 'typical' or ideal sperm, as the classification is based upon the morphology of those which can penetrate through cervical mucus and are found at the site of fertilization. Sperm morphology is evaluated using stained smears. The Sperm-Papanicolau and Quick-Diff techniques are those most commonly used. They have differing fixation effects upon the sperm such that standard morphology classification ranges for parameters such as head size must be specific to the stain employed.

Morphology defects are classified into those of the head; neck/midpiece; tail and abnormally retained cytoplasmic material (often referred to as cytoplasmic droplet). According to the strict criteria classification, any sample having <4% 'typical' or 'normal' forms is considered teratozoospermic [3, 5].

The Teratozoospermia Index, or TZI, combines the four separate morphology assessments of head, midpiece, tail and retained cytoplasm to provide a single number that represents the average number of defects per abnormal morphology sperm. ESHRE recommends multiparametric assessment which allows the calculation of the TZI [5].

Elements Other Than Spermatozoa

Scrutiny of untreated semen or stained smears allows non-sperm cells in the ejaculate to be identified, both in physiological and pathological conditions. The most important cells to identify are leucocytes, of which there should be less than 1 million/ml [3]. Staining (by Sperm-Papanicolau or immunohistochemistry) should be specific for leucocytes to allow differentiation from other spherical cells in semen. Raised leukocyte count in semen is indicative of genital tract infection, should be investigated further and treated, if appropriate, with antibiotics.

Semen may also contain immature germ cells, red blood cells, epithelial cells (released from accessory glands, ducts, and luminal surfaces of the genitourinary tract), prostatic corpuscles, and various other debris. It is normal to observe small amounts of debris in samples. If bacteria are present, then a sample should be sent for microbiological analysis. This requires centrifugation to remove the bacteria from seminal plasma which usually supresses the bacterial culture growth if conventionally streaked on an agar plate.

Vitality Tests

These tests are used in cases where less than 40% of sperm are motile, in order to differentiate spermatozoa that are immotile but alive from those that are dead [5].

The most frequently employed tests are the eosin-nigrosin test and the hypo-osmotic swelling (HOS) test. The eosin test utilizes a vital staining system (eosin Y alone or in combination with nigrosin) which distinguishes between the live (unstained) spermatozoa and the dead (stained) cells. The dead cells, with damaged plasma membranes, are permeable to the dye.

The HOS test evaluates the percentage of spermatozoa with swelling of the tail after incubation in a hypo-osmotic solution. This test is based on the fact that spermatozoa with an intact membrane, when suspended in a hypo-osmotic solution (below 150 mOsm/L), allow the passage of water molecules across the plasma membrane to achieve osmotic equilibrium. They consequently swell, especially at the level of the tail, and show specific morphologic changes. Conversely, dead cells allow the passage of water freely in both directions, and do not show any swelling. Since the HOS test can distinguish dead from vital cells, it has garnered renewed interest in identifying vital sperm for intracytoplasmic injection (ICSI) for *in vitro* fertilization.

Antisperm Antibodies

Seminal antisperm antibodies are auto-antibodies directed against sperm-surface antigens. Their presence is usually suggested by motile sperm being adherent to each other in clumps. Antisperm antibodies are thought to indicate that the blood–testis barrier has been compromised at some stage by illness or trauma. There are a number of different *in vitro* tests available, though they are not routinely used due to difficulties with standardization [3, 5].

Reporting Basic Semen Analysis Results

Results of a basic semen analysis are reported in many different ways. It is important to understand that the WHO reference

ranges are not absolute cut-offs for diagnosis; they are statistical limits below which natural conception may be less likely or take longer. Many patients with hypogonadotropic hypogonadism on gonadotropin treatment can achieve spontaneous pregnancies with sperm densities below 15 or 10 M/ml (for more information see Chapter 10.3.2.3, 'Gonadotrophin Induction of Spermatogenesis').

Secondly, uncertainty of the result related to the limited number of cells examined in any assessment means that absolute value cut-offs are not meaningful in practice [17]. The following table (Table 10.2.3.1), which takes into account accuracy of the result, is suggested as a useful alternative aide to interpretation when WHO methods are followed (taken from [17]).

Genetic Testing After Basic Semen Analysis

If severe oligozoospermia (<5×10^6/ml) or non-obstructive azoospermia are found and assisted conception therapies are planned, then further genetic tests should always be undertaken. These would include karyotype and Y-chromosome microdeletions, and cystic fibrosis transmembrane conductance regulator (CFTR) gene mutation analysis should be offered to all males with congenital bilateral absence of the vas deferens or cystic fibrosis [7] (see Chapter 10.3.3).

Advanced Semen Analysis Methods

Beyond a 'basic semen analysis' there are a multitude of further fertility analyses that can optionally be performed on patients or their samples. These tests are often not widely available as they are not a required part of a standard WHO semen analysis, and often the evidence for their use or interpretation is limited. These more complex and costly analytical tools should be considered of secondary or tertiary importance, and are to be carried out in specific cases only after standard semen analyses have been performed. For some men, results from basic semen analysis may indicate the need for more advanced tests looking in more detail at the seminal plasma and sperm functionality.

Biochemical Study of Seminal Plasma

Biochemical analysis of the seminal fluid can make a notable contribution to the differential diagnosis of azoospermia. Thus, in primary hypospermatogenic azoospermia, with normal androgenic production, there are no abnormalities in the secretions of the accessory glands and seminal plasma biochemistry is normal. In contrast, subjects with azoospermia due to congenital absence of the vas deference and seminal vesicle or bilateral obstruction of the ejaculatory ducts, there is a relatively high concentration of citric acid and very low or undetectable fructose. Subjects with obstruction of the efferent ducts have extremely low levels of free carnitine and α-glucosidase, while fructose and citric acid levels remain normal [18].

Sperm Function Tests

Many sperm function tests can be used in parallel with sperm analysis to establish a complete picture of the fertilizing capacity of sperm and perhaps guide further management including assisted reproduction [19, 20].

Sperm Migration and Interaction with the Female Genital Tract

Sperm–cervical mucous interaction tests (the *in vivo* postcoital test, or PCT) and the *in vitro* assay of sperm migration in cervical mucous provide valuable information about the condition of cervical mucous, and related hormonal functions in the female partner, and also sperm survival after intercourse. However, these tests are difficult to standardize and so have fallen out of routine use.

Sperm Kinetics and CASA

Many endeavours have been made to find an objective method to read sperm kinetics. Computer Assisted Semen Analysis (CASA) methods employ digital image acquisition and analysis to analyse sperm kinetic parameters such as velocity, linearity, lateral head displacement, and beat cross frequency. Although CASA may provide objective data on sperm motility, it is essentially still a research tool at the time of writing [21].

Table 10.2.3.1 Basic semen analysis results

Parameter		Pathological	Borderline	Normal	Unit
Concentration		**<10**	10–20	**20–250**	10^6/ml
Volume		**<1.5**	1.5–1.9	**2.0–6.0**	ml
Total sperm count		**<20**	20–79	>**80**	10^6/ejaculate
Motility	Motile	**<40**	40–59	>**60**	%
	Progressive	**<35**	35–49	>**50**	%
	Rapid			>**25**	%
Vitality		**<40**	40–59	>**60**	%
Morphology: typical forms		**<4**	4–13	>**14**	%
TZI*		**>1.80**	1.61–1.80	<**1.60**	

* TZI—the Teratozoospermia Index or TZI, combines the four separate morphology assessments of head, midpiece, tail and retained cytoplasm to provide a single number that represents the average number of defects per abnormal morphology sperm.

Other Advanced Sperm Function Tests

There are many other tests to investigate different aspects sperm functions, including sperm intracellular calcium signalling in response to progesterone challenge, sperm capacitation, acrosome reaction, sperm-oocyte interaction, and oocyte activation assays. Currently these are research methods without proven clinical application. The interested reader can refer to [22–28] for more details.

Sperm Nuclear Function

In recent years there has been great interest in the value of assessing sperm DNA packaging and DNA fragmentation/damage as a predictor of both natural conception and the assisted reproduction outcomes. Currently there is no standardization in different tests of sperm chromatin structure and sperm DNA quality. Current evidence does not support the routine use of sperm DNA testing [29–31]. It is also worth noting that there appears to be a relationship between sperm DNA and miscarriage for certain assays [31, 32]. With a number of assays in use in different laboratories and no clearly defined clinically relevant thresholds, the situation for these tests is rapidly evolving and routine use widely debated [30, 33–35].

Quality Control

Clinicians rely on accurate and well-standardized semen analysis results from the diagnostic andrology laboratory that can be used to inform the diagnosis and treatment of patients with fertility disorders. Unfortunately, evaluations of sperm parameters continue to demonstrate marked differences both within and between laboratories [36–38], a point usually not appreciated by clinicians outside Reproductive Medicine. Effective quality control is therefore mandatory in identifying random and systematic laboratory errors and to ensure that all analytical and non-analytical procedures are sufficient to ensure the most accurate results. The WHO manual and ESHRE recommendations both recommend strict quality control procedures and that all laboratories providing a semen diagnostic service should participate in an accredited EQA scheme [3, 5]. To implement these recommendations in practice, clinicians should only request and accept results from EQA-accredited andrology laboratories.

REFERENCES

1. Barratt CLR, Björndahl L, De Jonge CJ, *et al*. The diagnosis of male infertility: an analysis of the evidence to support the development of global WHO guidance-challenges and future research opportunities. *Hum Reprod Update*, 2017; **23**: 660–80.
2. Oehninger S, Kruger T, Seracchioli R, Porcu E, Flamigni C. The diagnosis of male infertility by semen quality [debate]. *Hum Reprod*, 1995; **10**: 1037–41.
3. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva: WHO, 2010.
4. Kvist U, Björndahl L, eds. *Basic Semen Analysis*. Oxford: Oxford University Press, ESHRE Monographs, 2002.
5. Björndahl L, Mortimer D, Barratt CLR, *et al*. *A Practical Guide to Basic Laboratory Andrology*. Cambridge: Cambridge University Press, 2010.
6. Castilla JA, Alvarez C, Aguilar J, González-Varea C, Gonzalvo MC, Martínez L. Influence of analytical and biological variation on the clinical interpretation of seminal parameters. *Hum Reprod*, 2006; **21**: 847–51.
7. Cooper TG, Keck C, Oberdieck U, Nieschlag E. Effects of multiple ejaculations after extended periods of sexual abstinence on total, motile and normal sperm numbers, as well as accessory gland secretions, from healthy normal and oligozoospermic men. *Hum Reprod*, 1993; **8**: 1251–8.
8. Hancock P, Woodward BJ, Muneer A, Kirkman-Brown JC. 2016 Laboratory guidelines for postvasectomy semen analysis: Association of Biomedical Andrologists, the British Andrology Society and the British Association of Urological Surgeons. *J Clin Pathol*, 2016; **69**: 655–60.
9. Pound N, Javed MH, Ruberto C, Shaikh MA, Del Valle AP. Duration of sexual arousal predicts semen parameters for masturbatory ejaculates. *Physiol Behav*, 2002; **76**: 685–9.
10. Tash JA, McGovern JH, Schlegel PN. Acquired hypogonadotropic hypogonadism presenting as decreased seminal volume. *Urology*, 2000; **56**: 669.
11. Barazani Y, Stahl PJ, Nagler HM, Stember DS. Management of ejaculatory disorders in infertile men. *Asian J Androl*, 2012; **14**: 525–9.
12. Shoshany O, Abhyankar N, Elyaguov J, Niederberger C. Efficacy of treatment with pseudoephedrine in men with retrograde ejaculation. *Andrology*, 2017; **5**: 744–8.
13. Pentyala S, Lee J, Annam S, *et al*. Current perspectives on pyospermia: a review. *Asian J Androl*, 2007; **9**: 593–600.
14. Munkelwitz R, Krasnokutsky S, Lie J, Shah SM, Bayshtok J, Ali Khan S. Current perspectives on hematospermia: a review. *J Androl*, 1997; **18**: 6–14.
15. Björndahl L. The usefulness and significance of assessing rapidly progressive spermatozoa. *Asian J Androl*, 2010; **12**: 33–5.
16. Kruger G, Acosta A, Simmons K, Swanson RJ, Matta RJ, Oehninger S. Predictive value of abnormal sperm morphology in *in vitro* fertilization. *Fertil Steril*, 1988; **49**: 112–17.
17. Björndahl L. What is normal semen quality? On the use and abuse of reference limits for the interpretation of semen analysis results. *Hum Fertil (Camb)*, 2011; **14**: 179–86.
18. Mann T, Lutwak-Mann C. *Biochemistry of Seminal Plasma and Male Accessory Fluids: Application to Andrological Problems*. New York: Springer-Verlag, 1981.
19. Lenzi A. Male infertility: evaluation of human sperm function and its clinical application. *J Endocrinol Invest*, 1995; **8**: 468–88.
20. Sigman M, Zini A. Semen analysis and sperm function assays: what do they mean? *Semin Reprod Med*, 2009; **27**: 115–23.
21. Gallagher M, Smith J, Kirkman-Brown J. CASA: tracking the past, and plotting the future. *Reprod Fertil Dev*, 2018; **30**: 867–74.
22. Kelly MC, Brown SG, Costello SM, *et al*. Single-cell analysis of [Ca2+]i signalling in sub-fertile men: characteristics and relation to fertilization outcome. *Hum Reprod*, 2018; **33**: 1023–33.
23. Cardona C, Neri QV, Simpson AJ, *et al*. Localization patterns of the ganglioside GM1 in human sperm are indicative of male fertility and independent of traditional semen measures. *Mol Reprod Dev*, 2017; **84**: 423–35.
24. Brucker C, Lipford GB. The human sperm acrosome reaction: physiology and regulatory mechanisms. An update. *Hum Reprod Update*, 1995; **1**: 51–5.

25. Visconti PE, Galantino-Homer H, Moore GD, *et al*. The molecular basis of sperm capacitation. *J Androl*, 1998; **19**: 242–8.
26. Bianchi E, Doe B, Goulding D, Wright GJ. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*, 2014; **508**: 483–7.
27. Liu DY, Baker HWG. A new test for the assessment of sperm-zona pellucida penetration: relationship with results of other sperm tests and fertilization *in vitro*. *Hum Reprod*, 1994; **9**: 489–94.
28. Yeste M, Jones C, Amdani SN, Patel S, Coward K. Oocyte activation deficiency: a role for an oocyte contribution? *Hum Reprod Update*, 2016; **22**: 23–47.
29. Cissen M, Wely MV, Scholten I, *et al*. Measuring sperm DNA fragmentation and clinical outcomes of medically assisted reproduction: a systematic review and meta-analysis. *PLoS One*, 2016; **11**: e0165125.
30. Simon L, Zini A, Dyachenko A, Ciampi A, Carrell DT. A systematic review and meta-analysis to determine the effect of sperm DNA damage on *in vitro* fertilization and intracytoplasmic sperm injection outcome. *Asian J Androl*, 2017; **19**: 80–90.
31. Kirkman-Brown JC, De Jonge C. Sperm DNA fragmentation in miscarriage—a promising diagnostic, or a test too far? *Reprod Biomed Online*, 2017; **34**: 3–4.
32. Robinson L, Gallos ID, Conner SJ, *et al*. The effect of sperm DNA fragmentation on miscarriage rates: a systematic review and meta-analysis. *Hum Reprod*, 2012; **27**: 2908–17.
33. Lefièvre L, Bedu-Addo K, Conner SJ, *et al*. Counting sperm does not add up any more: time for a new equation? *Reproduction*, 2007; **133**: 675–84.
34. Krawetz SA. Paternal contribution: new insights and future challenges. *Nat Rev Genet*, 2005; **6**: 633–42.
35. Wu TF, Chu DS. Sperm chromatin: fertile grounds for proteomic discovery of clinical tools. *Mol Cell Proteomics*, 2008; **7**: 1876–86.
36. Cooper TG. Internal quality control of semen analysis. *Fertil Steril*, 1992; **58**: 172–8.
37. Neuwinger J, Cooper TG, Knuth UA, Nieschlag E. External quality control in the andrology laboratory: an experimental multicenter trial. *Fertil Steril*, 1990; **54**: 308–14.
38. Tomlinson M. Is your andrology service up to scratch? *Hum Fertil (Camb)*, 2010; **13**: 194–200.

FURTHER READING

Björndahl L, Mortimer D, Barratt CLR, *et al*. *A Practical Guide to Basic Laboratory Andrology*. Cambridge: Cambridge University Press, 2010.

10.3

Klinefelter's Syndrome

Claus H. Gravholt

Introduction

Klinefelter syndrome (KS) is the most common sex chromosome aneuploidy and was first described by HF Klinefelter in 1942, publishing nine cases of a syndrome characterized by gynaecomastia, azoospermia, hyalinized small testes, elevated levels of follicle-stimulating hormone (FSH), and hypogonadism [1]. Jacobs and Strong demonstrated in 1959 the extra X chromosome in the karyotype of KS [2].

Knowledge concerning KS expanded through chromosome surveys on large number of unselected newborns, with follow-up through childhood and adolescence performed in the seventies and eighties [3–5]. These studies on newborns focused mainly on development in infancy, childhood, and adolescence, leaving the natural history of KS in adulthood less well described. The typical man suffering from KS is described as tall, with narrow shoulders, broad hips, sparse body and facial hair, gynaecomastia, small testes, androgen deficiency and reduced intelligence [6], but a less severe phenotype is present in many patients [7]. Typical features are listed in **Table 10.3.1**. However, this current clinical picture may only represent a 'tip of the iceberg' glimpse of the full phenotypic spectrum, since only about 25% of all KS males are ever diagnosed [8]. If and when all KS males are diagnosed, the entire literature on KS may well have to be re-written.

Genetic Background

KS is characterized by the presence of one Y chromosome and two X chromosomes in a phenotypical male. In the diagnosed Danish population of KS patients 90% had the 47,XXY karyotype, while mosaic forms are also seen [9, 10]. Non-mosaic 47,XXY most often occurs due to non-disjunction of the sex chromosomes during the first or second meiotic division of gametogenesis in either parent. There is evidence that failure of the paternal XY or maternal XX chromosomes to pair and recombine in the usual way is a causative factor, as is also the case for autosomal trisomies [7, 11]. Paternal meiosis I errors account for approximately 50% of 47,XXY karyotypes, while meiosis II errors does not lead to 47,XXY karyotypes [11]. Of the remaining maternally derived karyotypes, 22% to 48% are due to meiosis I errors, 9% to 29% to meiosis II errors, 7% to unknown meiotic errors, and 3% to 16% to postzygotic mitotic non-disjunction [7, 11, 12]. In prenatally ascertained cases, 54% are paternally derived and 46% are maternally derived [12]. The phenotype does not appear to be influenced by the parental origin of the extra X chromosome, though few studies have specifically addressed this issue [13]. The molecular basis of KS remains poorly understood, although new studies point towards complex genotype-phenotype relations dependent on both changes in DNA methylation profile and RNA expression, likely tissue dependent [14–16] (see online **Figure 10.3.1**). The expression of X-linked genes in the pseudoautosomal regions of the X and Y chromosomes that escape inactivation may also be involved [7], but the identity and function of the critical genes have not been determined. Presence of an extra copy of the short stature homeobox-containing gene on chromosome X (*SHOX*) is probably the basis for the increased height in KS [17]. *SHOX* is a homeodomain transcription factor and brain natriuretic peptide and fibroblast growth factor receptor 3 are transcriptional targets of *SHOX*, and it has been shown that extra copies lead to excessive height gain [17].

A KS diagnosis should always be based on clinical findings combined with a confirmatory cytogenetic evaluation. On the other hand, a phenotypically normal man with very low-grade mosaicism (probably below 5–10%) should not have the diagnosis of KS.

Table 10.3.1 Abnormalities and diseases associated with Klinefelter syndrome

Feature	Frequency (%) or increased risks
Infertility [92]	>99
Azoospermia [92]	>95
Decreased bitesticular testis volume (4–8 ml; normal range: 25–60 ml)) [18, 35, 93]	>95
Spermatozoa retrievable after TESE	30–50
Decreased beard growth and pubic hair# [92]	30–80
Abdominal adiposity [43]	~50
Decreased muscle mass and strength [35, 45]	~40
The metabolic syndrome [43]	46
Type 2 diabetes [43]	10–39
Osteopenia [45, 93]	~40
Osteoporosis [45, 94]	5–10
Mitral valve prolapse [95, 96]	0–50
Ischaemic heart disease [41, 42]	~1.5 fold*
Deep venous thrombosis and pulmonary embolism [41, 48, 97]	3–6-fold*
Autoimmunity [98–100]	Increased risk of several autoimmune diseases
Tremor (Parkinson-like symptoms) [101, 102]	>25
Breast cancer [103–105]	~4 fold*
Osteoarthritis [41]	4-fold increased risk*
Learning disability [81]	>75
Delayed speech development [81]	>40
Decreased penile size [81, 106, 107] incl. micropenis neonatally?	10–25
Mediastinal cancer [108]	Increased risk
Gynaecomastia [35, 81, 109]	28–75
Cryptorchidism [18, 41, 81]	27–37
Increased gonadotropin levels# [18, 92, 93]	>75
Decreased testosterone levels# [18, 92, 93]	>75
Increased height [35, 81]	>30
Psychiatric disturbances [81, 110]	>25
Congenital malformations (heart, cleft palate, and inguinal hernia) [41, 96, 111]	Increased risk
Fractures [41, 97]	Increased risk (2–40-fold)*
Autism spectrum disorder [112, 113]	30–50

TESE, testicular sperm extraction; * above normal male frequency; # in the untreated condition.

Adapted with permission from Gravholt CH, Chang S, Wallentin M, Fedder J, Moore P, Shakkebaek A. *Klinefelter Syndrome: Integrating Genetics, Neuropsychology, and Endocrinology.* Copyright © 2018 Endocrine Society. doi: 10.1210/er.2017-00212

Underdiagnosis and delayed diagnosis of KS is a major problem [8]. Importantly, early diagnosis permits identification of speech problems and scholastic difficulties that require special measures during childhood and adolescence to improve educational outcomes. Moreover, early diagnosis facilitates prevention or amelioration of the long-term consequences of gonadal insufficiency.

Summarizing epidemiological data across countries and ethnicities yields a prevalence estimate of 152 per 100 000 [8], but with a major discrepancy between the prenatal and postnatal prevalence, with only 25% of the expected number of KS being diagnosed postnatally and less than 10% of these being diagnosed before puberty. Prenatal diagnosis of a fetus with 47,XXY karyotype requires professional genetic counselling to inform the pregnant couple about the relatively good prognosis, but at the moment in Europe 75% of couples expecting a child with KS choose termination [8]. Since only 25% of the expected number of men with KS are diagnosed, ascertainment bias is a major problem when interpreting data from different studies on different populations of KS patients.

Congenital Malformations

There are no specific stigmata at birth, but an increased incidence of congenital malformations has been reported [18]. Minor congenital abnormalities were found in 26%, with clinodactyly of the fifth finger as the most frequent [19] and major congenital abnormalities were found in 18% of KS boys, with cleft palate, inguinal hernia and cryptorchidism being the most frequent [19]. In a large clinical study 27% of KS patients had a history of maldescent of testes compared to 8% of the total number of patients attending the same clinic [18].

Testicular Development

The typical testicular histology includes hyalinization of seminiferous tubules, loss of germ cells, and Leydig cell hyperplasia. In about 40–50% of KS patients, occasional focal spermatogenesis may be found with the possibility of surgically retrieved spermatozoa being used for intracytoplasmic sperm injection (ICSI) into oocytes to achieve *in vitro* fertilization [20]. The cause of the hyalinization of the testes with subsequent hypogonadism and infertility is unknown. There is a progressive loss of spermatogonia from infancy [21] while hyalinization of the seminiferous tubules probably does not occur until mid-puberty [22]. At the beginning of puberty [23], testes enlarge to approximately 4 ml and thereafter shrink in adulthood [24, 25]. Testes may be malfunctioning already during intrauterine life, since micropenis seen in some newborn males with KS may be a result of decreased testosterone production *in utero* [26]. Some find the normal surge in testosterone seen in the first 1 to 6 months of life (minipuberty) attenuated in KS boys [26, 27], while others find it similar to controls [28]. The levels of FSH, luteinizing hormone (LH) and testosterone are normal during the prepubertal period, but after the onset of puberty a rise in FSH and LH and a decline in testosterone occur, compared to normal boys [24], along with a decline in insulin-like factor 3 (INSL3), a marker of Leydig cell function [29], while 17β-oestradiol is normal despite the relatively lower levels of testosterone [30]. Even though levels of testosterone for many patients are within the normal range, levels of gonadotrophins are usually elevated [6, 30, 31], indicating a relative or compensated hypogonadism that is reflected by the increased pituitary drive, and hence may be one of the causes of Leydig cell hyperplasia.

Fertility

KS patients are usually considered infertile, but with the recent development of the testicular sperm extraction (TESE) and intracytoplasmic sperm injection (ICSI) methodologies, it is now possible to extract viable sperm from the testes by surgical biopsy and inject it directly into an extracted ovum. Successful extraction rates of sperm in clinical practice are around 30–45% [32], and about half of these experience a pregnancy [33] with fewer live childbirths [34] (also see Chapter 10.3.3, 'Management of Idiopathic Male Infertility'). Specialized genetic counselling and options of prenatal diagnosis or even preimplantation genetic diagnosis should be offered to couples seeking infertility treatment where the male is suffering from KS, although so far, there does not seem to be an increased risk of KS in the offspring.

Hypogonadism

Currently it is thought that all KS males will develop hypogonadism, some already in puberty, while others develop hypogonadism later in adult life [18, 35–37]. Hypogonadism may delay or reduce the development of normal male secondary sexual characteristics with less growth of beard, muscle bulk, and secondary body hair [6]. Sexuality may be affected—fewer 47,XXY men are married, experience greater sexual dissatisfaction in general, while more acknowledge unconventional sexual experiences and demonstrate a less masculine gender role compared to a control group [38, 39].

Hypogonadism may also contribute to the neuro-cognitive phenotype seen in KS. The long-term consequences of hypogonadism in KS are difficult to separate from the gene-dose effects of having an extra X chromosome, because studies on comparable hypogonadal conditions (e.g. hypogonadotropic hypogonadism) are lacking due to their infrequency.

Gynaecomastia

In KS the prevalence of gynaecomastia is markedly increased, up to 50% [6] in some series, although the true prevalence probably is lower [40]. The decreased testosterone level in concert with a normal level of oestradiol may cause gynaecomastia [24], because of an altered oestradiol/testosterone ratio. Testosterone treatment seldom leads to a regression of gynaecomastia, but most KS patients will need to have the breast tissue removed surgically.

Diabetes and the Metabolic Syndrome

Epidemiological studies on both morbidity [41] and mortality [42] show a fourfold increased risk of diabetes, and clinical studies describe a strikingly high frequency of the metabolic syndrome and insulin resistance in KS [43, 44], with increased cholesterol levels. KS subjects accumulate excessive amounts of body fat, especially truncal fat, and typically have a low level of physical fitness [30, 45].

Cardiovascular Disease

Mortality from cardiovascular diseases is increased in KS [10, 46], even though the mortality from ischaemic heart disease is reduced [10]. The risk of hypostatic leg ulcers may be significantly increased [47]. In addition, the risk of deep venous thrombosis and pulmonary embolism is 4–6-fold higher than in control males [41, 48]. Dysfunction of the fibrinolytic system has been proposed as a reason for this and increased activity of plasminogen activator inhibitor 1(PAI-1) in KS patients was described in one study [49], but further studies are needed.

Osteoporosis

Hypogonadism is a known cause of secondary osteoporosis in both females and males [50]. A majority of studies show a significant reduction in bone mineral density (BMD) in KS males compared to normal men [51–54]. Although fragility fractures is not as commonly reported in the few clinical studies of adults with KS as would be expected from the BMD deficits, nevertheless epidemiological data consistently show increased risks of hospital admissions with osteoporotic fractures (forearm, hip, and spinal fractures) [55] and increased mortality from hip fractures [10]. Hence, the reported reduction in BMD is likely to signal an important and potentially preventable clinical problem in KS patients.

Cancer

The overall risk of cancer is not increased in KS, but the specific risk of breast cancer is increased compared to the general male population [56, 57]. In addition, there is a significantly increased risk of mediastinal germ cell tumours [58, 59], an increased mortality from lung cancer, and non-Hodgkin lymphoma, but a significantly reduced mortality from prostate cancer.

Cognitive Disturbances

Global intelligence is somewhat lower in KS [4, 5, 60, 61], with decreased verbal intelligence, delayed development of speech, and with a high proportion of dyslexia and educational problems in many, but not all KS [62–65]. One study on adults with sex chromosome abnormalities, including KS and 47,XYY syndrome, detected at birth, showed that the decreased verbal IQ and reading impairment persisted in adulthood [66], but also that a large variability in educational and vocational achievement was present. The reason for the delay in speech and decreased verbal intelligence is unclear, but MRI studies have shown consistent changes in different nuclei of the brain [67–69] (see online Figure 10.3.2), although these changes have not been linked with specific changes in the neuropsychologic profile [70–72].

Psychiatric Diseases

Increased incidence of psychiatric illness were reported in KS primarily based on screening for sex chromosome abnormalities in penal institutions, psychiatric hospitals and institutions for the mentally retarded [73]. In a long-term follow-up study of KS males identified in a chromosome survey in Edinburgh, an increased frequency of referral to psychiatric treatment was seen [5]. A multinational survey for sex chromosome aberrations among patients

with schizophrenia found a four- to fivefold excess of KS compared with general population data [74]. An epidemiological study on hospital admissions in KS showed a significantly increased risk of discharge with a psychiatric diagnosis [55], and patients may show an increase in schizophrenia-spectrum pathology [75]. Overexpression of X-linked genes escaping X-inactivation is thought to be involved in some of the psychiatric disturbances seen in KS [76].

Criminality

One study found no increased frequency of criminal behaviour in KS judged by self-reported penalties [77]. Another study following 34 KS patients for 10 and 20 years showed increased criminal behaviour at 10 years follow-up, compared with both hypogonadal men (from other causes than KS) and with a eugonadal control group [60] but after 20 years of follow-up there was no significant difference in criminality [78]. Recently published data on 32 KS patients diagnosed prepubertally or during puberty showed a large proportion of patients with severe psychosocial problems, with 69% having problems with controlling aggression, 28% offending the law and 18% having convictions [79]. A recent large epidemiological study (n = 934) showed increased criminality specifically related to arson, sexual abuse and burglary compared to a large background population (n = 88 979) [80].

Treatment

Optimal treatment and care of patients with KS involves a multidisciplinary team including speech therapists, psychologists, general practitioners, paediatricians, endocrinologists, urologists, and infertility specialists. Although most clinicians believe that testosterone treatment has a positive impact, both physically and psychologically, no randomized controlled trials on the efficacy and safety of testosterone replacement therapy have been performed in KS, mainly due to small sample sizes in most clinics and lack of research funding support. Current practice is largely based on observational data from KS patients and evidence extrapolated from other hypogonadal conditions, so that some areas of management remain controversial and without clear consensus.

Infants with KS are rarely diagnosed because they lack KS-specific stigmata. However, some KS boys have micro penis, which anecdotally can respond to treatment with topical testosterone cream or single injections of intramuscular testosterone in some cases. The most serious problem in early childhood is the delay of speech development affecting perhaps half of the boys with KS [5]. Careful observation for speech delay is needed in order to refer these boys to speech therapists promptly. The same holds true for learning disabilities, which were observed in 77% of boys with KS followed from birth to adulthood [81].

At puberty, if **both** LH and FSH rise above normal, some centres recommend initiating testosterone treatment with the aim of promoting a full development of male secondary sexual characteristics and ensuring sufficient increase in muscle bulk and BMD [82], although such recommendations are not based on high level evidence. Testosterone treatment in pubertal KS boys has also been reported to increase energy, endurance, better mood and concentration and relations to others [83]. Retrospective chart data on a group of KS boys diagnosed before and during puberty (n = 32, whereof 17 received testosterone treatment) showed increased psychosocial problems in periods without testosterone treatment [84]. In a minority of KS adolescents with particular behavioural issues, (such as aggression, socialization, and communication problems), initiation of testosterone treatment should proceed cautiously and preferably with supports from psychologists with neurocognitive therapy experience.

Treatment in a large group of young hypogonadal men of mixed origins (wherein some had KS) has been shown to have a positive impact on fat mass, muscle mass, and muscle strength, as well as sexual function, with some improvements in positive aspects of mood [85].

There is general agreement that KS patients with testosterone levels at the lower end of the normal range, and with elevated FSH and LH levels, indicating hypogonadism to be present [37], should be treated if they complain of hypogonadal symptoms (lack of libido, erectile problems, decreased energy, etc.) [86–88]. Based on our own experience, we believe that all KS patients (asymptomatic as well as symptomatic) should receive testosterone treatment if their gonadotrophins are elevated, even if their testosterone levels are within the lower end of the normal range. Most, if not all, KS patients, who at first sight may seem asymptomatic, but after a more comprehensive assessment (see Table 10.3.1) using a combination of questionnaires with focus on quality of life [89], body composition, sleep quality [113], and biochemistry, will be found to have some features related to hypogonadism. In this context it is important to remember that many KS patients are not good rapporteurs concerning their own health, and that physicians should take due care to cover all relevant areas when interviewing these patients. This is supported by the frequent finding of excessive truncal fat, low muscle mass, decreased bone mass and microarchitectural bone changes [43, 87], reduced haemoglobin [35], presence of gynaecomastia [35], decreased sexual function [88], and depressive symptoms [89] in untreated KS patients. However, while there is good evidence that testosterone treatment improves these features in hypogonadal men (including KS) with low testosterone, there is currently no evidence from randomized controlled trials (RCTs) that testosterone treatment can reverse or improve these features in KS men with low or low-normal testosterone and elevated LH (see Chapter 10.4.4, 'Benefits of Testosterone Treatment'). More data are required before our practice outlined earlier is more widely adopted. Interestingly, only 48.6% of 1155 KS patients in the national Danish registry were ever prescribed testosterone, starting only at the age of 30.4 years [90].

We believe that testosterone treatment should aim for normalization in LH and testosterone levels in the mid-normal range by transdermal treatment and low-normal preinjection nadir values of testosterone when treatment is by injection therapy. In the authors' experience, LH and testosterone can be normalized in virtually all KS patients if the testosterone dose is carefully titrated, using recommended regimes for physiological replacement of testosterone. It may be inferred that many KS patients are not sufficiently treated, because LH often remain elevated. However, concerns have been raised that this approach (aim to normalize LH) may lead to excessive T exposure in some patients, especially from injectable therapy.

As with other forms of hypogonadism, improvements in subjective symptoms and other patient important outcomes should be key goals of testosterone replacement therapy in KS (in addition to normalization of biochemical endpoints). Treatment is recommended to be continued lifelong in order to prevent osteoporosis, obesity, metabolic syndrome and diabetes, and to ensure optimal cognitive development.

Temporary discontinuation of exogenous testosterone as well as treatment with human chorionic gonadotrophin (hCG) and aromatase inhibitors to increase intratesticular testosterone, are empirical approaches commonly employed to increase the chance of sperm recovery by TESE [91], although they have not been proven to be efficacious. These practices should be carefully audited and confirmed by evidence, preferably in a randomized trial setting and should not be extrapolated to influence the imperative of regular testosterone replacement therapy in KS. Currently, there is no evidence to support withholding testosterone treatment to young men with KS that have still not attempted TESE and may not do so for years.

Conclusions

Klinefelter syndrome is a frequent cause of hypogonadism and infertility. It is the most common sex chromosome disorder in man, with 47,XXY being the most common karyotype. The syndrome affects one in every 660 males but is severely underdiagnosed; only 25–50% of the expected number of cases is diagnosed and only a minority of them before puberty.

In children the main clinical findings are delayed speech development and learning disabilities as well as increased height and cryptorchidism. In adults the main reasons for diagnosing KS are infertility and hypogonadism. KS is associated with a significant increase in mortality and morbidity from a number of diseases.

In accordance with accepted practice for hypogonadism, there is general agreement that **symptomatic** KS men with unequivocally low testosterone should receive testosterone replacement. However, the exact indications for testosterone treatment in apparently **asymptomatic** KS men with low-normal testosterone and elevated FSH and LH remain uncertain. While some clinicians (including the authors) routinely initiate testosterone treatment in all KS men with elevated LH and FSH, aiming to counter the presumptive harmful effects of compensated hypogonadism, not all authorities agree with this approach due to the current lack of high level evidence from RCTs.

REFERENCES

1. Klinefelter HF, Reifenstein EC, Albright F. Syndrome characterized by gynecomastia, aspermatogenesis without a-leydigism, and increased excretion of follicle-stimulating hormone. *J Clin Endocrinol*, 1942; **2**: 615–27.
2. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*, 1959; **183**: 302–3.
3. Nielsen J, Sorensen AM, Sorensen K. Follow-up until age 7 to 11 of 25 unselected children with sex chromosome abnormalities. *Birth Defects Orig Artic Ser*, 1982; **18**: 61–97.
4. Robinson A, Bender BG, Linden MG, Salbenblatt JA. Sex chromosome aneuploidy: the Denver Prospective Study. *Birth Defects Orig Artic Ser*, 1990; **26**: 59–115.
5. Ratcliffe S. Long term outcome in children of sex chromosome abnormalities. *Arch Dis Child*, 1999; **80**: 192–5.
6. Smyth CM, Bremner WJ. Klinefelter syndrome. *Arch Intern Med*, 1998; **158**: 1309–14.
7. Simpson JL, de la Cruz F, Swerdloff RS, *et al*. Klinefelter syndrome: expanding the phenotype and identifying new research directions. *Genet Med*, 2003; **5**: 460.
8. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *J Clin Endocrinol Metab*, 2003; **88**: 622–6.
9. Gravholt CH, Chang S, Wallentin M, Fedder J, Moore P, Skakkebæk A. Klinefelter syndrome: integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*, 2018; **39**: 389–423.
10. Swerdlow AJ, Higgins CD, Schoemaker MJ, Wright AF, Jacobs PA, United Kingdom Clinical Cytogenetics Group. Mortality in patients with Klinefelter syndrome in Britain: a cohort study. *J Clin Endocrinol Metab*, 2005; **90**: 6516–22.
11. Thomas NS, Hassold TJ. Aberrant recombination and the origin of Klinefelter syndrome. *Hum Reprod Update*, 2003; **9**: 309–17.
12. Manning MA, Hoyme HE. Diagnosis and management of the boy with Klinefelter syndrome. *Adolesc Med*, 2003; **13**: 367–74.
13. Jacobs PA, Hassold TJ, Whittington E, *et al*. Klinefelter's syndrome: an analysis of the origin of the additional sex chromosome using molecular probes. *Ann Hum Genet*, 1988; **52**: 93–109.
14. Sharma A, Jamil MA, Nuesgen N, *et al*. DNA methylation signature in peripheral blood reveals distinct characteristics of human X chromosome numerical aberrations. *Clin Epigenetics*, 2015; **7**: 76.
15. Winge SB, Dalgaard MD, Jensen JM, *et al*. Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible involvement of long non-coding RNAs in gonocyte maturation *Hum Mol Genet*, 2018; **27**: 430–9.
16. Skakkebaek A, Nielsel MM, Trolle C, *et al*. DNA hypermethylation and differential gene expression associated with Klinefelter syndrome. *Sci Rep*, 2018; **8**: 13740.
17. Ottesen AM, Aksglaede L, Garn I, *et al*. Increased number of sex chromosomes affects height in a nonlinear fashion: a study of 305 patients with sex chromosome aneuploidy. *Am J Med Genet A*, 2010; **152A**: 1206–12.
18. Lanfranco F, Kamischke A, Zitzmann M, Nieschlag E. Klinefelter's syndrome. *Lancet*, 2004; **364**: 273–83.
19. Robinson A, Lubs HA, Nielsen J, Sorensen K. Summary of clinical findings: profiles of children with 47,XXX and 47,XYY karyotypes. *Birth Defects Orig Artic Ser*, 1979; **15**: 261–6.
20. Denschlag D, Tempfer C, Kunze M, Wolff G, Keck C. Assisted reproductive techniques in patients with Klinefelter syndrome: a critical review. *Fertil Steril*, 2004; **82**: 775–9.
21. Muller J, Skakkebaek NE, Ratcliffe SG. Quantified testicular histology in boys with sex chromosome abnormalities. *Int J Androl*, 1995; **18**: 57–62.
22. Wikström AM, Raivio T, Hadziselimovic F, Wikström S, Tuuri T, Dunkel L. Klinefelter syndrome in adolescence: onset of puberty is associated with accelerated germ cell depletion. *J Clin Endocrinol Metab*, 2004; **89**: 2263–70.
23. Christiansen P, Andersson AM, Skakkebaek NE. Longitudinal studies of inhibin B levels in boys and young adults with Klinefelter syndrome. *J Clin Endocrinol Metab*, 2003; **88**: 888–91.

24. Salbenblatt JA, Bender BG, Puck MH, Robinson A, Faiman C, Winter JS. Pituitary-gonadal function in Klinefelter syndrome before and during puberty. *Pediatr Res*, 1985; **19**: 82–6.
25. Wikström AM, Dunkel L. Klinefelter syndrome. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 239–50.
26. Ross JL, Samango-Sprouse C, Lahlou N, Kowal K, Elder FF, Zinn A. Early androgen deficiency in infants and young boys with 47,XXY Klinefelter syndrome. *Horm Res*, 2005; **64**: 39–45.
27. Lahlou N, Fennoy I, Carel JC, Roger M. Inhibin B and anti-Müllerian hormone, but not testosterone levels, are normal in infants with nonmosaic Klinefelter syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 1864–8.
28. Aksglaede L, Petersen JH, Main KM, Skakkebaek NE, Juul A. High normal testosterone levels in infants with non-mosaic Klinefelter's syndrome. *Eur J Endocrinol*, 2007; **157**: 345–50.
29. Bay K, Hartung S, Ivell R, *et al*. Insulin-like factor 3 serum levels in 135 normal men and 85 men with testicular disorders: relationship to the luteinizing hormone-testosterone axis. *J Clin Endocrinol Metab*, 2005; **90**: 3410–18.
30. Bojesen A, Kristensen K, Birkebaek NH, *et al*. The metabolic syndrome is frequent in Klinefelter's syndrome and is associated with abdominal obesity and hypogonadism. *Diabetes Care*, 2006; **29**: 1591–8.
31. Lanfranco F, Kamischke A, Zitzmann M, Nieschlag E. Klinefelter's syndrome. *Lancet*, 2004; **364**: 273–83.
32. Garolla A, Selice R, Menegazzo M, *et al*. Novel insights on testicular volume and testosterone replacement therapy in Klinefelter patients undergoing testicular sperm extraction. A retrospective clinical study. *Clin Endocrinol (Oxf)*, 2018; **88**: 711–18.
33. Corona G, Pizzocaro A, Lanfranco F, *et al*. Sperm recovery and ICSI outcomes in Klinefelter syndrome: a systematic review and meta-analysis. *Hum Reprod Update*, 2017; **23**: 265–75.
34. Vloeberghs V, Verheyen G, Santos-Ribeiro S, *et al*. Is genetic fatherhood within reach for all azoospermic Klinefelter men? *PLoS One*, 2018; **13**: e0200300.
35. Chang S, Skakkebæk A, Trolle C, *et al*. Anthropometry in Klinefelter syndrome—multifactorial influences due to CAG length, testosterone treatment and possibly intrauterine hypogonadism. *J Clin Endocrinol Metab*, 2015; **100**: E508–17.
36. Zitzmann M, Depenbusch M, Gromoll J, Nieschlag E. X-chromosome inactivation patterns and androgen receptor functionality influence phenotype and social characteristics as well as pharmacogenetics of testosterone therapy in Klinefelter patients. *J Clin Endocrinol Metab*, 2004; **89**: 6208–17.
37. Aksglaede L, Andersson AM, Jorgensen N, *et al*. Primary testicular failure in Klinefelter's syndrome: the use of bivariate luteinizing hormone-testosterone reference charts. *Clin Endocrinol (Oxf)*, 2007; **66**: 276–81.
38. Schiavi RC, Theilgaard A, Owen DR, White D. Sex chromosome anomalies, hormones, and sexuality. *Arch Gen Psychiatry*, 1988; **45**: 19–24.
39. Skakkebaek A, Moore PJ, Pedersen AD, *et al*. The role of genes, intelligence, personality, and social engagement in cognitive performance in Klinefelter syndrome. *Brain Behav*, 2017; **7**: e00645.
40. Vorona E, Zitzmann M, Gromoll J, Schuring AN, Nieschlag E. Clinical, endocrinological, and epigenetic features of the 46,XX male syndrome, compared with 47,XXY Klinefelter patients. *J Clin Endocrinol Metab*, 2007; **92**: 3458–65.
41. Bojesen A, Juul S, Birkebaek NH, Gravholt CH. Morbidity in Klinefelter syndrome: a Danish register study based on hospital discharge diagnoses. *J Clin Endocrinol Metab*, 2006; **91**: 1254–60.
42. Bojesen A, Juul S, Birkebaek N, Gravholt CH. Increased mortality in Klinefelter syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 3830–4.
43. Bojesen A, Kristensen K, Birkebaek NH, *et al*. The metabolic syndrome is frequent in Klinefelter's syndrome and is associated with abdominal obesity and hypogonadism. *Diabetes Care*, 2006; **29**: 1591–8.
44. Ishikawa T, Yamaguchi K, Kondo Y, Takenaka A, Fujisawa M. Metabolic syndrome in men with Klinefelter's syndrome. *Urology*, 2008; **71**: 1109–13.
45. Bojesen A, Birkebæk N, Kristensen K, *et al*. Bone mineral density in Klinefelter syndrome is reduced and primarily determined by muscle strength and resorptive markers, but not directly by testosterone. *Osteoporos Int*, 2011; **22**: 1441–50.
46. Bojesen A, Juul S, Birkebaek N, Gravholt CH. Increased mortality in Klinefelter syndrome. *J Clin Endocrinol Metab*, 2004; **89**: 3830–4.
47. Campbell WA, Price WH. Venous thromboembolic disease in Klinefelter's syndrome. *Clin Genet*, 1981; **19**: 275–80.
48. Zoller B, Ji Jii, Sundquist J, Sundquist K. High risk of venous thromboembolism in Klinefelter syndrome. *J Am Heart Assoc*, 2016; **5**: e003567.
49. Zollner TM, Veraart JC, Wolter M, *et al*. Leg ulcers in Klinefelter's syndrome--further evidence for an involvement of plasminogen activator inhibitor-1. *Br J Dermatol*, 1997; **136**: 341–4.
50. Khosla S, Melton LJ 3rd, Atkinson EJ, O'Fallon WM, Klee GG, Riggs BL. Relationship of serum sex steroid levels and bone turnover markers with bone mineral density in men and women: a key role for bioavailable estrogen. *J Clin Endocrinol Metab*, 1998; **83**: 2266–74.
51. Devogelaer JP, De Cooman S, Nagant D. Low bone mass in hypogonadal males; effect of testosterone substitution therapy, a densitometric study. *Maturitas*, 1992; **15**: 17.
52. Behre HM, Kliesch S, Leifke E, Link TM, Nieschlag E. Long-term effect of testosterone therapy on bone mineral density in hypogonadal men. *J Clin Endocrinol Metab*, 1997; **82**: 2386–90.
53. van den Bergh JP, Hermus AR, Spruyt AI, Sweep CG, Corstens FH, Smals AG. Bone mineral density and quantitative ultrasound parameters in patients with Klinefelter's syndrome after long-term testosterone substitution. *Osteoporos Int*, 2001; **12**: 55–62.
54. Bojesen A, Birkebæk N, Kristensen K, *et al*. Bone mineral density in Klinefelter syndrome is reduced and primarily determined by muscle strength and resorptive markers, but not directly by testosterone. *Osteoporos Int*, 2010; **22**: 1441–50.
55. Bojesen A, Juul S, Birkebaek NH, Gravholt CH. Morbidity in Klinefelter syndrome: a Danish register study based on hospital discharge diagnoses. *J Clin Endocrinol Metab*, 2006; **91**: 1254.
56. Hultborn R, Hanson C, Köpf I, Verbiené I, Warnhammar E, Weimarck A. Prevalence of Klinefelter's syndrome in male breast cancer patients. *Anticancer Res*, 1997; **17**: 4293–7.
57. Swerdlow AJ, Schoemaker MJ, Higgins CD, Wright AF, Jacobs PA. Cancer incidence and mortality in men with Klinefelter syndrome: a cohort study. *J Natl Cancer Inst*, 2005; **97**: 1204–10.

58. Hasle H, Mellemgaard A, Nielsen J, Hansen J. Cancer incidence in men with Klinefelter syndrome. *Br J Cancer*, 1995; **71**: 416–20.
59. Williams LA, Pankratz N, Lane J, *et al*. Klinefelter syndrome in males with germ cell tumors: a report from the Children's Oncology Group. *Cancer*, 2018; **124**: 3900–8.
60. Nielsen J, Johnsen SG, Sorensen K. Follow-up 10 years later of 34 Klinefelter males with karyotype 47,XXY and 16 hypogonadal males with karyotype 46,XY. *Psychol Med*, 1980; **10**: 345–52.
61. Skakkebaek A, Wallentin M, Gravholt CH. Neuropsychology and socioeconomic aspects of Klinefelter syndrome: new developments. *Curr Opin Endocrinol Diabetes Obes*, 2015; **22**: 209–16.
62. Stewart DA, Netley CT, Park E. Summary of clinical findings of children with 47,XXY, 47,XYY, and 47,XXX karyotypes. *Birth Defects Orig Artic Ser*, 1982; **18**: 1–5.
63. Rovet J, Netley C, Keenan M, Bailey J, Stewart D. The psychoeducational profile of boys with Klinefelter syndrome. *J Learn Disabil*, 1996; **29**: 180–96.
64. Boada R, Janusz J, Hutaff-Lee C, Tartaglia N. The cognitive phenotype in Klinefelter syndrome: a review of the literature including genetic and hormonal factors. *Dev Disabil Res Rev*, 2009; **15**: 284–94.
65. Geschwind DH, Boone KB, Miller BL, Swerdloff RS. Neurobehavioral phenotype of Klinefelter syndrome. *Ment Retard Dev Disabil Res Rev*, 2000; **6**: 107–16.
66. Bender BG, Linden MG, Harmon RJ. Neuropsychological and functional cognitive skills of 35 unselected adults with sex chromosome abnormalities. *Am J Med Genet*, 2001; **102**: 309–13.
67. Bryant DM, Hoeft F, Lai S, *et al*. Neuroanatomical phenotype of Klinefelter syndrome in childhood: a voxel-based morphometry study. *J Neurosci*, 2011; **31**: 6654–60.
68. Goddard MN, van RS, Rombouts SA, Swaab H. White matter microstructure in a genetically defined group at increased risk of autism symptoms, and a comparison with idiopathic autism: an exploratory study. *Brain Imaging Behav*, 2016; **10**: 1280–8.
69. Lentini E, Kasahara M, Arver S, Savic I. Sex differences in the human brain and the impact of sex chromosomes and sex hormones. *Cereb Cortex*, 2013; **23**: 2322–36.
70. Skakkebaek A, Gravholt CH, Rasmussen PM, *et al*. Neuroanatomical correlates of Klinefelter syndrome studied in relation to the neuropsychological profile. *NeuroImage Clin*, 2013; **4**: 1–9.
71. Skakkebaek A, Bojesen A, Kristensen MK, *et al*. Klinefelter syndrome has increased brain responses to auditory stimuli and motor output, but not to visual stimuli or Stroop adaptation. *Andrology*, 2014; **2**: 632–40.
72. Wallentin M, Skakkebæk A, Bojesen A, *et al*. Klinefelter syndrome has increased brain responses to auditory stimuli and motor output, but not to visual stimuli or Stroop adaptation. *Neuroimage Clin*, 2016; **11**: 239–51.
73. Forssman H, Hambert G. Incidence of Klinefelter's syndrome among mental patients. *Lancet*, 1963; **1**: 1327.
74. Delisi LE, Friedrich U, Wahlstrom J, *et al*. Schizophrenia and sex chromosome anomalies. *Schizophr Bull*, 1994; **20**: 495.
75. van Rijn S, Aleman A, Swaab H, Kahn R. Klinefelter's syndrome (karyotype 47,XXY) and schizophrenia-spectrum pathology. *Br J Psychiatry*, 2006; **189**: 459–60.
76. Ross NL, Wadekar R, Lopes A, *et al*. Methylation of two Homo sapiens-specific X-Y homologous genes in Klinefelter's syndrome (XXY). *Am J Med Genet B Neuropsychiatr Genet*, 2006; **141**: 544.
77. Gotz MJ, Johnstone EC, Ratcliffe SG. Criminality and antisocial behaviour in unselected men with sex chromosome abnormalities. *Psychol Med*, 1999; **29**: 953.–62
78. Nielsen J, Pelsen B. Follow-up 20 years later of 34 Klinefelter males with karyotype 47,XXY and 16 hypogonadal males with karyotype 46,XY. *Hum Genet*, 1987; **77**: 188–92.
79. Simm PJ, Zacharin MR. The psychosocial impact of Klinefelter syndrome--a 10 year review. *J Pediatr Endocrinol Metab*, 2006; **19**: 499–505.
80. Stochholm K, Bojesen A, Jensen AS, Juul S, Gravholt CH. Criminality in men with Klinefelter's syndrome and XYY syndrome: a cohort study. *BMJ Open*, 2012; **2**: e000650.
81. Ratcliffe S. Long term outcome in children of sex chromosome abnormalities. *Arch Dis Child*, 1999; **80**: 192–5.
82. Rogol AD, Tartaglia N. Considerations for androgen therapy in children and adolescents with Klinefelter syndrome (47, XXY). *Pediatr Endocrinol Rev*, 2010; **8** Suppl 1: 145–50.
83. Nielsen J, Pelsen B, Sorensen K. Follow-up of 30 Klinefelter males treated with testosterone. *Clin Genet*, 1988; **33**: 262–9.
84. Simm PJ, Zacharin MR. The psychosocial impact of Klinefelter syndrome--a 10 year review. *J Pediatr Endocrinol Metab*, 2006; **19**: 499–505.
85. Wang C, Swerdloff RS, Iranmanesh A, *et al*., Transdermal testosterone gel improves sexual function, mood, muscle strength, and body composition parameters in hypogonadal men. *J Clin Endocrinol Metab*, 2000; **85**: 2839–53.
86. Cherrier MM, Asthana S, Plymate S, *et al*. Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*, 2001; **57**: 80–8.
87. Shanbhogue VV, Hansen S, Jorgensen NR, Brixen K, Gravholt CH. Bone geometry, volumetric density, microarchitecture, and estimated bone strength assessed by HR-pQCT in Klinefelter syndrome. *J Bone Miner Res*, 2014; **29**: 2474–82.
88. Corona G, Petrone L, Paggi F, *et al*. Sexual dysfunction in subjects with Klinefelter's syndrome. *Int J Androl*, 2010; **33**: 574–80.
89. Skakkebaek A, Moore PJ, Chang S, Fedder J, Gravholt CH. Quality of life in men with Klinefelter syndrome: the impact of genotype, health, socioeconomics, and sexual function. *Genet Med*, 2018; **20**: 214–22.
90. Chang S, Christiansen CF, Bojesen A, Juul S, Münster A-MB, Gravholt CH. Klinefelter syndrome and testosterone treatment: a national cohort study on thrombosis risk. *Endocr Connect*. 2020; **9**(1): 34–43.
91. Schiff JD, Palermo GD, Veeck LL, Goldstein M, Rosenwaks Z, Schlegel PN. Success of testicular sperm extraction [corrected] and intracytoplasmic sperm injection in men with Klinefelter syndrome. *J Clin Endocrinol Metab*, 2005; **90**: 6263–7.
92. Smyth CM, Bremner WJ. Klinefelter syndrome. *Arch Intern Med*, 1998; **158**: 1309–14.
93. Aksglaede L, Skakkebaek NE, Almstrup K, Juul A. Clinical and biological parameters in 166 boys, adolescents and adults with nonmosaic Klinefelter syndrome: a Copenhagen experience. *Acta Paediatr*, 2011; **100**: 793–806.
94. van den Bergh JP, Hermus AR, Spruyt AI, Sweep CG, Corstens FH, Smals AG. Bone mineral density and quantitative ultrasound parameters in patients with Klinefelter's syndrome after long-term testosterone substitution. *Osteoporos Int*, 2001; **12**: 55–62.
95. Fricke GR, Mattern HJ, Schweikert HU, Schwanitz G. Klinefelter's syndrome and mitral valve prolapse. an echocardiographic study in twenty-two patients. *Biomed Pharmacother*, 1984; **38**: 88–97.

96. Andersen NH, Bojesen A, Kristensen K, *et al*. Morbidity and mortality in Klinefelter syndrome (47,XXY). *Clin Endocrinol (Oxf)*, 2008; **69**: 785–91.
97. Swerdlow AJ, Higgins CD, Schoemaker MJ, Wright AF, Jacobs PA. Mortality in patients with Klinefelter syndrome in Britain: a cohort study. *J Clin Endocrinol Metab*, 2005; **90**: 6516.
98. Seminog OO, Seminog AB, Yeates D, Goldacre MJ. Associations between Klinefelter's syndrome and autoimmune diseases: English national record linkage studies. *Autoimmunity*, 2015; **48**: 125–8.
99. Panimolle F, Tiberti C, Granato S, *et al*. Screening of endocrine organ-specific humoral autoimmunity in 47,XXY Klinefelter's syndrome reveals a significant increase in diabetes-specific immunoreactivity in comparison with healthy control men. *Endocrine*, 2016; **52**: 157.
100. Sawalha AH, Harley JB, Scofield RH. Autoimmunity and Klinefelter's syndrome: when men have two X chromosomes. *J Autoimmun*, 2009; **33**: 31–4.
101. Harlow TL, Gonzalez-Alegre P. High prevalence of reported tremor in Klinefelter syndrome. *Parkinsonism Relat Disord*, 2009; **15**: 393–5.
102. Koegl-Wallner M, Katschnig-Winter P, Pendl T, *et al*. Tremor associated with Klinefelter syndrome—a case series and review of the literature. *Parkinsonism Relat Disord*, 2014; **20**: 323–7.
103. Swerdlow AJ, Schoemaker MJ, Higgins CD, Wright AF, Jacobs PA. Cancer incidence and mortality in men with Klinefelter syndrome: a cohort study. *J Natl Cancer Inst*, 2005; **97**: 1204–10.
104. Hultborn R, Hanson C, Köpf I, Verbiené I, Warnhammar E, Weimarck A. Prevalence of Klinefelter's syndrome in male breast cancer patients. *Anticancer Res*, 1997; **17**: 4293–7.
105. Ji Jii, Zoller B, Sundquist J, Sundquist K. Survival of cancer in patients with rheumatoid arthritis: a follow-up study in Sweden of patients hospitalized with rheumatoid arthritis 1 year before diagnosis of cancer. *Int J Cancer*, 2016; **139**: 754.
106. Lahlou N, Fennoy I, Ross JL, Bouvattier C, Roger M. Clinical and hormonal status of infants with nonmosaic XXY karyotype. *Acta Paediatr*, 2011; **100**: 824–9.
107. Zinn AR, Ramos P, Elder FF, *et al*. Androgen receptor CAGn repeat length influences phenotype of 47,XXY (Klinefelter) syndrome. *J Clin Endocrinol Metab*, 2005; **90**: 5041–6.
108. Hasle H, Mellemgaard A, Nielsen J, Hansen J. Cancer incidence in men with Klinefelter syndrome. *Br J Cancer*, 1995; **71**: 416–20.
109. Salbenblatt JA, Bender BG, Puck MH, *et al*. Pituitary-gonadal function in Klinefelter syndrome before and during puberty. *Pediatr Res*, 1985; **19**: 82–6.
110. van Rijn S, Aleman A, Swaab H, Kahn R. Klinefelter's syndrome (karyotype 47, XXY) and schizophrenia-spectrum pathology. *Br J Psychiatry*, 2006; **189**: 459–61.
111. Stewart DA, Netley CT, Park E. Summary of clinical findings of children with 47,XXY, 47,XYY, and 47,XXX karyotypes. *Birth Defects Orig Artic Ser*, 1982; **18**: 1–5.
112. van Rijn S, Swaab H, Aleman A, Kahn RS. Social behavior and autism traits in a sex chromosomal disorder: Klinefelter (47XXY) syndrome. *J Autism Dev Disord*, 2008; **38**: 1634–41.
113. Ross JL, Roeltgen DP, Kushner H, *et al*. Behavioral and social phenotypes in boys with 47,XYY syndrome or 47,XXY Klinefelter syndrome. *Pediatrics*, 2012; **129**: 769–78.
114. Fjermestad KW, Stokke S. Sleep problems and life satisfaction as predictors of health in men with sex chromosome aneuploidies. *Behav Med*, 2017; **44**: 116–22.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

10.4

Male Adult Hypogonadism

10.4.1 Aetiology

Alvin M. Matsumoto and Radhika Narla

Introduction

Once the diagnosis of hypogonadism is suspected in a man with signs and symptoms of androgen deficiency and confirmed with consistently low early morning total and/or free testosterone concentrations drawn in the fasted state on at least two separate occasions, the aetiology of hypogonadism should be determined [1].

Measurements of circulating gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are usually performed in same blood sample as a confirmatory testosterone, to ascertain whether low testosterone is due a primary disorder of the testes or secondary to a defect in the hypothalamus and/or pituitary gland. Subsequently, a working diagnosis of the aetiology of the hypogonadism should be based on the patient's past medical history, clinical presentation, and further laboratory testing.

Aetiologies of hypogonadism can be classified into two broad categories: **primary hypogonadism** and **secondary hypogonadism**. Primary hypogonadism is usually caused by an intrinsic structural or congenital defect within the testes, resulting in impaired testosterone and sperm production and elevated LH and FSH concentrations (as a result of reduced negative feedback suppression; FSH concentrations are usually more elevated than LH concentrations because active cell-division during spermatogenesis is more sensitive to damage than testosterone production) [1, 2].

Secondary hypogonadism is caused by disease or dysfunction leading to inadequate hypothalamic gonadotropin-releasing hormone and/or pituitary gonadotropin secretion and stimulation of the testes [3], resulting in reduced testosterone and spermatogenesis, and low or inappropriately normal LH and FSH concentrations Key differences between the categories are highlighted in Table 10.4.1.1.

In addition to identifying the level of the HPT axis deficit, it is important to determine whether the cause of hypogonadism is organic or functional [1, 4]. Table 10.4.1.2 summarizes all the categories of hypogonadism. Organic hypogonadism, or classical hypogonadism, is caused by a congenital, structural, or infiltrative disorder that results in permanent hypothalamic, pituitary, or testicular damage. In contrast, functional hypogonadism is caused by conditions that suppress gonadotropin and testosterone concentrations that are potentially reversible with treatment of the underlying cause(s) (e.g. with weight loss or discontinuation of opioid medications) [4]. Finally, some aetiologies of hypogonadism are caused by defects both in the testes and hypothalamus and/or pituitary (i.e. combined primary and secondary hypogonadism).

Prior to initiating testosterone treatment, it is important determine whether a patient has secondary hypogonadism as some hypothalamic–pituitary causes require specific management besides testosterone therapy. In contrast to men with primary hypogonadism

Table 10.4.1.1 Differences between primary and secondary hypogonadism

	Primary hypogonadism	Secondary hypogonadism
Condition	Testicular failure	Hypothalamus–pituitary failure
Gonadotropin levels	Elevated	Low or normal
Frequency observed	Less common	More common
Spermatogenesis	Usually impaired	Impaired if severe
Organic or functional aetiology	Usually organic (structural or intrinsic testicular defect), permanent	Usually functional (extrinsic gonadotropin suppression of a structurally intact hypothalamus and pituitary), potentially reversible

Table 10.4.1.2 Aetiology of Male Hypogonadism

Primary hypogonadism	Secondary hypogonadism
Organic	
Congenital	
Klinefelter syndrome and variants	Congenital hypogonadotropic hypogonadism
Cryptorchidism (uncorrected bilateral)	Complex genetic syndromes*
Congenital anorchia	Inactivating LH beta mutations
Myotonic dystrophy	
Noonan syndrome	
Testicular adrenal rest tumours (congenital adrenal hyperplasia)	
Testosterone biosynthetic defects	
Complex genetic syndromes	
Inactivating mutations of LH receptor	
Acquired	
Bilateral orchidectomy testicular damage, torsion, radiation	Hypothalamic/pituitary or parasellar tumours
Cancer chemotherapy (alkylating agents)	Hypothalamic/pituitary radiation
Orchitis (mumps)	Pituitary apoplexy
Autoimmune testicular failure	Severe head trauma, pituitary stalk injury/section
Advanced older age*	Infiltrative disorders
Hodgkin's disease*	Hypothalamic/pituitary infection (tuberculosis, fungal)
Testicular cancer*	Immune check point inhibitors, hypophysitis, hypophysectomy
	Hemochromatosis*—that of primary hypogonadism, transfusion-related iron overload
	Sickle cell disease*
Functional	
Hepatic cirrhosis*	Constitutional delayed puberty
Renal failure (CKD 4–5)*	Diabetes mellitus
	Obesity
	Obstructive sleep apnoea
	Opioid use
	Hyperprolactinemia
	Malnutrition
	Excessive exercise (energy expenditure > intake)
	Glucocorticoid excess (Cushing syndrome)*
	Sex steroid use (anabolic steroids, progestins, oestrogens)*
	Chronic systemic illness, organ failure (liver, lung, heart)*
	Alcohol abuse*
	Severe acute illness or surgery*

* Combined primary and secondary hypogonadism.

who have permanent impairment of spermatogenesis, men with secondary hypogonadism can respond to gonadotropin therapy to stimulate testosterone and sperm production, and treat the infertility (see Chapter 10.5, 'Management of Idiopathic Male Infertility'). Also, although uncommon, secondary hypogonadism from a pituitary or hypothalamic tumour can be associated with deficiency and/or excessive secretion of other anterior pituitary hormones or pituitary mass effects (e.g. visual field defects or headaches) that require management in addition to testosterone treatment [1–3]. Finally, the cause of secondary hypogonadism is often functional and potentially reversible so that testosterone treatment may not always be required (e.g. weight loss with diet or bariatric surgery can improve testosterone concentrations in obesity-related hypogonadism) [1, 4].

Causes of Primary Hypogonadism

Organic Congenital Aetiologies

Klinefelter Syndrome

Klinefelter syndrome (KS) results from one or more extra X chromosomes, most commonly by a 47,XXY karyotype; or mosaicism in most of the remainder (a 47,XXY karyotype present in some tissues or cells and a normal 46,XY karyotype in others). KS occurs in approximately 150 per 100 000 males and is the most common cause of primary hypogonadism [5]. Men with KS have classical phenotypic characteristics: very small firm testes (less than 2.5 cm or 6 ml), varying degrees of testosterone deficiency, gynaecomastia, long legs relative to arms, azoospermia, and infertility. Some men may have milder signs and symptoms. A karyotype confirms the diagnosis [1–3, 5, 6] (see Chapter 10.3, 'Klinefelter's Syndrome').

Men with classical KS have azoospermia, but both mosaic and classical KS men been shown to be capable of conceiving children through testicular sperm extraction. KS men also have a higher incidence of male breast cancer, extragonadal germ cell tumours, and non-Hodgkin's lymphoma, decreased bone mass, higher degree of insulin resistance, and autoimmune disorders [2, 5, 6].

Cryptorchidism

Cryptorchidism is failure of one or both testes to descend through the inguinal canal into the scrotum. Most commonly, it is associated with selective impairment of sperm production and increased FSH with normal testosterone production [1–3]. However, rarely (e.g. in men with uncorrected bilateral cryptorchidism), both testosterone and sperm production are impaired with elevated FSH and LH concentrations, consistent with primary hypogonadism. Bilateral cryptorchidism may be present in other causes of both primary hypogonadism (e.g. Klinefelter syndrome) and secondary hypogonadism (e.g. Kallmann syndrome) or other complex congenital genetic syndromes (e.g. Prader–Willi syndrome). Patients with cryptorchidism have an approximately 5–10-fold increased testicular cancer risk, especially within an undescended testes; increased risk persists despite surgical relocation of the testis in the scrotum (orchiopexy) [2].

Congenital Anorchia

Congenital anorchia, also known as vanishing testes syndrome, is a rare disorder characterized by the absence of one or both testes in

a phenotypic male (suggesting that functioning testes were present during fetal genital development). It is believed to be caused by testicular torsion during fetal life after sufficient testosterone exposure to produce masculinization of the reproductive tract. Males with congenital anorchia present with primary hypogonadism, delayed puberty, and on physical exam, testes are not palpable. In contrast to bilateral cryptorchidism (discussed earlier), in congenital anorchia, testosterone concentrations do not increase with hCG stimulation and anti-Müllerian hormone concentrations are very low or undetectable. A response to hCG stimulation would raise the possibility of intra-abdominal testes and cryptorchidism [2].

Myotonic Dystrophy

Myotonic dystrophy is an autosomal dominant disorder that presents with progressive muscle weakness and wasting, myotonia (involuntary sustained muscle contraction), male pattern balding, and other congenital anomalies. There are two types of myotonic dystrophy, type 1, and type 2. Type 1 myotonic dystrophy is more prevalent and results from an expansion of trinucleotide cardiotocography (CTG) repeats in the *DMPK* gene. Type 2 myotonic dystrophy is a neurologically milder form caused by expansion of CTG repeats in *CNBP* gene. In both types, the pathogenesis is a disruption of the normal mRNA metabolism and this affects many genes and causes multisystem effects including hypogonadism. While the exact mechanism is unknown, testicular atrophy and primary hypogonadism is observed in both types [2, 3].

Noonan Syndrome

Noonan syndrome, previously referred to as the male Turner syndrome, is an autosomal dominant or occasionally sporadic genetic disorder that is caused by mutations of genes that are part of the Ras-MAPK signalling pathway, such as PTPN11, SOS1, BRAF, and others. While the exact mechanism is unknown, alterations in encoded proteins leads to multisystem dysregulation including hypogonadism. The clinical features of Noonan syndrome include short stature, certain facial features (low set ears, downward slanting eyes, strabismus, high arched palate), short web neck, skeletal abnormalities (shield-like chest, pectus excavatum or carinatum, cubitis valgus, and joint laxity), intellectual disability, congenital heart defects (pulmonic stenosis and hypertrophic cardiomyopathy), and primary hypogonadism [2].

Congenital Adrenal Hyperplasia (CAH)

Men with CAH have mutations in genes coding for enzymes essential for steroid biosynthesis. The most common form of CAH is autosomal recessive 21α-hydroxylase deficiency caused by a mutation of the *CYP21* gene leading to low cortisol and aldosterone concentrations that increase adrenocorticotrophic hormone (ACTH) concentrations (due to reduced cortisol negative feedback) and cause adrenal hyperplasia with accumulation of proximal steroid hormone precursors (17-hydroxyprogesterone and progesterone, and adrenal androgens (androstenedione and DHEA) [2, 7]. Men with CAH may develop testicular adrenal rest tumours (TARTs) that can cause compression of surrounding testicular tissue and primary hypogonadism. TARTs are palpable on examination or identified by ultrasound and decrease in size with adequate glucocorticoid treatment to suppress ACTH stimulation. Also, if glucocorticoid suppression is inadequate, men with CAH can then present with secondary hypogonadism. Due to oversecretion of adrenal androgens, these men maintain virilization but increased progesterone as well as testosterone and oestradiol concentrations from conversion of adrenal androgen precursors can suppress LH and FSH concentrations leading to impaired spermatogenesis and secondary hypogonadism [2, 3, 7].

Testosterone Biosynthetic Defects

Rare testosterone biosynthetic enzyme defects, including 17α-hydroxylase/17,20-lyase, 17β-hydroxysteroid dehydrogenase type 3/17-ketoreductase and 3β-hydroxysteroid dehydrogenase type 2 deficiencies caused by mutations in the *CYP17, HSD17B3*, and *HSD3B2* genes, respectively, usually present at birth with a partial virilized near-female phenotype, or ambiguous genitalia, i.e. 46,XY disorder of sex development (DSD). However, males with incomplete deficiencies in these biosynthetic enzymes may present as a near-male phenotype with delayed puberty, hypospadias, gynaecomastia, and primary hypogonadism [7].

Complex Genetic Syndromes

Numerous complex genetic syndromes associated with distinct morphological developmental manifestations and congenital anomalies, may also present with primary hypogonadism.

While the mechanism is unknown, Prader–Willi can present with either primary or secondary hypogonadism and can have either unilateral or bilateral cryptorchidism and obesity that may contribute to hypogonadism [8]. Men with Down's syndrome (trisomy 21) usually present with isolated impairment of sperm production with normal or selective high FSH, but occasionally, they may present with elevated FSH and LH concentrations [2].

Luteinizing Hormone Receptor Mutations

Patients with inactivating mutations of LH receptors resulting in Leydig cell aplasia or hypoplasia usually present with a near-female phenotype or ambiguous genitalia and cryptorchidism (46,XY DSD). Men with milder forms have micropenis, hypospadias, undervirilization, delayed puberty, or primary hypogonadism.

Organic Acquired Aetiologies

Bilateral Orchiectomy, Testicular Damage, or Torsion and Radiation

Bilateral orchiectomy results in rapid and severe primary hypogonadism often associated with hot flushes [1–3]. Direct blunt trauma or torsion to one or both testicles may be associated with vascular compromise that can lead to development of testicular atrophy. Exposure of the testes to the large amounts of ionizing radiation (>600–800 cGy) that is used in treatment for certain types of leukaemia can also result in testicular damage (especially spermatogenesis) and primary hypogonadism [9].

Cancer Chemotherapy

Combination cancer chemotherapy regimens that contain alkylating agents (e.g. cyclophosphamide) used in the treatment of Hodgkin's or non-Hodgkin's lymphoma; cyclophosphamide therapy for systemic rheumatic disorders; or platinum drugs (e.g. cisplatin) used in the treatment of testicular cancer can cause testicular atrophy, impaired spermatogenesis, and testosterone production with elevated gonadotropin concentrations [10].

Orchitis

Viral infections of the testes, such as mumps orchitis, most commonly result in testicular atrophy and isolated impairment of spermatogenesis and elevated FSH concentrations, but if severe, can cause primary hypogonadism. Orchitis is the most common complication of mumps in pubertal boys and men. Testis pain and swelling usually occurs 1–2 weeks after the initial painful swelling of the parotid glands; however, orchitis is subclinical in half the cases [1–3]. HIV infection, and/or associated opportunistic infections may involve the testes directly leading to primary hypogonadism [2, 3].

Autoimmune Polyglandular Syndrome

Autoimmune polyglandular syndrome type 1 is a rare autosomal recessive disorder that is caused by a mutation in the autoimmune regulator (*AIRE*) gene, characterized by autoimmune hypoparathyroidism, primary adrenal insufficiency, mucocutaneous candidiasis and, rarely, primary hypogonadism. Autoimmune polyglandular syndrome type 2 is a polygenic disorder associated with HLA-DR3 and DR4 that is characterized by autoimmune primary adrenal insufficiency, thyroid disease (Hashimoto's or Graves' disease), type 1 diabetes mellitus and rarely, primary hypogonadism associated with steroid cell autoantibodies (SCA) [11].

Advanced Older Age

In men, there is a gradual decline in testosterone concentrations with ageing, at a rate of approximately 1% per year after age 30 years [1, 2, 4, 12]. Sex-hormone-binding globulin (SHBG) concentrations increase with ageing, resulting in a greater decrease in free testosterone concentrations and an increasing prevalence of testosterone concentrations in the hypogonadal range in older men [4, 12]. However, the prevalence of clinical hypogonadism (low testosterone concentrations with clinical manifestations of androgen deficiency) is lower, 2% (with only sexual symptoms) to 9% [4].

With ageing, there is also a progressive increase in LH and FSH concentrations. However, in large part because of age-related comorbidities, gonadotropin concentrations do not increase to above normal until advanced age (70 years or older). Therefore, healthy advanced age men usually manifest primary hypogonadism, while older men with poorer health status and comorbid illness(es) have hormone pattern of secondary hypogonadism. There is evidence of both testicular failure (reduced testosterone response to hCG stimulation and histologically reduced Leydig cell numbers) and hypothalamic dysfunction (disordered LH pulses) with ageing [12].

Recently, The Testosterone Trials demonstrated that short-term testosterone treatment in older men with symptomatic hypogonadism (but without recognizable organic HPT axis pathology) improved sexual function (sexual activity, libido, and erectile dysfunction), anaemia, bone density; slightly improved walking distance, mood, and depressive symptoms; but did not improve cognitive function and vitality. However, testosterone did increase non-calcified coronary artery plaque volume with no differences in cardiovascular adverse events [1]. The significance of this last finding is not known and there is a need for longer term, larger randomized trials. Meanwhile, in older men, it is important to consider functional causes of low testosterone concentrations that are potentially reversible during work-up for hypogonadism aetiology before considering starting testosterone [1, 4, 12].

Cancer

Advanced Hodgkin's disease or testicular cancer can also present with primary hypogonadism before chemotherapy or radiation therapy; the cause is unknown. While the pattern is that of primary hypogonadism, secondary hypogonadism can also be observed [2, 3].

Functional Acquired Aetiologies

Hepatic Cirrhosis and Chronic Kidney Disease

In alcoholic cirrhosis, direct toxic effects of alcohol may lead to testicular atrophy and elevated FSH and LH concentrations; however, gonadotropin suppression can also be observed. Patients with cirrhosis are often treated with spironolactone that acts primarily as an androgen receptor antagonist, but also may reduce testosterone production. Furthermore, oestradiol and estrone concentrations may be elevated due to reduced hepatic clearance and contribute to secondary hypogonadism by suppressing LH and FSH concentrations. In severe cirrhosis and liver failure, reduced hepatic protein production, protein-calorie malnutrition, and complicating infections may also lead to gonadotropin suppression. Hepatic cirrhosis is considered a functional cause of hypogonadism because gonadal function may recover after liver transplant, at least partially, depending on the chronic immunosuppression regimen used [2, 3].

Chronic kidney disease may also produce both primary and secondary hypogonadism. However, in late stage chronic kidney disease (CKD), the LH, and FSH concentrations are high in large part, due to reduced renal clearance of gonadotropins [1, 2]. Primary testicular failure in CKD as evidenced by reduced Leydig cell testosterone response to hCG is due to numerous factors, including direct toxic effects of uraemia, systemic inflammation, and chronic nutritional deficiency. These factors, in addition to hyperprolactinaemia (from reduced prolactin clearance), contribute to alterations in pulsatile gonadotropin secretion. Renal transplantation usually reverses hypogonadism, depending on the presence of comorbid conditions and use of certain medications such as glucocorticoid immunosuppression [1–3]. See Chapter 10.6 'Hypothalamo–Pituitary–Testicular Axis Function in Systemic Diseases and Effects of Medications'.

Causes of Secondary Hypogonadism

Organic Congenital Aetiologies

Congenital hypogonadotropic hypogonadism (CHH) is a relatively rare condition that causes severe secondary hypogonadism with estimated prevalence 1/4000 to 1/10 000 males. Males with CHH present with delayed puberty or incomplete sexual maturation (eunuchoidism), severe androgen deficiency with prepubertal testosterone concentrations, impairment of sperm production, and varying but usually severe degrees of gonadotropin deficiency.

Gonadotropin deficiency is caused by gonadotropin-releasing hormone (GnRH) deficiency (as evidenced most commonly by absent or less commonly by deficient pulsatile LH secretion) or rarely, GnRH receptor defects. CHH can be divided into three subgroups: Kallmann syndrome, normosmic CHH, and CHH associated with complex genetic syndromes forms [13].

Approximately 60% of CHH men have anosmia or hyposmia due to defective development of the olfactory bulbs known as Kallmann syndrome (e.g., due to KAL1, FGFR1/FGF8, PROKR2/PROK2 mutations). Kallmann syndrome patients may have other developmental midline defects, such as cleft lip or palate, or other craniofacial defects, sensorineural hearing loss, digital skeletal anomalies, unilateral renal agenesis, and neurological defects such synkinesia or mirror movements [14].

Individuals with normosmic CHH have normal olfaction and are difficult to differentiate from those with constitutional delay of puberty (CDP) which is much more common. Presence of micropenis, cryptorchidism, careful history, and signs from family members would make the clinician favour normosmic CHH rather than CDP. Certain gene mutations that typically cause isolated CHH include KISS1R (kisspeptin 1 receptor gene), KISS1 ligand, and GnRH receptor. When the diagnosis of CHH is suspected, it is recommended to refer to a clinical geneticist for further evaluation [2, 14, 15].

CHH can also occur in a number of complex genetic syndromes that are characterized by a constellation of developmental and/or hormonal abnormalities. The CHARGE (**C**oloboma or central nervous system (CNS) anomalies, **H**eart anomalies, choanal **A**tresia, growth **R**etardation, **G**enital defect (cryptorchidism and hypospadias), and **E**ar anomalies) that is caused by a mutation in the *CHD7* gene may present with either Kallmann syndrome or normosmic CHH [15]. CHH is part of numerous congenital dysmorphology syndromes, many of which are associated with neurological abnormalities and obesity that may contribute to secondary hypogonadism, such as Prader–Willi syndrome that can manifest either with primary or secondary hypogonadism [2].

Although rare, inactivating mutations of the LHβ subunit gene result in prepubertal secondary hypogonadism manifested by failure to undergo puberty with severe androgen deficiency [16].

Organic Acquired Aetiologies

Hypothalamic, Pituitary, or Parasellar Tumours and Radiation

Tumours of the hypothalamus, pituitary, or parasellar area including pituitary macroadenomas, pituitary cysts, meningiomas, craniopharyngiomas can decrease GnRH and/or LH and FSH secretion. Hypopituitarism following treatment for these or other tumours with surgery or radiation therapy can also result in secondary hypogonadism. Although secondary hypogonadism associated with some pituitary macroadenomas may be reversed after surgical resection, more commonly, gonadotropin deficiency is irreversible. In addition, microprolactinomas can suppress GnRH secretion and reduce LH pulsatility independent of its mass effect (see later section on hyperprolactinaemia).

Depending on the neuroanatomical location, size, and associated destruction and functional state, hypothalamic or pituitary tumours may be associated with anterior pituitary hormone deficiencies (e.g. secondary adrenal insufficiency and hypothyroidism) or hypersecretion (e.g. hyperprolactinaemia, Cushing syndrome, or acromegaly); diabetes insipidus or hypothalamic dysfunction; tumour mass effects (e.g. headache, visual field defects, or cerebrospinal fluid rhinorrhoea) that require evaluation and management independent of testosterone treatment for secondary hypogonadism.

Pituitary Apoplexy/Severe Head Trauma/Pituitary Stalk Injury/Hypopituitarism

Pituitary apoplexy is sudden haemorrhage or infarction within a pituitary tumour that can lead to compression and destruction of the pituitary gland, manifested by sudden onset of headache, nausea, visual defects, and acute hypopituitarism [2]. It is rare and occurs in only 2–12% of patients with pituitary adenomas. The acute impairment of pituitary function usually persists after the acute event particularly if surgery is required [17].

Infiltrative Disorders

Hypopituitarism leading to decreased anterior pituitary hormone secretion, including LH and FSH, can be caused by infiltrative inflammatory or granulomatous disorders of the pituitary. The typical diseases are lymphocytic hypophysitis, sarcoidosis, and Langerhans cell histiocytosis. Tuberculosis and fungal infections causing basilar meningitis can affect hypothalamic and pituitary function that results in secondary hypogonadism [1–3].

Immune Checkpoint Inhibitors and Hypophysitis

Recently, immune checkpoint inhibitors have become a mainstay in cancer treatment especially for metastatic melanoma and lung cancer. However, these can cause autoimmune effects and lead to endocrinopathies including hypophysitis and secondary hypogonadism. One new class of immunomodulating antibodies directed against CTLA4 (anti-CTLA4-mAbs) has been shown to have an incidence of hypophysitis ranging up to 17%. While most often secondary adrenal insufficiency and hypothyroidism occur, the incidence of hypogonadotropic hypogonadism with this type of hypophysitis was approximately 7.5% [18].

Haemochromatosis and Transfusion-Related Iron Overload

Iron overload caused either by hereditary hemochromatosis, a genetic disorder of excessive gastrointestinal iron absorption caused by mutations in the *HFE* gene, or due to chronic blood transfusions for haematologic conditions (e.g. β-thalassemia major, sickle cell disease) may involve excessive iron deposition in both the pituitary gonadotrope and testes, but more commonly presents with secondary hypogonadism [2, 3]. In more recent studies, the prevalence of secondary hypogonadism associated with haemochromatosis is approximately 6% and hypogonadism is more likely in men who have hepatic cirrhosis, ferritin concentrations exceeding 1500 ng/ml and complications of iron deposition such as diabetes mellitus [19]. Early treatment of iron overload with therapeutic phlebotomy or chelation therapy may reverse hypogonadism however, pituitary dysfunction is more commonly irreversible.

Functional Aetiologies

Constitutional Delayed Puberty (CPD)

As discussed, CPD is the most common cause of delayed puberty, affecting approximately 3% of adolescents. Also often referred to as 'late bloomers', this is thought to be a transient delay in production of GnRH from hypothalamus leading to decreased LH and FSH secretions that resolves eventually with older age [2].

Diabetes Mellitus

Low free testosterone and inappropriately normal or low LH and FSH concentrations occur in 30–40% of men with type 2 diabetes mellitus. The mechanisms contributing to secondary hypogonadism are likely to be multifactorial and include obesity, poor glycaemic control, systemic inflammation, and diabetic complications (e.g. CKD) [2–4]. Some of the mechanisms that have been observed with poorly controlled diabetes mellitus and hypogonadism include higher glucose values decreasing LH pulses, increase in inflammatory markers such as C-reactive protein (CPR), tumour necrosis factor-α (TNF-α) and interleukin-1β (IL-1β) suppressing GnRH release in hypothalamus and LH secretion [4].

Obesity

In men with mild to moderate obesity (body mass index, BMI, 30–40), SHBG concentrations are usually suppressed, resulting in low total testosterone concentrations; however, free testosterone concentrations are usually normal unless obesity is associated with comorbid illness such as diabetes or obstructive sleep apnoea in which case, free testosterone concentrations are low and gonadotropin concentrations are inappropriately normal or low, consistent with secondary hypogonadism [1–4]. Men with severe or morbid obesity (BMI >40) have more consistently low free testosterone concentrations and secondary hypogonadism. Gonadotropin suppression is likely due to increases in pro-inflammatory cytokines (e.g. tumour necrosis factor-α or interleukin-1) and dysregulated leptin signalling [4]. Because obesity associated with insulin resistance suppresses SHBG concentrations by reducing hepatic SHBG production, it is critical to measure free testosterone in the work-up of an obese man with hypogonadism [1]. Substantial weight loss following bariatric surgery increases free testosterone and gonadotropins [4].

Obstructive Sleep Apnoea (OSA)

Men with untreated or inadequately treated obstructive sleep apnoea are also more likely to have reduced testosterone and gonadotropin concentrations as well as symptoms of low energy, low libido, and fatigue [1–4]. Secondary hypogonadism is likely due in part to obesity and age, but also to the significant stress associated with nocturnal oxygen desaturation that results in arousal and sleep fragmentation [4].

Opioid Use

Use of opioid medications, especially long-acting opioids, for the treatment of chronic pain or opioid use disorder may cause secondary hypogonadism by inhibiting pulsatile hypothalamic GnRH secretion (primarily via interaction with μ-opioid receptors). Opioid-induced hypogonadism usually occurs in men taking high dosages of opioids (e.g. 100–200 mg of oral morphine equivalents daily, intrathecal morphine sulphate, oral methadone greater than 30 mg daily or with transdermal fentanyl patches) [20]. Tapering off opioids may reverse secondary hypogonadism, although this may not always be possible.

Hyperprolactinaemia

Hyperprolactinaemia can suppress pulsatile GnRH secretion that can be corrected with D_2 dopamine receptor agonist treatment. The most common causes of hyperprolactinaemia are: prolactinoma (most commonly, prolactin-secreting macroadenomas in men); pituitary stalk interruption (preventing endogenous hypothalamic dopamine-induced inhibition of prolactin secretion); hypothalamic disease; central nervous system-active medications (D_2 dopamine receptor antagonists, e.g. older antipsychotics such as phenothiazines, some atypical antipsychotics such as risperidone, and gastrointestinal promotility agents such as metoclopramide); CKD (as a result of decreased prolactin clearance). Treatment of the prolactinoma with normalization of serum prolactin concentrations is generally associated with an increase in serum testosterone concentrations unless there is tissue destruction or vascular compromise from very large macroprolactinomas. Discontinuation or switching of a medication that causes hyperprolactinaemia or treatment with D2 dopamine receptor agonists (e.g. cabergoline or bromocriptine) if possible, can reverse secondary hypogonadism [1–3].

Malnutrition/Excessive Exercise

Severe energy deficit due to starvation, protein-calorie malnutrition, and eating disorders (e.g. anorexia nervosa) where there is substantial weight loss can suppress hypothalamic pulsatile GnRH secretion. Refeeding, nutritional resuscitation, or treatment of eating disorders, respectively, restores HPT function. High intensity training or exercise seen in athletes or in the military that results in very high energy expenditure, usually associated with inadequate energy intake can also cause a state of energy deficit that decreases LH, FSH, and testosterone concentrations. Once excessive exercising is stopped or reduced and/or energy intake is increased, the HPT recovers usually in several weeks to months [2, 3].

Glucocorticoid Excess (Cushing Syndrome)

Glucocorticoid excess caused by either by exogenous glucocorticoid treatment or endogenous Cushing syndrome predominantly suppresses hypothalamic GnRH secretion; but also has direct inhibitory effects on the testes, inducing apoptosis in Leydig cells [2, 3]. Glucocorticoid-induced hypogonadism typically occurs with chronic high-dose glucocorticoid treatment, but testosterone suppression can occur in men taking doses as low as prednisone 7.5 mg daily [1, 2]. Restoration of normal HPT axis function and reversal of hypogonadism occurs with discontinuation of glucocorticoid administration or treatment of Cushing syndrome.

Sex Steroids

Androgens, progestins or oestrogens can suppress gonadotropin concentrations by negative feedback on the hypothalamus and/or

pituitary. Androgenic anabolic steroids (AASs, synthetic androgens) are taken in various combinations by athletes to enhance performance and non-athletes to enhance muscle bulk and strength. Because AASs do not cross-react in testosterone assays, testosterone levels are suppressed while taking these steroids. Patients taking AASs often present normal virilization but small testes on physical exam and have profoundly suppressed gonadotropins and testosterone that may persist with symptoms of androgen deficiency for months to years after discontinuation [21]. Progestins (e.g. megestrol acetate), an appetite stimulant that is used in patients with wasting syndromes such as cancer cachexia and HIV/AIDS wasting also cause severe suppression of gonadotropins and testosterone. Exposure to oestrogen-containing substances or oestrogen-secreting tumours of the testes or adrenal gland also induce secondary hypogonadism that is reversible with discontinuation of the offending substance or resection of the tumour, respectively [2].

Chronic Systemic Illness, Organ Failure

As previously discussed, chronic illnesses, especially hepatic cirrhosis, chronic kidney disease, congestive heart failure or chronic obstructive pulmonary disease (COPD) can cause combined primary and secondary hypogonadism. In severe late- or end-stage organ failure, secondary hypogonadism is predominant, and the degree of hypogonadism depends on the severity of illness and multiple other factors, such as malnutrition, systemic inflammation, medication exposure (e.g. prednisone use in COPD patients) [1–3].

Predominantly secondary more than primary hypogonadism also occurs in other chronic systemic illnesses such as, chronic active alcohol use disorder, HIV disease, and sickle cell disease [2].

Acute Illness or Surgery

Acute or critical illness or surgery requiring hospitalization often causes a hormonal pattern where secondary hypogonadism predominates with suppressed pulsatile GnRH secretion [1–3]. Testosterone and gonadotropin suppression improve once the acute illness is resolved; however, the duration of recovery is dependent on the severity of the acute illness, underlying chronic comorbid illness, age, and medications used during the illness and recovery (e.g. opioids or glucocorticoids) [22]. Therefore, evaluation for hypogonadism should not be performed until a patient is fully recovered, which may take several months.

REFERENCES

1. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
2. Matsumoto AM, Bremner WJ. Testicular disorders. In: Melmed S, Polansky KS, Larsen PR, Kronenberg HM, eds. *Williams Textbook of Endocrinology*, 13th edn. New York, NY: Elsevier; 2016: 688–777.
3. Basaria S. Male hypogonadism. *Lancet*, 2014; **383**: 1250–63.
4. Grossmann M, Matsumoto AM. A Perspective on middle-aged and older men with functional hypogonadism: focus on holistic management. *J Clin Endocrinol Metab*, 2017; **102**: 1067–75.
5. Gravholt CH, Chang S, Wallentin M, Fedder J, Moore P, Skakkebæk A. Klinefelter syndrome: integrating genetics, neuropsychology, and endocrinology. *Endocr Rev*, 2018; **39**: 389–423.
6. Groth KA, Skakkebæk A, Høst C, Gravholt CH, Bojesen A. Clinical review: Klinefelter syndrome—a clinical update. *J Clin Endocrinol Metab*, 2013; **98**: 20–30.
7. Auchus RJ. The genetics, pathophysiology, and management of human deficiencies of P450c17. *Endocrinol Metab Clin North Am,* 2001; **30**: 101–19.
8. Radicioni AF, Di Giorgio G, Grugni G, *et al*. Multiple forms of hypogonadism of central, peripheral or combined origin in males with Prader–Willi syndrome. *Clin Endocrinol (Oxf)*, 2012; **76**: 72–7.
9. Shapiro E, Kinsella TJ, Makuch RW, Fraass BA, Glatstein E, Rosenberg SA, Sherins RJ. Effects of fractionated irradiation of endocrine aspects of testicular function. *J Clin Oncol*, 1985; **3**: 1232–9.
10. Rivkees SA, Crawford JD. The relationship of gonadal activity and chemotherapy-induced gonadal damage. *JAMA*, 1988; **259**: 2123–5.
11. Betterle C, Dal Pra C, Mantero F, Zanchetta R. Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction. *Endocr Rev*, 2002; **23**: 327–64. Review. Erratum in: *Endocr Rev*, 2002; **23**: 579.
12. Matsumoto, Alvin M. Andropause: clinical implications of the decline in serum testosterone levels with aging in men. *J Gerontol A Biol Sci Med*, 2002; **57**: M76–M99.
13. Brioude F, Bouligand J, Trabado S, *et al*. Non-syndromic congenital hypogonadotropic hypogonadism: clinical presentation and genotype-phenotype relationships. *Eur J Endocrinol*, 2010; **162**: 835–51.
14. Young J. Approach to the male patient with congenital hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2012; **97**: 707–18.
15. Vissers LE, van Ravenswaaij CM, Admiraal R, *et al*. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*, 2004; **36**: 955–7.
16. Basciani S, Watanabe M, Mariani S, *et al*. Hypogonadism in a patient with two novel mutations of the luteinizing hormone β-subunit gene expressed in a compound heterozygous form. *J Clin Endocrinol Metab*, 2012; **97**: 3031–8.
17. Briet C, Salenave S, Bonneville J-F, Laws ER, Chanson P. Pituitary apoplexy. *Endocr Rev*, 2015; **36**: 622–45.
18. Byun DJ, Wolchok JD, Rosenberg LM, Girotra M. Cancer immunotherapy—immune checkpoint blockade and associated endocrinopathies. *Nat Rev Endocrinol*, 2017; **13**: 195–207.
19. McDermott JH, Walsh CH. Hypogonadism in hereditary hemochromatosis. *J Clin Endocrinol Metab*, 2005; **90**: 2451–5.
20. Abs R, Verhelst J, Maeyaert J, *et al*. Endocrine consequences of long-term intrathecal administration of opioids. *J Clin Endocrinol Metab*, 2000; **85**: 2215–22.
21. Coward RM, Rajanahally S, Kovac JR, Smith RP, Pastuszak AW, Lipshultz LI. Anabolic steroid induced hypogonadism in young men. *J Urol*, 2013; **190**: 2200–5.
22. Woolf PD, Hamill RW, McDonald JV, Lee LA, Kelly M. Transient hypogonadotropic hypogonadism caused by critical illness. *J Clin Endocrinol Metab*, 1985; **60**: 444–50.

10.4.2 Types of Treatment

Giulia Rastrelli, Mario Maggi, and Giovanni Corona

Introduction

This chapter will present the medications specifically approved for treatment of male hypogonadism (testosterone preparations and gonadotropins) and those not yet approved but currently under investigation as possible new therapeutic options.

Pharmacologic Treatments

Testosterone Preparations

In 1935, Karoly Gyula David and Ernst Laqueur extracted and purified an androgenic steroid from bull testes, which was named **testosterone** (T). In the same year, Butenandt's group in Gottingen and Ruzicka and Wettstein in Basel simultaneously described the chemical synthesis of T [1]. It soon became apparent that native T could not be given effectively by oral or parenteral routes, because of a prompt hepatic metabolism. Hence, a series of chemical modifications were developed to improve T bioavailability and pharmacokinetics, essentially by retarding the rate of liver catabolism [2, 3].

Nowadays many T formulations are available including oral, injectable(intramuscular or subcutaneous implantation), transdermal (gels, axillary solution, and patches) or transmucosal (buccal and nasal systems) formulations. In the following sections, the specific characteristics of the T preparations available on the market will be analysed in detail (Table 10.4.2.1). It should be noted that available T preparations have been approved for clinical use based only on their efficacy in raising T levels into the physiological range [4–7], rather than achieving clinical benefits. The efficacy of T preparations in improving the clinical features of hypogonadism has only been properly evaluated in more recent post-marketing randomized clinical trials (RCTs).

Oral Testosterone Preparations

Alkylation in 17-α position in ring A of the T molecule prevents rapid breakdown in the liver, hence, 17α-methyl-T was synthesized for oral use in 1935. However, this compound, along with all other 17α methyl T derivates, is associated with liver toxicity and are now no longer in clinical use [2, 3]. Unfortunately, these products are still available on the black market and continue to be abused as anabolic steroids [8]. In contrast, alkylation at the 1-α position as well as esterification of the 17α-hydroxyl group also reduce hepatic first-pass effect, but these modifications are not associated with hepatotoxicity. Hence, oral T formulation based on these chemical modifications are can be prescribed for clinical use (Table 10.4.2.1).

Testosterone Undecanoate

In the late 1970s, a new orally effective T formulation, based on 17β-esterification of native T with a long aliphatic side chain, undecanoic acid (oral T undecanoate, TU), was introduced to the market (Table 10.4.2.1). This lipophilic T ester allows absorption via the intestinal lymphatic system, avoiding the first-pass effect in the liver [2, 3]. Recently, this preparation has been reformulated, in a mixture of castor oil and propylene glycol laureate (*T undecanoate caps*), to increase stability of the product at ambient temperature [3]. Oral TU is formulated in 40 mg gel capsules. However, its bioavailability is poor on an empty stomach and the absorption is dependent on the dietary fat content [2, 3, 9]. A minimal amount of 20 mg of dietary fat is required for optimal oral TU absorption [2, 3, 9]. One or two 40 mg capsules should be taken thrice daily during meals, limiting its acceptability for compliant T treatment [2, 3, 9].

Parenteral Testosterone Preparations

Subdermal Implantation of T Pellets

T pellets are currently available only in a limited number of countries, including the United States and the United Kingdom. The pellets consist of pure crystals of 100 mg and 200 mg of T compressed into short rods and are placed, under local anaesthesia, into the subdermal fat layer of the skin usually in the lower abdominal wall or upper buttock area [2, 3, 9] (Table 10.4.2.1). The recommended dosage includes the implantation of 3 to 6 pellets (600–1200 mg) subcutaneously every 4 to 6 months. T pellet implantation provides the longest and the most stable form of T release, but possible adverse side effects include implant site infections and/or pellet extrusion occurring in 5–10% of cases.

Intramuscular Injectable Preparations

Injectable formulations of T are based on 17β-esterification of native T with fatty acids of variable aliphatic or aromatic chain lengths, and are administered in an oily base. Following intramuscular injection, the oily base acts as a depot, because it serves to retard the release of the T ester. Following release, the esters are rapidly hydrolysed to release native T into the circulation. Therefore, the duration of action of esterified T esters is largely dependent on the length of the ester side chain (which determines the lipophilic property/fat solubility of the T ester), as well as on the volume of the oily base.

T propionate (TP) has a short side chain ester and a short duration of action, requiring the administration of at least two to three injections per week (usually 50 mg every 2 or 3 days). **T cypionate (TC)** and **enanthate (TE)** have longer side chain esters and a longer duration of action, and can be injected every 10–14 days at doses of 200–250 mg.

The relatively short half-life of these T esters can lead to wide fluctuations in T plasma concentrations across the injection interval, in so-called peaks and troughs. Circulating T concentrations are commonly supraphysiological in the first few days after administration and then decline to subphysiological levels in the days prior to the next scheduled injection. This can be perceived as unpleasant variations in mood or energy by patients and represents a risk factor for erythrocytosis [10]. However, due to due to their long-established usage and relatively low cost compared to other formulations, these shorter-acting T esters continue to be prescribed widely in some countries.

In 2004, a **long-lasting injectable formulation of TU** was introduced to the market [2, 3, 9] (Table 10.4.2.1). The greater fat solubility rendered by the longer fatty acid side chain and the relatively large volume of oily base in the formulation increases the duration of action compared to other T esters [11] (Table 10.4.2.1). In most countries, injectable TU is available as a 1000 mg dose in 4 ml castor oil. This formulation maintains relatively stable circulating T concentrations when injected at 12 weekly intervals, following a week 6 loading dose. Once steady-state T concentrations have been achieved, usually after the third dose, the injection interval may need to be adjusted to every 10–14 weeks to achieve therapeutic serum T trough concentrations of 10–15 nmol/L. In the United States, TU is marketed as 750 mg is 3 ml castor oil. This formulation requires 10-weekly injections, following an initial week 4 loading dose. Injectable TU preparations result in lower plasma T fluctuations compared to other intramuscular T preparations with shorter half-lives, reducing the risk of polycythemia and improving patient compliance [11].

An uncommon adverse effect specific to injectable T formulations in oily base reported in post-marketing studies is pulmonary oil microembolisms due to leakage of the oil-based depot into the systemic circulation [2, 3, 9]. This can manifest with transient cough and respiratory distress immediately after the injection which generally does not require any intervention apart from reassurance. In addition, along with other parenteral preparations, injectable TU should be avoided in men at increased bleeding risk, due to, for example, coagulopathies or thrombocytopaenia.

Transdermal Testosterone Preparations

Testosterone Patches

Transdermal T formulations have been available since the mid-1990s, initially as self-adherent T patches designed to be applied to the relatively thin, highly vascularized scrotal skin. Scrotal patches gained limited acceptance given they are relatively large, require shaving, and have relatively poor adherence to the skin. Moreover, scrotal systems tend to disproportionally increase dihydrotestosterone (DHT) due to the relatively high expression of 5α-reductase in scrotal skin. Subsequently, patches for application on non-scrotal skin were developed [2, 3, 9] (Table 10.4.2.1). Non-scrotal patches require alcohol-based absorption enhancers which not uncommonly leads to skin irritation at the application site which, in addition to the requirement of daily application, may limit compliance [2, 3, 9].

Testosterone Gels

Given drawbacks associated with T patches, transdermal T formulations in hydroalcoholic gels with less irritative penetration enhancers have become available in the early 2000s. Transdermal gels are available in different concentrations (1%, 1.62%, and 2%) either as sachets (usually containing 50 mg of T) or as metered-dose pumps to allow dose titrations. Worldwide, transdermal gels constitute the largest market share of T prescriptions [2, 3, 9] (Table 10.4.2.1). Like patches, T gels require daily administration usually to upper arms or shoulders. Once applied, T gel is rapidly absorbed by the stratum corneum and subcutaneous tissues which serve as a reservoir from which T is gradually released into systemic circulation, leading to relatively stable circulating T concentrations over the subsequent 24 hours (for dosing regimens see Table 10.4.2.1). However, T levels can be variable from day to day and dose adjustment is frequently required. To ensure optimal absorption, the skin should be kept dry for 4–6 hours following gel application [2, 3, 9]. Due to the potential risk of T transfer, men are instructed to avoid close skin contact with others for the first few hours following gel application. To try and reduce transference risk, an alcohol-based T solution (2%) developed for axillary application is available in some countries, but its acceptability has been relatively low [2, 3, 9] (Table 10.4.2.1)

Transmucosal Testosterone Preparations

Transbuccal Testosterone Preparations

In some countries, a buccal preparation in tablet form is available, designed to adhere to the gum or inner cheek to provide a controlled and sustained release of T through the buccal mucosa. The tablet does not dissolve completely and must be removed and replaced every 12 hours. Each buccal system contains 30 mg of T. This formulation can maintain physiological T concentrations, but local irritation (such as gum oedema, ulceration, and gingivitis) can occur [2, 3, 9] (Table 10.4.2.1).

Transnasal Testosterone Preparations

A gel containing 5.5 mg of T in 122.5 mg of gel for nasal administration is available in some countries, including the United States and Canada. The recommended dose is 11 mg of T (corresponding to two pump actuations, one per nostril) administered three times daily [2, 3, 9] (Table 10.4.2.1).

Monitoring Testosterone Treatment

International guidelines [12, 13], recommend that patients receiving T treatment should be reviewed 3–6 months after initiation of therapy and yearly thereafter. Clinical assessment includes monitoring for improvement in hypogonadal symptoms and evaluation for side effects of T treatment (Table 10.4.2.1). Serum T should be measured to assess adequacy of T dosing, noting that timing of the blood draw and circulating T targets vary depending on the T formulation used (Table 10.4.2.1). Haematocrit should be monitored and if above 0.54, T therapy should be withheld until it declines to the normal range (see Chapter 10.4.5, 'Risks of T Treatment'). For men of more than 55 years of age who, following individualized discussion regarding advantages and disadvantages, opt for prostate monitoring, digital rectal exam, and prostate-specific antigen testing can be performed at regular intervals. In hypogonadal men at high risk of osteoporotic fracture, bone mineral density is measured at baseline, while subsequent measurements are individualized.

Androgens Other Than Testosterone

Dehydroepiandrosterone

Dehydroepiandrosterone (DHEA) is produced by the adrenal glands and circulates as a sulphated form (DHEAS). DHEA is a weak androgen receptor agonist and largely serves a steroid precursor of T. A meta-analysis of 25 RCTs including 1353 older men treated with DHEA at a dose varying from 25 to 100 mg per day for a mean of 36 weeks found that DHEA did not improve sexual function, quality of life, bone health or metabolic profile [14]. Therefore, use of DHEA for treatment of male hypogonadism is not recommended.

Table 10.4.2.1 Available testosterone preparations for the treatment of adulthood hypogonadism. NA = not available

Formulation	Trade names	Chemical structure	Half-life	Standard dosage	Monitoring T concentrations		Side effects
					When to measure	When to adjust the dose or frequency	
Testosterone preparations							
Oral							
Testosterone undecanoate	Andriol® Andriol Testocaps®	17-α-hydroxyl-ester	4 h	120–240 mg 2–3 times daily	3–5 h after ingestion	NA	Need for multiple assumptions per day with fatty meals
Mesterolone	Proviron®	1α-methyl-4,5α-dihydrotestosterone	12 h	50–100 mg 2–3 times daily	NA	NA	NA
Parental							
Testosterone enanthate	Testoviron Depot® Delatestryl® Testoenant®	17-α-hydroxyl-ester	4–5 days	250 mg every 2–3 weeks	Midway between injections	TT<14 nmol/L TT>25 nmol/L	Pain at injection site Cough after injection Fluctuation in symptoms
Testosterone cypionathe	Delatestril®	17-α-hydroxyl-ester	8 days	200 mg every 2–3 weeks	Midway between injections	TT<14 nmol/L TT>25 nmol/L	
Testosterone propionate	Testovis®	17-α-hydroxyl-ester	20 h	100 mg every 2 days	NA	NA	
Testosterone undecanoate in castor oil	Nebido® Aveed ®(US)*	17-α-hydroxyl-ester	34 days	1000 mg every 10–14 weeks *750 mg every 10 weeks	The week prior the next injection	TT nadir below or above the low-mid normal range	
Surgical implants	Testopel® Testoimplant®	Native testosterone	–	4–6 200 mg implants lasting up to 6 months	At the end of dosing interval	TT nadir below or above the low-mid normal range	Infection Expulsion of implant
Transdermal							
Testosterone patches	*Not scrotal:* Androderm® Andropatch® Testopatch®	Native testosterone	10 h	50–100 mg/day	3–12 h after application	TT below or above the low-mid normal range	Itching and skin irritation at application site
Testosterone gel 1–2%	*1% gel:* Androgel® Testogel® Testim® *2% gel:* Testostop® Tostrex® (also known as Fortesta®, Tostran® and Itnogen®, available only in Europe) *1.6% gel* Androgel (available only in US)	Native testosterone	6 h	50–100 mg/day	2–8 h after application	TT below or above the low-mid normal range	Skin irritation at application site Stickiness or slow drying Potential risk of transfer to person in close contact
Underarm testosterone (testosterone solution 2%)	Axiron®	Native testosterone	NA	60–120 mg/day	2–8 h after application	TT below or above the low-mid normal range	
Dihydrotestosterone	Andractim ®	Native dihydrotestosterone	NA	5 or 10 g/day	NA	NA	NA

Transmucosal							
Testosterone buccal	Striant®	Native testosterone	12 h	30 mg/twice daily	Immediately before or after application	NA	Gum Irritation Taste alterations
Testosterone nasal	Natesto®	Native testosterone	10–100 minutes	11 mg three times/daily	NA	NA	Nasal discomfort Rhinorrhoea Epistaxis Nasal congestion Parosmia
Androgens other than testosterone							
Oral							
Dehydroepiandrosterone	Galenic preparation	Native dehydroepiandrosterone	12 h	25–100 mg daily	NA	NA	NA
Mesterolone	Proviron®	1α-methyl-4,5α-dihydrotestosterone	12 h	50–100 mg 2–3 times daily	NA	NA	NA
Transdermal							
Dihydrotestosterone	Andractim ®	Native dihydrotestosterone	NA	5 or 10 g/day	NA	NA	NA

Dihydrotestosterone (DHT) Gels

In some European countries, DHT is available as a hydroalcoholic 2.5% gel with a dosage of 5 or 10 g/day [2, 3, 9] (Table 10.4.2.1). It has been used in the context of gynaecomastia or microphallus, but evidence for benefit is limited [2, 3, 9] (Table 10.4.2.1). Moreover, DHT cannot be aromatized to oestradiol. This limits DHT use, given increasing evidence that some of the biological actions attributed to T (e.g. maintenance of bone mass, prevention of adipose tissue accumulation, and possibly maintenance of libido) are dependent on its conversion to oestradiol [15]. Moreover, evidence regarding benefits and risks of DHT is limited, due to the paucity of RCTs. In one RCT of 120 men aged 50–70 years with borderline-low T levels, DHT treatment (125–250 mg daily for 6 months) was associated with a modest improvement in sexual function, but no effects on body weight or lipid parameters [16]. In an RCT of 114 healthy (not hypogonadal) men DHT gel (70 mg/daily) modestly improved body composition and measures of muscle performance, without evidence of adverse prostate effects [17, 18]. However, bone mineral density and sexual desire decreased, likely due to the relative lack of oestrogenic action.

Mesterolone

Mesterolone (1α-methyl-4,5α-dihydrotestosterone) is a 1α-methylated derivative of 5α-dihydrotestosterone (DHT) available for oral administration (Table 10.4.2.1) [2, 3]. However, similar to DHT, Mesterolone cannot be aromatized to oestradiol, limiting its clinical use.

Alternative Treatments to Testosterone

GnRH/Gonadotrophin Therapy

Gonadotropin-releasing hormone (GnRH) and gonadotrophin therapy is most commonly used to stimulate spermatogenesis (see Chapter 10.4.3, 'Gonadotrophin Induction of Spermatogenesis'). These agents require testicular responsiveness, and are therefore only effective in hypogonadotropic hypogonadism. Due to cost and need for parental therapy, they are generally not a practical treatment option for T replacement in hypogonadism. with limited evidence supporting their clinical use in this context.

Only two studies have evaluated the efficacy of luteotropic hormone (LTH) stimulation in adult men [19–21]. These RCTs have used human chorionic gonadotrophin (hCG) due to its longer half-life compared to LTH in doses used to induce spermatogenesis. In a placebo-controlled RCT, hCG modestly increased lean mass and decreased fat mass and modestly improved lipid parameters, but had no effect on sexual function [20, 21]. A very small study (including ten men receiving hCG) reported improvements in sexual function comparable to T treatment, but this study lacked a control group [22].

Antioestrogens

This category includes the selective oestrogen receptor (ER) modulators (SERMs) and the aromatase inhibitors (AIs). These orally active agents antagonize the oestradiol-mediated negative feedback that normally supresses GnRH and gonadotrophin section and lead to a reflex rise in circulating LTH, FSH, and T. However, the agents require an intact gonadal axis and are not effective in organic hypogonadotropic hypogonadism due to established hypothalamic–pituitary disease. In addition, as discussed next, they may interfere with oestrogenic signalling and to date, there are limited clinical data supporting their use. It should be noted that both SERMs and AIs are not approved for use in male hypogonadism. For all these reasons, their use should be considered experimental and restricted to clinical trials.

SERMs

SERMs are non-steroidal compounds that bind to the oestrogen receptor and exert both agonistic and/or antagonist actions, depending on the target tissue.

SERMs are classified into three generations. The first generation includes the triphenylethylene derivatives, clomiphene citrate (CC), tamoxifen, and toremifene. The second generation includes the benzothiophene derivatives, of which only raloxifene is available for clinical use. More recently, third generation compounds such as basedoxifene, ospemifene, and lasofoxifene have become available, but have not been evaluated in men.

CC is a non-racemic mixture composed of two stereoisomers: the trans-isomer enclomiphene (ECC), representing 62% and the cis-isomer zuclomiphene representing 38% of the mixture [3]. ECC has a pure antioestrogen effect with a half-life of 10 hours. Zuclomiphene has mixed oestrogenic and antioestrogenic actions and a longer half-life of 30 days. CC is approved for use in female infertility to induce ovulation. Due to its antioestrogenic actions, CC antagonizes the oestradiol-mediated negative central gonadal axis feedback resulting in increased gonadotropin secretion, which stimulates oocyte development. The same mechanism can be exploited in men, whose endogenous T levels significantly increase after using CC, ECC, or other SERMs [3].

Besides CC and ECC, tamoxifen, toremifene, and raloxifene have been evaluated in clinical studies using doses listed in Table 10.4.2.2 [3]. A meta-analysis of available studies has reported that in men with low to borderline T concentrations who do not have organic gonadal axis pathology, SERMs increase gonadotropins and circulating T, generally into the mid-normal range for healthy young men. However, so far, high-quality studies assessing their effects on androgen deficiency symptoms and end organ effects are lacking and currently available reports do not provide conclusive evidence of efficacy [3]. One placebo-controlled 8-week RCT in 17 men with lowered T concentration (less than 9.4 nmol/L) and erectile dysfunction (ED) [23] reported that CC at a dose of 50 mg daily had no effect on erectile function, despite a marked increase in T (+10.0 nmol/L vs. baseline and +7.7 nmol/L vs. placebo). Another 12-week RCT in 26 men with lowered serum T (less than 10.4 nmol/L) [24] compared CC 25 mg daily with anastrozole 1 mg similarly found no improvement in hypogonadal symptoms, including ED, were observed. In contrast to exogenous T, which suppresses gonadotropins and spermatogenesis, SERMs increase gonadotropins and may stimulate/preserve spermatogenesis in patients with functional secondary hypogonadism. To date, the largest RCT involving a SERM randomized 256 obese men with low serum T to either enclomiphene, T, or placebo. As expected ECC increased, but T therapy decreased, gonadotropins, while both ECC and T therapy increased serum T concentrations compared to placebo [25]. ECC maintained, whereas T modestly reduced sperm counts from a normal baseline [25]. However, to date, robust data on symptomatic improvements and fertility are

Table 10.4.2.2 Available preparations alternative to testosterone for the treatment of adulthood hypogonadism. NA= not available

Formulation	Trade names	Chemical structure	Half-life	Standard dosage
GnRH				
Gonadotropin-releasing hormone (GnRH)	Lutrelef® Factrel® Lutrepulse®	Native GnRH	10–40 minutes	100–400 ng/kg every 90–120 min
Gonadotropins				
Human chorionic gonadotropin (hCG) *Extractive* *Recombinant* Luteotropic hormone (LTH) *Recombinant*	Novarel® Chorex® Gonic® Choron® Chorigon® Pregnyl® Gonasi® Prophasi® Ovidrel® Ovitrelle®	hCG purified from the urine of pregnant women Human recombinant hCG Human recombinant LTH	23–38 h 29 h 4.5 h	1000–2000 IU 3 times/week NA NA
Follicular stimulating hormone (FSH) *Extractive* *Recombinant* *Follitropin alfa* *Follitropin beta* *Follitropin delta*	Menogon® Meropur® Fostimon® Fertinorm® Bravelle® Gonal-f® Cinnal-f® Fertilex® Ovaleap® Bemfola® Follistim® Puregon® Follitropin (generic)	FSH purified from the urine of pregnant women Human recombinant FSH Human recombinant FSH	NA	75–150 IU 3 times/week 75–150 IU 3 times/week 75–150 IU 3 times/week
Antiestrogen preparations				
Selective estrogen receptor modulators (serm)				
Clomiphene Enclomiphene Tamoxifen Toremifen Raloxifene	Clomid® Androxal® Nolvadex® Istuba® Valodex® Fareston® Evista® Optruma®	*Triphenylethylene derivative mixture of the trans-isomer, enclomiphene (62%) and the cis-isomer, zuclomiphene (38%)* *Trans-isomer with pure antioestrogen effect* *(Z)-2-[4-(1,2-Diphenyl-1-butenyl) phenoxy]-N,N-dimethylethanamine* *4-Chlorotamoxifen* *Benzothiophene derivative*	12 h 10 h 5–7 days 5–6 days 27.7 h	25–50 mg/day 12.5–25 mg/day 100–150 mg/week 80 mg/daily 60 mg/daily
Aromatase Inhibitor				
Letrozole Anastrazole	Femara® Estriol® Arimidex®	$C_{17}H_{11}N_5$ $C_{17}H_{19}N_5$	2 days 46.8 h	2.5–17.5 mg/week 1 mg/day

not available. ECC has not been approved for clinical use and is not commercially available.

Raloxifene and toremifene, both SERMs with partial agonistic activity in bone, have been evaluated for their effect on skeletal endpoints. In a 6 month RCT of older men with serum bioavailable T less than 4.5 nmol/L, treatment with raloxifene 60 mg daily was associated with an improvement in markers of bone remodelling compared to placebo, but only in the subgroup of men with the lowest oestradiol and T concentrations [26]. Similar results were reported in a 6 months RCT of 43 healthy older men treated with raloxifene 120 mg daily [27]. Effects of SERMs on bone have also been assessed in men receiving androgen deprivation therapy (ADT) for prostate cancer, who have castrate concentrations of both T and oestradiol. In a large (n = 646 men receiving ADT) RCT, toremifene 80 mg daily for 2 years was associated with a significantly lower incidence of morphometric (according to the Genant semiquantitative scoring method) vertebral fractures than placebo (relative risk reduction of −50 ([−1.5; −75]%; p <0.05) [28]. However, men receiving toremifene had higher incidence of venous thromboembolic events (VTE; 2.6% vs. 1.1%) that was significant among older men (80 years old or older), which led the manufacturer to discontinue further evaluation of toremifene in men with prostate cancer receiving ADT [2, 29]. While an increased risk of VTE has not been reported in younger men without prostate cancer, existing studies have not been designed or powered to evaluate such risks. However, SERMs should be avoided in men at risk for VTE until more definitive data are available.

Aromatase Inhibitors

AIs inhibit aromatase, an enzyme whose physiologic function is to convert T into oestradiol. Based on their structure, AIs are classified into steroidal (testolactone and exemestane) and non-steroidal (anastrozole and letrozole) compounds. Exemestane, anastrozole, and letrozole are approved for the treatment of oestrogen receptor positive breast cancer in women. AIs profoundly reduce serum oestradiol (by >98%), and hence, in a mechanism similar to SERMs, remove oestradiol-mediated central gonadal feedback with consequent increase in gonadotropin secretion in turn stimulating testicular T production and possibly spermatogenesis. However, to date beneficial effects of AIs on spermatogenesis and fertility have not been reported. Similar to SERMs AIs increase circulating T into the mid-normal range, with an average difference of 7 nmol/L compared to placebo arm [3]. Effect in clinical outcomes such as sexual function, bone parameters, muscle strength, and metabolic outcomes have not been adequately evaluated to date. Effects of AIs on sexual function have been evaluated by three small short-term placebo-controlled RCTs [30–32]. None of the studies reported an improvement in sexual function. While this may have been due to lack of power, the findings are consistent with evidence that oestradiol plays a role in sexual function in men. Three RCTs [32–34] evaluated the effect of AIs on body composition, lipids, and muscle strength. None of them showed any difference in fat or lean mass and lipids in the AI arm as compared with placebo. Given the profound reduction of circulating oestradiol, detrimental effects of AIs on bone health are a potential concern. While two small RCTs did not report an effect on bone density [32, 35] the largest RCT (n = 69) using anastrozole reported a significant reduction in lumbar spine bone mineral density (BMD) at 52 weeks compared to placebo (change from baseline BMD: −0.020 ± 0.005 vs. 0.009 ± 0.006 for anastrozole and placebo group, respectively; $P = 0.0014$) [34]. At present, the available data do not support a role for AIs in the treatment of men with lowered T, despite their ability to increase circulating T levels [3]. However, the limited available evidence is not sufficient for drawing conclusions on their clinical utility, and further studies are required.

Conclusions

The optimal treatment of adult hypogonadism should be based on the clinical context, treatment availability, and individual patient needs and expectations. In the majority of cases, T substitution represents the best treatment. The use of transdermal T preparations and long-acting injectable T undecanoate have become the most popular forms of the replacement, providing assured efficacy and safety. When fertility is required, in the case of secondary hypogonadism, the use of gonadotrophins is appropriate and quite successful. The unlicensed use of antioestrogens in the management of male hypogonadism is not supported by good evidence on efficacy and carries concerns regarding long-term safety.

REFERENCES

1. Nieschlag E, Nieschlag S. Testosterone deficiency: a historical perspective. *Asian J Androl*, 2014; **16**: 161–8.
2. Corona G, Rastrelli G, Vignozzi L, Maggi M. Emerging medication for the treatment of male hypogonadism. *Expert Opin Emerg Drugs*, 2012; **17**: 239–59.
3. Rastrelli G, Maggi M, Corona G. Pharmacological management of late-onset hypogonadism. *Expert Rev Clin Pharmacol*, 2018; **11**: 439–58.
4. Handelsman DJ, Conway AJ, Boylan LM. Pharmacokinetics and pharmacodynamics of testosterone pellets in man. *J Clin Endocrinol Metab*, 1990; **71**: 216–22.
5. Behre HM, Abshagen K, Oettel M, Hübler D, Nieschlag E. Intramuscular injection of testosterone undecanoate for the treatment of male hypogonadism: phase I studies. *Eur J Endocrinol*, 1999; **140**: 414–19.
6. Wang C, Swerdloff R, Kipnes M, *et al*. New testosterone buccal system (Striant) delivers physiological testosterone levels: pharmacokinetics study in hypogonadal men. *J Clin Endocrinol Metab*, 2004; **89**: 3821–9.
7. Wang C, Ilani N, Arver S, McLachlan RI, Soulis T, Watkinson A. Efficacy and safety of the 2% formulation of testosterone topical solution applied to the axillae in androgen-deficient men. *Clin Endocrinol (Oxf)*, 2011; **75**: 836–43.
8. Nieschlag E, Vorona E. Medical consequences of doping with anabolic androgenic steroids: effects on reproductive functions. *Eur J Endocrinol*, 2015; **173**: R47–58.
9. Shoskes JJ, Wilson MK, Spinner ML. Pharmacology of testosterone replacement therapy preparations. *Transl Androl Urol*, 2016; **5**: 834–43.
10. Ohlander SJ, Varghese B, Pastuszak AW. Erythrocytosis following testosterone therapy. *Sex Med Rev*, 2018; **6**: 77–85.

11. Corona G, Maseroli E, Maggi M. Injectable testosterone undecanoate for the treatment of hypogonadism. *Expert Opin Pharmacother*, 2014; **15**: 1903–26.
12. Wang C, Nieschlag E, Swerdloff R, *et al.*; International Society of Andrology; International Society for the Study of Aging Male; European Association of Urology; European Academy of Andrology; American Society of Andrology. Investigation, treatment, and monitoring of late-onset hypogonadism in males: ISA, ISSAM, EAU, EAA, and ASA recommendations. *Eur Urol*, 2009; **55**: 121–30.
13. Bhasin S, Brito JP, Cunningham GR, *et al.* Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
14. Corona G, Rastrelli G, Giagulli VA, *et al.* Dehydroepiandrosterone supplementation in elderly men: a meta-analysisstudy of placebo-controlled trials. *J Clin Endocrinol Metab*, 2013; **98**: 3615–26.
15. Finkelstein JS, Lee H, Burnett-Bowie SA, *et al.* Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*, 2013; **369**: 1011–22.
16. Kunelius P, Lukkarinen O, Hannuksela ML, Itkonen O, Tapanainen JS. The effects of transdermal dihydrotestosterone in the aging male: a prospective, randomized, double blind study. *J Clin Endocrinol Metab*, 2002; **87**: 1467–72.
17. Idan A, Griffiths KA, Harwood DT, *et al.* Long-term effects of dihydrotestosterone treatment on prostate growth in healthy, middle-aged men without prostate disease: a randomized, placebo-controlled trial. *Ann Int Med*, 2010; **153**: 621–32.
18. Sartorius GA, Ly LP, Handelsman DJ Male sexual function can be maintained without aromatization: randomized placebo-controlled trial of dihydrotestosterone (DHT) in healthy, older men for 24 months. *J Sex Med*, 2014; **11**: 2562–70.
19. Corona G, Rastrelli G, Reisman Y, Sforza A, Maggi M. The safety of available treatments of male hypogonadism in organic and functional hypogonadism. *Expert Opin Drug Saf*, 2018; **17**: 277–292.
20. Liu PY, Wishart SM, Handelsman DJ. A double blind, placebo-controlled, randomized clinical trial of recombinant human chorionic gonadotropin on muscle strength and physical function and activity in older men with partial age-related androgen deficiency. *J Clin Endocrinol Metab*, 2002; **87**: 3125–35.
21. Liu PY, Wishart SM, Celermajer DS, et al. Do reproductive hormones modify insulin sensitivity and metabolism in older men? A randomized, placebo-controlled clinical trial of recombinant human chorionic gonadotropin. *Eur J Endocrinol*, 2003; **148**: 55–66.
22. La Vignera S, Condorelli RA, Cimino L, Russo GI, Morgia G, Calogero AE. Late-onset hypogonadism: the advantages of treatment with human chorionic gonadotropin rather than testosterone. *Aging Male*, 2016; **19**: 34–9.
23. Guay AT, Bansal S, Heatley GJ. Effect of raising endogenous testosterone levels in impotent men with secondary hypogonadism: double blind placebo-controlled trial with clomiphene citrate. *J Clin Endocrinol Metab*, 1995; **80**: 354–52.
24. Helo S, Ellen J, Mechlin C, *et al.* A randomized prospective double-blind comparison trial of clomiphene citrate and anastrozole in raising testosterone in hypogonadal infertile men. *J Sex Med*, 2015; **12**: 1761–9.
25. Kim ED, McCullough A, Kaminetsky J. Oral enclomiphene citrate raises testosterone and preserves sperm counts in obese hypogonadal men, unlike topical testosterone: restoration instead of replacement. *BJU Int*, 2016; **117**: 677–85.
26. Doran PM, Riggs BL, Atkinson EJ, Khosla S. Effects of raloxifene, a selective estrogen receptor modulator, on bone turnover markers and serum sex steroid and lipid levels in elderly men. *J Bone Miner Res*, 2001; **16**: 2118–25.
27. Uebelhart B, Herrmann F, Pavo, Draper MW, Rizzoli R. Raloxifene treatment is associated with increased serum estradiol and decreased bone remodeling in healthy middle-aged men with low sex hormone levels. *J Bone Miner Res*, 2004; **19**: 1518–24.
28. Smith MR, Morton RA, Barnette KG, *et al.* Toremifene to reduce fracture risk in men receiving androgen deprivation therapy for prostate cancer. *J Urol*, 2010; **184**: 1316–21.
29. Smith MR, Morton RA, Barnette KG, *et al.* Toremifene to reduce fracture risk in men receiving androgen deprivation therapy for prostate cancer. *J Urol*, 2013; **18**(1 Suppl): S45–50.
30. Leder BZ, Rohrer JL, Rubin SD, Gallo J, Longcope C. Effects of aromatase inhibition in elderly men with low or borderline-low serum testosterone levels. *J Clin Endocrinol Metab*, 2004; **89**: 1174–80.
31. Herzog AG, Farina EL, Drislane FW, *et al.* A comparison of anastrozole and testosterone versus placebo and testosterone for treatment of sexual dysfunction in men with epilepsy and hypogonadism. *Epilepsy Behav*, 2010; **17**: 264–71.
32. Loves S, de Jong J, van Sorge A, *et al.* Somatic and psychological effects of low-dose aromatase inhibition in men with obesity-related hypogonadotropic hypotestosteronemia. *Eur J Endocrinol*, 2013; **169**: 705–14.
33. Dias JP, Melvin D, Simonsick EM, *et al.* Effects of aromatase inhibition vs. testosterone in older men with low testosterone: randomized-controlled trial. *Andrology*, 2016; **4**: 33–40.
34. Burnett-Bowie SA, McKay EA, Lee H, Leder BZ. Effects of aromatase inhibition on bone mineral density and bone turnover in older men with low testosterone levels. *J Clin Endocrinol Metab*, 2009; **94**: 4785–92.
35. Leder BZ, Finkelstein JS. Effect of aromatase inhibition on bone metabolism in elderly hypogonadal men. *Osteoporos Int*, 2005; **16**: 1487–94.

10.4.3 Induction of Spermatogenesis by Gonadotrophin Treatment

Michael Zitzmann

Introduction—General Principles

Testicular dysfunction often compromises both production of testosterone and spermatogenesis. These functions are dependent on

gonadotropin action in men, with luteinizing hormone (LH) stimulating testosterone production in Leydig cells and follicle-stimulating hormone (FSH) supporting spermatogenesis by stimulating Sertoli cell function (see Chapter 10.1.1). Testicular disorders, causing primary hypogonadism, do not present a condition in which a therapy with gonadotropin-releasing hormone (GnRH) or gonadotropins is an option to induce spermatogenesis. Rather, morbidities with impaired secretion of GnRH (tertiary hypogonadism) or lack of production of gonadotropins (secondary hypogonadism) in the presence of an intact, yet unstimulated testis, present an option for the therapy with these hormones [1–4].

Mutations in a large set of genes causing the secondary and tertiary hypogonadism have been reported [5, 6], but also other, prepubertal or postpubertally acquired pathologies of the pituitary and hypothalamus can give rise to a similar a clinical picture (see Section 2, 'Pituitary and Hypothalamic Diseases').

Hypogonadism is usually treated with replacement of testosterone to correct androgen deficiency. However, in hypogonadotropic hypogonadism (HH), it is also possible to achieve fertility as well, by temporarily switching testosterone treatment to gonadotropins: human chorionic gonadotropin (hCG) in combination with FSH (purified urinary FSH or recombinant human FSH [rhFSH]) or, alternatively, pulsatile GnRH in case of tertiary hypogonadism [7–13]. Once paternity has been achieved, men should switch back to more convenient and cost-effective testosterone replacement, unless men wishing to father a second child within a short time frame.

Hormonal treatment designed to achieve fertility in hypogonadotropic hypogonadal men usually do not increase sperm counts into the normal range (sperm concentration ≥15 mill/ml or a total sperm count ≥38 mill/ejaculate as per current World Health Organization definitions). Normalization of sperm counts is not necessary to achieve fertility. In a study of 24 men with IHH receiving gonadotropin therapy, 71% of a total of 40 initiated pregnancies were conceived with sperm concentrations ranging from 1 to 20 mill/ml [14]. Similar results were seen in men with IHH receiving pulsatile GnRH combined hCG/hMG therapy [15, 16]. The length of treatment required to induce spermatogenesis and achieve fertility (ranging from a few months to several years) depends on cofactors discussed next [17].

Spermatogenesis can be induced even in hypogonadotropic patients with a very small testicular volume (of less than 3 ml), but this may require treatment for 18–24 months [15, 16, 18]. During therapy, testicular volume can be monitored carefully by ultrasound sonography, in order to detect subtle increases, which precede the first appearance of sperm [19].

A review of 42 patients with HH reported that a history of surgically corrected testicular maldescent (uni- or bilateral) hampers spermatogenesis and requires a longer course of treatment, but again, this does not preclude fertility. This is especially the case in unilateral maldescent, but there are also reports of men with previous bilateral cryptorchidism achieving paternity with gonadotrophin treatment. On average, patients with unilateral maldescent required 5 months (range 1–16 months) and patients with bilateral maldescent required 13 months (12–22 months) of treatment until induction of spermatogenesis, that is appearance of sperm in the ejaculate. In comparison patients without a history of maldescent required 4.5 months of treatment (2–18 months) [16, 17]. Similar findings have been reported in other studies [20–24]. Since induction of spermatogenesis may take 2 years or occasionally even longer, men with HH desiring paternity should plan well in advance to initiate timely treatment. In repeatedly treated patients, stimulation of spermatogenesis tended to be faster, leading to a reduced time to pregnancy [16]. In one study [25], 75 men with idiopathic hypogonadotropic hypogonadism (IHH) (with 72 desiring fertility) were treated at two academic andrology centres for a total of 116 courses of therapy from 1981–2008; semen analysis and testicular examination were performed every 3 months. A total of 38 men became fathers, including five through assisted reproduction. The median time to achieve first sperm appearance was 7.1 months (95% confidence interval [CI] 6.3–10.1) and for conception was 28.2 months (95% CI 21.6–38.5). The median sperm concentration at conception for unassisted pregnancies was 8.0 m/ml (95% CI 0.2–59.5). Multivariate correlated time-to-event analyses showed that larger testis volume predicted faster induction of spermatogenesis and unassisted pregnancy. In general, advanced age was a negative predictor, but a definitive threshold in terms of 'too old age' does not seem to exist.

It has been recently shown that the underlying pathology strongly influences the treatment response in HH. A history of puberty was a predictive marker for successful of gonadotropin substitution, while, for example patients with Kallmann syndrome had lower chances for testicular growth and they also had lower sperm output [17].

The induction of fertility in men with secondary or tertiary hypogonadism is highly effective but obviously needs patience in subgroups with unfavourable confounders [17, 19, 26]. Of note, recent evidence suggests that central hypogonadism may be a reversible condition in up to 10% men [27], and should be considered in men experiencing testicular growth despite testosterone treatment. In such men, gonadotropin treatment to induce spermatogenesis or testosterone replacement may no longer be necessary; however, these men require periodic clinical monitoring, as the gonadal axis recovery may only be temporary.

Gonadotropin Therapy

Gonadotrophin therapy is effective in promoting fertility in patients with secondary as well as tertiary hypogonadism. Historical hCG and hMG, both purified extractions from urine was used. In contemporary practice, however, purified urinary FSH and recombinant human FSH have replaced hMG in most countries [19, 28].

Pulsatile GnRH Therapy

In general, GnRH therapy is not frequently used since gonadotrophin therapy (see next) is successful in both secondary and tertiary hypogonadism. Pulsatile GnRH delivery is more expensive and demanding and not generally available except in the few centres with a research interest in GnRH physiology [29].

Treatment with GnRH requires subcutaneous pulsatile application using a portable pump via a butterfly needle placed in the abdominal wall and changed every 2 days. The dose usually ranges from 5 to 20 μg/120 min (corresponding to 100–400 ng/kg body weight/120 min). Low-dose pulsatile GnRH therapy (2 μg/150 min) may

be tried but is commonly not sufficient to elicit a pituitary response [30]. The induction of spermatogenesis is evidenced by the appearance of sperm in the ejaculate. Spermatogenesis (with counts ranging from 1.2–15.3 mill/ml) was achieved after an average of 4 months of treatment, as shown during GnRH substitution in patients with idiopathic (i.e. tertiary) HH or Kallmann's syndrome [16].

Similar results have been demonstrated in other studies [21, 31–33]. Despite low sperm counts, pregnancies can be achieved, the duration until conception being on average 6–7 months. During pulsatile GnRH treatment testicular size increased significantly, from an initial mean size of 6.8 ± 2.2 ml to 14.9 ± 3.2 ml after 5–12 months of treatment [16].

Generally, GnRH therapy is successful in inducing virilization and spermatogenesis in tertiary hypogonadism. However, a small subset of tertiary IHH men usually those who showed a complete absence of prior pubertal development (presumed GnRH receptor gene defects), does not achieve a satisfactory response to GnRH [34].

Treatment with hCG/Highly Purified Urinary Human FSH (Urinary-hFSH)

Improved purification methods have provided highly purified urinary FSH with enhanced specific activity in comparison to hMG (10 000 IU/mg of protein vs. 150 IU/mg of protein for hMG) [35].

In 14 prepubertal males with isolated HH or panhypopituitarism (i.e. secondary hypogonadism), complete virilization was achieved in all patients; in seven of eight patients willing to provide ejaculates spermatogenesis was achieved [36]. The subcutaneous application makes self-administration feasible: in 60 men with different forms of HH (16 with Kallmann syndrome, 19 with IHH, 25 with hypopituitarism) highly purified urinary FSH (150 IU thrice/week) and hCG (2500 IU twice/week) were administered. Results were comparable to other studies and the treatment was well tolerated [19, 37].

Treatment with hCG/Recombinant Human FSH (rhFSH)

rhFSH has advantages over urinary preparations in terms of purity, specific activity, consistent composition, and constant supply. Multiple dose pharmacokinetics showed an elimination half-life of 48 ± 5 h and demonstrated that serum FSH increases in a dose-dependent fashion. No intrinsic LH activity was detected [38–40].

Early case reports suggested rhFSH effectiveness in inducing spermatogenesis in HH [15, 41] in combination with hCG-treatment.

In a study of ten men with HH due to hypothalamic or pituitary disorders, combined therapy with recombinant FSH (150 IU s.c. thrice/week) and hCG (2000 IU 2–3 times/week), seven had evidence of spermatogenesis after a median of 6 months. Five achieved a sperm output of more than 1.5 mill/ml. Mean testicular volume increased by 4.2 ml. Three pregnancies were achieved during FSH treatment. Its efficacy is comparable to urinary FSH in restoring normal fertility in men with gonadotropin deficiency [19, 42].

A combined analysis of data from four studies using rhFSH/hCG to induce spermatogenesis in men with HH (n = 100) revealed that spermatogenesis could be achieved in 68 men after a maximal duration of treatment of 18 months [43].

Table 10.4.3. Modern modalities of gonadotropin substitution therapy in men to achieve spermatogenesis and maintain androgenicity

Substance	Form of application	Dosage
GnRH pulsatile	Subcutaneous by minipump	5–20 µg/Pulse every 2 h
or alternatively		
Human chorion-gonadotropin (hCG)	Subcutaneous or intramuscular	1500–3000 IE 2 2 to 3 times per week depending on testosterone levels achieved
in combination with		
Highly purified or recombinant FSH	Subcutaneous or intramuscular	150 IE 3 ml 3 times per week

hCG preparations are only approved for i.m. application, however, it long clinical experience that a s.c. administration of hCG-preparations is therapeutically very effective and safe. But this is, strictly speaking, an off-label use.

There is no evidence of any significant difference in terms of various clinical outcomes between GnRH and gonadotrophin therapy in hypogonadotropic patients seeking fertility. A study with 36 patients with tertiary hypogonadism (IHH or Kallmann syndrome), comparing GnRH (n = 18) with gonadotrophin (n = 18) therapy, reported no significant difference in effectiveness with respect to sperm counts. A 2-year comparison including 16 patients showed no advantage of either therapy concerning acceleration and/or enhancement of testicular growth, onset of spermatogenesis or increment of sperm output [31].

Conclusion

Gonadotropin substitution therapy is highly effective in men with HH in achieving fertility. See Table 10.4.3.

Practical Relevance

- In tertiary or secondary hypogonadism, fertility and androgenization of patients can be achieved by administration of gonadotropins. A rarely used alternative method is pulsatile GnRH application in tertiary hypogonadism.

REFERENCES

1. Albano C, Smitz J, Camus M, Bennink HC, Van Steirteghem AC, Devroey P. Pregnancy and birth in an in-vitro fertilization cycle after controlled ovarian stimulation in a woman with a history of allergic reaction to human menopausal gonadotrophin. *Hum Reprod*, 1996; **11**: 1632–4.
2. Barberi M, Ermini B, Morelli MB, Ermini M, Cecconi S, Canipari R. Follicular fluid hormonal profile and cumulus cell gene expression in controlled ovarian hyperstimulation with recombinant FSH: effects of recombinant LH administration. *J Assist Reprod Genet*, 2012; **29**: 1381–91.
3. Finkel DM, Phillips JL, Snyder PJ. Stimulation of spermatogenesis by gonadotropins in men with hypogonadotropic hypogonadism. *N Engl J Med*, 1985; **12**: 651–5.

4. Ferlin A, Vinanzi C, Selice R, Garolla A, Frigo AC, Foresta C. Toward a pharmacogenetic approach to male infertility: polymorphism of follicle-stimulating hormone beta-subunit promoter. *Fertil Steril*, 2011; **96**:1344–9.
5. Gromoll, J., Simoni, M., Nieschlag, E., 1996. An activating mutation of the follicle-stimulating hormone receptor autonomously sustains spermatogenesis in a hypophysectomized man. *J Clin Endocrinol Metab*, 1996; **81**: 1367–70.
6. Quaynor SD, Bosley ME, Duckworth CG, *et al*. Targeted next generation sequencing approach identifies eighteen new candidate genes in normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Mol Cell Endocrinol*, 2016; **437**: 86–96.
7. Blumenfeld Z, Frisch L, Conn PM. Gonadotropin-releasing hormone (GnRH) antibodies formation in hypogonadotropic azoospermic men treated with pulsatile GnRH-diagnosis and possible alternative treatment. *Fertil Steril*, 1988; **50**: 622–9.
8. Casarini L, Lispi M, Longobardi S, *et al*. LH and hCG action on the same receptor results in quantitatively and qualitatively different intracellular signalling. *PLoS One*, 2012; **7**: e46682.
9. Cassatella D, Howard SR, Acierno JS, *et al*. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*, 2018; **178**: 377–88.
10. Claustrat B, David L, Faure A, Francois R. Development of anti-human chorionic gonadotropin antibodies in patients with hypogonadotropic hypogonadism. A study of four patients. *J Clin Endocrinol Metab*, 1983; **57**: 1041–7.
11. de Roux N, Young J, Brailly-Tabard S, Misrahi M, Milgrom E, Schaison G. The same molecular defects of the gonadotropin-releasing hormone receptor determine a variable degree of hypogonadism in affected kindred. *J Clin Endocrinol Metab*, 1999; **84**: 567–72.
12. Efesoy O, Cayan S, Akbay E. The efficacy of recombinant human follicle-stimulating hormone in the treatment of various types of male-factor infertility at a single university hospital. *J Androl*, 2009; **30**: 679–84.
13. European Metrodin HP Study Group. Efficacy and safety of highly purified urinary follicle-stimulating hormone with human chorionic gonadotropin for treating men with isolated hypogonadotropic hypogonadism. *Fertil Steril*, 1998; **70**: 256–62.
14. Burris AS, Clark RV, Vantman DJ, Sherins RJ. A low sperm concentration does not preclude fertility in men with isolated hypogonadotropic hypogonadism after gonadotropin therapy. *Fertil Steril*, 1988; **50**: 343–7.
15. Kliesch S, Behre HM, Nieschlag E. Recombinant human follicle-stimulating hormone and human chorionic gonadotropin for induction of spermatogenesis in a hypogonadotropic male. *Fertil Steril*, 1995; **63**: 1326–8.
16. Büchter D, Behre HM, Kliesch S, Nieschlag E. Pulsatile GnRH or human chorionic gonadotropin:human menopausal gonadotropin as effective treatment for men with hypogonadotropic hypogonadism: a review of 42 cases. *Eur J Endocrinol*, 1998; **139**: 298–303.
17. Rohayem J, Sinthofen N, Nieschlag E, Kliesch S, Zitzmann M. Causes of hypogonadotropic hypogonadism predict response to gonadotropin substitution in adults. *Andrology*, 2016; **4**: 87–94.
18. Burris AS, Rodbard HW, Winters SJ, Sherins RJ. Gonadotropin therapy in men with isolated hypogonadotropic hypogonadism: the response to human chorionic gonadotropin is predicted by initial testicular size. *J Endocrinol Metab*, 1988; **66**: 1144–51.
19. Zitzmann M, Nieschlag E. Hormone substitution in male hypogonadism. *Mol Cell Endocrinol*, 2000; **161**: 73–88 (IF 2.369).
20. Saal W, Happ J, Cordes U, Baum RP, Schmidt M. Subcutaneous gonadotropin therapy in male patients with hypogonadotropic hypogonadism. *Fertil Steril*, 1991; **56**: 319–24.
21. Delemarre-van de Waal H. Induction of testicular growth and spermatogenesis by pulsatile, intravenous administration of gonadotrophin-releasing hormone in patients with hypogonadotrophic hypogonadism. *Clin Endocrinol*, 1993; **38**: 473–80.
22. Jones TH, Darne JF. Self-administered subcutaneous human menopausal gonadotropin for the stimulation of testicular growth and the initiation of spermatogenesis in hypogonadotropic hypogonadism. *Clin Endocrinol*, 1993; **38**: 203–8.
23. Kamischke A, Nieschlag E. Analysis of medical treatment of male infertility. *Hum Reprod*, 1999; **14** (Suppl. 1): 1–23.
24. Kamischke A, Behre HM, Bergmann M, Simoni M, Schäfer T, Nieschlag E. Recombinant human follicle stimulating hormone for treatment of male idiopathic infertility: a randomized, double-blind, placebo-controlled, clinical trial. *Hum Reprod*, 1998; **13**: 596–603.
25. Liu PY, Baker HW, Jayadev V, Zacharin M, Conway AJ, Handelsman DJ. Induction of spermatogenesis and fertility during gonadotropin treatment of gonadotropin-deficient infertile men: predictors of fertility outcome. *J Clin Endocrinol Metab*, 2009; **94**: 801–8.
26. Rohayem J, Hauffa BP, Zacharin M, Kliesch S, Zitzmann M, 'German Adolescent Hypogonadotropic Hypogonadism Study Group'. Testicular growth and spermatogenesis: new goals for pubertal hormone replacement in boys with hypogonadotropic hypogonadism? -a multicentre prospective study of hCG/rFSH treatment outcomes during adolescence. *Clin Endocrinol (Oxf)*, 2017; **86**: 75–87.
27. Raivio T, Falardeau J, Dwyer A, *et al*. Reversal of idiopathic hypogonadotropic hypogonadism. *N Engl J Med*, 2007; **357**: 863–73.
28. Levalle O, Zylbersztein C, Aszpis S, *et al*. Recombinant human follicle-stimulating hormone administration increases testosterone production in men, possibly by a Sertoli cell-secreted non-steroid factor. *J Clin Endocrinol Metab*, 1998; **83**: 3973–6.
29. Bals-Pratsch M, Knuth UA, Honigl W, Klein HM, Bergmann M, Nieschlag E. Pulsatile GnRH-therapy in oligozoospermic men does not improve seminal parameters despite decreased FSH levels. *Clin Endocrinol (Oxf)*, 1989; **30**: 549–60.
30. Happ J, Ditscheid W, Krause U. Pulsatile gonadotropin-releasing hormone therapy in male patients with Kallmann's syndrome or constitutional delay of puberty. *Fertil Steril*, 1985; **43**: 599–608.
31. Liu L, Banks SM, Barnes KM, Sherins RJ. Two-year comparison of testicular responses to pulsatile gonadotropin-releasing hormone and exogenous gonadotropins from the inception of therapy in men with isolated hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 1988; **67**: 1140–5.
32. Schopohl J. Pulsatile gonadotropin releasing hormone versus gonadotropin treatment of hypothalamic hypogonadism in males. *Hum Reprod*, 1993; **8** (Suppl. 2): 175–9.
33. Mao JF, Liu ZX, Nie M, *et al*. Pulsatile gonadotropin-releasing hormone therapy is associated with earlier spermatogenesis compared to combined gonadotropin therapy in patients with congenital hypogonadotropic hypogonadism. *Asian J Androl*, 2017; **19**: 680–5.

34. Pitteloud N, Hayes FJ, Dwyer A, Boepple PA, Lee H, Crowley WF Jr. Predictors of outcome of long-term GnRH therapy in men with idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2002; **87**: 4128–36.
35. Biffoni M, Marcucci I, Ythier A, Eshkol A. Effects of urinary gonadotrophin preparations on human in-vitro immune function. *Hum Reprod*, 1998; **13**: 2430–4.
36. Barrio R, de Luis D, Alonso M, Lamas A, Moreno JC. Induction of puberty with human chorionic gonadotropin and follicle-stimulating hormone in adolescent males with hypogonadotropic hypogonadism. Fertil Steril, 1999; **71**: 244–8.
37. Burgues S, Calderon MD. Subcutaneous self-administration of highly purified follicle stimulating hormone and human chorionic gonadotropin for the treatment of mal hypogonadotropic hypogonadism. Spanish Collaborative Group on Male Hypogonadotropic Hypogonadism. *Hum Reprod*, 1997; **12**: 980–6.
38. Mannaerts B, Fauser B, Lahlou N, *et al*. Serum hormone concentrations during treatment with multiple rising doses of recombinant follicle stimulating hormone (Puregon) in men with hypogonadotropic hypogonadism. *Fertil Steril*, 1996; **65**: 406–410.
39. Laml T, Obruca A, Fischl F, Huber FC. Recombinant luteinizing hormone in ovarian hyperstimulation after stimulation failure in normogonadotropic women. *Gynecol Endocrinol*, 1999; **13**: 98–103.
40. Lejeune H, Sanchez P, Chuzel F, Langlois D, Saez JM. Time-course effects of human recombinant luteinizing hormone on porcine Leydig cell specific differentiated functions. *Mol Cell Endocrinol*, 1998; **25**: 59–69.
41. Knuth UA, Honigl W, Bals-Pratsch M, Schleicher G, Nieschlag E. Treatment of severe oligospermia with human chorionic gonadotropin:human menopausal gonadotropin: a placebo-controlled, double blind trial. *J Clin Endocrinol Metab*, 1987; **65**: 1081–7.
42. Liu PY, Turner L, Rushford D, *et al*. Efficacy and safety of recombinant human follicle stimulating hormone (Gonal-F) with urinary human chorionic gonadotrophin for induction of spermatogenesis and fertility in gonadotrophin-deficient men. *Hum Reprod*, 1999; **14**: 1540–5.
43. Warne DW, Decosterd G, Okada H, Yano Y, Koide N, Howles CM. A combined analysis of data to identify predictive factors for spermatogenesis in men with hypogonadotropic hypogonadism treated with recombinant human follicle-stimulating hormone and human chorionic gonadotropin. *Fertil Steril*, 2009; **92**: 594–604.

FURTHER READING

Mougdal NR, Sairam MR. Is there a true requirement for follicle-stimulating hormone in promoting spermatogenesis and fertility in primates? *Hum Reprod*, 1998; **13**: 916–19.
Mougdal NR, Murthy GS, Prasanna Kumar KM, *et al*. Responsiveness of human male volunteers to immunization with ovine follicle stimulating hormone vaccine: results of a pilot study. *Hum Reprod*, 1997; **12**: 457–63.
Nieschlag E, Bernitz S, Topert M. Antigenicity of human chorionic gonadotrophin preparations in men. *Clin Endocrinol (Oxf)*, 1982; **16**: 483–8.
Nieschlag E, Simoni M, Gromoll J, Weinbauer GF. Role of FSH in the regulation of spermatogenesis: clinical aspects. *Clin Endocrinol*, 1999; **51**: 139–46.
Philip M, Arbelle JE, Segev Y, Parvari R. Male hypogonadism due to a mutation in the gene for the beta-subunit of follicle-stimulating hormone. *NEJM*, 1998; **338**: 1729–32.
Recombinant Human FSH Product Development Group. Recombinant follicle stimulating hormone: development of the first biotechnology product for the treatment of infertility. *Hum Reprod Update*, 1998; **4**: 862–81.
Redfearn A, Hughes EG, O'Connor M, Dolovich J. Delayed-type hypersensitivity to human gonadotropin: case report. *Fertil Steril*, 1995; **64**: 855–6.
Schaison G, Young J, Pholsena M, Nahoul K, Couzinet B. Failure of combined follicle-stimulating hormone-testosterone administration to initiate and:or maintain spermatogenesis in men with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 1993; **77**: 1545–9.
Singh M, Das SK, Suri S, Singh O, Talwar GP. Regain of fertility and normality of progeny born during below protective threshold antibody titers in women immunized with the HSDhCG vaccine. *Am J Reprod Immunol*, 1998; **39**: 395–8.
Sokol RZ, McClure RD, Peterson M, Swerdloff RS. Gonadotropin therapy failure secondary to human chorionic gonadotropin-induced antibodies. *J Clin Endocrinol Metab*, 1981; **52**: 929–32.
Talwar GP, Gupta SK. Antibody response against three epitopic domains on human chorionic gonadotropin (hCG) in women and rodents immunized with a b hCG-based immunocontraceptive vaccine. *J Clin Immunol*, 1994; **14**: 162–8.
Thau RB, Goldstein M, Yamamoto Y, Burrow GN, Phillips D, Bardin CW. Failure of gonadotropin therapy secondary to chorionic gonadotropin-induced antibodies. *J Clin Endocrinol Metab*, 1988; **66**: 862–7.
Tüttelmann F, Laan M, Grigorova M, Punab M, Sõber S, Gromoll J. Combined effects of the variants FSHB-211G>T and FSHR 2039A>G on male reproductive parameters. *J Clin Endocrinol Metab*, 2012; **97**: 3639–47.
Vicari E, Mongioi A, Calogero AE, *et al*. Therapy with human chorionic gonadotrophin alone induces spermatogenesis in men with isolated hypogonadotrophic hypogonadism—long-term follow-up. *Int J Androl*, 1992; **15**: 320–9.

10.4.4 Benefits of Testosterone Treatment

Shehzad Basaria and Thiago Gagliano-Jucá

Introduction

Male hypogonadism is a clinical syndrome that arises from disruption of the hypothalamic–pituitary–testicular axis resulting in failure to produce physiological concentrations of

testosterone and/or normal production of sperm [1]. Primary hypogonadism results from testicular dysfunction, while secondary hypogonadism arises due to disease or dysfunction of the hypothalamus and/or the pituitary. In primary hypogonadism, serum concentrations of gonadotropins are increased, testosterone concentrations are decreased (hypergonadotropic hypogonadism), and spermatogenesis is impaired, with impairment of spermatogenesis generally greater than that of testosterone production. In secondary hypogonadism, testicular testosterone secretion and spermatogenesis are similarly impaired, but concentrations of gonadotropins are low or inappropriately normal (hypogonadotropic hypogonadism) [2].

Men seek consultation in the clinics typically as a result of symptoms and signs related to low serum testosterone levels, although some men might present with infertility as their chief complaint. Testosterone is produced by the Leydig cells of the testes and is the predominant androgen in men; it is essential for virilization and maintenance of the male phenotype. As a result, prepubertal onset of androgen deficiency results in small testes and micropenis, high-pitched voice, eunuchoidal habitus, gynaecomastia, decreased body hair, and low libido [2]. If the onset is post-pubertal, size of the phallus, body proportions, and voice remain normal; these patients may present with gynaecomastia, reduced hair growth, low libido, and erectile dysfunction [1, 2]. Additionally, as testosterone has several non-reproductive functions, which include impact on bone mass, body composition, muscle strength, erythropoiesis, and mood, men with low testosterone may present with low bone mass, reduced muscle strength, unexplained anaemia, and decreased energy [1, 2].

Even though randomized trials of testosterone replacement in men with organic hypogonadism are few (it is unethical for these patients with profound androgen deficiency to be randomized to a placebo arm), agreement exists—on the basis of clinical experience and open-label trials—that testosterone replacement therapy (TRT) is associated with beneficial effects. This statement on accepted efficacy of TRT applies only to classic/organic hypogonadism, and must not be extrapolated to the age-related decline in testosterone levels, also known as 'late-onset hypogonadism' or 'age-related low testosterone'. Testosterone levels peak in the second and third decade of life and then gradually decline (in some men) with advancing age without a clear inflection point [3–6]. The age-related decline in testosterone levels has been associated with defects at all levels of the hypothalamic–pituitary–testicular axis, and the trajectory of this decline is influenced by body mass index (BMI), comorbid conditions, medications and genetic factors [7–9]. In the European Male Ageing Study (EMAS), three sexual symptoms (poor morning erection, low sexual desire, and erectile dysfunction) were associated with testosterone levels below 11 nmol/L (317 ng/dl) or free testosterone levels below 220 pmol/L (63 pg/ml) [6]. Men defined as having late-onset hypogonadism by these criteria tended to be older and have a higher BMI, lower muscle mass, lower bone mineral density (BMD), lower haemoglobin levels, and slower gait speed, which suggests that men with late-onset hypogonadism have end-organ deficits similar to those observed in patients with organic androgen deficiency [9, 10]. Nonetheless, randomized trials of TRT in men with late-onset hypogonadism have shown modest benefits at best. Furthermore, long-term safety of TRT, particularly its impact on cardiovascular disease and prostate cancer risk, remains unknown. Therefore, testosterone treatment for age-related low testosterone is not approved by the American Food and Drug Administration (FDA) and remains 'off-label'. Here we discuss the data on the efficacy of testosterone treatment in both organic hypogonadism (most studies included younger men but some also included middle-aged and older men with organic hypogonadism) and in men with purely age-related low testosterone (Figure 10.4.4.1 Online-only).

Benefits of Testosterone Replacement

Sexual Function

Hypogonadal men experience fewer sexual thoughts and spontaneous erections, and have lower overall sexual activity than eugonadal men [11]. However, hypogonadal and eugonadal men have similar erectile responses to visual erotic stimuli [11]. These observations have led to the generally accepted view that testosterone influences spontaneous but not sexually aroused erections, whereas its association with libido remains its most consistent effect.

Testosterone therapy in young hypogonadal men improves many aspects of sexual function, including overall sexual activity, sexual desire, intensity of sexual feelings, sexual fantasies, as well as the number and duration of night-time and spontaneous erections [11–15]. However, TRT does not improve ejaculatory function in hypogonadal men with ejaculatory dysfunction [16]. A meta-analysis of randomized trials showed that TRT improved sexual motivation, number of successful intercourses, erectile function scores, and overall sexual satisfaction in men with baseline serum testosterone concentrations lower than 12.0 nmol/L (356 ng/dl), whereas no effect was seen in men with higher (>12 nmol/L) serum testosterone concentrations [17].

Even though sexual symptoms are consistently associated with low serum testosterone levels in older men, intervention trials have reported inconsistent efficacy in this population. This might be because some trials enrolled men with low-normal or normal serum testosterone levels, and only a few studies investigated the effects of TRT in older men with sexual dysfunction and unequivocally low testosterone levels. Additionally, most earlier trials did not enrol participants based on rigorously defined criteria for sexual dysfunction, thus making it difficult to draw firm conclusions regarding the efficacy of TRT on sexual function. Nevertheless, positive effects of TRT on libido have been reported more consistently than those on erectile function in randomized trials in older men. Indeed, in men older than 60 years with age-related low testosterone, serum testosterone concentrations do not correlate strongly with erectile function, which suggests aetiologies other than testosterone deficiency may be responsible for erectile dysfunction [2]. More recently, however, the Testosterone Trials (TTrials), a coordinated set of seven trials involving 790 men aged 65 years or older with unequivocally low testosterone concentration at baseline (below 9.54 nmol/L (275 ng/dL)), randomized men with at least one manifestation of hypogonadism (decreased libido, difficulty walking, or low vitality) to transdermal testosterone gel or placebo gel treatment for 12 months [18]. The therapy-induced increase in serum testosterone concentration was associated with modest increases in sexual activity (Figure 10.4.4.2a), sexual desire and erectile function [18]. These findings were further confirmed in a recent meta-analysis of randomized, placebo-controlled testosterone trials in adult men with morning

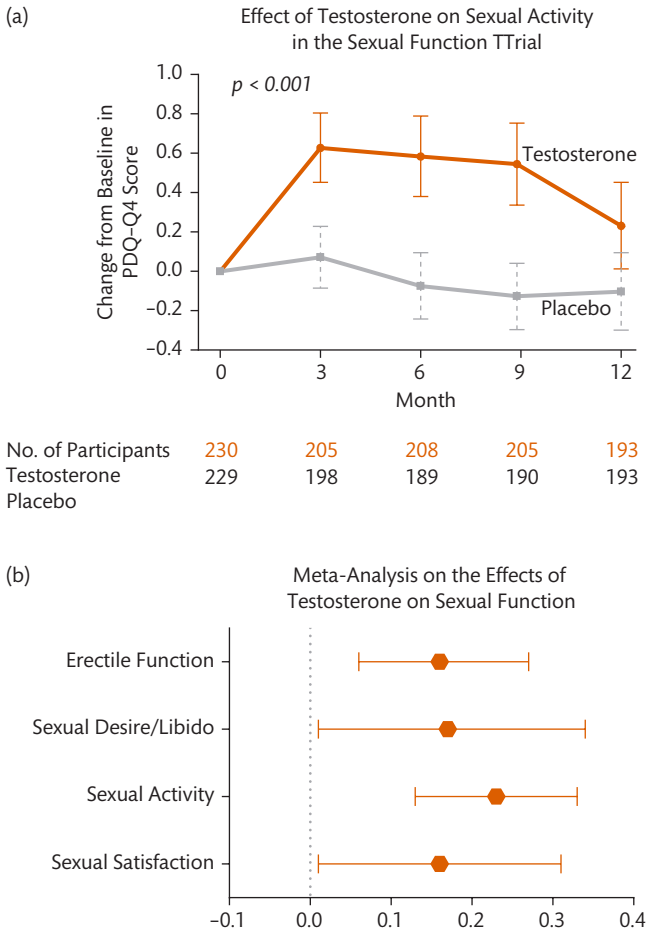


Figure 10.4.4.2 Effects of testosterone replacement on sexual function in hypogonadal men. Panel A shows change from baseline in overall sexual activity with testosterone treatment in older men assessed by the Psychosexual Daily Questionnaire (data presented as means; error bars are 95%CI). Panel B shows the effects of testosterone replacement on various aspects of sexual function summarized in a meta-analysis of randomized testosterone trials in men with low serum testosterone concentrations (≤300 ng/dl) and one or more symptoms of androgen deficiency (data displayed as standardized mean difference and 95%CI).

Panel A Source data from Snyder PJ *et al*. Effects of Testosterone Treatment in Older Men. *N Engl J Med*. 2016;374(7):611–24. Copyright © 2016, Massachusetts Medical Society. Panel B Source data from Ponce *et al., J Clin Endocrinol Metab*. 2018;103(5):1745–54.

testosterone levels below 10.4 nmol/L (300 ng/dl) and one or more symptoms or signs of hypogonadism [19] (Figure 10.4.4.2b).

When men present with predominant complaint of erectile dysfunction and their serum testosterone concentrations are only mildly reduced (generally from a non-organic aetiology), phosphodiesterase-5 inhibitors (PDE-5 inhibitors) might be the first line of therapy as they are more efficacious than TRT based on the magnitude of improvement as assessed by validated questionnaires. In small clinical trials, patients with erectile dysfunction and low serum testosterone concentrations who do not respond to monotherapy with PDE-5 inhibitors, addition of testosterone to PDE-5 inhibitors have shown some benefit [20]. However, in a large randomized, controlled trial, no additional benefit was seen when TRT was added to participants after their dose of sildenafil was carefully optimized [21] (Figure 10.4.4.3 Online-only).

Muscle Mass, Muscle Strength, and Physical Function

Testosterone is a potent anabolic hormone which plays a major role in regulating body composition, maximal voluntary muscle strength and muscle power [22, 23]. In epidemiological studies, low testosterone concentrations are associated with higher fat mass, lower lean body mass, and lower muscle strength [24]. Additionally, men with low testosterone levels are more prone to having worse physical performance than men with normal testosterone levels [24, 25], and are more predisposed to become frail and experience falls [5, 26].

The most noticeable and consistent effect of TRT in hypogonadal men is the increase in muscle mass and maximal voluntary strength [27–29]. In a trial in young hypogonadal men, TRT for 10 weeks induced a mean increase in lean mass of 5.0 kg (8.9%), as well as improvements in chest press (22%) and leg press strength (45%) [27] (Figure 10.4.4.4a). In studies that included adult men of all ages (including older men), modest increments in lean mass (1–3 kg) and muscle strength have been observed [29–31]. Similarly, modest increases in lean mass and muscle strength were seen in trials of TRT in intermediate-frail and frail older men as well as in older men with mobility limitations [32–34].

Despite the consistent observation of improvements in lean mass and maximal voluntary muscle strength, improvements in measures of physical function have been inconsistent or of a small magnitude [30, 32–35]. In a large trial of community-dwelling older men with low to low-normal serum testosterone, TRT for 3 years improved loaded and unloaded stair climbing power [30]. Loaded stair climbing power also improved in a 6-month trial of TRT in older men with mobility limitation and low testosterone [34] (Figure 10.4.4.4b), with a greater proportion of men receiving testosterone (compared with placebo) demonstrating improvement that was considered to be clinically meaningful [33]. However, there was no significant improvement in gait speed [34]. Similarly, clinical trials in older men have also not demonstrated improvements in other measures of physical function, such as the sit-to-stand test, the get up and go test, the 6-minute walk test and the Short Physical Performance Battery score [32].

The conflicting findings relating to the efficacy of TRT in improving measures of physical function in older men were further highlighted by the findings of the TTrials. In the pre-determined analysis of the cohort enrolled in the *Physical Function Trial* of the TTrials (387 men with self-reported difficulty walking or climbing stairs ***plus*** a gait speed <1.2 m/s on the 6-minute walk test), TRT did not result in significant improvements in the primary endpoint of walking distance in men treated with testosterone for 12 months compared to placebo [18, 35] (Figure 10.4.4.4c). In other words, in carefully selected men who had both unequivocal low serum testosterone concentrations and physical dysfunction, TRT was not beneficial. However, a sub-analysis that included all randomized men (including those who did not have physical dysfunction) demonstrated significant improvements in the 6-minute walk test [18]. Additionally, men treated with testosterone were more likely to report that their walking ability had improved [18].

Another functional parameter that has been evaluated in some testosterone trials is aerobic capacity (VO_2peak). In the TEAAM trial, TRT for 3 years in men 60 years and older with low to low-normal serum testosterone concentrations (below 13.9 nmol/L; 400 ng/dl) attenuated the expected age-related decline in aerobic

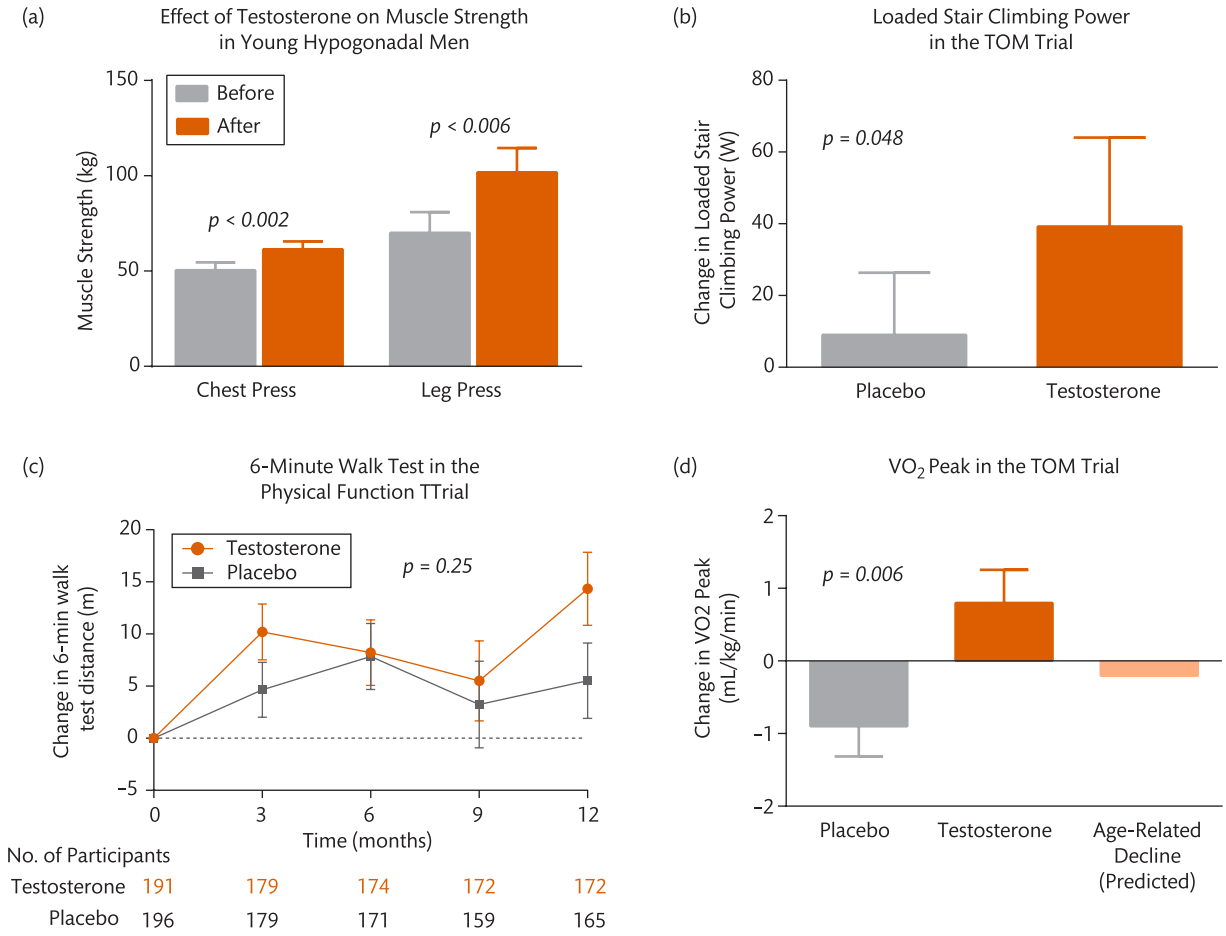


Figure 10.4.4.4 Effects of testosterone administration on measures of muscle performance and physical function in interventional trials. Panel A shows the effects of testosterone replacement for 12 weeks on voluntary maximum strength in the chest press and leg press exercises in young hypogonadal men (bars are means and error bars SEM). Panel B shows the effects of testosterone treatment for 6 months on loaded stair climbing power in older men with mobility limitations (data displayed as means and error bars are 95% CI). Panel C shows the change in walk distance in the 6-minute walk test in older men with mobility limitations participating in the Physical Function Trial of the TTrials (data displayed as means and error bars are 95% CI). Panel D shows mean changes in VO_2peak after 6 months of testosterone treatment or placebo in mobility-limited older men (error bars are SEM). The anticipated age-related decline in the VO_2peak for the 6-month period is shown for reference.

Data in Panel A derived from Bhasin *et al., J Clin Endocrinol Metab*. 1997;82(2):407–13. Data in Panel B derived from Basaria *et al. N Engl J Med*. 2010;363(2):109–22. Panel C adapted with permission from Bhasin *et al., Lancet Diabetes Endocrinol*. 2018;6(11):879–890. Panel D adapted with permission from Storer *et al., J Clin Endocrinol Metab*. 2016;101(6):2562–9.

capacity [36]. However, this effect was modest (a 0.91 ml/kg/min difference from placebo) and its clinical meaningfulness remains uncertain. Similar findings were also seen in older men with mobility limitations [37] (Figure 10.4.4.4d).

The reasons why the gains in muscle mass and muscle power consistently demonstrated in clinical trials of TRT have not translated into improvements in physical function remain unclear. It is conceivable that adjunctive physical exercise training, together with TRT may be necessary to achieve more meaningful improvements in functional performance; this avenue deserves further investigation. In the interim, TRT for the sole purpose of improving physical performance is not recommended.

Bone Mineral Density and Bone Quality

Testosterone has anabolic effects on the male skeleton; it increases bone mass and strength via its direct action on the androgen receptor and indirectly through its conversion to oestradiol [38, 39]. During puberty, increase in the secretion of gonadal steroids stimulate osteoblast activity, markedly increasing BMD [40]. Pre-pubertal androgen deficiency is associated with reduced cortical and trabecular bone mass [40, 41]; similarly, compared to eugonadal men, hypogonadal men also have lower cortical and trabecular bone mass [41]. Additionally, lower bioavailable testosterone levels have also been associated with lower volumetric BMD, bone geometry and bone quality in older men [42]. Indeed, suppression of testosterone production with androgen deprivation therapy for prostate cancer is associated with a rapid loss of bone mass [43, 44].

Testosterone therapy in young, hypogonadal men is associated with improvement in vertebral BMD [31, 45–47]. Small trials in men with unequivocal organic hypogonadism have shown that TRT is associated with an increase of 5–8% in spinal and 4% in hip BMD, with maximum improvement observed by 24 months [29, 31], even though normalization of BMD may not be achieved in this timeframe [31]. Nonetheless, BMD in some hypogonadal men continues to increase even after long-term testosterone treatment [45], supporting a sustained beneficial role of TRT on male skeleton.

The age-related decline in sex hormones has been associated with reduction in BMD and increased risk of osteoporotic fractures [42, 48]. However, earlier clinical trials in this patient population reported inconsistent effects of TRT on BMD and bone quality. A meta-analysis of randomized trials found a significantly greater increase in lumbar spine but not femoral BMD in older men receiving intramuscular testosterone compared to placebo (Figure 10.4.4.5a); interestingly use of transdermal testosterone did not show benefits [49]. More recently, in the Bone Trial of the TTrials, TRT in men with age-related low testosterone showed an increase in volumetric BMD and estimated bone strength on quantitative computerized tomography scan compared to men receiving placebo [50], however, these effects appear modest when assessed with dual energy X-ray absorptiometry (DEXA) scan (Figure 10.4.4.5b), a tool more likely to be utilized by clinicians.

To date, no randomized trial has evaluated the effects of TRT on fracture rate. Future studies should be powered to determine whether improvements in BMD from TRT translate into reduced fracture risk. In the interim, bone-specific osteoporosis therapy with agents that have proven antifracture efficacy should remain the main therapeutic approach for men who are at a high risk for fracture [51].

Mood

Epidemiological studies have demonstrated an association between lower testosterone levels and chronic low-grade depression in men [52, 53]. However, in young and middle-age men with refractory depression, placebo-controlled trials of testosterone have not consistently shown a beneficial effect of testosterone, either in men with total testosterone concentrations below 12.1 nmol/L (350 ng/dl) [54, 55] or in men with normal testosterone levels [56, 57]. In a previous meta-analysis, TRT in hypogonadal men resulted in greater improvements in depression scores than did placebo, an effect that was not seen in eugonadal men [58]. However, in a recent meta-analysis including 27 randomized trials of testosterone administration in hypogonadal or eugonadal men reporting depressive symptoms (determined with psychometrically validated depression scales), testosterone treatment was associated with a significant reduction in depressive symptoms compared to placebo [59]. It is not clear, however, if these improvements are clinically relevant. Importantly, one should not conclude that testosterone treatment induces remission of major depression or that it might improve the response to antidepressants in these patients.

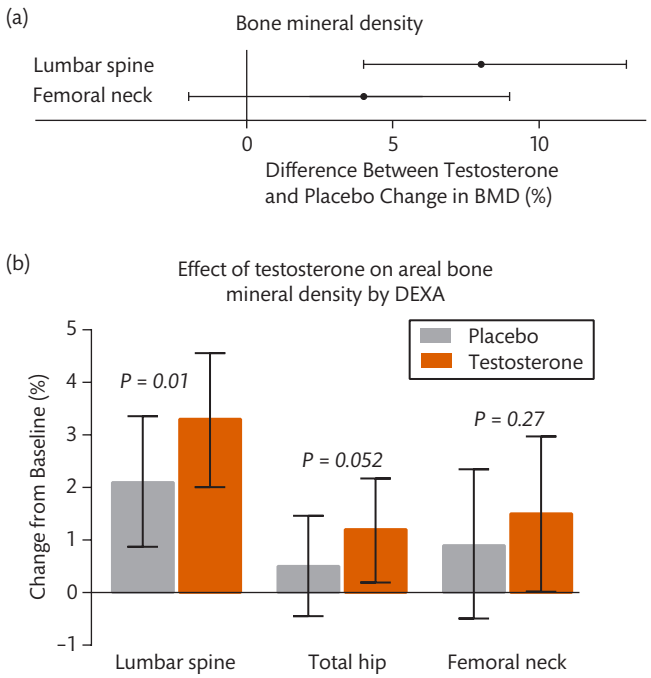


Figure 10.4.4.5 The effects of testosterone therapy on bone health in intervention trials. Panel A shows the effects of testosterone therapy on lumbar and femoral bone mineral density in a meta-analysis of randomized trials. Panel B shows the effects of testosterone replacement for 12 months on bone mineral density of the spine, hip, and femur, as assessed by DEXA scan.

Data in Panel A derived from a meta-analysis by Tracz *et al., J Clin Endocrinol Metab.* 2006;91(6):2011–16. Panel B reproduced with permission from Snyder *et al. JAMA Intern Med.* 2017;177(4):471–9. Copyright © 2017, American Medical Association.

Vitality

In older hypogonadal men with sexual dysfunction, decreased vitality, and/or mobility limitation, higher endogenous testosterone levels have not been associated with improved vitality or depression scores. In the *Vitality Trial* of the TTrials [18], TRT for 12 months did not improve vitality in older men who were carefully selected for low vitality. However, men receiving testosterone had small but significant improvements in mood and lower severity of depression. Other randomized controlled trials in older men have failed to show a clear benefit on vitality and health-related quality of life with testosterone therapy [60, 61]. However, in a randomized trial of adult hypogonadal men (mean age 55 years) with decreased energy or decreased sex drive, TRT for 9 months significantly improved energy assessed by a new questionnaire known as hypogonadism energy diary (HED) [62]. The reasons for these conflicting findings are not clear, but it is conceivable that the currently available tools (questionnaires) are not able to capture changes in energy in a meaningful way.

Cognition

Cohort studies have reported weak associations of testosterone levels with measures of visuospatial cognition and verbal fluency [63], but the findings have not been consistent [64]. Other studies have reported a complex relationship between serum androgen levels and spatial ability [65, 66]. A recent study found an association between the age-related decline in serum testosterone levels with impairment in cognitive function [67]. Additionally, in a meta-analysis of prospective cohort studies of older men, low serum testosterone was associated with increased risk of being diagnosed with Alzheimer's disease [68].

Earlier clinical trials evaluating the impact of testosterone therapy on cognitive function have been small and not well-designed, and reported mixed findings, either in cognitively healthy older men [69, 70] or in older men with impaired cognition or Alzheimer's disease [71, 72]. More recently, a secondary analysis of the TEAAM trial evaluated the effects of testosterone administration for 3 years on multiple domains of cognitive function in a large cohort of cognitively healthy older men; testosterone therapy was **not** associated with improved cognitive function [73]. These findings are similar to those reported by the *Cognitive Function Trial* of the TTrials, in which TRT for 12 months in hypogonadal older men was *not* associated with improvements in any domain of cognitive function [74] (Figure 10.4.4.6).

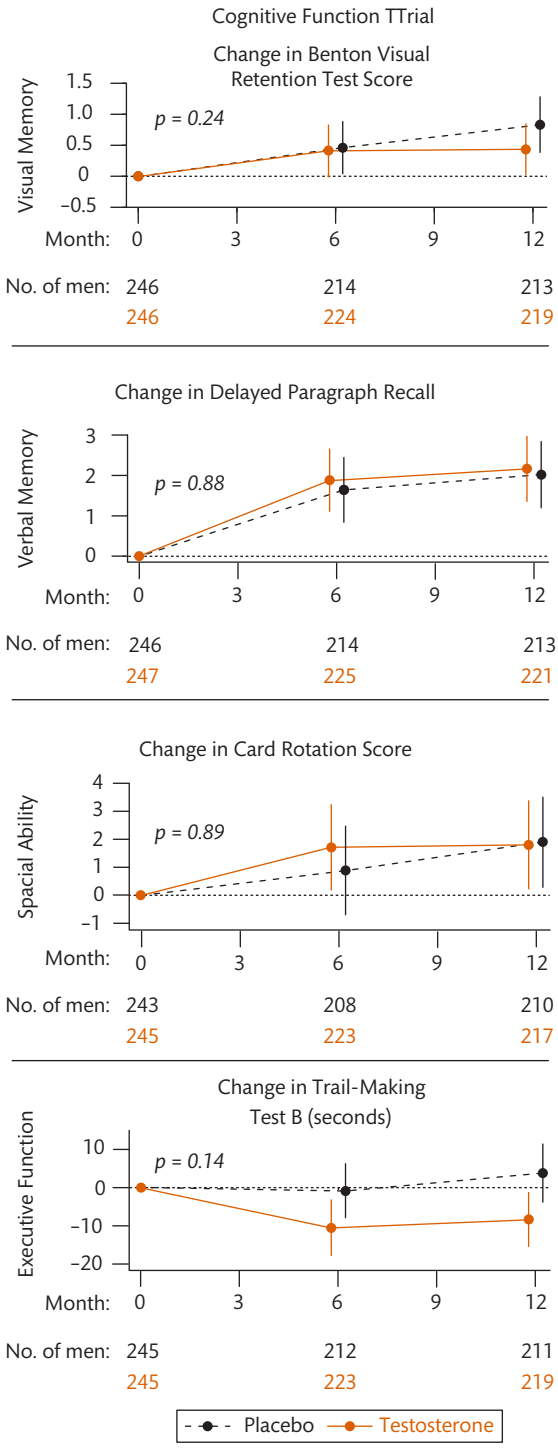


Figure 10.4.4.6 Effects of testosterone therapy on cognitive function in older men. Panels show adjusted mean change from baseline at 6 and 12 months for men with age-associated memory impairment by treatment group in cognition domains in the Cognitive Function Trial of the TTrials.

Source data from Resnick SM et al. Testosterone Treatment and Cognitive Function in Older Men With Low Testosterone and Age-Associated Memory Impairment. JAMA. 2017;317(7):717–727. Copyright © 2017, American Medical Association.

Based on the cumulative data, the current evidence suggests that testosterone therapy does not benefit cognitive function. In fact, TRT with supraphysiological doses may even have a negative impact on cognitive function [75]. These data suggest that testosterone therapy should not be initiated in men solely for the purposes of improving cognition.

Anaemia

Testosterone plays an important role in erythropoiesis. It stimulates erythropoiesis via multiple mechanisms which include (i) increasing iron availability [76–78] via suppression of hepatic secretion of hepcidin [79, 80], the liver-derived master iron regulatory peptide hormone that controls systemic iron homeostasis [81]; (ii) stimulation of erythroid progenitor cells and; (iii) stimulation of erythropoietin (EPO) secretion [77, 82]. Indeed, before the discovery of EPO, androgens were widely used in the treatment of anaemias of various aetiologies [77]. Additionally, erythrocytosis is the most common adverse event observed in men receiving TRT [1] and anaemia is a common adverse effect of androgen deprivation therapy in men with prostate cancer [83]. Moreover, lower testosterone levels are associated with lower haematocrit in men [84] and an increased risk of anaemia [85].

In the TTrials, TRT significantly increased haematocrit and haemoglobin levels in men with anaemia from known causes as well as in men with unexplained anaemia [86]. Furthermore, changes in haemoglobin concentrations in the testosterone group were associated with changes in walking speed [35]. In the TEAAM trial, testosterone-induced attenuation of the age-related decline in aerobic capacity (VO_2peak) was also associated with changes in haemoglobin [36]. These observations highlight the importance of haemoglobin concentrations to cardiovascular capacity and physical function and suggest that testosterone-induced increments in haemoglobin contribute to improvement in some aspects of quality of life in hypogonadal men. However, it should be noted that anaemia alone is not an approved indication to initiate TRT.

Metabolic Health

A large body of evidence suggests that testosterone plays an important role in the maintenance of metabolic health in men. Population studies have shown that higher endogenous serum testosterone concentrations are associated with a lower risk of metabolic syndrome and diabetes [87–90]. Indeed, initiation of androgen deprivation therapy for prostate cancer is associated with worsening of metabolic parameters and increased risk of developing metabolic syndrome and diabetes [91–96]. Similarly, acute interruption of TRT in hypogonadal men worsens insulin sensitivity [97]. Additionally, endogenous testosterone levels were positively associated with insulin sensitivity assessed with hyperinsulinaemic-euglycemic clamp [98]. Data suggests that adiposity modulates the beneficial role of TRT on insulin sensitivity. In a prospective cohort study, men with total testosterone concentration in the lower quartiles had an increased risk of developing diabetes [90]; however, this higher risk was dependent on waist circumference. This finding was later corroborated by a cross-sectional study [99]. Conversely, the association of endogenous testosterone concentrations and risk of incident diabetes was independent of adiposity in other cross-sectional studies [89], suggesting that low testosterone is itself a risk factor for metabolic dysfunction.

Notwithstanding the protective effect of **endogenous** testosterone on male metabolic health reported in population studies, large randomized trials of TRT have reached conflicting conclusions. In the TTrials, TRT for 12 months in older men resulted in modest

improvements in fasting insulin and homeostasis model assessment of insulin resistance (HOMA-IR) compared to placebo [100]. To the contrary, in a trial in older non-diabetic men with low to low-normal serum testosterone with an intervention duration of 3 years, testosterone treatment did not improve insulin sensitivity (assessed by the octreotide insulin suppression test) compared to placebo [101].

Findings from randomized trials evaluating the effects of TRT on glucose metabolism in men **with type 2 diabetes** and low serum testosterone are also conflicting. In a study of diabetic men aged 35 to 70 years with HbA1c ≤8.5% and baseline testosterone ≤346 ng/dl, testosterone treatment for 40 weeks did not improve insulin sensitivity (assessed by HOMA-IR) nor HbA1c compared to placebo [102]. Conversely, in a small trial of diabetic men aged 30 to 65 years with HbA1c ≤8% and hypogonadotropic hypogonadism, TRT for 24 weeks significantly improved insulin sensitivity as assessed by hyperinsulinaemic-euglycemic clamp [103].

In summary, despite the robust epidemiological data supporting a positive impact of **endogenous** testosterone level on metabolic parameters, trials of TRT have yielded conflicting results both in diabetic and in community-dwelling older non-diabetic men. Therefore, the current evidence does not support initiating TRT solely with the intent of improving glycaemic control.

Conclusion

Testosterone therapy in young hypogonadal men with known disease of hypothalamic-pituitary-testicular axis is generally beneficial and safe, and improves several health outcomes. However, the benefits of testosterone therapy in older men with age-related low testosterone levels are modest and mainly limited to sexual function. Despite testosterone-induced increases in lean mass in older men, improvements in physical function have been small and inconsistent. Additionally, even though testosterone treatment has been demonstrated to consistently increase bone density, no trials have evaluated the effects of TRT on fracture rate; TRT is therefore not recommended solely for the purposes of improving physical function or bone health in this population. Similarly, randomized trials in older men evaluating mood and cognition have yielded mixed results. Moreover, the long-term safety of TRT in men with age-related low testosterone has not been determined. Therefore, testosterone therapy is mainly indicated in men with **organic hypogonadism** (most of these men are young but may include men of all ages); as for men with age-related low testosterone levels, TRT might be indicated for **select** men who have sexual dysfunction.

REFERENCES

1. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
2. Basaria S. Male hypogonadism. *Lancet*, 2014; **383**: 1250–63.
3. Bhasin S, Pencina M, Jasuja GK, *et al*. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. *J Clin Endocrinol Metab*, 2011; **96**: 2430–9.
4. Feldman HA, Longcope C, Derby CA, *et al*. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. *J Clin Endocrinol Metab*, 2002; **87**: 589–98.
5. Yeap BB, Alfonso H, Chubb SA, *et al*. Reference ranges and determinants of testosterone, dihydrotestosterone, and estradiol levels measured using liquid chromatography-tandem mass spectrometry in a population-based cohort of older men. *J Clin Endocrinol Metab*, 2012; **97**: 4030–9.
6. Wu FC, Tajar A, Beynon JM, *et al*. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*, 2010; **363**: 123–35.
7. Mohr BA, Bhasin S, Link CL, O'Donnell AB, McKinlay JB. The effect of changes in adiposity on testosterone levels in older men: longitudinal results from the Massachusetts Male Aging Study. *Eur J Endocrinol*, 2006; **155**: 443–52.
8. Wu FC, Tajar A, Pye SR, *et al*. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metab*, 2008; **93**: 2737–45.
9. Tajar A, Huhtaniemi IT, O'Neill TW, *et al*. Characteristics of androgen deficiency in late-onset hypogonadism: results from the European Male Aging Study (EMAS). *J Clin Endocrinol Metab*, 2012; **97**: 1508–16.
10. Tajar A, Forti G, O'Neill TW, *et al*. Characteristics of secondary, primary, and compensated hypogonadism in aging men: evidence from the European Male Ageing Study. *J Clin Endocrinol Metab*, 2010; **95**: 1810–18.
11. Kwan M, Greenleaf WJ, Mann J, Crapo L, Davidson JM. The nature of androgen action on male sexuality: a combined laboratory-self-report study on hypogonadal men. *J Clin Endocrinol Metab*, 1983; **57**: 557–62.
12. Alexander GM, Swerdloff RS, Wang C, *et al*. Androgen-behavior correlations in hypogonadal men and eugonadal men. I. Mood and response to auditory sexual stimuli. *Horm Behav*, 1997; **31**: 110–19.
13. Bancroft J, Wu FC. Changes in erectile responsiveness during androgen replacement therapy. *Arch Sex Behav*, 1983; **12**: 59–66.
14. Boloña ER, Uraga MV, Haddad RM, *et al*. Testosterone use in men with sexual dysfunction: a systematic review and meta-analysis of randomized placebo-controlled trials. *Mayo Clin Proc*, 2007; **82**: 20–8.
15. Hirshkowitz M, Moore CA, O'Connor S, Bellamy M, Cunningham GR. Androgen and sleep-related erections. *J Psychosom Res*, 1997; **42**: 541–6.
16. Paduch DA, Polzer PK, Ni X, Basaria S. Testosterone replacement in androgen-deficient men with ejaculatory dysfunction: a randomized controlled trial. *J Clin Endocrinol Metab*, 2015; **100**: 2956–62.
17. Isidori AM, Giannetta E, Gianfrilli D, *et al*. Effects of testosterone on sexual function in men: results of a meta-analysis. *Clin Endocrinol (Oxf)*, 2005; **63**: 381–94.
18. Snyder PJ, Bhasin S, Cunningham GR, *et al*. Effects of testosterone treatment in older men. *N Engl J Med*, 2016; **374**: 611–24.
19. Ponce OJ, Spencer-Bonilla G, Alvarez-Villalobos N, *et al*. The efficacy and adverse events of testosterone replacement therapy in hypogonadal men: a systematic review and meta-analysis of randomized, placebo-controlled trials. *J Clin Endocrinol Metab*, 2018.
20. Shabsigh R, Kaufman JM, Steidle C, Padma-Nathan H. Randomized study of testosterone gel as adjunctive therapy

to sildenafil in hypogonadal men with erectile dysfunction who do not respond to sildenafil alone. *J Urol*, 2008; **179**: S97–102.
21. Spitzer M, Basaria S, Travison TG, *et al.* Effect of testosterone replacement on response to sildenafil citrate in men with erectile dysfunction: a parallel, randomized trial. *Ann Intern Med*, 2012; **157**: 681–91.
22. Bhasin S, Storer TW, Berman N, *et al.* The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. *N Engl J Med*, 1996; **335**: 1–7.
23. Bhasin S, Woodhouse L, Casaburi R, *et al.* Testosterone dose-response relationships in healthy young men. *Am J Physiol Endocrinol Metab*, 2001; **281**: E1172–81.
24. Schaap LA, Pluijm SM, Deeg DJ, *et al.* Low testosterone levels and decline in physical performance and muscle strength in older men: findings from two prospective cohort studies. *Clin Endocrinol (Oxf)*, 2008; **68**: 42–50.
25. O'Donnell AB, Travison TG, Harris SS, Tenover JL, McKinlay JB. Testosterone, dehydroepiandrosterone, and physical performance in older men: results from the Massachusetts Male Aging Study. *J Clin Endocrinol Metab*, 2006; **91**: 425–31.
26. Cawthon PM, Ensrud KE, Laughlin GA, *et al.* Sex hormones and frailty in older men: the osteoporotic fractures in men (MrOS) study. *J Clin Endocrinol Metab*, 2009; **94**: 3806–15.
27. Bhasin S, Storer TW, Berman N, *et al.* Testosterone replacement increases fat-free mass and muscle size in hypogonadal men. *J Clin Endocrinol Metab*, 1997; **82**: 407–13.
28. Brodsky IG, Balagopal P, Nair KS. Effects of testosterone replacement on muscle mass and muscle protein synthesis in hypogonadal men—a clinical research center study. *J Clin Endocrinol Metab*, 1996; **81**: 3469–75.
29. Snyder PJ, Peachey H, Berlin JA, *et al.* Effects of testosterone replacement in hypogonadal men. *J Clin Endocrinol Metab*, 2000; **85**: 2670–7.
30. Storer TW, Basaria S, Traustadottir T, *et al.* Effects of testosterone supplementation for 3 years on muscle performance and physical function in older men. *J Clin Endocrinol Metab*, 2017; **102**: 583–93.
31. Katznelson L, Finkelstein JS, Schoenfeld DA, Rosenthal DI, Anderson EJ, Klibanski A. Increase in bone density and lean body mass during testosterone administration in men with acquired hypogonadism. *J Clin Endocrinol Metab*, 1996; **81**: 4358–65.
32. Srinivas-Shankar U, Roberts SA, Connolly MJ, *et al.* Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2010; **95**: 639–50.
33. Travison TG, Basaria S, Storer TW, *et al.* Clinical meaningfulness of the changes in muscle performance and physical function associated with testosterone administration in older men with mobility limitation. *J Gerontol A Biol Sci Med Sci*, 2011; **66**: 1090–9.
34. Basaria S, Coviello AD, Travison TG, *et al.* Adverse events associated with testosterone administration. *N Engl J Med*, 2010; **363**: 109–22.
35. Bhasin S, Ellenberg SS, Storer TW, *et al.* Effect of testosterone replacement on measures of mobility in older men with mobility limitation and low testosterone concentrations: secondary analyses of the Testosterone Trials. *Lancet Diabetes Endocrinol*, 2018; **6**: 879–90.
36. Traustadottir T, Harman SM, Tsitouras P, *et al.* Long-term testosterone supplementation in older men attenuates age-related decline in aerobic capacity. *J Clin Endocrinol Metab*, 2018; **103**: 2861–9.
37. Storer TW, Bhasin S1, Travison TG, *et al.* Testosterone attenuates age-related fall in aerobic function in mobility limited older men with low testosterone. *J Clin Endocrinol Metab*, 2016; **101**: 2562–9.
38. Finkelstein JS, Lee H, Leder BZ, *et al.* Gonadal steroid-dependent effects on bone turnover and bone mineral density in men. *J Clin Invest*, 2016; **126**: 1114–25.
39. Vanderschueren D, Laurent MR, Claessens F, *et al.* Sex steroid actions in male bone. *Endocr Rev*, 2014; **35**: 906–60.
40. Bonjour JP, Theintz G, Buchs B, Slosman D, Rizzoli R. Critical years and stages of puberty for spinal and femoral bone mass accumulation during adolescence. *J Clin Endocrinol Metab*, 1991; **73**: 555–63.
41. Finkelstein JS, Klibanski A, Neer RM, *et al.* Osteoporosis in men with idiopathic hypogonadotropic hypogonadism. *Ann Intern Med*, 1987; **106**: 354–61.
42. Khosla S, Melton LJ 3rd, Robb RA, *et al.* Relationship of volumetric BMD and structural parameters at different skeletal sites to sex steroid levels in men. *J Bone Miner Res*, 2005; **20**: 730–40.
43. Basaria S, Lieb J 2nd, Tang AM, *et al.* Long-term effects of androgen deprivation therapy in prostate cancer patients. *Clin Endocrinol (Oxf)*, 2002; **56**: 779–86.
44. Greenspan SL, Coates P, Sereika SM, Nelson JB, Trump DL, Resnick NM. Bone loss after initiation of androgen deprivation therapy in patients with prostate cancer. *J Clin Endocrinol Metab*, 2005; **90**: 6410–17.
45. Behre HM, Kliesch S, Leifke E, Link TM, Nieschlag E. Long-term effect of testosterone therapy on bone mineral density in hypogonadal men. *J Clin Endocrinol Metab*, 1997; **82**: 2386–90.
46. Leifke E, Körner HC, Link TM, Behre HM, Peters PE, Nieschlag E. Effects of testosterone replacement therapy on cortical and trabecular bone mineral density, vertebral body area and paraspinal muscle area in hypogonadal men. *Eur J Endocrinol*, 1998; **138**: 51–8.
47. Wang C, Swerdloff RS, Iranmanesh A, *et al.* Effects of transdermal testosterone gel on bone turnover markers and bone mineral density in hypogonadal men. *Clin Endocrinol (Oxf)*, 2001; **54**: 739–50.
48. Mellstrom D, Johnell O, Ljunggren O, *et al.* Free testosterone is an independent predictor of BMD and prevalent fractures in elderly men: MrOS Sweden. *J Bone Miner Res*, 2006; **21**: 529–35.
49. Tracz MJ, Sideras K, Boloña ER, *et al.* Testosterone use in men and its effects on bone health. A systematic review and meta-analysis of randomized placebo-controlled trials. *J Clin Endocrinol Metab*, 2006; **91**: 2011–16.
50. Snyder PJ, Kopperdahl DL, Stephens-Shields AJ, *et al.* Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone: a controlled clinical trial. *JAMA Intern Med* 2017; **177**: 471–9.
51. Watts NB, Adler RA, Bilezikian JP, *et al.* Osteoporosis in men: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2012; **97**: 1802–22.
52. Seidman SN, Araujo AB, Roose SP, *et al.* Low testosterone levels in elderly men with dysthymic disorder. *Am J Psychiatry*, 2002; **159**: 456–9.
53. Shores MM, Sloan KL, Matsumoto AM, Moceri VM, Felker B, Kivlahan DR. Increased incidence of diagnosed depressive

illness in hypogonadal older men. *Arch Gen Psychiatry*, 2004; **61**: 162–7.
54. Seidman SN, Spatz E, Rizzo C, Roose SP. Testosterone replacement therapy for hypogonadal men with major depressive disorder: a randomized, placebo-controlled clinical trial. *J Clin Psychiatry*, 2001; **62**: 406–12.
55. Pope HG, Jr., Cohane GH, Kanayama G, Siegel AJ, Hudson JI. Testosterone gel supplementation for men with refractory depression: a randomized, placebo-controlled trial. *Am J Psychiatry*, 2003; **160**: 105–11.
56. Wang C, Alexander G, Berman N, *et al*. Testosterone replacement therapy improves mood in hypogonadal men—a clinical research center study. *J Clin Endocrinol Metab*, 1996; **81**: 3578–83.
57. Seidman SN, Miyazaki M, Roose SP. Intramuscular testosterone supplementation to selective serotonin reuptake inhibitor in treatment-resistant depressed men: randomized placebo-controlled clinical trial. *J Clin Psychopharmacol*, 2005; **25**: 584–8.
58. Zarrouf FA, Artz S, Griffith J, Sirbu C, Kommor M. Testosterone and depression: systematic review and meta-analysis. *J Psychiatr Pract*, 2009; **15**: 289–305.
59. Walther A, Breidenstein J, Miller R. Association of testosterone treatment with alleviation of depressive symptoms in men: a systematic review and meta-analysis. *JAMA Psychiatry*, 2018; **76**: 31–40.
60. Bhasin S, Apovian CM, Travison TG, *et al*. Effect of protein intake on lean body mass in functionally limited older men: a randomized clinical trial. *JAMA Intern Med*, 2018; **178**: 530–41.
61. Basaria S, Harman SM, Travison TG, *et al*. Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: a randomized clinical trial. *JAMA*, 2015; **314**: 570–81.
62. Brock G, Heiselman D, Knorr J, Ni X, Kinchen K. 9-month efficacy and safety study of testosterone solution 2% for sex drive and energy in hypogonadal men. *J Urol*, 2016; **196**: 1509–15.
63. Moffat SD, Zonderman AB, Metter EJ, Blackman MR, Harman SM, Resnick SM. Longitudinal assessment of serum free testosterone concentration predicts memory performance and cognitive status in elderly men. *J Clin Endocrinol Metab*, 2002; **87**: 5001–7.
64. LeBlanc ES, Wang PY, Janowsky JS, *et al*. Association between sex steroids and cognition in elderly men. *Clin Endocrinol (Oxf)*, 2010; **72**: 393–403.
65. Silverman I, Kastuk D, Choi J, Phillips K. Testosterone levels and spatial ability in men. *Psychoneuroendocrinology*, 1999; **24**: 813–22.
66. Hooven CK, Chabris CF, Ellison PT, Kosslyn SM. The relationship of male testosterone to components of mental rotation. *Neuropsychologia*, 2004; **42**: 782–90.
67. Hsu B, Cumming RG, Waite LM, *et al*. Longitudinal relationships between reproductive hormones and cognitive decline in older men: the concord health and ageing in men project. *J Clin Endocrinol Metab*, 2015; **100**: 2223–30.
68. Lv W, Du N, Liu Y, *et al*. Low testosterone level and risk of Alzheimer's disease in the elderly men: a systematic review and meta-analysis. *Mol Neurobiol*, 2016; **53**: 2679–84.
69. Vaughan C, Goldstein FC, Tenover JL. Exogenous testosterone alone or with finasteride does not improve measurements of cognition in healthy older men with low serum testosterone. *J Androl*, 2007; **28**: 875–82.
70. Cherrier MM, Asthana S, Plymate S, *et al*. Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*, 2001; **57**: 80–8.
71. Cherrier MM, Matsumoto AM, Amory JK, *et al*. Testosterone improves spatial memory in men with Alzheimer disease and mild cognitive impairment. *Neurology*, 2005; **64**: 2063–8.
72. Lu PH, Masterman DA, Mulnard R, *et al*. Effects of testosterone on cognition and mood in male patients with mild Alzheimer disease and healthy elderly men. *Arch Neurol*, 2006; **63**: 177–85.
73. Huang G, Masterman DA, Mulnard R, *et al*. Effects of long-term testosterone administration on cognition in older men with low or low-to-normal testosterone concentrations: a prespecified secondary analysis of data from the randomised, double-blind, placebo-controlled TEAAM trial. *Lancet Diabetes Endocrinol*, 2016; **4**: 657–65.
74. Resnick SM, Matsumoto AM, Stephens-Shields AJ, *et al*. Testosterone treatment and cognitive function in older men with low testosterone and age-associated memory impairment. *JAMA*, 2017; **317**: 717–27.
75. Maki PM, Ernst M, London ED, *et al*. Intramuscular testosterone treatment in elderly men: evidence of memory decline and altered brain function. *J Clin Endocrinol Metab*, 2007; **92**: 4107–14.
76. Shahani S, Braga-Basaria M, Maggio M, Basaria S. Androgens and erythropoiesis: past and present. *J Endocrinol Invest*, 2009; **32**: 704–16.
77. Shahidi NT. Androgens and erythropoiesis. *N Engl J Med*, 1973; **289**: 72–80.
78. Coviello AD, Kaplan B, Lakshman KM, Chen T, Singh AB, Bhasin S. Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab*, 2008; **93**: 914–19.
79. Bachman E, Feng R, Travison T, *et al*. Testosterone suppresses hepcidin in men: a potential mechanism for testosterone-induced erythrocytosis. *J Clin Endocrinol Metab*, 2010; **95**: 4743–7.
80. Bachman E, Travison TG, Basaria S, *et al*. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*, 2014; **69**: 725–35.
81. Ganz T. Hepcidin and iron regulation, 10 years later. *Blood*, 2011; **117**: 4425–33.
82. Moriyama Y, Fisher JW. Effects of testosterone and erythropoietin on erythroid colony formation in human bone marrow cultures. *Blood*, 1975; **45**: 665–70.
83. Gagliano-Jucá T, Pencina KM, Ganz T, *et al*. Mechanisms responsible for reduced erythropoiesis during androgen deprivation therapy in men with prostate cancer. *Am J Physiol Endocrinol Metab*, 2018.
84. Paller CJ, Shiels MS, Rohrmann S, *et al*. Association between sex steroid hormones and hematocrit in a nationally representative sample of men. *J Androl*, 2012; **33**: 1332–41.
85. Ferrucci L, Maggio M, Bandinelli S, *et al*. Low testosterone levels and the risk of anemia in older men and women. *Arch Intern Med*, 2006; **166**: 1380–8.
86. Roy CN, Snyder PJ, Stephens-Shields AJ, *et al*. Association of testosterone levels with anemia in older men: a controlled clinical trial. *JAMA Intern Med*, 2017; **177**: 480–90.
87. Muller M, Grobbee DE, den Tonkelaar I, Lamberts SW, van der Schouw YT. Endogenous sex hormones and metabolic syndrome in aging men. *J Clin Endocrinol Metab*, 2005; **90**: 2618–23.

88. Ding EL, Song Y, Malik VS, Liu S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA*, 2006; **295**: 1288–99.
89. Selvin E, Feinleib M, Zhang L, *et al*. Androgens and diabetes in men: results from the Third National Health and Nutrition Examination Survey (NHANES III). *Diabetes Care*, 2007; **30**: 234–8.
90. Vikan T, Schirmer H, Njolstad I, Svartberg, J. Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. *Eur J Endocrinol*, 2010; **162**: 747–54.
91. Braga-Basaria M, Dobs AS, Muller DC, *et al*. Metabolic syndrome in men with prostate cancer undergoing long-term androgen-deprivation therapy. *J Clin Oncol*, 2006; **24**: 3979–83.
92. Smith JC, Bennett S, Evans LM, *et al*. The effects of induced hypogonadism on arterial stiffness, body composition, and metabolic parameters in males with prostate cancer. *J Clin Endocrinol Metab*, 2001; **86**: 4261–7.
93. Shahani S, Braga-Basaria M, Basaria S. Androgen deprivation therapy in prostate cancer and metabolic risk for atherosclerosis. *J Clin Endocrinol Metab*, 2008; **93**: 2042–9.
94. Basaria S, Muller DC, Carducci MA, Egan J, Dobs AS. Hyperglycemia and insulin resistance in men with prostate carcinoma who receive androgen-deprivation therapy. *Cancer*, 2006; **106**: 581–8.
95. Keating NL, O'Malley AJ, Smith MR. Diabetes and cardiovascular disease during androgen deprivation therapy for prostate cancer. *J Clin Oncol*, 2006; **24**: 4448–56.
96. Gagliano-Jucá T, Burak MF, Pencina KM, *et al*. Metabolic changes in androgen-deprived nondiabetic men with prostate cancer are not mediated by cytokines or aP2. *J Clin Endocrinol Metab*, 2018; **103**: 3900–8.
97. Yialamas MA, Dwyer AA, Hanley E, Lee H, Pitteloud N, Hayes FJ. Acute sex steroid withdrawal reduces insulin sensitivity in healthy men with idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*, 2007; **92**: 4254–9.
98. Pitteloud N, Mootha VK, Dwyer AA, *et al*. Relationship between testosterone levels, insulin sensitivity, and mitochondrial function in men. *Diabetes Care*, 2005; **28**: 1636–42.
99. Hsu B, Cumming RG, Naganathan V, *et al*. Associations between circulating reproductive hormones and SHBG and prevalent and incident metabolic syndrome in community-dwelling older men: the Concord Health and Ageing in Men Project. *J Clin Endocrinol Metab*, 2014; **99**: E2686–91.
100. Mohler ER, 3rd, Ellenberg SS, Lewis CE, *et al*. The effect of testosterone on cardiovascular biomarkers in the testosterone trials. *J Clin Endocrinol Metab*, 2018; **103**: 681–8.
101. Huang G, Pencina KM1, Li Z, *et al*. Long-term testosterone administration on insulin sensitivity in older men with low or low-normal testosterone levels. *J Clin Endocrinol Metab*, 2018; **103**: 1678–85.
102. Gianatti EJ, Dupuis P, Hoermann R, *et al*. Effect of testosterone treatment on glucose metabolism in men with type 2 diabetes: a randomized controlled trial. *Diabetes Care*, 2014; **37**: 2098–107.
103. Dhindsa S, Ghanim H, Batra M, *et al*. Insulin resistance and inflammation in hypogonadotropic hypogonadism and their reduction after testosterone replacement in men with type 2 diabetes. *Diabetes Care*, 2016; **39**: 82–91.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

10.4.5 Risks of Testosterone Treatment

Adrian Dobs and Swaytha Yalamanchi

Introduction

Testosterone has increasingly been prescribed for middle-aged and older men with relatively non-specific symptoms that mimic androgen deficiency, many without a formal diagnosis of hypogonadism. It has been speculated that this trend is at least in part due to direct-to-consumer advertising in the United States and the availability of more convenient forms of testosterone replacement such as gels and patches [1–5]. However, there is evidence that since 2013, testosterone prescribing may have declined, possibly due to concerns regarding adverse effects [6].

There is also concern that a substantial proportion of this increase in testosterone therapy is being prescribed for age-related decline in testosterone levels, which is not an approved indication for use [7]. While benefits in young men with organic (classic) hypogonadism are well-established with low frequency of adverse events, limited data on efficacy and safety are available in older men. As will be discussed, adequately powered randomized controlled trials to examine serious adverse effects and key patient-important outcomes are generally lacking. In this chapter we will outline the available data on the effects of exogenous testosterone on erythrocytosis, lower urinary tract symptoms, prostate cancer, and cardiovascular disease risks in older men. The potential beneficial effects of testosterone therapy on sexual function, body composition, physical function, and bone health are discussed elsewhere (see Chapter 10.4.4, 'Benefits of Testosterone Treatment').

Erythrocytosis

The stimulatory effects of androgens on erythropoiesis have been recognized since the early twentieth century [8]. In clinical studies, men were noted to have higher red cell mass than women and animal studies consistently showed that androgens promote haemoglobin synthesis.

With respect to mechanisms, it has been hypothesized that testosterone stimulates erythropoietin (EPO) secretion and recalibrates the set point of EPO relative to haemoglobin by increasing iron utilization for erythropoiesis [9]. It has also been hypothesized that upregulation of the GATA-1 transcription factor (also termed erythroid transcription factor) and GATA-dependent genes could increase EPO sensitivity and stimulate stress erythropoiesis [10]. The association of circulating EPO concentrations with

testosterone in human studies however has been conflicting [9, 11–14]. Hepcidin, which plays a major role in the negative regulation of iron intestinal absorption and release of iron from macrophages, is another mechanism whereby testosterone can promote erythropoiesis via suppression of hepcidin transcription [15]. Aromatization of testosterone to oestradiol may be required in the regulation of hepcidin transcription [9, 16–18].

An increase in haematocrit may be of benefit to men with anaemia and androgens were previously used to treat various forms of refractory anaemia [8]. Data from the Testosterone Trials (TTrials) demonstrated that among men aged 65 years or older with low testosterone levels (defined as <9.54 nmol/L), exogenous testosterone increases haemoglobin levels in those with unexplained anaemia [19]. However, the pro-erythropoietic actions of testosterone increase the potential for an excessive rise in haematocrit with treatment.

Observational data from the Framingham study demonstrated a 35% increase in cardiovascular mortality in individuals with the highest quintile of haematocrit (46–65% in men) compared to the third quintile of haematocrit (42–43%) (odds ratio 1.35, $P = 0.05$) [20]. Elevated haematocrit has also been associated with increased prevalence of atherosclerosis, coronary artery disease, unstable angina, and myocardial infarction, as well as increased risk of venous thrombosis [21].

In a meta-analysis of 19 randomized control trials of testosterone (n = 651 men treated with testosterone and 433 men with placebo), erythrocytosis defined as haematocrit >50% was the most commonly reported adverse effect (OR 3.69, 95% CI 1.82, 7.51) [22]. Coviello *et al.* performed a secondary analysis of data in a testosterone dose-response study in younger (19–35 years, n = 61) and older men (60–75 years, n = 60) who received long-acting gonadotropin-releasing hormone (GnRH) agonist therapy to suppress endogenous testosterone production and then were given one of five weekly intramuscular doses of testosterone enanthate (20, 50, 125, 300, or 600 mg) [11]. Haemoglobin and haematocrit increased in a dose-dependent linear fashion in both groups of men, though increases in these levels were greater in older compared to younger men. The study suggested that testosterone causes a dose-dependent effect of erythropoiesis that may be more pronounced in older men. The type of testosterone formulation may also impact on the risk of erythrocytosis (haematocrit >52%) with the highest incidence reported with injections such as testosterone cypionate and enanthate (wide range of risk reported, 3–18% with transdermal administration and up to 44% with injections) [23–25]. In the TTrials, a rise in haemoglobin concentration to >17.5 g/dl was seen in only 2% of men who were not anaemic at baseline and were treated with testosterone gel, mostly when serum testosterone was elevated [19].

Another systematic review and meta-analysis (2010) including 51 studies (randomized and non-randomized) with transdermal, oral, and intramuscular formulations with a follow up duration ranging from three months to three years reported that the relative risk of erythrocytosis was three fold higher in testosterone treated men compared to non-treated men [26]. In subgroup analyses, intramuscular route and high testosterone treatment doses were associated with greater increases in haemoglobin and haematocrit. However, a recent meta-analysis [27] including three randomized placebo-controlled trials using exclusively transdermal testosterone for 12–52 weeks in symptomatic hypogonadal (testosterone ≤10.41 nmol/L) men, an eight fold higher risk (relative risk: 8.14; 95% CI, 1.87, 35.40; n = 1579) of erythrocytosis (haematocrit >54%) was reported. Patients predisposed to erythrocytosis due to smoking, hypoventilation syndrome, sleep apnoea, living at high altitude, end stage renal disease, and lung disease may be at higher risk of polycythemia [28].

Evidence-based guidelines suggest baseline haematocrit >48% (>50% for men living at high altitudes) should be a relative contraindication to commencing testosterone treatment, as such men may be at higher risk of experiencing a haematocrit rise to a potentially harmful level [29]. Should men with elevated haematocrit be considered for testosterone treatment, further evaluation of possible causes of erythrocytosis and thorough discussion about risk of polycythemia on treatment is recommended before starting. Currently, it is recommended that both testosterone and haematocrit be monitored at 3–6-month intervals for the first year after initiation, and annually thereafter.

If haematocrit exceeds 54% during treatment, testosterone should be withheld until haematocrit normalizes—this may take several weeks or months. Therapeutic phlebotomy is effective in rapidly lowering the haematocrit in the immediate management of erythrocytosis in the context of testosterone pharmacotherapy. Thereafter periodic phlebotomy may allow continuation of testosterone therapy, when clinically indicated.

Venous Thromboembolism (VTE)

VTE is associated with morbidity and mortality [30, 31]. The association of testosterone with VTE risk is unclear. It has been hypothesized that testosterone may increase the risk of VTE via erythrocytosis-induced viscosity; other postulated mechanisms include platelet aggregation, increased thromboxane A_2 concentration in platelets, and an increase in circulating oestrogens [32]. A population-based case control study demonstrated that the risk of thromboembolism was elevated in the first six months of therapy, translating to ten additional VTE events/year above the base rate of 15.8 per 10 000 person-years [33]. In 2014, the FDA required manufacturers to include information on VTE risk on testosterone products based on post-marketing data, noting that VTE may occur even in the absence of an excessive rise in haematocrit. It is possible that men who developed VTE while on testosterone may have an undiagnosed susceptibility for thromboembolism, which becomes unmasked by the testosterone treatment. Glueck *et al.* (2014) reported that among 42 individuals (38 men and 4 women) who developed thrombotic events after starting testosterone treatment (median 5 months), 40 had evidence of thrombophilia-hypofibrinolysis [32]. Men with VTE were more likely to have Factor V Leiden heterozygosity and high levels of both Factor VIII and factor XI compared to controls [32]. Similar results were seen among individuals hospitalized for pulmonary embolism [34]. Currently available data are limited to uncontrolled anecdotal case reports and short-term studies. A meta-analysis of randomized controlled data (mean duration 42 weeks, mean age 47 years) including different cohorts and different formulations/doses of testosterone did not demonstrate an increased risk of VTE with testosterone (Mantel–Haenszel-odds ratio 1.96 [0.75, 5.17], $P = 0.17$) [35]. Screening all patients before starting testosterone for thrombophilia (including but not limited

to Factor V Leiden and prothrombin mutations) is not currently recommended, but a thorough personal and family history of thromboembolic events is important. While known thrombophilia is a contraindication to testosterone therapy in current guidelines, a history of VTE is not because there are too few testosterone-associated VTE events in RCTs to make meaningful inferences [29]; clearly more data are needed.

Prostate

Lower Urinary Tract Symptoms (LUTS)

Androgens play an important role in prostate development and growth; this had led to concerns that exogenous testosterone may increase prostate volume and exacerbate LUTS. LUTS is a term used to describe a range of symptoms due to storage (irritative) and/or voiding (obstructive) disturbances related to problems of the lower urinary tract (including bladder irritation and prostate enlargement due to benign prostatic hypertrophy) common in ageing men. The International Prostate Symptoms Score (IPSS) is a validated measure consisting of seven questions designed to evaluate symptom severity (0–7 mild symptoms; 8–19 moderate symptoms; 20–35 severe symptoms) and response to therapy.

In the TTrials (n = 778 participants), exogenous testosterone for 12 months was not associated with a difference in men with moderately severe LUTS (defined as IPSS >19) [36]. Exogenous testosterone has not been demonstrated to cause *de novo* or worsen mild to moderate LUTS—a consistent finding in systematic reviews and meta-analyses [27, 37, 38]. However, it is important to bear in mind that men with severe LUTS (IPSS score >19) have invariably been excluded from entering testosterone trials; therefore it is unknown whether exogenous testosterone will exacerbate LUTS in this population

Consequently, all guidelines recommend against starting testosterone in men with severe LUTS (AUA/IPSS>19) [29].

Prostate Cancer

The relationships between testosterone and prostate cancer is poorly understood. Currently there is not clear evidence that endogenous testosterone levels are associated with increased risk of prostate cancer [39], nor is there any evidence that exogenous testosterone causes the development of prostate cancer de novo [26].

Traditionally, prostate cancer has been considered to be a contraindication to testosterone treatment, partly on the basis of studies in the 1940s that demonstrated clinical/biochemical improvements with castration therapy and acceleration of prostate cancer progression during testosterone treatment [40]. In the 1980s, the saturation model for androgen action in the prostate was proposed on the basis of androgen receptor binding studies of benign prostate tissue. The model suggests that, at an intraprostatic androgen concentration threshold corresponding to a relatively low concentration of circulating testosterone, prostate androgen receptors may become saturated and are relatively insensitive to further increases in serum testosterone. Thus androgens may have limited ability to stimulate prostate cancer growth [40–42]. However, the saturation model has been criticized as being overly simplistic—it does not account for complexities in the interchange between serum/intraprostatic androgens, nor for the functional heterogeneity and transformation of androgen receptor signalling (modified by transcriptional activators/repressors) in prostate cancer cells. Additionally, as evidenced by the established role of androgen deprivation therapy (medically or surgically) in advanced prostate cancer [43], androgen receptor signalling clearly plays an important role in the growth of metastatic prostate cancer.

The majority of available data on exogenous testosterone and incident prostate cancer are limited to small and prospective studies insufficiently powered to assess prostate cancer risk. It has been estimated that to detect a 30% increase in the incidence of prostate cancer, a trial would require 6000 men in each arm of a randomized placebo controlled trial receiving testosterone treatment for 5 years [44]. A meta-analysis in 2010 of a variety of safety endpoints in 51 randomized controlled trials did not detect an increase in prostate cancer [26]. More recently, Corona *et al.* (2018) identified a total of seven published systematic reviews and meta-analyses on the relationship between testosterone and prostate safety since 2005 [35]. Both parenteral and transdermal preparations were reported to be associated with short-term PSA elevation within the normal range, but no increased risk of histologically positive prostate biopsy or prostate cancer.

Reports of increased prostate cancer in men on testosterone may be related to detection bias. Men on testosterone undergo increased surveillance, thus increasing the risk of detecting subclinical prostate cancer; additionally, testosterone-induced rises in PSA levels may increase the rate of prostate biopsy. Testosterone treatment increase the likelihood of detecting PSA levels above the threshold of 4 ng/ml, raising the risk of prostate biopsy by up to 1.8-fold, and consequently the detection of prostate cancer [22].

Recognizing the importance of minimizing the risks and expense of unnecessary testing, especially the potential harms of prostate biopsy, current guidelines recommend discussing the benefits and risks of prostate cancer screening and monitoring with hypogonadal men aged 55–69 years being considered for testosterone therapy, who have a life expectancy greater than 10 years, in order to reach a joint decision with the clinician on prostate evaluation. In hypogonadal men 40–69 years old who are at increased risk of prostate cancer (such as African Americans or those with a first-degree relative with prostate cancer), prostate cancer screening and monitoring should be offered after discussing the risks. It is not recommended to screen men <40 years (due to low risk for prostate cancer in this age group), nor screening men >70 years (as the risk of dying from prostate cancer if diagnosed is not high). Among men who choose prostate screening/monitoring, PSA and digital rectal exam should be performed before initiating and 3–12 months after starting treatment and thereafter in accordance with local guidelines for prostate cancer screening depending on the age and race of the patient. A urological consultation is recommended if there is a confirmed PSA >4.0 ng/ml or an abnormal prostate on digital rectal examination at any time, and a confirmed increase in PSA to >1.4 ng/ml above baseline during the first year of treatment.

The Endocrine Society classifies men with metastatic prostate cancer and/or breast cancer (considered to be hormone-dependent cancers) at very high risk of serious adverse outcomes and recommends against testosterone in this group. Men with unevaluated prostate nodules or induration, or unevaluated PSA >4 ng/ml (>3 ng/ml in men at high risk for prostate cancer) are classified as

moderate to high risk of adverse outcomes and testosterone is not recommended in these individuals without prior urological evaluation. The Endocrine Society refrained from offering a general recommendation on testosterone treatment in patients with adequately treated organ-confined prostate cancer. However, the American Urological Society states that hypogonadal patients with a history of prostate cancer should be informed that there is inadequate evidence to quantify the risk-benefit ratio of testosterone therapy; in the setting of treated prostate cancer, testosterone therapy should be a negotiated decision based on perceived potential benefit of treatment vs. limited knowledge of potential risks [45].

Cardiovascular System

Multiple studies have described an inverse relationship between endogenous testosterone and adverse cardiovascular outcomes independent of traditional risk factors. The evidence overall suggests that physiologic testosterone levels may be beneficial, while lower testosterone concentrations have been associated with increased cardiovascular risks [46, 47]. However, data are largely observational, and may therefore be influenced by unadjusted variables and/or unknown confounders, making it difficult to designate testosterone as a mediator (as opposed to a risk marker) in the pathway leading to cardiovascular disease. Indeed, obesity and a high comorbid burden may be associated with central suppression of the hypothalamic–pituitary–testicular axis, suggesting that in some men, low testosterone could therefore simply be a biomarker of poor health including cardiovascular disease (see Chapter 10.6, 'HPT Axis Function in Systemic Diseases and Effects of Medications').

Definitive data on cardiovascular risk associated with testosterone use do not exist.

No placebo-controlled randomized control trials in men on testosterone have been designed with primary cardiovascular endpoints and much of our knowledge is derived from trials aimed to explore other outcomes.

One of the first trials to raise concern about adverse cardiovascular events was the Testosterone in Older Men with Mobility Limitations (TOM) trial [48]. This randomized placebo-controlled trial included community-dwelling older men (mean age 74 years) with limitations in mobility, and serum total testosterone levels 3.47–12.15 nmol/L. Although cardiovascular outcomes were not a pre-specified endpoint, the study was terminated early because of a higher frequency of self-reported 'cardiovascular-related' events in the treatment arm. Because of early termination, small sample size, lack of predominance of a single event (events ranged from oedema to myocardial infarction (MI) and death), and small number of cardiovascular events, it is difficult to interpret this data. Two retrospective cohort studies reported an association of exogenous testosterone with an increased risk of cardiovascular disease. The first demonstrated that veterans started on testosterone therapy after coronary angiography had a higher percentage of cardiovascular events at three years compared to men who were not treated (19.9% vs. 25.7%) [49]. The second study compared the incidence rate of acute non-fatal myocardial infarctions in the 90 days post-testosterone prescription with the rate in the one year prior to the prescription [50]. The authors reported 2.9-fold of excess risk of cardiovascular evens in men ≥65 years of age and in younger men with a prior history of heart disease post-prescription as compared to pre-prescription. However, several other studies have not shown an increased cardiovascular risk associated with testosterone treatment [26, 51, 52]. In particular, most meta-analyses have not demonstrated a significant association with adverse cardiovascular events/mortality and major adverse cardiovascular effects [53, 54]. The FDA (31 January 2014) issued a statement, despite the conflicting and inconclusive evidence, to warn against the possibility of increased cardiovascular risk with exogenous testosterone and mandated that manufacturers of testosterone products change their labelling to clarify approved uses and to include information about the possibility of increased risk of myocardial infarction and strokes. In contrast, the European Medicines Agency issued a report stating that the signal for an increased risk of cardiovascular associated with use of testosterone remains weak and inconclusive [55].

An overview of systematic reviews of randomized controlled trials published up to 2016 identified six systematic reviews, each including a meta-analysis, that showed no overall significant association between testosterone and cardiovascular events [54]. Among these six meta-analyses [22, 26, 56–59], two showed an increased risk only in subgroup analyses (one with oral testosterone [58] and the other in men ≥65 years of age in the first year [59]). A subsequent systematic review and meta-analysis included 39 randomized controlled trials and 10 observational studies similarly demonstrated no increase in the risk of myocardial infarction, stroke, or mortality in randomized controlled trials [60]. It has been suggested [54] that in order to detect a true difference in cardiovascular risk in a randomized controlled trial among men receiving testosterone and controls (assuming P value of 0.05 and power of 80%) the study would require at least 17 664 participants in each arm. Despite the obvious difficulties, such a study, mandated by the FDA and led by industry, is planned [54]. In summary, the very low quality of available evidence precludes definitive conclusions on the association between testosterone and cardiovascular adverse outcomes.

In accordance with the aforementioned data, current guidelines acknowledge that there is no conclusive evidence that exogenous testosterone is associated with increased cardiovascular risk in hypogonadal men [29]. However, as a precautionary good clinical practice statement (or opinion), exogenous testosterone is not recommended in those with uncontrolled heart failure, myocardial infarction, or stroke with the past 6 months. Although testosterone has been explored as a potential treatment for heart failure [61], testosterone may cause fluid retention and oedema and use is not currently recommended in men with uncontrolled heart failure [48].

Conclusion

Age-related decline in testosterone levels may be physiologic, due to disorders at one or more levels of the hypothalamic–pituitary–gonadal system or due to gonadal axis suppression by accumulation of age-related comorbidities, including obesity. Testosterone treatment is generally recommended only in those considered to have organic (classical hypogonadism). However, the vast majority of symptomatic older men with low testosterone do not have any identifiable aetiology usually associated with organic hypogonadism. Further compounding the diagnostic difficulties, management decisions in older patients often have to be made on the

basis of limited data on the risks/benefits of testosterone. Due to a possible association with erythrocytosis, testosterone is not recommended in men with an elevated haematocrit at baseline or at increased risk for thrombophilia. VTE risk, if it does exist, may be highest in the first 6 months of treatment. Available data do not suggest worsening of low to moderate LUTS with testosterone, though data are lacking in men with severe LUTS. Testosterone is not recommended in men who have untreated prostate cancer. Among hypogonadal men with cured prostate cancer, the risk of testosterone treatment is unclear. Data with regards to cardiovascular risk are inconclusive, but presently testosterone initiation is not recommended in men with uncontrolled heart failure, myocardial infarction, or stroke within the last 6 months or thrombophilia.

There are no adequately powered randomized controlled trial data on all the aforementioned safety endpoints, limiting our ability to ascertain the true level of risk associated with testosterone. Further research is urgently needed to understand the benefits as well as risks of testosterone in men with age-related hypogonadism.

REFERENCES

1. Layton JB, Li D, Meier CR, *et al*. Testosterone lab testing and initiation in the United Kingdom and the United States, 2000 to 2011. *J Clin Endocrinol Metab*, 2014; **99**: 835–42.
2. Baillargeon J, Urban RJ, Kuo YF, *et al*. Risk of myocardial infarction in older men receiving testosterone therapy. *Ann Pharmacother*, 2014; **48**: 1138–44.
3. Handelsman DJ. Global trends in testosterone prescribing, 2000–2011: expanding the spectrum of prescription drug misuse. *Med J Aust*, 2013; **199**: 548–51.
4. Abadilla KA, Dobs AS. Topical testosterone supplementation for the treatment of male hypogonadism. *Drugs*, 2012; **72**: 1591–603.
5. Layton JB, Kim Y, Alexander GC, Emery SL. Association between direct-to-consumer advertising and testosterone testing and initiation in the united states, 2009–2013. *JAMA*, 2017; **317**: 1159–66.
6. Baillargeon J, Kuo YF, Westra JR, Urban RJ, Goodwin JS. Testosterone prescribing in the united states, 2002–2016. *JAMA*, 2018; **320**: 200–2.
7. Nguyen CP, Hirsch MS, Moeny D, Kaul S, Mohamoud M, Joffe HV. Testosterone and 'age-related hypogonadism'—FDA concerns. *N Engl J Med*, 2015; **373**: 689–91.
8. Shahani S, Braga-Basaria M, Maggio M, Basaria S. Androgens and erythropoiesis: past and present. *J Endocrinol Invest*, 2009; **32**: 704–16.
9. Bachman E, Travison TG, Basaria S, *et al*. Testosterone induces erythrocytosis via increased erythropoietin and suppressed hepcidin: evidence for a new erythropoietin/hemoglobin set point. *J Gerontol A Biol Sci Med Sci*, 2014; **69**: 725–35.
10. Guo W, Li M, Bhasin S. Testosterone supplementation improves anemia in aging male mice. *J Gerontol A Biol Sci Med Sci*, 2014; **69**: 505–13.
11. Coviello AD, Kaplan B, Lakshman KM, Chen T, Singh AB, Bhasin S. Effects of graded doses of testosterone on erythropoiesis in healthy young and older men. *J Clin Endocrinol Metab*, 2008; **93**: 914–19.
12. Ip FF, di Pierro I, Brown R, Cunningham I, Handelsman DJ, Liu PY. Trough serum testosterone predicts the development of polycythemia in hypogonadal men treated for up to 21 years with subcutaneous testosterone pellets. *Eur J Endocrinol*, 2010; **162**: 385–90.
13. Maggio M, Snyder PJ, Ceda GP, *et al*. Is the haematopoietic effect of testosterone mediated by erythropoietin? the results of a clinical trial in older men. *Andrology*, 2013; **1**: 24–8.
14. Rochira V, Zirilli L, Madeo B, Maffei L, Carani C. Testosterone action on erythropoiesis does not require its aromatization to estrogen: insights from the testosterone and estrogen treatment of two aromatase-deficient men. *J Steroid Biochem Mol Biol*, 2009; **113**: 189–94.
15. Guo W, Bachman E, Li M, *et al*. Testosterone administration inhibits hepcidin transcription and is associated with increased iron incorporation into red blood cells. *Aging Cell*, 2013; **12**: 280–91.
16. Yang Q, Jian J, Katz S, Abramson SB, Huang X. 17beta-estradiol inhibits iron hormone hepcidin through an estrogen responsive element half-site. *Endocrinology*, 2012; **153**: 3170–8.
17. Ikeda Y, Tajima S, Izawa-Ishizawa Y, *et al*. Estrogen regulates hepcidin expression via GPR30-BMP6-dependent signaling in hepatocytes. *PLoS One*, 2012; **7**: e40465.
18. Hou Y, Zhang S, Wang L, *et al*. Estrogen regulates iron homeostasis through governing hepatic hepcidin expression via an estrogen response element. *Gene*, 2012; **511**: 398–403.
19. Roy CN, Snyder PJ, Stephens-Shields AJ, *et al*. Association of testosterone levels with anemia in older men: a controlled clinical trial. *JAMA Intern Med*, 2017; **177**: 480–90.
20. Gagnon DR, Zhang TJ, Brand FN, Kannel WB. Hematocrit and the risk of cardiovascular disease—the Framingham study: a 34-year follow-up. *Am Heart J*, 1994; **127**: 674–82.
21. Hajjar RR, Kaiser FE, Morley JE. Outcomes of long-term testosterone replacement in older hypogonadal males: a retrospective analysis. *J Clin Endocrinol Metab*, 1997; **82**: 3793–6.
22. Calof OM, Singh AB, Lee ML, *et al*. Adverse events associated with testosterone replacement in middle-aged and older men: a meta-analysis of randomized, placebo-controlled trials. *J Gerontol A Biol Sci Med Sci*, 2005; **60**: 1451–7.
23. Rhoden EL, Morgentaler A. Risks of testosterone-replacement therapy and recommendations for monitoring. *N Engl J Med*, 2004; **350**: 482–92.
24. Dobs AS, Hoover DR, Chen MC, Allen R. Pharmacokinetic characteristics, efficacy, and safety of buccal testosterone in hypogonadal males: a pilot study. *J Clin Endocrinol Metab*, 1998; **83**: 33–9.
25. Snyder PJ, Peachey H, Berlin JA, *et al*. Effects of testosterone replacement in hypogonadal men. *J Clin Endocrinol Metab*, 2000; **85**: 2670–7.
26. Fernandez-Balsells MM, Murad MH, Lane M, *et al*. Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2010; **95**: 2560–75.
27. Ponce OJ, Spencer-Bonilla G, Alvarez-Villalobos N, *et al*. The efficacy and adverse events of testosterone replacement therapy in hypogonadal men: a systematic review and meta-analysis of randomized, placebo-controlled trials. *J Clin Endocrinol Metab*, 2018. doi: 10.1210/jc.2018-00404.
28. Scherber RM, Mesa RA. Elevated hemoglobin or hematocrit level. *JAMA*, 2016; **315**: 2225–6.
29. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.

30. Grosse SD, Nelson RE, Nyarko KA, Richardson LC, Raskob GE. The economic burden of incident venous thromboembolism in the United States: a review of estimated attributable healthcare costs. *Thromb Res*, 2016; **137**: 3–10.
31. Cohen AT, Agnelli G, Anderson FA, *et al*. Venous thromboembolism (VTE) in europe. the number of VTE events and associated morbidity and mortality. *Thromb Haemost*, 2007; **98**: 756–64.
32. Glueck CJ, Wang P. Testosterone therapy, thrombosis, thrombophilia, cardiovascular events. *Metabolism*, 2014; **63**: 989–94.
33. Martinez C, Suissa S, Rietbrock S, *et al*. Testosterone treatment and risk of venous thromboembolism: population based case-control study. *BMJ*, 2016; **355**: i5968.
34. Prince M, Glueck CJ, Shah P, *et al*. Hospitalization for pulmonary embolism associated with antecedent testosterone or estrogen therapy in patients found to have familial and acquired thrombophilia. *BMC Hematol*, 2016; **16**: 6.
35. Corona G, Rastrelli G, Reisman Y, Sforza A, Maggi M. The safety of available treatments of male hypogonadism in organic and functional hypogonadism. *Expert Opin Drug Saf*, 2018; **17**: 277–92.
36. Snyder PJ, Ellenberg SS, Farrar JT. Testosterone treatment in older men. *N Engl J Med*, 2016; **375**: 90.
37. Kathrins M, Doersch K, Nimeh T, Canto A, Niederberger C, Seftel A. The relationship between testosterone-replacement therapy and lower urinary tract symptoms: a systematic review. *Urology*, 2016; **88**: 22–32.
38. Kohn TP, Mata DA, Ramasamy R, Lipshultz LI. Effects of testosterone replacement therapy on lower urinary tract symptoms: a systematic review and meta-analysis. *Eur Urol*, 2016; **69**: 1083–90.
39. Endogenous Hormones and Prostate Cancer Collaborative Group, Roddam AW, Allen NE, Appleby P, Key TJ. Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. *J Natl Cancer Inst*, 2008; **100**: 170–83.
40. Bell MA, Campbell JD, Joice G, Sopko NA, Burnett AL. Shifting the paradigm of testosterone replacement therapy in prostate cancer. *World J Mens Health*, 2018; **36**: 103–9.
41. Traish AM, Muller RE, Wotiz HH. A new procedure for the quantitation of nuclear and cytoplasmic androgen receptors. *J Biol Chem*, 1981; **256**: 12028–33.
42. Traish AM, Muller RE, Wotiz HH. A new exchange procedure for the quantitation of prostatic androgen receptor complexes formed *in vivo*. *J Steroid Biochem*, 1985; **23**: 405–13.
43. Basaria S. Testosterone therapy in older men with late-onset hypogonadism: a counter-rationale. *Endocr Pract*, 2013; **19**: 853–63.
44. Bhasin S, Singh AB, Mac RP, Carter B, Lee MI, Cunningham GR. Managing the risks of prostate disease during testosterone replacement therapy in older men: recommendations for a standardized monitoring plan. *J Androl*, 2003; **24**: 299–311.
45. Mulhall JP, Trost LW, Brannigan RE, *et al*. Evaluation and management of testosterone deficiency: AUA guideline. *J Urol*, 2018; **200**: 423–32.
46. Elagizi A, Kohler TS, Lavie CJ. Testosterone and cardiovascular health. *Mayo Clin Proc*, 2018; **93**: 83–100.
47. Sharma R, Oni OA, Gupta K, *et al*. Normalization of testosterone level is associated with reduced incidence of myocardial infarction and mortality in men. *Eur Heart J*, 2015; **36**: 2706–15.
48. Basaria S, Coviello AD, Travison TG, *et al*. Adverse events associated with testosterone administration. *N Engl J Med*, 2010; **363**: 109–22.
49. Vigen R, O'Donnell CI, Baron AE, *et al*. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *JAMA*, 2013; **310**: 1829–36.
50. Finkle WD, Greenland S, Ridgeway GK, *et al*. Increased risk of non-fatal myocardial infarction following testosterone therapy prescription in men. *PLoS One*, 2014; **9**: e85805.
51. Shores MM, Smith NL, Forsberg CW, Anawalt BD, Matsumoto AM. Testosterone treatment and mortality in men with low testosterone levels. *J Clin Endocrinol Metab*, 2012; **97**: 2050–8.
52. Srinivas-Shankar U, Roberts SA, Connolly MJ, *et al*. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2010; **95**: 639–50.
53. Corona G, Rastrelli G, Monami M, *et al*. Hypogonadism as a risk factor for cardiovascular mortality in men: a meta-analytic study. *Eur J Endocrinol*, 2011; **165**: 687–701.
54. Onasanya O, Iyer G, Lucas E, Lin D, Singh S, Alexander GC. Association between exogenous testosterone and cardiovascular events: an overview of systematic reviews. *Lancet Diabetes Endocrinol*, 2016; **4**: 943–56.
55. European Medicines Agency. No consistent evidence of an increased risk of heart problems with testosterone medicines. http://www.ema. europa.eu/ema/index jsp?curl=pages/news_ and_events/news/2014/11/news_detail_00221 8jsp&mid=WC0b01ac058004d5c1 (accessed: 21 November 2014).
56. Haddad RM, Kennedy CC, Caples SM, *et al*. Testosterone and cardiovascular risk in men: a systematic review and meta-analysis of randomized placebo-controlled trials. *Mayo Clin Proc*, 2007; **82**: 29–39.
57. Xu L, Freeman G, Cowling BJ, Schooling CM. Testosterone therapy and cardiovascular events among men: a systematic review and meta-analysis of placebo-controlled randomized trials. *BMC Med*, 2013; **11**: 108-7015-11-108.
58. Borst SE, Shuster JJ, Zou B, *et al*. Cardiovascular risks and elevation of serum DHT vary by route of testosterone administration: a systematic review and meta-analysis. *BMC Med*, 2014; **12**: 211.
59. Albert SG, Morley JE. Testosterone therapy, association with age, initiation and mode of therapy with cardiovascular events: a systematic review. *Clin Endocrinol (Oxf)*, 2016; **85**: 436–43.
60. Alexander GC, Iyer G, Lucas E, Lin D, Singh S. Cardiovascular risks of exogenous testosterone use among men: a systematic review and meta-analysis. *Am J Med*, 2017; **130**: 293–305.
61. Toma M, McAlister FA, Coglianese EE, *et al*. Testosterone supplementation in heart failure: a meta-analysis. *Circ Heart Fail*, 2012; **5**: 315–21.

10.5

Management of Idiopathic Male Infertility

Herman Tournaye and Biljana Popovic-Todorovic

Introduction

According to the World Health Organization (WHO), the definition of infertility is 'the inability of a sexually active, non-contracepting couple to achieve spontaneous pregnancy in 1 year [1]. About 15% of all couples are infertile and seek medical treatment for fertility [2]. Male infertility is wholly or partly the cause in 20–70% of infertile couples [3]. Excellent fertility status in a female partner might compensate for reduced fertility of a male partner. Alternatively, poor fertility in a male partner could hamper or eliminate conception altogether, irrespective of the female partner's fecundity [3].

Epidemiology

In 2010, around 48.5 million couples worldwide were considered to have primary infertility [4]. The WHO surmises that the real prevalence of infertility is probably 2.5 times higher [5, 6]. It is estimated that more than 30 million men worldwide may be infertile [7].

Causes of Male Infertility

Male infertility has traditionally been divided into three major aetiological categories: pretesticular, testicular, and post-testicular. A new clinically based aetiological construct which describes the underlying causes of male fertility in terms of hypothalamic–pituitary axis function, quantitative and qualitative spermatogenesis impairments, and ductal obstruction or dysfunction (**Table 10.5.1**) has been proposed recently [3]. Since increasing numbers of underlying genetic bases for defects in male reproduction are being discovered, this classification can be further subdivided into genetic, non-genetic, and environmental categories (for detailed description of the classification, please consult [3]).

Genetics of Male Infertility

Spermatogenesis can be impaired by gross chromosomal anomalies identifiable by karyotype analysis or by submicroscopic deletions and gene mutations [3]. The gross chromosomal anomalies most commonly include Klinefelter's syndrome (47XXY) which is a direct cause of male reproductive failure in 15% of azoospermic men and frequency is one per 600 men [8] and 46XX male syndrome (de la Chapelle syndrome) is seen in one per 20,000 male neonates and causes azoospermia (see Chapter 10.3, 'Klinefelter Syndrome'). Structural changes that might alter the natural composition and integrity of the Y chromosome are isodicentricism, truncation, or ring formation [9] where the degree of spermatogenesis defect depends on the proportion of cells with the aberrant Y chromosome. Structural chromosomal anomalies such as Robertsonian translocations, inversions, and reciprocal translocations are more commonly seen in oligozoospermic men than in normospermic men, with a frequency of 4–8%, which is ten times more frequent than in the general population [3]. Y-chromosomal microdeletions are the most frequent known molecular genetic causes of severe impairment of spermatogenesis, accounting for around 10% of cases of non-obstructive azoospermia and 3–5% of severe oligozoospermia [3].

Spermatogenesis is one of the most complex cellular differentiation processes in which germ cells transform from spermatogonia to spermatozoa while undergo meiosis [10] (Chapter 10.1.1, 'Male Reproductive Physiology and Spermatogenesis'). Spermatogenesis is thought to be orchestrated by a up to 2000 genes, of which 600 to 900 seem to be exclusively expressed in the male germ cells [11–13]. Accordingly, genetic defects underlying abnormal spermatogenesis have been actively sought, initially through a candidate gene single nucleotide polymorphisms (SNP) [14] and latterly by high-throughput genome-wide association studies (GWAS), comparative genomic hybridization-arrays (array-CGH), and next generation sequencing (NGS).

Table 10.5.1 New classification for male reproductive impairment

Hypothalamic pituitary axis	
Genetic	Congenital hypogonadotropic hypogonadism with anosmia (e.g. Kallmann's syndrome) or normosmia
Non-genetic	CNS malignancy or transphenoidal resection or radiation (ablative);haemochromatosis, sarcoidosis, tuberculosis or fungal (infiltrative); secreting and non-secreting pituitary adenoma (compressive); and exogenous androgen or testosterone use (suppressive)
Spermatogenic, quantitative	
Genetic	Y-chromosomal microdeletions in the AZFa, AZFb, AZFc subregions of the long armAZF region, 47,XXY Klinefelter's syndrome, 46,XX male syndrome, or isodicentric Y chromosomes, partial androgen insensitivity syndrome (mild form), chromosomal structural anomalies (translocation, inversions), X chromosomal TEX11 mutation
Non-genetic	Varicocele (grade 3), previous cytotoxic chemotherapy or radiotherapy, previous testicular torsion leading to loss of testis, bilateral mumps orchitis, bilateral testis malignancy and orchiectomy, and systemic illness (liver or renal insufficiency)
Presumed non-genetic	Idiopathic oligozoospermia or azoospermia and cryptorchidism or testicular dysgenesis syndrome
Spermatogenic, qualitative	
Genetic	Globozoospermia, immotile cilia syndrome, stump-tail syndrome, macrocephalic sperm head, and advanced paternal age
Non-genetic	Oxidative stress or DNA damage
Presumed non-genetic	Phospholipase C ζ deficiency, idiopathic asthenozoospermia, idiopathic teratozoospermia, autoimmune
Ductal obstruction or dysfunction	
Genetic	Congenital absence of the vas deferens with normal renal anatomy
Non-genetic	Previous vasectomy, idiopathic epididymal occlusion, bilateral inguinal hernia repair, ejaculatory duct obstruction, diabetes mellitus with vasal peristaltic deficiency, spinal cord injury, multiple sclerosis, neural tube defects, retroperitoneal lymph node dissection, pelvic surgery, and ejaculatory or erectile dysfunction
Presumed non-genetic	Congenital bilateral absence of the vas with unilateral renal agenesis, Young's syndrome

The results of these research effort to date however have been largely unsuccessful in identifying common genetic factors in male infertility [15]. A few rare single gene defects have been described in association with azoospermia and oligozoospermia [16] and teratozoospermia and asthenozoospermia [17]. One recently identified mutation associated with non-obstructive azoospermia is the X-chromosome meiotic gene *TEX11* [16]. Approximately 80% of the patients with macrozoospermia and globozoospermia are associated with mutations in one single gene, AURKC for macrozoospermia and DPY19L2 for globozoospermia [17]. In addition, four genes have been linked to asthenozoospermia: *CATSPER1, CATSPER2, SEPT12*, and *SLC26A8* but all of the four mutations are rare [14].

Idiopathic Male Infertility

A recent prospective clinical-epidemiological study, conducted in 8518 couples trying to achieve pregnancy for more than 1 year [18], revealed that 20.4% (n = 1737) of men had total sperm numbers below 39 million per ejaculate, the lower limit of normal proposed by the WHO. In 60% of these men, the primary causal factor could not be assigned [18].

In general, 30–40% of male infertility cases have no clear identified cause, a heterogeneous group referred to as idiopathic male infertility [2]. Despite the fact that these men have no history of diseases affecting fertility while physical examinations and endocrine, genetic, and biochemical tests are normal, their semen analysis often reveals non-specific abnormal semen parameters such as 'oligozoospermia' (reduced sperm count), 'asthenozoospermia' (reduced sperm motility), 'teratozoospermia' (reduced percentage of sperm with normal morphology). Combinations are common; most frequently 'oligoasthenoteratozoospermia' or 'OAT syndrome'. It can be seen from Table 10.5.1 that idiopathic male infertility patients are mostly classified in the presumed non-genetic causes. However, it is likely that in the years to come, novel genetic mutations associated with male infertility will be identified as more powerful molecular genetic diagnostic methods such as whole genome sequencing are applied to studies of large cohorts of patients. This could enable rapid identification of many more genetic alterations (including copy number variations) associated with male infertility [14, 16].

Diagnosis

Standardized investigation of the male partner of an infertile couple should start with an accurate medical history focusing on past medical (e.g. cryptorchidism, mumps orchitis, TB, STI, respiratory disease) and surgical history (e.g. orchidopexy, herniorrhaphy), sexual history (including libido, erection, and ejaculation) histories medication usage, dietary supplements, and anabolic steroids exposure, family history of infertility.

Physical examination should include assessment of secondary sex characteristics (distribution of pubic hair, body proportions, voice changes, and the presence of gynaecomastia) and genitalia

examination including penis, testes (position, size, consistency), other scrotal contents (including vas, epididymis, varicoceles, cysts). Enhanced, out-of-proportion upper body muscular development can indicate anabolic steroid misuse. Body weight, body-mass index, and waist circumference should be recorded. Routine semen analysis is the best method to assess male fertility and should be performed according to WHO guidelines [1] (see Chapter 10.2.3, 'Diagnostic Semen Analysis').

Management of Idiopathic Male Infertility

Presently, there is no specific treatment for idiopathic male infertility. Thus, the management of idiopathic male infertility is based on maximizing the chance of spontaneous conception in the couple and/or artificially by assisted reproductive technology (ART).

Various lifestyle factors may contribute to a decrease in sperm quality and male fecundity [19, 20]. Although smoking, marijuana use, high alcohol consumption, obesity have been reported to be associated with reduced sperm quality, a direct negative causal impact remains unproven [19]. In addition, there are no good intervention studies on the possible benefits of lifestyle change on semen quality [20, 21]. Nevertheless, it is generally sound to recommend a healthy lifestyle including diet, exercise, avoidance of smoking and illicit drugs, for infertile men seeking conception [19].

Empirical Medical Therapy for Idiopathic Male Infertility

In the absence of any targeted therapy, a number of empirical drugs, including hormonal preparations and antioxidants have been used off-label to treat infertile men without clear benefits [22]. Hormonal treatments include the use of gonadotropins, aromatase inhibitors and selective oestrogen receptor modulators (SERMs).

Gonadotropins

Gonadotropins are the only class of medications approved for the medical management of male factor infertility in case of hypogonadotropic hypogonadism [22] (see Chapter 10.4.3, 'Gonadotrophin Induction of Spermatogenesis') while their use in the idiopathic male infertility is controversial.

A Cochrane review including six randomized controlled trials involving 456 patients with idiopathic male subfertility treated with FSH [23], and a more recent meta-analysis [26] including 15 controlled clinical studies (614 men treated with FSH and 661 treated with placebo or untreated), suggested a possible beneficial effect on pregnancy rate in the treatment group compared to placebo or no treatment. The limited data of the Cochrane review [23] showed that the live birth rate per couple randomly assigned (27% vs. 0%; Peto odds ratio (OR) 9.31, 95% confidence interval (CI) 1.17 to 73.75, one study, 30 participants, very low-quality evidence) and the spontaneous pregnancy rate per couple randomly assigned (16% vs. 7%; Peto OR 4.94, 95% CI 2.13 to 11.44, five studies, 412 participants, I(2) = 0%, moderate-quality evidence) were significantly higher in men receiving gonadotrophin treatment than in men receiving placebo or no treatment. In the more recent meta-analysis by Santi *et al.*, (2015), there was a significantly higher treatment-related spontaneous pregnancy rate, with an OR of 4.5 (CI: 2.17–9.33) and OR of 1.60 (CI: 1.08–2.37) for pregnancy after ART [24]. However, it must be recognized that both meta-analyses included studies of low to moderate quality, flawed by methodological issues: significant heterogeneity due to differing types of gonadotropins used (hCG, hMG, purified FSH, and recombinant FSH), different treatment duration, and no reporting on live birth rates. The latest systematic review and meta-analysis by Omar *et al.*, (2019) in idiopathic male infertility reiterated that the quality of included studies is low, found some improvements in semen parameters in the FSH-treated patients, but due to methodological flaws it was difficult to draw conclusions on pregnancy and live births [25]. Therefore, based on current evidence, the use of gonadotrophins in the management of idiopathic male infertility cannot be recommended.

SERMs

An empirical therapy for idiopathic male infertility is the use of SERM (which may have mixed/partial agonistic and antagonistic properties in different tissues). The basic rationale is to inhibit the negative feedback of oestrogen on the hypothalamus and pituitary, thereby increasing endogenous gonadotropin secretion in men where these levels are low. In turn, FSH and LH stimulate Leydig cells with the possible increase in testosterone production and spermatogenesis [26]. Clomiphene citrate (CC) and tamoxifen are the most commonly used, CC in doses of 25–50 mg daily and tamoxifen in 10–30 mg doses. Based on low-quality meta-analysis involving 11 trials (where 10 out of 11 were published before 1997) including only few randomized controlled trials (RCTs), small studies, majority not accounting for the female fertility status, the evidence of a beneficial effect of SERMs in idiopathic male infertility is questionable [26]. The use of SERMs for idiopathic male infertility is not recommended.

Androgens

The use of androgens is sadly still being misused for treatment of idiopathic male infertility [27]. Exogenous testosterone has a suppressive effect on spermatogenesis, being a highly effective male contraceptive [27] (see Chapter 10.8, 'Hormonal Male Contraception'). Androgens are contraindicated in men seeking conception [19].

Antioxidants

Contemporary studies have found elevated reactive oxygen species reactive oxygen species (ROS) levels in ejaculated spermatozoa in 30% to 80% of infertile men [28]. It has been postulated, based on *in vitro* studies, that ROS damage the sperm membrane, decreasing sperm motility and its ability to fertilize the oocyte. In addition, ROS can alter the sperm DNA, resulting in the passage of defective paternal DNA on to the conceptus [28].

The use of antioxidants, assessed in 48 RCTs which investigated: magnesium, zinc, folic acid, N-acetylcysteine, coenzyme Q10, vitamins E and C, selenium, docosahexaenoic acid (DHA), various carnitines, and pentoxifylline, was summarized in a Cochrane review on in 4170 subfertile men [29]. Again, the methodological quality of most of these trials is poor and often focused only on

sperm parameters or DNA damage as the main outcome measure, rather than pregnancy rates. Despite suggestions that oral antioxidant therapy might improve seminal oxidative status and possibly sperm motility in infertile men, [30] there is no evidence to support any benefits on pregnancy outcomes.

Assisted Reproduction for Idiopathic Male Infertility

Since there is no evidence that empirical medical treatment of idiopathic male infertility is effective, most infertile couples with significant male factors should be advised to consider evidence-based methods of assisted reproduction techniques [19]. The flow of ART is usually from less invasive, low-cost, low-technology such as intrauterine insemination (IUI) to high-cost, more invasive assisted reproductive technologies of *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI) [31, 32].

Intrauterine Insemination

IUI is one of the most frequently used fertility treatments for couples with male subfertility. Intrauterine insemination is a technique which is used to increase the number of motile sperm that are morphologically normal at the site of fertilization by injecting a small volume of prepared semen transcervically into the uterine cavity at the expected time of ovulation. The rationale behind this procedure is to bypass the cervix and to bring the semen closer to the oocyte released into the Fallopian tube.

The increasing use of IUI in idiopathic male infertility is mainly the result of the refinement of techniques for the preparation of washed motile spermatozoa, as used in IVF procedures [31]. Retrospective data from the European Society of Human Reproduction report a live birth of 8.9% following IUI for couples with subfertility including unexplained infertility and mild male factor subfertility [33].

The most recent Cochrane review on assisted reproductive techniques for male subfertility [34] compared IUI versus timed intercourse in natural (unstimulated) cycles and found no significant difference between pregnancy rates per woman. Neither were there significant differences between pregnancy rates per couple for IUI with ovarian stimulation versus IUI in natural cycles. However, the quality of studies included in this Cochrane review was low and no firm conclusions could be drawn [34].

In 2013 the United Kingdom's National Institute for Clinical Excellence (NICE) guidelines on fertility treatment advised against routinely offering IUI to subfertile couples but to proceed immediately to treatment with *IVF* after a period of expectant management of two years in women under 36 years of age [35]. It is interesting that the majority of the fertility clinics in the United Kingdom did not adhere to these guidelines [36]. Indeed, the most recent recommendation from a systematic review [35] regarding the treatment of couples with idiopathic male factor infertility with a total motile count (TMC = volume × density × percentage progressive motile spermatozoa/100) of >10 million is to offer at least three cycles of IUI [37]. It is however accepted that IUI is not a treatment option for couples with exceeding poor sperm quality and IVF or ICSI is the most appropriate treatment [19].

IVF/ICSI

Assisted reproductive techniques (IVF and ICSI) aim to increase the probability of fertilization by bringing the spermatozoa closer to or even within the oocyte(s), thereby bypassing some functional deficits of the male gametes. In IVF/ICSI gamete interaction and fertilization take place outside the woman's body.

Conventional IVF involves insemination of the oocyte with a number of processed motile spermatozoa.

In ICSI the rate of fertilization is enhanced by microinjecting one single spermatozoon directly into the oocyte's cytoplasm. The introduction of ICSI in 1992 has been an unparalleled revolutionary treatment of male infertility [38]. Almost three decades later, ICSI is now by far the most used infertility treatment for both male and non-male infertility [39].

When the TMC is too low to perform IUI (post processing TMC below 1 million) or when pregnancy is not achieved after 3–4 IUIs in couples with moderate male factor subfertility with at least 1 million motile sperm after semen preparation, the next treatment option is IVF [19].

The success rates of IVF/ICSI are increasing in comparison with IUI where pregnancy rates have remained unchanged over the years. After three IVF/ICSI cycles, the cumulative live birth rate in the population of patients with male infertility was 74% when the age of the female partner is below 37 years and 57% between 37 and 40 years [19].

An important debated issue is whether ICSI should be preferred to conventional IVF for circumventing male factor infertility. Even if the sperm quality is normal or near normal, ICSI has overtaken conventional IVF in routine clinical practice in an effort to maximize fertilization rates [40]. A meta-analysis including 11 studies with a total of 901 couples (female age range 30–35 years) with 11,767 sibling oocytes showed that the pooled relative risk (RR) of a mature oocyte fertilizing was higher with ICSI than with conventional IVF (RR 1.49, 95% confidence interval [CI] 1.35–1.65.) The pooled RR of total fertilization failure was significantly higher with conventional IVF than with ICSI (RR 8.22, 95% CI 4.44–15.23) [41].

Although there are currently no studies directly comparing either ICSI vs. IUI or IVF vs. ICSI for idiopathic male factor infertility, as shown by the most recent Cochrane review [34], the therapeutic mindset of preferred use of ICSI can be explained by the fact that about five out of ten patients will have their own child within three treatment cycles [42–44].

Conclusions

The diagnosis of idiopathic male infertility, made by exclusion after the standard work-up in the male partner, is more than an assessment of the semen. Its management is closely linked to the fertility status (and age) of the female partner and would often benefit from a multidisciplinary approach, consisting of a urologist or clinical andrologist, gynaecologist, and endocrinologist. Given the assumption that idiopathic male infertility may have a genetic background, the future role of a geneticist will probably become more important as genome-wide carriership screening enter the diagnostic scene. The treatment of men with idiopathic infertility should commence with advocating a healthy lifestyle in an attempt to increase natural fecundity. There is however no

proven benefit of food supplements and oral antioxidant preparations. Based on current evidence, empiric hormonal treatment has no role in unexplained male infertility. ART have gained considerable momentum in the management of idiopathic male infertility: IUI and/or ICSI are increasingly used to shorten time to pregnancy in couples with idiopathic/unexplained male infertility. However, more prospective trials focusing on cumulative live birth rates of these procedures are needed.

REFERENCES

1. WHO. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva: World Health Organization, 2010.
2. Nieschlag E, Behre HM, Nieschlag S. *Male Reproductive Health and Dysfunction*. Berlin Heidelberg: Springer, 2010.
3. Tournaye H, Krausz C, Oates RD. Novel concepts in the aetiology of male reproductive impairment. *Lancet Diabetes Endocrinol*, 2017; **5**: 544–53.
4. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med* 2012b; **9**: e1001356
5. WHO. Infertility is a global public health issue. Available at: http://www.who.int/reproductivehealth/topics/infertility/perspective/en/(accessed: 18 April 2016).
6. Rutstein SO, Shah IH. *Infecundity, Infertility, and Childlessness in Developing Countries*. DHS comparative reports no. 9. Geneva: World Health Organization, 2004
7. Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. *Reprod Biol Endocrinol*, 2015; **13**: 37.
8. Ghorbel M, Gargouri Baklouti S, Ben Abdallah F, *et al*. Chromosomal defects in infertile men with poor semen quality. *J Assist Reprod Genet*, 2012; **29**: 451–56.
9. Kim JW, Park SY, Ryu HM, *et al*. Molecular and clinical characteristics of 26 cases with structural Y chromosome aberrations. *Cytogenetic Genome Res*, 2012; **136**: 270–7.
10. Neto FTL, Bach PV, Najari BB, Li PS, Goldstein M. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol*, 2016; **59**: 10–26.
11. Chalmel F, Lardenois A, Evrard B *et al*. Global human tissue profiling and protein net work analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum Reprod*, 2012; **27**: 3233–48.
12. Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med*, 2008; **14**: 1197–213.
13. Yan W. Male infertility caused by spermiogenic defects: lessons from gene knockouts. *Mol Cell Endocrinol*, 2009; **306**: 24–32.
14. Bracke A, Peeters K, Punjabi U, Hoogewijs D, Dewilde S. A search for molecular mechanisms underlying male idiopathic infertility. *Reprod Biomed Online*, 2018; **36**: 327–39.
15. Krausz C, Escamilla AR, Chianese C. Genetics of male infertility: from research to clinic. *Reproduction*, 2015; **150**: R159–74.
16. Mitchell MJ, Metzler-Guillemain C, Aminata T, Coutton C, Arnoult C, Ray PF. Single gene defects leading to sperm quantitative anomalies. *Clin Genet*, 2017; **91**: 208–16.
17. Ray PF, Toure A, Mitchell MJ. Genetic abnormalities leading to qualitative defects of sperm morphology or function. *Clin Genet*, 2017; **91**: 217–32.
18. Punab M, Poolamets O, Paju P, *et al*. Causes of male infertility: a 9-year prospective monocentre study on 1737 patients with reduced total sperm counts. *Hum Reprod*, 2017; **32**: 18–23.
19. Tournaye H, Krausz C, Oates RD. Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabetes Endocrinol*, 2017; **5**: 554–64.
20. Homan GF, Davies M, Norman R. The impact of lifestyle factors on reproductive performance in the general population and those undergoing infertility treatment: a review. *Hum Reprod Update*, 2007; **13**: 209–23.
21. Kumar S, Kumari A, Murarka S. Lifestyle factors in deteriorating male reproductive health. *Indian J Exp Biol*, 2009; **47**: 615–24.
22. Chehab M, Madala A, Trussell JC. On-label and off-label drugs used in the treatment of male infertility. *Fertil Steril*, 2015; **103**: 595–604.
23. Attia AM, Abou-Setta AM, Al-Inany HG. Gonadotrophins for idiopathic male factor subfertility. *Cochrane Database Syst Rev*, 2013; **6**: CD005071.
24. Santi D, Granata ARM, Simoni M. FSH treatment of male idiopathic infertility improves pregnancy rate: a meta-analysis. *Endocr Connect*, 2015; **4**: R46–58.
25. Omar MI, Pal RP, Kelly BD, *et al*. Benefits of empiric nutritional and medical therapy for semen parameters and pregnancy and live birth rates in couples with idiopathic infertility: a systematic review and meta-analysis. *Eur Urol*, 2019; **75**: 615–25.
26. Chua ME, Escusa KG, Luna S, Tapia LC, Dofitas B, Morales M. Revisiting oestrogen antagonists (clomiphene or tamoxifen) as medical empiric therapy for idiopathic male infertility: a meta-analysis. *Andrology*, 2013; **1**: 749–57.
27. Ko EY, Siddiqi K, Brannigan RE, Sabanegh ES. Empirical medical therapy for idiopathic male infertility: a survey of the American Urological Association. *J Urol*, 2012; **187**: 973–8.
28. Tremellen K. Oxidative stress and male infertility—a clinical perspective. *Hum Reprod Update*, 2008; **s**: 243–58.
29. Showell MG, Mackenzie-Proctor R, Brown J, Yazdani A, Stankiewicz MT, Hart RJ. Antioxidants for male subfertility. *Cochrane Database Syst Rev*, 2014; **30**: CD007411.
30. Ross C, Morriss A, Khairy M, *et al*. A systematic review of the effect of oral antioxidants on male infertility. *Reprod Biomed Online*, 2010; **20**: 711–23.
31. National Collaborating Centre for Women's and Children's Health. *Clinical Guideline 11, Fertility: Assessment and Treatment for People with Fertility Problems*. Commissioned by the National Institute for Clinical Excellence (NICE). London: NICE, 2004.
32. Royal College of Obstetrics and Gynaecology (RCOG). *Infertility Guidelines Group. The Management of Infertility in Secondary Care*. London: RCOG, 1998.
33. Kupka MS, Ferraretti AP, de Mouzon J, *et al*. Assisted reproductive technology in Europe, 2010: results generated from European registers by ESHRE†. *Hum Reprod*, 2014; **29**: 2099–113.
34. Cissen M, Bensdorp A, Cohlen BJ, Repping S, de Bruin JP, van Wely M. Assisted reproductive technologies for male subfertility. *Cochrane Database Syst Rev*, 2016; **2**: CD000360.
35. National Institute for Health and Clinical Excellence. Assessment and Treatment for People with Fertility Problems, 2013. Clinical guideline [CG156]. Available at: https://www.nice.org.uk/guidance/cg156 (accessed: 17 March 2020).
36. Kim D, Child T, Farquhar C. Intrauterine insemination: a UK survey on the adherence to NICE clinical guidelines by fertility clinics. *BMJ Open*, 2015; **5**: e007588–8.

37. Cohlen B, Bijkerk A, Van der Poel S, Ombelet W. IUI: review and systematic assessment of the evidence that supports global recommendations. *Hum Reprod Update*, 2018; **24**: 300–19.
38. Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *Lancet*, 1992; **340**: 17–18.
39. Evers JLHH. Santa Claus in the fertility clinic. *Hum Reprod*, 2016; **31**: 1381–2.
40. Boulet SL, Mehta A, Kissin DM, Warner L, Kawwass JF, Jamieson DJ. Trends in use of and reproductive outcomes associated with intracytoplasmic sperm injection. *JAMA*, 2015; **313**: 255–63.
41. Johnson LNC, Sasson IE, Sammel MD, Dokras A. Does intracytoplasmic sperm injection improve the fertilization rate and decrease the total fertilization failure rate in couples with well-defined unexplained infertility? A systematic review and meta-analysis. *Fertil Steril*, 2013; **100**: 704–11.
42. Osmanagaoglu K, Tournaye H, Camus M, Vandervorst M, Van Steirteghem A, Devroey P. Cumulative delivery rates after intracytoplasmic sperm injection: 5-year follow-up of 498 patients. *Hum Reprod*, 1999; **14**: 2651–5.
43. Malizia BA, Hacker MR, Penzias AS. Cumulative live birth rates after *in vitro* fertilization. *N Engl J Med*, 2009; **360**: 236–43.
44. Gnoth C, Maxrath B, Skonieczny T, Friol K, Godehardt E, Tigges J. Final ART success rates: a 10 years survey. *Hum Reprod*, 2011; **26**: 2239–46.

10.6

Hypothalamo–Pituitary–Testicular Axis Function in Systemic Diseases and Effects of Medications

Gary Wittert, Bu B. Yeap, and Mathis Grossmann

Introduction

Organic pathology of the hypothalamic-pituitary-testicular (HPT) axis resulting in low serum testosterone concentrations (i.e. 'pathological hypogonadism') occurs in under 2.5% of the male population. However, epidemiological studies have shown that with increasing age up to 40% of men have serum testosterone concentrations that are low compared to those found in young healthy males [1]. In many older men low testosterone may be attributable to acute and chronic health conditions external to the HPT axis, medication use and lifestyle behaviours (i.e. 'functional hypogonadism'). This chapter describes the effects of acute and chronic disorders and medications on the HPT axis. The important message to take away is that a low serum testosterone concentration may be a 'barometer' of overall poor physical and psychological health and well-being and when found should prompt assessment for and optimization of ill health rather than the knee-jerk prescription of testosterone.

Undernutrition

Both under- and overnutrition can disrupt the function of the HPT axis. For example, a study undertaken in non-obese men showed that a 7-day period of overfeeding decreased serum testosterone by 6%. Following a subsequent 3-week period of calorie restriction (50% of energy requirements) sufficient to reduce body weight by 7.5% and fat mass by 17.8%, serum testosterone concentrations decreased by a further 11%. After 14 days of refeeding, testosterone increased by 6% on average. Over these three varying food intake periods, there was a 14% decrease, followed by a 41% increase and then a 40% decrease respectively, in serum sex-hormone-binding-globulin (SHBG) [2].

In states of chronic energy deficit, either because food intake is insufficient or physical activity is excessive relative to nutrient intake, there is a positive association between fat mass and serum testosterone that appears to be mediated by leptin. The starvation-associated leptin decline signals energy deficits to the hypothalamus, with subsequent HPT axis suppression intended to limit procreation and promote survival. Restoration of fat mass leads to an increase in both leptin and serum testosterone concentrations [3].

Obesity

Obesity is the clinical condition most strongly associated with lowered testosterone concentrations in men [4].

In men aged 18–55 years with class 1–2 obesity (BMI 30–40 kg/m^2), but who are otherwise healthy, reductions in total testosterone primarily reflect obesity-associated reductions in its circulating

carrier protein SHBG [5]. These men have no clinical evidence of clinical androgen deficiency or of defective spermatogenesis. Men with class 3 obesity (BMI >40 kg/m^2) have reductions in total and free testosterone concentrations with decreased luteinizing hormone (LH) diurnal concentrations and pulse amplitudes compared to lean controls [6]. Despite these changes, studies have not found evidence of clinical androgen deficiency, in these men [5, 6].

The obesity-hypogonadism relationship has subsequently been examined in large population-based studies of middle-aged and older men with varying degrees of comorbid burden that by and large excluded men with classical hypogonadism. These studies have confirmed that obesity is the single most important factor associated with low testosterone, overriding the effects of age and comorbidities [4].

In many of these epidemiological studies, low testosterone was associated with androgen deficiency-like symptoms and clinical markers of androgen deficiency such as sexual dysfunction, reduced lean mass, reduced physical performance, reduced bone density, and anaemia. However, observational studies cannot clarify whether low testosterone is causing these clinical features, or whether instead obesity and associated comorbidities lead to HPT axis suppression and non-specific symptoms and signs that mimic androgen deficiency.

Although variations in circulating testosterone can influence fat mass, current evidence suggests that the effects of obesity on testosterone concentrations are more substantial than the effects of testosterone on adiposity. Weight loss is associated with increases in serum testosterone in proportion to the amount of weight lost. Following bariatric surgery (Roux-en Y Gastric Bypass and Sleeve Gastrectomy) total testosterone, gonadotropins and free testosterone all increase, suggesting a genuine reactivation of the HPT axis [7].

In contrast, changes in circulating testosterone concentrations induced either by testosterone treatment [8] or testosterone deprivation [9] have limited effects on overall body weight, although significant changes in body composition occur. Reductions in body weight with testosterone treatment have been reported in registry studies, this has to date not been demonstrated in randomized controlled trials (RCTs) [8].

Therefore, in obese men presenting with lowered testosterone, after classical hypogonadism has been excluded by an individualized approach, the first line therapeutic approach should be to treat the obesity and its related comorbidities (particularly depression and sleep apnoea), optimize health-related behaviours to induce weight loss and, bariatric surgery (preferably a Roux-en-Y gastric bypass) when indicated for those with intractable and severe obesity [10]. Given the lack of long-term, large RCTs powered for patient-important outcomes and long-term safety, testosterone treatment cannot be routinely recommended for men with obesity-associated non-classical hypogonadism [11].

Type 2 Diabetes and the Metabolic Syndrome

Observational studies consistently show that men with type 2 diabetes (T2D) and/or the metabolic syndrome have modestly reduced (−1.6 mmol/L) circulating testosterone concentrations, with 30–50% of them having low concentrations relative to reference ranges based on healthy young men, after adjustment for age and obesity [12]. However, some studies suggest it is the low SHBG that underlies this relationship [13]. There is a feedforward effect of low testosterone concentrations associated with obesity and the metabolic syndrome so as to increase the likelihood of developing T2D [1]. Androgen deficiency-like symptoms in diabetic men with lowered testosterone are non-specific and their presence is more closely correlated with increasing age rather than diabetes status in itself [14].

Low testosterone concentrations in men with T2D are independently associated with insulin resistance [15]. Experimental evidence suggests that testosterone promotes the commitment of pluripotent stem cells into the myogenic lineage, inhibits their differentiation into adipocytes [16] and regulates the metabolic functions of mature adipocytes and myocytes in ways that reduce insulin resistance. For example, testosterone has been reported to promote lipolysis in male rat adipocytes [17], and to prevent abdominal fat accumulation in men by regulating adipocyte triglyceride uptake and lipoprotein lipase activity [18]. Moreover, testosterone promotes skeletal muscle glucose uptake by activating the glucose metabolism-related glucose transporter type 4 (GLUT4) signalling pathway in skeletal muscle cells [19]. Reducing testosterone concentrations to near zero by androgen deprivation therapy in men with prostate cancer leads to insulin resistance [9], and testosterone therapy reduces fat mass and increases lean body mass [8].

The evidence from randomized controlled clinical trials of testosterone treatment on glycaemic outcomes are inconclusive. A meta-analysis of seven RCTs including 833 men with type 2 diabetes and/or the metabolic syndrome concluded that testosterone treatment may modestly improve surrogate measures of insulin resistance but had no effect on glycaemic control (assessed by HbA1c) or on androgen deficiency-like symptoms [20]. While larger, longer-term studies are needed, current guidelines recommend against testosterone treatment in men with type 2 diabetes and low testosterone as means of improving glycaemic control [11].

Cardiovascular Disease

The presence of cardiovascular disease (CVD) is associated with lower endogenous testosterone concentrations. In the acute setting of hospitalization for unstable angina or myocardial infarction, this reflects a variable but generally moderate suppression of the HPT axis which may be secondary to the acute event, physiological stress and accompanying use of medications such as opioids for analgesia [21]. Testosterone concentration at the time of acute admission appears to have limited utility as a predictor for outcomes and routine measurement of serum testosterone is not recommended in this context. Assessment for underlying hypogonadism in men with symptoms and signs suggestive of androgen deficiency should only be conducted after recovery from the acute cardiovascular event. In one study, endogenous testosterone concentrations had largely recovered when re-assessed 6 months after the acute admission [21].

Testosterone concentrations are reduced in community-dwelling men with prevalent CVD [22–24] by a central mechanism of action reflecting reduction of HPT axis function. Of note, relative reductions in HPT axis function resulting in lower endogenous testosterone concentrations are associated with incidence of cardiovascular events such as stroke, independently of conventional cardiovascular risk factors [25]. However, studies might not have accounted for sleep apnoea, depression, and other confounders.

Testosterone treatment was associated with adverse cardiovascular events in one clinical trial, but other trials found no excess of such adverse events [26–28]. Thus, ongoing uncertainty exists over the effects of testosterone treatment on the cardiovascular system. Whether testosterone treatment has beneficial, neutral, or harmful effects on the cardiovascular system remains to be clarified by the conduct of adequately powered clinical trials (refer to Chapter 10.4.5, 'Risks of Testosterone Treatment').

Chronic Obstructive Airways Disease

In community-dwelling men higher testosterone concentrations are associated with better indices of pulmonary function, independently of age, smoking status, and other covariates [29]. Conversely, men with chronic obstructive airways disease commonly exhibit lower testosterone concentrations compared to men without apparent pulmonary disease [30], which may be the result of chronic hypoxia, inflammation, and to some extent the use of glucocorticoid medication. Of note, men with chronic obstructive airways disease exhibit clinical features that overlap androgen deficiency, particularly loss of lean muscle mass [31]. Further research is required to establish whether testosterone therapy will safely and efficaciously improve clinically relevant health outcomes in men with chronic obstructive airways disease in the absence of pathological hypogonadism.

Chronic Kidney Disease

Serum testosterone is low in men in men with chronic kidney disease associated with the decreasing estimated glomerular filtration rates (eGFR) and lower haemoglobin concentrations. There is an independent relationship between the degree of visceral adiposity and lower testosterone concentrations among men with chronic kidney disease stages 3 to 5 not undergoing dialysis [32]. With end-stage renal disease testosterone concentrations are in the hypogonadal range in 40–60% of men [33]. The most common mechanism is decreased testosterone synthesis in Leydig cells resulting in hypergonadotropic hypogonadism [34] although in some men with significant hyperprolactinaemia (common in renal failure), gonadotrophins may not be elevated. Testosterone concentrations increase rapidly following recovery from renal transplantation [35] particularly in patients under the age of 50 [33]. Haemodialysis is thought not to improve testosterone concentrations, but this may depend on the technique used [36] and in some centres, well-dialysed patients have normal serum testosterone.

Chronic Liver Disease

In men with chronic liver disease, serum testosterone is reduced in proportion to the severity of the disease. Serum testosterone is reduced in up to 90% of men with cirrhosis, and an independent predictor of mortality [37]. SHBG concentrations are increased in moderate to severe disease but are low normal in end-stage disease. Serum oestradiol is increased. Multiple mechanisms contribute to HPT axis dysregulation, including non-specific illness effects, porta-systemic shunting and altered sex steroid clearance [38]. Aetiology-dependent toxins such as alcohol and iron deposition (in haemochromatosis) contribute. Increased oestradiol is thought to mediate gynaecomastia and spider naevi, whereas low testosterone contributes to sarcopaenia and testicular atrophy. In clinical trials, testosterone treatment robustly increases lean mass, but effects on patient-important outcomes are unknown [39]. Successful liver transplantation markedly improves HPT axis dysregulation, suggesting the effects are functional [40].

Cancer

Primary hypogonadism because of toxicity of prior cancer therapy (particularly in childhood) may merit testosterone supplementation to treat symptoms and signs of androgen deficiency. Cancer can also negatively impact on central HPT axis function, particularly in the presence of metastatic or advanced disease and when anorexia and cachexia are present [41]. Lower testosterone concentrations can also reflect use of opioid analgesia and are frequently associated with symptoms such as weight loss, depression, and poor sense of well-being [42]. In an RCT of late stage cancer patients, seven weeks of testosterone therapy at standard replacement dose increased lean body mass (compared with placebo recipients who lost lean mass) and improved quality of life [43].

Androgen deprivation therapy (ADT) for prostate cancer has deleterious effects on metabolic parameters, and body composition [11, 18].

Inflammatory Disorders

Men with hypogonadism are at increased risk of developing autoimmune disorders particularly those affecting the musculoskeletal system [44]. Testosterone concentrations are lower in men with rheumatoid arthritis than in healthy men, but in those who respond to disease-modifying treatment, serum testosterone increase over a 2-year period [45]. Inflammatory bowel disease is associated with low testosterone concentrations in men; the extent to which this is the disease itself or the treatment is unclear [46]. The testosterone-suppressive effects of acute and chronic inflammation have been reported to be mediated by inflammatory cytokines on steroidogenesis in Leydig cells [47, 48].

Thyroid Disease

Thyroid hormones regulate several aspects of testicular function, including steroidogenesis and spermatogenesis [49]. Small studies in adult men with hyperthyroidism have reported that circulating SHBG, gonadotrophins, total testosterone and oestradiol are increased compared to euthyroid controls, whereas the opposite changes occur in hypothyroid men [50]. Both hyper- and hypothyroidism are associated with erectile dysfunction and reduced libido, while premature ejaculation is increased in hyperthyroid men. Delayed ejaculation more frequent in hypothyroid men. These alterations generally normalize once euthyroidism is restored [49]. Reversible abnormalities in spermatogenesis have been described in both hyper- and hypothyroid men. Gynaecomastia is considered to be the consequence of increases in the oestradiol to testosterone ratio as a result of the increased SHBG [50].

Infections

In men with severe septic shock HPT axis function is suppressed with resultant low LH and testosterone concentrations [51]. Older men hospitalized for respiratory tract infections have also been found to have lower testosterone concentrations with either normal or elevated LH [52]. Thus, where pre-existing androgen deficiency is suspected on clinical grounds assessment for pathological hypogonadism should be deferred until after recovery from an acute infection. Of note, impaired HPT axis function with secondary hypogonadism has also been reported as chronic sequelae of central nervous system infections [53].

There has been considerable interest in the effect of human immunodeficiency virus (HIV) infection on the function of the HPT axis [54]. Early studies commented on the ability of HIV manifesting as the acquired immunodeficiency syndrome (AIDS) to involve multiple endocrine systems including the HPT axis with biochemical features of secondary hypogonadism in affected men [55]. Low testosterone concentrations in men with AIDS appeared to correlate with weight loss and with depressive symptom scores [55]. Furthermore, in RCTs involving men with AIDS wasting syndrome, testosterone improved depressive symptoms scores and increased lean and muscle mass with a corresponding improvement in well-being [56]. Other testosterone RCTs in men with HIV and low testosterone concentrations have reported increases in lean and muscle mass and reductions in fat mass [57, 58]. In the current era of highly active antiretroviral therapy, only a minority of HIV-infected men are found to have low testosterone concentrations in the presence of low or normal LH and increased adiposity, with an even smaller proportion having elevated LH [54]. A current guideline suggests consideration of short-term testosterone therapy in HIV-infected men with low testosterone concentrations and weight loss (when other causes have been excluded) to benefit body composition noting that the evidence to support this recommendation was of low quality [59].

Disorders of Sleep

The production of testosterone is dependent on sleep rather than an endogenous rhythm entrained to the light dark cycle. Circulating testosterone begins to increase with the onset of sleep, peaks at the first rapid eye movement (REM) sleep episode and remains stable thereafter until waking. If REM latency is increased, then the rise in testosterone will be slower [60]. Therefore, the effect of sleep restriction on serum testosterone will depend on the timing of occurrence and duration. For shift workers serum testosterone will peak at the time of waking and provided the minimum sleep requirements are met, concentrations will be normal.

Testosterone deficiency is associated with reduced slow-wave sleep and overall sleep efficiency that is improved by treatment with testosterone [60].

Obstructive sleep apnoea (OSA) is not an independent cause of low testosterone after adjustment for the degree of obesity [60, 61] and treatment of OSA with continuous positive airways pressure does not in and of itself increase serum testosterone [62]. There is a widespread belief, reflected in the recent US Endocrine Society guideline for treatment of hypogonadism [59], that untreated sleep apnoea is a contraindication to testosterone treatment. However, recent data have shown that, testosterone replacement therapy in obese men with severe OSA leads to a slight transient worsening of some indices of sleep-disordered breathing that recover after 3–4 months of continued treatment [63]. That said, the prevalence of OSA has increased dramatically in association with the obesity epidemic, and OSA, even if severe may not be associated with subjective sleepiness [64]. When clinically suspected it should be actively excluded even if not considered a contraindication to initiation of testosterone treatment.

From a practical standpoint lower serum testosterone concentrations, poor energy, sexual dysfunction, obesity, and OSA often coexist and are difficult to disentangle. Treatment of OSA by weight reduction and/or continuous positive airway pressure (CPAP) may lead to significant symptomatic improvement without resorting to testosterone replacement.

Psychiatric Disorders

Eating Disorders

Anorexia nervosa is associated with hypogonadism. Although adolescent and young men considered to comprise 10–15% of patients diagnosed with anorexia nervosa [65] the proportion with the disorder is likely to be closer to 30% and many cases are overlooked [66]. Men with anorexia more frequently use exercise as a compensating mechanism as compared to women and may look muscular rather than particularly thin. Presentation may be with a dominant concern about muscle size. Hypogonadism resolves with weight gain, implying it is functional.

Approximately 36% of the patients suffering from bulimia nervosa are male. This is not a condition in and of itself known to be specifically associated with low testosterone.

Major Depressive Disorder and Anxiety Disorder

Compared with healthy men, serum testosterone concentrations are lower in older men with major depressive disorder (MDD) but not in those with MDD and comorbid anxiety [67]. In a longitudinal cohort study of community-dwelling middle-aged and older men, severe depression was associated with lowered serum testosterone [23], and a very large population-based cohort study showed an inverse association of serum testosterone with depressive symptom burden [68]. There is recent evidence that the reduced testosterone in MDD may be limited to a subgroup of atypical depressives [69].

Although symptoms such as dysphoria, lack of motivation, fatigue, and lethargy, and decreased libido that present in organic hypogonadism may improve with testosterone treatment, these symptoms, in the context of hypogonadism constitute a different construct to MDD-associated low testosterone and evidence to implicate testosterone in the pathogenesis of MDD is lacking.

Post-traumatic stress disorder is associated with sexual dysfunction but there are no associations with testosterone although, dehydroepiandrosterone (DHEA) has been reported to be higher in men with post-traumatic stress disorder (PTSD) as compared to healthy men in one small study [70].

Psychoses

Disorders such as schizophrenia are not associated with low testosterone in and of themselves. Antipsychotic medications induce

increases in prolactin and obesity, both of which reduce testosterone concentrations by a central mechanism of action.

Medications

Medications may lower serum testosterone concentrations by mechanisms that impair: hypothalamic gonadotrophic-releasing hormone (GnRH) and/or pituitary gonadotroph LH production, either directly or consequent to increased prolactin; Leydig cell function and pathways of steroidogenesis in the testis; binding of testosterone to the androgen receptor or induce enzymes that metabolize testosterone. Some drugs affect the function of either 5-α reductase or aromatase altering biotransformation of testosterone to dihydrotestosterone and oestradiol, respectively.

Medications (Long acting GnRH agonists, 5-α reductase and aromatase inhibitors and androgen receptor antagonists) specifically designed to inhibit one or more components of the HPT axis for therapeutic purposes are discussed elsewhere in this volume. Table 10.6.1 lists medications that may affect the HPT axis, and the mechanism/s whereby this effect occurs.

Table 10.6.1 Medications and mechanisms of effect on the hypothalamo–pituitary–testicular axis

	Hypothalamic/ pituitary	Steroidogenesis	Degradation CYP3A4	Biotransformation	Androgen receptor	SHBG
Analgesics						
Opioids	Direct ↑ Prolactin			↓ aromatase		
Psychotropics						
^Antipsychotics	↑ Prolactin					
SSRI antidepressants	The effects are unresolved. Some studies suggest a decrease while others no effect.					
Clomipramine	↑ Prolactin					
Lithium	↑ Prolactin					
Antiepileptics						
Phenytoin			↑↑			↑↑
Carbamazepine			↑↑			↑↑
Topiramate			↑↑			
Anti-inflammatory						
Glucocorticoids	Direct	↓↓↓	↑↑			
Antimicrobial						
Azole antifungals		↓↓↓	↓↓			
Rifampicin			↑↑			
Clarithromycin		↓↓	↓↓			
Chloramphenicol		↓↓	↓↓			
Protease inhibitors		↓↓	↓↓			
Cardiovascular						
Verapamil, Diltiazem			↓			
Spironolactone		↓			++	
Atenolol						
Lipid-lowering						
Statins (Simvastatin)		↓				
Diabetes						
Glitazones		↓	↑↑			
Gastrointestinal						
Cimetidine					+	
Other						
Modafinil			↑↑			
Capsaicin			↑↑			

The medications listed are those associated with reductions in serum testosterone concentrations and the mechanism/s by which this occurs. CP3A4 it is an enzyme of the cytochrome P-450 complex involved in the degradation of testosterone.

In practice opioid prescribed for analgesia is the medication most frequently implicated in in low testosterone concentrations with low or normal LH in men [71]. There is a worldwide scourge of overprescribing opioids for chronic musculoskeletal pain [72] and the appropriate management is to withdraw the opioids. Treatment with testosterone has not been shown to affect self-reported pain or opioid dosage [73].

Although associated with major adverse consequences for spermatogenesis (see Chapter 10.4.3), there are limited data relating to the effects of chemotherapeutic agents on serum testosterone concentration in men. In one study [74], 8.8% of adult survivors of cancer treatment during childhood or adolescence had low serum testosterone concentrations; none had been commenced on replacement therapy. Based on the Children's Oncology Group Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancer (COG-LTFU) Guidelines, the risk is greatest with high dose alkylating agents or testicular radiation above 20 Gy. The testis has greater resistance to radiotherapy post puberty. Pubertal status does not affect the risk of hypogonadism associated with chemotherapy. Follow-up should be lifelong. Mild, subclinical Leydig cell dysfunction can become overt due to interaction with any of the disorders described earlier. Treatment with testosterone is beneficial for overt hypogonadism but there is no evidence of benefit from mild or subclinical Leydig cell dysfunction [75].

Glucocorticoids remain widely prescribed and the magnitude of effect to reduce serum testosterone concentration depends dose, duration, and individual patient sensitivity. Long-term treatment with even modest doses of glucocorticoid leads to reduced bone density and muscle mass and increased risk of fracture. Treatment with testosterone has been shown, independent of the starting plasma concentration, to increase muscle mass and strength, reduce fat mass, and improve bone density at the lumbar spine in chronic steroid treated men [76]. To date, however, studies have been small, short-term, and there are no data on fracture outcomes. There has been no trial since 2003 and clearly further study of this neglected area is warranted. Until then, in the absence of contraindications, it is reasonable to consider treatment with testosterone on an individual patient basis.

Antipsychotic medications lower serum testosterone by increasing prolactin. This is best managed wherever possible using antipsychotics with either no or minimal effect on serum prolactin. These include aripiprazole, clozapine, olanzapine, and quetiapine. OSA occurs with substantially increased frequency and severity in people affected by schizophrenia treated with antipsychotics and should be diagnosed and treated. Testosterone treatment is beneficial for men with schizophrenia and hypogonadism and may ameliorate negative symptoms.

Selective serotonin reuptake inhibitors (SSRI) are the medications most commonly prescribed to treat depression. It is unlikely that they have clinically significant effects on serum testosterone concentrations, however they are associated with sexual dysfunction. Those antidepressants that primarily modulate adrenergic (e.g. desvenlafaxine, vortioxetine), dopaminergic (e.g. bupropion) or melatonin (agomelatine) receptor systems attenuate the risk of emergent sexual dysfunction during treatment of depression.

Medications used for the treatment of epilepsy may disrupt the HPT axis either by increasing the degradation of sex steroids by induction of CYP3A4 (e.g. phenytoin, carbamazepine, topiramate) or by increasing SHBG (e.g. phenytoin, carbamazepine). This appears not to occur with newer drugs used for seizure control (e.g. lamotrigine and levetiracetam).

REFERENCES

1. Araujo AB, Wittert GA. Endocrinology of the aging male. *Best Pract Res Clin Endocrinol Metab*, 2011; **25**: 303–19.
2. Muller MJ, Enderle J, Pourhassan M, *et al*. Metabolic adaptation to caloric restriction and subsequent refeeding: the Minnesota Starvation Experiment revisited. *Am J Clin Nutr*, 2015; **102**: 807–19.
3. Wabitsch M, Ballauff A, Holl R, *et al*. Serum leptin, gonadotropin, and testosterone concentrations in male patients with anorexia nervosa during weight gain. *J Clin Endocrinol Metab*, 2001; **86**: 2982–8.
4. Tajar A, Forti G, O'Neill TW, *et al*. Characteristics of secondary, primary, and compensated hypogonadism in aging men: evidence from the European Male Ageing Study. *J Clin Endocrinol Metab*, 2010; **95**: 1810–18.
5. Glass AR, Swerdloff RS, Bray GA, Dahms WT, Atkinson RL. Low serum testosterone and sex-hormone-binding-globulin in massively obese men. *J Clin Endocrinol Metab*, 1977; **45**: 1211–19.
6. Vermeulen A, Kaufman JM, Deslypere JP, Thomas G. Attenuated luteinizing hormone (LH) pulse amplitude but normal LH pulse frequency, and its relation to plasma androgens in hypogonadism of obese men. *J Clin Endocrinol Metab*, 1993; **76**: 1140–6.
7. Corona G, Rastrelli G, Monami M, *et al*. Body weight loss reverts obesity-associated hypogonadotropic hypogonadism: a systematic review and meta-analysis. *Eur J Endocrinol*, 2013; **168**: 829–43.
8. Isidori AM, Giannetta E, Greco EA, *et al*. Effects of testosterone on body composition, bone metabolism and serum lipid profile in middle-aged men: a meta-analysis. *Clin Endocrinol (Oxf)*, 2005; **63**: 280–93.
9. Cheung AS, Hoermann R, Dupuis P, Joon DL, Zajac JD, Grossmann M. Relationships between insulin resistance and frailty with body composition and testosterone in men undergoing androgen deprivation therapy for prostate cancer. *Eur J Endocrinol*, 2016; **175**: 229–37.
10. Grossmann M, Matsumoto AM. A perspective on middle-aged and older men with functional hypogonadism: focus on holistic management. *J Clin Endocrinol Metab*, 2017; **102**: 1067–75.
11. Yeap BB, Grossmann M, McLachlan RI, *et al*. Endocrine Society of Australia position statement on male hypogonadism (part 1): assessment and indications for testosterone therapy. *Med J Aust*, 2016; **205**: 173–8.
12. Ding EL, Song Y, Malik VS, Liu S. Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA*, 2006; **295**: 1288–99.
13. Bhasin S, Jasjua GK, Pencina M, *et al*. Sex hormone-binding globulin, but not testosterone, is associated prospectively and independently with incident metabolic syndrome in men: the framingham heart study. *Diabetes Care*, 2011; **34**: 2464–70.
14. Kapoor D, Aldred H, Clark S, Channer KS, Jones TH. Clinical and biochemical assessment of hypogonadism in men with type

2 diabetes: correlations with bioavailable testosterone and visceral adiposity. *Diabetes Care*, 2007; **30**: 911–17.
15. Grossmann M, Thomas MC, Panagiotopoulos S, *et al*. Low testosterone levels are common and associated with insulin resistance in men with diabetes. *J Clin Endocrinol Metab*, 2008; **93**: 1834–40.
16. Singh R, Artaza JN, Taylor WE, Gonzalez-Cadavid NF, Bhasin S. Androgens stimulate myogenic differentiation and inhibit adipogenesis in C3H 10T1/2 pluripotent cells through an androgen receptor-mediated pathway. *Endocrinology*, 2003; **144**: 5081–8.
17. Xu XF, De Pergola G, Bjorntorp P. Testosterone increases lipolysis and the number of beta-adrenoceptors in male rat adipocytes. *Endocrinology*, 1991; **128**: 379–82.
18. Marin P, Oden B, Bjorntorp P. Assimilation and mobilization of triglycerides in subcutaneous abdominal and femoral adipose tissue *in vivo* in men: effects of androgens. *J Clin Endocrinol Metab*, 1995; **80**: 239–43.
19. Sato K, Iemitsu M, Aizawa K, Ajisaka R. Testosterone and DHEA activate the glucose metabolism-related signaling pathway in skeletal muscle. *Am J Physiol Endocrinol Metab*, 2008; **294**: E961–8.
20. Grossmann M, Hoermann R, Wittert G, Yeap BB. Effects of testosterone treatment on glucose metabolism and symptoms in men with type 2 diabetes and the metabolic syndrome: a systematic review and meta-analysis of randomized controlled clinical trials. *Clin Endocrinol (Oxf)*, 2014; **83**: 344–51.
21. Pesonen E, Pussinen P, Huhtaniemi I. Adaptation to acute coronary syndrome-induced stress with lowering of testosterone: a possible survival factor. *Eur J Endocrinol*, 2016; **174**: 481–9.
22. Yeap BB, Knuiman MW, Divitini ML, *et al*. Differential associations of testosterone, dihydrotestosterone and oestradiol with physical, metabolic and health-related factors in community-dwelling men aged 17–97 years from the Busselton Health Survey. *Clin Endocrinol (Oxf)*, 2014; **81**: 100–8.
23. Shi Z, Araujo AB, Martin S, O'Loughlin P, Wittert GA. Longitudinal changes in testosterone over five years in community-dwelling men. *J Clin Endocrinol Metab*, 2013; **98**: 3289–97.
24. Yeap BB, Alfonso H, Chubb SA, *et al*. Reference ranges and determinants of testosterone, dihydrotestosterone, and estradiol levels measured using liquid chromatography-tandem mass spectrometry in a population-based cohort of older men. *J Clin Endocrinol Metab*, 2012; **97**: 4030–9.
25. Yeap BB. Testosterone and its metabolites: differential associations with cardiovascular and cerebrovascular events in men. *Asian J Androl*, 2018; **20**: 109–14.
26. Basaria S, Coviello AD, Travison TG, *et al*. Adverse events associated with testosterone administration. *N Engl J Med*, 2010; **363**: 109–22.
27. Srinivas-Shankar U, Roberts SA, Connolly MJ, *et al*. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2010; **95**: 639–50.
28. Snyder PJ, Bhasin S, Cunningham GR, *et al*. Effects of testosterone treatment in older men. *N Engl J Med*, 2016; **374**: 611–24.
29. Mohan SS, Knuiman MW, Divitini ML, *et al*. Higher serum testosterone and dihydrotestosterone, but not oestradiol, are independently associated with favourable indices of lung function in community-dwelling men. *Clin Endocrinol (Oxf)*, 2015; **83**: 268–76.
30. Atlantis E, Fahey P, Cochrane B, Wittert G, Smith S. Endogenous testosterone level and testosterone supplementation therapy in chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis. *BMJ Open*, 2013; **3**: e003127.
31. Velema MS, Kwa BH, de Ronde W. Should androgenic anabolic steroids be considered in the treatment regime of selected chronic obstructive pulmonary disease patients? *Curr Opin Pulm Med*, 2012; **18**: 118–24.
32. Cobo G, Cordeiro AC, Amparo FC, Amodeo C, Lindholm B, Carrero JJ. Visceral adipose tissue and leptin hyperproduction are associated with hypogonadism in men with chronic kidney disease. *J Ren Nutr*, 2017; **27**: 243–8.
33. Reinhardt W, Kubber H, Dolff S, Benson S, Fuhrer D, Tan S. Rapid recovery of hypogonadism in male patients with end stage renal disease after renal transplantation. *Endocrine*, 2018; **60**: 159–66.
34. Fugl-Meyer KS, Nilsson M, Hylander B, Lehtihet M. Sexual function and testosterone level in men with conservatively treated chronic kidney disease. *Am J Mens Health*, 2017; **11**: 1069–76.
35. Eckersten D, Giwercman A, Pihlsgard M, Bruun L, Christensson A. Impact of kidney transplantation on reproductive hormone levels in males: a longitudinal study. *Nephron*, 2018; **138**: 192–201.
36. Cigarran S, Coronel F, Florit E, *et al*. Testosterone deficiency in dialysis patients: difference between dialysis techniques. *Nefrologia*, 2017; **37**: 526–30.
37. Grossmann M, Hoermann R, Gani L, *et al*. Low testosterone levels as an independent predictor of mortality in men with chronic liver disease. *Clin Endocrinol (Oxf)*, 2012; **77**: 323–8.
38. Sinclair M, Grossmann M, Gow PJ, Angus PW. Testosterone in men with advanced liver disease: abnormalities and implications. *J Gastroenterol Hepatol*, 2015; **30**: 244–51.
39. Sinclair M, Grossmann M, Hoermann R, Angus PW, Gow PJ. Testosterone therapy increases muscle mass in men with cirrhosis and low testosterone: a randomised controlled trial. *J Hepatol*, 2016; **65**: 906–13.
40. Handelsman DJ, Strasser S, McDonald JA, Conway AJ, McCaughan GW. Hypothalamic-pituitary-testicular function in end-stage non-alcoholic liver disease before and after liver transplantation. *Clin Endocrinol (Oxf)*, 1995; **43**: 331–7.
41. Wiechno PJ, Poniatowska GM, Michalski W, *et al*. Clinical significance of androgen secretion disorders in men with a malignancy. *Med Oncol*, 2017; **34**: 123.
42. Dev R, Hui D, Del Fabbro E, *et al*. Association between hypogonadism, symptom burden, and survival in male patients with advanced cancer. *Cancer*, 2014; **120**: 1586–93.
43. Wright TJ, Dillon EL, Durham WJ, *et al*. A randomized trial of adjunct testosterone for cancer-related muscle loss in men and women. *J Cachexia Sarcopenia Muscle*, 2018; **9**: 482–96.
44. Baillargeon J, Al Snih S, Raji MA, *et al*. Hypogonadism and the risk of rheumatic autoimmune disease. *Clin Rheumatol*, 2016; **35**: 2983–7.
45. Tengstrand B, Carlstrom K, Hafstrom I. Gonadal hormones in men with rheumatoid arthritis—from onset through 2 years. *J Rheumatol*, 2009; **36**: 887–92.
46. O'Toole A, Winter D, Friedman S. Review article: the psychosexual impact of inflammatory bowel disease in male patients. *Aliment Pharmacol Ther*, 2014; **39**: 1085–94.
47. Tremellen K, McPhee N, Pearce K, Benson S, Schedlowski M, Engler H. Endotoxin-initiated inflammation reduces

testosterone production in men of reproductive age. *Am J Physiol Endocrinol Metab*, 2018; **314**: E206–13.
48. Leisegang K, Henkel R. The *in vitro* modulation of steroidogenesis by inflammatory cytokines and insulin in TM3 Leydig cells. *Reprod Biol Endocrinol*, 2018; **16**: 26.
49. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*, 2010; **31**: 702–55.
50. Meikle AW. The interrelationships between thyroid dysfunction and hypogonadism in men and boys. *Thyroid*, 2004; **14** Suppl 1: S17–25.
51. Fourrier F, Jallot A, Leclerc L, *et al*. Sex steroid hormones in circulatory shock, sepsis syndrome, and septic shock. *Circ Shock*, 1994; **43**: 171–8.
52. Iglesias P, Prado F, Macias MC, *et al*. Hypogonadism in aged hospitalized male patients: prevalence and clinical outcome. *J Endocrinol Invest*, 2014; **37**: 135–41.
53. Schaefer S, Boegershausen N, Meyer S, Ivan D, Schepelmann K, Kann PH. Hypothalamic-pituitary insufficiency following infectious diseases of the central nervous system. *Eur J Endocrinol*, 2008; **158**: 3–9.
54. Rochira V, Guaraldi G. Hypogonadism in the HIV-infected man. *Endocrinol Metab Clin North Am*, 2014; **43**: 709–30.
55. Dobs AS, Dempsey MA, Ladenson PW, Polk BF. Endocrine disorders in men infected with human immunodeficiency virus. *Am J Med*, 1988; **84**(3 Pt 2): 611–16.
56. Grinspoon S, Corcoran C, Askari H, *et al*. Effects of androgen administration in men with the AIDS wasting syndrome. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*, 1998; **129**: 18–26.
57. Bhasin S, Storer TW, Javanbakht M, *et al*. Testosterone replacement and resistance exercise in HIV-infected men with weight loss and low testosterone levels. *JAMA*, 2000; **283**: 763–70.
58. Bhasin S, Parker RA, Sattler F, *et al*. Effects of testosterone supplementation on whole body and regional fat mass and distribution in human immunodeficiency virus-infected men with abdominal obesity. *J Clin Endocrinol Metab*, 2007; **92**: 1049–57.
59. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
60. Wittert G. The relationship between sleep disorders and testosterone in men. *Asian J Androl*, 2014; **16**: 262–5.
61. Ruchala M, Brominska B, Cyranska-Chyrek E, Kuznar-Kaminska B, Kostrzewska M, Batura-Gabryel H. Obstructive sleep apnea and hormones—a novel insight. *Arch Med Sci*, 2017; **13**: 875–84.
62. Knapp A, Myhill PC, Davis WA, *et al*. Effect of continuous positive airway pressure therapy on sexual function and serum testosterone in males with type 2 diabetes and obstructive sleep apnoea. *Clin Endocrinol (Oxf)*, 2014; **81**: 254–8.
63. Hoyos CM, Killick R, Yee BJ, Grunstein RR, Liu PY. Effects of testosterone therapy on sleep and breathing in obese men with severe obstructive sleep apnoea: a randomized placebo-controlled trial. *Clin Endocrinol (Oxf)*, 2012; **77**: 599–607.
64. Adams R, Appleton S, Taylor A, McEvoy D, Wittert G. Are the ICSD-3 criteria for sleep apnoea syndrome too inclusive? *Lancet Respir Med*, 2016; **4**: e19–20.
65. Schorr M, Miller KK. The endocrine manifestations of anorexia nervosa: mechanisms and management. *Nat Rev Endocrinol*, 2017; **13**: 174–86.
66. Zhang C. What can we learn from the history of male anorexia nervosa? *J Eat Disord*, 2014; **2**: 138.
67. Giltay EJ, van der Mast RC, Lauwen E, Heijboer AC, de Waal MWM, Comijs HC. Plasma testosterone and the course of major depressive disorder in older men and women. *Am J Geriatr Psychiatry*, 2017; **25**: 425–37.
68. Kische H, Pieper L, Venz J, *et al*. Longitudinal change instead of baseline testosterone predicts depressive symptoms. *Psychoneuroendocrinology*, 2018; **89**: 7–12.
69. Rodgers S, Grosse Holtforth M, Hengartner MP, *et al*. Serum testosterone levels and symptom-based depression subtypes in men. *Front Psychiatry*, 2015; **6**: 61.
70. Lehrner A, Flory JD, Bierer LM, Makotkine I, Marmar CR, Yehuda R. Sexual dysfunction and neuroendocrine correlates of posttraumatic stress disorder in combat veterans: preliminary findings. *Psychoneuroendocrinology*, 2016; **63**: 271–5.
71. Vuong C, Van Uum SH, O'Dell LE, Lutfy K, Friedman TC. The effects of opioids and opioid analogs on animal and human endocrine systems. *Endocr Rev*, 2010; **31**: 98–132.
72. Hagemeier NE. Introduction to the opioid epidemic: the economic burden on the healthcare system and impact on quality of life. *Am J Manag Care*, 2018; **24**(10 Suppl): S200–S6.
73. Basaria S, Travison TG, Alford D, *et al*. Effects of testosterone replacement in men with opioid-induced androgen deficiency: a randomized controlled trial. *Pain*, 2015; **156**: 280–8.
74. Yoon JY, Park HJ, Ju HY, *et al*. Gonadal and sexual dysfunction in childhood cancer survivors. *Cancer Res Treat*, 2017; **49**: 1057–64.
75. Kenney LB, Cohen LE, Shnorhavorian M, *et al*. Male reproductive health after childhood, adolescent, and young adult cancers: a report from the Children's Oncology Group. *J Clin Oncol*, 2012; **30**: 3408–16.
76. Crawford BA, Liu PY, Kean MT, Bleasel JF, Handelsman DJ. Randomized placebo-controlled trial of androgen effects on muscle and bone in men requiring long-term systemic glucocorticoid treatment. *J Clin Endocrinol Metab*, 2003; **88**: 3167–76.

10.7

Management of Sexual Dysfunction

Vincenzo Rochira, Cesare Carani, and Antonio R.M. Granata

Introduction

Sexuality is a complex construct encompassing not only the act of sexual intercourse. but also includes the physiological, behavioural, and relational aspects of human sexual life, which are variously influenced by psychological (e.g. sexual fantasies, desire, arousal, psychosexual orientation, and the choice of the sexual object), as well as social and organic (vascular, nervous, and endocrine) factors [1, 2].

From a male sexual functional point of view, a normal penile erection may be defined as an erection that permits the penetration of a lubricated vagina without additional assistance. Concerning the erectile mechanism, the haemodynamic changes in the penis require a high degree of central and peripheral neurovascular coordinated control and an intact endocrine system [1, 3, 4].

Penile Anatomy

The penis consists of two paired and elongated sponge-like bodies, the *corpora cavernosa*. These are ventrally joined to the *corpus spongiosum*, all of which are covered by a tough fibroelastic sheath, the *tunica albuginea*. The *corpus spongiosum* surrounds the urethra and distally forms the glans [3] (**Figure 10.7.1a**).

The erectile tissue of the *corpora cavernosa* is composed of sinusoidal spaces, also known as lacunae, lined by endothelial cells. The lacunae are surrounded by trabecular tissue, consisting of an intricate network of both smooth muscle cells and fibroblasts joined by collagen and elastin. The trabeculae have both structural and contractile functions [3]. The structural function is provided by the extracellular, fibroelastic component of collagen and elastin; contraction of the smooth muscle component is associated with the flaccid state of the penis while relaxation leads to lacunae engorgement and penile erection (**Figure 10.7.1b**).

Arterial inflow to the penis is supplied by the helicine arteries which branch out from the two deep cavernosal arteries. When the penis is flaccid, the helicine arteries are contracted. During erection arterial relaxation causes an increase in blood flow and a consequent enlargement of the lacunar spaces. The structural elasticity of the trabeculae allows the increase in penile size from the flaccid to the erected state [3] (**Figure 10.7.1b**). Blood drainage from the *corpora cavernosa* is provided by the circumflex and emissary veins which end in the deep dorsal penile vein.

Innervation of the penis comes from the thoracolumbar spinal centre of erection (T12–L2), the sacral spinal centre of erection (S2–S4), the hypogastric plexus, and the pelvic, cavernous, and pudendal nerves [3, 4].

Erectile Physiology

An erection occurs as a consequence of haemodynamic changes in the penile structures induced by integration of one or more central and peripheral neurological signalling events. Erections are mediated by the spinal centres, which can be activated by genital sensory stimulation and by neural pathways from the brain. Normal erection depends on multiple regulatory inputs including psychological, neurological, and hormonal components that are closely integrated and reinforce each other [1, 2, 4] (**Figure 10.7.2**). For example, psychological stimuli (i.e. visual, auditory, olfactory, and internal imagery—imagination/fantasy) increase the sensitivity to reflexive (i.e. tactile) stimuli (**Figure 10.7.2**). *Vice versa*—undesired tactile stimuli (i.e. painful stimulation) might inhibit central psychogenic activity (**Figure 10.7.2**).

Central erotogenic stimulation promotes sexual arousal, a subjective state which stimulates the man to search for sexual stimulation and sexual intercourse [2].

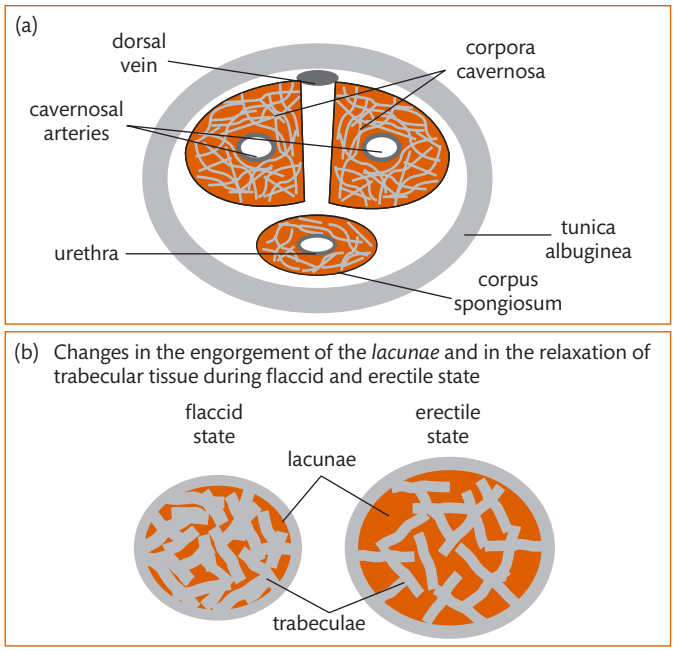


Figure 10.7.1 Schematic representation of the gross penile anatomy (a) and of the different engorgement of the tissue of *corpora cavernosa* and *corpus spongiosum* during the flaccid and the erectile state (b).

The main areas of the central nervous system involved in the promotion of the male sexual response are the limbic system, the preoptic region, and the hypothalamus, while the sympathetic pathway that exerts an inhibitory control on erection corresponds to the sympathetic thoracolumbar spinal erection centre (T12–L2) [1, 2, 4].

In addition to psychogenic erections, direct physical stimulation of the penis and perineal skin induces a reflexogenic erection by activation of the sacral spinal reflex pathway. This results from the inhibition of the thoracolumbar sympathetic centre of erection (T12–L2), a stimulation of the parasympathetic centre of erection (S2–S4), and the activation of both parasympathetic and non-adrenergic non-cholinergic (NANC) pathways [4] (Figure 10.7.3).

Furthermore, sleep-related erections occur during rapid eye movement (REM) sleep; however, the precise physiological mechanism involved in sleep-related erections is not thoroughly understood.

The endocrine system also contributes to penile erection. Androgens are the hormones involved in erectile physiology. Most of the actions of testosterone on the brain seem to be mediated by oestrogens after aromatization [5–7]. Dihydrotestosterone also plays a role on penile erection as suggested by erectile dysfunction as a side effect of 5-α-reductase-inhibitors treatment for benign prostatic hyperplasia and hair loss [8].

Prenatal and perinatal brain androgenization is a prerequisite for normal male sexual function in some species, and androgens are necessary for secondary sexual maturation in mammals [1].

In adult men, androgens are required to maintain sexual behaviour; lack of testosterone frequently produces loss of libido and sexual dysfunction [1, 9]. In hypogonadal men testosterone replacement therapy consistently increases spontaneous sexual interest and sexual arousal thereby facilitating erectile function and increasing sexual activities [9]. Note however that most cases of erectile dysfunction (without a major loss of sexual interest) are predominantly associated with vascular diseases and not testosterone deficiency (see next). Similarly, erections in response to visual erotic stimuli are not affected by the lack of testosterone [10]. In contrast, nocturnal (sleep-related) or early morning erections are androgen-dependent, being impaired in hypogonadal men and restored by testosterone replacement therapy [11].

Local Mechanisms of Erection

Neurological activation or inhibition of penile erection is mediated by modulation of neurotransmitter release in the penile vasculature [3, 12] (Figure 10.7.2). Knowledge of neurotransmitters responsible for the local regulation of erection has recently been extended by

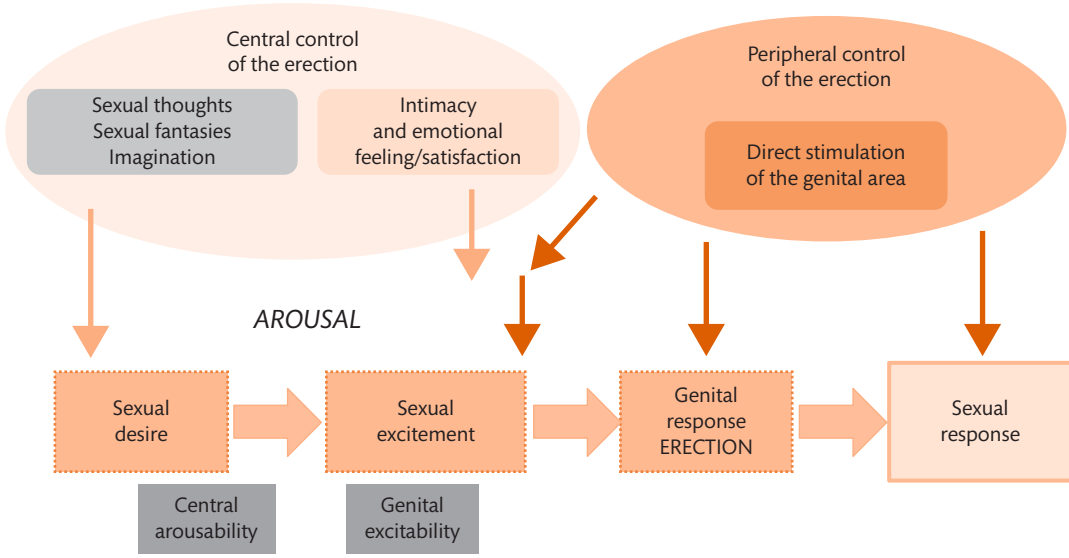


Figure 10.7.2 Mechanisms involved in the central and peripheral control of penile erection during the various phases of male sexual response from initiation to the achievement of a full erection.

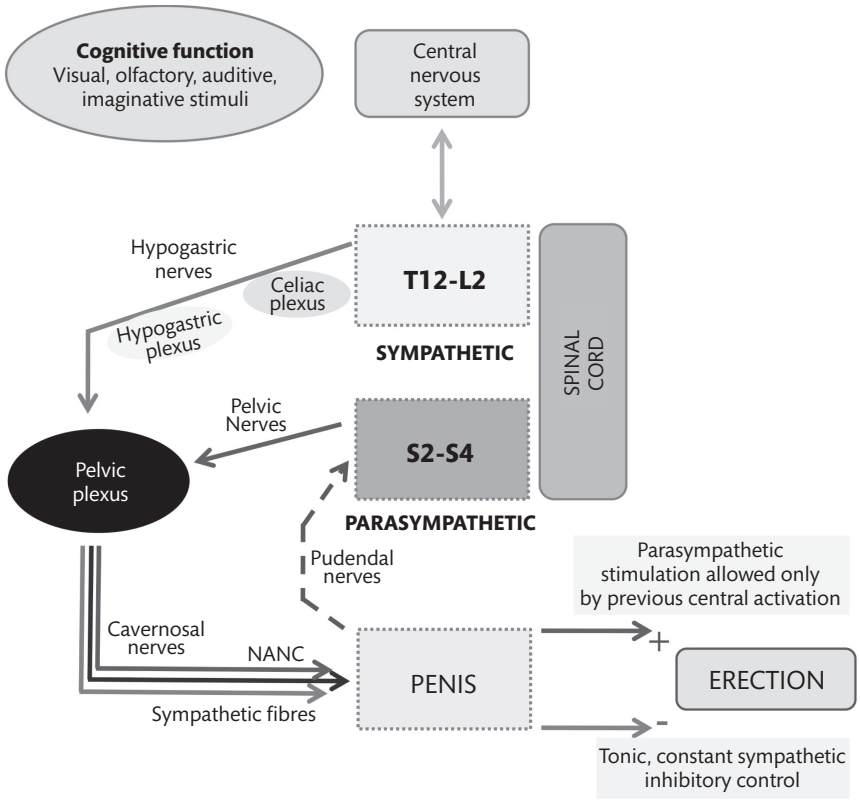


Figure 10.7.3 Central and peripheral nervous pathways involved in the control of erection.
T10–L2: toracholumbar spinal centre; S2–S4: sacral spinal centre; NANC: non-adrenergic-noncholinergic neurons/fibres.

the identification of NANC fibres, which are the terminal effectors of the nervous system in controlling erection [3, 12]. The activation of cholinergic fibres stimulates postganglionic NANC nerve neurons to synthesize and release nitric oxide, which is the principal neurotransmitter involved in the haemodynamic promotion of penile erection [3, 12]. The sympathetic system exerts a constant negative control on erection by inhibiting postganglionic NANC neurons [3, 12]. Nitric oxide is released by both neuronal fibre endings and endothelial cells, but under physiological conditions neuronal nitric oxide plays the dominant role [13].

Nitric oxide is a cleavage product whose synthesis from amino acid L-arginine is catalysed by nitric oxide synthase (NOS). Nitric oxide relaxes the trabecular smooth muscle cells of corpora cavernosa by increasing the synthesis of cyclic guanosine monophosphate (cGMP) via a direct stimulation of guanylate cyclase [3, 12, 13] (Figure 10.7.4). The cGMP activates protein kinase G, which then causes a reduction of transmembrane Ca^{2+}influx, and membrane hyperpolarization. This sequence of events induces smooth muscle relaxation of trabeculae (Figure 10.7.4). Residual cGMP is catabolized by phosphodiesterase 5 (PDE5) [3, 12, 13] (Figure 10.7.4).

Based on animal studies, it has been suggested that testosterone may directly promote erections through the penile nitric oxide pathway by stimulating both neuronal NOS synthesis/release and by modulating PDE5 activity [14–16]. However, the relevance of this work to human erectile physiology is uncertain.

Other substances, including neurotransmitters, hormones, prostaglandins, and other peptides also play a role in the local control of erection. In particular, prostaglandin E_1 and vasoactive intestinal peptide (VIP) relax the smooth muscle, while substance P and prostaglandin F_2-α cause contraction [3, 12]. Prostaglandin E_1 induces smooth muscle relaxation by reducing intracellular Ca^2+ uptake, as a result of the activation of adenylate cyclase and a consequent increase of intracellular cyclic adenylate monophosphate (cAMP), which activates protein kinase A (Figure 10.7.4) [3, 12].

Vascular Response

Erection and detumescence are linked haemodynamic events regulated by smooth muscle relaxation and contraction, respectively. Therefore, the tone of the *corpora cavernosa* smooth muscle is the major determinant in controlling the flaccid or erect state of the penis. The haemodynamic events involved in penile erection are as follows [3, 12, 13]: (1) the resistance of intracavernosal arterioles decreases by relaxation of the trabecular smooth muscle cells; (2) the dilatation of the arterial bed (particularly helicine arteries) increases the arterial flow causing [17] the engorgement of sinusoids and the engorgement of lacunae; (4) these events cause an increase in penile tumescence/rigidity and penile length; (5) the stretching of the poorly distensible tunica albuginea and the expansion of the lacunae activate a veno-occlusive mechanism with reduced venous outflow due to compression of the subtunical and emissary veins [3, 12, 13].

Erectile Dysfunction

Erectile dysfunction had its first official definition in 1993: the inability to attain and/or maintain a penile erection sufficient to

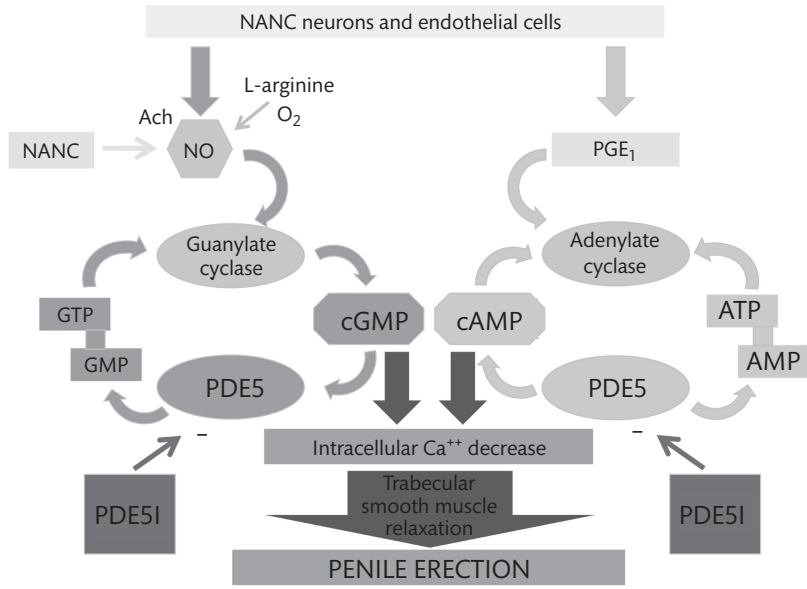


Figure 10.7.4 Signalling pathways mediating the erectile response.

Ach, acetylcholine; O_2, oxygen; NO, nitric oxide; NANC, non-adrenergic-noncholinergic neurons; PGE_1, prostaglandin E_1; GTP, guanosine triphosphate; GMP, guanosine monophosphate; cGMP, cyclic guanosine monophosphate; cAMP, cyclic adenosine monophosphate; ATP, adenosine triphosphate; AMP, adenosine monophosphate; PDE5, phosphodiesterase type 5; PDE5I, phosphodiesterase type 5 Inhibitors; Ca^{++}, calcium.

permit satisfactory sexual performance (Box 10.7.1) [18]. A more recent definition in 2000 [19] stated that erectile dysfunction is the persistent or recurrent inability to attain, or to maintain until completion of sexual activity, an adequate erection.

Erectile dysfunction *per se*, should be differentiated from other aspects of sexual dysfunctions such as loss of libido, decreased spontaneous erections, or disorders of ejaculation, and disorders of orgasm. Loss of libido can be the consequence of both hypogonadism [1, 2, 20, 21] and psychogenic erectile dysfunction [22]; on the other hand, quality of spontaneous (especially nocturnal) erections is related to organic rather than psychogenic erectile dysfunction [23, 24]. It is important to be reminded that erection constitutes only part of the overall multifaceted process of normal sexual function, which also includes psychological, behavioural, and relational components that can contribute to the overall picture of sexual dysfunctions.

Aetiologies and Epidemiology of Erectile Dysfunction

Erectile dysfunction (ED) may be caused by psychological, organic (vascular, metabolic, neurological, hormonal), and iatrogenic disorders [13, 25, 26]. Up until the 1960s, psychological disorders were considered the most frequent cause of erectile dysfunction [26]; at present this is true only in young men; about 70% of men less than 35 years old presenting with erectile dysfunction, the main cause is psychogenic [13, 25, 26]. In middle-aged and older men, organic disorders now account for two-thirds of the causes of erectile dysfunction [13, 26]. In the Massachusetts Male Aging Study, the combined prevalence of minimal, moderate, and complete impotence was 52% in the general population. The prevalence of complete impotence tripled from 5 to 15% between subject ages 40 and 70 years. Erectile dysfunction correlated positively with age, cigarette smoking, depression, diabetes mellitus, and cardiovascular diseases [26, 27].

Among organic causes, vascular disorders are the most frequent [26]; neurological causes occur in 3–10% of the cases [26]. Although hormone abnormalities (e.g. hypogonadism, hyperprolactinaemia, hyperthyroidism, acromegaly) can be detected in 3–5% of patients with ED [9, 13, 26], low T frequently **coexists** with ED in men with chronic diseases (such as obesity, diabetes mellitus, HIV infection, cardiovascular/peripheral vascular disease, and chronic kidney disease) [28–31], rather than being the overriding *cause* of the erectile failure [32].

Although erectile dysfunction may be exclusively of psychological origin, a psychological disorder often arises secondarily to organic erectile dysfunction [33]; therefore, organic and psychogenic erectile dysfunction often coexist in many patients.

Box 10.7.1 Definitions of erectile dysfunction

1993: The inability to attain and/or maintain a penile erection sufficient to permit satisfactory sexual performance [12]

2000: The persistent or recurrent inability to attain, or to maintain until completion of the sexual activity, an adequate erection [13]

Psychogenic Erectile Dysfunction

Psychogenic erectile dysfunction is frequently associated with generalized trait anxiety, situational anxiety (e.g. performance anxiety), relationship conflicts, disorders involving sexuality (psychosocial sexual inhibition, sex-preference conflicts, history of childhood sexual abuse, disorders of sexual orientation), fear of pregnancy, fear of sexually transmitted infections, sexual performance anxiety, fear of erectile failure, and decreased libido [23, 26, 29, 34]. The suggested mechanism in primary psychogenic erectile dysfunction is an activation of the sympathetic nervous

system with increased adrenergic and noradrenergic vasoconstrictive tone [3, 13, 26, 33].

Psychogenic erectile dysfunction can also occur in primary psychiatric diseases, such as depression [24, 34]. Loss of libido is an important symptom of depression, which may lead secondarily to erectile dysfunction and is possibly worsened by antidepressant drug administration [23, 34]. Accordingly, serotoninergic antidepressants may lead to mild to severe sexual dysfunction (i.e. decreased libido and delayed orgasm frequently) in >60% of patients, while other sexual dysfunctions such as anorgasmia and arousal difficulties occur less frequently (30%) [35] (**Box 10.7.2**). In contrast, noradrenergic, dopaminergic, or melatonergic antidepressants do not cause sexual dysfunction but the clinical profile of patients receiving these compounds could be different [35].

Vasculogenic Erectile Dysfunction

Vasculogenic erectile dysfunction is caused by reduced arterial inflow into the penis or an impaired veno-occlusive mechanism, or both, and its frequency increases with ageing. Arterial insufficiency is a common cause of erectile dysfunction and results from many disorders, including thrombotic and thromboembolic occlusion of the terminal aorta and the iliac, hypogastric, pudendal, and penile arteries. Atherosclerosis is the most common cause of arterial occlusion; therefore, arterial penile insufficiency shares the same risk factors for atherosclerosis, including age, cigarette smoking, hypertension, dyslipidaemia, diabetes mellitus, and the metabolic syndrome [28, 36]. Arterial penile insufficiency may also be due to congenital anomalies, surgery, and perineal or pelvic trauma [13, 37].

Impairment of the veno-occlusive mechanism of erection may cause erectile dysfunction because of venous leakage with increased venous blood outflow. Failure of the veno-occlusive mechanism may occur at many levels: tunica albuginea, trabecular tissue, endothelial cells, nerve fibres, and veins [37].

Diseases of the tunica albuginea include fibrosis, Peyronie's disease, penile fracture, diabetes, trauma, and congenital tunical abnormalities (e.g. reduced tunica thickness) [3, 13, 26].

Trabecular compliance is impaired when changes in the smooth muscle/collagen ratio occurs. Morphological and functional changes in cavernous smooth muscle may be due to smooth muscle

Box 10.7.2 Principal drugs affecting erection according to the strength of the body of evidence

(In parenthesis): 1: good evidence; 2: moderate evidence; 3: poor evidence

Antihypertensive drugs
- *Diuretics*
  - Thiazide diuretics (1)
  - Spironolactone (1)
- *Sympatholytics*
  - β-blockers (1)
- *Central agents*
  - Methyldopa (1)
  - Clonidine (1)
- *Calcium-antagonists (2)*
- *ACE inhibitors (2)*

Psychotropic drugs
- *Antidepressant drugs*
  - Monoamine oxidase inhibitors (2)
  - Tricyclics (1)
  - Selective serotonin reuptake inhibitors (1)
  - Serotonin noradrenaline reuptake inhibitors (2)
  - Lithium (3)
- *Antipsychotics*
  - Chlorpromazine (3)
  - Fluphenazine (2)
  - Loxapine (3)
  - Mesoridazine (3)
  - Molindone (3)
  - Thioridazine (3)
  - Thiothixene (3)
  - Trifluoperazine (3)
  - Haloperidol (2)
  - Clozapine (2)
  - Olanzapine (2)
  - Quetiapine (2)
  - Risperidone (2)
  - Aripiprazole (3)
  - Ziprasidone (3)
- *Others*
  - Sulpiride (3)
  - Barbiturates (3)
  - Benzodiazepines (2)
  - Anticonvulsant (2)
  - Diphenylhydantoin (1)
  - Ketoconazole (3)
  - Disopyramide (2)

Gastrointestinal drugs
- *H^2 blockers*
  - Cimetidine (3)
  - Ranitidine (1)
- Metoclopramide (1)

Cytotoxic agents
- Methotrexate (2)
- Mitotane (1)

Hormonal drugs
- LHRH-agonists or antagonists (1)
- Antiandrogens
  - Flutamide (2)
  - Cyproterone acetate (1)
  - 5-α-reductase inhibitors [Finasteride] (2)*
- Oestrogens (at high doses) (2)
- Progestins (3)
- Corticosteroids (3)

Miscellaneous
- Digoxin (2)
- Clofibrate (3)
- Gemfibrozil (3)

Recreational drugs
- *Illicit drugs*
  - Cocaine (2)
  - Opioids (1)
  - Cannabinoids (Cannabis/Marijuana/Hashish) (2)
- *Alcohol [high consumption] (1)*

* see the specific paragraph at the end of this chapter for further details

atrophy with replacement by fibrotic tissue (as in atherosclerosis and diabetes), and to local functional disorders with impaired neurotransmitter release or receptor function [3]. *In vitro* studies suggest that testosterone exerts a trophic effect on the penile nerves, vasculature and cavernosal tissue and regulates nitric oxide release by enhancing both the expression and activity of NOS [16], but the significance of these animal data to human physiology and clinical erectile dysfunction remains unclear.

Venous drainage may be impaired by congenital abnormal or ectopic veins and by pathological shunts (traumatic, post-priapism, surgical, or congenital) between the *corpus cavernosum* and the *corpus spongiosum* of the glans penis [13, 25, 26].

Endocrine Erectile Dysfunction

Among the endocrine disorders, men with organic hypogonadism due to established hypothalamic–pituitary commonly present with low libido and may have erectile dysfunction [9]. However, in older men with age-related low testosterone in the absence of recognizable hypothalamus–pituitary–testicular HPT axis pathology, sexual dysfunction is commonly caused by coexisting neurovascular disease, rather than due to androgen deficiency. Organic hypogonadism is not a common finding in patients with ED, occurring in about 2–32% of cases depending on how hypogonadism is defined, age of patients and different clinical settings [38]. In general, there is a lack of association between serum testosterone levels and erectile function [38]. In men over 50, ED and hypogonadism are common but independently distributed rather than causally linked disorders [32]. Among men presenting with erectile dysfunction, low serum T below 300 ng/dl (10.4 nmol/L) may be detected in 12% of 1022 consecutive ED patients, including 4% under and 15% over the age of 50 [39]. The prevalence of low serum T <200 ng/dl, 7 nmol/L was even lower: 1.8%, 0.8% before 50 years, and 2.6% after this age [39]. Even then, this association does not prove causality, given the evidence is observational [38] (see Chapter 10.4.2, 'Types of Treatment').

Hyperprolactinaemia with or without hypotestosteronaemia may also be associated with loss of libido and erectile dysfunction [9], but does not appear to modify sleep-related erections and response to visual erotic stimulation, suggesting that hyperprolactinaemia negatively affects libido and sexual behaviour directly [40]. Hyperprolactinaemia is common (30–65%) in patients with chronic kidney disease (CKD); however the specific contribution of hyperprolactinaemia to the very high frequency (60–80%) of erectile dysfunction in men with CKD remains unclear due to concomitant prevalent vascular abnormalities [30, 31]. Among the other endocrinopathies associated with erectile dysfunction are adrenal insufficiency, acromegaly, Cushing's syndrome, hyperthyroidism, and hypothyroidism [9, 41–43]. Again, the exact role and the mechanisms by which these conditions contribute to erectile malfunction are unclear.

Neurological Erectile Dysfunction

Brain, spinal cord, and neuropathic diseases can cause neurogenic erectile dysfunction. Lesions in various brain areas may induce erectile and sexual dysfunction (e.g. Parkinson's disease, cerebrovascular or expansive lesions of either temporal lobes or limbic system) [13, 44]. Congenital (e.g. spina bifida, syringomelia) or acquired (traumas, neoplasia, and inflammatory disorders as multiple sclerosis) spinal cord diseases are the most frequent causes of neurogenic erectile dysfunction [13, 44]. Injuries to lower spinal cord segments (lumbar and sacral) often result in complete erectile dysfunction, while lesions of the upper spinal cord segments (cervical and thoracic) do not affect reflexogenic erections [4, 44]. In the peripheral nervous system, erectile dysfunction can result from traumatic (pelvic fracture) or surgical (radical prostatectomy, cystoprostatectomy, proctocolectomy) injuries to the pudendal and cavernous nerves [44]. Peripheral neuropathy in diabetes mellitus (the most common cause), uraemia, amyloidosis, vitamin deficiency (folic acid, B_6, and B_{12}), and alcoholism can cause erectile dysfunction [44].

Iatrogenic Erectile Dysfunction

Drugs

Erectile function can be affected by many drugs acting through different mechanisms (i.e. penile arterial flaw, central action on libido, changes in hormonal secretion/action) (**Box 10.7.2**) [3, 25, 26]. Data on drugs and erectile dysfunction are often collected by uncontrolled studies and/or case reports; therefore, a cause–effect relationship is not always proven.

Erectile dysfunction is commonly associated with the use of antihypertensive drugs (thiazide diuretics, β-blockers). Some drugs affecting the hypothalamus–pituitary–testicular axis (oestrogens, progestins, antiandrogens, both luteinizing hormone-releasing-hormone agonists, and antagonists) can impair erectile function by decreasing gonadotropin and testosterone release. The peripheral bioavailability of androgens is decreased by digitalis, cimetidine, and spironolactone. Hyperprolactinaemia, and the often-related low libido and erectile dysfunction, can result from antipsychotic drugs (phenothiazines, sulpiride, and related drugs) and H_2-receptor antagonists. Alcohol abuse can cause low testosterone serum levels, peripheral neuropathy, and chronic liver damage [3, 25, 26].

Surgery

Pelvic surgery, such as radical prostatectomy, cystoprostatectomy, proctocolectomy, is a frequent cause of erectile dysfunction due to pelvic vessel and/or pelvic nerve lesions [13, 25, 26, 37, 44], occurring in 20–80% of patients undergoing the above-mentioned surgical procedures, the prevalence depending on several factors such as the extent of surgery, the surgical technique used, the skills of the surgeon [13, 25, 26, 37, 44]. Also external pelvic radiotherapy can cause erectile dysfunction, with delayed onset 2–5 years after treatment. Comparison between patients with prostate cancer who underwent radical prostatectomy or radiotherapy shows higher prevalence after the former treatment (79% versus 63%) [45], with apparently lower frequency when a robot-assisted radical prostatectomy is performed [46]. Renal transplantation and vascular surgery on the aortoiliac arteries may also cause erectile dysfunction [31].

Trauma

Pelvic fractures can cause erectile dysfunction if injury to a vessel or to nerves involved in the erectile mechanism occurs [37, 44].

Diagnosis

More than one cause can be simultaneously responsible for erectile dysfunction at the time of the diagnosis, even if they may have occurred at different times.

The diagnostic approach to erectile dysfunction should include several sequential steps which may be influenced by the expertise (psychologist/sexual medicine, endocrinologist, urologist) of the clinician (Figure 10.7.5).

General Clinical Interview

A general medical interview and examination can identify systemic diseases, medications/drugs or previous surgery/trauma which can cause erectile dysfunction [13, 23, 25, 26, 34].

Sexological Interview

The sexological interview constitutes a main step, because it may indicate either psychogenic or organic erectile dysfunction [13, 23, 25, 26, 34].

Some information concerning sexual orientation, the presence of one, more, or no constant partner, sexual habits, and frequency of sexual intercourse is required [23, 34]. Other information (e.g. cultural level, religious beliefs, or profession) is also helpful [34].

A sexological interview can reveal psychogenic erectile dysfunction suggested by: abrupt onset, full morning erections, full erections following visual erotic stimulation, full erections with masturbation, inconstant occurrence, occurrence only with some form of sexual intercourse, and occurrence with one partner, but not with another [23, 26, 34].

Conversely, organic erectile dysfunction is often characterized by a progressive loss of erectile function and by an almost constant presence of the problem [13, 23, 25, 26, 34].

Libido needs to be enquired carefully. When a patient complains of low libido, the physician should try to determine if what is occurring is loss of libido, or unwillingness to face erectile dysfunction [13, 23, 25, 26, 34]. Low libido justifies an evaluation of testosterone and prolactin levels, and consideration of depression, or the use of neuroleptics or antiandrogens [13, 21, 23, 25, 26]. Some structured interview instruments for use with patients complaining of erectile dysfunction or other sexual problems are currently available [26, 47, 48]. The International Index of Erectile Function (IIEF) is a validated and extensively used instrument in research and in the clinical work-up of individual patients with erectile dysfunction [34], while others are more suitable for research [48]. The short version of the IIEF (IIEF-5; 5 items) is a self-administered questionnaire that may be used as a diagnostic tool in evaluating erectile function (EF domain), ascertaining if erectile dysfunction is present (EF score ≤25) or not (EF score >25), as well as assessing the degree of erectile function impairment (mild, moderate or severe for EF scores of 17–25, 11–16, 6–10, respectively) [17, 48, 49]. The long version of the IIEF (IIEF-15; 15 items) provides further information also on the domains concerning sexual desire, orgasmic function, intercourse satisfaction, and overall satisfaction [17, 48]. Other questionnaires (e.g. the Aging Male Symptom Scale for male hypogonadism) [50] are not recommended due to poor clinical discriminating power [34, 48].

The interview should also be directed to check other symptoms of hypogonadism such as the decrease of the volume of the ejaculate, the presence of hot flushes and the changes in body hair and beard growth and distribution. See Chapter 10.4.4, 'Benefits of Testosterone Treatment'.

Psychological Evaluation

A psychological evaluation of the patients affected by erectile dysfunction is almost always helpful because of the very frequent involvement of psychological factors, even when the causes are organic [23, 26]. Psychological evaluation should ideally be performed formally by a psychologist with experience/training in psychosexual problems. However, when this is not readily available, self-completed, brief questionnaires are available in order to provide a picture of the patient's personality and anxiety levels, and to detect clinical depression [33, 47].

Physical Examination

The physical examination should focus on the presence of signs of hypogonadism (i.e. reduced facial and body hair, reduced volume of the testes, reduced muscle mass, and increased body fat), the presence of penile fibrosis, the peripheral pulses and the presence of signs of peripheral and autonomic nervous system abnormalities.

Blood Chemistry

Erectile dysfunction is associated with almost all severe endocrinological diseases, particularly hypogonadism. For this reason, measuring serum morning fasting total testosterone is mandatory in all patients with erectile dysfunction, especially in those with low sexual desire or symptoms and signs of hypogonadism [23, 51] (Figure 10.7.5). Further diagnostic procedures are needed in case of low serum total testosterone [51] (see Chapter 10.4.4, 'Benefits of Testosterone Treatment'). Prolactin should be measured in all subjects who complain of low libido and have documented low serum testosterone and low or inappropriately normal gonadotropins [9] (Figure 10.7.5) (see Chapter 2.3.9, 'Prolactinomas and Hyperprolactinaemia').

Fasting blood glucose should be checked because of the high prevalence of erectile dysfunction in diabetics. Cholesterol and triglycerides should also be measured in order to detect risk factors for arteriosclerosis (Figure 10.7.5).

Further Investigations

Further investigations aiming to better define the cause and mechanism of erectile dysfunction may be pursued in selected patients managed under specialized clinical settings. Documenting the presence of full erections under defined circumstances may be useful to rule out an organic cause of impotence. Visual erotic stimulation by movies, slides, or magazines provides an inexpensive test, where the erectile response is self-scored by the patient. A full erection in response to visual erotic stimuli indicates that both vascular and nerve supplies involved in erection are intact. However, erection in response to visual erotic stimuli may not occur (despite intact neurovascular system) because of embarrassment or because the patient is accustomed/desensitized to visual erotic stimulation. Furthermore, a full erection can occur during visual erotic

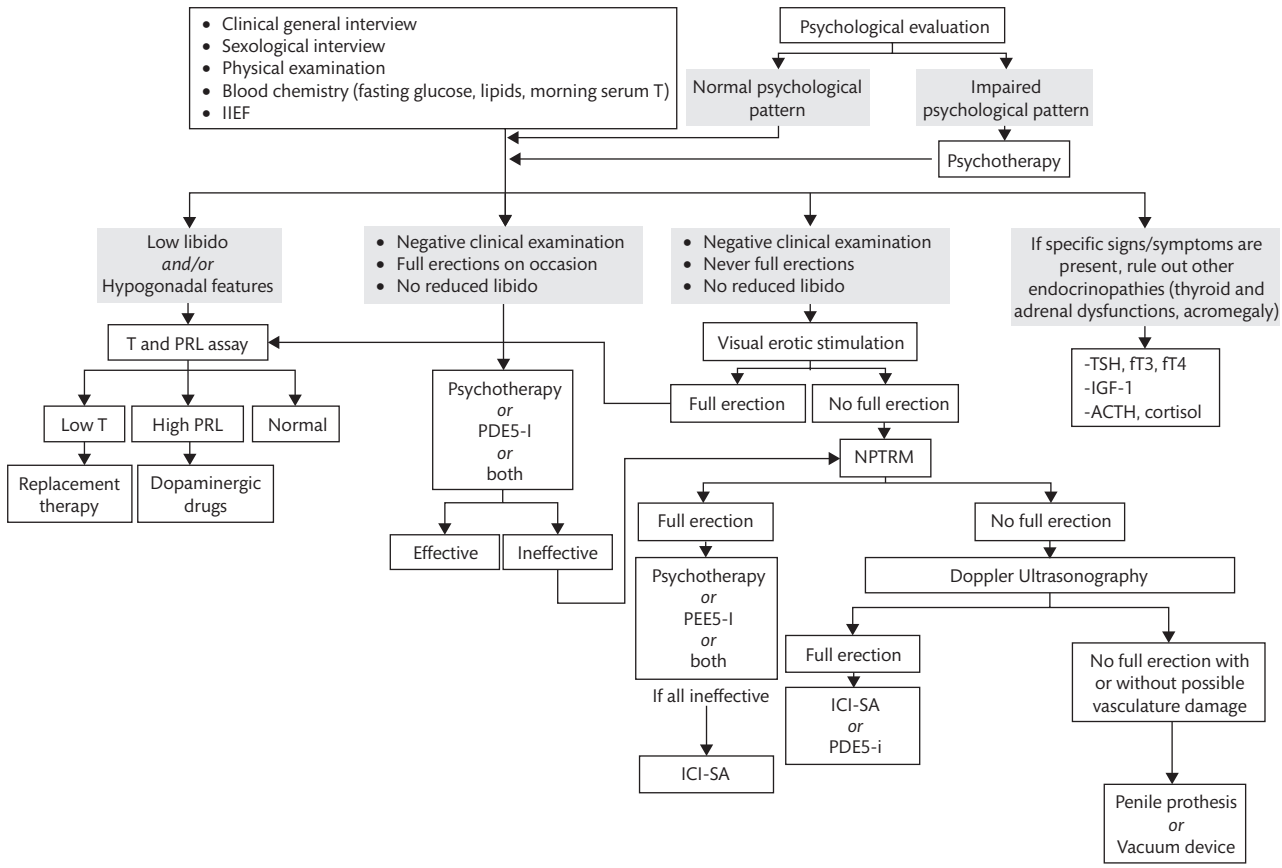


Figure 10.7.5 Diagnostic and therapeutic algorithm of erectile dysfunction.

T, testosterone; PRL, prolactin; NPTRM, nocturnal penile tumescence and rigidity monitoring; ICI, intracavernosal injection; ICSI, intracavernosal self-injection; PSV, peak systolic velocity; RI, resistance index; PDV, peak diastolic velocity.

stimulation in hypogonadal men, which renders the test uninformative in excluding erectile dysfunction due to low testosterone and/or high prolactin levels [10, 40].

Nocturnal penile tumescence and rigidity monitoring (NPTRM) may help to differentiate between psychogenic (normal NPTRM) and organic (impaired NPTRM) erectile dysfunction [24]. Furthermore, devices (e.g. RigiScan®) recording and measuring nocturnal penile tumescence and rigidity reduce the possibility of false positive diagnosis of psychogenic erectile dysfunction due to normal increase of tumescence but impaired rigidity [24].

Impaired NPTRM can occur because of hormonal (namely, low testosterone) vascular and neurological pathologies [10, 11, 14, 24, 40]. Thus, an impaired response cannot identify the specific cause underlying organic erectile dysfunction. The usefulness of NPTRM is also limited by the possibility of impaired sleep-related erections in men without organic pathologies but suffering from depression or sleep disturbances. Despite these drawbacks, there is general agreement that NPTRM is a valuable method to study the integrity of the erectile mechanism without psychological interferences [52].

To investigate the integrity of the vasculature within the penis, the simplest test is the direct intracavernosal injection of vasoactive drugs, useful both for obtaining information about the possible vasculature damage and the effectiveness of these drugs as potential treatment for erectile dysfunction [23, 33, 34, 53]. Intracavernosal injection of vasoactive drugs directly promotes penile erection by increasing arterial blood inflow and the subsequent achievement of a full erection indicates a normal veno-occlusive mechanism [23, 33, 34, 53]. Although many substances have been used for this test, to date prostaglandin E_1 and papaverine hydrochloride are the drugs most frequently used, both alone or added to phentolamine. This is not an appropriate test to evaluate the arterial penile vessels [53] because a full erection in response to this test is reached by as many as 20% of patients with borderline low arterial inflow, according to penile Doppler parameters (see next) [54].

Penile Doppler ultrasonography is the most accurate tool to investigate arterial haemodynamics in the penile arteries after an intracavernosal injection of a vasoactive drug [23, 26, 34]. The cavernosal artery peak blood flow velocity (normal: above 25–30 cm/sec) is a measure of vascular integrity of penile blood vessels which can reveal local atherosclerotic disease within the penis, that may reflect the general status of other small-to-medium-sized arteries in the patient [55]. Thus, investigation and detection of vasculogenic erectile dysfunction may prompt further investigation of unsuspected generalized cardiovascular diseases in the patient [56].

Management of Erectile Dysfunction

In the management of erectile dysfunction, the physician should consider the goal(s) that the patient (and his partner) is seeking and determine which therapy is best suited to the patient's psychological pattern and sexual habits (Figure 10.7.5).

General and Psychosexual Counselling

General counselling should provide the patient with information on lifestyle risk factors (e.g. obesity, smoking, alcohol, and recreational drug abuse) of erectile dysfunction and on how to modify them if present. As adjunctive intervention, lifestyle modification may lead to small improvement of both the erectile function and/or the hormonal secretion, especially testosterone [26, 34, 36].

Ideally, every patient suffering from erectile dysfunction, not only with psychogenic aetiologies, can potentially benefit from some behavioural and couple-oriented psychosexual counselling to tackle common contributory psychological factors in organic erectile dysfunction as well as any relationship issues of the couple. In general, the physician managing patients with erectile dysfunction should be skilled in providing simple, elementary psychosexual support (behavioural advice, sex education information) but should be able to recognize who and when to refer to sexologist, psychiatrist or psychologist/counsellor for specialized care [34, 36]. Pharmacological treatments for erectile dysfunction may be administered even in patients undergoing psychotherapy (cognitive-behavioural interventions) and/or drugs prescribed by the psychiatrist [34, 36].

Oral Non-Hormonal Drugs

Several oral agents (e.g. yohimbine, phentolamine, and apomorphine) were used in the past for the treatment of erectile dysfunction, but most of them are now of historical interest only [12, 26].

Type 5 Phosphodiesterase Inhibitors (PDE5Is)

Sildenafil, vardenafil, tadalafil, and avanafil are the drugs currently used as first-line treatment of erectile dysfunction; they belong to the class of orally-active inhibitors of PDE5, the enzyme that metabolizes and inactivates cGMP and cAMP in the penile blood vessels (Figure 10.7.2). PDE5Is decrease PDE5 enzyme activity and the consequent increase of cGMP and cAMP that causes the efflux of intracellular $Ca^{2}+$ to induce penile vascular smooth muscle relaxation (Figure 10.7.2) [3, 12, 13, 26, 57].

PDE5Is are effective for both psychogenic and organic erectile dysfunction but their action is contingent upon sexual arousal and central erectile neurological stimulation [13, 25, 26, 34, 36, 57]. The dose should be titrated to the maximum tolerated dose (Table 10.7.1). The starting dose is chosen by the clinician and it is not mandatory to start with the lowest dose, except in older people (>65 years) [57]. PDE5Is do not differ from each other in terms of efficacy, the only differences being the time of onset and duration of action, and the dose range (Table 10.7.1) [57]. PDE5Is are preferably used on-demand, except for tadalafil, which is effective at the same dose also when administered once daily due to its longer duration of action (Table 10.7.1) [26, 34, 36, 57]. Daily therapy may be advisable in case of psychogenic erection, post-radical prostatectomy, couples discomforted by the need for programming sexual activities [36, 57]. PDE5Is therapy should be stopped when there is no response (defined as lack of efficacy after at least 8 maximum doses of the drug with sexual stimulation) as well as in patients with psychogenic erectile dysfunction after having restored normal erections or when other drugs contraindicated in PDE5Is therapy are indispensable [20, 26, 34, 58]. Non-responders to PDE5Is are common among patients with type 2 diabetes mellitus and those who underwent radical prostatectomy [34, 36].

The safety profile of PDE5Is is excellent when the drugs are used at the right dosage and within the appropriate clinical context [12, 26, 34, 36, 58]. Their adverse effects are usually transient and have minor importance; these include headache, flushing, rhinitis, dyspepsia, dizziness, blue vision, and abnormal vision (change in brightness perception) (Table 10.7.1) [26, 34, 36, 58]. Other minor side effects, such as back pain, myalgia have been described occasionally [12, 58]. Absolute contraindications to the use of PDE5Is are the concomitant administration of nitrate medications for ischaemic heart disease or as recreational drug (i.e. poppers) as well as other potent vasodilators such as nicorandil [34, 36, 58]. Retinitis pigmentosa, an inherited disorder of retinal PDE6, is also an absolute contraindication [26, 34, 36]. The dosage of PDE5Is should be reduced in case of older patients (>65 years), impaired hepatic and/or renal function and use of potent cytochrome P450 3A4 (CYP3A4) inhibitors (e.g. ritonavir, squinavir, cimetidine, erythromycin) [34, 36, 58]. PDE5Is might further reduce blood pressure in men taking antihypertensive drugs, but the decrease is usually too small to lead to symptomatic hypotensive episodes [34, 36, 58]. PDE5Is are safe in patients with cardiovascular diseases with the caveats on vasodilator medications; in particular they do not increase the overall incidence and the occurrence of further cardiovascular events (i.e. myocardial infarction); *vice versa* they seem to exert beneficial effect on the cardiac arterial bed [34, 36, 58].

Sildenafil and vardenafil have a reduced maximal plasma concentration when taken after a high fat meal and should ideally be administered on an empty stomach in which an acidic environment increases the bioavailability [59]. Alcohol consumption should be avoided or moderated when taking any PDE5Is in order to minimize the hypotensive effects of these drugs [59].

Misuse of PDE5Is drugs for recreational purposes is increasing in young men [60, 61] especially among men who have sex with men, HIV infected patients and sex workers [29, 62, 63].

Hormonal Therapy

Testosterone should be reserved only for patients with an established diagnosis of hypogonadism [34, 36, 51], where improvement in libido and, concomitantly, erectile function can be expected [20] (see Chapters 10.4.2, 10.4.4, and 10.4.5). As mentioned earlier, this represent a minority of patients presenting with erectile dysfunction as the predominant complaint. Before testosterone treatment is considered, plans for paternity should be ascertained. Patients should be informed that testosterone treatment, while not being a fully reliable contraceptive, suppresses spermatogenesis, and reduces fertility.

Phosphodiesterase 5 inhibitors can improve erectile function in eugonadal and hypogonadal men [14, 57, 59, 64]. However, the efficacy and the place of combination therapy with testosterone and PDE5Is in the management of ED and/or hypogonadism is debated.

Table 10.7.1 Pharmacokinetic of PDE5 inhibitors

	Tmax (h) range (median)	T1/2 (h) range (median)	Dose (mg)	Way/Onset of action (h)	Duration of action (h)	Food effect on drug availability	Headache	Flushing	Dyspepsia	Altered colour perception	Dizziness
Sildenafil	1–2 (1)	3–5 (3.82)	25–100	On-demand/1 h	4	++	++	++	+	+++	+
Vardenafil	1–2 (1)	4–5 (3.94)	5–20	On-demand/1 h	4	+	+++	++	+	+++	+
Tadalafil	0.5–6 (2)	17–50 (10–24)	5–20	On-demand; daily/1 h	36	–	++	+	++	+/–	+
Avanafil	0.5–0.75 (1/2)	6–17 (5.0)	50–200	On-demand/1/2 h	6	–	++	+	+/–	+/–	+/–

PDE5, phosphodiesterase 5; Tmax: time required for maximum serum drug concentration to be reached; T1/2: time required for drug serum concentration to be reduced by one-half of its maximum concentration.

Several uncontrolled clinical studies have shown an apparent synergic effect between testosterone and PDE5Is in patients with ED, not all of whom had low T in the hypogondal range [14, 65–67]. However, a randomized controlled trial in men with ED and T levels <330 ng/dl failed to demonstrate further improvements in erectile function with the addition of T to an optimized regimen of PDE5I (Sildenafil) [64]. Similarly, addition of T to impotent men treated with PDE5 (tadalafil 10 mg daily for 4 weeks) did not produce any further improvements in erectile function unless the baseline T levels were ≤ 300 ng/dl [68]. Based on current evidence, a reasonable approach in the older ED patients with low T (but without a diagnosis of organic hypogonadism) is to optimize sexual function using (unless contraindicated) a PDE5I in the first instance, and consider add-on T treatment on a case by case basis, educating the man about the lack of definitive evidence.

Low libido and/or erectile dysfunction in hyperprolactinaemic men with or without hypogonadism can be successfully treated with dopaminergic drugs; testosterone alone may be ineffective in these patients if the hyperprolactinaemia remains uncorrected [9, 40]. In case of hypogonadism and hyperprolactinaemia due to CKD, the recovery of normal renal function following renal transplantation is able to improve or even restore normal function of the testis with beneficial effects also on erectile function (in the absence of pelvic complication due to surgery on nerves and vessels), especially in younger patients [30, 31]. See Chapter 10.6, 'Hypothalamo–Pituitary–Testicular Axis Function in Systemic Diseases and Effects of Medications'.

Intracavernosal Self-Injection Therapy

This therapy is effective in more than 70% of men with psychogenic and neurogenic erectile dysfunction, and is also effective in men with mild to moderate vasculogenic erectile dysfunction [57, 69, 70]. The injection is given, usually by the patient prior to sexual activity, at the penile base on a lateral side.

The main side effect linked with regular therapeutic use of intracavernosal injections is the development of fibrous plaques as in Peyronie's disease, which are reported to occur in up to 2–4% of self-injecting men after one year of injections at a rate of once per week [57]. An ischaemic, veno-occlusive priapism, that is a rigid and painful erection lasting more than 4 hours and characterized by absent cavernous blood flow [71] can occur when an excessive dose of prostaglandin E1 is used, but this can be easily avoided by patient education and careful dose titration starting with lower doses (5–10 mcg) initially [57, 69]. Priapism is a medical emergency, requiring urgent treatment to relieve pain and avoid penile fibrosis, persistent erectile dysfunction, and possible major penile deformity [72]. Treatment of priapism consists of blood aspiration from corpora cavernosa or intra corpora cavernosa administration of sympathomimetic drugs (most frequently phenylephrine or etilefrine) or both measures to restore penile flaccidity. If treatment is delayed by 48–72 hours after priapism onset, there may be worsening of the erectile response when intracavernosal self-injection is restarted [72].

Transurethral Therapy

Transurethral administration of prostaglandin E1 (250–1000 mcg) [57, 73] is less effective than intracavernosal injections; the occurrence of full erection ranges from 10% to 65% of patients. Furthermore, the transurethral absorption of prostaglandin E_1 can lead to more acute systemic side effects (dizziness, sweating, and hypotension), than intracavernosal injections, but the absence of the risk of penile fibrosis is noteworthy [57].

External Devices

Among external devices available for purchase on the market, vacuum pumps are the most commonly. They consist of a suction pump connected to a cylinder, which is placed over the penis. The negative pressure causes increased blood flow into the penis. A tension ring is placed at the base of the penis when the erection is reached; the pump is then removed. Adverse effects of the vacuum device include pain, blocked and painful ejaculation, haematoma, ecchymosis, petechiae, and ischaemic penile injury [26, 74]. Depending on individual patient's and partner's preferences, external penile devices may be an effective alternative to medications or intracavernosal injections for some couples to achieve intercourse [74].

Extracorporeal Shockwave Therapy

There is recent literature describing the use of low intensity extracorporeal shockwave therapy applied to the penis in men with erectile dysfunction. Shockwave therapy should improve penile haemodynamics, but a widespread acceptance of this treatment has not yet been reached [75].

Penile Prosthesis

Thanks to the increased availability and effectiveness of intracavernosal self-injection, oral drugs, and the pump devices, penile prosthesis are less commonly employed as a surgical treatment for erectile dysfunction. It remains the last resort treatment for erectile dysfunction that is refractory to all the above-mentioned therapies. There are semi-rigid and inflatable penile prostheses, which provide penile rigidity continuously or on-demand, respectively. The invasiveness of this treatment continues to be a deterrent for the patient, even though the surgical complications have decreased in recent years [26, 76]. Perioperative infection, poor erection, penile deformity, device failure, and penile glans trauma during sex are amongst the complications of penile prostheses [26, 76].

Vascular Surgery

Vascular surgery is reserved for venous leakage and focal arterial block. The treatment of venous leakage by ligation of the superficial and deep dorsal veins is disappointing with respect to long-term effectiveness [26, 77]. In contrast, arterial surgery—anastomosis of the inferior epigastric artery to the penile dorsal artery or to the deep dorsal vein—seems more promising [26, 77].

Post-Finasteride Syndrome

Post-finasteride syndrome is a controversial entity that has been recently described in men (usually young adults) taking finasteride for counteracting hair loss [78]. It is characterized by a multitude of sexual (low libido, erectile dysfunction, sexual anhedonia (i.e. reduced motivation or ability to experience pleasure), loss of pleasure during orgasm) physical (gynaecomastia, fatigue, weakness, skin lesions), and psychological (memory and cognitive impairment, depressed mood, anxiety, suicidal ideation, no emotional changes, insomnia), symptoms which may be present in different

combinations [78, 79]. Much of the information comes from patients' self-reports [79]. The most rigorous study to date has failed to identify consistent objective parameters validating the existence of this 'syndrome' [78]. From a pathophysiological perspective, the side effects of finasteride may be due, at least in part, to its antiandrogen action (sexual symptoms) but this mechanism does not explain psychoneurological symptoms, and/or their persistence after cessation of finasteride. Recently, changes in some neurosteroids have been found in cerebrospinal fluid of patients with the post-finasteride 'syndrome' compared to controls but whether they are cause or consequence is not known [80].

The management of this 'syndrome' is complex since symptoms seem not to reverse in most of the patients after finasteride discontinuation and/or testosterone therapy [78, 79]. No data on the possible beneficial effects of exogenous androgens administration are available. Furthermore, there are no predictive factors that can identify patients at risk of developing this 'syndrome', except that a pre-existing mood disturbance seems to be frequent [79]. At present there is no known effective treatment for this poorly understood condition.

Conclusions

Due to advances in the understanding of the physiological mechanisms of penile erection and the availability of PDE5Is, the awareness, diagnosis and management of erectile dysfunction has improved greatly in the last two decades for an increasing number of patients. Significant benefits on sexual activities and satisfaction can now be extended also to some patients with chronic diseases (e.g. chronic kidney disease, HIV infection, diabetes, cardiovascular disease). However, despite these welcomed improvements, treatment for sexual/erectile dysfunction remain unsatisfactory for a significant proportion of men (and their partners). Consultations for erectile dysfunction may alert clinicians and patients to the presence of undiagnosed cardiovascular diseases. Finally, the recovery of a satisfying sexual relationship often results in improved self-esteem and overall quality of life in a previously under-served group of patients.

REFERENCES

1. Rochira V, Zirilli L, Madeo B, Balestrieri A, Granata AR, Carani C. Sex steroids and sexual desire mechanism. *J Endocrinol Invest*, 2003; **26**(3 Suppl): 29–36.
2. Bancroft J. The endocrinology of sexual arousal. *J Endocrinol*, 2005; **186**: 411–27.
3. Gratzke C, Angulo J, Chitaley K, *et al*. Anatomy, physiology, and pathophysiology of erectile dysfunction. *J Sex Med*, 2010; **7**(1 Pt 2): 445–75.
4. Giuliano F. Neurophysiology of erection and ejaculation. *J Sex Med*, 2011; **8** Suppl 4: 310–15.
5. Rochira V, Balestrieri A, Madeo B, *et al*. Congenital estrogen deficiency: in search of the estrogen role in human male reproduction. *Mol Cell Endocrinol*, 2001; **178**: 107–15.
6. Finkelstein JS, Lee H, Burnett-Bowie SA, *et al*. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med*, 2013; **369**: 1011–22.
7. Rochira V, Madeo B, Diazzi C, Zirilli L, Daniele S, Carani C. Estrogens and male reproduction. In: De Groot LJ, Chrousos G, Dungan K, Feingold KR, Grossman A, Hershman JM, *et al*., eds. Endotext. South Dartmouth (MA): MDText.com, Inc., 2000.
8. Corona G, Tirabassi G, Santi D, *et al*. Sexual dysfunction in subjects treated with inhibitors of 5alpha-reductase for benign prostatic hyperplasia: a comprehensive review and meta-analysis. *Andrology*, 2017; **5**: 671–8.
9. Maggi M, Buvat J, Corona G, Guay A, Torres LO. Hormonal causes of male sexual dysfunctions and their management (hyperprolactinemia, thyroid disorders, GH disorders, and DHEA). *J Sex Med*, 2013; **10**: 661–77.
10. Carani C, Granata AR, Bancroft J, Marrama P. The effects of testosterone replacement on nocturnal penile tumescence and rigidity and erectile response to visual erotic stimuli in hypogonadal men. *Psychoneuroendocrinology*, 1995; **20**: 743–53.
11. Granata AR, Rochira V, Lerchl A, Marrama P, Carani C. Relationship between sleep-related erections and testosterone levels in men. *J Andrology*, 1997; **18**: 522–7.
12. Andersson KE. Mechanisms of penile erection and basis for pharmacological treatment of erectile dysfunction. *Pharmacol Rev*, 2011; **63**: 811–59.
13. Lue TF. Erectile dysfunction. *N Engl J Med*, 2000; **342**: 1802–13.
14. Rochira V, Balestrieri A, Madeo B, Granata AR, Carani C. Sildenafil improves sleep-related erections in hypogonadal men: evidence from a randomized, placebo-controlled, crossover study of a synergic role for both testosterone and sildenafil on penile erections. *J Andrology*, 2006; **27**: 165–75.
15. Morelli A, Corona G, Filippi S, *et al*. Which patients with sexual dysfunction are suitable for testosterone replacement therapy? *J Endocrinol Invest*, 2007; **30**: 880–8.
16. Traish AM, Goldstein I, Kim NN. Testosterone and erectile function: from basic research to a new clinical paradigm for managing men with androgen insufficiency and erectile dysfunction. *Eur Urol*, 2007; **52**: 54–70.
17. Rosen RC, Cappelleri JC, Gendrano N, 3rd. The International Index of Erectile Function (IIEF): a state-of-the-science review. *Int J Impotence Res*, 2002; **14**: 226–44.
18. NIH Consensus Conference. Impotence. NIH Consensus Development Panel on Impotence. *JAMA*, 1993; **270**: 83–90.
19. Association AP. Diagnostic and Statistical Manual of Mental Disorders Text Revision (DSM-IV-TR), 4th edn. Washington, DC: American Psychiatric Publishing, 2000.
20. Corona G, Isidori AM, Buvat J, *et al*. Testosterone supplementation and sexual function: a meta-analysis study. *J Sex Med*, 2014; **11**: 1577–92.
21. Wu FC, Tajar A, Beynon JM, *et al*. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*, 2010; **363**: 123–35.
22. Nguyen HMT, Gabrielson AT, Hellstrom WJG. Erectile dysfunction in young men—a review of the prevalence and risk factors. *Sex Med Rev*, 2017; **5**: 508–20.
23. Mulhall JP, Giraldi A, Hackett G, *et al*. The 2018 revision to the process of care model for evaluation of erectile dysfunction. *J Sex Med*, 2018; **15**: 1280–92.
24. Jannini EA, Granata AM, Hatzimouratidis K, Goldstein I. Use and abuse of Rigiscan in the diagnosis of erectile dysfunction. *J Sex Med*, 2009; **6**: 1820–9.
25. McVary KT. Clinical practice. Erectile dysfunction. *N Engl J Med*, 2007; **357**: 2472–81.

26. Yafi FA, Jenkins L, Albersen M, *et al*. Erectile dysfunction. *Nat Rev Dis Primers*, 2016; **2**: 16003.
27. Feldman HA, Goldstein I, Hatzichristou DG, Krane RJ, McKinlay JB. Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *J Urol*, 1994; **151**: 54–61.
28. Corona G, Mannucci E, Schulman C, *et al*. Psychobiologic correlates of the metabolic syndrome and associated sexual dysfunction. *Eur Urol*, 2006; **50**: 595–604; discussion 604.
29. Santi D, Brigante G, Zona S, Guaraldi G, Rochira V. Male sexual dysfunction and HIV—a clinical perspective. *Nat Rev Urol*, 2014; **11**: 99–109.
30. Rochira V. Hypogonadism in systemic diseases. In: Simoni MH, Huhtaniemi IT, eds. *Endocrinology of the Testis and Male Reproduction*, 1st edn. Cham: Springer International Publishing, 2017: 829–79.
31. Edey MM. Male sexual dysfunction and chronic kidney disease. *Front Med*, 2017; **4**: 32.
32. Korenman SG, Morley JE, Mooradian AD, *et al*. Secondary hypogonadism in older men: its relation to impotence. *J Clin Endocrinol Metab*, 1990; **71**: 963–9.
33. Granata A, Bancroft J, Del Rio G. Stress and the erectile response to intracavernosal prostaglandin E1 in men with erectile dysfunction. *Psychosom Med*, 1995; **57**: 336–44.
34. Hackett G, Kirby M, Wylie K, *et al*. British Society for Sexual Medicine guidelines on the management of erectile dysfunction in men—2017. *J Sex Med*, 2018; **15**: 430–57.
35. Montejo AL, Montejo L, Navarro-Cremades F. Sexual side-effects of antidepressant and antipsychotic drugs. *Curr Opin Psychiatry*, 2015; **28**: 418–23.
36. Mulhall JP, Giraldi A, Hackett G, *et al*. The 2018 revision to the process of care model for management of erectile dysfunction. *J Sex Med*, 2018; **15**: 1434–45.
37. Johnsen NV, Kaufman MR, Dmochowski RR, Milam DF. Erectile dysfunction following pelvic fracture urethral injury. *Sex Med Rev*, 2018; **6**: 114–23.
38. Mikhail N. Does testosterone have a role in erectile function? *Am J Med*, 2006; **119**: 373–82.
39. Buvat J, Bou Jaoude G. Significance of hypogonadism in erectile dysfunction. *World J Urol*, 2006; **24**: 657–67.
40. Carani C, Granata AR, Fustini MF, Marrama P. Prolactin and testosterone: their role in male sexual function. *Int J Androl*, 1996; **19**: 48–54.
41. Carani C, Isidori AM, Granata A, *et al*. Multicenter study on the prevalence of sexual symptoms in male hypo- and hyperthyroid patients. *J Clin Endocrinol Metab*, 2005; **90**: 6472–9.
42. Granata A, Tirabassi G, Pugni V, *et al*. Sexual dysfunctions in men affected by autoimmune Addison's disease before and after short-term gluco- and mineralocorticoid replacement therapy. *J Sex Med*, 2013; **10**: 2036–43.
43. Lotti F, Rochira V, Pivonello R, *et al*. Erectile dysfunction is common among men with acromegaly and is associated with morbidities related to the disease. *J Sex Med*, 2015; **12**: 1184–93.
44. Fode M, Krogh-Jespersen S, Brackett NL, Ohl DA, Lynne CM, Sonksen J. Male sexual dysfunction and infertility associated with neurological disorders. *Asian J Andrology*, 2012; **14**: 61–8.
45. Potosky AL, Davis WW, Hoffman RM, *et al*. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. *J Natl Cancer Inst*, 2004; **96**: 1358–67.
46. Tavukcu HH, Aytac O, Atug F. Nerve-sparing techniques and results in robot-assisted radical prostatectomy. *Invest Clin Urol*, 2016; **57**(Suppl 2): S172–S84.
47. Petrone L, Mannucci E, Corona G, *et al*. Structured interview on erectile dysfunction (SIEDY): a new, multidimensional instrument for quantification of pathogenetic issues on erectile dysfunction. *Int J Impotence Res*, 2003; **15**: 210–20.
48. Hatzichristou D, Kirana PS, Banner L, *et al*. diagnosing sexual dysfunction in men and women: sexual history taking and the role of symptom scales and questionnaires. *J Sex Med*, 2016; **13**: 1166–82.
49. Cappelleri JC, Rosen RC, Smith MD, Mishra A, Osterloh IH. Diagnostic evaluation of the erectile function domain of the International Index of Erectile Function. *Urology*, 1999; **54**: 346–51.
50. Rosen RC, Wu F, Behre HM, *et al*. Quality of life and sexual function benefits of long-term testosterone treatment: longitudinal results from the registry of hypogonadism in men (RHYME). *J Sex Med*, 2017; **14**: 1104–15.
51. Bhasin S, Brito JP, Cunningham GR, *et al*. Testosterone therapy in men with hypogonadism: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
52. Qin F, Gao L, Qian S, Fu F, Yang Y, Yuan J. Advantages and limitations of sleep-related erection and rigidity monitoring: a review. *Int J Impotence Res*, 2018; **30**: 192–201.
53. Pescatori ES, Hatzichristou DG, Namburi S, Goldstein I. A positive intracavernous injection test implies normal veno-occlusive but not necessarily normal arterial function: a hemodynamic study. *J Urol*, 1994; **151**: 1209–16.
54. Elhanbly S, Schoor R, Elmogy M, Ross L, Hegazy A, Niederberger C. What nonresponse to intracavernous injection really indicates: a determination by quantitative analysis. *J Urol*, 2002; **167**: 192–6.
55. Uddin SMI, Mirbolouk M, Dardari Z, *et al*. Erectile dysfunction as an independent predictor of future cardiovascular events: the multi-ethnic study of atherosclerosis. *Circulation*, 2018; **138**: 540–2.
56. Foresta C, Ferlin A, Lenzi A, Montorsi P. The great opportunity of the andrological patient: cardiovascular and metabolic risk assessment and prevention. *Andrology*, 2017; **5**: 408–13.
57. Hatzimouratidis K, Salonia A, Adaikan G, *et al*. Pharmacotherapy for erectile dysfunction: recommendations from the fourth International Consultation for Sexual Medicine (ICSM 2015). *J Sex Med*, 2016; **13**: 465–88.
58. Yafi FA, Sharlip ID, Becher EF. Update on the safety of phosphodiesterase type 5 inhibitors for the treatment of erectile dysfunction. *Sex Med Rev*, 2018; **6**: 242–52.
59. Hong JH, Kwon YS, Kim IY. Pharmacodynamics, pharmacokinetics and clinical efficacy of phosphodiesterase-5 inhibitors. *Exp Opinion Drug Metab Toxicol*, 2017; **13**: 183–92.
60. Smith KM, Romanelli F. Recreational use and misuse of phosphodiesterase 5 inhibitors. *J Am Pharm Assoc*, 2005; **45**: 63–72; quiz 3–5.
61. Bechara A, Casabe A, De Bonis W, Helien A, Bertolino MV. Recreational use of phosphodiesterase type 5 inhibitors by healthy young men. *J Sex Med*, 2010; **7**: 3736–42.
62. De Ryck I, Van Laeken D, Noestlinger C, Platteau T, Colebunders R. The use of erection enhancing medication and party drugs among men living with HIV in Europe. *AIDS Care*, 2013; **25**: 1062–6.
63. Hammoud MA, Jin F, Lea T, Maher L, Grierson J, Prestage G. Off-label use of phosphodiesterase type 5 inhibitor erectile dysfunction medication to enhance sex among gay and bisexual men in Australia: results from the FLUX study. *J Sex Med*, 2017; **14**: 774–84.

64. Spitzer M, Basaria S, Travison TG, *et al*. Effect of testosterone replacement on response to sildenafil citrate in men with erectile dysfunction: a parallel, randomized trial. *Ann Int Med*, 2012; **157**: 681–91.
65. Aversa A, Isidori AM, Spera G, Lenzi A, Fabbri A. Androgens improve cavernous vasodilation and response to sildenafil in patients with erectile dysfunction. *Clin Endocrinol*, 2003; **58**: 632–8.
66. Shabsigh R, Kaufman JM, Steidle C, Padma-Nathan H. Randomized study of testosterone gel as adjunctive therapy to sildenafil in hypogonadal men with erectile dysfunction who do not respond to sildenafil alone. *J Urol*, 2004; **172**: 658–63.
67. Rosenthal BD, May NR, Metro MJ, Harkaway RC, Ginsberg PC. Adjunctive use of AndroGel (testosterone gel) with sildenafil to treat erectile dysfunction in men with acquired androgen deficiency syndrome after failure using sildenafil alone. *Urology*, 2006; **67**: 571–4.
68. Buvat J, Montorsi F, Maggi M, *et al*. Hypogonadal men nonresponders to the PDE5 inhibitor tadalafil benefit from normalization of testosterone levels with a 1% hydroalcoholic testosterone gel in the treatment of erectile dysfunction (TADTEST study). *J Sex Med*, 2011; **8**: 284–93.
69. Porst H. The rationale for prostaglandin E1 in erectile failure: a survey of worldwide experience. *J Urol*, 1996; **155**: 802–15.
70. Urciuoli R, Cantisani TA, Carlini M, Giuglietti M, Botti FM. Prostaglandin E1 for treatment of erectile dysfunction. *Cochrane Database Syst Rev*, 2004; **2**: CD001784.
71. Berger R, Billups K, Brock G, *et al*. Report of the American Foundation for Urologic Disease (AFUD) Thought Leader Panel for evaluation and treatment of priapism. *Int J Impotence Res*, 2001; **13** Suppl 5: S39–43.
72. Salonia A, Eardley I, Giuliano F, *et al*. European Association of Urology guidelines on priapism. *Eur Urol*, 2014; **65**: 480–9.
73. Padma-Nathan H, Hellstrom WJ, *et al*. Treatment of men with erectile dysfunction with transurethral alprostadil. Medicated Urethral System for Erection (MUSE) Study Group. *N Engl J Med*, 1997; **336**: 1–7.
74. Wassersug R, Wibowo E. Non-pharmacological and non-surgical strategies to promote sexual recovery for men with erectile dysfunction. *Transl Androl Urol*, 2017; **6**(Suppl 5): S776–S94.
75. Chung E, Wang J. A state-of-art review of low intensity extra-corporeal shock wave therapy and lithotripter machines for the treatment of erectile dysfunction. *Expert Rev Med Devices*, 2017; **14**: 929–34.
76. Trost LW, McCaslin R, Linder B, Hellstrom WJ. Long-term outcomes of penile prostheses for the treatment of erectile dysfunction. *Expert Rev Med Devices*, 2013; **10**: 353–66.
77. Trost LW, Munarriz R, Wang R, Morey A, Levine L. External mechanical devices and vascular surgery for erectile dysfunction. *J Sex Med*, 2016; **13**: 1579–617.
78. Basaria S, Jasuja R, Huang G, *et al*. Characteristics of men who report persistent sexual symptoms after finasteride use for hair loss. *J Clin Endocrinol Metab*, 2016; **101**: 4669–80.
79. Giatti S, Diviccaro S, Panzica G, Melcangi RC. Post-finasteride syndrome and post-SSRI sexual dysfunction: two sides of the same coin? *Endocrine*, 2018; **61**: 180–93.
80. Melcangi RC, Santi D, Spezzano R, *et al*. Neuroactive steroid levels and psychiatric and andrological features in post-finasteride patients. *J Steroid Biochem Mol Biol*, 2017; **171**: 229–35.

10.8

Hormonal Male Contraception

Stephanie T. Page and Maritza T. Farrant

Introduction

Effective male contraception requires the reduction of motile sperm in the ejaculate to levels that reliably prevent fertilization. To date, all approved forms of male contraception aim to reduce female exposure to spermatozoa via mechanical blockade (condoms and vasectomy). The last half of the twentieth century saw the introduction of reliable, reversible hormonal contraceptive options for women, yet men continue to rely on less effective and irreversible family planning methods, with condoms, withdrawal, rhythm, and male sterilization, accounting for one-quarter of all contraceptive use worldwide [1]. Despite advances in female contraceptive methods and availability, in 2012 nearly 40% of all pregnancies worldwide were unintended and 50% of these ended in termination [2]. Men's role in furthering effective family planning has been underexplored and the development of more efficacious, reliable, safe, and reversible male contraceptive methods could have an enormous public health impact [3].

Reversible suppression of the hypothalamic–pituitary–testicular axis, resulting in abrogation of spermatogenesis, provides the foundation for reversible male hormonal contraception (MHC) regimens. High levels of intratesticular testosterone (and adequate intratesticular oestradiol) are required for normal human spermatogenesis. Administration of appropriately dosed exogenous androgens +/– progestins to men serves two purposes simultaneously. First, due to negative hypothalamic–pituitary feedback leading to suppression of GnRH and gonadotropin secretion, testicular steroidogenesis is inhibited. At the same time, peripheral androgen action is maintained to support non-gonadal tissues and prevent the sequalae of male hypogonadism (such as hot flashes, impairment of sexual function, bone density loss, and reduction in lean body mass). Progestins are used to augment androgen-induced hypothalamic–pituitary suppression, as androgens by themselves may not be sufficiently potent and at high doses exogenous androgens may also have direct, stimulatory intratesticular effects.

Here we provide an overview of various regimens and formulations that have been evaluated as potential MHC, including novel modified steroids that show promise as reversible contraceptive methods for men.

Physiology of Male Hormonal Contraception

Testosterone (T) is essential for maintenance of male secondary sexual characteristics including hair pattern, sexual function, muscle, and bone mass, as well as spermatogenesis. In eugonadal men, starting at puberty, pulsatile release of gonadotropin-releasing hormone (GnRH) from the hypothalamus stimulates the release of gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH), from the pituitary gland. LH binds to Leydig cells in the interstitium of the testes, stimulating T synthesis within the testes and generating high local concentrations of intratesticular T (between 50 and 100 times serum T concentrations) that are essential for spermatogenesis. Within the testes 1–2% of T is metabolized to produce both oestradiol (E2), and dihydrotestosterone (DHT). All three sex steroids are secreted into the circulation and inhibit FSH and LH secretion from the pituitary. Nurtured by Sertoli cells and stimulated by FSH, spermatogenesis occurs in the seminiferous tubules in this high T environment. FSH also stimulates inhibin B production from Sertoli cells that supresses FSH secretion by negative feedback.

Reversible MHC interrupts this naturally occurring negative feedback loop. The administration of an exogenous androgen (+/– a progestin), often at a supraphysiogical dose, results in suppression of GnRH and gonadotropin production through

interactions with the androgen, oestrogen, and progesterone receptors in the hypothalamus and pituitary gland. Importantly, intratesticular testosterone concentrations plummet without LH stimulation of Leydig cells, depriving developing spermatogonia of local testicular endocrine and paracrine signals required for completion of spermatogenesis. In the acute phase following administration of the MHC, there is blockade in the differentiation of Type A to Type B spermatogonia. Because the existing type B spermatogonia complete their maturation process despite the altered intratestiscular hormonal milieu, MHC requires 8–16 weeks for maximal suppression of mature sperm in the ejaculate. Therefore, markedly decreased sperm concentrations are only achieved after 3–4 months of MHC administration. Following cessation of exogenous steroid administration, the hormonal axis recovers in a matter of days, while sperm concentrations require 3–6 months to return to pretreatment levels [4]. Adjuncts to exogenous androgens, including progestins, are used to augment the androgen-mediated hypothalamic–pituitary suppression, providing further reduction of gonadotropin production, and, consequently, spermatogenesis.

Testosterone Formulations for MHC

Steady-state serum concentrations of serum testosterone at least within normal range are required for maintaining systemic eugonadism. In addition, exogenous androgens inhibit gonadotropin production, contributing to at least partial suppression of spermatogenesis. Injectable androgens have been the mainstay of MHC (see following sections), but alternative oral, transdermal, and implantable androgen delivery systems hold promise for novel MHC regimens in the future.

Oral androgens undergo significant first-pass metabolism, alter hepatic protein secretion, and have limited bioavailability. Early studies of oral testosterone undecanoate (TU), which is safe, as part of MHC were encouraging but limited by a requirement for twice daily dosing [5]. Transdermal T patches are approved in many countries to treat male hypogonadism, but, even when combined with a progestin, are not effective for male contraception [6, 7], likely due to the due to failure to consistently maintain high normal T concentrations. In addition, skin irritation occurs frequently at the patch application sites often leading to discontinuation [8]).

Several proof-of-concept studies have demonstrated that transdermal T-gels may be effective as MHC when combined with a progestin. The combination of T-gel (10 g daily) with nestosterone (NES), a non-androgenic progestin, led to severe oligospermia (<1 million sperm/ml, see sperm targets next) in 89% of men with minimal adverse effects, compared to 23% in the group receiving T-gel alone [9]. The median serum total and free T concentrations were maintained within the adult male range throughout the 24-week treatment period with combination therapy. Based upon these promising results, a combined Nes-Tes gel is slated for Phase 2 testing in a multinational MHC efficacy trial in 2018. Similarly, T-gel combined with depomedroxyprogesterone acetate (DMPA) injections resulted in severe oligospermia in 90% of men treated for 24 weeks [10]. Sexual function was preserved during treatment, and the majority of men reported satisfaction with the regimen [11]. Skin irritation associated with T-gels is uncommon, while potential downsides for some users include the need for daily application and care to avoid transferring the gel to female partners and children.

The subdermal T-implant, originally developed in 1937, lasts 3–6 months and achieves steady testosterone concentrations. Clinical use of T-implants and pellets as androgen replacement therapy increased in the 1990s [12]. T-pellets were used successfully in combination with long-acting progestin injections (DMPA) in a proof-of-concept MHC efficacy study [13]. T-pellets are generally well-tolerated, but their insertion requires a minor surgical procedure and there is a small risk of painful self-extrusion and infection [14], making T-pellets less attractive for some settings and users.

Androgen-Alone MHC Studies

Systematic dose-finding and feasibility studies using testosterone for MHC began in the 1970s [15]. The largest experience with androgen-alone regimens arose from landmark WHO-sponsored studies in the 1990s. The first [16] enrolled 271 fertile men in seven countries given 200 mg testosterone enanthate (TE) intramuscularly (IM) weekly, twofold the physiologic dose. Sixty-five per cent (65%) of these men became azoospermic, and an additional 30% became severely oligospermic. The 119 men who became azoospermic then discontinued other birth control methods and continued the injections as their only method of contraception with their female partner for 1 year. There was one pregnancy during the efficacy phase (0.8 conceptions [95% CI 0.02–4.5] per 100 person-years), a rate comparable with female contraceptive methods [17] demonstrating supraphysiologic testosterone-induced azoospermia to be a highly effective contraceptive.

The second international WHO study [18] was designed to evaluate the contraceptive effects of suppression to near azoospermia. In this nine-country study, 399 healthy men (123 from Asian and 276 from non-Asian countries) in stable heterosexual relationships were recruited and again received 200 mg/week IM TE. Of the men who completed the 6-month sperm suppression phase, 98% suppressed to the target threshold (<3 million (M) sperm/ml). The combined pregnancy rate for the men whose sperm concentration suppressed from 0 to 3 M/ml was 1.4 per 100 person-years. The combined pregnancy rate for men whose sperm concentration suppressed to 0–1 M/ml was 0.7 per 100 person-years, providing the benchmark for subsequent MHC studies stipulating a sperm concentration of ≤1 M/ml for entrance into an efficacy phase (wherein the experimental method is relied upon for contraception [19]).

Testosterone undecanoate (TU) in castor oil is a promising longer-acting injectable T formulation approved for use in many countries to treat male hypogonadism. Two studies in China demonstrated that IM TU is an effective hormonal contraceptive. In the first study [20], 308 men were enrolled in a 6-month suppression phase with a loading dose of 1000 mg IM TU followed by 500 mg monthly. 97% of men had sperm concentrations suppress to <3 M/ml during the suppression phase, and 296 couples were enrolled in a 6-month efficacy phase. There was only one pregnancy in the 280

couples who completed the efficacy phase. Based on these promising results, a longer follow-up study [21] recruited 1045 healthy Chinese men using the same regimen and suppression criteria in a 24-month efficacy phase study. 893 couples completed the suppression phase and 95% of these enrolled in the efficacy phase. Nine pregnancies occurred, with a resulting cumulative pregnancy rate of 0.55 per 100 person/years. Thus, using a long-acting injectable T formulation, these studies confirmed the efficacy findings of the shorter-acting WHO testosterone-alone trials. The most common side effects associated with these supraphysiologic doses of IM TU included injection site discomfort, acne, and acute, post-injection coughing.

Testosterone Plus Progestin Combinations

Progestins suppress gonadotropin secretion and may have additional direct, intratesticular effects, resulting in more rapid and complete suppression of spermatogenesis when a progestin is added to T for MHC [22]. The addition of a progestin also allows for lower, more physiological T dosing, reducing androgenic side effects and minimizing any long-term risks associated with supraphysiologic androgen exposure. The combination of T plus a progestin achieves severe oligospermia in 85–90% of men [10, 23–27]. Turner *et al.* demonstrated contraceptive efficacy using combination MHC in a proof-of-concept Australian study enrolling 55 couples, using T-pellets plus IM DMPA [13]. No pregnancies occurred in 426 person-months (95% CI contraceptive failure rate, 0–8%/year).

Based on results of preliminary work administering IM TU plus IM norethisterone enanthate [26] demonstrating an 8-week dosing interval induced severe oligozoospermia in 100% of men, a multicentre efficacy study [28] of this regimen was undertaken. Of the 320 participating couples, 96% of men (95% CI 92.8–97.9) suppressed to a sperm concentration ≤1 M/ml within 24 weeks. During the 1-year efficacy phase, four pregnancies occurred among 266 participants, giving a pregnancy rate of 1.57/100 continuing users (95% CI, 0.59–4.14). This study was discontinued early due to sponsor safety concerns (not those of the Data Safety Monitoring Board). There was one suicide during the efficacy phase judged by the investigators to be unrelated to the study medications, one case of depression (judged probably related to the study), and one attempted paracetamol overdose (judged possibly related). Additional adverse effects of concern included site injection pain, mood changes, increased libido, and acne. Adverse events surrounding mood, libido, and injection site pain were much more frequent in one of the ten study centres.

Testosterone Plus GnRH Antagonists

Systemically dosed gonadotropin-releasing hormone (GnRH) antagonists, acting on the pituitary, suppress circulating gonadotropin concentrations within hours of administration. It has been hypothesized that the addition of a GnRH antagonist to androgen-based MHC could potentiate suppression of gonadotropins and, consequently, spermatogenesis. However, GnRH antagonists are expensive to synthesize and difficult to deliver. There have been no efficacy trials of MHC regimens that include a GnRH antagonist, but studies of an androgen plus short acting GnRH antagonists (given daily) have not generally demonstrated superiority over androgens alone [29, 30] although one small study suggested that suppression induced by a GnRH antagonist plus testosterone might be sufficiently maintained with a physiologic dose of testosterone-alone. A similar study using 19-nortestosterone did not replicate these results [31]. A longer-acting GnRH antagonist, acyline, administered by subcutaneous injection, suppresses LH within hours, and lowers serum T concentrations to castrate levels within 24 hours, with effects lasting 14–15 days after a single injection [32]. However, a study of daily transdermal T-gel plus IM DMPA with or without acyline showed a similar degree and rapidity of suppression to severe oligozoospermia between groups [10].

Novel Synthetic Androgens

With the goals of minimizing androgen-related side effects, maximizing efficacy, and optimizing delivery systems for users with varied preferences, novel synthetic androgens are currently in development for MHC. Derivatives of 19-nortestosterone that retain the androgen effects of T but may have a longer half-life, negligible hepatotoxicity with oral administration, and progestogenic activity [33] have shown promise in recent clinical trials. In addition, ideal androgens for MHC are: (1) resistant to 5α-reductases, enzymes highly expressed in the prostate gland which convert T to the more potent androgen dihydrotestosterone (DHT) and are implicated in prostate cancer; and (2) aromatizable to metabolites which bind and activate the human oestrogen receptor. The role of oestrogen in male physiology has been increasingly recognized; oestrogen deficiency in men may result in bone loss, accumulation of body fat, decreased sexual function, and vasomotor symptoms (VMS) [34].

7α-methyl-19-nortesterone (MeNT) an attractive 'prostate-sparing' androgen, is aromatizable yet resistant to 5α-reductase [35, 36]. In a small 6-month study [37] MeNT implants administered to hypogonadal men maintained most androgen-dependent functions without prostate effects, although there was some bone loss suggesting oestrogenic activity may not have been adequate. As a contraceptive, an initial 1-year pilot study of MeNT implants was encouraging, with 8/11 of men developing severe oligozoospermia [38]. However, subsequent studies of MeNT plus a progestin were stymied by an implant design flaw, preventing sustained and adequate release of MeNT, which will need to be overcome before further human testing [39].

Dimethandrolone undecanoate (DMAU) and 11β-methyl-19-nortestosterone-17β-dodecylcarbonate (11β-MNTDC) are orally bioavailable 19-nortestosterone derivatives that bind and potently activate both androgen and progesterone receptors. Both DMAU and 11β-MNTDC are de-esterified to their active metabolites, dimethandrolone (DMA) and 11β-methyl-19-nortestosterone (11β-MNT), respectively, *in vivo*, and neither are reduced by 5α-reductase [40] nor likely aromatized to oestrogenically active metabolites. In castrated rodents, administration of DMAU demonstrated favourable changes in androgenic endpoints and maintenance of bone mineral density [41] with doses comparable to exogenous testosterone, while in rabbits oral DMAU given daily was an effective, reversible contraceptive [42]. These encouraging preclinical data supported the first-in-human trials of oral

DMAU and 11β-MNTDC. Placebo-controlled Phase I studies of escalating single doses of oral DMAU have assessed pharmacokinetics, safety and food effects in men [43, 44] and have demonstrated that oral DMAU is well-tolerated and significantly suppresses gonadotropin and T production in men, properties required for effective MHC. Of note, concomitant administration with a high-fat meal markedly improves oral DMAU/DMA pharmacokinetics and similar results were reported for single doses of 11β-MNTDC [45]. Recently, a one month, double-blind Phase I trial of daily oral DMAU administered to healthy men demonstrated doses of 200–400 mg are well-tolerated, markedly suppress gonadotropins, and suppress serum T to near castrate levels without significant, acute hypogonadal side effects [46]. These exciting results support the further development of DMAU as a single-agent oral male contraceptive pill and investigations are ongoing.

Contraceptive Efficacy, Sperm Targets, and Variable Responsiveness to MHC

In healthy men, sperm concentrations in the ejaculate are 15–150 million per millilitre (M/ml). Sperm concentrations of greater than 15 M/ml are associated with normal fertility [47]. The absence of spermatozoa in the ejaculate, azoospermia, is the ideal target of MHC, analogous to anovulation. However, the early WHO clinical efficacy studies demonstrated that oligospermia (<3 M/ml) and severe oligospermia (≤1 M/ml) are efficacious in the prevention of pregnancy [16, 18, 28].

An integrated analysis [22] reviewed MHC studies of at least 3 months' treatment duration, including a total of 1756 healthy men. Restricting the analysis to contraceptive efficacy studies employing optimized regimens (testosterone plus progestin) the authors concluded 96% of men suppress sperm output to <1 M/ml by 6 months using MHC. Current expert opinion based on phase II dose-finding trials [19] recommends a sperm concentration threshold of ≤1 M/ml as suitable for reliable contraception.

One question that has plagued the field is why some men, 5–10%, do not suppress their sperm production completely. Initial MHC studies with supraphysiologic T-alone found that 60% of non-Chinese and more than 90% of Chinese men became azoospermic [16, 18]. High rates of azoospermia among Chinese men have been replicated using monthly TU injections [3, 20, 21]. The explanation for differences in susceptibility to hormonally induced azoospermia between individual men and across ethnicities remains largely unknown, but may include both genetic and environmental factors [48]. In some men gonadotropin production may be incompletely suppressed thus allowing for persistent spermatogenesis but analyses of gonadotropin concentrations in multiple MHC studies have not supported this explanation [22, 49]. However, it remains possible that there are very small differences in serum gonadotropin concentrations or responsiveness in men who fail to reach severe oligospermia [50]. Rodent studies have demonstrated that spermatogenesis can be maintained by persistent, gonadotropin-independent testicular steroidogenesis [51, 52] although this has yet to be demonstrated in men.

Reversibility of MHC

Following cessation of short-term MHC in studies of reproductively normal men, hormone and sperm count recovery is nearly universal. In a multivariate, integrated analysis of individual data gathered from approximately 90% of all published androgen or androgen-progestin MHC trials, the median times for sperm to recover to thresholds of 20 or 10 M/ml were 3.4 months (95% CI 3.2–3.5) and 3.0 months (95% CI 2.9–3.1), respectively [4]. The typical probability of recovery to 20 M/ml was 90% within 12 months and 100% within 24 months. Various covariables were associated with time to recovery, but not the extent of recovery, including shorter duration of treatment, shorter-acting T preparations, higher sperm concentrations at baseline, and lower blood concentrations of LH at baseline [4]. Treatment duration appears to be a clinically important covariable [21, 53] but understanding its importance is limited by lack of 'long-term' MHC studies (>30 months treatment duration) highlighting a need for post-marketing surveillance [54]. There are no systematic studies of eugonadal men treated with androgens beyond 30 months, although users of androgenic-anabolic steroids generally have spontaneous sperm recovery within 12 months [55]. MHC efficacy data are limited for men of African or Hispanic origin; longer treatment studies in more diverse populations of men are required [54], although one study in African men demonstrated full reversibility [24].

Acceptability

Survey data suggest that men, women, and heterosexual couples across a wide range of countries find hypothetical MHC acceptable and appealing. In an extensive cross-cultural, multicentre population survey of more than 9000 men, the willingness to use a hypothetical MHC averaged 55% (range 29–71%) [56, 57]. A smaller four-country study yielded similar results [58]. A small Australian study [59], reported 75% of men willing to try an MHC. In this study, the three most popular choices for method of administration of MHC were (in descending order) an oral pill, a 3-monthly injection, or a 2-yearly injection. Female partners from a variety of cultures also indicate high level of interest in MHC. In a survey of 1894 women attending family planning clinics in Scotland, China and South Africa, 65% felt the responsibility for contraception falls too much on women and more than 90% in South Africa and Scotland thought that a 'male pill' was a good idea, with Chinese women only slightly less positive. Overall, 40–78% indicated they would support and trust their male partners in stable relationships to use MHC [60].

Most study participants in MHC trials would be willing to use the method under evaluation were it available, although such data have inherent responder bias given the self-selection of participants volunteering in clinical trials. A Chinese acceptability study [61] of monthly IM testosterone undecanoate (TU) carried out concurrently with, but independently from, a clinical efficacy trial, indicated the majority of participants were satisfied with monthly injections. More recently, as part of a six-month, double-blind, randomized controlled trial of T and NES transdermal gels, 56% of men were satisfied with the method, and 51% reported that they would recommend a transdermal gel-based MHC [11]. More than 75% of

men, and their female partners, who participated in the most recent MHC efficacy study of long-acting T and progestin injections would use the method were it available [28].

Together, these studies support the development of effective MHC but further work is needed to more clearly define the potential market. Such work is challenging without an available product, as the introduction of new contraceptive methods may influence behaviours and acceptability in unforeseen ways. Models of limited uptake suggest that MHC could influence unplanned pregnancy rates [62].

Adverse Effects

Potential adverse effects from MHC include alterations in sexual function, mood, body composition, and other changes in androgen-sensitive tissues. Side effects of exogenous androgen administration include reversible testicular atrophy and acne, and these can occur with T containing MHC. Androgen administration can also modestly lower high-density lipoprotein-cholesterol (HDL-C), as has been observed in some MHC studies [54], although the clinical significance of alterations in HDL-C are unclear [63].

Potential adverse effects of MHC related to mood and sexual function are difficult to quantify as very few studies of prototype MHC have included a placebo group. A recent placebo-controlled androgen-progestin trial [53] to assess sperm suppression and safety, demonstrated changes in mood, libido, acne, weight gain, and night sweats that were statistically more prevalent than in the active compared to placebo groups. Adverse effects were self-reported by 93% of men receiving active treatment versus 81% of placebo-treated men. Although mood changes have been reported as mild in T-only studies, the most recent MHC efficacy study combing TU and a long-acting progestin [28] raised concerns regarding the potential for adverse mood effects of progestin-containing regimens, a finding that requires further study in future MHC trials.

The influence of MHC on long-term disease risk in men is unknown. There has been no evidence to suggest that MHC increases the risk of prostate disease or cardiovascular events, but studies to date have been done in healthy men who are young or middle-aged (18–50 years), have been short in duration (1–2.5 years), and underpowered to identify cardiovascular, thrombotic, or prostatic events. Longer-term studies will be required, at a minimum, as well as post-marketing evaluation if and when MHC is approved.

Conclusion

MHC has the potential to offer a new effective and reversible family planning option for men and couples. Androgens provide the basis of MHC via suppression of the hypothalamic–pituitary–testicular axis and, as a result, spermatogenesis; adding a progestin improves sperm suppression and can reduce androgen dosing and androgen-related side effects. Short-term side effects of T-progestin regimens are mild, but there are concerns regarding potential effects on mood and long-term disease risks. Novel, orally bioavailable, synthetic androgens with both androgen and progestin activity, are in clinical trials and are promising candidates as single-agent MHC, including as a once daily 'male pill'.

REFERENCES

1. Ross J, Hardee K. Use of male methods of contraception worldwide. *J Biosoc Sci*, 2017; **49**: 648–63.
2. Sedgh G, Singh S, Hussain R. Intended and unintended pregnancies worldwide in 2012 and recent trends. *Stud Fam Plann*, 2014; **45**: 301–14.
3. Wang C, Festin MP, Swerdloff RS. Male hormonal contraception: where are we now? *Curr Obstet Gynecol Rep*, 2016; **5**: 38–47.
4. Liu PY, Swerdloff RS, Christenson PD, *et al*. Rate, extent, and modifiers of spermatogenic recovery after hormonal male contraception: an integrated analysis. *Lancet*, 2006; **367**: 1412–20.
5. Meriggiola MC, Bremner WJ, Costantino A, Pavani A, Capelli M, Flamigni C. An oral regimen of cyproterone acetate and testosterone undecanoate for spermatogenic suppression in men. *Fertil Steril*, 1997; **68**: 844–50.
6. Buchter D, von Eckardstein S, von Eckardstein A, *et al*. Clinical trial of transdermal testosterone and oral levonorgestrel for male contraception. *J Clin Endocrinol Metab*, 1999; **84**: 1244–9.
7. Gonzalo IT, Swerdloff RS, Nelson AL, *et al*. Levonorgestrel implants (Norplant II) for male contraception clinical trials: combination with transdermal and injectable testosterone. *J Clin Endocrinol Metab*, 2002; **87**: 3562–72.
8. Hair WM, Kitteridge K, O'Connor DB, Wu FC. A novel male contraceptive pill-patch combination: oral desogestrel and transdermal testosterone in the suppression of spermatogenesis in normal men. *J Clin Endocrinol Metab*, 2001; **86**: 5201–9.
9. Ilani N, Roth MY, Amory JK, *et al*. A new combination of testosterone and nestorone transdermal gels for male hormonal contraception. *J Clin Endocrinol Metab*, 2012; **97**: 3476–86.
10. Page ST, Amory JK, Anawalt BD, *et al*. Testosterone gel combined with depomedroxyprogesterone acetate is an effective male hormonal contraceptive regimen and is not enhanced by the addition of a GnRH antagonist. *J Clin Endocrinol Metab*, 2006; **91**: 4374–80.
11. Amory JK, Page ST, Anawalt BD, Matsumoto AM, Bremner WJ. Acceptability of a combination testosterone gel and depomedroxyprogesterone acetate male contraceptive regimen. *Contraception*, 2007; **75**: 218–23.
12. Handelsman DJ, Mackey MA, Howe C, Turner L, Conway AJ. An analysis of testosterone implants for androgen replacement therapy. *Clin Endocrinol (Oxf)*, 1997; **47**: 311–16.
13. Turner L, Conway AJ, Jimenez M, *et al*. Contraceptive efficacy of a depot progestin and androgen combination in men. *J Clin Endocrinol Metab*, 2003; **88**: 4659–67.
14. Kelleher S, Howe C, Conway AJ, Handelsman DJ. Testosterone release rate and duration of action of testosterone pellet implants. *Clin Endocrinol (Oxf)*, 2004; **60**: 420–8.
15. Roth MY, Page ST, Bremner WJ. Male hormonal contraception: looking back and moving forward. *Andrology*, 2016; **4**: 4–12.
16. World Health Organization Task Force on Methods for the Regulation of Male Fertility. Contraceptive efficacy of testosterone-induced azoospermia in normal men. *Lancet*, 1990; **336**: 955–9.
17. Kubba A, Guillebaud J, Anderson RA, MacGregor EA. Contraception. *Lancet*, 2000; **356**: 1913–19.
18. World Health Organization Task Force on Methods for the Regulation of Male, Fertility. Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil Steril*, 1996; **65**: 821–9.

19. Aaltonen P, Amory JK, Anderson RA, *et al*. 10th Summit Meeting consensus: recommendations for regulatory approval for hormonal male contraception. *J Androl*, 2007; **28**: 362–3.
20. Gu YQ, Wang XH, Xu D, *et al*. A multicenter contraceptive efficacy study of injectable testosterone undecanoate in healthy Chinese men. *J Clin Endocrinol Metab*, 2003; **88**: 562–8.
21. Gu Y, Liang X, Wu W, *et al*. Multicenter contraceptive efficacy trial of injectable testosterone undecanoate in Chinese men. *J Clin Endocrinol Metab*, 2009; **94**: 1910–15.
22. Liu PY, Swerdloff RS, Anawalt BD, *et al*. Determinants of the rate and extent of spermatogenic suppression during hormonal male contraception: an integrated analysis. *J Clin Endocrinol Metab*, 2008; **93**: 1774–83.
23. Anderson RA, Kinniburgh D, Baird DT. Suppression of spermatogenesis by etonogestrel implants with depot testosterone: potential for long-acting male contraception. *J Clin Endocrinol Metab*, 2002; **87**: 3640–9.
24. Anderson RA, Van Der Spuy ZM, Dada OA, *et al*. Investigation of hormonal male contraception in African men: suppression of spermatogenesis by oral desogestrel with depot testosterone. *Hum Reprod*, 2002; **17**: 2869–77.
25. Kinniburgh D, Zhu H, Cheng L, Kicman AT, Baird DT, Anderson RA. Oral desogestrel with testosterone pellets induces consistent suppression of spermatogenesis to azoospermia in both Caucasian and Chinese men. *Hum Reprod*, 2002; **17**: 1490–501.
26. Meriggiola MC, Costantino A, Saad F, *et al*. Norethisterone enanthate plus testosterone undecanoate for male contraception: effects of various injection intervals on spermatogenesis, reproductive hormones, testis, and prostate. *J Clin Endocrinol Metab*, 2005; **90**: 2005–14.
27. Anawalt BD, Herbst KL, Matsumoto AM, *et al*. Desogestrel plus testosterone effectively suppresses spermatogenesis but also causes modest weight gain and high-density lipoprotein suppression. *Fertil Steril*, 2000; **74**: 707–14.
28. Behre HM, Zitzmann M, Anderson RA, *et al*. Efficacy and safety of an injectable combination hormonal contraceptive for men. *J Clin Endocrinol Metab*, 2016; **101**: 4779–88.
29. Bagatell CJ, Matsumoto AM, Christensen RB, Rivier JE, Bremner WJ. Comparison of a gonadotropin releasing-hormone antagonist plus testosterone (T) versus T alone as potential male contraceptive regimens. *J Clin Endocrinol Metab*, 1993; **77**: 427–32.
30. Swerdloff RS, Bagatell CJ, Wang C, *et al*. Suppression of spermatogenesis in man induced by Nal-Glu gonadotropin releasing hormone antagonist and testosterone enanthate (TE) is maintained by TE alone. *J Clin Endocrinol Metab*, 1998; **83**: 3527–33.
31. Behre HM, Kliesch S, Lemcke B, von Eckardstein S, Nieschlag E. Suppression of spermatogenesis to azoospermia by combined administration of GnRH antagonist and 19-nortestosterone cannot be maintained by this non-aromatizable androgen alone. *Hum Reprod*, 2001; **16**: 2570–7.
32. Herbst KL, Coviello AD, Page S, Amory JK, Anawalt BD, Bremner WJ. A single dose of the potent gonadotropin-releasing hormone antagonist acyline suppresses gonadotropins and testosterone for 2 weeks in healthy young men. *J Clin Endocrinol Metab*, 2004; **89**: 5959–65.
33. Cook CE, Kepler JA. 7alpha,11beta-Dimethyl-19-nortestosterone: a potent and selective androgen response modulator with prostate-sparing properties. *Bioorg Med Chem Lett*, 2005; **15**: 1213–16.
34. Taylor AP, Lee H, Webb ML, Joffe H, Finkelstein JS. Effects of testosterone and estradiol deficiency on vasomotor symptoms in hypogonadal men. *J Clin Endocrinol Metab*, 2016; **101**: 3479–86.
35. LaMorte A, Kumar N, Bardin CW, Sundaram K. Aromatization of 7 alpha-methyl-19-nortestosterone by human placental microsomes *in vitro*. *J Steroid Biochem Mol Biol*, 1994; **48**: 297–304.
36. Beri R, Kumar N, Savage T, Benalcazar L, Sundaram K. Estrogenic and progestational activity of 7alpha-methyl-19-nortestosterone, a synthetic androgen. *J Steroid Biochem Mol Biol*, 1998; **67**: 275–83.
37. Anderson RA, Wallace AM, Sattar N, Kumar N, Sundaram K. Evidence for tissue selectivity of the synthetic androgen 7 alpha-methyl-19-nortestosterone in hypogonadal men. *J Clin Endocrinol Metab*, 2003; **88**: 2784–93.
38. von Eckardstein S, Noe G, Brache V, *et al*. A clinical trial of 7 alpha-methyl-19-nortestosterone implants for possible use as a long-acting contraceptive for men. *J Clin Endocrinol Metab*, 2003; **88**: 5232–9.
39. Nieschlag E, Kumar N, Sitruk-Ware R. 7alpha-methyl-19-nortestosterone (MENTR): the population council's contribution to research on male contraception and treatment of hypogonadism. *Contraception*, 2013; **87**: 288–95.
40. Attardi BJ, Hild SA, Koduri S, *et al*. The potent synthetic androgens, dimethandrolone (7alpha,11beta-dimethyl-19-nortestosterone) and 11beta-methyl-19-nortestosterone, do not require 5alpha-reduction to exert their maximal androgenic effects. *J Steroid Biochem Mol Biol*, 2010; **122**: 212–18.
41. Attardi BJ, Marck BT, Matsumoto AM, Koduri S, Hild SA. Long-term effects of dimethandrolone 17beta-undecanoate and 11beta-methyl-19-nortestosterone 17beta-dodecylcarbonate on body composition, bone mineral density, serum gonadotropins, and androgenic/anabolic activity in castrated male rats. *J Androl*, 2011; **32**: 183–92.
42. Attardi BJ, Engbring JA, Gropp D, Hild SA. Development of dimethandrolone 17beta-undecanoate (DMAU) as an oral male hormonal contraceptive: induction of infertility and recovery of fertility in adult male rabbits. *J Androl*, 2011; **32**: 530–40.
43. Surampudi P, Page ST, Swerdloff RS, *et al*. Single, escalating dose pharmacokinetics, safety and food effects of a new oral androgen dimethandrolone undecanoate in man: a prototype oral male hormonal contraceptive. *Andrology*, 2014; **2**: 579–87.
44. Ayoub R, Page ST, Swerdloff RS, *et al*. Comparison of the single dose pharmacokinetics, pharmacodynamics, and safety of two novel oral formulations of dimethandrolone undecanoate (DMAU): a potential oral, male contraceptive. *Andrology*, 2017; **5**: 278–85.
45. Wu S, Yuen F, Swerdloff RS, *et al*. Single, escalating dose pharmacokinetics, safety, and food effects of a new oral androgen 11β-methyl-19-nortestosterone 17β-dodecylcarbonate in man: a potential oral male hormonal contraceptive. *J Clin Endocrinol Metab*, 2019; **104**: 629–38.
46. Thirumalai A, Ceponis J, Amory JK, *et al*. Pharmacokinetic and pharmacodynamic effects of 28 days of oral dimethandrolone undecanoate in healthy men: a prototype male pill. *J Clin Endocrinol Metab*, 2019; **104**: 423–32.
47. Cooper TG, Noonan E, von Eckardstein S, *et al*. World Health Organization reference values for human semen characteristics. *Hum Reprod Update*, 2010; **16**: 231–45.
48. Handelsman DJ, Farley TM, Peregoudov A, Waites GM, *et al*. Factors in nonuniform induction of azoospermia by testosterone enanthate in normal men. World Health Organization Task Force on Methods for the Regulation of Male Fertility. *Fertil Steril*, 1995; **63**: 125–33.

49. McLachlan RI, Robertson DM, Pruysers E, *et al*. Relationship between serum gonadotropins and spermatogenic suppression in men undergoing steroidal contraceptive treatment. *J Clin Endocrinol Metab*, 2004; **89**: 142–9.
50. Simoni M, Huhtaniemi IT. *Endocrinology of the Testis and Male Reproduction*, Vol. 1. Cham: Springer International, 2017.
51. Page ST, Amory JK, Bremner WJ. Advances in male contraception. *Endocr Rev*, 2008; **29**: 465–93.
52. Oduwole OO, Huhtaniemi IT. Feasibility of male hormonal contraception: lessons from clinical trials and animal experiments. *Curr Mol Pharmacol*, 2014; **7**: 109–18.
53. Mommers E, Kersemaekers WM, Elliesen J, *et al*. Male hormonal contraception: a double-blind, placebo-controlled study. *J Clin Endocrinol Metab*, 2008; **93**: 2572–80.
54. Piotrowska K, Wang C, Swerdloff RS, Liu PY. Male hormonal contraception: hope and promise. *Lancet Diabetes Endocrinol*, 2017; **5**: 214–23.
55. de Souza GL, Hallak J. Anabolic steroids and male infertility: a comprehensive review. *BJU Int*, 2011; **108**: 1860–5.
56. Heinemann K, Saad F, Wiesemes M, Heinemann LA. Expectations toward a novel male fertility control method and potential user types: results of a multinational survey. *J Androl*, 2005; **26**: 155–62.
57. Heinemann K, Saad F, Wiesemes M, White S, Heinemann L. Attitudes toward male fertility control: results of a multinational survey on four continents. *Hum Reprod*, 2005; **20**: 549–56.
58. Martin CW, Saad F, Wiesemes M, White S, Heinemann L. Potential impact of hormonal male contraception: cross-cultural implications for development of novel preparations. *Hum Reprod*, 2000; **15**: 637–45.
59. Weston GC, Schlipalius ML, Bhuinneain MN, Vollenhoven BJ. Will Australian men use male hormonal contraception? A survey of a postpartum population. *Med J Aust*, 2002; **176**: 208–10.
60. Glasier AF, Anakwe R, Everington D, *et al*. Would women trust their partners to use a male pill? *Hum Reprod*, 2000; **15**: 646–9.
61. Zhang L, Shah IH, Liu Y, Vogelsong KM, Zhang L. The acceptability of an injectable, once-a-month male contraceptive in China. *Contraception*, 2006; **73**: 548–53.
62. Dorman E, Perry B, Polis CB, *et al*. Modeling the impact of novel male contraceptive methods on reductions in unintended pregnancies in Nigeria, South Africa, and the United States. *Contraception*, 2018; **97**: 62–9.
63. Hovingh GK, Rader DJ, Hegele RA. HDL re-examined. *Curr Opin Lipidol*, 2015; **26**: 127–32.

10.9

Management of Gynaecomastia

Glenn D. Braunstein

Introduction

The term gynaecomastia is used to describe the unilateral or bilateral benign enlargement of the glandular tissue in the male breast. Various studies have defined gynaecomastia as glandular tissue exceeding 0.5–2.0 cm [1, 2]. The condition may present as an acute growth with pain and tenderness, or may be an incidental, painless finding detected during a physical examination. It is relatively common and needs to be distinguished from other causes of breast enlargement in males including fat deposition (lipomastia or pseudogynaecomastia) and breast carcinoma. Although gynaecomastia itself is benign, and often occurs as part of normal puberty or senescence, or as an adverse reaction to some drugs, its presence may indicate an underlying pathological abnormality. Therefore, it is important to have an understanding of its pathophysiology, causes, and diagnostic approach to arrive at a rational plan for managing gynaecomastia and any associated conditions.

Prevalence of Gynaecomastia

Transient gynaecomastia has been noted in 60–90% of newborns, presumably due to the transplacental transfer of maternal oestrogens during pregnancy. Physiological gynaecomastia also occurs during puberty, ranging from 4% to 69% in different epidemiological studies, with a median prevalence of about one-third of adolescent boys exhibiting some degree of breast glandular tissue enlargement. Clinical and autopsy studies have shown that about one-third to two-thirds of adult men may exhibit gynaecomastia [1–3].

Pathophysiology

Breast tissue at birth is identical in both males and females, and their responses to hormonal stimulation is the same. Oestrogens stimulate growth and branching of the ducts, hyperplasia of the ductal epithelium, fibroblast proliferation in the periductal areas, and increase in the vascularity of the breasts. Insulin-like growth factor may help sensitize the glandular tissue to the effects of oestrogens. Acinar development, a requirement for lactation to occur under the influence of prolactin, is generally not seen in males because luteal phase progesterone levels are required. Androgens antagonize the effects of oestrogen on breast glandular tissue. Thus, gynaecomastia occurs when there is an imbalance between the stimulatory effects of oestrogens and the inhibitory effects of androgens [1].

An alteration of this balance can occur when oestrogen secretion is increased, androgen secretion decreased, exogenous oestrogen or oestrogen-like substances are present, exogenous antiandrogens are present, or there are abnormalities in androgen receptor function.

About 95% of testosterone, 15% of oestradiol, and less than 5% of estrone are secreted by the testes in adult males. The majority of the circulating oestradiol and estrone is produced in peripheral tissues from androgen precursors, testosterone, and androstenedione, the latter being secreted by the adrenal glands. This conversion is mediated by the aromatase enzyme complex that is widely distributed throughout the body, being present in the testes, adrenals, fat, skin, liver, muscles, bone, and kidneys. Aromatase activity increases with age, reflecting the age-related increase in fat mass, and also the increased aromatase activity per fat cell. Testosterone and androstenedione are interconverted in extragonadal tissues by 17-ketosteroid reductase, as is oestradiol and estrone. Approximately one-half to two-thirds of the sex steroid hormones are transported in the blood bound to sex hormone-binding globulin (SHBG), a low capacity, high affinity transport protein; about one-third bound to albumin, a high capacity, low affinity binding protein; with 1–2% being unbound or free to enter target tissues. Since the steroids bound to albumin may disassociate and also enter target tissues, the combination of free and albumin-bound steroids is termed 'bioavailable'. When testosterone enters target tissues, it may bind directly to the androgen receptor or be converted to dihydrotestosterone (DHT) by 5α-reductase, which also binds to the androgen receptor. Following receptor binding, the androgen-androgen receptor complex is translocated into the

nucleus, binds to DNA, and transcribes androgen action in the target tissues. Similarly, oestrogens bind to oestrogen receptors to bring about oestrogenic effects. Alterations in any of these activities may result in a change in the oestrogen/androgen relationship to a sufficient degree to stimulate breast glandular tissue growth [1, 2].

Conditions Associated with Gynaecomastia

Box 10.9.1 lists the conditions that have been associated with gynaecomastia and the primary pathophysiologic mechanisms identified for pathologic gynaecomastia. Combined results from several series of adult patients seeking consultation at plastic surgical or endocrine clinics have shown that 25% of patients have acute or persistent pubertal gynaecomastia, 25% were classified as having idiopathic gynaecomastia, 10–20% related to medications or drugs, 8% from cirrhosis or malnutrition, 8% from primary hypogonadism, 3% from testicular tumours, 2% secondary hypogonadism, 1.5% from hyperthyroidism, and 1% from kidney disease (2). A more recent survey at a single medical centre of 786 men with gynaecomastia over the age of 18 years reported that 25% had persistent pubertal gynaecomastia, 37% with idiopathic, 10.8% associated with medicine intake, 10% had used anabolic steroids or suffered from substance abuse, 10% had hypogonadism, 2.9% hepatic insufficiency, 2.3% other endocrine disorders, 0.8% had a testicular tumour, and 0.1% renal insufficiency (3).

Pubertal Gynaecomastia

Pubertal gynaecomastia has an onset usually between the ages of 10 and 12 years, with a peak occurrence between the ages of 13 and 14, when Tanner genital or pubic hair stages 3 to 5 are reached. Gynaecomastia usually occurs within one year of the boy's peak height velocity. The development of gynaecomastia may be asynchronous, with one breast enlarging weeks to months before the other enlarges. Although most boys who develop gynaecomastia will have it resolved in 6–18 months, it may persist beyond puberty in up to 20% or higher of late adolescents to young men, accounting for the high proportion of patients with gynaecomastia presenting to endocrine or plastic surgical clinics for evaluation and treatment [4].

The cause of pubertal gynaecomastia has not been fully elucidated. Some studies have shown an imbalance in oestrogens/androgens at the start of or during pubertal gynaecomastia while others have not, possibly due to the investigator missing the peaks of increased oestrogen production, which may have antedated the appearance of the breast growth. The rate of change of oestrogen and androgen production during puberty may also be a factor, since oestrogen levels rise 3-fold across puberty and androgens increase 30-fold, with the oestrogen levels reaching adult concentrations before the testosterone concentration peaks. The activity of aromatase may be increased in pubertal males who develop gynaecomastia in comparison to those who do not, as aromatase activity was shown to be increased in pubic skin fibroblasts in some boys with pubertal gynaecomastia. However, a randomized, placebo-controlled trial of the aromatase inhibitor anastrozole in pubertal gynaecomastia did not show any improvement over placebo, suggesting that aromatization of androgens in fat or other tissues is not of major importance in the genesis of pubertal gynaecomastia [5]. Similarly, studies have shown that serum oestradiol levels largely determine the oestrogen

Box 10.9.1 Conditions associated with gynaecomastia and their pathophysiologic mechanisms

Physiologic
Neonatal
Pubertal
Ageing

Pathologic
Decreased testosterone action
- Primary hypogonadism
- Secondary hypogonadism
- Defects in testosterone synthesis
  - Congenital
  - Drug-induced (e.g. ketoconazole)
- Decreased androgen receptor function
  - Congenital androgen receptor defects
  - Drug-induced (e.g. spironolactone, cimetidine)

Increased oestrogen
- Increased aromatization of androgens
  - Germ cell tumours (testicular and extragonadal)
  - Sertoli cell (sex cord) tumours
  - Leydig cell tumours
  - Feminizing adrenocortical tumours
  - Disorders of sexual development
  - Obesity
  - Hyperthyroidism
  - Liver disease
  - Androgen insensitivity syndrome
  - Refeeding after starvation
  - Primary aromatase-excess
- Displacement of oestrogen from SHBG
  - Spironolactone
  - Ketoconazole
- Decreased oestrogen metabolism
  - Cirrhosis
- Exogenous sources
  - Topical oestrogen creams, lotions, or patches
  - Ingestion of oestrogen
  - Environmental exposure
- Human chorionic gonadotrophin-producing tumours
  - Eutopic (germ cell tumours)
  - Ectopic (e.g. lung, liver, gastric, renal)

Drug-induced
- Good evidence
  - Oestrogens
  - Spironolactone
  - Cimetidine
  - Ketoconazole
  - Growth hormone
  - Gonadotropins
  - Antiandrogens
  - 5α-reductase inhibitors
- Fair evidence
  - Calcium channel blockers
  - Antipsychotics
  - Omeprazole
  - HIV drugs
  - Alkylating agents
  - Anabolic steroids
  - Alcohol
  - Opioids
  - Lavender and tea tree oil

Enhanced breast tissue sensitivity
Chronic renal insufficiency and dialysis
Idiopathic

Modified with permission from Mathur R, Braunstein G. Gynecomastia: Pathomechanisms and Treatment Strategies. *Hormone Research* 1997;48:95–102. Copyright © 1997 Karger Publishers.

levels in normal breast tissue. Blood levels of two other hormones, leptin, and insulin-like growth factor-1 (IGF-1), have been found to be elevated in boys with pubertal gynaecomastia in comparison to those without. The finding that pubertal gynaecomastia occurs in a temporal association with peak height velocity, the time when IGF-1 levels also peak, lends credence to the suggestion that oestrogens and IGF-1 both play a role in the genesis of pubertal gynaecomastia [6]. Also, the observation that gynaecomastia may occur during growth hormone therapy, which elevates the levels of IGF-1, supports a role of IGF-1 in the development of pubertal gynaecomastia. In addition, polymorphisms of the leptin and oestrogen receptor-β genes and the G protein-coupled receptor (GPR30), a membrane bound receptor mediating non-classical oestradiol signalling, have been implicated in an increased susceptibility for pubertal gynaecomastia [7].

Ageing

The frequent finding of asymptomatic gynaecomastia in older men is probably multifactorial. Body composition changes during ageing, with fat mass increasing while lean body mass decreases. Since fat contains aromatase and aromatase levels per fat cell increases with age, there is enhanced conversion of androgens to oestrogens as part of the ageing process. Additionally, in many men, testosterone levels decrease over time and SHBG concentrations increase, resulting in a lowering of the free and bioavailable fractions of testosterone, and, thus, less antagonism of oestrogen at the breast glandular tissue level. A rise in luteinizing hormone (LH) levels that accompanies the drop in testosterone in some older men also may stimulate enhanced aromatization of testosterone to oestradiol in the testes. Also, older men are more likely to take medications (**Box 10.9.1**) that may be associated with gynaecomastia [1, 2, 8].

Drug-Induced

Many different drugs, including some illicit substances, have been associated with gynaecomastia in adults, but most of the reports, often anecdotal or small clinical series, relied on just a temporal relationship between taking the drug and the subsequent occurrence of gynaecomastia. An evidence-based review placed drugs into three categories based upon the level of evidence. Good evidence required a systematic review of randomized controlled trials, or randomized placebo-controlled trial, or prospective cohort studies with or without concurrent controls plus a good pathophysiological explanation. Fair evidence required retrospective studies or case-control studies or case series with a good pathophysiological explanation. Isolated case reports alone were classified as poor level of evidence. Drugs with good evidence of association with gynaecomastia include oestrogens, spironolactone, cimetidine, ketoconazole, growth hormone, gonadotropins, antiandrogen therapies (flutamide, bicalutamide, nilutamide, enzalutamide, cyproterone acetate, goserelin, leuprorelin) and 5α-reductase inhibitors (finasteride, dutasteride). Drugs with fair evidence of an association with gynaecomastia include some antipsychotics (haloperidol, risperidone), calcium channel blockers (verapamil, nifedipine), omeprazole, HIV antiretroviral drugs (especially protease inhibitors), alkylating agents, anabolic steroids, lavender, and tea tree oil, alcohol, and opioids. All of the other drugs reported to be associated with gynaecomastia currently fall in the poor level of evidence category [9].

Drugs such as cimetidine, spironolactone, flutamide, bicalutamide, nilutamide, enzalutamide, and cyproterone acetate are known to bind to the androgen receptor and inhibit androgen action.

Drug-induced gynaecomastia generally has its onset as early as a week after starting the offending medication, with the majority of patients experiencing gynaecomastia within 9 months of continuous drug usage, although onset between 2 and 3 years of continuous drug usage have been reported.

Hypogonadism

Primary hypogonadism due to intrinsic testicular failure results in a low testosterone and elevation of LH and often follicle-stimulating hormone (FSH). Oestrogen antagonism at the breast glandular tissue level is reduced. The most common genetic cause is Klinefelter syndrome (XXY genotype) characterized by small firm testicles, gynaecomastia, and varying degrees of hypogonadism. The degree of hypogonadism and gynaecomastia is related in part to the number of CAG repeats in the androgen receptor gene—those with longer repeats are less virilized [10]. Patients with Klinefelter syndrome are at increased risk for the development of breast cancer as well as germ cell tumours, which may cause gynaecomastia through the secretion of human chorionic gonadotropin (hCG) [11]. Other causes of primary hypogonadism include anorchia, cryptorchidism, Leydig cell aplasia, Noonan syndrome, myotonic dystrophy, viral orchitis, granulomatous diseases, and defects in testosterone biosynthesis.

Secondary hypogonadism results from hypothalamic or pituitary dysfunction, and ranges from panhypopituitarism to specific defects in gonadotropin production or secretion. Genetic causes include defects in the *KAL1* gene that leads to anosmia and hypogonadotropic hypogonadism and mutations in the *FGFR1, KISS1, KISS1R, GnRH1, PROK2*, and *PROKR2* genes that are associated with secondary hypogonadism without defects in the sense of smell. Hypogonadotropic hypogonadism also is found with defects in the genes encoding leptin, leptin receptor, and the β-subunit of LH. Several complex congenital syndromes have secondary hypogonadism as part of the clinical picture [1]. These include Prader–Willi, Laurence–Moon Bardet–Biedl, Mobius, and Lowe syndromes, as well as cerebellar ataxia. Acquired causes may result from tumours, infection, vascular issues, and infiltrative diseases involving the pituitary or hypothalamus. The loss of hypothalamic gonadotropin-releasing hormone (GnRH) production or destruction of or defects in the gonadotrophs in the anterior pituitary results in inadequate LH secretion that in turn leads to lowered testosterone secretion.

Prolactin-secreting pituitary tumours in men causes hypogonadism through several mechanisms. Macroadenomas may destroy enough gonadotrophs to result in inadequate LH secretion. Additionally, the elevated prolactin may inhibit hypothalamic GnRH secretion and, thus, inadequate gonadotroph stimulation, reducing LH stimulation of testicular testosterone production. Finally, the high prolactin may directly suppress testosterone release from the testes. Lowering the prolactin through the use of a dopamine agonist often results in a rise in testosterone as long as sufficient gonadotrophs are functional. Of note, prolactin, which stimulates milk production in breasts that have been primed by oestrogens and progesterone, does not stimulate breast glandular growth and has not been directly associated with the development of gynaecomastia.

Testicular Tumours

Approximately 95% of testicular tumours are germ cell neoplasms and close to 10% of patients have gynaecomastia at presentation due to the secretion of hCG which inhibits testicular 17,20-lyase and 17-hydroxylase activity and stimulates Leydig cell aromatase activity, leading to a relative increase of oestradiol to testosterone [12]. Following treatment, close to 15% of patients will develop at least transient gynaecomastia due to therapy-related hypogonadism Ectopic production of hCG by non-trophoblastic neoplasms, such as lung, gastric, renal, or hepatocellular carcinomas, may rarely be associated with gynaecomastia if the tumour secretes intact hCG rather than the biologically inactive hCGβ-subunit, through the same mechanism as with the germ cell trophoblastic tumours (eutopic production).

Leydig cell tumours secrete increased quantities of oestradiol resulting in gynaecomastia in 20–30% of males harbouring this tumour, 90% of which are benign. The elevated oestradiol inhibits LH secretion leading to a decrease in testosterone production, contributing to the oestrogen/androgen imbalance.

The rare large-cell-calcifying Sertoli cell (sex cord) testicular tumours exhibit excessive aromatase activity and are associated with gynaecomastia. These tumours occur sporadically and in patients with two autosomal dominant disorders, Peutz–Jeghers syndrome, and Carney complex. Peutz–Jeghers is due to a mutation in the *STK11 or LKB1* genes and results in prepubertal gynaecomastia, gastrointestinal hamartomatous polyps, and mucocutaneous macules. Carney complex also is associated with prepubertal gynaecomastia, pigmented lentigines, blue nevi, myxomas, Cushing syndrome, and thyroid nodules including thyroid cancer [13].

Feminizing Adrenocortical Tumours

Gynaecomastia is found in virtually all males with feminizing adrenocortical tumours, three-quarters of which are malignant. Over half of these patients have a palpable tumour and about half of the patients have testicular atrophy. These tumours have high levels of aromatase activity and secrete increased amounts of oestrogens. The high oestrogen levels also lead to secondary hypogonadism with a reduction in testosterone [14].

Disorders of Sexual Development

Disorders of sexual development in which both testicular and ovarian tissue is present may also be associated with gynaecomastia due to the secretion of oestradiol from the ovarian tissue, which suppresses LH and testicular production of testosterone. These disorders are usually diagnosed in childhood because of the presence of ambiguous genitalia.

Hyperthyroidism

Estimates of the prevalence of gynaecomastia in males with hyperthyroidism vary from less than 10% to as high as 40%. Hyperthyroidism is associated with an increase in LH levels, excessive aromatization of androgens to oestrogens, and elevated SHBG levels [15]. Given SHBG binds testosterone with higher affinity than oestradiol, increased SHBG concentrations lead to a raised free E2/free T ratio, further aggravating the gynaecomastia.

Liver Disease

The gynaecomastia associated with severe liver disease or cirrhosis is multifactorial. About two-thirds of patients with cirrhosis exhibit breast glandular enlargement due to increased adrenal production of androstenedione, increased aromatization of the androstenedione to estrone, increased conversion of estrone to oestradiol, raised SHBG levels and the use of spironolactone to treat the fluid retention of cirrhosis [16, 17].

Obesity

Breast enlargement is common in obese males and may be due to fat deposition (pseudogynaecomastia or lipomastia), gynaecomastia, or a combination of the two. In adults, a close correlation has been found between the prevalence of gynaecomastia and the body mass index and the amount of adipose tissue [16]. As previously noted, fat tissue contains the aromatase enzyme complex and it has been found that oestrogen production rates increase in proportion to the percentage increase in body weight above normal [18]. Thus, it has been hypothesized that the gynaecomastia found in some obese patients is due to the local production of oestrogens from androgen precursors such as androstenedione by the periglandular fat tissue in the breast, a process termed intracrinology [19]. However, it has been shown that breast tissue oestradiol levels are primarily derived from the blood and not locally produced [20]. In addition, aromatase inhibitors have not been found to be effective in treating gynaecomastia not associated with the aromatase-excess syndrome (see next). Finally, there is conflicting data in adolescents, with some studies showing a correlation between body mass index and the prevalence of gynaecomastia, while others show no correlation or even a negative correlation. It is possible that much of the gynaecomastia reported in obese males essentially reflects lipomastia/pseudogynaecomastia and not true gynaecomastia.

Refeeding After Starvation

Testicular atrophy, decreased libido, and erectile dysfunction were noted to be a common occurrence in American prisoners of war following the Second World War, and about 10% exhibited gynaecomastia. Studies on patients with chronic disease and weight loss have shown that the gonadotropins are low as is testosterone. During refeeding, the gonadotropins rise as does the testosterone, and many men developed transient breast tenderness, pain, and gynaecomastia about 2–3 months following initiation of refeeding, a process that some have termed a 'second puberty'. Refeeding gynaecomastia resolves within one to two years. This mechanism probably accounts in part for the gynaecomastia that can occur in men with chronic renal failure after they start dialysis, and some patients with treated congestive heart failure [1, 2].

Chronic Renal Insufficiency

Men with chronic kidney disease often have gynaecomastia that may be multifactorial. Leydig cell dysfunction with low testosterone production and elevated LH levels may be present. Renal failure reduces the clearance of LH and prolactin and the elevated prolactin may contribute to the reduction in testosterone production [21, 22]. As already noted, improvement in nutrition and health status with dialysis or transplantation may lead to a transient 'refeeding' gynaecomastia.

Androgen Receptor Defects

Congenital inactivating mutations in the androgen receptor gene results in a spectrum of clinical findings ranging from the complete androgen insensitivity ('testicular feminization') syndrome to partial androgen insensitivity syndrome (PAIS). In the absence of androgen receptor function, affected individuals whose genotype is XY will appear to be phenotypic females at birth with open labioscrotal folds, a short vagina, absent uterus, and fallopian tubes, abdominal or inguinal testes. During puberty, there will be absence of pubic or axillary hair and, because of unopposed oestrogen action, gynaecomastia will develop. Patients with PAIS show a spectrum of clinical findings from some degree of ambiguous genitalia (e.g. hypospadias), to normal phenotypic males with gynaecomastia being the only clinical manifestation. In all of these syndromes, testosterone is elevated as is oestradiol (from aromatization of the elevated testosterone). LH is also elevated since suppression of GnRH from the androgen component is absent. The gynaecomastia is the result of unopposed oestrogen effect on the breast glandular tissue [23].

Patients with Kennedy syndrome (spinobulbar muscular atrophy), an X-linked disorder that leads to proximal muscle weakness with fasciculations and tremor and gynaecomastia is the result of an increased number of CAG repeats in the androgen receptor gene that leads to PAIS.

Congenital Aromatase-Excess Syndrome

The rare familial prepubertal gynaecomastia or aromatase-excess syndrome is an autosomal dominant constitutive activation of the aromatase gene (*CYP19A1*) due to heterozygous inversions, duplications, and polymorphisms. Boys with these mutations exhibit prepubertal or peripubertal gynaecomastia, accelerated growth and bone maturation in childhood, with diminished height in adulthood due to premature fusion of the epiphyseal plates in the long bones, elevated oestradiol and estrone from the excessive aromatization and low testosterone [4, 24].

Diagnostic Evaluation

Breast Examination

It is important to establish whether the patient has true gynaecomastia versus pseudogynaecomastia (lipomastia), breast carcinoma, or other conditions that can lead to unilateral or bilateral breast enlargement. Gynaecomastia cannot be diagnosed from a photograph. It requires careful palpation. One method is to have the patient lie flat on his back with his hands locked behind his head. The examiner separates his or her thumb and forefinger, places them on either side of the breast, and slowly and gently brings them together. True gynaecomastia will be felt as a firm or rubbery disc of tissue that is concentric to the nipple-areolar complex, while in patients with pseudogynaecomastia the thumb and forefinger will not meet resistance until they come to the nipple. If gynaecomastia is of recent onset, tenderness to palpation may be present [8].

In contrast, breast cancer, which is rare in men, generally presents as a unilateral, hard, or firm non-tender mass that is eccentric and may be fixed to the underlying tissues. Dimpling of the skin, nipple retraction, nipple discharge, and axillary lymphadenopathy may also be found.

Other lesions affecting the breast include hematomas, lipomas, dermoid cysts, neurofibromas, and lymphangiomas. These are usually unilateral and generally do not present as a mass that is concentric to the nipple-areolar complex.

History

It is important to document when the patient became aware of the gynaecomastia, whether it was unilateral or bilateral, was or is associated with pain and tenderness, and whether the breast tissue has continued to grow. A family history of gynaecomastia suggests a genetic aetiology. A medication and drug history is essential. Has the patient taken any of the drugs known to be associated with gynaecomastia, and is there environmental exposure to oestrogens or oestrogen-like agents? Over-the-counter preparations should be considered as some herbal remedies have been adulterated with hormones. Is there a temporal relationship between starting a medication and the occurrence of breast enlargement? A full pubertal history should be obtained including age of onset of axillary and pubic hair, growth spurt, acne, sexual function, and whether the patient recalls pubertal gynaecomastia. Symptoms of hyperthyroidism, liver, or renal disease should be sought. Recent weight loss or weight gain should be recorded and the reasons for the weight change evaluated.

Physical Examination

In addition to the breast examination, a full physical examination should be performed looking for signs of pituitary tumours, hyperthyroidism, liver or renal insufficiency, abdominal masses, testicular size and tumours, incomplete sexual development, and neurological abnormalities.

Laboratory Studies

Most breast masses in males represent gynaecomastia, especially if it is bilateral and extends concentrically beneath the nipple-areolar complex, and do not require imaging for diagnosis. However, if it is unclear whether the breast mass is gynaecomastia or breast carcinoma, an ultrasound, mammogram, breast MRI, or needle aspiration of the mass could be performed to make the distinction.

If a specific condition associated with gynaecomastia is suspected, appropriate laboratory and imaging studies should be performed. Chronic, asymptomatic gynaecomastia discovered incidentally on a physical examination is rarely associated with laboratory abnormalities. However, the prevalence of compensated or uncompensated primary hypogonadism increases with ageing, and, therefore, it is reasonable to measure a serum testosterone level. If it is low or low normal, remeasurement of total or free testosterone with a serum LH may provide the diagnosis [8].

Gynaecomastia occurring in the pubertal age range generally does not require any evaluation other than a careful physical examination and ascertaining whether the adolescent is using anabolic steroids. Most often reassurance and a follow-up examination at 3 monthly intervals until it resolves or is determined to be persistent is all that is necessary.

In an adult, the onset of acute, symptomatic gynaecomastia should warrant an evaluation of hepatic and renal function tests, thyroid-stimulating hormone (TSH), total testosterone, LH, hCG, and oestradiol. Figure 10.9.1 provides a means to arrive at the probable diagnosis based upon the results.

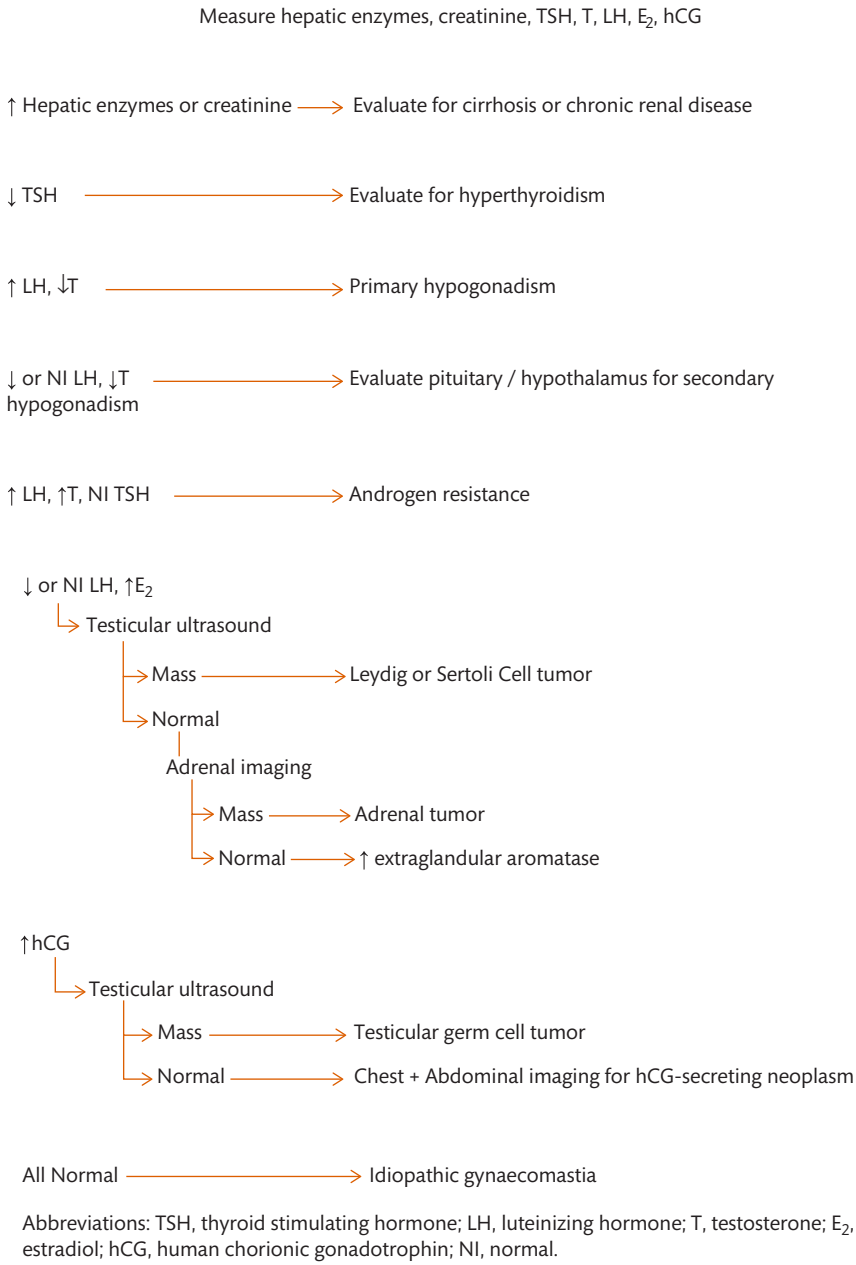


Figure 10.9.1 Laboratory evaluation for gynaecomastia.
Measure hepatic enzymes, creatinine, TSH, T, LH, E_2, hCG.
Abbreviations: TSH, thyroid stimulating hormone; LH, luteinizing hormone; T, testosterone; E_2, oestradiol; hCG, human chorionic gonadotrophin; NI, normal.

Treatment

Treatment depends upon the duration of the gynaecomastia, the underlying aetiology, if known, and whether the patient is bothered by its presence due to pain, tenderness, or embarrassment.

The longer the gynaecomastia is present, the less reversible it becomes. Histological studies have shown that there are three phases that the breast tissue exhibits. During the acute phase, which lasts up to about 6 months, there is proliferation of the ducts and hyperplasia of the ductal epithelium, along with an increase in the stromal and periductal connective tissue, and a marked increase in vascularity accompanied by periductal oedema. It is during this phase that pain and tenderness is present. If the inciting cause is treated (e.g. the medication stopped or a testicular tumour removed), regression of the gynaecomastia takes place. Also, it is during this stage that spontaneous resolution also is most likely. Gynaecomastia that has been present for over a year generally shows some hyalinization of the stroma, reduction of the ductal epithelial hyperplasia and dilation of the ducts. This fibrotic stage of gynaecomastia is less likely to respond to medical therapy or resolve on its own, although spontaneous resolution of gynaecomastia present for over 2 years has been noted in epidemiological studies. Between the acute florid stage and chronic fibrotic stages is an intermediate stage with mixtures of both components.

Since the majority of adolescents who develop pubertal gynaecomastia have spontaneous resolution of the condition, counselling, and observation generally is all that is needed. It is important to rule out confounding factors such as the use of anabolic steroids, and to perform a careful physical examination including the genitalia to determine that the gynaecomastia has occurred at the anticipated time during the pubertal process.

Several medical therapies have been examined to treat gynaecomastia. However, few of the trials have been double-blind and placebo-controlled, which is important when studying a condition that has a high spontaneous resolution rate. Additionally, the numbers of subjects in the individual trials generally have been low, the causes of the gynaecomastia have been mixed, and some have included patients with long-standing gynaecomastia which, because of the fibrosis in the breast, are unlikely to show improvement.

Medical therapies have included androgens, selective oestrogen receptors modulators (SERMS), and aromatase inhibitors.

In hypogonadal men with painful gynaecomastia, testosterone replacement may resolve the breast enlargement. However, long-standing gynaecomastia is unlikely to respond. Testosterone therapy for eugonadal men with gynaecomastia is not useful. Testosterone has also had some success in treating gynaecomastia associated with cirrhosis. The testosterone metabolite, dihydrotestosterone, which is not aromatizable has been used as a gel in uncontrolled studies to treat gynaecomastia. However, there is insufficient published data to recommend its use. The impeded androgen, danazol, has shown some efficacy in limited trials.

Several selective oestrogen receptor modulators, including clomiphene citrate, tamoxifen, and raloxifene, have been tried with some success. The best studied has been tamoxifen in doses of 10 mg twice daily for three months, used in a non-regulatory agency approved ('off label') manner. Analysis of six studies including 109 patients indicated that this treatment is effective in reducing the size of gynaecomastia in up to 80% of individuals and is safe, although it may not lead to complete resolution of the gynaecomastia [25].

Although theoretically ideal for treating a condition where there is an imbalance in oestrogen to androgens in breast tissue, the results from aromatase inhibitor trials have been disappointing. A double-blind, placebo-controlled trial in adolescents with pubertal gynaecomastia did not show superior results compared to the placebo arm [5]. However, aromatase inhibitors have been effective in patients with the aromatase-excess syndrome; further studies with some of the newer, more potent aromatase inhibitors in patients with recent-onset gynaecomastia should be performed.

For long-standing gynaecomastia, or gynaecomastia that does not respond to medical therapy, surgical removal of the breast tissue should be considered in patients with persistent pain and tenderness or who are psychologically distraught over the condition. There are several plastic surgical techniques that have been developed, but most include blunt dissection and removal of the glandular tissue along with liposuction to contour the breasts. Exuberant gynaecomastia often requires skin excision in addition. Complications of surgery include dehiscence of the wound, sloughing of tissue, contour irregularity, infection, seroma, or haematoma formation. Surgery should not be performed in adolescents until the patient is at the end of puberty because of the risk of regrowth of breast tissue if performed during the time that the oestrogen/androgen balance is fluctuating [26]. It is important to have the procedure performed by a breast plastic surgeon with extensive experience in performing gynaecomastia surgery in order to minimize the complication rate and maximize patient satisfaction.

Prevention

Men receiving antiandrogen treatment for prostate cancer have a high risk of developing gynaecomastia. Indeed, those receiving bicalutamide monotherapy, have about a 70% chance of developing the condition. Concurrent treatment of these individuals with tamoxifen reduces this rate to about 10%. Similarly, prophylactic radiotherapy also prevents antiandrogen-induced gynaecomastia in the majority of the patients [8].

REFERENCES

1. Narula HS, Carlson HE. Gynaecomastia—pathophysiology, diagnosis, and treatment. *Nat Rev Endocrinol*, 2014; **10**: 684–98.
2. Braunstein GD. Gynecomastia. *N Engl J Med*, 1993; **328**: 490–5.
3. Mieritz MG, Christiansen P, Jensen MB, *et al.* Gynaecomastia in 786 adult men: clinical and biochemical findings. *Euro J Endocrinology*, 2017; **176**: 555–66.
4. Ma NS, Geffner ME. Gynecomastia in prepubertal and pubertal men. *Curr Opin Pediatr*, 2008; **20**: 465–70.
5. Plourde PV, Reiter EO, Jou H-C, *et al.* Safety and efficacy of anastrozole for the treatment of pubertal gynecomastia: a randomized, double-blind, placebo-controlled trial. *J Clin Endocrinol Metab*, 2004; **89**: 4428–33.
6. Mieritz MG, Raket LL, Hagen CP, *et al.* A longitudinal study of growth, sex steroids and insulin-like growth factor I in boys with physiological gynaecomastia. *J Clin Endocrinol Metab*, 2015; **100**: 3752–59.
7. Korkmaz HA, Edgunlu T, Eren E, *et al.* GPR30 gene polymorphisms are associated with gynecomastia risk in adolescents. *Horm Res in Pediatrics*, 2015; **83**: 177–82.
8. Braunstein GD. Gynecomastia. *N Engl J Med*, 2007; **357**: 1229–37.
9. Deepinder F, Braunstein GD. Drug-induced gynecomastia: an evidence-based review. *Expert Opin Drug Saf*, 2012; **11**: 779–95.
10. Zitzmann M, Nieschlag E. The CAG repeat polymorphism within the androgen receptor gene and maleness. *Int J Androl*, 2003; **26**: 76–83.
11. Sokol RZ. It's not all about the testes: medical issues in Klinefelter patients. *Fert Steril*, 2012; **98**: 261–5.
12. Hassan HC, Cullen IM, Casey RG, Rogers E. Gynaecomastia: an endocrine manifestation of testicular cancer. *Andrologia*, 2008; **40**: 152–7.
13. Gourgari E, Saloustros E, Stratakis CA. Large-cell calcifying Sertoli cell tumors of the testes in pediatrics. *Curr Opinion Pediatr*, 2012; **24**: 518–22.
14. Chentii F, Bekkaye I, Azzoug S. Feminizing adrenocortical tumors: literature review. *Indian J Endocrinol Metab*, 2015; **19**: 332–9.
15. Meikle AW. The interrelationships between thyroid dysfunction and hypogonadism in men and boys. *Thyroid*, 2004; **14**(Suppl. 1): S17–25.
16. Sinclair M, Grossmann M, Gow PJ, Angus PW. Testosterone in men with advanced liver disease: abnormal ties and implications. *J Gastroenterol Hepatol*, 2015; **30**: 244–51.

17. Niewoehner CB, Nuttall FQ. Gynecomastia in a hospitalized male population. *Am J Med*, 1984; **77**: 633–7.
18. Braunstein GD. Aromatase and gynecomastia. *Endocrine-Related Cancer*, 1999; **6**: 315–24.
19. Africander D, Storbeck K-H. Steroid metabolism in breast cancer. Where are we and what are we missing? *Mol Cell Endocrinol*, 2018; **466**: 86–97.
20. Depypere HT, Bolca S, Bracke M, Delanghe J, Comhaire F, Bondeel PH. The serum estradiol concentration is the main determinant of the estradiol concentration in normal breast tissue. *Maturitas*, 2015; **81**: 42–5.
21. Klang E, Kanana N, Grossman A, *et al*. Quantitative CT assessment of gynecomastia in the general population and in dialysis, cirrhotic, and obese patients. *Acad Radiol*, 2018; **25**: 626–35.
22. Iglesias P, Carrero JJ, Diez, JJ. Gonadal dysfunction in men with chronic kidney disease: clinical features, prognostic implications and therapeutic options. *J Nephrol*, 2012; **25**: 31–42.
23. Hiort O. Clinical and molecular aspects of androgen insensitivity. *Endocr Develop*, 2013; **24**: 33–40.
24. Fukami M, Tsuchiya T, Volbach H, *et al*. Genomic basis of aromatase excess syndrome: recombination and replication—mediated rearrangements leading to *CYP19A1* overexpression. *J Clin Endo Metab*, 2013; **98**: E2013–21.
25. Gruntmanis U, Braunstein GD. Treatment of gynecomastia. *Curr Opin Investig Drugs*, 2001; **2**: 643–9.
26. Rahmani S, Turton P, Shaaban A, Dall B. Overview of gynecomastia in the modern era and the Leeds Gynaecomastia Investigation algorithm. *Breast J*, 2011; **17**: 246–55.

10.10

Exogenous Factors and Male Reproductive Health

10.10.1 Environmental Influences on Male Reproductive Health

Jorma Toppari

Introduction

Adverse trends in male reproductive health have raised concerns about environmental effects. Reproductive toxicology has long been one of the main requirements of chemical testing that aims to demonstrate product safety. The crucial role of hormonal regulation of the development and functions of reproductive organs brought endocrine-disrupting chemicals, also referred to as endocrine disruptors, into the limelight of reproductive toxicology. Endocrine disruptors have been on the top of agenda in chemical regulation in the European Union, the United States, and elsewhere with reproductive health being a major focus. What is the role of endocrine disruptors in male reproductive health problems? Male reproductive health and endocrine disruption was first considered with oestrogenic chemicals [1], but antiandrogens have attracted more attention recently. Comprehensive reports have been produced by World Health Organization [2], and Endocrine Society [3].

Environmental effects are not limited to chemical exposures. Irradiation, heat, and other physical effects can contribute to reproductive problems. Lifestyle and general health also affect reproductive function. Consideration of all these factors together is necessary for full assessment of possible environmental effects on reproduction.

As for many other health outcomes, foundations of reproductive health are largely laid early in development. In normal male development, structural integrity of reproductive organs is established already during gestational weeks 8–15, and the perinatal period is important for testicular and penile growth. Late prepuberty and puberty finalize sexual and reproductive development. Adverse effects during these developmental windows can produce irreversible damage. In adult age, reproductive health can also be impaired, but the effects are often temporary and therefore reversible. For example, cancer therapy can induce sterility for several years, after which spermatogenesis recovers if spermatogonial stem cells have not been destroyed. Environmental influences can be analysed in these two contexts: developmental effects and adult effects. As said, the former can cause structural anomalies and permanent functional deficiencies, whereas the latter cause functional defects that can be reversible. There is robust evidence of adverse effects of endocrine disruptors on the male reproductive system in experimental animals (**Table 10.10.1.1**). In contrast, human evidence is based on epidemiological observational studies where single exposures have been correlated with various outcomes and the results are less clear. Occupational and accidental exposures are informative exceptions as described next.

Contemporary Effects

Cancer chemotherapy and irradiation have direct toxic effects on spermatogenic cells. Spermatogonia are actively proliferating and therefore particularly susceptible. Critical doses of X-irradiation have been quantified decades ago [4]. Gonadal exposure to higher than 4–6 Gy can produce permanent azoospermia, whereas recovery can occur within 5 years after lower doses, and the recovery time is inversely correlated to the dose. Even exposure to 0.15 Gy produces a significant but transient decline in sperm production, and there is no safe radiation dose. Fractionation of irradiation leads to even worse disruption than a single dose. Alkylating cancer drugs, such as cyclophosphamide, can cause permanent azoospermia, while non-alkylating agents (e.g. vincristine and methotrexate) are less toxic for spermatogenic cells.

Table 10.10.1.1 Effects of endocrine disruptors on male reproductive system in experimental animals

EDC	Observations
Synthetic oestrogen Diethylstilboestrol (DES)	Reduction in testis weight; testicular lesions; reduction in the number of spermatogonia with multinucleate cells in lumina of testis; distension and overgrowth of the rete testis; underdevelopment of the epididymal duct epithelium; epididymal cysts; reduction in epithelial height in the vas deferens; convolution of the extra-epididymal vas; distension and reduction in epithelial height of the efferent ducts; decreased androgen receptor expression in testis, epithelium of the rete testis, caput and cauda epididymis and vas deferens; inflammatory disease of the accessory sex glands; nodular enlargements of the seminal vesicles and/or prostate; decreased testosterone levels; increased gonadotrophin levels; decreased pituitary response to GnRH; cryptorchidism; sterility
Antifouling agent (Tributyltin)	Increased anogenital distance; reduced number of Sertoli cells and gonocytes in fetal testis
Phytoestrogens (genistein, daidzein)	Impaired erectile function; decreased testosterone levels; increased testis weight; reduction in epithelial height of the efferent ducts
Surfactants Alkyl phenol ethoxylates (*p*-tert-octylphenol, *p*-nonylphenol)	Changes in testis weight; decreased seminiferous tubule diameter; decreased epididymal weight; decreased total cauda epididymal sperm count; reduction in epithelial height of the efferent ducts
Plasticizers Phthalate esters (DEHP, BBP, DINP, DBP)	Histopathological changes of testis; decreased testis weight; multinucleated gonocytes; haemorrhagic testis; lesion of the rete testis; reduced anogenital distance; nipple retention; cryptorchidism; reduced accessory sex organ weights; agenesis of epididymis and vas deferens; agenesis of the seminal vesicles and coagulating glands; agenesis of bulbourethral glands; agenesis of gubernacular cords; cleft phallus and hypospadias; delayed preputial separation; reduced testicular testosterone levels; increased serum testosterone levels; reduced serum inhibin B levels; increased plasma LH levels; reduced daily sperm production; reduced fecundity; reduced fertility
Chlorinated pesticides (DDE)	Reduced anogenital distance; hypospadias; reduced accessory sex organ weights; poorly organized testis; nipple retention; decreased plasma testosterone levels; delayed preputial separation; abnormally small penis
Pollutants (Dioxins)	Decreased testis weight; delayed testis descent; decreased daily sperm production; decreased sperm numbers; epididymal malformations; reduced accessory sex organ weights; reduced anogenital distance; dose-related tendencies to decrease plasma testosterone and dihydrotestosterone (DHT); altered sex behaviour
Dicarboximide Fungicides (Vinclozolin, Procymidone, Prochloraz)	Decreased testis weight; abnormal morphology of seminiferous tubules; increased number of apoptotic germ cells in testis; decreased sperm number and daily sperm production; reduced elongated spermatid content per testis; increased sperm head abnormalities; low ejaculated sperm count; spermatogenic granuloma; epididymal granulomas; chronic inflammation of epididymis; cryptorchidism; hypospadias; cleft phallus; reduced anogenital distance; nipple retention; reduced accessory sex organ weights; glandular atrophy and chronic inflammation of prostate; agenesis of prostate; reduced secretion and chronic inflammation of seminal vesicles; reduction of erections during the ex-copula penile reflex test; increase in seminal emissions during the ex-copula penile reflex tests; decreased testosterone levels; decreased fertility
Herbicides (Linuron)	Decreased testis weight; testicular and epididymal malformations; reduced spermatid number; decreased anogenital distance; nipple retention; epispadias; reduced accessory sex organ weights; delayed preputial separation
UV screens (4-MBC, OMC)	Reduced sperm count; decreased testosterone level; decreased weight of ventral prostate; delayed preputial separation

Based on Toppari J, Adamsson A, Boas M, Juul A, Main KM, Skakkebaek NE, Virtanen HE: Endocrine disrupters and child health. Possible developmental early effects of endocrine disrupters on child health. WHO 2012, 83 pp. ISBN 978 92 4 150376 1 and Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, Toppari J, Zoeller RT. EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocr Rev*. 2015 Dec;36(6):E1–E150. For references, see Gore *et al*.

Heat is another well-known disruptor of spermatogenesis. In undescended testes, spermatogonia undergo apoptosis over time, although body temperature is only few degrees higher than in the scrotum [5]. Extended bathing in a hot tub (>15 min at temperature >36°C) or working in very hot environment can affect spermatogenesis [6, 7], whereas results on sauna effects at temperatures as high as 80°C are equivocal [7, 8]. Variable results on the effects of tight underwear or sportswear have also been published [7, 9, 10]. Obesity and sitting work, such as truck driving have also been linked to increased scrotal temperature and impaired sperm production, however, many other factors may have contributed to these results [7].

Occupational exposures to specific toxicants can be much larger than those in the general population. The classic example of male infertility associated with exposure to 1,2.dibromo-3-chloropropane (DBCP) received much attention in the 1970s. Toxicity of DBCP was found only after an alert by the wives of men working in the chemical factory producing this nematocide. The young couples noticed their common fertility problem and started to suspect that it might be related to occupational exposures. Indeed, subsequent investigations demonstrated a positive association between the extent of exposure and the degree of suppression of sperm production. While in many of the men, sperm production recovered after removal from exposure, some remained permanently infertile [11, 12]. Interestingly, animal experiments using dibromochloropropane (DBPC) clearly showed reproductive toxicity much earlier, but this was ignored [13]. This is an example where ordinary citizens spotted the problem, while regulatory authorities and occupational health professionals initially failed to notify and prevent it.

In addition to DBCP, other chlorinated pesticides such as chlordecone, carbaryl, and ethylene dibromide have similarly been associated with reproductive toxicity [14].

Developmental Effects

Genetic disorders of male reproductive health have taught us much about the mechanisms of development and function of reproductive organs, thereby guiding the search for environmental

influences [15]. Hormones regulate reproductive development and male-specific differentiation depends on normal androgen action. Androgen insensitivity due to inactivating androgen receptor mutations and defects in androgen biosynthesis lead to complete sex reversal or varying degrees of undermasculinization. The lack of normal androgen action causes maldescent of the testes, hypospadias, poor sperm production, and a marked increase in the risk of testicular germ cell cancer [16]. Fetal exposure to antiandrogenic chemicals can cause similar effects. There is substantial epidemiological evidence suggesting that all the above-mentioned male reproductive disorders may share similar developmental risk factors leading Skakkebaek and coworkers to coin the term testicular dysgenesis syndrome (TDS) to emphasize the possibility of a common aetiological or pathogenic pathway (see **Figure 10.10.1.1**) [16–19]. In the case of reproductive disorders that manifest during or beyond young adulthood (e.g. testicular cancer or defects in sperm production) potential developmental origins of these conditions are difficult to ascertain given that fetal development occurred more than 20 years earlier. Undescended testes (cryptorchidism) and hypospadias can be observed immediately after birth, which allows an earlier opportunity to study the preceding fetal exposures without a 20- or 30-year time gap. Cryptorchidism is a very frequent birth defect (1–8% of full-term newborns), whereas hypospadias occurs in less than one percent of boys, and therefore most of the studies on environmental exposures and reproductive birth defects have focused on cryptorchidism.

Cryptorchidism

Occupational and epidemiological studies have shown an association between pesticide exposure *in utero* and risk of cryptorchidism [20–22], but other studies did not confirm these associations (reviewed in Virtanen *et al.* [23]). Exposure measurements are expensive and few large-scale studies can meaningfully analyse direct exposure-outcome relationships. However, there are studies showing an association between the levels of organochlorine pesticides, polybrominated diphenyl ethers (PBDEs; used as flame retardants), polychlorinated biphenyls (PCBs; used as flame retardants and insulators), and dioxins (pollutant from e.g. incineration and traffic) in either breast milk or child's adipose tissue and the risk of cryptorchidism [24–28]. In studies where pesticide levels were measured in maternal serum or in cord serum, no significant association with cryptorchidism was found [25, 29, 30, 31]. Levels of PCBs in maternal or cord serum, placenta, umbilical cord, or boy's adipose tissue were not associated with the risk of cryptorchidism in most studies [33]. When breast milk levels of 106 environmental chemicals were combined in an analysis, PBDEs and dioxins were identified as risk factors for cryptorchidism [33]. Levels of perfluorinated compounds (used in surface materials, such as water-proof shoes and clothes) in amniotic fluid or cord blood were not associated with the risk of cryptorchidism [34, 35]. All the above-mentioned chemicals are persistent and measurable even long after exposure. Fetal exposure to some of these chemicals, such as dioxins can be extrapolated also from samples taken years after birth when the duration of breast feeding is taken into account, whereas exposures that took place largely after birth and non-persistent chemicals need to be measured at the right time [26]. Exposure to persistent chemicals has been declining after most of them have been banned and governments have committed to phase out these harmful chemicals according to the Stockholm convention (https://www.pops.int).

There are fewer studies on the association of exposure to non-persistent chemicals and the risk of cryptorchidism than those of persistent compounds. Antiandrogenic phthalates (used as plasticizers) have been high on the list of suspected contributors to cryptorchidism. Risk of cryptorchidism was positively associated with phthalate levels in maternal urine during pregnancy [36], but no positive association was found when phthalates were measured in amniotic fluid, cord serum, or breast milk [25, 37, 38]. Serum levels of bisphenol A (used widely in plastics and tin coatings) were associated with the risk of cryptorchidism in one study [39], while no

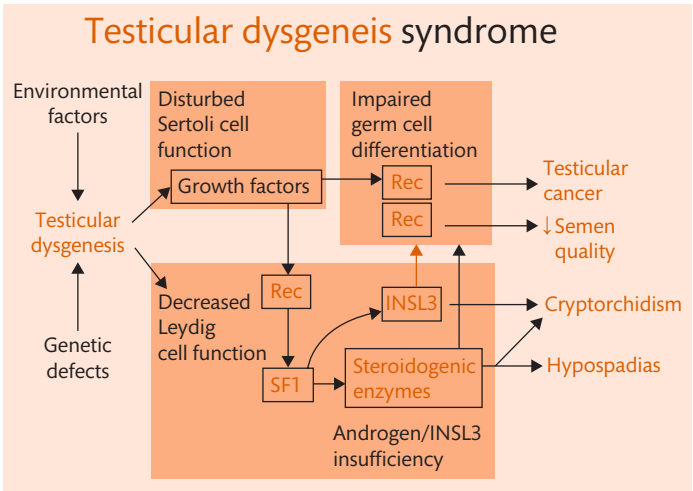


Figure 10.10.1.1 Genetic and environmental factors can disrupt the function of fetal Sertoli cells and Leydig cells that communicate with each other by growth factors. Steroidogenic factor-1 is an important transcription factor in the Leydig cells controlling steroidogenesis and production of insulin-like peptide-3 (INSL3). Insufficient production of androgens and INSL3 affect developing germ cells and external genitalia causing testicular germ cell cancer, poor semen quality, cryptorchidism, or hypospadias.

association was found when bisphenol A (BPA) levels were measured in cord blood [40, 41].

A problem in many of the studies is their small sample size that limits their statistical power, which may give false negative results. Furthermore, there can be a publication bias that reduces the number of publications that show only negative results. Meta-analyses are difficult in this area, because the matrices in which the exposures have been measured are variable (breast milk, maternal serum and urine, amniotic fluid, placenta, adipose tissue, serum or urine of the child, etc.), number of chemicals is large, and outcome measures are poorly standardized.

Hypospadias

Hypospadias can be associated with cryptorchidism and a shortened anogenital distance, but usually present as an isolated birth defect [42, 43]. Anogenital distance has been used as a post-natal read-out of androgen effects *in utero* [44]. Epidemiological studies on hypospadias have been more challenging because the prevalence of hypospadias is low. Most studies have assessed persistent chemicals. In ecological and occupational studies, exposure to pesticides has been associated with the risk of hypospadias. According to a meta-analysis, the risk ratio was 1.36 (95%CI 1.04–1.77) after maternal exposure to pesticides and 1.19 (95%CI 1.00–1.41) after paternal exposure [45]. A positive association between fetal exposure to pesticides and hypospadias was also reported in two recent studies [20, 46], whereas no association was observed in others [47–49].

Case–control studies have reported either positive or no association between hypospadias and pesticides or the sum of PCBs in maternal serum or in the boy's serum (reviewed in Virtanen *et al.* [32] and Vand Der Zanden *et al.* [50]). A positive association of placental concentrations of pesticides, BPA, and benzophenone (UV screen) with cryptorchidism and hypospadias together was reported [51, 52]. Meta-analyses combining data from different exposures measured in different matrices did not reveal any positive associations with either hypospadias or cryptorchidism [53].

Testicular Cancer

Testicular cancer usually refers to testicular germ cell tumours (TGCTs) since more than 90% of these cancers have a germ cell origin [54]. They appear after puberty and the incidence peaks within the reproductive age. Thus, TGCTs are the most common cancers of young men in Western countries, and the incidence has increased rapidly in all parts of world [55, 56]. Testicular cancer is preceded by germ cell neoplasia in situ (GCNIS) that develops either into seminoma or non-seminoma. GCNIS cells are very similar to fetal gonocytes, and it is likely that the TGCTs have a fetal origin [54]. Genetic studies have identified several genes, such as *KIT* and *CDC27*, whose polymorphisms are associated with an increased risk of testicular cancer [57]. Cryptorchidism, hypospadias, and other genital malformations are risk factors for testicular cancer [58].

Many studies on environmental effects on testicular cancer have focused on contemporary exposures, which have not produced much evidence on increased risks, perhaps because the relevant exposure time during fetal and early childhood development have not been documented. However, some case–control studies have reported an association between testicular cancer and exposure to organochlorine pesticides, while there are conflicting results regarding the association with exposure to PCBs [59–63].

Maternal samples may reflect exposures during pregnancy better than samples from adult offsprings. Measurement of DDT enantiomers ratios gives information on the timing of exposure because DDT conformation changes over time, thereby revealing when DDT has entered body [64, 65]. This approach helped to identify higher DDT exposures in mothers of cryptorchid boys than in controls in a U.S. pregnancy cohort when chemical analysis were made in postpartum serum samples [66].

Measurements have been made in mothers' serum also in samples collected years after pregnancy, and extrapolations of fetal exposures to persistent chemicals in such samples are possible, although exposures after pregnancy confound the findings. Swedish studies of men with testicular cancer compared men's serum chemical levels and then maternal serum chemicals in a case–control setting [67–69]. Mothers of men with testicular cancer had significantly higher serum levels of PCBs, hexachloro benzene, PBDEs, and cis- and transnonachlordanes than controls, whereas only cis-nonachlordane concentration was higher in serum of cases themselves as compared to controls.

Semen Quality

Semen quality encompasses many aspects of sperm counts and seminal plasma constituents. Sperm concentration and total sperm counts can be measured very accurately (see Chapter 10.2.3, 'Semen analysis'). Assessment of sperm morphology and motility is less accurate and not well standardized, thus hampering comparisons between studies. Several meta-analyses have indicated decline of sperm concentrations in developed countries since 1940s [70–73]. Although, there are regional differences both in timing and degree of this decline, the current level of sperm concentration of young men in different areas of the world is surprisingly similar [74]. However, both the existence of a secular decline in sperm counts and its possible association with environmental effects have been disputed and the debate continues [73, 75–77].

Sperm production depends on the number of Sertoli cells in the testis, because each Sertoli cell can host and support only a finite number of germ cells. Sertoli cells cease proliferating in puberty, and therefore the sperm production capacity is determined before puberty. Sertoli cells actively multiply first in fetal development, then after birth during the so-called minipuberty and finally during late prepuberty [78]. Toxicants, heat and irradiation (and illnesses, etc.) can reduce sperm production (see earlier), but otherwise, sperm production rate in an individual remains remarkably stable during adult life. Since fetal and childhood testicular developments are decisive windows for determining lifelong sperm production capacity, it is possible that mechanisms operating during these early critical periods could account for the putative decline in adult sperm counts at the population level.

Accidental exposure to high levels of persistent pollutants such as PCBs and 2,3,7,8,-tetrachlorodibenzo-p-dioxin (TCDD) as fetus or infant has been associated with reduced semen quality in adulthood [79–81]. Serum concentrations of TCDD and polychlorinated dibenzo-p-dioxins (PCDDs) toxic equivalents (TEQs) at the age of 8 to 9 years were inversely associated with sperm concentration, total sperm count and total motile sperm count at the age of 18 to 19 years in a prospective Russian study, although the TCDD exposure was much lower than in the Seveso study by Mocarelli [82]. These studies suggest that dioxins have adverse developmental effects on the testis.

Perfluororinated compounds, prefluorooctanoic acid (PFOA), and perfluorooctanesulfonate (PFOS) have been of concern because of their toxicity and endocrine-disrupting properties [2]. Prenatal exposure to background levels of PFOA was inversely associated with semen quality (sperm concentration, total sperm count), whereas no association with PFOS, PCBs, or p,p'-DDE was observed [83, 84]. There is also some evidence that contemporary (i.e. adult), exposure to phthalates is associated adversely with semen quality [85–87]. Results on bisphenol A remain inconclusive [88].

Hormone Levels

Reproductive hormone levels can reflect toxic effects on the testis, pituitary, or hypothalamus. The levels need to be considered in relation with each other and with other findings. Increased follicle-stimulating hormone (FSH) levels and decreased inhibin B levels in adult men who had been exposed to high levels of TCDD perinatally in the Seveso accident pointed to adverse testicular effects of dioxins [89]. Prenatal exposure to background levels of PFOA was associated positively with adult FSH and luteinizing hormone (LH) levels, whereas no significant associations with hormone levels were found for PFOS, PCBs, and p,p'-DDE [83, 90]. Prenatal exposure to phthalates has also been associated with increased serum concentration of FSH and LH in adult men [91–93]. Phthalate and PBDE concentrations in breast milk were positively associated with LH and LH-free testosterone ratio in the male offspring at minipuberty (3 months of age) [27, 38].

Mixture Effects

Exposures to environmental chemicals (unless accidental) seldom occur alone. We are constantly exposed to a complex mixture of chemicals that can act additively, synergistically, and sometimes in opposite directions. Male reproductive disorders can be seen when several antiandrogenic compounds act together at concentrations below their individual no-adverse-effect levels [94, 95]. Dioxins add to the overall exposure load, although their mechanism of action is clearly different from that of antiandrogens [96]. Clear evidence on adverse effects of endocrine disruptors on human male reproductive health is still limited. However, the ample data from the animal experiments (Table 10.10.1.1), well-known mechanisms of action (aryl hydrocarbon receptor agonism, androgen receptor antagonism, and inhibition of androgen synthesis), and emerging evidence on mixture effects give us a reasonable basis for risk assessment in chemical regulation. However, current safety margins may not guarantee health and safety, because the net effects of combined exposures of multiple chemicals can be harmful, even though individual chemicals in the mixture are present at lower concentrations than their no-observed-effect-levels that are used for determination of accepted daily doses of exposure [97]. Insightful regulation of chemicals is believed to be able to produce considerable healthcare savings in future reproductive healthcare [98].

Conclusions

Experimental evidence from animal studies has clearly shown reproductive toxicity of many endocrine disruptors (Table 10.10.1.1), but the risk for human reproduction is still controversial unproven out. Many endocrine disruptors have been banned, and therefore our exposure to several harmful compounds is decreasing. It takes, however, several generations before we can be rid of the most persistent organic pollutants, such as DDE, because they are transferred from mother to offspring via breast-feeding [99]. Multiple new chemicals are being introduced all the time, and it is important that appropriate resources and strategies are mandated to their safety testing in order to avoid adverse reproductive outcomes in humans.

REFERENCES

1. Toppari J, Larsen JC, Christiansen P, *et al.* Male reproductive health and environmental xenoestrogens. *Environ Health Perspect*, 1996; **104** (Suppl 4): 741–803.
2. WHO. State of the science of endocrine disrupting chemicals 2012, United Nations Environment Programme and the World Health Organization, 2013. Available at: https://www.who.int/ceh/publications/endocrine/en/ (accessed: 19 March 2020).
3. Gore AC, Chappell VA, Fenton SE, *et al.* EDC-2: the Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocr Rev*, 2015; **36**: E1.
4. Rowley MJ, Leach DR, Warner GA, Heller CG. Effect of graded doses of ionizing radiation on the human testis. *Radiat Res*, 1974; **59**: 665–78.
5. Kollin C, Stukenborg JB, Nurmio M, *et al.* Boys with undescended testes: endocrine, volumetric and morphometric studies on testicular function before and after orchidopexy at nine months or three years of age. *J Clin Endocrinol Metab*, 2012; **97**: 4588–95.
6. Procopé BJ. Effect of repeated increase of body temperature on human sperm cells. *Int J Fertility*, 1965; **10**: 333–9.
7. Jung A, Schuppe HC. Influence of genital heat stress on semen quality in humans. *Andrologia*, 2007; **39**: 203–15.
8. Garolla A, Torino M, Miola P, *et al.* Twenty-four-hour monitoring of scrotal temperature in obese men and men with a varicocele as a mirror of spermatogenic function. *Hum Reprod*, 2015; **30**: 1006–13.
9. Wang C, McDonald V, Leung A, *et al.* Effect of increased scrotal temperature on sperm production in normal men. *Fertil Steril*, 1997; **68**: 334–9.
10. Mínguez-Alarcón L, Gaskins AJ, Chiu YH, *et al.* Type of underwear worn and markers of testicular function among men attending a fertility center. *Hum Reprod*, 2018; **33**: 1749–56.

11. Whorton D, Krauss RM, Marshall S, Milby TH. Infertility in male pesticide workers. *Lancet*, 1977; **2**: 1259.
12. Potashnik G, Ben-Aderet N, Israeli R, Yanai-Inbar I, Sober I. Suppressive effect of 1,2-dibromo-3-chloropropane on human spermatogenesis. *Fertil Steril*, 1978; **30**: 444–7.
13. Torkelson TR, Sadek SE, Rowe VK, *et al.* Toxicologic investigations of 1,2-dibromo-3-chloropropane. *Toxicol Appl Pharmacol*, 1961; **3**: 545.
14. Giwercman A, Bonde JP. Declining male fertility and environmental factors. *Endocrinol Metab Clin N Am*, 1998; **27**: 807–30.
15. Toppari J, Virtanen HE, Main KM, Skakebaek NE. Cryptorchidism and hypospadias as a sign of testicular dysgenesis syndrome (TDS): environmental connection. *Birth Defects Res A Clin Mol Teratol*, 2010; **88**: 910–19.
16. Skakkebaek NE, Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod*, 2001; **16**: 972–8.
17. Skakkebaek NE, Rajpert-De Meyts E, Buck Louis GM, *et al.* Male reproductive disorders and fertility trends: influences of environment and genetic susceptibility. *Physiol Rev*, 2016; **96**: 55–97.
18. Dalgaard MD, Weinhold N, Edsgärd D, *et al.* A genome-wide association study of men with symptoms of testicular dysgenesis syndrome and its network biology interpretation. *J Med Genet*, 2012; **49**: 58–65.
19. Juul A, Almstrup K, Andersson AM, *et al.* Possible fetal determinants of male infertility. *Nat Rev Endocrinol*, 2014; **10**: 553–62.
20. Garcia J, Ventura M, Requena M, Hernandez AF, Parron T, Alarcon R. Association of reproductive disorders and male congenital anomalies with environmental exposure to endocrine active pesticides. *Reprod Toxicol*, 2017; **71**: 95.
21. Weidner IS, Moller H, Jensen TK, Skakkebaek NE. Cryptorchidism and hypospadias in sons of gardeners and farmers. *Environ Health Perspect*, 1998; **106**: 793–6.
22. Andersen, HR, Schmidt, IM, Grandjean, P, *et al.* Impaired reproductive development in sons of women occupationally exposed to pesticides during pregnancy. *Environ Health Perspect*, 2008; **116**: 566–72.
23. Virtanen HE, Toppari J. Epidemiology and pathogenesis of cryptorchidism. *Hum Reprod Update*, 2008; **14**: 49–58.
24. Damgaard IN, Skakkebaek NE, Toppari J, *et al.* Persistent pesticides in human breast milk and cryptorchidism. *Environ Health Perspect*, 2006; **114**: 1133–8.
25. Brucker-Davis, F, Wagner-Mahler, K, Delattre, I, *et al.* Cryptorchidism at birth in Nice area (France) is associated with higher prenatal exposure to PCBs and DDE, as assessed by colostrum concentrations. *Hum Reprod*, 2008; **23**: 1708–18.
26. Koskenniemi JJ, Virtanen HE, Kiviranta H, *et al.* Association between levels of persistent organic pollutants in adipose tissue and cryptorchidism in early childhood: a case-control study. *Environ Health*, 2015; **14**: 78.
27. Main KM, Kiviranta H, Virtanen HE, *et al.* Flame retardants in placenta and breast milk and cryptorchidism in newborn boys. *Environ Health Perspect*, 2007; **115**: 1519–26.
28. Buck Louis GM, Sundaram R, Schisterman EF, *et al.* Persistent environmental pollutants and couple fecundity: the LIFE study. *Environ Health Perspect*, 2013; **121**: 231–6.
29. Longnecker MP, Klebanoff MA, Brock JW, *et al.* Maternal serum level of 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene and risk of cryptorchidism, hypospadias, and polythelia among male offspring. *Am J Epidemiol*, 2002; **155**: 313–22.
30. Trabert B, Longnecker MP, Brock JW, Klebanoff MA, McGlynn KA. Maternal pregnancy levels of trans-nonachlor and oxychlordane and prevalence of cryptorchidism and hypospadias in boys. *Environ Health Perspect*, 2012; **120**: 478–82.
31. Pierik FH, Klebanoff MA, Brock JW, Longnecker MP. Maternal pregnancy serum level of heptachlor epoxide, hexachlorobenzene, and beta-hexachlorocyclohexane and risk of cryptorchidism in offspring. *Environ Res*, 2007; **105**: 364.
32. Virtanen HE, Main KM, Toppari J. Association of endocrine disrupting chemicals with male reproductive health. In: Huhtaniemi IT, ed. *Encyclopedia of Endocrine Diseases*. Academic Press/Elsevier, 2019: 802–11.
33. Krysiak-Baltyn K, Toppari J, Skakkebaek NE, *et al.* Association between chemical pattern in breast milk and congenital cryptorchidism: modelling of complex human exposures. *Int J Androl*, 2012; **35**: 294–302.
34. Vesterholm Jensen D, Christensen J, Virtanen HE, *et al.* No association between exposure to perfluorinated compounds and congenital cryptorchidism: a nested case-control study among 215 boys from Denmark and Finland. *Reproduction*, 2014; **147**: 411–17.
35. Toft, G, Jönsson, BA, Bonde, JP, *et al.* Perfluorooctane sulfonate concentrations in amniotic fluid, biomarkers of fetal Leydig cell function, and cryptorchidism and hypospadias in Danish boys (1980–1996). *Environ Health Perspect*, 2016; **124**: 151.
36. Swan SH. Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. *Environ Res*, 2008; **108**: 177.
37. Jensen MS, Anand-Ivell R, Nørgaard-Pedersen B, *et al.* Amniotic fluid phthalate levels and male fetal gonad function. *Epidemiology*, 2015; **26**: 91–9.
38. Main KM, Mortensen GK, Kaleva MM, *et al.* Human breast milk contamination with phthalates and alterations of endogenous reproductive hormones in infants three months of age. *Environ Health Perspect*, 2006; **114**: 270–6.
39. Komrowska MD, Hermanowicz A, Czyzewska U, *et al.* Serum bisphenol A level in boys with cryptorchidism: a step to male infertility? *Int J Endocrinol*, 2015; 973154.
40. Fenichel P, Dechaux H, Harthe C, *et al.* Unconjugated bisphenol A cord blood levels in boys with descended or undescended testes. *Hum Reprod*, 2012; **27**: 983–90.
41. Chevalier N, Brucker-Davis F, Lahlou N, *et al.* A negative correlation between insulin-like peptide 3 and bisphenol A in human cord blood suggests an effect of endocrine disruptors on testicular descent during fetal development. *Hum Reprod*, 2015; **30**: 447–53.
42. Thankamony A, Lek N, Carroll D, *et al.* Anogenital distance and penile length in infants with hypospadias or cryptorchidism: comparison with normative data. *Environ Health Perspect*, 2014; **122**: 207–11.
43. Van Der Horst HJ, De Wall LL. Hypospadias, all there is to know. *Eur J Pediatr*, 2017; **176**: 435–41.
44. Dean A, Sharpe RM. Clinical review: anogenital distance or digit length ratio as measures of fetal androgen exposure: relationship to male reproductive development and its disorders. *J Clin Endocrinol Metab*, 2013; **98**: 2230–8.
45. Rocheleau CM, Romitti PA, Sanderson WT, *et al.* Maternal occupational pesticide exposure and risk of hypospadias in the

National Birth Defects Prevention Study. *Birth Defects Res A Clin Mol Teratol*, 2011; **91**: 927–36.
46. Kalfa, N, Paris, F, Philibert, P, *et al*. Is hypospadias associated with prenatal exposure to endocrine disruptors? A French collaborative controlled study of a cohort of 300 consecutive children without genetic defect. *Eur Urol*, 2015; **68**: 1023–90.
47. Nassar N, Abeywardana P, Barker A, Bower C. Parental occupational exposure to potential endocrine disrupting chemicals and risk of hypospadias in infants. *Occup Environ Med*, 2010; **67**: 585–9.
48. Rocheleau CM, Romitti PA, Dennis LK. Pesticides and hypospadias: a meta-analysis. *J Pediatr Urol*, 2009; **5**: 17–24.
49. Carmichael SL, Yang W, Roberts EM, *et al*. Hypospadias and residential proximity to pesticide applications. *Pediatrics*, 2013; **132**: e1216–26.
50. Vand Der Zanden LF, Van Rooij IA, Feitz WF, *et al*. Aetiology of hypospadias: a systematic review of genes and environment. *Hum Reprod Update*, 2012; **18**: 260–83.
51. Fernandez, MF, Olmos, B, Granada, A, *et al*. Human exposure to endocrine-disrupting chemicals and prenatal risk factors for cryptorchidism and hypospadias: a nested case-control study. *Environ Health Perspect*, 2007; **115** Suppl 1: 8–14.
52. Fernandez, MF, Arrebola, JP, Jimenez-Diaz, I, *et al*. Bisphenol A and other phenols in human placenta from children with cryptorchidism or hypospadias. *Reprod Toxicol*, 2016; **59**: 89–95.
53. Bonde JP, Flachs EM, Rimborg S, *et al*. The epidemiologic evidence linking prenatal and postnatal exposure to endocrine disrupting chemicals with male reproductive disorders: a systematic review and meta-analysis. *Hum Reprod Update*, 2016; **23**: 104–25.
54. Rajpert-De Meyts E, McGlynn KA, Okamoto K, Jewett MA, Bokemeyer C. Testicular germ cell tumours. *Lancet*, 2016; **387**: 1762–74.
55. Znaor A, Lortet-Tieulent J, Jemal A, Bray F. International variations and trends in testicular cancer incidence and mortality. *Eur Urol*, 2014; **65**: 1095–106.
56. Trabert B, Chen J, Devesa S, Bray F, McGlynn KA. International patterns and trends in testicular cancer incidence, overall and by histologic subtype, 1973–2007. *Andrology*, 2015; **3**: 4–12.
57. Litchfield K, Summersgill B, Yost S, *et al*. Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. *Nat Commun*, 2015; **6**: 5973.
58. Trabert B, Zugna D, Richiardi L, McGlynn KA, Akre O. Congenital malformations and testicular germ cell tumors. *Int J Cancer*, 2013; **133**: 1900–4.
59. Giannadrea F, Gandini L, Paoli D, Turci R, Figa-Talamnca I. Pesticide exposure and serum organochlorine residuals among testicular cancer patients and healthy controls. *J Environ Sci Health B*, 2011; **46**: 780–7.
60. McGlynn KA, Quraishi SM, Graubard BI, Weber JP, Rupertone MV, Erickson RI. Persistent organochlorine pesticides and risk of testicular germ cell tumors. *J Natl Cancer Instit*, 2008; **100**: 663–71.
61. McGlynn KA, Quraishi SM, Graubard BI, Weber JP, Rupertone MV, Erickson RI. Polychlorinated biphenyls and risk of testicular germ cell tumors. *Cancer Res*, 2009; **69**: 1901–9.
62. Purdue MP, Engel LS, Langseth H, *et al*. Prediagnostic serum concentrations of organochlorine compounds and risk of testicular germ cell tumors. *Environ Health Perspect*, 2009; **117**: 1514–19.
63. Toppari, J, Adamsson, A, Boas, M, *et al*. Endocrine disrupters and child health. Possible developmental early effects of endocrine disrupters on child health. WHO 2012: 83 Available at: https://www.who.int/iris/bitstream/10665/78101/1/9789241505031_eng.pdf (accessed:19 March 2020).
64. Shen H, Virtanen HE, Main KM, *et al*. Enantiomeric ratios as an indicator of exposure processes for persistent pollutants in human placentas. *Chemosphere*, 2006; **62**: 390.
65. De Jager C, Farias P, Barraza-Villareal A, *et al*. Reduced seminal parameters associated with environmental DDT exposure and p,p-DDE concentrations in men in Chiapas, Mexico: a cross-sectional study. *J Androl*, 2006; **27**: 16–27.
66. Cohn BA, Cirillo PM, Christianson RE. Prenatal DDT exposure and testicular cancer: a nested case-control study. *Arch Environ Occup Health*, 2010; **65**: 127–34.
67. Hardell L, Bavel B, Lindströms G, Eriksson M, Carlberg M. In utero exposure to persistent organic pollutants in relation to testicular cancer risk. *Int J Androl*, 2006; **29**: 228–34.
68. Hardell L, Van Bavel B, Lindström G, *et al*. Increased concentrations of polychlorinated biphenyls, hexachlorobenzene, and chlordanes in mothers of men with testicular cancer. *Environ Health Perspect*, 2003; **111**: 930–4.
69. Hardell, L, Van Bavel, B, Lindström, G, *et al*. Concentrations of polychlorinated biphenyls in blood and the risk for testicular cancer. *Int J Androl*, 2004; **27**: 282–90.
70. Carlsen E, Giwercman A, Keiding N, Skakkebaek NE. Evidence for decreasing quality of semen during past 50 years. *BMJ*, 1992; **305**: 609–13.
71. Swan SH, Elkin EP, Fenster L. Have sperm densities declined? A reanalysis of global trend data. *Environ Health Perspect*, 1997; **105**: 1228–32.
72. Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934–1996. *Environ Health Perspect*, 2000; **108**: 961–6.
73. Levine H, Jørgensen N, Martino-Andrade A, *et al*. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update*, 2017; **23**: 646–59.
74. Virtanen HE, Jørgensen N, Toppari J. Semen quality in the 21(st) century. *Nat Rev Urol*, 2017; **14**: 120–30.
75. Fisch H. Declining worldwide sperm counts: disproving a myth. *Urol Clin N Am*, 2008; **35**: 137–46.
76. Ravanos K, Petousis S, Margioula-Siarkou C, *et al*. Declining sperm counts… or rather not? A mini review. *Obstet Gynecol Surv*, 2018; **73**: 595–605.
77. Toft G, Rignell-Hydbom A, Tyrkiel E, *et al*. Semen quality and exposure to persistent organochlorine pollutants. *Epidemiology*, 2006; **17**: 450–8.
78. Mäkelä JA, Koskenniemi JJ, Virtanen HE, Toppari J. Testis development. *Endocr Rev*, 2019; **40**: 857–905.
79. Guo YL, Hsu PC, Hsu CC, Lambert GH. Semen quality after prenatal exposure to polychlorinated biphenyls and dibenzofurans. *Lancet*, 2000; **356**: 1240–1.
80. Mocarelli P, Geerthoux PM, Needham LL, *et al*. Perinatal exposure to low doses of dioxin can permanently impair human semen quality. *Environ Health Perspect*, 2011; **119**: 713–18.
81. Mocarelli P, Geerthoux PM, Patterson DG, Jr., *et al*. Dioxin exposure, from infancy through puberty, produces endocrine disruption and affects human semen quality. *Environ Health Perspect*, 2008; **116**: 70.
82. Minguez-Alarco L, Sergeyev O, Burns JS. A longitudinal study of peripubertal serum organochlorine concentrations and semen parameters in young men: the Russian Children's Study. *Environ Health Perspect*, 2017; **125**: 460.
83. Vested A, Ramlau-Hansen CH, Olsen SF, *et al*. Associations of in utero exposure to perfluorinated alkyl acids with human semen

quality and reproductive hormones in adult men. *Environ Health Perspect*, 2013; **121**: 453.
84. Vested A, Ramlau-Hansen CH, Olsen SF, *et al*. In utero exposure to persistent organochlorine pollutants and reproductive health in the human male. *Reproduction*, 2014; **148**: 635–46.
85. Cai H, Zheng W, Zheng P, *et al*. Human urinary/seminal phthalates or their metabolite levels and semen quality: a meta-analysis. *Environ Res*, 2015; **142**: 486–94.
86. Axelsson J, Rylander L, Rignell-Hydbom A, Jönsson BA, Lindh CH, Giwercman A. Phthalate exposure and reproductive parameters in young men from the general Swedish population. *Environ Int*, 2015; **85**: 54–60.
87. Thurston SW, Mendiola J, Bellamy AR, *et al*. Phthalate exposure and semen quality in fertile US men. *Andrology*, 2015; **4**: 623–8.
88. Minguez-Alarco L, Hauser R, Gaskins AJ. Effects of bisphenol A on male and couple reproductive health: a review. *Fertil Steril*, 2016; **106**: 864–70.
89. Mocarelli P, Geerthoux PM, Needham LL, *et al*. Perinatal exposure to low doses of dioxin can permanently impair human semen quality. *Environ Health Perspect*, 2011; **119**: 713–18.
90. Vested A, Giwercman A, Bonde JP, Toft G. Persistent organic pollutants and male reproductive health. *Asian J Androl*, 2014; **16**: 71–80.
91. Longnecker MP, Gladen BC, Bucul-Uicap LA, *et al*. In utero exposure to the antiandrogen 1,1-dichloro-2,2-bis(p-chlorophenyl) ethylene (DDE) in relation to anogenital distance in male newborns from Chiapas, México. *Am J Epidemiol*, 2007; **165**: 1015–22.
92. Axelsson J, Rylander L, Rignell-Hydbom A, Lindh CH, Jönsson BA, Giwercman A. Prenatal phthalate exposure and reproductive function in young men. *Environ Res*, 2015; **138**: 264.
93. Haugen TB, Tefre T, Malm G, *et al*. Differences in serum levels of CB-153 and p,p'-DDE, and reproductive parameters between men living south and north in Norway. *Reprod Toxicol*, 2011; **32**: 261–7.
94. Rider CV, Furr JR, Wilson VS, Gray LE. A mixture of seven antiandrogens induces reproductive malformations in rats. *Int J Androl*, 2008; **31**: 249–82.
95. Christiansen S, Scholze M, Axelstad M, Boberg J, Kortenkamp A, Hass U. Combined exposure to anti-androgens causes markedly increased frequencies of hypospadias in the rat. *Int J Androl*, 2008; **31**: 241–8.
96. Rider CV, Furr J, Wilson VS, Gray LE, Jr. Cumulative effects of in utero administration of mixtures of reproductive toxicants that disrupt common target tissues via diverse mechanisms of toxicity. *Int J Androl*, 2010; **33**: 443–62.
97. Kortenkamp A, Faust M. Regulate to reduce chemical mixture risk. *Science*, 2018; **361**: 224–6.
98. Hauser R, Skakkebaek NE, Hass U, *et al*. Male reproductive disorders, diseases, and costs of exposure to endocrine-disrupting chemicals in the European Union. *J Clin Endocrinol Metab*, 2015; **100**: 1267–77.
99. Stockholm Convention on persistent organic pollutants (POPs). UNEP, 2009. Available at: https://www.pops.int (accessed: 19 March 2020).

SECTION 11

Management of the Transgender Patient

11.1

Introduction to Transgender and Gender Diverse People

Jon Arcelus and Walter Pierre Bouman

Introduction

Gender identity is an essential part of any person, and expressing it is a basic human need and the right of each individual, regardless of age, gender, ethnicity, religion, disability and sexual orientation [1]. The liberalization of society, particularly in the Western developed world, has led to more positive and less restrictive attitudes and views in the population regarding the expression of one's gender identity [2–4]. The latter is reflected in a significant growth in literature concerning transgender health, especially in the last few decades, as well as in the growing interest shown in this topic by the media [5–9]. Clinicians, researchers, educators, journalists, programme makers as well as global organizations, such as the World Professional Association of Transgender Health (WPATH; wpath.org), the European Professional Association of Transgender Health (EPATH; epath.eu), Global Action for Trans* Equality (GATE; transactivists.org), and Transgender Europe (TGEU; tgeu.org) continue to highlight the plight of transgender people and the need for a world wherein people of all gender identities and gender expressions have access to evidence-based healthcare, social services, justice, and equality. In many parts of the world, transgender people remain effectively marginalized and clearly discriminated against, with limited or no access to adequate (transgender) health services [10–12].

This chapter provides an introduction to transgender health. It starts with a brief historical and cultural overview concerning being transgender, followed by a description of current language and terminology in transgender health, the classification, and prevalence of identifying as transgender. It further describes transgender health services, including assessment and gender-affirming medical treatments. The chapter concludes with a synopsis of mental health problems in the transgender population. Other routes into the literature are provided by textbooks [13, 14] and many of the books, chapters, and articles cited elsewhere in this chapter.

Historical Background

Descriptions of the adoption of varying gender roles have been reported throughout history and across cultures. As early as the fifth century BC, Herodotus described the disease of the Scythians (morbus feminarium), which was thought to be divine retribution for their pillaging of the temple, in line with his view of diseases as being from divine causes [15]. Hippocrates described transgender women using the term 'Anandrii', and believed the disease was due to excessive riding on horseback. At the end of the nineteenth century, Edvard Westermarck, a Finnish philosopher and sociologist, reported that in nearly every part of North America there have been, throughout history, men dressing themselves in female clothes and performing female societal roles [16]. The history of transgender men has perhaps been generally less visible, although there are many accounts of assigned females at birth living as men, working and marrying without attention, sometimes only found to be assigned females at birth at death and, at other times, suffering great adversity and even death upon discovery. During the American Civil War (1861–1865) people assigned female at birth are known to have worn what was traditionally men's clothing and fought as soldiers. Many may have worn men's clothes as this was their only means of fighting and participating in the war effort. Some of them may have been transgender men and continued to live as men throughout their lives. There are many other people, male and female, some famous and others less so, who are thought

to have identified as transgender throughout history (for further reading see [17–19]).

Equally, there are many more examples across the world and among different cultures of gender role change [20, 21]. Different terms are used when talking about people expressing themselves as 'other' from male assigned at birth and female assigned at birth. The people from the Pueblo Indians of New Mexico use the term 'Mujerados', the term 'Hijras' is used in India [22], 'rhawaja sira' in Pakistan [23], 'Fa'afafine' in Samoa [24], 'Kathooey' in Thailand [25], 'Travestis' in Brazil [22], 'Mahu' in Hawaii [25], and 'Mahoos' in Tahiti [26]. Most of these terms have been used for people who were assigned male at birth, based on their sexual characteristics, but who live their lives as women. Terms and descriptions for transgender men (people assigned female at birth), compared with their female transgender counterparts throughout cultures and in history, are comparatively rare. Individuals assigned female at birth are often underrepresented in medical and legal discourse [27–29].

In addition to the aforementioned terms for what we often may understand to be transgender women, Non-Western societies have also generally been more tolerant and inclusive of people identifying outside the binary concept of being male or female. There are many cultures where gender was not commonly binarized. Such cultures include the indigenous Chukchi people in Siberia [30], the Waria in Indonesia [31], First Nation tribes of the United States and Canada [32], and the Machi of Latin America [33], among others. They show the heterogeneity of gender identity and expression and demonstrate especially that the Western notion of a gender binary, which is deeply rooted in Christianity, is only one of many possible perspectives [34]. Most societies, however, have only two distinct classes of gender according to a binary system, male or female, that correspond with the biological sexes of male and female. When a baby is born, society allocates the child to one of two possible genders, on the basis of their external genitalia. The assigned gender is subsequently specified on one's birth certificate and passport and most countries require that a passport gender marker is either female or male. A few countries, such as Australia, Bangladesh, Denmark, Germany, India, Nepal, and New Zealand allow passports to have a non-binary gender marker, called X (unspecified), T (transgender or third gender), E (eunuch), or O (other), depending on the country [35]. Having a non-binary marker on one's passport may make it difficult to travel to a country whose passports do not give that option [36]. Only relatively recently there is a growing awareness and understanding that a sizeable group of people in terms of raw numbers in Western and North American societies identify as non-binary [37–41].

As our understanding of gender identity and gender expression evolves, it is important to have a clear understanding of the meaning of the terminology and use of appropriate language in the field of transgender health. Language and terminology are addressed in the next paragraph.

Language and Terminology

In order to be able to describe who transgender and cisgender people are, we need first to define the terms 'transgender' and 'cisgender'. The word 'transgender' is an adjective often used as an umbrella term to describe anyone whose gender identity, expression, or behaviour is different from the assigned gender at birth based on the sexual characteristics. The term 'trans' is often used as an abbreviation for transgender. An asterisk used in the term 'trans*' indicates the inclusion of all gender diverse groups and includes non-binary people who self-describe or self-identify in a number of ways (e.g. as pangender, agender, polygender, bigender, neutrois, and so on. For further information, see [40]). The term 'cisgender' refers to a person whose assigned gender at birth is in line with their gender identity, or, in other words, someone who is happy to remain the gender assigned at birth. The term 'cis' is also used as an abbreviation for cisgender. So, a transgender woman is a person who was assigned male at birth, based on their sexual characteristics, but who identifies as female. Equally, a transgender man is a person assigned female at birth, based on their sexual characteristics, but who identifies as male. A cisgender woman is a person who was assigned female at birth, based on their sexual characteristics, and who identifies as female. Cisgender men are people who were assigned male at birth, based on their sexual characteristics, and who identify as male. Also, many trans people consider themselves men or women and no longer identify as a trans person following gender-affirming medical interventions, which may include cross-sex hormone treatment and surgery, and which change the body irreversibly. This process is known as transition [42–45].

Terminology has changed significantly over the past decade in particular, and many new terms have been introduced, while older ones have been discarded and are no longer deemed politically correct [36, 46]. It is important, both for respect and for clarity, to use the correct language. The community often referred to as the transgender or trans* community is an extraordinarily diverse group of people. Defining and quantifying the transgender population is problematic as there are currently few measurable and/or standardized criteria (e.g. physical, social, political) regarding what might or should constitute a transgender person. Furthermore, there is a relative invisibility in which many transgender people exist in their daily lives [47]. Moreover, language and terminology are important when attempting to address and target a community for access to healthcare and education, health promotion, disease prevention and service development [48–50].

People, including our patients, expect gender pronouns to be used according to the gender that they identify with. This applies to anyone, whether they identify as cisgender, transgender, non-binary, or otherwise. People who identify as female generally prefer that the pronouns 'she' and 'her' and the adjective 'her' are used, while people who identify as male prefer the pronouns 'he' and 'him' and the adjective 'his'. So, for a trans man, this would equally be 'he' or 'him'; and for a trans woman, 'she' or 'her'. Preferred pronouns can change over time, as a person comes out as transgender and transitions or discovers more about what feels right. If you are unsure as to how to address someone, it is generally a good idea to ask which pronouns (he or she or they; him or her or them) someone prefers and to use these pronouns accordingly [51]. There are trans people who do not identify within the binary spectrum of male and female, and they tend to prefer the pronouns 'they' and 'them' and the adjective 'their'. It is always better to ask rather than getting it wrong and potentially jeopardizing the quality of the patient–doctor relationship [36, 52]. Definitions and descriptions of frequently used and relevant terms in transgender healthcare are provided in alphabetical order in the following glossary in **Box 11.1.1**.

Box 11.1.1 Glossary of transgender-related nomenclature

Affirmed gender: This term is used to describe transgender people post-transition. This term indicates that the gender identity is now in line with the social gender role and the physical characteristics of the person.

Differences of sex development (DSD): Refers to people born with a sexual anatomy and/or chromosome pattern that does not fit the typical sexual anatomy and/or chromosome pattern of male and female. In some cases, the appearance of the genitalia at birth may not be clear and the assigned gender and gender role (boy or girl) given may not be consistent with the gender identity of the person. The inconsistencies in the development of the sexual organs may be associated with atypical sex chromosomes, such as Klinefelter syndrome (XXY), or Jacob's syndrome (XYY). The term Intersex has also been used instead of DSD. There remains controversy as to which term is more appropriate, DSD or intersex. We would suggest asking what a person prefers rather than assuming which term is best.

Gender and sex: These terms are often used interchangeably, but the meaning of both terms is different. The term 'sex' refers to the biological differences between male and female. This includes biological indicators such as sex chromosomes, anatomy of the reproductive system and secondary sexual characteristics, such as body hair. The term 'Gender' is more complex to define. It is related to the internal awareness or personal identification of being male, female, or other gender. It is also closely related to the cultural, social, and psychological differences between being male or female, which are part of gender role.

Gender-affirming treatment: Transgender people who wish to transition permanently to their experienced gender in order to affirm their gender may undergo treatment. This treatment includes hormone treatment and gender-affirming surgery.

Gender-affirming surgery: This term is used to describe the surgical procedures required in order for a transgender person to change their body to reflect their gender identity. This may include chest reconstructive surgery, breast augmentation surgery, gender-affirming genital surgery, and other surgeries such as vocal cord surgery, facial surgery, and so on.

Gender-affirming genital surgery: This refers to the surgical procedures required in order for a transgender person to change their genitals to reflect their gender identity. In the past, the terms sex reassignment surgery (SRS) and genital reconstructive surgery have been used. The term gender surgery is also used. Colloquially, 'bottom' surgery can be used.

Gender dysphoria: This term is generally used *as a symptom* which denotes the personal experience of dissonance between one's gender and phenotype [148]. The distress, which often accompanies gender dysphoria, is frequently manifested as depression, anxiety, or self-harm, and may also be related to feelings of non-acceptance, discrimination, and victimization by people and society. As a consequence of gender dysphoria many people access transgender health services in order to be considered for gender-affirming treatment. The term gender dysphoria (GD) is also used **as a diagnosis** in the Diagnostic and Statistical Manual (DSM-5) of the American Psychiatric Association [55].

Gender expression and gender role: Refers to all of the external characteristics and behaviours that are socially defined as either masculine or feminine. A person can express their gender through their behaviour, mannerism, clothing, hairstyle, voice, or body characteristics. Some of those expressions are culturally bound, for example, outside of Western cultures, it is not uncommon for men to wear skirts and skirt-like garments; however, in Europe and North America, the wearing of a skirt is usually associated with being a woman. Other stereotypical gender expectations are that men should not cry and that women are gentle.

Gender identity: Refers to the sense of oneself as male or female [150]. Whether a person looks like a man or a woman may be easy to identify. However, only the individual knows the gender they identify with. Some people identify themselves with the sex they were assigned at birth, however, not everyone does. In those cases, the term transgender (as an adjective) or trans is used [90].

Gender identity disorder: An obsolete term, which is considered a poignant remnant of the historic pathologization of transgender people. Gender identity disorder has been removed from the DSM-IV-TR [54] and ICD-10 [53], and no longer features in current nosological classification systems of transgender people [55–59].

Gender recognition certificate: In the United Kingdom, transgender people who have undergone a permanent change of social gender status can obtain legal recognition in the form of a Gender Recognition Certificate (GRC), which allows them to change their birth certificate. Obtaining a GRC does not necessarily require hormone treatment or gender confirmation surgery [151].

Misgender/Misgendering: When somebody makes incorrect assumptions about your gender or refuses to accept your gender and uses language that makes this apparent, such as using pronouns or gendered language like 'sir' or 'madam'.

Passing: Being seen or read as a certain gender. The opposite meaning is conveyed by the terms 'to be read' or 'to be clocked', which means not passing.

Non-binary gender: This term is used to describe people who do not identify as being part of the binary gender (male or female). Non-binary identifying can also be defined as identifying as either having a gender which is in-between or beyond the two categories 'man' and 'woman', as fluctuating between 'man' and 'woman', or as having no gender, either permanently or some of the time. As part of the non-binary spectrum people can identify themselves as genderqueer, neutrois, pangender, polygender, third gender, agender, gender fluid, etcetera. The pronoun that non-binary people often prefer is 'they' and the title 'Mx'.

Transition: The process of changing one's gender presentation permanently to correspond with one's felt gender. It is usually the time when a person begins to live as their felt gender. This can take the form of a social gender role transition (by coming out to people, asking people to use the right pronoun and name, changing name legally, changing the way one presents and looks, etcetera). Transitioning may or may not include gender-affirming medical interventions, such as cross-sex hormone treatment and surgery.

Trans*: This term has been used as the most inclusive umbrella term to include anyone who does not identify as cisgender. In other words, the asterisk used at the end of trans* indicates inclusion of all gender diverse groups and includes genderqueer and non-binary people who self-describe or self-identify in a number of ways (e.g. as pangender, agender, polygender, bigender, gender fluid, neutrois, and so on).

Transsexual: The term transsexual (as an adjective) has been used since 1949 to refer to people who had a clear sense of being '[born] in the wrong body' [149]. The terms 'transsexual' and 'transsexualism' have been used as a diagnosis in the International Classification of Diseases and Health Related Problems (ICD-10; [53]), but are being replaced by a different nomenclature, namely 'gender incongruence'. The terms 'transsexual' and 'transsexualism' are often considered old-fashioned and stigmatizing, and this terminology is increasingly being replaced with terms, such as 'transgender' and 'trans', which most clinicians and members of the trans* community deem more acceptable. Older trans people seem more inclined to use 'transsexual', and to distinguish between this binary identity and other non-binary or intermittent gender expressions. Also, after physically transitioning, many transsexual people consider themselves men or women and no longer identify as a transsexual person.

Finally, we will use the term 'transgender' to refer to individuals who wish to be socially recognized as a gender distinct from their assigned sex at birth, with or without the desire for body modification. For reasons of order and containment, we will limit the content of this chapter to populations—however inconsistently defined—that have either transitioned from one gender to another or who present with a desire to do so. Although we have striven to be as consistent and comprehensive as possible regarding the use of terminology, the listings are by no means exhaustive and continue to develop. For further reading we refer to more extensive glossaries of relevant textbooks [36, 43, 51].

Classification

Over the years, different terms have been used to indicate gender identity-related phenomena. People who experience distress because they do not identify with their gender assigned at birth were formerly known as transsexual people [53], or people with a gender identity disorder [54]. Recently, the name of the diagnosis changed to gender dysphoria (GD) [55], and in the forthcoming 11th version of the International Classification of Diseases of the World Health Organization (WHO), the proposed term for the diagnosis is **gender incongruence** [56–59]. People with GD (as a symptom) have a marked incongruence between the gender they have been assigned at birth and their experienced/expressed gender. This discrepancy is the core component of the diagnosis. There must also be evidence of distress regarding this incongruence for a diagnosis of GD [55]. Experienced gender may include alternative gender identities beyond binary stereotypes. Consequently, the distress is not limited to a desire to simply be of the other gender, but may include a desire to be of an alternative gender, provided that it differs from the individual's assigned gender at birth [40, 55]. The diagnostic criteria for a diagnosis of GD are:

(A) A marked incongruence between one's experienced/expressed gender and assigned gender, of at least 6 months' duration, as manifested by at least two of the following: (1) A marked incongruence between one's experienced/expressed gender and primary and/or secondary sex characteristics (or in young adolescents, the anticipated secondary sex characteristics); (2) A strong desire to be rid of one's primary and/or secondary sex characteristics because of a marked incongruence with one's experienced/expressed gender (or in young adolescents, a desire to prevent the development of the anticipated secondary sex characteristics); (3) A strong desire for the primary and/or secondary sex characteristics of the other gender; (4) A strong desire to be of the other gender (or some alternative gender different from one's assigned gender); (5) A strong desire to be treated as the other gender (or some alternative gender different from one's assigned gender); (6) A strong conviction that one has the typical feelings and reactions of the other gender (or some alternative gender different from one's assigned gender).

(B) The condition is associated with clinically significant distress or impairment in social, occupational, or other important areas of functioning [55].

The American Psychiatric Association (APA) has been criticised for retaining identifying as transgender as a diagnosis in the current mental disorders classification (e.g. [60, 61]). Both the DSM-5 work group of the APA and the ICD-11 working group of the WHO wrestled with the challenges of reducing stigma (which underlies the call for removal of the gender diagnoses) and maintaining access to care (which requires the existence of a diagnosis in order to obtain gender-affirming medical treatment) [62]. The APA chose to retain a diagnosis of GD in the psychiatric classification to ensure access to care and treatment due to existing health insurance infrastructures, mainly in North America, while the WHO took a clear position that issues related to transgender identity should not be classified as mental disorders in the ICD-11. Consequently, the WHO opted for a non-pathologizing reclassification of the newly coined term gender incongruence outside their Mental and Behavioural Disorders chapter; and effectively 'declassified' being transgender as a mental disorder. Individuals with gender incongruence may, or may not, experience distress and they may, or may not, want to live as 'the other gender' (see [59, 63] for an overview). Treatment preferences may also differ from the standard cross-sex hormone treatment and gender-affirming surgeries [41, 64].

Prevalence

Estimating prevalence rates of identifying as transgender is complicated. Most studies investigating prevalence rates in the transgender population have been based on clinical populations [36]. There is growing evidence that an increasing number of people come out as transgender identifying [5, 65, 66], but not every transgender individual chooses to, or is in a position to seek gender-affirming medical interventions [66]. There also remain significant barriers to accessing transgender healthcare [67–69]. The time of the study also affects the prevalence rates, as older studies collected information in an era where society was less tolerant towards transgender people and transgender health services were lacking. An example is the low prevalence rates found in some studies from the 1960s onwards, for example 0.40 per 100 000 [70]; 1.42 per 100 000 [71]. Studies from non-Western countries, such as Thailand and Singapore, where transgender people have a higher visibility and genital affirming medical interventions are more generally available, produce higher prevalence rates. For example, Tsoi found a prevalence of 23.60 transsexual people (as defined at that time) per 100 000 people [72], while Totman reports a prevalence of 333.33 transgender females per 100 000 people [73]. Some prevalence studies have focused on those undergoing gender-affirming genital surgery only, which some transgender people choose not to for a variety of reasons [64]. For example, prevalence rates of 10 per 100 000 were found in Spain [74], 4.28 per 100 000 in Belgium [75], and 16.67 per 100 000 in Sweden [76]. Other prevalence studies are based on transgender people accessing transgender health services or taking cross-sex hormone treatment. These studies have provided prevalence rates of 8.05 per 100 000 people [77], 6.77 [78], 4.42 [79], or 3.88 per 100 000 [80]. Arcelus *et al.* investigated the meta-analytical prevalence of transsexual people and found rates of 4.6 per 100 000 [81]. There was a difference regarding the rates of transsexual women (6.8 per 100 000) and men (2.6 per 100 000). This study found a clear increase in prevalence rates over time. However, the reviewed studies are mainly based on clinical populations. Recently, population studies have investigated prevalence rate of trans* people in a more inclusive way. Unsurprisingly, these studies have found higher rates of people who feel incongruent with their gender assigned at birth.

Population studies from the Netherlands and Flanders in Belgium found rates varying from 0.7 to 1.1% in people assigned male at birth and 0.6 to 0.8% in those assigned female at birth [38, 39]. A meta-regression study of population-based probability samples from the United States found prevalence rates of transgender people of 0.39%, which accounts for approximately 1 million people in North America [82]. The high rates of gender incongruity in people does not necessarily means a high rate of referrals to transgender health services, but may explain the recent increase in demand for these services [65, 67, 83].

Transgender Health Services

In the past two decades, there has been a move away from national, centralized, specialized transgender health service in Western Europe to the gradual introduction and development of regional transgender health services. This has occurred in the Netherlands, Sweden, Finland, Italy, Spain, and the United Kingdom, among others [84–86]. In small countries such as Belgium, Denmark, and Serbia centralized specialist centres remain. Some Northern European countries, most notably the Netherlands, Belgium, and Sweden, which historically have been academic and clinical leaders in the field of transgender healthcare, have moved transgender health services away from psychiatry and based these clinical services within the department of medicine (endocrinology). These developments will not only reduce stigma associated with being trans*, but also serve as exemplary clinical service development models for education and training. Moreover, they will stimulate research in a multidisciplinary setting firmly grounded in physical healthcare [87–89]. The majority of adult transgender health services in Europe consist of a multidisciplinary team or clinical network of clinical professionals, including mental health practitioners (psychologists, psychiatrists, and psychotherapists/counsellors), endocrinologists, speech and language therapists, and surgeons. The assessment team usually makes recommendations to the endocrinologist and surgeon regarding the commencement of gender-affirming medical interventions [45, 90].

In the United States, on the other hand, a loose network of medical and mental health providers exists, often affiliated with the WPATH. The first transgender health clinic in China (Hong Kong) opened as recent as 2016 (personal communication Dr Irene Lam). However, there remain many countries with no transgender health services, while other countries only provide these privately.

WPATH developed the Standards of Care (SoC) for gender transition, which are used globally and currently revised and expected to be published as the WPATH SoC 8th edition in 2020. In the current WPATH SoC7, mental health professionals are tasked with determining whether transgender people who want to undergo gender-affirming treatments meet eligibility criteria, have the capacity to provide informed consent, and have adequately anticipated the psychosocial impact of their transition [90, 91]. Furthermore, the WPATH SoC provides evidence-based recommendations regarding assessment and gender-affirming medical treatment that clinical services for transgender people should follow. WPATH SoC7 has also attempted to improve access to care by including the informed consent model for cross-sex hormone prescription. In multidisciplinary clinics providing transgender care, clinicians specialized in transgender healthcare can diagnose and assess patients with GD [55], who may benefit from treatment with cross-sex hormones, without referral from a mental health practitioner. Patients with coexisting mental health conditions should be referred to mental health services when appropriate. The SoC is closely aligned with the Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline of the Endocrine Society [92]. The latter established an evidence-based framework for endocrine treatment for transgender people, and reaffirms the core role of the endocrinologists in transgender healthcare.

Finally, WPATH advocates for the depathologization of transgender identity, the medical necessity of transgender care, and improvement of access to legal gender change [1, 90].

Assessment of Gender Concerns

A careful and comprehensive assessment procedure, generally comprising one or two appointments [45, 83] is necessary, particularly as most gender-affirming medical interventions are not easily reversible [13]. The assessment should include the age and circumstances when the patient became first aware of a sense of difference from peers regarding the sex assigned at birth, as well as experiences of low self-esteem or negative effect related to that sense of difference. Self-harm and suicidality should always be assessed, as should protective factors such as social and family support [93, 94]. Self-harm, suicidality and suicide are significantly increased in the transgender population [95–99]. Any history of peripubertal and/or pubertal distress due to the anticipation and/or emergence of unwanted secondary sex characteristics, as well as body image, should be explored [100–103]. Past experiences of gender-related stigmatization, discrimination, harassment, and violence should also be addressed [91, 104, 105]. Fertility preservation and child wish must be discussed, particularly with young people, prior to starting cross-sex hormone treatment [106, 107]. Autistic spectrum presentations have been found to be overrepresented among people with GD, although the relationship is far from clear [108, 109]. It is important to clarify of what each patient hopes to achieve and what difficulties may need to be overcome. Some individuals may have considerable anxiety regarding the social transition to another gender role, and may need some support and information, as well as counselling/psychotherapy, regarding the potential gains and losses they may experience by making such a change [110]. Exploring the options for gender expression, as well as the potential risks and benefits of transition, can be some of the tasks of such psychotherapy, although one must note that mandatory psychotherapy (rather than supportive assessment) for transgender people has been shown to be harmful [111, 112]. Psychotherapists may discuss some of the challenges and negotiations that occur in relationships and may explore the impact of stigma and both external and internalized transphobia. The mental health practitioner may provide information, preparation, and support regarding cross-sex hormone and surgical treatments if these are requested [110]. Eligibility for both gender-affirming cross-sex hormone treatment and surgery requires persistent GD, a documented diagnosis of GD or gender incongruence according to established classification criteria [55, 59], and the capacity to give informed consent. In addition, any significant medical or

psychiatric coexisting condition must be sufficiently controlled, so that the risks associated with gender-affirming medical interventions are minimized and there is no interference with the patient's ability to safely adhere to the treatment regimen [90, 91].

Gender-Affirming Medical Interventions

Transgender people may seek any one of a number of gender-affirming interventions, including cross-sex hormone treatment, surgery, facial hair removal, interventions for the modification of speech and communication, and behavioural adaptations, such as genital tucking or packing, or chest binding [42, 90]. There is flexibility in this process, given that some people do not pursue all of these interventions or choose to do so in a different sequence [64]. Generally, the endocrinologist initiates, administers, and monitors hormone treatment, although in some countries, like the United Kingdom, the general practitioner does so guided by the advice of the transgender specialist [42, 45]. A significant proportion (around 20%) of transgender people obtain cross-sex hormones via the internet [113] prior to their first appointment at a transgender health service, partly due to difficulties in accessing services, including long waiting lists. Further detailed information regarding hormone treatment will be provided in the subsequent sections. Gender-affirming surgery is the surgical procedure (or procedures) by which a transgender person's physical appearance and the function of their existing sexual characteristics are altered to resemble that socially associated with their identified gender. For transgender women, genital reconstruction usually involves the surgical construction of a vagina, clitoris, labia, and urinary tract, with removal of the penis and testicles [114]. Breast augmentation, chondrolaryngoplasty (reduction of the Adam's apple), voice surgery, hair implants, and facial feminization surgery are also part of gender-affirming surgeries for transgender women [115]. For transgender men, genital reconstruction may involve construction of a penis, through either phalloplasty or metoidioplasty [116]. Gender-affirming genital surgery may also involve other medically necessary ancillary procedures, such as vaginectomy, hysterectomy, and bilateral salpingo-oophorectomy [117]. Most transgender men choose to undergo a bilateral mastectomy and chest reconstruction (the shaping of a male-contoured chest).

Before genital reconstructive surgery, the WPATH SoC7 also recommend 12 months of continuous hormone treatment (unless hormones are clinically contra-indicated) and 12 continuous months of living in a social gender role that is congruent with the patient's gender identity [90]. Postoperative regret following gonadectomy is low and reported as 0.6% for transgender women and 0.3% for transgender men [66]. For further information regarding gender-affirming surgeries see [13] and [15].

Mental Health and Transgender People

Research investigating mental health problems in transgender people has primarily focused on those attending transgender health services [74, 88, 104, 118–123]. These studies found that the main mental health problems in transgender people accessing transgender health services relate to depression [123], anxiety [13], and self-harm [96, 97]. The high level of mental health problems in the transgender population compared with their cisgender counterparts has been linked to gender minority stress [124–126]. The prevalence of severe psychiatric disorders does not appear to be more common than in the cisgender population [127].

Research investigating the role of cross-sex hormone treatment found that people on cross-sex hormone treatment experience better self-esteem [104, 128], lower symptoms of depression [74, 120, 121, 123, 129–131], lower symptoms of anxiety [46, 74, 104, 119–121, 129, 130], less self-harm behaviour [96, 97], and improved quality of life [122, 128, 132, 133, 134]. Improved body image and a reduction in eating disorders psychopathology have also been found in transgender people taking cross-sex hormone treatment [100, 103, 135–137]. Research investigating the role of gender-affirming surgeries shows improved quality of life, body image, and mental health indicators following surgery [130, 138, 139]. Some studies are controlled for age and gender, but lack a longitudinal design. Few longitudinal studies have taken place to evaluate the effect of gender-affirming medical treatment on mental health and quality of life [140–144]. These studies show a clear improvement of mental health indicators and quality of life, reaching similar levels to those in the cisgender population, but they lack a control group. Other factors (than cross-sex hormone treatment) linked to improved mental health and quality of life consist of social support [93, 104, 118, 145], interpersonal function [146], and lack of transphobia experiences [118, 147]. There is a growing body of scientific evidence that gender-affirming medical interventions are beneficial to binary transgender people, but longitudinal and robustly designed outcome studies are lacking. There is a growing number of non-binary-identifying people attending transgender health services too, with initial evidence indicating that mental health and physical problems are significantly more prevalent in this group than in their binary transgender counterparts. Evidence-based medical treatment options in non-binary people remain unexplored [41].

Conclusions

The last decade has seen a significant worldwide increase in the visibility of transgender and gender-diverse identifying people with a concomitant rise in those seeking gender-affirming medical interventions. Transgender healthcare is a relatively new and multidisciplinary specialty providing assessment, treatment, and support for transgender and gender-diverse identifying patients. There is a growing body of evidence that cross-sex hormone treatment and gender-affirming surgery have a positive beneficial effect of the mental health, body image, and quality of life of the vast majority of transgender and gender-diverse identifying people. The role of the endocrinologist is crucial in the provision of an adequate standard of care in transgender healthcare. It is vital that transgender healthcare constitutes an intrinsic part of education and training within the medical and specialist curriculum.

REFERENCES

1. Knudson G, Green J, Tangpricha V, *et al*., on behalf of the WPATH Executive Committee and Board of Directors. Identity Recognition Statement of the World Professional Association for Transgender Health (WPATH). *Int J Transgend*, 2018; **19**: 355–6

2. Brown S, Kucharska J, Marczak M. Mental health practitioners' attitudes towards transgender people: a systematic review of the literature. *Int J Transgend*, 2018; **19**: 4–24.
3. Flores AR. Attitudes toward transgender rights: perceived knowledge and secondary interpersonal contact. *Polit Groups Identities*, 2015; **3**: 398–416.
4. Kanamori Y, Cornelius-White JHD. Big changes, but are they big enough? Healthcare professionals' attitudes toward transgender persons. *Int J Transgend*, 2016; **17**: 165–75.
5. Bouman WP, de Vries ALC, T'Sjoen G. Gender dysphoria and gender incongruence: an evolving inter-disciplinary field. *Int Rev Psychiatry*, 2016; **28**: 1–4.
6. Bouman WP, de Vries ALC, T'Sjoen G. *Gender Dysphoria and Gender Incongruence*. New York, NY: Routledge, 2016.
7. Frey JD, Poudrier G, Thomson JE, Hazen A. A historical review of gender-affirming medicine: focus on genital reconstructive surgery. *J Sex Med*, 2017; **14**: 991–1002.
8. Miller B, Behm-Morawitz E. Exploring social television, opinion leaders, and Twitter audience reactions to Diane Sawyer's coming out interview with Caitlyn Jenner. *Int J Transgend*, 2017; **18**: 140–53.
9. Sweileh WM. Bibliometric analysis of peer-reviewed literature in transgender health (1900–2017). *BMC Int Health and Hum Rights*, 2018; **18**: 16.
10. Reisner SL, Poteat T, Keatley J, *et al*. Global health burden and needs of transgender populations: a review. *Lancet*, 2016; **388**: 412–36.
11. Shires DA, Jaffee K. Factors associated with health care discrimination experiences among a national sample of female-to-male transgender individuals. *Health Soc Work*, 2015; **40**: 134–41.
12. Winter S, Diamond M, Green J, *et al*. Transgender people: health at the margins of society. *Lancet*, 2016; **388**: 390–400.
13. Bouman WP, Arcelus J, eds. *The Transgender Handbook—A Guide for Transgender People, Their Families and Professionals*. New York, NY: Nova Publishers, 2017.
14. Ettner R, Monstrey S, Coleman E. *Principles of Transgender Medicine and Surgery*, 2nd edn. New York, NY: Routledge, 2016.
15. Thomas R. *Herodotus in Context: Ethnography, Science and the Art of Persuasion*. Cambridge: Cambridge University Press, 2000.
16. Westermarck EA. *The History of Human Marriage*. London: MacMillan Publishers, 1891.
17. Herdt GH. *Third Sex, Third Gender: Beyond Sexual Dimorphism in Culture and History*. New York, NY: Zone Books, 1993.
18. Stryker S. *Transgender History*. Berkeley, CA: Seal Press, 2008.
19. Lobdell BL. *A Strange Sort of Being*. Jefferson, NC: McFarland & Company, 2011.
20. Nanda S. Cross-cultural issues. In: Rowland DL, Incrocci L, eds. *Handbook of Sexual and Gender Identity Disorders*. Hoboken, NJ: Wiley, 2008: 457–85.
21. Vincent B, Manzano A. History and cultural diversity. Chapter 2. In: Richards C, Bouman WP, Barker W-J, eds. *Genderqueer and Non-Binary Genders*. London: Palgrave MacMillan, 2018.
22. Diehl A, Vieira DL, Zaneti MM, *et al*. Social stigma, legal and public health barriers faced by the third gender phenomena in Brazil, India and Mexico: travestis, hijras and muxes. *Int J Soc Psychiatry*, 2017; **63**: 389–99.
23. Nisar MA. (Un)becoming a man: legal consciousness of the third gender category in Pakistan. *Gend Soc*, 2017; **32**: 59–81.
24. Vanderlaan DP, Vokey JR, Vasey PL. Is transgendered male androphilia familial in non-Western populations? The case of a Samoan village. *Arch Sex Behav*, 2013; **42**: 361–70.
25. Gooren LJ, Sungkaew T, Giltay EJ, Guadamuz TE. Cross-sex hormone use, functional health and mental well-being among transgender men (Toms) and Transgender Women (Kathoeys) in Thailand. *Cult Health Sex*, 2015; **17**: 92–103.
25. Odo C, Hawelu A. Eo na Mahu o Hawai'i: the extraordinary health needs of Hawai'i's Mahu. *Pacific Health Dialog*, 2001; **8**: 327–34.
26. Burns C. *Trans Britain: Our Journey from the Shadows*. London: Unbound, 2017.
27. Baer JA. Feminist theory and the law. In: Goodin RE, ed. *The Oxford Handbook of Political Science*. Oxford: Oxford University Press, 2011: 305–20.
28. Holdcroft A. Gender bias in research: how does it affect evidence-based medicine? *J R Soc Med*, 2007; **100**: 2–3.
29. McGregor AJ, Hasnain M, Sandberg K, Morrison MF, Berlin M, Trott J. How to study the impact of sex and gender in medical research: a review of resources. *Biol Sex Differ*, 2016; 7(suppl. 1), 46.
30. Balzer MM. Sacred genders in Siberia. In: Harvey G, ed. *Shamanism*. London: Routledge, 2003: 242–61.
31. Peletz M. Transgenderism and gender pluralism in southeast Asia since early modern times. *Curr Anthropol*, 2006; **47**: 309–40.
32. Driskill Q. Doubleweaving two-spirit critiques: building alliances between native and queer studies. *GLQ: J Lesbian and Gay Stud*, 2010; **16**: 69–92.
33. Bacigalupo AM. *Shamans of the Foye Tree: Gender, Power, and Healing Among Chilean Mapuche*. Austin, TX: University of Texas Press, 2007.
34. Wood ME. How we got this way: the sciences of homosexuality and the Christian right. *J Homosex*, 2000; **38**: 19–40.
35. Reineck K. Running from the gender police: reconceptualizing gender to ensure protection for non-binary people. *Mich J Gend Law*, 2017; **24**: 265–322.
36. Arcelus J, Bouman WP. Language and terminology. In: Bouman WP, Arcelus J, eds. *The Transgender Handbook: A Guide for Transgender People, Their Families and Professionals*. New York, NY: Nova Science, 2017: 1–12.
37. Herman JL, Flores AR, Brown TNT, Wilson BDM, Conron KJ. *Age of Individuals who identify as transgender in the United States*. Los Angeles, CA: The Williams Institute, 2017.
38. Kuyper L, Wijsen C. Gender identities and gender dysphoria in the Netherlands. *Arch Sex Behav*, 2014; **43**: 377–85.
39. Van Caenegem E, Wierckx K, Elaut E, *et al*. Prevalence of gender nonconformity in Flanders, Belgium. *Arch Sex Behav*, 2015; **44**: 1281–7.
40. Richards C, Bouman WP, Seal L, Barker M-J, Nieder TO, T'Sjoen G. Non-binary or genderqueer genders. *Int Rev Psychiatry*, 2016; **28**: 95–102.
41. Richards C, Bouman WP, Barker M-J. *Genderqueer and Non-binary Genders*. London: Palgrave Macmillan, 2018.
42. Ahmad S, Barrett J, Beaini A, *et al*. Gender dysphoria services: a guide for general practitioners and other healthcare staff. *Sex Relationsh Ther*, 2013; **28**: 173–86.
43. Bouman WP, Arcelus J. *A Short Guide to Transgender Health*. London: Jessica Kingsley, 2019.
44. Parkes G, Barrett J, Beail N, *et al*. Gender dysphoria. Chapter 22. In: Fletcher RJ, Barnhill J, Cooper S, eds. *Diagnostic Manual—Intellectual Disability DM-ID-2*, 2nd edn. New York, NY: National Association for the Dually Diagnosed, 2016: 507–20.

45. Wylie K, Barrett J, Besser M, *et al*. Good practice guidelines for the assessment and treatment of adults with gender dysphoria. *Sex Relationsh Ther*, 2014; **29**: 154–214.
46. Bouman WP, Suess Schwend A, Motmans J, *et al*. Language and transgender health. *Int J Transgend*, 2017; **18**: 1–6.
47. Meier SC, Labuski CM. The demographics of the transgender population. In: Baumle AK, ed. *International Handbook on the Demography of Sexuality*. Dordrecht, the Netherlands: Springer, 2013: 289–327.
48. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV transgender women: a systematic review and meta-analysis. *Lancet*, 2013; **13**: 214–22.
49. Safer JD, Coleman E, Feldman J, *et al*. Barriers to health care for transgender individuals. *Curr Opin Endocrinol Diabetes Obes*, 2016; **23**: 168–71.
50. Wylie K. Appreciation of diversity and nomenclature within clinical practice. *J Sex Med*, 2015; **12**: 581–3.
51. Richards C, Barker M. *Sexuality & Gender for Mental Health Professionals*. London: Sage, 2013.
52. Ha JF, Longnecker N. Doctor-patient communication: a review. *Ochsner J*, 2010; **10**: 38–43.
53. World Health Organization (WHO). *International Classification of Diseases* (ICD-10). Geneva: World Health Organization, 1992.
54. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: APA, 2000.
55. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Washington, DC: APA, 2013.
56. Beek T, Cohen-Kettenis PT, Bouman WP, *et al*. Gender incongruence of childhood: clinical utility and stakeholder agreement with the World Health Organization's proposed ICD-11 criteria. *PloS One*, 2017; **12**: e0168522.
57. Beek T, Cohen-Kettenis PT, Bouman WP, *et al*. Gender incongruence of adolescence and adulthood: acceptability and clinical utility of the World Health Organization's proposed ICD-11 criteria. *PloS One*, 2016; **11**: e0160066.
58. Beek TF, Cohen-Kettenis PT, Kreukels BPC. Gender incongruence/gender dysphoria and its classification history. *Int Rev Psychiatry*, 2016; **28**: 5–12.
59. Drescher J, Cohen-Kettenis P, Winter S. Minding the body: situating gender identity diagnoses in the ICD-11. *Int Rev Psychiatry*, 2012; **24**: 568–77.
60. Bouman WP, Bauer GR, Richards C, Coleman E. WPATH Consensus Statement on Considerations on the Role of Distress (Criterion D) in the DSM Diagnosis of Gender Identity Disorder. *Int J Transgend*, 2010; **12**: 100–6.
61. Lev AI. Gender dysphoria: two steps forward, one step back. *Clin Soc Work J*, 2013; **41**: 288–96.
62. Drescher J. Controversies in gender diagnoses. *LGBT Health*, 2013; **1**: 9–13.
63. Cohen-Kettenis PT, Pfäfflin F. The DSM diagnostic criteria for gender identity disorder in adolescents and adults. *Arch Sex Behav*, 2010; **39**: 499–513.
64. Beek T, Kreukels BPC, Cohen-Kettenis PT, Steensma TD. Partial gender request and underlying motives of applicants for gender affirming interventions. *J Sex Med*, 2015; **12**: 2201–5.
65. Aitken M, Steensma TD, Blanchard R, *et al*. Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *J Sex Med*, 2015; **12**: 756–63.
66. Wiepjes CM, Nota NM, de Blok CJM, *et al*. The Amsterdam cohort of gender dysphoria study (1972–2015): trends in prevalence, treatment, and regrets. *J Sex Med*, 2018; **15**: 582–90.
67. McNeil J, Bailey L, Ellis S, Morton J, Regan M. *Trans Mental Health Study 2012*. Edinburgh: the Scottish Transgender Alliance, 2012.
68. Scheim AI, Zong X, Giblon R, Bauer GR. Disparities in access to family physicians among transgender people in Ontario, Canada. *Int J Transgend*, 2017; **18**: 343–52.
69. Stewart L, O'Halloran P, Oates J. Provider quality and barriers to care for transgender people: an analysis of data from the Wisconsin transgender community health assessment. *Int J Transgend*, 2018; **19**: 46–58.
70. Pauly IB. The current status of the change of sex operation. *J Nerv Ment Dis*, 1968; **147**: 460–71.
71. Wålinder J. Transsexualism: definition, prevalence and sex distribution. *Acta Psychiatrica Scandinavica*, 1968; **203**: 255–7.
72. Tsoi WF. The prevalence of transsexualism in Singapore. *Acta Psychiatrica Scandinavica*, 1988; **78**: 501–4.
73. Totman R. *The Third Sex: Kathoey, Thailand's Ladyboys*. London: Souvenir Press, 2003.
74. Gómez-Gil E, Zubiaurre-Elorza L, Esteva I, *et al*. Hormone-treated transsexuals report less social distress, anxiety and depression. *Psychoneuroendocrinology*, 2012; **37**: 662–70.
75. De Cuypere G, Van Hemelrijck M, Michel A, *et al*. Prevalence and demography of transsexualism in Belgium. *Eur Psychiatry*, 2007; **22**: 137–41.
76. Dhejne C, Öberg K, Arver S, Landén M. An analysis of all applications for sex reassignment surgery in Sweden, 1960–2010: prevalence, incidence, and regrets. *Arch Sex Behav*, 2014; **43**: 1535–45.
77. van Kesteren PJ, Gooren LJ, Megens JA. An epidemiological and demographic study of transsexuals in the Netherlands. *Arch Sex Behav*, 1966; **25**: 589–600.
78. Judge C, O'Donovan C, Callaghan G, Gaoatswe G, O'Shea D. Gender dysphoria—prevalence and co-morbidities in an Irish adult population. *Front Endocrinol*, 2014; **5**: 87.
79. Bakker A, van Kesteren PJ, Gooren LJ, Bezemer PD. The prevalence of transsexualism in The Netherlands. *Acta Psychiatr Scand*, 1992; **87**: 237–8.
80. Gómez-Gil E, Trilla Garcia A, Godás Sieso T, *et al*. Estimación de la prevalencia, incidencia y razón de sexos del transexualismo en Cataluña según la demanda asistencial. *Actas Esp Psiquiatr*, 2006; **34**: 295–302.
81. Arcelus J, Bouman WP, Van Den Noortgate W, Claes L, Witcomb G, Fernandez-Aranda F. Systematic review and meta-analysis of prevalence studies in transsexualism. *Eur Psychiatry*, 2015; **30**: 807–15.
82. Meerwijk EL, Sevelius JM. Transgender population size in the United States: a meta-regression of population-based probability samples. *Am J Public Health*, 2017; **107**: e1–e8.
83. Jones BA, Brewin N, Richards C, Van Eijk M, Stephenson-Allen A, Arcelus J. Investigating the outcome of the initial assessment at a national transgender health service: time to review the process? *Int J Transgend*, 2017; **18**: 427–32.
84. Caldarera A, Pfäfflin F. Transsexualism and sex reassignment surgery in Italy, *Int J Transgend*, 2011; **13**: 26–35.
85. Combs R, Turner L, Whittle S. *The Gender Identity Services in England Mapping Project Report*. London: Department of Health, 2008.
86. TGEU. *Overdiagnosed but Underserved*. Berlin: TGEU, 2017.

87. Arcelus J, Bouman WP. (2015). Current and future direction of gender dysphoria and gender incongruence research. *J Sex Med*, 2015; **12**: 2226–8.
88. Heylens G, Elaut E, Kreukels BPC, *et al*. Psychiatric characteristics in transsexual individuals: multicenter study in four European countries. *Br J Psychiatry*, 2014; **204**: 151–6.
89. T'Sjoen G, Motmans J, Arcelus J, Bouman WP. The need of patient involvement in transgender healthcare research. *J Sex Med*, 2017; **14**: 1494–5.
90. Coleman E, Bockting W, Botzer M, *et al*. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgend*, 2012; **13**: 165–232.
91. Byne W, Karasic D, Coleman E, *et al*. Gender dysphoria in adults: an overview and primer for psychiatrists. *Transgend Health*, 2018; **3**: 57–70.
92. Hembree WC, Cohen-Kettenis PT, Gooren L, *et al*. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 3869–903.
93. Davey A, Bouman WP, Arcelus J, Meyer C. Social support and psychological wellbeing: a comparison of patients with gender dysphoria and matched controls. *J Sex Med*, 2014; **11**: 2976–85.
94. Trujillo MA, Perrin PB, Sutter M, Tabaac A, Benotsch EG. The buffering role of social support on the associations among discrimination, mental health, and suicidality in a transgender sample. *Int J Transgend*, 2017; **18**: 39–52.
95. Arcelus J, Claes L, Witcomb GL, Marshall E, Bouman WP. Risk factors for non-suicidal self-injury among trans youth. *J Sex Med*, 2016; **13**: 402–12.
96. Claes L, Bouman WP, Witcomb G, Thurston M, Fernandez-Aranda F, Arcelus J. Non-suicidal self-injury in trans people: associations with psychological symptoms, victimization, interpersonal functioning, and perceived social support. *J Sex Med*, 2015; **12**: 168–79.
97. Davey A, Arcelus J, Meyer C, Bouman WP. Self-injury among trans individuals in transition and matched controls: prevalence and associated factors. *Health Soc Care Community*, 2016; **24**: 485–94.
98. Dhejne C, Lichtenstein P, Boman M, Johansson ALV, Långström N, Landén M. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. *PloS One*, 2011; **6**: e16885.
99. Marshall E, Claes L, Bouman WP, Witcomb GL, Arcelus J. Non-suicidal self-injury and suicidality in trans people: a systematic review of the literature. *Int Rev Psychiatry*, 2016; **28**: 58–69.
100. Jones BA, Haycraft E, Brewin N, Claes L, Bouman WP, Arcelus J. Risk factors for eating disorder psychopathology within the treatment seeking transgender population: the role of cross-sex hormone treatment. *Eur Eat Disord Rev*, 2018; **26**: 120–8.
101. Röder M, Barkmann C, Richter-Appelt H, Schulte-Markwort M, Ravens-Sieberer U, Becker I. Health-related quality of life in transgender adolescents: associations with body image and emotional and behavioral problems. *Int J Transgend*, 2018; **19**: 78–91.
102. Van de Grift TC, Cohen-Kettenis PT, Elaut E, *et al.* A network analysis of body satisfaction of people with gender dysphoria. *Body Image*, 2016; **17**: 184–90.
103. Witcomb GL, Bouman WP, Brewin B, Richards C, Fernandez-Aranda F, Arcelus J. Body image and eating disorders psychopathology in trans individuals: a matched control study. *Eur Eat Disord Rev*, 2015; **23**: 287–93.
104. Bouman WP, Claes L, Marshall E, *et al*. Socio-demographic variables, clinical features and the role of pre-assessment cross-sex hormones in older trans people. *J Sex Med*, 2016; **13**: 711–19.
105. Mizock L, Woodrum TD, Riley J, Sotilleo EA, Yuen N, Ormerod AJ. Coping with transphobia in employment: strategies used by transgender and gender-diverse people in the United States. *Int J Transgend*, 2017; **18**: 282–94.
106. Broughton D, Omurtag K. Care of the transgender or gender-nonconforming patient undergoing *in vitro* fertilization. *Int J Transgend*, 2017; **18**: 372–5.
107. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. *Int Rev Psychiatry*, 2016; **28**: 112–19.
108. Glidden D, Bouman WP, Jones BA, Arcelus J. Gender dysphoria and autism spectrum disorder: a systematic review of the literature. *Sex Med Rev*, 2016; **4**: 3–14.
109. Nobili A, Glazebrook C, Bouman WP, *et al.* (2018). Autistic traits in treatment-seeking transgender adults. *J Autism Dev Disord*, 2018; **48**: 3984–94.
110. Murjan S, Bouman WP. Transgender—Living in a gender different from that assigned at birth. Chapter 12. In: Richards C, Barker MJ, eds. *The Palgrave Handbook of the Psychology of Sexuality and Gender*. London: Palgrave Macmillan, 2015: 198–215.
111. American Psychological Association. *Guidelines for Psychological Practice with Transgender and Gender Nonconforming People*. Washington, DC: APA, 2015.
112. Mizock L, Lundquist C. Missteps in psychotherapy with transgender clients: promoting gender sensitivity in counselling and psychological practice. *Psychol Sex Orientat Gend Divers*, 2016; **3**: 148–55.
113. Mepham NJ, Bouman WP, Arcelus J, Hayter M, Wylie KR. People with gender dysphoria who self-prescribe cross sex hormones: prevalence, sources and side effects knowledge. *J Sex Med*, 2014; **11**: 2995–3001.
114. Selvaggi G, Ceulemans P, De Cuypere G, *et al*. Gender identity disorder: general overview and surgical treatment for vaginoplasty in male-to-female transsexuals. *Plast Reconstr Surg*, 2005; **116**: 135–45.
115. Colebunders B, Verhaeghe W, Bonte K, D'Arpa S, Monstrey S. Male-to-female gender reassignment surgery. In: Ettner R, Monstrey S, Coleman E, eds. *Principles of Transgender Medicine and Surgery*. New York: Routledge, 2016: 250–78.
116. Christopher N, Ralph D, Garaffa G. Genital reconstructive surgery for transgender men. In: Bouman WP, Arcelus J, eds. *The Transgender Handbook*. New York, NY: Nova Publishers, 2017.
117. Colebunders B, D'Arpa S., Weijers S, Lumen N, Hoebeke P, Monstrey S. Female-to-male gender reassignment surgery. In: Ettner R, Monstrey S, Coleman E, eds. *Principles of Transgender Medicine and Surgery*. New York, NY: Routledge, 2016: 279–318.
118. Bouman WP, Davey A, Meyer C, Witcomb GL, Arcelus J. Predictors of psychological well-being among trans individuals. *Sex Relationsh Ther*, 2016; **31**: 357–73.
119. Bouman, W. P., Claes, L., Brewin, N., *et al.* Transgender and anxiety: a comparative study between transgender people and the general population. *Int J Transgend*, 2016; **18**: 16–26.
120. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health. *Psychoneuroendocrinology*, 2014; **39**: 65–73.

121. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere, G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. *J Sex Med*, 2014; **11**: 119–126.
122. Gorin-Lazard A, Baumstarck K, Boyer L, *et al*. Is hormonal therapy associated with better quality of life in transsexuals? A cross-sectional study. *J Sex Med*, 2012; **9**: 531–41.
123. Witcomb GL, Bouman WP, Claes L, Brewin N, Crawford J, Arcelus J. Levels of depression in transgender people and its predictors: results of a large matched control study with transgender people accessing clinical services. *J Affect Disord*, 2018; **235**: 308–15.
124. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay and bisexual populations: conceptual issues and research evidence. *Psychol Bull*, 2003; **129**: 674–97.
125. Millet N, Longworth J, Arcelus J. Prevalence of anxiety symptoms and disorders in the transgender population: a systematic review of the literature. *Int J Transgend*, 2017; **18**: 27–38.
126. Reisner SL, Greytak EA, Parsons JT, Ybarra ML. Gender minority social stress in adolescence: disparities in adolescent bullying and substance use by gender identity. *J Sex Res*, 2015; **52**: 243–56.
127. Dhejne C, van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: a review of the literature. *Int Rev Psychiatry*, 2016; **28**: 44–57.
128. Gorin-Lazard A, Baumstarck K, Boyer L, *et al*. Hormonal therapy is associated with better self-esteem, mood, and quality of life in transsexuals. *J Nerv Ment Dis*, 2013; **201**: 996–1000.
129. Colton-Meier SL, Fitzgerald KM, Pardo ST, Babcock J. The effects of hormonal gender affirmation treatment on mental health in female-to-male transsexuals. *J Gay Lesbian Ment Health*, 2011; **15**: 281–99.
130. Davis SA, Meier SC. Effects of testosterone treatment and chest reconstruction surgery on mental health and sexuality in female-to-male transgender people. *Int J Sex Health*, 2014; **26**: 113–28.
131. Keo-Meier CL, Herman LI, Reisner SL, Pardo ST, Sharp C, Babcock JC. Testosterone treatment and MMPI-2 improvement in transgender men: a prospective controlled study. *J Consult Clin Psychol*, 2015; **83**: 143–56.
132. Gómez-Gil E, Zubiaurre-Elorza L, de Antonio ED, Guillamon A, Salamero M. Determinants of quality of life in Spanish transsexuals attending a gender unit before genital sex reassignment surgery. *Qual Life Res*, 2014; **23**: 669–76.
133. Murad MH, Elamin MB, Garcia MZ, *et al*. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol*, 2010; **72**: 214–31.
134. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. *Transgend Health*, 2016; **1**: 21–31.
135. Fisher AD, Castellini G, Bandini E, *et al*. Cross-sex hormonal treatment and body uneasiness in individuals with gender dysphoria. *J Sex Med*, 2014; **11**: 709–19.
136. Jones BA, Haycraft E, Murjan S, Arcelus J. Body dissatisfaction and disordered eating in trans people: a systematic review of the literature. *Int Rev Psychiatry*, 2016; **28**: 81–94.
137. van de Grift TC, Elaut E, Cerwenka SC, *et al*. Effects of medical interventions on gender dysphoria and body image. *Psychosom Med*, 2017; **79**: 815–23.
138. Ainsworth T, Spiegel J. Quality of life of individuals with and without facial feminization surgery or gender reassignment surgery. *Qual Life Res*, 2010; **19**: 1019–24.
139. Castellano E, Crespi C, Dell'Aquila R, *et al*. Quality of life and hormones after sex reassignment surgery. *J Endocrinol Invest*, 2015; **38**: 1373–81.
140. De Cuypere G, Elaut E, Heylens G, *et al*. Long-term follow-up: psychosocial outcome of Belgian transsexuals after sex reassignment surgery. *Sexologies*, 2006; **15**: 126–33.
141. Johansson A, Sundbom E, Höjerback T, Bodlund, O. A five-year follow-up study of Swedish adults with gender identity disorder. *Arch Sex Behav*, 2010; **39**: 1429–37.
142. Ruppin U, Pfäfflin F. Long-term follow-up of adults with gender identity disorder. *Arch Sex Behav*, 2015; **44**: 1321–9.
143. Smith YL, van Goozen SH, Cohen-Kettenis PT. Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: a prospective follow-up study. *J Am Acad Child Adolesc Psychiatry*, 2001; **40**: 472–81.
144. Smith YLS, Van Goozen SHM, Kuiper AJ, Cohen-Kettenis P. Sex reassignment: outcomes and predictors of treatment for adolescent and adult transsexuals. *Psychol Med*, 2005; **35**: 89–99.
145. Boza C, Nicholson Perry K. Gender-related victimization, perceived social support, and predictors of depression among transgender Australians. *Int J Transgend*, 2014; **15**: 35–52.
146. Davey A, Bouman WP, Meyer C, Arcelus J. Interpersonal functioning among treatment-seeking trans individuals. *J Clin Psychol*, 2015; **71**: 1173–85.
147. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Prevalence of and risk and protective factors for depression in female-to-male transgender Ontarians: Trans PULSE Project. *Can J Commun Ment Health*, 2011; **30**: 135–55.
148. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: APA, 1994.
149. Cauldwell DO. Psychopathia transsexualis. *Sexology*, 1994; **16**: 274–80.
150. Diamond LM, Pardo ST, Butterworth MR. Transgender experience and identity. In: Schwartz SJ, Luyckx K, Vignoles VL, eds. *Handbook of Identity Theory and Research*. New York, NY: Springer, 2011: 629–47.
151. Great Britain. *Gender Recognition Act 2004*. London: Department of Constitutional Affairs, 2004.

11.2

Endocrine Treatment of Transgender Youth

Daniel Klink

Introduction

The endocrine treatment of transgender adolescents consists of two phases: pubertal suspension or gonadal suppression followed by the addition of the desired hormones. During the first phase (further) pubertal development is halted and adolescents can further explore their gender identity and prepare for the next phase. To start treatment The Standards of Care of WPATH [1] mentions the following minimum criteria for puberty suppressing hormones:

1. The adolescent has demonstrated a long-lasting and intense pattern of gender non-conformity or gender dysphoria (GD).
2. GD emerged or worsened with the onset of puberty.
3. Any coexisting psychological, medical, or social problems that could interfere with treatment (e.g. that may compromise treatment adherence) have been addressed, such that the adolescent's situation and functioning are stable enough to start treatment.
4. The adolescent has given informed consent and, particularly when the adolescent has not reached the age of medical consent, the parents or other caretakers or guardians have consented to the treatment and are involved in supporting the adolescent throughout the treatment process.

The minimum ages when puberty suppression should start and when the treatment with gender-affirming hormones (GAH) should commence, remains an issue of debate. At first, an age of 12 and 16 years, respectively, was recommended [2]. These criteria have more juridical roots rather than they are based on endocrine factors. The treatment protocol was developed in the Netherlands, where children may take medical decisions together with their caretakers from the age of 12 years. Subsequently, patients are considered to be legally adult to make medical decisions at age 16. However, boys and girls enter puberty at different ages. Also, psychological maturity differs between adolescents [3]. Therefore, not age but onset of puberty is now mentioned in the Standards of Care. Of course, it is up to the clinicians to assess readiness for treatment, even in older adolescents who have already reached Tanner stage 5. As for starting sex hormones in transgender adolescents the present recommended age by the Endocrine Society still remains 16 years. However, the need for re-evaluating is recognized by the medical profession [4] and the recommended age of starting GAH may shift in the future.

Gonadotropin-Releasing Hormone Analogue (GnRHa) Treatment

Background

Gonadotropin-releasing hormone (GnRH) was isolated and analysed in 1971 [5, 6]. GnRH is an oligopeptide that consists of 10 amino acids: Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly [7, 8]. In 1973 the first GnRHa, were introduced. These analogues reduce the levels of gonadotropins leading to a decrease of gonadal stimulation and subsequent fall of sex hormone levels in the blood to prepubertal levels. They have been used in the treatment of central precocious puberty since 1981 [9, 10]. The benefits of GnRHa have been well-established and their use in central precocious puberty is regarded both safe and effective, with no long-term adverse effects [11]. However, although there is now a broad experience in the transgender population, somatic data is scarce and limited to reported studies of a few centres worldwide [12].

GnRHa treatment will lead to regression of the first stages of the already developed sex characteristics. In adolescents with GD assigned female at birth (transgender boys), the present breast tissue will become weak and may disappear completely. In adolescents

with GD assigned boy at birth (transgender girls), testicular volume will regress to a lower volume. When treatment is started in adolescents in later phases of pubertal development, the various physical changes of pubertal development, such as a late stage of breast development in transgender boys and lowering of the voice and facial hair in transgender girls, will not regress completely, although any further progression will be stopped [3].

Treatment Protocol

In clinical practice, transgender boys usually can start when in Tanner stage Breast 2 and transgender girls when they have a testicular volume of 6–8 ml (see Table 11.2.1). Also, adolescents who have already physically matured but are not yet ready to decide on hormones can use GnRHa to inhibit unwanted pubertal functions, such as breast formation and menses in transgender boys or further male phenotype development and erections in transgender girls to bridge the period until the adolescent's gender identity is more stable [4].

The use of GnRHa in transgender adolescents is off-label but the general safety and efficacy of triptorelin 3.75 mg every 4 weeks subcutaneously have been studied [3, 13]. Anthropometry and body development, hormonal status and metabolic parameters were followed prospectively in late pubertal transgender girls (n = 49)) and transgender boys (n = 67) during 12 months of triptorelin monotherapy [13].

Table 11.2.1 Baseline and follow-up protocol during GnRHa treatment

Start GnRHa at Tanner stage 2–3	Follow-up frequency
Clinical assessment	Every 3–6 months
Height, weight, sitting height	
BMI-SDS	
Tanner stage	
Blood pressure	
Laboratory	
LH, FSH, E2/T	Every 6–12 months
25OH vitamin D	Upon indication
Metabolic profile (upon indication)#	Every 24 months
Liver function	
Fasting glucose and insulin	
Lipids	
Karyotype (upon indication*)	Once
Imaging	
Bone density using DXA	Every 12–24 months
Bone age (upon indication)	Every 12–24 months

Abbreviations: BMI-SDS Body Mass Index Standard Deviation Score; E2, oestradiol; FSH, follicle-stimulating hormone; LH, luteinizing hormone; T, testosterone; DXA, dual-energy X-ray absorptiometry

* in premenarchial transgender boys when physical examination of external genitals is atypical

clinical indication of compromised metabolic health such as obesity and/or prolonged GnRHa monotherapy is expected.

Modified with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, Rosenthal SM, Safer JD, Tangpricha V, T'Sjoen GG. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2017; 102:3869–903; and from de Vries AL, Klink D and Cohen-Kettenis PT. What the Primary Care Pediatrician Needs to Know About Gender Incongruence and Gender Dysphoria in Children and Adolescents. *Pediatr Clin North Am.* 2016 Dec;63(6):1121–35.

Effects on the Hypothalamic–Pituitary–Gonadal (HPG) Axis

Puberty was adequately suppressed with a decrease of testicular volume from 14 ml to 9 ml in 33 transgender girls. In transgender boys, when early in puberty, Tanner Breast 2, breast tissue fully regressed to stage 1 (n = 4) and when menarche had already started menses ceased and breast tissue became unstimulated. Luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels decreased in 3 months to prepubertal levels in both transgender girls as transgender boys. As for sex hormones, testosterone decreased to below detection level in transgender girls within 3 months of therapy. In transgender boys, oestradiol decreased to early pubertal levels [13].

Effects on Anthropometry and Metabolic Profile

In both transgender girls and transgender boys during the first year of GnRHa treatment height standard deviation score (SDS) calculated for sex assigned at birth significantly decreased, reflecting in transgender girls arrested linear growth and in transgender boys the attainment of final height. Changes in body mass index (BMI) SDS were not clinically relevant. With respect to body composition, the lean body mass percentage decreased and fat percentage significantly increased. Glutamyl transferase, aspartate aminotransferase (AST), alanine aminotransferase (ALT), and creatinine levels did not significantly change from baseline to 12 months of treatment but alkaline phosphatase decreased, most likely reflecting the decrease in growth velocity [13].

Effects on Bone Mass

Although bone mass accrual during puberty is regulated by genetic factors and environmental factors such as physical activity and adequate supply of nutrients (calcium, vitamin D), gonadal hormones are key [14–16]. Thus, the hypogonadal state induced by GnRHa monotherapy affects bone mineral density (BMD) [3, 17, 18]. In 30 transgender girls who started GnRHa at a median age of 14.7 years and were on GnRHa monotherapy for 1.7 years the bone mass defined as bone mineral apparent density (BMAD) of the lumbar spine remained stable whereas the BMAD of the hip significantly decreased. Since during puberty normally bone mass accrual occurs *Z*-scores inherently decreased from −0.37 to −1.15 and from −0.5 to −0.95, respectively. Likewise, in transgender boys (n = 41, median age start GnRHa 15.1 years, median duration GnRHa monotherapy 1.3 years) BMAD of the lumbar spine and the hip region decreased. Corresponding Z-scores decreased from 0.14 to −0.62 and from −0.07 to −0.39, respectively [17].

Adverse Events

GnRHa is generally well tolerated with the exception of hot flushes early in treatment [3]. However, hypertension in transgender adolescents under triptorelin treatment was reported in three cases in a cohort of 138 subjects. They were all transgender boys. Hypertension was reversible upon cessation of triptorelin but in one case increased intracranial pressure developed which required the temporary use of acetazolamide [19]. GnRHa induced hypertension

is an uncommon side effect and has only been reported incidentally in girls [20, 21]. Indeed, the association that females seem to be more susceptible than males also is described in adult patients [22, 23].

Monitoring GnRHa Treatment

Based on the evidence published, the guidelines for follow-up have been revised recently [4] and are summarized in Table 11.2.1. However, the cohorts studied were relatively older and more matured populations, mid-teens, and Tanner stage G/M4 and up, which coincides with a relatively shorter duration of an induced hypogonadal state. There are currently no publications available focusing on treatment of the young and less matured (Tanner 2 or 3) adolescent with GD and therefore the effects of prolonged gonadal suppression (i.e. 3 to 4 years; short- or long-term are unknown). Thus, vigilance remains warranted.

Gender-Affirming Hormone Treatment

Hormone therapy in adolescents with GD generally has two treatment regimes. When GnRHa treatment had started in the early stages of pubertal development, the 'new' puberty is induced with a dosage scheme that is also common in hypogonadal patients. Alternatively, when GnRHa treatment had started in physically matured transgender persons and thus the duration of the hypogonadal state was limited, hormones can be given at a higher start dose and more rapidly increased until maintenance dosage (Table 11.2.2). An additional advantage of GnRHa treatment is that hormones do not have to be administered in supraphysiological dosages, which would otherwise be needed to suppress endogenous sex steroid production [4].

Oestrogen Treatment in Transgender Girls

The natural 17β-oestradiol has a less thrombogenic profile compared to synthetic oestrogens, such as ethinyl oestradiol and is therefore preferred. 17β-oestradiol is available in both oral and dermal preparations. Both formulae are off-label for a pubertal induction in transgender girls. Dosage schemes are summarized in Table 11.2.2. For oral 17β-oestradiol it is recommended to start at a dosage of 5 μg/kg/day, followed by 6-monthly increments of 5 μg/kg until a maintenance dosage of 2 mg or higher is reached. A second treatment regime suitable for transgender girls who commenced GnRHa treatment in late puberty (Tanner 4 and up) is recommended. After a preferably short period of gonadal suppression varying from 3 to 6 months, oestrogens can be given at a daily start dosage of 1 mg and increased to 2 mg after 6 months [24, 25]. Transdermal 17b-estradiol may be an alternative for oral 17b-estradiol. Although it is increasingly used for pubertal induction in hypogonadal adolescent females (mostly Turner syndrome), the lack of low-dose oestrogen patches may be a problem. To achieve the appropriate start dosage of 6.25–12 μg/24 hour individuals may need to cut patches. The patch is applied twice a week with increments 12.5 μg/24 hour every 6 months. Adult

Table 11.2.2 Hormone therapy

Dosing scheme for pubertal induction		Dosing scheme when GnRHa started at T_{4-5}	
Transgender girls			
p.o.17β-oestradiol therapy	Increase every 6 months	p.o. 17β-oestradiol therapy	
5 μg/kg/day		1 mg/day	First 6 months
10 μg/kg/day		Adult dosage 2–6 mg/day	
15 μg/kg/day			
20 μg/kg/day			
Adult dosage 2–6 mg/day			
TD 17β-oestradiol therapy	Increase every 6 months		
6.5–12.5 μg/day			
25 μg/day			
37.5 μg/day			
Adult dosage 50–200 μg/day			
Transgender boys			
i.m. Testosterone* injection	Increase every 6 months	i.m. Testosterone* injection	First 6 months
25 mg/m²/2 weeks		75 mg/2 weeks	
50 mg/m²/2 weeks		Adult dosage 200–250 mg/2–3 weeks	
75 mg/m²/2 weeks			
Adult dosage 200–250 mg/2–3 weeks			

* Testosterone formula: testosterone-ester mixture (Sustanon 250), testosterone mono-esters (testosterone enanthate, testosterone cypionate). p.o., per os; TD, transdermal i.m., intramuscular.

Reproduced with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, Rosenthal SM, Safer JD, Tangpricha V, T'Sjoen GG. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2017; 102:3869–903.

maintenance varies from 50 to 200 μg/24 hour [26]. Currently data on dermal oestrogen therapy in adolescent transgender girls is not available and only the effects of the addition of oral 17β-oestradiol were studied prospectively in transgender girls (n=28; minimum duration of oestrogen therapy 1 year) [25].

Breast Development and Effects on the HPG Axis

When oestrogen treatment had started at a median age of 16.0 years after median duration of 24.8 months of triptorelin monotherapy, breast development had started within 3 months and after 1-year median Tanner stage was Breast 3. In the next 3 years there was progression to Breast 5 (n = 16) with a variability of all breast stages. The gonadotropins remained low and when the adult dose of 2 mg oestradiol daily was used during a median duration of 2 years the median serum oestradiol was 100 pmol/L (range 24–380 pmol/L). A change in prolactin levels was not seen [25].

Effects on Anthropometry and Metabolic Profile

In the first 12–24 months, waist circumference SDS (female reference) decreased and while hip circumference increased, the hip circumference SDS did not change. The changes in body shape resulted in a decrease of waist/hip ratio (WHR) SDS from 0.48 to 0.04 SDS reflecting a change towards the affirmed phenotype. BMI increased but BMI-SDS did not. When bone age was less than 15 years at the start of oestradiol median height gain was 6.8 cm after 3 years of oestrogen therapy. Overall final height was 182.7 cm corresponding to +1.9 SD for Dutch adult women. Regarding metabolic parameters, haemoglobin, haematocrit, HbA1c, liver enzymes and creatine did not change [25].

Effects on Bone Mass

After start of oestrogen therapy bone mass accrual reassumed [17] and was described during the first two years of oestrogen treatment in 28 transgender girls after a median duration of GnRHa treatment of 2.5 years. The bone mass defined as BMAD in the lumbar spine increased with 8–12% with a coinciding significant *Z*-score increase. The BMAD of the hip and *Z*-score did not increase [18].

Androgen Treatment in Transgender Boys

For pubertal induction the use of testosterone-esters injections is recommended. The start dosage is 25 mg/m² every two weeks IM; and is increased with 25 mg/m² every 6 months. The maintenance dosages vary from 200 mg every 2 weeks for testosterone monoesters, such as testosterone enanthate, to 250 mg per 3–4 weeks for testosterone-esters mixture. For transgender boys who started treatment in late pubertal stage testosterone can be started at 75 mg IM every 2 weeks, followed by the maintenance dosage after 6 months (Table 11.2.2) [4]. It is advised to continue GnRHa at least until maintenance dosage of testosterone is reached and is preferred to continue until gonadectomy. The effects of androgens and triptorelin have been studied prospectively in ten transgender boys [3].

Effects on the HPG Axis

Virilization of the body occurred with lowering of the voice, increased muscular development, particularly in the upper body, facial and body hair growth, and clitoral growth [3, 4].

Effects on Anthropometry and Metabolic Profile

With respect to metabolic profile, both carbohydrate metabolism and lipids did not change during 2 years of testosterone therapy (n = 10) [3].

Effects on Bone Mass

After start of GAH the BMD increases both at the lumbar spine as in the hip [17]. During the first 2 years of GAH therapy BMAD of the lumbar spine increased with 4–9% coinciding with an increase of *Z*-score. The degree of bone mass accrual depended on the bone age. A bone age younger than 14 years (female reference) favoured a higher increase. The BMAD of the hip increased with 5–8% resulting in an increase of *Z*-score [18].

Monitoring Gender-Affirming Hormone Therapy

Guidelines for follow-up during pubertal induction are summarized in Table 11.2.3. Once pubertal induction is completed, guidelines for transgender adults can be followed. However, BMD should be monitored closely until the age of 25–30 or until peak bone mass has been reached [4].

Long-Term Outcome of Gender-Affirming Therapy in Adolescence Using GnRHa

The psychological benefits of gender-affirming treatment for young adolescents with GD using GnRHa have been established [27, 28]. One year after surgery the GD was alleviated, psychological functioning had steadily improved and well-being was similar to or better than same-age young adults from the general population [28]. In addition, some somatic outcome in young adulthood has also become available [17, 29].

Effects on Bone Mass

The effect of gender-affirming treatment in adolescence on bone mass in young adulthood has been studied. In transgender girls (n = 15; median age at start of GnRHa 14.9 years; duration of gonadal suppression 1.5 years) the areal BMD *Z*-score of the lumbar spine was −0.77 at start GnRHa. During GnRHa monotherapy areal BMD remained stable. However, since the physiological development would be bone mass accrual, *Z*-score—as expected—decreased to −0.1. After commencing treatment, oestrogens areal BMD increased but *Z*-score decreased further to −1.36 at the age of 22. Three transgender women had a T-score <−2.5.

In transgender boys (n = 19; median age start GnRHa 15.0 years; median duration GnRHa monotherapy) during GnRHa monotherapy both areal BMD of the lumbar spine and of the hip decreased. Corresponding Z-scores decreased from 0.17 to −0.72 and from 0.36 to −0.35, respectively. Subsequently, between start testosterone therapy and age 22, areal BMD of the lumbar spine and hip increased significantly. Z-score of the lumbar spine showed a trend of increase to −0.33 whereas Z-score of the hip remained the same at −0.35. At age 22 years, a BMD Z-score of the lumbar spine was lower when compared to start GnRHa but this was not significant. No subject had a T-score <−2.5.

Concluding, at age 22 the *Z*-score of the lumbar spine was under pretreatment level, implying a possible delay in or loss of peak

Table 11.2.3 Baseline and follow-up protocol during pubertal induction

	Follow-up frequency
Clinical assessment	Every 3–6 months
Height, weight, sitting height	
BMI-SDS	
Tanner stage	
Blood pressure	
Laboratory	Every 6–12 months
E2, prolactin in transgender girls	
T in transgender boys	
25OH vitamin D	
Metabolic profile (upon indication)	
Haemoglobulin, haematocrit*	
Liver function*†	
Serum electrolytes (potassium)†	
Fasting glucose and insulin#	
Lipids*	
Imaging	
Bone density using DXA	Every 12–24 months
Bone age (upon indication)	Every 12–24 months

Abbreviations: BMI-SDS Body Mass Index Standard Deviation Score; E2, oestradiol; T, testosterone; DXA, dual-energy X-ray absorptiometry.

clinical indication of compromised metabolic health such obesity; *transgender boys using testosterone; † transgender girls using cyproterone acetate.

Modified with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, Rosenthal SM, Safer JD, Tangpricha V, T'Sjoen GG. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2017; 102:3869–903; and from de Vries AL, Klink D and Cohen-Kettenis PT. What the Primary Care Pediatrician Needs to Know About Gender Incongruence and Gender Dysphoria in Children and Adolescents. *Pediatr Clin North Am*. 2016 Dec;63(6):1121–35.

bone mass. This was more pronounced in transgender women than in transgender men. Taken into consideration that even before the start of GnRHa transgender girls had a Z-score below reference population, transwomen who have started gender-affirming therapy in adolescence may be more at risk for lesser bone health than transgender men [17]. One case report has been published on long-term BMD development and it was shown that absolute BMD and Z-scores of a transgender man, treated with GnRHa in his adolescence was in the normal range at age 35. However, pretreatment data was not provided [30].

Anthropometry and Metabolic Profile

The change in body shape and body composition from the start of treatment with GnRHa until 22 years was studied in 192 transgender individuals. At a mean age of 14.5 years 71 transgender girls started GnRHa, to be followed by the addition of 17β-oestradiol (mean age 16.4 years). During treatment till the age of 22 waist and hip circumference increased with 8 cm and 17 cm, respectively. WHR decreased from 0.81 to 0.77. Regarding metabolic profile only body composition was studied and it was found that the percentage of total body fat increased with 9% and thus the percentage of lean body mass decreased with 9%. Percentage of body fat increased in the android region with 9% and the gynoid region with 11%.

Transgender men (n = 121) who started GnRHa at a mean age of 15.3 years, followed by the addition of testosterone (mean age 16.9 years) were studied. From start treatment to the age of 22, transgender men showed increases in waist and hip circumference of 6 cm and 5 cm, respectively. WHR increased from 0.77 to 0.80. Percentage of total body fat decreased with 3% to 27% and percentage of lean body mass inversely increased to 73%. Percentage of body fat decreased in the gynoid region with 5% with no change in the android region.

When compared at 22 years to age-matched peers, it was found that in young adult transgender women, SDS for WHR, body fat, and lean body mass showed greater similarity to ciswomen than to cismen. In transgender men SDS for WHR, body fat, and lean body mass were within reference values for ciswomen and cismen. in transgender men, an earlier Tanner stage at start of treatment appeared to be associated with a closer resemblance of body shape to their affirmed sex at 22 years, and this tended to be the same in transgender women [29].

Alternatives to GnRHa Treatment

GnRHa treatment can be costly and when resources are limited, alternative treatment regimens should be considered. Antiandrogens such as cyproterone acetate (CPA) or spironolactone can be used in transgender girls. Transgender boys can be treated with progestogens [4].

Induction of Hypogonadal State with Antiandrogens in Transgender Girls

Similar to transgender women, in late pubertal transgender girls endogenous androgen production can be suppressed using antiandrogens. CPA is dosed 25–50 mg daily and spironolactone dosages varies from 100 to 300 mg per day [4]. The effects of prolonged CPA monotherapy were studied retrospectively [31, 32].

Effects on HPG Axis

After 6 months of CPA 50 mg once daily testosterone decreased and remained stable at 7.8 nmol/L. LH and FSH however were not suppressed. Prolactin increased almost twofold but none developed galactorrhoea. Clinically more than half of the subjects reported reduced shaving frequency and in 29.6% breast development (Tanner stage Breast 2–3) was reported [31].

Effects on Anthropometry and Metabolic Profile

Weight development did not demonstrate an increase of BMI-SDS. Metabolic parameters such as lipid profile and glucose homeostasis were not negatively affected [31]. However, there was a significant increase in fat mass and decrease in lean mass. The fat tissue accumulation in transgender girls was more than in their age-matched male peers [32].

Effects on Bone Mass

Twenty-one transgender girls were started on 50 mg CPA once daily at a chronological age of 16.3 years (bone age 17.1 years) and changes in bone were studied over a period of 10.6 months.

Already at start of CPA, areal BMD was lower in transgender girls as compared to their peers of the same natal sex. During the study period, areal BMD of total hip decreased, coinciding with a *Z*-score decrease. At the lumbar spine, areal BMD remained stable, but *Z*-scores decreased [32].

Adverse Events

Only a transient increase of liver enzymes was seen in 15% of the study subjects. The levels remained under the threshold of three times the upper limit and therefore treatment was not stopped. The use of CPA is associated in adults with depression [33] but the only reported side effect in adolescents was fatigue [31].

Pubertal Induction in Antiandrogen-Treated Transgender Girls

For the start of oestrogens, the same dosage schemes (Table 11.2.2) as in GnRHa-treated transgender girls can be used. When choosing a dosage scheme, factors such as pretreatment sexual maturation level and duration of hypogonadal state should be taking into consideration. Data on the addition of oestrogens to antiandrogenic therapies in transgender adolescents is limited to two published studies. In one study the subjects received CPA [31] and in the other study spironolactone [34] was used.

Breast Development and Effects on HPG Axis

When 17β-oestradiol was started at 0.5 mg orally once daily and subsequently increased to 0.75 mg after 6 months, initiation or further progression of breast development occurred. After 12 months of oestrogen therapy 66.7% of the participants reached Tanner Breast 3 and 9.5% reached Tanner Breast 4. During this period testosterone decreased significantly to 0.06 nmol/L. LH and FSH became more suppressed. Prolactin levels decreased and the mean 17β-oestradiol level was 121 pmol/L [31].

Effects on Anthropometry and Metabolic Profile

The addition of oestrogens to CPA monotherapy did not result in a change of BMI-SDS. With regards to metabolic parameters, lipid profile and glucose homeostasis did not change [31]. Oestrogens were added in transgender girls (n = 44; age range 14–25 years) who received spironolactone (dosage 50–200 mg daily) in different methods, namely, oral (dosage 1–8 mg daily), intramuscular (dosage 20–80 mg monthly) or transdermal (dosage 0.025–0.200 mg weekly). For all methods there were no changes reported in BMI, metabolic parameters, lipid profile. The use of spironolactone did not change potassium levels [34].

Adverse Events

Serious adverse events were not reported [31, 34]. The adverse event 'fatigue' as reported during CPA monotherapy resolved in almost all transgender girls. Although transient ALT and AST increases were observed sporadically, interruption of treatment was never needed. Also there was a decline in prolactin levels [31]. Prolactin levels in spironolactone treated transgender girls were not affected [34].

The Use of Progestational Agents in Transgender Boys

In postmenarche adolescent transgender boys as alternative for GnRHa to stop or decrease menses frequency, progestogens such as lynestrenol [35] or medroxyprogesterone acetate [36] can be used. Effects of lynestrenol monotherapy were reported [32, 35].

Effects on HPG Axis

LH decreased but levels of FSH did not change. Testosterone levels decreased by almost 30% whereas oestradiol decreased by almost 60% and remained stable afterwards. In the first 6 months metrorrhagia occurred in 50% but reduced to 18% in the following 6 months. Percentage of amenorrhoea was not reported [35].

Effects on Anthropometry and Metabolic Profile

Weight increased during the first 6 months of lynestrenol monotherapy but returned to baseline value after 12 months. Haemoglobin and haematocrit increased but remained in the normal male range. Serum creatinine significantly increased during the first 6 months but remained stable afterwards. Total cholesterol and triglyceride levels did not change; however, mean high density-lipoprotein (HDL) decreased significantly and mean low-density lipoprotein (LDL) levels increased significantly in the first 6 months of lynestrenol monotherapy. No significant changes in haemoglobin A1c (HbA1c), glucose levels, insulin levels, or homeostasis model assessment (HOMA) index were reported (n = 42) [35]. When body composition was studied, it was found that during the first year of treatment lean mass, but not fat mass, significantly increased coinciding with a decreased body fat percentage. Compared to their age-matched female peers transgender boys became more lean. Also, pQCT measurements demonstrated that the muscle area increased at the non-dominant forearm and left lower leg (n = 44) [32].

Effects on Bone Mass

Changes in bone mass were studied during lynestrenol monotherapy (dose 5–10 mg daily, duration 11.6 months) in transgender boys (n = 44; age at start 16.2 years). Bone mass—as defined areal BMD—of the hip increased with an increase of *Z*-score. In the lumbar spine, areal BMD increased while *Z*-scores remained stable, suggesting an acquisition similar to age-matched female peers. Based on pQCT scans, both trabecular and cortical bone parameters at the radius and the tibia increased during the study period, to the same extent as in age-matched control girls [32].

Adverse Events

Fatigue (7%), headache (12%), and hot flushes (10%) were reported. In addition, in one patient, ALT levels transiently increased moderately after 12 months of lynestrenol but normalized after adding testosterone [35].

Pubertal Induction in Transgender Boys Treated Previously with Progestational Agents

Testosterone can be added to progestogens following the dosage schemes summarized in Table 11.2.2. In choosing which dosage scheme is appropriate, one should consider duration of endogenous sex steroid exposure prior to start progestogens. The clinical effects and effects on metabolic parameters in adolescent transgender boys have been investigated retrospectively [34, 35].

Effects on HPG Axis

Testosterone therapy increased mean testosterone levels significantly in the first months, already at the lowest dose (50 mg/2 weeks) and further increased in the next months to reach testosterone values well within the male reference range. Some patients exceeded the male upper reference due to blood sampling close to the last testosterone

injection. There was a non-significant rise in oestradiol levels most likely representing the effect of aromatization of the injected testosterone esters. Complete suppression of both gonadotropins was now achieved after the addition of testosterone to lynestrenol [35].

Anthropometry and Metabolic Profile

Clinically, there was weight gain as both BMI [34] and BMI-SDS increased [35]. Haemoglobin and haematocrit increased but mostly remained within the normal male range [35]. Rarely, haematocrit increased to supraphysiological levels [34]. ALT, AST, creatinine increased but remained in the normal range. Lipid profile was more unfavourable with an increase of cholesterol and LDL and a decrease of HDL. Glucose homeostasis parameters HbA1c [34, 35] and insulin, glucose, or HOMA index [35] were not affected.

Adverse Events

The prevalence of acne significantly increased in the first 6 months of testosterone therapy requiring treatment with oral retinoic acid in some individuals. In addition, fatigue was reported [35]. Despite sporadic incidence of polycythemia, no serious cardiovascular adverse events were reported [34].

Therapy Monitoring

In contrast to GnRHa treatment there are no international guidelines for alternative treatment regimes in transgender adolescents and clinicians are for therapy monitoring referred to protocols for adults [4]. Although data in adults are more abundant, protocols cannot be readily translated to adolescents. Firstly, adolescents may not respond the same to therapeutic interventions. Indeed, where in adult transgender women CPA and oestradiol resulted in increased body weight and body fat [37, 38], in adolescent transgender girls these changes were not observed [31]. Secondly, although adolescents reach sexual maturation in their teens, other tissues such as bone and brain are not fully developed until young adulthood [39, 40]. Therefore, it is recommended to follow the guidelines for monitoring gender-affirming hormone therapy in adolescents as summarized in Table 11.2.3 for, at a minimum, during the endocrine transition process.

Conclusion

Early medical intervention in transgender youth has been performed for about 20 years. After at first considered very controversial, nowadays treatment is generally accepted and endorsed by international endocrine communities. The limited availability of somatic outcome and the specific therapeutic needs of the transgender adolescent warrants a multidisciplinary approach provided in specialized centres.

REFERENCES

1. World Professional Association for Transgender Health (WPATH). Standards of care for the health of transsexual, transgender and gender non-conforming people, 7th version. 2011. Available at: http://www.wpath.org/ (accessed: 3 March 2020).
2. Hembree WC, Cohen-Kettenis PT, Delemarre-van de Waal HA, *et al.* Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Endocrinol Metab*, 2009; **94**: 3132e54.
3. Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. *Eur J Endocrinol*, 2006; **155**: S131e7.
4. Hembree WC, Cohen-Kettenis PT, Gooren L, *et al.* Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*, 2017; **102**: 3869–903.
5. Schally AV, Nair RM, Redding TW, Arimura A. Isolation of the luteinizing hormone and follicle-stimulating hormone-releasing hormone from porcine hypothalami. *J Biol Chem*, 1971; **246**: 7230–6.
6. Schally AV, Arimura A, Baba Y, *et al.* Isolation and properties of the FSH and LH-releasing hormone. *Biochem Biophys Res Commun*, 1971; **43**: 393–9.
7. Matsuo H, Baba Y, Nair RM, Arimura A, Schally AV. Structure of the porcine LH- and FSH-releasing hormone. I. The proposed amino acid sequence. *Biochem Biophys Res Commun*, 1971; **43**: 1334–9.
8. Baba Y, Matsuo H, Schally AV. Structure of the porcine LH- and FSH-releasing hormone. II. Confirmation of the proposed structure by conventional sequential analyses. *Biochem Biophys Res Commun*, 1971; **44**: 459–63.
9. Comite F, Cutler GB, Jr., Rivier J, Vale WW, Loriaux DL, Crowley WF, Jr. Short-term treatment of idiopathic precocious puberty with a long-acting analogue of luteinizing hormone-releasing hormone. A preliminary report. *N Engl J Med*, 1981; **305**: 1546–50.
10. Laron Z, Kauli R, Zeev ZB, Comaru-Schally AM, Schally AV. D-TRP5-analogue of luteinising hormone releasing hormone in combination with cyproterone acetate to treat precocious puberty. *Lancet*, 1981; **2**: 955–6.
11. Carel JC, Eugster EA, Rogol A, *et* al. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics*, 2009; **123**: e752–62.
12. T'Sjoen G, Arcelus J, Gooren L, Klinkd D, Tangpricha V. Endocrinology of transgender medicine. *Endocr Rev*, 2019; **40**: 97–117.
13. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and safety of gonadotropin-releasing hormone agonist treatment to suppress puberty in gender dysphoric adolescents. *J Sex Med*, 2016; **13**: 1125–32.
14. Berger C, Goltzman D, Langsetmo L, *et al.* Peak bone mass from longitudinal data: implications for the prevalence, pathophysiology, and diagnosis of osteoporosis. *J Bone Miner Res*, 2010; **25**: 1948–57.
15. De Schepper J, Derde MP, Van den Broeck M, Piepsz A, Jonckheer MH. Normative data for lumbar spine bone mineral content in children: influence of age, height, weight, and pubertal stage. *J Nucl Med*, 1991; **32**: 216–20.
16. Mughal MZ, Khadilkar AV. The accrual of bone mass during childhood and puberty. *Curr Opin Endocrinol Diabetes Obes*, 2011; **18**: 28–32.
17. Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *J Clin Endocrinol Metab*, 2015; **100**: E270–5.

18. Vlot MC, Klink DT, den Heijer M, Blankenstein MA, Rotteveel J, Heijboer AC. Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. *Bone*, 2017; **95**: 11–19.
19. Klink, D, Bokenkamp, A. Dekker, C; Rotteveel J. Arterial hypertension as a complication of triptorelin treatment in adolescents with gender dysphoria. *Endocrinol Metab Int J*, 2015; **2**: 8.
20. Calcaterra V, Mannarino S, Corana G, *et* al. Hypertension during therapy with triptorelin in a girl with precocious puberty. *Indian J Pediatr*, 2013; **80**: 884–5.
21. Siomou E, Kosmeri C, Pavlou M, Vlahos A, Argyropoulou M, Siamopoulou A. Arterial hypertension during treatment with triptorelin in a child with Williams-Beuren syndrome. *Pediatr Nephrol*, 2014; **29**: 1633–6.
22. Bonfirraro G, Sanna B, De Marco A, Chieffi O, Bonfirraro PP, Orlando V. [Use of a GnRH analogue in the treatment of certain forms of endometrial hyperplasia associated with menometrorrhagia]. *Minerva Ginecol*, 1995; **47**: 467–70.
23. Levine GN, D'Amico AV, Berger P, *et al.* Androgen-deprivation therapy in prostate cancer and cardiovascular risk: a science advisory from the American Heart Association, American Cancer Society, and American Urological Association: endorsed by the American Society for Radiation Oncology. *Circulation*, 2010; **121**: 833–40.
24. de Vries AL, Klink D and Cohen-Kettenis PT. What the primary care pediatrician needs to know about gender incongruence and gender dysphoria in children and adolescents. *Pediatr Clin North Am*, 2016; **63**: 1121–35.
25. Hannema SE, Schagen SEE, Cohen-Kettenis PT, Delemarre-van de Waal HA. Efficacy and safety of pubertal induction using 17beta-estradiol in transgirls. *J Clin Endocrinol Metab*, 2017; **102**: 2356–63.
26. Karberg-Lindgren C, Kristro MB, Norjavaara E. Physiological estrogen replacement therapy for puberty induction in girls: a clinical observational study. *Horm Res Paediatr*, 2014; **81**: 239–244.
27. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med*, 2011; **8**: 2276–83.
28. de Vries AL, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, 2014; **134**: 696–704.
29. Klaver M, Mutsert R, Wiepjes C, Twisk J, *et al.* Early hormonal treatment affects body composition and body shape in young transgender adolescents. *J Sex Med*, 2018; **15**: 251–60.
30. Cohen-Kettenis PT, Schagen SE, Steensma TD, *et al.* Puberty suppression in a gender dysphoric adolescent: a 22-year follow-up. *Arch Sex Behav*, 2011; **40**: 843e7.
31. Tack LJW, Heyse R, Craen M, *et al.* Consecutive cyproterone acetate and estradiol treatment in late-pubertal transgender female adolescents. *J Sex Med*, 2017; **14**: 747–57
32. Tack L JW, Craen M, Lapauw B, *et al.* Proandrogenic and antiandrogenic progestins in transgender youth: differential effects on body composition and bone metabolism. *J Clin Endocrinol Metab*, 2018; **103**: 2147–56.
33. Seal LJ, Franklin S, Richards C, Shishkareva A, Sinclaire C, Barrett J. Predictive markers for mammoplasty and a comparison of side effect profiles in transwomen taking various hormonal regimens. *J Clin Endocrinol Metab*, 2012; **97**: 4422–8.
34. Jarin J, Pine-Twaddell E, Trotman G, *et al.* Cross-sex hormones and metabolic parameters in adolescents with gender dysphoria. *Pediatrics*, 2017; **139**: DOI: 10.1542/peds.2016-3173.
35. Tack LJW, Craen M, Dhondt K, Vanden Bossche H, Laridaen J, Cools M. Consecutive lynestrenol and cross-sex hormone treatment in biological female adolescents with gender dysphoria: a retrospective analysis. *Biol Sex Differ*, 2016; **7**: 14.
36. Lynch MM, Khandheria MM, Meyer WJ. Retrospective study of the management of childhood and adolescent gender identity disorder using medroxyprogesterone acetate. *Int J Transgenderism*, 2015; **16**: 201–8.
37. Wierckx K, Van Caenegem E, Schreiner T, *et al.* Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European Network for the Investigation of Gender Incongruence. *J Sex Med*, 2014; **11**: 1999–2011.
38. Van Caenegem E, Wierckx K, Taes Y, *et al.* Preservation of volumetric bone density and geometry in trans women during cross-sex hormonal therapy: a prospective observational study. *Osteoporos Int*, 2015; **26**: 35–47.
39. Berger C, Goltzman D, Langsetmo L, *et al.* Peak bone mass from longitudinal data: implications for the prevalence, pathophysiology, and diagnosis of osteoporosis. *J Bone Miner Res*, 2010; **25**: 1948–57.
40. Tamnes CK, Herting MM, Goddings AL, *et al.* Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *J Neurosci*, 2017; **37**: 3402–12.

11.3

Hormone Therapy in Transgender Women

Vin Tangpricha and Craig Sineath

Introduction

Recent studies estimate that 0.6% of the adult population and 0.7% of the adolescent population currently identify as transgender, equalling about 1.4 million transgender people in the United States [1]. As this number continues to rise and treatment options become more readily available, more transgender patients will be presenting to their physicians for medical and surgical treatment for gender affirmation [2, 3].

When a transgender woman is ready to pursue medical transition, she must meet with a physician to be prescribed gender-affirming hormone therapy. Traditionally, this physician was an endocrinologist; however, to improve access to hormone therapy, there has been a shift towards other medical providers prescribing hormone therapy. With the ease of prescribing hormone therapy and monitoring its potential side effects and complications as outlined in this chapter, more primary care physicians should become comfortable prescribing hormone replacement therapy (HRT). The most recent Endocrine Society guidelines for hormone therapy for gender dysphoric/gender incongruent persons allow for hormone therapy to be initiated by non-endocrinologists who are able to make and confirm the diagnosis of gender dysphoria. Several other barriers exist to transgender people receiving hormone therapy. One study demonstrated that barriers to transgender people receiving care included bias within the medical field and lack of provider awareness and education about interacting with and treating transgender people [4].

Medical transition via hormone therapy in transgender women has the goal of reversing some of the effects of androgens and providing feminizing effects on the body of the person and allowing the person's physical body to be more congruent with their gender identity. Effects of these therapies include softening of the skin, redistribution of body fat, breast development, and decreased hair growth. Atrophy of androgen dependent tissue (i.e. seminal vesicles and prostate) also occur.

These effects are achieved medically via supplementation with oestrogens and antiandrogens, or surgically by removal of the androgen-producing testes from the body via bilateral orchiectomy. This chapter outlines the use of hormone therapy in transgender women, their wanted and adverse effects on the body, how they should be monitored in the clinical setting, and future directions in the medical hormonal treatment of transgender women.

Initial Evaluation

Before being prescribed gender-affirming hormone therapy, an initial evaluation should be conducted by a physician. According the recent guidelines released by the Endocrine Society, this evaluation should include well-documented gender dysphoria/incongruence, capacity for informed consent, age of majority, and well-controlled mental health concerns if present [5].

This visit should also be a place for the patient to discuss goals of therapy with the physician. Many people presenting with gender dysphoria desire total feminization from hormone therapy, but this is not always the case. It is important to understand where the person is in their transition and what they hope to gain from taking hormone therapies. Some transgender people may not be ready for full medical transition due to not being 'out' to family members or fear of discrimination at work. It is also important to note that many transgender people describe psychological improvement when starting even low doses of hormone therapies before any physical effects take place [6]. Furthermore, some transgender people may identify as gender nonbinary which means that they wish to acquire some but not all of the features of a different gender to better align with their gender identity.

This initial visit should also include a discussion on risks and benefits of hormone therapy with the person. Also, a complete history and physical should be taken with an emphasis of modifying adverse events such as thromboembolic disease (i.e. smoking, sedentary lifestyle, obesity, and underlying personal or family history of blood-clotting disorders). While hormone therapy for transgender people is considered relatively safe under medical supervision, there are some adverse effects that may occur that should be discussed with each individual person [7–9].

These are discussed later in this chapter. For monitoring purposes throughout the transgender person's transition, baseline labs should be drawn at this visit for comparison after the person starts therapy. These should include serum oestradiol and serum testosterone. Once a person has begun therapy, the goal is to keep these values within the normal female reference range. Having levels above this range increases the risk of adverse outcomes such as venous thromboembolism (VTE), ischaemic stroke, and lipid abnormalities [5].

Gender-Affirming Hormone Therapy

Gender-affirming hormone therapy for transgender women requires a dual approach: antiandrogens are needed to reduce the effects of testosterone on the body and oestrogens to increase the supply of feminizing hormone. Progestins have often been added to regimens in the past. These were thought to augment breast development when prescribed with oestrogens; however, there have been no data to show any benefit or additional feminization [3]. In fact, progestins, when prescribed with oestrogens, have been shown to increase the risk of cardiovascular and cerebrovascular disease [6]. Therefore, physicians should counsel all individuals that the risks of progestins out-weigh any potential perceived benefits of such therapy.

The effects of hormones on the physical changes will vary from person to person. There are a variety of formulations currently available on the market and the exact regimen for each person should be tailored based on their physiologic response to the medications and the available local pharmacy formulary. Care should be taken to ensure the levels of these hormones do not exceed normal physiologic levels that occur in reference cisgender females, since many of the adverse events related to HRT may be worsened at higher hormonal levels. Table 11.3.1 outlines the various medications currently available and recommended for hormone therapy in transgender feminine individuals.

Oestrogen Agents

Many forms of oestrogen are available on the market including oral, intramuscular, and transdermal formulations. Ethinyl oestradiol is the form of oral oestrogen commonly used in birth control and was formerly the mainstay in oestrogen replacement for gender-affirming hormone therapy. Guidelines currently recommend against the use of this agent for hormone therapy in transgender people due to a strong associated increase in the risk of cardiovascular disease and venous thrombosis [3, 10]. Oral 17 β-oestradiol, which must be taken daily, has been shown to have less of these complications and is commonly used for gender transition. Conjugated oestrogens such as Premarin are also not recommended because levels are oestradiol are not able to be monitored in blood which may result in inadvertent supraphysiologic levels of oestrogens in the circulation.

Injectable and transdermal (patch) formulations of oestradiol avoid first-pass metabolism from the liver. This property theoretically allows the medication to have a more favourable profile in terms of adverse effects. However, no head to head studies have been conducted to demonstrate the superiority of these compounds to oral preparations of oestrogen.

Testosterone-Lowering Agents

Many drugs exist for testosterone-lowering agents in gender-affirming hormone therapy. There is currently no standard for prescribing these medications.

Spironolactone, a mineralocorticoid receptor antagonist with an added property of inhibiting androgen receptors, is commonly prescribed by physicians as part of gender-affirming hormone therapy [3]. This is the primary antiandrogen prescribed by physicians in the United States. Cyproterone acetate (CPA) is a progestin and androgen receptor antagonist and is widely used in Europe due to historical reasons.

Gonadotropin-releasing hormone (GnRH) is a hormone that is cyclically released in the hypothalamus and causes the pituitary gland to release follicle-stimulating hormone and luteinizing hormone. When a GnRH agonist medication is given, this disrupts the cyclic nature of the hormone and causes the pituitary to halt release of follicle-stimulating hormone (FSH) and luteinizing hormone (LH), resulting in a decline in testosterone production. These medications, such as leuprolide, are relatively expensive, but are used as first-line treatment in the United Kingdom [3].

5-alpha reductase inhibitors block the conversion of testosterone to its more active form. These medications are not recommended by the Endocrine Society since they do not lower testosterone concentrations and have been associated with sexual dysfunction and depression [5].

Table 11.3.1 Preferred options for gender-affirming hormone therapy in transgender women

	Mode of administration	Type	Total dose	Frequency
Oestrogen	Oral	Oestradiol (17β-oestradiol valerate)	2–6 mg	Once or twice daily
	Parenteral (IM)	Oestradiol valerate	5–30 mg	Every 1–2 weeks
		Oestradiol cypionate	2–10 mg	Every week
	Transdermal	Oestradiol patch	0.1–0.4 mg	New patch every 3–5 days
Antiandrogens	Oral	Spironolactone	100–400 mg	Once or twice daily
GnRH agonists	Intramuscular	Leuprolide	3.75–7.5 mg	Monthly
	Implant	Histrelin	50 mg	Implanted every 12 months
	Implant (subcutaneous)	Goserelin	3.6 mg	Implanted monthly

* The medications listed in this table are not FDA approved for gender transition due to no studies being available.

Source data from Hembree, W.C., *et al.*, Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. *Endocr Pract*, 2017. 23(12): p. 1437. Copyright © 2017 AACE.

It is important to note that after surgical removal of the testes, the main androgen-producing organ, patients no longer need to take antiandrogen medications and may require a reduction in their oestrogen dosage [3].

Feminizing Effects of HRT

The goals of gender-affirming hormone therapy for transgender women, outlined next, are to reduce the virilizing effects of testosterone (i.e. decrease terminal hair growth, oiliness of the skin, muscle mass, and voice changes) and to enhance the feminizing effects of oestrogen (i.e. redistribution of body fat, breast growth, softening of the skin). Many of the physical results of taking hormone therapy are irreversible (Table 11.3.2).

Breast Development

A recent study suggests that most breast development in patients taking feminizing hormones occurs within the first 6 months of therapy, and only modest growth occurs after 12 months [11]. Furthermore, there has been no correlation seen with breast development based on dose or even type of hormone regimen given, and most transgender patients report insufficient breast growth from HRT [12]. However, the transgender women studied here were also taking cyproterone which may have interfered with the breast development. The histologic changes that occur in breast tissue of a transgender women taking feminizing hormones appear to be identical to cisgender women with formation of ducts, lobules, and acini [13]. More research is still needed to understand breast development in transgender women taking hormones. Patients who are not fully satisfied with the breast development that occurs from hormone therapy may choose to have breast augmentation surgery.

Table 11.3.2 Feminizing effects of HRT

Effect	Expected onset	Expected maximum onset
Redistribution of body fat	3–6 months	2–3 years
Decrease in muscle mass and strength	3–6 months	1–2 years
Softening of skin/decreased oiliness	3–6 months	Unknown
Decreased sexual desire	1–3 months	3–6 months
Decreased spontaneous erections	1–3 months	3–6 months
Male sexual dysfunction	Variable	Variable
Breast growth	3–5 months	2–3 years
Decreased testicular volume	3–6 months	2–3 years
Decreased sperm production	Unknown	>3 years
Decreased terminal hair growth	6–12 months	>3 years
Scalp hair	Variable	–
Voice changes	None	–

Body Fat Distribution

After starting feminizing hormone therapy transgender women experience an increase in body fat and changes in body fat distribution. One study showed that transgender women experience a body fat increase of 18% the android region (abdomen, chest, and shoulders), 42% in the leg region, and 35% in the gynoid region (hips, thighs, and buttocks). Transgender women also had a decrease in the waist-to-hip ratio, mainly due to an increase in hip circumference [14].

Dermatological Effects

Transgender women on hormone therapy see several changes in their skin including hair growth and sebum production. While hormone therapy does not completely reverse the androgen effects of hair growth that have already taken place if a person has already undergone puberty, there is a decrease in facial and body hair growth and density [15]. Hair-shaft diameter also decreases and has been shown to reach its maximum reduction in size after four months of oestrogen plus antiandrogen therapy [16]. These changes are often more evident on the abdominal hair compared to facial hair due to having lower hair density and hair-shaft diameter. Many transgender women opt to undergo laser hair removal to correct the masculinizing effects of testosterone. Transgender women on oestrogen and antiandrogen agents also have a decrease in skin sebum production to levels that are almost undetectable. One study showed transgender women had significantly less oily skin after four months of therapy [16]. Patients with acne often will notice less lesions due to the decrease in sebum production.

Voice

Feminizing hormone therapy has no effect on the voice. Therefore, if a patient has undergone virilizing puberty where their voice has deepened, starting hormone therapy will not provide any feminizing effects on the voice (i.e. increasing the tone). Many transgender women choose to see a voice therapist to learn techniques for making their voice sound more feminine. Some patient may also choose to undergo vocal cord surgery to tighten their vocal cords and make their voice higher pitched.

Genitalia and Sexual Desire

Patients on feminizing hormone therapy see a reduction in testicular and prostate size and decreased erectile function [6]. Also, testosterone levels are attributed to maintaining a person's sexual desire. In feminizing hormone therapy, testosterone levels are generally kept at low levels, resulting in a reduction in sexual drive [17].

Monitoring

According to the Endocrine Society's 2017 guidelines, a person on gender-affirming hormone therapy should be monitored every 3 months in the first year and every 6–12 months thereafter with a physical exam and labs. Serum testosterone levels should be kept below 50 ng/dl and oestradiol levels should be kept in the physiologic range of about 200 pg/ml. Keeping the hormone levels within these parameters has been shown to be relatively safe and have less adverse effects than if the patient were to achieve supraphysiologic levels of oestrogen.

Cardiovascular Safety

The greatest health concern with feminizing hormone therapy is thromboembolic disease, which is increased in patients on oestrogen [7]. The risk of this adverse event is even more pronounced in patients with compounding risk factors including age greater than 40, tobacco use, sedentary lifestyle, obesity, and pre-existing underlying thrombotic disease [6]. Physicians should counsel their transgender women on these risks and encourage lifestyle changes such as smoking cessation to reduce these adverse events. Furthermore, the risk of thromboembolic disease may be decreased in individuals who are taking oestrogens transdermally or intramuscularly compared to those on oral pills [3].

There is also some evidence that the lipid profile of a person changes when they are taking feminizing hormone therapy. Improved measures include a decrease in low-density lipoprotein (LDL) and an increase in HDL [18]. Hypertriglyceridemia has also been shown to be associated with feminizing hormone therapy [19]. Oral oestrogen therapy seems to have a larger impact on this when compared to transdermal therapies here as well; however, the clinical implications of this increase in triglycerides has yet to be studied.

Studies have also demonstrated a relationship with feminizing hormone therapy increased blood pressure [18]. However, it is thought that this change may be more so related to changes in bodyweight and fat distribution when a person is taking gender-affirming hormone therapy than actual direct effects of the medications.

Hepatic Safety

Use of oestrogens has been linked to transient elevation in liver enzymes with the rare complication of hepatotoxicity [6]. As mentioned previously, use of injectable and transdermal formulations of oestrogen avoid first-pass metabolism through the liver and may have lower risk of this complication. Neither the Endocrine Society nor the World Professional Association for Transgender Health (WPATH) recommend routine evaluation of liver enzymes in patients on feminizing hormone therapy.

Endocrine Safety

Prolactin levels should also be monitored in transgender women on HRT at baseline, one year after starting hormone therapy, and then every two years indefinitely [5]. Oestrogen is a strong stimulator of prolactin formation and release from the pituitary gland, and there have been case reports of patients on gender-affirming feminizing hormone therapy who have developed prolactinomas [20]. However, the risks of hyperprolactinemia have been debated and may be increased in transgender women taking cyproterone as compared to those taking spironolactone [21].

Electrolyte Changes

Patients on spironolactone, traditionally used as a potassium-sparing diuretic, have the added risk of hyperkalaemia. Serum electrolytes should be monitored every three months when beginning this therapy, and then every 12 months thereafter [5].

Bone Health

Sex hormones are integral to the preservation of bone density. Use of oestrogens in HRT has shown to increase bone density and reduce the formation of osteoporosis [3]. There have been no evidence-based guidelines released on monitoring the bone density of transgender women. The Endocrine Society recommends that clinicians consider doing a bone-mineral density test at baseline before starting therapy and then screen for osteoporosis in low-risk patients at 60 years of age or in patients who are not adherent to hormone therapy [5].

Oncology

Very little research has been published on cancer in transgender individuals and there are currently no guidelines for cancer screening specifically in transgender patients. While it is rare, there have been several cases reported of prostate cancer in patients on feminizing hormone therapy, and effects of receiving a bilateral orchiectomy on development of prostate cancer in this patient population has yet to be established [22]. There have also been cases of breast cancer development in transgender women reported in the literature [13]. While no evidence-based guidelines have been released, the Endocrine Society recommends that physicians continue with current guidelines for cancer screening recommended for cisgender populations for all anatomy that is present in the person. This would include prostate and breast cancer screening in this population.

Mental Health

Transgender patients have unique mental health needs. Diagnoses of depression and anxiety occur at disproportionately high prevalences in patients with gender dysphoria, and transgender people are at increased risk of suicide compared to their cisgender peers [23, 24]. A mental health professional is often very helpful in addressing any coexisting mood or social situations before and during the hormone transition process. While many psychological symptoms may be directly linked to a person's gender dysphoria and improve after starting gender-affirming treatment, the transgender population remains a vulnerable population for chronic psychiatric conditions due to the unique daily stress and trauma they experience as a result of stigma and discrimination [25]. It is important for transgender people to have access to culturally competent mental health services throughout their lives to meet their unique healthcare needs.

Other Considerations

Fertility Preservation

Taking oestrogens has been shown to reduce spermatogenesis in transgender women. However, research on the exact effects of dosing, length of treatment, and reversibility of hormone therapy on fertility have yet to be established [26]. Since it is known that taking gender-affirming feminizing hormone therapy has currently unpredictable and potentially irreversible dramatic effects on a person's fertility, it is important to counsel all transgender women starting this therapy on options for fertility preservation. These options include sperm banking for patients who are able to provide sperm of sufficient quality and quantity. However, some transgender women experience such severe dysphoria concerning their genitals, and this is not an option. There are several procedures that can be done by specialists (i.e. urologists) that include surgical sperm extraction and testicular tissue cryopreservation [27]. It is important to note

that while these technologies are available, most are expensive and are not feasible for a large proportion of transgender people (more details can be found in Chapter 11.5).

Lactation

There are case reports of functional lactation achieved in transgender patients. One such case report describes a transwoman on domperidone, oestradiol, progesterone, with breast pump stimulation who was able to achieve lactation with the ability to feed her child solely on her own milk for 6 weeks [28]. Breastfeeding as a transgender woman has not been described well in the literature, but is a promising and expanding field of research.

REFERENCES

1. Flores AR, Herman JL, Gates, GJ, Brown TNT. *How Many Adults Identify as Transgender in the United States?* Los Angeles, CA: The Williams Institute, 2016.
2. Safer JD, Tangpricha V. Care of the transgender patient. *Ann Intern Med*. 2019; **171**(1): ITC1-ITC16. doi:10.7326/AITC201907020.
3. Safer JD, Tangpricha V. Care of transgender persons. *N Engl J Med*. 2019; 381(25): 2451–60. doi:10.1056/NEJMcp1903650.
4. Puckett JA, Cleary P, Rossman K, Newcomb ME, Mustanski B. Barriers to gender-affirming care for transgender and gender nonconforming individuals. *Sex Res Social Policy*, 2018; **15**: 48–59.
5. Hembree WC, Cohen-Kettenis PT, Gooren L, *et al*. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. *Endocr Pract*, 2017; **23**: 1437.
6. WPATH. WPATH standards of care for the health of transgender and gender nonconforming people. 2012. Available at: https://www.wpath.org/publications/soc (accessed: 5 March 2020).
7. Weinand JD, Safer JD. Hormone therapy in transgender adults is safe with provider supervision: a review of hormone therapy sequelae for transgender individuals. *J Clin Transl Endocrinol*, 2015; **2**: 55–60.
8. Wierckx K, Van Caenegem E, Schreiner T, *et al*. Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. *J Sex Med*, 2014; **11**: 1999–2011.
9. Unger CA. Hormone therapy for transgender patients. *Translational Andrology and Urology*, 2016; **5**: 877–84.
10. Toorians AW, Thomassen MC, Zweegman S, *et al*. Venous thrombosis and changes of hemostatic variables during cross-sex hormone treatment in transsexual people. *J Clin Endocrinol Metab*, 2003; **88**: 5723–9.
11. de Blok CJM, Klaver M, Wiepjes CM, *et al*. Breast development in transwomen after 1 year of cross-sex hormone therapy: results of a prospective multicenter study. *J Clin Endocrinol Metab*, 2018; **103**: 532–8.
12. Wierckx K, Gooren L, T'Sjoen G. Clinical review: breast development in trans women receiving cross-sex hormones. *J Sex Med*, 2014; **11**: 1240–7.
13. Sonnenblick EB, Shah AD, Goldstein Z, Reisman T. Breast imaging of transgender individuals: a review. *Curr Radiol Rep*, 2018; **6**: 1.
14. Klaver M, de Blok CJM, Wiepjes CM, *et al*. Changes in regional body fat, lean body mass and body shape in trans persons using cross-sex hormonal therapy: results from a multicenter prospective study. *Eur J Endocrinol*, 2018; **178**: 165–73.
15. Gao Y, Maurer T, Mirmirani P. Understanding and addressing hair disorders in transgender individuals. *Am J Clin Dermatol*, 2018; **19**: 517–27.
16. Giltay EJ, Gooren LJ. Effects of sex steroid deprivation/administration on hair growth and skin sebum production in transsexual males and females. *J Clin Endocrinol Metab*, 2000; **85**: 2913–21.
17. Wierckx K, Elaut E, Van Hoorde B, *et al*. Sexual desire in trans persons: associations with sex reassignment treatment. *J Sex Med*, 2014; **11**: 107–18.
18. Elbers JM, Giltay EJ, Teerlink T, *et al*. Effects of sex steroids on components of the insulin resistance syndrome in transsexual subjects. *Clin Endocrinol (Oxf)*, 2003; **58**: 562–71.
19. Maraka S, Ospina NS, Rodriguez-Gutierrez R, *et al*. Effect of sex steroids on lipids, venous thromboembolism, cardiovascular disease and mortality in transgender individuals: a systematic review and meta-analysis. In: Endocrine Society's 98th Annual Meeting and Expo, 2016, Boston, MA.
20. Cunha FS, Domenice S, Câmara VL, *et al*. Diagnosis of prolactinoma in two male-to-female transsexual subjects following high-dose cross-sex hormone therapy. *Andrologia*, 2015; **47**: 680–4.
21. Bisson JR, Chan KJ, Safer JD. Prolactin levels do not rise among transgender women treated with estradiol and spironolactone. *Endocr Pract*, 2018; **24**: 646–51.
22. Deebel NA, Morin JP, Autorino R, *et al*. Prostate cancer in transgender women: incidence, etiopathogenesis, and management challenges. *Urology*, 2017; **110**: 166–71.
23. Becerra-Culqui TA, Liu Y, Nash R, *et al*. Mental health of transgender and gender nonconforming youth compared with their peers. *Pediatrics*, 2018; **141**(5).
24. World Health Organization. *Mental Health Action Plan 2013—2020*. Geneva: WHO, 2013.
25. James SE, Herman JL, Rankin S, *et al*. The Report of the 2015 U.S. Transgender Survey. National Center for Transgender Equality, 2016. Available at: https://www.transequality.org/sites/default/files/docs/USTS-Full-Report-FINAL.PDF (accessed 5 March 2020).
26. Rowlands S, Amy JJ. Preserving the reproductive potential of transgender and intersex people. *Eur J Contracept Reprod Health Care*, 2018; **23**: 58–63.
27. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. *Int Rev Psychiatry*, 2016; **28**: 112–19.
28. Reisman T, Goldstein Z. Case report: induced lactation in a transgender woman. *Transgend Health*, 2018; **3**: 24–6.

11.4

Hormone Therapy in Transgender Men

Guy T'Sjoen and Justine Defreyne

Introduction

As the number of trans persons actively seeking psychological, endocrinological, and/or surgical treatment is increasing [1, 2] it is important to get healthcare professionals acquainted with transgender care. Currently, transgender care is often not a (strong) part of the medical curriculum [3–5], which may lead to miscommunication, misinformation, and not referring trans persons to the appropriate care providers. This may lead to people taking hormones without the follow-up of a physician or undergoing clandestine gender-affirming surgery. Due to real or perceived stigma and fear of medicalization when reaching out to healthcare providers, trans persons may encounter barriers when accessing healthcare [6]. As gender-affirming hormonal and surgical treatment has been shown to reduce or even resolve feelings of gender dysphoria in trans persons [7–17], gender-affirming care should be accessible to all trans persons.

In most countries, hormonal treatment is ideally prescribed under the supervision of endocrinologists. However, many endocrinologists may feel uneasy and unskilled when working with the transgender population. This chapter aims to summarize the endocrine treatment for transgender men wishing physical transition.

Initial Evaluation

The Endocrine Society guideline 'Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline' recommends that a medical professional confirms the diagnosis of gender dysphoria in transgender persons prior to the initiation of gender-affirming hormone therapy. Medical professionals should document that the gender dysphoria has been persistent and that the individual is of legal age and able to make an informed decision and consent for treatment. If significant medical or mental health concerns are present, they must be reasonably well-controlled, but the presence of mental health concerns does not necessarily exclude patients from gender-affirming hormones [18]. If hormone therapy is contraindicated due to serious health conditions, the practitioner should ensure access to non-hormonal interventions for gender dysphoria, which can be discussed with a qualified mental health professional (see Chapter 11.1).

Transgender people seeking gender-affirming hormone therapy expect to change their physical appearance to better match their gender identity and expression, which can range from an androgynous presentation to maximum virilization [18]. Therefore, hormone prescribing physicians should always individualize the hormone therapy based on the patient's goals, the risk/benefit ratio of the treatment, the presence of other medical conditions, taking into account possible social and economic issues [19]. The physician who prescribes the hormones is responsible for discussing the expected effects of testosterone therapy and the possible adverse health effects, including the possible reduction in fertility (which should ideally be discussed before the patient is referred to the endocrinology department). He/she should also provide ongoing medical monitoring with regular physical and laboratory examination and communicate as needed with the patient's primary care provider, mental health professional, surgeon, and/or other healthcare professionals involved [18, 19]. Gender-affirming hormonal treatment (and surgery) has been shown to reduce and in some cases even resolve gender dysphoria and mental health problems. In addition, gender-affirming treatment increases self-reported quality of life (QOL). Satisfaction rates among transgender persons are high and gender-affirming treatment is rarely regretted [20–23]. Ideally, gender-affirming hormonal therapy is only available by prescription. However, one must realize that there is a black market for steroids, including testosterone analogues. In addition, in non-Western countries or areas where gender-affirming care is less easily available, gender-affirming hormone therapy is often self-prescribed without supervision by a medical professional. Available evidence from the United States and Europe suggest that hormone therapy initiated and monitored under the supervision of a medical professional is associated with very low rates of adverse events [24–27].

Guidelines, including the World Professional Association for Transgender Health (WPATH) Standards of Care edition 7 (SOC

7) and the Endocrine Society guideline advises to screen transgender people for conditions that can be aggravated by hormone therapy, before initiating gender-affirming care. Medical history should be assessed thoroughly, including history of thromboembolic diseases (e.g. deep vein thrombosis, pulmonary embolism), arterial hypertension, polycythemia and sleep apnoea [18]. Testosterone therapy is usually continued lifelong after oophorectomy, unless medical contraindications arise [19].

Gender-Affirming Hormone Therapy

Testosterone Agents

Hormonal therapy in transgender men is aimed at inducing virilization and consists of testosterone agents, which is preferably administered intramuscularly or transdermally. If patients are screened adequately and follow-up visits are provided regularly, testosterone therapy is safe on the longer term [28–30]. Testosterone therapy is contraindicated in case of (desired) pregnancy or lactation. In order to maintain virilization and avoid hypogonadism symptoms (e.g. osteoporosis or vasomotor symptoms), testosterone therapy must be continued lifelong. Testosterone regimens in transgender men follow the general principle of hormone replacement therapy of male hypogonadism, aiming at cisgender male reference ranges for serum testosterone levels [18]. Most commonly prescribed are injectable testosterone esters, used in dosages of 200–250 mg IM every 2–3 weeks [18, 31]. Long-acting testosterone undecanoate 1000 mg IM every 10–12 weeks is also being used for treatment of transgender men [32]. In the United States, both the patient and provider must undergo risk evaluation and mitigation strategy (REMS) training in order to be able to initiate testosterone undecanoate therapy, due to the potential risk of oil pulmonary embolism. The use of oral testosterone (testosterone undecanoate 160–240 mg/day), axillary solutions, patches, nasal sprays, buccal tablets, or pellets is rarely reported for treatment in transgender men. If transgender men wish to self-inject testosterone, subcutaneous administration is also an option at a at a median dosage of 75 mg weekly [31, 33]. Other options include topical androgen gel 1% (25–100 mg per day) or transdermal patches (2.5–10 mg/day) [18]. As long as patients adhere to the prescribed dose controlled by their physician, there is no difference between the different testosterone formulations regarding short-term safety, compliance, metabolic parameters, body composition and general life satisfaction [34]. The dose of testosterone may be increased (e.g. from 200 mg of testosterone esters to 250 mg, biweekly or from testosterone undecanoate once every 12 weeks to once every 10 weeks) if trough serum total testosterone levels do not reach cisgender male levels after six months of testosterone therapy, with subsequent monitoring of serum haematocrit levels. The dose of testosterone may be decreased if serum haematocrit levels rise above male reference ranges or if desired by the patient (see monitoring—cardiovascular safety) (Table 11.4.1).

As transgender men often describe menstrual bleeding as a discomfort [35, 36] medication to induce amenorrhea should be discussed with every transgender man early in the transition process. If suppression of the menses is desired or menstrual bleeding does not cease, a progestational agent or a gonadotropin-releasing hormone (GnRH) analogue can be added to the treatment regimen. Suppression of menstrual bleedings is less likely in people treated with transdermal or oral testosterone [24] and enquiring about the persistence of menstrual bleeding and the degree of inconvenience due to this should be part of the assessment. The most commonly used agents include oral lynestrenol 5–10 mg daily or medroxyprogesterone 5–10 mg daily. The use of progestational agents is contraindicated in case of pregnancy (history of) breast cancer or any other gynaecological cancer, unexplained vaginal bleeding, severe arterial disease, liver insufficiency, and severe liver or bile diseases. Possible side effects include gastrointestinal discomfort, water- and salt retention, weight gain, decreased sexual desire and spotting. GnRH analogues are rarely used in adults because of the costs of the therapy. Another rarely used option is endometrial ablation [26]. After hystero-oophorectomy, progestational therapy can be stopped.

Virilizing Effects of Testosterone Therapy

Hormone therapy with testosterone induces virilization; it causes the voice to break, sometimes menses stop, facial and body hair

Table 11.4.1 Preferred options for gender-affirming hormone therapy in transgender men

	Mode of administration		Type	Dosage	Frequency
Testosterone	Parenteral	Intramuscular	Testosterone esters	200–250 mg	Every 2–3 weeks
			Testosterone undecanoate[a]	1000 mg	Every 10–12 weeks
		Subcutaneous	Testosterone esters	75–125 mg	Every week
	Transdermal		Androgen gel 1%[b]	25–100 mg	Once daily
			Transdermal patches	2.5–10 mg	Once daily
Progestational agents	Oral		Lynestrenol	5–10 mg	Once daily
			Medroxyprogesterone	5–10 mg	Once daily
			Nomegestrol	5–10 mg	Once daily
	Parenteral		Medroxyprogesterone	150 mg	Once every three months

[a] One thousand milligrams initially followed by an injection at 6 weeks, then at 12-week intervals.
[b] Avoid cutaneous transfer to other individuals.

Adapted with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, *et al.* Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2017; Progestational agents. Copyright © 2017 Oxford University Press.

Table 11.4.2 Virilizing effects of testosterone therapy

Effect	Expected onset	Expected maximum onset
Cessation of menses	1–6 months	n/a
Skin oiliness/acne	1–6 months	1–2 years
Clitoral enlargement	1–6 months	1–2 years
Vaginal atrophy	3–6 months	1–2 years
Body fat redistribution	1–6 months	2–5 years
Deepened voice	6–12 months	1–2 years
Facial/body hair growth	6–12 months	4–5 years
Increased muscle mass/strength	6–12 months	2–5 years
Scalp hair loss	6–12 months	Variable

Adapted with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, *et al.* Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2017; Progestational agents. Copyright © 2017 Oxford University Press.

appears in a male pattern, the clitoris enlarges (mean 3.83 +/– 0.42 cm after 2 years of testosterone therapy) and the musculature becomes bulkier and more pronounced, leading to an overall good passability (being perceived by others as a cisgender man) in daily life [7, 9]. It is important to discuss the possibilities and limitations of testosterone therapy before and during gender-affirming hormone therapy; height and bone structure will not change if testosterone therapy is initiated after puberty. In addition, testosterone therapy will not affect the degree of subcutaneous fat, unless physical activity is initiated or increased (Table 11.4.2) [37].

Dermatological Effects

In the pilosebaceous unit of the skin, both androgen and oestrogen receptors are expressed, enabling testosterone to have effects on both the skin and hair of transgender men. Testosterone therapy leads to an increase in acne both at the face and back after four months, with peak severity after 6 months [30]. Long-term data (10-year follow-up) shows no to mild acne in 93.9% of transgender men [38]. Testosterone therapy also leads to a desirable increase in body and facial hair. The increase is already apparent during the first year of treatment, although after 12 months of testosterone therapy, the facial and abdominal hair profile is still less compared to cisgender men [39]. Long-term data (10-year follow-up) shows further increasing body and facial hair growth after the first year, with an increase of the Ferriman–Gallwey score from median 0.5 (before hormone therapy) to 12 after 1 year, while long-term testosterone treatment (10 years) resulted in a median score of 24 [38]. While body and facial hair increases, male pattern baldness (MPB) may occur with thinning of hair related to the duration of testosterone therapy. One retrospective study found that 38.3% of transgender men had MPB type II-V [39] and another study by Wierckx *et al.* reported a 17% incidence of developed androgenetic alopecia based on the Norwood-Hamilton classification after 1 year of treatment [40]. Studies with longer follow-up data report an incidence of 32.7% of mild frontotemporal hair loss and 31% of moderate to severe androgenetic alopecia after 10 years of testosterone therapy [41]. If desired, oral finasteride 1 mg daily can be added to the treatment regimen [41].

Voice

Testosterone causes hypertrophy of muscle cells with a reduction of surrounding fat cells. These effects are also exerted at the level of the larynx, which results in acoustic changes [42]. In high dose, testosterone is responsible for the development of the male voice characteristics. Androgens act virilizing to the voice in birth-assigned females from a concentration of 150 μg/dl. Repeated exposure to concentrations higher than 200 μg/dl results in irreversible changes, which occur predominantly during the first three months of testosterone therapy [43]. Often, the voice of transgender men cannot be distinguished from cisgender men, although transgender men are more likely to experience problems with pitch quality (incidence 10%) [44, 45]. In transgender men in whom the voice was perceived as male, overall well-being was higher compared to those in whom the voices were less gender congruent [46].

Genitalia

Testosterone therapy causes the clitoris to grow. Clitoral length already increases by 60% after only 3 months of testosterone therapy (mean clitoral length after three months: 3.19 ± 0.54 cm) and continues to enlarge during the testosterone therapy (mean clitoral length after 2 years: 3.83 ± 0.42 cm). As long-term follow-up data are not available, it is uncertain if further growth past the second year of testosterone therapy is to be expected [7]. Long-term testosterone use causes vaginal and cervical atrophy, with decreased vaginal secretions which can result in vaginal dryness or itching, and painful penetration reported by some patients [47]. If necessary, vaginal dryness can be helped by simple over-the-counter lubricants [48].

Sexual Desire

In cisgender men, numerous studies have shown a relationship between sexual desire and circulating levels of testosterone [49–51], which has also been described in cisgender women [52, 53]. Testosterone therapy in transgender men may facilitate sexual desire and arousal. However, to date, studies focusing on sexual functioning in trans men treated with testosterone who did not (yet) undergo genital gender-affirming surgery are scarce. Wierckx *et al.* [54] reported higher sexual desire in transgender men who recently initiated testosterone therapy. Costantino *et al.* [51] reported an increase in frequency of desire, masturbation, sexual fantasies, and arousal after 1 year of testosterone administration. Studies on trans persons (without substratification of trans men and trans women) show an increased sexual QOL in persons on gender-affirming hormones [55].

Monitoring

See Box 11.4.1.

Cardiovascular Safety

Testosterone is one of the most anabolic hormones in the human body; it has a positive effect on protein synthesis, stimulates fat oxidation and reduces lipid storage [56]. Despite this, cardiovascular mortality is higher in adult men, compared to women [57]. This difference may be caused by sex hormone differences, although at physiological levels, both testosterone and oestradiol are thought to be involved in maintaining insulin sensitivity [58].

Box 11.4.1 Suggested guideline for the monitoring of testosterone therapy in transgender men

- Provide adequate follow-up every 3 to 4 months during the first year of testosterone therapy and one to two times per year thereafter.
- Monitor for appropriate signs of virilization, serum sex steroid levels, and development of adverse reactions.
- For testosterone esters, the testosterone level should be measured midway between injections with a target level of 400–700 ng/dl (13.9–24.3 nmol/L) of total serum testosterone.
- For testosterone undecanoate, testosterone should be measured just before the following injection. If the level is <400 ng/dl (13.9 nmol/L), increase the dosing interval.
- For transdermal testosterone, the testosterone level can be measured at least 2 hours after application. It is not advised to measure serum testosterone levels within the first week of testosterone therapy.
- Measure serum haematocrit or haemoglobin at baseline and every three to four months during the first year and one to two times per year thereafter.
- Monitor weight, blood pressure, and serum lipid levels at regular intervals.
- Screening for osteoporosis should be conducted in transgender men who stop testosterone therapy, are not compliant with hormone therapy or who develop risks for bone loss.
- If cervical tissue is present, monitoring is recommended as in cisgender females.
- Conduct annual breast examinations if mastectomy was performed. If mastectomy has not been performed, monitoring is recommended as in cisgender females (by mammography).

Adapted with permission from Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, *et al.* Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2017; Progestational agents. Copyright © 2017 Oxford University Press.

Studies examining the effects of the administration of androgens on insulin resistance and glucose intolerance in cisgender men showed conflicting results [59–62]. Despite this, there is currently no evidence for an increase in adverse cardiovascular outcomes in transgender men taking testosterone therapy on short and medium (10 years) follow-up time, although data on older transgender men are lacking [63–65]. However, a systematic review of the literature concluded that the level of evidence was too low to allow an interpretation of the risk of testosterone therapy on morbidity (stroke, myocardial infarction, venous thrombo-embolism) and mortality in transgender men [66]. Testosterone therapy does alter independent cardiac risk factors, including the lipid profile (decrease in high-density lipoprotein cholesterol and increase in triglycerides and low-density lipoprotein cholesterol), increasing inflammation parameters and (minimally) increasing systolic blood pressure and increasing serum haematocrit levels (to cisgender male reference values). To date, there is scarce evidence that these changes in risk factors result in altered morbidity [67–69]. The majority of the studies on transgender men report no changes in insulin resistance or diastolic blood pressure [30, 34, 64, 70], although one prospective study on 43 transgender men reported an increased incidence of metabolic syndrome after 1 (16.3%) and 2 years (18.6%) of testosterone ester therapy every 3 weeks [71]. One prospective study on 13 transgender men on testosterone esters every 2 weeks, reported diminished peripheral glucose uptake after 4 months [72]. The Endocrine Society guideline [18] suggests monitoring weight, blood pressure, and lipids at regular intervals. If cardiovascular risk factors emerge, they should be managed according to established population-based guidelines.

Testosterone therapy is associated with an increase in serum haemoglobin (+ 4.9–12.5% range) and haematocrit (+ 4.4–17.6% range) levels during the first year of treatment, with the most pronounced increase during the first three months (+2.7% Hct, 95% CI 1.94–3.29). Serum haematocrit levels remain stable after the first year of testosterone therapy [32, 38, 67]. As serum haematocrit levels can be found in the reference range of the perceived gender as from three months after the initiation of gender-affirming hormonal treatment, it is advised to consult the reference range for men in transgender men after the initiation of testosterone treatment [68]. Although theoretically, testosterone therapy increases the risk for erythrocytosis, clinically significant erythrocytosis is very rare [30, 67, 68]. In a large prospective study by the European Network for the Investigation of Gender Incongruence (ENIGI), which included 192 transgender men, the maximum measured haematocrit level was 54.0% and none of the transgender men experienced a thromboembolic event during follow-up. There are no reasons to assume that the observed mild increase in serum haematocrit levels is associated with an increased thrombotic risk in the short term [68]. Transgender men on testosterone undecanoate exhibit lower erythrocytosis rates compared to transgender men on testosterone esters or gel. Changing the treatment to testosterone undecanoate seems a valid option if the hormone prescribing physician and/or the patient are concerned about elevated serum haematocrit levels. This may prevent unnecessary interruptions in hormonal treatment [68]. The Endocrine Society guideline [18] suggests measuring haematocrit or haemoglobin at baseline and every 3 months for the first year and then one to two times a year.

Hepatologic Safety

An increase in liver enzymes can be observed during testosterone therapy [18], but this is mainly observed in transgender men taking oral 17-alkylated testosterone, which is no longer recommended [73, 74]. Two prospective studies reported no significant increase in liver enzymes in transgender men during gender-affirming hormone therapy [32, 75]. In addition, in the prospective study by Wierckx *et al.*, none of the 53 transgender men had liver enzyme values exceeding twice the upper limit according to male reference ranges and only 1.9% of the transgender men had liver enzyme values exceeding twice the upper limit according to female reference ranges, during one year of follow-up [30]. The Endocrine Society guideline [18] states that hepatotoxicity is not anticipated in transgender men taking parenteral or transdermal testosterone.

Bone Health

Many studies have assessed bone mineral density (BMD) in transgender men and results have recently been meta-analysed by Singh-Ospina *et al.* [76]. This meta-analysis revealed a similar BMD compared to cisgender females prior to testosterone therapy [76–79] and no decrease in BMD after the initiation of testosterone therapy [76, 77, 80]. There are two possible explanations for this: the aromatization of testosterone to oestradiol and the androgen-mediated increase in muscle mass, which stimulates bone formation according

to the Mechanostat theory [81]. In the study by Van Caenegem *et al.* [82], both muscle mass and strength were positively associated with trabecular and cortical parameters and bone size.

There are very limited data on the risk of osteoporotic fractures in transgender men [76]; only one cohort study of 53 transgender women and 53 transgender men reported hard outcomes such as fracture results. There were no fractures after 12 months of follow-up [30].

After ovariectomy, testosterone therapy becomes crucial to preserve BMD, as it prevents oestrogen-deprivation related bone loss in the short (<2 years) [34, 38, 77, 78, 80, 83–85] and long term (>10 years) [82, 86–88] and can even result in a prospective increase in BMD at cortical sites [84]. In addition, testosterone therapy also results in an increased cortical thickness [89] compared to postmenopausal cisgender women. However, bone loss has been described in transgender men who are not compliant with their testosterone therapy, who did not receive an adequate dose of testosterone or who stopped testosterone therapy completely [83, 86]. The Endocrine Society guideline [18] suggests screening for osteoporosis in those who stop testosterone treatment, are not compliant with hormone therapy, or who develop risks for bone loss.

Oncology

With more countries abandoning laws requiring gonadectomy for legal gender affirmation in transgender people, the number of transgender men who do not wish for hystero-oophorectomy is increasing. Both healthcare practitioners and transgender men have expressed concern about the impact of long-term testosterone therapy on its oncological risk. However, high-quality empirical data assessing cancer incidence and mortality among transgender people are lacking primarily because of an absence of large-scale prospective studies.

Breast Cancer

It is estimated that, in cisgender women, the incidence of breast carcinoma after prophylactic mastectomy is probably less than 2% [90]. In the current literature, seven cases of breast cancer in transgender men taking testosterone therapy have been reported [91–94]. In addition, Van Renterghem *et al.* [95] examined mastectomy specimens of 148 transgender men. They did not report invasive cancer, but apocrine metaplasia was seen in 23.6%, lactational changes in 2%, columnar cell changes in 37.2%, sclerosing adenosis in 4.7%, fibroadenoma in 4.1%, usual ductal hyperplasia in 27%, flat epithelial atypia in 0.7% and atypical ductal hyperplasia in 3.4%. In cisgender women, an increased risk of breast carcinoma has been reported in both pre- and postmenopausal women with higher serum testosterone levels. However, no causal relationship or association between testosterone therapy and the incidence of breast carcinoma in transgender men has been found and routine histopathological examination of mastectomy specimens may assess whether there is an increased risk of breast carcinoma in transgender men, compared to cisgender women [95]. Gooren *et al.* even found a lower incidence rate of breast cancer in transgender men, compared to cisgender females (5.9 versus 155 per 100 000 person-years) [91]. It should also be noted that breast cancer can still occur after mastectomy has been performed, thus adequate screening is still necessary postmastectomy [92, 94]. To date, no large studies on breast cancer risk have been performed in this population. The Endocrine Society [18] recommends conducting sub- and periareolar annual breast examinations if mastectomy has been performed. If mastectomy is not performed, then mammography should be considered, as recommended by the American Cancer Society.

Cervical Cancer

In cisgender women, the presence of human papilloma virus (HPV) increases the risk for cervical cancer [96]. However, it is unsure whether serum testosterone is related to cervical cancer in cisgender women; in premenopausal women, a positive correlation was found between free testosterone and higher risk of invasive cervical carcinoma, whereas in postmenopausal women, there was a positive correlation with total testosterone [97]. Whether testosterone therapy is associated with an increased risk for cervical cancer in transgender men remains unknown to date. In the literature, only two case reports of transgender men diagnosed with cervical cancer have been published [98, 99]. Both cases were diagnosed upon gender-affirming surgery. If cervical tissue is present, the Endocrine Society guideline [18] suggests monitoring as recommended by the American College of Obstetricians and Gynecologists [100].

Endometrial and Ovarian Cancer

Observational studies have reported an association between endometrial or ovarian cancer and higher total serum testosterone levels in postmenopausal women. The proposed explanation is twofold: testosterone can be aromatized into oestradiol, which results in lower testosterone levels and an increased in oestrogen levels, which may stimulate endometrial and ovarian epithelium proliferation, and testosterone can also be converted into dihydrotestosterone (DHT) by 5α-reductase. DHT and testosterone can both increase endometrial and ovarian epidermal growth. However, as DHT also has antiproliferative effects on the endometrial and ovarian cancer cells, higher testosterone levels may also lead to a decreased cancer risk [97]. In addition, cisgender women with polycystic ovary syndrome (PCOS) are at higher risk for endometrial cancer due to androgen-mediated endometrial epithelium proliferation [97]. Whether testosterone therapy induces PCOS morphology of the ovarian cortex in transgender men remains unsure, as results are contradictory [101, 102]. In the current literature, only one case of endometrial cancer in a transgender man after 7 years of testosterone therapy [98] and three cases of ovarian cancer in transgender men receiving testosterone therapy have been reported, of which two cases were tested for the endometrial epidermal growth factor receptor; both were positive [103, 104].

However, in Europe, in the past, many transgender men included in published studies underwent hystero-oophorectomy after 12–18 months of gender-affirming hormone therapy, which reduces the cohort of transgender men at risk for endometrial and ovarian cancers and increases the need for large long-term cohort studies on cancer risk in transgender men who do not undergo this surgery.

Other Cancers

To date, only eight published papers assessed cancer incidence or mortality in transgender men, these articles examined different populations (the Netherlands, Belgium, Sweden, and the United States) in different time periods (1989–2016) [24, 28, 65, 91, 105–107].

Although large representative prospective studies of transgender populations across multiple clinical sites are not available to date,

there is evidence that transgender people are more frequently exposed to cancer risk factors such as smoking, obesity and lack of/inadequate cancer screening due to perceived barriers for accessing healthcare [97, 108]. These factors result in transgender people less frequently accessing healthcare for preventive care, such as cervical smears in transgender men [109]. However, to date, the number of cancers reported in transgender men is relatively low compared to transgender women [105, 106] which may be due to the younger age of transgender men in the majority of the published articles. However, as long as large-scale epidemiological data on long-term cancer risk in transgender people has not been published, it is advised to adhere to the general screening protocols.

Psychological Monitoring

Depression, suicide attempts and anxiety rates in the transgender sample are high (36–63%, 32%, and 33–47%, respectively) [110–113]. Avoidant coping is more frequent in the beginning of the gender-affirming process [112], which in turn can increase symptoms of depression and anxiety. In addition, social support is directly and indirectly related to depression and anxiety [112]. Therefore, it is important to enquire about psychosocial well-being during each contact and to refer to a psychologist/psychiatrist if necessary. If additional risk factors are present, the WPATH Standards of Care [19] warn for a possible increased risk of destabilization of certain psychiatric disorders (including bipolar, schizoaffective, and other disorders that may include manic or psychotic symptoms) in transgender men on too high dosage of testosterone. In addition, the WPATH Standards of Care also warn for aggression or expansive mood at the beginning of an injection cycle with testosterone esters [19]. However, in transgender men who adhere to the prescribed dose of testosterone, a decrease of coexistent psychopathology has been observed upon the initiation of gender-affirming hormone therapy [114] and the absence of increase in aggression has been observed [115].

Other Considerations

Fertility Preservation

Hysterectomy and bilateral oophorectomy in transgender men lead to an irreversible loss of natural reproductive capacity; therefore, it is important for healthcare professionals to discuss reproductive options beforehand [116]. As many transgender men desire children [117], it is important for healthcare providers to get acquainted with the fertility needs of transgender men and to be able to discuss reproductive options at each stage of the gender-affirming process and the impact of hormonal therapy on fertility. There have been case reports of transgender men who became pregnant after stopping testosterone therapy [118] and it is possible to stop testosterone therapy in order to preserve germ cells, but most transgender men experienced this as emotionally difficult [36]. Therefore, we advise freezing of germ cells prior to the initiation of gender-affirming therapy if there is a genetically related parental desire. If germ cells have been frozen, it may remain possible for transgender men to have genetically related children, even after gonadectomy. However, the preservation of oocytes, ovarian tissue, or embryos implicates extra costs for the retrieval procedure and an annual cost for the preservation, financed by the patient. Further details regarding fertility options in trans men are discussed in De Roo *et al.* [119] and Chapter 11.5.

REFERENCES

1. Wylie K, Knudson G, Khan SI, Bonierbale M, Watanyusakul S, Baral S. Serving transgender people: clinical care considerations and service delivery models in transgender health. *Lancet*, 2016; **388**: 401–11.
2. Kuyper L, Wijsen C. Gender identities and gender dysphoria in the Netherlands. *Arch Sex Behav*, 2014; **43**: 377–85.
3. Cruz TM. Assessing access to care for transgender and gender nonconforming people: a consideration of diversity in combating discrimination. *Soc Sci Med*, 2014; **110**: 65–73.
4. Snelgrove JW, Jasudavisius AM, Rowe BW, Head EM, Bauer GR. Completelyout-at-sea with two-gender medicine: a qualitative analysis of physician-side barriers to providing healthcare for transgender patients. *BMC Health Serv Res*, 2012; **12**: 110.
5. Sequeira GM, Chakraborti C, Panunti BA. Integrating lesbian, gay, bisexual, and transgender (LGBT) content into undergraduate medical school curricula: a qualitative study. *Ochsner J*, 2012; **12**: 379–82.
6. Bockting W, Robinson B, Benner A, Scheltema K. Patient satisfaction with transgender health services. *J Sex Marital Ther*, 2004; **30**: 377.
7. Fisher AD, Castellini G, Ristori J, *et al*. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *J Clin Endocrinol Metab*, 2016; **101**: 4260–9.
8. Gorin-Lazard A, Baumstarck K, Boyer L, *et al*. Is hormonal therapy associated with better quality of life in transsexuals? A cross-sectional study. *J Sex Med*, 2012; **9**: 531–41.
9. Fisher AD, Castellini G, Bandini E, *et al*. Cross-sex hormonal treatment and body uneasiness in individuals with gender dysphoria. *J Sex Med*, 2014; **11**: 709–19.
10. Smith YLS, Van Goozen SHM, Kuiper AJ, Cohen-Kettenis PT. Sex reassignment: outcomes and predictors of treatment for adolescent and adult transsexuals. *Psychol Med*, 2005; **35**: 89–99.
11. Newfield E, Hart S, Dibble S, Kohler L. Female-to-male transgender quality of life. *Qual Life Res*, 2006; **15**: 1447–57.
12. Murad MH, Elamin MB, Garcia MZ, *et al*. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*, 2010; **72**: 214–31.
13. Wierckx K, Elaut E, Van Hoorde B, *et al*. Sexual desire in trans persons: associations with sex reassignment treatment. *J Sex Med*, 2014; **11**: 107–18.
14. Rakic Z, Starcevic V, Maric J, Kelin K. The outcome of sex reassignment surgery in Belgrade: 32 patients of both sexes. *Arch Sex Behav*, 1996; **25**: 515–25.
15. Ainsworth TA, Spiegel JH. Quality of life of individuals with and without facial feminization surgery or gender reassignment surgery. *Qual Life Res*, 2010; **19**: 1019–24.
16. Gómez-Gil E, Zubiaurre-Elorza L, Esteva I, *et al*. Hormone-treated transsexuals report less social distress, anxiety and depression. *Psychoneuroendocrinology*, 2012; **37**: 662–70.
17. De Cuypere G, T'Sjoen G, Beerten R, *et al*. Sexual and physical health after sex reassignment surgery. *Arch Sex Behav*, 2005; **34**: 679–90.

18. Hembree WC, Cohen-Kettenis PT, Gooren L, *et al*. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 3869–903.
19. WPATH. WPATH standards of care. *Int J Transgenderism*, 2012; **13**: 165–232.
20. Defreyne J, Motmans J, T'Sjoen G. Healthcare costs and quality of life outcomes following gender affirming surgery in trans men: a review. *Expert Rev Pharmacoecon Outcomes Res*, 2017; **17**: 543–56.
21. Kuhn A, Bodmer C, Stadlmayr W, Kuhn P, Mueller MD, Birkhäuser M. Quality of life 15 years after sex reassignment surgery for transsexualism. *Fertil Steril*, 2009; **92**: 1685–9.e3.
22. Jones BA, Haycraft E, Murjan S, Arcelus J. Body dissatisfaction and disordered eating in trans people: a systematic review of the literature. *Int Rev Psychiatry*, 2016; **28**: 81–94.
23. Murad MH, Elamin MB, Garcia MZ, *et al*. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*, 2010; **72**: 214–31.
24. Wierckx K, Mueller S, Weyers S, *et al*. Long-term evaluation of cross-sex hormone treatment in transsexual persons. *J Sex Med*, 2012; **9**: 2641–51.
25. Weinand JD, Safer JD. Hormone therapy in transgender adults is safe with provider supervision: a review of hormone therapy sequelae for transgender individuals. *J Clin Transl Endocrinol*, 2015; **2**: 55–60.
26. Gorton RN, Erickson-Schroth L. Hormonal and surgical treatment options for transgender men (female-to-male). *Psychiatr Clin North Am*, 2017; **40**: 79–97.
27. Trum HW, Hoebeke P, Gooren LJ. Sex reassignment of transsexual people from a gynecologist's and urologist's perspective. *Acta Obstet Gynecol Scand*, 2015; **94**: 563–7.
28. Asscheman H, Giltay EJ, Megens JAJ, van Trotsenburg MAA, Gooren LJG. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *Eur J Endocrinol*, 2011; **164**: 635–42.
29. Cavanaugh T, Hopwood R, Lambert C. Informed consent in the medical care of transgender and gender-nonconforming patients. *AMA J Ethics*, 2016; **18**: 1147–55.
30. Wierckx K, Van Caenegem E, Schreiner T, *et al*. Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. *J Sex Med*, 2014; **11**: 1999–2011.
31. Olson J, Schrager SM, Clark LF, Dunlap SL, Belzer M. Subcutaneous testosterone: an effective delivery mechanism for masculinizing young transgender men. *LGBT Heal*, 2014; **1**: 165–7.
32. Jacobeit JW, Gooren LJ, Schulte HM. Safety aspects of 36 months of administration of long-acting intramuscular testosterone undecanoate for treatment of female-to-male transgender individuals. *Eur J Endocrinol*, 2009; **161**: 795–8.
33. Spratt DI, Stewart II, Savage C, *et al*. Subcutaneous injection of testosterone is an effective and preferred alternative to intramuscular injection: demonstration in female-to-male transgender patients. *J Clin Endocrinol Metab*, 2017; **102**: 2349–55.
34. Pelusi C, Costantino A, Martelli V, *et al*. Effects of three different testosterone formulations in female-to-male transsexual persons. *J Sex Med*, 2014; **11**: 3002–11.
35. Mitu K. Transgender reproductive choice and fertility preservation. *AMA J Ethics*, 2016; **18**: 1120.
36. Armuand G, Dhejne C, Olofsson JI, Rodriguez-Wallberg KA. Transgender men's experiences of fertility preservation: a qualitative study. *Hum Reprod*, 2017; **32**: 383–90.
37. Gooren LJ. Management of female-to-male transgender persons: medical and surgical management, life expectancy. *Curr Opin Endocrinol Diabetes Obes*, 2014; **21**: 233–8.
38. Mueller A, Haeberle L, Zollver H, *et al*. Effects of intramuscular testosterone undecanoate on body composition and bone mineral density in female-to-male transsexuals. *J Sex Med*, 2010; **7**: 3190–8.
39. Giltay EJ, Gooren LJG. Effects of sex steroid deprivation/administration on hair growth and skin sebum production in transsexual males and females. *J Clin Endocrinol Metab*, 2000; **85**: 2913–21.
40. Wierckx K, Van de Peer F, Verhaeghe E, *et al*. Short-and long-term clinical skin effects of testosterone treatment in trans men. J Sex Med, 2014; **11**: 222–9.
41. Moreno-Arrones OM, Becerra A, Vano-Galvan S. Therapeutic experience with oral finasteride for androgenetic alopecia in female-to-male transgender patients. *Clin Exp Dermatol*, 2017; **42**: 743–8.
42. Evans S, Neave N, Wakelin D, Hamilton C. The relationship between testosterone and vocal frequencies in human males. *Physiol Behav*, 2008; **93**: 783–8.
43. Bultynck C, Pas C, Defreyne J, Cosyns M, den Heijer M, T'Sjoen G. Self-perception of voice in transgender persons during cross-sex hormone therapy. *Laryngoscope*, 2017; **127**: 2796–804.
44. Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice*, 1999; **13**: 424–46.
45. Cosyns M, Borsel J, Wierckx K, *et al*. Voice in female-to-male transsexual persons after long-term androgen therapy. *Laryngoscope*, 2014; **124**: 1409–14.
46. Watt SO, Tskhay KO, Rule NO. Masculine voices predict well-being in female-to-male transgender individuals. *Arch Sex Behav*, 2017; **127**: 2796–804.
47. Schlatterer K, Von Werder K, Stalla GK. Multistep treatment concept of transsexual patients. *Exp Clin Endocrinol Diabetes*, 1996; **104**: 413–19.
48. Leiblum S, Bachmann G, Kemmann E, Colburn D, Swartzman L. Vaginal atrophy in the postmenopausal woman. *JAMA*, 1983; **249**: 1983–2198.
49. Nakamura A, Watanabe M, Sugimoto M, *et al*. Dose-response analysis of testosterone replacement therapy in patients with female to male gender identity disorder. *Endocr J*, 2013; **60**: 275–81.
50. Allan C a, Forbes E a, Strauss BJG, McLachlan RI. Testosterone therapy increases sexual desire in ageing men with low-normal testosterone levels and symptoms of androgen deficiency. *Int J Impot Res*, 2008; **20**: 396–401.
51. Costantino A, Cerpolini S, Alvisi S, Morselli PG, Venturoli S, Meriggiola MC. A prospective study on sexual function and mood in female-to-male transsexuals during testosterone administration and after sex reassignment surgery. *J Sex Marital Ther*, 2013; **39**: 321–35.
52. Buster JE, Kingsberg SA, Aguirre O, *et al*. Testosterone patch for low sexual desire in surgically menopausal women: a randomized trial. *Obstet Gynecol*, 2005; **105**(5 Pt 1): 944–52.
53. Heiman JR. Treating low sexual desire—new findings for testosterone in women. *N Engl J Med*, 2008; **359**: 2047–9.

54. Wierckx K, Elaut E, Van Caenegem E, *et al*. Sexual desire in female-to-male transsexual persons: exploration of the role of testosterone administration. *Eur J Endocrinol*, 2011; **165**: 331–7.
55. Bartolucci C, Gómez-Gil E, Salamero M, *et al*. Sexual quality of life in gender-dysphoric adults before genital sex reassignment surgery. *J Sex Med*, 2015; **12**: 180–8.
56. Shadid S. Fatty acid and glucose metabolism in upper body obesity: effects of diet and exercise compared to pioglitazone administration. Maastricht University; 2007.
57. Jousilahti P, Vartiainen E, Tuomilehto J, Puska P. Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation*, 1999; **99**: 1165–72.
58. Livingstone C, Collison M. Sex steroids and insulin resistance. *Clin Sci*, 2002; **102**: 151–66.
59. Cohen JC, HICKMAN R. Insulin resistance and diminished glucose tolerance in powerlifters ingesting anabolic steroids. *J Clin Endocrinol Metab*, 1987; **64**: 960–3.
60. Mårin P, Krotkiewski M, Björntorp P. Androgen treatment of middle-aged, obese men: effects on metabolism, muscle and adipose tissues. *Eur J Med*, 1992; **1**: 329–36.
61. Singh AB, Hsia S, Alaupovic P, *et al*. The effects of varying doses of T on insulin sensitivity, plasma lipids, apolipoproteins, and C-reactive protein in healthy young men. *J Clin Endocrinol Metab*, 2002; **87**: 136–43.
62. Elbers JMH, Giltay EJ, Teerlink T, *et al*. Effects of sex steroids on components of the insulin resistance syndrome in transsexual subjects. *Clin Endocrinol (Oxf)*, 2003; **58**: 562–71.
63. Gooren LJ, Wierckx K, Giltay EJ. Cardiovascular disease in transsexual persons treated with cross-sex hormones: reversal of the traditional sex difference in cardiovascular disease pattern. *Eur J Endocrinol*, 2014; **170**: 809–19.
64. Maraka S, Singh-Ospina N, Rodriguez-Gutierrez R, *et al*. Sex steroids and cardiovascular outcomes in transgender individuals: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2017; **102**: 3914–23.
65. Van Kesteren PJM, Asscheman H, Megens JAJ, Gooren LJG. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. *Clin Endocrinol (Oxf)*, 1997; **47**: 337–43.
66. Elamin MB, Garcia MZ, Murad MH, Erwin PJ, Montori VM. Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses. *Clin Endocrinol (Oxf)*, 2010; **72**: 1–10.
67. Velho I, Fighera TM, Ziegelmann PK, Spritzer PM. Effects of testosterone therapy on BMI, blood pressure, and laboratory profile of transgender men: a systematic review. *Andrology*, 2017; **5**: 881–8.
68. Defreyne J, Vantomme B, Van Caenegem E, *et al*. Prospective evaluation of hematocrit in gender-affirming hormone treatment: results from European Network for the Investigation of Gender Incongruence. *Andrology*, 2018; **6**: 446–54.
69. Gooren LJ, Giltay EJ. Men and women, so different, so similar: observations from cross-sex hormone treatment of transsexual subjects. *Andrologia*, 2014; **46**: 570–5.
70. Berra M, Armillotta F, D'emidio L, *et al*. Testosterone decreases adiponectin levels in female to male transsexuals. *Asian J Androl*, 2006; **8**: 725–9.
71. Colizzi M, Costa R, Scaramuzzi F, *et al*. Concomitant psychiatric problems and hormonal treatment induced metabolic syndrome in gender dysphoria individuals: a 2 year follow-up study. *J Psychosom Res*, 2015; **78**: 399–406.
72. Polderman KH, Gooren LJ, Asscheman H, Bakker A, Heine RJ. Induction of insulin resistance by androgens and estrogens. *J Clin Endocrinol Metab*, 1994; **79**: 265–71.
73. Bird D, Vowles K, Anthony PP. Spontaneous rupture of a liver cell adenoma after long term methyltestosterone: report of a case successfully treated by emergency right hepatic lobectomy. *BJS*, 1979; **66**: 212–3.
74. Westaby D, Paradinas FJ, Ogle SJ, Randell JB, Murray-Lyon I. Liver damage from long-term methyltestosterone. *Lancet*, 1977; **310**: 261–3.
75. Mueller A, Kiesewetter F, Binder H, Beckmann MW, Dittrich R. Long-term administration of testosterone undecanoate every 3 months for testosterone supplementation in female-to-male transsexuals. *J Clin Endocrinol Metab*, 2007; **92**: 3470–5.
76. Singh-Ospina N, Maraka S, Rodriguez-Gutierrez R, *et al*. Effect of sex steroids on the bone health of transgender individuals: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2017; **102**: 3904–13.
77. Van Caenegem E, Wierckx K, Taes Y, *et al*. Body composition, bone turnover, and bone mass in trans men during testosterone treatment: 1-year follow-up data from a prospective case-controlled study (ENIGI). *Eur J Endocrinol*, 2015; **172**: 163–71.
78. Haraldsen IR, Haug E, Falch J, Egeland T, Opjordsmoen S. Cross-sex pattern of bone mineral density in early onset gender identity disorder. *Horm Behav*, 2007; **52**: 334–43.
79. Miyajima T, Kim YT, Oda H. A study of changes in bone metabolism in cases of gender identity disorder. *J Bone Miner Metab*, 2012; **30**: 468–73.
80. Wiepjes CM, Vlot MC, Klaver M, *et al*. Bone mineral density increases in trans persons after 1 year of hormonal treatment: a multicenter prospective observational study. *J Bone Miner Res*, 2017; **32**: 1252–60.
81. Frost HM. Bone's mechanostat: a 2003 update. *Anat Rec*, 2003; **275**: 1081–101.
82. Van Caenegem E, Wierckx K, Taes Y, *et al*. Bone mass, bone geometry, and body composition in female-to-male transsexual persons after long-term cross-sex hormonal therapy. *J Clin Endocrinol Metab*, 2012; **97**: 2503–11.
83. Goh HH V, Ratnam SS. Effects of hormone deficiency, androgen therapy and calcium supplementation on bone mineral density in female transsexuals. *Maturitas*, 1997; **26**: 45–52.
84. Turner A, Chen TC, Barber TW, Malabanan AO, Holick MF, Tangpricha V. Testosterone increases bone mineral density in female-to-male transsexuals: a case series of 15 subjects. *Clin Endocrinol (Oxf)*, 2004; **61**: 560–6.
85. Meriggiola MC, Armillotta F, Costantino A, *et al*. Effects of testosterone undecanoate administered alone or in combination with letrozole or dutasteride in female to male transsexuals. *J Sex Med*, 2008; **5**: 2442–53.
86. Van Kesteren P, Lips P, Gooren LJG, Asscheman H, Megens J. Long-term follow-up of bone mineral density and bone metabolism in transsexuals treated with cross-sex hormones. *Clin Endocrinol (Oxf)*, 1998; **48**: 347–54.
87. Ruetsche AG, Kneubuehl R, Birkhaeuser MH, Lippuner K. Cortical and trabecular bone mineral density in transsexuals after long-term cross-sex hormonal treatment: a cross-sectional study. *Osteoporos Int*, 2005; **16**: 791–8.
88. Schlatterer K, Auer DP, Yassouridis A, Von Werder K, Stalla GK. Transsexualism and osteoporosis. *Exp Clin Endocrinol Diabetes*, 1998; **106**: 365–8.

89. Lips P, Van Kesteren PJ, Asscheman H, Gooren LJG. The effect of androgen treatment on bone metabolism in female-to-male transsexuals. *J Bone Miner Res*, 1996; **11**: 1769–73.
90. Willemsen HW, Kaas R, Peterse JH, Rutgers EJT. Breast carcinoma in residual breast tissue after prophylactic bilateral subcutaneous mastectomy. *Eur J Surg Oncol*, 1998; **24**: 331–2.
91. Gooren L, Bowers M, Lips P, Konings IR. Five new cases of breast cancer in transsexual persons. *Andrologia*, 2015; **47**: 1202–5.
92. Nikolic D V, Djordjevic ML, Granic M, *et al*. Importance of revealing a rare case of breast cancer in a female to male transsexual after bilateral mastectomy. *World J Surg Oncol*, 2012; **10**: 280.
93. Burcombe RJ, Makris A, Pittam M, Finer N. Breast cancer after bilateral subcutaneous mastectomy in a female-to-male transsexual. *The Breast*, 2003; **12**: 290–3.
94. Shao T, Grossbard ML, Klein P. Breast cancer in female-to-male transsexuals: two cases with a review of physiology and management. *Clin Breast Cancer*, 2011; **11**: 417–19.
95. Van Renterghem S, Van Dorpe J, T'Sjoen G, Monstrey S, Verbeke S, van Bockstal M. Histopathology of transgender breast specimens: expect the unexpected! *Virchows Arch*, 2017; **471**: S60–S60.
96. Gillison ML, Chaturvedi AK, Lowy DR. HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. *Cancer*, 2008; **113**: 3036–46.
97. Braun H, Nash R, Tangpricha V, Brockman J, Ward K, Goodman M. Cancer in transgender people: evidence and methodological considerations. *Epidemiol Rev*, 2017; **39**: 93–107.
98. Urban RR, Teng NNH, Kapp DS. Gynecologic malignancies in female-to-male transgender patients: the need of original gender surveillance. *Am J Obstet Gynecol*, 2011; **204**: e9–12.
99. Driák D, Samudovsky M. Could a man be affected with carcinoma of cervix? The first case of cervical carcinoma in trans-sexual person (FtM)—case report. *ACTA Medica (Hradec Kralove)*, 2005; **48**: 53.
100. Saslow D, Solomon D, Lawson HW, *et al*. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin*, 2012; **62**: 147–72.
101. Grynberg M, Fanchin R, Dubost G, *et al*. Histology of genital tract and breast tissue after long-term testosterone administration in a female-to-male transsexual population. *Reprod Biomed Online*, 2010; **20**: 553–8.
102. Ikeda K, Baba T, Noguchi H, *et al*. Excessive androgen exposure in female-to-male transsexual persons of reproductive age induces hyperplasia of the ovarian cortex and stroma but not polycystic ovary morphology. *Hum Reprod*, 2012; **28**: 453–61.
103. Dizon DS, Tejada-Berges T, Koelliker S, Steinhoff M, Granai CO. Ovarian cancer associated with testosterone supplementation in a female-to-male transsexual patient. *Gynecol Obstet Invest*, 2006; **62**: 226–8.
104. Hage JJ, Dekker J, Karim RB, Verheijen RHM, Bloemena E. Ovarian cancer in female-to-male transsexuals: report of two cases. *Gynecol Oncol*, 2000; **76**: 413–15.
105. Wierckx K, Elaut E, Declercq E, *et al*. Prevalence of cardiovascular disease and cancer during cross-sex hormone therapy in a large cohort of trans persons: a case-control study. *Eur J Endocrinol*, 2013; **169**: 471–8.
106. Dhejne C, Lichtenstein P, Boman M, Johansson AL V, Långström N, Landén M. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. *PLoS One*, 2011; **6**: e16885.
107. Brown GR, Jones KT. Incidence of breast cancer in a cohort of 5,135 transgender veterans. *Breast Cancer Res Treat*, 2015; **149**: 191–8.
108. Fredriksen-Goldsen KI, Cook-Daniels L, Kim H-J, *et al*. Physical and mental health of transgender older adults: an at-risk and underserved population. *Gerontologist*, 2013; **54**: 488–500.
109. Peitzmeier SM, Khullar K, Reisner SL, Potter J. Pap test use is lower among female-to-male patients than non-transgender women. *Am J Prev Med*, 2014; **47**: 808–12.
110. Clements-Nolle K, Marx R, Katz M. Attempted suicide among transgender persons: the influence of gender-based discrimination and victimization. *J Homosex*, 2006; **51**: 53–69.
111. Bockting WO, Miner MH, Swinburne Romine RE, Hamilton A, Coleman E. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health*, 2013; **103**: 943–51.
112. Budge SL, Adelson JL, Howard KAS. Anxiety and depression in transgender individuals: the roles of transition status, loss, social support, and coping. *J Consult Clin Psychol*, 2013; **81**: 545.
113. Reisner SL, Poteat T, Keatley J, *et al*. Global health burden and needs of transgender populations: a review. *Lancet*, 2016; **388**: 412–36.
114. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. *J Sex Med*, 2014; **11**: 119–26.
115. Defreyne J, T'Sjoen G, Bouman WP, Brewin N, Arcelus J. Prospective evaluation of self-reported aggression in transgender persons. *J Sex Med*, 2018; **15**: 768–76.
116. Selvaggi G, Dhejne C, Landén M, Elander A. The 2011 WPATH Standards of Care and penile reconstruction in female-to-male transsexual individuals. *Adv Urol*, 2012; **2012**: 581712.
117. Wierckx K, Van Caenegem E, Pennings G, *et al*. Reproductive wish in transsexual men. *Hum Reprod*, 2011; **27**: 483–7.
118. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. *Obstet Gynecol*, 2014; **124**: 1120–7.
119. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people Fertility options in transgender people. *Int Rev Psychiatry*, 2016; **28**: 112–19.

11.5

Fertility Options for Transgender Persons

Chloë De Roo and Guy T'Sjoen

Introduction

Persons with gender dysphoria inherently possess normal reproductive capacity [1]. Treatment options to facilitate the transition and to live in accordance with the identified gender often comprises both hormone therapy and/or surgical interventions [2]. Unfortunately, both these options have a negative effect on fertility [2, 3]. Healthcare professionals should address the consequences for future fertility with their patients before treatments have started.

Although there are only few studies that investigated the desire of transgender people to have children, they all conclude that approximately half of transgender women and half of transgender men wish to have children [4–6].

The relevance of this topic is also reflected in the fact that transgender people with children score significantly better on mental health and vitality scores than transgender people without children [4]. In transgender women, parenting was even identified as a suicide protective factor [7]. Apart from an apparent child wish, a small majority of transgender men and transgender women would actually have cryopreserved their gametes, or would have seriously considered doing it, if the technique would have been available [4]. Especially lesbian and bisexual transgender women were interested in using their own cryopreserved sperm to fulfil a future child wish [4]. Regardless of their personal desire, the majority of transgender people clearly expressed the opinion that fertility preservation techniques should be discussed and offered [6].

On the contrary and quite striking is the fact that some of the transgender patients are willing to completely sacrifice their fertility, since fertility preservation was considered not important enough to postpone their transition process [6] as most of transgender people are in favour of a fast transition [5]. Also cost, physical discomfort and the invasiveness of the procedures were identified as burdens for fertility preservations [8, 9]. However, decisional regret and quality of life following declining fertility preservation needs to be further explored [8].

The World Professional Association for Transgender Health (WPATH) Standards of Care recommend in their seventh version to discuss fertility options with patients prior to any treatment or medical intervention, especially before gender-affirming surgery [2]. The impact of each treatment on fertility as well as a possible fertility preservation option to maintain the possibility of having future genetically related children, should be addressed.

The Effect of Gender-Affirming Therapy on Fertility

Gender-Affirming Surgery

Gender-affirming surgery with penectomy and orchidectomy or hysterectomy with bilateral oophorectomy leads to irreversible sterility [4, 10].

Pubertal Suppression

Gonadotropin releasing agonists (GnRH-a) preventing development of secondary sexual characteristics, inhibit germ cell maturation [11]. However, discontinuation results in normal pubertal progression [1, 12]. Hagen *et al.* (2012) showed that GnRH-a administration in girls with precocious puberty resulted in a significant decline of anti-Müllerian hormone (AMH—a marker for ovarian follicle reserve) after 3 months with recovery to pretreatment AMH serum levels 6 months after interruption of the treatment [11].

Hormone Treatment Therapy in Transgender Women

Clinically, prolonged oestrogen treatment results in reduction of the testicular volume [13]. Additionally, oestrogens have a suppressive effect on sperm motility and density in a (cumulative)

dose-dependent manner [13]. Hamada *et al.* (2014) clearly demonstrated a poor semen quality in transgender women following feminizing therapy. Their results show a high incidence of oligozoosperma, asthenozoospermia, and teratozoospermia [7, 14]. If semen samples of such poor quality are used, assisted reproduction techniques, such as *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI) are needed [15].

Payer *et al.* (1979) described the histological effects on Leydig cells of treatment with oestrogens alone or with medroxyprogesterone acetate [13]. Three morphologies were distinguished: Leydig cells very similar to untreated Leydig cells (type 1); the absence of Leydig cells in presence of cells with increased microfilaments and abundant smooth endoplasmic reticulum and lipid drops (type 2); and complete absence of any cell type but with varying amounts of microfilaments and pigmentation (type 3) [13]. The persisting spermatogonia showed the typical features of the so-called pale type-A spermatogonia [16]. Because of the low mitotic rate of these cells, the pale type-A spermatogonia survive disturbances of the endocrine balance, as well as other noxious stimuli such as radio- and chemotherapy [16]. Therefore, these cells are regarded as the stem cells of the human testis [16].

These histological findings explain why feminizing hormonal therapy induces hypospermatogenesis, ultimately leading to azoospermia [10]. This impact is—to a certain extent—heterogenous possibly due to different types of cross-sex hormone therapy [17] and reversible upon cessation of oestrogen therapy based on the presence of the spermatogonial stem cells [16]. Thresholds for hormone treatment in transgender women to avoid permanent reproductive impairment have not been established [1].

Androgen Therapy in Transgender Men

Masculinizing hormonal therapy will in most cases lead to a reversible amenorrhea but ovarian follicles are not depleted from the tissue [18, 19]. Increased androgen levels may however adversely affect follicle growth, mostly affecting the more matured follicle stages as these follicles are surrounded by an androgen-sensitive theca cell layer [19–21]. Apparently, primordial follicles in cryopreserved and xenotransplanted ovarian cortical tissue do not lose their potential to resume growth and maturation [22].

An ongoing debate is still ongoing whether or not masculinizing therapy induces polycystic ovarian syndrome (PCOS) [20, 21, 23]. Descriptive histology by Pache *et al.* (1991) showed some evidence for altered morphology induced by testosterone treatment comprising of enlarged ovaries, collagenized ovarian cortex, theca interna hyperplasia, and stromal hyperplasia [20, 24]. The idea that androgens induce high AMH levels, which is a particular feature of PCOS, is however questionable [25]. Moreover, a (3D) transvaginal ultrasound study in trans men treated with long-term testosterone therapy showed no increased polycystic ovarian morphology [26].

One must realize that hormonal interventions do not exclude possible pregnancy in transgender men [27, 28]. This implies that exogenous testosterone is not an adequate mean of birth control. Testosterone has teratogenic effects on the fetus, therefore transgender men must avoid pregnancy while on testosterone therapy. In the absence of surgery, contraception should be discussed. Data on the choice of the correct contraceptive are currently lacking. Progesterone-only pills, a (progesterone-releasing) intrauterine device or barrier methods of birth control in order to avoid the use of oestrogen treatment are suggested [29].

The Current Fertility Preservation Options for Transgender Men

Current fertility preservation options include cryopreservation of embryos, oocytes, or ovarian tissue. The theoretical options are presented in Table 11.5.1, including a short description of every technique, considerations in favour of and against the technique and the potential use in case of a (future) male or a female partner. These options are based on the known fertility preservation options for patients undergoing gonadotoxic treatment [30] transposed to the specific needs of transgender patients. It is preferable to bank gametes before using hormones. For transgender men and transgender women already using hormones, an interruption of hormone treatment is recommended at least 3 months to restore possible therapy-induced effects.

Embryo Cryopreservation

The embryo cryopreservation technique consists of a hormonal stimulation and oocyte aspiration for IVF/ICSI techniques to create embryos that are subsequently cryopreserved for future embryo transfer [15, 31]. This is an appropriate option for post-pubertal transgender men with a male partner and offers the possibility of a genetically related child. However, if desired, a sperm donor can also be used to create embryos [31]. This fertility preservation method requires a controlled ovarian stimulation and thus female hormone exposure [31]. Additionally, frequent vaginal ultrasound monitoring is needed during the ovarian stimulation phase and a transvaginal surgical procedure is performed for oocyte aspiration [10, 31]. This can be a physical and psychological burden for many transgender men, hereby limiting their future reproductive options [4]. Although technically possible, genital exams in transgender men may need to be postponed until a well trusted doctor-patient relationship is established [32].

Nowadays embryo freezing is a routine procedure in the field of assisted reproduction. In case of an embryo transfer, a recipient uterus (surrogacy mother) is required, especially when the uterus has been removed in the transgender man [10, 15]. In case of a female partner, partner donation—where an oocyte of the transgender man is inseminated with donor sperm and subsequently transferred to the female partner—is a possibility.

Oocyte Cryopreservation

Oocyte cryopreservation includes hormonal stimulation and oocyte retrieval for oocyte vitrification [31]. Therefore, the same considerations as for controlled ovarian stimulation for embryo cryopreservation have to be taken into account. Cryopreservation of oocytes does not require fertilization before cryopreservation. Therefore, there is no need for a partner or for the use of donor sperm at the moment of the fertility preservation procedure. Again it is an established technique and future use of the cryopreserved oocytes requires the use of partner sperm or donor sperm and a recipient uterus which can be the female partner or a surrogate if a male partner is present [10].

Table 11.5.1 Fertility preservation options in transgender men

Technique	Description	Considerations	Future use
Embryo cryopreservation	Controlled ovarian stimulation for oocyte retrieval and fertilization to obtain embryo's for cryopreservation	• Established method • Controlled ovarian stimulation • Vaginal procedure • Postpubertal • Partner or donor sperm	*Male partner* Use of partner's sperm prior to cryopreservation, need of a surrogate mother *Female partner* Fertilization by donor sperm prior to cryopreservation, implantation into the partner's uterus
Oocyte cryopreservation	Controlled ovarian stimulation to obtain oocytes for cryopreservation	• Established method • Controlled ovarian stimulation • Vaginal ultrasounds • Postpubertal • No partner required	*Male partner* Use of partner's sperm, need of a recipient uterus (7) *Female partner* Fertilization by donor sperm, implantation into the partner's uterus
Ovarian tissue cryopreservation	Surgical excision of ovarian tissue for cryopreservation	• Experimental • Prepubertal or postpubertal • No controlled ovarian stimulation • Possible at moment of transitioning surgery • No partner required	*Male partner* *In vitro* maturation and use of partner's sperm, need of a recipient uterus (7) *Female partner* *In vitro* maturation, fertilization by donor sperm, implantation into the partner's uterus

Reproduced with permission from De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. International Review of Psychiatry. *Int Rev Psychiatry*. 2016 Feb;28(1):112–19.

Ovarian Tissue Cryopreservation

Ovarian tissue cryopreservation is the resection and cryopreservation of the ovary [31]. The technique requires a surgical procedure but does not include an ovarian stimulation. Since hormonal therapy does not deplete the ovary, resection of the tissue can be performed at the time of gender-affirming surgery [18]. At the moment, it is a promising but experimental procedure [31].

Future use of the frozen tissue can be either transplantation of thawed tissue or *in vitro* maturation of primordial follicles. Transplantation of ovarian cortical strips can however include unwanted side effects by restoring female hormone activity. It makes natural conception theoretically possible in case the uterus was not removed or if this transplanted ovarian tissue is stimulated in order to obtain mature oocytes for IVF/ICSI techniques. *In vitro* maturation of primordial follicles (the abundant follicle stage in ovarian cortex) would prevent the recovery of female hormone activity after transplantation, but *in vitro* maturation starting from these immature follicles is not yet possible [4, 15, 33]. Like oocyte cryopreservation, once a mature oocyte is obtained, the use of partner sperm or donor sperm and a recipient uterus upon thawing of the oocytes for future use (female partner or surrogate mother) enables fertility treatment.

The Current Fertility Preservation Options for Transgender Women

The fertility preservation options for transgender women include cryopreservation of sperm collected through ejaculation or direct testicular extraction and cryopreservation of immature testicular tissue. An overview is provided in Table 11.5.2.

Sperm Cryopreservation

The cryopreservation of ejaculated sperm (through masturbation or vibratory stimulation) is the simplest and most reliable method of male fertility preservation [31]. Transgender women may find it difficult to masturbate in order to produce a semen sample for preservation. Even the fact that they have semen samples stored would remind them of their male past and would therefore not make them feel as a complete woman [5, 6]. Depending on the sperm quality, the cryopreserved spermatozoa can be used for future intrauterine insemination or to perform IVF/ICSI in case of a female partner [10]. The need for IVF/ICSI however creates the necessity to start controlled ovarian stimulation in the female partner, followed by oocyte aspiration. The obtained embryo can subsequently be transferred in the partner's uterus. In case of a male partner, a donor oocyte and a surrogate mother are both necessary.

Surgical Sperm Extraction

In case of surgical sperm extraction, a percutaneous aspiration of sperm from the testis of epididymis is performed [31]. Again, this is an established method in daily IVF practice. Although a solution for the transgender women for whom masturbation is a burden, one must not forget that this is a surgical procedure [31]. The obtained spermatozoa can be used for future IVF or ICSI procedures in case of a female partner. In case of a male partner, an oocyte donor and surrogate mother again are both necessary in order to fulfil their child wish [15].

Testicular Tissue Cryopreservation

For cryopreservation of immature testicular sperm, a surgical biopsy of testicular tissue from prepubertal or postpubertal transgender women is performed [31]. This option overcomes the need of masturbation and is possible in prepubertal boys [31]. It is a surgical procedure that can be combined with the sex reassignment surgery. In contrast with the other two options, this is an experimental method. For future use, an *in vitro* maturation procedure or retransplantation is necessary, followed by assisted reproduction techniques. Retransplantation can however restore the male endocrine environment, which clearly is an undesired effect for transgender women.

Table 11.5.2 Fertility preservation options in transgender women

Technique	Description	Considerations	Future use
Sperm cryopreservation	Cryopreservation of ejaculated sperm through masturbation or vibratory stimulation	• Established technique • Masturbation • Postpubertal	*Male partner* Need of a donor oocyte and surrogate mother *Female partner* Intrauterine insemination or IVF/ICSI depending on sperm quality followed by embryo transfer in partner
Surgical sperm extraction	Percutaneous aspiration of sperm from testis or epididymis	• Established technique • No masturbation • Surgical procedure • Postpubertal	*Male partner* Need of a donor oocyte and surrogate mother *Female partner* IVF/ICSI treatment followed by embryo transfer in partner
Immature testicular tissue cryopreservation	Surgical biopsy to of testicular tissue	• Experimental • Prepubertal or postpubertal • Possible at moment of sex reassignment surgery	*Male partner* *In vitro* maturation and need of a donor oocyte and surrogate mother *Female partner* *In vitro* maturation and IVF/ICSI followed by embryo transfer in partner

Reproduced with permission from De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. International Review of Psychiatry. *Int Rev Psychiatry.* 2016 Feb;28(1):112–19.

Future Fertility Preservation Options for Transgender Patients

Current research focuses on optimizing the *in vitro* maturation of immature oocytes and spermatogonial stem cells. An optimized culture model would allow to use the currently banked ovarian or testicular tissue without the need of transplantation. This would solve the unwanted side effect of endocrine restoration by the transplanted tissue.

For trans men, the *in vitro* maturation of obtained cumulus-enclosed oocytes collected during ovarian tissue processing is very promising. These oocytes originate most probably from antral follicles punctured through manipulation of the tissue during the cryopreservation procedure. In combination with freezing of the ovarian cortex, cryopreservation of *in vitro* matured oocytes might further broaden the reproductive perspectives for trans men.

Apart from a testicular biopsy in trans women, to obtain spermatogonial stem cells, research to obtain artificial gametes through stem cells is ongoing [15, 34]. This possibility would be a true revolution for those patients who have not stored their own gametes and currently need oocyte or sperm donation to fulfil their future child wish [3].

Transgender Gestation

In cases where transgender men retain their ovaries and uterus, they may regain fertility after discontinuing androgen therapy. It is a fact that transgender men are becoming pregnant and are having babies, regardless of prior testosterone use [28]. This also emphasizes the need for specialized obstetric care, addressing the specific needs of pregnant transgender men. This care includes a gender-neutral environment and specific attention for increasing gender dysphoria during and after pregnancy. Clinical concerns should also be addressed, such as interruption of testosterone therapy when trying to conceive and postpartum considerations including options for chest (breast) feeding or how and when to reinitiate testosterone [35].

In transgender women, being pregnant and giving birth is still impossible. The Swedish research unit of Brännström and his colleagues conducted a series of uterus transplants and reported a first live birth in 2014 [36]. This opens the possibility for assisted gestation for transgender women [37]. However, there are important medical concerns regarding uterus transplantations if introduced to transgender people [3, 7]. A difficult surgical procedure would be needed in order to change the anatomy of the male pelvis with the intention to perform a successful uterus transplantation. Moreover, immunosuppressive therapy would be necessary and is possibly contraindicated during a pregnancy [3], but that *in se* would not be any different from a uterus transplantation in a cisgender female patient.

Trans Parenting and Children

The above-mentioned options clearly show the opportunities (and limits) for transgender patients with a present or future child wish. All these possibilities, however, are strictly regulated by national legislations. Apart from legislation, some healthcare professionals still need to be convinced about the necessity and the ethical acceptability to preserve fertility in this patient group [10]. The underlying question is whether trans parenting has a negative influence on the gender identity and the sexual orientation of the child [38, 39]. Few studies have addressed this question and conclusive evidence is scarce. Although the results from these studies are reassuring, long-term follow-up studies are undoubtedly needed. None of the studies published so far showed that children suffer to such an extent that would warrant a prohibition of trans parenting [40]. A transgender identity as a reason to interrupt contact between the transgender parent and his or her children, as is the case in some countries, is documented to be harmful for the children [38, 41]. It is shown that a child having a transgender parent may experience more transient and mild harassment than those who do not [38, 41].

Children who were younger at the time of their parent's transitioning, showed to adapt better and maintained healthier relationships with both the transitioning and the non-transitioning parent

in the study of White and Ettner (2007). A less conflicted relationship between child and parents is also predicted by a positive relationship between the two parents [39, 42]. In case of transitioning of the transgender parent before the birth of a child, it is important to disclose the transgender identity of the parent early in childhood, rather than later in the life of the child. The possibility that certain specific circumstances concerning the birth of the child are disclosed by someone else than the parents should be avoided at all-time as this can be traumatic for the child [43].

Conclusion

Fertility and fertility preservation are important topics to discuss before planning any treatment to facilitate a transgender person to live according to the gender they identify themselves with. A patient should be clearly informed upon diagnosis, before starting hormonal treatment or gender-affirming surgery. The first information on fertility preservation should be given by healthcare professionals within the transgender care unit. Next, the patient should be referred to a specialized fertility centre to discuss in more detail his or her options.

The current available options for fertility preservation are very promising. One must however realize that the banking of gametes cannot guarantee future treatment. Upon an apparent child wish post-transitioning, a couple should undergo the screening procedure according to the protocol of the centres for assisted reproduction. Furthermore, not all theoretical reproductive options are possible and not all forms of medically assisted reproduction are available in certain countries. Additionally, medically assisted reproduction, although considered to be safe, is not without health risks and they are often expensive.

The use of the cryopreserved gametes will therefore depend on their quality, success rate of the technique and choice of partner, apart from the centre's policy and national legislation. Fertility in transgender patients also raises the need for appropriate and adapted care before conception, during pregnancy, and after giving birth.

Transgender individuals should be counselled on reproductive issues by professionals prior to initiating treatment in order to have a clear overview of the effects of treatment, the preservation possibilities, and what to expect from it. Transgender people should be referred to specialized centres for assisted reproduction to discuss possibilities. Even if a patient does not always have a clear view on his or her future child wish, it is very important that the patients have access to clear and detailed information so that a well-informed choice can be made.

REFERENCES

1. Finlayson C, Johnson EK, Chen D, *et al*. Proceedings of the working group session on fertility preservation for individuals with gender and sex diversity. *Transgend Health*, 2016; **1**: 99–107.
2. Coleman E, Bockting W, Botzer M, *et al*. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *International Journal of Transgenderism*, 2012; **13**: 165–232.
3. T'Sjoen G, Van Caenegem E, Wierckx K. Transgenderism and reproduction. *Curr Opin Endocrinol Diabetes Obes*, 2013; **20**: 575–9.
4. Wierckx K, Van Caenegem E, Pennings G, *et al*. Reproductive wish in transsexual men. *Hum Reprod*, 2012; **27**: 483–7.
5. Wierckx K, Stuyver I, Weyers S, *et al*. Sperm freezing in transsexual women. *Arch Sex Behav*, 2012; **41**: 1069–71.
6. De Sutter P, Kira K, Verschoor A, Hotimsky A. The desire to have children and the 101 preservation of fertility in transsexual women: a survey. *International Journal of Transgenderism*, 2003; **6**: 3.
7. Hamada A, Kingsberg S, Wierckx K, *et al*. Semen characteristics of transwomen referred for sperm banking before sex transition: a case series. *Andrologia*, 2014: **47**: 832–8.
8. Chen D, Simons L, Johnson EK, Lockart BA, Finlayson C. Fertility preservation for transgender adolescents. *J Adolesc Health,* 2017; **61**: 120–3.
9. Nahata L, Tishelman AC, Caltabellotta NM, Quinn GP. Low fertility preservation utilization among transgender youth. *J Adolesc Health*, 2017; **61**: 40–4.
10. De Sutter P. Gender reassignment and assisted reproduction: present and future reproductive options for transsexual people. *Hum Reprod*, 2001; **16**: 612–14.
11. Hagen CP, Sorensen K, Anderson RA, Juul A. Serum levels of antimüllerian hormone in early maturing girls before, during, and after suppression with GnRH agonist. *Fertil Steril*, 2012; **98**: 1326–30.
12. Johnson EK, C F. Preservation of fertility potential for gender and sex diverse individuals. *Transgender Health*, 2016; **1**: 41–4.
13. Payer AF, Meyer WJ, Walker PA. The ultrastructural response of human Leydig cells to exogenous estrogens. *Andrologia*, 1979; **11**: 423–36.
14. Lubbert H, Leo-Rossberg I, Hammerstein J. Effects of ethinyl estradiol on semen quality and various hormonal parameters in a eugonadal male. *Fertil Steril*, 1992; **58**: 603–8.
15. Ettner R, Monstrey S, Eyler AE. *Principles of Transgender Medicine and Surgery*. New York, NY: Haworth Press, 2007.
16. Schulze C. Response of the human testis to long-term estrogen treatment: morphology of Sertoli cells, Leydig cells and spermatogonial stem cells. *Cell Tissue Res*, 1988; **251**: 31–43.
17. Schneider F, Neuhaus N, Wistuba J, *et al*. Testicular functions and clinical characterization of patients with gender dysphoria (GD) undergoing sex reassignment surgery (SRS). *J Sex Med*, 2015; **12**: 2190–200.
18. Van Den Broecke R, Van Der Elst J, Liu J, Hovatta O, Dhont M. The female-to-male transsexual patient: a source of human ovarian cortical tissue for experimental use. *Hum Reprod*, 2001; **16**: 145–7.
19. De Roo C, Lierman S, Tilleman K, *et al*. Ovarian tissue cryopreservation in female-to-male transgender people: insights into ovarian histology and physiology after prolonged androgen treatment. *Reprod Biomed Online*, 2017; **34**: 557–66.
20. Pache TD, Chadha S, Gooren LJ, *et al*. Ovarian morphology in long-term androgen-treated female to male transsexuals. A human model for the study of polycystic ovarian syndrome? *Histopathology*, 1991; **19**: 445–52.
21. Caanen MR, Soleman RS, Kuijper EA, *et al*. Antimüllerian hormone levels decrease in female-to-male transsexuals using testosterone as cross-sex therapy. *Fertil Steril*, 2015; **103**: 1340–5.
22. Van den Broecke R, Liu J, Handyside A, *et al*. Follicular growth in fresh and cryopreserved human ovarian cortical grafts

transplanted to immunodeficient mice. *Eur J Obstet Gynecol Reprod Biol*, 2001; **97**: 193–201.
23. Grynberg M, Fanchin R, Dubost G, *et al*. Histology of genital tract and breast tissue after long-term testosterone administration in a female-to-male transsexual population. *Reprod Biomed Online*, 2010; **20**: 553–8.
24. Spinder T, Spijkstra JJ, van den Tweel JG, *et al*. The effects of long term testosterone administration on pulsatile luteinizing hormone secretion and on ovarian histology in eugonadal female to male transsexual subjects. *J Clin Endocrinol Metab*, 1989; **69**: 151–7.
25. Caanen MR, Soleman RS, Kuijper EA, *et al*. Antimüllerian hormone levels decrease in female-to-male transsexuals using testosterone as cross-sex therapy. *Fertil Steril*, 2015; **103**: 1340–45.
26. Caanen MR, Schouten NE, Kuijper EAM, *et al*. Effects of long-term exogenous testosterone administration on ovarian morphology, determined by transvaginal (3D) ultrasound in female-to-male transsexuals. *Hum Reprod*, 2017; **32**: 1457–64.
27. T'Sjoen G, Van Caenegem E, Wierckx K. Transgenderism and reproduction. *Curr Opin Endocrinol Diabetes Obes*, 2013; **20**: 575–9.
28. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. *Obstet Gynecol*, 2014; **124**: 1120–7.
29. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. *Int Rev Psychiatry*, 2016; **28**: 112–19.
30. Kim J, Deal AM, Balthazar U, Kondapalli LA, Gracia C, Mersereau JE. Fertility preservation consultation for women with cancer: are we helping patients make high-quality decisions? *Reprod Biomed Online*, 2013; **27**: 96–103.
31. Wallace SA, Blough KL, Kondapalli LA. Fertility preservation in the transgender patient: expanding oncofertility care beyond cancer. *Gynecol Endocrinol*, 2014: **30**: 868–71.
32. Steever J. Cross-gender hormone therapy in adolescents. *Pediatr Ann*, 2014; **43**: e138–44.
33. Dewailly D, Andersen CY, Balen A, *et al*. The physiology and clinical utility of anti-M√ºllerian hormone in women. *Hum Reprod Update*, 2014; **20**: 804.
34. Duggal G, Heindryckx B, Deroo T, De Sutter P. Use of pluripotent stem cells for reproductive medicine: are we there yet? *Vet Q*, 2014; **34**: 42–51.
35. Obedin-Maliver J, Makadon HJ. Transgender men and pregnancy. *Obstet Med*, 2016; **9**: 4–8.
36. Brännström M, Johannesson L, Bokström H, *et al*. Livebirth after uterus transplantation. *Lancet*, 2015; **385**: 607–16.
37. Murphy TF. Assisted gestation and transgender women. *Bioethics*, 2015; **29**: 389–97.
38. T'Sjoen G, Van Caenegem E, Wierckx K. Transgenderism and reproduction. *Curr Opin Endocrinol Diabetes Obes*, 2013; **20**: 575–9.
39. White T, Ettner R. Disclosure, risks and protective factors for children whose parents are undergoing a gender transition. *Journal of Gay & Lesbian Psychotherapy*, 2004; **8**: 129–45.
40. Murphy TF. The ethics of fertility preservation in transgender body modifications. *J Bioeth Inq*, 2012; **9**: 311–16.
41. Green R. Sexual identity of 37 children raised by homosexual or transsexual parents. *Am J Psychiatry*, 1978; **135**: 692–7.
42. White T, Ettner R. Adaptation and adjustment in children of transsexual parents. *Eur Child Adolesc Psychiatry*, 2007; **16**: 215–21.
43. Chiland C, Clouet A-M, Golse B, Guinot M, Wolf J. A new type of family: transmen as fathers thanks to donor sperm insemination. A 12-year follow-up exploratory study of their children. *Neuropsychiatrie de L'Enfance et de l'Adolescence*, 2013; **61**: 365–70.

SECTION 12

Endocrine Responses to Systemic Diseases or Substance Misuse

12.1

Endocrinology of Systemic Disease

12.1.1 The Endocrine Response to Stress

David Henley, Thomas Upton, and Stafford L. Lightman

Introduction—What is Stress?

In the face of a real or perceived threat, organisms mount coordinated and specific hormonal, autonomic, immune, and behavioural responses that allow escape or adaptation [1–3]. Responses that are either inadequate or excessive in terms of specificity, intensity or duration may result in a range of psychological or physical pathologies [2–5]. This concept of 'threat' and the subsequent 'response' is what we typically understand as 'stress'; however, as this construct is so broad and difficult to define [2] 'stress' can be very difficult to investigate or study objectively [6].

The concept of homeostasis emerged in the early twentieth century [4] as the ideal steady state for all physiological processes. Stress may be considered the state where this ideal is threatened. More easily appreciated however are those factors, both intrinsic and extrinsic, that challenge homeostasis (termed stressors) and the complex physiologic, hormonal, and behavioural actions that restore balance, the stress response [1]. The importance of endocrine systems in the stress response was emphasized by the pioneering work of Hans Selye and others [7] who described the need for multiple, integrated systems responding in a coordinated fashion following stressor exposure. It was noted that exposure to noxious stimuli caused non-specific activation of the hypothalamic–pituitary–adrenal (HPA) and sympathoadrenomedullary (SAM) axes. Continued exposure resulted in lasting and damaging effects on various endocrine, immune, and other systems although recovery was possible provided the stress was terminated [8]. Other potential stressors exist including exertion, physical extremes, trauma, injury, and psychological stress. Psychological stressors are some of the most potent stimuli of the endocrine stress response particularly when they involve elements of novelty, uncertainty, and unpredictability. The anticipation of a potentially stressful event may be as potent an activator of the stress response as the event itself [8].

Anatomy and Physiology of the Endocrine Response to Stress

The HPA and SAM axes are the principal endocrine effector arms of the stress response (Figure 12.1.1.1). However, several other hormone axes and neurotransmitter systems are either directly stress-responsive themselves, or act to modulate these systems.

The HPA Axis

Corticotropin-releasing hormone (CRH) promotes synthesis and release of anterior pituitary adrenocorticotropin (ACTH) [9]. Hypophysiotropic CRH neurons project from the paraventricular hypothalamic nucleus (PVN) to the median eminence. CRH is also widely distributed throughout the central nervous system (CNS) where it has important effects on behaviour and cognitive processing. Within the brainstem, interactions with sympathetic and parasympathetic centres influence autonomic functioning while within limbic and para-limbic regions, such as the amygdala, CRH influences the expression of mood and anxiety-type behaviours [10]. Arginine vasopressin (AVP), synthesized in parvocellular cells of the PVN, acts synergistically with CRH to stimulate the release of ACTH [10].

ACTH acts directly on the adrenal cortex to promote release of glucocorticoids into the circulation [1, 3, 10]. Glucocorticoids, in general, have two fundamental roles in the stress response. First, during stress free periods, basal levels prepare the organism for future stress exposure. The circadian rise occurs prior to the onset of daily activity and thus in humans starts around 3:00 a.m. This

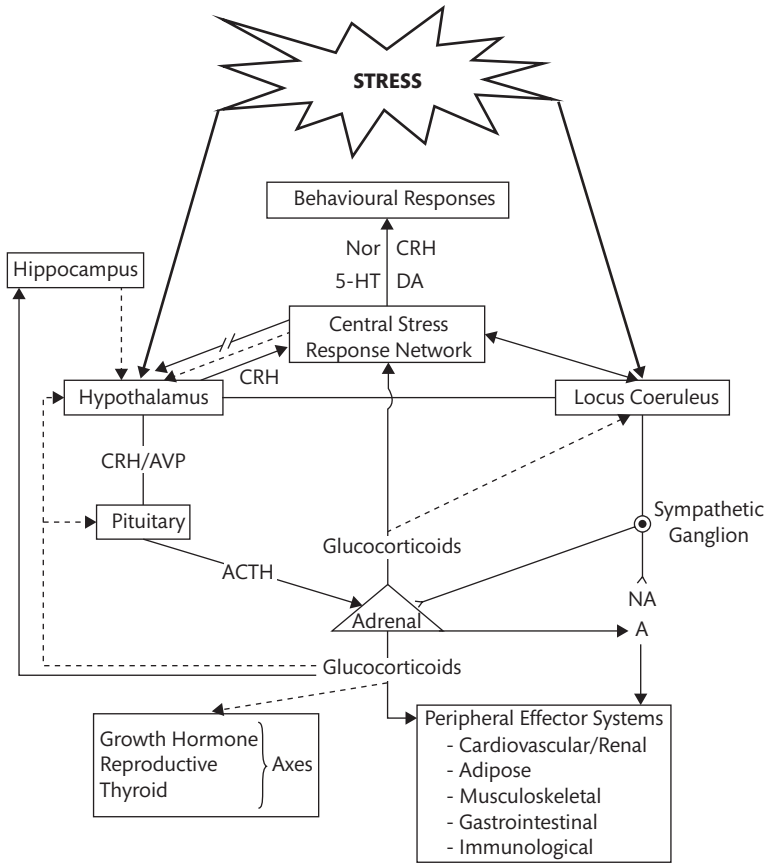


Figure 12.1.1.1 Chronic stress response. Simplified overview of the chronic stress response and its two main effector arms, the hypothalamic–pituitary–adrenal axis and the sympathoneural/sympathoadrenomedullary system. Note the glucocorticoid feedforward and feedback regulatory loops, reciprocal interaction of CRH, and the locus coeruleus, together with the putative central stress response network in effecting peripheral and central adaptive responses. Components of the central brain stress response network include: parvocellular neurons in the paraventricular nuclei, central nucleus of the amygdala, bed nuclei of the stria terminalis, Barrington's nucleus, ventral tegmental area, dorsal raphe, locus coeruleus and the A1/A2 medullary noradrenergic cell groups. CRH, corticotropin-releasing hormone; AVP, arginine vasopressin; ACTH, adrenocorticotrophin; NA, noradrenaline; A, adrenaline; 5-HT, 5-hydroxytryptamine (serotonin); DA, dopamine. Solid lines indicate stimulation; dashed lines indicate inhibition; broken line indicates indirect projections.

anticipatory activity results in energy storage and conservation by promoting glucose and fat uptake, opposing energy utilization, and thus prepares the organism for the activities of the next waking day. Glucocorticoids also prime the immune system for future activation and promote memory formation of previous stressors so that repeat exposure to the same or similar stressor results in a rapid and efficient response [11].

The second role of the HPA response is to modulate events at the time of stress exposure itself. Initially glucocorticoids enhance the cardiovascular effects of catecholamines and AVP, promote energy provision and utilization, influence and enhance appropriate stress-related behaviours, and stimulate certain aspects of the immune response [11]. However, and very importantly, once the stress response has been initiated some of the principal actions of glucocorticoids are to suppress and restrain the activity of these systems, in particular the SAM and immune responses. In doing so, glucocorticoids help regulate a stress response that is appropriate in terms of intensity and duration, ensuring responses are 'switched off' when the stress has been successfully dealt with [2, 11].

Glucocorticoid secretion is controlled by a complex feedback system that involves direct inhibitory action on the hypothalamus and anterior pituitary, reducing releasing hormone (CRH and ACTH) production and consequently limiting glucocorticoid release into the circulation. A further level of feedback occurs at the hippocampus, a site that is also important in memory formation. A subset of hippocampal neurons project to the hypothalamus where they release the neurotransmitter γ-aminobenzoic acid (GABA), inhibiting CRH release, thus contributing to negative feedback effect on cortisol secretion [1].

In the brain, the glucocorticoid receptor (GR) and the mineralocorticoid receptor (MR) are both involved in the feedback system. GRs are found throughout the brain but are most abundant in hypothalamic CRH neurons and pituitary corticotrophs, while MR expression is highest in the hippocampus. GR is occupied during periods of intermediate to high glucocorticoid secretion (e.g. during the circadian peak and following stress) while the high affinity MR is extensively bound even during periods of basal secretion [12]. Therefore, MR is thought to regulate tonic neurocognitive HPA activity while GR (in coordination with MR) mediates the response to stress.

Three time domains of corticosteroid feedback have been described [13]. Fast, rate sensitive feedback occurs within seconds

to minutes, during the period of increasing plasma corticosteroid concentrations, and probably controls the rate and magnitude of ACTH and corticosteroid response to stimuli. This is likely mediated by membrane associated MRs via rapid, non-genomic mechanisms [14]. Disruption of fast feedback has been demonstrated in ageing humans and in depressed patients, and thus may have a role in the maintenance of homeostasis [14]. Intermediate feedback (2–10 hr) limits the response of the system to repeated stimulation within a relatively short period of time, while slow feedback (over hours to days) may have the same role during prolonged stress [13].

Circadian and Ultradian Rhythms of HPA Activity

ACTH and cortisol fluctuate in their secretory activity. The classic circadian (24-h) rhythm describes the pattern of HPA activity with plasma hormone concentrations falling to nadir around midnight, before rising in the early hours of the morning, reaching a peak soon after waking. Concentrations gradually fall again throughout the day toward the nadir. However, the circadian rhythm is subserved by an underlying ultradian (less than 24-h) rhythm of secretory pulses which can only be detected by frequent blood sampling.

Mathematical modelling shows the episodic secretion of ACTH and cortisol is an inherent feature of the pituitary–adrenal axis [15]. Modulation of the amplitude of both ACTH and cortisol secretory pulses gives rise to their respective nyctohemeral rhythms [16]. Under physiological conditions ACTH secretion is characterized by episodic pulses of activity separated by intervals of low basal (non-pulsatile) secretion.

Cortisol is synthesized and secreted from zona fasciculata cells of the adrenal cortex in response to ACTH stimulation. There is a high temporal concordance between ACTH and cortisol secretion peaks, with the latter lagging those of ACTH by 10 min [16–18] (Figure 12.1.1.2). Secretory bursts of both hormones are episodic in that they are independent events produced randomly over time [16, 19]. Sexual diergism in ACTH pulsatility has been demonstrated with males showing greater pulse frequency (18 vs. 10 per 24-h), mean peak amplitude and area under the 24-h profile [20]. Cortisol secretory bursts occur more frequently in the early hours of the morning (shortly before waking) and least frequently during late afternoon [19].

The pulsatile, ultradian rhythms of the HPA axis arise as a consequence of dynamic feedback and feedforward loops and the interactions of glucocorticoids at the molecular level [21]. Corticosteroid pulsatility is important in that it provides an optimally sensitive HPA axis equilibrium with scope for digital, in addition to analogue, signal for tissue GRs [22]. Hippocampal GR and MR receptors translocate rapidly from the cytoplasm to the nucleus and bind DNA in response to a corticosteroid pulse [23]. Within a 1-hour interpulse interval, GR dissociates rapidly from DNA and disappears from the nucleus, in contrast MR remains bound to DNA. Therefore, changes in pulse frequency will have differential effects on MR and GR binding to DNA. Given the variety of cellular transcription factors and molecular chaperones there is scope for multiple cell specific responses to different digital signals [21].

The Sympathoneural and Sympathoadrenomedullary Axis

The hallmark sympathetic 'fight or flight' response is characterized by global activation of the SAM system. Typical features include accelerated heart rate, increased blood pressure, and rapid breathing. Fear, vigilance, sensory arousal, and motor activation often occur with trembling, goose bumps, and piloerection.

The sympathetic nervous system originates from nuclei in the lower brainstem that use noradrenaline as their principal neurotransmitter (Figure 12.1.1.1). Noradrenergic nuclei, centred on the locus coeruleus (LC), project downward to the intermediolateral columns of the spinal cord. Cell bodies send preganglionic fibres to the paraspinal ganglia chain from where postganglionic fibres give rise to sympathetic nerves that supply the heart, blood vessels, lungs, gut, kidneys, and other organs. These nerves principally release noradrenaline from terminals close to their site of action. Other preganglionic fibres innervate the adrenal medulla, regulating release of adrenaline into the general circulation.

Catecholamines, the effector hormones of the system, act through specific cell surface receptors whose wide distribution account for their rapid effect on multiple physiological processes [1, 3]. Catecholamine surges promote glucose release, immune activation, and increased blood flow to essential organs while inhibiting non-essential activities like digestion. Together this produces a 'state of emergency' which can rapidly attend to sudden changes in physiological balance [3]. The SAM response has the ability to trigger purely in anticipation of a stressful event and interact with other stress-responsive systems [3]. Interaction may occur either through neural connections or as a consequence of increased blood flow that allows rapid transport of other messengers (such as hormones and cytokines) to their respective sites of action [3].

Acute Stress

The stress response system has evolved as both an early warning system capable of recognizing potential or existing threats, and as a response system that can initiate and drive the necessary processes required to escape or confront the threat. The response is dynamic, beginning rapidly with brain and behavioural activation, followed quickly by physiological activation. Positive-feedback and feedforward loops enhance and reinforce these processes as well as recruiting other effector arms of the stress response. Slower acting hormone systems moderate the response by effectively providing a brake and ensuring the stress response is appropriate in both intensity and duration [11, 24].

Changes in the environment that represent either real or potential threats are recognized by parts of the brain responsible for receiving, integrating, interpreting, and then relaying information on to areas responsible for coordinating the necessary response. Brain activation occurs within milliseconds and proceeds over seconds to minutes as the response continues to unfold. Stereotypical orienting behaviour gradually gives way to more goal-directed behaviour that is specific to the stressor being faced and the environment in which it is occurring [24].

Activation of the autonomic nervous system occurs within seconds and is enhanced by withdrawal of parasympathetic activity. Within minutes of the onset of this cascade of events, the appearance of ACTH signals recruitment of the HPA axis into the process [1, 10]. Cortisol levels begin to rise within 2–5 min [25], and levels peak about 15–20 minutes after the onset of the stress [26] (Figure 12.1.1.3). Early actions of the HPA system provide additional energy resources, while slower gene-related effects over minutes to hours restrain ongoing actions of the stress response which, if left unchecked, may prove to be unsustainable [27].

Chronic Stress and the Brain

There is a marked change in the hypothalamic response to chronic stress with a greater role for AVP [28]. In addition to colocalization of AVP and CRH in neurosecretory axon terminals [29] there are increases in hypothalamic AVP synthesis, proportion of CRH neurons coexpressing AVP, and in the ratio of AVP to CRH immunoreactivity in neurosecretory vesicles. AVP stimulated ACTH secretion is less sensitive to glucocorticoid feedback than is CRH [30]. Pituitary changes with chronic stress paradigms include a reduction in CRH receptor numbers and sustained elevations in V1b (AVP) receptor mRNA [29]. It appears in some chronic stress paradigms that CRH has a permissive role whereas AVP is the dynamic mediator of ACTH secretion.

It is important for survival that the HPA axis responds adequately during chronic stress. Rodent stress models reveal three basic patterns of response, depending on the type of stress [31]: (1) desensitization of ACTH responses to sustained stimulus, but hyper-responsiveness to a novel stress despite elevated plasma glucocorticoid levels; (2) corticotroph hyper-responsiveness to a novel stimulus, with no desensitization to the primary repeated stress; (3) small and transient increases in basal ACTH, followed by marked hyporesponsiveness to novel stimuli. The level of response is determined by the differential regulation of CRH and AVP. The increase in AVP during chronic stress (where glucocorticoid negative feedback downregulates CRH and ACTH responses) appears to be an important mediator of ACTH release upon new demand.

In the presence of increased plasma glucocorticoid levels during chronic stress, decreased sensitivity of glucocorticoid feedback is critical for the maintenance of ACTH responses. It appears that the increase in number of pituitary V1b receptors is the main determining factor for the responsiveness of the corticotroph during adaptation to chronic stress [31].

Involvement of the limbic system in HPA axis regulation is complex (Figure 12.1.1.1). The role of limbic structures is both region- and stimulus-specific; they all express both GR and MR and they all exert their effects via subcortical intermediaries [32]. Typically, the hippocampus and anterior cingulate/prelimbic cortex inhibit stress-induced HPA axis activation, whereas the amygdala and possibly the infralimbic cortex may enhance glucocorticoid secretion [32]. Furthermore, the HPA axis is also subject to glucocorticoid-independent inhibition from neuronal sources. For example, the PVN is richly innervated by GABAergic neurons from the bed nucleus of the stria terminalis, medial preoptic area, dorsomedial hypothalamus, and lateral hypothalamic area. However, the degree to which these GABAergic inhibitory circuits respond to neural vs. glucocorticoid inhibition has not been fully elucidated [32].

The concept of a **central stress response network** [25] recruited by glucocorticoids in chronic stress defines a critical role for extrahypothalamic CRH neuronal cell groups, in particular the amygdala. Elevated glucocorticoid concentrations acting in a feed-forward manner at the amygdala increase CRH expression and secretion, which is tightly coupled to hypersensitivity of the HPA axis to stressors. CRH acts on receptors in structures throughout the brain, especially monoaminergic cell groups that widely innervate the forebrain, resulting in behavioural changes (e.g. more cautious, more ready to be diverted from tasks at hand, adopt alternative strategies, enjoy rewards and remember fearful situations) that make the organism chronically exposed to stress more capable of adapting to the stressful conditions (Figure 12.1.1.1).

Reciprocal neural connections exist between CRH and the LC/noradrenergic neurons of the central stress system, with each one stimulating the other [4] (Figure 12.1.1.1). Chronic stress increases CRH content in the LC. Thus, CRH may induce mechanisms that result in HPA axis facilitation via increased catecholaminergic input to CRH cells, increasing the capacity for CRH response to acute stress during periods of chronic stress when the corticosteroid feedback signal is high [30].

Clinical Manifestations of Chronic Stress

Throughout the history of medicine, reference has been made to the influence of stress, particularly in the form of negative emotions and psychological distress, on physical health [6]. Relevant examples include psychiatric conditions (depression and post-traumatic stress disorder) [1], coronary heart disease (CHD), immune-mediated conditions including asthma, chronic diseases such as osteoporosis, diabetes, and dementia, as well as premature death [1, 6]. Why some individuals manifest stress as psychiatric illness when others are either more prone to physical disease or even resistant to the effects of stress exposure is not well understood.

There is a protective role for the stress response in the short term [2]. The associated learning and adaptation (a process that requires plasticity of brain responses) that follows stress exposure is critical to the longer-term health and survival of the individual. However, when these responses occur in excess of the body's requirements or continue for longer than necessary, damaging effects occur [2].

Psychosocial Stress

There is an association between socioeconomic status (SES) and health at every level of the SES hierarchy [33]. The Whitehall study of CHD mortality [34] found employment grade was a stronger predictor of subsequent risk of CHD death than any other major coronary risk factor. Depression and depressive symptoms are both inversely related to SES status and depression is linked to health outcomes, particularly CHD [33]. There are two mechanisms by which higher placement in the SES hierarchy can reduce stress and its somatic consequences: (1) by diminishing the likelihood that individuals will experience negative events; (2) through greater social and psychological resources to cope with stressful life events, therefore being less susceptible to the subjective experience of stress [33].

Mood Disorders

Melancholic depression has been described as the prototypic example of chronic activation of the stress system (both HPA axis and SAM) [4]. Cortisol secretion is increased, plasma ACTH response to exogenous CRH decreased and autopsy studies show marked increase in the number of PVN CRH and AVP neurons [1]. Depression is also associated with increased pituitary vasopressinergic responsivity and the LC of depressed patients contains elevated CRH concentrations [35]. Repeated stress that causes frequent surges in blood pressure and catecholamine release is associated with accelerated atherosclerosis and an increased risk of myocardial infarction. Patients with melancholic depression develop varying degrees of atherosclerosis and cardiovascular disease [1] and depressed people with chronic HPA hyperactivity may have reduced life expectancy predominantly as a result of increased cardiovascular mortality [1, 6, 36]. Patients with melancholic depression may develop metabolic syndrome, osteoporosis and Th_1 immunosuppression [1] consistent with chronic hyperactivation of the stress system. Hypercortisolism is also associated with other mood and affective disorders including anorexia nervosa, chronic anxiety, obsessive-compulsive disorder, chronic alcoholism, and other situations such as childhood sexual abuse [1]. Hyperactivity of the LC and other central noradrenergic centres have been shown to influence anxiety and behavioural arousal, with dysregulation of this system postulated as contributing to the pathogenesis of mood disorders, particularly depression, and with noradrenaline levels being an important predictor of outcome in major depression [1].

Animal experiments of chronic stress provide evidence that glucocorticoid overexposure affects the hippocampus with respect to neuronal viability and function—decreased neurogenesis, degenerative loss in pyramidal neurons, reduced dendritic branching and atrophy [35, 37]. This led to the so-called glucocorticoid cascade hypothesis [35] where stress-induced HPA activation and elevated glucocorticoid levels were purported to act in a feedforward manner causing hippocampal damage, resulting in disinhibition of glucocorticoid negative feedback, further rise in glucocorticoid levels and accumulating damage to the hippocampus. This has been supported in principle by the fact that patients with Cushing's disease exhibit both hippocampal atrophy and depression, both of which are reversed with treatment. In addition, depressed patients experience cognitive dysfunction consistent with hippocampal damage and antidepressant treatment may enhance neurogenesis [38, 39]. However, although reduced hippocampal volumes have been seen on MRI scans of depressed patients, significant histologic damage has not been found on post-mortem studies [35]. Thus, despite compelling animal data linking stress-induced hypercortisolism with modulation of neurogenesis in the pathogenesis of depression, evidence for translation to human depression is inconclusive, and is an active area of ongoing research. HPA dysregulation appears well before clinical symptomatology and is a predictor of treatment resistance in depression. Failure to normalize HPA axis responses with treatment is a strong predictor of relapse [40].

Obesity and the Metabolic Syndrome

Chronic stress has been linked to obesity and the metabolic syndrome (characterized by the combination of central obesity, insulin resistance, dyslipidaemia, and hypertension [41]). Glucocorticoids regulate adipocyte differentiation and stress-induced excess cortisol is associated with increased abdominal fat accumulation [41]. In humans, chronic stress-induced increases in cortisol, catecholamines, and the inflammatory cytokine IL-6 in combination with associated suppression of the growth hormone-, gonadal- and thyroid-axes, produces a hormonal milieu conducive to the development of visceral obesity, hypertension, atherosclerosis, osteoporosis and immune dysfunction [41]. Corticosteroids stimulate behaviours that are mediated by dopaminergic mesolimbic 'reward' pathways and the central stress response network [25]. Glucocorticoids stimulate caloric intake and consumption of 'comfort foods' which may result in metabolic feedback signals that dampen brain stress responses [25]. In the current era of chronic social stress, together with the availability of high calorie palatable foods (acquired with ever decreasing physical effort), this adaptive mechanism proposed to enable many species to survive may be occurring at a significant (maladaptive) metabolic cost to contemporary humans.

Sleep Disorders

In animal models with varying degrees of sleep deprivation there is a consistent pattern of cognitive impairment, namely in learning and retention [42]. This is associated with increased brain levels of pro-inflammatory cytokines (IL-1β mRNA), hippocampal oxidative stress, and structural changes. Clinical studies have confirmed elevated evening cortisol and day time growth hormone levels with increased sympathetic nervous activity in both total and partial sleep deprivation [43]. Resultant increased insulin resistance and reduced glucose tolerance increases diabetes risk. This is further compounded by dysregulation of the neuroendocrine control of appetite, promoting obesity.

Obstructive sleep apnoea (OSA) represents a chronically stressed state. OSA is characterized by intermittent upper airway obstruction and subsequent hypoxia during sleep. A cyclical sequence of events consisting of upper airway obstruction, progressive hypoxaemia, autonomic, and electroencephalogram (EEG) arousal occurs. This will prompt the individual to open and clear the airway to reverse the asphyxia, followed by successive relaxation of the airway and subsequent constriction, resulting in fragmented sleep [44]. Morbidity and mortality from OSA are primarily due to cardiovascular disease. OSA is also associated with insulin resistance and the metabolic syndrome.

HPA axis dysfunction in OSA that is altered with continuous positive airways pressure (CPAP) therapy [18]. Obese male subjects with moderately severe or severe OSA had ACTH and cortisol levels measured every 10 minutes over 24 hours both pre- and 3 months post-CPAP under basal conditions [18]. Hormone secretory characteristics estimated from multiparameter deconvolution analysis found no change in the number of predicted secretory episodes, secretion pulse height, or frequency. However, there was an increased mean pulse mass pretreatment due to a longer duration of the individual secretory episodes, and a significant reduction in pulsatile and total ACTH and cortisol production post-CPAP [45]. This is consistent with impaired fast feedback affecting pulse duration [14] and may be due to metabolic/hypoxic insults on the hippocampus, alterations in hippocampal MR expression due to SAM hyperactivity or result from an AVP effect on ACTH pulse duration. Following the single breath 35% CO_2 stress test [26], there was a markedly exaggerated cortisol response to CO_2 pre-CPAP which was reduced to normal levels after treatment supporting HPA axis

hyper-responsiveness in untreated OSA [45]. It is therefore likely that the activation of the stress system in OSA contributes to the metabolic complications of this condition.

Other Effects

CRH hypersecretion and HPA axis activation influences the activity of many organ systems. CRH hyperactivity is associated with gastro-intestinal symptoms such as pain, increased gut motility, and diarrhoea—typical features of the irritable bowel syndrome, a condition commonly associated with stress [1]. Similarly, glucocorticoids inhibit the growth axis and it has been postulated that the severe growth retardation associated with psychosocial abuse or deprivation during childhood is, in part, related to chronic HPA axis activation [1].

By contrast, chronic hypoactivation of the HPA axis may also result in disease states. Post-traumatic stress disorder, chronic fatigue syndrome and atypical depression [1, 36] are associated with CRH hypoactivity and reduced cortisol production. Similarly, immune dysregulation is an important consequence of altered HPA axis activity. Differential levels of hypothalamic CRH in the high CRH-producing Fischer and lower CRH-producing Lewis rats are associated with enhanced immune response and resistance to infections and tumours in the Lewis rats, but also an increased susceptibility to some autoimmune conditions [6]. In human studies, rheumatoid arthritis appears to be associated with HPA axis hypoactivation [46] with blunted cortisol diurnal rhythms and reduced ACTH and cortisol levels.

HPA Response in Critical Illness and Major Surgery

Critical illness and major surgical procedures are examples of extreme stressors that result in marked elevation of circulating glucocorticoid concentrations. In spite of this, high concentrations of ACTH in these patients appears only to be transient [47]. High frequency plasma sampling reveals ongoing coordinated and pulsatile ACTH/cortisol interactions remain [48], albeit with lower ACTH pulse amplitude [49]. The altered dynamics of the HPA axis in these situations are likely explained by combinations of increased sensitivity of the adrenal to ACTH [50], altered corticosteroid-binding globulin function [51], and reduced cortisol metabolism [52].

Adjuvant glucocorticoid supplementation to improve mortality in patients with septic shock has been controversial for decades. Two large clinical trials [53, 54] attempting to answer this question provide conflicting evidence with regard to mortality but some suggestion of short-term benefit in terms of improved shock resolution and shorter duration of mechanical ventilation. Clearly there remains much that is unknown about the dynamics of the stress response in critical illness.

Summary

Stress may be considered as a real or perceived threat to homeostasis. The two primary arms of the stress response are the HPA axis and the SAM systems. These two systems are interlinked and regulated by complex feedback and feedforward processes. The acute stress response is protective and promotes survival in the short term. However, prolonged activation of the stress response is implicated in the pathogenesis of illness, in particular mood and affective disorders, but also obesity, metabolic syndrome, and obstructive sleep apnoea.

REFERENCES

1. Charmandari E, Tsigos C, Chrousos G. Endocrinology of the stress response. *Annu Rev Physiol*, 2005; **67**: 259–84.
2. McEwen BS. Protective and damaging effects of stress mediators. *N Engl J Med*, 1998; **338**: 171–9.
3. Habib KE, Weld KP, Rice KC, *et al.* Oral administration of a corticotropin-releasing hormone receptor antagonist significantly attenuates behavioral, neuroendocrine, and autonomic responses to stress in primates. *Proc Natl Acad Sci U S A*, 2000; **97**: 6079–84.
4. Chrousos GP, Gold PW. The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis. *JAMA J Am Med Assoc*, 1992; **267**: 1244–52.
5. Vanitallie TB. Stress: a risk factor for serious illness. *Metabolism*, 2002; **51**: 40–5.
6. Sternberg EM. Emotions and disease: from balance of humors to balance of molecules. *Nat Med*, 1997; **3**: 264–7.
7. Szabo S, Tache Y, Somogyi A. The legacy of Hans Selye and the origins of stress research: a retrospective 75 years after his landmark brief 'letter' to the Editor of Nature. *Stress*, 2012; **15**: 472–8.
8. Levine S. Influence of psychological variables on the activity of the hypothalamic-pituitary-adrenal axis. *Eur J Pharmacol*, 2000; **405**: 149–60.
9. Vale W, Spiess J, Rivier C, Rivier J. Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and beta-endorphin. *Science*, 1981; **213**: 1394–97.
10. Harbuz MS, Lightman SL. Stress and the hypothalamo-pituitary-adrenal axis: acute, chronic and immunological activation. *J Endocrinol*, 1992; **134**: 327–39.
11. Sapolsky RM, Romero LM, Munck AU. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocr Rev*, 2000; **21**: 55–89.
12. Reul JM, De Kloet ER. Two receptor systems for corticosterone in rat brain: microdistribution and differential occupation. *Endocrinology*, 1985; **117**: 2505–11.
13. Keller-Wood ME, Dallman MF. Corticosteriod inhibition of ACTH secretion. *Endocr Rev*, 1984; **5**: 1–24.
14. Atkinson HC, Wood SA, Castrique ES, Kershaw YM, Wiles CC, Lightman SL. Corticosteroids mediate fast feedback of the rat hypothalamic-pituitary-adrenal axis via the mineralocorticoid receptor. *Am J Physiol Endocrinol Metab*, 2008; **294**: E1011–22.
15. Walker JJ, Spiga F, Waite E, *et al.* The origin of glucocorticoid hormone oscillations. *PLoS Biol*, 2012; **10**: e1001341.
16. Veldhuis JD, Iranmanesh A, Johnson ML, Lizarralde G. Amplitude, but not frequency, modulation of adrenocorticotropin secretory bursts gives rise to the nyctohemeral rhythm of the corticotropic axis in man. *J Clin Endocrinol Metab*, 1990; **71**: 452–63.

17. Gallagher TF, Yoshida K, Roffwarg HD, Fukushima DK, Weitzman ED, Hellman L. ACTH and cortisol secretory patterns in man. *J Clin Endocrinol Metab*, 1973; **36**: 1058–68.
18. Henley DE, Leendertz JA, Russell GM, *et al.* Development of an automated blood sampling system for use in humans. *J Med Eng Technol*, 2009; **33**: 199–208.
19. Veldhuis JD, Iranmanesh A, Lizarralde G, Johnson ML. Amplitude modulation of a burstlike mode of cortisol secretion subserves the circadian glucocorticoid rhythm. *Am J Physiol*, 1989; **257**: E6–E14.
20. Horrocks PM, Jones AF, Ratcliffe WA, *et al.* Patterns of ACTH and cortisol pulsatility over twenty-four hours in normal males and females. *Clin Endocrinol (Oxf)*, 1990; **32**: 127–34.
21. Lightman SL, Conway-Campbell BL. The crucial role of pulsatile activity of the HPA axis for continuous dynamic equilibration. *Nat Rev Neurosci*, 2010; **11**: 710–18.
22. Lightman SL, Wiles CC, Atkinson HC, *et al.* The significance of glucocorticoid pulsatility. *Eur J Pharmacol*, 2008; **583**: 255–62.
23. Conway-Campbell BL, McKenna MA, Wiles CC, Atkinson HC, de Kloet ER, Lightman SL. Proteasome-dependent down-regulation of activated nuclear hippocampal glucocorticoid receptors determines dynamic responses to corticosterone. *Endocrinology*, 2007; **148**: 5470–7.
24. Eriksen HR, Olff M, Murison R, Ursin H. The time dimension in stress responses: relevance for survival and health. *Psychiatry Res*, 1999; **85**: 39–50.
25. Dallman MF, Pecoraro NC, La Fleur SE, *et al.* Glucocorticoids, chronic stress, and obesity. *Prog Brain Res*, 2006; **153**: 75–105.
26. Kaye J, Buchanan F, Kendrick A, *et al.* Acute carbon dioxide exposure in healthy adults: evaluation of a novel means of investigating the stress response. *J Neuroendocrinol*, 2004; **16**: 256–64.
27. Oster H, Challet E, Ott V, *et al.* The functional and clinical significance of the 24-hour rhythm of circulating glucocorticoids. *Endocr Rev*, 2017; **38**: 3–45.
28. Lightman SL. The neuroendocrinology of stress: a never ending story. *J Neuroendocrinol*, 2008; **20**: 880–4.
29. Scott LV, Dinan TG. Vasopressin and the regulation of hypothalamic-pituitary-adrenal axis function: implications for the pathophysiology of depression. *Life Sci*, 1998; **62**: 1985–98.
30. Dallman MF. Stress update. Adaptation of the hypothalamic-pituitary-adrenal axis to chronic stress. *Trends Endocrinol Metab*, 1993; **4**: 62–9.
31. Aguilera G. Regulation of pituitary ACTH secretion during chronic stress. *Front Neuroendocrinol*, 1994; **15**: 321–50.
32. Herman JP, Ostrander MM, Mueller NK, Figueiredo H. Limbic system mechanisms of stress regulation: hypothalamo-pituitary-adrenocortical axis. *Prog Neuropsychopharmacol Biol Psychiatry*, 2005; **29**: 1201–13.
33. Adler NE, Boyce T, Chesney MA, *et al.* Socioeconomic status and health: the challenge of the gradient. *Am Psychol*, 1994; **49**: 15–24.
34. Marmot MG, Rose G, Shipley M, Hamilton PJS. Employment grade and coronary heart disease in British civil servants. *J Epidemiol Community Health*, 1978; **32**: 244–9.
35. Swaab DF, Bao AM, Lucassen PJ. The stress system in the human brain in depression and neurodegeneration. *Ageing Res Rev*, 2005; **4**: 141–94.
36. Miller DB, O'Callaghan JP. Neuroendocrine aspects of the response to stress. *Metab Clin Exp*, 2002; **51**: 5–10.
37. McEwen BS. Central effects of stress hormones in health and disease: understanding the protective and damaging effects of stress and stress mediators. *Eur J Pharmacol*, 2008; **583**: 174–85.
38. Thomas RM, Peterson DA. Even neural stem cells get the blues: evidence for a molecular link between modulation of adult neurogenesis and depression. *Gene Expression*, 2007; **14**: 183–93.
39. Anacker C, Zunszain PA, Cattaneo A, *et al.* Antidepressants increase human hippocampal neurogenesis by activating the glucocorticoid receptor. *Mol Psychiatry*, 2011; **16**: 738–50.
40. Holsboer F. The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology*, 2000; **23**: 477–501.
41. Kyrou I, Chrousos GP, Tsigos C. Stress, visceral obesity, and metabolic complications. *Ann N Y Acad Sci*, 2006; **1083**: 77–110.
42. McEwen BS. Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load. *Metab Clin Exp*, 2006; **55**: S20–3.
43. Knutson KL, Van Cauter E. Associations between sleep loss and increased risk of obesity and diabetes. *Ann N Y Acad Sci*, 2008; **1129**: 287–304.
44. Buckley TM, Schatzberg AF. Review: on the interactions of the hypothalamic-pituitary-adrenal (HPA) axis and sleep: normal HPA axis activity and circadian rhythm, exemplary sleep disorders. *J Clin Endocrinol Metabolism*, 2005; **90**: 3106–14.
45. Henley DE, Russell GM, Douthwaite JA, *et al.* Hypothalamic-pituitary-adrenal axis activation in obstructive sleep apnea: the effect of continuous positive airway pressure therapy. *J Clin Endocrinol Metab*, 2009; **94**: 4234–42.
46. Eijsbouts AMM, Kempers MJ, van den Hoogen FH, *et al.* Hypothalamic-pituitary-adrenal axis activity in patients with rheumatoid arthritis. *Clin Exp Rheumatol*, 2005; **23**: 658–64.
47. Boonen E, Van Den Berghe G. Endocrine responses to critical illness: novel insights and therapeutic implications. *J Clin Endocrinol Metab*, 2014; **99**: 1569–82.
48. Gibbison B Keenan DM, Roelfsema F, *et al.* Dynamic pituitary-adrenal interactions in the critically ill after cardiac surgery. *J Clin Endocrinol Metabol*, 2020; **105**(5): 1–16. https://doi.org/10.1210/clinem/dgz206
49. Gibbison B, Spiga F. Dynamic pituitary-adrenal interactions in response to cardiac surgery. *Crit Care*, 2014; **43**: 791–800.
50. Miller T, Gibbison B, Russell GM. Hypothalamic–pituitary–adrenal function during health, major surgery, and critical illness. *BJA Educ*, 2017; **17**: 16–21.
51. Gagliardi L, Ho JT, Torpy DJ. Corticosteroid-binding globulin: the clinical significance of altered levels and heritable mutations. *Mol Cell Endocrinol*, 2010; **316**: 24–34.
52. Boonen E, Vervenne H, Meersseman P, *et al.* Reduced cortisol metabolism during critical illness. *N Engl J Med*, 2013; **368**: 1477–88.
53. Venkatesh B, Finfer S, Cohen J, *et al.* Adjunctive glucocorticoid therapy in patients with septic shock. *N Engl J Med*, 2018; **378**: 797–808.
54. Annane D, Renault A, Brun-Boisson C, *et al.* Hydrocortisone plus fludrocortisone for adults with septic shock. *N Engl J Med*, 2018; **378**: 809–18.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

12.1.2 Endocrinology in the Critically Ill

Greet Van den Berghe and Lies Langouche

Introduction

Critical illness is any condition requiring support of failing vital organs without which death would ensue. It is characterized by striking alterations in the hypothalamic-anterior-pituitary axes that follow a bi-phasic neuroendocrine response to critical illness. Acutely, an actively secreting pituitary together with the development of target-organ resistance results in low concentrations of peripheral effector hormones. These endocrine alterations may reduce energy and substrate expenditure, likely beneficial for short-term survival [1]. About 30% of critically ill patients do not recover within a few days, and enter a chronic phase of critical illness during which they remain dependent on vital-organ support. The high mortality observed during this prolonged phase is usually attributed to non-resolving failure of multiple organ systems and vulnerability to infectious complications, rather than to the type or severity of the initial disease for which patients were originally admitted to the intensive care unit (ICU). During the prolonged phase of illness, low serum levels of peripheral effector hormones are caused by uniform suppression of the neuroendocrine axes, primarily of hypothalamic origin. The prolonged phase of critical illness is further characterized by persistent hypercatabolism, despite feeding, which leads to a substantial loss of lean body mass in the presence of relative preservation of adipose tissue. This 'wasting syndrome' is likely to compromise vital functions and delay recovery, and as such contribute to the increased morbidity and mortality.

The different patterns of the neuroendocrine responses between the acute and prolonged phase of critical illness indicate the importance of a thorough understanding of the pathophysiology underlying these neuroendocrine changes. Indeed, erroneous extrapolation of the changes observed in the acute-disease state to the prolonged phase of critical illness has misled investigators to apply certain endocrine treatments that unexpectedly increased rather than decreased mortality [2, 3].

The Somatotropic Axis

Growth hormone (GH) is essential for growth during childhood, and for several mainly anabolic functions throughout life. The release of GH is stimulated by the hypothalamic GH-releasing hormone (GHRH), and is inhibited by somatostatin. Ghrelin, produced both in peripheral tissues and in the hypothalamic arcuate nucleus, is the third key factor in the complex physiological control of pulsatile GH release. The pulsatile nature of GH release, with peak serum levels alternating virtually undetectable troughs, is important for its metabolic effects. Apart from its direct actions, GH also exerts indirect effects that are mediated mainly through stimulation of insulin-like growth factor 1 (IGF-1) production, of which the bioactivity in turn is regulated by several IGF-binding proteins (IGFBPs).

The Somatotropic Axis in Acute Critical Illness

The first hours to days after an acute insult, such as surgery, trauma, or infection, are hallmarked by a dramatically changed GH profile. The GH pulse frequency is increased, and peak levels and interpulse concentrations are high. Concomitantly, a state of peripheral GH resistance develops, believed to be triggered by cytokines. Serum concentrations of IGF-1, GH-dependent IGFBP-3, and the acid-labile subunit (ALS) of the ternary complex are low during the acute phase of critical illness, in spite of the clearly enhanced GH secretion. An enhanced clearance of IGF-1, in part related to elevated circulating levels of small IGFBPs such as IGFBP-1, IGFBP-2, and IGFBP-6, also contributes to its low serum levels. These events are preceded by a decrease in serum growth hormone-binding protein (GHBP), reflecting reduced level of GH receptor expression in peripheral tissues [1, 4].

It remains unclear which factor ultimately controls the stimulation of GH release in response to acute stress. Nevertheless, it can be inferred that reduced negative feedback inhibition, caused by reduced expression of the GH receptor and subsequent low levels of circulating IGF-1, is the primary event inducing the abundant release of GH in the acute phase of illness. The high GH levels may then exert direct lipolytic, insulin-antagonizing, and immune-stimulating actions, resulting in increased fatty acid and glucose levels in the circulation, whereas the indirect, IGF-1-mediated effects of GH are attenuated. This would prioritize essential substrates such as glucose, free fatty acids, and amino acids toward survival rather than costly anabolism, which is mainly mediated by IGF-1 and considered less vital at this time. Therefore, from a teleological point of view, the response within the GH axis to acute illness seems highly appropriate in the struggle for survival.

The Somatotropic Axis in Prolonged Critical Illness

When recovery is not achieved within a few days and patients enter a prolonged phase of critical illness, the non-pulsatile GH fraction remains somewhat elevated, but although the number of pulses is still high, the pulsatile release of GH is strongly suppressed [1]. Nocturnal GH serum concentrations are scarcely elevated compared with the healthy, non-stressed condition, and substantially lower than in the acute phase of stress. Furthermore, although GH resistance may be partially reversed in the chronic phase, as indicated by increased serum levels of GHBP, the levels of IGF-1, IGFBP-3, and ALS are even lower in prolonged critically ill patients [1]. Furthermore, the pulsatile fraction of GH secretion correlates strongly with circulating levels of IGF-1, IGFBP-3, and ALS in this phase, meaning that the smaller the GH pulses, the lower the circulating levels of GH-dependent IGF-1 and ternary-complex-binding proteins. This clearly no longer represents a state of GH resistance, but rather suggests that loss of pulsatile GH release contributes to

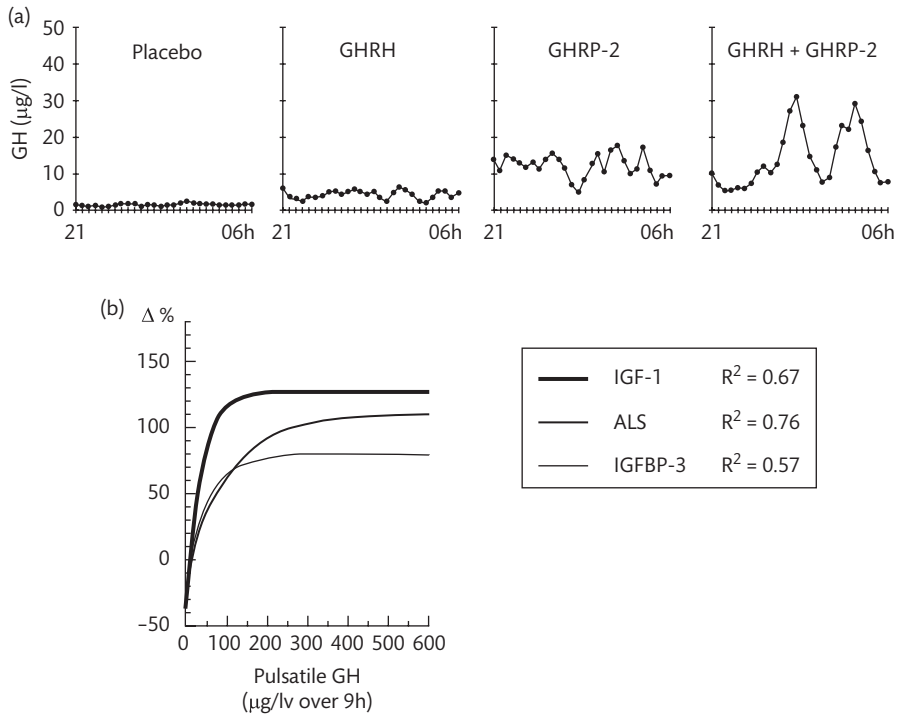


Figure 12.1.2.1 Effects of a growth hormone secretagogue on the somatotropic axis in prolonged critical illness. (A) Nocturnal serum GH profiles with continuous infusion of placebo, GHRH, GHRP-2 or GHRH + GHRP-2. (B) Exponential regression lines between pulsatile GH secretion and the changes in circulating IGF-1, ALS, and IGFBP-3.

the low levels of IGF-1 and binding proteins. Since the robust release of GH in response to GH secretagogues (GHSs) (Figure 12.1.2.1) excludes a possible inability of the somatotropes to synthesize GH, the origin of the relative hyposomatotropism is likely located within the hypothalamus. Furthermore, the release of GH in response to GHRH injection appears to be less pronounced than to GHRP-2 injection in prolonged critical illness, suggesting that hypothalamic deficiency or inactivity of endogenous GHRP-like GHSs is a more plausible cause of the hyposomatotropism than is GHRH deficiency [1]. However, both elevated and decreased concentrations of plasma ghrelin have been reported in critically ill patients [5, 6]. Of note, both studies only quantified total ghrelin, whereas the active—acetylated was clearly decreased [7].

The chronic relative GH deficiency is believed to contribute to the pathogenesis of the 'wasting syndrome' that characterizes prolonged critical illness. This is suggested by the observation that low serum levels of IGF-1 and ternary-complex-binding proteins are closely correlated to biochemical markers of impaired anabolism, such as low serum osteocalcin and leptin concentrations during prolonged critical illness [1].

Therapeutic Interventions: Treatment with GH During Critical Illness

Based on the assumption of sustained GH resistance in the presence of normal pituitary function during the catabolic condition of prolonged critical illness, pharmacological doses of GH were administered in an attempt to restore anabolism in ICU patients. However, a large multicentre study found that, instead of improving outcome, this intervention increased both morbidity and mortality [3]. Since it is clear now that the GH resistance of acute illness is at least partially resolved in the prolonged phase, it is likely that administration of such high doses may have evoked toxic side effects, exceeding any possible beneficial effect of this therapy. Indeed, high doses of GH can induce supranormal IGF-1 levels, excessive fluid retention, hypercalcaemia, and pronounced insulin resistance with hyperglycaemia.

Another treatment option might be the combined administration of GH and IGF-1, which are additive in their anabolic actions and neutralize each other's side effects [8]. Also, treatment with hypothalamic-releasing factors may be more effective and safer. Indeed, infusions of GHSs not only restored the pulsatile GH secretion, but also evoked an increase of IGF-1, IGFBP-3, and ALS, which is indicative of a restored peripheral responsiveness [1]. Evidence supporting a beneficial role of GHS comes from animal studies reporting that ghrelin supplementation reduces catabolic loss in critical illness. [9].

The Thyroid Axis

Thyroid hormones (TH) play a key role in the regulation of energy and substrate metabolism, and are essential for the stimulation of normal growth and development. The hypothalamic thyrotropin-releasing hormone (TRH) stimulates the pituitary to produce thyroid-stimulating hormone (TSH). TSH in turn drives the thyroid gland to synthesize and secrete TH. Deiodinases (D) are

responsible for the peripheral activation of thyroxine (T_4) to either the biological active tri-iodothyronine (T_3) (D1 and D2) or to the inactive reverse T_3 (rT_3) (D3). TRH and TSH secretion are controlled by negative feedback from the TH.

The Thyroid Axis in Acute Critical Illness

Early after the onset of severe physical stress, there is a rapid decrease in serum levels of T_3 and an increase of rT_3 levels, predominantly because of altered peripheral conversion of T_4 due to reduced D1 activity and increased D3 activity (Figure 12.1.2.2). TSH and T_4 levels are elevated briefly and subsequently return to normal, though in the more severely ill patients T_4 levels may also decrease (Figure 12.1.2.2). Although mean serum levels of TSH are normal, the normal nocturnal TSH surge is absent. Low T_3 levels persist beyond TSH normalization, a constellation referred to as 'the low T_3 syndrome'. The magnitude of the T_3 reduction within 24 hours reflects the severity of illness and correlates with mortality. Cytokines might play a role in the pathogenesis of the low T_3 syndrome. Other factors include low concentrations of TH binding proteins and inhibition of TH binding by compounds such as fatty acids and bilirubin, and low selenium levels which reduce deiodinase activity [10].

The thyroidal alterations observed during acute critical illness are strikingly similar to those seen with fasting. As critically ill patients suffer from poor nutritional intake, part of the observed thyroidal alterations could be evoked by this illness-induced fasting.

A large randomized controlled trial (RCT) reported that late initiation of parenteral nutrition, thereby generating a large caloric deficit, was associated with faster recovery and fewer complications, as compared with early full feeding [11]. Circulating levels of TSH, total T_4, T_3 and the T_3 to rT_3 ratio were all further reduced by late parenteral nutrition support [12]. Statistically, the further reduction in circulating T_4 with late feeding reduced part of the outcome benefit of nutrient restriction, whereas the further reduction of T3/rT3 ratio statistically explained part of the outcome benefit. The authors concluded that peripheral inactivation of T_3 with nutrient restriction during critical illness could be a beneficial adaptation, whereas the central lowering of T_4 could be deleterious. The benefit

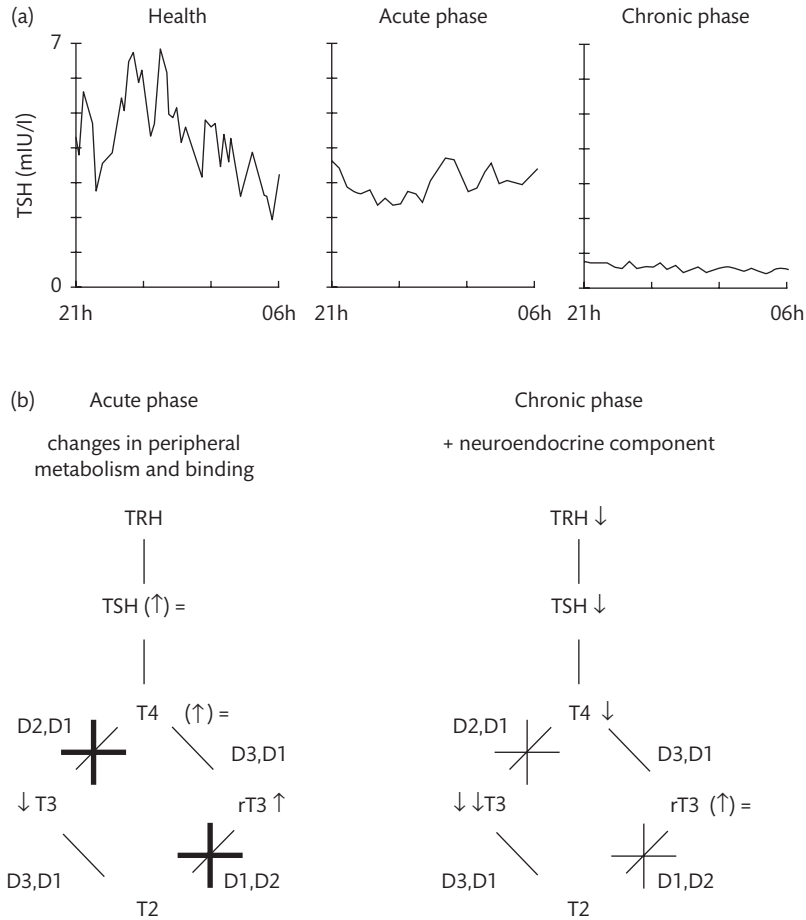


Figure 12.1.2.2 The thyroid axis in critical illness. (a) Nocturnal serum concentration profiles of thyrotropin (TSH) in critical illness are abnormal and differ between the acute and prolonged phase. (b) Overview of the major changes within the thyroid axis during the acute and prolonged phase of critical illness (black: normal regulation, grey: alterations induced by critical illness).

Panel A modified with permission from Van den Berghe G, de Zegher F, Bouillon R. Clinical review 95: Acute and prolonged critical illness as different neuroendocrine paradigms. *J Clin Endocrinol Metab* 1998; 83:1827–34 Copyright © 1998, Oxford University Press (Ref 1). Panel B reproduced with permission from Van den Berghe G. Novel insights into the neuroendocrinology of critical illness. *Eur J Endocrinol* 2000; 143:1–13 [45].

of the fasting-induced reduction in T_3 could, besides less energy expenditure, also optimize bacterial killing capacity via a direct impact of high D_3 activity locally within granulocytes [13, 14].

The Thyroid Axis in Prolonged Critical Illness

Patients who remain in the ICU and enter a prolonged phase of critical illness, now receiving full enteral and/or parenteral nutrition, show a different set of changes within the thyroid axis. In addition to the absent nocturnal TSH surge, the pulsatility in the TSH secretion pattern is dramatically reduced (Figure 12.1.2.2), with even lower serum levels of both T_3 and T_4. In human and an animal model of critical illness, reduced TRH expression in the hypothalamus occur, indicating a predominantly central origin of the suppressed thyroid axis [15, 16] that can be reactivated by TRH administration [17] (Figure 12.1.2.3). Furthermore, reduced GH secretagogue action may also be involved, as the pulsatility of the TSH secretion pattern is only improved when TRH is infused together with GHRP. Interestingly, simultaneous infusion of TRH and GHRP-2 not only increased TSH, T_4, and T_3 levels, but also prevented the rise in rT_3 seen with TRH alone (Figure 12.1.2.3). These results suggest that peripheral deiodinase activity is under joint control of the somatotropic and thyroid axes [1, 18].

The regulation of TH action at the level of the thyroid hormone receptor (TR) is also changed during prolonged critical illness. Alternative splicing gives rise to two TR isoforms, with TR-1 being a bona fide T_3 receptor, and TR-2 acting as a dominant negative isoform. An inverse correlation was observed between the T_3/rT_3 ratio and the TR-1/TR-2 ratio in liver biopsies of prolonged critically ill patients. Furthermore, sicker and older patients presented with higher TR-1/TR-2 ratios compared to less sick and younger ones [19]. Furthermore, in peripheral tissues such as muscle, lung, and liver-residing macrophages, D2 expression is upregulated [20–22]. These findings indicate that during prolonged critical illness, peripheral tissues seems to adapt to the low T_3 levels by increasing cellular sensitivity and local tissue levels.

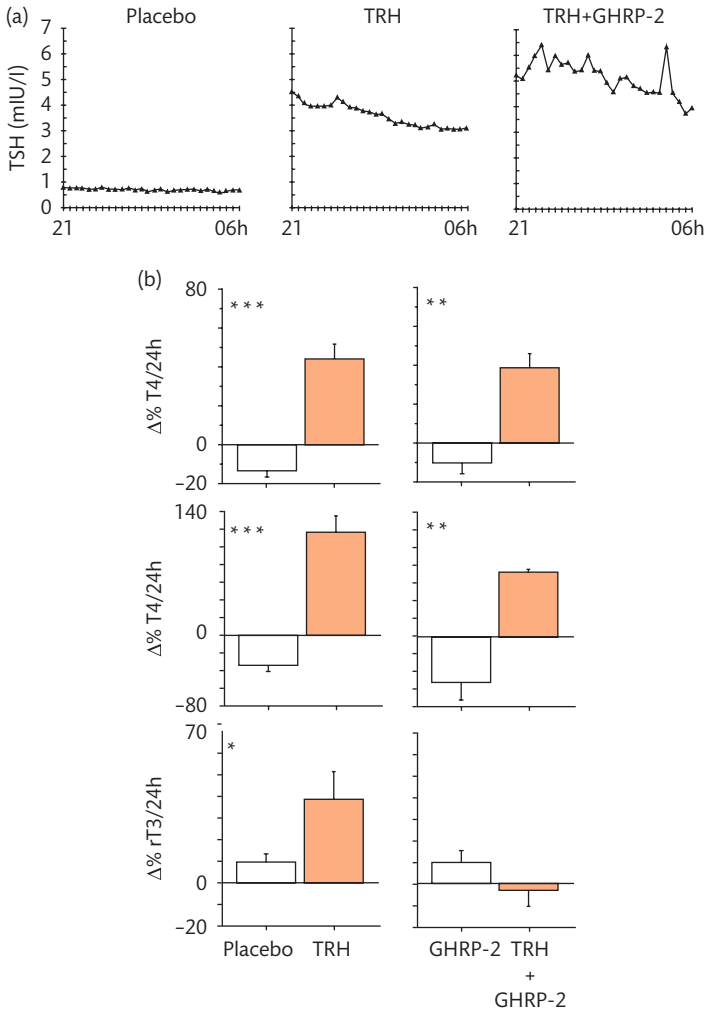


Figure 12.1.2.3 Effects of TRH and GHRP-2 on the thyroid axis in prolonged critical illness. (a) Nocturnal serum TSH profiles with continuous infusion of placebo, TRH, or TRH + GHRP-2. (b) Continuous administration of TRH, alone or together with GHRP-2, induces a significant rise in serum T_4 and T_3 within 24h. Reverse T_3 is increased after infusion of TRH alone, but not after co-infusion with GHRP-2.

Modified with permission from Van den Berghe G, de Zegher F, Bouillon R. Clinical review 95: Acute and prolonged critical illness as different neuroendocrine paradigms. J Clin Endocrinol Metab 1998; 83:1827-1834. Copyright © 1998, Oxford University Press (Ref 1). As adapted from Greet Van den Bergheet al. Neuroendocrinology of Prolonged Critical Illness: Effects of Exogenous Thyrotropin-Releasing Hormone and Its Combination with Growth Hormone Secretagogues. *The Journal of Clinical Endocrinology & Metabolism* (1998) 83 (2): 309–319, doi:10.1210/jcem.83.2.4575

Therapeutic Interventions: Treatment with Thyroid Hormone or Releasing Factors During Prolonged Critical Illness

The acute changes within the thyroid axis, uniformly present in all types of acute illnesses, and concomitant fasting, is considered a beneficial and adaptive response that does not warrant intervention. Nevertheless, it remains controversial whether and when correction of the low serum and tissue concentrations of T_3 in critically ill patients is beneficial. Pioneering studies using T_4 administration have so far failed to demonstrate clinical benefit within the ICU [23], although this is likely due to the impaired conversion of T_4 to T_3 caused by reduced D1 activity. Short-term intravenous T_3 administration to patients during elective coronary bypass grafting has been shown to improve postoperative cardiac function [24], but the used doses resulted in supranormal serum levels. Administration of replacement T_3 doses during paediatric cardiac surgery improved postoperative cardiac function [25]. However, a benefit of T_3 treatment in iatrogenic, dopamine-induced hypothyroidism does not provide evidence of clinical benefit for low T_3 levels that are characteristic of prolonged critical illness.

Rather than administration of TH, a safer method for treatment of illness-associated hypothyroidism may be the infusion of hypothalamic-releasing factors, since this preserves the normal feedback systems. Indeed, by continuous infusion of TRH in combination with a GH secretagogue, not only TH levels were restored to normal physiological levels, but also markers of hypercatabolism were reduced [1].

The Gonadal and Lactotropic Axis

Hypothalamic gonadotropin-releasing hormone (GnRH) is secreted in pulses and stimulates the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the pituitary. In women, LH mediates ovarian androgen production, whereas FSH drives the aromatization of androgens to oestrogens in the ovary. In men, LH stimulates the testicular production of androgens, whereas the combined action of FSH and testosterone supports spermatogenesis. In turn, sex steroids exert negative feedback control on GnRH and gonadotropin secretion. Several other hormones and cytokines are involved in the complex regulation of the gonadal axis [26]. Prolactin is produced and secreted by the pituitary in a pulsatile and diurnal pattern. The main function of prolactin is to stimulate lactation, but it is also presumed to have immune-enhancing properties. Physiological regulation of prolactin secretion is largely under the control of dopamine, although it can be modulated by several other prolactin inhibiting and releasing factors [27].

The Gonadal and Lactotropic Axis in Acute Critical Illness

As most female patients in the ICU are in the menopausal stage, clinical data on the changes within the female gonadal axis are scarce. Acute physical stress in men causes an immediate fall in serum testosterone, even though LH levels are elevated. This observation suggests an immediate suppression of anabolic androgen production in Leydig cells, as an attempt to reduce energy consumption and conserve substrates for more vital functions. The exact cause remains unclear, but again, inflammatory cytokines might play a role [1]. Acute physical stress causes prolactin levels to rise, which may contribute to altered immune function during critical illness. This increase is possibly mediated by vasoactive intestinal peptide, oxytocin, and dopaminergic pathways, but also cytokines or as-yet uncharacterized factors may be involved [1].

The Gonadal and Lactotropic Axis in Prolonged Critical Illness

When critical illness prolongs, more dramatic changes develop within the male gonadal axis. Circulating levels of testosterone become extremely low, while mean LH concentrations and pulsatile LH release are suppressed. Exogenous GnRH is only partially and transiently effective in correcting these abnormalities, suggesting also peripheral alterations within the male gonadal axis. Indeed, an increased aromatization of adrenal androgens to oestrogens in critically ill patients has been observed [28].

In the prolonged phase of critical illness, the pulsatile fraction of prolactin release becomes suppressed and mean serum prolactin levels are reduced compared to the acute phase. Exogenous dopamine further suppresses prolactin secretion and concomitantly aggravates T-lymphocyte dysfunction and impaired neutrophil chemotaxis [1, 2]. This suggests that the blunted prolactin secretion contributes to the immune suppression or increased susceptibility to infections associated with prolonged critical illness.

Therapeutic Interventions: Sex Steroid Substitution Therapy During Critical Illness?

It remains unclear whether the profound hypoandrogenism seen in male critically ill patients reflects adaptation or pathology. Pioneering studies evaluating the use of androgens in prolonged critical illness failed to demonstrate any conclusive clinical benefit [29, 30]. Exogenous pulsatile GnRH administration in prolonged critically ill men partially overcomes the hypogonadotropic hypogonadism. Moreover, when GnRH pulses were given together with GHRP2 and TRH infusion, target-organ responses, and anabolic effects followed. Despite its immune-enhancing properties, prolactin is currently not available for therapy. Further studies are required to evaluate the therapeutic potential of TRH-induced prolactin release for optimizing immune function during prolonged critical illness.

The Adrenal Axis

Diurnal cortisol release from the adrenal cortex is induced by adrenocorticotropic hormone (ACTH or corticotropin) which is produced by the pituitary under the control of the hypothalamic corticotropin releasing hormone (CRH). In turn, cortisol exerts negative feedback control on both CRH and ACTH. Although only free cortisol is biologically active, more than 90% of circulating cortisol is bound to binding proteins such as corticosteroid-binding globulin (CBG) and, to a lesser extent, albumin.

The Adrenal Axis in Acute Critical Illness

In the early phase of critical illness, the diurnal variation in cortisol secretion is lost and cortisol levels rise in response to the

severity of illness, although plasma ACTH concentrations are already low when admitted into the ICU [31]. In addition to the hypercortisolism, CBG levels are substantially decreased, resulting in proportionally much higher increases in free cortisol. The stress-induced hypercortisolism suppresses anabolism and promotes catabolism, to ensure acute provision of energy to vital organs such as the brain. In addition, it offers hemodynamic advantages in the fight-and-flight reflex by induction of fluid retention and sensitization of the vasopressor response to catecholamine, and it protects against excessive inflammation by suppression of the inflammatory response.

An appropriate activation of the hypothalamic–pituitary–adrenal axis and cortisol response to critical illness appears to be essential for survival, since both very high and very low cortisol levels have been associated with increased mortality.

The Adrenal Axis in Prolonged Critical Illness

In the prolonged phase of critical illness, hypercortisolism is usually sustained, but serum ACTH levels remain low (Figure 12.1.2.4). The dissociation of ACTH and cortisol release suggests a change in adrenal sensitivity to ACTH or the emergence of ACTH-independent pathways such as cytokines, neuropeptides, or vasoactive factors [32]. However, it was recently demonstrated that daytime cortisol production was only mildly elevated, whereas nocturnal production was even suppressed [31, 33]. Moreover, cortisol breakdown was substantially reduced attributable to suppressed expression and activity of A-ring reductases in the liver 11β-hydroxysteroid dehydrogenase type 2 in kidney [31].

Whether the persisting elevation in cortisol is exclusively beneficial remains uncertain. In theory, it could increase the susceptibility to infectious complications, impair wound healing and myopathy, complications that are frequently observed during prolonged critical illness. Teleological, sustaining elevated cortisol levels could be an energy-sparing mechanism during sustained stress. However, this may reduce ACTH levels through negative feedback which in turn might induce adrenal atrophy, which predisposes to adverse outcome. In severely ill critically ill patients, plasma cortisol levels are no longer high beyond 4 weeks in the ICU, at a time when ACTH levels are also low [34]. Post-mortem examination of adrenal glands harvested from

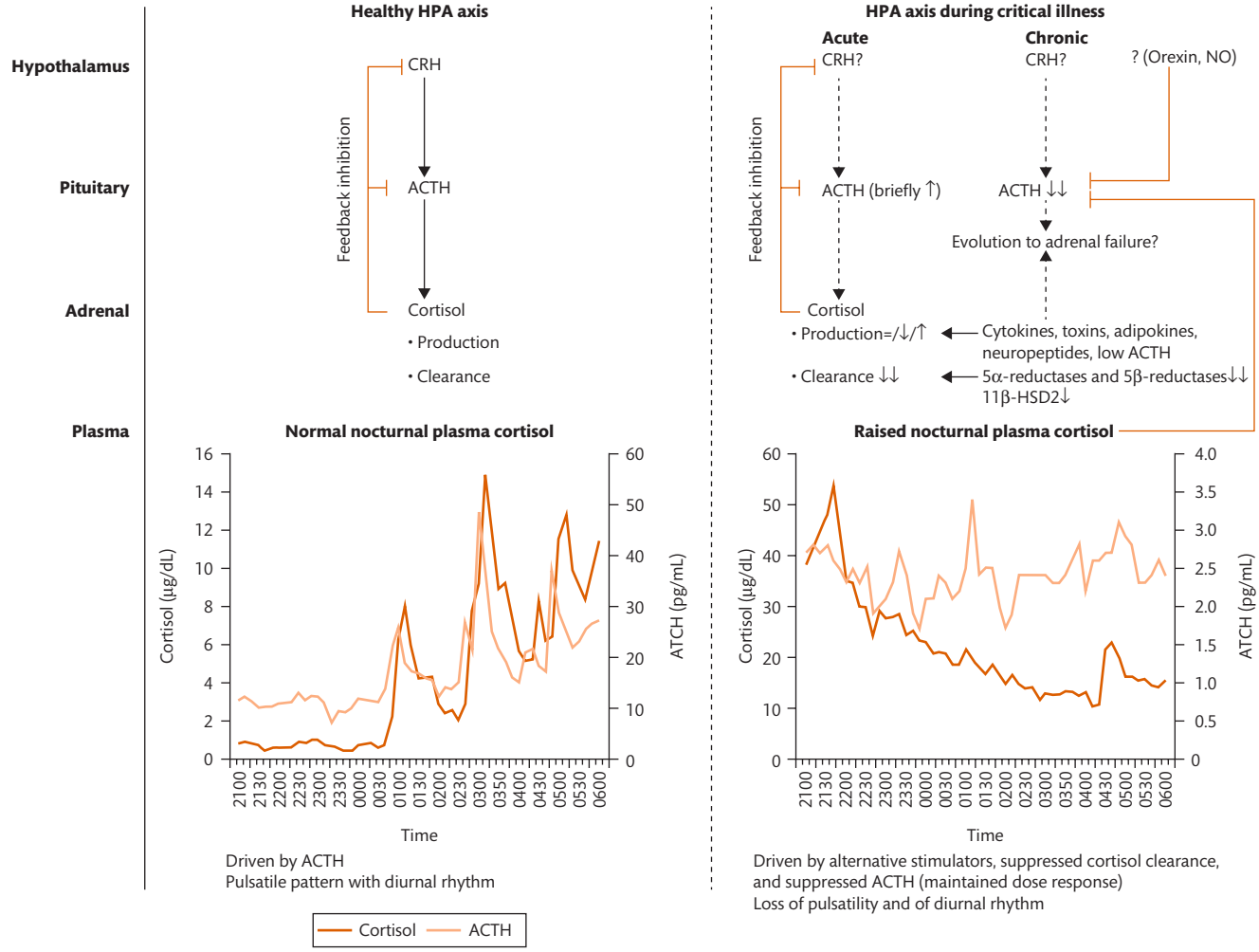


Figure 12.1.2.4 Overview of the hypothalamic–pituitary axis regulation in health and during critical illness. Solid lines represent proven regulation. Dotted lines represent regulation that is still speculated and needing additional research. Question marks represent an unknown part of the pathway.

prolonged critically ill patients reveal distorted architecture, lipid droplet depletion, and suppressed ACTH-regulated gene expression suggesting that sustained lack of ACTH may contribute to the risk of adrenal insufficiency in long-stay ICU patients [35].

Therapeutic Interventions: Treatment with Thyroid Hormone or Releasing Factors During Prolonged Critical Illness

In patients with previously diagnosed primary or central adrenal insufficiency, or previously treated with systemic glucocorticoids, glucocorticoid treatment should be started or continued during critical illness. A condition labelled 'relative' adrenal failure also has been suggested by experts, indicating a 'relative exhaustion' or 'insufficiently activated' adrenal cortex, insufficient to cope with the level of stress of septic shock. However, the results of clinical trials of high doses of glucocorticoids in critically ill patients are conflicting with some reporting beneficial [36, 37], ineffective [38, 39], or harmful outcomes [2]. Glucocorticoid treatment downregulates expression of the glucocorticoid receptor, and thus glucocorticoid sensitivity, in the liver but not in muscle [40]. Therefore, steroid-induced side effects such as insulin resistance and catabolism, may develop in muscle, whereas suppression of the hepatic glucocorticoid receptor may be harmful as observed in an animal model [41]. Another controversial issue regarding 'relative' adrenal failure in acute sepsis is the dose and duration of treatment once it has been initiated. Indeed, high dose long-term glucocorticoid administration will conceivably worsen the loss of lean tissue, increase the risk of polyneuropathy and myopathy, extend the ICU dependence, and increase the susceptibility to potentially lethal complications.

The term 'relative' adrenal insufficiency therefor became quite controversial. In the recently published guidelines on the diagnosis and management of critical illness-related corticosteroid insufficiency, the task force was unable to reach agreement on how to reliably diagnose such insufficiency [42]. The ACTH stimulation test is unreliable because of a 40% increased cortisol distribution volume in critical illness [31, 34].

Implications for Clinical Practice

The anterior pituitary responds biphasic to severe stress of illness and trauma (Figure 12.1.2.5). In the acute phase it is actively secreting, but target organs become resistant and concentrations of most peripheral effector hormones are low. In contrast, prolonged critical illness is characterized by a uniform suppression, predominantly of hypothalamic origin, of the neuroendocrine axes. These alterations contribute to low serum levels of the respective target-organ hormones. Although the differentiation between beneficial and harmful neuroendocrine responses to critical illness is difficult, it is important before considering any therapeutic intervention. The hypercatabolic reaction during acute critical illness is probably beneficial and, as such, provides no evidence that supports intervention. In prolonged critical illness, however, sustained hypercatabolism may compromise vital functions, cause weakness, and delay or hamper recovery. Theoretically, during this phase, a strategy of therapeutic intervention to correct these abnormalities could improve survival. In view of the adverse outcome of single hormone treatment strategies, high doses of either GH or glucocorticoids appeared to aggravate insulin resistance and hyperglycaemia that usually develop during critical illness. Importantly, although it had long been widely accepted that stress-induced hyperglycaemia might be beneficial to organs that largely rely on glucose for energy supply but do not require insulin for glucose uptake, strict blood glucose control with intensive insulin therapy has shown otherwise when maintained for at least a few days and avoiding excess hypoglycaemia [43, 44].

Although co-infusion of GHRP, TRH, and GnRH at least partially restores the three pituitary axes and reinitiates anabolism, the effect on survival remains unknown. However, the interaction among the different endocrine axes underlines the importance of jointly correcting all hypothalamic–pituitary defects rather than applying a single hormone treatment. However, evidence from appropriately designed and powered clinical trials is still lacking, making these and other interventions in the critically ill still experimental.

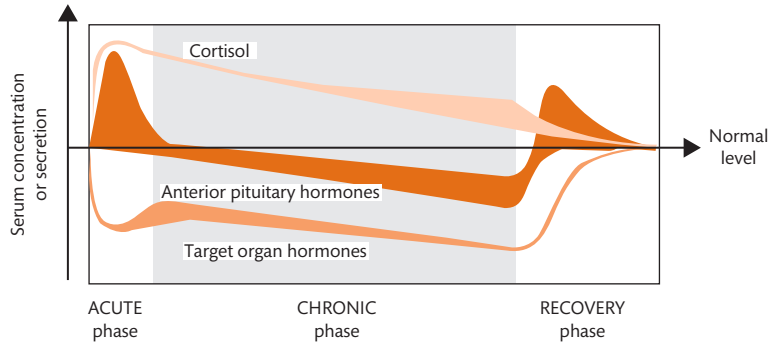


Figure 12.1.2.5 Simplified concept of the pituitary-dependent changes during the course of critical illness. In the acute phase, the secretory activity of the anterior pituitary is maintained or amplified, whereas anabolic target-organ hormones are inactivated. Cortisol levels are elevated. In the prolonged phase, impaired hormone secretion from the anterior pituitary allows the respective target-organ hormones to decrease proportionally over time, with cortisol being a notable exception. The onset of recovery is characterized by restored sensitivity of the anterior pituitary to reduced feedback control.

REFERENCES

1. Van den Berghe G, de Zegher F, Bouillon R. Clinical review 95: acute and prolonged critical illness as different neuroendocrine paradigms. *J Clin Endocrinol Metab*, 1998; **83**: 1827–34.
2. Roberts I, Yates D, Sandercock P, *et al.* Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. *Lancet*, 2004; **364**: 1321–8.
3. Takala J, Ruokonen E, Webster NR, *et al.* Increased mortality associated with growth hormone treatment in critically ill adults. *N Engl J Med*, 1999; **341**: 785–92.
4. Ross R, Miell J, Freeman E, *et al.* Critically ill patients have high basal growth hormone levels with attenuated oscillatory activity associated with low levels of insulin-like growth factor-I. *Clin Endocrinol (Oxf)*, 1991; **35**: 47–54.
5. Koch A, Sanson E, Helm A, Voigt S, Trautwein C, Tacke F. Regulation and prognostic relevance of serum ghrelin concentrations in critical illness and sepsis. *Crit Care*, 2010; **14**: R94.
6. Nematy M, O'Flynn JE, Wandrag L, *et al.* Changes in appetite related gut hormones in intensive care unit patients: a pilot cohort study. *Crit Care*, 2006; **10**: R10.
7. Crona D, MacLaren R. Gastrointestinal hormone concentrations associated with gastric feeding in critically ill patients. *JPEN J Parenter Enteral Nutr*, 2012; **36**: 189–96.
8. Kupfer SR, Underwood LE, Baxter RC, Clemmons DR. Enhancement of the anabolic effects of growth hormone and insulin-like growth factor I by use of both agents simultaneously. *J Clin Invest*, 1993; **91**: 391–6.
9. Hill NE, Murphy KG, Singer M. Ghrelin, appetite and critical illness. *Curr Opin Crit Care*, 2012; **18**: 199–205.
10. Van den Berghe G. Non-thyroidal illness in the ICU: a syndrome with different faces. *Thyroid*, 2014; **24**: 1456–65.
11. Casaer MP, Mesotten D, Hermans G, *et al.* Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*, 2011; **365**: 506–17.
12. Langouche L, Vander Perre S, Marques M, *et al.* Impact of early nutrient restriction during critical illness on the nonthyroidal illness syndrome and its relation with outcome: a randomized, controlled clinical study. *J Clin Endocrinol Metab*, 2013; **98**: 1006–13.
13. Boelen A, Kwakkel J, Fliers E. Beyond low plasma T3: local thyroid hormone metabolism during inflammation and infection. *Endocr Rev*, 2011; **32**: 670–93.
14. Boelen A, Boorsma J, Kwakkel J, *et al.* Type 3 deiodinase is highly expressed in infiltrating neutrophilic granulocytes in response to acute bacterial infection. *Thyroid*, 2008; **18**: 1095–103.
15. Fliers E, Guldenaar SE, Wiersinga WM, Swaab DF. Decreased hypothalamic thyrotropin-releasing hormone gene expression in patients with nonthyroidal illness. *J Clin Endocrinol Metab*, 1997; **82**: 4032–6.
16. Mebis L, Debaveye Y, Ellger B, *et al.* Changes in the central component of the hypothalamus-pituitary-thyroid axis in a rabbit model of prolonged critical illness. *Crit Care*, 2009; **13**: R147.
17. Bacci V, Schussler GC, Kaplan TB. The relationship between serum triiodothyronine and thyrotropin during systemic illness. *J Clin Endocrinol Metab*, 1982; **54**: 1229–35.
18. Debaveye Y, Ellger B, Mebis L, Darras VM, Van den Berghe G. Regulation of tissue iodothyronine deiodinase activity in a model of prolonged critical illness. *Thyroid*, 2008; **18**: 551–60.
19. Thijssen-Timmer DC, Peeters RP, *et al.* Thyroid hormone receptor isoform expression in livers of critically ill patients. *Thyroid*, 2007; **17**: 105–12.
20. Mebis L, Langouche L, Visser TJ, Van den Berghe G. The type II iodothyronine deiodinase is up-regulated in skeletal muscle during prolonged critical illness. *J Clin Endocrinol Metab*, 2007; **92**: 3330–3.
21. Kwakkel J, Surovtseva OV, de Vries EM, Stap J, Fliers E, Boelen A. A novel role for the thyroid hormone-activating enzyme type 2 deiodinase in the inflammatory response of macrophages. *Endocrinology*, 2014; **155**: 2725–34.
22. Ma SF, Xie L, Pino-Yanes M, *et al.* Type 2 deiodinase and host responses of sepsis and acute lung injury. *Am J Respir Cell Mol Biol*, 2011; **45**: 1203–11.
23. Brent GA, Hershman JM. Thyroxine therapy in patients with severe nonthyroidal illnesses and low serum thyroxine concentration. *J Clin Endocrinol Metab*, 1986; **63**: 1–8.
24. Klemperer JD, Klein I, Gomez M, *et al.* Thyroid hormone treatment after coronary-artery bypass surgery. *N Engl J Med*, 1995; **333**: 1522–7.
25. Bettendorf M, Schmidt KG, Grulich-Henn J, Ulmer HE, Heinrich UE. Tri-iodothyronine treatment in children after cardiac surgery: a double-blind, randomised, placebo-controlled study. *Lancet*, 2000; **356**: 529–34.
26. Spratt DI. Altered gonadal steroidogenesis in critical illness: is treatment with anabolic steroids indicated? *Best Pract Res Clin Endocrinol Metab*, 2001; **15**: 479–94.
27. Samson WK, Taylor MM, Baker JR. Prolactin-releasing peptides. *Regul Pept*, 2003; **114**: 1–5.
28. Spratt DI, Morton JR, Kramer RS, Mayo SW, Longcope C, Vary CP. Increases in serum estrogen levels during major illness are caused by increased peripheral aromatization. *Am J Physiol Endocrinol Metab*, 2006; **291**: E631–8.
29. Angele MK, Ayala A, Cioffi WG, Bland KI, Chaudry IH. Testosterone: the culprit for producing splenocyte immune depression after trauma hemorrhage. *Am J Physiol*, 1998; **274**: C1530–6.
30. Ferrando AA, Sheffield-Moore M, Wolf SE, Herndon DN, Wolfe RR. Testosterone administration in severe burns ameliorates muscle catabolism. *Crit Care Med*, 2001; **29**: 1936–42.
31. Boonen E, Vervenne H, Meersseman P, *et al.* Reduced cortisol metabolism during critical illness. *N Engl J Med*, 2013; **368**: 1477–88.
32. Bornstein SR, Engeland WC, Ehrhart-Bornstein M, Herman JP. Dissociation of ACTH and glucocorticoids. *Trends Endocrinol Metab*, 2008; **19**: 175–80
33. Boonen E, Meersseman P, Vervenne H, *et al.* Reduced nocturnal ACTH-driven cortisol secretion during critical illness. *Am J Physiol Endocrinol Metab*, 2014; **306**: E883–92.
34. Peeters B, Meersseman P, Vander Perre S, *et al.* Adrenocortical function during prolonged critical illness and beyond: a prospective observational study. *Intensive Care Med*, 2018; **44**: 1720–9.
35. Boonen E, Langouche L, Janssens T, *et al.* Impact of duration of critical illness on the adrenal glands of human intensive care patients. *J Clin Endocrinol Metab*, 2014; **99**: 4214–22.
36. Annane D, Sebille V, Charpentier C, *et al.* Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA*, 2002; **288**: 862–71.

37. Annane D, Renault A, Brun-Buisson C, *et al.* Hydrocortisone plus fludrocortisone for adults with septic shock. *N Engl J Med*, 2018; **378**: 809–18.
38. Sprung CL, Annane D, Keh D, *et al.* Hydrocortisone therapy for patients with septic shock. *N Engl J Med*, 2008; **358**: 111–24.
39. Venkatesh B, Finfer S, Cohen J, *et al.* Adjunctive glucocorticoid therapy in patients with septic shock. *N Engl J Med*, 2018; **378**: 797–808.
40. Peeters RP, Hagendorf A, Vanhorebeek I, *et al.* Tissue mRNA expression of the glucocorticoid receptor and its splice variants in fatal critical illness. *Clin Endocrinol (Oxf)*, 2009; **71**: 145–53.
41. Jenniskens M, Weckx R, Dufour T, *et al.* The hepatic glucocorticoid receptor is crucial for cortisol homeostasis and sepsis survival in humans and male mice. *Endocrinology*, 2018; **159**: 2790–802.
42. Annane D, Pastores SM, Rochwerg B, *et al.* Guidelines for the diagnosis and management of critical illness-related corticosteroid insufficiency (CIRCI) in critically ill patients (Part I): Society of Critical Care Medicine (SCCM) and European Society of Intensive Care Medicine (ESICM) 2017. *Crit Care Med*, 2017; **45**: 2078–88.
43. Van den Berghe G, Wouters P, Weekers F, *et al.* Intensive insulin therapy in critically ill patients. *N Engl J Med*, 2001; **345**: 1359–67.
44. Van den Berghe G, Wilmer A, Hermans G, *et al.* Intensive insulin therapy in the medical ICU. *N Engl J Med*, 2006; **354**: 449–61.
45. Van den Berghe G. Novel insights into the neuroendocrinology of critical illness. *Eur J Endocrinol*, 2000; **143**: 1–13.
46. Boonen E, Bornstein SR, Van den Berghe G. New insights into the controversy of adrenal function during critical illness. *Lancet Diabetes Endocrinol*, 2015; **3**: 805–15.

12.1.3 Hormones and the Kidney

Melissa Nataatmadja, Yeoungjee Cho, and David W. Johnson

Introduction

Chronic kidney disease (CKD) is a major, growing, global public health problem that affects approximately 10–15% of the world's population [1]. Since the kidneys play a critical role in the synthesis, elimination, and actions of a variety of hormones (Table 12.1.3.1), endocrine abnormalities are common in CKD. This chapter will review the common endocrine systems affected in kidney disease.

Renin–Angiotensin–Aldosterone System

The renin–angiotensin system (RAS) plays a vital role in the maintenance of blood pressure homeostasis. Within the kidney, the juxtaglomerular apparatus synthesizes renin, which converts angiotensinogen to angiotensin I [2]. Angiotensin I is in turn converted by angiotensin converting enzyme (ACE) to angiotensin II, which binds to angiotensin II type 1 (AT1R) and type 2 receptors (AT2R) to promote vasoconstriction, renal sodium retention, and secretion of aldosterone, vasopressin, and endothelin [2]. Inappropriate intrarenal RAS activation, usually without elevated plasma renin activity or plasma angiotensin II levels, has been observed in a variety of animal models of kidney disease [2]. Consequently, it has been hypothesized that the beneficial effects of RAS-modifying medications, such as ACE inhibitors and angiotensin receptor blockers (ARB), on CKD progression and proteinuria may occur largely at a local intrarenal level [2].

CKD patients, especially those with advanced renal disease, may experience worsening of their renal function and/or hyperkalaemia following commencement of an ACE inhibitor or ARB, especially if they are poorly hydrated, unwell, known to have bilateral renal artery stenosis, or taking certain medications (e.g. diuretics or non-steroidal anti-inflammatory drugs [NSAIDs]). Current guidelines permit a reduction in estimated glomerular filtration rate (eGFR) of up to 25% within the first month or an increase in serum potassium of up to 6 mmol/L before consideration should be given to reducing or ceasing RAS blockers [3].

Aldosterone, a mineralocorticoid produced by the adrenal glands, activates the apical epithelial sodium channel (ENaC) in the distal nephron to facilitate sodium reabsorption by the principal cells and reduced potassium reabsorption and increased H^+ excretion by the α-intercalated cells [4]. It has also been shown to contribute to kidney damage in CKD, including tubulointerstitial fibrosis and inflammation, glomerulosclerosis, podocyte injury and mesangial cell proliferation [4]. These effects occur independently of angiotensin II and are accentuated in the presence of salt loading [4]. 'Aldosterone breakthrough', with elevated levels of aldosterone, can be observed in patients receiving ACE inhibitor or ARB therapy, and the addition of mineralocorticoid receptor blockers can successfully reduce proteinuria and blood pressure, although these effects are counterbalanced by increased risks of hyperkalaemia and gynaecomastia [5]. Consequently, the role of mineralocorticoid receptor blockers in preventing CKD progression currently remains uncertain [5].

Table 12.1.3.1 Commonly affected hormones in kidney disease

Hormones synthesized by the kidney	Hormones catabolized by the kidney	Hormones that act on the kidney
• Erythropoietin • Fibroblast growth factor 23 (FGF-23) • IGF-binding proteins (IGFBP) • Insulin-like growth factor-1 (IGF-1) • Renin	• Atrial natriuretic peptide (ANP) • Catecholamines • Corticosteroids • Glucagon • Growth hormone • Insulin • Prolactin • Sex hormones • Thyroid hormone • Vasopressin	• 1,25 dihydroxycholecalciferol ($1,25(OH)_2D_3$) • Aldosterone • Angiotensin II • ANP • Calcitonin • Catecholamines • Growth hormone • IGF-1 • Insulin • Parathyroid hormone (PTH) • Thyroid hormone • Vasopressin

Erythropoietin and Anaemia of Renal Disease

Anaemia is present in approximately 50% of patients with CKD and the vast majority of patients with end-stage kidney disease (ESKD). It is usually normocytic and normochromic and is associated with left ventricular hypertrophy, cardiovascular events, excess hospitalizations, reduced quality of life, and mortality. The main cause is deficiency of erythropoietin (EPO), a 34 kDa glycoprotein hormone that is produced by type 1 cortical fibroblasts within the kidney [6]. This hemopoietic growth factor regulates the proliferation and differentiation of erythroid progenitor cells through binding with EPO receptors on endothelial cell surfaces and induces stimulation of Jak2 tyrosine kinase and subsequent activation of multiple cellular pathways leading to inhibition of apoptosis [6]. Prior to its availability in the late 1980s, up to 25% of dialysis patients were transfusion-dependent [7].

In addition to EPO deficiency, a number of other factors may contribute to anaemia in patients with CKD, including absolute or functional iron deficiency, gastrointestinal bleeding, uraemia, hyperparathyroidism, inflammation, nutrient deficiencies (vitamin B_{12}, folate, vitamin C, and carnitine), ACE inhibitors, ARBs, aluminium overload, and erythropoiesis-stimulating agent (ESA)-induced antibody-mediated pure red cell aplasia [7]. Elevated serum hepcidin levels, perhaps due to increased expression by inflammatory cytokines and reduced renal clearance, may also contribute to anaemia of CKD by binding to ferroportin (FPN1), a cellular iron exporter expressed in reticuloendothelial system macrophages and duodenal enterocytes, resulting in internalization, loss of function, reticuloendothelial cell iron blockade, impaired dietary iron absorption, and ultimately functional iron deficiency [7].

The use of ESAs has become commonplace in the treatment of anaemia associated with CKD and can be administered either subcutaneously or intravenously (if the patient is if receiving haemodialysis). Human recombinant EPO (alpha, beta, omega, and delta) may be synthesized in human or animal cells and differ based on their carbohydrate moieties, though all have a similar half-life of 6 to 8 hours intravenously and 19–24 hours subcutaneously [8]. More recently, long acting ESAs have been developed, including darbepoetin alpha (half-life 25 hours if intravenous and 48 hours if subcutaneous) and continuous erythropoiesis receptor activator (CERA—half-life approximately 130 hours), which allow less frequent dosing [8]. While cost and ease of administration may influence the choice of ESA used, there is low certainty evidence that the use of different ESAs may make little or no difference to anaemia correction or adverse events [9].

The haemoglobin target to aim for with ESA treatment of anaemia of CKD has been the subject of considerable debate. The accumulated trial evidence to date suggests that targeting normal or high haemoglobin levels reduces the need for blood transfusion, does not have any other clear benefits on patient-level outcomes, and increases the risks of stroke, hypertension and vascular access thrombosis [10]. It remains uncertain whether the increased risk of adverse outcomes at normal or near-normal haemoglobin levels following ESA therapy is related to haemoglobin level, ESA dose-related toxicity, or other factors (e.g. greater iron requirements, inflammatory state, or concurrent illness). Consequently, the current Kidney Disease Improving Global Outcomes (KDIGO) Guideline recommendations state, 'in initiating and maintaining ESA therapy, we recommend balancing the potential benefits of reducing blood transfusions and anaemia-related symptoms against the risks of harm in individual patients (e.g. stroke, vascular access loss, hypertension) [11]. Specifically, the guidelines recommend initiating ESA therapy when the haemoglobin concentration is less than 100 g/L where correctable causes of anaemia have been addressed, that the haemoglobin target concentration with ESA use should not be greater than 130 g/L and, in general, the target should not be greater than 115 g/L [11]. As a result, most nephrologists will now use alternative treatments for anaemia (such as iron supplementation) first and thereafter resort to ESA administration with the primary goal of avoiding blood transfusion without raising haemoglobin levels too high.

Emerging novel therapies, such as orally active hypoxia inducible factor (HIF) stabilizers, EPO mimetic peptides and activin receptor ligand traps (e.g. sotatercept), may offer new treatment approaches for anaemia of CKD in the future [8].

Sexual Dysfunction

Sexual dysfunction is a highly prevalent, but under-recognized, complication of CKD [12]. The cause of sexual dysfunction in these patients is likely multifactorial in nature, due to not only their CKD, but also to comorbid illnesses and complications of CKD, such as diabetes, hyperprolactinemia, hyperparathyroidism, sex hormone imbalances, autonomic neuropathy, cardiovascular and peripheral vascular disease, veno-occlusive dysfunction, depression, and medication side effects [12, 13].

Box 12.1.3.1 Sex hormone abnormalities in men with CKD

- Reduced free and total testosterone
- Reduced Leydig cell response to HCG
- Normal sex hormone-binding globulin
- Elevated total plasma oestrogen
- Normal oestradiol
- Elevated LH
- Reduced LH per secretory burst
- Elevated follicle stimulating hormone (FSH) (lesser degree than LH, so LH/FSH increased)
- Reduced inhibin
- Appropriate increase in LH and FSH to clomiphene (competitive oestrogen/testosterone receptor blocker → prevents negative feedback to hypothalamus)

Male Sexual Dysfunction

Sexual dysfunction in CKD is likely due to a combination of structural and functional hypogonadism. Men with CKD exhibit detectable abnormalities in the pituitary–gonadal axis, even when GFR is only mildly reduced (see Box 12.1.3.1) [12]. Reduced free and total testosterone levels are observed in early CKD, and worsen with progressive falls in GFR [12]. These low testosterone levels also contribute to osteoporosis, anaemia, and reduced muscle mass [12]. Men with CKD have a blunted Leydig cell response to testosterone stimulation by human chorionic gonadotrophin (HCG) [12], and there is evidence for an inhibitory factor in uremic serum that is able to block the luteinizing hormone (LH) receptor at the level of the Leydig cells [12]. This blunted Leydig cell response usually resolves following kidney transplantation, although institution of dialysis treatment generally fails to normalize pituitary–gonadal disturbances and these abnormalities usually progress [12].

Men with CKD also experience impaired spermatogenesis, with decreased ejaculate volume, oligo- or azoospermia, impaired sperm maturation, and low sperm motility [12]. In addition, there may be interstitial fibrosis and calcification within the testes and seminiferous tubules and a failure of Leydig and Sertoli cells to show compensatory hypertrophy and hyperplasia [12]. Though oestrogen is usually elevated in ESKD, oestradiol levels are usually within the normal range despite elevated LH, suggesting functional gonadotrophin deficiency or resistance [12]. Follicle-stimulating hormone (FSH) elevation in men with ESKD is more variable compared to LH elevation. Increased elevation of FSH likely reflects a greater degree of damage to the seminiferous tubules and, thus, lower levels of inhibin production by Sertoli cells to feedback on FSH [12], and is associated with a lower recovery of spermatogenic function post-renal transplantation [12].

Erectile dysfunction in men with CKD is often related to vascular disease (including disease of cavernosal, proximal iliac and pudendal arteries causing effective 'steal' syndrome), autonomic neuropathy and multiple commonly used medications (see Box 12.1.3.2).

Clinical management of sexual dysfunction in CKD involves a thorough assessment for possible causes and contributors, and a multipronged treatment approach. Potentially reversible factors, such as medication effect, must be excluded, and non-organic aetiology (such as depression) should be considered and addressed. Anaemia should be corrected, hyperparathyroidism treated, and nutritional state optimized [12]. ESAs may enhance sexual function and normalize the pituitary–gonadal axis in CKD, with some studies demonstrating reduced serum LH and FSH concentrations, reduced prolactin levels, and increased plasma testosterone levels [12]. In addition, zinc supplementation has been shown in increase testosterone levels in haemodialysis patients [12].

Box 12.1.3.2 Common medications used in CKD that are implicated in male erectile dysfunction

- Alpha blockers
- Beta blockers
- Methyldopa
- Clonidine
- Antidepressants
- Metoclopramide

Testosterone replacement in uremic men does not usually produce a significant improvement in libido, sexual potency, or erectile function [12], although it is successful in increasing testosterone levels and reducing LH and FSH [14]. Potential benefits must be weighed up against risks, including adverse effects on lipid profile, hepatotoxicity, polycythemia, prostatic enlargement, and risk of prostate cancer.

Phosphodiesterase-5 inhibitors (PDE5i), such as sildenafil, are frequently used as first line therapy for erectile dysfunction with normal libido [15]. Although PDE5is increase testosterone and reduce LH and FSH release, they are generally not effective in restoring libido and potency [15, 16]. Its use is contraindicated in patients using nitrates, and should be used with caution in patients with coronary artery disease, and haemodialysis patients prone to hypotension.

Alprostadil treatment, vacuum devices, and penile prostheses may be useful in those patients with neurogenic or vascular causes of impotence [12].

Renal transplantation likely provides the best treatment for male sexual dysfunction through restoration of the normal hypothalamic-pituitary axis, normalization of levels of gonadotrophin releasing hormone (GnRH), LH release, testosterone, and prolactin, and improvements in sperm density and motility (though not morphology) [12]. Improvements in libido and global sexual dysfunction are variable post-kidney transplantation in both male and female recipients [12].

Female Sexual Dysfunction

Women with CKD also display disturbances in the pituitary–gonadal axis at an early stage (see Box 12.1.3.3), and sexual dysfunction becomes increasingly common as CKD progresses.

Box 12.1.3.3 Sex hormone abnormalities in women with CKD

- Failure to increase basal body temperature at expected time of ovulation
- Absence of preovulatory LH and oestradiol surge
- Elevated LH and FSH
- Appropriate increase in LH and FSH to clomiphene (competitive oestrogen/testosterone receptor blocker → prevents negative feedback to hypothalamus)
- Reduced oestradiol

The most common disturbance of reproductive function seen in women with CKD is menstrual disturbance and amenorrhea. Although the preovulatory LH and oestradiol surges are often absent, administration of clomiphene increases LH and FSH secretion, suggesting an intact negative oestradiol feedback pathway [12]. Anovulation and infertility are common in patients with CKD and ESKD [17] and frequent anovulation leads to failure of progesterone secretion by the corpus luteum. Thus, for women who wish to resume menses, a progestogen agent can be administered towards the end of the menstrual cycle, and for women with menorrhagia, administration of a progestogen throughout the menstrual cycle can terminate flow [12]. Over time, chronic proliferation of the endometrium due to frequent anovulation and lack of cyclical progestational endometrial changes may predispose to endometrial hyperplasia and malignancy [17]. Intermittent use of a prostagen may be implemented to avoid this and these patients should be followed by a gynaecologist.

Women also report decreased libido and reduced ability to achieve an orgasm [12]. Oestrogen-deficient patients may experience improved sexual desire and restoration of normal menses, as is the case in patients without CKD, although concerns remain about adverse cardiovascular, cerebrovascular, and malignancy-related side effects. Sildenafil, while possibly useful for women with isolated sexual arousal disorder, does not appear to be useful in those women with decreased libido [18]. Vaginal atrophy and dyspareunia can occur due to chronically low oestradiol levels, and local oestrogen treatments and lubricant use may be useful for symptom management [19].

As is the case for men, clinical management of sexual dysfunction in women with CKD should include optimization of dialysis, treatment of anaemia and hyperparathyroidism, review of medications, and exclusion of non-organic causes. Renal transplantation reverses most hormonal abnormalities in premenopausal women, and can restore regular menstrual cycles [12].

Successful pregnancy while receiving dialysis treatment is uncommon, with low conception rates (2%) and increased risks of miscarriage, intrauterine death, low birth weight and premature delivery [20]. Frequent, long dialysis sessions, such as in nocturnal dialysis, aiming for >36 hours per week increases the likelihood of a successful pregnancy [20]. Kidney transplantation is the most effective means of restoring both sexual desire and fertility in women with CKD [12].

Prolactin

Hyperprolactinemia is common in patients with ESKD, and is primarily due to increased production, with some reduced renal clearance and secondary hyperparathyroidism [12]. The clinical significance of hyperprolactinemia in men with CKD is poorly understood, and the effects of this on libido and sexual potency are inconsistent [21]. In women with CKD, hyperprolactinemia may contribute to sexual dysfunction and galactorrhoea [12].

Prolactin secretion in patients with CKD is resistant to suppression by dopamine infusion or oral L-dopa, and is similarly resistant to stimulation by arginine infusion, thyrotropin-releasing hormone infusion, and insulin-induced hypoglycaemia [12]. Bromocriptine, though effective in reducing prolactin levels, rarely results in restoration of normal menses in women with CKD, and is poorly tolerated due to its side effect profile [12]. Suppression of parathyroid hormone (PTH) with calcitriol has been shown to reduce prolactin and increase testosterone levels [22], although the reported effects on sexual function are mixed [12].

Growth Hormone and Insulin-Like Growth Factor

In the setting of CKD and kidney function impairment, growth hormone (GH) levels are elevated due to increased pituitary release as well as reduced renal metabolic clearance [23]. The elevated GH levels do not translate to a greater GH effect. There is GH resistance which arises from reduced tissue expression of GH receptors and from impairment in the signalling of the GH receptor [23]. The levels of insulin-like growth factor (IGF), as quantified by immunoassays, are generally normal in CKD. However, the biological activity of IGF-1 is reduced. The bioactivity of IGF is influenced by a family of IGF-binding proteins (IGFBPs), the production of which is increased and the clearance of which is reduced in CKD [23, 24].

Growth restriction is a common complication of paediatric CKD and is associated with altered GH/IGF-1 function, altered renin–angiotensin system programming, and renal structural abnormalities, such as low nephron count [25]. The pubertal height gain in children with CKD is approximately 65% that of healthy children [25]. Recombinant human GH (rhGH) is a commonly used treatment in affected patients, with increase in height being positively associated with older age, higher GFR and higher pretreatment growth rate [25]. Following kidney transplantation, rhGH has been used to effectively increase 'catch up' growth with no reported increases in allograft failure rate or adverse events, although routine use is not recommended [26]. The use of rhGH should be avoided in critically ill patients, where it has been associated with significant increases in mortality through an unclear mechanism [27]. Other risks of rhGH use include increased risks of idiopathic intracranial hypertension and type 2 diabetes, for which patients with CKD are already at increased risk.

Low serum IGF-1 concentration has been associated with body composition changes, muscle wasting, low bone mineral density, and increased mortality risk in patients with CKD, although replacing IGF-1 has not been shown to attenuate this risk [28]. While rhIGF-1 treatment has been demonstrated to produce an increase in growth velocity in children with GH-receptor deficiency or antibodies, its usefulness in patients with CKD is less clear given their state of IGF-1 resistance. Combination therapy of rhGH and rhIGF-1 is effective for growth promotion in uremic animal studies, but has not yet been evaluated in human studies [28].

Thyroid

Thyroid abnormalities, including hypothyroidism, goitre, thyroid nodules, and thyroid malignancies, are more common patients with CKD or kidney transplants compared to the general population [29, 30].

Patients with CKD appear to be at particularly increased risk of developing subclinical hypothyroidism and manifest a variety of thyroid axis abnormalities, including blunted response to

thyrotropin-releasing hormone (TRH), and disturbed diurnal variation of thyroid-stimulating hormone (TSH) with reduced or absent evening peaks and shorter and smaller pulsatile releases [30]. Reduced GFR also leads to impaired iodide excretion, total body iodide accumulation and reduced thyroid hormone production [31]. Despite reduced total thyroxine (T_4) levels, free T_4 is usually normal in patients with ESKD, likely due to reduced binding to serum carrier proteins as a result of elevated inflammatory cytokines (e.g. IL-1β and TNF-α), uremic toxins (e.g. indoxyl sulphate and hippuric acid) and concomitant medications (e.g. furosemide, NSAIDs, and heparin) [31]. Dialysis does not produce any significant effect on thyroid hormone metabolism [32]. Total and free triiodothyronine (T_3) are also often reduced despite preserved clearance of T_3 due to decreased conversion of T_4 to T_3 [30]. This may represent a metabolic adaptation to conserve energy and minimize protein-nitrogen catabolism, such that thyroid hormone therapy should not be used except in patients with diagnosed hypothyroidism [33]. Hypothyroidism may in turn increase the risk of CKD through effects on cardiac output, peripheral vascular resistance, and intrarenal vasoconstriction, resulting in reduced renal plasma flow and GFR [30]. In those patients with true hypothyroidism, thyroid hormone replacement may play a role in slowing the decline of kidney function in patients with CKD [30] and may even result in an increase in GFR over time [30]. Treating hypothyroidism may also mitigate some of the excess cardiovascular risk associated with CKD [30]. When using thyroid hormone replacement in patients with CKD, it is important to note that T_3 and T_4 may be bound in the gastrointestinal tract by commonly used CKD medications, such as kayexalate [34] and aluminium-based phosphate binders [35].

Hyperthyroidism occurs in patients with CKD with a similar frequency to that observed in the general population. However, biochemical features of hyperthyroidism may be masked due to reduced T_4 binding to carrier proteins and reduced T_4 conversion to T_3, such that a normal or low T_3 may not exclude hyperthyroidism [36].

CKD may also coexist with thyroid disease in the case of autoimmune disease, such as Hashimoto's thyroiditis and Grave's disease, which may be associated with glomerulonephritis [30]. In addition, some thyroid disease treatments have been associated kidney disease. Propylthiouracil has been linked to the development of antineutrophil cytoplasmic antibody (ANCA)-positive glomerulonephritis, while radioiodine may induce membranous nephropathy [30].

In patients with thyroid malignancies or hyperthyroidism where treatment with radioactive iodine is warranted, it is important to note that renal impairment interferes with body excretion of radioactive iodine Na^{131} and a prolonged half-life is seen in both dialysis-dependent and non-dialysis-dependent patients with ESKD due to reduced clearance [31]. Thus, patients receiving maintenance haemodialysis and peritoneal dialysis require significant reductions in dosage to avoid complications of excessive radiation [30].

Insulin Resistance

Insulin resistance increases the risk of developing kidney disease, while the kidney plays an important role in glucose homeostasis. Renal gluconeogenesis accounts for approximately 40% of all gluconeogenesis and 20% of overall endogenous glucose release in the body [37]. Like hepatic gluconeogenesis, renal gluconeogenesis is not appropriately suppressed by insulin in patients with diabetes [38].

Kidney disease also predisposes to insulin resistance and diabetes, and insulin resistance exists even in the early stages of kidney disease [39]. This is mainly due to reduced peripheral target tissue sensitivity to insulin, and patients with CKD often exhibit elevated glucose levels during fasting despite hyperinsulinemia [39]. Glycogen synthase activity is preserved, despite an attenuated increase in insulin-stimulated glucose transport from muscles [40]. *In vitro* studies of skeletal muscle from uremic patients have indicated that skeletal muscle receptor number, insulin binding, β-subunit receptor autophosphorylation, and tyrosine kinase activity are preserved, suggesting that the observed peripheral insulin resistance is likely due to a post-receptor defect in the insulin pathway [39]. There are many other factors commonly present in patients with CKD that interfere with the action of insulin, and may be amenable to treatment with synbiotics, medications, or dialysis, resulting in improved glucose tolerance [39]. Some of these include: (1) metabolic acidosis (reduces post-receptor signalling in insulin-signalling pathway); (2) vitamin D deficiency (involved in stimulating expression of insulin receptors); (3) secondary hyperparathyroidism (may result in suppressed insulin-stimulated glucose uptake via the cAMP pathway); (4) hyperuricaemia (reduces glucose uptake by peripheral tissues via reduced insulin-mediated nitric oxide release); and (5) gut dysbiosis (leads to bacterial production of uremic toxins, including indoxyl sulphate and p-cresyl sulphate, which alter insulin signalling) [39].

Insulin resistance is a major risk factor for cardiovascular morbidity and mortality in patients with CKD, especially when combined with other risk factors, such as smoking and physical inactivity [39]. As a result, early and intensive management of diabetes and impaired glucose tolerance should be performed in appropriate individuals, together with risk factor modification.

CKD Mineral and Bone Disorder, Hyperparathyroidism, and Vitamin D

CKD mineral and bone disorder (CKD-MBD) has been defined by the KDIGO Guidelines group as any one or a combination of the following: (1) Laboratory abnormalities of calcium, phosphorus, PTH or vitamin D metabolism; (2) Bone abnormalities in turnover, mineralization, volume, linear growth or strength; and (3) Calcification of the vasculature or other soft tissues [41]. The condition occurs early in the course of CKD, affects the vast majority of patients as kidney disease progresses and is a strong, independent predictor of bone fracture, CKD progression, cardiovascular disease, and death [42].

Secondary hyperparathyroidism is common in patients with CKD and is characterized by elevated serum FGF-23, PTH and phosphate levels, normal or low serum calcium levels, and low serum $1,25(OH)_2D_3$ concentrations, with increased bone resorption [43, 44]. FGF-23, a phosphaturic hormone produced by osteoblasts and osteocytes, precedes other laboratory biomarkers of CKD-MBD and rises exponentially as eGFR declines [44]. FGF-23 binds and forms a complex with its receptor, FGFR1, in the presence of

Klotho, a transmembrane protein, to promote decreased proximal tubule expression of sodium-phosphate cotransporters (leading to increased phosphate reabsorption and hyperphosphataemia), decreased kidney $1,25(OH)_2D_3$ synthesis and decreased PTH gene expression and parathyroid cell proliferation [43, 44]. Uncontrolled hyperparathyroidism can lead to osteodystrophy, vascular calcification and calciphylaxis (uremic calcific arteriolopathy), and is associated with increased morbidity and mortality [43, 44].

Treatment of secondary hyperparathyroidism in patients with CKD consists of efforts to lower PTH and control serum calcium and phosphate levels using dietary modification, medications, and dialysis. While using a variety of medications to correct serum PTH, phosphorus, and calcium levels is currently considered standard medical care, a recent systematic review and meta-analysis of randomized trials found that such drug effects on these biochemical markers were imprecisely correlated with cardiovascular and all-cause mortality and were consistent with the possibility of no association [45]. Thus, it remains uncertain as to whether any of the currently recommended interventions are of net clinical benefit [46].

Calcium-based phosphate binders have traditionally been used as first line treatment to reduce serum phosphate concentrations due to their low cost and availability. Unfortunately, their use has been associated with hypercalcaemia, vascular calcification, and calciphylaxis [47]. Aluminium-based binders have been associated with significant side effects, including aluminium-induced bone disease, microcytic anaemia, and dementia, although some investigators have argued that aluminium binders have only played a secondary role to aluminium–contaminated dialysis water and that no cases of aluminium binder-associated toxicity have been reported in the era of ultrapure dialysis water quality [48]. Nevertheless, some guidelines recommend that aluminium-based binders should be used only for short periods of time with close monitoring [49]. Non-calcium phosphate binders, including lanthanum carbonate, sevelamer hydrochloride, and sucroferric oxyhydride, are effective at lowering phosphate, although their use is often restricted by higher cost and difficult access in many healthcare systems [47]. A recently published network meta-analysis of randomized trials of phosphate binders in adults with CKD reported that there was no evidence that phosphate binder treatment reduced mortality compared with placebo [50]. However, patients treated with sevelamer had a lower mortality than those treated with calcium binders, although it remained uncertain whether this reflected a benefit of sevelamer, a harm of calcium or both [50].

Active vitamin D compounds (calcitriol and its prodrugs) increase gut absorption of calcium and phosphorus and reduce PTH synthesis by activating the parathyroid vitamin D receptor. These drugs can cause hypercalcaemia and hyperphosphataemia and so serum calcium and phosphate levels must be monitored. There is however no evidence that these agents alter patient-level outcomes [42, 51].

Calcimimetics, such as cinacalcet, modulate the calcium sensing receptor (CaSR) resulting in reduced PTH levels by increasing sensitivity to calcium [52]. A cumulative meta-analysis of randomized controlled trials demonstrated that cinacalcet reduced the need for parathyroidectomy by approximately 50%, but did not improve cardiovascular or all-cause mortality [52]. Cinacalcet is therefore especially useful for patients who are awaiting parathyroid surgery, or those who are not medically stable enough to receive a parathyroidectomy. Hypocalcaemia is a common side effect and should be monitored.

Prescription of bisphosphonates for the purpose of managing CKD-MBD and reducing bone resorption should be avoided, except as a bridge to definitive management of secondary hyperparathyroidism, as they may cause low-turnover bone disease and have also been reported to be nephrotoxic [53]. Moreover, the benefits and harms of bisphosphonates in patients with CKD have not been clearly established [53].

Parathyroidectomy, usually with preservation of remnant gland or autotransplantation of parathyroid tissue, is indicated in severe secondary hyperparathyroidism, typically with PTH levels >100pmol/L, or tertiary hyperparathyroidism when medical management has failed. It remains the most effective treatment for secondary or tertiary hyperparathyroidism and, in appropriate candidates, may reduce morbidity and mortality [54].

REFERENCES

1. Bello AK, Levin A, Tonelli M, *et al.* Assessment of global kidney health care status. *JAMA*, 2017; **317**: 1864–81.
2. Yang T, Xu C. Physiology and pathophysiology of the intrarenal renin-angiotensin system: an update. *JASN*, 2017; **28**: 1040–9.
3. Johnson D, Fawcett K, Harvie B, *et al. Chronic Kidney Disease Management in General Practice*, 3rd edn. Adelaide: Kidney Health Australia; 2015.
4. Briet M, Schiffrin EL. Aldosterone: effects on the kidney and cardiovascular system. *Nat Rev Nephrol*, 2010; **6**: 261–73.
5. Bolignano D, Palmer SC, Navaneethan SD, Strippoli GF. Aldosterone antagonists for preventing the progression of chronic kidney disease. *Cochrane Database Syst Rev*, 2014; **4**: CD007004.
6. Lacombe C, Mayeux P. The molecular biology of erythropoietin. *Nephrol Dial Transplant*, 1999; **14** Suppl 2: 22–8.
7. Johnson DW, Pollock CA, Macdougall IC. Erythropoiesis-stimulating agent hyporesponsiveness. *Nephrology (Carlton, Vic)*, 2007; **12**: 321–30.
8. Locatelli F, Del Vecchio L, Luise MC. Current and future chemical therapies for treating anaemia in chronic kidney disease. *Exp Opin Pharmacother*, 2017; **18**: 781–8.
9. Palmer SC, Saglimbene V, Mavridis D, *et al.* Erythropoiesis-stimulating agents for anaemia in adults with chronic kidney disease: a network meta-analysis. Cochrane Database Syst Rev, 2014; **12**: CD010590.
10. Palmer SC, Navaneethan SD, Craig JC, *et al.* Meta-analysis: erythropoiesis-stimulating agents in patients with chronic kidney disease. *Ann Intern Med*, 2010; **153**: 23–33.
11. Kidney Disease: Improving Global Outcomes (KDIGO) Anemia Work Group. KDIGO Clinical Practice Guideline for Anemia in Chronic Kidney Disease. *Kidney Int Suppl*, 2012; **2**: 279–335.
12. Palmer BF, Clegg DJ. Gonadal dysfunction in chronic kidney disease. *Rev Endocr Metab Disord*, 2017; **18**: 117–30.
13. Vecchio M, Palmer SC, Tonelli M, Johnson DW, Strippoli GF. Depression and sexual dysfunction in chronic kidney disease: a narrative review of the evidence in areas of significant unmet need. *Nephrol Dial Transplant*, 2012; **27**: 3420–8.
14. van Coevorden A, Stolear JC, Dhaene M, van Herweghem JL, Mockel J. Effect of chronic oral testosterone undecanoate

administration on the pituitary-testicular axes of hemodialyzed male patients. *Clin Nephrol*, 1986; **26**: 48–54.
15. Vecchio M, Navaneethan SD, Johnson DW, *et al.* Interventions for treating sexual dysfunction in patients with chronic kidney disease. *Cochrane Database Syst Rev*, 2010; **12**: CD007747.
16. Vecchio M, Navaneethan SD, Johnson DW, *et al.* Treatment options for sexual dysfunction in patients with chronic kidney disease: a systematic review of randomized controlled trials. *CJASN*, 2010; **5**: 985–95.
17. Lim VS, Henriquez C, Sievertsen G, Frohman LA. Ovarian function in chronic renal failure: evidence suggesting hypothalamic anovulation. *Ann Intern Med*, 1980; **93**: 21–7.
18. Berman JR, Berman LA, Toler SM, Gill J, Haughie S. Safety and efficacy of sildenafil citrate for the treatment of female sexual arousal disorder: a double-blind, placebo controlled study. *J Urol*, 2003; **170**(6 Pt 1): 2333–8.
19. Weisinger JR, Bellorin-Font E. Outcomes associated with hypogonadism in women with chronic kidney disease. *Adv Chron Kidney Dis*, 2004; **11**: 361–70.
20. Tangren J, Nadel M, Hladunewich MA. Pregnancy and end-stage renal disease. *Blood Purification*, 2018; **45**: 194–200.
21. Gomez F, de la Cueva R, Wauters JP, Lemarchand-Beraud T. Endocrine abnormalities in patients undergoing long-term hemodialysis. The role of prolactin. *Am J Med*, 1980; **68**: 522–30.
22. Blumberg A, Wildbolz A, Descoeudres C, *et al.* Influence of 1,25 dihydroxycholecalciferol on sexual dysfunction and related endocrine parameters in patiens on maintenance hemodialysis. *Clin Nephrol*, 1980; **13**: 208–14.
23. Mahesh S, Kaskel F. Growth hormone axis in chronic kidney disease. *Pediatr Nephrol (Berlin, Germany)*, 2008; **23**: 41–8.
24. Oh Y. The insulin-like growth factor system in chronic kidney disease: pathophysiology and therapeutic opportunities. *Kidney Res Clin Pract*, 2012; **31**: 26–37.
25. Haffner D, Schaefer F, Nissel R, Wuhl E, Tonshoff B, Mehls O. Effect of growth hormone treatment on the adult height of children with chronic renal failure. German Study Group for Growth Hormone Treatment in Chronic Renal Failure. *N Engl J Med*, 2000; **343**: 923–30.
26. Fine RN, Stablein D. Long-term use of recombinant human growth hormone in pediatric allograft recipients: a report of the NAPRTCS Transplant Registry. *Pediatr Nephrol (Berlin, Germany)*, 2005; **20**: 404–8.
27. Takala J, Ruokonen E, Webster NR, *et al.* Increased mortality associated with growth hormone treatment in critically ill adults. *N Engl J Med*, 1999; **341**: 785–92.
28. Feldt-Rasmussen B, El Nahas M. Potential role of growth factors with particular focus on growth hormone and insulin-like growth factor-1 in the management of chronic kidney disease. *Semin Nephrol*, 2009; **29**: 50–8.
29. Karamchandani D, Arias-Amaya R, Donaldson N, Gilbert J, Schulte KM. Thyroid cancer and renal transplantation: a meta-analysis. *Endocr Relat Cancer*, 2010; **17**: 159–67.
30. Iglesias P, Bajo MA, Selgas R, Diez JJ. Thyroid dysfunction and kidney disease: an update. *Rev Endocr Metab Disord*, 2017; **18**: 131–44.
31. Kaptein EM. Thyroid hormone metabolism and thyroid diseases in chronic renal failure. *Endocr Rev*, 1996; **17**: 45–63.
32. Sakurai S, Hara Y, Miura S, *et al.* Thyroid functions before and after maintenance hemodialysis in patients with chronic renal failure. *Endocrinologia Japonica*, 1988; **35**: 865–76.
33. Docter R, Krenning EP, de Jong M, Hennemann G. The sick euthyroid syndrome: changes in thyroid hormone serum parameters and hormone metabolism. *Clin Endocrinol*, 1993; **39**: 499–518.
34. McLean M, Kirkwood I, Epstein M, Jones B, Hall C. Cation-exchange resin and inhibition of intestinal absorption of thyroxine. *Lancet (London, England)*, 1993; **341**: 1286.
35. Liel Y, Sperber AD, Shany S. Nonspecific intestinal adsorption of levothyroxine by aluminum hydroxide. *Am J Med*, 1994; **97**: 363–5.
36. Wartofsky L, Burman KD. Alterations in thyroid function in patients with systemic illness: the 'euthyroid sick syndrome'. *Endocr Rev*, 1982; **3**: 164–217.
37. Gerich JE, Meyer C, Woerle HJ, Stumvoll M. Renal gluconeogenesis: its importance in human glucose homeostasis. *Diabetes Care*, 2001; **24**: 382–91.
38. Meyer C, Stumvoll M, Nadkarni V, Dostou J, Mitrakou A, Gerich J. Abnormal renal and hepatic glucose metabolism in type 2 diabetes mellitus. *J Clin Invest*, 1998; **102**: 619–24.
39. Koppe L, Pelletier CC, Alix PM, *et al.* Insulin resistance in chronic kidney disease: new lessons from experimental models. *Nephrol Dial Transplant*, 2014; **29**: 1666–74.
40. Bak JF, Schmitz O, Sorensen SS, Frokjaer J, Kjaer T, Pedersen O. Activity of insulin receptor kinase and glycogen synthase in skeletal muscle from patients with chronic renal failure. *Acta Endocrinologica*, 1989; **121**: 744–50.
41. Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Work Group. KDIGO Clinical Practice Guideline for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). *Kidney Int Suppl*, 2009; **76**: 1–130.
42. O'Shea S, Johnson DW. Review article: addressing risk factors in chronic kidney disease mineral and bone disorder: can we influence patient-level outcomes? *Nephrology (Carlton, Vic)*, 2009; **14**: 416–27.
43. Silver J, Naveh-Many T. FGF-23 and secondary hyperparathyroidism in chronic kidney disease. *Nat Rev Nephrol*, 2013; **9**: 641–9.
44. Lee G, Krishnasamy R, Hawley CM, Johnson DW. The impact of fibroblast growth factor-23 on the cardiovascular system in chronic kidney disease. *Exp Rev Endocrinol Metab*, 2015; **10**: 565–8.
45. Palmer SC, Teixeira-Pinto A, Saglimbene V, *et al.* Association of drug effects on serum parathyroid hormone, phosphorus, and calcium levels with mortality in CKD: a meta-analysis. *Am J Kidney Dis*, 2015; **66**: 962–71.
46. Johnson DW. Attempting to build solid recommendations on a shaky evidence base. *Semin Dial*, 2011; **24**: 22–4.
47. Chan S, Au K, Francis RS, Mudge DW, Johnson DW, Pillans PI. Phosphate binders in patients with chronic kidney disease. *Aust Prescr*, 2017; **40**: 10–4.
48. Mudge DW, Johnson DW, Hawley CM, *et al.* Do aluminium-based phosphate binders continue to have a role in contemporary nephrology practice? *BMC Nephrol*, 2011; **12**: 20.
49. National Kidney Foundation. K/DOQI clinical practice guidelines for bone metabolism and disease in chronic kidney disease. *Am J Kidney Dis*, 2003; **42**(4 Suppl 3): S1–201.
50. Palmer SC, Gardner S, Tonelli M, *et al.* Phosphate-binding agents in adults with CKD: a network meta-analysis of randomized trials. *Am J Kidney Dis*, 2016; **68**: 691–702.
51. Palmer SC, McGregor DO, Craig JC, Elder G, Macaskill P, Strippoli GF. Vitamin D compounds for people with chronic kidney disease requiring dialysis. *Cochrane Database Syst Rev*, 2009; **4**: CD005633.

52. Palmer SC, Nistor I, Craig JC, *et al.* Cinacalcet in patients with chronic kidney disease: a cumulative meta-analysis of randomized controlled trials. *PLoS Med*, 2013; **10**: e1001436.
53. Toussaint ND, Elder GJ, Kerr PG. Bisphosphonates in chronic kidney disease; balancing potential benefits and adverse effects on bone and soft tissue. *Clin J Am Soc Nephrol*, 2009; **4**: 221–33.
54. Tominaga Y, Matsuoka S, Uno N, Sato T. Parathyroidectomy for secondary hyperparathyroidism in the era of calcimimetics. *Ther Apher Dial*, 2008; **12** Suppl 1: S21–6.

12.1.4 The Endocrinology of Liver Disease

Jacob George and Mohammed Eslam

Introduction

The liver plays a fundamental role in metabolic regulation and energy homeostasis. Thus, it is also an inevitable target of endocrine diseases with intimate reciprocal feedback mechanisms. Liver and endocrine diseases are both common in the general population and hence it is to be expected that diseases of both frequently coexist and modulate the phenotypic manifestations or progression of the other. Of particular interest, the common liver diseases, non-alcoholic fatty liver disease, viral hepatitis and haemochromatosis have unique extrahepatic endocrine manifestations while the liver is frequently a target of a variety of endocrinopathies, including hypothyroidism, hypercortisolaemia, polycystic ovarian syndrome (PCOS), hypopituitarism, growth hormone deficiency, and hypogonadism. In the following chapter, we summarize this bidirectional relationship (Table 12.1.4.1).

Endocrine Associations of Common Hepatic Diseases

Metabolic associated fatty liver disease-Associated Endocrine Disorders

Metabolic associated fatty liver disease (MAFLD) afflicts more than 20–30% of the world population, paralleling the dramatic escalation in global obesity prevalence [1]. Thus, MAFLD is a leading cause for end-stage liver disease, cancer, and liver transplantation in Western countries [1]. The association between MAFLD and the endocrine system is bidirectional. MAFLD is seen in association with a variety of endocrinopathies, while conversely MAFLD increases the risk for multiple endocrine disorders.

Diabetes

The close association between MAFLD and diabetes is robust, so much so that which is the chicken and which is the egg cannot be easily discerned. A recent large prospective study including around 133 000 Taiwanese non-diabetic subjects who had two or more health examinations during 1996–2014 concluded that MAFLD diagnosed by ultrasonography was significantly and independent of other common risk factors such as age, sex, family history of diabetes (T2DM) or obesity, associated with a twofold increase in incident T2DM [2]. This result was confirmed in a recent meta-analysis of 20 observational studies comprising around 117 000 non-diabetic individuals diagnosed with MAFLD by either ultrasonography or abnormal serum liver enzymes, indicating an almost twofold increase in incident T2DM over a median follow-up of five years [3].

Consistently, data demonstrating a reduced incidence of T2DM after improvement or resolution of MAFLD provide strong evidence for a causal relationship between MAFLD and glucose abnormalities [4, 5]. The first study of large cohort of South Korean individuals (n = 13 000) demonstrated a significant reduction in the risk of incident T2DM in individuals in whom fatty liver resolved over time. The association was independent of other potential confounding factors. A second study included ~4000 non-diabetic Japanese subjects followed for 10 years. Improvement of MAFLD based on ultrasonography was again independently associated with a nearly 70% risk reduction of incident T2DM [5]. Notably, the severity of MAFLD was associated with an increased risk of T2DM. Finally, in a recent study of 396 non-diabetic patients with biopsy-proven MAFLD, subjects with advanced liver fibrosis stages 3–4 were at higher risk of T2DM, compared to those with fibrosis stages 0–2 (51% vs. 31%), over a mean follow-up of 18.4 years. Among subjects with fibrosis stages 0–2, fat score was an independent and principal predictor of T2DM risk. Notably, histological features of steatohepatitis were not associated with T2DM risk.

The converse association of T2DM with MAFLD has also been convincingly demonstrated. For example, the prevalence of MAFLD is high among patients with T2DM ranging from 45% to 75% in hospital-based and 30% to 70% in population-based series. Furthermore, patients with T2DM have higher risk of steatohepatitis and for the development of serious liver-related complications (cirrhosis, liver failure, and hepatocellular cancer (HCC)) [6].

Pathophysiological mechanisms linking MAFLD and T2DM is the shared abnormal metabolic milieu seen that is the substrate for both diseases. This includes overnutrition and certain dietary factors such as consumption of an energy dense Western-type diet low in fibre and high in processed foods that can lead to insulin resistance (IR) and low grade systemic chronic inflammation. Peripheral IR leads to ectopic intramuscular lipid accumulation and reduced muscle glucose uptake while hepatic IR promotes hepatic lipid deposition; both coordinately linked to the development of T2DM and MAFLD. Furthermore, a Western-type diet leads to disturbed intestinal permeability with leakage of microbial and other content into the portal circulation leading to the activation of multiple pro-inflammatory innate immune receptors such as toll-like receptor (TLR)2 and TLR4 and inflammasomes, that triggers the development of T2DM, and as

Table 12.1.4.1 Summary of the main interactions between liver and endocrine disorders

Endocrine disease	Hepatic manifestations	Liver disease	Endocrine manifestations
Thyroid dysfunction		**MAFLD**	
Hyperthyroidism	Elevation of liver enzymes Jaundice: uncommon Few case reports of fulminant hepatitis		Diabetes mellitus Hypothyroidism Polycystic ovarian syndrome Osteoporosis
Hypothyroidism	Hepatic steatosis Gallbladder stones Elevated AST Myxoedema ascites		
Antithyroid drugs	Elevation of liver enzymes Idiosyncratic reaction can lead to fulminant hepatitis		
Adrenal dysfunction		**Hepatitis C**	
Addison's disease (adrenal insufficiency)	Mildly elevated aminotransferases		Insulin resistance and diabetes mellitus Thyroid disorders including autoimmune thyroiditis, hypothyroidism, and thyroid cancer Osteoporosis Gonadal dysfunction
Cushing's syndrome (adrenal excess)	Features of metabolic syndrome including hepatic steatosis		
Sex hormones		**Hereditary hemochromatosis**	
Oestrogens	Cholestasis. Hepatic peliosis Hepatic adenomas Focal nodular hyperplasia, Haemangiomas. Budd–Chiari syndrome		Diabetes Hypopituitarism Hypogonadism Thyroid dysfunction Adrenal dysfunction Parathyroid defects Osteoporosis
Androgens	Elevated aminotransferase levels cholestasis Adenomas Rarely hepatocellular carcinoma		

well hepatic IR and lipid deposition. Alterations in gut microbiota and changes to the activation of cytokine/adipocytokine cascades are another potential player in the pathogenesis.

Based on the aforementioned details, ideally, any therapeutic target for MAFLD, should not only improve liver injury, but also ameliorate the associated systemic metabolic milieu to reduce the risk of MAFLD-related extrahepatic disorders, such as diabetes and cardiovascular disease. Further randomized controlled studies will be required to determine if we need to change current clinical practice guidelines and incorporate screening for MAFLD in T2DM patients and vice versa.

Thyroid Disease

Hypothyroidism, a condition characterized by the failure of the thyroid to produce enough thyroid hormones (THs), has been strongly associated with MAFLD. A recent retrospective, case-control study found that the prevalence of hypothyroidism was higher in biopsy-proven MAFLD compared to controls (21% vs. 9.5%; P <0.01). Notably, the prevalence was even higher in patients with steatohepatitis than those with simple steatosis (25% vs. 12.8%, P = 0.03) [7]. This finding was confirmed in a larger study of patients with hypothyroidism (n = 2000) that suggested that overt and subclinical hypothyroidism are both associated with an increased risk of MAFLD, independent of other potential confounding by known metabolic risk factors [8].

On the other hand, clinical hypothyroidism may be an independent risk factor for the development of MAFLD. A recent cross-sectional study of 425 subjects with biopsy-proven MAFLD suggested that the presence of low-normal thyroid function and subclinical hypothyroidism are independent predictors of steatohepatitis and advanced fibrosis [9]. Another prospective study from the Rotterdam cohort based on 9419 individuals with a median follow-up of around 10 years assessed for the presence of MAFLD by ultrasonography. This work concluded that increased levels of free T_4 were associated with a decreased risk of MAFLD, independent of other possible confounding factors. Subjects in the lowest free T_4 tertile had a 1.31-times greater risk of developing MAFLD, compared to those in the highest tertile. Consistently, those individuals with increased thyroid-stimulating hormone (TSH) levels were at a greater risk of developing MAFLD [8]. Hence, screening of patients with MAFLD with annual TSH and vice versa, screening patients with hypothyroidism for MAFLD should be considered.

Mechanistically, a recent study has suggested that hypothyroidism induces MAFLD through hepatic and extrahepatic mechanisms [10]. A mild reduction in THs leads to impaired insulin secretion, suppression of lipolysis, and increased shuttling of fatty acids to the liver, where they induce MAFLD. Surprisingly this study also suggested that a severe reduction in serum TH levels could protects against the development of MAFLD via a

constitutive suppression of lipolysis [10]. Further, mechanistic studies will be required.

Polycystic Ovarian Syndrome (PCOS)

Numerous studies have suggested that patients with PCOS are at increased risk of MAFLD. A recent retrospective longitudinal cohort study assessed NAFLD rates in 63 120 women with PCOS and 121 064 age- and BMI-matched patients utilizing a primary care database in the United Kingdom. This study concluded that patients with PCOS have a 2.23-fold increased risk of MAFLD [11]. Consistently, a recent meta-analysis that incorporated 7 studies from 6 different countries demonstrated that patients with PCOS had a 3.93-fold increased risk of MAFLD, independent of BMI [12]. Notably, most of the current evidence suggests that IR, rather than androgen excess mediates the association. The prevalence of MAFLD among women with PCOS has been shown to vary in different countries ranging from 32.9%, 71%, and 73.3% among Chinese, Australian, and Brazilian women with PCOS, respectively [11]. Though, it is less clear, the converse also seems to be true. Again, these data suggest that screening for the presence of MAFLD among women with PCOS should be considered.

Osteoporosis

Osteoporosis is a prevalent complication in patients with chronic liver disease, especially in end-stage disease, where a prevalence rate of up to 55% has been reported. Osteoporosis in liver disease has been particularly linked to chronic cholestasis, MAFLD, alcoholism, and haemochromatosis.

Several cross-sectional and case-control studies involving both male and female adults and adolescents have found an association between MAFLD and low bone mineral density or osteoporosis [13]. A recent retrospective study involving 7797 Chinese men aged 40 years or older demonstrated that the presence of MAFLD diagnosed by ultrasonography was associated with a 2.5-fold increased risk of osteoporotic fractures [14]. Alarmingly, this risk association has also been reported in the paediatric MAFLD population [15]. Notably, in that study, the degree of bone mineral density loss correlated with increasing severity of MAFLD, being higher in patients with steatohepatitis compared to those with simple steatosis. However, these data should be interpreted with caution, due to the limited sample size (n = 38).

The pathophysiological links between MAFLD and low bone mineral density are not completely understood. However, both disorders share common risk factors such as low levels of physical activity, obesity, IR, T2DM, vitamin D_3 deficiency, and chronic systemic inflammation. Many of these common risk factors are also present in patients with chronic cholestasis, alcoholism, and haemochromatosis.

Elucidation of the core molecular pathways and mechanisms between liver and bone disorders would help to identify novel therapeutic targets to ameliorate both conditions. Further prospective studies and on other ethnicities (as most current reports are from Asia) are required before any recommendation to change current guidelines on screening can be recommend. However, collectively, the findings argue for increased awareness of osteopenia/osteoporosis and the importance of monitoring and evaluation for this among patients with MAFLD.

Hepatitis C-Associated Endocrine Disorders

Hepatitis C (HCV) infects ~3% (170 million) of the world's population [16]. Of these, 70–80% develop chronic disease with a risk of developing liver complications including cirrhosis and cancer, with an estimated liver-related mortality of 350 000 people per year [17]. HCV can be considered not only as a hepatotropic virus, but also a systemic disease with diverse extrahepatic manifestations, including endocrine disorders—mainly IR and type 2 diabetes mellitus, but also thyroid disease and gonadal dysfunction [18].

IR and Type 2 Diabetes Mellitus

Multiple levels of evidences suggest a strong link between HCV and IR/T2DM [19]. T2DM is more prevalent in those with HCV-related cirrhosis than those with cirrhosis resulting from conditions other than chronic hepatitis C (CHC). Consistently, a large study included 9841 subjects from the third National Health and Nutrition Examination Survey (USA) showed that patients with HCV have a threefold higher risk of developing T2DM. Similarly, a longitudinal study demonstrated that patients with HCV were 11 times more likely to develop T2DM than those without HCV infection. Even, in the context of organ transplantation, T2DM is more frequent in liver and kidney transplantation among HCV-infected compared with HCV non-infected patients. Conversely, the prevalence of HCV infection in patients with T2DM ranges from 5% to 12% and is higher than in the general population [19].

CHC patients are also at increased risk of IR and impaired glucose tolerance, the key pathophysiological underpinning T2DM, first demonstrated by Hui and colleagues [20]. Insulin resistance is present in 30–70% of individuals with CHC and is interestingly associated with peripheral (muscle and adipose tissue), rather than hepatic IR [21]. HCV-related IR correlates with an increased risk of liver fibrosis, liver cirrhosis-related complications, such as portal hypertension and variceal bleeding, HCC, high serum HCV RNA levels, and an impaired response to interferon-based therapy before the current era of direct-acting antivirals (DAA) for HCV cure [22–24]. Notably, cure of HCV infection with either Peg-IFN/RBV or with DAAs is associated with an amelioration of IR [25].

Several molecular pathways have been implicated in mediating HCV-induced IR and T2DM. This includes enhanced intrahepatic and systemic expression of pro-inflammatory cytokines, changes in adipocytokines and generation of reactive oxygen species that possibly influence insulin signalling and increase both hepatic and peripheral (muscle and adipose tissue) IR. Furthermore, HCV-induced hepatic steatosis can promote an increase in lipid oxidation and endogenous glucose production, and hinder glucose disposal by skeletal muscles by decreasing glucose oxidation [19].

Thyroid Disorders

Thyroid disorders are among the most frequent endocrine extrahepatic manifestation of HCV; this prevalence is further increased in patients with HCV-induced cryoglobulinemia. Circulating thyroid autoantibodies such as antithyroglobulin antibody (AbTG), antithyroid peroxidase antibody (AbTPO), and antithyroid microsomal antibody (ATMA) are detected in 80–85% of cases and are higher than their prevalence in the general population. Autoimmune thyroid disease is the most frequent HCV-related

endocrine disorder, with hypothyroidism present in about 30% of HCV-infected patients. The mechanisms of HCV-related autoimmune thyroid disease is not fully understood, but is suggested to be a consequence of T-cell-mediated autoimmunity.

Of historical interest, before the era of DAA's, interferon therapy for HCV was associated with an increased risk of thyroid dysfunction; up to 25–30% of HCV patients developed autoimmune thyroid diseases (both hypo- and hyperthyroidism) during Peg-IFN/RBV treatment, and about half of these patients need thyroid hormone replacement therapy. Data are limited on whether DAAs have any effect on thyroid function.

A higher prevalence of papillary thyroid cancer has also been reported in CHC patients, especially in those with autoimmune thyroid disorders. A recent meta-analysis suggests that HCV increases the risk of thyroid cancer 2.86-fold after adjusting for the heterogeneity of the included studies. Longitudinal follow-up studies are however required to determine outcomes [26]. Assessment of TSH and thyroid autoantibodies, thyroid ultrasonography, and cytology are part of the diagnostic work-up of HCV-related thyroid disease.

Osteoporosis

Several studies have suggested that HCV-infected patients have an increased risk of osteoporosis compared to uninfected patients. A recent report of around 2500 HCV-positive patients from the Swiss cohort suggest that HCV was associated with a 1.43-fold increased risk of osteoporosis/fracture [27]. Notably, this association does not seem to be dependent on persistent HCV RNA and was not ameliorated by viral eradication. Consistently, a meta-analysis that included 362 285 participants suggested that HCV infection is associated with a pooled relative risk for osteoporosis of 1.53 [26]. This risk is exacerbated in HIV/HCV coinfected patients who exhibit a threefold increased risk of fracture incidence compared with uninfected controls, and a 1.2–2.4-fold increased fracture risk compared with HIV mono-infected individuals.

Sex Hormones

Though information is scanty, there are reports that suggest an association between CHC and sexual dysfunction; this was more prevalent in males compared to females. Consistent with this observation serum sex hormones were reported to be altered in HCV-infected patients and a high-prevalence of sexual dysfunction (28–74%) has been reported in multiple studies. The mechanisms of this association are unclear, but several hypotheses have been postulated. These include altered hypothalamic-pituitary function that might play a role in sex hormone regulation and, consequently, the impairment of sexual function. Psychological factors also might play a role; given the well know association of HCV infection with socioeconomic disadvantage [28]. In the past, the use Peg-IFN/RBV treatment was associated with an increased risk of sexual dysfunction, but this does not seem to be replicated in the era of interferon free DAA-based therapy.

On the other hand, there are several lines of evidence that indicate that women of reproductive age (premenopausal) have a more favourable HCV natural history, with slower fibrosis progression and a higher response rate to Peg-IFN/RBV-based antiviral therapy. This is consistent with human and experimental evidence indicating a protective role for oestrogens against the development of liver fibrosis and cancer [29] that is opposite to the effect of testosterone.

Hereditary Haemochromatosis

Haemochromatosis (HH) is a genetic disorder of iron overload that results, if untreated, in end-organ damage including of the liver and many endocrine organs. The latter includes diabetes, hypopituitarism, hypogonadism, thyroid dysfunction, adrenal dysfunction, parathyroid defects, and osteoporosis.

The reported prevalence of diabetes among patients with HH varies according to the method of HH diagnosis. It was originally reported to be present in 40–63%, but with the improvement of earlier diagnosis and treatment, the prevalence of T2DM is approximately 13–23%. The pathophysiology of diabetes associated with HH is not clear, but seems to be multifactorial. Iron overload can induce hepatic dysfunction leading to hepatic IR; excess iron can also lead to pancreatic β-cell oxidative stress and apoptosis, with insulin-deficiency. Notably, phlebotomy, which is the gold standard treatment for HH, has a variable impact on diabetes control. In early HH, patients who have not developed complications/organ damage, normalization of glucose tolerance leads to an improvement of insulin secretory capacity. In contrast, no significant improvement is noticed in HH patients with cirrhosis or diabetes [30].

Hypogonadism is the most frequent non-diabetic-endocrinopathy in HH patients with prevalence ranging from 10 to 100%. The vast majority of individuals present clinically with features of androgen deficiency, including loss of libido and impotence, reduced body hair, rarely testicular atrophy, and gynaecomastia with or without altered fat distribution. The main laboratory characteristics reported in male hypogonadic HH patients are subnormal testosterone levels with low or inappropriately normal basal gonadotropin concentrations and a blunted or no response of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) to luteinizing hormone-releasing hormone (LHRH) administration, indicating primary pituitary dysfunction. Similar studies in females are limited by sample size. When the diagnosis is confirmed, the recommended treatment is testosterone replacement in males and cyclical oestrogen and progesterone therapy in fertile females. Mechanistically, the original defect seems to be principally at pituitary level, with iron deposition leading to abnormal hormonal secretion. Iron deposition induced defects in hypothalamic or gonadal dysfunction with and without pituitary defects have also been reported.

Though panhypopituitarism, as well as a single tropin defects, has been reported in HH, it is relatively uncommon and a recovery of the pituitary–thyroid and adrenal axes following therapeutic phlebotomy has been reported. Similarly, it was suggested from autopsy studies that iron tends to deposit in the thyroid gland in HH patients. However, cases of thyroid dysfunction occur rarely and are usually observed at an end-stage of the disease where subjects already have cirrhosis and/or other endocrine disorders. Excess iron can also deposit in the adrenal glands, but data on adrenal dysfunction in HH patients is limited; they seem to have preserved glucocorticoid function [30].

Decreased bone mineral density and osteoporosis are relatively common in HH patients, with osteoporosis prevalence in approximately 25–34% and osteopenia in 40–79%. While parathyroid dysfunction has been suggested as a potential causative factor, it is more likely that the effect is multifactorial. Notably, improvements in bone mineral density have been reported with decreases in serum of ferritin [30].

Hepatic Manifestations of Endocrine Disorders

Thyroid Function

The association between liver and thyroid function is bidirectional. Several liver diseases have varying effects on thyroid hormone metabolism. Conversely, alterations in thyroid function can lead to liver dysfunction. The THs thyroxine and triiodothyronine regulate hepatocyte metabolic rate, development, and function. On the other hand, the liver is the main site for thyroid hormone metabolism and hence plays a key role in regulating their systemic effects. Consequently, thyroid or liver dysfunction may alter the functions of the other. Furthermore, some diseases such as infiltrative disorders and autoimmune affect both organs concomitantly [31].

Hyperthyroidism

Thyrotoxicosis is commonly associated with some degree of liver injury, which can be classified as hepatocellular or cholestatic. Clinically, it manifested by elevations in aminotransferase levels and occurs in about a third of patients. Similarly, elevations of alkaline phosphatase are reported in two third of cases while elevations in γ-glutamyl transpeptidase are observed in less than 20% of patients; progression of liver damage is infrequent. Jaundice is an uncommon presentation and if present, is usually secondary to the systemic effects of thyrotoxicosis, such as heart failure or septic shock. A few case reports have described fulminant liver failure as the initial manifestation of thyrotoxicosis. Histologically, mild liver injury with centrilobular cholestasis is observed and abnormalities return to normal with the treatment of thyroid disease. Mechanistically, the effects are likely due to hypoxia, oxidative stress, and increases in bile acid production.

Hypothyroidism

Hypothyroidism can cause hepatic steatosis, gallbladder stones, and elevated aspartate aminotransferase levels of muscular origin. It can also cause myxoedema ascites, a high protein concentration ascites; severe ascites is a very rare complication. Liver histology is usually normal, and only rarely has fibrosis been described; cholestasis has been reported. Nevertheless, liver manifestations usually regress to normal with thyroid hormone replacement therapy.

Antithyroid Drugs and Liver Abnormalities

Alterations of liver function have also been observed with antithyroid drugs. Hepatotoxicity and increased aspartate aminotransferase and alanine aminotransferase is observed more frequently with propylthiouracil, and is dose dependent. This idiosyncratic reaction that can develop at any time and the course is usually benign, despite cases of fulminant hepatic failure have been reported. In the United States, propylthiouracil is the second leading cause of non-acetaminophen–associated drug-induced liver injury requiring liver transplantation.

Adrenal Dysfunction and the Liver

Addison's disease (adrenal insufficiency) is associated with mildly elevated aminotransferases, which generally return to normal after appropriate glucocorticoid replacement therapy. The mechanisms for this elevation are unclear, but might be related to changes in body weight. Cushing's syndrome (adrenal excess) is linked with features of metabolic syndrome, including hepatic steatosis, MAFLD, IR, and hypertension. This can also be caused by up-regulation of the renin–angiotensin system.

Sex Hormones and the Liver

Oestrogens

The use of oral contraceptives is associated with hepatic disorders. Oestrogens increase the risk of developing cholestasis, hepatic peliosis, hepatic adenomas, focal nodular hyperplasia, and haemangiomas. Oral contraceptives can also be a risk factor for the development of Budd–Chiari syndrome by contributing to the underlying pro-thrombotic state and reasoning the risk of venous thrombosis (of the hepatic veins). Notably, the incidence of these complications has decreased with the use of currently available contraceptives that have lower doses of oestrogen and progesterone.

Androgens

The risk of liver dysfunction following the use of androgenic hormones is well described and includes elevated aminotransferase or cholestasis that can on occasions be severe. Hepatic adenomas and even hepatocellular carcinoma in association with androgen use has been described.

Conclusion

Liver and endocrine disorders are both common and intimately associated, often with shared pathogenic mechanisms (Figure 12.1.4.1). Hence, it is incumbent on both gastroenterologists and endocrinologists to consider these reciprocal interrelationships in

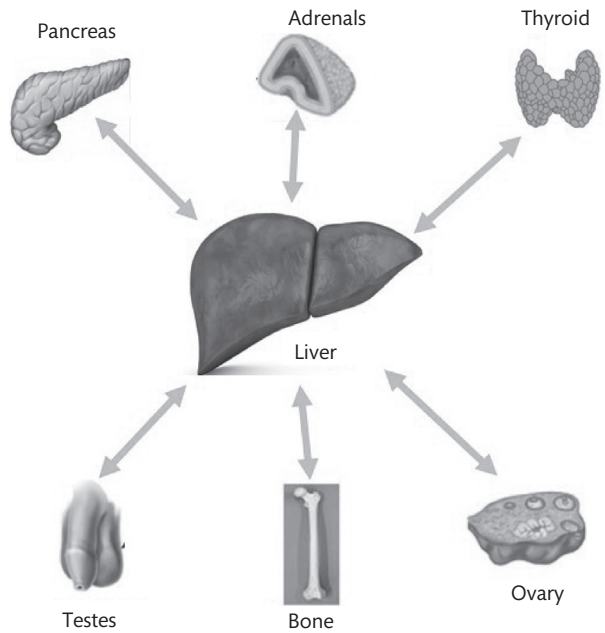


Figure 12.1.4.1 Intimate reciprocal associations and feedback mechanisms between the liver and endocrine organs. Common liver diseases such as non-alcoholic fatty liver disease, viral hepatitis, and haemochromatosis have extrahepatic endocrine manifestations including diabetes, thyroid, gonadal and adrenal dysfunction, and osteoporosis. Likewise, the liver is commonly involved in a variety of endocrinopathies.

daily medical practice, in order to provide better clinical care. It is particularly important to be aware of, and to deal with the hepatic adverse effects of some endocrine therapies, particularly the use of oral contraceptive formulations. Understanding the mechanistic basis of the interrelationships is an avenue for ongoing research that is pivotal for developing novel therapeutic targets, particularly for fatty liver disease. Importantly, prospective studies are required to reveal if we need to change current guidelines on screening for endocrine disease in patients with liver dysfunction and *vice versa*, at least in high risk groups.

ACKNOWLEDGEMENTS

ME and JG are supported by the Robert W. Storr Bequest to the Sydney Medical Foundation, University of Sydney; a National Health and Medical Research Council of Australia (NHMRC) Program Grant (APP1053206) and Project Grants (APP1107178 and APP1108422).

REFERENCES

1. Younossi Z, Anstee QM, Marietti M, *et al.* Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*, 2018; **15**: 11–20.
2. Chen SCC, Tsai SP, Jhao JY, *et al.* Liver fat, hepatic enzymes, alkaline phosphatase and the risk of incident type 2 diabetes: a prospective study of 132,377 adults. *Sci Rep*, 2017; **7**: 4649.
3. Ballestri S, Zona S, Targher G, *et al.* Nonalcoholic fatty liver disease is associated with an almost twofold increased risk of incident type 2 diabetes and metabolic syndrome. Evidence from a systematic review and meta-analysis. *J Gastroenterol Hepatol*, 2016; **31**: 936–44.
4. Yamazaki H, Tsuboya T, Tsuji K, *et al.* Independent association between improvement of nonalcoholic fatty liver disease and reduced incidence of type 2 diabetes. *Diabetes Care*, 2015; **38**: 1673–9.
5. Sung KC, Wild SH, Byrne CD. Resolution of fatty liver and risk of incident diabetes. *J Clin Endocrinol Metab*, 2013; **98**: 3637–43.
6. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat Rev Gastroenterol Hepatol*, 2013; **10**: 330–44.
7. Pagadala MR, Zein CO, Dasarathy S, *et al.* Prevalence of hypothyroidism in nonalcoholic fatty liver disease. *Dig Dis Sci*, 2012; **57**: 528–34.
8. Bano A, Chaker L, Plompen EPC, *et al.* Thyroid function and the risk of nonalcoholic fatty liver disease: the Rotterdam study. *J Clin Endocrinol Metab*, 2016; **101**: 3204–11.
9. Kim D, Kim W, Joo SK, *et al.* Subclinical hypothyroidism and low-normal thyroid function are associated with nonalcoholic steatohepatitis and fibrosis. *Clin Gastroenterol Hepatol*, 2018; **16**: 123–31.
10. Ferrandino G, Kaspari RR, Spadaro O, *et al.* Pathogenesis of hypothyroidism-induced NAFLD is driven by intra- and extrahepatic mechanisms. *Proc Natl Acad Sci U S A*, 2017; **114**: E9172–80.
11. Kumarendran B, O'Reilly MW, Manolopoulos KN, *et al.* Polycystic ovary syndrome, androgen excess, and the risk of nonalcoholic fatty liver disease in women: a longitudinal study based on a United Kingdom primary care database. *Plos Med*, 2018; **15**: e1002542.
12. Ramezani-Binabaj M, Motalebi M, Karimi-Sari H, *et al.* Are women with polycystic ovarian syndrome at a high risk of non-alcoholic fatty liver disease; a meta-analysis. *Hepat Mon*, 2014; **14**: e23235.
13. Targher G, Lonardo A, Rossini M. Nonalcoholic fatty liver disease and decreased bone mineral density: is there a link? *J Endocrinol Invest*, 2015; **38**: 817–25.
14. Li M, Xu Y, Xu M, *et al.* Association between nonalcoholic fatty liver disease (NAFLD) and osteoporotic fracture in middle-aged and elderly Chinese. *J Clin Endocrinol Metab*, 2012; **97**: 2033–8.
15. Pardee PE, Dunn W, Schwimmer JB. Non-alcoholic fatty liver disease is associated with low bone mineral density in obese children. *Aliment Pharmacol Ther*, 2012; **35**: 248–54.
16. World Health Organization (WHO). Hepatitis C. Fact Sheet. Available at: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c (accessed: 10 March 2020).
17. Hoofnagle JH. Course and outcome of hepatitis C. *Hepatology*, 2002; **36**: S21–9.
18. Khattab MA, Eslam M, Alavian SM. Hepatitis C virus as a multifaceted disease: a simple and updated approach for extrahepatic manifestations of hepatitis C virus infection. *Hepatitis Monthly*, 2010; **10**: 258–69.
19. Eslam M, Khattab MA, Harrison SA. Insulin resistance and hepatitis C: an evolving story. *Gut*, 2011; **60**: 1139–51.
20. Hui JM, Sud A, Farrell GC, *et al.* Insulin resistance is associated with chronic hepatitis C and virus infection fibrosis progression. *Gastroenterology*, 2003; **125**: 1695–704.
21. Milner KL, van der Poorten D, Trenell M, *et al.* Chronic hepatitis C is associated with peripheral rather than hepatic insulin resistance. *Gastroenterology*, 2010; **138**: 932–U170.
22. Eslam M, Ampuero J, Jover M, *et al.* Predicting portal hypertension and variceal bleeding using non-invasive measurements of metabolic variables. *Ann Hepatol*, 2013; **12**: 588–98.
23. Eslam M, Aparcero R, Kawaguchi T, *et al.* Meta-analysis: insulin resistance and sustained virological response in hepatitis C. *Aliment Pharmacol Ther*, 2011; **34**: 297–305.
24. Khattab MA, Eslam M, Mousa YI, *et al.* Association between metabolic abnormalities and hepatitis C-related hepatocellular carcinoma. *Ann Hepatol*, 2012; **11**: 487–94.
25. Serfaty L, Forns X, Goeser T, *et al.* Insulin resistance and response to telaprevir plus peginterferon alpha and ribavirin in treatment-naive patients infected with HCV genotype 1. *Gut*, 2012; **61**: 1473–80.
26. Wang P, Jing ZH, Liu CJ, *et al.* Hepatitis C virus infection and risk of thyroid cancer: a systematic review and meta-analysis. *Arab J Gastroenterol*, 2017; **18**: 1–5.
27. Kovari H, Rauch A, Kouyos R, *et al.* Hepatitis C infection and the risk of non-liver-related morbidity and mortality in HIV-infected persons in the Swiss HIV cohort study. *Clin Infect Dis*, 2017; **64**: 490–7.
28. Fusco F, D'Anzeo G, Rossi A, *et al.* Erectile dysfunction in patients with chronic viral hepatitis: a systematic review of the literature. *Expert Opin Pharmacother*, 2013; **14**: 2533–44.
29. Bernabucci V, Villa E. The role played by gender in viral hepatitis. *Scand J Clin Lab Invest*, 2014; **74**: 90–4.
30. Pelusi C, Gasparini DI, Bianchi N, *et al.* Endocrine dysfunction in hereditary hemochromatosis. *J Endocrinol Invest*, 2016; **39**: 837–47.
31. Burra P. Liver abnormalities and endocrine diseases. *Best Pract Res Clin Gastroenterol*, 2013; **27**: 553–63.

12.1.5 Endocrine Abnormalities in HIV Infection

Steven K. Grinspoon and Takara L. Stanley

Introduction

Approximately 37 million people worldwide are living with human immunodeficiency virus (HIV) infection [1]. As access to antiretroviral therapy has improved, the endocrine manifestations of HIV have evolved (Table 12.1.5.1). Whereas untreated individuals may develop undernutrition, wasting, and end-organ effects of opportunistic infections (OIs) such as primary adrenal insufficiency secondary to adrenal destruction, individuals treated with combination antiretroviral therapy (cART) may develop generalized or abdominal obesity, peripheral fat loss, insulin resistance, hyperlipidaemia, and non-alcoholic fatty liver disease (NAFLD). Individuals with treated HIV also have increased risk of cardiovascular disease (CVD) compared to the general population. Some individuals with HIV infection may experience gonadal dysfunction and bone mineral loss. The purpose of this chapter is to review the endocrine manifestations of HIV infection, including pathogenesis and treatment.

Anthropometric Effects of HIV Infection

Depending on the severity of infection and the use of antiretroviral therapy, HIV may have significant effects on body weight, fat distribution, and, in children, growth. These changes often accompany and may contribute to the HIV-associated metabolic and endocrine abnormalities described next.

Wasting and Undernutrition

HIV was originally termed 'slim disease' due to the cachexia that accompanies untreated infection. Before the advent of cART, 'AIDS wasting syndrome', now called HIV-related wasting, was common. Wasting is an AIDS-defining condition, characterized by involuntary weight loss of more than 10% of usual body weight, accompanied by diarrhoea or weakness and fever lasting 30 days in the absence of other illness. The aetiology of wasting is multifactorial and includes immune dysregulation, inflammatory response to HIV and OIs, increased energy expenditure, and protein catabolism [2]. In practice, wasting is less common today but may still be found in individuals with untreated HIV. Additionally, because HIV is prevalent in many areas where malnutrition is also common, HIV infection and malnutrition often coexist in a phenotype similar to wasting but with distinct aetiology [2]. Unsurprisingly, malnutrition predicts poorer outcomes even in treated HIV infection [2].

Changes in Fat Distribution Associated with Combination Antiretroviral Therapy (cART)

With the advent of cART, individuals with HIV infection began to demonstrate changes in body fat distribution, including increased abdominal fat accumulation, dorsocervical fat accumulation ('buffalo hump'), lipoatrophy of the face and limbs, and,

Table 12.1.5.1 Major endocrine abnormalities in HIV infection

	Untreated HIV and/or advanced disease	cART-treated HIV
Body composition	Wasting and/or malnutrition	Generalized obesity and/or visceral fat accumulation, dorsocervical fat accumulation, facial and peripheral lipoatrophy
Adrenal function	Possible primary or secondary adrenal insufficiency	Typically normal
GH/IGF-1 axis	GH resistance	Relative GH deficiency
Gonadal function	Hypogonadism and menstrual irregularities, greater prevalence in untreated or advanced disease. Increased SHBG	
Thyroid function	Subclinical hypothyroidism (non-auto-immune); increased TBG	
Bone density and calcium homeostasis	Reduced BMD and increased fracture risk. High prevalence of vitamin D deficiency. Risk of osteomalacia with TDF, efavirenz, or severe vitamin D deficiency. Increased risk of osteonecrosis	
Glucose metabolism	May be altered due to pancreatic process	Increased risk of glucose intolerance and diabetes
Lipid metabolism	Increased triglyceride, decreased LDL, increased de novo lipogenesis	Same as untreated, plus increased LDL
Cardiovascular disease	Unknown	Increased
Other	Increased prevalence of non-alcoholic fatty liver disease	

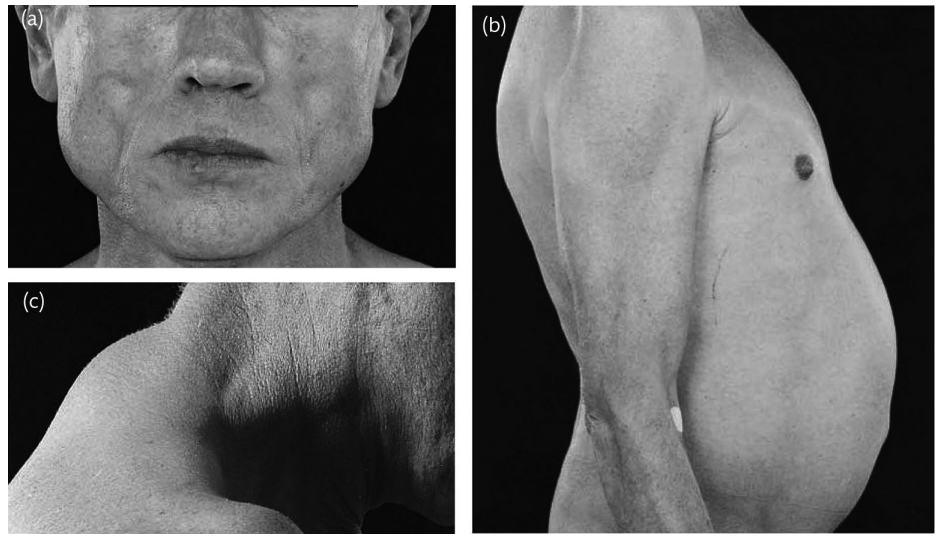


Figure 12.1.5.1 Facial fat atrophy (a), abdominal and breast lipohypertrophy with gluteal fat atrophy (b), and dorsocervical fat accumulation (c) in individuals with HIV infection.

Modified with permission from Grinspoon S, Carr A. Cardiovascular risk and body-fat abnormalities in HIV-infected adults. *N Engl J Med.* 2005;352(1):48–62. (38), Copyright © 2005 Massachusetts Medical Society. All rights reserved.

less commonly, multiple discrete lipomas. These changes, often termed 'HIV lipodystrophy', do not represent a single syndrome but rather are characterized by differing degrees of lipoatrophy and lipohypertrophy in individual patients. For example, patients may experience peripheral fat atrophy, abdominal fat accumulation with visceral fat gain and central subcutaneous fat loss, or a combination of both (Figure 12.1.5.1). Additionally, studies of cART initiation consistently show increases in weight and fat mass with treatment. Although weight gain with cART is associated with lower mortality for those who are initially of normal weight or less [3], generalized obesity is increasingly common in cART-treated patients [4].

The aetiology of altered fat accumulation in HIV is multifactorial. Older agents in the protease inhibitor (PI) and nucleoside reverse transcriptase inhibitors (NRTI) classes exert direct effects on adipocyte differentiation, glucose metabolism, and mitochondrial function. With newer antiretroviral agents, extreme phenotypes of lipodystrophy are less common, but generalized obesity is of increasing concern [4]. Randomized studies generally have not demonstrated substantial benefits of one antiretroviral regimen over another with respect to preventing ongoing fat gain, although weight gain appears to be more significant with integrase inhibitors [4].

Ectopic fat accumulation in other areas, including the liver, muscle, and epicardium, is also common in HIV and may contribute to adverse metabolic profile. NAFLD appears to be common in HIV and may be related in part to use of antiretrovirals with mitochondrial toxicity [5, 6].

Chronic inflammation and immune activation in HIV may contribute to abnormal fat deposition and altered adipocyte function. Additionally, HIV viral proteins alter adipocyte function and may contribute to lipohypertrophy [4]. The increased dorsocervical fat and visceral fat accumulation seen among many cART-treated patients are similar to the phenotype of Cushing syndrome, but neither facial atrophy nor laboratory evaluation of these patients is consistent with true Cushing syndrome.

As described next, visceral fat accumulation is associated with reduced GH levels, dyslipidaemia, insulin resistance, and impaired glucose metabolism. In addition, patients with HIV and associated changes in fat distribution demonstrate elevated C-reactive protein (CRP) and decreased adiponectin compared to patients without such changes [7], as well as increased intramyocellular lipid and hepatic steatosis, all of which may also contribute to increased cardiometabolic risk. The medical and surgical interventions used for obesity in the general population may also be appropriate and effective in HIV but require further study [4].

Effects of HIV on Growth and Body Composition in Children

Children and adolescents with HIV often have decreased height-for-age and weight-for-age, and growth failure in children with HIV is associated with disease progression and mortality [8, 9]. Combined antiretroviral therapy (ART) improves but may not normalize growth in children with HIV, particularly in the context of coexisting malnutrition [9]. Abdominal fat accumulation and/or peripheral fat atrophy are also described in children receiving antiretroviral therapy. As in adults, the severity of these changes depends on the antiretroviral agents used as well as the appropriateness of paediatric dosing.

Growth Hormone/IGF-1 Axis

Physiology of the GH/IGF-1 Axis in HIV Infection

GH secretion is commonly altered in individuals with HIV, often related to changes in body composition. Patients with HIV-associated weight loss often demonstrate elevated GH levels and reduced IGF-1, consistent with GH resistance. Endogenous overnight GH secretion in this population is inversely associated with fat mass, and IGF-1 levels are directly associated with caloric intake. In contrast, patients with HIV and abdominal fat accumulation may have

decreased GH production. Men with HIV and abdominal fat accumulation demonstrate reductions in both mean overnight endogenous GH secretion and peak GH levels following GH-releasing hormone (GHRH)/arginine stimulation testing as compared to both healthy volunteers and men with HIV and normal body composition. As a result, a large percentage of men with HIV infection and visceral fat accumulation demonstrate relative GH deficiency. Although fewer studies describe GH secretion in women with HIV, increased abdominal adiposity is a negative predictor of peak stimulated GH in women as well. The aetiology of relative GH deficiency in individuals with HIV infection and abdominal fat accumulation is multifactorial and includes increased somatostatin tone, increased free fatty acids, and decreased ghrelin.

Growth Hormone Therapy for AIDS Wasting

In patients with AIDS, lean body mass may decline significantly and disproportionately to weight. Moreover, lean body mass correlates with survival. High-dose GH has been used as an anabolic therapy to increase lean mass in patients with wasting. Numerous studies of recombinant human GH (rhGH) at supraphysiologic doses of 0.1 mg/kg, or at a fixed dose of 6 mg, demonstrate increases in lean mass along with corresponding improvement in muscle function and/or exercise capacity. The relatively high doses of GH in these studies are thought to be necessary to overcome the GH resistance found in individuals with wasting. However, these supraphysiologic doses may also result in numerous side effects, including hyperglycaemia, arthralgia, and fluid retention, limiting long-term use of GH in this population.

Augmenting GH for Abdominal Fat Accumulation in HIV Infection

Though not currently approved by regulatory agencies in Europe or the United States for this indication, GH has been shown to reduce visceral fat in patients with HIV-associated abdominal fat accumulation. However, use of rhGH exacerbates insulin resistance, limiting its use in HIV-associated visceral adiposity.

An alternative strategy to increase GH levels in individuals with HIV-associated visceral adiposity is through use of GHRH analogue. In theory, GHRH may have two advantages over the use of rhGH: it maintains GH pulsatility, thus more closely mimicking physiologic secretion, and it preserves the negative feedback of IGF-1 on pituitary GH release. A large study of synthetic GHRH (tesamorelin) daily for six months demonstrated a 15% reduction in visceral fat, decreased triglyceride and total cholesterol, and increased high density lipoprotein (HDL), without changes in fasting or 2-hour glucose or insulin measured by oral glucose tolerance test [4]. A 6-month extension demonstrated that beneficial effects on visceral fat and lipids were maintained over 12 total months of treatment, without effect on insulin sensitivity, but that patients who discontinued treatment had rapid reaccumulation of visceral adipose tissue. In some countries, tesamorelin has received regulatory approved for use to reduce visceral adiposity in HIV-infected adults [4]. Although supraphysiologic levels of IGF-1 are uncommon with tesamorelin use, monitoring of IGF-1 is useful before and 6–12 months after initiation of therapy to confirm physiologic dosing [4].

GH/IGF-I Axis in HIV-Infected Children

Untreated children with HIV tend to exhibit GH resistance, sometimes accompanied by reductions in IGF-1 [10]. With cART, GH resistance improves, and both IGF-1 and IGFBP-3 generally increase; improvements in IGF-1 are associated with increases in lean body mass [10]. As in adults, adolescents with HIV-associated visceral adiposity demonstrate decreased GH response to GHRH-arginine testing [11]. Therapies targeting the GH axis have not been studied widely in children.

Gonadal Function

Pubertal Progression in Children with HIV

In conjunction with poor growth, children with HIV may have delay in pubertal progression [8]. Immune activation and chronic inflammation likely contribute to pubertal delay, whereas inadequate or delayed cART, OIs, and concomitant malnutrition are exacerbating factors [8]. Later age at cART initiation, lower CD4+ T-cell count, and higher viral load are associated with greater delays [8]. Pubertal delay may contribute to the reductions in bone mineral density seen in children and adults with HIV.

Gonadal Function in Men

Hypogonadism is common in men with untreated HIV, particularly in the advanced stages of disease and among patients with HIV-associated weight loss. In an early study, 6% of asymptomatic HIV-infected men demonstrated hypogonadism compared to 50% of men with CDC-defined AIDS [12]. The prevalence of hypogonadism in men with HIV who are receiving effective cART is much lower than in untreated men but still higher than the general population [13]. On average, men with HIV have lower morning free testosterone levels, and the slope of decrease in free testosterone with age is similar among men with and without HIV [13]. Of note, sex hormone binding globulin (SHBG) is elevated by 40–50% in individuals with HIV. Screening for hypogonadism in HIV should occur in the morning, using either a reliable free testosterone assay or calculation of bioavailable testosterone through assessment of total testosterone and SHBG. Hypogonadism should be confirmed with a second assessment.

The mechanisms underlying hypogonadism in men with HIV are shown in Box 12.1.5.1. Hypogonadism is most commonly pituitary

Box 12.1.5.1 Mechanisms of gonadal dysfunction in HIV

Primary hypogonadism
Infiltrative disorders of the testes
Cytomegalovirus
Toxoplasmosis
Kaposi's sarcoma
Germ cell neoplasm
Lymphoma
Idiopathic

Secondary hypogonadism
Severe illness
Malnutrition
Infiltrative disorders of the pituitary/hypothalamus
Cytomegalovirus
Toxoplasmosis
Lymphoma
Adenohypophyseal necrosis
Medications (megestrol acetate, glucocorticoids)

or hypothalamic in origin, related to an effect of severe acute illness or undernutrition on gonadotropin production. Less commonly, men with HIV may develop primary hypogonadism due to anatomic destruction of testicular tissue by opportunistic infection. Hypothalamic and/or pituitary destruction from opportunistic infection (e.g. cytomegalovirus (CMV)) severe enough to cause panhypopituitarism and gonadal failure have also been reported in a small number of patients. Medications may also suppress the pituitary gonadal axis. Megestrol acetate suppresses gonadotropin secretion because of its glucocorticoid-like properties, and ketoconazole may cause primary hypogonadism by inhibiting enzymes in testicular steroidogenesis.

Androgen Therapy in Men with HIV Infection

Hypogonadism in men with HIV is associated with decreased lean body mass and diminished exercise capacity as well as increased indices of depression. Physiologic testosterone replacement increases lean mass and improves patient report of overall quality of life, appearance, and well-being. In men with HIV-associated abdominal fat accumulation, testosterone levels may also be decreased, but testosterone therapy in this population has no effect on visceral adiposity in spite of reductions in total and subcutaneous abdominal fat mass [14].

In addition to monitoring testosterone levels, laboratory monitoring during testosterone administration should include prostate specific androgen (PSA) measurement, HDL, which may decrease during therapy, and haematocrit, which may increase. Among patients who achieve stable weight and/or experience improved virologic control, discontinuation of testosterone, and reassessment of gonadal function by morning measurement of bioavailable testosterone level may be appropriate, as androgen concentrations may improve with nutritional and immunologic recovery.

Gonadal Function in Women with HIV Infection

Amenorrhea is common among women with advanced HIV disease, with the prevalence approaching 40% in women with severe HIV-related wasting. In women without an AIDS-defining illness, oligomenorrhea may be up to ten times more prevalent and amenorrhea up to seven times more prevalent than in HIV-negative controls [15]. Compared to the HIV-negative population, women with HIV have lower oestrogen levels and higher SHBG levels [16] and may have earlier menopause.

In addition, androgen deficiency is highly prevalent among women with HIV, particularly among those with wasting. In a cohort of women studied prior to the cART era, over 50% of women with wasting and more than one-third of normal weight patients demonstrated serum free testosterone concentrations below the lower limit of normal for healthy age-matched women. A more recent cohort also demonstrates lower testosterone and dehydroepiandrosterone sulphate (DHEA-S) concentrations in women with HIV compared to controls. Importantly, androgen deficiency in women with HIV is associated with decreased bone mineral density and may also be associated with increased cardiovascular risk [16]. The aetiology of the androgen deficiency is unknown but may relate to intra-adrenal shunting away from androgen production toward cortisol synthesis; in women with HIV-related wasting, DHEA-S levels are low compared to controls and correlate with decreased androgen levels.

Multiple studies have investigated physiologic transdermal testosterone replacement (150–300 mcg daily) in women with HIV, with some studies showing improvements in lean body mass, bone mineral density at the hip, sexual function, and depression indices. Transdermal testosterone has not received regulatory approval for the treatment of androgen deficiency in women with HIV. DHEA supplementation has also been investigated as a means of increasing testosterone and dihydrotestosterone as well as improving mood, but the safety profile and effects of DHEA on body composition are not known [17].

Adrenal Function

Adrenal function is typically normal in individuals with effectively treated HIV, but the adrenal axis may be affected in advanced HIV disease. In these cases, subclinical impairment of adrenal function is more common than frank adrenal insufficiency. Adrenal impairment is most often associated with anatomic destruction of the adrenal glands or anterior pituitary due to OIs, or to use of medications that suppress adrenal function. In contrast, increased cortisol concentrations may also be seen in HIV-infected patients, typically due to stress activation of the hypothalamic–pituitary–adrenal axis or intra-adrenal shunting toward cortisol synthesis.

Adrenal Insufficiency

Although adrenal insufficiency is rare in individuals with virologic control from cART, adrenal impairment is relatively common in patients with progressive HIV infection or AIDS. Impaired adrenal reserve may precede clinical adrenal insufficiency, as asymptomatic men with HIV demonstrate progressively increased plasma corticotropin (ACTH) concentrations over time in spite of normal cortisol responses to synthetic ACTH [18]. Individuals with advanced HIV disease are at higher risk for frank adrenal insufficiency. In a relatively large study of 93 patients with AIDS or AIDS-related complex (ARC), 4% of patients exhibited clinical adrenal insufficiency, whereas 54% had subnormal cortisol responses to synthetic ACTH indicating marginal adrenal reserve [19]. Adrenal insufficiency in individuals with HIV is most often related to tissue destruction of the adrenal glands by OIs (Box 12.1.5.2). Cytomegalovirus (CMV), mycobacterium avium intracellulare (MAI), and cryptococcus may damage adrenal tissue. Glandular destruction does not typically exceed more than 50% of adrenal tissue, however, and is therefore unlikely to result in frank adrenal insufficiency in most cases. Kaposi sarcoma, lymphoma, and haemorrhage may also cause adrenal damage in advanced HIV. Secondary adrenal insufficiency related to pituitary infiltration from disseminated toxoplasma gondii, cryptococcus, and CMV has also been reported in patients with AIDS.

Assessment and Treatment of Adrenal Insufficiency

Routine assessment of adrenal function is not indicated in asymptomatic individuals with HIV but should be performed in patients with advanced disease who have significant fatigue, inanition, hypotension, or hyponatraemia. Patients with known disseminated CMV or MAI are at increased risk. Testing should proceed with either a morning cortisol concentration or with cosyntropin administration (0.25 mg of ACTH 1–24). The cosyntropin test is misleading,

Box 12.1.5.2 Mechanisms of adrenal dysfunction in HIV

Primary adrenal insufficiency
Cytomegalovirus
Mycobacterium tuberculosis
Mycobacterium avium intracellulare
Cryptococcus neoformans
Kaposi's sarcoma
Haemorrhage
Lymphoma

Secondary adrenal insufficiency
Adenohypophyseal necrosis
Cytomegalovirus
Toxoplasmosis gondii
Lymphoma
Medications (megestrol acetate, ketoconazole, rifampin, opiates, ritonavir, glucocorticoids)

Hypercortisolism
Stress response to illness
Intra-adrenal shunting toward cortisol synthesis
Cytokine modulation
Glucocorticoid resistance
Medications (concurrent use of ritonavir and fluticasone)

however, if secondary adrenal insufficiency is of relatively acute onset, in which case testing with metyrapone or insulin tolerance testing is indicated. Of note, the insulin tolerance test should not be performed in certain circumstances, for example if patients are older, or have known heart disease or seizures. Long-term therapy includes mineralocorticoid and glucocorticoid replacement in primary adrenal insufficiency and glucocorticoid administration alone in secondary adrenal insufficiency.

Medication Effects

Medications such as megestrol acetate, ketoconazole, rifampin, and opiates are known to affect adrenal function. Ketoconazole inhibits multiple cytochrome-P450 dependent steroidogenic enzymes, including the side chain cleavage and 11-hydroxylase enzymes, and decreases cortisol synthesis in a dose dependent manner. Fluconazole and itraconazole are much less likely to cause adrenal insufficiency. Rifampin, an antituberculous agent, increases the metabolism of cortisol and may precipitate adrenal insufficiency in the setting of known hypoadrenalism or decreased adrenal reserve. In addition, megestrol acetate, a synthetic progestational agent approved for use as an appetite stimulant in patients with HIV-related wasting, decreases adrenal function through a glucocorticoid-like effect on the hypothalamic–pituitary–adrenal (HPA) axis. Chronic use of megestrol acetate may result in Cushingoid symptoms, and rapid withdrawal of megestrol acetate may cause adrenal insufficiency. Furthermore, megestrol acetate use can aggravate underlying glucose intolerance and diabetes mellitus in patients with HIV.

Hypercortisolism, HPA Activation, and Glucocorticoid Resistance

Individuals with HIV may demonstrate increased cortisol levels. Although this is most often due to a stress response to illness, studies have also demonstrated intra-adrenal shunting toward cortisol synthesis, potentially due to 17,20 lyase dysfunction. Cytokine modulation of the HPA axis is another potential mechanism for hypercortisolism in patients with HIV. Anomalous cases of clinical adrenal insufficiency with normal or elevated cortisol concentrations have also been reported in individuals with HIV in association with glucocorticoid resistance due to abnormal glucocorticoid receptors [20]. Finally, Cushing's syndrome has been described in individuals with HIV concurrently taking fluticasone and ritonavir, secondary to ritonavir's inhibition of CYP3A4, which prevents metabolism of fluticasone. Discontinuation of fluticasone in this setting can lead to severe adrenal insufficiency. Interactions between ritonavir and other steroids have also been reported.

Thyroid Function

Abnormal thyroid function may accompany HIV, although subtle laboratory abnormalities are more common that overt hyper- or hypothyroidism. Thyroid binding globulin (TBG) levels increase in both adults and children with HIV, correlating positively with severity of illness. Several studies indicate that subclinical hypothyroidism, indicated by elevated thyroid-stimulating hormone (TSH) with normal T_3 and T_4 levels, is also relatively common in individuals with HIV, with prevalence ranging from 3% to 12% [21]. Antithyroid peroxidase antibodies are commonly negative in these patients [21], arguing against an autoimmune aetiology. In addition to subclinical hypothyroidism, the prevalence of isolated low free thyroxine levels (without increased TSH) may also be increased in HIV infection [21]. Paediatric data demonstrate a high prevalence (approximately 20%) of isolated low free T_4 in children with HIV, with free T_4 levels directly associated with CD4+ count [22]. Development of hyperthyroidism due to Graves' disease has also been reported in individuals with HIV several months following initiation of cART, potentially as a late manifestation of immune reconstitution [21].

In patients with advanced HIV disease, overt hypothyroidism may result from infiltration of the pituitary or thyroid by OIs. *Pneumocystis jirovecii* may cause a pneumocystis thyroiditis characterized by painful gland enlargement; depending on the extent of necrosis, thyroid function may remain normal, or patients may develop hypothyroidism, sometimes preceded by a brief period of hyperthyroidism [21]. In addition, the severity of HIV disease, as well as undernutrition and weight loss, may lead to thyroid function test abnormalities similar to those in non-thyroidal illness (i.e. 'euthyroid sick syndrome'). In advanced HIV, however, both T_3 and reverse T_3 (rT_3) tend to be reduced in direct association to serum albumin and surrogate measures of muscle mass, in contrast to the rT_3 elevations typically found in non-thyroidal illness.

Bone Health, Calcium, and Vitamin D

Calcium and Vitamin D

Vitamin D insufficiency is relatively common in individuals with HIV, as in the general population, and vitamin D deficiency may be associated with more rapid disease progression [23]. People with HIV share many risk factors for vitamin D deficiency with the general population, and, additionally, some types of cART reduce vitamin D. Use of efavirenz, a non-nucleoside reverse

transcriptase inhibitor (NNRTI), is associated with vitamin D deficiency [24]. In addition, *in vitro* studies demonstrate that PIs inhibit both 25-hydroxylation and 1-alpha hydroxylation, while also mildly inhibiting 24-hydroxylation, with a net effect of decreasing 1,25 dihydroxyvitamin D production [25]. Guidelines recommend checking 25-hydroxyvitamin D levels in individuals with HIV and risk factors for deficiency, including dark skin, malnutrition, or malabsorption, obesity, efavirenz use, and chronic kidney disease [26].

Severe HIV infection may be associated with hypocalcaemia, with mechanisms including malabsorption, vitamin D deficiency, abnormal protein binding, medication effects, hypomagnesaemia, and altered parathyroid function. The intestinal tract of patients with advanced HIV is often the target of OIs, which may result in malabsorption of calcium and vitamin D. Hypocalcaemia associated with severe illness most often results from hypoalbuminemia and abnormal protein binding.

Bone Density and Fracture Risk in HIV Infection

Bone mineral density (BMD) is decreased in both children and adults with HIV [27, 28]. In children, lower BMD appears to be related to increasing severity of disease as well as decreased height-for-age, weight-for-age, and IGF-1 levels [27]. In adults, decreased bone density has been associated with decreased gonadal steroids, decreased vitamin D levels, lower IGF-1, lower body mass index (BMI) and/or weight loss, visceral fat accumulation, duration of HIV, and increasing severity of disease. In a large longitudinal analysis, decreased BMD over time was predicted by low body weight, albumin, corticosteroid use, and menopause, whereas strength training was protective [29]. With regard to the effect of antiretrovirals on bone density, one NRTI, tenofovir disoproxil fumarate (TDF), has a clear effect to reduce bone density. Additionally, PIs 'boosted' with ritonavir appear to decrease BMD to a greater degree than other regimens [26].

Reduced BMD in individuals with HIV translates into increased fracture risk. In an epidemiologic study in a large healthcare system, women with HIV were more likely to sustain vertebral and wrist fractures than non-HIV-infected patients, and men with HIV were more likely to sustain any fracture compared to non-HIV-infected patients. The overall fracture prevalence in the combined cohort was 2.87 fractures per 100 persons in the HIV-infected group vs. 1.77 fractures per 100 persons in the uninfected controls [30]. A study of Canadian women demonstrated a similar increase in risk, with women with HIV 1.7 times more likely to sustain low-impact fractures than uninfected women [31].

Screening and Treatment of Osteopenia and Osteoporosis in HIV

Strategies for screening and management of osteopenia and osteoporosis are shown in **Box 12.1.5.3** [26]. For individuals with HIV and osteopenia or osteoporosis, contributing endocrine conditions or medications should be considered. Hypogonadism may contribute to reduced bone density, and physiologic testosterone replacement in women with HIV and reduced androgen levels significantly increased bone density over 18 months. For individuals receiving TDF, switching to tenofovir alafenamide (TAF) significantly improves BMD by 2–3% after 144 weeks [32]. The bisphosphonates alendronate and zoledronic acid have also proven safe and effective to increase BMD in patients with HIV [26]. There have not yet been studies assessing the effects of exercise, androgen, or bisphosphonate on fracture risk in the HIV-infected population.

Box 12.1.5.3 Summary of recommended management of bone health in HIV [26]

Screening

25-hydroxyvitamin D deficiency

Indications: low BMD, history of fracture, or major risk factor(s) for vitamin D deficiency, including efavirenz use, dark skin, malnutrition or malabsorption, obesity, or chronic kidney disease

Frequency: per clinical judgement

Fracture risk assessment by FRAX score

Indications: men 40–49 yo and premenopausal women ≥40 yo without major risk for fragility fracture

Frequency: every 2–3 years, or with development of a new clinical risk factor

Bone density assessment with DXA

Indications: (1) men 40–49 yo and premenopausal women ≥40 yo with intermediate- or high-risk stratification by FRAX; (2) all postmenopausal women; (3) all men ≥50 y; (4) all adults with major fragility fracture risk factors regardless of age

Frequency: Consider repeat every 1–2 years if advanced osteopenia, every 5 years if moderate osteopenia, or up to every 15 years if normal bone density

Management

Lifestyle optimization

Interventions: Increased weight bearing exercise, smoking cessation, avoidance of excessive alcohol use, fall prevention

Indications: All patients with HIV

Calcium and vitamin D optimization

Interventions: (1) Calcium: ensure 1000 mg daily for men 50–70 yo, or 1200 daily for women ≥51 yo and men ≥71 yo; (2) Vitamin D: ensure 1000 IU daily if 25-hydroxyvitamin D levels normal; if insufficient or deficient, treat with higher doses until serum 25-hydroxyvitamin D >75 nmol/L (30 ng/ml)

Indications: All patients with HIV

Treatment of osteoporosis

Intervention: (1) If possible, discontinue or prescribe alternatives for medications that cause bone loss, including glucocorticoids and TDF. (2) Alendronate or zoledronic acid

Indications: Bisphosphonates should be prescribed according to country- or region-specific guidelines for treatment of osteoporosis in the general population

Osteomalacia and Osteonecrosis

Osteomalacia may occur in individuals using efavirenz or TDF or in those with severe vitamin D deficiency [26]. Additionally, individuals with HIV are at risk for osteonecrosis, typically occurring at the femoral head [33]. Though the mechanisms are unknown, case-control studies demonstrate a significantly increased risk with exposure to antiretroviral medications [33]. Other identified risk factors include alcohol consumption, use of corticosteroids, a prior AIDS-defining illness, and low CD4+ count. Although the absolute incidence of osteonecrosis remains relatively low, estimated at approximately 0.2 to 0.6 cases per 100 person-years, individuals with HIV appear to have increased risk over the general population [34].

Salt and Water Balance

In general, patients on effective cART are not expected to have abnormalities in fluid balance or electrolytes, although these may occur in advanced illness. Recent evidence suggests more subtle perturbations of the renin angiotensin aldosterone system (RAAS), however, with increased aldosterone in individuals with HIV compared to non-infected individuals [35]. RAAS activation appears to be associated with visceral adiposity and insulin resistance in HIV, although more study is needed to determine if therapeutic targeting of RAAS improves body composition or metabolic health.

Sodium

Sodium and water balance may be disturbed in advanced HIV disease, most commonly manifesting as hyponatraemia secondary to syndrome of inappropriate antidiuretic hormone secretion (SIADH). Certain medications such as vidarabine, miconazole, and pentamidine are associated with hyponatraemia of unknown aetiology. Hypernatraemia and nephrogenic diabetes insipidus have been reported with foscarnet therapy and also in association with CMV infection of the hypothalamus.

Potassium

Diarrheal OIs may lead to hypokalaemia in patients with advanced HIV disease. In addition, TDF has been associated with development of renal Fanconi's syndrome, with accompanying hypokalaemia, hypophosphatemia, and acidosis. Foscarnet therapy for CMV has also been associated with hypokalaemia, hypomagnesaemia, and hypophosphatemia.

Lipid Metabolism

HIV is associated with abnormalities in lipid metabolism, with decreased HDL and decreased low-density lipoprotein (LDL) occurring early in the disease [36]. Triglyceride levels also increase in HIV infection in conjunction with increased *de novo* hepatic lipogenesis [36]. These lipid abnormalities are seen in children as well as adults. In patients receiving cART, hypertriglyceridemia may be more pronounced, particularly in patients with altered fat distribution (Figure 12.1.5.2) [37]. Moreover, in contrast to patients with untreated HIV, in whom LDL is often low, patients receiving cART may demonstrate an increase in LDL to pre-HIV levels or beyond [36, 38]. The lipid profile often depends on the specific components of the antiretroviral regimen, with some PIs demonstrating particularly adverse effects. In one large cohort comparing antiretroviral naïve patients to those on different cART regimens, use of PI and NRTI was associated with 27% prevalence of hypercholesterolemia (total cholesterol ≥ 6.2 mmol/L) and 40% prevalence of hypertriglyceridemia (triglyceride ≥ 2.3 mmol/L), compared to 8% and 15% prevalence, respectively, in antiviral naïve patients [39]. A significant effect of PIs on lipids is also seen in paediatric cohorts. Newer classes of cART, such as integrase inhibitors and fusion inhibitors, as well as newer agents in the NRTI, PI, and NNRTI classes, may have fewer adverse effects on lipid profile.

Individuals with HIV should have lipid panels obtained at the initiation of care and routinely thereafter [40]. As in the general population, lifestyle changes and exercise are an important initial treatment strategy for dyslipidaemia, and formal exercise programmes effectively reduce triglyceride and total cholesterol in individuals with HIV [41]. If an individual is receiving a cART regimen with particularly adverse effects on lipid, such as a regimen with 'boosted' PI, switching to another regimen with more favourable lipid effects can be considered if feasible. If pharmacotherapy is needed, medical treatment of dyslipidaemia in HIV-infected patients currently follows the same principles as treatment of lipid abnormalities in the general population, although the known increase in cardiovascular risk in HIV may argue for more aggressive treatment. Interactions between lipid-lowering agents and antiretrovirals may alter therapeutic options [41, 42]. For instance, many PIs are metabolized by CYP3A4, affecting HMG-CoA reductase concentrations. Lovastatin and simvastatin are contraindicated with many PIs. Rosuvastatin may also interact with specific PIs and may require dose adjustment. Pravastatin and pitavastatin are not metabolized by CYP3A4 and may be given safely [43]. A study comparing treatment in HIV-infected vs. uninfected individuals showed similar LDL-lowering benefit in both groups [43]. In addition, studies suggest that statins improve endothelial function in HIV-infected patients [42] and reduce non-calcified plaque volume and high-risk plaque features. A multinational study investigating the benefit of statins on primary CVD prevention in the HIV-infected population is ongoing.

Many patients with HIV demonstrate significant hypertriglyceridemia (>500–900 mg/dl or >5.65–10 mmol/L), which can be treated with gemfibrozil or fenofibrate [41]. Omega-3 fatty acids may also be effective.

Glucose Homeostasis

In HIV-infected individuals naïve to antiretrovirals, glucose abnormalities are rare and caused most often by pancreatic damage and reduced insulin secretion from pentamidine administration or hyperglycaemia from megestrol acetate. For individuals treated with cART, however, impaired glucose tolerance is relatively common, particularly in patients with HIV-associated fat redistribution. In one cohort, 35% of patients with HIV lipodystrophy had impaired glucose tolerance (IGT), and 7% had frank type 2 diabetes that had been previously undiagnosed (see Figure 12.1.5.2) [37]. A more recent study demonstrated a 14% prevalence of diabetes in a large group of HIV-infected individuals receiving cART (with or without lipodystrophy) compared to a 5% prevalence in uninfected controls [44]. With the development of newer antiretroviral agents, the incidence of diabetes appears to be decreasing in HIV, suggesting that much of the dysglycaemia in HIV is due to specific medication effects.

With the high prevalence of IGT and diabetes in HIV-infected individuals, evaluation of glucose homeostasis is important in this population, particularly in patients with changes in fat distribution. Notably, haemoglobin A1c values may underestimate glycaemia in HIV-infected individuals, particularly those receiving NRTIs, such that fasting glucose may be a more accurate method for diagnosing diabetes in this population [45]. As in the general population, lifestyle modification is first line for individuals with glucose intolerance and early diabetes. Additionally, if an individual is receiving

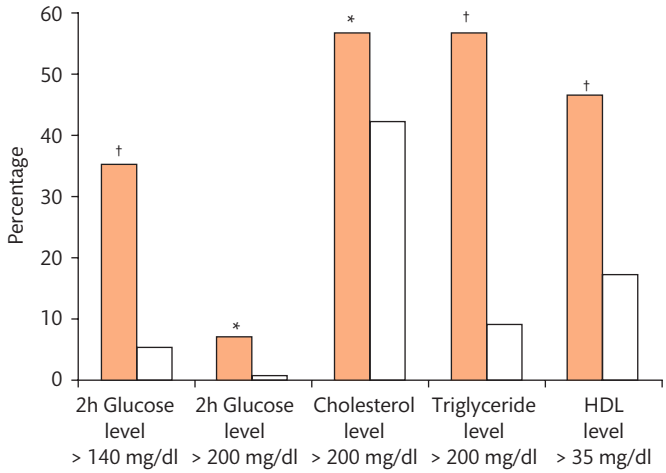


Figure 12.1.5.2 Percentages of impaired glucose tolerance, diabetes, and elevated lipids in a cohort of individuals with HIV lipodystrophy (dark bars) vs. Framingham controls (white bars) matched for sex, age, and BMI. *p <0.05, †p <0.001 for comparison of unadjusted odds ratios between groups.

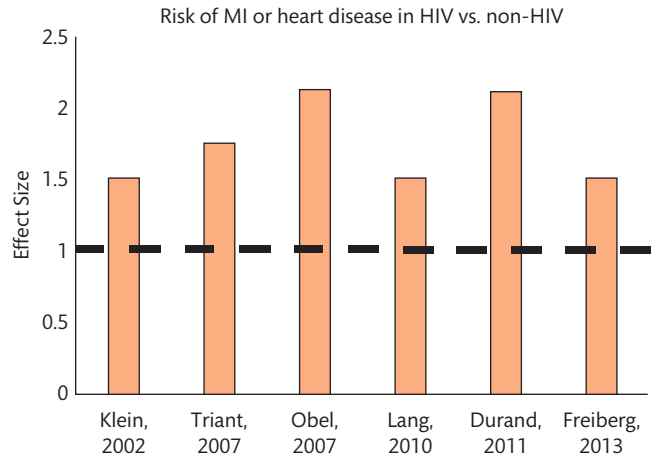


Figure 12.1.5.3 Relative risks for myocardial infarction (MI) or heart disease in various cohort studies in individuals with HIV infection compared to those without. In all studies, the HIV group was comprised, at least in part, of individuals treated with cART [47, 50–54]. The dotted line at 1 would indicate no increased risk conferred by HIV.

lopinavir/ritonavir, zidovudine, or stavudine, all of which have particularly adverse effects on glucose homeostasis, a switch in cART could be considered if feasible. For treatment of insulin resistance and diabetes in HIV, metformin is first line and may also modestly improve body composition and cardiovascular risk. Metformin may interact with dolutegravir, requiring reduced metformin dose. In patients with peripheral lipoatrophy, thiazolidinediones (TZDs) are an alternative potential strategy to both improve insulin resistance and increase limb fat, but use of TZDs is limited due to adverse cardiovascular effects as well as weight gain. The other antidiabetic agents available for general use (sulfonylureas, glucagon-like peptide 1 (GLP-1) analogues, gliflozins, dipeptidyl peptidase 4 (DPP-IV) inhibitors, and meglitinides) can also be used [45], although currently none of these is extensively studied in the HIV population. For all of these agents, drug monographs should be consulted to assess for cytochrome P450 interactions that would affect antiretroviral concentrations. Insulin therapy is the preferred second-line medication, following metformin, if HbA1c is over 8.5% [45].

Cardiovascular Risk

Numerous studies have demonstrated increased rates of coronary heart disease and approximately twice the rate of myocardial infarction in HIV-infected individuals compared to the general population (Figure 12.1.5.3). Although increased CVD risk relates in part to traditional risk factors such as dyslipidaemia, dysglycaemia, and smoking, HIV-related mechanisms such as immune activation, increased microbial translocation, and virus-mediated changes in cholesterol metabolism also seem to play a large role [46]. In a large veterans' cohort in the United States, the hazard ratio of myocardial infarction was 1.48 (95% confidence interval 1.27–1.72) in individuals with HIV after adjusting for comorbidities, substance use, and traditional risk factors as measured by Framingham Risk Score [47]. Multiple markers of immune activation, including soluble CD163 and soluble CD14, are elevated in HIV and strongly associated with increased CVD risk [46]. These markers decrease with cART but remain elevated above levels found in non-infected controls. Additionally, serum markers of microbial translocation are elevated in HIV and are associated with measures of subclinical atherosclerosis, suggesting that impaired gut integrity in HIV may contribute to chronic inflammation and CVD risk [46]. Trials to reduce immune activation and inflammation in hopes of decreasing CVD risk are an active area of investigation in HIV. Some evidence has suggested that certain antiretroviral agents may increase the risk of CVD, but there is not clear evidence that newer cART regimens increase risk. A study of continuous cART versus an intermittent regimen based on CD4+ count demonstrated higher cardiovascular risk in the intermittent regimen [48], in association with higher systemic inflammatory markers [49]. These results suggest that sustained cART, by achieving viral suppression and reducing systemic inflammation, may actually lower CVD risk, but further research is needed.

Given the adverse cardiometabolic effects of HIV just described, effective treatment of risk factors for cardiovascular disease is crucial for HIV-infected patients. As in the general population, prevention strategies include smoking cessation as well as screening and treatment for dyslipidaemia, hypertension, and disordered glucose metabolism [42].

Summary

HIV is associated with a number of endocrine abnormalities, largely dependent on the stage of disease and the medications used for therapy. In individuals with advanced HIV disease or AIDS, OIs and the medications used to treat them often play a role by damaging endocrine organs and altering hormone synthesis. Weight loss is a predictor of mortality in this group, and clinicians should be aware of strategies to increase lean body mass, including

physiologic androgen replacement and potentially GH. In contrast, HIV-infected individuals treated with cART may present quite differently, with visceral adiposity, relative GH deficiency, hyperlipidaemia, and altered glucose homeostasis. In these patients, management of hyperlipidaemia, impaired glucose metabolism, and CVD risk is of primary concern. HIV-infected patients generally demonstrate decreased bone density, increased prevalence of subclinical thyroid disease, and increased cardiovascular risk as evidenced by increased myocardial infarction rate as well as increased carotid intima-media thickness. Children with HIV also demonstrate many of these abnormalities and may also present with growth failure, particularly in severe disease.

ACKNOWLEDGEMENTS

This work was partially supported by NIH Grant P30-DK040561. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center For Research Resources or the National Institutes of Health. The authors would also like to thank all of the participants in HIV-research for their dedication to advancing the medical care of HIV.

REFERENCES

1. World Health Organization GHOD. HIV/AIDS 2018. Available at: http://www.who.int/gho/hiv/en/ (accessed: 10 March 2020).
2. Koethe JR, Heimburger DC, PrayGod G, Filteau S. From wasting to obesity: the contribution of nutritional status to immune activation in HIV infection. *J Infect Dis*, 2016; **214** Suppl 2: S75–82.
3. Yuh B, Tate J, Butt AA, *et al.* Weight change after antiretroviral therapy and mortality. *Clin Infect Dis*, 2015; **60**: 1852–9.
4. Lake JE, Stanley TL, Apovian CM, *et al.* Practical review of recognition and management of obesity and lipohypertrophy in human immunodeficiency virus infection. *Clin Infect Dis*, 2017; **64**: 1422–9.
5. Guaraldi G, Squillace N, Stentarelli C, *et al.* Nonalcoholic fatty liver disease in HIV-infected patients referred to a metabolic clinic: prevalence, characteristics, and predictors. *Clin Infect Dis*, 2008; **47**: 250–7.
6. Macias J, Mancebo M, Merino D, *et al.* Changes in liver steatosis after switching from efavirenz to raltegravir among human immunodeficiency virus-infected patients with nonalcoholic fatty liver disease. *Clin Infect Dis*, 2017; **65**: 1012–19.
7. Samaras K, Gan SK, Peake PW, Carr A, Campbell LV. Proinflammatory markers, insulin sensitivity, and cardiometabolic risk factors in treated HIV infection. *Obesity (Silver Spring)*, 2009; **17**: 53–9.
8. Williams PL, Jesson J. Growth and pubertal development in HIV-infected adolescents. *Curr Opin HIV AIDS*, 2018; **13**: 179–86.
9. Seth A, Malhotra RK, Gupta R, *et al.* Effect of antiretroviral therapy on growth parameters of children with HIV infection. *Pediatr Infect Dis J*, 2018; **37**: 85–9.
10. Chantry CJ, Hughes MD, Alvero C, *et al.* Insulin-like growth factor-1 and lean body mass in HIV-infected children. *J Acquir Immune Defic Syndr*, 2008; **48**: 437–43.
11. Vigano A, Mora S, Brambilla P, *et al.* Impaired growth hormone secretion correlates with visceral adiposity in highly active antiretroviral treated HIV-infected adolescents. *AIDS*, 2003; **17**: 1435–41.
12. Dobs AS, Dempsey MA, Ladenson PW, Polk BF. Endocrine disorders in men infected with human immunodeficiency virus. *Am J Med*, 1988; **84**(3 Pt 2): 611–16.
13. Slama L, Jacobson LP, Li X, *et al.* Longitudinal changes over 10 years in free testosterone among HIV-infected and HIV-uninfected men. J Acquir Immune Defic Syndr, 2016; **71**: 57–64.
14. Bhasin S, Parker RA, Sattler F, *et al.* Effects of testosterone supplementation on whole body and regional fat mass and distribution in human immunodeficiency virus-infected men with abdominal obesity. *J Clin Endocrinol Metab*, 2007; **92**: 1049–57.
15. Chirgwin KD, Feldman J, Muneyyirci-Delale O, Landesman S, Minkoff H. Menstrual function in human immunodeficiency virus-infected women without acquired immunodeficiency syndrome. *J Acquir Immune Defic Syndr Hum Retrovirol*, 1996; **12**: 489–94.
16. Karim R, Mack WJ, Kono N, *et al.* Gonadotropin and sex steroid levels in HIV-infected premenopausal women and their association with subclinical atherosclerosis in HIV-infected and -uninfected women in the women's interagency HIV study (WIHS). *J Clin Endocrinol Metab*, 2013; **98**: E610–18.
17. Rabkin JG, Ferrando SJ, Wagner GJ, Rabkin R. DHEA treatment for HIV+ patients: effects on mood, androgenic and anabolic parameters. *Psychoneuroendocrinology*, 2000; **25**: 53–68.
18. Findling JW, Buggy BP, Gilson IH, Brummitt CF, Bernstein BB, Raff H. Longitudinal evaluation of adrenocortical function in patients with the human immunodeficiency virus. *J Clin Endocrinol Metab*, 1994; **79**: 1091–6.
19. Membreno L, Irony I, Dere W, Klein R, Biglieri EG, Cobb E. Adrenocortical function in acquired immune deficiency syndrome. *J Clin Endocrinol Metab*, 1987; **65**: 482–7.
20. Norbiato G, Bevilacqua M, Vago T, *et al.* Cortisol resistance in acquired immunodeficiency syndrome. *J Clin Endocrinol Metab*, 1992; **74**: 608–13.
21. Hoffmann CJ, Brown TT. Thyroid function abnormalities in HIV-infected patients. *Clin Infect Dis*, 2007; **45**: 488–94.
22. Hirschfeld S, Laue L, Cutler GB, Jr., Pizzo PA. Thyroid abnormalities in children infected with human immunodeficiency virus. *J Pediatr*, 1996; **128**: 70–4.
23. Havers F, Smeaton L, Gupte N, *et al.* 25-Hydroxyvitamin D insufficiency and deficiency is associated with HIV disease progression and virological failure post-antiretroviral therapy initiation in diverse multinational settings. *J Infect Dis*, 2014; **210**: 244–53.
24. Welz T, Childs K, Ibrahim F, *et al.* Efavirenz is associated with severe vitamin D deficiency and increased alkaline phosphatase. *AIDS*, 2010; **24**: 1923–8.
25. Cozzolino M, Vidal M, Arcidiacono MV, Tebas P, Yarasheski KE, Dusso AS. HIV-protease inhibitors impair vitamin D bioactivation to 1,25-dihydroxyvitamin D. *AIDS*, 2003; **17**: 513–20.
26. Brown TT, Hoy J, Borderi M, *et al.* Recommendations for evaluation and management of bone disease in HIV. *Clin Infect Dis*, 2015; **60**: 1242–51.
27. Jacobson DL, Spiegelman D, Duggan C, *et al.* Predictors of bone mineral density in human immunodeficiency virus-1 infected children. *J Pediatr Gastroenterol Nutr*, 2005; **41**: 339–46.
28. Borderi M, Gibellini D, Vescini F, *et al.* Metabolic bone disease in HIV infection. *AIDS*, 2009; **23**: 1297–310.
29. Jacobson DL, Spiegelman D, Knox TK, Wilson IB. Evolution and predictors of change in total bone mineral density over time in

HIV-infected men and women in the nutrition for healthy living study. *J Acquir Immune Defic Syndr*, 2008; **49**: 298–308.
30. Triant VA, Brown TT, Lee H, Grinspoon SK. Fracture prevalence among human immunodeficiency virus (HIV)-infected versus non-HIV-infected patients in a large U.S. healthcare system. *J Clin Endocrinol Metab*, 2008; **93**: 3499–504.
31. Prior J, Burdge D, Maan E, *et al.* Fragility fractures and bone mineral density in HIV positive women: a case-control population-based study. *Osteoporos Int*, 2007; **18**: 1345–53.
32. Brown TT, Yin MT, Gupta SK, *et al.* Combined effects of bisphosphonates & TDF to TAF switch in HIV+ adults with low BMD. Conference on Retroviruses and Opportunistic Infections; 4–7 March 2018; Boston, MA, 2018.
33. Ho YC, Shih TT, Lin YH, *et al.* Osteonecrosis in patients with human immunodeficiency virus type 1 infection in Taiwan. *Jpn J Infect Dis*, 2007; **60**: 382–6.
34. Morse CG, Mican JM, Jones EC, *et al.* The incidence and natural history of osteonecrosis in HIV-infected adults. *Clin Infect Dis*, 2007; **44**: 739–48.
35. Srinivasa S, Fitch KV, Wong K, *et al.* RAAS activation is associated with visceral adiposity and insulin resistance among HIV-infected patients. *J Clin Endocrinol Metab*, 2015; **100**: 2873–82.
36. Grunfeld C, Kotler DP, Arnett DK, *et al.* Contribution of metabolic and anthropometric abnormalities to cardiovascular disease risk factors. *Circulation*, 2008; **118**: e20–8.
37. Hadigan C, Meigs JB, Corcoran C, *et al.* Metabolic abnormalities and cardiovascular disease risk factors in adults with human immunodeficiency virus infection and lipodystrophy. *Clin Infect Dis*, 2001; **32**: 130–9.
38. Grinspoon S, Carr A. Cardiovascular risk and body-fat abnormalities in HIV-infected adults. *N Engl J Med*, 2005; **352**: 48–62.
39. Friis-Moller N, Weber R, Reiss P, *et al.* Cardiovascular disease risk factors in HIV patients--association with antiretroviral therapy. Results from the DAD study. *AIDS*, 2003; **17**: 1179–93.
40. Aberg JA, Gallant JE, Ghanem KG, *et al.* Primary care guidelines for the management of persons infected with HIV: 2013 update by the HIV Medicine Association of the Infectious Diseases Society of America. *Clin Infect Dis*, 2014; **58**: 1–10.
41. Dube MP, Stein JH, Aberg JA, *et al.* Guidelines for the evaluation and management of dyslipidemia in human immunodeficiency virus (HIV)-infected adults receiving antiretroviral therapy: recommendations of the HIV Medical Association of the Infectious Disease Society of America and the Adult AIDS Clinical Trials Group. *Clin Infect Dis*, 2003; **37**: 613–27.
42. Stein JH, Hadigan CM, Brown TT, *et al.* Prevention strategies for cardiovascular disease in HIV-infected patients. *Circulation*, 2008; **118**: e54–60.
43. Silverberg MJ, Leyden W, Hurley L, *et al.* Response to newly prescribed lipid-lowering therapy in patients with and without HIV infection. *Ann Intern Med*, 2009; **150**: 301–13.
44. Brown TT, Cole SR, Li X, *et al.* Antiretroviral therapy and the prevalence and incidence of diabetes mellitus in the multicenter AIDS cohort study. *Arch Intern Med*, 2005; **165**: 1179–84.
45. Monroe AK, Glesby MJ, Brown TT. Diagnosing and managing diabetes in HIV-infected patients: current concepts. *Clin Infect Dis*, 2015; **60**: 453–62.
46. Nou E, Lo J, Hadigan C, Grinspoon SK. Pathophysiology and management of cardiovascular disease in patients with HIV. *Lancet Diabetes Endocrinol*, 2016; **4**: 598–610.
47. Freiberg MS, Chang CC, Kuller LH, *et al.* HIV infection and the risk of acute myocardial infarction. *JAMA Intern Med*, 2013: **173**: 614–22.
48. El-Sadr WM, Lundgren JD, Neaton JD, *et al.* CD4+ count-guided interruption of antiretroviral treatment. *N Engl J Med*, 2006; **355**: 2283–96.
49. Duprez DA, Neuhaus J, Kuller LH, *et al.* Inflammation, coagulation and cardiovascular disease in HIV-infected individuals. *PLoS One*, 2012; **7**: e44454.
50. Klein D, Hurley LB, Quesenberry CP, Jr., Sidney S. Do protease inhibitors increase the risk for coronary heart disease in patients with HIV-1 infection? *J Acquir Immune Defic Syndr*, 2002; **30**: 471–7.
51. Triant VA, Lee H, Hadigan C, Grinspoon SK. Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease. *J Clin Endocrinol Metab*, 2007; **92**: 2506–12.
52. Obel N, Thomsen HF, Kronborg G, *et al.* Ischemic heart disease in HIV-infected and HIV-uninfected individuals: a population-based cohort study. Clin Infect Dis, 2007; **44**: 1625–31.
53. Lang S, Mary-Krause M, Cotte L, *et al.* Increased risk of myocardial infarction in HIV-infected patients in France, relative to the general population. *AIDS*, 2010; **24**: 1228–30.
54. Durand M, Sheehy O, Baril JG, Lelorier J, Tremblay CL. Association between HIV infection, antiretroviral therapy, and risk of acute myocardial infarction: a cohort and nested case-control study using Quebec's public health insurance database. *J Acquir Immune Defic Syndr*, 2011; **57**: 245–53.

12.1.6 The Endocrinology of Anorexia Nervosa

Karen K. Miller

Introduction

Sir William Gull, a personal physician to Queen Victoria, coined the term 'anorexia nervosa' in the 1860s to describe a syndrome of wasting in young female patients, for which he could discern no organic cause, such as tuberculosis, which was a common cause of such signs and symptoms at the time [1]. He surmised that anorexia nervosa was of primary psychiatric origin—a construct that persists today. The most recent psychiatric diagnostic criteria (published in 2013) are delineated in the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (*DSM-5)*, which defines anorexia nervosa as a disorder characterized by food restriction leading to low weight, an intense fear of gaining weight or becoming fat (or persistent behaviours designed to prevent weight gain), and altered body image (such that the patient is convinced that she is

overweight when she is not). With the recognition that there are a wide range of physical manifestations of this psychiatric syndrome, these revised diagnostic criteria relaxed the weight criteria and removed the amenorrhoea requirement that had been included in the prior diagnostic criteria [2]. Moreover, DSM-V added a number of additional diagnostic criteria, including 'atypical anorexia nervosa', in which patients share the psychiatric characteristics of anorexia nervosa, but without low weight [2].

Anorexia nervosa is a common disorder, particularly in adolescent girls and young women. It has a prevalence of 0.3 to 3% in women [3] and it ranks third in prevalence among chronic disorders in adolescent girls [4]. However, it should be noted that the disorder also affects older women and men, though much less frequently. Therefore, very little is known about the endocrine complications of the disorder in these groups, and this chapter will focus on endocrine dysregulation in girls and women with anorexia nervosa.

Anorexia nervosa is biologically a disorder in which chronic starvation affects multiple endocrine systems, including classic hypothalamic–pituitary, adipokine, and satiety pathways. Much of this dysregulation is adaptive to the starved state and reverses with nutritional recovery. However, it remains relevant because the psychiatric disorder is chronic in approximately 50% of women [5], and because endocrine dysregulation can have long lasting consequences on skeletal, and other aspects of, health.

Hypothalamic–Pituitary Axis Dysregulation

Electrolyte disturbances, particularly **hyponatraemia** and hypokalaemia, occur in 20% of women with anorexia nervosa [6]. Hyponatraemia can be caused by syndrome of inappropriate antidiuresis (SIAD) and/or primary polydipsia and rarely results in seizures. **Diabetes insipidus** has also rarely been reported. Hypokalaemia is associated with purging behaviours (self-induced vomiting, laxative overuse, and/or diuretic use). See Table 12.1.6.1 [7].

Table 12.1.6.1 Endocrine dysfunction in women and adolescent girls with anorexia nervosa

	Women and adolescent girls with anorexia nervosa	Implicated in bone loss
Hypothalamic amenorrhoea	✓	✓
GH resistance	✓	
↑GH	✓	
↓GF-1	✓	✓
Hypercortisolaemia	✓	✓
Appetite hormone dysregulation	✓	
↑PYY	✓	✓
↓Leptin	✓	
↑Ghrelin	✓	

Abbreviations: IGF-1, insulin-like growth factor I; PYY, peptide YY.

Reproduced with permission from Miller KK. Endocrine effects of anorexia nervosa. *Endocrinol Metab Clin North Am.* 2013;42(3):515–28. Copyright © 2013 Elsevier Inc. [7].

Non-thyroidal illness is a common complication of this disorder [8]—and is adaptive to the starved state. For this reason and because of the potential for levothyroxine abuse to induce weight loss, levothyroxine therapy should only be prescribed for frank hypothyroidism.

Hypothalamic amenorrhoea is a common—but not universal—complication of anorexia nervosa and results in relative oestrogen, androgen, and progesterone deficiency. Our understanding of the interplay of starvation, the hypothalamic–pituitary–gonadal axis and fat stores has advanced since the 1970s, when Drs Frisch and McArthur published their seminal work describing a 'minimum weight for height' required for menarche [9]. In the same decade, Boyar *et al.*, published important work demonstrating luteinizing hormone (LH) apulsatility in women with anorexia nervosa, which recovered to normal with weight gain [10]. However, Dr Boyar's group also noted that this seemingly clear construct, in which starvation results in near total cessation of gonadotropin-releasing hormone (GnRH) secretion by the hypothalamus, does not capture the complexity of the hypothalamic–pituitary–gonadal (HPG) axis in women with anorexia nervosa. In addition to the absence of GnRH—and therefore LH—pulsations, they described a number of other patterns, including LH pulsatility only at night, as seen in normal late puberty [10]. They also demonstrated that the severity of derangement of LH pulsatility is not directly related to the length of time the patient has been suffering from the disease or the degree of low weight. Moreover, they demonstrated that weight gain did not reliably result in return of menses [11]. In fact, approximately 15% of women with anorexia nervosa remain amenorrhoeic despite weight gain to a 'healthy' level [12]. Whether the lack of recovery of reproductive function in such women reflects hypothalamic or pituitary atrophy due to prolonged suppression, a resetting of the HPG axis, continued psychologic stress, lack of sufficient fat mass despite weight gain and/or other factors, is unknown.

The importance of adequate fat mass to initiate and sustain normal reproductive function has mostly focused on the adipokine leptin, a 16 kDa peptide secreted by adipocytes which is permissive (necessary) for normal reproductive function. Leptin is an anorexigenic hormone. However, its primary function appears to be as a signal of inadequate energy stores to the reproductive, and other, systems. Evidence in humans of leptin's key role in regulating reproduction include that congenital leptin deficiency is complicated by hypogonadotropic hypogonadism [13]. Rodent data demonstrate that leptin deficiency is a critical link between starvation and hypothalamic amenorrhoea. In an elegant study, Ahima *et al.* showed that whereas acute starvation delayed the rodent equivalent of ovulation (oestrous), simultaneous administration of leptin to 'normal' blood levels (levels characteristic of non-starved mice) resulted in normal reproductive function [14]. Although, mean serum leptin levels are lower in women with anorexia nervosa who experience amenorrhoea than in those without amenorrhoea [15], there is considerable overlap in individual levels between these two groups. This demonstrates the considerable complexity and variable intraindividual vulnerability of the female reproductive system to stress and also highlights the lack of utility of leptin levels in clinical practice when, for example, setting weight goals for recovery in women with hypothalamic amenorrhoea due to anorexia nervosa [15]. In a study of eight healthy-weight women with hypothalamic amenorrhoea (BMI 18.8–24.4 kg/m^2), leptin administration resulted in ovulatory cycles

in three. However, because of the resultant weight loss (to be expected because leptin is an anorexigenic hormone) [16], it cannot be used in women with anorexia nervosa in whom nutritional repletion must be the primary goal. Other putative mediators of reproductive function have been studied. Although the role of kisspeptin, the endogenous ligand for the hypothalamic kisspeptin receptor, has not been studied in women with anorexia nervosa, acute administration of kisspeptin increases gonadotropin release in healthy-weight women with hypothalamic amenorrhoea [17].

In addition to hypo-oestrogenaemia, the effects of hypogonadism in women with anorexia nervosa also include androgen and progesterone deficiency. Both testosterone and free testosterone levels are reduced in women with anorexia nervosa, with a further reduction in free testosterone in women with the syndrome who also take oral contraceptives [18]. Low androgen levels result from a reduction of ovarian, not adrenal, precursors [19], consistent with the diagnosis of hypothalamic amenorrhoea, though there are conflicting data as to whether adrenal pre-androgen levels are normal or low in girls and women with anorexia nervosa [18, 20]. When considering the end organ effects of hypothalamic amenorrhoea in girls and women with anorexia nervosa, the effects of oestrogen, progesterone, and androgen deficiencies should all be considered. If one extrapolates from men with regard to the effects of testosterone deficiency, one could surmise that women with relative androgen deficiency might experience decreased muscle mass and bone mineral density (BMD), and depressed mood. However, testosterone levels are ten to 20 times lower in healthy women than men and therefore direct extrapolation is not possible. A small randomized study of low-dose transdermal testosterone (300 mcg daily) administration demonstrated increases in lean body mass but not BMD [21], and a cross-sectional study demonstrated an inverse association between severity of depression and anxiety symptoms and free and total testosterone levels [22, 23]. However, a small, randomized, placebo-controlled trial of low-dose transdermal testosterone showed no improvement in weight, eating disorder cognition or depression symptom severity. Therefore, there is no evidence that testosterone administration ameliorates such symptoms. Moreover, there are no FDA-approved testosterone preparations that deliver appropriate female doses, and the long-term effects of testosterone replacement therapy in women are unknown.

GH resistance is another endocrinopathy observed in girls and women with anorexia nervosa. That GH levels are relatively high in association with acute and chronic fasting was first reported in a classic paper by Roth *et al.* in 1963 in *Science*—the same publication in which the insulin tolerance test was described [24]. Later it was shown that despite the relatively high GH concentrations in women with anorexia nervosa (basal levels four times higher and pulsatile release 20-times greater than in healthy controls) [25], insulin-like growth factor 1 (IGF-1) levels (total and bioactive) are relative low, consistent with a state of resistance at the level of the liver to the effects of GH [26]. This is because nutrition and insulin are required for IGF-1 production by GH in the liver [27].

Hypercortisolaemia is present in approximately one-third of girls and women with anorexia nervosa, as originally shown by Boyar and colleagues [8]. Some studies have shown that hypercortisolism may not be dexamethasone-suppressible and can, but does not always, reverse with restoration of weight. Despite low weight, the consequences of hypercortisolaemia in girls and women with anorexia nervosa share some similarities with those with Cushing's syndrome, with disproportionate accumulation of trunk fat during recovery [28] and bone loss [29], the latter remained significant after controlling for months of amenorrhoea, age, and IGF-1 levels (though low body weight, a determinant of both high cortisol and low BMD was not included in the statistical models). A positive result of hypercortisolaemia—perhaps mediated by the relative increase in trunk fat observed—is that girls with anorexia nervosa and relatively higher cortisol levels are more likely to regain menses during the following year [30].

Satiety Pathway Dysregulation

Some appetite-regulating hormone pathways are also dysregulated in girls and women with anorexia nervosa (Table 12.1.6.1). One would expect low levels of anorexigenic hormones and high levels of orexigenic hormones in individuals with low weight, and in some cases, this is what is seen in girls and women with anorexia nervosa. For example, serum leptin levels are relatively suppressed, as would be predicted from low fat mass and anorexigenic effects of leptin [31, 32]. Moreover, as one would expect in a low-weight population, mean systemic levels of ghrelin, an orexigenic hormone, are elevated in girls and women with anorexia nervosa [33–35]. However, there are some contradictory studies, including one which demonstrated low mean ghrelin levels in women with binge-purge-type anorexia nervosa [36] and another which reported a greater reduction in ghrelin levels during a euglycemic hyperinsulinaemic clamp than controls [37]. The authors hypothesized that this exaggerated response contributes to a feeling of satiety in such women and as such could be contributing to the maintenance of the disorder.

Moreover, contrary to what would be predicted in a low-weight, low-caloric-intake state, mean systemic levels of peptide YY (PYY), an anorexigenic gut hormone, is high both in girls and women with anorexia nervosa compared with controls [38, 39]. The cause and result of this elevation are unknown and raise the possibility that NPY-PYY appetite system dysregulation may contribute to the pathophysiology of anorexia nervosa. In support of this unproven hypothesis, PYY levels correlate with the degree of eating disorder thinking and behaviour [40], suggesting a possible relationship between the psychologic features of anorexia nervosa and PYY levels.

Bone Loss

The global endocrine dysregulation delineated earlier contributes to severe bone loss in women with anorexia nervosa—both those with classic low-weight anorexia nervosa and those with normal-weight atypical anorexia nervosa (Table 12.1.6.2). The vast majority of women with low-weight anorexia nervosa have Z-scores and or T-scores of less than or equal to –1.0, and 40% have osteoporosis [41]. In addition, BMD in such women declines by a mean 2.5% annually [42]. Bone loss occurs disproportionally at primarily trabecular sites (e.g. lateral and posteroanterior (PA) spine), and only 15% of women with anorexia nervosa have normal BMD at all skeletal sites [6]. This results in a non-spinal fracture rate seven times expected [43], and 30% of such women report a history of at least one fracture [6]. Healthy adolescence is characterized by a substantial increase in BMD and an increase in markers of bone metabolism, which is not experienced by girls with anorexia nervosa [44],

Table 12.1.6.2 Therapies for low bone mineral density in adolescent girls and women with anorexia nervosa

Therapy	Age	*n*	Change in BMD	Treatment length (months)	Refs
Oral oestrogen/progestin	16–42	48	No change versus placebo	19 (mean)	[55]
	12–21	50	No change versus placebo	12	[56]
	18–38	60	No change versus placebo	9	[57]
	11–17	112	No change versus placebo	12	[58]
Physiologic oestrogen replacement	12–18	110	Increase in spine BMD versus placebo (2.6% versus 0.3%). Stable hip BMD versus decline in placebo (0.004% versus −1.2%)	18	[59]
Recombinant human IGF-1	18–38	60	Increase in spine BMD versus placebo (1.1% versus −0.6%)	9	[57]
Recombinant human IGF-1 plus oral contraceptive	18–38	60	Increase in spine BMD versus placebo (1.8% versus −1.0%)	9	[57]
Testosterone	19–49	77	No change versus placebo	12	[21]
DHEA	14–28	61	No change in lumbar BMD in either DHEA or oral oestrogen/progestin group. Similar increase in hip BMD (1.7%) in both DHEA and oral oestrogen/progestin group, but change was not significant after controlling for weight gain	12	[20]
DHEA plus oral contraceptive	13–27	94	Maintenance of spine and hip BMD versus decrease in placebo group	18	[61]
Oral bisphosphonates	12–21	32	No change versus placebo, although both groups gained weight and increased spine and hip BMD	12	[69]
	19–49	77	3–4% increase in spine BMD and 2% increase in hip BMD versus placebo	12	[21]
Teriparatide	32–58	32	Increase in spine BMD versus placebo (6% versus 0.6%). No change in hip BMD versus placebo	6	[70]
Weight gain and resumption of menses	18–40	75	1.8% increase in hip BMD, 3.1% increase in spine BMD compared with baseline	12	[43]

DHEA, Dehydroepiandrosterone; hIGF1, human insulin-like growth factor 1.

Adapted with permission from Schorr M, and Miller KK. The endocrine manifestations of anorexia nervosa: mechanisms and management. *Nat Rev Endocrinol*. 2017;13(3):174–86.

resulting in low BMD compared to healthy girls [45] and an elevated fracture risk [46]. In one study 70% of amenorrhoeic, 67% of eumenorrheic, and 65% of women on oral contraceptive pills with atypical anorexia nervosa (psychiatric symptoms characteristic of anorexia nervosa in the absence of low weight) had a BMD Z score <−1 at least one skeletal site [46]. Prior low weight and/or amenorrhoea conferred a high risk for low BMD, whereas having a history of neither appeared to be relatively protective [47].

Bone microarchitecture and strength (as modelled by finite element analysis) are also adversely affected in adolescent girls and women with anorexia nervosa [48–52], before dual energy X-ray absorptiometry (DXA) shows evidence of low BMD [51]. Recovery of bone mass does occur with weight recovery, particularly in adolescents in whom intervention is early and adults who recover menses with their weight gain [42]. However, nutritional restoration is not easy to achieve for many girls and women suffering from this psychiatric disorder, and even when it does occur, persistent low BMD is common [53]. In addition, weight recovery may take years to achieve. These factors provide a rationale for early intervention—to restore weight and normal neuroendocrine function, as well as to develop therapeutic strategies to prevent and treat skeletal compromise.

Bone metabolism in women with anorexia nervosa is characterized by an uncoupling of bone formation (which is low) and bone resorption (which is high) [31], which provides clues to the pathophysiology of the disease and a rationale for putative treatment strategies. In adolescent girls with anorexia nervosa, the characteristic increase in bone turnover (bone formation and resorption) observed in adolescence—and leading to rapid bone accrual toward peak BMD—is not present [54]. These differences in bone turnover in girls versus adults with anorexia nervosa highlight very real differences in skeletal physiology in these groups that inform preventative and therapeutic choices (i.e. findings in one age group cannot be generalized to the other).

Given the well-known link between oestrogen-deficiency and bone loss during menopause, hypogonadism in anorexia nervosa is an excellent candidate for an important contributing factor to bone loss in girls and women with anorexia nervosa—and an obvious therapeutic target. In fact, in amenorrhoeic low-weight women with anorexia nervosa, the mean spine T-score in amenorrhoea group was −1.9 compared to −0.9 in women with comparable low weight but without amenorrhoea [15]. This suggests that gonadal steroids are important for skeletal integrity in the setting of low weight and that nutritional deprivation plus gonadal steroid deficiency is particularly detrimental to bone. Whether there is a component of relative androgen and/or progesterone deficiency or whether it is simply oestrogen-deficiency that contributes to bone loss is unknown. In adolescent girls and women with anorexia nervosa, oral replacement strategies that involve oestrogens and progestins (both postmenopausal hormone therapy [55] and oral contraceptives) [56–58] have demonstrated unimpressive results. The former appears to be entirely ineffective [55], whereas the latter may stem

bone loss but does not increase BMD [57]. This is contrast to the effects of transdermal oestrogen and progestin at replacement doses (100 μg of 17β-oestradiol with cyclic progesterone transdermally in older adolescents (with a bone age ≥15 years), which was shown in a randomized, placebo-controlled study to result in significant increases in BMD, though it did not result in normalization of BMD [59]. Low-dose testosterone administration stimulates bone formation in women with anorexia nervosa [21, 60], but does not increase BMD [21]. Dehydroepiandrosterone (DHEA) administration to adolescent girls with anorexia nervosa resulted in increases in BMD that were explained by increases in weight [20], but a subsequent study of DHEA plus ethinyl oestradiol plus a progestin demonstrated spine and total body BMD stability compared to placebo, which was associated with bone loss [61].

As might be expected, duration of illness (time since onset of anorexia nervosa) and body mass index are inversely associated with BMD. Muscle mass (as measured in most studies as lean body mass by DXA) is the body composition parameter that most strongly correlates with BMD [44]. Weaker associations are seen with total fat mass, but bone marrow fat, which is high in anorexia nervosa, is strongly inversely associated with BMD [62]. Whether this provides a clue to the mechanisms underlying bone loss in anorexia nervosa is unknown. It is possible that the increase in bone marrow fat is mechanistically related with endocrine dysregulation resulting in mesenchymal stem cell precursor differentiation toward the adipocyte, rather than osteoblast pathway. Alternatively, impaired osteoblastogenesis may be the cause, not result, of accumulation of bone marrow fat.

Although components of nutritional intake have not been implicated in bone loss in anorexia nervosa, there is controversy about the impact of vitamin D deficiency. In studies in both girls and women with anorexia nervosa, mean vitamin D and calcium intake have been shown to comparable to healthy controls of comparable age (in part due to the ingestion of supplements) and low vitamin D levels have even been shown to be more common in healthy controls. The conclusions drawn from these studies are that vitamin D deficiency is unlikely to be an adequate explanation for the lower BMD observed in anorexia nervosa [63, 64]. However, when vitamin D deficiency is present, it may contribute to bone loss. In one study, a majority of females with anorexia nervosa had 25-hydroxy-vitamin D levels, and very low levels (<12 ng/ml) were associated with lower hip BMD than those with vitamin D levels ≥ 20 ng/ml [65]. Moreover, in another study, improvements in spine BMD with weight restoration occurred only in individuals who had vitamin D levels ≥ 30 ng/ml This suggests that it is important to ensure normal vitamin D levels to optimize improvements in bone health with weight gain or pharmacologic therapies [66].

GH is well-known to be anabolic to bone, primarily responsible for linear growth and important for maintenance of bone mass in adults. However much of GH's bone-anabolic actions are mediated through IGF-1, systemic levels of which are reduced in anorexia nervosa due to the state of GH resistance. Evidence of a detrimental effect of relative IGF-1 deficiency on BMD in anorexia nervosa include an association of IGF-1 levels with markers of bone formation [44] and bone microarchitectural parameters [49]. Recombinant human IGF-1 (rhIGF-1) replacement therapy is effective at increasing BMD in women with anorexia nervosa but not to normal (1.4% at the spine—2.8% when combined with oral contraceptives) over 9 months [57]. Of note, rhIGF-1 is not FDA-approved to treat low BMD in women with anorexia nervosa, and studies in adolescents with anorexia nervosa are lacking.

Bone loss is a well-known complication of hypercortisolaemia in women and men with Cushing's syndrome. Whether hypercortisolaemia is driven in part by relatively higher ghrelin levels is unknown. Moreover, whether the functional hypercortisolaemia observed in anorexia nervosa contributes to the bone loss observed in this syndrome has not been fully determined. However, there is evidence of such an effect from studies. For example, overnight frequent samples for cortisol negatively correlate with BMD in both women and girls with anorexia nervosa [29, 67]. It is unknown how to translate these observations into a therapeutic strategy other than to work toward nutritional and psychiatric recovery, which result in normalization of cortisol levels.

There is also a complex interplay between bone loss and appetite hormone dysregulation. Most studies suggest indirect effects. For example, although there is little evidence that hypoleptinaemia directly causes bone loss, normalization of leptin may improve BMD through normalization of reproductive function. In contrast, PYY levels are elevated in women and girls with anorexia nervosa and strongly inversely associated with BMD, even after controlling for weight [68], raising the question of whether there may be direct effects. However, no studies establishing causality have been published.

Other traditional pharmacologic and non-pharmacologic interventions developed to treat postmenopausal osteoporosis have also been tested in girls and women with anorexia nervosa. Antiresorptive and anabolic agents have been studied in small, randomized, placebo-controlled studies. Risedronate increases BMD in women with anorexia nervosa, but one year of therapy does not normalize BMD [21]. In contrast, alendronate was not found to be effective at increasing BMD in adolescent girls with anorexia nervosa after controlling for weight gain [69]. This is yet another example of the non-generalizability of results in adults and adolescents. It is important to note that bisphosphonates are not FDA-approved for use in girls or women with anorexia nervosa, and although the limited data available are reassuring about teratogenicity of bisphosphonates, they should be avoided in individuals at risk of becoming pregnant in the near future. Anabolic therapies have particular appeal in anorexia nervosa because of the decrement in bone formation observed. A pilot study demonstrated increases in BMD at the spine 6–10% with 6 months of teriparatide as compared with placebo [70]. Although these data are very promising, teriparatide is not FDA-approved for the treatment of low BMD in anorexia nervosa, and an FDA warning states that teriparatide should not be used in patients at risk for the development of osteosarcoma, including adolescents. Finally, although weight-bearing exercise appears to improve BMD in women who are have recovered from anorexia nervosa [71], exercise is deleterious for skeletal health in low-weight women with anorexia nervosa [72], likely because it exacerbates the negative energy balance.

Conclusions

Endocrine complications of chronic starvation, as seen in girls and women with anorexia nervosa, are myriad and include SIAD,

hypothalamic amenorrhoea (resulting in relative oestrogen, androgen, and progesterone deficiency), GH resistance (with resultant relative IGF-1 deficiency) and hypercortisolaemia, as well as alterations in anorexigenic and orexigenic hormone levels and pathways. These endocrine consequences of starvation contribute to bone loss, which is nearly universal in girls and women with anorexia nervosa. A multidisciplinary approach to psychiatric and nutritional rehabilitation, with a physician, therapist, and nutritionist, is the key to recovery from this psychiatric disease. The combination of weight restoration and psychiatric recovery results in normalization of endocrine axis dysregulation in many, but not all, patients. Recovery of skeletal health often accompanies weight recovery, especially if it is accompanied by recovery of reproductive function. However, some degree of BMD impairment may persist after recovery from anorexia nervosa, and the disorder is chronic in many patients. Therefore, further research into preventative and treatment strategies for bone loss in girls and women with anorexia nervosa is warranted.

REFERENCES

1. Gull WW. Address in medicine delivered before the Annual Meeting of the British Medical Association at Oxford. *Lancet*, 1868: **2**: 171–6.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Washington, DC: American Psychiatric Publishing; 2013.
3. Hoek HW, van Hoeken D. Review of the prevalence and incidence of eating disorders. *Int J Eat Disord*, 2003; **34**: 383–96.
4. Lucas AR, Beard CM, O'Fallon WM, Kurland LT. 50-year trends in the incidence of anorexia nervosa in Rochester, Minn: a population-based study. *Am J Psychiatry*, 1991; **148**: 917–22.
5. Fairburn CG, Harrison PJ. Eating disorders. *Lancet*, 2003; **361**: 407–16.
6. Miller KK, Grinspoon SK, Ciampa J, Hier J, Herzog D, Klibanski A. Medical findings in outpatients with anorexia nervosa. *Arch Intern Med*, 2005; **165**: 561–6.
7. Miller KK. Endocrine effects of anorexia nervosa. *Endocrinol Metab Clin North Am*, 2013; **42**: 515–28.
8. Boyar RM, Hellman LD, Roffwarg H, *et al.* Cortisol secretion and metabolism in anorexia nervosa. *N Engl J Med*, 1977; **296**: 190–3.
9. Frisch RE, McArthur JW. Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance or onset. *Science*, 1974; **185**: 949–51.
10. Boyar RM, Katz J, Finkelstein JW, *et al.* Anorexia nervosa. Immaturity of the 24-hour luteinizing hormone secretory pattern. *N Engl J Med*, 1974; **291**: 861–5.
11. Katz JL, Boyar R, Roffwarg H, Hellman L, Weiner H. Weight and circadian luteinizing hormone secretory pattern in anorexia nervosa. *Psychosom Med*, 1978; **40**: 549–67.
12. Jocoangeli F, Masala S, Staar Mezzasalma F, *et al.* Amenorrhea after weight recovery in anorexia nervosa: role of body composition and endocrine abnormalities. *Eat Weight Disord*, 2006; **11**: e20–6.
13. Strobel A, Issad T, Camoin L, Ozata M, Strosberg AD. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat Genet*, 1998; **18**: 213–15.
14. Ahima RS, Prabakaran D, Mantzoros C, Qu D, Lowell B, Maratos-Flier E, Flier JS. Role of leptin in the neuroendocrine response to fasting. *Nature*, 1996; **382**: 250–2.
15. Miller KK, Grinspoon S, Gleysteen S, *et al.* Preservation of neuroendocrine control of reproductive function despite severe undernutrition. *J Clin Endocrinol Metab*, 2004; **89**: 4434–8.
16. Welt CK, Chan JL, Bullen J, *et al.* Recombinant human leptin in women with hypothalamic amenorrhea. *N Engl J Med*, 2004; **351**: 987–97.
17. Jayasena CN, Nijher GM, Chaudhri OB, *et al.* Subcutaneous injection of kisspeptin-54 acutely stimulates gonadotropin secretion in women with hypothalamic amenorrhea, but chronic administration causes tachyphylaxis. *J Clin Endocrinol Metab*, 2009; **94**: 4315–23.
18. Miller KK, Lawson EA, Mathur V, *et al.* Androgens in women with anorexia nervosa and normal-weight women with hypothalamic amenorrhea. *J Clin Endocrinol Metab*, 2007; **92**: 1334–9.
19. Lawson EA, Misra M, Meenaghan E, *et al.* Adrenal glucocorticoid and androgen precursor dissociation in anorexia nervosa. *J Clin Endocrinol Metab*, 2009; **94**: 1367–71.
20. Gordon CM, Grace E, Emans SJ, *et al.* Effects of oral dehydroepiandrosterone on bone density in young women with anorexia nervosa: a randomized trial. *J Clin Endocrinol Metab*, 2002; **87**: 4935–41.
21. Miller KK, Meenaghan E, Lawson EA, *et al.* Effects of risedronate and low-dose transdermal testosterone on bone mineral density in women with anorexia nervosa: a randomized, placebo-controlled study. *J Clin Endocrinol Metab*, 2011; **96**: 2081–8.
22. Miller KK, Wexler TL, Zha AM, *et al.* Androgen deficiency: association with increased anxiety and depression symptom severity in anorexia nervosa. *J Clin Psychiatry*, 2007; **68**: 959–65.
23. Kimball A, Schorr M, Meenaghan E, et al. A randomized placebo-controlled trial of low-dose testosterone therapy in women with anorexia nervosa. *J Clin Endocrinol Metab*. 2019; **104**(10): 4347–4355.
24. Roth J, Glick SM, Yalow RS, Bersonsa. Hypoglycemia: a potent stimulus to secretion of growth hormone. *Science*, 1963; **140**: 987–8.
25. Stoving RK, Veldhuis JD, Flyvbjerg A, *et al.* Jointly amplified basal and pulsatile growth hormone (GH) secretion and increased process irregularity in women with anorexia nervosa: indirect evidence for disruption of feedback regulation within the GH-insulin-like growth factor I axis. *J Clin Endocrinol Metab*, 1999; **84**: 2056–63.
26. Stoving RK, Chen JW, Glintborg D, *et al.* Bioactive insulin-like growth factor (IGF) I and IGF-binding protein-1 in anorexia nervosa. *J Clin Endocrinol Metab*, 2007; **92**: 2323–9.
27. Fazeli PK, Lawson EA, Prabhakaran R, *et al.* Effects of recombinant human growth hormone in anorexia nervosa: a randomized, placebo-controlled study. *J Clin Endocrinol Metab*, 2010; **95**: 4889–97.
28. Grinspoon S, Thomas L, Miller K, Pitts S, Herzog D, Klibanski A. Changes in regional fat redistribution and the effects of estrogen during spontaneous weight gain in women with anorexia nervosa. *Am J Clin Nutr*, 2001; **73**: 865–9.
29. Lawson EA, Donoho D, Miller KK, *et al.* Hypercortisolemia is associated with severity of bone loss and depression in hypothalamic amenorrhea and anorexia nervosa. *J Clin Endocrinol Metab*, 2009; **94**: 4710–16.

30. Misra M, Prabhakaran R, Miller KK, *et al.* Role of cortisol in menstrual recovery in adolescent girls with anorexia nervosa. *Pediatr Res*, 2006; **59**(4 Pt 1): 598–603.
31. Grinspoon S, Baum H, Lee K, Anderson E, Herzog D, and Klibanski A. Effects of short-term recombinant human insulin-like growth factor I administration on bone turnover in osteopenic women with anorexia nervosa. *J Clin Endocrinol Metab*, 1996; **81**: 3864–70.
32. Hebebrand J, Blum WF, Barth N, *et al.* Leptin levels in patients with anorexia nervosa are reduced in the acute stage and elevated upon short-term weight restoration. *Mol Psychiatry*, 1997; **2**: 330–4.
33. Misra M, Miller KK, Kuo K, *et al.* Secretory dynamics of ghrelin in adolescent girls with anorexia nervosa and healthy adolescents. *Am J Physiol Endocrinol Metab*, 2005; **289**: E347–56.
34. Monteleone P, Serritella C, Martiadis V, Scognamiglio P, Maj M. Plasma obestatin, ghrelin, and ghrelin/obestatin ratio are increased in underweight patients with anorexia nervosa but not in symptomatic patients with bulimia nervosa. *J Clin Endocrinol Metab*, 2008; **93**: 4418–21.
35. Usdan LS, Khaodhiar L, Apovian CM. The endocrinopathies of anorexia nervosa. *Endocr Pract*, 2008; **14**: 1055–63.
36. Germain N, Galusca B, Grouselle D, *et al.* Ghrelin and obestatin circadian levels differentiate bingeing-purging from restrictive anorexia nervosa. *J Clin Endocrinol Metab*, 2010; **95**: 3057–62.
37. Karczewska-Kupczewska M, Straczkowski M, Adamska A, *et al.* Increased suppression of serum ghrelin concentration by hyperinsulinemia in women with anorexia nervosa. *Eur J Endocrinol*, 2010; **162**: 235–9.
38. Misra M, Miller KK, Tsai P, *et al.* Elevated peptide YY levels in adolescent girls with anorexia nervosa. *J Clin Endocrinol Metab*, 2006; **91**: 1027–33.
39. Pfluger PT, Kampe J, Castaneda TR, *et al.* Effect of human body weight changes on circulating levels of peptide YY and peptide YY3-36. *J Clin Endocrinol Metab*, 2007; **92**: 583–8.
40. Lawson EA, Eddy KT, Donoho D, *et al.* Appetite-regulating hormones cortisol and peptide YY are associated with disordered eating psychopathology, independent of body mass index. *Eur J Endocrinol*, 2011; **164**: 253–61.
41. Grinspoon S, Thomas E, Pitts S, *et al.* Prevalence and predictive factors for regional osteopenia in women with anorexia nervosa. *Ann Intern Med*, 2000; **133**: 790–4.
42. Miller KK, Lee EE, Lawson EA, *et al.* Determinants of skeletal loss and recovery in anorexia nervosa. *J Clin Endocrinol Metab*, 2006; **91**: 2931–7.
43. Rigotti NA, Neer RM, Skates SJ, Herzog DB, Nussbaum SR. The clinical course of osteoporosis in anorexia nervosa. A longitudinal study of cortical bone mass. *JAMA*, 1991; **265**: 1133–8.
44. Soyka LA, Misra M, Frenchman A, Miller KK, Grinspoon S, Schoenfeld DA, Klibanski A. Abnormal bone mineral accrual in adolescent girls with anorexia nervosa. *J Clin Endocrinol Metab*, 2002; **87**: 4177–85.
45. Misra M, Aggarwal A, Miller KK, *et al.* Effects of anorexia nervosa on clinical, hematologic, biochemical, and bone density parameters in community-dwelling adolescent girls. *Pediatrics*, 2004; **114**: 1574–83.
46. Faje AT, Fazeli PK, Miller KK, *et al.* Fracture risk and areal bone mineral density in adolescent females with anorexia nervosa. *Int J Eat Disord*, 2014; **47**: 458–66.
47. Schorr M, Thomas JJ, Eddy KT, *et al.* Bone density, body composition, and psychopathology of anorexia nervosa spectrum disorders in DSM-IV vs DSM-5. *Int J Eat Disord*, 2017; **50**: 343–51.
48. Schorr M, Miller KK. The endocrine manifestations of anorexia nervosa: mechanisms and management. *Nat Rev Endocrinol*, 2017; **13**: 174–86.
49. Lawson EA, Miller KK, Bredella MA, *et al.* Hormone predictors of abnormal bone microarchitecture in women with anorexia nervosa. *Bone*, 2010; **46**: 458–63.
50. Faje AT, Karim L, Taylor A, *et al.* Adolescent girls with anorexia nervosa have impaired cortical and trabecular microarchitecture and lower estimated bone strength at the distal radius. *J Clin Endocrinol Metab*, 2013; **98**: 1923–9.
51. Bredella MA, Misra M, Miller KK, *et al.* Distal radius in adolescent girls with anorexia nervosa: trabecular structure analysis with high-resolution flat-panel volume CT. *Radiology*, 2008; **249**: 938–46.
52. Walsh CJ, Phan CM, Misra M, *et al.* Women with anorexia nervosa: finite element and trabecular structure analysis by using flat-panel volume CT. *Radiology*, 2010; **257**: 167–74.
53. Herzog W, Minne H, Deter C, *et al.* Outcome of bone mineral density in anorexia nervosa patients 11.7 years after first admission. *J Bone Miner Res*, 1993; **8**: 597–605.
54. Misra M, Soyka LA, Miller KK, *et al.* Serum osteoprotegerin in adolescent girls with anorexia nervosa. *J Clin Endocrinol Metab*, 2003; **88**: 3816–22.
55. Klibanski A, Biller BM, Schoenfeld DA, Herzog DB, Saxe VC. The effects of estrogen administration on trabecular bone loss in young women with anorexia nervosa. *J Clin Endocrinol Metab*, 1995; **80**: 898–904.
56. Golden NH, Lanzkowsky L, Schebendach J, Palestro CJ, Jacobson MS, Shenker IR. The effect of estrogen-progestin treatment on bone mineral density in anorexia nervosa. *J Pediatr Adolesc Gynecol*, 2002; **15**: 135–43.
57. Grinspoon S, Thomas L, Miller K, Herzog D, Klibanski A. Effects of recombinant human IGF-I and oral contraceptive administration on bone density in anorexia nervosa. *J Clin Endocrinol Metab*, 2002; **87**: 2883–91.
58. Strokosch GR, Friedman AJ, Wu SC, Kamin M. Effects of an oral contraceptive (norgestimate/ethinyl estradiol) on bone mineral density in adolescent females with anorexia nervosa: a double-blind, placebo-controlled study. *J Adolesc Health*, 2006; **39**: 819–27.
59. Misra M, Katzman D, Miller KK, *et al.* Physiologic estrogen replacement increases bone density in adolescent girls with anorexia nervosa. *J Bone Miner Res*, 2011; **26**: 2430–8.
60. Miller KK, Grieco KA, Klibanski A. Testosterone administration in women with anorexia nervosa. *J Clin Endocrinol Metab*, 2005; **90**: 1428–33.
61. Divasta AD, Feldman HA, Giancaterino C, Rosen CJ, Leboff MS, Gordon CM. The effect of gonadal and adrenal steroid therapy on skeletal health in adolescents and young women with anorexia nervosa. *Metab Clin Exp*, 2012; **61**: 1010–20.
62. Bredella MA, Fazeli PK, Miller KK, *et al.* Increased bone marrow fat in anorexia nervosa. *J Clin Endocrinol Metab*, 2009; **94**: 2129–36.
63. Hadigan CM, Anderson EJ, Miller KK, *et al.* Assessment of macronutrient and micronutrient intake in women with anorexia nervosa. *Int J Eat Disord*, 2000; **28**: 284–92.
64. Misra M, Tsai P, Anderson EJ, *et al.* Nutrient intake in community-dwelling adolescent girls with anorexia nervosa and in healthy adolescents. *Am J Clin Nutr*, 2006; **84**: 698–706.
65. Gatti D, El Ghoch M, Viapiana O, *et al.* Strong relationship between vitamin D status and bone mineral density in anorexia nervosa. *Bone*, 2015; **78**: 212–15.

66. Giollo A, Idolazzi L, Caimmi C, *et al.* Vitamin D levels strongly influence bone mineral density and bone turnover markers during weight gain in female patients with anorexia nervosa. *Int J Eat Disord*, 2017; **50**: 1041–9.
67. Misra M, Miller KK, Almazan C, *et al.* Alterations in cortisol secretory dynamics in adolescent girls with anorexia nervosa and effects on bone metabolism. *J Clin Endocrinol Metab*, 2004; **89**: 4972–80.
68. Utz AL, Lawson EA, Misra M, *et al.* Peptide YY (PYY) levels and bone mineral density (BMD) in women with anorexia nervosa. *Bone*, 2008; **43**: 135–9.
69. Golden NH, Iglesias EA, Jacobson MS, *et al.* Alendronate for the treatment of osteopenia in anorexia nervosa: a randomized, double-blind, placebo-controlled trial. *J Clin Endocrinol Metab*, 2005; **90**: 3179–85.
70. Fazeli PK, Wang IS, Miller KK, *et al.* Teriparatide increases bone formation and bone mineral density in adult women with anorexia nervosa. *J Clin Endocrinol Metab*, 2014; **99**: 1322–9.
70. Waugh EJ, Woodside DB, Beaton DE, Cote P, Hawker GA. Effects of exercise on bone mass in young women with anorexia nervosa. *Med Sci Sports Exerc*, 2011; **43**: 755–63.

12.2

Endocrine Complications of Substance Misuse

12.2.1 Endocrinology and Alcohol

Marc Walter and Margit G. Proescholdt

Introduction

Alcohol has widespread effects on the endocrine system, potentially impairing endocrine function. Endocrine impairment is more obvious and frequent in those who have developed alcoholic liver disease. Endocrine dysfunction not only results from hepatic failure or chronic malnutrition, but also from direct, toxic actions of alcohol on the endocrine organs as well as impairing regulation at the hypothalamic–pituitary level. In the absence of liver disease, the adverse effects of alcohol on the endocrine organs are mostly reversible upon reduction or cessation of alcohol consumption. There is increasing evidence that the endocrine system itself may play a crucial role in the pathogenesis of addictive behaviour.

Studies on alcohol–endocrine interactions in humans are difficult and often report inconclusive or controversial results. Difficulties arise from the highly heterogenic groups of alcohol-dependent persons differing in amount and duration of alcohol consumption, periods of abstinence, confounding variables (e.g. age, gender, nutritional status, smoking, concomitant drugs) and comorbidities (e.g. liver disease), respectively. The mechanisms of alcohol–endocrine interactions are poorly understood, due to multifactorial causes of disease and the complexity of endocrine regulation. However, a better understanding of the impact of alcohol on the endocrine system is required to guide public health recommendations on alcohol consumption.

Hypothalamic–Pituitary–Adrenal (HPA) Axis and Alcohol-Induced Hypercortisolism

Excessive alcohol intake increases cortisol secretion acutely and chronically and is primarily mediated through the activation of hypothalamic corticotropin-releasing hormone (CRH). In addition, hepatic dysfunction may result in an impaired peripheral clearance of cortisol and thus contribute to the hypercortisolaemic state in chronic alcoholism. As a consequence, some patients with chronic alcohol consumption may develop a physiologic, non-neoplastic hypercortisolism (formerly pseudo-Cushing's syndrome) with clinical and endocrine features that may be indistinguishable from true Cushing's disease (pathologic, neoplastic hypercortisolism), which is usually caused by either an adrenocorticotrophic hormone (ACTH)-secreting neoplasm or adrenal neoplastic disease.

Due to the similar clinical and endocrine features, differentiation of alcohol-induced hypercortisolism from true Cushing's syndrome may be difficult; therefore, a thorough history and clinical examination are often considered best. Possible endocrine findings in alcohol-induced hypercortisolism are described in **Box 12.2.1.1**. Importantly, both the physical and the endocrine abnormalities improve after discontinuation of alcohol use. Thus, abstinence from alcohol not only is curative but also an important diagnostic tool [1].

Box 12.2.1.1 Alcohol-induced hypercortisolism (formerly pseudo-Cushing's syndrome)

May be indistinguishable from true Cushing's syndrome
History and physical examination best way to differentiate from true Cushing's syndrome
Endocrine abnormalities resolve with discontinuation of alcohol use

Endocrine findings

- ↑ late-night salivary cortisol and urinary corticosteroids
- Mostly normal or ↑ plasma ACTH
- Rarely adrenal nodular disease and subnormal plasma ACTH
- Often abnormal overnight dexamethasone suppression test
- Abnormal dexamethasone-CRH test during active alcohol consumption
- Likelihood of no ACTH response to DDAVP (in contrast to true Cushing's syndrome)

↑: increased.
DDAVP, desmopressin acetate; ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone.

Hypothalamic–Pituitary–Gonadal (HPG) Axis and Alcohol

HPG Axis and Alcohol in Men

Chronic alcohol abuse may affect all levels of the hypothalamic–pituitary–gonadal (HPG) axis resulting in hypogonadism. The full spectrum of hypogonadism (testicular atrophy, loss of libido, impotence, infertility), osteoporosis, decreased muscle, anaemia, altered immune function, depressed mood, and decreased vitality, may occur [2, 3].

Alcohol-induced (AI) hypogonadism arises from direct (toxic) and indirect effects on the gonads and on the hypothalamic–pituitary unit [3]. AI hypogonadism is more frequently observed in advanced alcoholic liver disease [4]. The prevalence of sexual disorders ranges from 8 to 58%. Abnormal sperm parameters (morphology, semen volume), and altered sex hormones are frequent findings [4, 5] (Box 12.2.1.2). In a recent study of young Danish men, reduced semen quality occurred with only 5 drinks/week, and was most pronounced in heavy drinkers (>25 drinks/week) [6].

Box 12.2.1.2 Alcohol-induced gonadal dysfunction in men

Sexual disorders
Altered sperm parameters
Feminization in advanced liver disease

Endocrine findings

- Acute alcohol exposure
  - LH, testosterone, and oestradiol ↓
- Chronic alcohol exposure
  - Testosterone*, free testosterone, and FAI → or ↓
  - LH, SHBG, oestradiol, and androstenedione → or ↑
- Abstinence
  - Testosterone and free testosterone → or ↑, LH ↑

↑: increased; →: normal; ↓: decreased
*: ↓ testosterone mostly seen in alcoholic liver disease
LH, luteinizing hormone; SHBG, sex hormone-binding globulin; FAI, free androgen index (testosterone/SHBG ×100).

Sex hormones levels are variable [4, 7–11] (Box 12.2.1.2). Blood testosterone levels are mostly normal in the absence of liver disease. By contrast, levels of luteinizing hormone (LH) are frequently inappropriately high or low indicating disturbance of adenohypophyseal feedback [7]. In the absence of significant hepatic or gonadal injury, abstinence may improve gonadal dysfunction [5, 12].

Feminization, as distinct from hypogonadism is manifested by gynaecomastia, female body habitus changes, spider angioma, palmar erythema, and changes in body hair patterns. Feminization is seen rarely in the absence of chronic alcoholic liver disease. Feminization results from the combined effects of altered enterohepatic circulation of biliary excreted steroids from portal hypertension, and increased conversion of testosterone to androstenedione, oestradiol, and oestrone [13].

A clinical evaluation of hypogonadism should always include a thorough history of alcohol consumption, information on the negative effects on the male reproductive system and the potential reversibility from abstinence.

HPG Axis and Alcohol in Women

In **premenopausal women**, chronic heavy alcohol use may contribute to a multitude of reproductive disorders, such as menstrual cycle irregularities, amenorrhoea, anovulation, loss of libido, early menopause, and increased risk of spontaneous abortions [14]. During early adolescence, alcohol consumption may delay puberty and adversely affect the maturation of the female reproductive system [15]. Binge drinking may be most detrimental at certain times, namely puberty. It can impair the cyclical selection of follicles for maturation, ovulation, implantation, and subsequent survival of the blastocyst [16].

Studies investigating effects on sex hormones are challenging because of the variation in hormone levels throughout the menstrual cycle. A recent large study observed higher luteal oestrogen and sex hormone-binding globulin (SHBG) concentrations and lower levels of free testosterone among women who consumed alcohol compared with non-drinkers. In this study, there were no significant associations between alcohol consumption and plasma androgen, oestrone, or oestradiol levels in the follicular phase [17]. However, the role of AI alterations of sex hormones on the female reproductive system is still unclear. Further studies are needed to clarify the relationships and underlying mechanisms.

In **postmenopausal women**, plasma oestradiol levels are elevated in alcoholic liver disease as well as in 'moderate' drinkers (1 to 28 drinks/week) in the absence of liver disease. However, so far, a firm relationship between 'moderate' alcohol consumption (no more than 1 drink/day) and oestrogen levels has not been established. By contrast, single dose alcohol administration to postmenopausal women on oestrogen replacement therapy (ERT) increases serum oestradiol levels markedly, which may increase the risk of breast cancer. Following a single oral dose of ethanol (0.75 g/kg bodyweight), serum oestradiol levels increased by 22% when oestradiol was administered by transdermal patch (0.15 mg), and by 300% by oral administration (1 mg/day), respectively [18]. Notably, ERT has been implicated in breast cancer and the risk of breast cancer is increased in postmenopausal women with concomitant alcohol use [19]. Therefore, postmenopausal women on ERT with concomitant

alcohol use should be informed about the increased risk for breast cancer and motivated to reduce alcohol intake.

Alcohol and Breast Cancer

Alcohol consumption is a risk factor for breast cancer in women. However, the underlying pathophysiological mechanisms are unclear. Recent hypotheses include perturbation of oestrogen metabolism and response, induction of mutagenesis by carcinogenic products of alcohol metabolism (i.e. acetaldehyde), and stimulation of oxidative damage through ethanol metabolism [20].

A causal association has been established between alcohol consumption and cancers of the oral cavity, pharynx, larynx, oesophagus, liver, colon, rectum, and breast. This association is highest for breast cancer than for any other type of cancer [21]. There is a 7–10% risk increase for each 10 g alcohol consumed per day (premenopausal and postmenopausal women). It has been suggested that breast tissue is more vulnerable to the carcinogenic effects of alcohol compared to other organs [20]. A recent meta-analysis has reported an odds ratio for breast cancer of 1.085 (95% confidence interval [CI] 1.015–1.160) for low-level (<21 g/d) drinkers, 1.374 (95% CI 1.319–1.431) for hazardous-level drinkers, and 1.336 (95% CI 1.228–1.454) for harmful-level (>40 g/d) drinkers [22]. There is evidence that early-life exposure to alcohol may contribute to the lifetime risk of breast cancer, implying a need for increased awareness and counselling regarding alcohol use among adolescent girls and young women [20].

Hypothalamic–Pituitary–Thyroid (HPT) Axis and Alcohol

Thyroid dysfunction is a frequent finding in alcohol-dependent patients. The mechanisms include direct toxic effects of alcohol and its metabolism on the thyroid gland, as well as central effects on the hypothalamus and/or pituitary gland [23]. AI thyroid dysfunction usually normalizes during abstinence.

The most consistent clinical and endocrine findings are described in Box 12.2.1.3 [14, 24]. In individuals where thyroid dysfunction persists into abstinence, other non-alcohol-related thyroid diseases should be excluded (e.g. autoimmune thyroid disease). In patients with pre-existing hyperthyroidism, acute alcohol intoxication may mimic a thyrotoxic crisis warranting immediate analysis of thyroid hormones and adequate medical treatment.

Box 12.2.1.3 Alcohol-induced thyroid dysfunction

Usually absence of overt clinical signs of hypothyroidism
Thyroid volume ↓, thyroid fibrosis ↑

Endocrine findings

- Usually normal basal TSH
- T_4 → or ↓, T_3 and fT_3 ↓
- Deiodination of T_4 to T_3 ↓
- Normalization of peripheral thyroid hormones during abstinence
- Frequently blunted TRH test, may persist into abstinence

T_4, total thyroxine; T_3, total triiodothyronine; fT_3, free triiodothyronine; TSH, thyroid-stimulating hormone; TRH, thyrotropin-releasing hormone.

Water and Electrolyte Balance and Alcohol

Water and electrolyte balance can be substantially disturbed by alcohol consumption. While regulation of fluid balance and electrolyte homeostasis is highly complex, this section will focus on the effects of alcohol on arginine vasopressin (AVP), the main regulator of blood and urine osmolality.

AVP levels are profoundly altered by alcohol during acute and chronic alcohol ingestion as well as during alcohol withdrawal and abstinence. Acute alcohol ingestion results in a biphasic AVP response. During ascending plasma alcohol concentrations, AVP levels decrease leading to increased diuresis of free water. During descending plasma alcohol concentrations, AVP levels increase, resulting in decreased diuresis [25]. The effects of chronic alcohol consumption on AVP levels are unclear. Overhydration from increased AVP levels occur during withdrawal. Despite increased AVP levels, plasma, and urine osmolalities may be normal, contrary to what would be expected in AVP-induced overhydration. Higher AVP levels relative to serum sodium levels may result from a lowering of the threshold for osmotic stimuli. Increased AVP levels return to normal within 4 to 10 days. By contrast, plasma AVP levels are reduced in long-term abstinent alcoholics in whom the free water clearance is increased, indicating there is persistent alterations in hormonal regulation of electrolyte and water balance [25, 26].

Since alcohol withdrawal may be associated with overhydration, fluid management, particularly during withdrawal should be undertaken with caution. In addition, alcohol withdrawal may cause substantial disturbances in electrolyte homeostasis, thus, blood electrolytes should be monitored closely [25].

Beer Potomania

'Beer potomania' is a specific hypo-osmolality syndrome related to massive consumption of beer. It carries a high risk of an osmotic demyelination syndrome (ODS) resulting from rapid correction of hyponatraemia [27]. Because beer has very little sodium (<2 mmol/L), the low osmole intake results in an excess of electrolyte-free water but no solute for diuresis. If AVP levels at admission are not decreased, patients are likely to quickly convert to a state of low antidiuretic hormone (ADH) secretion. Exogenous solute intake may then lead to a brisk diuresis resulting in a rapid increase of serum sodium which in turn may lead to ODS. To prevent ODS, fluid management within the first 48 hrs is crucial. Common findings and recommendations for management of beer potomania are described in Box 12.2.1.4. ODS may manifest clinically 2 to 3 days after the initial insult, therefore, the patient's management must be defined at admission [27].

Hypertension and Alcohol

Alcohol consumption and hypertension (HT)—HT defined as systolic blood pressure (BP) >140 mmHg or diastolic BP <90 mmHg—is among the top five risk factors responsible for the growing global non-communicable diseases (NCD) burden [28]. Notably, chronic heavy drinking is among the most common reversible causes of HT and heavy drinking causes approximately

Box 12.2.1.4 Clinical findings and management of beer potomania

Typical history and clinical findings
- Excess beer drinking, recent binge drinking or illness
- Severe hyponatraemia (100–120 mmol/L), hypokalaemia
- Blood urea nitrogen ↓, urine Na^+ ↓
- Brisk diuresis in response to solute intake
- Mild neurologic symptoms (confusion)

High risk upon rapid serum sodium correction
- Osmotic demyelination syndrome (ODS)
- Symptoms range from mild confusion to coma and death

Management*
- Intensive care status
- Restrictive and careful fluid management
- Serum Na^+ every 2 hrs
- Serum Na^+ ↑ in first 24 hrs <10 mmol/L
- Serum Na^+ ↑ in first 48 hrs <18 mmol/L
- If necessary relower Na^+ levels

↓: decreased; ↑: increased; Na^+: sodium level
*: for more detail see: Sanghvi SR, *et al.* Beer potomania: an unusual cause of hyponatraemia at high risk of complications from rapid correction. *American journal of kidney diseases: the official journal of the National Kidney Foundation.* 2007;50(4):673–80.

16% of HT cases worldwide [29]. The underlying mechanisms are not well understood [30].

Whereas 1–2 drinks/day may have a neutral effect on long-term BP, regular alcohol intake at doses greater than 2 drinks/day increases BP in a linear dose-response relationship. More than 2 drinks/per day, each additional alcoholic drink will raise BP by 1.5 mm Hg and the relative risk (RR) for HT being 1.7 for 50 g alcohol/day and 2.5 for 100 g alcohol/day [29]. By contrast, acute alcohol intake causes peripheral vasodilatation causing a fall of BP. Hypotension may also develop in those with AI autonomic neuropathy and/or late-stage cardiomyopathy [30]. Reduction of alcohol intake lowers BP in a dose-dependent manner with an apparent threshold effect at two 2 drinks/day. The strongest effect was found in people who drank more than 6 drinks per day and reduced their intake by about 50% [28]. Upon cessation or significant reduction of alcohol intake, BP falls within 2–4 weeks. The effects on BP reduction are similar to that of other health behaviour changes such as physical activity, weight loss, and diets. Therefore, implementation of effective alcohol interventions in people who drink more than 2 drinks/day would be an effective means to reduce the disease burden from both alcohol and hypertension, particularly in countries with a substantial alcohol–attributable risk [28].

Growth Hormone/Insulin-like Growth Factor-1 Axis and Alcohol

Growth hormone (GH), is essential for the growth of all tissues in the body and acute and chronic alcohol exposure has various effects on the GH/IGF-1 axis.

Alcohol interferes with the normal, pulsatile release pattern of GH with the major secretory episode occurring during the first period of slow-wave sleep (SWS), inducing a significant and dose-dependent decrease of the nocturnal GH surge. While studies show a robust relationship between sleep disturbances and decreased GH secretion, alcohol-dependent individuals release less GH during SWS compared to normal control subjects [31]. GH responses to various challenges, e.g. apomorphine (dopamine receptor agonist), diazepam (allosteric modulator of $GABA_A$ receptor), sumatriptan (serotonin receptor agonist) are reduced not only in early, but also in long-term abstinence (1–2 years), indicating persistent impairment [14, 31].

The liver is the major source of IGF-1 which mediates the anabolic actions of GH. The regulation of IGF-1 in the liver by GH is inhibited by glucocorticoids. IGF-1 synthesis is reduced in long-term alcohol users with and without evidence of significant liver disease leading to progressive loss of muscle protein [32].

Parathyroid Hormone and Alcohol

Parathyroid hormone (PTH) regulates serum calcium levels through its effects on bone, kidney, and intestine. The activation of vitamin D by PTH increases calcium absorption by the intestine. PTH is stimulated by decreased serum calcium, and mild decreases in serum magnesium, whereas severe decreases in serum magnesium concentrations result in an inhibition of PTH.

Vitamin D deficiency is a common finding in alcohol-dependent patients, indicating a necessity for substitution. PTH levels are usually within the normal range but show a negative correlation with vitamin D levels. Calcium levels in general are also within the normal range. A recent study shows a negative correlation between calcium levels and alcohol craving, implying calcium supplementation as a potentially useful intervention to reduce craving and relapse in alcohol dependence. However, the role of calcium in addictive behaviour as well as the pros and cons of vitamin D and calcium supplementation in alcohol-dependent patients are still under debate and future studies are needed for clarification [33].

Chronic alcohol ingestion is often associated with a disturbance of electrolyte homeostasis (calcium, magnesium, phosphorus, and potassium), which is mainly due to poor intake, vomiting, diarrhoea, and increased urinary loss. Severe hypomagnesaemia results both in reduced secretion of PTH and end-organ (in bone and kidney) resistance to PTH, leading to hypocalcaemia und hyperphosphatemia. Transient hypoparathyroidism secondary to magnesium deficiency can be reversed by magnesium supplementation alone which corrects the calcium abnormalities [34].

Bone Disease and Alcohol

Alcohol is one of the leading risk factors for male osteoporosis, whereas in women it is an infrequent risk factor.

The alcohol-associated decrease in bone mass and strength is mainly due to an imbalance in bone remodelling with a predominant decrease in bone formation. The pathophysiology of alcohol's effects on bone is not fully understood, multifactorial in nature, includes direct and indirect mechanisms and is influenced by many confounders and comorbidities [35, 36].

Current evidence suggests that the doses (U-shaped relationship), duration of alcohol use, and the type of alcoholic beverage consumed influence bone loss. Whereas light alcohol consumption (8–10 g alcohol/day) is without effect, that of moderate alcohol consumption (2–3 drinks/day) may be positive or negative depending on sex, age, and hormonal status, and the type of beverage consumed. By contrast, chronic heavy drinking (100–200 g of alcohol/day) unequivocally reduces bone mass and bone mineral density. Alcohol consumption greater than 4 drinks/day in men and 2 drinks/day in women may increase fracture and trauma risk. Alcohol may delay the healing of fractures and associated complications. There is evidence from animal studies that binge drinking may influence bone remodelling. Thus, teenagers with binge drinking may decrease their peak bone mass and be at risk for later skeletal pathophysiology [36].

Furthermore, chronic alcohol abuse can rarely lead to aseptic osteonecrosis (ON), mainly of the femoral head, but also of other locations such as the knee. ON appears to be a late bone complication mainly seen in patients with late-stage liver disease. Therapy of choice for ON of the femoral heads is total hip arthroplasty [35]. For bone health, the European Foundation for Osteoporosis, and the National Osteoporosis Foundation of the United States has recommended that alcohol consumption be limited to 1 drink/day for women and 2 drinks/day for men [36].

Diabetes Mellitus and Alcohol

Since T1D usually manifests in childhood and adolescence, alcohol is more likely to affect the development and course of T2D although patients with T1D may benefit from limiting alcohol intake [37].

Current evidence suggests a J-shaped relationship between alcohol intake and the risk of developing T2D. Accordingly, maximum daily alcohol doses of 22 g in men and 24 g in women lowers the risk, whereas higher alcohol doses increased the risk for T2D both in women and men [37]. Potential mechanisms include decreased fasting insulin and HBA_1 concentrations among non-diabetic subjects and improved insulin sensitivity among women.

There is also strong evidence that alcohol may increase diabetes-induced complications ('double trouble' [37]) (See Table 12.2.1.1 online only) [37].

The American Diabetes Association recommends not more than 2 drinks/day for diabetic men and not more than 1 drink/day for diabetic women [29]. When pharmacologic treatment includes metformin, patients are to avoid consuming excessive amounts of alcohol because of the increased risk of a potentially life-threatening lactic acidosis.

Role of the HPA Axis in the Development and Maintenance of Alcohol Use Disorders

Changes in the activity of certain brain hormones have long been suggested to play an important role in the development and maintenance of addictive behaviour although proof of causality is still outstanding [38]. Whereas many neuroendocrine hormones have been implicated in addictive behaviour, the following section focuses on the HPA axis.

The HPA axis is one of the main stress response pathways and extensive research has shown that problematic alcohol use and dependence result in serious perturbations of the HPA axis. The effects of alcohol on the HPA axis are modulated by vulnerability factors (genetic factors, early-life environment, current stress, gene-environment interactions) and alcohol use-associated factors (dose, recent drinking history, expectancies) [39]. The dynamic interactions between alcohol and HPA axis function vary with respect to the stages of progression towards alcohol dependence, withdrawal, and recovery [39–42]. Table 12.2.1.2 (online only) shows the alcohol-associated changes on HPA axis regulation and their potential impact on development and maintenance of addictive behaviour. Altogether, there is ample evidence that the HPA axis—with particular emphasis on the CRH system—plays a key role in facilitating and maintaining substance use disorders and may therefore qualify as a major target for treatment.

Summary

Alcohol can impair endocrine function by affecting regulation at all levels of the endocrine system. Since endocrine regulation is particularly important during puberty and adolescence, binge drinking behaviour places this group at risk for later diseases. The dysregulation of the endocrine system particularly of the HPA axis may play a role in the development and maintenance of addictive disorders. Endocrine dysfunction is potentially reversible by abstinence in the absence of liver disease. The level of evidence supporting claims that low to moderate alcohol consumption may have positive effects on certain diseases, is not high due to methodological concerns [43].

Current recommendations are aimed at limiting regular alcohol intake to not more than one drink/day for women and not more than two drinks/day for men. Non-drinkers should not be advised to begin drinking (paucity of outcome data, abuse risk) and problem drinkers should be advised to abstain permanently or significantly reduce their alcohol intake and seek professional help therefore [29].

REFERENCES

1. Findling JW, Raff H. Diagnosis of endocrine disease: differentiation of pathologic/neoplastic hypercortisolism (Cushing's syndrome) from physiologic/non-neoplastic hypercortisolism

(formerly known as pseudo-Cushing's syndrome). *Eur J Endocrinol*, 2017; **176**: R205–16.
2. Fraietta R, Zylberstejn DS, Esteves SC. Hypogonadotropic hypogonadism revisited. *Clinics*, 2013; **68** (Suppl. 1): 81–8.
3. Emanuele MA, Emanuele N. Alcohol and the male reproductive system. *Alcohol Res Health*, 2001; **25**: 282–7.
4. Villalta J, Ballesca JL, Nicolas JM, Martinez de Osaba MJ, Antunez E, Pimentel C. Testicular function in asymptomatic chronic alcoholics: relation to ethanol intake. *Alcohol Clin Exp Res*, 1997; **21**: 128–33.
5. Schiavi RC, Stimmel BB, Mandeli J, White D. Chronic alcoholism and male sexual function. *Am J Psychiatry*, 1995; **152**: 1045–51.
6. Jensen TK, Gottschau M, Madsen JO, *et al.* Habitual alcohol consumption associated with reduced semen quality and changes in reproductive hormones; a cross-sectional study among 1221 young Danish men. *BMJ Open*, 2014; **4**: e005462.
7. Hasselblatt M, Krieg-Hartig C, Hufner M, Halaris A, Ehrenreich H. Persistent disturbance of the hypothalamic-pituitary-gonadal axis in abstinent alcoholic men. *Alcohol Alcoholism*, 2003; **38**: 239–42.
8. Heinz A, Rommelspacher H, Graf KJ, Kurten I, Otto M, Baumgartner A. Hypothalamic-pituitary-gonadal axis, prolactin, and cortisol in alcoholics during withdrawal and after three weeks of abstinence: comparison with healthy control subjects. *Psychiatry Res*, 1995; **56**: 81–95.
9. Walter M, Gerhard U, Gerlach M, Weijers HG, Boening J, Wiesbeck GA. Controlled study on the combined effect of alcohol and tobacco smoking on testosterone in alcohol-dependent men. *Alcohol Alcoholism*, 2007; **42**: 19–23.
10. Vatsalya V, Issa JE, Hommer DW, Ramchandani VA. Pharmacodynamic effects of intravenous alcohol on hepatic and gonadal hormones: influence of age and sex. *Alcohol Clin Exp Res*, 2012; **36**: 207–13.
11. Watts EL, Appleby PN, Albanes D, *et al.* Circulating sex hormones in relation to anthropometric, sociodemographic and behavioural factors in an international dataset of 12,300 men. *PloS One*, 2017; **12**: e0187741.
12. Sermondade N, Elloumi H, Berthaut I, *et al.* Progressive alcohol-induced sperm alterations leading to spermatogenic arrest, which was reversed after alcohol withdrawal. *RBMO*, 2010; **20**: 324–7.
13. Van Thiel DH. Feminization of chronic alcoholic men: a formulation. *Yale J Biol Med*, 1979; **52**: 219–25.
14. Rachdaoui N, Sarkar DK. Pathophysiology of the effects of alcohol abuse on the endocrine system. *Alcohol Res Curr Rev*, 2017; **38**: 255–76.
15. Emanuele MA, Wezeman F, Emanuele NV. Alcohol's effects on female reproductive function. *Alcohol Res Health*, 2002; **26**: 274–81.
16. Gill J. The effects of moderate alcohol consumption on female hormone levels and reproductive function. *Alcohol Alcoholism*, 2000; **35**: 417–23.
17. Hirko KA, Spiegelman D, Willett WC, Hankinson SE, Eliassen AH. Alcohol consumption in relation to plasma sex hormones, prolactin, and sex hormone-binding globulin in premenopausal women. *Cancer Epidemiol Biomarkers Prev*, 2014; **23**: 2943–53.
18. Purohit V. Moderate alcohol consumption and estrogen levels in postmenopausal women: a review. *Alcohol Clin Exp Res*, 1998; **22**: 994–7.
19. Hvidtfeldt UA, Tjonneland A, Keiding N, *et al.* Risk of breast cancer in relation to combined effects of hormone therapy, body mass index, and alcohol use, by hormone-receptor status. *Epidemiology*, 2015; **26**: 353–61.
20. Liu Y, Nguyen N, Colditz GA. Links between alcohol consumption and breast cancer: a look at the evidence. *Women's Health*, 2015; **11**: 65–77.
21. Boffetta P, Hashibe M. Alcohol and cancer. *Lancet Oncol*, 2006; **7**: 149–56.
22. Zeisser C, Stockwell TR, Chikritzhs T. Methodological biases in estimating the relationship between alcohol consumption and breast cancer: the role of drinker misclassification errors in meta-analytic results. *Alcohol Clin Exp Res*, 2014; **38**: 2297–306.
23. Hegedus L, Rasmussen N, Ravn V, Kastrup J, Krogsgaard K, Aldershvile J. Independent effects of liver disease and chronic alcoholism on thyroid function and size: the possibility of a toxic effect of alcohol on the thyroid gland. *Metabolism*, 1988; **37**: 229–33.
24. Hermann D, Heinz A, Mann K. Dysregulation of the hypothalamic-pituitary-thyroid axis in alcoholism. *Addiction*, 2002; **97**: 1369–81.
25. Ragland G. Electrolyte abnormalities in the alcoholic patient. *Emerg Med Clin North Am*, 1990; **8**: 761–73.
26. Doring WK, Herzenstiel MN, Krampe H, *et al.* Persistent alterations of vasopressin and N-terminal proatrial natriuretic peptide plasma levels in long-term abstinent alcoholics. *Alcohol Clin Exp Res*, 2003; **27**: 849–61.
27. Sanghvi SR, Kellerman PS, Nanovic L. Beer potomania: an unusual cause of hyponatremia at high risk of complications from rapid correction. *Am J Kidney Dis*, 2007; **50**: 673–80.
28. Roerecke M, Kaczorowski J, Tobe SW, Gmel G, Hasan OSM, Rehm J. The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. *Lancet Public Health*, 2017; **2**: e108–e20.
29. O'Keefe EL, Di Nicolantonio JJ, O'Keefe JH, Lavie CJ. Alcohol and CV health: Jekyll and Hyde J-curves. *Prog Cardiovasc Dis*, 2018; **61**: 68–75.
30. Grogan JR, Kochar MS. Alcohol and hypertension. *Arch Fam Med*, 1994; **3**: 150–4.
31. Lands WE. Alcohol, slow wave sleep, and the somatotropic axis. *Alcohol*, 1999; **18**: 109–22.
32. Santolaria F, Gonzalez-Gonzalez G, Gonzalez-Reimers E, *et al.* Effects of alcohol and liver cirrhosis on the GH-IGF-I axis. *Alcohol Alcoholism*, 1995; **30**: 703–8.
33. Schuster R, Koopmann A, Grosshans M, Reinhard I, Spanagel R, Kiefer F. Association of plasma calcium concentrations with alcohol craving: new data on potential pathways. *Eur Neuropsychopharmacol*, 2017; **27**: 42–7.
34. Hermans C, Lefebvre C, Devogelaer JP, Lambert M. Hypocalcaemia and chronic alcohol intoxication: transient hypoparathyroidism secondary to magnesium deficiency. *Clin Rheumatol*, 1996; **15**: 193–6.
35. Mikosch P. Alcohol and bone. *Wiener Medizinische Wochenschrift*, 2014; **164**: 15–24.
36. Maurel DB, Boisseau N, Benhamou CL, Jaffre C. Alcohol and bone: review of dose effects and mechanisms. *Osteoporosis Int*, 2012; **23**: 1–16.
37. Munukutla S, Pan G, Deshpande M, Thandavarayan RA, Krishnamurthy P, Palaniyandi SS. Alcohol toxicity in diabetes

and its complications: a double trouble? *Alcohol Clin Exp Res*, 2016; **40**: 686–97.
38. Kenna GA, Swift RM, Hillemacher T, Leggio L. The relationship of appetitive, reproductive and posterior pituitary hormones to alcoholism and craving in humans. *Neuropsychol Rev*, 2012; **22**: 211–28.
39. Stephens MA, Wand G. Stress and the HPA axis: role of glucocorticoids in alcohol dependence. *Alcohol Res Curr Rev*, 2012; **34**: 468–83.
40. Blaine SK, Sinha R. Alcohol, stress, and glucocorticoids: from risk to dependence and relapse in alcohol use disorders. *Neuropharmacology*, 2017; **122**: 136–47.
41. Heilig M, Koob GF. A key role for corticotropin-releasing factor in alcohol dependence. *Trends Neurosci*, 2007; **30**: 399–406.
42. Walter M, Gerhard U, Gerlach M, Weijers HG, Boening J, Wiesbeck GA. Cortisol concentrations, stress-coping styles after withdrawal and long-term abstinence in alcohol dependence. *Addict Biol*, 2006; **11**: 157–62.
43. Naimi TS, Stockwell T, Zhao J, *et al.* Selection biases in observational studies affect associations between 'moderate' alcohol consumption and mortality. *Addiction*, 2017; **112**: 207–14.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

12.2.2 Use and Abuse of Performance-Enhancing Hormones in Sport

Peter Sonksen and Richard I.G. Holt

Introduction

Anecdotes suggest that attempts to improve athletic performance in both man and animals date from the times of ancient Greece and Rome and probably before. The status of an athletic winner has always been sufficient to encourage people to cheat and in recent times winners have been rewarded with generous sponsorships and can accumulate great wealth.

Fair play and ethical behaviour are among the fundamental principles of 'Olympians' laid out in the Olympic Charter [1] which aims to achieve a hypothetical 'level playing field' where winning reflects true ability, dedication, and training rather than mastering the science for using performance-enhancing substances to cheat—the process known as 'doping'. Doping not only damages the integrity of sport but may cause significant harm to athletes. The origin of the term 'doping' is uncertain; some sources suggest that it was originally derived from the African Kaffirs word 'dop', an alcoholic drink made from grape skins that was used as a stimulant in battle, while others believe it came from the slang term for opium.

The true prevalence of doping is unknown and reports vary widely from 1% to 90% because of the secrecy surrounding it. The huge financial rewards for professional sportsmen and women and the need for corporate sponsors and TV broadcast act as a great temptation to cheat. When coupled with the ever-increasing pharmacopoeia of performance-enhancing substances, the athlete's drive to win may create an overwhelming urge to use prohibited substances.

The International Olympic Committee (IOC) Medical Commission

It was the deaths of the Danish cyclist Knud Jensen at the Rome Olympiad in 1960 and the British cyclist Tommy Simpson before the cameras in the 1967 Tour de France that made people sit up and realize that something had to be done to control the misuse of substances in the quest of winning. The IOC, aware of the growing problem, first placed the responsibility for dealing with this on the Sports Federations but when it was clear that this was ineffective, formed a 'Medical Commission' in 1967. The remit was to oversee the introduction of an effective antidoping system. The IOC Medical Commission established a list of banned substances that was regularly updated as 'The Prohibited List'. The Medical Commission also developed procedures as well as methods for detecting drugs and certifying the process. It was not until 1982, however, that the first dedicated IOC laboratory was created in Los Angeles with a grant from the organizing committee for the 1984 Olympic Games.

This laboratory was accredited by the IOC in November 1983 and detected one adverse analytical finding of an anabolic steroid (methandienone) at the Winter Olympic Games at Sarajevo in February 1984 and 12 positives at the Los Angeles Summer Games of 1984.

The IOC Laboratories and the 'Prohibited List'

Subsequently the IOC established and accredited a network of antidoping laboratories around the world. Each laboratory had to participate annually in a 'Quality Assurance' programme where they had to analyse a set of 'blind' samples and come up with the right answers. Failure to do so resulted in immediate suspension and the need for re-accreditation. This laboratory network was effective and the sports federations had to use these laboratories while running and paying for their own testing programmes.

The Creation of the World Anti-Doping Agency (WADA) and the Court of Arbitration for Sport (CAS)

The IOC system was not without critics, the main issue being the potential conflict of interest between the organizers of the Olympic Games and the system run to prevent doping. This led to the creation of WADA in 1999, composed and funded equally by the sport's movement and national governments. WADA took over responsibility for managing the network of IOC laboratories as well as producing the annually updated list of banned substances and methods.

To resolve legal disputes in elite sport, the CAS came into being in June 1984. Initially created and funded by the IOC and reformed in 1994 to make it independent of the IOC, CAS was verified in 2003 by a Swiss Federal Tribunal. It is the 'high court' for all sporting disputes, including doping violations.

The World Anti-Doping Agency List of Prohibited Substances

Initially when WADA was created and took over the official list of prohibited substances, they said that they would only include substances (and methods) that had been proven to be performance enhancing but it was not long before they adopted a more pragmatic approach, previously used by the IOC.

The decision to include a substance or method in the prohibited list depends on the substance meeting two of three criteria:

- The substance or method enhances sporting performance or has the potential to do so
- The substance or method may cause harm to the athlete or has the potential to do so
- The substance or method violates the spirit of sport

The discovery of new performance-enhancing drugs often arises after athletes have discovered desirable properties often using a version of the 'trial of one' methodology [2]. Once others have started using them and confirmed their efficacy, their use spreads by word-of-mouth and clandestine publications. Others are detected by laboratories through routine analysis of urine samples by mass spectrometry as a part of dope testing and their ergogenic properties are assumed by their spreading popularity.

A good recent example of this is meldonium. Developed in Latvia, it was initially used by Russian troops in Afghanistan to improve endurance and subsequently adopted by athletes. In 2015, when meldonium started appearing in athlete's urine samples, it was not known to be performance enhancing but was added to the list of substances monitored by WADA. Its use by athletes became extensive, largely in the Eastern European bloc countries [3]. At the 2015 Baku European Games, 13 medallists or competition winners were taking meldonium and most of the athletes taking meldonium did not declare it. It was subsequently added to the WADA list of banned substances operational from 1 January 2016. On 7 March 2016, the former number one tennis player Maria Sharapova announced that she had failed a drug test due to the detection of meldonium; she received a 2-year ban from the International Tennis Federation [4].

Some drugs are banned always, while others are only banned in competition while others are only banned in certain sports. A summary of the 2018 section dealing with endocrine-related substances is reproduced in **Box 12.2.2.1**.

Box 12.2.2.1 Summary of hormones and drugs appearing on the 2018 WADA list of prohibited substances

Substances and methods prohibited at all times

S1 anabolic agents:

1 Anabolic androgenic steroids (AAS)
   - a Exogenous AAS
   - b Endogenous AAS when administered exogenously

2 Other anabolic agents: including but not limited to: clenbuterol, SARMs

S2 peptide hormones, growth factors, related substances, and mimetics:

1 Erythropoietins (EPO)
   - 1.1 EPO receptor agonists
   - 1.2 Hypoxia-inducing factor (HIF) activating agents
   - 1.3 GATA inhibitors
   - 1.4 TGF-beta (TGF-β) inhibitors
   - 1.5 Innate repair receptor agonists

2 Peptide hormones and hormone modulators
   - 2.1 Chorionic gonadotrophin (CG) and luteinizing hormone (LH) and their releasing factors
   - 2.2 Corticotrophins and their releasing factors
   - 2.3 Growth hormone (GH), its fragments and releasing factors

3 Growth factors and growth factor modulators

S3 β-2 agonists:

Except:

- Inhaled salbutamol—restricted amount
- Inhaled formoterol—restricted amount
- Inhaled salmeterol—restricted amount

S4 hormone and metabolic modulators:

1 Aromatase inhibitors
2 Selective oestrogen receptor modulators (SERMS)
3 Other antioestrogen substances
4 Agents modifying myostatin function
5 Metabolic modulators
   - 5.1 Activators of the AMP-activated protein kinase (AMPK)
   - 5.2 Insulins and insulin-mimetics
   - 5.3 Meldonium
   - 5.4 Trimetazidine

Modified with permission from The World Anti-Doping Agency (WADA). Code International Standard: Prohibited List. 1 Jan 2018. Available at: https://www.wada-ama.org/sites/default/files/prohibited_list_2018_en.pdf [Accessed 2 July 2019].

Endocrinology and Doping

Testosterone and Anabolic Androgenic Steroids

Testosterone and anabolic androgenic steroids (AAS) are among the most commonly misused performance-enhancing drugs representing 43% of Adverse Analytical Findings in WADA-accredited laboratories in 2016, although this may be just because they have been easier to detect. AAS misuse is not confined to elite athletes and indeed use by amateur bodybuilders and weightlifters for appearance purposes outnumbers elite athletes. It is estimated that 6.4% of men and 1.6% of women have taken AAS.

The effects of testosterone in animal husbandry has been long known but it was Charles Édouard Brown-Séquard who is usually credited with the preparation and use of a watery testicular extract that he found to improve strength and well-being in himself and other older men [5].

The next important event was the isolation and synthesis of testosterone independently in 1935 by the German Adolf Butendant and the Croatian Leopold Ruzicka, for which they were awarded The Nobel Prize in Chemistry in 1939 [6]. There were rumours that testosterone was used for performance-enhancement and aggression in the German Army but there are no facts to support this. Likewise, there is no evidence that testosterone was used in the 1936 Olympic Games in Germany. The first well-documented use of testosterone as a performance-enhancing drug was with the German rowing team in 1952 [7].

Testosterone use in athletics was also suspected in 1952 at the Summer Olympics in Helsinki. The Soviet Union competing for the first time at the Olympics dominated weightlifting and hearsay attributed this to testosterone. This suspicion was later confirmed when a Soviet doctor told a US colleague that they were experimenting with testosterone [8].

Because of unwanted virilizing side effects, there was a drive by the pharmaceutical industry to alter the molecule to improve its therapeutic profile—altering the pharmacokinetics and increasing the anabolic effects while reducing the androgenic effects. This resulted in many new anabolic steroids, methandrostenolone (Dianabol) developed by CIBA being one of the first. Within 2 years 'almost every world-class athlete knows of its performance-enhancing possibilities' [8]. Synthetic androgens predominantly have one of two substitutions on the D ring of native testosterone, either esterification at the 17-β-hydroxy group or alkylation at the 17-α position (**Figure 12.2.2.1**). Many variations on the structure of testosterone has led to a long list of so-called designer steroids.

In the late 1990s, topical preparations of testosterone became available. Many further manipulations of the basic steroid structure have been undertaken to alter the anabolic and androgenic actions as well as to avoid detection by increasingly sophisticated antidoping technology. The extent of this can be judged by the list of known variants in the latest WADA Prohibited Substances list [9].

State-Sponsored Doping

State-sponsored doping in the former German Democratic Republic (GDR) has been well documented by Franke and Berendonk, including secret files recovered from the Stasi following the unification of Germany [10]. In the 1960s the GDR was a relatively obscure country with a 'Cold War' image and dominated by the 'Iron Curtain'. GDR politicians decided that they could enhance their international prestige through athletic success. They instituted a system of early selection of potential athletes and systematic doping. They used mainly a chlor-substituted version of methandrostenolone (Oral-Turinabol). In 1968 they started administering Oral-Turinabol to women athletes as well as to men. In the 1972 Olympic Games this small country with only 17 million people were placed third in the medal tables (66 medals) behind the United States (99 medals) and the Soviet Union (94 medals); next was West Germany (62 million people) with 40 medals. In swimming at the Olympic Games in Montreal in 1976 GDR athletes won Gold in 11 of the 13 events.

Monitoring the performance of their athletes closely and adjusting doping protocols allowed trainers to study not only how performance improved with dosage but also how long the improved performance lasted after the anabolic steroid was discontinued. They were then skilled at maximizing performance while ensuring that their athletes did not test positive when competing in major events outside GDR. After one GDR athlete tested positive in 1977 a rule was introduced that all athletes needed to provide a urine sample that would be tested and need to be found 'clean' in a central laboratory (actually accredited by the IOC) before being allowed to compete abroad [10].

Interestingly, while this was occurring in the GDR, there was considerable debate about whether anabolic steroids were performance enhancing. The science was hampered because undertaking double-blind placebo-controlled trials in elite athletes is considered unethical.

Other methods were therefore needed to determine 'efficacy' of potential performance-enhancing drugs. In the GDR, trials of one and carefully executed case–control studies were the most frequently used methodologies and provided the most persuasive evidence. Despite the scepticism shown by Western scientists, the benefits were realized by athletes whose performance improved dramatically. For example, between 1956 and 1980, the winning distance in the Olympic women's shot put increased from 15.28 m to 22.41 m (**Figure 12.2.2.2**). Since the introduction of effective testing in the early 1980s, the winning distance in the women's shot put has fallen progressively to the point where the winner in Rio de Janeiro in 2012 would not have made the final in Moscow in 1980 and has plateaued since then.

A further episode of state-sponsored doping occurred more recently at the Winter Olympic Games in Sochi and as a result 111 of 389 Russian athletes were banned from the 2016 Summer Olympic Games in Rio de Janeiro and all their Paralympic athletes from the Paralympic Games: in addition because the IOC were not satisfied with the corrective measures required, Russia was banned from the 2018 Winter Olympic Games in South Korea [11].

Why Do Athletes Misuse Anabolic Androgenic Steroids?

AAS increase muscle mass and strength [12] through several mechanisms including the stimulation of protein synthesis, recruitment of satellite cells, production of cytokines, and increase in androgen

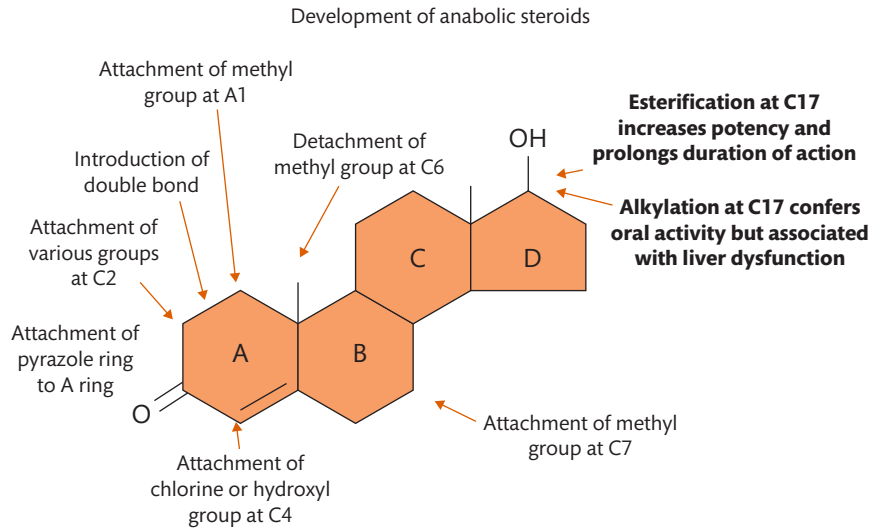


Figure 12.2.2.1 Development of androgenic anabolic steroids. As the half-life of testosterone is short, manipulations at the C17 position of the D ring have led to compounds that are either more potent and longer lasting (esterification) or orally acting (alkylation), marked in bold. Further alterations have been tried to alter the relative androgenic and anabolic actions.

receptor number. Clinical trials have shown that AAS improve both muscle mass and strength in a dose-dependent manner and these effects are additive to resistance training alone. The full potential of AAS, however, has almost certainly not been realized in these trials because of the limited doses used because of medical and ethical concerns about potential side effects. By contrast, doping athletes are reported to take combinations of AAS in doses that are up to 30 times higher than the physiological replacement dose. Although there is little hard evidence that AAS are performance enhancing, since performance in many athletic sports is dependent on lean body mass and strength, it is likely that they do have ergogenic effects.

There are many websites and printed manuals giving details about the use and supply of AAS. Steroid use usually involves techniques such as 'cycling' and 'stacking'. AAS appear to lose effectiveness with time through adaptation of the androgen receptor and athletes have learned to obtain greater benefit through using the steroids intermittently in a cycle over 5–10 weeks. Stacking involves the use of two or more steroids concomitantly to obtain a synergistic effect.

Adverse Effects of Androgenic Anabolic Steroids

The most marked side effect of AAS use is virilization. In both men and women these manifests as acne, an increase in body hair and male pattern baldness. In men, prostate enlargement may occur leading to urinary hesitancy and poor flow and there are concerns about long-term risk of prostate cancer. Gynaecomastia may develop as excessive testosterone is converted to oestradiol under the action of the aromatase enzyme. In women the virilizing effects are more marked including hirsutism, a decrease in breast size, clitoromegaly, and enlargement of the larynx causing deepening of the voice; these changes can be permanent.

Administration of AAS may reduce fertility through suppression of luteinizing hormone (LH) and follicle-stimulating hormone

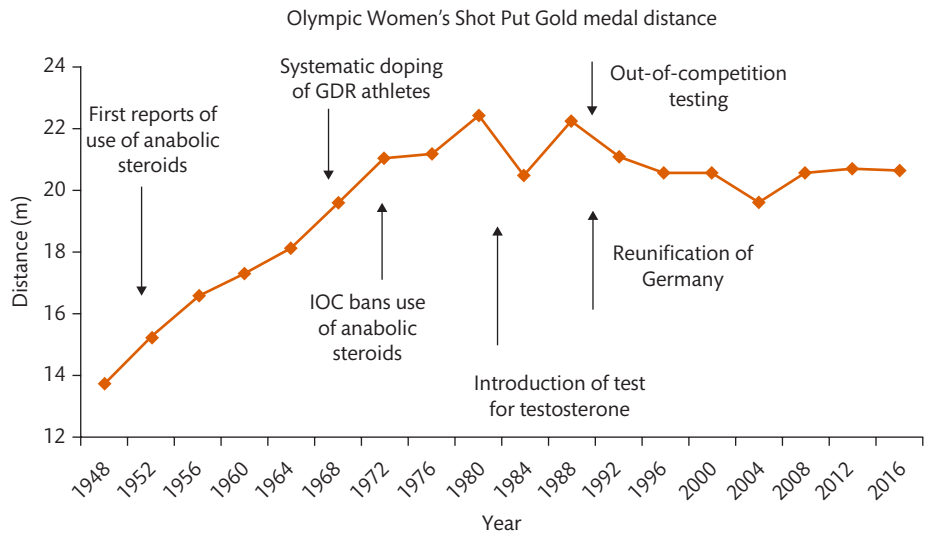


Figure 12.2.2.2 The winning distance in the women's Olympic Shot Put since 1948.

(FSH) secretion. In men sperm count falls and testicular atrophy may occur. In women there is reduced endogenous oestrogen production, ovulation is impaired, and irregular or absent menstrual cycles may ensue. These effects on reproduction function are prolonged taking many months to return to normal after AAS discontinuation.

Hepatotoxicity can occur with orally administered AAS due to a 'first pass' effect on the liver and cholestasis can progress to hepatocellular necrosis. The incidence of myocardial infarction is increased in young men taking AAS as is the risk of developing hypertrophic cardiomyopathy in genetically predisposed individuals. AAS misuse is also associated with the development of dyslipidaemia. The concentration of high-density lipoprotein (HDL) cholesterol is reduced while the concentration of low-density lipoprotein (LDL) cholesterol increases. The use of AAS may be associated with increased irritability, aggressiveness, and rarely mania. Dependency may develop in long-term users and depression may occur after AAS withdrawal.

Detection of AAS Misuse

Detecting AAS and many other drugs in urine is relatively simple and effective for two reasons: first, they are 'foreign' substances and should not be there at all and secondly, their metabolism and pharmacokinetics often mean that they can be traced for some time in urine.

Clenbuterol

Clenbuterol, a long-acting β2 agonist, is used as a partitioning agent in cattle where it promotes dietary intake towards muscle growth at the expense of fat [13]. It is an anabolic agent and a drug of abuse in sport. Eating contaminated meat can lead to the accumulation of levels in the human body that will provoke an adverse analytical finding [14]. It is also a bronchodilator and used to treat asthma in many countries (but not in the United Kingdom). B2 agonists are favourites of cross-country skiers where often they race at low environmental temperatures and may develop cold-induced bronchoconstriction. Clenbuterol is on the WADA list of prohibited substances and has a very long half-life in the body. In 1992 Katrin Krabbe of the GDR, along with two teammates, tested positive for clenbuterol and this effectively ended her career [15].

Growth Hormone

Growth hormone (GH) was initially promoted as a performance-enhancing drug in the first edition of *The Underground Steroid Handbook* authored by Dan Duchesne in 1982. It described GH as the newest and most potent anabolic substance of interest to athletes. This was seven years before the results of the first two double-blind controlled trials of GH administration to people with GH deficiency were published in the peer-reviewed scientific literature. These showed (among other things) the dramatic effects of GH on body composition [16, 17]. By then, however, it had become widely known in elite sport as a 'doping agent' often used in combination with testosterone. This was dramatically demonstrated to the world by the sprinter Ben Johnson in the Seoul Olympic Games of 1988 when he beat his arch rival Carl Lewis to win Gold in the 100 metres only to lose the medal a few days later when his urine tested positive for the anabolic steroid stanozolol; he later admitted in court to have taken GH as well [18].

In addition to the confession of Ben Johnson, other events pointed to its widespread misuse. For example, athletes have been caught smuggling vials of GH into countries where games are held [19] and pharmacy burglaries and hijacking of supply trucks have often seen specific targeting of GH.

Why Do Athletes Abuse Growth Hormone?

GH is best considered, like clenbuterol, as a 'partitioning agent' ('something which when included in an animal's nutrition increases the proportion of food energy that goes into the production of meat rather than fat') and results in changes in body composition with increases in lean tissue and matching reduction in fat mass. This is achieved through stimulation of protein synthesis at the expense of increased lipolysis [13]. Most of normal linear growth is determined by GH and IGF-I while androgens, oestrogens and other steroid hormones contribute little before puberty [13]. Post-puberty, GH, IGF-I, testosterone, and oestrogens are responsible for building and maintaining body composition. Excess GH and/or testosterone results in increased lean body mass in a dose-dependent manner. As strength is closely related to lean body mass it is easy to see why both testosterone and GH are misused by elite athletes to enhance performance although properly controlled trials to demonstrate this in athletes are very scarce since they are limited by ethical constraints.

In addition to regulating body composition, GH together with insulin are critical for the storage and timely release of metabolic fuel from depots and used in sport [13, 20–23].

The side effects of GH use are uncommon and occur only after prolonged use. These include a myopathy and development of diabetes.

Detection of Growth Hormone Misuse

GH misuse has been and still is difficult to detect. By 1989 the IOC recognized the need for a GH test and somewhat reluctantly for the first time put money into research for a method to detect GH misuse. There are now two methods available and approved by WADA, but both lack sensitivity and are probably only catching a small minority of those doping with GH.

The first method works by demonstrating an abnormally high ratio of the 22 kD isoform of GH to total GH. Recombinant GH (rhGH) contains only the 22 kD isoform and when injected leads to the suppression of all other isoforms that are normally present [21]. The second approach is based on two GH-sensitive 'markers' (IGF-I and P-III-NP), which when included in a discriminant function, are able to detect the prior use of GH for up to 14 days [22]. The tests are complimentary; the isoform test, introduced at the Athens Olympic Games of 2004, has a short 'window of opportunity (24 hours)' and is most potent when used for 'out of competition testing' while the marker method, introduced at the London Olympic Games of 2012, can detect prior GH use for longer. At the 2012 London Paralympic Games, two athletes tested positive by the 'marker' method who were not detected by the 'isoform' test. The athletes subsequently admitted that they took GH and were disqualified. This finding indicates that either they took rhGH more than 24 hours before they were tested, or they had taken 'pituitary-derived' GH which is not picked up by the isoform test since it still contains several isoforms. Pituitary-derived GH is believed to be available on the black market but although those misusing it may avoid testing positive by the isoform test, they will be at risk of contracting Jacob–Creutzfeldt disease.

Insulin-Like Growth Factor-I (IGF-I)

Although IGF-I is synthesized by most cells in the body, the majority of circulating IGF-I is produced by the liver under the control of GH and insulin. It has profound effects on cell proliferation and differentiation in many tissues as well as metabolic effects broadly like insulin. Unlike insulin, however, IGF-I stimulates protein synthesis whereas insulin's protein anabolic actions are mediated through inhibition of protein breakdown [24].

No doubt supplies are available on the black market and it may well be used as a performance-enhancing peptide and is on the WADA list of prohibited substances. In a double-blind randomized controlled study in 30 male and 26 female recreational athletes, recombinant human IGF-I (rhIGF-I) combined with IGF Binding Protein-3 (rhIGFBP-3) improved aerobic performance, as assessed by maximal oxygen consumption (VO_2 max), by 7%. In contrast to GH, there was no significant change in body composition [25].

Insulin

Insulin's major function is to control glucose, fat, and protein metabolism and in its absence type 1 diabetes develops with uncontrolled breakdown of body protein and fat to fuel gluconeogenesis and ketogenesis. Insulin normally regulates glucose metabolism primarily by inhibiting hepatic glucose production from glycogenolysis and gluconeogenesis (the 'Chalonic' action of insulin [26]). Contrary to common assumptions glucose entry into cells is primarily dependent on the glucose concentration in the blood stream and extracellular fluid and is only influenced by insulin at very high concentrations of insulin. Glycogen content of muscles is an important component of athletic performance in many sports. Infusions of insulin at rates sufficient to raise insulin concentrations to the high physiological levels and concomitant infusions of glucose to maintain normoglycaemia results in increases in muscle glycogen proportional to the rates of glucose infusion required to maintain normoglycaemia [27] (Figure 12.2.2.3). The use of insulin in such a way in people without diabetes was banned by the IOC in Nagano at the Winter Games in 1998.

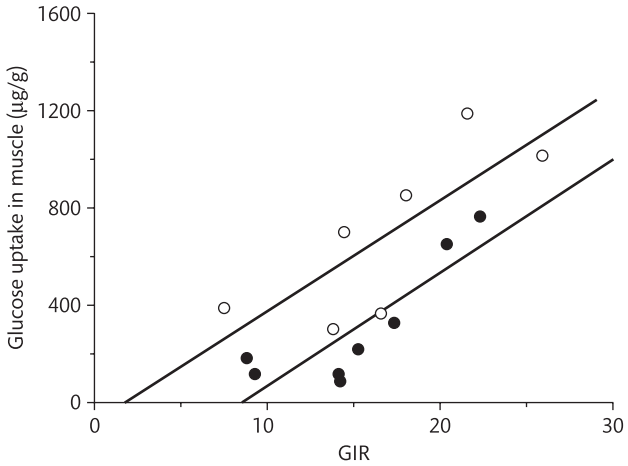


Figure 12.2.2.3 Relationship between glucose uptake in muscle glycogen and glucose infusion rate and the effects of insulin.

Reproduced with permission from Marin P, Hijgh-Kristiansen I, Jansson S, Krotkiewski M, Holm G. Uptake of glucose carbon in muscle glycogen and adipose tissue triglycerides *in vivo* in humans. *Am J Physiol*. 1992;263 *Endocrinol. Metab* 26:E473–80. Copyright © 1992 the American Physiological Society.

Erythropoietin

Receptor Gene Mutation and Blood Transfusion

Erythropoietin (EPO) is a hormone produced by the peritubular cells in the proximal tubules of the kidney and stimulates erythrocyte production by the bone marrow. Of interest to sport and doping is a family with a mutation in the EPO receptor who have erythrocytosis with asymptomatic polycythaemia due to continued bone marrow stimulation. The original phenotype (Eero Antero Mäntyranta) was a successful Finnish cross-country skier winning seven Olympic Medals [28, 29].

Prevalence of Erythropoietin Administration

Blood doping was first prohibited by the IOC in 1985. Before the ban was introduced, it was openly and commonly used by middle and long-distance runners as well as cyclists. Since its prohibition, there have been several high-profile cases involving athletes who have used either EPO or blood transfusion. These include Niklas Axelsson who tested positive for EPO in 2000 and Tyler Hamilton who used a homologous blood transfusion in 2004. The Spanish Operación Puerto in 2006 investigated allegations of blood doping in hundreds of athletes while several members of the Astana Team in the 2007 Tour de France tested positive for homologous blood transfusion leading to the withdrawal of the team. The German speed skater and fivefold Olympic Gold medallist Claudia Pechstein was banned for two years in 2009 for alleged blood doping.

Why do Athletes Abuse Erythropoietin?

Exogenous EPO, which is available as a recombinant protein or as an analogue, is abused by endurance athletes as the increase in red blood cells improves oxygen transport to muscles. Its administration results in a slow and sustained increase in erythrocyte volume, which is associated with improved performance. Endogenous EPO production can be induced by training at high altitude and this leads to a similar increase in erythrocyte volume and performance. However, only about 50% of competitive athletes respond to altitude training and it is notable that non-responders do not improve the aerobic capacity.

An alternative method used by athletes to expand erythrocyte volume is autologous or homologous blood transfusion. Blood is removed and erythrocytes are harvested, stored, and then reinfused later. In contrast to EPO administration, following a blood transfusion, the erythrocyte volume is only increased for a few weeks

Adverse Effects of Erythropoietin and Blood Transfusion

The main adverse effect of an increased erythrocyte volume is an increased risk of thrombotic events. However, there are also additional risks of infection associated with blood transfusion, particularly when this occurs in an unregulated fashion. Mismatched blood transfusions have occurred with serious consequences.

Detection of Erythropoietin Misuse

EPO testing (like that for GH) is difficult. The earliest test of EPO misuse was based on detecting immuno-specific bands in urine subjected to isoelectric focussing. This method separates natural EPO from synthetic varieties based on charge carried by the molecules. This method was first trialled in the Sydney Olympic Games in 2000 and has since been further refined and validated. It remains the mainstay of EPO misuse detection. This is used together with a blood test that examines the profile of red cells looking specifically at the proportion of young reticulocytes.

The Athlete's Biological Passport

Introduced by WADA in 2013, the athlete's biological passport involves individual athlete's 'profiling' of several biological variables in blood samples and the building of a statistical model representing each athlete. If a given sample produces a concentration of one or more key variables that are outside the 'normal limits' of their individual statistical model then that is considered a positive or can be used to monitor a suspicious athlete more closely. There are still concerns about the sensitivity and the reliability of the method [30]. There are plans to add an 'endocrine module' covering GH and other growth factors.

Synergistic Performance-Enhancing Properties of Testosterone, Growth Hormone, and Insulin

Although generally considered the most important, testosterone is not the only hormone regulating body composition and strength. GH plays an equal if not a more important role [13]. In fact, testosterone, GH and insulin work together in determining and regulating body composition while GH and insulin are the key substances regulating the flow of nutrients needed before, during and after an athletic event [20, 31].

Because of the ethical difficulties in giving GH or testosterone to athletes for research purposes and despite strong evidence of their anabolic effects, very little good evidence exists of their performance-enhancing properties. The one excellent randomized and well-controlled study in 96 recreational athletes carried out by Meinhardt *et al*. in Australia showed a 3.9% increase in sprint capacity with GH alone and an 8.3% increase when GH was coadministered with testosterone in men [32]. One short-term randomized and controlled study (6 days) in abstinent anabolic steroid-dependent men of average age 32 years showed a significant effect of GH on strength, peak power output, fat-free mass index and VO_2 peak [33]. In elderly unathletic men, Giannoulis *et al.* used a double-blind randomized placebo-controlled trial that looked at placebo versus GH alone, testosterone alone or testosterone and GH combined and found positive effects of GH alone on a number of exercise-related end points but more marked effects when GH was combined with testosterone [34].

Weissberger *et al*. using a randomized double-blind placebo-controlled design examined the effects of three months subcutaneous GH alone versus placebo on body composition and some strength and functional measures in elderly patients awaiting and undergoing total hip replacement. Preoperative GH treatment resulted in improvements in lean body mass and skeletal muscle mass that were sufficient to offset postoperative losses seen in the placebo group. The data showed also that GH preserved or improved muscle strength and walking ability [35].

REFERENCES

1. International Olympic Committee. Olympic Charter, 2019. Available at: https://stillmed.olympic.org/media/Document%20Library/OlympicOrg/General/EN-Olympic-Charter.pdf (accessed: 10 March 2020).
2. Ridenour TA, Hall DL, Bost JE. A small sample randomized clinical trial methodology using N-of-1 designs and mixed model analysis. *Am J Drug Alcohol Abuse*, 2009; **35**: 260–6.
3. Görgens C, Guddat S, Dib J, Geyer H, Schänzer W, Thevis M. Mildronate (Meldonium) in professional sports—monitoring doping control urine samples using hydrophilic interaction liquid chromatography—high resolution/high accuracy mass spectrometry. *Drug Test Anal*, 2015; **7**: 973–9.
4. Maria Sharopova. Available at: https://en.wikipedia.org/wiki/Maria_Sharapova (accessed: 10 March 2020).
5. Brown-Séquard FRS. Note on the effects produced on man by subcutaneous injections of a liquid obtained from the testicles of animals. *Lancet*, 1889; **134**: 105–7.
6. The Nobel Prize. The Nobel Prize in Chemistry 1939. Available at: https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1939/ (accessed: 10 March 2020).
7. Reinold M. The myth of the Nazi steroid. *The International Journal of the History of Sport*, 2014; **31**: 871–83
8. Kremenik M, Onodera S, Nagao M, Yuzuki O, Yonetani S. A historical timeline of doping in the Olympics (part I 1896–1968). *Kawasaki J Med Welf*, 2006; **12**: 19–28.
9. WADA. Prohibited substances list. Available at: https://www.wada-ama.org/en/content/what-is-prohibited (accessed: 10 March 2020).
10. Franke WW, Berendonk B. Hormonal doping and androgenization of athletes: a secret program of the German Democratic Republic government. *Clin Chem*, 1997; **43**: 1262–79.
11. Russia banned from Winter Olympics over state-sponsored doping. 2017. Available at: https://www.theguardian.com/sport/2017/dec/05/russian-olympic-committee-banned-winter-games-doping (accessed: 10 March 2020).
12. Bhasin S, Woodhouse L, Casaburi R, *et al.* Testosterone dose-response relationships in healthy young men. *Am J Physiol Endocrinol Metab*, 2001; **281**: E1172–81.
13. Sonksen P. Determination and regulation of body composition in elite athletes. *BrJSports Med*, 2016; **52**: 219–29.
14. Thevis M, Geyer L, Geyer H, *et al.* Adverse analytical findings with clenbuterol among U-17 soccer players attributed to food contamination issues. Drug Test Anal, 2013; **5**: 372–6.
15. Katrin Krabbe. Available at: https://en.wikipedia.org/wiki/Katrin_Krabbe (accessed: 10 March 2020).
16. Jorgensen JO, Pedersen SA, Thuesen L, *et al.* Beneficial effects of growth hormone treatment in GH-deficient adults. *Lancet*, 1989; **1**: 1221–5.

17. Salomon F, Cuneo RC, Hesp R, Sonksen PH. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. *N Engl J Med*, 1989; **321**; 1797–803.
18. Online articles: Ben Johnson, Seoul Olympic Games 1988 when he beat his arch rival Carl Lewis. Available at: http://articles.chicagotribune.com/1989-06-13/sports/8902090266_1_dianabol-hormones-injections; https://www.nytimes.com/1989/03/02/sports/track-and-field-coach-tells-a-canadian-inquiry-johnson-used-steroids-since-81.html (accessed: 10 March 2020).
19. We cannot ban Chinese swimmers—Fina. Available at: http://news.bbc.co.uk/1/hi/sport/47618.stm (accessed: 10 March 2020).
20. Sonksen PH. Insulin, growth hormone and sport. *J Endocrinol*, 2001; **170**: 13–25.
21. Powrie JK, Bassett EE, Rosen T, *et al.* Detection of growth hormone abuse in sport. *Growth Horm IGF Res*, 2007; **17**: 220–6.
22. Holt RIGI, Erotokritou-Mulligan I, Sonksen PH, Sönksen PH. The history of doping and growth hormone abuse in sport. *Growth Horm IGF Res*, 2009; **19**: 320–6.
23. Widdowson WM, Healy M-L, Sönksen PH, Gibney J. The physiology of growth hormone and sport. *Growth Horm IGF Res*, 2009; **19**: 308–19.
24. Umpleby AM, Russell-Jones DL. The hormonal control of protein metabolism. *Baillieres Clin Endocrinol Metab*, 1996; **10**: 551–70.
25. Guha N, Nevitt SP, Francis M, *et al.* The effects of recombinant human insulin-like growth factor-I/insulin-like growth factor binding protein-3 administration on body composition and physical fitness in recreational athletes. *J Clin Endocrinol Metab*, 2015; **100**: 3126–31.
26. Umpleby AM, Sonksen PH. The chalonic action of insulin in man. In: Garrow JS, Halliday D, eds. *Substrate and Energy Metabolism in Man*. London: John Libby & Co, 1985: 169–78.
27. Marin P, Hijgh-Kristiansen I, Jansson S, Krotkiewski M, Holm G. Uptake of glucose carbon in muscle glycogen and adipose tissue triglycerides *in vivo* in humans. *Am J Physiol*, 1992; **263**: E473–80.
28. de la Chapelle A, Traskelin AL, Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. *Proc Natl Acad Sci U S A*, 1993; **90**: 4495–9.
29. Eero Mäntyranta. Available at: http://www.olympic.org/eero-mantyranta (accessed: 10 March 2020).
30. Lundby C, Robach P, Saltin B. The evolving science of detection of 'blood doping'. *Br J Pharmacol*, 2012; **165**: 1306–15.
31. Widdowson WM, Healy M-L, Sönksen PH, Gibney J. The physiology of growth hormone and sport. *Growth Horm IGF Res*, 2009; **19**: 308–19.
32. Meinhardt U, Nelson AE, Hansen JL, *et al.* The effects of growth hormone on body composition and physical performance in recreational athletes: a randomized trial. *Ann Intern Med*, 2010; **152**: 568–77.
33. Graham MR, Baker JS, Evans P, *et al.* Physical effects of short-term recombinant human growth hormone administration in abstinent steroid dependency. *Horm Res*, 2008; **69**: 343–54.
34. Giannoulis MG, Sonksen PH, Umpleby M, *et al.* The effects of growth hormone and/or testosterone in healthy elderly men: a randomized controlled trial. *J Clin Endocrinol Metab*, 2006; **91**: 477–84.
35. Weissberger AJ, Anastasiadis AD, Sturgess I, Martin FC, Smith MA, Sönksen PH. Recombinant human growth hormone treatment in elderly patients undergoing elective total hip replacement. *Clin Endocrinol (Oxf)*, 2003; **58**: 99–107.

12.2.3 Effect of Opioids on Adrenal and Reproductive Endocrinology

Eleni Armeni, Ashley B. Grossman, and Bernard Khoo

History of Opioids

Derived from the ancient Greek word *ὀπός*, which can be translated as vegetable juice, opium refers to the dry latex derived from the Opium Poppy (*Papaver somniferum*). References to the use of the Opium Poppy because of its medicinal properties can be found in ancient Greek, Egyptian, early Roman as well as Indian and Chinese texts [1, 2]. Widely available nowadays, opioids represent a class of drugs consisting of opiates (opioids derived from *P. somniferum)* like opium, morphine, diamorphine, codeine and pethidine, and synthetic/semi-synthetic opioids like methadone, fentanyl, tramadol, buprenorphine, and oxycodone [3].

Endogenous Opioids

Endogenous opioids are neuropeptides produced in the brain which interact with opioid receptors. Opioid receptors can be classified as follows.

- **μ-receptors (mu):** These are localized in the brainstem and medial thalamus. They are further divided into μ1-receptors, which regulate analgesia, and μ2-receptors, which regulate euphoria, pruritus, sedation, vomiting, respiratory depression, anorexia, and physical dependence.
- **κ-receptors (kappa):** These are localized in the limbic and other diencephalic areas, brain stem, and spinal cord. They are largely involved in the regulation of pain in women as well as the regulation of POMC secretion in the arcuate nucleus [4] and the gonadotropin-releasing hormone (GnRH) pulsatility clock [5].
- **δ-receptors (delta):** These receptors are found in highest densities in the olfactory bulb, neocortex, nucleus accumbens, and the caudate putamen in the brain, as well as spinal sites where they are associated with the regulation of spinal analgesia.

- **Nociceptin/orphanin FQ receptor (N/OFQ receptor):** This is a receptor which shares 60% homology with the classical opioid receptors and is the most recently discovered member of the opioid receptor family. It is widely distributed in the central nervous system (CNS), especially in the amygdala, subthalamic nucleus, hypothalamus, substantia nigra, and thalamus. The N/OFQ receptor appears to have a pro-nociceptive effect and also regulates immune function [6].

The main subtypes of endogenous opioids are endorphins, enkephalins, and dynorphins, endomorphins, and nociceptin/orphanin FQ. Endorphins are derived from the prohormone pro-opiomelanocortin (POMC), as shown in Figure 12.2.3.1: β-endorphin is involved in pain control via μ-receptors and to a lesser extent via **δ-receptors**. Other products of POMC processing include adrenocorticotrophic hormone (ACTH) which acts on the adrenal cortex inducing steroidogenesis and cortisol release; α-melanocyte-stimulating hormone (MSH) exerts its effect on melanocytes inducing pigmentation, on immune/inflammatory cells resulting in anti-inflammatory and antipyretic effects as well as on the central nervous system regulating food intake, energy homeostasis, and erectile activity. γ-MSH exerts an anti-inflammatory response on macrophages and regulates the autonomic nervous system.

Enkephalin-related peptides are derived from proenkephalin (PENK), which produces one copy of leu-enkephalin, six copies of met-enkephalin, an octapeptide and a heptapeptide [7, 8]. The *PENK* gene is expressed mainly in the posterior pituitary, in neuronal axon terminals as well as in neuronal and non-neuronal system tissues such as the gastrointestinal, the visual systems, and the placenta [7]. Enkephalins bind and activate μ-receptors and δ-receptors.

Dynorphin peptides are potent endogenous ligands, for example β-neoendorphin, dynorphin, leu-enkephalin, rimorphin, and leumorphin, which are derived from the prodynorphin peptide (PDYN). Dynorphins are highly selective for the κ-receptor, inducing analgesia, psychomimesis, dysphoria, and diuresis as well as psychostimulant effects [9].

The pronociceptin peptide (PNOC) is processed into N/OFQ is the endogenous ligand for the N/OFQ receptor. N/OFQ has a pro-nociceptive/antiopioid effect and regulates immune function. N/OFQ also stimulates ACTH and cortisol release, and enhances the effects of stress on these hormones [6, 10].

Endomorphin-1 and -2 are endogenous opioid peptides derived from a recently discovered propeptide, proMexneurin [11]. The endomorphins have a potent antinociceptive effect on acute and neuropathic pain *in vivo* via μ-receptors.

Exogenous Opioids

As noted earlier, opioids have been used for thousands of years as first-line drugs to manage pain of various severity. Commonly prescribed opioids include codeine, fentanyl, hydrocodone, hydromorphone, methadone, morphine, oxycodone, oxymorphone, and tapentadol. These opioids can be classified according to their duration of action, as follows: (a) rapid onset, like fentanyl; (b) short-acting, like codeine, buprenorphine, oxymorphone,

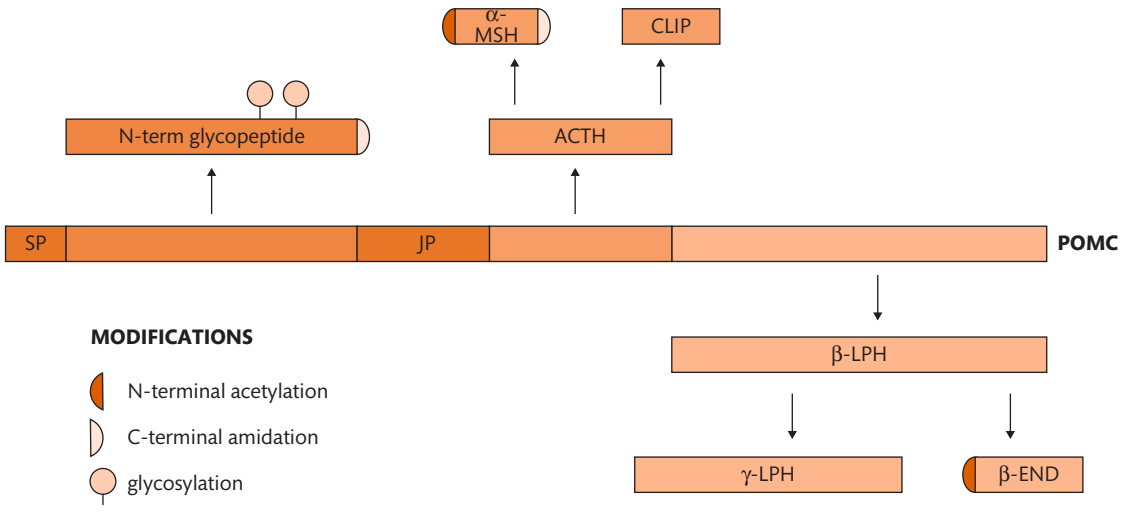


Figure 12.2.3.1 Pro-opiomelanocortin gene expression and post-translational modification.

Reproduced with permission from Lim CT, Khoo B. Normal Physiology of ACTH and GH Release in the Hypothalamus and Anterior Pituitary in Man. In: De Groot LJ, Chrousos GP, Dungan K, Feingold KR, Grossman A, Hershman JM, Koch C, Korbonits M, McLachlan R, New M, Purnell J, Rebar R, Singer F, Vinik A, eds. Endotext: MDText.com, Inc.; 2017.

oxycodone, hydrocodone and morphine; (c) long-acting, including transdermal fentanyl, methadone, extended-release oxycodone, extended-release oxymorphone or extended-release morphine [12, 13]. Their effects are exerted through activation of the opioid receptors either as agonists or partial agonists. Current efforts are underway to develop analogues with selective actions that have enhanced analgesic activity and reduced effects on the respiratory system. Strategies currently being used include exploiting biased signalling, targeting multiple or heterodimerized opioid receptors, modifying drug availability to receptors according to pH, and allosteric enhancement of endogenous opioid activity [14]. At present it is unclear how much the desirable effects of opioids can be disentangled from their potential for addiction using these strategies.

The possibility of substance abuse and dependency should always be considered during opioid treatment. Prior to initiation of opioid treatment, clinicians should carefully estimate the lowest effective dose and uptitrate the dose gradually. Appropriate justification regarding possible benefits and risks is necessary when administering doses equal or higher than 50 mg of morphine equivalents per day. Doses higher than 90 mg of morphine equivalents per day should be avoided. Immediate-release rather than long-acting opioids should be preferred when initiating opioid therapy for chronic pain [15].

The overreliance on the use of prescription opioids for the management of acute and chronic pain has reached epidemic proportions. Where patients are unable to obtain legal supplies of prescription opioids such as oxycodone there is substitution with illegally obtained opioids, such as fentanyl, which leads to increased risks of overdose and continued addiction [16]. The risk of addiction is dose-dependent and associated with chronic therapy (more than 3 months). This risk has been estimated as 5.5% in patients treated with a very low dose of non-prescribed opioid (<20 morphine equivalents daily) [17]. In addition, chronic treatment with methadone and buprenorphine is licensed for opiate use disorder; these are effective at reducing mortality and facilitating recovery, but entail the continuing side effects of opioid therapy [18]. Endocrine side effects of opioids are therefore increasingly encountered by physicians. The following summarizes the effects of opioids on two clinically important endocrine systems. For more information regarding the effects of opioids on other endocrine systems, see [19].

Effects of Opioids on the Hypothalamo–Pituitary–Adrenal (HPA) Axis

The HPA axis is a key pathway for the regulation of the body's response to stress. Endogenous opioids represent another such stress response pathway with a role in modulating pain [20], and also act to modulate the HPA axis. It is therefore not surprising that exogenous opioid treatment modulates the HPA axis. As noted earlier, POMC is processed to ACTH and β-endorphin, which inhibits POMC/ACTH release. The endogenous opioids β-endorphin and dynorphin are thought to regulate the POMC neurons via postsynaptic μ-receptors and presynaptic μ-, κ-, and δ-receptors [21]. Similarly, activation of opioid receptors by exogenously administered opioids inhibits POMC neurons postsynaptically and decreases the biosynthesis of POMC [22]. Consequently, there is a suboptimal ACTH and cortisol response to a corticotrophin-releasing hormone (CRH) stimulus in volunteers given morphine [23], but they may also suppress the hypothalamic release of the other ACTH-releasing factor vasopressin. Opioids also directly act on the adrenal gland as judged by inhibition of dehydroepiandrosterone sulphate (DHEAS) production [24, 25].

Acute administration of opioids suppresses ACTH, β-endorphin, and cortisol levels at baseline and secretion in response to CRH [23]. Patients under chronic treatment with both oral or transdermal opioid analgesics present with various degrees of cortisol deficiency [26–28]. The prevalence of secondary adrenal insufficiency in patients treated with opioids has not been systematically evaluated. A small cross-sectional analysis evaluated the potential association between long-term opioid analgesia and the HPA axis, confirming low cortisol levels and suboptimal response to synthetic ACTH(1–24) (*Synacthen*) in up to 6.25% of treated cases [29]. Similarly, adrenal insufficiency is estimated to occur in approximately 15% of patients treated with chronic intrathecal opioid administration [30].

In addition to the inhibitory effect on cortisol production, exogenous opioid treatment alters the circadian rhythm of the HPA axis, and appears to have a 'flattening' effect with more marked inhibition of morning cortisol levels relative to afternoon cortisol levels [31], However, while opioid peptides significantly inhibit the HPA axis [32], the endogenous 'tone' is not responsible for the circadian rhythmicity of ACTH and cortisol. The high doses of naloxone required to demonstrate endogenous opioid tone suggest the presence of non-μ receptors [33]. Prenatal exposure to opioids results in altered regulation of the HPA axis in offspring, according to animal studies, and more specifically: (a) changes in the baseline secretion of ACTH; (b) altered ACTH secretion by the hypothalamus following physiological and psychological stressors; (c) alterations to the activation of the glucocorticoid receptor [34]. The long-term impact of any alterations to HPA axis function in children born to opiate users is unknown.

The impact of opioids on the HPA axis seems to be reversible, with small studies describing increases in cortisol levels following successful withdrawal from opioid treatment. However, further longitudinal studies are required to determine the time course response of the HPA axis following the withdrawal of opioid administration [35].

No agreed clinical guidelines for the assessment and treatment of hypoadrenalism in opioid users exist. The following recommendations are therefore based on the authors' clinical experience and allied guidance on hypoadrenalism.

1. The clinical manifestations of hypoadrenalism can be non-specific (e.g. tiredness, weight loss, nausea, anorexia, hypotension). Therefore, adrenal function should be routinely evaluated in patients chronically treated with opioids, using 9 am cortisol levels, with repeated blood tests every 3–4 months.
2. 9:00 a.m. serum cortisol levels less than 100 nmol/L are strongly suggestive of adrenal insufficiency and in this case, replacement should be initiated. A 9:00 a.m. cortisol level higher than 380 nmol/L generally indicates normal adrenal function and does not require further investigation, but note that thresholds may be assay-dependent [36].
3. The cortisol response to *Synacthen*, according to locally established cut-offs, can be used to establish the diagnosis of hypoadrenalism in cases where the 9:00 a.m. cortisol level is

between 100 and 380 nmol/L. Insulin tolerance or glucagon stimulation tests are rarely required.

4. The specific characteristics of local cortisol assays [36], concomitant use of exogenous glucocorticoids (e.g. those used in management of chronic back pain), or presence of conditions affecting cortisol binding globulin such as oral oestrogen therapy or pregnancy, should be considered when interpreting the results.
5. Glucocorticoid replacement can be provided as short-acting corticosteroids (e.g. immediate-release hydrocortisone in 2–3 divided doses) or as longer acting agents (e.g. prednisolone given once a day) [37].
6. If the patient presents with clinical signs of acute hypoadrenalism, parenteral hydrocortisone should be administered in accordance with guidance on the acute management of hypoadrenalism [38].
7. The response of the HPA axis to withdrawal of opiates can be monitored by regular 9 am serum cortisol levels or short *Synacthen* tests as needed, with withdrawal of glucocorticoid support when euadrenalism is established.

Influence of Opioids on the Function of the Hypothalamo–Pituitary–Gonadal (HPG) and Prolactin Axes

The adverse impact of opiates on gonadal function has long been known. Referring to opium, Charles Alexander Bruce, a British explorer, described the effects on people in India as follows in 1843: 'This vile drug has kept, and does now keep down the population: the women have fewer children than those of other countries . . . the enfeebled opium-eaters of Assam, who are more effeminate than women' [39]. Hypogonadism is present in 21% of female and 75% of male patients on long-term treatment with opioids [40].

The endogenous opioid system influences the function of the HPG axis through interactions with hypothalamic and pituitary hormones, as well as directly on the gonads. One key mechanism involves suppression of pulsatile GnRH secretion during times of physiological stress, specifically via POMC processing, β-endorphin release, and activation of μ-opioid receptors. A second mechanism involves the termination of GnRH pulses via dynorphin, κ-receptors, and the so-called KNDy neurons in the arcuate nucleus which secrete kisspeptin [5]. A third, more indirect, mechanism involves inhibition of dopaminergic pathways, leading to de-repression of prolactin secretion [41] which in turn leads to a suppression of gonadotropin secretion [42].

With respect to male patients, acute administration of opioids leads to suppression of luteinizing hormone (LH) secretion, demonstrating their tonic inhibitory effect on GnRH pulsatility [43]. Naloxone antagonism leads to stimulation of LH and testosterone secretion [44]. The opiate effect on pituitary hormone secretion is reversible following administration of opioid antagonists. Opioid-induced androgen deficiency is diagnosed as low levels of testosterone in patients on treatment with opioids. Long-term opioid treatment reduces testosterone [45], a dose-dependent effect that is not influenced by the type of administered opioids (i.e. commonly prescribed opioids, heroin, or methadone) [46]. The impact of the dose was particularly pronounced in men treated with short-acting opioids. A longer duration of opioid action increases the risk of androgen deficiency more than 3-fold when compared with administration of short-acting agents [47]. With respect to erectile function, there are direct inhibitory effects, as the occurrence of erectile dysfunction is independent of serum testosterone levels in people taking opioids [48]. It is thought that central μ-receptors inhibit the oxytocinergic neurons in the paraventricular nucleus of the hypothalamus, directly inhibiting penile erection [49].

Endogenous opioid tone varies during the menstrual cycle, due to phase-dependent β-endorphin tone, opioid receptor expression/binding affinities, and the key influences of oestrogen and progesterone on endogenous opioid tone itself. Indeed, endogenous opioid tone plays a key role in the transition between stages of the menstrual cycle. In the early follicular phase, the endogenous opioid tone is low, enabling high-frequency pulsatile GnRH secretion. As oestradiol levels rise in the mid-follicular phase, leading to increased endogenous opioid tone and less frequent GnRH pulsatility, in turn leading to increased LH pulse amplitude which prepares the way for the ovulatory LH 'spike'. Luteal secretion of progesterone induces a high endogenous opioid tone, consequently reducing the frequency of GnRH and LH pulses (once every 5–6 hours instead of once an hour during the follicular phase). With menses, progesterone and oestradiol levels fall, endogenous opioid activity declines, leading to a restart of high-frequency pulsatile GnRH secretion in the follicular phase [50]. The effect of opioids (and opioid antagonists) on HPG axis function in animal models is blunted in prepubertal animals, suggesting that the effect of opioids on the HPG axis may differ according to the stage of sexual development [19].

Stress conditions (e.g. excessive exercise, weight loss, anorexia nervosa) have been speculated to increase endogenous opioid tone which in turn suppresses GnRH pulsatility, gonadotrophin, and oestradiol secretion. One hypothesis is that this drives the 'functional hypothalamic amenorrhoea' that is observed in such patients. However, the success of opioid antagonism as a therapy for this condition is variable: in one study, naloxone blockade was not successful in restoring cycles in patients with weight-related amenorrhoea [51], but in other hands naltrexone restored cycles in approximately 50% of patients [52]. It seems that weight-related amenorrhoea, and presumably the related exercise-induced amenorrhoea, are not opioid-dependent as they are naloxone insensitive, at least in short-term studies, although hyperprolactinaemic amenorrhoea may be opioid-mediated [51].

Interestingly, dysregulation of the endogenous opioid system appears to have a key role in polycystic ovarian syndrome (PCOS). Increased endogenous opioid activity has been linked with a reduction in GnRH pulsatility and elevated LH/follicle-stimulating hormone (FSH) ratios. Moreover, enhanced opioid activity stimulates pancreatic insulin secretion, inhibits insulin clearance by the liver and decreases peripheral insulin sensitivity, hence contributing to the hyperinsulinaemia observed in PCOS. This hyperinsulinaemia in turn synergizes with LH to induce hyperandrogenism, and also reduces hepatic sex-hormone-binding globulin (SHBG) production leading to elevated free androgens and the hyperandrogenic manifestations of PCOS. Elevated endogenous opioid tone is also conjectured to drive other issues seen in PCOS such as depression and poor sleep quality. Administration of opioid antagonists, like

naloxone and naltrexone, may help reverse the hormonal imbalances observed in PCOS cases [42, 53].

As a result of this complexity of regulation, the effects of opioid treatment in women can vary largely throughout the reproductive life of the woman as well as within the menstrual cycle. Exogenous opioid peptides suppress LH and FSH secretion when given acutely to normal and postmenopausal women [54]. Exogenous opioids have been shown to be associated with a reduced FSH/LH ratio, and with a lower age threshold of FSH/LH ratio inversion (a marker of menopause) indicating a shorter duration of fertility [55]. Naloxone increases LH pulse frequency in women; unlike the regulation of the HPA axis, as this is naloxone-sensitive it is presumably μ-receptors that mediate the effect [56]. Administration of the opioid antagonist naltrexone has positive effects on both the HPG and HPA axes, by inducing elevation of LH levels as well as levels of ACTH and cortisol, respectively [42].

Endogenous opioids are also known to directly influence oocyte maturation, preimplantation embryo development, endometrial, and myometrial function. β-endorphin and met-enkephalin are present within ovarian follicular fluid and influence oocyte maturation in a paracrine fashion by binding to δ-, κ-, and μ-receptors. Granulosa cells also express μ-receptors which is thought to stimulate angiogenesis. Endometrial fluid contains β-endorphin and met-enkephalin produced presumably by endometrial tissues. μ-receptors in the endometrial tissues are known to regulate oestradiol receptor expression. At least in animal models, morphine can inhibit the development of preimplantation embryos and reduce uterine receptivity to implantation. Opioids also inhibit/relax myometrial contraction, which is important for sperm transport in the uterus and Fallopian tubes [42]. However, the exact impact of exogenous opioids on oocyte maturation, fertilization, embryonic development, and implantation in humans is as yet unknown.

Finally, animal studies report adverse effects on the reproductive system of newborns following intrauterine exposure to opioids. Prenatal exposure to an opioid environment induces pre- and postnatal alterations of the hormonal milieu, and potentially changes of sexually dimorphic brain regions regulating reproductive and play behaviour [34].

No agreed clinical guidelines for the assessment and treatment of hypogonadism in opioid users exist. The following recommendations are therefore based on the authors' clinical experience and allied guidance on hypogonadism.

1. Physicians treating patients who are using opioids chronically should routinely enquire after symptoms and signs of potential hypogonadism during clinical reviews. For male patients, diminished libido, decreased frequency of morning erections and erectile dysfunction are the most specific symptoms [57]. For female patients, oligo- or amenorrhoea together with symptoms of hypo-oestrogenization (hot flushes, sleep disturbance, dyspareunia) are the most specific symptoms [58].
2. If hypogonadism in a male patient is suspected, evaluate with two separate fasting measurements of 9am testosterone with LH/FSH and prolactin levels. Determination of free testosterone by equilibrium dialysis where available or by Vermeulen calculation utilizing simultaneous measurements of SHBG and albumin may be helpful in the assessment of cases where patients have extremes of SHBG concentration, such as in obesity (which lowers SHBG), or with liver disease (which tends to increase SHBG) [57].
3. If hypogonadism in a female patient is suspected, evaluate with oestradiol, LH/FSH and prolactin levels. Human chorionic gonadotropin (HCG) testing is mandatory to exclude pregnancy.
4. Consider pituitary imaging with MRI to exclude a substantial hypothalamic/pituitary lesion if the blood tests indicate a secondary hypogonadism or hyperprolactinaemia [59].
5. If opioid-induced hypogonadism is confirmed, hormone replacement with testosterone [57] or oestrogen ± progestogen [59] should be considered.
6. Patients with long-standing hypogonadism should be assessed for osteopenia/osteoporosis with dual-energy X-ray absorptiometry (DEXA) and treated accordingly [59].
7. We recommend that physicians collaborate with pain control and addiction specialists to withdraw, if possible, patients from chronic opioid use prior to attempts at conception and/or fertility treatment to mitigate their effects on fertility, conception, and prenatal development, as well as on the HPA axis.

Effects of Opioids on the Posterior Pituitary and Sympathoadrenal System

Exogenous opioid peptides, contrary to earlier expectation, inhibit the release of vasopressin in the human [60], in agreement with animal studies showing a direct inhibitory effect on release from the hypothalamus. The significance of this induced free water diuresis in chronic opiate users remains unclear but seems unlikely to have a major impact. Similarly, opioids inhibit the release of adrenaline and noradrenaline from the sympathoadrenomedullary system, probably at κ- or δ-receptors. There appears to be an endogenous opioid tone which chronically inhibits these systems and which is more active during any form of physical or mental stress-induced activation [61–63]. Again, the significance of this in terms of chronic opioid use is undetermined.

Conclusions

Endogenous opioid pathways are fundamental to the regulation of many endocrine axes, the most clinically important of which are the HPA and HPG axes. Chronic opioid users are at high risk of developing hypoadrenalism and hypogonadism, but even short-term use may lead to suppression of the HPA and HPG axes and cause diagnostic confusion [64, 65]; therefore, physicians and other health workers involved in their treatment should ensure that a robust screening and treatment protocol is in place.

REFERENCES

1. Brook K, Bennett J, Desai SP. The chemical history of morphine: an 8000-year journey, from resin to de-novo synthesis. *J Anesth Hist*, 2017; **3**: 50–5.

2. Stefano GB, Pilonis N, Ptacek R, Kream RM. Reciprocal Evolution of Opiate Science from Medical and Cultural Perspectives. *Med Sci Monit*, 2017; **23**: 2890–6.
3. Vallejo R, Barkin RL, Wang VC. Pharmacology of opioids in the treatment of chronic pain syndromes. *Pain Physician*, 2011; **14**: E343–60.
4. Zhang X, van den Pol AN. Direct inhibition of arcuate proopiomelanocortin neurons: a potential mechanism for the orexigenic actions of dynorphin. *J Physiol*, 2013; **591**: 1731–47.
5. Weems PW, Lehman MN, Coolen LM, Goodman RL. The roles of neurokinins and endogenous opioid peptides in control of pulsatile LH secretion. In: Litwack G, ed. *Vitamins and Hormones*, Vol. **107**. San Diego, CA: Academic Press, 2018: 89–135.
6. Meunier JC. Nociceptin/orphanin FQ and the opioid receptor-like ORL1 receptor. *Eur J Pharmacol*, 1997; **340**: 1–15.
7. Zagon IS, Wu Y, McLaughlin PJ. Opioid growth factor and organ development in rat and human embryos. *Brain Res*, 1999; **839**: 313–22.
8. Noda M, Teranishi Y, Takahashi H, *et al.* Isolation and structural organization of the human preproenkephalin gene. *Nature*, 1982; **297**: 431–4.
9. Liu-Chen LY. Agonist-induced regulation and trafficking of kappa opioid receptors. *Life Sci*, 2004; **75**: 511–36.
10. Devine DP, Watson SJ, Akil H. Nociceptin/orphanin FQ regulates neuroendocrine function of the limbic-hypothalamic-pituitary-adrenal axis. *Neuroscience*, 2001; **102**: 541–53.
11. Matus-Ortega ME, Leff Gelman P, Calva-Nieves JC, *et al.* Mexneurin is a novel precursor of peptides in the central nervous system of rodents. *FEBS Lett*, 2017; **591**: 1627–36.
12. Yaksh T, Wallace M. Opioids, analgesia and pain management. Chapter 18. In: Chabner B, Knollmann B, eds. *The Parmacological Basis of Therapeutics*, 12th edn. New York: McGraw-Hill, 2011.
13. Lacy CF. *Drug Information Handbook*, 21st edn. Hudson, OH: Lexi-Comp, 2012.
14. Darcq E, Kieffer BL. Opioid receptors: drivers to addiction? *Nat Rev Neurosci*, 2018; **19**: 499–514.
15. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain—United States, 2016. *JAMA*, 2016; **315**: 1624–45.
16. Frank RG, Pollack HA. Addressing the fentanyl threat to public health. *N Engl J Med*, 2017; **376**: 605–7.
17. Busse JW, Craigie S, Juurlink DN, *et al.* Guideline for opioid therapy and chronic noncancer pain. *Can Med Assoc J*, 2017; **189**: E659–66.
18. Sordo L, Barrio G, Bravo MJ, *et al.* Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies. *BMJ*, 2017; **357**: j1550.
19. Vuong C, Van Uum SH, O'Dell LE, Lutfy K, Friedman TC. The effects of opioids and opioid analogs on animal and human endocrine systems. *Endocr Rev*, 2010; **31**: 98–132.
20. Ferdousi M, Finn DP. Stress-induced modulation of pain: role of the endogenous opioid system. *Prog Brain Res*, 2018; **239**: 121–77.
21. Pennock RL, Hentges ST. Differential expression and sensitivity of pre- and postsynaptic opioid receptors regulating hypothalamic proopiomelanocortin neurons. *J Neurosci*, 2011; **31**: 281–8
22. Garcia de Yebenes E, Pelletier G. Opioid regulation of proopiomelanocortin (POMC) gene expression in the rat brain as studied by in situ hybridization. *Neuropeptides*, 1993; **25**: 91–4
23. Allolio B, Schulte HM, Deuβ U, Kallabis D, Hamel E, Winkelmann W. Effect of oral morphine and naloxone on pituitary-adrenal response in man induced by human corticotropin-releasing hormone. *Acta Endocrinologica*, 1987; **114**: 509–14.
24. Rittmaster RS, Cutler GB, Jr., Sobel DO, *et al.* Morphine inhibits the pituitary-adrenal response to ovine corticotropin-releasing hormone in normal subjects. J Clin Endocrinol Metab, 1985; **60**: 891–5.
25. Daniell HW. DHEAS deficiency during consumption of sustained-action prescribed opioids: evidence for opioid-induced inhibition of adrenal androgen production. *J Pain*, 2006; **7**: 901–7.
26. Mussig K, Knaus-Dittmann D, Schmidt H, Morike K, Haring HU. Secondary adrenal failure and secondary amenorrhoea following hydromorphone treatment. *Clin Endocrinol (Oxf)*, 2007; **66**: 604–5.
27. Oltmanns KM, Fehm HL, Peters A. Chronic fentanyl application induces adrenocortical insufficiency. *J Intern Med*, 2005; **257**: 478–80.
28. Rhodin A, Stridsberg M, Gordh T. Opioid endocrinopathy: a clinical problem in patients with chronic pain and long-term oral opioid treatment. *Clin J Pain*, 2010; **26**: 374–80.
29. Gibb FW, Stewart A, Walker BR, Strachan MW. Adrenal insufficiency in patients on long-term opioid analgesia. *Clin Endocrinol*, 2016; **85**: 831–5.
30. Abs R, Verhelst J, Maeyaert J, *et al.* Endocrine consequences of long-term intrathecal administration of opioids. *J Clin Endocrinol Metab*, 2000; **85**: 2215–22.
31. Facchinetti F, Volpe A, Farci G, *et al.* Hypothalamus-pituitary-adrenal axis of heroin addicts. *Drug Alcohol Depend*, 1985; **15**: 361–6.
32. Gaillard RC, Grossman A, Smith R, Rees LH, Besser GM. The effects of a met-enkephalin analogue on ACTH, beta-LPH, beta-endorphin and MET-enkephalin in patients with adrenocortical disease. *Clin Endocrinol (Oxf)*, 1981; **14**: 471–8.
33. Grossman A, Gaillard RC, McCartney P, Rees LH, Besser GM. Opiate modulation of the pituitary-adrenal axis: effects of stress and circadian rhythm. *Clin Endocrinol (Oxf)*, 1982; **17**: 279–86.
34. Byrnes EM, Vassoler FM. Modeling prenatal opioid exposure in animals: current findings and future directions. *Front Neuroendocrinol*, 2017; **28**: 30058–4.
35. Bunce SC, Harris JD, Bixler EO, *et al.* Possible evidence for re-regulation of HPA axis and brain reward systems over time in treatment in prescription opioid-dependent patients. *J Addict Med*, 2015; **9**: 53–60.
36. Sbardella E, Isidori AM, Woods CP, *et al.* Baseline morning cortisol level as a predictor of pituitary–adrenal reserve: a comparison across three assays. *Clin Endocrinol (Oxf)*, 2017; **86**: 177–84.
37. Pazderska A, Pearce SH. Adrenal insufficiency—recognition and management. *Clin Med (Lond)*, 2017; **17**: 258–62.
38. Arlt W, the Society for Endocrinology Clinical Committee. Society for Endocrinology Endocrine Emergency Guidance: emergency management of acute adrenal insufficiency (adrenal crisis) in adult patients. *Endocr Connect*, 2016; **5**: G1–G3.
39. Suppression of the opium trade. *HC Debates*, 1843: **68**: cc362–469.
40. Fraser LA, Morrison D, Morley-Forster P, *et al.* Oral opioids for chronic non-cancer pain: higher prevalence of hypogonadism in men than in women. *Exp Clin Endocrinol Diabetes*, 2009; **117**: 38–43.

41. Delitala G, Grossman A, Besser GM. The participation of hypothalamic dopamine in morphine-induced prolactin release in man. *Clin Endocrinol (Oxf)*, 1983; **19**: 437–44.
42. Böttcher B, Seeber B, Leyendecker G, Wildt L. Impact of the opioid system on the reproductive axis. *Fertil Steril*, 2017; **108**: 207–13.
43. Delitala G, Grossman A, Besser M. Differential effects of opiate peptides and alkaloids on anterior pituitary hormone secretion. *Neuroendocrinology*, 1983; **37**: 275–9.
44. Delitala G, Giusti M, Mazzocchi G, Granziera L, Tarditi W, Giordano G. Participation of endogenous opiates in regulation of the hypothalamic-pituitary-testicular axis in normal men. *J Clin Endocrinol Metab*, 1983; **57**: 1277–81.
45. Cicero TJ, Bell RD, Wiest WG, Allison JH, Polakoski K, Robins E. Function of the male sex organs in heroin and methadone users. *N Engl J Med*, 1975; **292**: 882–7.
46. Bawor M, Bami H, Dennis BB, *et al.* Testosterone suppression in opioid users: a systematic review and meta-analysis. *Drug Alcohol Depend*, 2015; **149**: 1–9.
47. Rubinstein A, Carpenter DM. Elucidating risk factors for androgen deficiency associated with daily opioid use. *Am J Med*, 2014; **127**: 1195–201.
48. Cioe PA, Friedmann PD, Stein MD. Erectile dysfunction in opioid users: lack of association with serum testosterone. *J Addict Dis*, 2010; **29**: 455–60.
49. Argiolas A, Melis MR. Central control of penile erection: role of the paraventricular nucleus of the hypothalamus. *Prog Neurobiol*, 2005; **76**: 1–21.
50. Leyendecker G, Waibeltreber S, Wildt L. The central control of follicular maturation and ovulation in the human. *Oxf Rev Reprod Biol*, 1990; **12**: 93–146.
51. Grossman A, Moult PJ, McIntyre H, *et al.* Opiate mediation of amenorrhoea in hyperprolactinaemia and in weight-loss related amenorrhoea. *Clin Endocrinol*, 1982; **17**: 379–88.
52. Leyendecker G, Waibel-Treber S, Wildt L. Pulsatile administration of gonadotrophin releasing hormone and oral administration of naltrexone in hypothalamic amenorrhoea. *Hum Reprod*, 1993; **8** Suppl 2: 184–8.
53. Eyvazzadeh AD, Pennington KP, Pop-Busui R, Sowers M, Zubieta JK, Smith YR. The role of the endogenous opioid system in polycystic ovary syndrome. *Fertil Steril*, 2009; **92**: 1–12.
54. Grossman A, Moult PJ, Gaillard RC, *et al.* The opioid control of LH and FSH release: effects of a met-enkephalin analogue and naloxone. *Clin Endocrinol (Oxf)*, 1981; **14**: 41–7.
55. Reece AS, Thomas MR, Norman A, Hulse GK. Dramatic acceleration of reproductive aging, contraction of biochemical fecundity and healthspan-lifespan implications of opioid-induced endocrinopathy-FSH/LH ratio and other interrelationships. *Reprod Toxicol*, 2016; **66**: 20–30.
56. Moult PJ, Grossman A, Evans JM, Rees LH, Besser GM. The effect of naloxone on pulsatile gonadotrophin release in normal subjects. *Clin Endocrinol (Oxf)*, 1981; **14**: 321–4.
57. Bhasin S, Brito JP, Cunningham GR, *et al.* Testosterone therapy in men with hypogonadism: an Endocrine Society Clinical practice guideline. *J Clin Endocrinol Metab*, 2018; **103**: 1715–44.
58. Rothman MS, Wierman ME. Female hypogonadism: evaluation of the hypothalamic–pituitary–ovarian axis. *Pituitary*, 2008; **11**: 163.
59. Gordon CM, Ackerman KE, Berga SL, *et al.* Functional hypothalamic amenorrhea: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 1413–39.
60. Grossman A, Besser GM, Milles JJ, Baylis PH. Inhibition of vasopressin release in man by an opiate peptide. *Lancet*, 1980; **2**: 1108–10.
61. Grossman A, Bouloux P, Price P, *et al.* The role of opioid peptides in the hormonal responses to acute exercise in man. *Clin Sci (Lond)*, 1984; **67**: 483–91.
62. Bouloux PM, Grossman A, Lytras N, Besser GM. Evidence for the participation of endogenous opioids in the sympathoadrenal response to hypoglycaemia in man. *Clin Endocrinol (Oxf)*, 1985; **22**: 49–56.
63. Bouloux PM, Grossman A, Al-Damluji S, Bailey T, Besser M. Enhancement of the sympathoadrenal response to the cold-pressor test by naloxone in man. *Clin Sci (Lond)*, 1985; **69**: 365–8.
64. Policola C, Stokes V, Karavitaki N, Grossman A. Adrenal insufficiency in acute oral opiate therapy. *Endocrinol Diabetes Metab Case Rep*, 2014; **2014**: 130071.
65. Lim CT, Khoo B. Normal physiology of ACTH and GH release in the hypothalamus and anterior pituitary in man. In: De Groot LJ, Chrousos GP, Dungan K, *et al.*, eds. Endotext: MDText.com, Inc., 2017.

SECTION 13

Endocrinology of Cancer

13.1

Endocrine Disorders Caused by Cancer or its Treatment

13.1.1 Metastatic Disease in Endocrine Organs

Thomas G. Papathomas and Vania Nosé

Introduction

Endocrine organs are classified into primary and secondary based on whether a hormone-secreting organ synthesizes the relevant hormone(s) as a primary function or not. This chapter will focus on metastatic disease in major endocrine organs, including adrenals, pituitary, thyroid and parathyroid glands, from malignancies other than haemato-lymphoid neoplasms. Their presence is best regarded as generalized involvement rather than metastatic spread. Herein, we discuss the prevalence and importance of recognition of metastases in endocrine organs and highlight various aspects of endocrine organ-specific metastatic disease.

Prevalence of Metastatic Disease in Major Endocrine Organs

The prevalence of metastatic malignancy varies in the clinical setting (see Table 13.1.1.1):

- common in the **adrenals** in patients with lung carcinoma
- uncommon in the **pituitary gland** in patients with breast and lung carcinomas
- rare in the **thyroid gland** in patients with renal cell and lung carcinomas
- very rare in the **parathyroid glands** in patients with breast carcinoma

Importance of Recognition of Metastatic Deposits in Endocrine Organs

Recognition of metastatic disease in endocrine organs is important for various reasons (see Figure 13.1.1.1).

Metastases might:

- be discovered as an incidentaloma in an unconfined anatomical space without any pressure effects (e.g. adrenal metastases)
- occur as a mass in an unconfined, or relatively unconfined, space producing pressure effects, e.g. on the optic chiasm causing homonymous hemianopia (e.g. pituitary metastasis)
- present as a space-occupying lesion in a relatively confined anatomical space, such as the superior mediastinum (e.g. thyroid metastases)
- present with endocrine gland insufficiency, which is usually detected at an advanced stage given the large reserve of function of endocrine organs (e.g. pituitary and adrenal metastases)
- be functional secreting a hormone or other substance relevant to the primary neoplasm (e.g. small cell carcinoma of lung origin or pancreatic neuroendocrine tumour metastases)
- cause the host tissue to secrete excessive amounts of hormone because of their presence (e.g. transient hyperthyroidism due to thyroid parenchyma destruction and hormone release)
- be mistaken clinically for a primary neoplasm of the gland in which they are found causing diagnostic and management difficulties (e.g. metastatic origin from an occult primary)
- be discovered on histopathological examination within a primary endocrine tumour (i.e. tumour-to-tumour metastasis)

Table 13.1.1.1 Prevalence of metastases in major endocrine organs

	Relatively common primary sites	Relatively rare primary sites
Adrenal glands	Lung, gastrointestinal tract	Kidney, breast, skin (melanoma)
Pituitary gland	Breast, lung	Kidney, prostate, gastrointestinal tract, skin (melanoma), thyroid
Thyroid gland	Kidney, lung	Breast, gastrointestinal tract
Parathyroid glands	Breast	Skin (melanoma), lung, kidney, thyroid

The Adrenal Gland

The adrenal gland is the fourth commonest anatomic site for metastatic deposits after lung, liver, and bone, whereas per unit weight is the most frequently involved site [1, 2].

The most common primary tumour metastatic to the adrenals arises from the **lung** [3–5]. Based on a 30-year period retrospective **systematic review** of 464 patients with adrenal metastasis at Queen Mary Hospital in Hong Kong [3], **lung cancer** (35.4%) was the most frequent followed by malignancies originating from upper gastrointestinal (GI) tract (26.4%), hepatobiliary (10.7%) and pancreas (6.9%), lower GI tract (5.4%), kidney (4.3%) and breast (2.9%). Data emerging from a **multicentric European study** including 317 patients indicated the leading sources of adrenal metastasis, as follows: **lung** (46.7%), lower GI tract (13.6%), kidney (11.7%), breast (3.5%) and skin (melanoma; 3.5%) [4]. The difference in the relative prevalence of primary tumours has been attributed to the geographic origin and/or source of data [4, 5].

In contrast to autopsy-based data [6], metastatic melanoma to the adrenal gland is a rare ante-mortem diagnosis [3, 4, 7]. This is usually associated with evidence of systemic disease and overall poor prognosis with almost one-fifth of cases being of unknown primary origin [7, 8].

Primary adrenal tumours can rarely receive metastatic deposits with **lung cancer** being the most frequent donor neoplasm in tumour-to-tumour metastatic setting [9–12].

Most adrenal metastases are identified at staging or follow-up of patients, who are either oligo-metastatic or suffering from widespread metastatic disease [1, 3, 5]. Isolated metastasis after surgical removal of a primary tumour is a rather infrequent event [3, 13]. Less than 2% of renal cancer patients who develop a metachronous metastasis will present with a solitary adrenal metastasis [14] being consistent with data deriving from the **Japanese Society of Renal Cancer** [15]. Albeit, cases of isolated metastasis have been increasingly reported in the medical literature, highlighting the unpredictable behaviour of diverse tumour types with a variable propensity for metastatic spread. When bilateral and extensively effacing the adrenal parenchyma, latent or overt adrenocortical insufficiency can be detected. Concomitant involvement of pituitary gland and unilateral adrenal destruction can also result in secondary insufficiency, while abdominal pain is another rare presentation due to intra-abdominal haemorrhage and haemoperitoneum [3].

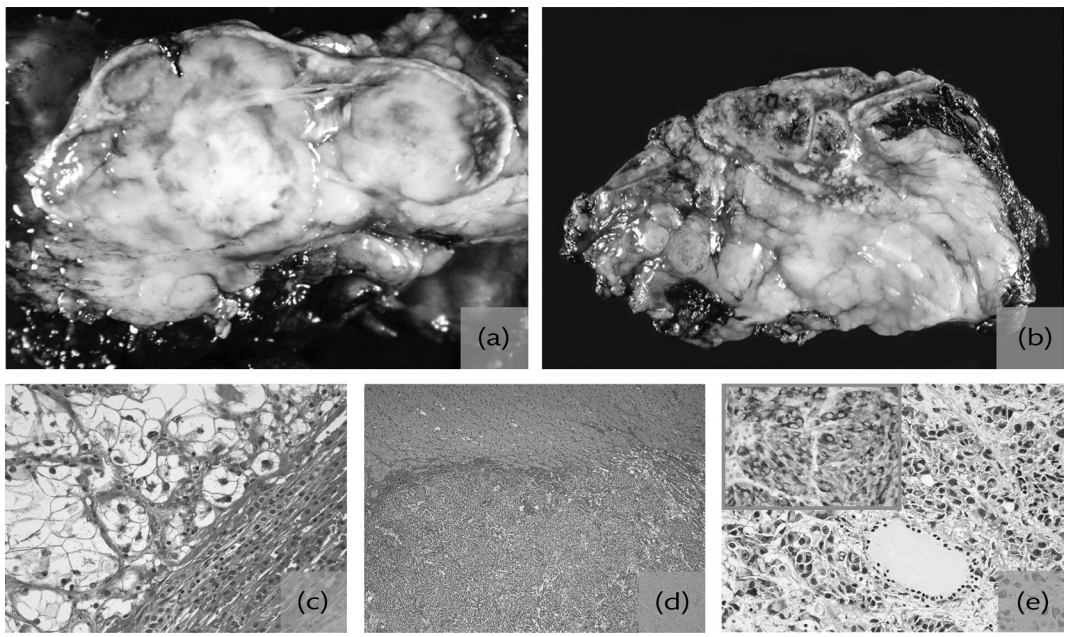


Figure 13.1.1.1 Transverse sections of adrenals either extensively replaced by confluent pale-white metastatic deposits of lung adenocarcinoma (a) or moderately expanded by two metastases of renal cell carcinoma, having a variegated appearance (b). Renal cell carcinoma metastatic to the adrenal gland simulating a primary adrenal tumour (c) and medullary thyroid carcinoma metastatic to the adrenal (d). Metastatic melanoma extensively effacing the thyroid gland parenchyma (e), which was confirmed on immunohistochemical grounds, i.e. S100, MELAN-A and HMB-45 (e; inset).

The Pituitary Gland

Pituitary tumours are intrasellar neoplasms with **pituitary adenomas** being the most common constituting up to 15% of all intracranial tumours, while tumours metastatic to the pituitary gland being rather rare [1]. Per data stemming from the **German Pituitary Tumour Registry**, the former account for 84.6% and the latter for 0.6% of 4122 surgical cases [16]. Only 0.4% of intracranial metastatic tumours are confined to the pituitary gland according to the **Brain Tumour Registry of Japan** [17], being consistent with previously published data supporting rare involvement of the pituitary gland by intracranial metastatic disease [18].

Based on a systematic review with pooled individual patient data analysis, He *et al.* [19] calculated the incidence of pituitary metastasis based on stratified patient populations, as follows: 0.87% among all intracranial metastases; 1.9% among all autopsied cancer cases; 11.56% among all breast cancer patients who had hypophysectomies and 12.83% among all autopsied breast cancer patients. This study confirmed **breast** and **lung cancer** as the most common primary malignancies accounting for two-thirds of the cases, while highlighted cancer of unknown primary (CUP) in approximately 3% of cases. To extend, it was shown that the odds ratio of a breast cancer patient having pituitary metastasis is ~7 times greater than a patient suffering from a different primary cancer.

Breast and lung are the commonest primary sites of metastatic origin in the rare tumour-to-tumour metastatic context [20].

Pituitary metastasis occurs mostly in patients suffering from systemic dissemination and less frequently being the only site of metastasis or the first manifestation of a latent primary tumour, as documented in 10% up to 40% of cases [17, 19, 21]. Although these tumours are clinically silent, a subset of patients is having pituitary-related symptoms [19, 21]. Diabetes insipidus remains the most frequent symptom at presentation, reflecting higher prevalence of metastatic deposits in the posterior lobe possibly due to lack of direct arterial blood supply to the anterior counterpart. Anterior hypopituitarism, pan-hypopituitarism, cranial nerve palsies, visual disturbance, and headaches can also occur. Extraordinary presentation of acromegaly due to somatotroph hyperplasia elicited by a metastatic GHRH-producing pancreatic neuroendocrine tumour has been documented [22].

The Thyroid Gland

Intrathyroidal metastasis is a relatively uncommon event in clinical practice, i.e. <0.2% of thyroid malignancies [23, 24]. Based on a systematic review and meta-analysis involving a total of 514 patients, Straccia *et al.* [23] evaluated the type of metastatic spread and the localization of the primary tumour. The overall rate of metachronous metastases was estimated at 69%, while the overall metastatic rate from organs located in the infradiaphragmatic region was 53%. **Renal cell carcinoma** was the most frequent malignancy followed by lung carcinoma, breast carcinoma, and colorectal adenocarcinoma. The **International Head and Neck Scientific Group** [24] confirmed the order of frequency in the *clinical series* and highlighted **lung cancer** as the most common primary tumour in the *autopsy series*. Both studies documented CUP in 2% of cases. Along these lines, Hegerova *et al.* [25] identified 97 patients over a 30-year period at the **Mayo Clinic** (Rochester, USA) with **kidney** (22%) and **lung** (22%) constituting the most frequent primary sites.

In accordance with a predilection of spread to endocrine tumours [26], metastatic disease can rarely involve primary thyroid tumours with follicular adenoma and papillary thyroid carcinoma being the commonest benign and malignant recipient respectively [27]. **Renal cell carcinoma** is the most common donor neoplasm followed by malignancies originating from lung, breast, and lower GI tract [27, 28].

Although most patients with a diagnosis of intrathyroidal metastasis have a prior history of malignancy, occult primary accounts for approximately 30% of cases [23]. Metastatic deposits can be asymptomatic or present with unspecific symptoms at an advanced stage, such as dysphagia, hoarseness, pain, and airway compromise. Extensive destruction of thyroid parenchyma can cause changes in thyroid function and regarded as symptom of advanced stage disease [23, 24].

The Parathyroid Glands

Metastatic disease to the parathyroid glands is rare, and poorly documented. Given evidence emerging from thyroidectomies performed for cancer and autopsy data, Shifrin *et al.* [26] suggested that metastatic involvement might be probably underreported [29, 30]. Approximately 4% of the former cases as well as ~12% of patients with various disseminated malignancies harboured metastatic infiltrates when multiple parathyroid glands were sampled [1, 30]. In a series of 392 consecutive parathyroidectomies performed for primary hyperparathyroidism, only one metastatic cancer was detected [26].

Per the **2017 WHO classification of endocrine neoplasia** and based predominantly on post-mortem data, the most common primary site is **breast** followed by skin (melanoma), lung and kidney [1, 26, 30]. Regarding cancer arising from the thyroid gland, the incidence rate of parathyroid involvement is ranging from 0.4% to 3.9% and with only a small subset corresponding to true metastasis [31, 32].

Metastatic disease can involve normal parathyroid tissue, hyperplastic parathyroid glands, and parathyroid adenomas. The patients might be asymptomatic or present with a mass in the neck, clinical signs of parathyroid disease and/or disseminated systemic metastasis. To exemplify, Torregrossa *et al.* [33] reported a metastatic renal cell carcinoma without clinical signs of parathyroid disease and/or disseminated metastasis. In the hyperplastic setting, Venkatraman *et al.* [34] described an incidental discovery of metastatic lung adenocarcinoma within a parathyroid gland resected for surgical management of primary hyperparathyroidism, while Shifrin *et al.* [35] a metachronous metastasis of neuroendocrine thymic carcinoma to a hyperplastic parathyroid gland in a patient with multiple endocrine neoplasia type 1 syndrome and widespread metastatic disease.

REFERENCES

1. Lloyd RV, Osamura RY, Kloppel G, Rosai J, eds. *WHO Classification of Tumours of Endocrine Organs*, 4th edn. IARC: Lyon, 2017.

2. Ernest E. *Lack: Tumours of the Adrenal Glands and Extraadrenal Paraganglia, Atlas of Tumour Pathology*, Series **4**(8). Washington, DC: Armed Forces Institute of Pathology, 2007.
3. Lam KY, Lo CY. Metastatic tumours of the adrenal glands: a 30-year experience in a teaching hospital. *Clin Endocrinol (Oxf)*, 2002; **56**: 95–101.
4. Moreno P, de la Quintana Basarrate A, Musholt TJ, *et al.* Adrenalectomy for solid tumor metastases: results of a multicenter European study. *Surgery*, 2013; **154**: 1215–22.
5. Sancho JJ, Triponez F, Montet X, Sitges-Serra A. Surgical management of adrenal metastases. *Langenbecks Arch Surg,* 2012; **397**: 179–94.
6. Patel JK, Didolkar MS, Pickren JW, Moore RH. Metastatic pattern of malignant melanoma. A study of 216 autopsy cases. *Am J Surg*, 1978; **135**: 807–10.
7. Mittendorf EA, Lim SJ, Schacherer CW, *et al.* Melanoma adrenal metastasis: natural history and surgical management. *Am J Surg*, 2008; **195**: 363–8.
8. Flaherty DC, Deutsch GB, Kirchoff DD, *et al.* Adrenalectomy for metastatic melanoma: current role in the age of nonsurgical treatments. *Am Surg*, 2015; **81**: 1005–9.
9. Matsukuma S, Kono T, Takeo H, Hamakawa Y, Sato K. Tumor-to-tumor metastasis from lung cancer: a clinicopathological postmortem study. *Virchows Arch,* 2013; **463**: 525–34.
10. Althoen MC, Siegel A, Tsapakos MJ, Seltzer MA. Lung cancer metastasis to an adrenal myelolipoma detected by PET/CT. *Clin Nucl Med*, 2011; **36**: 922–4.
11. Pathmanathan N, Murali R. Composite phaeochromocytoma with intratumoural metastatic squamous cell carcinoma. *Pathology*, 2003; **35**: 263–5.
12. Matsumoto K, Nosaka K, Shiomi T, Matsuoka Y, Umekita Y. Tumor-to-tumor metastases in Cowden's disease: an autopsy case report and review of the literature. *Diagn Pathol*, 2015; **10**: 172.
13. Paul CA, Virgo KS, Wade TP, Audisio RA, Johnson FE. Adrenalectomy for isolated adrenal metastases from non-adrenal cancer. *Int J Oncol*, 2000; **17**: 181–7.
14. Antonelli A, Cozzoli A, Simeone C, *et al.* Surgical treatment of adrenal metastasis from renal cell carcinoma: a single-centre experience of 45 patients. *BJU Int*, 2006; **97**: 505–8.
15. Miyao N, Naito S, Ozono S, *et al.* Late recurrence of renal cell carcinoma: retrospective and collaborative study of the Japanese Society of Renal Cancer. *Urology*, 2011; **77**: 379–84.
16. Saeger W, Lüdecke DK, Buchfelder M, Fahlbusch R, Quabbe HJ, Petersenn S. Pathohistological classification of pituitary tumors: 10 years of experience with the GermanPituitary Tumor Registry. *Eur J Endocrinol*, 2007; **156**: 203–16.
17. Habu M, Tokimura H, Hirano H, *et al.* Pituitary metastases: current practice in Japan. *J Neurosurg*, 2015; **123**: 998–1007.
18. Marsh JC, Garg S, Wendt JA, Gielda BT, Turian JV, Herskovic AM. Intracranial metastatic disease rarely involves the pituitary: retrospective analysis of 935 metastases in 155 patients and review of the literature. *Pituitary*, 2010; **13**: 260–5.
19. He W, Chen F, Dalm B, Kirby PA, Greenlee JD. Metastatic involvement of the pituitary gland: a systematic review with pooled individual patient data analysis. *Pituitary*, 2015; **18**: 159–68.
20. Hoellig A, Niehusmann P, Flacke S, Kristof RA. Metastasis to pituitary adenoma: case report and review of the literature. *Cent Eur Neurosurg*, 2009; **70**: 149–53.
21. Komninos J, Vlassopoulou V, Protopapa D, *et al.* Tumors metastatic to the pituitary gland: case report and literature review. *J Clin Endocrinol Metab*, 2004; **89**: 574–80.
22. Sanno N, Teramoto A, Osamura RY, *et al.* A growth hormone-releasing hormone-producing pancreatic islet cell tumor metastasized to the pituitary is associated with pituitary somatotroph hyperplasia and acromegaly. *J Clin Endocrinol Metab*, 1997; **82**: 2731–7.
23. Straccia P, Mosseri C, Brunelli C, *et al.* Diagnosis and treatment of metastases to the thyroid gland: a meta-analysis. *Endocr Pathol*, 2017; **28**: 112–20.
24. Nixon IJ, Coca-Pelaz A, Kaleva AI, *et al.* Metastasis to the thyroid gland: a critical review. *Ann Surg Oncol*, 2017; **24**: 1533–9.
25. Hegerova L, Griebeler ML, Reynolds JP, Henry MR, Gharib H. Metastasis to the thyroid gland: report of a large series from the Mayo Clinic. *Am J Clin Oncol*, 2015; **38**: 338–42.
26. Shifrin A, LiVolsi V, Shifrin-Douglas S, *et al.* Primary and metastatic parathyroid malignancies: a rare or underdiagnosed condition? *J Clin Endocrinol Metab*, 2015; **100**: E478–81.
27. Stevens TM, Richards AT, Bewtra C, Sharma P. Tumors metastatic to thyroid neoplasms: a case report and review of the literature. *Patholog Res Int*, 2011; **2011**: 238693.
28. Medas F, Calò PG, Lai ML, Tuveri M, Pisano G, Nicolosi A. Renal cell carcinoma metastasis to thyroid tumor: a case report and review of the literature. *J Med Case Rep*, 2013; **7**: 265.
29. Chrisoulidou A, Mandanas S, Mitsakis P, *et al.* Parathyroid involvement in thyroid cancer: an unforeseen event. *World J Surg Oncol*, 2012; **10**: 121.
30. Horwitz CA, Myers WP, Foote FW Jr. Secondary malignant tumors of the parathyroid glands. Report of two cases with associated hypoparathyroidism. *Am J Med*, 1972; **52**: 797–808.
31. Tang W, Kakudo K, Nakamura Y, *et al.* Parathyroid gland involvement by papillary carcinoma of the thyroid gland. *Arch Pathol Lab Med*, 2002; **126**: 1511–14.
32. Papi G, Corrado S, Fadda G, *et al.* Parathyroid gland involvement by thyroid cancer: results from a large series of thyroidectomies performed in two Italian university hospitals and review of the literature. *J Thyroid Res*, 2014; **2014**: 685425.
33. Torregrossa L, Rotondo MI, Insilla AC, *et al.* Metastasis of renal cell carcinoma to the parathyroid gland 16 years after radical nephrectomy: a case report. *Oncol Lett*, 2016; **12**: 3224–8.
34. Venkatraman L, Kalangutkar A, Russell CF. Primary hyperparathyroidism and metastatic carcinoma within parathyroid gland. *J Clin Pathol*, 2007; **60**: 1058–60.
35. Shifrin AL, LiVolsi VA, Zheng M, *et al.* Neuroendocrine thymic carcinoma metastatic to the parathyroid gland that was reimplanted into the forearm in patient with multiple endocrine neoplasia type 1 syndrome: a challenging management dilemma. *Endocr Pract*, 2013; **19**: e163–7.

13.1.2 Paraneoplastic Endocrine Syndromes

David W. Ray

Introduction

Production of hormones usually occurs in specialized endocrine glands. Such hormone production is typically under control from higher centres, ultimately the brain, and also subject to complex negative feedback. This results in tight regulation of circulating hormone levels, and affords a mechanism for influencing diverse tissue function throughout the body. Inappropriate hormone production by non-endocrine tissue causes a spectrum of rare syndromes which are important as they are not only a management challenge, but also because they shed light on the regulation of tissue-specific gene expression with widespread ramifications for understanding human physiology. In addition, inappropriate expression of a peptide may be useful as a tumour marker, as for example human chorionic gonadotrophin (hCG) or α-fetoprotein. This chapter will address the basic mechanisms of ectopic hormone production and will further discuss specific clinical syndromes.

The Origin of Ectopic Hormones

The use of sophisticated techniques has revealed that hormones may be expressed 'ectopically' in a variety of normal tissues other than in specialized glands [1]. It is, therefore, less than surprising that tumours arising from these tissues can give rise to ectopic hormone producing syndromes. If the hormone production in the tumour remained under physiological control ectopic hormone production would not pose a clinical problem. Sometimes aberrant control has an obvious mechanism, for example, lack of specific neural connection, physical distance from a portal system or absence of receptors for hormonal modulators of gene expression.

Dysdifferentiation Theory

This proposes that neoplastic change occurs in progenitor cells, rather than terminally differentiated ones [2]. The transformed progenitor cell subsequently undergoes differentiation but at each stage experiences a partial block. If this theory holds it would predict a tumour with a mixed population of cells at different stages in development. Depending on how the blocks to normal differentiation were arranged a majority of cells may develop into an endocrine cell type, and give rise to a typically endocrine-type tumour. This model may explain why certain hormones tend to be produced by certain tissues, since the tumour would retain some features of its parent tissue and would tend to transcribe the same genes as its parent tissue though may express these aberrantly (Figure 13.1.2.1).

Extrapituitary expression of the adrenocorticotrophic hormone (ACTH) gene (*POMC*) is described in detail next as a model for ectopic hormone syndromes. Ectopic ACTH syndrome shares many features with other ectopic hormone syndromes, and as many of the underlying mechanisms of expression are common they have a wider application.

Ectopic ACTH Syndrome

The human *POMC* gene is encoded in three exons on chromosome 2. The first exon is non-coding, the second contains the signal peptide which targets the protein product to the regulated secretion pathway and the third exon encodes the majority of the mature protein, including ACTH. In pituitary corticotroph cells, the only cells in health that express the gene at high level, the mature mRNA from the *POMC* gene is of 1200 nucleotides. A third *POMC* transcript has also been described which is longer than the pituitary form (about 1500 nucleotides). This arises from a site, or multiple sites, within the 5' flanking region of the human *POMC* promoter. This mRNA species therefore includes the entire coding region of the peptide, and it does appear to give rise to a secreted peptide product. This 'long' form of the *POMC* mRNA is found in extrapituitary tissues and tumours.

Dysdifferentiation hypothesis for ectopic hormone production

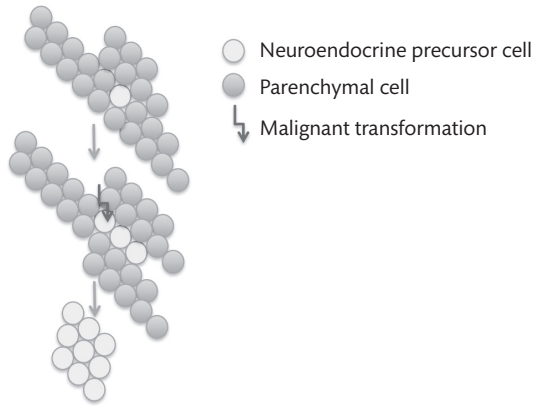


Figure 13.1.2.1 Dysdifferentiation results in aberrant gene expression. The malignant transformation of a progenitor cell results in a clonal population of cells that have undergone atypical differentiation. These cells share some features with their progenitor cells and may continue to express genes associated with these immature, incompletely differentiated cells. Malignant transformation of a cell can lead to persistent, or enhanced, hormone gene expression in the clone of proliferating cells.

Regulation of *POMC* Gene Expression

Expression of the *POMC* gene appears to be predominantly controlled at the level of gene transcription [3]. Expression of the *POMC* gene in the anterior pituitary corticotroph cell requires the action of a tightly restricted transcription factor, a member of the T-box family, termed Tpit [4]. This factor acts with the homeodomain protein PitX1 and promotes recruitment of SRC family coactivators to the *POMC* promoter, leading to enhanced gene transcription [5].

Corticotropin-releasing hormone (CRH) acts on pituitary corticotroph cells to increase cyclic adenosine monophosphate (cAMP) accumulation and activates mitogen-activated protein kinases. There is also evidence of activation of the orphan nuclear receptor nerve growth factor–induced clone B (or Nur 77) [6] leading to enhanced *POMC* transcription through the recruitment of SRC coactivators to nerve growth factor–induced clone B [5, 7]. As nerve growth factor-induced clone B and Tpit act synergistically, this suggests the formation of a regulatory complex on the *POMC* promoter with Tpit, NGFI-B and SRC coactivators [5]. It is important that expression of Tpit promotes corticotroph cell differentiation and that its expression is more limited than that of *POMC*. Therefore, there is no Tpit expression in hypothalamic *POMC*-expressing neurons, suggesting that Tpit is specific for corticotroph-specific expression of *POMC*; other mechanisms are responsible for expression elsewhere. Tpit expression has been found specifically in human pituitary corticotroph adenomas [4].

Glucocorticoids repress transcription of the *POMC* gene by binding to two DNA elements in the 5′ flanking region of the promoter. The more proximal element, an imperfect palindrome 63 nucleotides upstream from the transcription start site, is thought to bind three glucocorticoid receptor molecules in an unusual trimer formation [8–10]. This conformation of receptors on DNA directs repression of transcription rather than enhancement. Further upstream, between −480 and −320, there is another glucocorticoid-regulated element, suggesting that these two DNA elements interact to achieve the full effect of glucocorticoid repression [11]. It is interesting that Tpit expression, essential to the corticotroph cell type and to *POMC* expression, is not affected by glucocorticoids, in contrast to *POMC*, which is repressed [12]. Because Tpit is not part of the mechanism allowing glucocorticoid repression of *POMC*, the lack of Tpit in ectopic ACTH syndrome causing tumours cannot explain the failure of glucocorticoid repression characteristic of the disorder. However, the mechanism underlying *POMC* induction by CRH in some well-differentiated carcinoid tumours causing ectopic ACTH syndrome is not yet defined. Although the expression of nerve growth factor–induced gene B and Tpit in such tumours is not definitively addressed, other pathways, such as mitogen-activated protein kinase activation, cAMP activation, and induction of c-fos may be important.

A number of other hypothalamic factors act on the pituitary corticotroph to influence *POMC* expression. However, their modes of action are not well defined. In particular, arginine vasopressin stimulates *POMC* expression rather weakly but augments CRH action. The intracellular pathways activated by arginine vasopressin appear to be protein kinase C dependent, but arginine vasopressin also potentiates the action of CRH on cAMP generation [13].

Evidence points to intrapituitary factors as important modulators of corticotroph function. One such factor is the proinflammatory cytokine leukaemia inhibitory factor, which signals through the Janus kinase/signal transducers and activators of transcription pathway [14]. Leukaemia inhibitory factor has been shown to act on the *POMC* gene through a specific response element that overlaps with the −166 CRH response element. In addition to stimulating *POMC*, transcription leukaemia inhibitory factor also appears to trigger a 'switch' in cell phenotype from proliferative to synthetic [15].

However, many other peptide growth factors and cytokines are capable of activating cAMP, mitogen-activated protein kinase and Janus kinase/signal transducers and activators of transcription signalling cascades and thus are potentially capable of regulating *POMC* expression in non-pituitary tissue. Although extrapituitary tissues lack expression of corticotroph-specific transcription factors, activation of common signalling cascades might be expected to result in *POMC* gene expression. In extrapituitary tissues, the *POMC* gene may be modified to render it transcriptionally silent. One such irreversible modification is DNA methylation. The loss of methylation in tumour tissue may allow transcription of the gene to be activated by the common signalling pathways described previously. There is some evidence that such changes in DNA methylation do occur in cell line models of ectopic ACTH syndrome [16]. It seems likely that *POMC* expression per cell is less in most extrapituitary tumours compared with the pituitary corticotroph, but this relative inefficiency of expression is compensated for by the greater number of cells expressing the gene in extrapituitary tumours.

Pathophysiology

ACTH immunoreactivity has been recognized to show size heterogeneity for many years, with the presence of high-molecular-weight forms being detected in human plasma. The ectopic ACTH syndrome was the first recognized of the ectopic hormone syndromes. In its most florid form it is rare, affecting 4.5% of patients with small cell lung cancer in one study, but there is evidence of derangement in the hypothalamic–pituitary–adrenal axis in most patients with small cell lung cancer. Analysis of tumour tissue surprisingly suggested the presence of immunoreactive ACTH, even in the absence of clinical features of hormone excess [17]. The ACTH was present predominantly in a high-molecular-weight form, of approximately 20 kDa, but this purified material could be cleaved to mature ACTH (4.5 kDa) by the action of trypsin. Further work identified the presence of immunoreactive ACTH-like peptide in a variety of normal tissues, suggesting that extrapituitary ACTH expression was less 'ectopic' than inappropriately regulated. The ACTH immunoreactivity was found to have no biological activity, and was assumed to be 'big' ACTH. However, identification of predominantly high-molecular-weight forms of ACTH in the circulation of patients with clinically apparent Cushing's syndrome does suggest that the precursors of ACTH may have some activity at the ACTH receptor.

POMC Processing

The *POMC* gene leads to the generation of a pre-pro-hormone, *POMC*. This protein undergoes a series of proteolytic cleavages at dibasic amino acid residues to give rise to a series of small molecules including ACTH, melanocyte-stimulating hormone (MSH) and β-endorphin. In the anterior pituitary ACTH is cleaved by the action of a specific protease termed PC1 (for pro-hormone

convertase type 1). In the rodent intermediate lobe melanotroph the *POMC* molecule undergoes more comprehensive digestion to give smaller fragments, MSH, β-endorphin and corticotrophin-like intermediate lobe peptide (CLIP) as a result of cleavage by pro-hormone convertase type 2 (PC2).

In the majority of extrapituitary tumours causing the ectopic ACTH syndrome processing of the pre-pro-hormone is incomplete. Therefore, the ectopic ACTH syndrome is characterized by the presence of high-molecular-weight forms of ACTH in the circulation [18]. It is likely that the extent of processing correlates with the degree of neuroendocrine differentiation of the tumour, and hormonal manifestations are probably only seen in tumours with significant hormone processing capacity. This lack of processing could result from a lack of the specific cleavage enzymes, PC1 and PC2, which are expressed only in specialized endocrine tissue, or with a switch from a regulated secretory pathway to a constitutive one. A number of small, highly differentiated, slow-growing tumours, typically bronchial carcinoid, have been characterized to process *POMC* in the neurointermediate lobe manner, giving rise to small fragments in the circulation, such as CLIP, and β–MSH. These have been used to aid diagnosis in some cases of Cushing's syndrome, although the series are too small to confidently extrapolate from.

Dysregulation of *POMC* Expression in Extrapituitary Tumours

In contrast to expression of the *POMC* gene in pituitary corticotroph cells, which is repressed by glucocorticoid as just discussed, expression in extrapituitary tumours is characteristically resistant to glucocorticoid [19]. This is the basis of the high-dose glucocorticoid suppression test used to distinguish eutopic from ectopic sources of ACTH in Cushing's syndrome. As the test has approximately 10% false-positive and 10% false-negative it has largely been superseded by sophisticated imaging and inferior petrosal sinus sampling for differential diagnosis. With the availability of recombinant corticotrophin-releasing hormone responses of extrapituitary tumours to this peptide have been measured. In general, only pituitary corticotrophs stimulate *POMC* expression in response to corticotrophin-releasing hormone, but exceptions are increasingly being identified.

Clinical Manifestation

The diagnosis of Cushing's syndrome and differential diagnosis of ACTH-dependent Cushing's syndrome are described elsewhere. Briefly, dynamic endocrine testing is required to diagnose Cushing's syndrome, and detection of ACTH using a sensitive two-site immunoradiometric assay (IRMA) makes the diagnosis of ACTH-dependent Cushing's syndrome. A variety of dynamic endocrine and imaging protocols are used to identify a pituitary or extrapituitary source of the ACTH excess. These all have variable sensitivity and specificity. The most reliable test is bilateral inferior petrosal sinus sampling, which, if performed when the patient is hypercortisolaemic, has an accuracy approaching 100%.

Most occult tumours are carcinoid, phaeochromocytoma, or medullary thyroid carcinoma and originate in the neck, chest, or abdomen (Box 13.1.2.1). Computed tomography (CT) or magnetic resonance (MR) scanning can be used to detect chest tumours in those patients with a normal chest radiograph. There have been some reports of success in using radioactive In-labelled octreotide scanning to identify occult neuroendocrine tumours, although experience is still limited. The development of new imaging techniques may identify previously occult primary tumours, with ^{11}C-methionine PET-CT imaging being used in a limited number of patients to good effect [20]. However, this requires highly specialized facilities for the synthesis of the labelled aminoacid, and its use. This should be regarded as a research tool until larger series of ectopic ACTH patients have been reported.

Management

Treatment is focused on two objectives. The first is control of the endocrine manifestation and the second is management of the underlying tumour. Individual patients will present with different priorities. The ideal treatment is curative resection of the primary tumour, which achieves both objectives. If this is not possible, due to a small, elusive primary tumour, patients with small, occult primary tumours may be managed by chemical or surgical adrenalectomy, and in many cases the primary tumour will not be life threatening. Patients with extensive carcinoma, for example, small cell carcinoma, in whom ACTH excess coexists, are best managed by chemotherapy, which indirectly reduces ACTH expression. The presence of clinical hypercortisolaemia in small cell lung carcinoma is linked to poor prognosis but chemotherapy should be tailored for the cell type and tumour stage regardless of the presence of hormone excess. The exception to this is those cases with florid Cushing's syndrome in whom a tissue diagnosis has yet to be obtained. In these individuals, it is prudent to start adrenolytic treatment with metyrapone while concluding investigation. In patients with aggressive Cushing's syndrome adrenalectomy is lifesaving.

Syndrome of Inappropriate Antidiuretic Hormone (ADH) Secretion (Schwartz–Bartter Syndrome)

Pathophysiology

The syndrome of inappropriate antidiuresis (SIADH) is the most common cause of hyponatraemia. It may be caused by a wide range of underlying disorders, but in three broad groups. These comprise malignancies, neurological disorders, and lung diseases. The latter two conditions cause hyponatraemia as a result of hypothalamic vasopressin, whose secretion comes under aberrant control from either neuronal inputs or circulating humoral factors. The first results from vasopressin expression in non-hypothalamic-pituitary tissue. The result of either source of overproduction is hyponatraemia with apparently inappropriate renal sodium excretion.

Box 13.1.2.1 Types of tumours causing ectopic adrenocorticotropic hormone secretion

- Small cell lung carcinoma
- Carcinoid tumours (bronchus, thymus, small intestine)
- Pancreatic islet cell tumour
- Pheochromocytoma
- Medullary carcinoma of the thyroid
- Carcinomas (breast, gastrointestinal tract [oesophageal, gastric, colorectal], ovarian, cervical, prostate)

Box 13.1.2.2 Tumours associated with syndrome of inappropriate antidiuretic hormone secretion

Tumour
• Small cell lung carcinoma
• Pancreas
• Duodenum
• Urethra
• Prostate
• Bladder
• Lymphoma and other

The vasopressin gene is expressed in a number of separate neuronal nuclei, and also in peripheral tissues in health. Regulation of vasopressin expression is dependent on site. For example, hyperosmolality increases vasopressin expression in the supraoptic nucleus and the magnocellular division of the paraventricular nucleus, but vasopressin mRNA in other sites, including the suprachiasmatic nucleus, is unaltered. Vasopressin expression in the suprachiasmatic nucleus, in contrast, is under diurnal regulation. Androgens upregulate expression of vasopressin in the striae terminalis, and glucocorticoids suppress expression in the parvocellular division of the paraventricular nucleus. Differential regulation, even within such anatomically closely related sites likely results from differential expression of hormone receptors in the cells, and different neuronal afferents. Vasopressin gene transcription is under positive regulation by cAMP and protein kinase C pathways. Much less is known about regulation of vasopressin outside the central nervous system (CNS), but glucocorticoids were shown to suppress its expression in a small cell lung carcinoma cell line.

Ectopic secretion of vasopressin occurs in squamous cell carcinoma, small cell carcinoma, neuroblastoma, and in undifferentiated carcinoma. A wide variety of non-tumour causes for the SIADH has also been defined (Box 13.1.2.2). In one series 16% of patients with small cell lung carcinoma had hyponatraemia (less than 130 mmol/L) at diagnosis, compared to 0% of patients with non-small cell carcinoma. Hyponatraemia was found to be an independent predictor of poor prognosis in extensive stage disease. *In vitro* studies found 7 of 11 tumours in culture produced vasopressin, 9 of 11 atrial natriuretic factor and 5 of 11 both hormones. All the cells studied from patients with hyponatraemia produced either one of the two hormones [21].

Box 13.1.2.3 Clinical features of hyponatraemia

- Headache
- Lethargy
- Weakness
- Nausea/vomiting
- Mood swings
- Confusion
- Drowsiness
- Hyporeflexia
- Positive Babinski's sign
- Convulsions
- Coma

Clinical Manifestation

Hyponatraemia presents with features of neuropsychiatric dysfunction in most cases (Box 13.1.2.3). The elderly and the young are more likely to be symptomatic than others. The absolute sodium concentration is less reliable as a predictor of symptoms than the rate of fall of sodium concentration, although almost all symptomatic patients will have a plasma sodium less than 120 mmol/L, and plasma sodium concentration greater than 125 mmol/L rarely have symptoms related to hyponatraemia specifically. Clinical features include lethargy, fatigue, impaired conscious level, coma, seizures, and psychosis. Hyponatraemia may cause death as a result of cerebral oedema, uncontrolled seizures, and the consequences of coma. Although in most cases mild hyponatraemia (over 125 mmol/L) is regarded as a straightforward condition which may not require specific treatment, hyponatraemia should not be regarded as benign.

A set of diagnostic criteria must be fulfilled before a secure diagnosis may be reached (Box 13.1.2.4). In practice it is useful to perform a bedside evaluation of the patient's extracellular fluid volume. This measure is tightly related to the total body sodium. Patients who have hyponatraemia in the absence of oedema or hypovolaemia are a select group who, in the absence of other endocrine, psychiatric, or pharmacological cause, are defined as having SIADH. Plasma atrial natriuretic factor (ANF) is usually decreased in this group. The underlying cause is then sought. Neurological, lung, drug-related, and miscellaneous causes result in dysregulation of vasopressin regulation in the hypothalamus, and should not, therefore, be regarded as true ectopic hormone secretion states. In contrast, a variety of tumours have been shown to be aberrantly secreting vasopressin, and considerably more to be expressing the vasopressin gene inappropriately. There is evidence from T_1-weighted MR scans of the pituitary that such ectopic vasopressin secretion results in central suppression of vasopressin synthesis.

On the basis of salt and/or water loading tests four subgroups of SIADH have been defined, but as these subgroups do not partition with underlying cause this classification is not useful in routine practice.

Management

The management of this disorder falls into two parts. The first is diagnosis and treatment of the underlying cause, and the second is removal of excess, free body water. Discussion of specific therapy of the variety of underlying tumours is beyond the scope of this account, but surgical cure or debulking, chemotherapy, and radiotherapy have all been applied. In general, the circulating vasopressin concentration bears a direct relationship to tumour bulk within a patient, but little relationship across a patient cohort, presumably reflecting intertumour differences in cellular differentiation. Decisions about the acute correction of hyponatraemia are complicated by the occurrence of both pontine and extrapontine myelinolysis as consequences of therapy. The risk of myelinolysis is linked to the rate of change in sodium concentration. Therefore, a prudent approach is always justified. In symptomatic patients, treat with frusemide and hypertonic saline until convulsions cease and conscious level improves. This is usually achieved by a rapid increase in sodium concentration of 10% (approximately 10 mmol/L). After such initial emergency treatment patients are best managed by water restriction. In asymptomatic patients, the condition

Box 13.1.2.4 Diagnostic criteria for the syndrome of inappropriate antidiuretic hormone secretion

- Hyponatraemia
- Inappropriately increased urine osmolality (>100 mOsm/kg)
- Persistent sodium excretion in urine (>20 mmol/L)
- Normal renal, thyroid, and adrenal function
- No hypovolemia, oedema, or diuretic use

is almost always chronic. These patients should be treated by water deprivation in the first instance regardless of the sodium concentration as treatment is likely to hold greater dangers than persisting hyponatraemia. Water restriction results in a decrease in urinary sodium excretion, often to less than 10 mmol/L, indicating that these patients do have intact mechanisms for sodium conservation.

A number of pharmacological approaches will antagonize the action of ADH on the renal tubule. The most commonly used is demeclocycline, in divided doses up to 1200 mg daily. In addition, lithium carbonate is effective but is harder to use and a more toxic alternative. These agents induce a state of nephrogenic diabetes insipidus, and so encourage loss of water. Alternatively, oral sodium supplementation, up to 3 g daily, with frusemide, 40–80 mg daily, results in net loss of free water. The entry of specific V2 receptor antagonists to the clinic offers a new approach to management. This especially relevant to patients with tumoural synthesis of ADH, in particular small cell bronchogenic carcinoma, in whom curative therapy is not possible. The vasopressin receptor antagonists (Vaptans) increase free eater clearance by blocking the action of ADH at the renal tubule. Conivaptan is available in the United States, and is used as an intravenous agent, while in Europe Tolvaptan used orally. Initial studies focused on short-term use, and demonstrated efficacy, but now longer-term follow up is available, showing therapeutic response out to 4 years [22]. A number of specific cautions have been raised including the need to exclude patients who are hypovolaemic, and not to use with other interventions targeting hyponatraemia, including fluid restriction. Monitoring of serum sodium is required, to avoid over-rapid correction, and there are drug interaction liabilities through CYP3A4, including grapefruit juice. The issues around long-term use of vaptans relate to drug cost, and the risk of overcorrection, requiring careful monitoring of serum sodium.

Humoral Hypercalcaemia of Malignancy (HHM)

Hypercalcaemia is a common complication of malignancy. It may result from the lytic effect of bony metastases, or the effect of tumour-derived humoral factors.

Pathophysiology

The humoral syndrome has been explained by the isolation and characterization of a peptide hormone, parathyroid hormone-related protein (PTHrP). This hormone is closely related in amino acid sequence to parathyroid hormone (PTH) in its N-terminal region (amino acids 1–34), but after residue 34 the two peptides have unique sequences. The discovery of PTHrP as the circulating mediator of hypercalcaemia in malignancy allowed the discarding of earlier theories about ectopic production of PTH as the cause. There are isolated reports of ectopic PTH production by tumours, but these are excessively rare.

PTHrP is seldom detectable in the circulation of normal subjects, but its expression has been shown in a number of normal tissues. PTHrP may be regarded as the product of the diffuse paracrine system, and may have evolved to perform quite different physiological roles compared to the structurally related PTH. Under these circumstances it is hard to call production of PTHrP from a tumour arising from any tissue truly ectopic, as no definite eutopic source for the peptide has been defined. However, humoral hypercalcaemia of malignancy (HHM) is most conveniently considered with the group of ectopic hormone secretion syndromes.

PTH and PTHrP share a common receptor, which they recognize through their homologous N terminals. The PTH/PTHrP receptor mediates the action of both peptides in bone and kidney, and is a member of the G-protein-coupled seven transmembrane receptor family. The common receptor explains how PTHrP is able to generate cAMP in membrane preparations of PTH-sensitive renal tubule, and further why the humoral hypercalcaemia of malignancy (HHM) syndrome resulted in hypercalcaemia with hypophosphataemia. In the past there was controversy about 1,25-dihydroxyvitamin D levels in primary hyperparathyroidism versus HHM. PTHrP and 1,25-dihydroxyvitamin D appear to be loosely correlated in HHM, suggesting that PTHrP shares with PTH the capacity to induce 1α-hydroxylase. Occasional reported discrepancies stem from the action of other tumour-derived, circulating factors, or from the metabolic consequences of malignancy.

PTHrP may also play a role in hypercalcaemia related to osseous metastases, in that even the hypercalcaemia associated with bony metastases has a significant humoral component. Further, expression of PTHrP by primary tumours is a predictor of development of bony, metastatic disease. Thus, the local production of PTHrP by bone micrometastases may facilitate bony invasion and destruction [23].

Hypercalcaemia in Haematological Malignancy

Hypercalcaemia occur in up to 30% of patients with multiple myeloma. Skeletal involvement causes extensive bone destruction with pain, and risk of pathological fracture. Histological evidence suggests that the bone disease is caused by increased osteoclastic activity, in the absence of significant osteoblastic activity. Loss of osteoblastic activity is also supported by the characteristically negative bone scan and suppressed circulating osteocalcin concentration. A number of cytokines, produced by activated immune cells, have been shown to have direct effects promoting bone resorption. Such cytokines include tumour necrosis factor-α, tumour necrosis factor-β, interleukin-1 and leukaemia inhibitory factor (LIF). A superseded generic term for these factors was 'osteoclast activating factor'. However, in three of nine patients with multiple myeloma complicated by hypercalcaemia there was an elevation in circulating PTHrP, suggesting that a mechanism similar to HHM may be operating in at least some patients with haematological malignancy associated hypercalcaemia [24].

Generally hypercalcaemia is rare in lymphoma, with the exception of adult T-cell leukaemia/lymphoma [25]. This disease occurs in Japan and the West Indies and is caused by infection with the human T-cell lymphotropic virus type I (HTLV-I). At least one-quarter of

patients will develop hypercalcaemia, which is associated with suppressed 1,25-dihydroxyvitamin D. Hypercalcaemia predicts outcome, and is implicated in causing patient mortality. There is strong evidence that the hypercalcaemia is mediated by PTHrP, and also by local cytokine production, particularly interleukin-1α.

Clinical Manifestation

Hypercalcaemia is the most common metabolic complication of malignant disease, and is the cause of much morbidity (Box 13.1.2.5). Most cases are due to humoral mechanisms, principally PTHrP, rather than direct damage of bone by malignant cells. This is clear from the observation that even patients who have bone metastases and hypercalcaemia have a poor correlation between extent of skeletal involvement and calcium concentration in the circulation. In clinical practice, the underlying tumour will usually be obvious by the time hypercalcaemia is noted, and the patient is usually obviously ill. Less commonly hypercalcaemia may be due to an occult malignancy, here the diagnosis is made by checking serum intact PTH which is invariably suppressed in true HHM. The presence of a non-parathyroid tumour associated with hypercalcaemia and elevated PTH suggests either concomitant hyperparathyroidism, or true ectopic PTH secretion—a very rare event.

The presentation of hypercalcaemia may be confusing, and may be attributed to the underlying disease process itself. In general, the clinical manifestations of hypercalcaemia correlate with the calcium concentration and the rapidity of its rise. Most people show clinical features when the total calcium concentration exceeds 3.0 mmol/L, and features are almost invariable at concentrations above 3.5 mmol/L. Patients may be non-specifically unwell, they may complain of constipation, nausea, vomiting, confusion, or dehydration. Relatives may be the first to notice a change in concentration, or increased sleeping. Hypercalcaemia often affects the gut with constipation, anorexia, nausea, and vomiting. Hypercalcaemia induces a diuresis, and so may cause profound dehydration, particularly in association with vomiting or drowsiness. The absence of clinical features in a patient with severe hypercalcaemia should prompt measurement of ionized calcium to ensure that the hypercalcaemia is not due to excessive binding of calcium to plasma proteins.

Management

The decision to start treatment depends on the calcium concentration and the presence of symptoms. In general, patients with calcium concentrations below 3.0 mmol/L do not require therapy, and those with concentrations above 3.5 mmol/L do. Those with calcium concentrations between 3.0 and 3.5 mmol/L should be treated if there are symptoms, otherwise a conservative approach is preferred with monitoring of calcium concentrations and checking for development of clinical features. An important further consideration is the underlying malignancy and its prognosis, for example, in a terminally ill patient for whom there is no further specific antitumour therapy possible it may be better to resist attempting to reduce serum calcium concentration but rather make the patient comfortable.

There is little evidence that hypercalcaemia is a significant cause of premature mortality in cancer, but it is a significant cause of morbidity. Even if the underlying malignancy is beyond cure, effective relief of hypercalcaemia can be a most useful palliative intervention.

The initial management consists of general measures designed to increase calcium clearance. Dehydration is very common with significant hypercalcaemia and should be corrected. This is best achieved using an intravenous infusion of 3–4 L of 0.9% sodium chloride given over 24 h. This will typically reduce calcium concentration by about 0.5 mmol/L. Clearly, this form of therapy should be used with caution in older people, and those with impaired cardiac or renal function. Following hydration, intravenous loop diuretics will enhance calcium excretion by inhibiting calcium resorption by the thick ascending loop of Henle. Frusemide 40–80 mg may be used by bolus intravenous injection to supplement saline infusion, but there is little to be gained from higher doses or continuous infusions of diuretic. Diuretics should not be used in the presence of persisting dehydration.

Thiazide diuretics should be withdrawn as they tend to reduce renal calcium clearance, and the patient, where possible, encouraged to keep mobile in order to reduce immobility associated calcium mobilization.

Dialysis, either peritoneal or haemodialysis, against a low calcium dialysate is effective at rapidly reducing serum calcium, and may be particularly useful in renal impairment.

Bisphosphonates are analogues of pyrophosphate which are resistant to phosphatase degradation. Gastrointestinal absorption of bisphosphonates is very poor and so usually they are given by intravenous infusion. The bisphosphonates are rapidly cleared from the circulation and concentrated in bone. They appear to inhibit osteoclast activity, and may induce osteoclast apoptosis. Their duration of action is greatly longer than predicted by their plasma half-life, reflecting their distribution and mode of action.

Usually a single infusion of 30–60 mg of pamidronate is sufficient as initial treatment, pamidronate is more effective than the first-generation bisphosphonate, etidronate, which it has superseded for this indication. Pamidronate often causes myalgia and a transient fever, which can be helped by pretreatment with paracetamol. At a dose of 90 mg pamidronate often causes infusion reactions. Clodronate (300 mg intravenously) is another bisphosphonate which is effective for acute management of hypercalcaemia, and newer bisphosphonates including alendronate, risedronate, and aminobutane bisphosphonate are becoming available. These newer drugs appear to have similar efficacy to pamidronate. The bisphosphonates are usually given by slow intravenous infusion in large volumes (above 500 ml) to prevent nephrotoxicity due to precipitation of calcium bisphosphonate. The calcium response is typically rapid with a steady decline over the first 24 hours and may last from days up to one month. The response to treatment should be monitored by checking serum calcium daily until its concentration reaches a plateau; thereafter, monitoring can be performed weekly, or on recurrence of symptoms.

Box 13.1.2.5 Symptoms and signs of hypercalcaemia

- Polyuria
- Thirst
- Nausea
- Anorexia
- Constipation
- Confusion
- Drowsiness
- Headache
- Coma

Bisphosphonates may be used to maintain normocalcaemia, but treatment must be adjusted individually. It is possible to give bisphosphonates by subcutaneous infusion, or orally, in the domiciliary setting as intermittent intravenous infusions require hospital admission. Either clodronate or alendronate are effective in this role. Oral etidronate appears ineffective in maintenance therapy and may cause osteomalacia when used chronically.

In the past calcitonin and/or mithramycin (now called plicamycin) were used, but these have been largely superseded by the bisphosphonate drugs. Calcitonin inhibits osteoclastic bone resorption and is a very safe drug. Further, it causes a rapid fall in serum calcium, within 6 h of administration. Calcitonin (4–8 U/kg) is given by intramuscular or subcutaneous injection every 6 hours. Unfortunately, the effect of calcitonin is transient, and rarely sufficient to normalize serum calcium. It is occasionally useful in severe hypercalcaemia when a rapid response is needed while awaiting the more sustained bisphosphonate effect.

Gallium nitrate has been used to treat hypercalcaemia; however, it is cumbersome to administer and is nephrotoxic. It is used as a continuous 5-day infusion at a dose of 200 mg/m²/day. The maximal hypocalcaemic effect may not be seen until 3 days after the end of the infusion.

Glucocorticoid treatment, usually intravenous 200–300 mg hydrocortisone per day for 3–5 days, has been used to treat malignancy associated hypercalcaemia, but is usually effective only in lymphoma, multiple myeloma, or granulomatous disease.

Non-Islet Cell Tumour Hypoglycaemia

Pathophysiology

Fasting hypoglycaemia may arise as a consequence of non-islet cell tumour formation. Such tumours do not express insulin, a hormone which appears to be very tightly regulated in its tissue distribution, but the insulin-related molecule IGF-II. The two insulin-like growth factors (IGFs) I and II are members of the insulin family of peptide hormones along with relaxin, and are capable of signalling both through the type 1 IGF receptor or the insulin receptor. IGF-I is the liver-derived circulating mediator of growth hormone (GH) action, and IGF-II may have a more important role in development. In normal subjects IGF-I and IGF-II circulate at much higher concentrations than insulin, and would, if unopposed, cause profound hypoglycaemia due to their actions through the insulin receptor. That this does not occur is due to the presence of high-affinity, high-capacity, circulating IGF-binding proteins, the most important of which is IGF-binding protein-3. The IGFs form a ternary complex with IGF-binding protein-3 and with another liver produced protein, the acid labile subunit (ALS). Formation of this ternary complex between the IGFs and their binding proteins results in very low concentrations of free IGFs and so limits their bioavailability.

A number of tumours, typically of mesenchymal origin, have been identified as the cause of non-islet cell hypoglycaemia (Table 13.1.2.1). The apparent mechanism is overproduction of IGF-II. The circulating IGF-II would be expected to cause few problems if it were effectively sequestered by IGF-binding proteins, but this does not occur [26]. The tumour-derived IGF-II has a higher molecular mass compared with mature IGF-II, as a result of impaired proteolytic processing of pro-IGF-II, and is also often abnormally glycosylated [27]. The high concentrations of IGF-II suppress pituitary GH secretion and so result in reduced hepatic production of the ALS and IGF-binding protein 3 [27]. The high-molecular-weight form of IGF-II derived from the tumour ('Big' IGF-II) binds to IGF protein-3 in a 50 kDa binary complex in contrast to the 150 kDa ternary complex that is usually formed between IGF-I or IGF-II, IGF-binding protein-3, and the ALS. The IGF-II in the 50 kDa complex has increased insulin-like activity compared to the IGF-II containing ternary complex, and because of its smaller size appears to have greater capillary permeability, which results in greater bioavailability of the IGF-II. In addition to the changes caused by suppression of GH secretion (which result in reduced hepatic production of IGF-binding protein-3), the IGF-binding protein-3 produced has a lower affinity for forming the ternary complex with the ALS and IGF-II. This altered affinity also tends to promote IGF-II–IGF-binding protein-3 binary complex formation. Further secondary events also stem from these changes, including increased circulating concentrations of other IGF-binding protein including IGF-binding protein-2 and IGF-binding protein-6.

Table 13.1.2.1 Tumours associated with non-islet cell tumour hypoglycaemia

Carcinoma	Mesenchymal
Hepatocellular, hepatoma	Fibroma, fibrosarcoma
Adrenocortical	Mesothelioma
Pancreatic	Rhabdomyosarcoma
Gastric	Neurofibroma, neurofibrosarcoma
Colon	Leiomyosarcoma
Lung (small cell, squamous)	Others
Kidney	Haemangiopericytoma
Prostate	Haematologic, lymphoma

As pro-IGF-II is complexed just to IGF-binding protein-3 in the 50-kDa binary complex rather than the 150-kDa ternary complex, it has a 30-fold shorter plasma half-life. Therefore, measurement of total IGF-II may not be significantly elevated despite the presence of hypoglycaemia. Serum levels of free IGF-II and pro-IGF-II are both consistently elevated.

Glucocorticoids result in a series of metabolic effects in non-islet cell tumour hypoglycaemia. Prednisolone directly suppresses tumour production of IGF-II, enhances formation of the ternary complex and so relieves hypoglycaemia. Resulting from these changes there is an increase in IGF-I, IGF-binding protein-3, and ALS concentrations. As glucose is normalized, a fall in IGF-binding protein-2 follows. Recombinant human GH also reverses hypoglycaemia and promotes formation of the ternary complex. GH directly increases concentrations of IGF-I, IGF-binding protein-3, and ALS; however, only prednisolone improved the ability of IGF-binding protein-3 to bind to the ALS to form the IGF-I containing ternary complex. Therefore, prednisolone and GH are both effective treatments for non-islet cell hypoglycaemia, but they work in different ways [28].

Clinical Manifestation

The diagnosis is based on the recognition that the patient's symptoms are due to hypoglycaemia, which involves a detailed history,

Table 13.1.2.2 Non-islet cell tumour hypoglycaemia versus insulinoma

	Non-islet cell	Insulinoma
IGF-1	×	×
IGF-2		×
IGFBP-3	¬	×
Insulin	¬	
Glucose	¬	¬
Growth hormone	¬	× or
β-Hydroxybutyrate	¬	¬

IGF, insulin-like growth factor; IGFBP, insulin-like growth factor–binding protein., increased; ¬, decreased; ×, equivocal.

and confirmation of either hypoglycaemia occurring during a symptomatic episode, or fasting hypoglycaemia. Clinical features include hunger, attacks of sweating, pallor, and dizziness relieved by eating and weight gain. The typical biochemical accompaniment to hypoglycaemia is suppressed insulin, suppressed ketone bodies (β-hydroxybutyrate) and suppressed GH. Total IGF-II may be elevated or normal, while free IGF-II and pro-IGF-II are usually elevated. IGF-binding protein-3 and ALS are both suppressed, as a consequence of low GH, and two other IGF-binding proteins IGF-binding protein-2 and IGF-binding protein-6 are both elevated (Table 13.1.2.2).

The majority of tumours are large by the time hypoglycaemia has occurred, and are often easily diagnosed clinically. It may be that the tumour is recognized before the episodes are ascribed as hypoglycaemia. The tumours are typically imaged with a combination of chest radiography, abdominal ultrasound, and CT. Histological confirmation of the diagnosis is by specific immunostaining of a biopsy for IGF-II, or identification of pro-IGF-II in tumour extract.

Management

Effective management requires either surgical excision or debulking of the tumour. This is best achieved by close consultation between surgeon, interventional radiologist, who may have much to contribute by, for example, tumour embolization and oncologist. These tumours tend not to be radiosensitive, although there are isolated case reports of therapeutic response. As these tumours are rare, management is tailored to the individual patient, and is usually based on pragmatic grounds.

As a first course of action, curative resection of the primary tumour should be the aim. This depends on the site of tumour and the physical condition of the patient. Resection of pleural tumours has been reported, but curative resection of hepatoma is difficult unless orthotopic liver transplantation is feasible. Even in cases where curative resection is not possible, debulking accompanied by tumour embolization by an experienced interventional radiologist can result in resolution of hypoglycaemia in the absence of further drug therapy. Even inoperable tumours, such as a limb fibrosarcoma in an elderly patient, can be rendered hormonally inactive by tumour embolization, which offers a rapid and simple resolution of hypoglycaemia.

In patients whose tumours cannot be physically reduced relief of hypoglycaemia can be achieved by using glucocorticoid, usually dexamethasone, either alone or with recombinant human GH [29]. Glucocorticoids reduce tumour production of IGF-II, enhance formation of the ternary complex and increase production of IGF-binding protein-3 and the ALS. GH has no direct effect on the tumour but increases hepatic production of IGF-binding protein-3 and the ALS and enhances sequestration of IGF-II within the ternary complex. Pragmatically glucocorticoids are best used in the initial stages as their effect is of rapid onset, they are easy to administer and are inexpensive. The wide spectrum of glucocorticoid side-effects will make longer-term use of glucocorticoids less desirable. In the medium-term recombinant human GH will provide effective relief of hypoglycaemia, with a better side-effect profile. Disadvantages of GH include the need for daily injections and cost.

Other Ectopic Hormones

The two pituitary hormones prolactin and GH are of interest in that they have a wide extrapituitary expression, and yet are very seldom the cause of clinically significant ectopic hormone syndromes. Prolactin is expressed in decidualized endometrium, T lymphocytes, mammary epithelial cells, skin, sweat glands, and in the brain. It is the same gene that is transcribed in all these cases, rather than the related placental lactogen type gene, but the regulation of the gene appears completely different [30]. Whereas in the pituitary lactotroph cell prolactin is under the transcriptional control of the pituitary specific factor Pit1, in extrapituitary tissues Pit1 is not expressed, and the pituitary promoter of the prolactin gene is in consequence silent. The gene is transcribed from an upstream promoter, which gives rise to a slightly longer mRNA, with a unique 5' end, but after processing results in a protein with the same amino acid sequence. Because the gene is transcribed from a different promoter, the control of gene transcription, its basal rate, and regulation by external signals is different. For example, in T lymphocytes prolactin gene transcription is responsive to the immunophilins including cyclosporin A. The function of this extrapituitary prolactin is subject to debate, and it is not clear why such widespread expression in health is accompanied by such rarity of overexpression in malignant disease, in contrast to ACTH or vasopressin expression. As prolactin receptors are found in such a variety of tissues that cannot be reconciled with an exclusive action on mammary milk production prolactin may well have a more diverse pattern of action than that so far determined.

GH is also found in extrapituitary tissues in health, again in cells of haemopoietic lineage. This expression has been suggested to result in paracrine signalling, although hard data are lacking.

Extrapituitary GH expression causing acromegaly has been described in tumours of the pancreas, lung, and ovary. Ectopic GH-releasing hormone causing acromegaly through somatotroph hyperplasia rather than adenoma formation (occurs in less than 1% of acromegaly cases) may arise from carcinoid tumours of the pancreas or lung, or from phaeochromocytomas (Box 13.1.2.6). Acromegaly due to ectopic production of growth hormone-releasing hormone (GHRH) can be successfully treated using long-acting somatostatin analogues (octreotide or lanreotide).

Box 13.1.2.6 Tumours associated with growth hormone-releasing hormone secretion

- Carcinoid tumours (e.g. bronchial)
- Pheochromocytomas
- Pancreatic islet tumours
- Adrenal adenomas
- Hypothalamic gangliocytomas
- Small cell lung carcinoma

REFERENCES

1. DeBold CR, Menerjee JK, Nicholson WE, Orth DN. Proopiomelanocortin gene is expressed in many normal human tissues and in tumours not associated with ectopic ACTH syndrome. *Mol Endocrinol*, 1988; **2**: 862–70.
2. Baylin SB, Mendelsohn G. Time dependent changes in human tumours: implications for diagnosis and clinical behaviour. *Semin Oncol*, 1982; **9**: 504–12.
3. Gagner JP, Drouin J. Tissue-specific regulation of pituitary proopiomelanocortin gene transcription by corticotropin-releasing hormone, 3′,5′-cyclic adenosine monophosphate, and glucocorticoids. *Mol Endocrinol*, 1987; **1**: 677–82.
4. Lamolet B, Pulichino AM, Lamonerie T, *et al.* A pituitary cell-restricted T box factor, Tpit, activates *POMC* transcription in cooperation with Pitx homeoproteins. *Cell*, 2001; **104**: 849–59.
5. Maira M, Couture C, Le Martelot G, *et al.* The T-box factor Tpit recruits SRC/p160 co-activators and mediates hormone action. *J Biol Chem*, 2003; **278**: 46523–32.
6. Philips A, Lesage S, Gingras R, *et al.* Novel dimeric Nur77 signaling mechanism in endocrine and lymphoid cells. *Mol Cell Biol*, 1997; **17**: 5946–51.
7. Maira M, Martens C, Batsche E, *et al.* Dimer-specific potentiation of NGFI-B (Nur77) transcriptional activity by the protein kinase A pathway and AF-1-dependent coactivator recruitment. *Mol Cell Biol*, 2003; **23**: 763–76.
8. Drouin J, Trifiro MA, Plante RK, *et al.* Glucocorticoid receptor binding to a specific DNA sequence is required for hormone-dependent repression of pro-opiomelanocortin gene transcription. *Mol Cell Biol*, 1989; **9**: 5305–14.
9. Drouin J, Sun YL, Nemer M. Glucocorticoid repression of pro-opiomelanocortin gene transcription. *J Steroid Biochem*, 1989; **34**: 63–9.
10. Drouin J, Sun YL, Chamberland M, *et al.* Novel glucocorticoid receptor complex with DNA element of the hormone-repressed *POMC* gene. *EMBO J*, 1993; **12**: 145–56.
11. Riegel AT, Lu Y, Remenick J, *et al.* Proopiomelanocortin gene promoter elements required for constitutive and glucocorticoid-repressed transcription. *Mol Endocrinol*, 1991; **5**: 1973–82.
12. Vallette-Kasic S, Figarella-Branger D, Grino M, *et al.* Differential regulation of proopiomelanocortin and pituitary-restricted transcription factor (TPIT), a new marker of normal and adenomatous human corticotrophs. *J Clin Endocrinol Metab*, 2003; **88**: 3050–6.
13. Abou-Samra AB, Harwood JP, Manganiello VC, *et al.* Phorbol 12-myristate 13-acetate and vasopressin potentiate the effect of corticotropin-releasing factor on cyclic AMP production in rat anterior pituitary cells. Mechanisms of action. *J Biol Chem*, 1987; **262**: 1129–36.
14. Ray DW, Ren SG, Melmed S. Leukemia inhibitory factor (LIF) stimulates proopiomelanocortin (*POMC*) expression in a corticotroph cell line. Role of STAT pathway. *J Clin Invest*, 1996; **97**: 1852–9.
15. Stefana B, Ray DW, Melmed S. Leukemia inhibitory factor induces differentiation of pituitary corticotroph function: an immuno-neuroendocrine phenotypic switch. *Proc Natl Acad Sci U S A*, 1996; **93**: 12502–6.
16. Newell-Price J, King P, Clark AJ. The CpG island promoter of the human proopiomelanocortin gene is methylated in nonexpressing normal tissue and tumors and represses expression. *Mol Endocrinol*, 2001; **15**: 338–48.
17. Saito E, Iwasa S, Odell WD. Widespread presence of larger molecular weight adrenocorticotrophin-like substances in normal rat exrapituitary. *Endocrinology*, 1983; **113**: 1010–19.
18. Stewart PM, Gibson S, Crosby SR, *et al.* ACTH precursors characterise the ectopic ACTH syndrome. *Clin Endocrinol*, 1994; **40**: 199–204.
19. Liddle GW, Nicholson WF, Island DP. Clinical and laboratory studies of ectopic tumoral syndromes. *Recent Progr Hormone Res*, 1969; **25**: 283.
20. Koulouri O, Steuwe A, Gillett D, *et al.* A role for 11C-methionine PET imaging in ACTH-dependent Cushing's syndrome. *Eur J Endocrinol*, 2015; **173**: M107–20.
21. Gross AJ, Steinberg SM, Reilly JG, *et al.* Atrial natriuretic factor and arginine vasopressin production in tumour cell lines from patients with lung cancer and their relationship to serum sodium. *Cancer Res*, 1993; **53**: 67–74.
22. Berl T, Quittnat-Pelletier F, Verbalis JG, *et al.* Oral tolvaptan is safe and effective in chronic hyponatraemia. *J Am Soc Nephrol*, 2010; **21**: 705–12.
23. Bundred NJ, Walls J, Ratcliffe WA. Parathyroid hormone related protein, bony metastases and hypercalcaemia of malignancy. *Ann R Coll Surg Engl*, 1996; **78**: 354–8.
24. Firkin F, Seymour JF, Watson AM, Grill V, Martin TJ. Parathyroid hormone related protein in hypercalcaemia associated with haematological malignancy. *Br J Haematol*, 1996; **94**: 486–92.
25. Prager D, Rosenblatt JD, Ejima E. Hypercalcaemia, parathyroid hormone related protein and human T-cell leukaemia virus infection. *Leuk Lymphoma*, 1994; **14**: 395–400.
26. Frystyk J, Skjoerboek C, Zapf J, Orskov H. Increased levels of circulating free insulin-like growth factors in patients with non-islet cell tumour hypoglycaemia. *Diabetologia*, 1998; **41**: 589–94.
27. Zapf J. Role of insulin-like growth factor II and IGP binding proteins in extrapancreatic tumor hypoglycaemia. *Horm Res*, 1994; **42**: 20–6.
28. Baxter RC, Holman SR, Corbould A, Stranks S, Ho PJ, Braund W. Regulation of the insulin-like growth factors and their binding proteins by glucocorticoid and growth hormone in non-islet cell tumor hypoglycaemia. *J Clin Endocrinol Metab*, 1995; **80**: 2700–8.
29. Teale JD, Blum WF, Marks V. Alleviation of non-islet cell tumour hypoglycaemia by growth hormone therapy is associated with changes in IGF binding protein 3. *Ann Clin Biochem*, 1992; **29**: 314–23.
30. Ben-Jonathan N, Mershon JL, Allen DL, Steinmetz RW. Extrapituitary prolactin: distribution, regulation, functions and clinical aspects. *Endocr Rev*, 1996; **17**: 639–69.

13.1.3 Long-Term Endocrine Sequalae of Cancer Therapy

Claire E. Higham and Robert D. Murray

Introduction

The worldwide incidence of cancer is increasing, and it is predicted that around 20 million people will have a cancer diagnosis by 2025 [1]. In contrast, mortality from cancer is on the decline, dropping by 26% in the United States between 1991 and 2015 [2]. Thus, there are an increasing number of patients living with and beyond cancer (LWBC) who are susceptible to both short- and long-term cancer therapy related side effects.

Long-term endocrine sequelae of cancer therapies occur in up to 50% of childhood cancer survivors and may occur decades after the treatment has been delivered [3]. With over 30 000 patients registered in the United Kingdom as long-term survivors of childhood cancer [4], the importance of developing appropriate models of surveillance, follow-up, and management are clear.

The majority of research into the long-term sequelae of cancer therapies has been performed in childhood survivors, and is dominated by observational studies. Data are now emerging with regards to survivorship in patients treated as adults, where susceptibility to endocrinopathies is thought less prevalent, but still significant

This chapter will focus on describing the long-term endocrine sequelae in children and adults treated for cancer, the underlying pathophysiology, diagnostic, management, and surveillance strategies.

Endocrine late effects include disturbances of growth and puberty, hypothalamo-pituitary (HP) dysfunction, hypogonadism, thyroid dysfunction, benign and malignant thyroid nodules, hyperparathyroidism, and bone effects.

Growth and Puberty

Diagnosis and treatment of cancer during childhood is associated with reduced height velocity and final height. There are several contributors to this including the impact of acute illness, poor nutritional status, and ongoing cancer therapy. In addition, perturbations of the endocrine system, including radiation-induced hypothyroidism, precocious puberty, and growth hormone deficiency (GHD) impact adversely on growth.

Survivors of childhood malignancies who received cranial irradiation achieve final heights significantly below those predicted from parental heights even with irradiation doses as low as 18 Gy. Although radiation-induced GHD is an obvious cause for the abrogated growth it is not universally present in all children with impaired growth velocity, implicating additional mechanisms. A subanalysis of the Childhood Cancer Survivor Study (CCSS) survivors with brain tumours revealed 40% of patients to have a final height below the tenth percentile [5].

Spinal irradiation has a negative impact on growth above that of cranial irradiation alone, and relating directly to a reduction in vertebral height [6]. Greater impairment of spinal growth compared with the limbs results in an increase in leg length to sitting height ratio. This disproportion may be further amplified by growth hormone (GH) replacement therapy as GH impacts favourably on growth of the long bones, whereas the irradiated spine remains relatively resistant [7].

Precocious puberty is common in childhood brain tumour survivors, occurring in up to 12%. This depends on tumour location, presence of hydrocephalus, and low-dose cranial radiotherapy. If the diagnosis is delayed and puberty allowed to progress unabated, this can have significant decrement to final height, particularly when concomitant GHD is present.

Time allowing, catch-up growth is frequently observed once active treatment has been completed and remission achieved.

Pituitary

Deficiency of one or more pituitary hormones is commonly seen following treatment for central nervous system (CNS) tumours and cranial irradiation [8, 9]. It remains unclear whether standard chemotherapy contributes to pituitary dysfunction, with no definitive agent or class of agent implicated. Novel immunotherapy agents, in particular checkpoint inhibitor therapies (anti-CTLA-4, anti-PD1) are associated with hypophysitis and subsequent persistent pituitary hormonal deficiencies [10].

In children and adolescents, tumours such as medulloblastoma, ependymoma, choroid plexus tumour, astrocytoma, low-grade glioma, and primitive neuroectodermal tumour (PNET) may result in HP dysfunction dependent on tumour localization and treatment [11]. The spectrum of brain tumours in the adult population is different (a higher percentage of high-grade glioma and meningioma) and overall survival significantly lower, although those with low-grade gliomas have a more than 70% 5-year survival. Up to 90% develop a variable degree of hypopituitarism post-treatment [12]. In addition, both children and adults treated with irradiation for head and neck carcinomas or CNS tumours distant to the HP region (e.g. retinoblastoma) are at risk if the HP axis is included within the radiation field [11]. Low dose cranial irradiation in children with acute lymphoblastic leukaemia (ALL) or as part of total body irradiation (TBI) in stem cell transplantation is also a risk, although appears less of a concern in adults [13].

Greater than 50% of childhood cancer survivors exposed to cranial radiotherapy will develop at least one pituitary hormone deficiency by 25 years post-exposure. Doses of radiation lees than 30 Gy

are commonly associated with isolated GHD. Higher radiation doses can result in further damage, commonly causing gonadotrophin deficiency with thyroid-stimulating hormone (TSH) and adrenocorticotrophic hormone (ACTH) deficiency less frequent. Diabetes insipidus has not been described as a consequence of HP radiotherapy [14, 15].

Time since irradiation, age at time of radiotherapy, fraction size and radiation schedule are important determinants of subsequent HP function [14, 15].

Growth Hormone Deficiency (GHD)

The somatotropic axis is the most vulnerable to radiation damage. Symptoms and signs of GHD in the LWBC population do not differ from those described in non-cancer populations. In prepubertal children poor growth and development predominates (see Section 7, 'Disorders of Growth and Development and Transition'), whereas weight gain, fatigue, reduced muscle mass, and poor quality of life are indicators in adults. Diagnosis is complex in these individuals as symptomatology may overlap with other cancer related complications [15, 16].

Dynamic testing of the GH axis is generally required in children, adolescents, and adults to confirm the diagnosis; testing should not differ from that in the non-cancer population, with an insulin tolerance test (ITT) being the gold standard. There should be the recognition that if only hypothalamic damage is present post-radiotherapy, then the GHRH-arginine testing may be falsely reassuring [16]. Dynamic testing is not required in patients with two or more additional pituitary hormone deficits.

In those with isolated GHD in childhood after cranial radiotherapy, re-testing of the GH axis is recommended following achievement of final height as partial recovery of the axis is possible and lower thresholds for GH treatment are used in the paediatric population [15–17]. Low IGF-I or IGF-BP3 does not reliably predict GHD post-cancer therapy but may be a surrogate marker for severity [15–17].

GH replacement in prepubertal children with poor growth and documented GHD following cancer therapy is well accepted and generally started 1–2 years following completion of cancer therapy. Adequate treatment improves final adult height, although full height potential may not be achieved as a result of other endocrine dysfunction, chronic illness, and spinal or direct bone irradiation effects. Data concerning the safety of GH in childhood cancer survivors are reassuring with regards to recurrence of primary, or emergence of secondary tumours, although caution and careful monitoring is advised [15–17].

GH replacement therapy in GH-deficient survivors of cancers arising in adulthood remains controversial and there are minimal safety data available. The data that are available are reassuring. The positive impact of GH replacement on quality of life is well described [18, 19].

Cranial irradiation doses of less than 50 Gy can result in precocious or early puberty in children. Both genders are affected with irradiation doses employed in the treatment of brain tumours (25–50Gy), whereas lower doses used for prophylaxis in treatment of ALL results in a predominance of girls developing precocious puberty. The mechanism responsible for early puberty is thought to be a disinhibition of cortical influences on the hypothalamus allowing gonadotropin-releasing hormone (GnRH) pulse frequency and amplitude to increase prematurely. The impact of early puberty in a child with radiation-induced GHD is to foreshorten the time available for GH therapy and thereby restrict the therapeutic efficacy of this intervention. These latter children may therefore require a combination of GnRH analogues alongside GH replacement to optimize their height potential.

Gonadotropin Deficiency

Gonadotropins are the second most frequently affected anterior pituitary hormones following cranial irradiation, although deficiency is uncommon with radiation doses less than 40 Gy. In children and adolescents, deficiency may present with pubertal delay, arrested puberty, or primary amenorrhoea. Adults experience secondary amenorrhea, impairment of libido, and/or erectile dysfunction.

Diagnosis relies on baseline blood sampling of LH, FSH, oestradiol (females) or testosterone (males). These reveal low or inappropriately normal gonadotropins in association with low peripheral total and/or bioactive sex hormones [20].

Treatment with sex hormones in gonadotropin deficient patients is recommended to induce puberty if required, to prevent cardiovascular complications, and reduce loss of bone mass in premenopausal women and men, in line with current guidance [21, 22].

ACTH Deficiency

The ACTH axis is more resilient to irradiation-induced damage than either the GH or gonadotropin axes. There are only occasional reports of ACTH deficiency following TBI (9.0–15.0 Gy) used as preconditioning before bone marrow transplant (BMT) or prophylactic cranial irradiation during treatment of ALL (18–24 Gy) [21]. Even with cranial radiation doses up to 50 Gy only around 3% of children develop ACTH deficiency, though the incidence increases dramatically with doses more than 50 Gy. ACTH deficiency tends to occur late necessitating continued awareness and screening beyond 10 years after treatment of the primary disease [23].

ACTH deficiency leads to non-specific symptoms including fatigue, weight loss, gastrointestinal symptoms (epigastric pain, nausea, and vomiting), and lack of resistance to infection. Symptoms are commonly mistaken as sequelae of the tumour/treatment or as psychological problems. Life-threatening adrenal crisis may develop in stressful situations such as infections or surgery. Morning plasma cortisol measurements within the reference range are reassuring (generally >200 nmol/L, although this is local assay, and reference range, dependent), but only dynamic stimulation testing allows definitive diagnosis of the condition when there is clinical suspicion [8].

Treatment of ACTH deficiency does not differ from standard recommendations for replacement therapy. ACTH deficiency should be diagnosed and treated prior to instigating treatment with thyroxine or GH to prevent triggering an adrenal crisis (reviewed in [8, 15]).

TSH Deficiency

Damage to the TSH axis can develop over time post-treatment, although it is considered the least vulnerable of the anterior pituitary axes to radiation-induced damage. Overt secondary hypothyroidism is uncommon with irradiation doses below 50 Gy [22]. Central hypothyroidism is suspected if there is a low free thyroxine (fT_4) in association with low or inappropriately normal TSH. TSH cannot therefore

be used for screening in central hypothyroidism and a high index of clinical suspicion is required along with consideration of both fT_4 and TSH levels. Slightly elevated TSH levels can occur in central hypothyroidism as a consequence of an alteration in the predominant form of TSH secreted, resulting in an alteration in the ratio of bioactive/immunoreactive TSH. There is no evidence to support routine use of the thyrotropin-releasing hormone (TRH) test or assessment of TSH surge to improve diagnostic sensitivity and specificity [24].

Treatment of central hypothyroidism with l-thyroxine substitution should be according to published guideline. Titration of dose relies on measurement of fT_4 levels rather than TSH. Patients should be screened for adrenal insufficiency before the initiation of l-thyroxine to avoid adrenal crisis [8, 15, 25].

Prolactin

Hyperprolactinemia can occur after cranial irradiation; it is usually mild, transient, and rarely symptomatic, although should be considered as a factor in impaired gonadotropin secretion, if present. It most frequently occurs in adult female cancer survivors treated with high cranial radiation dose. Increasing or very high levels of prolactin are unusual and should prompt pituitary imaging [15].

Thyroid

Thyroid pathology following cancer therapy can relate to abnormalities of TSH secretion (hypopituitarism), a direct effect on the thyroid gland, or a combination of both (for example patients with radiation exposure to the head and neck region as well as the supraclavicular or chest are at risk for both primary and secondary thyroid dysfunction). Thyroid anomalies occur following cancer treatment in childhood and adulthood. Primary thyroid abnormalities are very common, and may present as primary hypothyroidism secondary to thyroid atrophy, autoimmune thyroid disease (hypothyroidism, Graves' disease, Graves' ophthalmopathy), thyroiditis, or thyroid nodules (benign and malignant).

Disorders of Thyroid Function

Direct irradiation of the thyroid gland causes both acute and chronic functional changes. Thyroid volume reduction has been observed during the first 6 months post-treatment and depends on received thyroid dose, with higher doses of radiation more likely to cause primary hypothyroidism (PH). The percentage of thyroid volume receiving more than 30–40 Gy is predictive of PH after radiotherapy and dose-volume constraints might reduce the risk [26].

In contrast, low dose radiotherapy increases the risk for autoimmune thyroid disorders which can manifest as either hypothyroidism (Hashimoto's disease) or hyperthyroidism (Graves' disease). The underlying pathophysiology for this is not well understood [27]. Gender (F > M) race (Caucasian survivors are 2.5× more likely to develop PH) and age at exposure (children and adolescents) are all determinants of developing thyroid dysfunction [28]. Following radiotherapy of more than 20 Gy to the thyroid, it is recommended that thyroid function tests (TFTs) are performed annually, or earlier if symptoms occur [15].

Conventional chemotherapeutic agents may play a role in thyroid dysfunction but this is less clearly defined; being described after the use of high-dose chemotherapy in stem cell transplantation, with alkylating agents, or after combined vincristine-cisplatin usage. Chemotherapy additionally may increase the risk of irradiation-induced hypothyroidism [29].

Newer chemotherapeutic agents such as the multitargeted receptor tyrosine kinase (RTK) inhibitors (e.g. sunitinib and sorafenib), and checkpoint inhibitors (e.g. PD1, PDL-1, and CTLA4 inhibitors) frequently result in thyroid dysfunction, which may require long-term thyroid hormone replacement [10, 30]. Regular monitoring of thyroid function is required during treatment with RTK and checkpoint inhibitor therapies.

Thyroid Nodules and Secondary Thyroid Malignancy

Thyroid nodules are common in the general population. Radiotherapy to the thyroid increases the prevalence of nodules, and the likelihood of thyroid nodules becoming malignant. The interval post-radiotherapy until the development of thyroid nodules and/or thyroid carcinoma is long (15–30 years) with a minimum latency period of 5 to 10 years. The risk is greater with young age at exposure, time since exposure, and dose of radiation received; with peak risk following a radiation dose between 10 and 30 Gy. Risk increases linearly with doses up to 20 Gy and decreases at higher doses of radiation, but remains significantly elevated even above 50 Gy. Most thyroid cancers histologically are papillary, with occasional cases of follicular, carcinoma. Prognosis is generally good, although there may be a significantly inferior survival in younger patients [31].

The appropriate screening for thyroid nodules following radiotherapy remains controversial. Routine ultrasound risks overdiagnosis of benign nodules and associated anxiety in a vulnerable group, whereas thyroid palpation risks missing smaller and more diffuse lesions. The Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium argue for an individualized surveillance strategy using a decision aid [32].

Parathyroid

Irradiation involving the neck increases the risk of primary hyperparathyroidism (PHPT). Higher radiation doses result in an increased probability of developing PHPT. PHPT tends to emerge decades after radiation exposure, and later than intrinsic thyroid disease, although the latency between radiotherapy and PHPT might be shorter in patients receiving high-dose radiotherapy for malignancies. Younger age at irradiation and female gender are additional risk factors.

The pathophysiology of primary hyperparathyroidism in this setting is unclear, but diagnosis and management follow standard guidance for primary hyperparathyroidism. A single parathyroid adenoma is often the underlying cause. It is suggested that annual calcium and parathyroid hormone (PTH) measurements are performed in at risk patients [15, 33].

Ovary

Premature ovarian failure (POF) can result from surgical intervention, chemotherapy regimens, and radiotherapy where the field includes ovarian tissue. Combinations of treatment with radiotherapy

and chemotherapy, particularly alkylating agents, imparts a greater risk of POF than either modality alone.

POF occurs in up to 83% of teenage and young adult cancer survivors following chemotherapy, although normal ovarian function may persist for a variable period. Important risk factors are alkylating agents, older age at chemotherapy, and cumulative dose. For example, in 620 young adult women receiving chemotherapy alone; 8%, 10%, 9%, and 5% of patients with Hodgkin's, non-Hodgkin's disease, breast and gastrointestinal cancer respectively developed POF [34].

POF results from direct therapeutic radiation to the ovary, or where the ovary is unavoidably within the radiation field, for example: spinal, abdominal, pelvic, vaginal, bladder, iliac lymph node irradiation, total lymphoid system, whole spine, or TBI. Relative risk of POF depends on radiation dose and age at initial therapy. In prepubertal girls, doses as low as 5 Gy may affect later ovarian function. The average doses required to induce sterility decline with age, being 20.3 Gy in infants, 18.4 Gy at age 10 years, and 16.5 Gy at age 20 years.

Radiotherapy related uterine damage is a further consideration. Radiotherapy can reduce uterine volume or impair growth, cause altered vasculature, and a reduced endometrium, dependent on age at irradiation and dosage received. A dose of 14–30 Gy has been shown to alter uterine vascularization and muscular elasticity, however, lower doses given during TBI can also be detrimental to the uterus. Radiation exposure of the uterus is associated with a markedly increased risk for miscarriage, preterm birth, and low birth weight [35].

Untreated POF leads to symptoms of oestrogen deficiency and prematurely reduced bone mineral density. Standard practice is to replace females with oestrogen plus progesterone (if there is potential for an endometrial lining). Replacement therapy is generally not recommended in those with hormone sensitive tumours, although the evidence to support this remains elusive. Optimal hormone replacement strategies have not been well studied. Regimes range from high doses of ethinyl oestradiol in the oral contraceptive pill to low doses of transdermal oestradiol, and will depend on patient preference and individual risk profile (reviewed in [14]).

Testes

The vast majority of primary gonadal late toxicity is due to chemotherapy and/or radiation, and occurs as a result of toxicity to Leydig cells, germ cells, or both.

Germ cells are extremely sensitive to toxicity from chemotherapy and radiotherapy. Chemotherapeutic agents include alkylating agents such as cyclophosphamide and procarbazine. Impaired fertility and reduced sperm count can occur with cyclophosphamide doses of 200–300 mg/m^2 in prepubertal boys and 100 mg/m^2 in adults. Doses exceeding 6 g in prepubertal boys are associated with the highest risk of azoospermia. It is noteworthy that recovery from azoospermia can occur many years after therapy [36]. Doses of testicular radiotherapy as low as 0.1 Gy have been associated with inhibition of spermatogenesis, which may recover; but doses exceeding 2 Gy most commonly lead to permanent impairment. Patients at risk therefore include those receiving direct testicular radiation, but potentially also those having TBI or pelvic radiotherapy.

Leydig cells are less sensitive to damage from chemotherapy or radiotherapy, with irradiation doses of up to 24 Gy tolerated by prepubertal testes and up to 30 Gy during puberty. Combination chemotherapy and radiotherapy regimens are more likely to lead to testosterone deficiency (e.g. cumulative cyclophosphamide doses of more than 20 mg/m^2 and 14 Gy of testicular radiotherapy). Subclinical testosterone deficiency is relatively common, and under-recognized.

As a consequence of the differential sensitivities of the Leydig and germ cells of the testis; puberty generally progresses normally in children and secondary sexual characteristics are maintained in most adults who received irradiation. Oligo- or azoospermia frequently accompany normal androgenization. Testicular volumes are small reflecting damage to the germinal epithelium, and cannot be relied upon for pubertal staging.

Available guidance suggests at risk patients should be screened using Tanner staging of pubertal development every 6 months until sexual maturity. Adults should be monitored annually for symptoms of androgen deficiency; decreased libido, reduction in spontaneous erections, gynaecomastia, loss of body hair, reduced muscle bulk, hot flushes/sweating, and reduced testicular volume. Diagnosis involves measurement of morning total testosterone and LH/FSH. Borderline results require further investigation with sex hormone-binding globulin (SHBG) and calculations of free testosterone levels.

Adolescents and adult cancer survivors with confirmed hypogonadism should be replaced with testosterone using standard treatment regimes. Monitoring follows standard guidelines [37]. Fertility in both males and females should be considered and early referral to fertility services for assessment is recommended.

Bone

Low bone mineral density (BMD) is well described as a consequence of treatment for childhood, adolescent, and adult cancers. Whether this low BMD contributes to an increased risk of fragility fracture in cancer survivors is less well defined.

Poor bone health in cancer survivors can relate to the original cancer, treatment modalities received, and hormonal deficiencies. Cachexia, anorexia, reduced weight-bearing exercise, and immobility also potentially contribute to low BMD and increased fracture risk. In addition, risk factors for some cancers (e.g. smoking, alcohol excess, type 2 diabetes) are well-recognized lifestyle risk factors for low BMD and fragility fracture. Cancer and the treatment for cancer during childhood and adolescence can halt bone accrual and prevent or delay the attainment of peak bone mass. This has the potential to be associated with increased fracture risk in later life.

Glucocorticoid (GC) use is common to a number of oncology regimens and supportive treatments. Glucocorticoid-induced bone loss and increased fracture risk is well described although understanding of the pathophysiology still remains incomplete [38]. GCs cause rapid onset bone loss, within 30 days as shown in ALL patients, and effects are dose dependent. In general terms, glucocorticoid-induced bone loss is, at least partially, reversible. Risk of fracture in children with ALL treated with GC therapy related to higher GC dose, recent average increase in GC dose and increase in GC dose intensity. Type of GC treatment might also be important, with

dexamethasone causing more bone morbidity than prednisolone ([39, 40] reviewed in [15]).

High-dose methotrexate (doses above 40 000 mg/m^2) in children is associated with osteopathy, bone pain, and increased fracture risk. Animal studies demonstrate reduced osteoblast proliferation, increased osteoclast formation and increased marrow adiposity with methotrexate [15, 41]. Other chemotherapy agents; doxorubicin, etoposide, vincristine, asparaginase, cyclosporin, tacrolimus, 6-mercaptopurine, and cyclophosphamide, have been postulated to cause direct bone toxicity, however, it is difficult to isolate direct effects of chemotherapy from other contributory factors such as glucocorticoids, radiotherapy, and chemotherapy-induced hypogonadism [15, 42].

Cranial radiotherapy in children with ALL is associated with low BMD, initially described as the presence of reduced lumbar volumetric BMD in patients receiving more than 18 Gy of cranial radiotherapy. Subsequent studies demonstrated BMD inversely correlated to the dose of irradiation received. This bone toxicity has been in part attributed to radiation-induced GHD. Furthermore, low total body and lumbar spine BMD was found in children with posterior fossa tumours treated with craniospinal radiotherapy or posterior fossa radiotherapy independent of the presence of GHD or treatment with chemotherapy [15, 43].

Focal radiotherapy in children and adults is associated with increased fracture risk within the radiation field, usually when doses exceed 40 Gy. Rib fractures are common following radiotherapy to the chest for hepatocellular carcinoma, breast, and lung cancers. Pelvic insufficiency fractures have been described in between 1.7% and 89% of patients receiving pelvic radiotherapy [44].

A full understanding of the pathophysiology underlying the direct effects of radiation on bone is still elusive. *In vitro* studies demonstrate radiation-induced reduction in osteoblast number, osteoblast cell cycle progression, and increased apoptosis. Data for osteoclasts are less clear. Both a reduction and increase in osteoclast number and activity have been described [44].

Patients undergoing haematopoietic stem cell transplant are at high risk of reduced BMD and fracture. This risk is multifactorial; use of pretransplantation chemotherapy, glucocorticoids, TBI, endocrine dysfunction, immobilization, and low body mass index (BMI). Bone loss occurs rapidly and mainly within the first year following haematopoietic stem cell transplantation (HSCT). There is some recovery of bone health with time but most patients do not fully normalize BMD and remain at increased risk of fracture [45].

Patients at risk of poor bone health should have an assessment of bone density and fracture risk assessment. Gold standard technique for measurement of BMD is a DEXA scan, although in children and those with small bone volume, QCT may be more appropriate, providing a volumetric rather than areal density. The impact of childhood cancer on ability to attain peak bone mass should be taken into account.

There is little evidence to guide the management of bone health in cancer survivors. Systematic review did not provide overall convincing evidence for the use of physical exercise training in improving BMD in childhood cancer patients [46].

Similarly, in adults, a meta-analysis of exercise programmes in breast and prostate cancer survivors; including over 800 patients, showed no overall benefit demonstrated on lumbar spine, total hip, or femoral neck BMD. There was a trend for a beneficial effect in those undertaking resistance/impact exercise [47].

Vitamin D insufficiency and deficiency have been reported in both childhood/adolescent and adult cancer survivors, although whether this exceeds that of the general background population is controversial. There are no data in cancer populations that replacement with vitamin D improves BMD or reduces fracture risk although standard practice to maximize bone health would be adequate calcium intake (700–1000 mg per day) and cholecalciferol 400–800 IU per day.

Low BMD and bone turnover can be improved in children with ALL and Non-Hodgkin lymphoma using bisphosphonate therapy during or following chemotherapy treatment. However, there are no data with regards to fracture and there are general concerns about using antiresorptive therapy in growing bones and in fertile young women—more long-term studies are required before these therapies can be recommended routinely [48].

Bisphosphonate and denosumab therapy have been shown to improve BMD and in the case of denosumab, reduce fracture risk, in post-menopausal breast cancer patients taking aromatase inhibitors and prostate cancer patients taking androgen deprivation therapy [49]. It is well demonstrated that bisphosphonate therapy improves hip and spine BMD and prevents vertebral fracture in glucocorticoid-induced osteoporosis, although there are no specific studies in patients LWBC.

For HSCT patients, improvements in BMD following 12 months of bisphosphonate therapy have been demonstrated, although again these studies were not sufficiently powered or of sufficient duration to comment on fracture prevention. There are no clinical data currently to support the use of antiresorptive therapy for insufficiency fractures secondary to focal radiotherapy [50].

Conclusion

The number of patients diagnosed and surviving cancer is increasing. Each patient with cancer has a unique phenotype and will receive an individualized treatment regimen, leading to specific risks of acute and long-term effects of cancer therapies, including the endocrine dysfunction described in this chapter. Metabolic disturbance, including diabetes, lipid abnormalities, and increased cardiovascular and cerebrovascular risk are also common sequelae to be considered. Future studies are required to determine contributory factors and biomarkers to identify patients at risk of endocrine and metabolic dysfunction in order to determine appropriate long-term surveillance strategies. Furthermore, randomized controlled trials of preventative and therapeutic interventions in this group are needed to maximize quality of life and minimize the current increased morbidity and mortality in this group.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, 2015; **136**: E359–86.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*, 2018; **68**: 7–30.
3. Brignardello E, Felicetti F, Castiglione A, *et al.* Endocrine health conditions in adult survivors of childhood cancer: the need for specialized adult-focused follow-up clinics. *Eur J Endocrinol*, 2013; **168**: 465–72.
4. Cancer CRUC. Cancer statistics for the UK. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers—heading-Four (accessed: 7 March 2020).
5. Gurney JG, Ness KK, Stovall M, *et al.* Final height and body mass index among adult survivors of childhood brain cancer: childhood cancer survivor study. *J Clin Endocrinol Metab*, 2003; **88**: 4731–9.
6. Shalet SM, Gibson B, Swindell R, Pearson D. Effect of spinal irradiation on growth. *Arch Dis Child*, 1987; **62**: 461–4.
7. Clayton PE, Shalet SM, Price DA. Growth response to growth hormone therapy following craniospinal irradiation. *Eur J Pediatr*, 1988; **147**: 597–601.
8. Higham CE, Johannsson G, Shalet SM. Hypopituitarism. *Lancet*, 2016; **388**: 2403–15.
9. Clement SC, Schouten-van Meeteren AY, Boot AM, *et al.* Prevalence and risk factors of early endocrine disorders in childhood brain tumor survivors: a nationwide, multicenter study. *J Clin Oncol*, 2016; **34**: 4362–70.
10. Sznol M, Postow MA, Davies MJ, *et al.* Endocrine-related adverse events associated with immune checkpoint blockade and expert insights on their management. *Cancer Treat Rev*, 2017; **58**: 70–6.
11. Armstrong GT, Liu Q, Yasui Y, *et al.* Long-term outcomes among adult survivors of childhood central nervous system malignancies in the Childhood Cancer Survivor Study. *J Natl Cancer Inst*, 2009; **101**: 946–58.
12. Kyriakakis N, Lynch J, Orme SM, *et al.* Pituitary dysfunction following cranial radiotherapy for adult-onset nonpituitary brain tumours. *Clin Endocrinol (Oxf)*, 2016; **84**: 372–9.
13. Brennan BM, Shalet SM. Endocrine late effects after bone marrow transplant. *Br J Haematol*, 2002; **118**: 58–66.
14. Chemaitilly W, Li Z, Huang S, *et al.* Anterior hypopituitarism in adult survivors of childhood cancers treated with cranial radiotherapy: a report from the St Jude Lifetime Cohort study. *J Clin Oncol*, 2015; **33**: 492–500.
15. Gebauer J, Higham C, Langer T, Denzer C, Brabant G. Long-term endocrine and metabolic consequences of cancer treatment: a systematic review. *Endocr Rev*, 2019; **40**: 711–67.
16. Sfeir JG, Kittah NEN, Tamhane SU, Jasim S, Chemaitilly W, Cohen LE, Murad MH. Diagnosis of GH deficiency as a late effect of radiotherapy in survivors of childhood cancers. *J Clin Endocrinol Metab*, 2018; **103**: 2785–93.
17. Tamhane S, Sfeir JG, Kittah NEN, *et al.* GH therapy in childhood cancer survivors: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2018; **103**: 2794–801.
18. Child CJ, Conroy D, Zimmermann AG, Woodmansee WW, Erfurth EM, Robison LL. Incidence of primary cancers and intracranial tumour recurrences in GH-treated and untreated adult hypopituitary patients: analyses from the Hypopituitary Control and Complications Study. *Eur J Endocrinol*, 2015; **172**: 779–90.
19. Murray RD, Darzy KH, Gleeson HK, Shalet SM. GH-deficient survivors of childhood cancer: GH replacement during adult life. *J Clin Endocrinol Metab*, 2002; **87**: 129–35.
20. Constine LS, Woolf PD, Cann D, *et al.* Hypothalamic-pituitary dysfunction after radiation for brain tumors. *N Engl J Med*, 1993; **328**: 87–94.
21. Gordon CM, Ackerman KE, Berga SL, *et al.* Functional hypothalamic amenorrhea: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 1413–39.
22. Bhasin S, Cunningham GR, Hayes FJ, *et al.* Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2010; **95**: 2536–59.
23. Littley MD, Shalet SM, Morgenstern GR, Deakin DP. Endocrine and reproductive dysfunction following fractionated total body irradiation in adults. *Q J Med*, 1991; **78**: 265–74.
24. Carter Y, Sippel RS, Chen H. Hypothyroidism after a cancer diagnosis: etiology, diagnosis, complications, and management. *Oncologist*, 2014; **19**: 34–43.
25. Parretti H, Okosieme O, Vanderpump M. Current recommendations in the management of hypothyroidism: developed from a statement by the British Thyroid Association Executive. *Br J Gen Pract*, 2016; **66**: 538–40.
26. Sommat K, Ong WS, Hussain A, *et al.* Thyroid V40 predicts primary hypothyroidism after intensity modulated radiation therapy for nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*, 2017; **98**: 574–80.
27. Kikawa Y, Takeuchi M, Sudo M, Iida Y, Kasagi K, Konishi J. Development of primary hypothyroidism with antithyroglobulin, antiperoxidase, and blocking-type thyrotropin receptor antibodies after radiation therapy for neuroblastoma. *J Pediatr*, 1996; **129**: 909–12.
28. Green DM, Brecher ML, Yakar D, *et al.* Thyroid function in pediatric patients after neck irradiation for Hodgkin disease. *Med Pediatr Oncol*, 1980; **8**: 127–36.
29. van Santen HM, Vulsma T, Dijkgraaf MG, *et al.* No damaging effect of chemotherapy in addition to radiotherapy on the thyroid axis in young adult survivors of childhood cancer. *J Clin Endocrinol Metab*, 2003; **88**: 3657–63.
30. Sabatier R, Eymard JC, Walz J, *et al.* Could thyroid dysfunction influence outcome in sunitinib-treated metastatic renal cell carcinoma? *Ann Oncol*, 2012; **23**: 714–21.
31. Iglesias ML, Schmidt A, Ghuzlan AA, *et al.* Radiation exposure and thyroid cancer: a review. *Arch Endocrinol Metab*, 2017; **61**: 180–7.
32. Clement SC, Kremer LCM, Verburg FA, *et al.* Balancing the benefits and harms of thyroid cancer surveillance in survivors of Childhood, adolescent and young adult cancer: recommendations from the International Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium. *Cancer Treat Rev*, 2018; **63**: 28–39.
33. Rao SD, Frame B, Miller MJ, Kleerekoper M, Block MA, Parfitt AM. Hyperparathyroidism following head and neck irradiation. *Arch Intern Med*, 1980; **140**: 205–7.
34. Metzger ML, Meacham LR, Patterson B, *et al.* Female reproductive health after childhood, adolescent, and young adult cancers: guidelines for the assessment and management of female reproductive complications. *J Clin Oncol*, 2013; **31**: 1239–47.
35. Critchley HO, Bath LE, Wallace WH. Radiation damage to the uterus—review of the effects of treatment of childhood cancer. *Hum Fertil (Camb)*, 2002; **5**: 61–6.
36. Kenney LB, Cohen LE, Shnorhavorian M, *et al.* Male reproductive health after childhood, adolescent, and young adult

cancers: a report from the Children's Oncology Group. *J Clin Oncol*, 2012; **30**: 3408–16.
37. Skinner R, Mulder RL, Kremer LC, *et al.* Recommendations for gonadotoxicity surveillance in male childhood, adolescent, and young adult cancer survivors: a report from the International Late Effects of Childhood Cancer Guideline Harmonization Group in collaboration with the PanCareSurFup Consortium. *Lancet Oncol*, 2017; **18**: e75–e90.
38. Hardy RS, Zhou H, Seibel MJ, Cooper MS. Glucocorticoids and bone: consequences of endogenous and exogenous excess and replacement therapy. *Endocr Rev*, 2018; **39**: 519–48.
39. Orgel E, Mueske NM, Wren TA, *et al.* Early injury to cortical and cancellous bone from induction chemotherapy for adolescents and young adults treated for acute lymphoblastic leukemia. *Bone*, 2016; **85**: 131–6.
40. Cummings EA, Ma J, Fernandez CV, *et al.* Incident vertebral fractures in children with leukemia during the four years following diagnosis. *J Clin Endocrinol Metab*, 2015; **100**: 3408–17.
41. Fan C, Georgiou KR, King TJ, Xian CJ. Methotrexate toxicity in growing long bones of young rats: a model for studying cancer chemotherapy-induced bone growth defects in children. *J Biomed Biotechnol*, 2011; **2011**: 903097.
42. Rellick SL, O'Leary H, Piktel D, *et al.* Bone marrow osteoblast damage by chemotherapeutic agents. *PLoS One*, 2012; **7**: e30758.
43. Gurney JG, Kaste SC, Liu W, *et al.* Bone mineral density among long-term survivors of childhood acute lymphoblastic leukemia: results from the St. Jude Lifetime Cohort Study. *Pediatr Blood Cancer*, 2014; **61**: 1270–6.
44. Higham CE, Faithfull S. Bone health and pelvic radiotherapy. *Clin Oncol (R Coll Radiol)*, 2015; **27**: 668–78.
45. Pundole X, Cheema HI, Petitto GS, Lopez-Olivo MA, Suarez-Almazor ME, Lu H. Prevention and treatment of bone loss and fractures in patients undergoing a hematopoietic stem cell transplant: a systematic review and meta-analysis. *Bone Marrow Transplant*, 2017; **52**: 663–70.
46. Braam KI, van der Torre P, Takken T, Veening MA, van Dulmen-den Broeder E, Kaspers GJ. Physical exercise training interventions for children and young adults during and after treatment for childhood cancer. *Cochrane Database Syst Rev*, 2016; **3**: CD008796.
47. Dalla Via J, Daly RM, Fraser SF. The effect of exercise on bone mineral density in adult cancer survivors: a systematic review and meta-analysis. *Osteoporos Int*, 2018; **29**: 287–303.
48. Lee JM, Kim JE, Bae SH, Hah JO. Efficacy of pamidronate in children with low bone mineral density during and after chemotherapy for acute lymphoblastic leukemia and non-Hodgkin lymphoma. *Blood Res*, 2013; **48**: 99–106.
49. Rachner TD, Coleman R, Hadji P, Hofbauer LC. Bone health during endocrine therapy for cancer. *Lancet Diabetes Endocrinol*, 2018; **6**: 901–10.
50. van den Blink QU, Garcez K, Henson CC, Davidson SE, Higham CE. Pharmacological interventions for the prevention of insufficiency fractures and avascular necrosis associated with pelvic radiotherapy in adults. *Cochrane Database Syst Rev*, 2018; **4**: CD010604.

13.1.4 Endocrine Complications of Biological Cancer Therapies

Carla Moran

Introduction

Improved understanding of cancer biology over the past decade has led to a revolution in the immunotherapy field, with development of drugs targeted towards molecules mediating 'immune-silencing' in the tumour environment [1]. These treatments, collectively termed immune checkpoint inhibitors, are antibodies directed against cytotoxic T-lymphocyte antigen 4 (CTLA-4), programmed cell death protein 1 (PD-1) or its ligand (PD-L1), which work by regulating the natural checkpoints of the immune system. They boost immune antitumour activity, resulting in higher treatment responses and better survival rates in a number of metastatic malignancies, than seen hitherto.

Immunotherapeutic agents can have varying on- and off-target toxicities related to their affinity and specificity, respectively. Toxicities common to all checkpoint inhibitors include fevers, chills, diarrhoea, colitis, skin rash, hepatitis, pneumonitis, eye, and endocrine disorders. Although endocrinopathies are common, they are difficult to detect but, unlike other immune-related adverse events (IrAEs), are rarely life-threatening.

CTLA-4 Inhibitors

Ipilimumab, a fully human monoclonal IgG1 antibody against CTLA-4 potentiates T-cell activation by blocking the interaction between CTLA-4 and the costimulatory ligand B7 expressed on antigen-presenting cells, thus enhancing the T-cell response to tumour (see **Figure 13.1.4.1**) [1]. It was the first checkpoint inhibitor to obtain approval for the treatment of advanced melanoma, but is associated with severe and life-threatening IrAEs in 10–15% of patients [2]. Endocrine adverse events are reported at a relatively low frequency (see **Table 13.1.4.1**), but these side effects last longer than

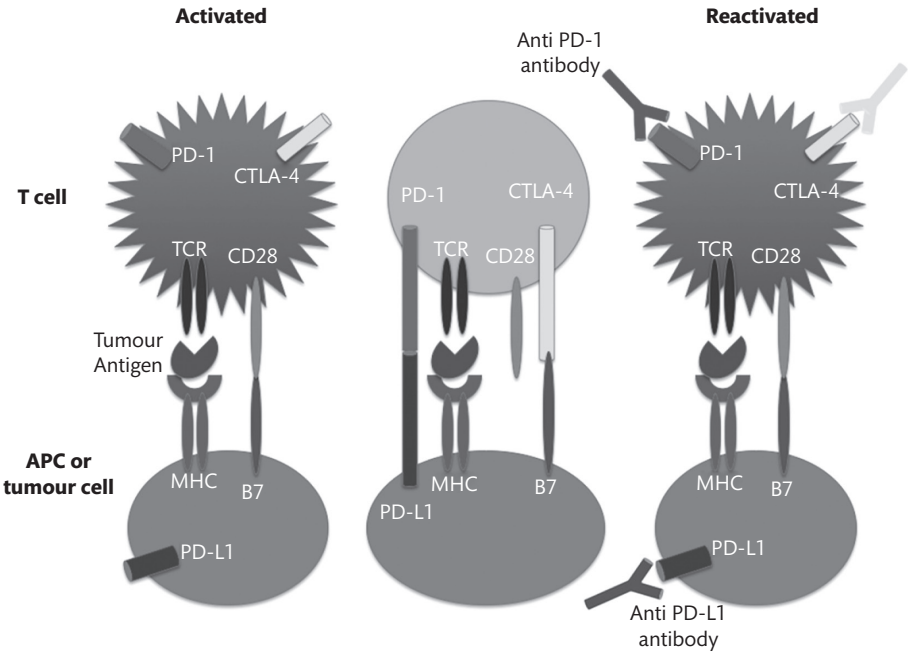


Figure 13.1.4.1 Mechanism of action of immune checkpoint inhibitors. On presentation of tumour antigen to T cells by the MHC complex on antigen-presenting or tumour cells, interaction of costimulatory molecules (CD28 and B7) results in T-cell activation. This process is halted by the interaction of PD-1 with its ligand, PD-L1, and CTLA-4 with B7. The presence of antibodies directed against either PD-(L)1 or CTLA-4 blocks their interaction with their partners on the APC or tumour cell and T-cell activation is therefore perpetuated. Abbreviations: TCR, T-cell receptor; CTLA-4, cytotoxic T-lymphocyte antigen-4; PD-(L)1, programmed cell death 1 (Ligand); CD28, cluster of differentiation 28; APC, antigen-presenting cell; MHC, major histocompatibility complex.

other IrAEs; moreover, although there may be an initial response to steroids, endocrine function does not usually recover.

Tremelimumab is another human monoclonal antibody (IgG2) against CTLA-4 currently being investigated for several types of malignancies. With a similar mechanism of action to ipilimumab, tremelimumab appears to have a similar toxicity profile, with one difference being that the latter has also been associated with Graves' disease (GD) [3].

CTLA-4 Inhibitors and Hypophysitis

Hypophysitis is the most common endocrinopathy associated with CTLA-4 antibody therapy, reported in 2–16% of treated individuals [2, 4–7]. A drug dose response relationship for this side effect has been reported in some [8], but not all, studies [9]. Time of onset is very variable, with hypophysitis reported as early as 3 weeks [5] and as late as 19 months [9], but typically occurring between 6 to 12 weeks (average 8.4 weeks) post-treatment [7]. Older age and male gender may be risk factors [7, 9].

Clinical Presentation, Investigations, and Initial Management of Hypophysitis

Algorithms for surveillance and management of hypophysitis following ipilimumab therapy have been proposed [10, 11]. Given the frequency of hypophysitis following CTLA-4 inhibitor therapy, baseline tests of pituitary function (LH, FSH, testosterone or oestradiol, thyroid-stimulating hormone (TSH), free T_4, adrenocorticotrophic

Table 13.1.4.1 Frequency of endocrine immune-related adverse events in patients treated with immune checkpoint inhibitors

	Hypophysitis	Adrenal insufficiency	Hyperthyroidism	Primary hypothyroidism	Thyroiditis
Anti-PD-1 therapy	0–1.9%	<1%	1–7.8%*	2–13%**	<1–9%
Anti-PD-L1 therapy	<1%	<1%	<5%	9–12%	1%
Anti CTLA-4 therapy	0–18%	<2%	<10%	0–7%	<5%
Combination therapy	0–12%	0–6%	4–12%#	13–22%##	0–16%

* subclinical hyperthyroidism occurs in an additional 13%.
** subclinical hypothyroidism occurs in an additional 13%.
# subclinical hyperthyroidism occurs in an additional 22%.
subclinical hypothyroidism occurs in an additional 5.6%.

hormone (ACTH), cortisol at 9:00 a.m.) would be helpful, both to document normal pituitary reserve and for comparison with later hormone measurements should development of hypophysitis be suspected. Thereafter, pituitary function should be evaluated if there are any clinical symptoms or signs suggestive of hypophysitis. However, since specific symptoms are often difficult to elicit, it is reasonable to undertake pituitary function testing prior to each cycle of ipilimumab administration and up to 16 weeks after initial treatment. If hypophysitis is suspected on clinical criteria outside these time points additional hormone measurements should be undertaken.

Most patients (80%) present with non-specific symptoms such as fatigue, muscle weakness, and headaches [5, 9]. Less common presenting symptoms include postural dizziness, weight loss, nausea, anorexia, vomiting, and generalized arthralgia. Hyponatraemia is the commonest reported biochemical abnormality (47–56%) at presentation [9].

TSH and gonadotrophin (luteinizing hormone and follicle-stimulating hormone) deficiency are very common (approximately 80%), as is ACTH deficiency (50–85%). Lack of ascertainment may explain the lower prevalence of growth hormone (GH) (25%) and prolactin deficiency [5, 7, 9]. Diabetes insipidus is very rare.

Visual fields should be tested by confrontation in all patients with suspected hypophysitis, and urgent formal visual field perimetry undertaken if an abnormality is suspected. Imaging of the pituitary gland, ideally by MRI with contrast, should be undertaken in order to determine if the typical appearance of hypophysitis (mild to moderate diffuse enlargement of the pituitary gland and homo- or heterogeneous gland enhancement following contrast administration) is present and also to exclude pituitary metastasis, or other, coincidental pituitary pathology (e.g. adenoma). Compression of the optic chiasm due to an enlarged pituitary gland has not been reported in anti-CTLA-4 induced hypophysitis. The pituitary gland typically decreases in size within 4 to 12 weeks from onset of hypophysitis with such change occurring irrespective of high-dose steroid administration [6, 7]. Most experts do not recommend routine administration of high-dose steroids in immune checkpoint inhibitor induced hypophysitis, however steroid treatment can be considered in selected circumstances, such as severe headache, pituitary enlargement that encroaches on the optic chiasm or associated with visual field defects.

As in conventional hypopituitarism, hormone replacement should be instituted, with particular care taken to exclude concomitant adrenal insufficiency prior to commencing thyroxine. Guidance on management of acute endocrine complications has been published [11]. Rates of recovery of hormone deficiencies are poorly defined; however, recovery of pituitary adrenal axis function is very rare (7, 9) and recovery of TSH and gonadotrophin deficiency is reported in approximately 50% of patients in some series [5, 7, 12].

CTLA4-Inhibitors and Other Endocrinopathies

Thyroid dysfunction was reported relatively infrequently (<10%) but possibly due to regular surveillance and improved detection of subclinical disease, recent studies have recorded thyroid dysfunction in 23% of cases [13]. Overt thyrotoxicosis is rare and is almost always due to thyroiditis, although GD has been reported. Primary adrenal insufficiency is very rare (<2%).

Pathogenesis

The CTLA-4 gene is also an important susceptibility locus for autoimmune endocrinopathies [14] and probably plays an important role in self-tolerance. CTLA-4 is expressed in murine and human pituitary glands, especially cells producing prolactin and TSH. Lymphocytic infiltration of the pituitary gland and development of pituitary antibodies is seen in mice injected with anti-CTLA-4 antibodies [15]. Although the pathogenesis of ipilimumab-induced hypophysitis is unknown, it is possible that administration of anti CTLA-4 antibody to susceptible individuals evokes an autoimmune process targeting the pituitary gland.

PD-1 Antibody Therapy

Programmed cell death-1 (PD-1) is an inhibitory receptor located on immune cells, which, by engaging its ligands (PD-L1 and PD-L2, located on antigen-presenting cells), acts as a negative regulator of immunity, downregulating the T-cell-mediated immune response [16, 17]. Pembrolizumab and nivolumab (humanized monoclonal antibodies targeting PD-1) exploit this mechanism by inhibiting the interaction between PD-1 and PD-L1/PD-L2, thereby perpetuating T-cell activation. In comparison to ipilimumab, they have been shown to exhibit greater clinical efficacy and a more favourable toxicity profile for some metastatic cancers [18–20].

Thyroid Dysfunction Following Anti-PD-1 Therapy

Thyroid dysfunction is common after treatment with PD-1 antibodies (3–39%) and tends to occur early (typically 3–8 weeks) (see Table 13.1.4.1) [21]. Although any form of thyroid dysfunction can occur, overt hypothyroidism (2–13%) seems more common than thyrotoxicosis (0–7.8%) (see Table 13.1.4.1) [13, 21–23].

Thyroiditis is likely to be the dominant cause of thyrotoxicosis; in one retrospective study, thyroiditis was the cause of dysfunction in all hyperthyroid patients and was also the suspected aetiology in the remainder of cases who were hypothyroid at presentation but may have had a preceding hyperthyroid phase [24].

High rates of subclinical thyroid dysfunction were reported by a single UK centre following treatment with either pembrolizumab or nivolumab; subclinical hypo- and hyperthyroidism occurred in 13% of cases, with the overall rate of thyroid dysfunction following anti-PD1 therapy being 39% [13].

Clinical Presentation, Investigations, Initial Management of Thyroid Dysfunction

As for CTLA-4 antibody-mediated endocrinopathies, recommendations for surveillance and management of thyroid dysfunction following administration of anti PD-(L)1 agents have been proposed [10, 11]. At a minimum, baseline assessment of thyroid function (TSH and fT_4) and periodic measurement of TSH and fT_4 (perhaps coinciding with cycles of treatment) is advised, with additional testing at intermediate timepoints if symptomatic.

If thyroid dysfunction develops, thyroid autoantibody status should be checked; in hypothyroidism, positive anti-TPO antibody levels can signify an autoimmune basis; in hyperthyroidism, positive anti-TSH receptor antibody status indicates GD (>95% specificity) rather than thyroiditis. At present, there is no evidence to suggest that thyroiditis or GD should be managed differently to that occurring in the conventional (non-immune checkpoint inhibitor) context.

Hyperglycaemia Following Anti-PD-1 Therapy

Development of diabetes mellitus, including diabetic ketoacidosis, has been described following anti-PD-1 therapy. In some cases biochemical measurements were strongly suggestive of immune-mediated insulin deficiency [25].

PD-L1/PD-L2 Inhibitors

Antibodies targeted against PD-L1 (such as atezolizumab, durvalumab, and avelumab) have shown efficacy against advanced non-small cell lung cancer and urothelial carcinoma [26]. Although their relative efficacy and toxicity has not been compared directly, indirect comparisons suggest endocrine IrAEs associated with atezolizumab are similar to those observed with anti-PD-1 antibody therapy (hypothyroidism 9%, thyrotoxicosis 3.6%, adrenal insufficiency and hypophysitis <1%) [27–30]. Durvalumab causes thyroid dysfunction in 5.5 to 12%, with the higher frequency possibly attributable to concomitant treatment with tremelimumab [31, 32]. Avelumab may have a lower frequency of endocrine IrAEs (adrenal insufficiency 0.5% [33, 34]; diabetes <0.1% [35]).

Pathogenesis

The interaction between PD1 and PDL1 evokes quiescence of autoreactive T cells, thereby maintaining peripheral tolerance and preventing autoimmunity. PD-1 deficient mice of differing genetic backgrounds develop different types of autoimmunity, presumably linked to varying target organ susceptibilities in different strains [34, 35].

Baseline positive thyroid autoantibody status may confer greater risk of anti-PD-1-mediated thyroid dysfunction [36], suggesting that this immunotherapy unmasks a latent autoimmune process in susceptible individuals.

Combination CTLA4 and PD-1 Blockade

The improved overall survival seen in metastatic melanoma after combination therapy with ipilimumab and nivolumab occurs at a cost of more frequent IrAEs [19, 20], and endocrinopathy is no exception. Rates of hypophysitis (4–12%) are higher than those observed following ipilimumab treatment alone, as is the incidence of thyroid dysfunction (30–50%) (see Table 13.1.4.1) [13, 37]. Adrenal insufficiency may also be more common (6.4% vs. 4.3%) (see Table 13.1.4.1).

Summary

It is now clear that immune checkpoint inhibitor agents are highly effective treatments for many advanced cancers. These agents frequently cause IrAEs, with thyroid dysfunction (following PD-(L)1 inhibitor-treatment) and hypophysitis (following CTLA-4 inhibitor therapy) being common. Although these side effects are not generally life-threatening, they can result in significant morbidity if not detected and treated promptly. These therapies provide remarkable insights into susceptibility and development of endocrine dysfunction, but further studies are needed to elucidate pathogenetic mechanisms for different endocrinopathies.

REFERENCES

1. Korman AJ, Peggs KS, Allison JP. Checkpoint blockade in cancer immunotherapy. *Adv Immunol*, 2006; **90**: 297–339.
2. Hodi FS, O'Day SJ, McDermott DF, *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*, 2010; **363**: 711–23.
3. Gan EH, Mitchell AL, Plummer R, Pearce S, Perros P. Tremelimumab-induced Graves' hyperthyroidism. *Eur Thyroid J*, 2017; **6**: 167–70.
4. Eggermont AM, Chiarion-Sileni V, Grob JJ, *et al.* Prolonged survival in stage III melanoma with ipilimumab adjuvant therapy. *N Engl J Med*, 2016; **375**: 1845–55.
5. Albarel F, Gaudy C, Castinetti F, *et al.* Long-term follow-up of ipilimumab-induced hypophysitis, a common adverse event of the anti-CTLA-4 antibody in melanoma. *Eur J Endocrinol*, 2015; **172**: 195–204.
6. Min L, Hodi FS, Giobbie-Hurder A, *et al.* Systemic high-dose corticosteroid treatment does not improve the outcome of ipilimumab-related hypophysitis: a retrospective cohort study. *Clin Cancer Res*, 2015; **21**: 749–55.
7. Faje AT, Sullivan R, Lawrence D, *et al.* Ipilimumab-induced hypophysitis: a detailed longitudinal analysis in a large cohort of patients with metastatic melanoma. *J Clin Endocrinol Metab*, 2014; **99**: 4078–85.
8. Corsello SM, Barnabei A, Marchetti P, De Vecchis L, Salvatopri R, Torino F. Endocrine side effects induced by immune checkpoint inhibitors. *J Clin Endocrinol Metab*, 2013; **98**: 1361–75.
9. Ryder M, Callahan M, Postow MA, Wolchok J, Fagin JA. Endocrine-related adverse events following ipilimumab in patients with advanced melanoma: a comprehensive retrospective review from a single institution. *Endocr Relat Cancer*, 2014; **21**: 371–81.
10. Haanen JBAG, Carbonnel F, Robert C, *et al.* ESMO Management of toxicities from immunotherapy: clinical practice guidelines for diagnosis, treatment and follow up. *Ann Oncol*, 2017; **28**: iv119–42.
11. Higham CE, Olsson-Brown A, Carroll P, *et al.* Acute management of the endocrine complications of checkpoint inhibitor therapy. *Endocr Connect*, 2018; **7**: G1–G7.
12. Blansfield JA, Beck KE, Tran K, *et al.* Cytotoxic T-lymphocyte-associated antigen-4 blockage can induce autoimmune hypophysitis in patients with metastatic melanoma and renal cancer. *J Immunother*, 2005; **28**: 593–8.
13. Morganstein DL, Lai Z, Spain L, *et al.* Thyroid abnormalities following the use of cytotoxic T-lymphocyte antigen-4 and

programmed death receptor protein-1 inhibitors in the treatment of melanoma. *Clin Endocrinol (Oxf)*, 2017; **86**: 614–20.
14. Vaidya B, Pearce S. The emerging role of the CTLA-4 gene in autoimmune endocrinopathies. *Eur J Endocrinol*, 2004; **150**: 619–26.
15. Iwama S, De Remigis A, Callahan MK, Slovian SF, Wolchok JD, Caturegli P. Pituitary expression of CTLA-4 mediates hypophysitis secondary to administration of CTLA-4 blocking antibody. *Sci Transl Med*, 2014; **6**: 230ra45.
16. Ishida Y, Agata Y, Shibahara K, Honjo T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J*, 1992; **11**: 3887–95.
17. Latchman Y, Wood CR, Chernova T, *et al.* PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat Immunol*, 2001; **2**: 261–8.
18. Robert C, Schachter J, Long GV, *et al.* Pembrolizumab versus ipilimumab in advanced melanoma. *N Engl J Med*, 2015; **372**: 2521–32.
19. Weber J, Mandala M, Del Vecchio M, *et al.* Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *N Engl J Med*, 2017; **377**: 1824–35.
20. Wolchok JD, Charion-Sileni V, Gonzalez R, *et al.* Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*, 2017; **377**: 1345–56.
21. De Filette J, Jansen Y, Schreuer M, *et al.* Incidence of thyroid-related adverse events in melanoma patients treated with pembrolizumab. *J Clin Endocrinol Metab*, 2016; **101**: 4431–9.
22. Ansell SM, Lesokhin L, Borello I, *et al.* PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med*, 2015; **372**: 311–19.
23. Reck M, Rodriguez-Abreu D, Robinson AG, *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*, 2016; **375**: 1823–33.
24. Orlov S, Salari F, Kashat L, Walfish PG. Induction of painless thyroiditis in patients receiving programmed death 1 receptor immunotherapy for metastatic malignancies. *J Clin Endocrinol Metab*, 2015; **100**: 1738–41.
25. Gaudy C, Clevy C, Monestier S, *et al.* Anti-PD1 Pembrolizumab can induce exceptional fulminant type 1 diabetes. *Diabetes Care*, 2015; **38**: e182–3.
26. Rosenberg JE, Hoffman-Censits J, Powles T, *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet*, 2016; **387**: 1909–20.
27. Hida T, Kaji R, Satouchi M, *et al.* Atezolizumab in Japanese Patients with previously treated advanced non-small-cell lung cancer: a subgroup analysis of the Phase 3 OAK study. *Clin Lung Cancer*, 2018; **19**: e405–15.
28. McDermott DF, Sosman JA, Sznol M, *et al.* Atezolizumab, an anti-programmed death-ligand 1 antibody, in metastatic renal cell carcinoma: long-term safety, clinical activity, and immune correlates from a phase Ia study. *J Clin Oncol,* 2016; **34**: 833–42.
29. Brahmer JR, Tukodi SS, Chow LQM, *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med*, 2012; **366**: 2455–65.
30. Kanie K, Iguchi G, Bando H, *et al.* Two cases of atezolizumab-induced hypophysitis. *J Endocr Soc*, 2018; **2**: 91–5.
31. Antonia SJ, Villegas D, Vicente DD, *et al.* Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer. *N Engl J Med*, 2017; **377**: 1919–29.
32. Antonia S, Goldberg SB, Balmanoukian A, *et al.* Safety and antitumour activity of durvalumab plus tremelimumab in non-small cell lung cancer: a multicentre, phase 1b study. *Lancet Oncol,* 2016; **17**: 299–308.
33. Gulley JL, Rajan A, Spigel DR, *et al.* Avelumab for patients with previously treated metastatic or recurrent non-small-cell lung cancer (JAVELIN Solid Tumor): dose-expansion cohort of a multicentre, open-label, phase 1b trial. *Lancet Oncol*, 2017; **18**: 599–610.
34. Nishimura H, Nose M, Hiai H, Minato N, Honjo T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. *Immunity*, 1999; **11**: 141–51.
35. Nishimura H, Okazaki T, Tanaka Y, *et al.* Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. *Science*, 2001; **291**: 319–22.
36. Osorio JC, Ni A, Chaft JE, *et al.* Antibody-mediated thyroid dysfunction during T-cell checkpoint blockade in patients with non-small-cell lung cancer. *Ann Oncol*, 2017; **28**: 583–9.
37. Postow MA, Chesney J, Pavlick AC, *et al.* Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N Engl J Med*, 2015; **372**: 2006–17.

13.2

Hormonal Therapy for Breast and Prostatic Cancers

13.2.1 The Breast

Lactation and Breast Cancer as an Endocrine Disease

Robert Clarke and Alice Greenhalgh

Normal Breast Development

The normal human breast is made up of two main tissue compartments—the stroma and the epithelium. The stroma consists of adipocytes, fibroblasts, blood vessels, inflammatory cells, and the extracellular matrix. Embedded within this stroma, the epithelium consists of several branching ductal-lobular systems forming tree-like structures from single major ducts of which there are estimated to be 12–16 that channel milk to their openings on the nipple [1]. This structure reflects the breast's specialist function to produce milk during lactation in order to feed and sustain offspring. Epithelial cells form terminal ductal lobuloalveolar units (TDLU), where milk is made, and connect to the nipple via a network of ducts. Both ducts and lobules are lined by a single layer of luminal epithelial cells which comprises hormone-sensing and milk-producing cells and their proliferative progenitors. The luminal cells are surrounded by a basal cell layer, consisting mostly of myoepithelial cells with contractile properties, as well as progenitor and multipotent stem cells, which are capable of giving rise to all epithelial cell types.

Human breast development is controlled by the endocrine system, specifically by the ovarian steroid hormones, oestrogen (E) and progesterone (Pg), and by the pituitary hormones prolactin (PRL) and growth hormone. There is evidence that other local and systemic proteins such as growth factors and cytokines play a role in tissue development at various stages.

Regulation of Adult Human Breast Development

Adult development of the human breast is regulated largely by ovarian steroid hormones which are first secreted at puberty. However, there is evidence of a requirement for other endocrine hormones such as growth hormone and PRL secreted by the pituitary gland, and placental lactogen secreted by the placenta during pregnancy. It has also become increasingly apparent that the effects of these systemic hormones are mediated locally by growth factors and cytokines synthesized and secreted by breast cells. In contrast to the endometrium, epithelial proliferation is maximal during the latter half of the menstrual cycle when both oestrogen and progesterone are secreted by the ovarian *corpus luteum*.

Ovarian Steroids: Oestrogen and Progesterone

E and Pg exert their effects following binding to receptors in the nucleus of the cell. The E receptor (ER) and the P receptor (PR) are nuclear receptors that act as ligand-activated transcription factors regulating cellular gene expression by both activation and repression of transcription. The hormone-sensing cells that express the classical ER and the PR are found within the luminal epithelial, but not the myoepithelial or stromal, cells of the human breast.

There are two isoforms of ER: ERα and ERβ, which are encoded by two different genes. ERα has historically been considered to be the most important in the development of the breast due to its expression pattern in the epithelium, its proliferative role, and the total lack of mammary epithelial development when it is knocked out in the mouse. In contrast, ERβ has a more subtle role in differentiation and the regulation of apoptosis, as the mammary gland of ERβ knockout mice develop normally. Unless indicated, ER refers to ERα in this review. There are two closely related isoforms of PR: PR-A and PR-B, both are transcribed from the same gene but with two different promoters. PR-B has an additional sequence of 164 amino acids at the amino terminus compared to PR-A. Both isoforms are equally expressed in breast epithelial cells and can be present as both hetero and homodimers.

E and P Effects on Mammary Development

In experimental studies utilizing ER and PR knockout animals it has been shown that, E regulates ductal elongation [2] and P regulates side branching and lobular development [3]. Other animal studies have demonstrated E to be a mitogenic stimulus during puberty promoting the elongation of the mammary ducts and expansion of the epithelium throughout the fat pad [4]. Mammary epithelial transplantation experiments in mice have also indicated that the growth factor amphiregulin, which is a ligand of the epidermal growth factor receptor, to be an essential mediator of E-induced proliferation during puberty [5].

Following puberty in the mouse, the effect of E on proliferation is weakened, although it is responsible for driving the strong mitogenic effects of P through its regulation of PR expression during this stage of development [6]. P is secreted in the human during the luteal phase of the menstrual cycle when proliferation is highest and therefore is regarded as the main proliferative hormone in both the mouse mammary gland and the normal human breast. P is also required for promoting much of the increase in breast epithelial proliferation during early to mid-pregnancy.

Normal Breast Development

Normal breast development can be grouped into three well-defined stages: embryonic, pubertal, and reproductive.

Embryonic Development

The rudiments of the gland are developed during embryogenesis when a mammary disc forms from the epidermal milk line and the parenchyma begins to invade the underlying mesenchyme. The invading epithelial cells become indented at the epithelial-stromal border, sprouting, and separating into 10–15 branches of epithelial ducts that each open separately onto the epidermal surface at the nipple. As these ducts grow into the mesenchyme they may bifurcate several times and may even form end-vesicles, a primitive form of the lobuloalveolar structures observed in the adult gland. Late in gestation, the skin above this area specializes to form the areolus and nipple from which secretions may be observed at parturition, presumably in response to maternal hormones.

Pubertal Development

In contrast to embryogenesis, all developmental stages of post-embryonic breast development are hormone dependent. Mammary gland development during childhood does little more than keep up with the overall growth of the child until the approach of puberty. Puberty triggers breast growth driven by ovulation and the formation of regular menstrual cycles. The mature gland undergoes cycles of proliferation and apoptosis over the course of each subsequent menstrual cycle. This results in a network of ducts leading from the nipple growing and dividing into bundles of primary and secondary ducts, lined with epithelial cells and ending with morphologically club-like structures called terminal end buds (TEBs). Equally, the mainly collagenous connective tissue of the breast expands and grows to accommodate the branching network of epithelium. It is from the TEBs that the TDLU, also called lobules, form and which comprise the principal, functional milk-producing glandular units of the breast (see Figure 13.2.1.1).

These lobules exist initially as alveolar buds that mature following menarche into a variable number of blind-ending, grape-like, secretory sacs known as acini, alveoli, or ductules which open into the intralobular terminal duct. At maturity, the lobular structure can be

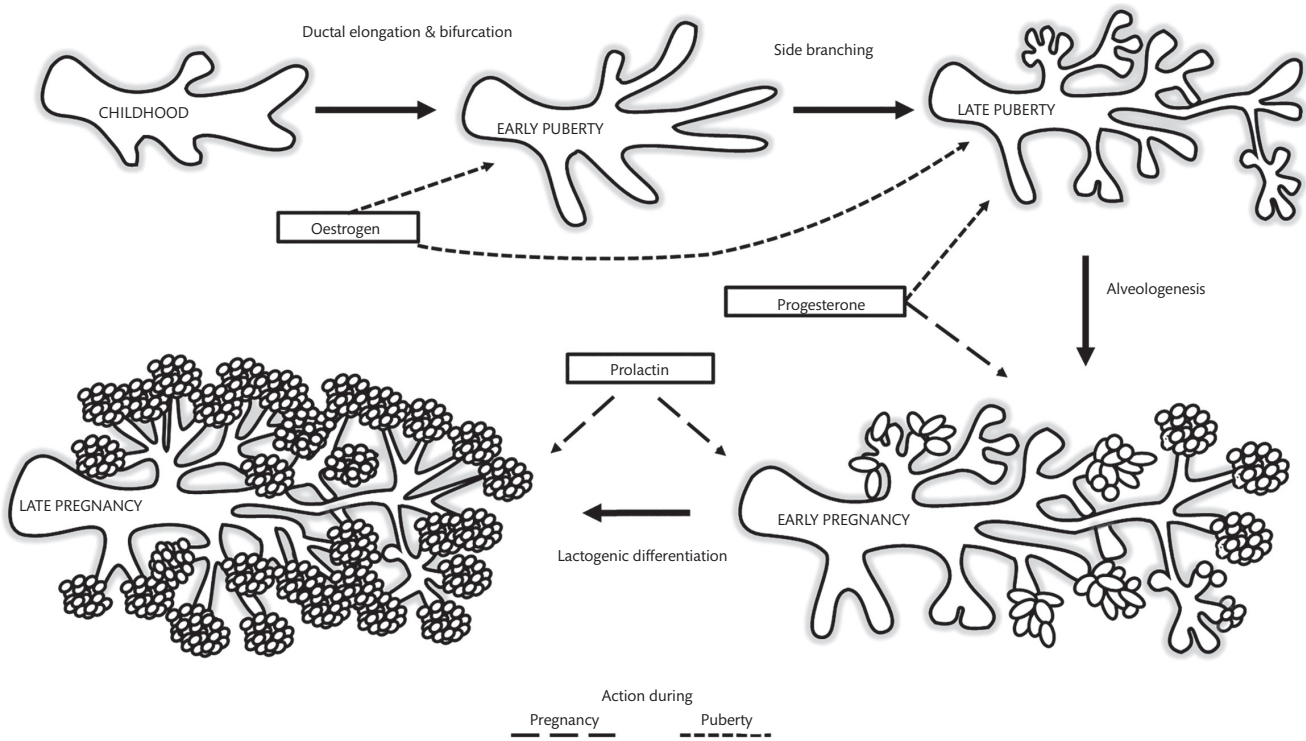


Figure 13.2.1.1 Diagrammatic representation of the lobular structures of the human breast throughout all stages of development.

imagined as resembling a bunch of grapes (see Figure 13.2.1.1). It is during the development of the TDLU that epithelial hyperplasias and carcinomas of the breast are thought to be initiated since this is the site at which they are most often observed histopathologically.

Reproductive Development

Epithelial development continues throughout adolescence and into adulthood. Despite many rounds of mammary remodelling events taking place during puberty and through adulthood, it is not until pregnancy that full mammary epithelial maturation occurs, as the breast progressively transforms into a fully mature functional organ. During pregnancy, mammary stem cells (MaSCs) and progenitor cells proliferate resulting in the formation of new ducts, elongation of the existing ducts, further epithelial branching, and formation and expansion of round hollow structures, which are individually called alveoli [7, 8].

In parallel with this growth and development of the epithelial component, the stromal elements also develop to accommodate this increase in volume and cell number.

The progressive maturation of the breast is regulated by the increased levels of the circulating lactogenic hormones, E, P, and PRL during pregnancy. Other hormones and growth factors are also involved in regulating mammary development during pregnancy. Placental lactogen, amphiregulin, Wnt, RANKL, CXCL12, IGF2, and stromal paracrine factors directly regulate mammary development. Contrastingly insulin, growth hormone, glucocorticoids, and fibroblast growth factors play a more indirect role [9]. Blood placental lactogen has been shown to be strongly correlated with breast growth, with a proposed mechanism of inducing the stem/progenitor cells to proliferate [10].

There are two main phases of human mammary development corresponding to the early and late stages of pregnancy. The early stage is characterized by increased side branching of the ductal tree corresponding to elevated P levels, resulting in the formation of ductules termed acini. P levels rise during the luteal phase of the menstrual cycle but particularly during pregnancy. During the second trimester, the lobules become further enlarged and increase in number. P works cooperatively with PRL to stimulate cellular differentiation at the alveolar sites (during alveologenesis) that are responsible for milk secretion during lactation. P is necessary in the mammary epithelium for side branching and alveologenesis [9], whereas the PRL is required for alveologenesis and milk secretion. The cooperation between P and PRL signalling becomes evident after side branches have been established during mid-pregnancy, when further alveologenesis requires epithelial PRL signalling [10]. This sequential action ensures that the discrete morphological steps occur in an ordered fashion; therefore, allowing all of the ducts to form before the alveoli bud, ensuring they have enough space to unfold and to be drained effectively [6].

The secretory alveoli formed during pregnancy signify the end of mammary gland differentiation. Each alveolus is surrounded by stroma and separated from it via the basement membrane. Secretion of milk by the epithelium during lactation also distends the alveoli and upon infant suckling, there is an increase in bound oxytocin causing the myoepithelial cells to contract releasing milk from the alveolus.

Following weaning, secretion of milk by the epithelial cells ceases and the lobules involute to resemble those present in the non-pregnant gland, although they may retain a larger number of individual alveoli per lobule than before. This intriguing process of expansion and regression, known as involution, can take place across multiple pregnancies during the reproductive lifespan, signifying that the epithelial cells of the mammary gland have substantial regenerative abilities.

A Mammary Stem Cell Niche

There is increasing evidence that mouse and human mammary epithelium is hierarchically arranged; from the multipotent stem cells to the lineage restricted, hormone-sensing, milk-producing, and myoepithelial cells [11–14].

The mammary gland is a dynamic structure, throughout a lifetime many changes in both structure and function occur, including recurring expansions relating to the hormonal changes prompted by the menstrual cycle. In addition, dramatic changes occur during pregnancy, lactation, and involution. MaSC are proposed to be the drivers of the constant mammary tissue remodelling events occurring through cycles of pregnancy, lactation, and involution. This quiescent, self-renewing, population of cells give rise to the differentiated ductal, alveolar, and myoepithelial cells [15].

A MaSC niche theory has been proposed encompassing both cellular and acellular supportive elements surrounding the stem cells. These supporting elements could, for example, include differentiated stem cell progeny their surrounding stromal cells which display cell surface molecules and can deliver soluble factors to regulate stem cell actions.

As previously discussed, the mammary gland is constantly remodelled in response to the female reproductive hormones. A self-renewing population of MaSCs is needed during these cycles of growth to differentiate into new epithelia. Consequently, MaSCs must respond to these hormonal signals; however, oestrogen receptor (ER) and progesterone receptor (PR) are only detected in a subset of luminal epithelial cells and have not yet been identified in the mouse or human MaSC-enriched basal cell population [16, 17]. Three epithelial subsets including basal stem/progenitor, luminal progenitor and mature luminal cells have been characterized so far with ER and PR being found to be enriched in mature luminal cells.

Therefore (ER^+PR^+) luminal cells can be described as essential niche cells that detect circulating hormone levels and signal, via paracrine mechanisms, to other niche populations, and to unipotent or multipotent MaSCs.

Hormones and Breast Cancer

There is a causal relationship between breast cancer and the ovarian steroid hormones. Early menarche, late first-term pregnancy, and late menopause are associated with a heightened breast cancer risk, this is coupled with research showing removing a woman's ovaries significantly reduces the risk of breast cancer. In addition, both hormone replacement therapy (HRT) and oral contraceptives can increase the risk of breast cancer, thus linking steroid hormones with breast carcinogenesis [18]. However, on the other hand, ER/PR+ tumours are associated with a more favourable breast cancer prognosis with receptor expression being most significant predictive marker for endocrine responsiveness [19].

The mammary stem cell theory that steroid receptor-positive cells signal via paracrine mechanisms to adjacent receptor-negative

cells, is supported by the inverse relationship between steroid receptor expression and proliferation in the normal breast [9, 20]. Contrastingly, in breast cancer the number of proliferating cells expressing steroid hormone receptors is increased [21], implying that steroid receptor-positive cells experience an early 'switch' to autocrine signalling mechanisms, or that these cells somehow gain the ability to divide [22]. However, the mechanisms that drive this molecular switch from paracrine to autocrine signalling are as yet unknown making this an interesting area for current research [23, 24].

Conclusions

This is a very interesting time for mammary gland scientists. New signalling pathways are being uncovered that dictate the lineage commitment, coupled with the development of improved humanized mouse models and three-dimensional culture models, the future of mammary gland biology is an exciting one. As new discoveries are made, the cellular targets of breast cancer will become better understood, thus providing oncologists with better tools to predict, prevent, and treat breast cancer.

REFERENCES

1. Moffat DF, Going JJ. Three dimensional anatomy of complete duct systems in human breast: pathological and developmental implications. *J Clin Pathol*, 1996; **49**: 48–52.
2. Hewitt SC, Harrell JC, Korach KS. Lessons in estrogen biology from knockout and transgenic animals. *Annu Rev Physiol*, 2005; **67**: 285–308.
3. Conneely OM, Mulac-Jericevic B, Lydon JP, De Mayo FJ. Reproductive functions of the progesterone receptor isoforms: lessons from knock-out mice. *Mol Cell Endocrinol*, 2001; **179**: 97–103.
4. Silberstein GB, Van Horn K, Shyamala G, Daniel CW. Essential role of endogenous estrogen in directly stimulating mammary growth demonstrated by implants containing pure antiestrogens. *Endocrinology*, 1994; **134**: 84–90.
5. Ciarloni L, Mallepell S, Brisken C. Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development. *Proc Natl Acad Sci U S A*, 2007; **104**: 5455–60.
6. Brisken C, O'Malley B. Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol*, 2010; **2**: a003178.
7. Sternlicht MD, Dunning AM, Moore DH, *et al.* Prognostic value of PAI1 in invasive breast cancer: evidence that tumor-specific factors are more important than genetic variation in regulating PAI1 expression. *Cancer Epidemiol Biomarkers Prev*, 2006; **15**: 2107–14.
8. Sternlicht MD, Kouros-Mehr H, Lu P, Werb Z. Hormonal and local control of mammary branching morphogenesis. *Differentiation*, 2006; **74**: 365–81.
9. Brisken C, Park S, Vass T, Lydon JP, O'Malley BW, Weinberg RA. A paracrine role for the epithelial progesterone receptor in mammary gland development. *Proc Natl Acad Sci U S A*, 1998; **95**: 5076–81.
10. Brisken C. Hormonal control of alveolar development and its implications for breast carcinogenesis. *J Mammary Gland Biol Neoplasia*, 2002; **7**: 39–48.
11. Shackleton M, Vaillant F, Simpson KJ, *et al.* Generation of a functional mammary gland from a single stem cell. *Nature*, 2006; **439**: 84–8.
12. Stingl J, Eirew P, Ricketson I, *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature*, 2006; **439**: 993–7.
13. Asselin-Labat ML, Sutherland KD, Barker H, *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol*, 2007; **9**: 201–9.
14. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev*, 2009; **23**: 2563–77.
15. Woodward WA, Chen MS, Behbod F, Rosen JM. On mammary stem cells. *J Cell Sci*, 2005; **118**(Pt 16): 3585–94.
16. Sleeman KE, Kendrick H, Robertson D, Isacke CM, Ashworth A, Smalley MJ. Dissociation of estrogen receptor expression and *in vivo* stem cell activity in the mammary gland. *J Cell Biol*, 2007; **176**: 19–26.
17. Asselin-Labat ML, Shackleton M, Stingl J, *et al.* Steroid hormone receptor status of mouse mammary stem cells. *J Natl Cancer Inst*, 2006; **98**: 1011–14.
18. Bhupathiraju SN, Grodstein F, Stampfer MJ, Willett WC, Hu FB, Manson JE. Exogenous hormone use: oral contraceptives, postmenopausal hormone therapy, and health outcomes in the nurses' health study. *Am J Public Health*, 2016; **106**: 1631–7.
19. Osborne CK, Schiff R, Arpino G, Lee AS, Hilsenbeck VG. Endocrine responsiveness: understanding how progesterone receptor can be used to select endocrine therapy. *Breast*, 2005; **14**: 458–65.
20. Clarke RB, Howell A, Potten CS, Anderson E. Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res*, 1997; **57**: 4987–91.
21. Anderson E. The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. *Breast Cancer Res*, 2002; **4**: 197–201.
22. Lange CA, Gioeli D, Hammes SR, Marker PC. Integration of rapid signaling events with steroid hormone receptor action in breast and prostate cancer. *Annu Rev Physiol*, 2007; **69**: 171–99.
23. Grimm SL, Hartig SM, Edwards DP. Progesterone receptor signaling mechanisms. *J Mol Biol*, 2016; **428**: 3831–49.
24. Hilton HN, Graham JD, Clarke CL. Minireview: progesterone regulation of proliferation in the normal human breast and in breast cancer: a tale of two scenarios? *Mol Endocrinol*, 2015; **29**: 1230–42.

13.2.2 Endocrine Treatment of Breast Cancer

Amna Sheri and Laura Morrison

Introduction

Endocrine manipulation has been recognized as a treatment modality for breast cancer for over 100 years. Oestrogen is an important promoter in the pathogenesis of breast cancer and endocrine response, is predominantly dependent on the presence of oestrogen receptor (ER), a protein which can be detected in about 70% of primary breast cancers.

A better understanding of the mechanisms that result in oestrogenic deprivation of breast cancer cells has enabled medical therapeutics to be developed which have largely replaced surgical ablative procedures (Box 13.2.2.1). These include hormone manipulation at the cellular level, via competition for ER in the breast tumour, using antioestrogens, such as tamoxifen or more recently selective oestrogen receptor modulators (SERMs), and 'pure' antioestrogens such as fulvestrant.

An alternative approach is to lower systemic oestrogen levels in premenopausal women by the use of luteinizing hormone releasing hormone (LHRH) agonists and in postmenopausal women by the use of aromatase inhibitors. Endocrine agents with more ill-defined mechanisms, such as progestogens, androgens, and corticosteroids can induce endocrine responses.

In patients with ER-positive advanced breast cancer, endocrine treatments generally achieve a response rate of between 20% and 40%. The median response duration to endocrine therapy in advanced disease is about 8–14 months but can last several years.

In patients with early stage ER-positive breast cancer adjuvant tamoxifen therapy given for 10 years delays local and distal relapse and prolongs survival. It also substantially reduces the incidence of contralateral breast cancer. The use of endocrine therapy in a chemopreventive role has been evaluated [1–4].

Antioestrogens

Tamoxifen

Approximately 80% of human breast carcinomas express ER. Tamoxifen is a non-steroidal oestrogen receptor antagonist which inhibits breast cancer growth by competitive antagonism of oestrogen at the receptor site. Its actions are complex due to partial oestrogenic agonist and antagonist effects depending on the specific end organ. It has been in use since its approval in 1978 and, although effective, the partial agonist effects may account for the development of tamoxifen resistance and disease progression after prolonged administration, in addition to specific adverse side effects on the gynaecological tract.

Adjuvant Tamoxifen

In early breast cancer, tamoxifen has been the gold standard of adjuvant endocrine therapy for both premenopausal and postmenopausal breast cancer for over two decades. In an overview of the effects of chemotherapy and hormonal therapy for early breast cancer by the Early Breast Cancer Trialist's Collaborative Group (EBCTG) [5], a 31% reduction in the annual breast cancer death rate with 5 years of adjuvant tamoxifen was reported at 15 years of follow-up. Breast cancer is relatively unusual in that the risk of recurrence remains substantial during the second decade and indeed may continue indefinitely. It is therefore highly significant that trials comparing allocation to a control arm or 5 years of tamoxifen have shown persistent benefits with a reduction in the 15-year probability of death from breast cancer about three times as great as the 5-year probability [1]. The benefit from tamoxifen treatment was confined to those patients with oestrogen receptor-positive cancers. Overall, this gave an 8% reduction in the 5-year mortality for patients with primary operable breast cancer. There was also a 47% reduction in the incidence of contralateral breast cancer at 5 years.

Adjuvant tamoxifen was traditionally administered for 5 years following data from the EBCTG. While earlier data in node-negative patients did not find any advantage in continuing tamoxifen beyond 5 years [6] newer data from two key trials suggests that 10 years of adjuvant treatment is beneficial. The Adjuvant Tamoxifen, Longer Against Shorter (ATLAS) study [7] randomized 12 894 women to 10 or 5 years of tamoxifen. In this study, there was a significant reduction in risk of breast cancer recurrence ($P = 0.02$), breast cancer mortality ($P = 0.01$) and overall mortality ($P = 0.01$) in women with ER-positive disease. Similar results were noted in the aTTom study [8].

Tamoxifen 20 mg/day for 10 years alone or in addition to chemotherapy is now the standard endocrine therapy of choice for early breast cancer in premenopausal women [9].

Side Effects/Toxicity of Tamoxifen

The side effects of tamoxifen are well documented. Short-term effects are usually mild and include hot flushes and altered menses. Tamoxifen acts as an oestrogen agonist on bone, increasing mineral density and thereby potentially reducing the risk of osteoporotic fractures in postmenopausal women. Similarly, it lowers serum cholesterol, potentially decreasing the risk of heart disease. The risk of developing deep vein thrombosis and pulmonary emboli is increased in women taking tamoxifen and caution must be used in patients with prothrombotic risk factors. The ATLAS study built on previous data showing an increased risk of endometrial cancer, due to the oestrogenic stimulation of the endometrium, with a longer duration of tamoxifen. The increased risk was predominantly confined to postmenopausal women who had not undergone hysterectomy and the risk was outweighed by the reduction in breast cancer mortality and recurrence. There were no deaths from endometrial cancer in this study.

Aromatase Inhibitors: Pharmacology

In contrast to tamoxifen, which antagonizes oestrogen at the receptor site, the oral aromatase inhibitors (AI) such as anastrozole), letrozole, and exemestane all reduce serum oestrogen levels in postmenopausal women by preventing the conversion of adrenal androgens into oestradiol and oestrone by the cytochrome P450 enzyme aromatase (Figure 13.2.2.1) [10].

Box 13.2.2.1 Endocrine treatment for breast cancer

- Antioestrogens
  - Tamoxifen
  - Other selective oestrogen modulators: raloxifene, lasofoxifene
  - Steroidal 'pure' antioestrogens: fulvestrant
- Oestrogen deprivation therapies
  - Luteinizing hormone releasing hormone analogues: goserelin
  - Aromatase inhibitors: letrozole, anastrozole, exemestane
- Other agents
  - Progestogens
  - Androgens: medroxyprogesterone acetate. Megestrol acetate
  - Corticosteroids
  - Oestrogens: diethylstilboestrol

While oestrogens are primarily synthesized in the ovary in premenopausal women, following the menopause mean plasma oestradiol levels fall from about 400–600pmol/L to around 25–50 pmol/L. These residual oestrogens come solely from peripheral aromatase conversion, particularly in the subcutaneous fat. Of note, AI are contraindicated in premenopausal women without additional ovarian suppression because the suppression of peripheral aromatase results in a reduced feedback to the hypothalamus and an increase in ovarian stimulation [10].

The currently approved 'third-generation' AI all powerfully inhibit oestrogen synthesis [10]. They are highly specific with almost no effect on cortisol or aldosterone levels. All are administered on a once daily basis.

Clinical Efficacy of AI in Advanced Breast Cancer

A number of clinical trials have now consistently shown improved efficacy of AI over tamoxifen [11–17]. See Table 13.2.2.1.

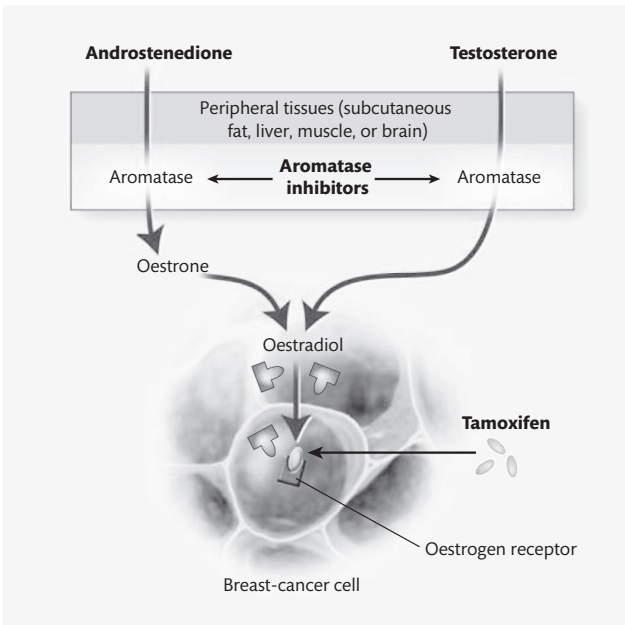


Figure 13.2.2.1 Mechanism of action of aromatase inhibitor and tamoxifen.

Reproduced with permission from Smith IE, Dowsett M. Aromatase inhibitors in breast cancer. *N Engl J Med*, 2003; 348:2431–42.

The largest of these, a randomized double-blind phase III trial in over 900 postmenopausal women with locally advanced or metastatic breast cancer, compared letrozole 2.5 mg with tamoxifen 20 mg daily [15, 16]. Patients treated with letrozole had a significantly higher objective tumour response rate (30% vs. 20%, P <0.001), clinical benefit rate (49% vs. 38%, P <0.001), and prolonged time to disease progression (median time to progression of 9.4 months vs. 6.0 months, hazard ratio 0.72, P <0.0001). Of particular note in this trial, nearly 20% of the patients had received prior tamoxifen in the adjuvant setting; in this subgroup, retreatment with tamoxifen had a low response rate of 8% compared with a 32% response rate with letrozole. The improvements in clinical efficacy for letrozole resulted in an early improvement in survival during the first 2 years [18], although with longer follow-up this difference was lost. This could be explained by the high number (>50%) of patients who prospectively crossed over to the alternate treatment at the time of progression. Significantly more patients benefited from second-line letrozole after progression on tamoxifen than the reverse. There were no significant differences in toxicity between the two treatments [16].

As a result of this data the third-generation AIs have now become the standard of care for first-line endocrine therapy in postmenopausal women with ER-positive advanced breast cancer.

Use of AI in Early Breast Cancer

The establishment of the efficacy and tolerability of AI in advanced breast cancer encouraged the development of a number of trials examining their use in the adjuvant setting.

Trials Comparing Adjuvant Tamoxifen Versus AI

A large meta-analysis undertaken by the (EBCTCG) [19] in 2015 compared the efficacy of tamoxifen versus AI in postmenopausal women with ER-positive early breast cancer. The analysis compared results from nine individual trials and reports on 31 920 women. A significant reduction in breast cancer mortality (RR 0.85) and recurrence, at both distal (RR 0.86) and local (RR 0.74) sites, in women receiving 5 years of AI versus 5 years of tamoxifen therapy (P <0.00001). Ten-year recurrence was 19.1% in the AI group and 22.7% in the tamoxifen group, suggesting a 3.6% 10-year gain of AI over tamoxifen. In terms of side-effect profile, there was a greater incidence of bone fracture in the AI group (8.2% vs. 5.5%; RR 1.30) and this risk was maintained throughout a 10-year follow-up period. There was a lower rate of endometrial cancer in the aromatase inhibitor group (0.4% vs. 1.2%).

Upfront AI Versus Switching Strategies

Earlier studies of tamoxifen versus tamoxifen followed by a switch to an AI demonstrated a survival benefit. However, this was not shown to be superior to commencing on upfront AI. The BIG-198 also looked at commencing on AI upfront followed by a switch to tamoxifen at year 2–3 and this was not inferior to 5 years of AI in this study. This therefore provides some evidence that patients who tolerate an AI poorly can be switched to tamoxifen safely if needed.

Extended Adjuvant Endocrine Therapy with AI After 5 Years of Tamoxifen

Hormone receptor-positive breast cancers have a chronic relapsing nature with risk of recurrence continuing indefinitely and

Table 13.2.2.1 Comparative of first-line trials of aromatase inhibitors versus tamoxifen in advanced breast cancer

Authors (reference)	Comparators	N	Response (%)	Clinical benefit (%)	Median time to progression (months)
Nabholtz *et al.* [12]	Anastrozole Tamoxifen	171 182	21 17	59 46	11.1 5.6
Bonneterre *et al.* [13, 14]	Anastrozole Tamoxifen	340 328	33 33	56 56	8.2 8.3
Mouridsen *et al.* [15, 16]	Letrozole Tamoxifen	453 454	30 20	49 38	9.4 6.0
Paridaens *et al.* [17]	Exemestane Tamoxifen	182 189	46 31	66 49	9.9 5.8

approximately half of all recurrences occurring between 5 and 15 years after surgery despite 5 years of adjuvant tamoxifen treatment. There is, therefore, a rationale for considering extended adjuvant endocrine therapy beyond 5 years. The MA.17 study [20], a double-blind, placebo-controlled trial evaluated whether 5 years of letrozole therapy after completion of 5 years of tamoxifen in postmenopausal women could lead to an improvement in disease-free survival (DFS). The positive results were most marked in patients with lymph node positive disease and those who were premenopausal at presentation.

Duration of AI Therapy

Treatment with an AI for 5 years either as monotherapy or after tamoxifen became the standard of care for postmenopausal women with early, receptor-positive breast cancer.

The MA.17R study, a double-blind, placebo-controlled trial, assessed the efficacy of extending AI therapy to 10 years [21]. The study enrolled 1918 women who had remained disease-free on adjuvant endocrine therapy with any AI. They were randomly assigned to receive placebo or continue receiving letrozole for a further 5 years. Most patients in this study had received 5 years of tamoxifen prior to starting AI so the study effectively compares 10 versus 15 years of adjuvant endocrine therapy. There was a significant improvement in the DFS and a reduction in the incidence of contralateral breast cancer as well as a 1% benefit in distant recurrence. Interestingly there was no difference in overall survival (OS). The extended therapy arm had an increased risk of osteoporotic fractures and this would need to be considered when deciding on the appropriate duration of treatment.

Two other studies, DATA and IDEAL [22, 23] have also looked at extended AI therapy in this group of patients. While the DATA study did show a reduction in DFS in the 6-year group they also noted a higher rate of side effects in the extended therapy arm. The IDEAL study, comparing 2.5 or 5 years of additional AI therapy did not show any significant difference between groups in relation to DFS or OS and

There is therefore a lack of clarity regarding which, if any, subgroups of patients will benefit from extended AI therapy. Decisions about extending therapy to 10 years will be influenced by the relative risk versus the potential toxicities in individual patients and is an area of ongoing research.

Toxicity/Side Effects of AI

In general, the third-generation AIs are well tolerated although symptoms of oestrogen withdrawal are common. The commonest side effects include hot flashes, musculoskeletal stiffness, and vaginal dryness. However, in the clinical trial setting the different side-effect profiles of tamoxifen and AI do not appear to impact on patient's quality of life [24]. Interestingly, the appearance of vasomotor or joint symptoms within the first 3 months of treatment has been associated with a greater decrease in breast cancer recurrence in the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial [25].

In trials comparing the third-generation AIs with tamoxifen, the adverse events associated with tamoxifen's oestrogenic properties, such as venous thromboembolism and endometrial cancer, were significantly less common in the AI groups [26]. However, an increased risk of osteoporosis and fractures has been observed with the AIs compared with tamoxifen [26, 27]. The American Society of Clinical Oncology (ASCO) guidelines recommend that postmenopausal women who receive an aromatase inhibitor should have their bone mineral density evaluated [28]. It is likely that the introduction of adjuvant bisphosphonates in postmenopausal women will also modify this risk in the future. Increases in cardiovascular events [29] and hypercholesterolaemia [26] have been observed with AIs.

The relatively short follow-up time and overall small numbers of cardiovascular events in these studies make it difficult to draw concrete conclusions about the causality of AIs and cardiovascular events. Given that most women presenting with early breast cancer can now expect long-term survival long-term vigilance of cardiovascular morbidity and mortality in these studies is warranted.

Current Role of AI

Cumulatively, the results of these adjuvant AI trials have led to a substantial increase in the use of AIs in early breast cancer, but also to some uncertainty as to optimal duration of the correct role of AI therapy.

At present the gold standard of care for postmenopausal women is to commence upfront AI as per the ASCO guidelines [30].

Ultimately risks of recurrence and relative toxicities must be considered when considering extended therapy. Who requires a longer duration of therapy is an area of active research.

Oestrogen Deprivation Therapies: Ovarian Ablation/Suppression

In 1896, Beatson published the first report of surgical oophorectomy as a treatment modality for advanced breast cancer in

premenopausal women. This was followed, in the 1920s, by radiotherapy-induced ovarian ablation. Since then ovarian ablation by either means has been used as a therapy for premenopausal patients with advanced breast cancer.

LHRH agonists, are currently used as an alternative to ovarian ablation for the treatment of advanced breast cancer in premenopausal women. The initial stimulation of luteinizing hormone can cause a short 'flare' in disease-related symptoms, followed by a complete inhibition of luteinizing hormone secretion, decreasing oestradiol levels to near castration levels. This effect is reversible on withdrawal of the LHRH agonist. The LHRH agonist goserelin 3.6 mg administered by deep subcutaneous injection every 4 weeks is licensed in the United Kingdom for the treatment of breast cancer.

The EBCTG [5] reviewed several trials involving women with ER-positive or ER-unknown early breast cancer randomized to ovarian ablation or ovarian suppression with an LHRH agonist. There was a definite effect of both ovarian ablation and suppression on recurrence and breast cancer mortality [1]. A more recent meta-analysis [31] focused on trials where oestrogen receptor status was known and used only LHRH agonists as the method of ovarian suppression. A particular benefit was observed when LHRH agonists were used after chemotherapy, either alone or with tamoxifen in women aged 40 years or younger in whom chemotherapy is less likely to induce permanent amenorrhea. Optimum duration of use is unknown.

Ovarian Suppression in Addition to Endocrine Therapy in Premenopausal Patients

A landmark study, the Suppression of Ovarian Function Trial (SOFT) [32] has examined the use of ovarian suppression in women with hormone-receptor-positive early breast cancer. In the SOFT study 3066 premenopausal women were randomized to tamoxifen alone, tamoxifen with ovarian suppression, or exemestane with ovarian suppression. There was no benefit of ovarian suppression across the whole study population however importantly a subgroup analysis did demonstrate benefit in women under 35 years and those who had been deemed high risk of recurrence and undergone chemotherapy. Of the 94% of women under 35 who had chemotherapy 67.1% remained breast cancer free at 5 years in the tamoxifen alone arm versus 75.9% in those who received ovarian suppression with tamoxifen and 83.2% in the exemestane and ovarian suppression group. Ovarian suppression is now recommended in premenopausal patients at high enough risk to receive adjuvant chemotherapy [30].

The Role of AIs in Premenopausal Women

In TEXT [33] premenopausal women were randomly allocated to receive ovarian suppression with either tamoxifen or exemestane. Data from the SOFT and TEXT have been combined to provide a larger study population for further analysis. The combined data demonstrated a relative reduction in risk of disease recurrence, second cancer, or death of 28% and a 34% with exemestane plus ovarian suppression versus tamoxifen with ovarian suppression. These figures compare favourably with results seen in postmenopausal women. Long-term follow-up to determine OS advantage is still pending.

Despite a relatively low adverse event rate and discontinuation level in the study population ovarian suppression is a difficult treatment to tolerate and the percentage of those intolerant to therapy is likely to be much greater in the real-world setting. More data is needed on the tolerance of ovarian suppression in this population. In younger patients receiving adjuvant ovarian suppression this may be given either with tamoxifen or an AI.

Other Agents

Steroidal Pure Antioestrogens: Fulvestrant Pharmacology

Fulvestrant is a novel type of oestrogen receptor antagonist which, unlike tamoxifen, has no known agonist effects. It is administered intramuscularly and does not appear to cause endometrial proliferation and is less likely than tamoxifen to cause thromboembolism. Fulvestrant binds to the oestrogen receptor, but, induces a different conformational shape with the receptor to that achieved by tamoxifen. Due to its unique mechanism of action, fulvestrant delays the emergence of acquired resistance compared with tamoxifen in an MCF-7 hormone-sensitive xenograft model [34].

Clinical Efficacy of Fulvestrant in Advanced Breast Cancer

In a small phase II study in advanced disease, fulvestrant was shown to produce remissions of 2 years in tamoxifen-resistant tumours in postmenopausal women [18]. Clinical data with fulvestrant in advanced breast cancer following resistance to AIs is limited but phase II trials in this setting have reported clinical benefit rates of between 19 and 52%.

The Evaluation of Faslodex versus Exemestane Clinical Trial (EFECT) [35] assessed the efficacy of fulvestrant versus exemestane in patients who had progressed on treatment with AIs and found no significant difference in the effectiveness or tolerability between treatments with a clinical benefit rate of 32.2% and 31.5%, respectively. Both treatments were well tolerated.

A second study (SoFEA) compared progression-free survival in patients who had progressed on an AI, treated with either fulvestrant plus continued anastrozole, or with fulvestrant alone [36]. No benefit was noted of fulvestrant either in combination or alone in this trial population.

An analysis of the SoFEA paper showed a clinical advantage with fulvestrant over exemestane in patients with an ESR1 mutation [37]. This is the subject of ongoing research but provides a potential biomarker to select patients who may benefit from specific treatment options.

The results of these trials will help define optimal sequencing of endocrine therapies, and in particular the effectiveness of fulvestrant alone or in combination with AIs [38].

Progestogens

Synthetic progestogens/androgens, such as medroxyprogesterone and megestrol acetate, have been used in the treatment of advanced breast cancer, although their main benefit is relief of metastatic bone pain. Their mechanism of action is unclear but they are known to cause steroidogenic side effects such as weight gain, and cardiovascular and thromboembolic complications. The new generation of

AIs have clearly been shown to be more effective and less toxic than progestogens and these agents have largely been replaced as breast cancer treatment options.

Androgens and Corticosteroids

Androgens are rarely used for the treatment of advanced breast cancer due to side effects. Less than 10% patients with advanced breast cancer respond to treatment with corticosteroids. However, corticosteroids are often effective at controlling symptoms, such as local oedema, and pain.

Oestrogens

High-dose oestrogen therapy was used in the treatment of advanced breast cancer until the introduction of tamoxifen in the 1970s. A phase II trial has explored the use of high-dose oestrogen therapy in highly refractory, advanced breast cancer [39]. The clinical benefit rate was 40% with a median duration of response of 9 months. Another trial compared low-dose oestradiol 6 mg/day with high-dose 30 mg/day [40] in women with AI-resistant advanced breast cancer. The lower dose was found to be equally as effective as the higher dose but with fewer side effects.

Abiraterone

Abiraterone acetate, an inhibitor of cytochrome P 17 is a key enzyme in androgen and oestrogen biosynthesis. Around 60–70% of breast cancers are thought to be androgen receptor-positive; however, the role of the androgen receptor in breast cancer remains incompletely understood and to date no role for abiraterone has been identified in treating hormone-receptor-positive breast cancer. A large, phase II trial did not demonstrate any difference between single agent exemestane or exemestane, abiraterone, and prednisolone. More work is needed to identify potential roles, if any, for abiraterone in breast cancer.

Endocrine Resistance

Despite adjuvant chemotherapy and endocrine therapy, a proportion of patients with ER-positive breast cancer will still relapse and ultimately die of the disease. Further developments depend on finding methods to prevent and overcome resistance to endocrine therapy.

Endocrine resistance may occur either initially (*de novo*) or subsequently (acquired). Exposure to long-term oestrogen deprivation and subsequent development of acquired resistance may be accompanied by adaptive increases in oestrogen receptor gene expression and intercellular signalling, resulting in hypersensitivity to low oestradiol levels.

HER2

HER2 has been implicated in the development of endocrine resistance. Combining endocrine therapy with agents targeted against HER2 such as trastuzumab or lapatinib has shown advantages in progression-free survival (PFS) and clinical benefit rates, although these are relatively modest [41, 42].

mTOR

The phosphoinositide 3 kinase (PI3K)/AKT/mammalian target of rapamycin (mTOR) pathway has also been investigated in numerous preclinical models.

In advanced breast cancer in postmenopausal women a phase III randomized trial, BOLERO-2 showed improved PFS (15.2 vs. 4.2 months) in patients receiving everolimus plus exemestane compared with exemestane alone [43]. This confirms the preclinical rationale for cotargeting these pathways.

This study led to the approval of everolimus plus exemestane in patients with ER+ve, HER2–ve breast cancer who had progressive disease with non-steroidal AIs.

Everolimus is currently also being evaluated in the adjuvant setting for early breast cancer (NCT01805271).

CDK4/6

Another pathway involving cyclin-dependent kinases 4 and 6 (CDK4 and CDK6) which promote tumour growth has also been implicated in the development of endocrine resistance. In the PALOMA-3 phase 3 study women were randomized to receive either palbociclib, an oral CDK4 and CDK6 inhibitor with fulvestrant or fulvestrant alone [44]. The combination arm showed significantly longer PFS (9.2 months vs. 3.8 months) and was well tolerated in the trial population.

The sister PALOMA-2 study investigated the addition of palbociclib to letrozole in the first-line setting for metastatic or advanced breast cancer and demonstrated improved PFS from 14.5 to 24.8 months [45]. This was the longest PFS ever reported in the first-line setting and has led to the adoption of letrozole + palbociclib as first-line treatment in metastatic breast cancer. Premenopausal women are given Zoladex alongside this combination.

Ribociclib, a CDK 4/6 inhibitor has recently been approved by the US Food and Drug Administration (FDA) for the use in combination with AI in postmenopausal women with hormone-receptor-positive, HER2 negative advanced breast cancer following results from the MONALEESA-2 study. In this study patients received either ribociclib + letrozole or placebo + letrozole. Improvements were seen in PFS and ORR [46].

Further progress will depend on understanding the mechanisms behind the development of endocrine resistance and the compensatory pathways that emerge in individual patients. The identification of biomarkers predictive of response will allow better selection of patients for treatment with specific combinations of targeted and endocrine therapies.

Conclusions

Adjuvant endocrine therapy for early breast cancer has led to significant improvements in both DFS and OS and is generally a well-tolerated treatment. The EBCTG [1] confirmed that 10 years of tamoxifen reduces both the annual recurrence rate and the breast cancer mortality rate.

In postmenopausal patients, AIs either upfront or in sequence offer further incremental benefits over tamoxifen. Ovarian suppression has shown clinical benefit in women aged under 35 and those at high risk of developing recurrent disease. Alongside this AIs have also demonstrated benefits in preventing disease recurrence in younger, premenopausal women in terms when used alongside ovarian suppression. AIs have had a major impact on the treatment of breast cancer and further progress depends on understanding the mechanisms that underlie the development of endocrine resistance.

Clinical trials examining the use of targeted agents as a means to overcome resistance are challenging given that compensatory pathways probably vary over time and from patient to patient. The identification of biomarkers will help to define which tumours are most likely to respond to particular treatments. Key aspects to be addressed in future will be the best trial design with appropriate target selection and patient selection with activation of the relevant target. Ultimately, this should help identify which patients benefit most from specific drug combinations and over the next few years we should learn whether this combination approach, of endocrine therapy with the new generation targeted agents, leads to significant improvements in the treatment and survival of hormone-positive breast cancer.

REFERENCES

1. Fisher B, Costantino JP, Wickerham DL, *et al.* Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst*, 2005; **97**: 1652–62.
2. Cuzick J, Forbes J, Edwards R, *et al.* First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. *Lancet*, 2002; **360**: 817–24.
3. Vogel VG, Costantino JP, Wickerham DL, *et al.* Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP study of tamoxifen and raloxifene (STAR) P-2 trial. *JAMA*, 2006; **295**: 2727–41.
4. La Croix AZ, Cummings SR, Delmas P, *et al.* Effects of 5 years of treatment with lasofoxifene on incidence of breast cancer in older women. *Cancer Res*, 2009; **69**: 11.
5. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, 2005; **365**: 1687–717.
6. Fisher B, Dignam J, Bryant J, Wolmark N. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. *J Natl Cancer Inst*, 2001; **93**: 684–90.
7. Peto R, Davies C on behalf of the atlas collaboration. ATLAS (Adjuvant Tamoxifen, Longer Against Shorter); Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS a randomised trial. *Lancet*, 2013; **381**: 805–16.
8. Gray RG, Rea DW, Handley K, *et al.* aTTom (adjuvant tamoxifen to offer more?): randomised trial of 10 versus 5 years of adjuvant tamoxifen among 6 934 women with estrogen receptor positive (ER+) or ER untested breast cancer—preliminary results. ASCO; 2008. *J Clin Oncol*, 2008; **26**(Suppl. 15S): 513.
9. Burnstein H, Temin S for the American Society of Clinical Oncology (ASCO). Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *J Clin Onc*, 2014; **32**: 2255–69.
10. Smith IE, Dowsett M. Aromatase inhibitors in breast cancer. *N Engl J Med*, 2003; **348**: 2431–42.
11. Hamilton A, Piccart M. The third-generation non-steroidal aromatase inhibitors: a review of their clinical benefits in the second-line hormonal treatment of advanced breast cancer. *Ann Oncol*, 1999; **10**: 377–84.
12. Nabholtz JM, Buzdar A, Pollak M, *et al.* Anastrozole is superior to tamoxifen as first-line therapy for advanced breast cancer in postmenopausal women: results of a North American multicenter randomized trial. Arimidex Study Group. *J Clin Oncol*, 2000; **18**: 3758–67.
13. Bonneterre J, Thurlimann B, Robertson JF, *et al.* Anastrozole versus tamoxifen as first-line therapy for advanced breast cancer in 668 postmenopausal women: results of the Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability study. *J Clin Oncol*, 2000; **18**: 3748–57.
14. Bonneterre J, Buzdar A, Nabholtz JM, *et al.* Anastrozole is superior to tamoxifen as first-line therapy in hormone receptor positive advanced breast carcinoma. *Cancer*, 2001; **92**: 2247–58.
15. Mouridsen H, Gershanovich M, Sun Y, *et al.* Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase III study of the International Letrozole Breast Cancer Group. *J Clin Oncol*, 2001; **19**: 2596–606.
16. Mouridsen H, Gershanovich M, Sun Y, *et al.* Phase III study of letrozole versus tamoxifen as first-line therapy of advanced breast cancer in postmenopausal women: analysis of survival and update of efficacy from the International Letrozole Breast Cancer Group. *J Clin Oncol*, 2003; **21**: 2101–9.
17. Paridaens R. First-line treatment for metastatic breast cancer with exemestane or tamoxifen in postmenopausal patients; a randomised phase III trial of the EORTC Breast Group. *Proc Am Soc Clin Oncol*, 2004.
18. Howell A, DeFriend DJ, Robertson JF, *et al.* Pharmacokinetics, pharmacological and anti-tumour effects of the specific anti-oestrogen ICI 182780 in women with advanced breast cancer. *Br J Cancer,* 1996; **74**: 300–8.
19. Early Breast Cancer Trials Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet*, 2015; **386**: 1341–52.
20. Goss PE, Ingle JN, Martino S, *et al.* A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med*, 2003; **349**: 1793–802.
21. Goss P, Ingle J, Pritchard K, et. al. Extending aromatase-inhibitor adjuvant therapy to 10 years. *N Engl J Med*, 2016; **375**: 209–19.
22. Tjan-Heilnen V, van Hellemond I, Peer P, *et al.* Extended adjuvant aromatase inhibition after sequential endocrine therapy (DATA): a randomised, phase 3 trial. *Lancet Oncol*, 2017; **18**: 1502–11.
23. Blok E, Kroep J, Meershohoek-Klein E, *et al.* Optimal duration of extended adjuvant endocrine therapy for early breast cancer; results of the IDEAL trial (BOOG 2006–05). *J Natl Cancer Inst*, 2018: **110**. DOI: 10.1093/jnci/djx134.
24. Fallowfield L, Cella D, Cuzick J, Francis S, Locker G, Howell A. Quality of life of postmenopausal women in the Arimidex, Tamoxifen, Alone or in Combination (ATAC) Adjuvant Breast Cancer Trial. *J Clin Oncol,* 2004; **22**: 4261–71.
25. Cuzick J, Sestak I, Cella D, Fallowfield L. Treatment-emergent endocrine symptoms and the risk of breast cancer recurrence: a retrospective analysis of the ATAC trial. *Lancet Oncol*, 2008; **9**: 1143–8.
26. Buzdar A, Howell A, Cuzick J, *et al.* Comprehensive side-effect profile of anastrozole and tamoxifen as adjuvant treatment for

early-stage breast cancer: long-term safety analysis of the ATAC trial. *Lancet Oncol*, 2006; **7**: 633–43.
27. Thurlimann B, Keshaviah A, Coates AS, *et al.* A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N Engl J Med*, 2005; **353**: 2747–57.
28. Hillner BE, Ingle JN, Chlebowski RT, *et al.* American Society of Clinical Oncology 2003 update on the role of bisphosphonates and bone health issues in women with breast cancer. *J Clin Oncol*, 2003; **21**: 4042–57.
29. Mouridsen H, Keshaviah A, Coates AS, *et al.* Cardiovascular adverse events during adjuvant endocrine therapy for early breast cancer using letrozole or tamoxifen: safety analysis of BIG 1–98 trial. *J Clin Oncol*, 2007; **25**: 5715–22.
30. Winer E, Hudis C, Burnstein H, *et al.* American Society of Clinical Oncology Technology Assessment on the use of aromatase inhibitors for post-menopausal women with hormone-receptor-positive breast cancer: status report 2004. Available at: http://www.asco.org (accessed: 1 July 2010).
31. Cuzick J, Ambroisine L, Davidson N, *et al.* Use of luteinising-hormone-releasing hormone agonists as adjuvant treatment in premenopausal patients with hormone-receptor-positive breast cancer: a meta-analysis of individual patient data from randomised adjuvant trials. *Lancet*, 2007; **369**: 1711–23.
32. Francis P, Regan M, Fleming G, *et al.* Adjuvant ovarian suppression in premenopausal breast cancer. *N Engl J Med*, 2015; **372**: 436–46.
33. Regan M, Pagania O, Fleming G, *et al.* International Breast Cancer Study Group SOFT and TEXT Investigators. Adjuvant treatment of premenopausal women with endocrine-responsive early breast cancer: design of the TEXT and SOFT trials. *Breast*, 2013; **22**: 1094–100.
34. Osborne CK, Coronado-Heinsohn EB, Hilsenbeck SG, *et al.* Comparison of the effects of a pure steroidal antiestrogen with those of tamoxifen in a model of human breast cancer. *J Natl Cancer Inst*, 1995; **87**: 746–50.
35. Chia S, Gradishar W, Mauriac L, *et al.* Double-blind, randomized placebo-controlled trial of fulvestrant compared with exemestane after prior nonsteroidal aromatase inhibitor therapy in postmenopausal women with hormone receptor-positive, advanced breast cancer: results from EFECT. *J Clin Oncol*, 2008; **26**: 1664–70.
36. Johnston S, Kilburn L, Ellis P, *et al.* Fulvestrant plus anastrozole or placebo versus exemestane alone after progression on non-steroidal aromatase inhibitors in postmenopausal patients with hormone-receptor-positive locally advanced or metastatic breast cancer (SoFEA): a composite, multicenter, phase 3 randomised trial. *Lancet Oncol*, 2013; **14**: 989–98.
37. Fribbens C, O'Leary B, Kilburn L, *et al.* Plasma ESR1 mutations and the treatment of estrogen receptor-positive advanced breast cancer. *JCO*, 2016; **34**: 2961–8.
38. Johnston S. Fulvestrant and the sequential endocrine cascade for advanced breast cancer. *Br J Cancer*, 2004; **90**(Suppl. 1): S15–18.
39. Mahtani RL, Stein A, Vogel CL. High dose estrogen as a salvage therapy for highly refractory metastatic breast cancer 'back to the future'. *Cancer Res*, 2009; **69**: 6129.
40. Ellis MJ, Dehdahti F, Kommareddy A, *et al.* A randomised phase 2 trial of low dose (6mg daily) versus high dose (30mg daily) estradiol for patients with estrogen receptor positive aromatase inhibitor resistant advanced breast cancer. *Cancer Res*, 2009; **69**: 16.
41. Kaufman B, Mackey JR, Clemens MR, *et al.* Trastuzumab plus anaztrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomised phase III TAnDEM study. *JCO*, 2009; **33**: 5529–37.
42. Johnston S, Pegraam M, Press M, *et al.* Lapatanib combined with letrozole vs letrozole alone for front line postmenopausal hormone receptor positive breast cancer: the first results from the EGF3008 trial. *Cancer Res*, 2009; **69**: 46.
43. Baselga J, Campone M, Piccart M, *et al.* Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med*, 2012; **366**: 520–9.
44. Turner N, Ro J, Andre F, *et al.* Palbociclib in hormone-receptor-positive advanced breast cancer. *N Engl J Med*, 2015; **373**: 209–19.
45. Finn R, Martin M, Rugo H, *et al.* Palbociclib and letrozole in advanced breast cancer. *N Engl J Med*, 2016; **375**: 1925–36.
46. O'Shaughnessy J, Petrakova K, Sonke G, *et al.* Ribociclib plus letrozole versus letrozole alone in patients with de novo HR+, HER2– advanced breast cancer in the randomized MONALEESA-2 trial. *Breast Cancer Res Treat*, 2018; **168**: 127–34.

13.2.3 Hormonal Therapy for Prostate Cancer

Molecular Basis of Efficacy and Therapeutic Bypass

Irina A. Vasilevskaya, Matthew J. Schiewer, and Karen E. Knudsen

Introduction

Despite tremendous advances in diagnosis and treatment, resulting in 51% decline in death rates from 1993 to 2016 [1], prostate cancer (PCa) remains the second leading cause of cancer-related death in American men. The American Cancer Society estimates 174 650 new cases and 31 620 deaths from the disease in 2019. Around 10% of men will be diagnosed with PCa during their lifetime, and for approximately 1 man in 40 this disease will be fatal [1]. As is the case for many cancers, ageing is one of the most significant risk factors for PCa. PCa develops mainly in older men, with a significantly higher frequency in African-American men. Diagnosis rarely occurs in men younger than 40 years old, with the majority of new cases diagnosed between the ages of 65 and 79. Current data suggest that only up to 10% of PCa cases are hereditary especially in early

onset disease [2], while other risk factors proposed in some studies, such as smoking and excess body weight, have not been universally accepted [1–3].

Androgens are indispensable for prostate development and normal function, including normal turnover of prostatic cells [4]. With age, the homeostatic balance between proliferation and cell death can become perturbed, giving rise to several abnormalities. For example, benign prostatic hyperplasia (BHP), occurring mostly in the transition zone of the prostate, is characterized by accelerated proliferation of stromal components and not considered to be a precursor of PCa [5]. Other disorders include proliferative inflammatory atrophy (PIA) and prostatic intraepithelial neoplasia (PIN). These selectively develop in the peripheral zone of the prostate and manifest through proliferation of luminal and basal epithelia (PIA) or hyperproliferation of luminal cells (PIN) [6]. High-grade PIN, in particular, exhibits some biochemical, genetic, and phenotypical characteristics of prostate cancer, and generally is accepted as the main premalignancy of the prostate [7].

The central feature of prostate cancer is a profound dependency on androgens. It was reported in 1941 by Huggins and Hodges that surgical castration, or orchiectomy, lead to significant decrease in primary tumour size and alleviation of symptoms in patients with metastatic prostate cancer. In the following years, feasibility of medical castration using inhibitors of steroid synthesis, which drastically reduce the levels of serum androgens, was established. Treatments leading to depletion of testosterone were appropriately termed androgen deprivation therapy (ADT). With the discovery and characterization of the androgen receptor (AR) as the crucial cellular mediator of androgen function during prostate tumorigenesis and disease progression, another promising avenue to combat PCa by directly targeting AR became the focus of vigorous research. As a result, a new class of therapeutics, androgen receptor antagonists, or antiandrogens, was developed and introduced into clinical usage [8].

Improvements in diagnostic capability, including development of the widely used test for prostate-specific antigen (PSA) in patient sera, contributed to earlier detection of PCa with the majority of new cases reported at the stage of local, organ-confined disease [1]. While localized tumours are successfully treated by either surgery or radiation (5-year survival rate of nearly 100%), patients diagnosed with locally advanced or metastatic disease harbour a less favourable prognosis (30% 5-year survival rate) [1]. As a first line of therapeutic intervention, patients with disseminated PCa undergo ADT alone or in combination with fist-generation antiandrogens (NCCN 2018 [9]). This approach is effective and results in substantial disease regression in the majority of patients. Unfortunately, these treatment strategies are not curative; recurrence is typically observed after 2–3 years and the disease progresses to a lethal stage termed castration-resistant prostate cancer (CRPC). CRPC retains dependency on AR, which is reactivated through multiple mechanisms and is able to sustain PCa growth despite low to undetectable levels of serum androgens [10]. Consequently, treatment of non-metastatic CPRC continues to include ADT, to maintain castrate levels of testosterone, with addition of second-generation AR antagonists (e.g. enzalutamide, apalutamide) or inhibitors of steroid metabolism (abiraterone), whereas combinatorial strategies for metastatic CRPC (mCRPC) include secondary hormonal therapy along with chemotherapy (docetaxel), immunotherapy (sipuleucel-T) or Radium-223 (NCCN 2018 [11]). Importantly, resistance to secondary hormonal therapy develops in mCRPC patients and results in a particularly aggressive disease course. Identifying the mechanisms that result in antiandrogen resistance continues to be an active area of research. Another detrimental result of persistent inhibition of AR signalling is emergence of tumours bearing some neuroendocrine characteristics, neuroendocrine PCa (NEPC), which may account for up to 15% of mCRPC tumours, while the *de novo* NEPC is extremely rare [12]. Tumours that no longer respond to hormonal therapy are treated with radiation, chemotherapy (platinum compounds in addition to docetaxel), and immunotherapy (NCCN 2018).

Essentially, PCa presents as a spectrum of diseases. The latest World Health Organization classification of PCa recognizes several types and subtypes of this malignancy which exhibit varying molecular, histopathological, and clinical characteristics [13]. Despite this heterogeneity, ADT and AR targeting remains the cornerstone of treatment for the locally advanced PCa, metastatic PCa, and CRPC, and more recently was utilized at the early stages of the disease [8, 14] (NCCN 2018). The subject matter of hormonal therapy in PCa has been extensively reviewed. Here, we are presenting an overview of the topic and latest data, with numerous excellent reviews cited throughout the chapter.

Androgen Receptor

AR Structure and Function

AR (NR3C4, nuclear receptor subfamily 3, group C, gene 4) is a ligand-dependent transcription factor belonging to a family of steroid hormone nuclear receptors that also includes oestrogen (ER) and glucocorticoid (GR) receptors, among others [15]. The AR gene is located within the X chromosome at Xq11-Xq12 and consists of 8 exons, spanning 2757 nucleotides, and 7 introns of varying length. The gene encodes for a 110 KDa protein comprised of 919 amino acid residues. Three major functional domains are recognized in AR (Figure 13.2.3.1a): the N-terminal transcriptional regulation domain (NTD), the DNA-binding domain (DBD) and the C-terminal ligand-binding domain (LBD). DBD and LBD are connected by short hinge region [15, 16].

The NTD is the largest and most variable domain of AR (amino acid residues 1–555). The NTD is a transactivation domain which binds to coactivators of the p160 family and basal transcription factors, TATA-box-binding protein (TBP), and TFIIF [15, 16]. The NTD also interacts with LBD, and this N-to-C-terminal interaction is requisite for maximal AR transcriptional activation function. The AR interacts with chromatin remodelling factors, such as components of the SWI/SNF complex, splicing factors, histone modifying enzymes involved with acetylation and methylation, DNA repair proteins, chaperones, cytoskeletal proteins, signal transduction proteins, as well as the ubiquitin/proteasome pathway [17]. Within the NTD, the region including amino acid residues from 142 to 485, termed activation function 1 (AF-1) is responsible for ligand-independent transcriptional activity of AR and is crucial for maximal AR activation. The NTD is characterized by the presence of variable polyglutamine (CAG) and

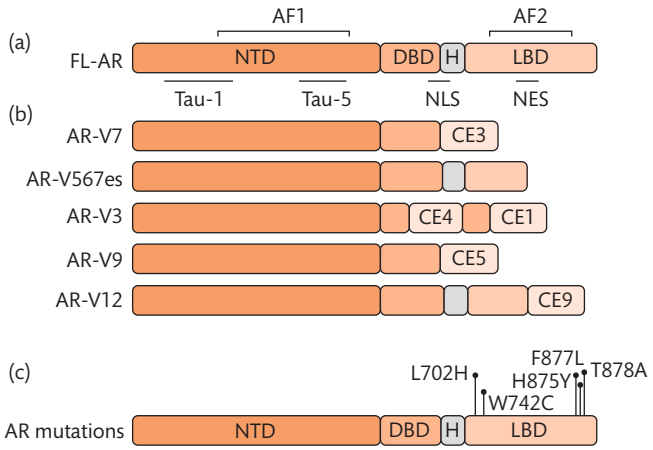


Figure 13.2.3.1 Androgen receptor structure (a), AR splice variants (b), and gain-of-function mutations (C). NTD, N-terminal transcription regulation domain; DBD, DNA-binding domain; H, hinge domain; LDB, C-terminal ligand-binding domain; AF1, activation function 1; AF2, activation function 2; NLS, nuclear localization signal; NES, nuclear export signal; CE1, cryptic exon 1; CE3, cryptic exon 3; CE5, cryptic exon 5; CE9, cryptic exon 9.

polyglycine (GGC) repeats, the lengths of which affect the conformation of the NTD and its interaction with protein partners, consequently impacting transcriptional activity of AR: shorter repeats are associated with higher AR activation, and longer repeats with reduced AR activation [15].

The DBD is a highly conserved cysteine-rich region (residues 556–623), which binds to AR-responsive elements (ARE) in the promoters and enhancers of target genes. Canonical ARE consists of two identical inverted half-sites 5'-AGAACA-3', separated by three nucleotides, and can be bound by AR and other nuclear receptors. The advent of next-generation sequencing and its application to ChIP (ChIP-Seq) greatly expanded the knowledge of the AR cistrome, identifying not only classic and AR-specific selective AREs (5'-GGTTCT-3' or canonical hexamer in direct orientation), but a variety of binding sites with nucleotide changes, half-sites and agonist- and antagonist-liganded AREs [15, 16], revealing the underlying basis of the capacity of AR to initiate diverse specific transcriptional programmes alone or in cooperation.

The LBD (residues 666–919) has the conformation of a three-fold sandwich comprised of eleven α-helices and four β-strands, forming a ligand-binding pocket. This domain also contains the ligand-dependent second activation function (AF2), part of which serves as a 'lid' closing after ligand binding [16].

In the inactive state, AR is sequestered in the cytoplasm in complex with inhibitory heat shock proteins (Hsp90, Hsp70, and p23). Ligand binding elicits conformational changes in AR and dissociation from chaperones, leading in turn to exposure of the nuclear localization signal situated between DBD and the hinge region of the protein. AR is then transported to the nucleus where it undergoes posttranscriptional modifications and dimerization, followed by binding to androgen receptor-responsive elements and coregulators through the zinc-finger of the DBD and AF1/AF2 regions, respectively (Figure 13.2.3.2a). The sum of these events leads to AR initiating a transcriptional programme that mediates a number of biological processes, including cell growth and proliferation, cell cycle progression, protein synthesis, and cell survival. AR-regulated target genes include *KLK3* (encoding PSA), and serum PSA is used in PCa diagnosis and monitoring PCa progression. Dissociation of ligand from the LBD results in the export of AR from the nucleus. In the healthy, non-transformed prostate, AR controls secretion of prostate-specific proteins, regulates lipid metabolism, and normal growth of prostatic epithelium in response to hormone [15, 16].

Androgen Receptor in Progression to CRPC

The molecular underpinnings of prostate tumorigenesis have been better defined in the last decade, as a result of rigorous scientific research and emergence of next-generation sequencing. Detailed analyses of genomic alterations in primary PCa revealed the most frequent genomic aberration—fusion of the AR-regulated TMPRSS2 gene promoter with the coding regions of ETS (E26 transcription specific) oncogene family members, mostly with ERG, which is present in nearly 60% of primary tumours, and also were detected in high-grade PIN cases [18]. These fusions amplify AR signal leading to accelerated growth of prostate epithelium. The ETS-fusion-negative tumours harbour mutually exclusive mutations of *SPOP, FOXA1*, and *IDH1* [18], and form a distinct group of PCa. ETS fusions and SPOP mutation occur early in PCa oncogenesis, followed by alterations of *CDKN1B, MED12, TP53,* and *PTEN*. Additional clinically relevant genes that are mutated in primary PCa, although at lower frequency, include: *ATM, CTNNB1, BRAF, HRAS*, and *AKT1* [18]. Together these genetic aberrations contribute to prostate tumorigenesis through deregulation of cell cycle, DNA repair, cell proliferation, and cell death.

The majority of hormone-sensitive PCa tumours respond well to ADT, but inevitably develop castration resistance, while still remaining AR-dependent [10]. Androgen withdrawal serves as a selective pressure that forces the initiation of numerous adaptive molecular alterations that allow PCa to survive in castrate conditions and to resume growth under control of reactivated AR. The molecular mechanisms of CRPC development include AR gene amplification and AR overexpression, altered expression and function of AR coactivators/corepressors, ligand-independent AR activation, gain-of-function mutations, constitutively active splice variants that lack the LBD, and intratumoral androgen synthesis [19] (Figure 13.2.3.2b). Gain-of-function mutations of the *AR* gene predominantly occur in the LBD (Figure 13.2.3.1c), and result in AR activation by alternative ligands, such as: progesterone, oestrogen, and antiandrogens (T878A, H875Y, W742C, F877L), or by glucocorticoids (H875Y, L702H). As such, *AR* gene mutations present a major clinical challenge [20]. The best described AR splice variant, constitutively, but weakly active AR-V7 (Figure 13.2.3.1b), confers resistance to ADT and is associated with short time to relapse after surgery, faster progression to CRPC and poor prognosis [20], and was recently shown to repress the expression of tumour suppressor genes [21]. AR splice variants AR-V567es, AR-V3, AR-V9, and AR-V12 are also clinically relevant: they are enriched in mCRPC and have been shown to contribute to a resistance to ADT [20]. In addition to AR axis dysregulation, other fundamental pathways are perturbed in CRPC. The most significant abnormalities enriched in CRPC include: hyperactivation of PI3K-AKT-mTOR pathway due to *PTEN* deletion/inactivation and genomic

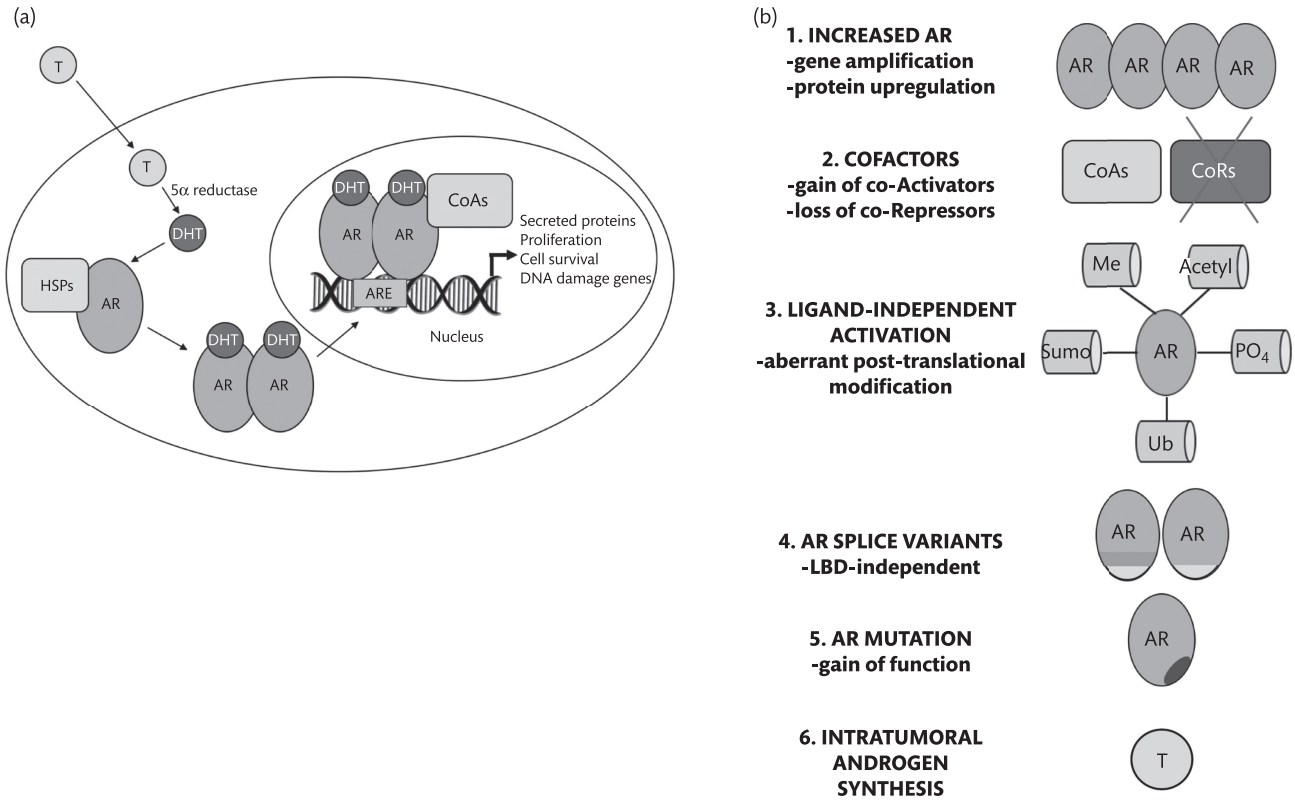


Figure 13.2.3.2 (a) Androgen receptor signalling in prostate cancer cells: testosterone (T) enters prostate cancer cells and is converted to the more potent androgen receptor (AR) ligand dihydrotestosterone (DHT) by 5α reductase. DHT then binds to the ligand-binding domain (LBD) of AR, eliciting conformation changes that release AR from inhibitory heat shock proteins (HSPs), thus permitting nuclear translocation. Once in the nucleus, AR binds to androgen-response elements (AREs) and recruits coregulators (including coactivators; CoAs) to elicit a transcriptional programme that includes genes that encode for secreted proteins, proliferation/cell survival genes, and DNA repair genes. (b) Mechanisms of androgen receptor reactivation in CRPC (castration-resistant prostate cancer). Cartoons indicate the six non-mutually exclusive mechanisms that have been reported to lead to androgen receptor (AR) reactivation in CRPC. Coactivators, CoAs; Corepressors, CoRs; Me, methylation; acetyl, acetylation; sumo, sumoylation; Ub, ubiquitylation; PO_4, phosphorylation; T, testosterone.

aberrations of kinases themselves, mutations of several DNA repair genes (*BRCA1, BRCA2*), deletions (*RB1, CDKN1*) or amplifications (*CDK4*) of cell cycle regulators, and changes in histone modifiers [22]. Collectively, these genomic aberrations constitute the distinct genetic makeup of CRPC, and may provide markers of progression as well as uncover novel nodes of therapeutic intervention.

Deregulated cell proliferation is a hallmark of cancer, and one mechanism of unchecked cell growth is loss of *RB1*. In PCa, deregulation of RB results in unique consequences. Knockdown of *RB1* results in an elevation of *AR* mRNA expression driven by the oncogenic transcription factor E2F1. This elevation of *AR* mRNA corresponds to aberrant heightened AR protein expression, which in turn results in castration-resistant AR function and CRPC *in vivo* [23]. While canonically, it is thought that RB loss results in uncontrolled cell proliferation, in the context of clinical PCa, there is no correlation between RB loss and a hyperproliferative phenotype [24]. Importantly, genomic aberrations (copy number alterations and structural variants) of *RB1* that render the RB protein inactive occur frequently in metastatic CRPC [25]. RB protein expression is heterogenous in PCa, *RB1* deletions are common in aggressive hormone therapy-sensitive PCa and more frequent in CRPC [25]. As such, deregulation of RB is frequent in PCa and results in deranged AR function and CRPC phenotypes.

Hormonal Therapy of Prostate Cancer

Androgen Deprivation

Regulation of steroid hormones: Testosterone (T) plays a central role in normal development of the male reproductive system, including the prostate, secondary sexual characteristics, and sexual behaviour. Levels of circulating T in healthy adult males can range from 2.6 to 9.6 ng/ml. Systemically levels of T are tightly regulated in a pulsatile manner and T biosynthesis is induced when the serum concentration falls below normal levels. T biosynthesis is a multistage process initiated in the hypothalamus, with the production of luteinizing hormone-releasing hormone (LHRH), also known as gonadotropin-releasing hormone (GnRH). GnRH induces the pituitary gland to release luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which in turn activate synthesis of T by Leydig cells in the testes [26]. Increased levels of testosterone lead to inhibition of LH and FSH release through a negative feedback loop mechanism, thus concluding the cycle of androgen regulation (Figure 13.2.3.3a). More than 95% of T is synthesized in the testes, with remaining T produced in the adrenal glands. Consequently, orchiectomy leads to dramatic decrease in circulating testosterone, with concentrations as low as 0.2 ng/ml [27]. In the prostate,

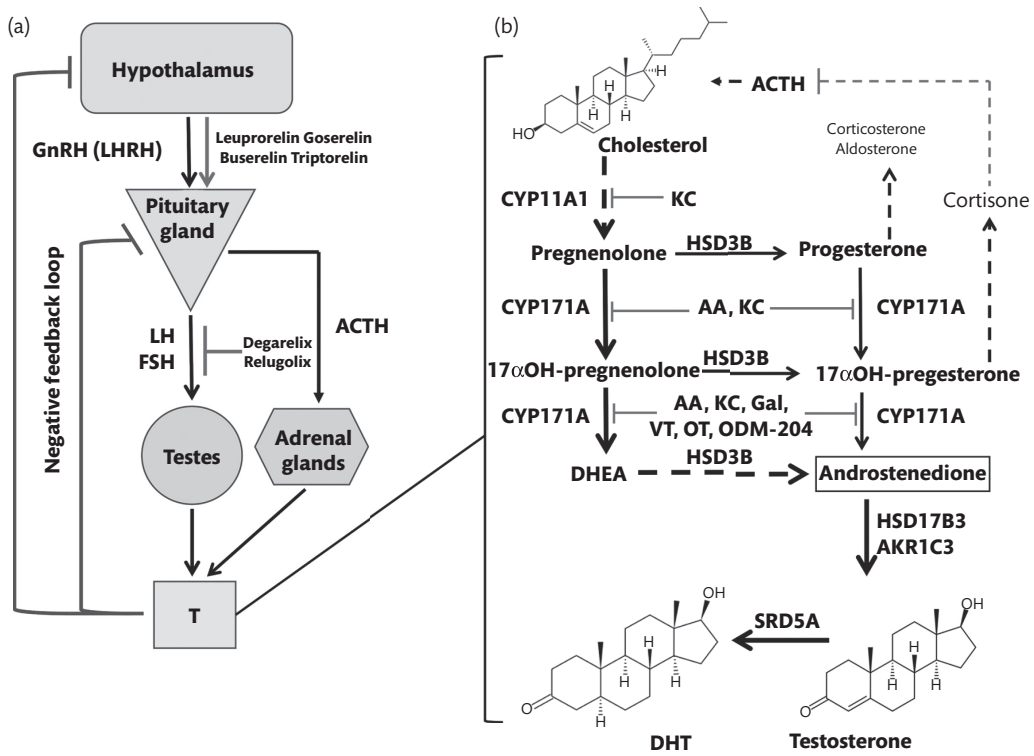


Figure 13.2.3.3 Regulation of testosterone biosynthesis. (a) Systemic regulation of testosterone (T) levels. (b) Scheme of T and dihydrotestosterone (DHT) biosynthesis from cholesterol. GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; FSH, follicle-stimulating hormone; ACTH, adrenocorticotropic hormone. KC, ketoconazole; AA, abiraterone acetate; Gal, galeterone; VT, VT-464; OT, orteronel. Thick arrows represent classical pathway of T biosynthesis, dashed lines depict multistep reactions.

testosterone is converted by 5α-reductase (*SRD5A*) to more potent metabolite, 5α-dihydrotestosterone (DHT), followed by activation of downstream signalling through AR.

Inhibitors of Testosterone Synthesis Through LHRH

Means of systemic testosterone depletion include both agonists and antagonists of GnRH. Agonists bind to and activate pituitary GnRH receptors, initially increasing LH and FSH release resulting in elevated testosterone levels. However, as a result of persisting activation, GnRH receptors become desensitized to the stimulus, which leads to a drop in LH and FSH secretion, effectively reducing androgen synthesis [26]. GnRH agonists are structurally similar to endogenous decapeptide hormone with slight modifications, such as amino acid substitution or alkylation. GnRH agonists currently used in clinic are leuprorelin, goserelin, triptorelin, and buserelin. Although all of the commercially available GnRH agonists were shown to achieve comparable castration levels of plasma testosterone, defined generally as less than 0.5 ng/ml [27], leuprorelin (also known as leuprolide) is the GnRH agonist most often prescribed in the clinic. Leuprolide was approved by the US Food and Drug Administration (FDA) in 1985, and serves as the primary ADT agent for both locally advanced and metastatic PCa (NCCN 2018). When delivered as subcutaneous implant, leuprolide causes a decrease of T to castrate level in 2–4 weeks, which can be maintained for up to 8 months, depending on formulation and treatment needs [28]. Adverse effects of GnRH agonists are due to low androgen levels, and include hot flashes, fatigue, anaemia, and decreased libido. Other detrimental effects, such as bone density loss, metabolic, cognitive, and cardiovascular events, including higher incidence of pulmonary embolism and stroke, were also observed upon treatment with GnRH agonists alone or in combination with antiandrogens. Initial elevation of T levels, so called testosterone flare, was considered a potential risk of driving disease progression, both at early stages and in advanced PCa. However, this view has evolved, as there is no definitive evidence that demonstrate direct consequences of testosterone flare on disease progression [29]. Nevertheless, to alleviate adverse effects of testosterone flare, pretreatment with or concomitant administration of first-generation antiandrogens are approved (NCCN 2018).

GnRH antagonists act as direct GnRH receptor blockers, effectively inhibiting the downstream signalling of the pituitary gland, leading to decreased T release by the testes without the negative effects of testosterone flare. The first GnRH antagonist, abarelix, was approved by FDA in 2004, but despite some beneficial features, was later withdrawn from the clinic in the United States. Currently, degarelix acetate is the only GnRH antagonist approved by FDA for a treatment of advanced PCa. Degarelix is a decapeptide analogue of endogenous GnRH, which was shown to suppress T production with an equal efficacy but more rapidly than leuprorelin, reaching castrate levels of the androgen within as few as 72 hours and maintaining them for up to a year when delivered monthly [30]. The main side effect of degarelix is injection site reaction. Multiple studies to interrogate efficacy and safety of degarelix versus GnRH agonists showed longer progression-free and overall survival in patients treated with the antagonist [30]. There is an ongoing controversy regarding the use of GnRH antagonists versus agonists, and

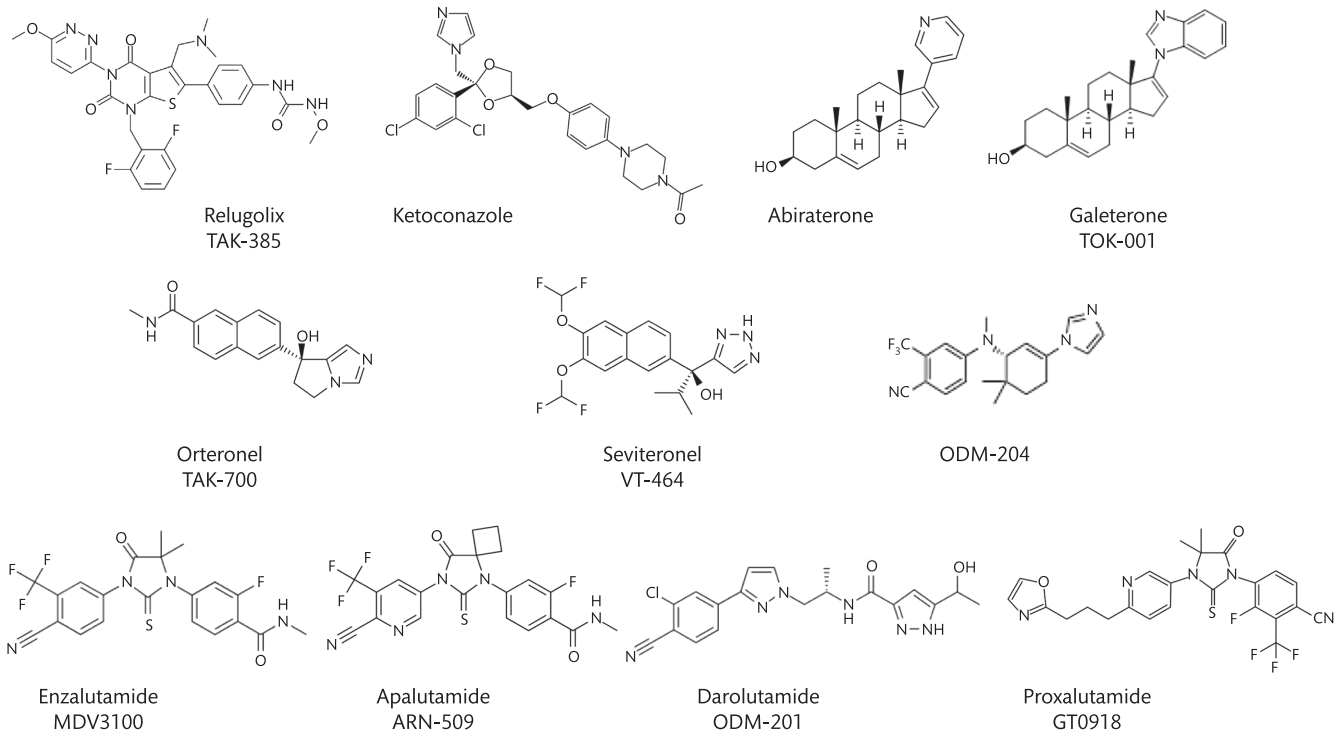


Figure 13.2.3.4 Chemical structure of GnRH antagonist relugolix, CYP17A1 inhibitors, and second-generation AR antagonists.
Taken from original publications and https://pubchem.ncbi.nlm.nih.gov.

degarelix is yet to equal leuprorelin for use in the clinic, but the results of clinical trials, completed and ongoing, indicate its future wider use in all stages of the PCa [31]. The main limiting factor for use of the peptide GnRH antagonists is low solubility and short half-life, mandating frequent injections. To address the disadvantage of peptide-based therapies, development of non-peptide GnRH antagonists is a field of current research effort. One of the non-peptide GnRH antagonists potentially effective in the treatment of PCa, is the oral thienopyridine compound relugolix, TAK-385 [32] (Figure 13.2.3.4). In phase I and II clinical trials, when given as a tablet once daily, relugolix demonstrated an adequate safety profile and achieved castrate T levels as quickly as degarelix, and much faster than leuprorelin (NCT02141659, NCT02135445, NCT02083185). Relugolix is currently in phase III clinical trial in men with advanced PCa, to evaluate retaining of castration in comparison to leuprorelin (NCT03085095). Both GnRH agonists and antagonists curtail gonadal T synthesis without impacting T production by adrenal glands.

Testosterone Biosynthesis

T is synthesized from cholesterol, the common precursor of steroid hormones, in a cascade of reactions mediated by several enzymes [33] (Figure 13.2.3.3b). Mitochondrial CYP11A1, a cholesterol side chain cleavage enzyme, converts cholesterol into pregnenolone in three sequential steps. The remainder of T biosynthesis occurs in the endoplasmic reticulum. 3β-hydroxysteroid dehydrogenase (HSD3B) catalyses the processing of pregnenolone into progesterone, which is crucial for the biosynthesis of all steroid hormones. Cytochrome P450 17α-hydroxylase/17,20-lyase (CYP17A1) possesses dual enzymatic activity, first acting as 17α-hydroxylase and then as 17,20-lyase to mediate conversion of pregnenolone to 17αOH-pregnenolone, and of 17αOH-pregnenolone to dehydroepiandrosterone (DHEA), respectively. 17β-hydroxysteroid dehydrogenase 3 (HSD17B3) mediates final step from androstenedione to testosterone. Once synthesized, T is metabolized into DHT by 5α-reductases [33]. Several small molecules, mostly inhibitors of CYP17A1, were developed and introduced to the clinic to further suppress gonadal T synthesis, which are also capable of inhibiting adrenal and intratumoral androgen synthesis. The problem arising from full inhibition of CYP17A1 is that its 17α-hydroxylase activity is necessary for biosynthesis of corticosteroids, namely cortisol, from 17αOH-progesterone, but not for mineralocorticoids produced downstream of progesterone (Figure 13.2.3.3b). Inhibition of 17α-hydroxylase leads to suppression of cortisol and abrogation of negative feedback loop to regulate pituitary adrenocorticotropic hormone (ACTH), in turn resulting in abnormal activation of steroidogenic pathway and excess of mineralocorticoids. Thus, there is a need for clinically active CYP17A1 inhibitors with strong specificity towards its 17,20-lyase function, which so far was not fulfilled despite several promising candidates briefly described next [34].

Inhibitors of Testosterone Synthesis

Ketoconazole, a synthetic imidazole compound (Figure 13.2.3.4), is a non-specific cytochrome P450 inhibitor approved for a treatment of multiple fungal infections. The inhibitory effects of ketoconazole on steroidogenesis were reported in 1982, leading to off-label use in CRPC patients. Ketoconazole potently inhibits CYP17A1, resulting in a substantial decrease of T levels within 2 to 8 hours, but the effect is short-lived [34, 35]. However, by also inhibiting CYP3A, CYP11A, and CYP24A, ketoconazole impinges on multiple

oxidative reactions catalysed by cytochromes, including potentially adversely modulating drug-drug interactions, although that remains controversial. Ketoconazole was also linked to a significant hepatic toxicity, which caused FDA to issue a safety announcement regarding the use of ketoconazole as an antifungal agent in 2013. In PCa, several studies of ketoconazole as monotherapy showed that, while not prolonging overall survival (OS), ketoconazole demonstrated acceptable toxicity and moderate efficacy in decreasing PSA levels and extending time to PSA progression. It also demonstrated clinical activity in combination with docetaxel [35]. While the newer and more effective agent, abiraterone acetate, has become the leading therapeutic for inhibition of T biosynthesis in mCRPC, ketoconazole is still considered a low-cost option for treatment of patients with non-metastatic CRPC or in cases of mCRPC when standard therapies are not available (NCCN 2018) [35].

Based on the earlier studies with ketoconazole, inhibition of CYP17A1 was identified as a valid approach to inhibit T biosynthesis in PCa treatment. Search for a less toxic alternative led to the development of abiraterone acetate in the mid-1990s. Abiraterone acetate (Figure 13.2.3.4), an oral pregnenolone derivative and irreversible inhibitor of CYP17A1, showed potent downregulation of both 17α-hydroxylase and 17,20-lyase activities in preclinical studies [36]. Treatment with abiraterone acetate suppresses cortisol synthesis and can result in mineralocorticoid excess syndrome (MES), which is minimized by concomitant administration of prednisone. The mechanism of abiraterone acetate function has been well studied. Clinical development of abiraterone acetate progressed rapidly from initial phase I trials in 2008 through phase II studies to a ground-breaking phase III trial, COU-AA-301 (NCT00638690) which demonstrated prolonged OS (4.6 months) with abiraterone acetate versus placebo in post-docetaxel mCRPC patients [37]. The following phase III trial in chemotherapy-naïve mCRPC patients showed similar effects of abiraterone acetate on OS (NCT00887198) [37]. After approval by FDA in 2011, abiraterone acetate was evaluated in combination with ADT in patients with hormone-sensitive metastatic PCa in two large randomized trials, LATITUDE (NCT01715285) and STAMPEDE (NCT00268476), both confirming significant improvement in progression-free survival (PFS) and OS in the ADT/abiraterone arm as compared to ADT alone [38, 39]. Subsequently, hundreds of clinical studies were completed worldwide, with over 100 active and recruiting trials just in the United States currently involving abiraterone, alone or as a complement to a standard ADT, in combination with chemotherapy, radiation, and second-generation antiandrogens at various stages of the disease. In addition to abiraterone-induced mineralocorticoid excess characterized by hypertension, fluid retention, and hypokalaemia [37, 38], and the undesirable side effects of continuous use of corticosteroids, a major limiting factor for abiraterone therapy is primary or acquired resistance. Molecular mechanisms of abiraterone resistance includes upregulation of intratumoral androgen synthesis, AR mutations, AR splice variants (AR-V7 and AR-V3, in particular), and glucocorticoid receptor overexpression, among others [40].

Galeterone (TOK-001) is an oral 17-heteroazole steroid analogue (Figure 13.2.3.4) which was developed as a potent inhibitor of CYP17A1 with a 3-fold selectivity towards 17,20-lyase activity. Subsequently, galeterone was confirmed as an antiproliferative agent for PCa *in vitro* and *in vivo*, and as an antagonist of both wild-type and mutant (T878A) AR [41]. In addition, galeterone prevents nuclear translocation and induces effective degradation of both FL-AR (full-length AR) and AR-V7 splice variant [34, 41]. These preclinical data demonstrating the ability of galeterone to disrupt androgen signalling on multiple levels, without significantly reducing cortisol, were highly encouraging. As a compound structurally similar to DHEA, galeterone was shown to be metabolized by HSD3B to Δ^4-galeterone, which is as effective against steroidogenesis and AR signalling as galeterone *in vitro* and *in vivo*. However, further conversion by SRD5A abrogates AR inhibitory activity of resulting metabolites, which potentially can compromise clinical efficacy of the drug [42]. The FDA granted Investigational New Drug status to galeterone in 2009, leading to clinical phase I and II trials, the ARMOR1 and two-part ARMOR2 trials (NCT00959959, NCT01709734) to assess safety and efficacy of oral galeterone in CRPC patients. Treatment was well tolerated and lead to a more than 50% PSA decline (PSA_{50}) in 70% of patients (combined ARMOR1 and ARMOR2 Part 1) [43]. Data from ARMOR2 Part 2 were presented at the twenty-third EACR Congress: galeterone not only lead to PSA_{50} levels in 75% of M0 and M1 treatment-naïve CRPC patients, but also resulted in measurable PSA decreases in the abiraterone-refractory cohort. More importantly, in 83% of the AR splice variant positive patients, as measured in circulating tumour cells, PSA_{50} levels were achieved, confirming preclinical data on efficacy of galeterone towards truncated AR variants [43]. However, ARMOR3-SV (NCT02438007), a phase III trial initiated in 2015 to compare galeterone with enzalutamide in mCRPC patients expressing AR-V7, was terminated due to a low probability of achieving the primary outcome of radiographic progression-free survival (rPFS) of more than 8 months.

Orteronel (TAK-700) is a non-steroidal imidazole derivative (Figure 13.2.3.4) with 5-fold increased selectivity towards CYP17 17,20-lyase. It has shown potent downregulation of DHEA and TT levels in *in vitro* and *in vivo* preclinical studies without cortisol suppression [34, 44]. Subsequently, in several phase I and phase II clinical trials in metastatic and non-metastatic CRPC, orteronel was well tolerated and demonstrated an effective decrease in TT, DHEA, and PSA levels as monotherapy and in combination with prednisone or docetaxel/prednisone [34, 45]. Orteronel then quickly moved into phase III trials. In the phase III ELM-PC 4 randomized trial (NCT01193244) orteronel was compared to a placebo, both given with prednisone, in chemotherapy-naïve patients [34, 37]. Primary endpoints for this trial were rPFS and OS. The phase III randomized ELM-PC 5 trial (NCT01193257) assessed the same combinations in mCRPC patients progressed after docetaxel, with OS as a primary endpoint [34, 37]. In both trials, orteronel demonstrated acceptable tolerance, significant PSA responses, and improved rPFS as compared to placebo. However, there was no improvement in OS, which led to a decision to unblind phase III studies and terminate further development of orteronel by the manufacturer, causing termination or withdrawal of several trials (clinicaltrials.gov). Despite not meeting the endpoint of OS in the aforementioned studies, orteronel has shown significant improvement in event-free and rPFS (8.5 vs. 2.9 months as compared to placebo) in patients with mCRPC who had stable disease after docetaxel therapy [46]. Currently, two phase III randomized clinical trials are still ongoing: one comparing LHRH agonist/orteronel and LHRH agonist/bicalutamide combinations in patients with newly diagnosed metastatic PCa (NCT01809691) and

another to assess addition of orteronel to dose-escalated radiation therapy in combination with standard ADT (GnRH agonists and antiandrogens prior to RT) in men with high-risk prostate cancer (NCT01546987). The primary outcome measure for both trials is OS. The results of these trials will indicate the feasibility of orteronel for use in both early and advanced PCa.

VT-464 (also marketed as seviteronel) is another novel non-steroidal oral triazole-based (Figure 13.2.3.4) small molecule inhibitor of CYP17A1 17,20-lyase activity. In preclinical studies, VT-464 showed an effective decrease of T concentrations *in vivo* without significant increases in progesterone and ensuing mineralocorticoid overproduction. VT-464 has demonstrated greater inhibition of androgen synthesis in CRPC and enzalutamide-resistant cell models as compared to abiraterone acetate. In addition, more potent inhibition of AR transactivation and decrease of AR-regulated proteins, including PSA, were observed, indicating that VT-464 acts as AR antagonist *in vitro* [34, 47]. Maity *et al.* [48] reported significant reduction of serum DHEA and PSA in mCRPC patient, and, using patient-derived xenograft (PDX) mouse models, demonstrated decreased tumour volume, downregulation of androgens, and inhibition of AR signalling equal or superior to abiraterone acetate. Notably, AR antagonism by VT-464 *in vitro* was demonstrated for full-length T878A AR mutant but not AR-V7, indicating efficacy against ligand-dependent AR functions. VT-464 has been evaluated in several recently completed open label phase I/II clinical trials for safety, tolerability, and efficacy in chemotherapy-naïve CRPC patients (NCT02361086) or previously treated with abiraterone, enzalutamide or chemotherapy (NCT02361086, NCT02012920, NCT02445976). In all studies, VT-464 was given orally once daily in 28-day cycles without steroid coadministration. Primary outcome measures included tolerability or proportion of patients reaching PSA_{50} levels (6 months) and median time to radiographic disease progression (10 months). The results of two latter trials have not yet been published.

ODM-204, novel non-steroidal compound, was synthesized with the purpose of incorporating a cytochrome binding moiety within an antiandrogen (Figure 13.2.3.4), thus enabling dual action against both CYP17A1 and AR. ODM-204 was shown to inhibit CYP17A1 and reduce testosterone *in vitro* with lower efficacy than abiraterone acetate, but as affectively as galeterone. When assessed for AR binging and antagonism, ODM-204 was superior to enzalutamide, the leading second-generation antiandrogen. The antiproliferative effects of ODM-204 were also demonstrated *in vitro* and confirmed *in vivo:* in VCaP xenografts, ODM-204 treatment led to reduction in tumour volume in intact and castrated mice with higher efficiency than abiraterone acetate [49]. On the basis of these promising preclinical studies, a phase I/II multicentre trial was initiated in 2015 to evaluate safety, tolerability, and pharmacokinetics of ODM-204 given in combination with prednisone to patients with metastatic CRPC (NCT02344017). Early results of this study have reported an acceptable safety profile and efficacy in decreasing serum testosterone and PSA levels, but the unsatisfactory pharmacokinetic features of ODM-204 led to a decision to cease further development of this compound [49].

While it is clear that targeting the AR signalling axis has clinical benefit in the management of PCa, the main limitation of ADT is the development of castration resistance. As such, a greater understanding of the mechanisms that lead to therapeutic bypass and means to thwart recurrent PCa are urgently needed.

AR Antagonists

Principle of Action

AR antagonists, or antiandrogens, are small molecules that compete with T and DHT for binding to the androgen receptor. Antiandrogens differ in chemical structure and efficacy, can exert varying side effects, and generally are divided into steroidal and non-steroidal compounds [8]. Steroidal antiandrogens, cyproterone acetate among others, were shown to effectively block binding of endogenous androgens to AR, but with partial agonistic activity paired with significant adverse effects prevented their approval by the FDA. Development of non-steroidal antiandrogens started in the 1960s, and since then efforts have been directed towards development of new non-steroidal antiandrogen therapeutics exhibiting higher affinity/specificity for AR, longer half-life, and fewer side effects [8].

First-Generation Antiandrogens

The first generation of non-steroidal AR antagonists (flutamide, nilutamide, and bicalutamide) bind to the LBD of AR, and thus compete for androgen binding to and activation of AR. Compared to T and DHT, these compounds have low relative affinity for the receptor, but based on preclinical studies and clinical trials, each of the aforementioned first-generation antiandrogens were FDA approved for use in PCa (flutamide, 1989; bicalutamide, 1995; nilutamide, 1996). However, these compounds also demonstrate partial AR agonism, evidenced by the clinical phenomenon of antiandrogen withdrawal response, in which upon discontinuation of use of these antiandrogens, serum PSA levels rise. As will be described next, second-generation antiandrogens have been rigorously developed to better target AR directly [8].

Second-Generation Antiandrogens

Enzalutamide (MDV-3100) is an oral non-steroidal antiandrogen which inhibits AR signalling by blocking the receptor and reducing AR nuclear transport, DNA binding, and recruitment of coactivators (Figure 13.2.3.4). In preclinical studies enzalutamide demonstrated up to 8-fold higher affinity to AR than bicalutamide, and induced tumour regression in mouse models of CRPC [50]. In a phase I/II clinical trial in patients that had received prior chemotherapy or were chem-naïve, enzalutamide was well tolerated, and demonstrated inhibition of DHT binding to AR, significant PSA response (41 weeks), and prolonged time to PSA progression (21 weeks) [51]. These Phase I/II studies were followed by two large, randomized phase III trials: AFFIRM (NCT00974311) [8, 37, 38], in mCRPC patients treated with prior docetaxel, and PREVAIL (NCT01212991) [8, 37, 38], in chemotherapy-naïve mCRPC patients. In both trials, enzalutamide demonstrated improved OS, increased time to PSA progression and rPFS [37, 52]. Based on the outcomes of these trials, enzalutamide was approved by FDA for a treatment of mCRPC patients. Subsequently, two more phase II randomized studies, STRIVE (NCT01664923) and TERRAIN (NCT01288911) in patients with or without metastatic disease showed superiority of enzalutamide over bicalutamide, with an average PFS advantage of 10 months [8, 37]. More recently enzalutamide was also tested as a monotherapy in phase III double-blind, randomized trial in non-metastatic CRPC (nmCRPC) patients, where enzalutamide

demonstrated significant improvements in metastasis-free survival (MFS), PSA progression and time to subsequent therapy (PROSPER, NCT02002924) [53]. In 2018, enzalutamide was approved by the FDA as a treatment for nmCRPC patients. In addition, two active trials are set to evaluate the safety and efficacy of enzalutamide in metastatic hormone-sensitive PCa (mHSPC): in the ARCHES trial, long-term enzalutamide + ADT will be compared to placebo + ADT (NCT02677896), with rPFS as primary outcome at 24 weeks; ENZAMET will measure OS with enzalutamide (with LHRH agonist) as first-line therapy in comparison to conventional non-steroidal antiandrogens (NCT02446405). Preliminary data from the ARCHES study show that enzalutamide + ADT significantly improves rPFS (NR versus 19.4 months on placebo), and PSA response (undetectable in 68.1% patients vs. 17.6% with placebo), although OS data have not been reported to date [54]. Main adverse side effects of enzalutamide treatment are cardiovascular events and seizures. Another major limiting factor for enzalutamide efficacy is the development of resistance, which is based on both AR-dependent and AR-independent molecular mechanisms. AR-dependent mechanisms include AR amplification and overexpression, AR splice variants (AR-V7), activating mutations (T878A, H875Y, F877L), and altered steroidogenesis. Main AR-independent mechanisms are overactivation of the glucocorticoid receptor and the PIK3-AKT pathway, induction of autophagy, EMT and gain of neuroendocrine characteristics [51]. Selected recruiting trials are indicated in **Table 13.2.3.1**.

Apalutamide (ARN-509) is a potent antiandrogen, similar in structure and with similar mechanisms of action to enzalutamide (**Figure 13.2.3.4**), which in preclinical studies showed maximal suppression of tumour growth at lower doses than enzalutamide. Apalutamide also has reduced blood–brain barrier penetrance, which may lead to alleviation of the adverse effects of enzalutamide on central nervous system [55]. In a phase I trial in mCRPC patients, apalutamide was well tolerated and a daily dose of 240 mg was recommended for future studies [8]. A Phase II trial expanded to three cohorts: patients with nmCRPC, and patients with mCRPC with or without prior abiraterone treatment. The primary endpoint in the nmCRPC cohort was a change in PSA levels at 12 weeks, and the secondary endpoint was median time to PSA progression. At 12 weeks, 89% of participants reached PSA_{50}, with median time to PSA progression of 24 months [8]. For the abiraterone-naïve and prior abiraterone mCRPC cohorts, PSA_{50} was achieved in 88% and 22% patients, with median time to PSA progression of 18.2 and 3.7 months, respectively [8].

Results of these studies have become the basis of a pivotal randomized, double-blind, placebo-controlled multicentre phase III trial, SPARTAN (NCT01946204) [56], in nmCRPC patients with continuing ADT. Primary outcome was MFS, while PFS, time to consequent chemotherapy, and OS were among secondary endpoints. Apalutamide showed significant improvement over placebo, with MFS of 40.5 versus 16.2 months. Time to PSA progression was not reached in apalutamide arm and was 3.7 months in the placebo arm. Adverse effects were slightly higher in the apalutamide arm, with the most prevalent being fatigue, rash, and hypertension. There were no grade 3/4 seizures recorded, since the patients with a history or predisposition to seizures were excluded from this study. After one year of additional follow-up, PFS2 (PFS on subsequent treatment from trial randomization) was evaluated and showed that apalutamide continued to be safe and superior to placebo: PFS2 of N/R versus 39.2 months, respectively [57]. Following the outcomes of the SPARTAN trial, apalutamide was approved by the FDA for a treatment on nmCRPC. A phase III randomized, placebo-controlled, double-blind study to determine if the addition of apalutamide to ADT (GnRH agonist or antagonist) improves rPFS or OS for participants with mHSPC (TITAN) is ongoing (NCT02489318). In another active phase III randomized, placebo-controlled, double-blind trial the apalutamide-abiraterone combination will be evaluated in chemotherapy-naïve mCRPC patients (NCT02257736). Selected recruiting trials are indicated in **Table 13.2.3.1**.

Darolutamide (OMD-201) is a non-steroidal antiandrogen structurally distinct from other AR antagonists (**Figure 13.2.3.4**). It was discovered during a compound library screen in AR-HEK293 cells using AR transactivation assay as a readout. In preclinical studies, darolutamide and its active metabolite, keto-darolutamide, demonstrated higher binding affinity to AR as compared to enzalutamide and apalutamide, greater inhibition of AR nuclear translocation, and stronger suppression of cell growth *in vitro* and *in vivo*. In addition, darolutamide blocked the activity of AR mutants, including F877L, which has been implicated in the resistance to enzalutamide and apalutamide. Darolutamide did not increase serum T levels and displayed negligible ability to cross blood–brain barrier [58]. All these features made darolutamide a very promising, more effective, and less toxic, novel AR antagonist for a treatment of PCa. In phase I studies, ARADES (NCT01429064) evaluated safety and bioavailability of darolutamide in mCRPC patients, including chemotherapy- and CYP17i-naïve patients; and ARAFOR (NCT01784757) assessed the efficacy of darolutamide in tablet form in chemotherapy-naïve men with progressive mCRPC [8, 59]. Phase II open label trial to compare darolutamide (1200 mg daily) to ADT (24 weeks of LHRH agonist or antagonist, including leuprolide, goserelin, triptorelin and degarelix) is currently recruiting hormone-naïve PCa patients (NCT02972060). Primary outcome will be PSA response at 24 weeks, defined as ≥80% reduction from the baseline, to demonstrate equal efficacy of darolutamide to standard ADT. A randomized, double-blind phase III trial, ARAMIS (NCT02200614), assessed the efficacy and safety of darolutamide versus placebo in men with high-risk nmCRPC, with MSF as the primary endpoint (up to 72 months), and OS and time to first chemotherapy included as secondary outcomes. Results from the ARAMIS trial showed median MSF of 40.4 in the darolutamide arm versus 18.4 months in the placebo arm; for all secondary endpoints darolutamide also demonstrated significant benefits, with adverse events similar in both cohorts [8, 60]. Another ongoing randomized, double-blind, placebo-controlled Phase III study, ARASENS (NCT02799602), is set to evaluate darolutamide in combination with standard ADT and docetaxel in mHSPC patients. Primary outcome measure is OS (up to 70 months), secondary endpoints are time to CRPC, time to subsequent therapy, and symptomatic skeletal event-free survival, among others. Selected recruiting trials are indicated in **Table 13.2.3.1**.

Proxalutamide (TG0918) is an oral non-steroidal antiandrogen which has shown higher affinity for AR than bicalutamide and enzalutamide, and more potency in blocking AR nuclear translocation and transcriptional activity (**Figure 13.2.3.4**). TG0918

Table 13.2.3.1 Selected recruiting clinical trials of second-generation antiandrogens

Phase	Disease stage	Experimental intervention(s)	Comparator intervention	Primary endpoint	Trial number
Enzalutamide					
Phase II/III	Oligometastatic CRPC	ADT + SBRT + enzalutamide	ADT + enzalutamide	rPFS	NCT02685397
Phase II/III	mCRPC previously treated with docetaxel and features of poor prognosis	ADT + abiraterone **OR** ADT + enzalutamide	ADT + cabazitaxel	CBR	NCT03295565
Phase III	Bone-metastatic CRPC	Radium-223 + enzalutamide	enzalutamide	rPFS	NCT02194842
Phase III	Homologous recombination-deficient CRPC	ADT + rucaparib	ADT+abiraterone **OR** enzalutamide **OR** docetaxel	rPFS	NCT02975934
Phase III	mCRPC	ADT + talazoparib) + enzalutamide	ADT + enzalutamide	rPFS	NCT03395197
Apalutamide					
Phase II	Hormone-naïve locally advanced or metastatic PCa	apalutamide **OR** apalutamide + abiraterone	ADT + abiraterone	PSA ≤ 0.2 ng/ml at 25 weeks	NCT02867020
Phase II	Localized high-risk PCa	neoadjuvant ADT + abiraterone **OR** neoadjuvant ADT + abiraterone + apalutamide	None	pCR or pnCR at prostatectomy	NCT02789878
Phase II	Hormone-naïve primary PCa	neoadjuvant ADT + abiraterone + apalutamide **OR** neoadjuvant ADT + abiraterone **OR** adjuvant ADT + abiraterone + apalutamide	No neoadjuvant or adjuvant therapy	Minimal residual disease and pCR at 2 years	NCT02903368
Phase II	Low risk primary PCa	active surveillance during and after 6 months apalutamide	Active surveillance	Time to definitive local treatment	NCT03088124
Phase II	Very high-risk primary PCa **OR** Hormone-naïve locally advanced or low volume metastatis **OR** Biochemical recurrence after prostatectomy with extrapelvic metastases	ADT + apalutamide **OR** ADT + abiraterone + apalutamide	None	Minimal residual disease and pCR at 2 years	NCT03436654
Phase II	Rising PSA after radical prostatectomy	ADT + bicalutamide followed by salvage RT **OR** ADT + abiraterone + apalutamide followed by salvage RT	None	PSA PFS	NCT03141671
Phase III	High-risk biochemically relapsed PCa	ADT + apalutamide **OR** ADT + abiraterone + apalutamide	ADT	PSA PFS	NCT03009981
Darolutamide					
Phase II	Hormone-naïve PCa (any stage)	Darolutamide	ADT	PSA reduction of ≥ 80% at 24 weeks	NCT02972060
Phase II	mCRPC after prior novel hormonal agent and non-progressive after taxane	Darolutamide	Placebo	rPFS at 12 weeks	NCT02933801

CRPC, castration-resistant prostate cancer; ADT, androgen deprivation therapy; SBRT, stereotactic body radiation therapy; rPFS, radiographic progression-free survival; mCRPC, metastatic castration-resistant prostate cancer; CBR, clinical benefit rate; PCa, prostate cancer; PSA, prostate-specific antigen; pCR, pathologic complete response; pnCR, pathologic near complete response; RT, radiation therapy; PFS, progression-free survival.

also induces downregulation of AR and inhibits proliferation of hormone therapy-sensitive and CRPC cell models *in vitro* and *in vivo* [61]. In an open label phase I/II dose-escalation trial (NCT02826772), GT0918 was given daily at multiple doses for 28 days (up to six cycles planned) to mCRPC patients who progressed after both hormonal and chemotherapy. The drug was well tolerated (at 500 mg dose-limiting toxicity was not achieved) and resulted in stable disease. The phase II/dose extension stage will further evaluate safety, efficacy, and antitumour activity of GT0918 in mCRPC patients [62].

Alternative Approaches

While the clinical data demonstrate that second-generation antiandrogens are likely to improve outcomes for PCa patients, resistance to enzalutamide has been well documented. The same AR activating mutations which result in enzalutamide resistance were

also shown to confer resistance to apalutamide, and AR splice variants diminish efficacy of all LBD-directed drugs in CRPC [20]. With the ultimate goal to maximally inhibit AR signalling axis by targeting wild-type, gain-of-function mutants, and splice variants of AR, multiple alternative approaches are being tested in preclinical and clinical settings.

Considerable efforts have been directed at targeting the NTD domain of AR [63]. One of the most developed compounds is EPI-001, a bisphenol A-derived small molecule antagonist of AR NTD that was identified in a screening of a library of marine sponge extracts for inhibition of both ligand-dependent and ligand-independent AR activation [63]. EPI-001 and its analogues (also called anitens) inhibit protein-protein interactions necessary for AR transcriptional activity by covalently binding to AF1 region of the NTD and blocking transcriptional activity of AR, its splice variants and gain-of-function mutants, and reduce the growth of CRPC xenografts [63]. EPI-001 was also shown to abrogate transcriptional activity of Tau1 and Tau5 domains in reporter-based assays, downregulate endogenous AR expression at the mRNA level and inhibit proliferation in multiple PCa cell lines [63]. Recently it was reported that EPI-001 can induce autophagy along with apoptosis in PCa cell lines, and that addition of autophagy inhibitors enhances antitumour effects of the drug in enzalutamide-resistant LNCaP cells [64]. Despite encouraging preclinical studies, a phase I dose-escalation trial of EPI-506 (prodrug of ralaniten, a.k.a. EPI-002) in mCRPC patients previously treated with abiraterone and/or enzalutamide (NCT02606123) was terminated due to excessive pill burden. EPI-506 was well tolerated but required high doses, which still resulted in minor and short-lived PSA declines. EPI-506 was also readily metabolized into inactive derivatives. Based on data obtained in this study, new aniten molecules EPI-7170 and EPI-7245, with 10–20-fold higher potency and improved metabolic stability were generated. EPI-7170 has shown antiproliferative activity in LNCaP and AR-V7 expressing LNCaP95 cells *in vitro*, and in LNCaP-derived xenografts in castrated mice [65]. IND-selection preclinical studies on the next generation anitens are underway. If successful, the NTD inhibitors might serve as valuable therapeutic agents with the potential to overcome antiandrogen resistance conferred by splice variants, activating mutations, or ligand-independent activation of AR.

Another actively pursued approach to abrogate AR signalling is the development of compounds that induce AR degradation and would be effective against all forms of AR. One example of these, ARV-110, an oral small molecule AR PROTAC (PROteolysis TArgeting Chimera) degrader, which recruits an E3 ubiquitin ligase to AR, followed by AR ubiquitination and proteasomal degradation. Treatment with ARV-110 has demonstrated robust reduction in AR protein levels in a high androgen environment, irrespective of AR mutation status and binding partners, and inhibited proliferation and AR-dependent gene expression in VCaP cells *in vitro* and *in vivo* [66]. The initial results were also confirmed in LNCaP cell line and patient-derived xenografts. In addition, ARV-110 effectively reduced AR-target gene expression in a long-term, castration-resistant and enzalutamide-resistant VCaP tumour model [66]. ARV-110 has completed IND-enabling studies, and a first-in-human phase I clinical trial has been approved by the FDA, and will investigate the safety and tolerability of ARV-110 in mCRPC patients that have progressed after at least two standard therapies (as reported at ASCO GU Symposium, 2019).

To achieve AR degradation, inhibition of AR chaperone proteins is also considered of significant interest. Numerous preclinical studies have demonstrated antitumour activity of heat shock protein (HSP) inhibitors in various cancers, including PCa. For example, HSP90 inhibitors (17-DMAG, onalespib, and ganetespib) have demonstrated potent downregulation of AR signalling and inhibition of tumour growth in multiple AR-dependent and AR-independent PCa models *in vitro, in vivo*, and *ex-vivo* [67]. However, this approach is not specific to AR, and the consequences of degradation of multiple HSP90 client proteins could cause additional toxicity. Furthermore, AR variants are resistant to HSP90 inhibitors, undermining their efficacy in the treatment of CRPC. Two trials (phase 1/2 NCT01685268 and phase II NCT01270880) were carried out to evaluate the efficacy of onalespib (AT13387) and ganetespib (STA-9090) in CRPC patients who were no longer responding to treatment with abiraterone and steroids, or in mCRPC patients previously treated with docetaxel, respectively. Results of the former trial were not reported, and in the latter, ganetespib showed only minimal clinical activity. Four active trials using onalespib and two with ganetespib are currently ongoing, but none is evaluating these drugs in the context of PCa. Unlike HSP90 inhibitors, drugs targeting HSP70, quercetin and VER155008, have demonstrated inhibitory effects on AR-V7 expression and proliferation in the LNCaP95 CRPC model [68]. More recently, development of HSP40/HSP70 inhibitors has identified a small chalcone molecule, C86, active against FL-AR, AR-V7, and GR *in vitro* and *in vivo* [69]. Notably, it was also active against H874Y AR mutant in 22Rv1 cells, suggesting that this drug could abrogate activity of other gain-of-function AR mutants in CRPC. Further research investigating HSP70 and HSP40 inhibitors could yield novel therapeutic agents to help manage progressive disease more effectively.

A novel approach to interfere with aberrant AR signalling was reported by Aboukameel, *et al.* [70]. They used selective inhibitors of nuclear export (SINE compounds) in preclinical studies. The central target was exportin-1 (XPO1), protein controlling nuclear localization of multiple clients, including tumour suppressor proteins, such as p53, RB, BRCA1, and p21, among others. Inhibitors of XPO1 covalently bind to XPO1 cargo-binding pocket, effectively interfering with normal nuclear export of proteins to the cytoplasm. This study revealed that increased expression of XPO1 correlated with higher AR-V7 levels. The silencing of XPO1 by RNAi or with the SINE agents selinexor and eltanexor (KTP-8602) resulted in downregulation of not only AR, AR-V7, and AR-V567es at mRNA and protein level, but also in inhibition of AR coregulators, including FOXA1, Src, and MED1, followed by abrogation of AR transcriptional activity. Eltanexor also suppressed PCa growth and induced apoptosis in AR-V7-expressing 22Rv1 and VCaP cell lines *in vitro* and *in vivo*, and enhanced the efficacy of abiraterone and enzalutamide. Application of SINE compounds to the clinic is underway. A recent phase II trial (NCT02215161) of first-generation drug, selinexor, in mCRPC patients following failure of abiraterone and/or enzalutamide was terminated due to high toxicity of the compound [71]. The second-generation agent, KTP-8602, is in clinical development. An active phase 1/2 trial will assess safety, tolerability, and efficacy of KPT-8602 in patients with several cancers, including mCRPC (NCT02649790). Preliminary results presented at the ASCO GU Symposium [72] showed that the combination of

eltanexor and abiraterone is well tolerated and demonstrates preliminary antitumour activity in patients with mCRPC: out of 21 evaluable patients, 2 exhibited partial response and 15 had stable disease. These data are encouraging, and further investigation of SINE agents is warranted.

Combinatorial Strategies

Based on the fact that the major mechanisms that lead to bypass of both first-line and subsequent AR-directed therapies result in reengagement of the AR signalling axis, it may be worthwhile to develop combinatorial means to either work in concert with ADT/antiandrogens or more fully and durably suppress AR function.

Targeting DNA Repair

The prevalence of DNA repair gene alterations has been determined to be elevated in advanced PCa as compared to primary PCa, and germline DNA repair aberrations have been reported to correspond with poor responses to hormone therapy, but there are conflicting data [73, 74]. Preclinical studies suggested that pharmacologically targeting PARP-1 (poly(ADP)-ribose polymerase-1) has the capacity to delay the onset of CRPC in a xenograft model, and potentially render CRPC tumours responsive to ADT *in vivo* [75]. This was reported as being associated with a decrease in AR residency on chromatin, and subsequently reduced AR transcriptional activity. Multiple PARP inhibitors (PARPi) (olaparib, rucaparib, niraparib, and talazoparib) are in clinical development [76] for PCa management and have been clinically tested in concert with AR-directed therapies. In the TOPARP [77] trial, patients continued an ADT regimen and received olaparib. Responses were enriched for either germline or somatic alterations that may impinge upon DNA repair competency. The ongoing PROFOUND trial (NCT02987543) is comparing olaparib to either abiraterone or enzalutamide in the context of continued ADT, in men with homologous recombination-deficient PCa, with a primary endpoint of PFS, and has the potential yield interesting results. Finally, a trial (NCT01972217) recently presented the 2018 ASCO Annual Meeting demonstrated that in mCRPC patients post-docetaxel therapy that received either olaparib or placebo in combination with abiraterone demonstrated that the combination resulted in an rPFS benefit, irrespective of homologous recombination status.

Immune Checkpoint Inhibitors

The concept of targeting immune checkpoints to activate the immune system to attack tumour cells is also currently under development in combination with AR-directed strategies. A phase II trial combining the anti-PD1 antibody pembrolizumab with enzalutamide demonstrated partial responses or stable disease [78]. The PD-L1 inhibitor atezolizumab is currently undergoing phase III investigation (IMbassador250, NCT03016312) in combination with enzalutamide versus enzalutamide alone in mCRPC patients. While the use of immune checkpoint inhibitors may currently elicit limited responses, it is anticipated that combinations with AR-directed strategies may be clinically useful in the short term.

Cell Cycle-Targeted Therapy

As stated earlier, part of the pro-tumorigenic function of AR is drive tumour cell proliferation. As such, targeting the cell cycle machinery in PCa is a rational approach. Preclinical evidence [79] suggests that combining CDK4/6 inhibition with AR-directed therapy may be a therapeutic strategy that may prove beneficial in

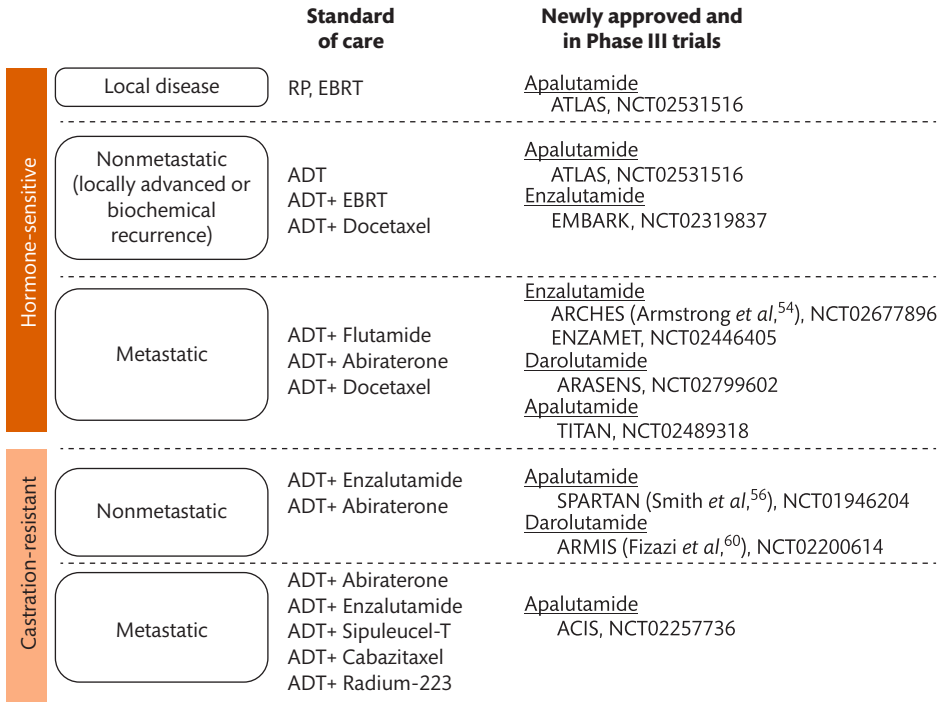


Figure 13.2.3.5 Hormonal therapy of prostate cancer—standard of care and novel approaches.

PCa. This postulate is currently under investigation in a phase Ib/II trial combining the CDK4/6 inhibitor ribociclib with docetaxel in the context of continued ADT, with the primary endpoints of maximally tolerated dose and PFS (NCT02494921). Furthermore, in a single arm phase II trial, the CDK4/6 inhibitor palbociclib is being investigated in the management of mCRPC patients continuing ADT, with the primary outcome being clinical benefit (NCT02905318).

BET Inhibition

As just discussed, AR signalling is aberrantly re-engaged in the transition to CRPC, and developing novel means to thwart castration-associated AR activity is an active category of research in the field of PCa. Bromodomain and extraterminal (BET) protein inhibitors (BETi) have been preclinically demonstrated to impact AR function, including AR-V7 activity [80], and is under clinical investigation (NCT02705469), including in combination with enzalutamide (NCT02711956).

Summary

While AR-directed therapies certainly prolong the lives of men with metastatic PCa, these strategies are not curative. ADT merely stalls disease progression, and serves as an advantageous adjuvant therapy throughout all stages of PCa (Figure 13.2.3.5). Due to tumour heterogeneity, it will be critically important to develop means to identify predictive markers of sensitivity/resistance to hormonal therapy to inform patient selection for the best life-prolonging agents. Furthermore, development of complementary strategies of targeting AR, or identifying therapeutic targets that may synergize with hormonal therapy for PCa remains a critical need.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*, 2019; **69**: 7–34.
2. Bratt O. Hereditary prostate cancer: clinical aspects. *J Urol,* 2002; **168**: 906–13.
3. Gann PH. Risk factors for prostate cancer. *Rev Urol*, 2002; **4 Suppl 5**: S3–S10.
4. Vander Griend DJ, Litvinov IV, Isaacs JT. Conversion of androgen receptor signaling from a growth suppressor in normal prostate epithelial cells to an oncogene in prostate cancer cells involves a gain of function in c-Myc regulation. *Int J Biol Sci*, 2014; **10**: 627–42.
5. Eaton CL. Aetiology and pathogenesis of benign prostatic hyperplasia. *Curr Opin Urol*, 2003; **13**: 7–10.
6. Eminaga O, Hinkelammert R, Abbas M, *et al.* High-grade prostatic intraepithelial neoplasia (HGPIN) and topographical distribution in 1,374 prostatectomy specimens: existence of HGPIN near prostate cancer. *Prostate*, 2013; **73**: 1115–22.
7. De Marzo AM, Haffner MC, Lotan TL, Yegnasubramanian S, Nelson WG. Premalignancy in prostate cancer: rethinking what we know. *Cancer Prev Res (Phila)*, 2016; **9**: 648–56.
8. Crawford ED, Schellhammer PF, McLeod DG, *et al.* Androgen receptor targeted treatments of prostate cancer: 35 years of progress with antiandrogens. *J Urol*, 2018; **200**: 956–66.
9. Mottet N, Bellmunt J, Bolla M, *et al.* EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol*, 2017; **71**: 618–29.
10. Coutinho I, Day TK, Tilley WD, Selth LA. Androgen receptor signaling in castration-resistant prostate cancer: a lesson in persistence. *Endocr Relat Cancer*, 2016; **23**: T179–97.
11. Nuhn P, De Bono JS, Fizazi K, *et al.* Update on systemic prostate cancer therapies: management of metastatic castration-resistant prostate cancer in the era of precision oncology. *Eur Urol*, 2019; **75**: 88–99.
12. Alanee S, Moore A, Nutt M, *et al.* Contemporary incidence and mortality rates of neuroendocrine prostate cancer. *Anticancer Res*, 2015; **35**: 4145–50.
13. Inamura K. Prostatic cancers: understanding their molecular pathology and the 2016 WHO classification. *Oncotarget*, 2018; **9**: 14723–37.
14. Tosco L, Briganti A, D'amico AV, *et al.* Systematic review of systemic therapies and therapeutic combinations with local treatments for high-risk localized prostate cancer. *Eur Urol*, 2019; **75**: 44–60.
15. Shafi AA, Yen AE, Weigel NL. Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacol Ther*, 2013; **140**: 223–38.
16. Tan MH, Li J, Xu HE, Melcher K, Yong EL. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin,* 2015; **36**, 3–23.
17. Heemers HV, Tindall DJ. Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev*, 2007; **28**: 778–808.
18. Cancer Genome Atlas Research, N. The molecular taxonomy of primary prostate cancer. *Cell,* 2015; **163**: 1011–25.
19. Knudsen KE, Penning TM. Partners in crime: deregulation of AR activity and androgen synthesis in prostate cancer. *Trends Endocrinol Metab*, 2010; **21**: 315–24.
20. Jernberg E, Bergh A, Wikstrom P. Clinical relevance of androgen receptor alterations in prostate cancer. *Endocr Connect*, 2017; **6**: R146–61.
21. Cato L, de Tribolet-Hardy J, Lee I, *et al.* ARv7 Represses tumor-suppressor genes in castration-resistant prostate cancer. *Cancer Cell*, 2019; **35**: 401–13.e6.
22. Robinson D, Van Allen EM, Wu YM, *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell*, 2015; **162**: 454.
23. Sharma A, Yeow WS, Ertel A, *et al.* The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *J Clin Invest*, 2010; **120**: 4478–92.
24. McNair C, Xu K, Mandigo AC, *et al.* Differential impact of RB status on E2F1 reprogramming in human cancer. *J Clin Invest*, 2018; **128**: 341–58.
25. Nava Rodrigues D, Casiraghi N, Romanel A, *et al.* RB1 heterogeneity in advanced metastatic castration-resistant prostate cancer. *Clin Cancer Res*, 2019; **25**: 687–97.
26. Stamatiades GA, Kaiser UB. Gonadotropin regulation by pulsatile GnRH: signaling and gene expression. *Mol Cell Endocrinol*, 2018; **463**: 131–41.
27. Nishiyama T. Serum testosterone levels after medical or surgical androgen deprivation: a comprehensive review of the literature. *Urol Oncol,* 2014; **32**: e17–28.
28. Meani D, Solarić M, Visapää H, Rosén RM, Janknegt R, Soče M. Practical differences between luteinizing hormone-releasing hormone agonists in prostate cancer: perspectives across the spectrum of care. *Ther Adv Urol*, 2018; **10**: 51–63.

29. Krakowsky Y, Morgentaler A. Risk of testosterone flare in the era of the saturation model: one more historical myth. *Eur Urol Focus*, 2019; **5**: 81–9.
30. Hosseini SA, Rajabi F, Akbari Sari A, Ayati M, Heidari S, Ghamary F. Degarelix for the treatment of advanced prostate cancer compared with GnRh-agonists: a systematic review and meta-analysis. *Med J Islam Repub Iran*, 2016; **30**: 317.
31. Salciccia S, Gentilucci A, Cattarino S, Sciarra A. GNRH-agonist or antagonist in the treatment of prostate cancer: a comparison based on oncological results. *Urologia*, 2016; **83**: 173–8.
32. Nakata D, Masaki T, Tanaka A, *et al.* Suppression of the hypothalamic-pituitary-gonadal axis by TAK-385 (relugolix), a novel, investigational, orally active, small molecule gonadotropin-releasing hormone (GnRH) antagonist: studies in human GnRH receptor knock-in mice. *Eur J Pharmacol*, 2014; **723**: 167–74.
33. Dai C, Heemers H, Sharifi N. Androgen signaling in prostate cancer. *Cold Spring Harb Perspect Med*, 2017; 7. DOI: 10.1101/cshperspect.a030452.
34. Alex AB, Pal SK, Agarwal N. CYP17 inhibitors in prostate cancer: latest evidence and clinical potential. *Ther Adv Med Oncol*, 2016; **8**: 267–75.
35. Patel V, Liaw B, Oh W. The role of ketoconazole in current prostate cancer care. *Nat Rev Urol*, 2018; **15**; 643–51.
36. Bryce A, Ryan CJ. Development and clinical utility of abiraterone acetate as an androgen synthesis inhibitor. *Clin Pharmacol Ther*, 2012; **91**: 101–8.
37. Roviello G, Sigala S, Sandhu S, *et al.* Role of the novel generation of androgen receptor pathway targeted agents in the management of castration-resistant prostate cancer: a literature based meta-analysis of randomized trials. *Eur J Cancer*, 2016; **61**: 111–21.
38. Purshouse K, Protheroe AS. Abiraterone acetate in combination with prednisone in the treatment of metastatic hormone-sensitive prostate cancer: clinical evidence and experience. *Ther Adv Urol*, 2019; **11**: 1756287218820804.
39. Shah H, Vaishampayan U. Therapy of advanced prostate cancer: targeting the androgen receptor axis in earlier lines of treatment. *Target Oncol,* 2018; **13**: 679–89.
40. Kita Y, Goto T, Akamatsu S, *et al.* Castration-resistant prostate cancer refractory to second-generation androgen receptor axis-targeted agents: opportunities and challenges. *Cancers (Basel)*, 2018; **10**.
41. Njar VC, Brodie AM. Discovery and development of Galeterone (TOK-001 or VN/124-1) for the treatment of all stages of prostate cancer. *J Med Chem*, 2015; **58**: 2077–87.
42. Alyamani M, Li Z, Berk M, *et al.* Steroidogenic metabolism of galeterone reveals a diversity of biochemical activities. *Cell Chem Biol*, 2017; **24**: 825–32, e826.
43. Montgomery B, Eisenberger MA, Rettig MB, *et al.* Androgen receptor modulation optimized for response (ARMOR) phase i and ii studies: galeterone for the treatment of castration-resistant prostate cancer. *Clin Cancer Res*, 2016; **22**: 1356–63.
44. Yamaoka M, Hara T, Hitaka T, *et al.* Orteronel (TAK-700), a novel non-steroidal 17,20-lyase inhibitor: effects on steroid synthesis in human and monkey adrenal cells and serum steroid levels in cynomolgus monkeys. *J Steroid Biochem Mol Biol*, 2012; **129**: 115–28.
45. Van Hook K, Huang T, Alumkal JJ. Orteronel for the treatment of prostate cancer. *Future Oncol,* 2014; **10**: 803–11.
46. Cathomas R, Crabb SJ, Mark M, *et al.* Orteronel switch maintenance therapy in metastatic castration resistant prostate cancer after first-line docetaxel: a multicenter, randomized, double-blind, placebo-controlled trial (SAKK 08/11). *Prostate*, 2016; **76**: 1519–27.
47. Toren PJ, Kim S, Pham S, *et al.* Anticancer activity of a novel selective CYP17A1 inhibitor in preclinical models of castrate-resistant prostate cancer. *Mol Cancer Ther*, 2015; **14**: 59–69.
48. Maity SN, Titus MA, Gyftaki R, *et al.* Targeting of CYP17A1 Lyase by VT-464 inhibits adrenal and intratumoral androgen biosynthesis and tumor growth of castration resistant prostate cancer. *Sci Rep*, 2016; **6**: 35354.
49. Peltola KJ, Bono P, Jones RH, *et al.* ODM-204, a novel dual inhibitor of CYP17A1 and androgen receptor: early results from phase I dose escalation in men with castration-resistant prostate cancer. *Eur Urol Focus*, 2020; **6**: 63–70.
50. Tran C, Ouk S, Clegg NJ, *et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science*, 2009; **324**: 787–90.
51. Linder S, van der Poel HG, Bergman AM, Zwart W, Prekovic S. Enzalutamide therapy for advanced prostate cancer: efficacy, resistance and beyond. *Endocr Relat Cancer*, 2018; **26**: R31–R52.
52. Bambury RM, Scher HI. Enzalutamide: development from bench to bedside. *Urol Oncol*, 2015; **33**: 280–8.
53. Hussain M, Fizazi K, Saad F, *et al.* Enzalutamide in men with nonmetastatic, castration-resistant prostate cancer. *N Engl J Med* 2018; **378**: 2465–74.
54. Armstrong AJ, Petrylak DP, Villers A, *et al.* Phase III study of androgen deprivation therapy (ADT) with enzalutamide (ENZA) or placebo (PBO) in metastatic hormone-sensitive prostate cancer (mHSPC): The ARCHES trial. In: *ASCO GU Symposium 2019*, Vol. Abstract 687. San Francisco, CA, 2019.
55. Clegg NJ, Wongvipat J, Joseph JD, *et al.* ARN-509: a novel antiandrogen for prostate cancer treatment. *Cancer Res*, 2012; **72**: 1494–503.
56. Smith MR, Saad F, Chowdhury S, *et al.* Apalutamide treatment and metastasis-free survival in prostate cancer. *N Engl J Med*, 2018; **378**: 1408–18.
57. Small EJ, Chowdhury S, Oudard S, *et al.* Updated analysis of progression-free survival with first subsequent therapy (PFS2) and safety in SPARTAN study of apalutamide (Apa) in patients (pts) with huge risk nonmetastatic castration-resistant prostate cancer (nmCRPC). In: *ASCO GU Symposium 2019*, Vol. Abstract 144. San Francisco, CA, 2019.
58. Moilanen AM, Riikonen R, Oksala R, *et al.* Discovery of ODM-201, a new-generation androgen receptor inhibitor targeting resistance mechanisms to androgen signaling-directed prostate cancer therapies. *Sci Rep*, 2015; **5**: 12007.
59. Fizazi K, Smith MR, Tombal B. Clinical development of darolutamide: a novel androgen receptor antagonist for the treatment of prostate cancer. *Clin Genitourin Cancer*, 2018; **16**: 332–40.
60. Fizazi K, Shore N, Tammela TL, *et al.* Darolutamide in nonmetastatic, castration-resistant prostate cancer. *N Engl J Med*, 2019; **380**: 1235–46.
61. Tong Y, Wu J, Yang J, *et al.* Proxalutamide (GT0918), a potent androgen receptor pathway inhibitor. In: *The 105th Annual AACR.*, Vol. Cancer Res 2014, **74**(19 Suppl): Abstract #614 (2014).
62. Vogelzang NJ, Chu DT, Gabrail NY, Nordquist LT, Stein MN. A phase I/II study of proxalutamide (GT0918), a potent androgen receptor blocker, in patients with mCRPC progressed after both hormonal therapy and chemotherapy. In: *ASCO GU Symposium 2019*, Vol. Abstract #169. San Francisco, CA, 2019.

63. Monaghan AE, McEwan IJ. A sting in the tail: the N-terminal domain of the androgen receptor as a drug target. *Asian J Androl*, 2016; **18**: 687–94.
64. Kranzbuhler B, Salemi S, Mortezavi A, Sulser T, Eberli D. Combined N-terminal androgen receptor and autophagy inhibition increases the antitumor effect in enzalutamide sensitive and enzalutamide resistant prostate cancer cells. *Prostate*, 2019; **79**: 206–14.
65. Le Moigne R, Mawji NR, Banuelos CA, *et al.* Next generation N-terminal domain androgen receptor inhibitors with improved potency and metabolic stability in castration-resistant prostate cancer models. In: *ASCO GU Symposium 2019*, Vol. Abstract 220. San Francisco, CA, 2019.
66. Neklesa T, Willard RR, Vitale N, *et al.* An oral androgen receptor PROTAC degrader for prostate cancer. In: *ASCO GU 2018*, Vol. Abstract 381 (San Francisco, 2019).
67. Jansson KH, Tucker JB, Kelly K, *et al.* High-throughput screens identify HSP90 inhibitors as potent therapeutics that target inter-related growth and survival pathways in advanced prostate cancer. *Sci Rep*, 2018; **8**: 17239.
68. Kita K, Shiota M, Tanaka M, *et al.* Heat shock protein 70 inhibitors suppress androgen receptor expression in LNCaP95 prostate cancer cells. *Cancer Sci*, 2017; **108**: 1820–7.
69. Moses MA, Kim YS, Rivera-Marquez GM, *et al.* Targeting the Hsp40/Hsp70 chaperone axis as a novel strategy to treat castration-resistant prostate cancer. *Cancer Res*, 2018; **78**: 4022–35.
70. Aboukameel A, Muqbil I, Erkan Baloglu E, *et al.* Down-regulation of AR splice variants through XPO1 suppression contributes to the inhibition of prostate cancer progression. *Oncotarget*, 2018; **9**: 35327–42.
71. Wei XX, Siegel AP, Aggarwal R, *et al.* A phase II trial of selinexor, an oral selective inhibitor of nuclear export compound, in abiraterone- and/or enzalutamide-refractory metastatic castration-resistant prostate cancer. *Oncologist*, 2018; **23**: 656–64.
72. Zhang J, Tagawa ST, Monk P, *et al.* Eltanexor (KPT-8602), a second-generation selective inhibitor of nuclear export (SINE) compound, in patients with metastatic castration-resistant prostate cancer (mCRPC). In: *ASCO GU Symposium 2019*, Vol. Abstract 197. San Francisco, CA, 2019.
73. Annala M, Struss WJ, Warner EW, *et al.* Treatment outcomes and tumor loss of heterozygosity in germline DNA repair-deficient prostate cancer. *Eur Urol*, 2017; **72**: 34–42.
74. Antonarakis ES, Lu C, Luber B, *et al.* Germline DNA-repair gene mutations and outcomes in men with metastatic castration-resistant prostate cancer receiving first-line abiraterone and enzalutamide. *Eur Urol*, 2018; **74**: 218–25.
75. Schiewer MJ, Goodwin JF, Han S, *et al.* Dual roles of PARP-1 promote cancer growth and progression. *Cancer Discov*, 2012; **2**: 1134–49.
76. Mateo J, Boysen G, Barbieri CE, *et al.* DNA repair in prostate cancer: biology and clinical implications. *Eur Urol*, 2017; **71**: 417–25.
77. Mateo J, Carreira S, Sandhu S, *et al.* DNA-repair defects and olaparib in metastatic prostate cancer. *N Engl J Med*, 2015; **373**: 1697–708.
78. Graff JN, Alumkal JJ, Drake CG, *et al.* Early evidence of anti-PD-1 activity in enzalutamide-resistant prostate cancer. *Oncotarget*, 2016; **7**: 52810–7.
79. Comstock CE, Augello MA, Goodwin JF, *et al.* Targeting cell cycle and hormone receptor pathways in cancer. *Oncogene*, 2013; **32**: 5481–91.
80. Markowski MC, De Marzo AM, Antonarakis ES. BET inhibitors in metastatic prostate cancer: therapeutic implications and rational drug combinations. *Expert Opin Investig Drugs*, 2017; **26**: 1391–7.

SECTION 14

Obesity, Dyslipidaemia and other Metabolic Disorders

14.1

Obesity

14.1.1 The Physiology of Bodyweight Regulation

Anthony P. Coll

Introduction

Understanding of the processes that govern body weight is highly relevant to clinical practice as disorders of energy homeostasis cause significant morbidity and mortality. Obesity, defined as excessive storage of energy as fat, is a serious medical and socioeconomic problem, with obese individuals at increased risk of developing a range of cardiovascular, metabolic, musculoskeletal, and malignant disorders [1]. At the other extreme, cachexia is a syndrome of negative energy balance where muscle and fat mass are progressively lost. It affects over a third of all cancer patients and is strongly associated with both reduced tolerance to anticancer therapy and reduced survival times [2].

Mammalian body weight has long been recognized and recorded to be determined by the net result of a number of highly regulated processes. Such regulatory systems have evolved though prehistory, in environments likely to favour highly adaptive systems with the capacity to maintain homeostasis and enable survival through times of food scarcity and extreme physical exertion.

The seeds of our current understanding were sown through the 20th century by many pioneering clinician scientists who noted that a steady body weight is often maintained in the face of marked variation in daily food intake and physical activity. This laid the crucial framework for the more mechanistic detail that was to come. A comprehensive account of this work lies beyond the scope of this chapter (see review by Bray [3]) but several milestone studies are worth highlighting. For example, in what has since become known as 'the Minnesota experiment', Keys *et al.* [4] limited food intake to deplete fat mass in a group of healthy, lean young men and demonstrated a reduction in metabolic rate and physical activity coupled with a marked increase in measures of hunger (**Figure 14.1.1.1**). Conversely, overfeeding of both lean and obese humans resulting in significant weight gain [5, 6] also leads to reduced appetite and increased energy expenditure, thereby opposing the maintenance of the new, heavier body weight and favouring a return to baseline weight.

Compelling evidence for significant heritability of human body weight came from Stunkard's landmark study of identical twins separated after birth and brought up in different families [7]. Adult body mass index, used as a surrogate for fat mass, did not match that of the adoptive family, but had a strong correlation with the identical twin, indicating that genetic influences on body mass index are substantial.

In parallel with these human studies, a series of key observations in rodent models gave additive insights. Kennedy's observed that rats, whether made obese through overfeeding or lean through underfeeding, tended to return to the growth trajectory of *ad libitum* fed controls after removal of the nutritional stress. This helped in formulating the hypothesis that there is a signal generated in proportion to fat mass which regulates body weight [8].

Animal data also highlighted the primacy of certain regions of the central nervous system in body weight homeostasis. For example, dramatic body weight changes evident after ablation of discrete hypothalamic regions, in particular the ventromedial hypothalamus, placed these anatomical sites as crucial 'control nodes', and laid the groundwork for a basic anatomical-behavioural map. Parabiosis experiments surgically connecting the circulations of such a lesioned animal with an intact, lean animal showed that the latter showed reduced food intake and weight gave direct evidence for circulating factor(s) present in the obese animal that influenced appetite and weight (reviewed in [9]).

A further, seminal series of parabiosis experiments by Doug Coleman using severely obese (*ob/ob*) mice and obese diabetic (*db/db*) mice determined that *ob/ob* mice lacked a then-unknown

Figure 14.1.1.1 Photograph of conscientious objectors during starvation experiment.
Reproduced with permission from the Hennepin County Library Special Collections.

circulating factor, whereas *db/db* mice produced the factor in excess but were unable to respond to it [10, 11]. Advances in molecular biology and biotechnology over the last 25 years [12] have led to identification of critical mechanisms underpinning many of these physiological observations and have indicated a biological basis for much interindividual difference in body weight.

Control of Long-Term Adipose Mass

Leptin

The discovery in 1994 that the protein product of the *obese (ob)* gene is the adipocyte-derived hormone leptin (the name derives from the Greek *leptos*, or 'thin') transformed study of the endocrinology of body weight regulation [13]. Not only did it herald a change in the view of the adipocyte from being a passive storage receptacle to being a *bona fide* endocrine cell, it also spearheaded re-exploration of how peripheral signals from metabolically relevant tissues communicate with the central nervous system. The identification of the leptin receptor [14] and the description of a mutant leptin receptor in the *db/db* mouse [15] soon followed, thereby defining key components of a regulatory circuit.

Leptin is a 167 amino acid peptide with a molecular weight of 16kD and is structurally related to the family of long-chain helical cytokines. It is produced by adipose tissue and secreted in proportion to total fat mass (Figure 14.1.1.2). Plasma levels show marked sexual dimorphism even allowing for differing fat masses, with leptin concentrations higher in females than in males for any given fat content. Leptin levels fall in response to fasting and reduced body fat stores but the mechanisms behind fasting-induced reduction in leptin, or indeed how the mass of the fat store is signalled to the adipose tissue mass and reflected in appropriate leptin production, remain to be determined. The primary role of the leptin–leptin receptor axis is to communicate a state of severe peripheral energy deficiency to the brain, thereby initiating a counter response. In rodents, falling leptin levels increase appetite, reduce sympathetic tone, suppress the thyroid and reproductive axes, and reduce energy expenditure [16].

Leptin acts via the long form of the leptin receptor (LepRb) although other forms of the receptor exist [17]. LepRb is a type I cytokine receptor of the IL-6 receptor family expressed predominantly in subpopulations of neurons in the hypothalamus, midbrain, and brainstem. Hypothalamic LepRb signalling in the lateral hypothalamic area, ventromedial, dorsomedial, and arcuate nuclei are particularly relevant for energy balance. Leptin binding to LepRb activates the associated JAK2 tyrosine kinase and subsequent tyrosine phosphorylation recruits and activates multiple downstream signalling pathways. Each of these pathways appears to subserve specific aspects of leptin's physiological actions with, for example, phosphorylation of LepRb Tyr1138 and then activation of STAT3 signalling particularly important in the regulation of body weight [18].

The Leptin-Melanocortin Pathway

A model that has held sway for some years holds that leptin's actions on two anatomically distinct subsets of arcuate neurons (POMC neurons and Agouti-related peptide (AgRP)/NPY neuron) play the dominant role in neural control of energy balance [19]. Neurons expressing proopiomelanocortin (POMC) are stimulated by leptin. The POMC pro-peptide is processed post-translationally to produce the melanocortin peptides such as α—and β-MSH, both potent suppressors of food intake (anorexigenic). The second neuronal population is suppressed by leptin and expresses AgRP and NPY. Both peptides stimulate food intake (orexigenic). From the arcuate nucleus, these two populations of neurons project to other brain areas that contain second order neurons expressing neuropeptides involved in energy homeostasis. One site of particular relevance is the paraventricular nucleus which is enriched with melanocortin 4 receptors (MC4R), the cognate receptor for both melanocortin

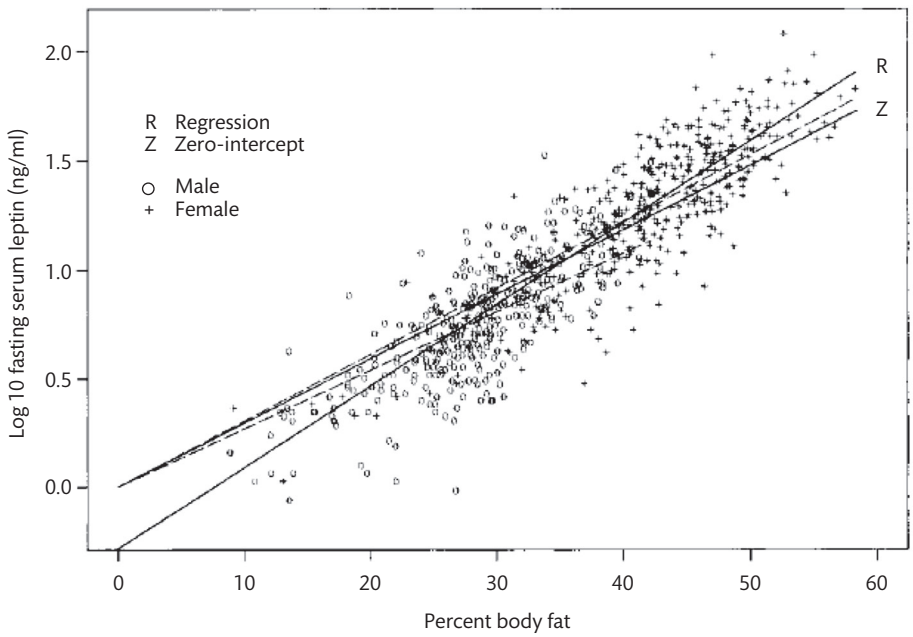


Figure 14.1.1.2 Fasting serum leptin levels plotted on a log scale against percentage body fat for women (+) and men (O). The dashed lines show the mean of the slopes for women (upper line) and men (lower line). The solid lines show two regression lines, one (Z) constrained to go through the origin (0,0), the second (R) where the intercept is estimated from the data.

peptides (agonist) and AgRP (antagonist). As a consequence of this circuitry, when there are suppressed levels of leptin, as in negative energy balance, an increase in orexigenic signals coupled with a decrease in anorexigenic signals strongly favours an increase in appetite and food intake to replenish diminished stores.

The central tenets of this model have held true over several decades of intense scrutiny; leptin's actions on energy balance are primarily, if not wholly, driven by its action within the central nervous system (CNS), and both POMC neurons and central MC4R continue to be active targets for pharmacological manipulation to treat obesity. However, ever more complexity is being revealed within these neural circuits and emerging data indicate that it may be AgRP neurons, rather than POMC neurons, that are the primary site of action for leptin's effects on energy balance [20].

Leptin and Melanocortin Deficiency in Humans

In the 1990s, molecular medicine and the study of patients with severe obesity yielded powerful insights into the fundamental physiological processes that regulate body weight. The first reports of a single gene disorder causing obesity came from detailed biochemical and genetic analysis of a woman with a complex polyendocrine syndrome and severe obesity from childhood, who was discovered to have mutations in both alleles encoding prohormone convertase 1(PC1), rendering her unable to process peptide prohormones such as pro-insulin and POMC [21]. Soon after, two children with severe, early-onset obesity and marked hyperphagia were found to have homozygous mutations in the leptin gene that rendered them totally leptin deficient [22]. Studies using the classic endocrine paradigm of phenotypic analysis first in the absence of hormone and then after replacement therapy defined the clinical phenotype of leptin deficiency and thereby generated key insights into leptin biology in humans. Treatment of congenital leptin deficiency with recombinant leptin significantly reduced hyperphagia, lead to a sustained reduction in fat mass while preserving lean mass weight and had a permissive role in the onset of puberty [23].

Leptin receptor deficiency has also been characterized in humans. Although the clinical phenotype was less severe than in congenital leptin deficiency, affected subjects were hyperphagic, had severe obesity, and also had delayed puberty due to hypogonadotropic hypogonadism. Interestingly, serum leptin levels were within the range predicted by the elevated fat mass and were similar to those seen in equally obese subjects with a normal leptin-receptor gene [24]. However, just as in rodents, the message was clear; reduced activity through the leptin–leptin receptor axis is a powerful signal to the brain to enact responses to reverse what is perceived to be a substantive diminution of energy stores.

Many more reports of humans with obesity due to defects in the neuronal pathways that lie downstream of leptin signalling have since followed [25], including POMC deficiency [26] and MC4R mutations [27, 28] (see Chapter 14.1.6, 'Assessment of Obesity in Children'). These monogenic obesity syndromes have largely validated the central paradigm of a neural control of appetite.

Leptin Therapeutics

Leptin has been successfully used as a therapy in a number of clinical scenarios. The profound effects in congenital leptin deficiency have been well documented [25]. As discussed in Chapter 15.10.2, 'Lipodystrophies and Severe Insulin Resistance', leptin therapy in patients with generalized lipodystrophy who lack leptin due to their deficiency of adipose tissue can also have dramatic

effects on the metabolic sequelae seen in this condition such as hypertriglyceridaemia and poorly controlled diabetes [29]. There have also been promising results for leptin treatment in weight related reproductive disorders such as hypothalamic amenorrhoea. Young women with low leptin levels and low adiposity can have their reproductive function restored with leptin therapy [30].

Leptin is also likely to be a useful adjunctive therapy in preventing weight gain after weight loss. As weight loss is characterized by a relatively leptin deficient state with a range of neurohormonal processes activated to restore the weight that has been lost, leptin supplementation is likely to be helpful in maintaining the weight reduced state [31].

Control of Short-Term Feeding and Satiety

Enteroendocrine Cells

The enteroendocrine cells (EECs) of the gastrointestinal tract play an important role in balancing food intake and energy balance in the short term [32]. Spread throughout the length of the intestine, EECs form a large and interconnected network of functional endocrine tissue that enable direct sensing of ingested nutrients to be coupled with hormonal release and communication with surrounding enteric neurons. Originally classified into distinct 'alphabetical' types based on location, morphology and assumed hormonal production (e.g. cholecystokinin (CCK) from I-cells) the ability to isolate and analyse single cells has revealed a much higher degree of complexity and overlap with many cells type able to produce and cosecrete a range of different peptide hormones.

More than 30 peptide hormones have been identified in the gastrointestinal tract and a full review of all their actions lies beyond the scope of this chapter (see further reading [33, 34]). However, a brief review of the physiology of L cells and the peptides they produce exemplifies how signals from EECs can both have important roles on local gut function and impact upon whole-body physiology [35] (**Figure 14.1.1.3**).

L cells are found throughout the gut but are particularly located in the distal small intestine and the colon. The primary peptides produced are those derived from post-translational processing of the proglucagon precursor molecule (GLP1, GLP2, and oxyntomodulin) and PYY, although most L cells seemingly have plurihormonal capacity [35].

GLP1 (30 amino acids) is a metabolically active gut hormone that has been extensively characterized [33, 34]. It has long been recognized to be a potent incretin hormone, augmenting glucose-stimulated insulin secretion after a meal. This effect on insulin secretion is maintained in many patients with type 2 diabetes mellitus and has led to widespread clinical use of GLP-1 analogues. GLP1 receptors

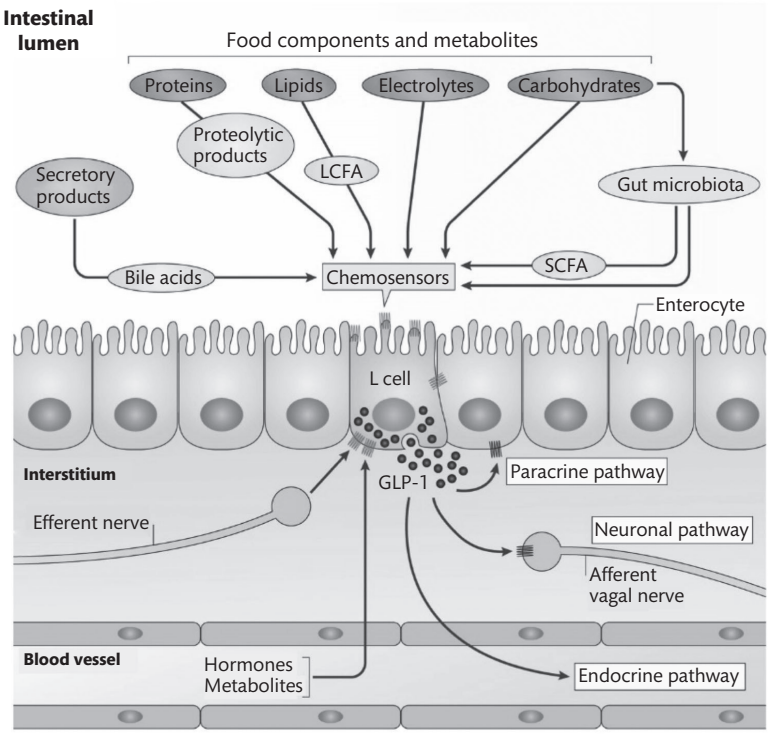


Figure 14.1.1.3 The sensory and secretory function of the L cell. Release of glucagon-like peptide 1 (GLP-1) from L cells is regulated by nutritional, hormonal, and neural signals. Food components and metabolites at the luminal side of the L cell are directly sensed by various G protein-coupled receptors that function as chemosensors and trigger exocytosis of GLP-1-containing granules at the basolateral side of the cell. GLP-1 can act through endocrine, paracrine and neuronal pathways to regulate physiological responses in local and/or remote tissues and cell types. These effects are consistent with the widespread and abundant expression of the GLP-1 receptor. LCFA, long-chain fatty acid; SCFA, short-chain fatty acid.

are found in many regions that regulate feeding and, in addition to its incretin effect, administration of GLP1 can reduce food intake in a dose dependent manner. Intriguingly, GLP1 is also expressed within the nucleus tractus solitarii (NTS) of the brainstem, a relay centre for gut derived signals controlling food intake, and it is likely centrally produced GLP1 also plays a role in energy balance [36].

Glucagon-like peptide-2 (GLP2) is another small (33 amino acid) proglucagon-derived peptide whose plasma levels rise rapidly after eating [35]. It has an important facilitatory role in maintaining gut integrity and performance, being able not only to induce proliferative and pathways in the bowel but also to increase mesenteric blood flow and activate proabsorptive pathways. Recombinant analogues of GLP2 have been successfully used in patients with intestinal failure [37].

Oxyntomodulin, a third peptide product of the proglucagon gene, is composed of the 29 amino acids of glucagon with an 8 amino acid carboxy terminal extension [34]. It is also released from the small intestine after a meal, and is active on GLP1R and glucagon receptors. It reduces gastric motility and improves glucose-dependent insulin secretion. Pharmacological administration reduces energy intake and weight, actions thought to be mediated both indirectly via the vagus nerve and directly in the CNS.

Peptide tyrosine tyrosine (PYY) levels in humans are lowest in the fasting state and rise after food ingestion. PYY 3–36, produced by DPP4-mediated cleavage of the larger peptide PYY 1–36, is the main circulating form and has an anorectic effect via central Y2 receptors within the hypothalamus and brainstem [43].

Therapeutic interest in gut hormones has been further stimulated by the remarkable effects on severe obesity of bariatric surgery [38], discussed in detail in Chapter 14.1.5, 'Metabolic Surgery'. This term encompasses a range of surgical procedures whose purpose is to manipulate the stomach and small intestine to induce weight loss. Common procedures include removal of a portion of the stomach (e.g. sleeve gastrectomy) or the creation of a small stomach pouch with the jejunum pulled up onto this pouch, thereby leaving the stomach and duodenum bypassed (Roux-en-Y procedure). Although there is little doubt that such surgery significantly reduces calorific intake and places the patient in negative energy balance, previous explanations that this was due wholly to restriction of stomach size and/or malabsorption from the modified intestine have more recently undergone a more nuanced review, with effects of these surgeries thought, at least in part, to be mediated through postoperative changes in gut hormones. Clinical trials recapitulating this state through parenteral administration of combinations of peptides to patients with obesity are underway [39].

Gut Microbiota and Body Composition

The human gut contains trillions of microbes existing in symbiosis with their human host [40]. The composition of the gut microbiome varies significantly between and within individuals and can be changed markedly by diet and medication. There are a large number of primary animal studies that suggest the microbiome can produce many bioactive substances potentially affecting energy balance [40]. The significance and true therapeutic potential in humans remain to be fully determined. Many studies to date have been descriptive, lacking the ability to determine if associated changes in microbiota and weight reflect causality. For example, one recent study investigating changes in gut microbiota in obese men after 7 days of amoxicillin or vancomycin found that despite considerably altered microbial composition there was no clinically relevant impact on metabolic health in the subject studies [41].

Methodologies are currently rapidly evolving to permit more rigorous assessment of causality in studies of the relationship between human microbiota and body weight, and it may well be that more focused manipulation of gut flora function will be of therapeutic benefit. One plausible mechanism for such an effect relates to enteroendocrine cells. A number of metabolites generated by colonic gut microbiota through their action on luminal foodstuffs have been implicated as ligands to activate L cells. These include short-chain fatty acids and indole, a product of tryptophan metabolism and a study in humans has shown that an increase in the short-chain fatty acid (SCFA) propionate in the colon can increase PYY and GLP-1 and reduce energy intake [42].

BAT, Thermogenesis, and Energy Expenditure

Mammals have an absolute requirement to maintain their core body temperature. Part of the response to a cold challenge that threatens body temperature is increased sympathetic drive to stimulate brown adipose tissue (BAT) in a process called non-shivering thermogenesis. BAT is composed of specialized adipocytes that break down energy substrate to generate heat in a futile metabolic cycle without generation of adenosine triphosphate (ATP) [34]. This involves expression of uncoupling protein 1 (UCP1), a mitochondrial protein that short-circuits the process of mitochondrial ATP generation. Brown fat has been studied for decades in rodents but its relevance to human metabolism had long been contested and debated [43].

The concept of a highly reactive energy sink through which energy in excess of current requirement (that would otherwise be laid down in fat stores) could be disposed of is not new. At the turn of the twentieth century, Neuman used the word 'luxuskonsumption' to describe the self-observed phenomenon that his weight remained unchanged over a period in which he had noted significant change in dietary intake, including what he considered excess consumption [44].

Work by Rothwell and Stock reporting that rats fed a 'cafeteria diet' appeared to gain less weight than expected due to burning of unaccounted calories in BAT refocused attention on the significance of 'diet-induced thermogenesis' (DIT) [45]. Nevertheless, BAT in humans was considered to be irrelevant outside the neonatal period, and the potential biological purpose of DIT continues to be debated [45].

More recent identification in healthy children and adults of functional BAT in the supraclavicular, neck, perirenal and intercostal regions (Figure 14.1.1.4) has sparked renewed interest in the expenditure side of the energy equation. In particular the notion that targeting pathways innervating and controlling BAT could be an effective and safe way to increase energy expenditure and combat obesity has garnered much attention.

One concept under intense preclinical scrutiny is that adipocytes found in classical white depots that have the ability to express UCP1 to become heat—generating, energy consuming cells may be upregulated as an energy sink [46]. These cells have been viewed as a potential third class of adipocyte sitting between brown and white, thus being named 'beige' or 'brite' ('brown-in-white'). Directing

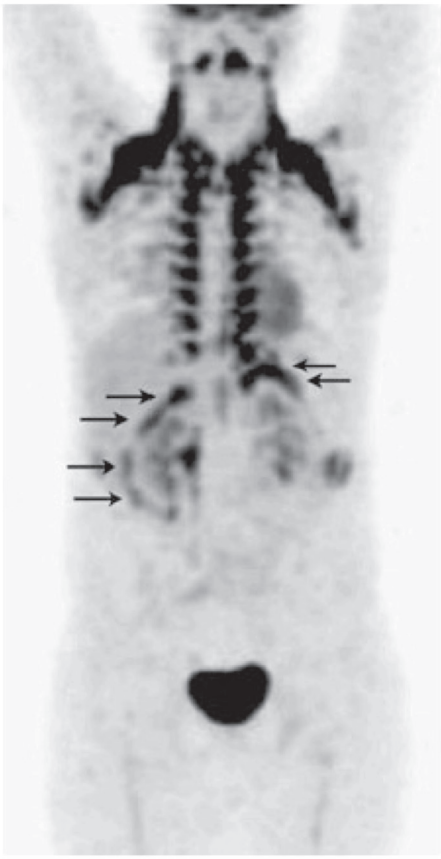


Figure 14.1.1.4 A 13-year-old boy with BAT FDG uptake in the neck, supraclavicular-axillary, paravertebral-intercostal, and mediastinal regions. Arrows indicate uptake around the lateral edge of the kidneys.
Reproduced with permission from Hong, T.S., Shammas, A., Charron, M. *et al. Pediatr Radiol* (2011) 41: 759. https://doi.org/10.1007/s00247-010-1925-y. Copyright © 2010, Springer-Verlag.

such cells into a more energy-consuming role, primarily through augmented sympathetic and adrenergic stimulation, is given the moniker of 'browning'.

Thyroid hormones regulate basal metabolic rate and dysregulation of the thyroid axis leads to marked alterations in energy balance [47]. Increased appetite is also common in hyperthyroidism but the majority of patients with hyperthyroidism exhibit weight loss, indicating that increased caloric intake fails to match the increased energy expenditure. Hypothyroidism, by contrast, is associated with decreased metabolic rate, often causing weight gain despite reduced food intake. Many of these effects have been attributed to direct actions of thyroid hormone on metabolically active tissues such as the liver, fat and muscle but it is now clear that thyroid hormones regulate energy balance and metabolism by acting at the central level, with distinct hypothalamic pathways modulating food intake and regulating thermogenesis in BAT [47].

Outside of thyroid hormone disorders, there are few data to support the concept of a reduced basal metabolic rate (the apocryphal 'slow metabolism') as a cause of severe obesity. However rare variants in kinase suppressor of Ras 2 (KSR2), an intracellular scaffolding protein involved in multiple signalling pathways, seen in individuals with severe, early-onset obesity have been reported to be associated not only with hyperphagia but also with a lower basal metabolic rate compared with subjects with other genetic forms of obesity [48].

Control of Skeletal Muscle Mass

Skeletal muscle is a highly plastic tissue which, throughout life, adapts to external stimuli and internal endocrine signals [46]. Muscle mass and strength peak in early adulthood in both males and females but begin to decline in midlife and fall away more rapidly from the fifth decade onwards [52]. The phenomena of hypertrophy in response to mechanical loading and atrophy on inactivity are well recognized. Muscle mass is also lost directly as a result of a many systemic inflammatory, malignant, and infectious disorders.

A key regulator of muscle mass is myostatin, a member of the TGFβ superfamily [50]. Myostatin strongly inhibits muscle mass development and loss of functional myostatin leads to hypermuscularity in genetically engineered mice or in humans [51] or in cattle such Belgian Blue that are reared for their 'double muscling' [52] (**Figure 14.1.1.5**).

Myostatin is secreted as an inactive peptide that undergoes post-secretion cleavage, forms a C-terminal dimer that then binds to activating type IIB receptor (ActRIIB), which in turn phosphorylates and activates Smad to affect changes in gene transcription. Active myostatin is also negatively regulated by follistatin, which binds directly with the mature dimer and inhibits its ability to bind to receptors. Data from transgenic mouse models indicate that high levels of follistatin can lead to a significant increase in muscle mass, comparable to those seen in myostatin knockout mice [53], and molecules that block signalling through this ActRIIB pathway are currently under active development as agents to treat muscular dystrophies and other disorders involving the loss of functional muscle mass [54].

Fat Cells—How Much and Where?

Adipose tissue is exquisitely adapted to store large amount of energy; however, it is also highly heterogeneous, with fat cells in distinct depots of the body having site-specific roles [55]. Evidence to date, however, indicate that genetically normal, functional fat cells do not harbour a mechanism which is a primary driver of whole-body obesity. What is increasingly clear from human data is that body fat content is primarily determined by neuronal pathways affecting appetitive drive and energy consumption [56, 57]. However, the related but distinct questions of *where* the fat stored and of *how much* can be safely accommodated there are more likely driven by processes autonomous to the peripheral tissues. Indeed, in a recent large genome-wide association study of the genetic architecture of fat distribution, genetic loci implicated were enriched for genes expressed in adipose tissue and for putative regulatory elements in adipocytes [58]. What determines an individual's 'fat mass storage capacity' remains to be fully defined. This has great clinical relevance in societies where overnutrition is commonplace, as excess energy that can no longer be safely stored in adipose depots leads to ectopic deposition of lipids in other tissues such as liver and muscle and the development of adverse metabolic sequelae. The processes

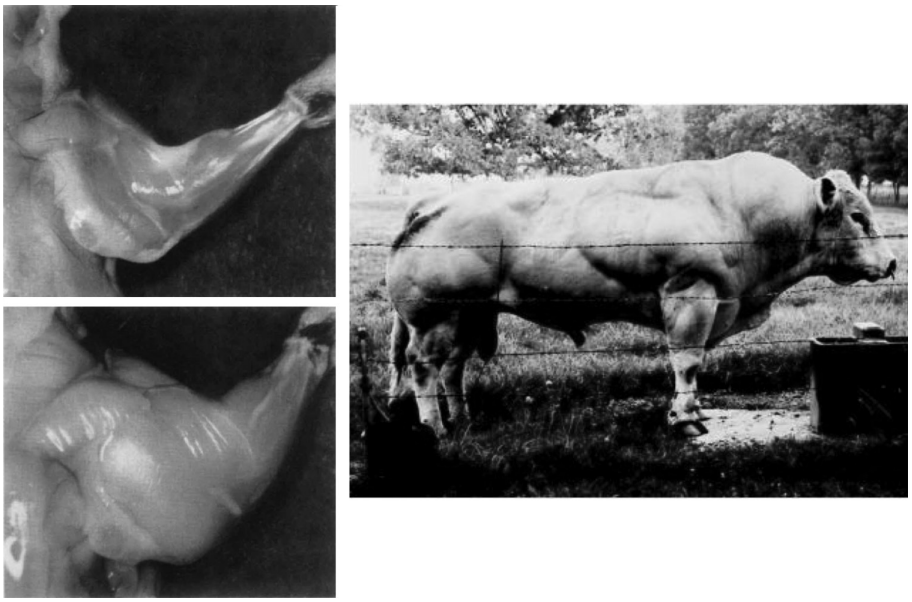


Figure 14.1.1.5 *Left panels*: Increased skeletal muscle mass in myostatin null upper arm (bottom) compared to wild type littermate (top). *Right*: A Belgian Blue bull showing the double muscling phenotype.

Left panels reproduced with permission from McPherron AC, Lawler AM, Lee SJ. Regulation of skeletal muscle mass in mice by a new TGF-p superfamily member. *Nature*. 1997 May 1;387(6628):83–90. Copyright © 1997, Springer Nature. Right panel reproduced with permission from McPherron AC, Lee SJ. Double muscling in cattle due to mutations in the myostatin gene. *Proc Natl Acad Sci USA*. 1997 Nov 11; 94(23): 12457–61.Copyright (1997) National Academy of Sciences, U.S.A.

underpinning these adverse consequences have elegantly been explored by O'Rahilly [59].

Cachexia

Cachexia is characterized by involuntary loss of body weight and muscle mass. It is distinct from starvation in that it is not readily reversible by nutritional support alone. It is a major feature of a range of malignant diseases and chronic, non-malignant disorders, with the depletion of skeletal muscle contributing significantly to functional impairment, therapy-related toxicity, and increased mortality [2]. Detailed metabolic studies of cachexia in humans are challenging due to the heavy disease burden borne by many of these patients and much of our current mechanistic understanding of this condition has derived from analysis of animal model systems.

Although the various contributors to the negative energy balance of cachexia likely varies among diseases and individuals, reduction in food intake is an important and often primary driver. A host of different pro-cachectic factors have been implicated including interleukin—6 (IL-6), tumour necrosis factor (TNF) and interferon gamma (IFNg). Working through the CNS these factors are thought to driver further neurohormonal changes, such as activation of the adrenal axis and the sympathetic nervous system, to favour catabolic changes that, in the face of coexisting poor food intake, can be considered aberrant and highly deleterious [60]. However, it is a reflection of the limited understanding of the key mechanisms involved that no effective therapeutic options based on these mechanisms are currently available.

One circulating peptide that has long been associated with a diverse range of cachexia-associated conditions is growth and differentiation factor 15 (GDF15). In humans, serum levels of GDF 15 are a robust biomarker of all-cause mortality and overexpression of GDF15 in animal models robustly suppresses food intake. Recent identification of the receptor for this peptide, expressed in brainstem loci well known to play important roles in energy balance, has reignited interest in this pathway as being potentially amenable to tractable therapies for both cachexia and obesity [61].

Further Reading—Integrating Physiology and Environment

In the face of ever rising obesity, there has been much thoughtful reflection as to why physiological systems that have served us well for millennia now appear to be overwhelmed, leading to a huge increase in obesity. The physiology of the overfed state and the evidence to date for the existence of a regulatory system that could defend against excessive fat mass expansion is expertly review by Ravussin and colleagues [62]. At the population level Speakman's article on the evolutionary genetics of human obesity makes a compelling case against the 'thrifty genotype' hypothesis (the evolutionary advantage of ingesting excess calories in times of plenty to enable survival in times of famine). He argues that genes predisposing to obesity are neutral and have been drifting over evolutionary time, with the more recent food abundance and release from predation combining to explain current trends in body weight [63]. Finally, readers concerned to understand why health-harming behaviours such as continued overeating in the presence of metabolic disease are so hard to change are directed to an excellent article by Marteau and colleagues on the automatic processing associated with human disease [64].

REFERENCES

1. Kopelman PG. Obesity as a medical problem. *Nature*, 2000; **404**: 635–43.
2. Baracos VE, Martin L, Korc M, Guttridge DC, Fearon KCH. Cancer-associated cachexia. *Nat Rev Dis Primers*, 2018; **4**: 17105.
3. Bray GA. Obesity: historical development of scientific and cultural ideas. *Int J Obes*, 1990; **14**: 909–26.
4. Keys AB, Brozek J, Henschel A. *The Biology of Human Starvation*. Minneapolis: University of Minnesota Press, 1950.
5. Sims EA, Danforth E Jr, Horton ES, Bray GA, Glennon JA, Salans LB. Endocrine and metabolic effects of experimental obesity in man. *Recent Prog Horm Res,* 1973; **29**: 457–96.
6. Leibel RL, Rosenbaum M, Hirsch J. Changes in energy expenditure resulting from altered body weight. *N Engl J Med*, 1995; **332**: 621–8.
7. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE. The body-mass index of twins who have been reared apart. *N Engl J Med*, 1990; **322**: 1483–7.
8. Kennedy GC. The hypothalamus and obesity. *Proc R Soc Med*, 1966; **59**: 1276–7.
9. King BM. The rise, fall, and resurrection of the ventromedial hypothalamus in the regulation of feeding behavior and body weight. *Physiol Behav*, 2006; **87**: 221–44.
10. Coleman DL, Hummel KP. Effects of parabiosis of normal with genetically diabetic mice. *Am J Physiol*, 1969; **217**: 1298–304.
11. Coleman DL. Effects of parabiosis of obese with diabetes and normal mice. *Diabetologia*, 1973; **9**: 294–8.
12. Krashes MJ. Untangling appetite circuits with optogenetics and chemogenetics. In: Harris RBS, ed. *Appetite and Food Intake: Central Control*, 2nd edn. Boca Raton, FL: CRC Press/ Taylor & Francis; 2017: 91–116.
13. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature*, 1994; **372**: 425–32.
14. Tartaglia LA, Dembski M, Weng X, *et al*. Identification and expression cloning of a leptin receptor, OB-R. *Cell*, 1995; **83**: 1263–271.
15. Chen H, Charlat O, Tartaglia LA, *et al*. Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell*, 1996; **84**: 491–95.
16. Ahima RS, Prabakaran D, Mantzoros C, *et al*. Role of leptin in the neuroendocrine response to fasting. *Nature*, 1996; **382**: 250–2.
17. Myers MG, Cowley MA, Munzberg H. Mechanisms of leptin action and leptin resistance. *Annu Rev Physiol*, 2008; **70**: 537–56.
18. Allison MB, Myers MG, Jr. 20 years of leptin: connecting leptin signaling to biological function. *J Endocrinol,* 2014; **223**: T25–35.
19. Cone RD. Studies on the physiological functions of the melanocortin system. *Endocr Rev*, 2006; **27**: 736–49.
20. Xu J, Bartolome CL, Low CS, *et al*. Genetic identification of leptin neural circuits in energy and glucose homeostases. *Nature*, 2018; **556**: 505–9.
21. Jackson RS, Creemers JW, Ohagi S, *et al*. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*, 1997; **16**: 303–6.
22. Montague CT, Farooqi IS, Whitehead JP, *et al*. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, 1997; **387**: 903–8.
23. Farooqi IS, Jebb SA, Langmack G, *et al*. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med*, 1999; **341**: 879–84.
24. Farooqi IS, Wangensteen T, Collins S, *et al*. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*, 2007; **356**: 237–47.
25. Farooqi IS, O'Rahilly S. 20 years of leptin: human disorders of leptin action. *J Endocrinol,* 2014; **223**: T63–70.
26. Krude H, Gruters A. Implications of proopiomelanocortin (POMC) mutations in humans: the POMC deficiency syndrome. *Trends Endocrinol Metab,* 2000; **11**: 15–22.
27. Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, O'Rahilly S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med*, 2003; **348**: 1085–95.
28. Alharbi KK, Spanakis E, Tan K, *et al*. Prevalence and functionality of paucimorphic and private MC4R mutations in a large, unselected European British population, scanned by meltMADGE. *Hum Mutat*, 2007; **28**: 294–302.
29. Oral EA, Chan JL. Rationale for leptin-replacement therapy for severe lipodystrophy. *Endocr Pract*, 2010; **16**: 324–33.
30. Welt CK, Chan JL, Bullen J, *et al*. Recombinant human leptin in women with hypothalamic amenorrhea. *N Engl J Med*, 2004; **351**: 987–97.
31. Rosenbaum M, Goldsmith R, Bloomfield D, *et al*. Low-dose leptin reverses skeletal muscle, autonomic, and neuroendocrine adaptations to maintenance of reduced weight. *J Clin Invest*, 2005; **115**: 3579–86.
32. Drucker DJ. Evolving concepts and translational relevance of enteroendocrine cell biology. *J Clin Endocrinol Metab*, 2016; **101**: 778–86.
33. Rehfeld JF. The new biology of gastrointestinal hormones. *Physiol Rev*, 1998; **78**: 1087–108.
34. Monteiro MP, Batterham RL. The importance of the gastrointestinal tract in controlling food intake and regulating energy balance. *Gastroenterology*, 2017; **152**: 1707–17.e1702.
35. Gribble FM, Reimann F. Signalling in the gut endocrine axis. *Physiol Behav,* 2017; **176**: 183–8.
36. Daniels D, Mietlicki-Baase EG. 36. Glucagon-Like Peptide 1 in the Brain: Where Is It Coming From, Where Is It Going? *Diabetes*, 2019; **68**: 15–17.
37. Jeppesen PB. Gut hormones in the treatment of short-bowel syndrome and intestinal failure. *Curr Opin Endocrinol Diabetes Obes*, 2015; **22**: 14–20.
38. Abdeen G, le Roux CW. Mechanism underlying the weight loss and complications of Roux-en-Y gastric bypass. Review. *Obes Surg,* 2016; **26**: 410–21.
39. Scott RV, Tan TM, Bloom SR. Can Bayliss and Starling gut hormones cure a worldwide pandemic? *J Physiol*, 2014; **592**: 5153–167.
40. Rosenbaum M, Knight R, Leibel RL. The gut microbiota in human energy homeostasis and obesity. *Trends Endocrinol Metab,* 2015; **26**: 493–501.
41. Reijnders D, Goossens GH, Hermes GD, *et al*. Effects of gut microbiota manipulation by antibiotics on host metabolism in obese humans: a randomized double-blind placebo-controlled trial. *Cell Metab*, 2016; **24**: 341.
42. Chambers ES, Viardot A, Psichas A, *et al*. Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut*, 2015; **64**: 1744–54.
43. Cannon B, Nedergaard J. Brown adipose tissue: function and physiological significance. *Physiol Rev*, 2004; **84**: 277–359.
44. Bray GA. Luxuskonsumption - Myth or Reality? *Obes Res*, 1995; **3**: 491–5.
45. Kozak LP. Brown fat and the myth of diet-induced thermogenesis. *Cell Metab*, 2010; **11**: 263–67.
46. Montanari T, Poscic N, Colitti M. Factors involved in white-to-brown adipose tissue conversion and in thermogenesis: a review. *Obes Rev*, 2017; **18:** 495–13.
47. Lopez M, Alvarez CV, Nogueiras R, Dieguez C. Energy balance regulation by thyroid hormones at central level. *Trends Mol Med,* 2013; **19**: 418–27.

48. Pearce LR, Atanassova N, Banton MC, *et al*. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell*, 2013; **155:** 765–77.
49. Cohen S, Nathan JA, Goldberg AL. Muscle wasting in disease: molecular mechanisms and promising therapies. *Nat Rev Drug Discov*, 2015; **14**: 58–74.
50. McPherron AC, Lawler AM, Lee SJ. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*, 1997; **387**: 83–90.
51. Schuelke M, Wagner KR, Stolz LE, *et al*. Myostatin mutation associated with gross muscle hypertrophy in a child. *N Engl J Med*, 2004; **350**: 2682–8.
52. Dickman S. Gene mutation provides more meat on the hoof. *Science*, 1997; **277**: 1922–3.
53. Lee SJ. Quadrupling muscle mass in mice by targeting TGF-beta signaling pathways. *PLoS One*, 2007; **2**: e789.
54. Bondulich MK, Jolinon N, Osborne GF, *et al*. Myostatin inhibition prevents skeletal muscle pathophysiology in Huntington's disease mice. *Sci Rep*, 2017; **7**: 14275.
55. Zwick RK, Guerrero-Juarez CF, Horsley V, Plikus MV. Anatomical, physiological, and functional diversity of adipose tissue. *Cell Metab,* 2018; **27**: 68–83.
56. Locke AE, Kahali B, Berndt SI, *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 2015; **518**: 197–206.
57. Turcot V, Lu Y, Highland HM, *et al*. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet*, 2018; 50: 26–41.
58. Shungin D, Winkler TW, Croteau-Chonka DC, *et al*. New genetic loci link adipose and insulin biology to body fat distribution. *Nature,* 2015; **518**: 187–96.
59. O'Rahilly S. Harveian Oration 2016: some observations on the causes and consequences of obesity. *Clin Med (Lond),* 2016; **16**: 551–64.
60. Flint TR, Fearon DT, Janowitz T. Connecting the metabolic and immune responses to cancer. *Trends Mol Med,* 2017; **23**: 451–64.
61. O'Rahilly S. GDF15-from biomarker to allostatic hormone. *Cell Metab,* 2017; **26**: 807–8.
62. Ravussin Y, Leibel RL, Ferrante AW, Jr. A missing link in body weight homeostasis: the catabolic signal of the overfed state. *Cell Metab,* 2014; **20**: 565–72.
63. Speakman JR. The evolution of body fatness: trading off disease and predation risk. *J Exp Biol*, 2018; **221**(Pt Suppl 1): pii: jeb167254.
64. Marteau TM, Hollands GJ, Fletcher PC. Changing human behavior to prevent disease: the importance of targeting automatic processes. *Science*, 2012; **337**: 1492–5.

14.1.2 Obesity as a Public Health Problem

Adrian Bauman

Public Health Approaches to Obesity

Obesity has been recognized as a disease by the American Heart Association, posing both challenges and opportunities [1]. This formalizes a definition (e.g. body mass index (BMI) ≥30), empowers health services to plan approaches to obesity management, and allows clinical therapy to be institutionalized. However, it is also a 'public health' or population-wide problem. The public health problem is illustrated by the rapidity of the increase in the prevalence of obesity, such that it is described as an epidemic, or more accurately as a pandemic (because it is globally pervasive). A public health approach posits that obesity has resulted from changes to human culture and technology that have led in turn to decreased total physical activity (energy expended) and increased food consumed (energy intake) especially in industrialized and rapidly developing nations.

Figure 14.1.2.1 depicts a public health approach to obesity, showing the stages in its classification as a 'population or public health problem' (left hand box), and the need for parallel clinical and population strategies to address the problem. Public health (primary prevention) aims to avoid the long-term morbidity, mortality, and healthcare costs attributable to obesity.

A Brief History of Public Health Thinking About Obesity

Bray (1995) [2] reviewed obesity-related medical writings from antiquity to the present, especially over the past three centuries; he noted clinical tomes with titles such as the 1727 *'Discourse Concerning the Causes and Effects of Corpulency Together with the Method for its Prevention and Cure'* but concluded that the upswing in obesity writings occurred around 1940, when it became more frequent in the medical literature.

The earliest papers that described *obesity as a public health problem* date to the 1950s, well before the oft-reported increases in adult obesity since the 1980s. For example, *Life Magazine* (1954) featured a special issue on obesity, describing it as a 'health crisis' affecting 20–30% of adult Americans over 'the desirable weight', and a 50% increased risk of mortality [3, 4]; these were curiously accurate estimates, considering the lack of any public health surveillance or epidemiological studies available at that time. The solutions proposed included reduced calorie intake, and prophetically, fewer meals consumed outside of the home [3]. By 1968, Huenemann reported that obesity was a problem among adolescents too, with obesity rates between 10% and 15% (although variously defined). She also described childhood obesity as persisting into adult life, and being associated with cardio-metabolic risk [5].

Obesity continued to be described as a public health issue during the 1970s and as 'the second greatest public health problem after tobacco' [6], antedating the subsequent 'obesity epidemic'. Epidemiological risks were characterized in the 1980s, also predating the recognition by public health agencies that obesity was

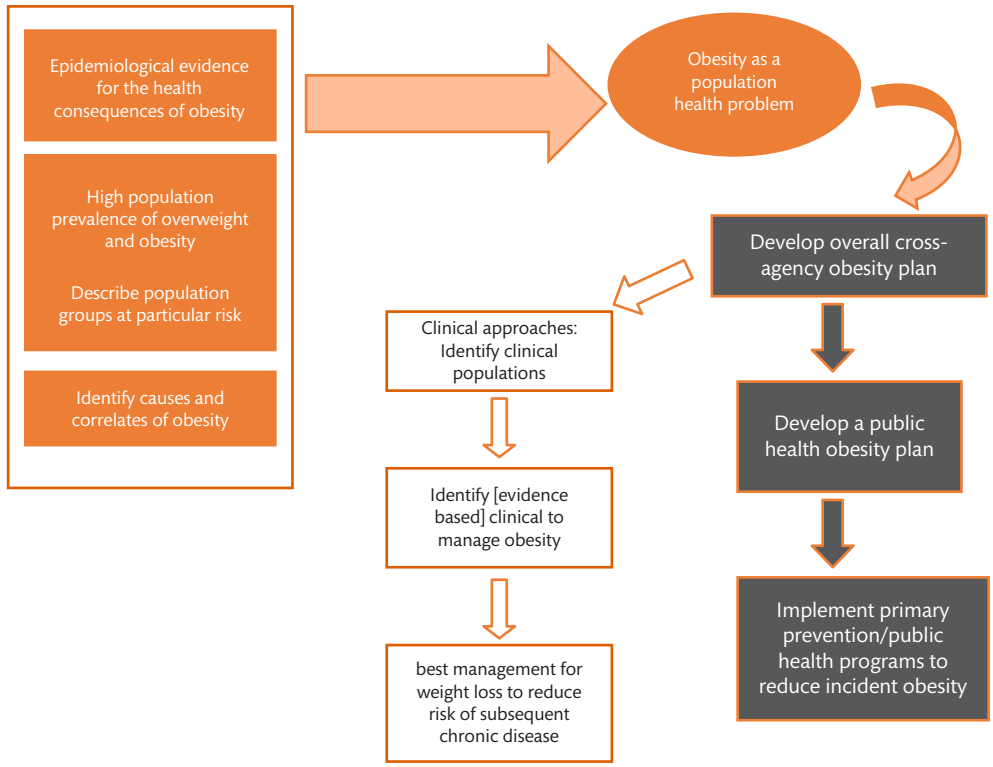


Figure 14.1.2.1 Stages in clinical and public health approaches to the problem of obesity.

a problem [7]. This illustrates the delays commonly observed between problem definition and public health policy response.

Prevalence and Trends in Obesity

Epidemiological studies of overweight and obesity can be characterized as (i) studies of trends in the prevalence of obesity; (ii) studies of the correlates and causes of obesity in populations; and (iii) aetiological studies of the health consequences of obesity.

The rapidly increasing rates of obesity, especially in high- and middle-income countries, have occurred since around 1980. For epidemiological and surveillance purposes, studies standardize assessment through the unitless metric, BMI, derived as weight (kg) divided by height in metres squared. Levels of BMI over 25 are classified as overweight and BMI values over 30 as obese. Obesity increases have mesmerized policymakers, the mass media, and the general public, although solutions are challenging and complex to implement.

In high-income countries, overweight and obesity may affect half to two-thirds of the adult population, and a quarter of adolescents [8]. Globally, objectively measured obesity among adults has increased between 1975 and 2016, with the highest current rates in Polynesian/Melanesian Pacific Island countries, in some middle Eastern nations, and in the USA [8]. Self-reported data on obesity trends adapted and updated from an OECD report are shown in Figure 14.1.2.2. For high-income countries such as the USA, the UK, and Australia, increases started in the 1980s, with obesity rates reaching between 20–30% of adults within two decades. Overall obesity rates have started to plateau in European countries, but seem to be still increasing in the USA [8]. Several countries in Europe (such as the Netherlands and France) and high-income countries in Asia (Japan, South Korea) showed a delayed pattern, with obesity increases not occurring until the mid-1990s or later. Other global patterns [8] show later onset in middle-income and rapidly developing countries in the twenty-first century, notably countries with large populations that are showing rapid industrialization and urbanization, such as China and India, who now report huge absolute numbers of overweight and obese adults.

Global childhood and adolescent obesity rates have increased from <1% in 1975 to around 6% of in 2016 [8]. There has been a concomitant global decline in underweight children. For children in Western Europe and the UK, a clear flattening off in prevalence has occurred, assessed through objectively measured weight in population surveillance systems. Nonetheless, as with adults, even if a plateau in obesity occurs, it still poses a long-term public health challenge in the form of future non-communicable disease risk. By contrast, in many middle-income countries, obesity rates are continuing to increase, and prevention efforts are urgently needed.

Obesity rates vary between and within countries. In high-income countries, the risk of obesity was originally higher among affluent adults and children, but is now 2–3 times more likely in the most socially disadvantaged groups; by contrast, in rapidly industrializing countries, such as China and India, the reverse pattern may still occur, obesity being more prevalent among the economically booming and wealthier middle classes [9]. Obesity in high-income countries varies by other demographic characteristics, with rates higher in rural than urban adults, and higher rates in specific population subgroups, such as indigenous populations in Australia, New Zealand, and North America. These demographic patterns, and the delineation of at-risk subgroups may be quite different in low- and

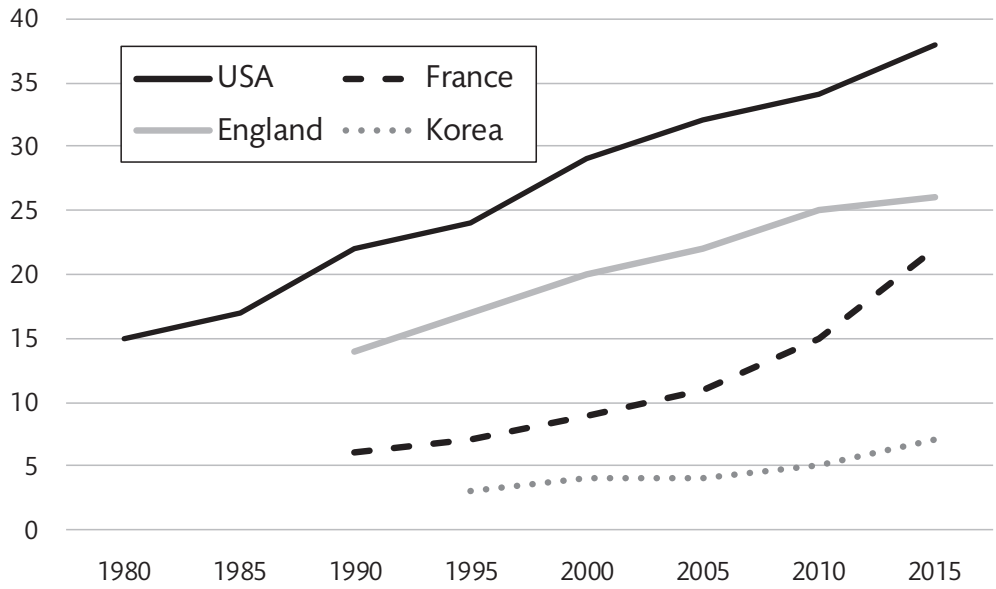


Figure 14.1.2.2 Trends in adult obesity rates selected countries.
Data derived from OECD (2013), Education at a Glance 2013: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/eag-2013-en

middle-income countries. In some countries, obesity rates are higher among women (such as in Africa), but this pattern is not universal.

Other correlates that are relevant to policy relate to the social and physical environments in which people live. Those who have ready access to inexpensive unhealthy foods are more likely to be obese than those with access to healthier food and physical activity choices in their local environments [10]. Social and cultural groups where obesity was or has become normative also show an increased likelihood of being overweight or obese. These environmental factors are mutable, but may interact with individual genetic propensity for obesity, through increasing focus on epigenetic mechanisms.

Health Risks and Costs of Obesity

The health risks of overweight and obesity are based on a combination of basic science (mechanistic evidence), clinical observations and trials, and long-term prospective epidemiological studies. These risks are reviewed in Chapter 14.1.3, 'Medical Complications of Obesity', but selected elements of the evidence base are of public health significance and will be discussed further.

The proportion of chronic disease attributable to obesity varies among countries and across epidemiological studies. This variation is contributed to by different methods for estimating attributable risk. Reported estimates for all-cause mortality ranged from 5% to 15%, up to 8% for all-cancer incidence, 7% to 44% for cardiovascular disease (CVD) incidence, and 3% to 83% for incident diabetes.

Some epidemiological studies noted the lack of increased cardiovascular risk for overweight adults, and sometimes even for grade 1 obesity. This has become known as the 'obesity paradox'. In a meta-analysis, Flegal (2013) observed that there was a slight decrease in all cause and cardiovascular deaths among the overweight and grade 1 obese [11, 12]. More recent large studies have challenged this view. A meta-analysis of 239 cohorts (10.6 million adults, from across the world, followed for ~13.7 years) has provided contrary evidence and has suggested that there is a continuous increased risk across all levels of overweight and obesity, with risk increasing at the usual established overweight level (BMI≥ 25) [13]. A similar fine-grained analysis of CVD incidence and mortality reached a similar conclusion, namely that all levels of overweight and obesity conferred increased risk [14].

An area of consistent evidence is the relationship between obesity and type 2 diabetes. Meta-analytic evidence from epidemiological cohorts suggests a 7–8-fold increase in risk of diabetes incidence among obese adults, and a threefold increase in risk among those who are overweight. The distribution of obesity appears to influence risk, with abdominal (central) obesity being a marker of greater risk, even among people with apparently healthy BMI levels. The population attributable fraction or obesity and diabetes is large, with up to 40–50% of diabetes potentially preventable if obesity were eliminated from the population [11]. In the large Swedish SOS trial, improvements in diabetes control occurred with weight loss when it occurred at any initial level of BMI, indicating the metabolic benefits occur across the spectrum of obesity [15].

Obesity is associated consistently with diabetes risk, but not all obese people develop diabetes. An expert consensus statement concluded that some of the risk is mediated through inflammatory biomarkers, and other metabolic factors that regulate glycaemic control and insulin resistance [16]. Further, being 'metabolically healthy' despite being obese does lower non-communicable disease (NCD) and mortality risk, irrespective of which definition of metabolic health is used. This may be further attenuated if the person is physically active, which may offset diabetes risk [17].

Another area of concern is the risk of maternal obesity or excessive weight gain in pregnancy, which may lead to adverse pregnancy-related and postnatal outcomes, as well as influencing the NCD risk in their offspring [18]. This is a public health problem, since effective interventions are seldom generalized and delivered to

the whole target population. This requires strategies beyond enrolling volunteer overweight pregnant women into lifestyle trials.

Although recent evidence suggests that increases in health risk occur across all levels of increased BMI or adiposity, the risks increase markedly for grade 2 and 3 obesity (BMI levels > 35 and >40 respectively). At these higher ends of the obesity spectrum, clinical therapy is essential, including pharmacological and surgical approaches. For the much large proportion who are overweight, upstream prevention approaches are essential to prevent incident obesity, prevent additional weight gain in those already overweight, and the challenge of weight loss, where possible, at the population level.

Furthermore, since obesity tracks across the lifespan, high BMI levels in adolescence may lead to chronic disease risk decades later. However, the media-fuelled alarmist concept that children may have shorter life expectancies than their parents [19] is probably an overstatement, as management continues to improve. Obese people are kept alive for longer, and the once-anticipated ever-increasing linear projections in obesity are unlikely in high-income countries. Nonetheless, this is not cause for public health complacency, especially in middle-income countries, where obesity prevalence is rapidly escalating, and globally, as the absolute numbers at risk of chronic disease still pose dire challenges for future healthcare spending.

Many studies have examined the *direct costs* (healthcare) of overweight and obesity, and the *indirect costs* (IC) (lost productivity, absenteeism, premature morbidity, and mortality) or both. The parameters vary substantially across studies, but are always impressively high as a total dollar value, or as a proportion of health costs or national GDP [20] and consistent increases in most countries are seen. Typically, studies report that between 3% and 10% of healthcare costs, or between 0.7 and 2.8% of total GDP is attributable to obesity [21]. Another method involves estimation of the proportion of costs attributable to obesity-related non-communicable disease and musculoskeletal disease. Around 15–30% of such disease costs may be due to obesity and overweight [22].

Population Level Causes of Obesity

It is unlikely that the rapid increases in obesity are wholly attributable to individual-level biological changes or genetic factors alone, and equally improbable that very similar patterns of increase coincided with the subsequent economic development in low- to middle-income countries. Furthermore, the reverse trend, seen when a country rapidly becomes less affluent and obesity decreases in a short time period, as was demonstrated in Cuba in the 1990s [23]. This does not preclude epigenetic contributions to obesity, but does point to a more complex web of metabolic, behavioural, environmental, cultural, and socioeconomic influences. These were summarized in the Foresight model prepared for the UK Department of Health [24]. This report identified groups of interrelated areas in the food system and the physical activity system that contributed directly or indirectly to the 'obesogenic environment' and to obesity. This systems map identified thematic groupings of likely causal factors [24]. These included:

- Individual-level factors can contribute, including individual physiology, responses to thermogenesis, response to early life exposures, and individual potential for epigenetic phenomena to upregulate obesity-related metabolism
- Dietary behaviours that exist within a food system. These are composed of food production choices and food distribution systems that influence patterns of food consumption; additionally, geographic and cultural factors related to food availability and choices; societal and individual norms related to food preferences, convenience, and choice; food marketing and distribution of low-cost energy-dense and nutrient-poor foods to the majority of the population
- Physical activity-related behaviours within a 'physical activity system', influencing active recreation, exercise, and sport; urban environments and transport systems that influence active travel and active commuting; influences on activity at workplaces and in domestic settings.

While genetic factors, individual-level choices, and metabolic factors may be amenable to clinical and individual-level interventions, the broader social, environmental, and societal level contributors to 'obesogenesis' provide a mandate for population-wide preventive interventions. Even if genetic factors are only expressed in adverse 'obesogenic' environments, then programmes, policies, and strategies to address these macro-level determinants are needed to address obesity in the immediate future.

Public Health Solutions

There is broad awareness of the extent and seriousness of the obesity problem, but it is very challenging to develop and implement feasible public health solutions at the population scale. The principles of public health solutions to obesity can be drawn from previous public health 'success areas', such as tobacco control and HIV prevention in many high-income countries. These have been distilled into the 'lessons for obesity' shown **Box 14.1.2.1**. They epitomize the components of a likely effective community-wide approach to obesity prevention. Note that some solutions lie outside the health sector, and that 'evidence-based action' requires public health actions across the inter-related obesogenic 'system' in society.

Fox (2013) suggested key obesity-relevant policy actions for the USA, namely increasing the availability of healthy foods (and reducing 'food deserts'); taxing unhealthy foods or subsidizing healthy foods (especially for disadvantaged groups) [25]. One target is sugar-sweetened beverages (SSBs), which include carbonated and other high-sugar, mass-produced and heavily marketed beverages, it is widely accepted that taxation of SSBs appear effective in reducing population SSB consumption, but may be politicised and not implemented through subtle food industry lobbying and influences on governmental views. Physical activity-relevant policies include improving public transport systems and the provision of places to be active, including sports fields, walking paths (sidewalks), and low-cost exercise facilities. The evidence base is mixed, but this public health approach describes 'best bets' in a comprehensive suite of obesity prevention policies [26].

For children and adolescents, obesity prevention strategies have been described by World Health Organization (WHO), starting with prenatal influences and preventing excess maternal weight

Box 14.1.2.1 Six lessons from tobacco: elements of successful public health relevant to obesity

1. Need to consider the social and environmental determinants [of obesity], which includes working outside the health sector, with a range of Government departments, non-Government agencies, and the private sector
2. Need sustained public education, and consistently disseminated obesity mass media and social media campaigns, and coordinated advocacy (especially against unhealthy food marketing)
3. Have a clear cross-agency (cross-sectoral) mandate, clearly identified leadership, sustained policy support and resources, and an implemented multiyear (national) obesity prevention plan
4. Implementing regulatory policies and taxation as appropriate (for SSBs or across a broader range of unhealthy foods) and policies that change the (food and physical activity related) environments and the social environment (social norms) that support obesity
5. Have the flexibility to implement actions, based on best available public health evidence (not waiting for controlled trial and meta-analytic evidence syntheses to precede public health actions)
6. Integrate health sector and health professional advice into the overall obesity prevention strategy

Data derived from Bonfiglioli, C.M.F., Allman-Farinelli, M.A., King, L., Bauman, A.E., 2008. Mapping solutions to obesity: Lessons from the Human Genome Project. *Australian and New Zealand Journal of Public Health* 32, 546–8; Capewell, S., Lloyd-Williams, F., 2018. The role of the food industry in health: Lessons from tobacco? *British Medical Bulletin* 125, 131–43; Heymann, E.P., Goldsmith, D., 2011. The battle against obesity: Lessons from tobacco. *The Lancet* 378, 2069.

gain in pregnancy. Subsequently, breastfeeding should be encouraged at least to 6 months of age [27]. Strategies for childhood obesity include restricting marketing of unhealthy foods and beverages, ideally through legislated action. Taxes on SSBs are recommended as an evidence-based approach to reducing consumption at the population level [27]. Front of pack food labelling is also recommended. School environments should prioritize healthier canteen food, restrict vending machines, and encourage physical activity within and outside of the school.

Integrating Clinical and Population Approaches

In summary, there is a clear place for clinical treatment of obesity, but given its population prevalence, upstream preventive efforts are also required. The latter are expensive, require formulation in a policy environment, and may pose challenges to implement. Ongoing advocacy for obesity prevention efforts, for the implementation of existing national obesity plans, and for ongoing surveillance to document success are essential ingredients. The biggest task for public health is to integrate efforts across Government and non-Government agencies, beyond the health sector alone.

REFERENCES

1. Kyle TK, Dhurandhar EJ, Allison DB. Regarding obesity as a disease: evolving policies and their implications. *Endocrinol Metab Clin*, 2016; **45**: 511–20.
2. Bray GA. The tide shifts again: the ebb and flow of history. *Obes Res*, 1995; **3**: 605–8.
3. Jolliffe N. Some basic considerations of obesity as a public health problem. *Am J Public Health Nations Health*, 1953; **43**: 989–92.
4. *Life Magazine*. The plague of overweight, 8th March 1954: 120–4
5. Huenemann RL. Consideration of adolescent obesity as a public health problem. *Public Health Rep,* 1968; **83**: 491–95.
6. Garrow J. Obesity—Public health enemy No 2? *Health Educ J,* 1986; **45**: 57–8.
7. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic (No. 894)*. Geneva: World Health Organization, 2000.
8. Lancet, NCD risk factor Collaboration. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*, 2017; **390**: 2627–42.
9. Dinsa GD, Goryakin Y, Fumagalli E, Suhrcke M. Obesity and socioeconomic status in developing countries: a systematic review. *Obes Rev,* ; **13**: 1067–79.
10. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*, 2006; **117**: 417–24.
11. Flegal K, Panagiotou OA, Graubard BI. Estimating population attributable fractions to quantify the health burden of obesity. *Ann Epidemiol*, 2015; **25**: 201–207.
12. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories a systematic review and meta-analysis. *JAMA*, 2013; **309**: 71–82.
13. Di Angelantonio E, Bhupathiraju SN, Wormser D, *et al*. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*, 2016; 388: 776–86.
14. Khan SS, Ning H, Wilkins JT, *et al*. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA Cardiol*, 2018; **3**: 280–7.
15. Sjöholm K, Pajunen P, Jacobson P, *et al*. Incidence and remission of type 2 diabetes in relation to degree of obesity at baseline and 2 year weight change: the Swedish Obese Subjects (SOS) study. *Diabetologia*, 2015; **58**: 1448–53.
16. Eckel RH, Kahn SE, Ferrannini E, *et al*. Obesity and type 2 diabetes: what can be unified and what needs to be individualized?. *J Clin Endocrinol Metab*, 2011; **96**: 1654–63.
17. Colberg SR, Sigal RJ, Yardley JE, *et al*. Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. *Diabetes Care*, 2016; **39**: 2065–79.
18. Godfrey KM, Reynolds RM, Prescott SL, *et al*. Influence of maternal obesity on the long-term health of offspring. *Lancet Diabetes Endocrinol*, 2017; **5**: 53–64.
19. Olshansky SJ, Passaro DJ, Hershow RC, *et al*. A potential decline in life expectancy in the United States in the 21st century. *N Engl J Med*, 2005; **352**: 1138–45.
20. Specchia ML, Veneziano MA, Cadeddu C, *et al*. Economic impact of adult obesity on health systems: a systematic review. *Eur J Public Health*, 2015; **25**: 255–62.
21. Withrow D, Alter DA. The economic burden of obesity worldwide: A systematic review of the direct costs of obesity. *Obes Rev*, 2011; **12**: 131–41.
22. Lette M, Bemelmans WJE, Breda J, Slobbe LCJ, Dias J, Boshuizen HC. Health care costs attributable to overweight calculated in a standardized way for three European countries. *Eur J Health Econ*, 2016; **17**: 61–9.

23. Franco M, Bilal U, Orduñez P, Benet M, *et al*. Population-wide weight loss and regain in relation to diabetes burden and cardiovascular mortality in Cuba 1980–2010: repeated cross sectional surveys and ecological comparison of secular trends. *BMJ*, 2013; **346**: f1515.
24. Foresight—tackling obesities: future choices—project report, 2nd edn. Government Office for Science, UK, 2007. Available at: https://www.gov.uk/government/collections/tackling-obesities-future-choices (accessed: 15 April 2020).
25. Fox AM, Horowitz CR. Best practices in policy approaches to obesity prevention. *J Health Care Poor Underserved*, 2013; **24**: 168.
26. Swinburn B, Gill T, Kumanyika S. Obesity prevention: a proposed framework for translating evidence into action. *Obes Rev*, 2005; **6**: 23–33.
27. World Health Organization. Taking action on childhood obesity. Geneva: World Health Organization, 2018. Available at: http://www.who.int/iris/handle/10665/274792 (accessed: 15 April 2020).

14.1.3 Medical Complications of Obesity

Friedrich C. Jassil and Rachel L. Batterham

Introduction

Obesity is a chronic progressive condition associated with multiple diseases and reduced life expectancy (Figure 14.1.3.1). Countries such as the United States and Canada, as well as several professional health bodies including the World Health Organization and the World Obesity Federation, have recognized obesity as a disease, on the basis that the pathogenesis of obesity fits the epidemiological model of a disease process. In this chapter, the latest epidemiological and clinical studies describing the associations and suggested links between obesity and its medical complications are summarized, and current understanding of potential underlying mechanisms are briefly described.

Candidate Mechanisms Linking Obesity and its Medical Complications

A large number of mechanisms have been proposed to account for the link between obesity and metabolic sequelae, but how these interact in humans, and which mechanisms play the dominant role, remains uncertain. Important themes in current understanding include inflammatory consequences of adipose tissue overload, and differences among different adipose depots.

Adipocytes play a crucial role metabolically, storing excess caloric intake in the form of highly energy-dense triglycerides. They also produce a panoply of factors that act in a paracrine, autocrine, and endocrine manner, impacting upon a multitude of physiological functions including energy and glucose homeostasis, reproduction, and immune responses [1]. Following periods of energy surplus, adipocytes expand in size and number to accommodate excess lipid storage. Hypertrophic adipocytes resulting from unhealthy increase in size produce pro-inflammatory adipokines and cytokines such as tumour necrosis factor-alpha (TNF-α) and interleukin-6 (IL-6), with concomitant downregulation of the anti-inflammatory protein adiponectin. Ultimately, these changes lead to chronic systemic low-grade inflammation which is associated with adverse clinical manifestations such as atherosclerosis, insulin resistance and hyperinsulinaemia [2]. It has been shown that individuals with obesity, particularly abdominal (central) obesity, exhibit greater levels of pro-inflammatory biomarkers compared to normal-weight counterparts [3]. Conversely, achieving at least 10% weight loss is beneficial in improving the obesity-associated unfavourable inflammatory markers.

Differences among different adipose depots are also believed to play an important role in obesity-related disease. In the face of continued positive energy balance, it is postulated that when subcutaneous adipose tissue reaches its maximal capacity to expand, excess substrate is redirected and accumulates as lipid in visceral depots, leading to abdominal obesity, and is also accumulated ectopically, in extra-adipose sites in organs such as the liver, muscles, pancreas, and heart [4]. Excess ectopic lipid leads to cell lipotoxicity and lipoapoptosis and impairment of the respective organ function. For instance, excess lipid accumulation in the liver results in non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH), in muscle leads to insulin resistance, in the pancreas is associated with β-cell failure, and in the heart is linked to cardiomyopathy [5].

Health Complications of Obesity

Type 2 Diabetes Mellitus

Being obese is associated with a sevenfold increase in the relative risk (RR) of developing T2D [6]. More worryingly, over the last few decades, the rise in prevalence of childhood obesity has been accompanied by a growing trend of incident T2D in youth, dispelling the notion of T2D as an adult-onset disease. For every ten years of living with obesity, there is a 7% higher chance of developing T2D for both men and women [7]. Importantly, adiposity gained during young adulthood has been linked to a greater RR of T2D incidence than if the same level of adiposity was gained during middle age [8].

The ability of weight loss interventions to ameliorate T2D provides further evidence for a causal role of obesity in mediating T2D. In the Diabetes Remission Clinical Trial (DiRECT), weight loss induced by lifestyle measures was associated with reversal to a non-diabetic state in a weight dependent manner. Of those achieving ≥15 kg weight loss at 12 months' follow-up, 86%

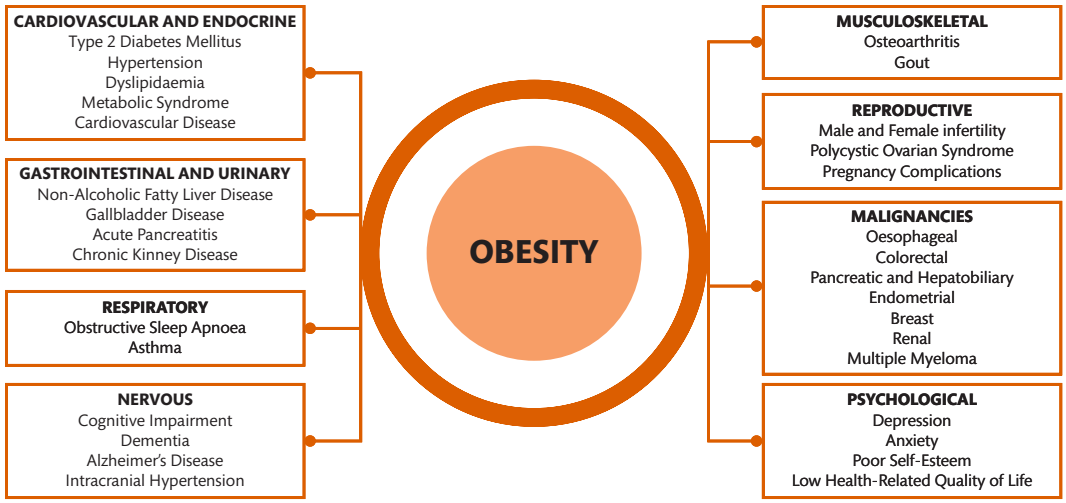


Figure 14.1.3.1 Health complications of obesity.

were able to achieve a glycated haemoglobin (HbA1c) <6.5% and stop all of their T2D medication compared to only 7% in those who were able to maintain 0–5 kg weight loss [9]. Similarly, following bariatric surgery, T2D remission has been shown to be dependent upon the magnitude of weight loss achieved post-surgery. Indeed, mechanistic studies undertaken in animals and humans have demonstrated that weight loss induced via lifestyle interventions or surgical means leads to restoration of the first-phase insulin response, possibly owing to reversal of intrapancreatic fat accumulation, which is suggested to drive β-cell dysfunction and de-differentiation [10].

Hypertension

Obesity also plays a key role in the pathogenesis of hypertension. According to data from the Third National Health and Nutrition Examination Survey (NHANES III), the prevalence of hypertension in men and women with Body Mass Index (BMI) ≥30 kgm^{-2} was 41.9% and 37.8%, respectively, compared to 14.9% and 15.2% in people with BMI ≤25 kgm^{-2} [11]. Also, data from the Johns Hopkins Precursors Study further revealed that obesity in early adulthood conferred a threefold risk of developing hypertension later in life even after adjustment for potential confounders [12]. It is known that blood pressure control is more difficult to achieve in hypertensive patients with obesity while weight loss resulting in better control. Furthermore, obesity is also a common cause of treatment-resistant hypertension. These highlight the importance of weight loss as an adjunct therapeutic approach in managing hypertension.

Dyslipidaemia

As discussed in more detail in Chapters 15.13.4, 'Diabetic Dyslipidaemia' and 14.2 'Lipoprotein Metabolism and Dyslipidaemia', many individuals with obesity have deranged lipid metabolism that manifests clinically as fasting and postprandial lipemia with elevated triglyceride, reduced high-density lipoprotein (HDL) cholesterol and slightly increased or normal level of low-density lipoprotein (LDL) cholesterol. Data from the NHANES III revealed a direct relationship between increased BMI and dyslipidaemia in adult (11). Similarly, in youth aged 12–19 years, 42.9% of those with BMI exceeding the 95th percentile were reported to have at least one abnormal lipid level [13]. In obesity, enhanced free fatty acid influx to the liver, brought about by insulin resistance, leads to continuous hepatic hypersecretion of very low-density lipoprotein (VLDL) particles. Once this triglyceride-rich lipoprotein is released into the circulation, it undergoes a sequence of lipoprotein changes through enzymatic actions leading to a reduction in circulating levels of antisclerotic HDL cholesterol and production of the atherogenic remnant of small dense LDL with consequent hypertriglyceridemia. Increased circulating small dense LDL can trigger the pro-inflammatory response that eventually promotes the development of atherosclerosis. Conversely, reduced HDL cholesterol levels impair the reverse cholesterol transport, a mechanism that serves as a protection against atherosclerosis [14].

Metabolic Syndrome

Metabolic syndrome (MetS) is a clinical condition diagnosed based on the simultaneous clustering of at least three of the following risk factors: abdominal obesity, raised triglycerides, reduced HDL cholesterol, elevated blood pressure, and raised fasting plasma glucose. This constellation of metabolic abnormalities are primarily mediated by insulin resistance. In the long-term, MetS increases the risk of T2D and cardiovascular disease (CVD) [15].

In a Multi-Ethnic Study of Atherosclerosis (MESA) involving 2748 adult men and women, of 1362 subjects with obesity without MetS at baseline, 44% developed MetS within 2 years. Further analyses have shown that greater obesity severity and duration were both associated with higher odds of incident MetS [16]. Indeed, growing evidence from prospective studies suggests that metabolically healthy obesity is a transient state before transitioning to a metabolically unhealthy phenotype, mainly driven by an increase in general and abdominal adiposity. Medical weight loss that leads to reduction in waist circumference has been shown to improve all components of metabolic abnormalities of MetS. Importantly, a recent systematic review and meta-analysis has concluded that children with overweight or obesity from birth to 6 years of age have a higher risk of developing MetS later in life [17].

Cardiovascular Disease

Obesity is a well-known independent risk factor for CVD. Obesity also increases CVD risk secondarily through worsening of other weight-related comorbidities including T2D, hypertension, dyslipidaemia, MetS, and NAFLD. Together, obesity and its comorbidities adversely affect the cardiovascular system, impacting upon cardiac structure, function and haemodynamics [18]. The higher incidence of coronary artery disease in people with obesity has been reported to be strongly associated with endothelial dysfunction and systemic inflammation [19]. Worryingly, robust evidence from epidemiological and post-mortem studies in children with obesity has shown that the atherogenic process begins in early childhood [20]. A meta-analysis of prospective studies including more than two million adults from Western and Eastern populations has demonstrated that relative to normal-weight individuals, the RR for ischaemic stroke and haemorrhagic stroke were 1.64 and 1.24, respectively, in individuals with obesity [21]. Also, data from the Framingham Heart Study showed that after adjustment for other established risk factors, obesity was associated with 5% and 7% increased risk of heart failure in men and women, respectively [22]. Moreover, in another report from the same study cohort, every one unit of increment in BMI was independently associated with 4% increased risk of atrial fibrillation observed in both genders [23]. Also, obesity doubled the risk of developing venous thromboembolism due to its hypercoagulability and pro-inflammatory state, and impaired fibrinolysis [24].

Cancer

Another major adverse health consequence of obesity is cancer. Recent reports from both the United States and United Kingdom populations revealed that overweight and obesity is now one of the leading causes of cancer, second only to tobacco smoking. Obesity is linked to many cancer types with strong evidence observed for oesophageal, colorectal, pancreatic, and hepatobiliary, endometrial (in premenopausal women), breast (in postmenopausal women) and renal cancers, in addition to multiple myeloma [25]. Many mechanisms and mediators have been proposed to underlie increased obesity-related risk of cancer development and progression, including insulin resistance with compensatory hyperinsulinaemia; pro-inflammatory changes in adipose secretome including upregulation of leptin and cytokines and downregulation of adiponectin; changes in sex hormones; increased circulating plasminogen activator inhibitor-1 and vascular endothelial growth factor; and changes in gut microbiota [26]. Importantly, obesity may impact upon cancer survival as well as quality of life, cancer progression, and recurrence, metastasis, treatment efficacy, and adverse treatment effects.

Non-Alcoholic Fatty Liver Disease

Excess adiposity is the major risk factor for NAFLD, characterized by intrahepatic triglyceride (IHTG) content of more than 5%. A subset of patients with obesity-induced NAFLD progress to NASH, a more aggressive form of fatty liver characterized by steatosis, inflammation and/or fibrosis which may further progress to cirrhosis and hepatocellular carcinoma (HCC). Worryingly, accumulating evidence has shown that HCC can develop in patients with NAFLD in the absence of cirrhosis [27]. In a pathophysiological study undertaken in subjects with obesity, 5% weight gain was associated with 20% increase in *de novo* intrahepatic of fatty acid synthesis from carbohydrates, hence increasing the IHTG content. In contrast, the rate of β-oxidation, a mechanism for removing IHTG, decreased substantially [28]. Both of these processes were mediated by impairment in insulin sensitivity as a result of weight gain. Although the export of IHTG from the liver was upregulated through higher production of triglyceride-rich VLDL particles following weight gain, this process was unable to adequately compensate for the increased IHTG production rate [28].

Gallbladder Disease

Obesity is one of the predisposing non-genetic risk factors for gallbladder disease, particularly gallstones. Analysis of data from published prospective studies revealed that for every five unit increase in BMI, the RR for gallbladder disease was 1.63 [29]. Hypersecretion of hepatic cholesterol is known to be one of the metabolic abnormalities occurring in obesity. Increased biliary cholesterol secretion leads to supersaturation of bile, resulting in formation of cholesterol gallstones [30]. Due to the increased prevalence of gallstones among people with obesity, the incidence of developing acute pancreatitis is thus higher. Unfortunately, if people with obesity do develop acute pancreatitis, its severity tends to be greater.

Chronic Kidney Disease

There is an association between obesity and renal impairment, even in the absence of the predisposing comorbidities such as T2D and hypertension. An analysis undertaken in more than 600 000 adults with normal renal function found that after a mean follow-up period of 6.8 years, participants with obesity exhibited 28% increased risk for new-onset reduced estimated glomerular filtration rate and 51% increased risk for albuminuria [31]. Obesity causes upregulation of the renal tubular sodium reabsorption, and this leads to renal vasodilation and glomerular hyperfiltration that act as feedback mechanisms to maintain sodium balance. Together, these haemodynamic changes result in glomerular hypertension, glomerulomegaly and glomerulosclerosis that eventually lead to a progressive loss of kidney function [32]. Weight loss intervention, especially bariatric surgery, have been shown to prevent further decline in renal function through normalization of glomerular filtration rate and reduction in albuminuria and proteinuria.

Reproductive Health and Pregnancy Outcomes

Obesity affects a significant number of women of childbearing age and has a substantial negative impact on overall reproductive health. It is well documented that obesity contributes to primary and secondary infertility and exacerbates many aspects of the polycystic ovarian syndrome phenotype [33]. Women with obesity are more likely to experience anovulation brought about by dysregulation of the hypothalamic–pituitary–ovarian axis and high circulating leptin levels. Furthermore, obesity also appears to affect oocyte development and quality, embryo implantation and endometrial receptiveness [33, 34]. As compared to pregnancy in normal weight, the risk of adverse maternal and fetal outcomes such as miscarriage, gestational diabetes, pregnancy-induced hypertension, prematurity, abnormal fetal growth and congenital anomalies are increased in obesity [33, 35]. Obesity also negatively impacts upon the male reproductive potential, primarily due to impaired spermatogenesis and sexual

dysfunction. Data from a meta-analysis provide clear evidence that men with obesity are more likely to exhibit poor semen quality such as oligospermia or azoospermia and increased proportion of sperm with DNA damage, abnormal morphology, and low mitochondrial membrane potential, compared to normal-weight men [36]. Furthermore, it has become evident that both maternal and paternal obesity are associated with adverse health consequences of the offspring such as increased risk of obesity, T2D, and CVD in later life, mediated by epigenetic alterations in sperm and *in utero* [35, 37].

Weight reduction increases the chances of pregnancy from natural conception and improves ovulation, menstrual cycles regulation, and an overall improvement in pregnancy outcomes with higher live birth rates. Similarly, in men with obesity, weight loss has been shown to be effective in reversing infertility by improving semen parameters and reproductive hormonal profile.

Osteoarthritis

Obesity is a recognized modifiable risk factor for osteoarthritis (OA). Every five unit increase in BMI is associated with 35% and 11% increased risk for knee and hip OA, respectively [38, 39]. In a large population-based cohort study of 1 764 061 subjects followed-up for a median of 4 years, the incidence rate of OA was reported to be higher in subjects with obesity compared to subjects with normal weight with the knee joint being greatly affected followed by the hand and hip [40]. OA in obesity was initially thought to be solely caused by biomechanical overload of the weight-bearing joints due to excessive body mass but new evidence is emerging that suggests increased expression of pro-inflammatory cytokines and adipokines (particularly leptin) and obesity-related dyslipidaemia may play a pivotal role. This explains why OA of non-weight-bearing joints such as those in the hand is also common among individuals with obesity [41]. The relationship between obesity and OA also appears to be bidirectional. OA leads to reduced mobility often contributing to further weight gain and a vicious cycle of increasing weight and worsening OA. Weight loss is a mainstay of treatment for OA, even a 5% weight reduction will promote improvement in physical disability. Importantly, patients with obesity who underwent joint replacement surgery for the treatment of severe or end-stage OA reported experiencing poorer post-surgery outcomes compared to patients without obesity.

Gout

Obesity increases the burden of gout. Data from the NHANES has shown the relationship between body weight and prevalence of gout appears to be dose-dependent and continues to persist even after adjustment for comorbidities and serum uric acid level [42]. It is postulated that obesity-associated insulin resistance and high leptin production promote hyperuricemia by inhibiting the renal excretion of uric acid. Furthermore, associated comorbidities such as elevated triglycerides and blood pressure increase plasma uric acid concentration [43]. In contrast, weight loss among patients with overweight and obesity diagnosed with gout has been shown to provide a beneficial effect in lowering uric acid level.

Obstructive Sleep Apnoea

Obesity is widely acknowledged as a major risk factor for obstructive sleep apnoea (OSA). Indeed, the incidence of OSA in adults, children, and adolescents parallels obesity prevalence. Individuals with obesity, particularly those with regional obesity, have a greater neck circumference due to parapharyngeal fat deposition resulting in greater collapsibility of the pharyngeal airway during sleep. Furthermore, intra-abdominal fat accumulation in obesity decreases lung volumes which further exacerbates OSA. Elevated circulating leptin levels and pro-inflammatory cytokines may also play a role in the pathogenesis of obesity-related OSA given their influence upon respiratory drive and neuromuscular control of the upper airways [44]. Interestingly, OSA also contributes to further weight gain as sleep disruption leads to altered energy expenditure, appetite hormones, and eating pattern.

Asthma

Obesity is recognized as an important risk factor for asthma in both children and adults. In fact, the prevalence and incidence of asthma are highly correlated with the magnitude of obesity. Obesity is associated with poor asthma control, frequent hospitalization, and decreased response to treatment. The mechanism by which obesity causes asthma is still not well understood, but genetic susceptibility, changes in lungs mechanics, airway inflammation, cardiometabolic abnormalities and diet and lifestyle have been suggested to be key contenders [45]. Indeed, weight loss via both non-surgical and surgical means lead to overall improvement in asthma control.

Cognitive Impairment and Dementia

Growing evidence supports the link between obesity and cognitive dysfunction encompassing poor executive function and short-term memory and increased risk of dementia, particularly Alzheimer's disease. The recent finding based on data from over 1.3 million adults from different regions indicates that high midlife BMI was directly associated with greater risk of dementia in later life [46]. Chronic inflammation, neurodegeneration, vascular disease, and metabolic alteration are among the potential mechanisms proposed to explain the link between obesity and cognitive decline [47].

Intracranial Hypertension

Intracranial hypertension is characterized by raised intracranial pressure that typically seen in women with obesity and of reproductive age [48]. Headaches, visual loss, pulsatile tinnitus, dizziness, and back and neck pain are among the common reported symptoms. The mechanism by which obesity causes intracranial hypertension is still unknown [48]. However, preliminary findings have suggested the beneficial effect of weight loss surgery in improving the associated symptoms and a reduction in intracranial pressure.

Psychological Comorbidities

From a psychological point of view, children and adolescent with obesity are more likely to develop depression, anxiety, lower self-esteem, and poor health-related quality of life compared with peers of healthy-weight. This is in part, due to weight-related stigma, teasing and bullying which can have a serious impact upon their mental health [49]. Adults with obesity report perceived discrimination in various settings such as work-related, healthcare-related,

and day-to-day interpersonal settings. Such weight-related stigmatization may negatively affect psychological well-being. It is therefore not surprising that people with obesity were 32% more likely to develop depression compared to their normal-weight counterparts [50].

Conclusion

Excess accumulation of adipose tissue leads to an adverse metabolic and inflammatory state that drives several disease processes, which, coupled with biomechanical effects of excess weight and obesity-associated stigma, lead to a multitude of comorbidities and premature death. A deeper understanding of how obesity contributes to such a myriad of health complications is therefore imperative and will pave the way for development of evidence-based therapeutic interventions. Importantly, the growing body of evidence that obesity-related health complications are present in young people with obesity highlights the need for more effective and robust prevention and therapeutic strategies that should be implemented as early as feasible.

REFERENCES

1. Trayhurn P, Beattie JH. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. *Proc Nutr Soc*, 2001; **60**: 329–39.
2. Greenberg AS, Obin MS. Obesity and the role of adipose tissue in inflammation and metabolism. *Am J Clin Nutr*, 2006; **83**: 461S–5S.
3. Park HS, Park JY, Yu R. Relationship of obesity and visceral adiposity with serum concentrations of CRP, TNF-alpha and IL-6. *Diabetes Res Clin Pract*, 2005; **69**: 29–35.
4. Smith U. Abdominal obesity: a marker of ectopic fat accumulation. *J Clin Invest*, 2015; **125**: 1790–2.
5. van Herpen NA, Schrauwen-Hinderling VB. Lipid accumulation in non-adipose tissue and lipotoxicity. *Physiol Behav*, 2008; **94**: 231–41.
6. Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. *Diabetes Res Clin Pract*, 2010; **89**: 309v19.
7. Abdullah A, Amin FA, Hanum F, *et al*. Estimating the risk of type-2 diabetes using obese-years in a contemporary population of the Framingham Study. *Glob Health Action*, 2016; **9**: 30421.
8. Wei GS, Coady SA, Reis JP, *et al*. Duration and degree of weight gain and incident diabetes in younger versus middle-aged black and white adults: ARIC, CARDIA, and the Framingham Heart Study. *Diabetes Care*, 2015; **38**: 2042–9.
9. Lean ME, Leslie WS, Barnes AC, *et al*. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet*, 2018; **391**: 541–51.
10. White MG, Shaw JA, Taylor R. Type 2 diabetes: the pathologic basis of reversible beta-cell dysfunction. *Diabetes Care*, 2016; **39**: 2080–8.
11. Brown CD, Higgins M, Donato KA, *et al*. Body mass index and the prevalence of hypertension and dyslipidemia. *Obes Res*, 2000; **8**: 605–19.
12. Shihab HM, Meoni LA, Chu AY, *et al*. Body mass index and risk of incident hypertension over the life course: the Johns Hopkins Precursors Study. *Circulation*, 2012; **126**: 2983–9.
13. Centers for Disease C, Prevention. Prevalence of abnormal lipid levels among youths—United States, 1999–2006. *MMWR*, 2010; **59**: 29–33.
14. Klop B, Elte JW, Cabezas MC. Dyslipidemia in obesity: mechanisms and potential targets. *Nutrients*, 2013; **5**: 1218–40.
15. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med*, 2006; **23**: 469–80.
16. Mongraw-Chaffin M, Foster MC, Kalyani RR, *et al*. Obesity severity and duration are associated with incident metabolic syndrome: evidence against metabolically healthy obesity from the multi-ethnic study of atherosclerosis. *J Clin Endocrinol Metab*, 2016; **101**: 4117–24.
17. Kim J, Lee I, Lim S. Overweight or obesity in children aged 0 to 6 and the risk of adult metabolic syndrome: a systematic review and meta-analysis. *J Clin Nurs*, 2017; **26**: 3869–80.
18. Bastien M, Poirier P, Lemieux I, Despres JP. Overview of epidemiology and contribution of obesity to cardiovascular disease. *Prog Cardiovasc Dis*, 2014; **56**: 369–81.
19. Rossi R, Iaccarino D, Nuzzo A, *et al*. Influence of body mass index on extent of coronary atherosclerosis and cardiac events in a cohort of patients at risk of coronary artery disease. *Nutr Metab Cardiovasc Dis*, 2011; **21**: 86–93.
20. Nadeau KJ, Maahs DM, Daniels SR, Eckel RH. Childhood obesity and cardiovascular disease: links and prevention strategies. *Nat Rev Cardiol*, 2011; **8**: 513–25.
21. Strazzullo P, D'Elia L, Cairella G, Garbagnati F, Cappuccio FP, Scalfi L. Excess body weight and incidence of stroke: meta-analysis of prospective studies with 2 million participants. *Stroke*, 2010; **41**: e418–26.
22. Kenchaiah S, Evans JC, Levy D, *et al*. Obesity and the risk of heart failure. *N Engl J Med*, 2002; **347**: 305–13.
23. Wang TJ, Parise H, Levy D, *et al*. Obesity and the risk of new-onset atrial fibrillation. *JAMA*, 2004; **292**: 2471–7.
24. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: a meta-analysis. *Circulation*, 2008; **117**: 93–102.
25. Kyrgiou M, Kalliala I, Markozannes G, *et al*. Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ*, 2017; **356**: j477.
26. O'Flanagan C, Bowers L, Allott E, Hursting S. Molecular and metabolic mechanisms underlying the obesity–cancer link. Energy balance and obesity: International Agency for Research on Cancer; 2017.
27. Marengo A, Rosso C, Bugianesi E. Liver cancer: connections with obesity, fatty liver, and cirrhosis. *Ann Rev Med*, 2016; **67**: 103–17.
28. Fabbrini E, Tiemann Luecking C, Love-Gregory L, *et al*. Physiological mechanisms of weight gain-induced steatosis in people with obesity. *Gastroenterology*, 2016; **150**: 79–81.e2.
29. Aune D, Norat T, Vatten LJ. Body mass index, abdominal fatness and the risk of gallbladder disease. *Eur J Epidemiol*, 2015; **30**: 1009–19.
30. Bonfrate L, Wang DQ, Garruti G, Portincasa P. Obesity and the risk and prognosis of gallstone disease and pancreatitis. *Best Pract Res Clin Gastroenterol*, 2014; **28**: 623–35.

31. Garofalo C, Borrelli S, Minutolo R, Chiodini P, De Nicola L, Conte G. A systematic review and meta-analysis suggests obesity predicts onset of chronic kidney disease in the general population. *Kidney Int*, 2017; **91**: 1224–35.
32. Hall ME, do Carmo JM, da Silva AA, Juncos LA, Wang Z, Hall JE. Obesity, hypertension, and chronic kidney disease. *Int J Nephrol Renovasc Dis,* 2014; **7**: 75–88.
33. Talmor A, Dunphy B. Female obesity and infertility. *Best Pract Res Clin Obstet Gynaecol,* 2015; **29**: 498–506.
34. Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. *Fertil Steril*, 2017; **107**: 840–7.
35. Chandrasekaran S, Neal-Perry G. Long-term consequences of obesity on female fertility and the health of the offspring. *Curr Opin Obstet Gynecol*, 2017; **29**: 180–7.
36. Campbell JM, Lane M, Owens JA, Bakos HW. Paternal obesity negatively affects male fertility and assisted reproduction outcomes: a systematic review and meta-analysis. *Reprod Biomed Online*, 2015; **31**: 593–604.
37. Raad G, Hazzouri M, Bottini S, Trabucchi M, Azoury J, Grandjean V. Paternal obesity: how bad is it for sperm quality and progeny health? *Basic Clin Androl*, 2017; **27**: 20.
38. Jiang L, Rong J, Wang Y, *et al.* The relationship between body mass index and hip osteoarthritis: a systematic review and meta-analysis. *Joint, Bone, Spine*, 2011; **78**: 150–5.
39. Jiang L, Tian W, Wang Y, *et al.* Body mass index and susceptibility to knee osteoarthritis: a systematic review and meta-analysis. *Joint, Bone, Spine*, 2012; **79**: 291–7.
40. Reyes C, Leyland KM, Peat G, Cooper C, Arden NK, Prieto-Alhambra D. Association between overweight and obesity and risk of clinically diagnosed knee, hip, and hand osteoarthritis: a population-based cohort study. *Arthritis Rheumatol*, 2016; **68**: 1869–75.
41. Thijssen E, van Caam A, van der Kraan PM. Obesity and osteoarthritis, more than just wear and tear: pivotal roles for inflamed adipose tissue and dyslipidaemia in obesity-induced osteoarthritis. *Rheumatology*, 2015; **54**: 588–600.
42. Juraschek SP, Miller ER, 3rd, Gelber AC. Body mass index, obesity, and prevalent gout in the United States in 1988–1994 and 2007–2010. *Arthritis Care Res (Hoboken)*, 2013; **65**: 127–32.
43. de Oliveira EP, Burini RC. High plasma uric acid concentration: causes and consequences. *Diabetol Metab Syndr,* 2012; **4**: 12.
44. Isono S. Obstructive sleep apnea of obese adults: pathophysiology and perioperative airway management. *Anesthesiology*, 2009; **110**: 908–21.
45. Boulet LP. Asthma and obesity. Clin Exp Allergy, 2013; **43**: 8–21.
46. Kivimaki M, Luukkonen R, Batty GD, *et al.* Body mass index and risk of dementia: analysis of individual-level data from 1.3 million individuals. *Alzheimers Dement*, 2018; **14**: 601–9.
47. Nguyen JC, Killcross AS, Jenkins TA. Obesity and cognitive decline: role of inflammation and vascular changes. *Front Neurosci*, 2014; **8**: 375.
48. Markey KA, Mollan SP, Jensen RH, Sinclair AJ. Understanding idiopathic intracranial hypertension: mechanisms, management, and future directions. *Lancet Neurol,* 2016; **15**: 78–91.
49. Rankin J, Matthews L, Cobley S, *et al.* Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. *Adolesc Health Med Ther*, 2016; **7**: 125–46.
50. Pereira-Miranda E, Costa PRF, Queiroz VAO, Pereira-Santos M, Santana MLP. Overweight and obesity associated with higher depression prevalence in adults: a systematic review and meta-analysis. *J Am Coll Nutr*, 2017; **36**: 223–33.

14.1.4 Dietary and Medical Management of Obesity

John P. Wilding and Jonathan Z.M. Lim

Introduction

Overweight and obese individuals have excess adiposity due to complex interactions among genetic, developmental, physiological, environmental, and psychological factors. Fat distribution, especially increased intraabdominal (visceral fat) may be as important as overall obesity in increasing risk of coronary artery disease, type 2 diabetes, and certain reproductive and gastrointestinal cancers. While scientists debate exactly how much visceral fat is too much, cross sectional data shows that weight gain and visceral fat increment of over twofold in men and over threefold in women between the ages of 25 and 65 years contributes to the increased risk of metabolic complications [1]. Mechanical complications include joint pain and arthritis, sleep apnoea, and higher obstetric risk. The psychosocial effects of obesity include an increased risk of mood disorders and the effects of stigmatization. The impact on health is so significant that several worldwide organizations including the American Medical Association and the World Obesity Federation have identified obesity as a disease [2, 3]. We discuss the evidence behind successful weight management strategies including dietary approaches, the role of physical activity and effective use of behaviour modification, and show how they can be supported by the use of pharmacotherapy.

Energy Balance and Energy Requirements

Imbalance between energy derived from intake of food and expenditure of energy to maintain body physiology and perform physical work account for changes in the body's energy stores, primarily composed of fat tissue. Factors contributing to energy imbalance include genetic predisposition, maternal obesity, environmental, and cultural influences. Weight gain is associated with an increase in energy expenditure, due to an associated increase in the metabolically active component of the body (i.e. fat-free mass). Conversely, weight loss leads to a decrease in energy expenditure, which may partly explain why weight loss tends to plateau in the context of ongoing dietary restriction. Manipulation of dietary quantity or composition may alter overall energy intake and/or energy expenditure with a corresponding change in fat-free mass and subsequently alter the energy stores of the body [4].

Up to 60–70% of an individual's total daily energy expenditure is accounted for by the resting energy expenditure (REE) required to maintain normal body physiology. REE varies according to age, sex, pubertal growth, ethnicity, pregnancy, lactation, and physical activity. The British Nutrition Foundation has established the average estimated dietary reference values according to age groups, based on the average healthy population in the United Kingdom [5] (Table 14.1.4.1).

Estimated Energy Requirement = REE × Physical Activity Level (PAL)

The Physical Activity Level (PAL) is estimated using a list of the physical activities undertaken within a 24-hour period. PAL is determined based upon the impact of basal activity on energy expenditure (a factor of 1.1) and the sum of all activities (sum of Δ PAL). A third of obese individuals (BMI > 30 kg/m^2) would fit into a description of an inactive person, defined as overall PAL of less than 1.4. Estimated values of PAL are listed in Table 14.1.4.2.

Overall, physical activity levels in the United Kingdom are low; the amount of time spent on any physical activity is on average 6.5 hours for men and 5.5 hours for women per week [6]. One in five individuals in UK spend less than 30 minutes undertaking moderate physical activity per week. Thirteen per cent (13%) of all UK men and women report being sedentary for longer than eight and a half hours per day [6]. The American College of Sports Medicine (ACSM) and the World Obesity Federation recommended that 150–250 min/week of moderate to vigorous physical activity with an energy equivalent of 1200–2000 kcal/week prevents a weight gain of greater than 3% [7, 8].

Table 14.1.4.1 UK Energy requirements based on the average energy required for people with a healthy weight who are moderately active. British Nutrition Foundation (2016)

	Males		Females	
Age (years)	MJ/d	Kcal	MJ/d	kcal
19–24	11.6	2772	9.1	2175
25–34	11.5	2749	9.1	2175
35–44	11.0	2629	8.8	2103
45–54	10.8	2581	8.8	2103
55–64	10.8	2581	8.7	2079
65–74	9.8	2342	7.7	1912
75+	9.6	2294	8.7	1840

Reproduced with permission from the British Nutrition Foundation (2017). Available from: https://www.nutrition.org.uk/attachments/article/907/Nutrition%20Requirements_Revised%20Oct%202017.pdf [Accessed on 3 July 2019]. Copyright 2017 © British Nutrition Foundation. Data obtained from Scientific Advisory Committee on Nutrition. Dietary Reference Values for Energy (2011). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/339317/SACN_Dietary_Reference_Values_for_Energy.pdf [Accessed 5 July 2019].

Dietary Interventions

Reducing energy intake below the requirement for weight maintenance leads, by definition, to weight loss. The most commonly recommended diets advise reduction of energy intake by about 500–600 kcal from estimated daily requirements and tend to recommend reduction in dietary fat [9]. There is controversy as to whether low-fat or low-carbohydrate diets are superior for weight loss. The positive energy balance due to increased intake of refined carbohydrate [10] is hypothesized to be a consequence of insulin-driven shift in fat partitioning towards storage and away from oxidation by metabolically active tissues such as heart, muscle, and liver. The POUNDS Lost trial [11] included over 800 obese participants who were placed on similar recommended energy restriction with either a low-carbohydrate or low-fat diet over 8 years. Prescribing very low carbohydrate diets appears to offer short-term benefit for weight loss [12], although the mechanisms are uncertain. Such diets may suppress appetite by promoting a rise in circulating ketones and may also be higher in protein which increases satiety, decreases overall energy intake, and increases energy expenditure [13].

Very-low-calorie diets (VLCD) restrict energy to 800 kcal/day or less. Low-calorie diets (LCD) have a deficit of between 800 and 1200 kcal/day. Anderson and colleagues [14] analysed 29 long-term observational studies using VLCDs compared to hypocaloric balanced diet. Negative deficit in energy was observed to produce rapid weight loss of between 15% and 25% of initial body weight within 3 to 6 months of treatment with VLCD or LCD diet and may maintain 8–9% one year after treatment [14]. Greater weight loss of up to 10–16% initial body weight was observed with VLCD, although it was not sustainable and gradual weight regain occurred within 2 years [15]. Additionally, during the weight loss maintenance phase, dietary adherence with meal replacements and prolonged refeeding phase improved overall weight loss maintenance by 3–5% at 12 months [16]. The recommendation for close medical supervision in VLCD and LCD arises from health concerns over theoretical loss of lean tissue mass, and higher incidence of gallstones [17].

Intermittent energy restriction entails a combination of regular food consumption alternating with days of caloric restriction, such as seen in the 5:2 diet. There is a suggestion that this could be as effective as continuous energy restriction, but data is limited by low sample size and lack of long-term follow-up [18]. Reducing calorie intake across different diets often is a result of secondary behaviour change with imposed dietary restrictions, which alter habitual eating habits. Over time weight regain mainly occurs due to dietary non-adherence, but also in part, explained by the individual finding

Table 14.1.4.2 Examples of low to intense physical activity. MET (metabolic equivalent) is the ratio of activity metabolic rate to resting metabolic rate. One MET is expressed as 1 kcal/kg/hour

Low intensity physical activity (approx. 1.4–2.0 METs) Light activity without exertion of pulse rate or breathing	Moderate intensity physical activity (approx. 3–6 METs) Moderate effort with acceleration of pulse rate	Heavy intensity physical activity (approx. >6 METs) Heavy intense effort accompanied by rapid pulse rate and breathing
• Sitting on couch • Hospital stay in bed • Office-based job at desk • Short-distance stroll	• Domestic and household chores • Washing and drying clothes • Cleaning within the house • Gardening • Carrying light loads • Recreation sport (swimming, bicycling at regular pace) • Physical work in agriculture or construction	• Continuous running • Climbing or hiking up a hill • Fast-paced cycling • Fast-paced swimming • Competitive athletic sports (e.g. tennis, badminton, hockey, basketball) • Carrying heaving loads (>30 kg)

alternative ways to satisfy their food cravings within the context of diet restriction.

Lean *et al.* [19] conducted a cluster-randomized trial in Scottish and North East England primary care centres in people with type 2 diabetes using total meal replacement (825–853 kcal/d) for 3–5 months, stepped food re-introduction (2–8 weeks) and structured dietary support and compared this to best practice care guidelines. After 12 months, 36 participants (24%) in the intervention group achieved >15% weight loss compared to none in the control group, and the importance of this was substantiated by the 68 participants (46%) who showed remission from type 2 diabetes. The 2-year results showed sustained remissions at 24 months for more than a third of patients with type 2 diabetes [20].

The Role of Physical Activity

Evidence from randomized controlled trials (RCTs) has demonstrated that structured exercise programmes ranging from 1000 to 4000 kcal/week energy expenditure without energy restriction is associated with reductions in visceral and abdominal adiposity [21]. A positive relationship between the amount of exercise (energy expenditure) and reduction in the components of abdominal obesity (waist circumference, visceral adipose tissue, and abdominal subcutaneous abdominal tissue) is observed across several studies varying in exercise-induced energy expenditure [22]. Overall there appears to be beneficial effects of aerobic exercise on blood glucose and insulin measures in obese individuals, which is more pronounced in those with impaired glucose tolerance and type 2 diabetes [23].

Behavioural Elements and Modification

Overweight and obese participants enrolled into intensive behavioural programmes typically achieve weight losses of between 8% and 10% [24]. The Australian Bureau of Statistics projected that the daily intake of an 'average person' comprises of one-third discretionary foods [25] deemed as nutrient poor and/or with high levels of sugar sweetened beverages. Behaviour modification for weight management is the core intervention for the treatment of obesity. Commonly applied strategies include establishing baseline dietary patterns [26] to facilitate realistic weight loss goals and becoming educated about the nutrients required and reducing energy dense foods and/or reducing portion sizes. Clinicians must be mindful of patients' or parents' attitude towards receiving information. Integrating motivational interview during the consultation encourages positivity and promotes self-reflection which is particularly useful to challenge the stigma associated with being overweight. The element of success in behavioural therapy is associated with lower levels of depression and disinhibition [27].

Weight Maintenance

While it is difficult to lose weight, maintaining weight loss is even more challenging. The energy deficit triggers a decrease in energy expenditure, a rise in hunger signals and reduction of satiety hormones, causing increased hunger and food intake. Formerly obese individuals are more likely to have low REE values after adjusting for fat-free mass, leading to the hypothesis that calorie restriction may only be sustained in a small proportion of overweight or obese individuals. Human studies on effects of hypocaloric diets on REE, demonstrated that REE could be significantly reduced by 10–15%. Furthermore, altered REE was sustained up to 2 months after achieving the new energy balance [28]. We are yet to understand how these effects change in the long-term or whether or not they are sustained.

The National Weight Control Registry (NWCR) consists of over 10 000 participants in the United States of people who have lost at least 30 lbs (13.6 kg) and maintained the weight loss for at least 1 year [29]. Over the 10-year longitudinal follow-up key, predictors of success include food intake modification, increased physical activity [30], regular portion meals including daily breakfast and decrease in leisure-time activity (watching less than 10 hours of TV per week). NWCR consists of 2886 participants, of which 78% women, reported mean weight loss of 23.8 kg (95% CI = 23.2, 24.4) at 5 years and 23.1 kg (95% CI = 22.3, 23.9) at 10 years [29]. However, it is important not to generalize these results to the overweight and obese population as this is not representative of a random sample. Maintainers of long-term weight loss were more likely to adhere with the recommended calorie restriction, reported lower fat content in diet and cultivated exercise strategies to maintain high levels of physical activity (~1 h/d). Low mood, depression and disinhibition influenced the likelihood of obese participants to comfort eat and significantly impacted upon their self-confidence [31].

Lifestyle Changes to Support Weight Management in Prediabetes and Type 2 Diabetes

The US Diabetes Prevention Programme studied the effects of intensive lifestyle intervention in over 3200 participants with impaired glucose tolerance. Intensive lifestyle intervention included 150 minutes of moderately-intense physical activity per week with an energy restriction of 1200–1500 kcals per day [32]. Furthermore, the lifestyle group received up to 16 behavioural modification sessions over 24 weeks. After an average of 2.8 years' follow-up, participants in the lifestyle group lost significantly more weight than the metformin and placebo groups (lifestyle = –5.6 kg; metformin = –2.1 kg; placebo = –0.01 kg). The incidence of diabetes was 58% lower in the lifestyle group, and 31% lower with metformin when compared to placebo. Similar intensive lifestyle intervention studies conducted in Finland, China, and India have largely supported these findings [33–35], and many countries, including the United Kingdom, have implemented programmes to support lifestyle changes in people at high risk of developing diabetes.

The Look AHEAD (Action for Health in Diabetes) is the largest RCT focused on intensive lifestyle changes through energy restriction (1200–1800 kcal/d), behaviour modification and increased physical activity (energy consumption) in 5145 overweight and obese individuals with type 2 diabetes [36] over a median follow-up of 9.6 years. Interestingly only 830 participants achieved ≥10% weight loss after one year, of those, 324 participants (39%) managed to maintain this weight loss after 8 years. Key predictors of successful maintenance of weight loss at 1, 4, and 8 years for participants include greater number of behavioural sessions attended, number of meal replacements used, and weekly minutes spent on physical activity [37].

Diet and lifestyle interventions are considered core components in management of obesity, however they may not effect meaningful long-term weight change [38]. As most patients may find it difficult to maintain lifestyle changes in the long-term, typically showing reduced rate of weight loss, weight plateau, and thereafter weight regain, it is worth considering the prospect of combining other strategies to support weight loss and weight maintenance, including the use of pharmacotherapy.

Pharmacotherapy for Obesity

Despite the option of bariatric surgery, it is not an appropriate intervention for most people with obesity due to high costs, limited availability, and risk of perioperative and long-term complications. Antiobesity pharmacotherapy is licensed only for patients with BMI ≥30 kg/m² or BMI ≥27 kg/m² in the presence of metabolic risk factors like diabetes. Used as an adjunct to lifestyle intervention, pharmacotherapy use as a treatment modality could serve a useful tool to help achieve greater weight loss and promote long-term weight maintenance.

There has been longstanding research interest in development of drugs to promote long-term weight loss. Theoretical approaches include inhibition of food intake centrally, either by inhibiting appetite stimulatory pathways, or enhancing satiety signals (some of which may originate in the GI tract), inhibiting absorption of food, and/or increasing energy expenditure. Despite recent advances, there remains concern of the long-term use of drugs to treat obesity, and many apparently good concepts were eventually found to be ineffective or to have significant adverse effects. Early attempts included use of amphetamine and its derivatives and mitochondrial uncoupling agents such as dinitrophenol, which is highly toxic. Fenfluramine/dexfenfluramine broadly target 5HT receptors, leading to concerns over cardiac valvulopathy. Rimonabant, a cannabinoid receptor antagonist, was withdrawn in Europe and never marketed in the United States, due to risk of increased suicidal ideation and low mood. These examples of antiobesity pharmacotherapy serve to remind us of the need for robust, long-term data on safety of all antiobesity agents [39].

Currently only four drugs (phentermine/topiramate, naltrexone/bupropion, orlistat and liraglutide) are US FDA-approved for use in chronic weight management; only orlistat, naltrexone/bupropion and liraglutide are approved in Europe. Four other centrally acting agents are limited to short-term use. Further advances in our understanding of appetite regulatory hormones and complex neurobiology interactions have led to much interest but much pessimism remains on the use of long-term pharmacotherapy to treat obesity.

Pancreatic Lipase Inhibitors

Orlistat

Orlistat is an intestinal lipase inhibitor chemically synthesized from lipstatin, a natural product isolated from the fungus *Streptomyces toxytricini.* It covalently binds to pancreatic lipase enzyme inhibiting absorption of approximately 30% of dietary fat. The prescribed dose is 120 mg three times daily with meals. In a multicentre randomized controlled trial, the orlistat group on average lost more than placebo group (10.2% [10.3 kg] vs. 6.1% [6.1 kg]); a weight loss difference of 3.9 kg at the end of 1 year [40]. Participants who continued with orlistat for a second year regained only half as much weight as compared to those crossed over to placebo arm. The XENDOS trial, a 4-year double-blind prospective study on orlistat use plus lifestyle changes found a reduced incidence of diabetes (6.2%) for participants taking orlistat, compared to placebo (9%) at 4 years [41]. Mean weight loss after 4 years was also significantly greater with orlistat (5.8 vs. 3.0 kg). Steatorrhoea, abdominal cramps, and flatulence limit its acceptability to many patients. Reduced absorption of fat-soluble vitamins was observed in clinical trials, but participants were asymptomatic and severe deficiencies do not develop.

Centrally Acting Antiobesity Drugs

Phentermine

Phentermine hydrochloride is an amphetamine-derived appetite suppressant, approved for short-term (up to 3 months) use by the FDA since 1977. It acts on the central nervous system to decrease hunger [42]. Phentermine monotherapy (30 mg daily) remains the most commonly prescribed therapy by US bariatric physicians. It

is a controlled drug available only on prescription for short-term (3 months) use in Europe. The attrition rates in controlled trials and presentation of data for study completers only led to scarcity of long-term data. Among the completers, commonest side effects were transient insomnia, irritability, and anxiety. Clinical use has not been associated with phentermine cravings, withdrawal, or abuse leading to psychological or physical impairment. Clinicians should not use phentermine in patients with coronary artery disease, stroke, arrhythmias, congestive cardiac failure, and uncontrolled hypertension [43].

Phentermine/Topiramate

The combination of low dose, controlled-release phentermine and topiramate (7.5/46 mg or 15/92 mg) are FDA-approved as an adjunct to hypocaloric diet. The effects of weight loss observed from use of topiramate as an anticonvulsant promoted its potential as a weight-loss agent. Combined with phentermine, clinical data demonstrates impact on reduced appetite, proposed to stem from centrally driven neurotransmitter pathways causing appetite suppression and improved satiety. Phentermine/topiramate 15/92mg ER use was associated with mean weight loss of 9.3% from baseline at 56 weeks [44]. Greater proportions of participants on the PHEN/TPM CR achieved clinically significant weight loss of 5% (79% vs. 30% in placebo) and 10% (53.9% vs. 11.5%) and 15% (31.9% vs. 6.6%) [45]. Although lowering doses of both drugs helps minimize adverse effects compared to doses needed for weight loss as monotherapy, adverse effects of dizziness, paraesthesiae, and dry mouth are common. This combination was not approved in Europe due to concerns over teratogenicity with topiramate, and ongoing concerns over the cardiac and neuropsychiatric safety of phentermine.

Lorcaserin

Lorcaserin is a selective $5HT2_C$ receptor agonist, promoting hypophagia and improved satiety. Activity at other serotonin receptors, including the 2B heart valve receptors, is minimal. Lorcaserin 10 mg twice daily is FDA-approved for chronic weight management adult obesity (BMI ≥ 30 kg/m² or ≥ 27 kg/m² in the presence of ≥ 1 weight-related comorbidity). Its combination with lifestyle intervention for obese, non-diabetic patients, 47.5% achieved ≥ 5% weight loss vs. 20.3% with placebo [46]. In a pooled-analysis of lorcaserin-treated obese participants, week 12 early responders from the BLOOM and BLOSSOM studies were observed to demonstrate better weight loss (10.8% weight loss of initial baseline weight at week 52) as compared to 2.7% for week 12 non-responders [46, 47]. Despite previous controversy around association of fenfluramine/dexfenfluramine with cardiac valvulopathy [48], no reports of valvulopathy were demonstrated with lorcaserin use. The 5-year post-marketing cardiovascular outcome trial provided reassurance to clinicians that long-term use of lorcaserin was not associated with any major cardiovascular events [49].

Naltrexone SR/Bupropion SR

Bupropion is a weak inhibitor of dopamine and norepinephrine reuptake. Naltrexone, an opioid antagonist has minimum effect on weight loss on its own, but naltrexone is thought to block the inhibitory effects of opioid receptors at the hypothalamic proopiomelanocortin (POMC) neurons, stimulating α-melanocyte stimulating hormone (α-MSH) release. α-MSH initiates a cascade, leading to reduced food intake. Forty-nine percent of participants treated with naltrexone/bupropion achieved 5% weight loss at 1 year, while 25% achieved 10% weight loss [50]. Mean reduction in body weight was −1.3% in placebo, −6.1% in naltrexone 32 mg plus bupropion group, and −5.0% in the naltrexone 16 mg plus bupropion group at 56 weeks [51, 52]. Despite the label containing a warning of suicidal ideation, an FDA requirement for antidepressants, clinical use have not demonstrated any serious concerns of increased depression or suicidality as compared to placebo [51, 52]. Most common adverse effects include nausea, constipation, diarrhoea, and headache. Approximately 1 in 4 patients discontinued naltrexone/bupropion as a result [52]. Its use is contraindicated in uncontrolled hypertension, opioid dependence, and pregnancy.

Liraglutide

Glucagon-like peptide (GLP)-1, released from intestinal L cells following a consumption of a meal, circulates naturally as $GLP\text{-}1_{7\text{-}36\ amide}$. Up to 50–70% of insulin response to a meal is due to the incretin effect, which denotes the much greater insulin secretion noted on oral glucose challenge compared to a parenteral isoglycaemic glucose administration. Among other therapeutic benefits, GLP-1 inhibits glucagon secretion from alpha cells, reduces gastric emptying and improves satiety, all of which improve glycaemia.

GLP-1 is also synthesized by neurones in the nucleus tractus solitarius (NTS) of the caudal brainstem. The NTS projects to hypothalamic and hindbrain neurones expressing GLP-1 receptors. In animal studies, intracerebroventricular GLP-1 inhibits food intake, while peripheral administration leads to increased uptake and activity within the mid and hindbrain areas involved in feeding regulation [53]. GLP-1 receptors are thus accessible both to GLP-1 as gut hormone peripherally and GLP-1 as neurotransmitter centrally, potentially explaining its appetite-suppressant satiety-inducing effects. Effects of GLP-1 are shown in **Figure 14.1.4.1**.

Liraglutide is a GLP-1 receptor agonist consisting of a polypeptide backbone structure derived from human GLP-1, following modification of the sequence of amino acids and addition of fatty acid chain to protect against rapid degradation by native dipeptidyl peptidase-4. The biochemical change renders liraglutide more effective as a GLP-1 agonist with higher affinity for albumin binding compared to its native predecessor. From 2014, liraglutide has been approved for use in chronic weight management by the FDA and EMA at a dose of 3.0 mg [54].

In SCALE obesity and prediabetes [55], 3731 patients with BMI >30 kg/m² or BMI >27 kg/m² with comorbidities were randomized into liraglutide 3.0 mg or placebo with diet, exercise, and lifestyle modification (54, 55). Intervention group demonstrated a weight loss of 8.4 kg +/− 7.3 kg; as compared to placebo (2.8 kg +/− 6.5 kg; P <0.001). Lorcaserin was withdrawn in early 2020 after concerns about increased cancer risk. Up to two-thirds (63.2%) of patients reported weight loss ≥5%, as compared to the 27.1% in placebo group; and 33.1% and 10.6%, respectively, reported weight loss ≥10% [55]. Liraglutide 3.0 mg used as an adjunct to diet and exercise, showed maintenance of the weight loss [54] (81.4% maintained the ≥5% run-in weight loss) as compared to participants on placebo (48.9%)

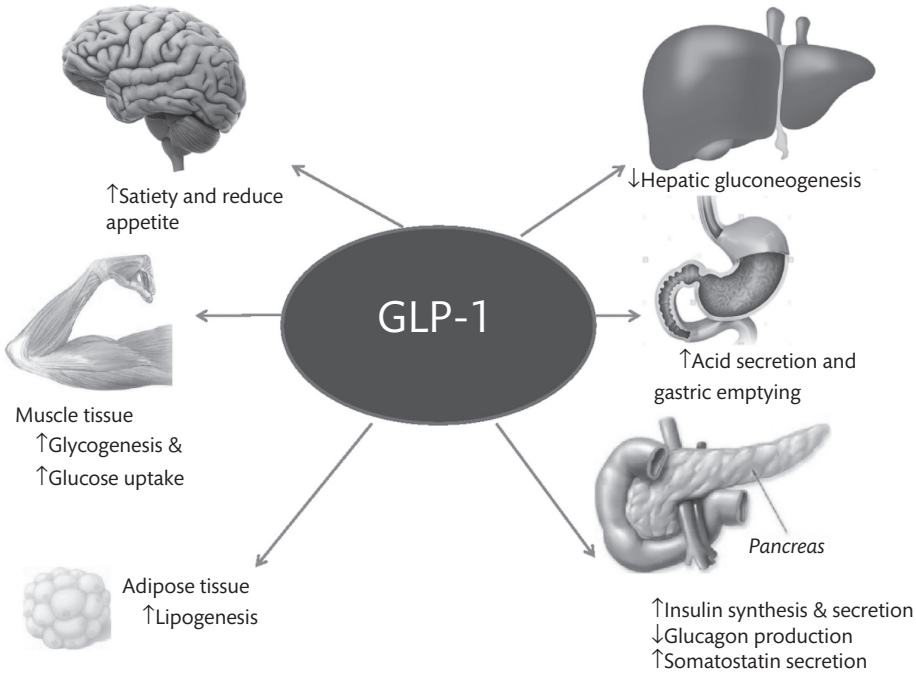


Figure 14.1.4.1 Summary of effects of GLP-1.

[54] over 56 weeks. Predominantly, participants report a mixture of gastrointestinal adverse effects of nausea, stomach cramps, flatus, diarrhoea, constipation, and fatigue. Despite these adverse effects, attrition rate associated with liraglutide 3.0 mg use is similar to that of phentermine/topiramate and naltrexone/bupropion, but slightly higher compared to lorcaserin and orlistat [56]. Reassuringly, post-hoc analysis did not demonstrate any firm association between liraglutide and pancreatitis [57].

Future Medical Therapies in Chronic Weight Management

Semaglutide is another GLP-1 receptor agonist for which phase II trial [58] results demonstrate that 83% of participants on the highest dose tested (0.4 mg once daily), attained ≥5% weight loss and 65% achieved ≥10% weight loss. The most common reported adverse event was dose-dependent nausea, similar to that of other GLP-1 agonists. Phase 3 trials using a weekly dosing regimen, including a large cardiovascular outcomes trial, are ongoing.

Another potential antiobesity agent in development is setmelanotide, a high potency melanocortin-4 receptor (MCR4) agonist, that is effective in obese individuals with pro-opiomelanocortin (POMC) deficiency [59]. This rare genetic disorder usually results from heterozygous defects within the MC4 receptor, which drives hyperphagia in infancy and lifelong severe obesity. Setmelanotide has also entered phase III trials for use in leptin receptor deficiency and Bardet-Biedl syndrome. Leptin replacement therapy (metreleptin, Myalept) is approved only for use in one extreme form of monogenic obesity, namely congenital leptin deficiency (CLD) [60].

Conclusion

Cumulative epidemiological, observational, and randomized clinical trial evidence has demonstrated that maintenance of moderate energy deficit of 500–600 kcal/d leads to 3–5% sustainable weight loss. In principle, lifestyle interventions in the form of dieting and/or exercise are the core tool in tipping energy balance towards a net deficit; use of meal replacements with a robust behavioural support package may result in greater weight loss, but longer-term maintenance remains difficult. Unsurprisingly, the key predictors of success in weight loss and weight maintenance are dietary adherence, maintenance of an increased energy expenditure through physical activity, and frequent attendance and participation in individual or group behavioural sessions. Given that many people find it difficult to maintain these crucial lifestyle changes, due to physiological responses to weight loss, including increased hunger and lower energy expenditure, physicians now have the option to consider the benefits and risks of introducing adjunct pharmacotherapy to induce and further support maintenance of weight loss. As our understanding of the complex neurobiology and role of different neurohormonal or enteroendocrine classes of antiobesity medications evolve, shared decision making with patients leading to individualized and personalized care would help bridge the gap of managing obesity. Pharmacotherapy use requires further long-term

data to support efficacy and more importantly patient safety. Future therapies may include new drug combinations designed to maintain efficacy while considering tolerability and safety of potential long-term adverse effects.

REFERENCES

1. Hunter GR, Gower BA, Kane BL. Age related shift in visceral fat. *Int J Body Compos Res*, 2010; **8**: 103–8.
2. Allison DB, Downey M, Atkinson RL, *et al*. Obesity as a disease: a white paper on evidence and arguments commissioned by the council of the Obesity Society. *Obesity*, 2008; **16**: 1161–77.
3. Bray GA, Kim KK, Wilding JPH. Obesity: a chronic relapsing progressive disease process. A position statement of the World Obesity Federation. *Obes Rev*, 2017; **18**: 715–23.
4. Muller MJ, Enderle J, Bosy-Westphal A. Changes in energy expenditure with weight gain and weight loss in humans. *Curr Obes Rep*, 2016; **5**: 413–23.
5. British Nutrition Foundation. Nutrition Requirements. Available at: https://www.nutrition.org.uk/nutritionscience/nutrients-food-and-ingredients/nutrient-requirements.html (accessed: 15 April 2020).
6. Townsend N, Wickramasinghe K, Williams J, Bhavnagar P, Rayner M. *Physical Activity Statistics 2015*. London: British Heart Foundation, 2015.
7. Donnelly JE, Blair SN, Jakicic JM, Manore MM, Rankin JW, Smith BK. American College of Sports Medicine Position Stand. Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults. *Med Sci Sports Exerc*, 2009; **41**: 459–71.
8. Office of Disease Prevention and Health Promotion. *The Physical Activity Guidelines for Americans*, 2nd edn, 2019. Available at: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf (accessed: 15 April 2020).
9. Guideline Development Group Obesity Partial update of CG43. Identification, assessment and management of overweight and obesity in children, young people and adults. National Institute for Health and Care Excellence Nov, 2004. Available at: https://www.nice.org.uk/guidance/cg43 (accessed: 15 April 2020).
10. Hu T, Yao L, Reynolds K, *et al*. The effects of a low-carbohydrate diet vs. a low-fat diet on novel cardiovascular risk factors: a randomized controlled trial. *Nutrients*, 2015; **7**: 7978–94.
11. Sacks FM, Bray GA, Carey VJ, *et al*. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med*, 2009; **360**: 859–73.
12. Bueno NB, de Melo IS, de Oliveira SL, da Rocha Ataide T. Very-low-carbohydrate ketogenic diet v. low-fat diet for long-term weight loss: a meta-analysis of randomised controlled trials. *Br J Nutr*, 2013; **110**: 1178–87.
13. Astrup A, Westerterp-Plantenga MS, Luscombe-Marsh ND, *et al*. The role of protein in weight loss and maintenance. *Am J Clin Nutr*, 2015; **101**: 1320S–9S.
14. Anderson JW, Hamilton CC, Brinkman-Kaplan V. Benefits and risks of an intensive very-low-calorie diet program for severe obesity. *Am J Gastroenterol*, 1992; **87**: 6–15.
15. Tsai AG, Wadden TA. The evolution of very-low-calorie diets: an update and meta-analysis. *Obesity*, 2006: **14**: 1283–93.
16. Johansson K, Neovius M, Hemmingsson E. Effects of anti-obesity drugs, diet, and exercise on weight-loss maintenance after a very-low-calorie diet or low-calorie diet: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr*, 2014; **99**: 14–23.
17. Johansson K, Sundström J, Marcus C, Hemmingsson E, Neovius M. Risk of symptomatic gallstones and cholecystectomy after a very-low-calorie diet or low-calorie diet in a commercial weight loss program: 1-year matched cohort study. *Int J Obes (Lond)*, 2014; **38**: 279–84.
18. Harvie MN, Pegington M, Mattson MP, *et al*. The effects of intermittent or continuous energy restriction on weight loss and metabolic disease risk markers: a randomized trial in young overweight women. *Int J Obes (Lond)*, 2011; **35**: 714–27.
19. Lean MEJ, Leslie WS, Barnes AC, *et al*. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet*, 2018; **391**: 541–51.
20. Lean MEJ, Leslie WS, Barnes AC, *et al*. Durability of primary care led weight management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster randomised trial. *Lancet Diabetes Endocrinol*, 2019; **7**: 344–55.
21. Ramage S, Farmer A, Eccles KA, McCargar L. Healthy strategies for successful weight loss and weight maintenance: a systematic review. *Appl Physiol Nutr Metab*, 2014; **39**: 1–20.
22. Anderson AS, Craigie AM, Caswell S, *et al*. The impact of a bodyweight and physical activity intervention (BeWEL) initiated through a national colorectal cancer screening programme: randomised controlled trial. *BMJ*, 2014; **348**: g1823.
23. Chudyk A, Petrella RJ. Effects of exercise on cardiovascular risk factors in type 2 diabetes: a meta-analysis. *Diabetes Care*, 2011; **34**: 1228–37.
24. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. *Psychiatr Clin North Am*, 2011; **34**: 841–59.
25. Australian Bureau of Statistics, 2013. *Australian Health Survey*: Updated Results 2011–2012. cat. no. 4364.0.55.003.
26. Grafenauer SJ, Tapsell LC, Beck EJ, Batterham MJ. Baseline dietary patterns are a significant consideration in correcting dietary exposure for weight loss. *Eur J Clin Nutr*, 2013; **67**: 330.
27. Kelley CP, Sbrocco G, Sbrocco T. Behavioral modification for the management of obesity. *Primary Care,* 2016; **43**: 159–75.
28. Santos I, Sniehotta FF, Marques MM, Carraça EV, Teixeira PJ. Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. *Obes Rev*, 2016; **18**: 32–50.
29. The National Weight Control Registry. Available at: http://www.nwcr.ws/ (accessed: 10 June 2018).
30. Hill JO, Wyatt H, Phelan S, Wing R. The national weight control registry: is it useful in helping deal with our obesity epidemic? *J Nutr Educ Behav*, 2005; **37**: 206–10.
31. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. *J Am Diet Assoc*, 2011; **111**: 92–102.
32. Diabetes Prevention Program Research G. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*, 2009; **374**: 1677–86.
33. Li G, Zhang P, Wang J, *et al*. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet,* 2008: **371**: 1783–9.
34. Lindstrom J, Peltonen M, Eriksson JG, *et al*. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). *Diabetologia*, 2013; **56**: 284–93.
35. Ramachandran A, Snehalatha C, Yamuna A, Mary S, Ping Z. Cost-effectiveness of the interventions in the primary prevention

of diabetes among Asian Indians: within-trial results of the Indian Diabetes Prevention Programme (IDPP). *Diabetes Care*, 2007; **30**: 2548–52.
36. Look AHEAD Research Group. Eight-year weight losses with an intensive lifestyle intervention: the Look AHEAD study. *Obesity,* 2013; **22**: 5–13.
37. Ryan D, Heaner M. Guidelines (2013) for managing overweight and obesity in adults. Preface to the full report. Obesity *(Silver Spring)*, 2014; **22** Suppl 2: S1–3.
38. Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. *Med Clin North Am*, 2018; **102**: 183–97.
39. Gadde KM, Apolzan JW, Berthoud HR. Pharmacotherapy for patients with obesity. *Clin Chem*, 2018; **64**: 118–29.
40. Sjöström L, Rissanen A, Andersen T, Boldrin M, Golay A, Koppeschaar HPF, Krempf M. Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. *Lancet*, 1998; **352**: 167–72.
41. Torgerson JS, Hauptman J, Boldrin MN, Sjöström L. XENical in the prevention of diabetes in obese subjects (XENDOS) study: a randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care*, 2004; **27**: 155–61.
42. Alexander M, Rothman Richard B, Baumann Michael H, Endres Christopher J, Brašić James R, Wong Dean F. Noradrenergic and dopaminergic effects of (+)-amphetamine-like stimulants in the Baboon *Papio anubis*. *Synapse*, 2005; **56**: 94–9.
43. FDA. Phentermine Hydrochloride Safety and Efficacy Prescribing Information. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/202088s005lbl.pdf (accessed: 16 April 2020).
44. Allison DB, Gadde KM, Garvey WT, *et al.* Controlled-release phentermine/topiramate in severely obese adults: a randomized controlled trial (EQUIP). *Obesity (Silver Spring)*, 2012; **20**: 330–42.
45. Garvey WT, Ryan DH, Look M, *et al.* Two-year sustained weight loss and metabolic benefits with controlled-release phentermine/topiramate in obese and overweight adults (SEQUEL): a randomized, placebo-controlled, phase 3 extension study. *Am J Clin Nutr*, 2012; **95**: 297–308.
46. Smith SR, Weissman NJ, Anderson CM, *et al.* Multicenter, placebo-controlled trial of lorcaserin for weight management. *N Engl J Med*, 2010; **363**: 245–56.
47. Fidler MC, Sanchez M, Raether B, *et al.* A one-year randomized trial of lorcaserin for weight loss in obese and overweight adults: the BLOSSOM trial. *J Clin Endocrinol Metab*, 2011; **96**: 3067–77.
48. Centres for Disease Control and Prevention (CDC). Cardiac valvulopathy associated with exposure to fenfluramine or dexfenfluramine: U.S. Department of Health and Human Services interim public health recommendations. *MMWR Morb Mortal Wkly Rep*, 1997; **46**: 1061–6.
49. Bohula EA, Wiviott SD, McGuire DK, *et al.* Cardiovascular safety of lorcaserin in overweight or obese patients. *N Engl J Med*, 2018; **379**: 1107–17.
50. Sherman MM, Ungureanu S, Rey JA. Naltrexone/bupropion ER (contrave): newly approved treatment option for chronic weight management in obese adults. *Pharmacy and Therapeutics*, 2016; **41**: 164–72.
51. Greenway FL, Fujioka K, Plodkowski RA, *et al.* Effect of naltrexone plus bupropion on weight loss in overweight and obese adults (COR-I): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet,* 2010; **376**: 595–605.
52. Apovian CM, Aronne L, Rubino D, Still C, Wyatt H, Burns C, Kim D, Dunayevich E. A randomized, phase 3 trial of naltrexone SR/bupropion SR on weight and obesity-related risk factors (COR-II). *Obesity (Silver Spring),* 2013; **21**: 935–43.
53. Kanoski SE, Hayes MR, Skibicka KP. GLP-1 and weight loss: unraveling the diverse neural circuitry. *Am J Physiol*, 2016; **310**: R885–R95.
54. Wadden TA, Hollander P, Klein S, Niswender K, Woo V, Hale PM, Aronne L. Weight maintenance and additional weight loss with liraglutide after low-calorie-diet-induced weight loss: the SCALE Maintenance randomized study. *Int J Obes (Lond),* 2013; **37**: 1443–51.
55. Pi-Sunyer X, Astrup A, Fujioka K, *et al.* A randomized, controlled trial of 3.0 mg of liraglutide in weight management. *N Engl J Med*, 2015; **373**: 11–22.
56. Khera R, Murad MH, Chandar AK, *et al.* Association of pharmacological treatments for obesity with weight loss and adverse events: a systematic review and meta-analysis. *JAMA*, 2016; **315**: 2424–34.
57. Jensen TM, Saha K, Steinberg WM. Is there a link between liraglutide and pancreatitis? A post hoc review of pooled and patient-level data from completed liraglutide type 2 diabetes clinical trials. *Diabetes Care*, 2015; **38**: 1058.
58. O'Neil PM, Birkenfeld AL, McGowan B, *et al.* A randomised, phase II, placebo—and active-controlled dose-ranging study of semaglutide for treatment of obesity in subjects without diabetes. Presented at ENDO, 2018 (OR12). Presented at ENDO, 2018 (OR12). Chicago, USA, 17–20 March 2018.
59. Collet T-H, Dubern B, Mokrosinski J, *et al.* Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency. *Mol Metab*, 2017; **6**: 1321–9.
60. Paz-Filho G, Wong ML, Licinio J. Ten years of leptin replacement therapy. *Obes Rev*, 2011; **12**: e315–23.

14.1.5 Metabolic Surgery

Francesco Rubino, Vivian Anastasiou, Luca Ferraro, Dalal Qanaq, and Ghassan Chamseddine

Introduction

Type 2 Diabetes currently affects more than 400 million people worldwide and this number is increasing rapidly. Despite constantly evolving drugs many patients with type 2 diabetes develop micro and macrovascular complications that increase risk of further illness and mortality. In the USA, only about half of the patients with type 2 diabetes achieve recommended levels of HbA_{1c}, and less than 20% achieved combined targets for HbA1c, low-density lipoprotein (LDL) cholesterol, and blood pressure [1]. Developing novel and more effective strategies to understand and treat diabetes is therefore an important priority for medical research.

Recently, 56 medical organizations worldwide have approved guidelines from the 2nd Diabetes Surgery Summit (DSS-II) that recommend consideration of bariatric/metabolic surgery as a standard treatment option for patients with obese type II diabetes [2, 3]. This conclusion is based on both biological and clinical evidence. In fact, in addition to positive effects from weight reduction, studies in both animals and humans have shown that gastrointestinal operations can induce weight-independent effects on glucoregulatory physiology [4, 5]. Furthermore, several randomized clinical trials have shown that surgery achieves better glycaemic control than lifestyle and available pharmaceutical interventions [6], even including long-lasting remission of hyperglycaemia.

Although the exact mechanisms of action of surgery remain unclear, the study of metabolic surgery provides the most compelling evidence so far in support of a critical role of the gastrointestinal system in glucose homeostasis. Further elucidation of the mechanisms of action of surgery and the role of the gut in metabolic regulation may provide clues about the cause of the disease and help identify novel targets for prevention and therapy of diabetes and obesity. To capitalize on the opportunities provided by surgery, however, will require a shift in mindsets across the broad spectrum of care and research, and addressing of misperceptions and misconceptions that still impede appropriate use of this effective therapeutic strategy. This chapter reviews biological, clinical, and economic evidence supporting a surgical approach to type 2 diabetes and discusses current evidence-based indications.

Surgical Treatment of Diabetes: The History So Far

In 1925, a report in *The Lancet* [7] described sudden normalization of glycaemia in a patient with diabetes undergoing a gastrointestinal operation to treat a peptic ulcer. Similar clinical observations documenting dramatic improvement or resolution of diabetes after gastrectomies for ulcer or cancer were reported in ensuing decades and became more common after the advent of bariatric or weight-loss surgery in the mid-1950s. During the 1980s and 1990s, resolution of diabetes was reported after various types of bariatric surgery procedures, including in a landmark report by Pories and coworkers documenting more than 80% resolution of hyperglycaemia in a series of 120 patients undergoing Roux-en-Y Gastric bypass surgery [8].

Experiments in rats conducted in the early 2000s investigated whether a modified form of gastric bypass surgery could directly influence glucose homeostasis. These experiments documented that manipulation of the anatomy of the gastrointestinal tract, similar to those imposed by bariatric surgery procedures, can exert weight-independent effects on glucose homeostasis [5].

Subsequent mechanistic studies in both animal models and humans corroborated these findings, implicating various mechanisms of gastrointestinal (GI) physiology, including changes in gut hormones, bile acids, microbiota, and nutrient sensing signals. Surgical teams from around the world started to use standard bariatric procedures and novel operations in mildly obese and non-obese humans with type 2 diabetes, documenting efficacy of surgery in patients with body mass index (BMI) levels above and below the traditional BMI 35 threshold for indication to traditional bariatric surgery. On the back of these studies, a multidisciplinary group of leading clinicians and scientists convened at the first Diabetes Surgery Summit (DSS-1) in 2007 recognized the biological and clinical rationale for repurposing conventional bariatric operations as a treatment for type 2 diabetes, a practice referred to as 'metabolic surgery', thereby establishing an agenda for research priorities. The DSS-1 inspired further mechanistic studies and randomized clinical trials that now provide the evidence supporting a role of surgery in diabetes. In September 2015, the introduction of surgery into standard care for type 2 diabetes was formally recommended by the participants of the second Diabetes Surgery Summit—DSS-2 [2]. Clinical guidelines established through the DSS-2 are now endorsed by over 56 organizations from worldwide and are consistent with other clinical guidelines for diabetes including NICE guidance, ADA Standards of Care and a recent ADA-EASD consensus algorithm for the treatment of hyperglycaemia [3].

Biological Rationale for Surgical Treatment of Type 2 Diabetes

Studies in humans and rodents show that various gastrointestinal (GI) procedures, and especially gastric bypass or sleeve gastrectomy cause dramatic changes in circulating levels of various gut hormones, including suppression of ghrelin [9], a hunger-inducing hormone of hormones and elevation of glucagon-like peptide-1 (GLP-1) and peptide YY (PYY), which can explain at least in part increased insulin production and decreased appetite. See Figure 14.1.5.1.

Studies show that Roux-en-Y gastric bypass (RYGB) and sleeve gastrectomy can also increase total circulating levels of bile acids as well as change bile acids composition [10], an effect that may contribute to the insulin-sensitizing effects of these operations. In fact, in addition to their role in facilitating digestion, bile acids are re-absorbed into the bloodstream and act as metabolic regulators by signalling through receptors in the nucleus (FXR receptor) or membrane (TGR-5 receptor) of cells of various organs and tissues and participating in the regulation of gut hormones, lipid metabolism, inflammation, and insulin sensitivity [11].

Gastric bypass surgery and duodenal-jejunal bypass (DJB), a stomach-sparing experimental form of metabolic surgery have also been shown to influence intestinal glucose transport and intestinal glucose metabolism in rodents and pigs, thus contributing to the systemic glucose-lowering effects of these procedures [12, 13]. Experiments in rodents also showed that DJB, an experimental,

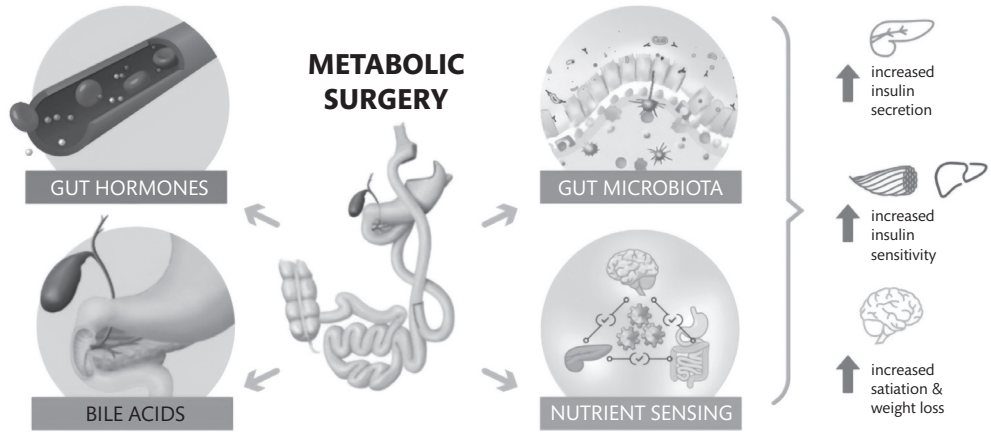


Figure 14.1.5.1 Metabolic surgery changes various mechanisms of GI physiology involved in metabolic regulation. Given its role in metabolic regulation, the GI tract constitutes a clinically and biologically meaningful target for the management of type 2 diabetes.

stomach-sparing form of gastric bypass surgery, enhances nutrient sensing mechanism that signal via the vagus nerve to the brainstem and to the liver to suppress hepatic glucose output [14].

GI surgery alters intestinal nutrient exposure and the acidity of the gut, which in turn can influence the local microbe population. In fact, experiments in humans and rodents have showed that bariatric/metabolic surgery can cause significant changes in gut microbiota. In one experimental study, bacteria extracted from gastric bypass-treated mice were transferred to non-operated, germ-free mice on a high fat diet, causing the recipients to gain less weight compared to matched animals that received similar transplants from sham-operated controls.

Other mechanisms have also been hypothesized. For example, it has been hypothesized that surgery may reduce putative nutrient-stimulated anti-incretin signals from the gut that are responsible for dysregulation of insulin sensitivity and/or secretion [15]. In aggregate, studies to date do not identify the exact mechanism responsible for diabetes remission, however, they provide evidence of a complex and critical role of the gastrointestinal tract in metabolic regulation. On the back of this evidence, in 2016 the DSS-2 recognized the gastrointestinal tract as a legitimate target for antidiabetes interventions.

Clinical Rationale for Surgical Treatment of Type 2 Diabetes

Various gastrointestinal operations originally developed as weight-loss intervention ('bariatric surgery'), can cause dramatic improvement of Type 2 Diabetes and even complete remission of hyperglycaemia, without need for ongoing medical therapy [6]. Remission of diabetes is a durable phenomenon after surgery. While recidivism of diabetes is possible, studies show that the disease recurs in a small subset of patients who have undergone RYGB, after a median disease-free period of 8.3 years [16]. Taking into account the 'legacy effect' of intensive control of diabetes on cardiovascular complications [17], even patients who experience diabetes recurrence after metabolic surgery are likely to attain long-lasting cardiovascular benefits. Although randomized studies with hard endpoints are needed, well matched case-controlled studies show that metabolic surgery diminishes cardiovascular (CV) risk as well as CV events such as myocardial infarction and stroke, cancer and overall mortality [6, 18–21].

Several clinical observations in these studies support the notion that both diabetes and CV effects of surgery are not merely a consequence of weight loss. In fact, diabetes remission appears soon after surgery, well before weight loss occurs. Furthermore, glucose homeostasis is better regulated after metabolic surgery compared to similar weight reduction obtained after changing lifestyle alone [6, 18–20]. Improvement of diabetes, but also amelioration of hard cardiovascular outcomes is not proportional to the amount of weight lost after surgery. Moreover, experimental operations or devices that partially reproduce the anatomy of bariatric/metabolic surgery induce diabetes remission with little or no weight reduction, uncoupling the weight loss effects from effects on glucose regulation. Moreover, uncommon complications of bariatric surgery such as hyperinsulinemia and hypoglycaemia, which typically occur years after surgery, suggest long-lasting enhancement of beta cell function and/or mass postoperatively. This is not an effect achievable by weight reduction alone. In addition, metabolic surgery achieves comparable diabetes remission rates in patients with BMI <30 kg/m^2 and BMI ≥ 35 kg/m^2 preoperatively [22]. Evidence of weight-independent effects on glucose homeostasis after gastrointestinal bypass operations and the role of the GI tract in glucose metabolism provide a rationale to considering a surgical approach to treat type 2 diabetes, even in patients who are only mildly obese [23]. Accordingly, the term 'metabolic' surgery has been coined in recent years to indicate a surgical approach intentionally aimed at treating type 2 diabetes and metabolic disease.

Clinical Outcomes of Metabolic Surgery

Over several decades, clinical practice and research have shown that bariatric/metabolic surgery is an effective and safe procedure for treatment of severe obesity and related diseases. Surgery results in major and sustained weight reduction, as well as improvement of glycaemic and lipid control, cardiovascular risk, and reduced incidence of microangiopathy complications, heart attacks, strokes, cancer, and death compared to patients of similar BMI who do not

undergo surgery [6, 18–21]. Nine long-term observational studies have shown major reduction in overall mortality among subjects with severe obesity undergoing bariatric/metabolic surgery [6, 20, 24, 25], with a remarkable 92% reduction of diabetes-specific related deaths [25].

Furthermore, 12 randomized clinical trials comparing surgical vs. non-surgical diabetes treatments have been recently reported. A DSS-II meta-analysis of such trials (level 1A evidence) revealed remarkable consistency in the results [2, 6] and showed that surgery outperforms conventional diabetes therapies (lifestyle + medications) in important clinical outcomes. In fact, surgically treated patients achieve better glycaemic control and diabetes remission, increased high-density lipoprotein (HDL) and lower triacylglycerol levels, and better quality of life while also dramatically reducing overall drug utilization. Furthermore, these benefits were associated with a low rate of complications and non-surgical related deaths during a follow-up period of 5 years, representing unanimous level 1a evidence. Moreover, great improvements in LDL levels and hypertension were found, although these outcomes were not as profound or consistent as the results for diabetes and its comorbidities, even for the patients with a preoperative BMI <35 kg/m² [2, 6, 26].

Safety of Metabolic Surgery

The development of minimally invasive laparoscopic approaches as the gold standard for the surgical treatment of morbid obesity has dramatically improved safety of bariatric surgery, with a ten-fold decrease in surgical mortality compared to equivalent open-surgery operations [27]. Furthermore, recent evidence has assessed bariatric/metabolic perioperative mortality to be as far lower than other routinely performed operations in general surgery. Mortality rates peribariatric/metabolic surgery are in fact similar to those of operations such as laparoscopic cholecystectomy or hysterectomy, and complication rates are even lower than those observed after other types of major general surgery procedures [28]. Although laparoscopic RYGB is a complex and technically challenging procedure, which requires adequate surgical skills and training, it has a perioperative 3.4% complication rate, which is less than that for laparoscopic operations such as hysterectomy, cholecystectomy, or appendectomy, as reported by a recent US National Registry [28]. Most common long-term complications of bariatric/metabolic surgery are summarized in Table 14.1.5.1. Among these, the most common is iron deficiency. Patients who undergo bariatric/metabolic procedures are usually on lifelong oral iron supplementation and might require repeated iron infusion in cases of severely lowered iron levels or related anaemia. There is also a risk for metabolic bone disease, osteoporosis, and fracture especially after malabsorption-inducing procedures. Moreover, some studies suggested that bariatric/metabolic surgery is associated with an increased relative risk of some unpredictable long-term outcomes, such as suicide, alcohol abuse and accidental injuries [6], although definitive evidence of a causal link to bariatric surgery is not available. These potential adverse events must be carefully weighed against the numerous potential benefits in terms of weight loss, resolution of comorbidities, and reduction of medication intake. Despite risks of long-term complications, all nine existing published series of long-term overall mortality have reported that death rates are reduced in patients who have undergone bariatric/metabolic procedures compared to non-surgical controls [6], [29, 24, 25]. Additionally, the AHRQ (Agency for Healthcare Research and Quality) review considered most of the surgical complications after bariatric surgery minor and not requiring complex management in most of the cases [30].

Table 14.1.5.1 Most common long-term complications of bariatric/metabolic surgeries, including Roux-en-Y Gastric Bypass (RYGB), vertical sleeve gastrectomy (VSG), laparoscopic adjustable gastric banding (LAGB), and biliopancreatic diversion (BPD) [6, 31–42]

	Most common with: (type of procedure)
Surgical complications	
Internal hernia/small bowel occlusion	RYGB and BPD
Stricture	RYGB, VSG
Band slippage/erosion	LAGB
Marginal ulcers	RYGB (rare report with BPD)
G-E Reflux/Barret oesophagus	VSG
Re-operation	All
Nutritional/metabolic complications	
Iron deficiency	All (++ RYGB, VSG)
Vitamin B_{12} deficiency	RYGB, VSG, BPD
Vitamin D deficiency	All
Vitamin A deficiency	BPD, RYGB
Anaemia	All (++ BPD, RYGB, VSG)
Post-prandial hypoglycaemia	RYGB, VSG, (BPD?)
Hypocalcaemia	++RYGB, (VSG, BPD)
Osteoporosis and fracture risk	All (++ BPD, RYGB)
Copper deficiency	BPD, RYGB
Zinc deficiency	All, (++ BPD, RYGB, LAGB)
Thiamine (B_1) deficiency (rare development of Wernicke encephalopathy)	BPD, VSG, RYGB

Data derived from Schauer, P.R., *et al.*, Clinical Outcomes of Metabolic Surgery: Efficacy of Glycemic Control, Weight Loss, and Remission of Diabetes. *Diabetes Care*, 2016. 39(6): p. 902–11.

Standard and Novel Procedures in Metabolic Surgery

Bariatric/metabolic procedures most commonly performed worldwide are RYGB, vertical sleeve gastrectomy (VSG), laparoscopic adjustable gastric banding (LAGB) and biliopancreatic diversion (BPD) (Figure 14.1.5.2). The efficacy on weight loss and diabetes remission is ordered as follows: BPD>RYGB>VSG>LAGB, whereas the inverse is true for safety. VSG has currently overtaken RYGB in many countries, due to a fairly simple learning curve, however, RYGB is considered to be the gold standard operation for resolution of metabolic diseases and diabetes by most surgeons. Nowadays several bariatric/metabolic operations are under development, including: (1) Proximal intestinal bypass procedures for type 2

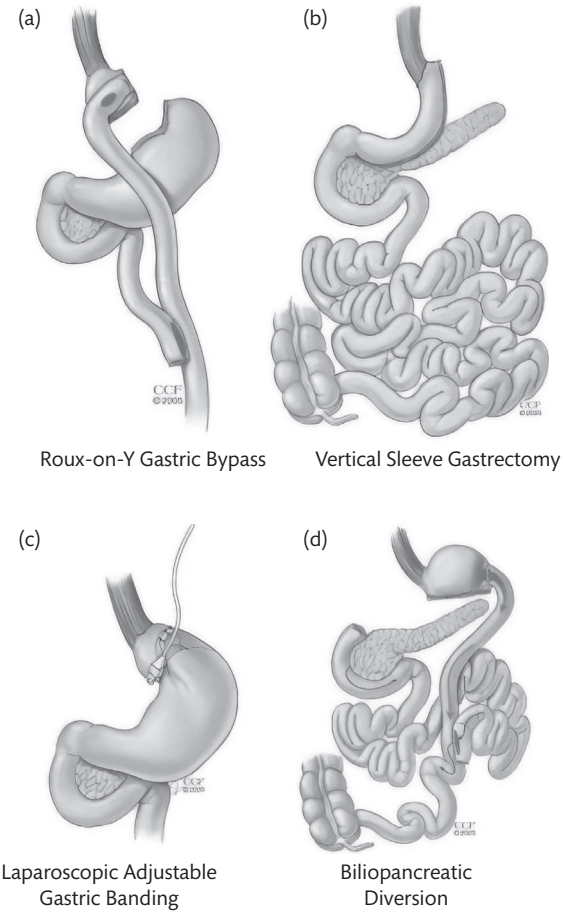


Figure 14.1.5.2 Bariatric/Metabolic Procedures most commonly performed: Roux-en-Y gastric bypass (RYGB): the stomach is divided, and a small proximal pouch is created. (a) A gastro-jejunal anastomosis is created. The remnant stomach, duodenum, and proximal jejunum are excluded from the transit of nutrients. Bile and biliopancreatic juices are diverted; (b) laparoscopic adjustable gastric banding (LAGB): an inflatable band is placed around the upper part of the stomach. The band adjusted by injecting saline into a subcutaneous port.; (c) vertical sleeve gastrectomy (VSG): a vertical gastric resection is performed along the smaller curvature using staples, leaving behind a 'sleeve-shaped' stomach, without rerouting the intestine; biliopancreatic diversion (BPD): the stomach is resected horizontally—classic BPD, or vertically—'BPD-duodenal switch'. The duodenum, jejunum, and part of the ileum are bypassed. Nutrients and biliopancreatic juice mix only within the distal 50–100 cm of the ileum—this segment is dubbed 'common channel'.

Reproduced with permission from Rubino, F., *et al.*, Metabolic Surgery in the Treatment Algorithm for Type 2 Diabetes: A Joint Statement by International Diabetes Organizations. Diabetes Care, 2016. 39(6): pp. 861–77; and from Cummings, D.E. and R.V. Cohen, Bariatric/Metabolic Surgery to Treat Type 2 Diabetes in Patients with a BMI <35 kg/m². *Diabetes Care*, 2016. 39(6): pp. 924–33.

diabetes (e.g. DJB surgery, endoscopically implanted endoluminal sleeves and endoscopic duodenal mucosal resurfacing); (2) Ileal interposition surgery to increase the secretion of L-cell peptide from the distal gut; (3) Endoscopic reduction of gastric volume for weight loss (e.g. new-generation gastric balloons, gastric plication, and gastric electrical stimulation). Most of these approaches are currently used primarily in clinical trials.

Cost-Effectiveness of Metabolic Surgery

All of the studies investigating economic aspects of bariatric/metabolic surgery have reported surgery for obesity or diabetes to be cost-effective with a mean price of $5000–10 000 per quality-adjusted-life-year (QALY). This figure is far below the average threshold of $50 000/QALY used to determine coverage by healthcare insurance [43, 44]. Additionally, bariatric/metabolic surgery seems more cost-effective than medication for glycaemic and lipid control, that are reported to have $41 384/QALY and $51 889/QALY, respectively [45]. The time to achieve return-on-investment differs significantly among studies. However, the total cost after bariatric/metabolic surgery seems to be compensated throughout time avoiding further expenditure for hospitalizations, medications, and complications related to the progression of the disease. Thus, insurance companies should be able to recoup the cost of surgery over time and, prospectively, the whole insurance coverage for bariatric/metabolic surgery may decrease the entire cost of the private healthcare system. It is reasonable to contemplate that the public healthcare system, such as the UK National Health Service (NHS), would experience similar benefits.

Current Indications for Surgical Treatment of Type 2 Diabetes

Historical international guidelines for bariatric surgery have limited indications to only patients with BMI ≥35 kg/m², based on a 1991 National Institutes of Health (NIH) consensus statement [30]. Such weight-centric criterion for surgery was not originally based on evidence of clinical or biological efficacy and is particularly inadequate to identify patients with greatest potential benefits from surgery when control of diabetes is the primary concern.

A recent meta-analysis has compared 94 579 surgical patients with type 2 diabetes, showing similar rates of disease remission (71% vs. 72% respectively) among patients with baseline BMI either lower or greater than 35 kg/m² [46]. Similar outcomes have been reported by a meta-analysis of 11 randomized clinical trials, in which no differences were found in the remission rate of diabetes among the six randomized clinical trials with preoperative BMI ≥35 kg/m², when compared with the five randomized clinical trials in which mean baseline BMI of the study cohorts was <35 kg/m² [22].

Randomized clinical trials to identify the correct indications for surgical treatment of diabetes was initially supported by the first Diabetes Surgery Summit in 2007 [47, 48]. In 2015 a multidisciplinary and international group of diabetes scholars, representing leading diabetes organizations reviewed available evidence for the 2nd Diabetes Surgery Summit and developed new guidelines that recognize metabolic surgery as standard of care option for type 2 diabetes. These guidelines recommend that metabolic surgery be considered for treating diabetes in patients with a BMI >30 kg/m² or BMI >27.5 kg/m² for Asian individuals. These guidelines (Figure 14.1.5.3) have been ratified by 56 organizations from around the world up to the present time including diabetes and surgical societies [2, 49].

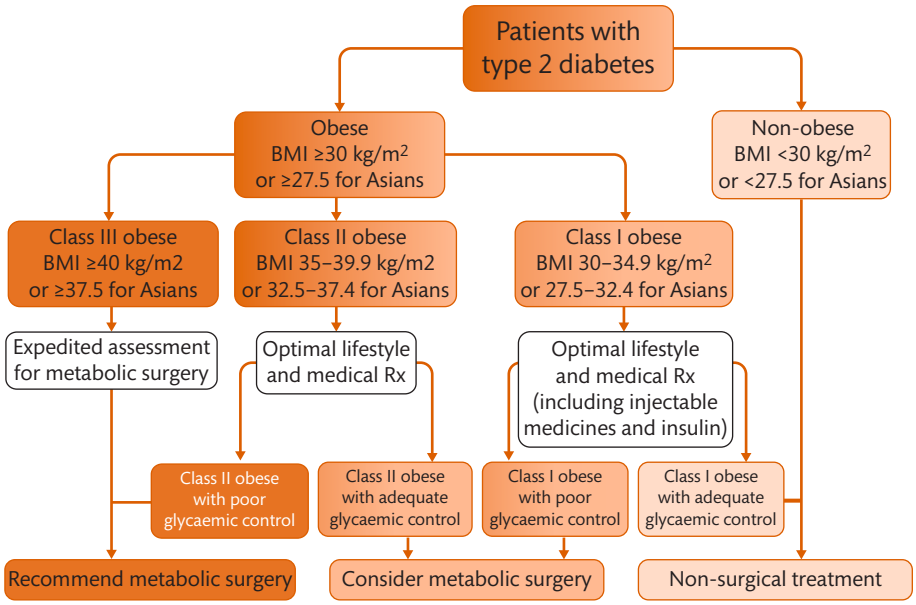


Figure 14.1.5.3 DSS-II international guidelines for metabolic surgery.

REFERENCES

1. Ali MK, Bullard KM, Gregg EW. Achievement of goals in U.S. Diabetes Care, 1999–2010. *N Engl J Med*, 2013; **369**: 287–8.
2. Rubino F, Nathan DM, Eckel RH, *et al*. Metabolic surgery in the treatment algorithm for type 2 diabetes: a joint statement by International Diabetes Organizations. *Diabetes Care*, 2016; **39**: 861–77.
3. American Diabetes Association . Standards of medical care in diabetes—2016. Abridged for primary care providers. *Clin Diabetes*, 2016. **34**: 3–21.
4. Batterham RL, Cummings DE. Mechanisms of diabetes improvement following bariatric/metabolic surgery. *Diabetes Care*, 2016; **39**: 893–901.
5. Rubino F, Marescaux J. Effect of duodenal-jejunal exclusion in a non-obese animal model of type 2 diabetes: a new perspective for an old disease. *Ann Surg*, 2004; **239**: 1–11.
6. Schauer PR, Mingrone G, Ikramuddin S, Wolfe B. Clinical outcomes of metabolic surgery: efficacy of glycemic control, weight loss, and remission of diabetes. *Diabetes Care*, 2016; **39**: 902–11.
7. Leyton O . Hypoglycopyresis. *Proc R Soc Med*, 1925; **18**(Med Sect): 11–16.
8. Pories WJ, Swanson MS, MacDonald KG, *et al*. Who would have thought it? An operation proves to be the most effective therapy for adult-onset diabetes mellitus. *Ann Surg*, 1995; **222**: 339–50; discussion 350–2.
9. Thaler JP, Cummings DE. Minireview: hormonal and metabolic mechanisms of diabetes remission after gastrointestinal surgery. *Endocrinology*, 2009; **150**: 2518–25.
10. Albaugh VL, Flynn CR, Cai S, Xiao Y, Tamboli RA, Abumrad NN. Early increases in bile acids post Roux-en-Y gastric bypass are driven by insulin-sensitizing, secondary bile acids. *J Clin Endocrinol Metab*, 2015; **100**: E1225–33.
11. Kuhre RE, Wewer Albrechtsen NJ, Larsen O, *et al*. Bile acids are important direct and indirect regulators of the secretion of appetite—and metabolism-regulating hormones from the gut and pancreas. *Mol Metab*, 2018. **11**: 84–95.
12. Saeidi N, Meoli L, Nestoridi E, *et al*. Reprogramming of intestinal glucose metabolism and glycemic control in rats after gastric bypass. *Science*, 2013; **341**: 406–10.
13. Pattou F, Daoudi M, Baud G. Roux-en-Y gastric bypass and intestinal glucose handling: a salty sweet operation. *Gastroenterology*, 2016; **151**: 210.
14. Duca FA, Bauer PV, Hamr SC, Lam TK. Glucoregulatory relevance of small intestinal nutrient sensing in physiology, bariatric surgery, and pharmacology. *Cell Metab*, 2015; **22**: 367–80.
15. Rubino F, Amiel SA. Is the gut the 'sweet spot' for the treatment of diabetes? *Diabetes*, 2014; **63**: 2225–8.
16. Arterburn DE, Bogart A, Sherwood NE, *et al*. A multisite study of long-term remission and relapse of type 2 diabetes mellitus following gastric bypass. *Obes Surg*, 2013; **23**: 93–102.
17. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*, 2008; **359**: 1577–89.
18. Sjöström L . Review of the key results from the Swedish Obese Subjects (SOS) trial—a prospective controlled intervention study of bariatric surgery. *J Intern Med*, 2013; **273**: 219–34.
19. Adams TD, Davidson LE, Litwin SE, *et al*. Health benefits of gastric bypass surgery after 6 years. *JAMA*, 2012; **308**: 1122–31.
20. Arterburn DE, Olsen MK, Smith VA, *et al*. Association between bariatric surgery and long-term survival. *JAMA*, 2015; **313**: 62–70.
21. Adams TD, Arterburn DE, Nathan DM, Eckel RH. Clinical outcomes of metabolic surgery: microvascular and macrovascular complications. *Diabetes Care*, 2016; **39**: 912–23.
22. Cummings DE, Cohen RV. Bariatric/metabolic surgery to treat type 2 diabetes in patients with a BMI <35 kg/m². *Diabetes Care*, 2016; **39**: 924–33.
23. Cummings DE, Cohen RV. Beyond BMI: the need for new guidelines governing the use of bariatric and metabolic surgery. *Lancet Diabetes Endocrinol*, 2014; **2**: 175–81.

24. Sjöström L, Narbro K, Sjöström CD, *et al*. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med*, 2007; **357**: 741–52.
25. Adams TD, Gress RE, Smith SC, *et al*. Long-term mortality after gastric bypass surgery. *N Engl J Med*, 2007; **357**: 753–61.
26. Vest AR, Heneghan HM, Agarwal S, Schauer PR, Young JB. Bariatric surgery and cardiovascular outcomes: a systematic review. *Heart,* 2012; **98**: 1763–77.
27. Longitudinal Assessment of Bariatric Surgery (LABS), Flum DR, Belle SH, *et al*. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med*, 2009; **361**: 445–54.
28. Aminian A, Brethauer SA, Kirwan JP, *et al*. How safe is metabolic/diabetes surgery? *Diabetes Obes Metab*, 2015; **17**: 198–201.
29. Arterburn DE, McDonell MB, Hedrick SC, Diehr P, Fihn SD . Association of body weight with condition-specific quality of life in male veterans. *Am J Med*, 2004; **117**: 738–46.
30. Maglione MA, Gibbons MM, Livhits M, *et al*. *Bariatric Surgery and Nonsurgical Therapy in Adults with Metabolic Conditions and a Body Mass Index of 30.0 to 34.9 kg/m²*. Rockville, MD: Agency for Healthcare Research and Quality (US), 2013.
31. Duvoisin C, Favre L, Allemann P, Fournier P, Demartines N, Suter M. Roux-en-Y gastric bypass: ten-year results in a cohort of 658 patients. *Ann Surg*, 2018; **268**: 1019–25.
32. Rousseau C, Jean S, Gamache P, *et al*. Change in fracture risk and fracture pattern after bariatric surgery: nested case-control study. *BMJ*, 2016; **354**: i3794.
33. Balsa JA, Botella-Carretero JI, Gómez-Martín JM, *et al*. Copper and zinc serum levels after derivative bariatric surgery: differences between Roux-en-Y Gastric bypass and biliopancreatic diversion. *Obes Surg*, 2011; **21**: 744–50.
34. Gehrer S, Kern B, Peters T, Christoffel-Courtin C, Peterli R. Fewer nutrient deficiencies after laparoscopic sleeve gastrectomy (LSG) than after laparoscopic Roux-Y-gastric bypass (LRYGB)-a prospective study. *Obes Surg*, 2010. **20**: 447–53.
35. Coupaye M, Puchaux K, Bogard C, *et al*. Nutritional consequences of adjustable gastric banding and gastric bypass: a 1-year prospective study. *Obes Surg*, 2009; **19**: 56–65.
36. Marinari GM, Murelli F, Camerini G, *et al*. A 15-year evaluation of biliopancreatic diversion according to the Bariatric Analysis Reporting Outcome System (BAROS). *Obes Surg*, 2004; **14**: 325–8.
37. Liu D, Ke Z, Luo J. Thiamine deficiency and neurodegeneration: the interplay among oxidative stress, endoplasmic reticulum stress, and autophagy. *Mol Neurobiol*, 2017; **54**: 5440–8.
38. Said HM. Intestinal absorption of water-soluble vitamins in health and disease. *Biochem J*, 2011; **437**: 357–72.
39. Matrana MR, Vasireddy S, Davis WE. The skinny on a growing problem: dry beriberi after bariatric surgery. *Ann Intern Med*, 2008; **149**: 842–4.
40. Via MA, Mechanick JI. Nutritional and micronutrient care of bariatric surgery patients: current evidence update. *Curr Obes Rep*, 2017; **6**: 286–96.
41. Parrott J, Frank L, Rabena R, Craggs-Dino L, Isom KA, Greiman L. American Society for Metabolic and Bariatric Surgery integrated health nutritional guidelines for the surgical weight loss patient 2016 update: micronutrients. *Surg Obes Relat Dis*, 2017; **13**: 727–41.
42. Salinari S, Mingrone G, Bertuzzi A, Previti E, Capristo E, Rubino F. Downregulation of insulin sensitivity after oral glucose administration: evidence for the anti-incretin effect. *Diabetes*, 2017; **66**: 2756–763.
43. Picot J, Jones J, Colquitt JL, *et al*. The clinical effectiveness and cost-effectiveness of bariatric (weight loss) surgery for obesity: a systematic review and economic evaluation. *Health Technol Assess,* 2009; **13**: 1–190, 215–357, iii–iv.
44. Keating C, Neovius M, Sjöholm K, *et al*. Health-care costs over 15 years after bariatric surgery for patients with different baseline glucose status: results from the Swedish Obese Subjects study. *Lancet Diabetes Endocrinol*, 2015; **3**: 855–65.
45. CDC Diabetes Cost-effectiveness Group. Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes. *JAMA*, 2002; **287**: 2542–51.
46. Panunzi S, De Gaetano A, Carnicelli A, Mingrone G. Predictors of remission of diabetes mellitus in severely obese individuals undergoing bariatric surgery: do BMI or procedure choice matter? A meta-analysis. *Ann Surg*, 2015; **261**: 459–67.
47. Rubino F, Kaplan LM, Schauer PR, *et al*. The Diabetes Surgery Summit consensus conference: recommendations for the evaluation and use of gastrointestinal surgery to treat type 2 diabetes mellitus. *Ann Surg*, 2010; **251**: 399–405.
48. Rubino F, Schauer PR, Kaplan LM, Cummings DE . Metabolic surgery to treat type 2 diabetes: clinical outcomes and mechanisms of action. *Annu Rev Med*, 2010; **61**: 393–411.
49. Cummings DE, Rubino F. Metabolic surgery for the treatment of type 2 diabetes in obese individuals. *Diabetologia*, 2018; **61**: 257–64.

14.1.6 Assessment of Obesity in Children

I. Sadaf Farooqi

Introduction

Obesity is defined by excess body fat. In adults, internationally accepted criteria rely on body mass index (BMI; weight in kg/height in metres²) to define overweight (BMI≥25 kg/m²), obesity (BMI≥30 kg/m²) and severe obesity (BMI≥40 kg/m²). However, in children, body composition changes with age, and differs by gender, and longitudinal data permitting prediction of the impact

of different weight trajectories in childhood and adolescence, is lacking. The US Centers for Disease Control (https://www.cdc.gov/) recommend use of gender-specific charts accounting for the change in BMI with age: children are diagnosed as overweight if the BMI is >85th percentile but <95th percentile and obese if the BMI is >95th percentile for age and gender. BMI data for children under 2 years of age is lacking; in this age group, a child is considered obese if the weight for recumbent length is >97.7th percentile. Extreme obesity is defined as a BMI >120% of the 95th percentile or >35 kg/m^2.

Using these definitions, the prevalence of childhood obesity has been shown to be increasing worldwide driven by inexpensive, easily available, energy-rich, highly palatable foods which promote increased energy intake, and factors which contribute to a decrease in energy expenditure such as sedentary lifestyles (television watching) and reduced physical activity at school and in leisure time. However, within a given environment, there is considerable variation in body weight and fat mass among individuals, with some children much more likely to gain weight than others [1, 2]. This variability is influenced by complex interactions between environmental and biological (genetic, developmental, and behavioural) factors [3]. As well as an increase in the mean BMI in many populations, the proportion of children at the upper end of the distribution, those with severe obesity, has also increased. Physicians thus need a systematic approach to the assessment of severe childhood-onset obesity.

Clinical Assessment

The assessment of severely obese children should include appropriate screening for potentially treatable endocrine and neurological conditions. Testing for underlying genetic conditions should also be considered, so that appropriate genetic counselling and, in some cases, specific treatment can be instituted. In most patients, specific endocrine and neurological (tumours) causes can be excluded by a careful clinical history, examination and investigations focusing on associated clinical features (**Box 14.1.6.1**). A short history of rapid weight gain may be a key clue in such cases. Obese children should also be assessed for the potential complications of severe obesity such as sleep apnoea, impaired glucose tolerance/type 2 diabetes, fatty liver disease, and hypertension as well as valgus/varus skeletal deformities and orthopaedic problems (see Chapter 14.1.3 'Medical Complications of Obesity').

Medical History

A specific weight history should be taken to establish the age of onset of obesity (photographs can be helpful), as obesity that begins in early childhood is more likely to be driven by a strong genetic component. Previous treatment for obesity and the amount of weight lost, regular diet inside and outside the home, and levels of physical activity should be noted. A careful history should be taken to identify hyperphagia, and associated food-seeking behaviours such as searching for/stealing food, waking at night to find food, or eating food others leave behind. Such behaviours can be distressing for families and may not be volunteered to healthcare professionals, particularly if parents fear they will be blamed for causing their child's obesity. A sensitive approach to history taking is an important aspect of the assessment of severe childhood-onset obesity. Hyperphagia and impaired satiety typically occur as a result of disruption of hypothalamic pathways involved in the regulation of energy balance. Pica, the eating of non-food items, is rare but can be indicative of disruption of neural circuits in the hypothalamus and amygdala. Specific genetic causes of this have not been defined to date. A clear history of hyperphagia should prompt genetic investigation, and neurological causes should be excluded where the history is short.

Developmental history is also important as several obesity disorders are associated with speech and language delay. Sometimes severely obese children are slow to walk for mechanical reasons; this seldom represents true motor delay, although muscle weakness in children with mutations in thyroid hormone receptor alpha (who are often overweight) is recognized [4]. Vision and hearing should be assessed as some genetic syndromes are associated with retinal dystrophy and hearing loss (as follows). A behavioural history is important as autistic behaviour, hyperactivity, and/or aggression are features of a subgroup of genetic disorders. Anxiety, mood disorders, and self-harm are also common in obese adolescents and should be tactfully considered. A general history should seek features suggestive of hypothalamic dysfunction (sleep disturbance, temperature dysregulation, abnormal thirst), delayed pubertal development, and frequent infections. A family history to identify potential consanguineous relationships, other family members with severe childhood obesity (some of whom may have had bariatric surgery) and the ethnic and geographical origin of family members should be noted.

Box 14.1.6.1 Features in the clinical history suggesting genetic and/or other causes for severe obesity

Age of onset—use growth charts and family photographs. Early onset (<5 years of age) suggests a genetic cause and warrants investigation.

Duration of obesity—short history suggests endocrine or central cause.

Hyperphagia—often denied, but sympathetic approach needed. Ask specific questions, such as waking at night to eat, demanding food very soon after a meal suggest hyperphagia. If severe, especially in children, suggests a genetic cause for obesity and warrants investigation.

Past history of damage to the CNS (infection, trauma, haemorrhage, radiation therapy, seizures) suggests hypothalamic obesity with or without growth hormone deficiency or hypothyroidism. Morning headaches, vomiting, visual disturbance, and excessive urination or drinking also suggests that the obesity may be caused by a tumour or mass in the hypothalamus.

Dry skin, constipation, intolerance to cold, or fatigue can suggest hypothyroidism. Mood disturbance and central obesity may suggest Cushing's syndrome. Frequent infections, cholestatic jaundice, hypoglycaemia, and fatigue may suggest ACTH deficiency due to POMC/PCSK1 mutations.

Developmental delay—check milestones, educational history, behavioural disorders. Consider craniopharyngioma or structural causes (often relatively short history) and genetic causes.

Visual impairment and deafness can suggest genetic causes.

Onset and tempo of pubertal development—onset can be early or delayed in children and adolescents. Primary hypogonadotropic hypogonadism or hypogenitalism associated with some genetic disorders.

Family history—consanguineous relationships, other children affected, family photographs useful. Severity may differ due to environmental effects and as yet unidentified genetic modifiers.

Treatment with certain drugs or medications e.g. glucocorticoids, oral contraceptives, antidepressants, and antipsychotics.

Medical Examination

Height should be measured using a stadiometer and weight measured in light clothing with shoes off. Waist circumference is taken as the mid-point between the lower rib margin and the iliac crest. Skinfold thickness has limited value and its measurement is uncomfortable for patients. An examination of the skin is important, however: thin, atrophic skin is a feature of excess corticosteroids; acanthosis nigricans (pigmented 'velvety' skin creases, especially in the axillae) suggests insulin resistance. Assessment for hypogonadism and any dysmorphic features is important.

Clinicians should use laboratory testing to evaluate obese patients who may be at high risk for cardiovascular disease and type 2 diabetes. Investigations to consider are fasting plasma glucose or 2-h postprandial glucose levels and serum lipid levels. Thyroid-stimulating hormone (TSH) may be helpful in excluding hypothyroidism. Other tests to consider depend on clinical assessment and include liver imaging for hepatic steatosis, polysomnography for patients with possible sleep apnoea; and head CT or MRI when pituitary or hypothalamic disorders are suspected. The use of these tests has recently been reviewed as part of US and European guidelines for the assessment of childhood obesity [5]. Genetic testing is now recognized as an important part of the assessment of severe childhood obesity [5]. The measurement of serum leptin is not recommended routinely, but should be undertaken in very young patients with severe obesity, as congenital leptin deficiency is a potentially treatable disorder.

Genetic Obesity Syndromes–An Overview

Genetic studies focusing on severe childhood obesity have led to the identification of several genes in which rare, highly penetrant variants cause obesity (Figure 14.1.6.1). While individually these disorders are rare, cumulatively at least 10% of children with severe obesity have rare chromosomal abnormalities and/or highly penetrant genetic mutations that drive their obesity [6]. This figure is likely to increase with wider access to genetic testing in clinical practice and adherence to clinical guidelines recommending that children with severe obesity before the age of 5 years and hyperphagia should be tested [5]. Some genetic obesity syndromes are associated with learning difficulties and developmental delay which mean they often come to medical attention at a young age. However, there is a large and increasing group of genetic disorders where severe obesity itself is the presenting feature [7, 8].

Obesity With Developmental Delay

Prader–Willi Syndrome

Prader–Willi syndrome (PWS) is one of the most prevalent genetic obesity syndromes affecting approximately 1 in 25 000 people [9]. Classical clinical features include hypotonia and failure to thrive in infancy, learning difficulties, short stature, the development of hyperphagia and severe obesity in childhood and hypogonadotropic hypogonadism [10]. Children with PWS have reduced lean body mass and increased fat mass, abnormalities which resemble those

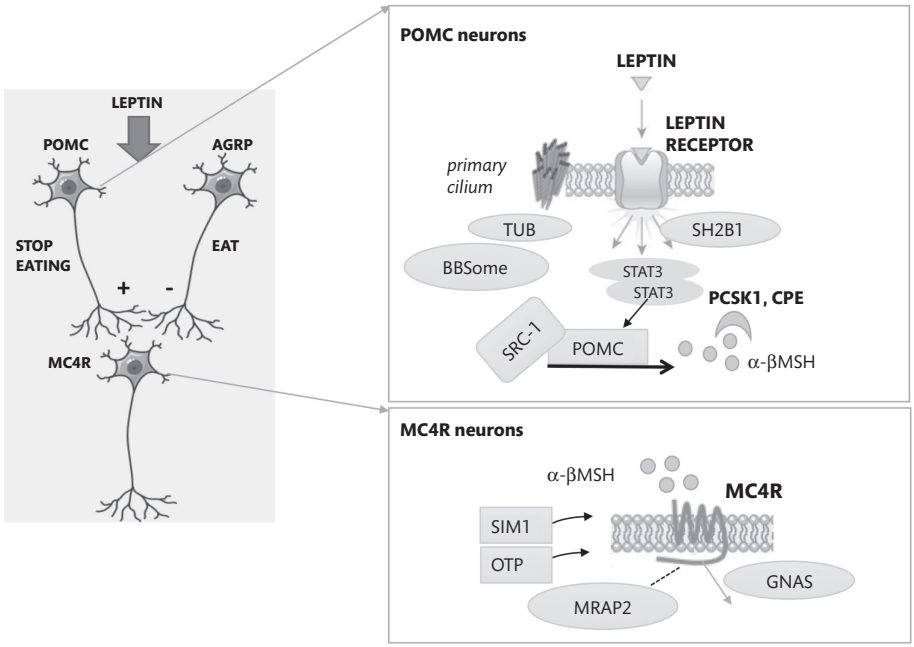


Figure 14.1.6.1 Mutations in multiple components of the leptin-melanocortin pathway are associated with severe childhood obesity. The adipose tissue-derived hormone leptin crosses the blood-brain barrier to activate/inhibit primary neurons expressing pro-opiomelanocortin (POMC) or agouti-related peptide (AGRP) which express the leptin receptor on their cell surface. POMC is post-translationally processed to yield POMC peptides, alpha and beta-MSH which are agonists at the melanocortin 4 receptor (MC4R) expressed on second order neurons; POMC stimulates MC4R signalling (+) while AGRP is a MC4R antagonist (–). In orange are genes affected by mutations associated with severe childhood obesity; all disrupt the development of, or signalling by, neurons in this circuit that reduce food intake. As a result, these obesity syndromes are characterized by hyperphagia, an increased drive to eat.

seen in growth hormone (GH) deficiency. GH treatment in children, adolescents, and adults decreases body fat and increases linear growth, muscle mass, fat oxidation and energy expenditure and as a result is often now commenced at a young age [11].

PWS is caused by sporadic deletion of a critical segment on the paternally inherited copy of chromosome 15q11.2-q12 or loss of the entire paternal chromosome 15 with 2 maternal copies (uniparental maternal disomy). Chromosomal abnormalities can be identified by a karyotype analysis or by the detection of differences in DNA methylation of the parental alleles. Chromosomal deletions that cause PWS vary in size and so disrupt a number of genes and non-coding ribonucleic acids (RNAs). The minimal genetic lesion associated with severe hyperphagia and obesity in PWS contains a cluster of non-coding small nucleolar RNAs (snoRNAs)—the SNORD116 gene cluster [12, 13]. A recent transcriptomic study of post-mortem hypothalamic tissue from PWS patients versus age- and BMI-matched controls [14] found that genes downregulated in PWS were involved in neurogenesis, neurotransmitter release, and synaptic function. Expression of brain-derived neurotrophic factor (BDNF), a major regulator of neuronal development and maturation, was found to be reduced in the hypothalamus of PWS patients. However, plasma BDNF levels were not reduced in people with PWS versus age-matched obese controls. In line with these findings, several imaging studies have identified structural abnormalities such as reduced cortical grey matter volume and abnormal gyrification in PWS patients [15]. This transcriptomic study confirmed findings from a previous histopathological study of the PWS hypothalamus which found a significantly reduced number of oxytocin neurons [16] in PWS patients. As low levels of oxytocin in the hypothalamus could contribute to some of the behavioural difficulties seen in patients as well as to the hyperphagia, clinical trials of intranasal oxytocin administration in PWS are being conducted [17]. Plasma ghrelin levels are markedly elevated in children and adults with PWS although the physiological relevance of this finding is unknown [18].

BDNF and TrkB Deficiency

A small number of children with hyperphagia, obesity, impaired short-term memory, hyperactivity, moderate to severe learning difficulties and impaired nociception who have mutations or chromosomal deletions that disrupt BDNF or its tyrosine kinase receptor tropomyosin-related kinase B (TrkB) [19, 20] have been reported. Given the severe developmental phenotype of these patients, it is not surprising that mutations often arise *de novo* and as such should be considered where both parents are of normal weight and IQ.

Albright's Hereditary Osteodystrophy

Albright's hereditary osteodystrophy (AHO) is characterized by short stature, obesity, skeletal defects, and impaired olfaction due to mutations in *GNAS1*, the gene encoding $G\alpha_s$ protein which mediates signalling by multiple G-protein coupled receptors (GPCRs). Imprinting at this locus results in a variable clinical phenotype in patients who carry loss-of-function mutations in *GNAS1*. *GNAS1* is imprinted in a tissue-specific manner, being expressed primarily from the maternal allele in some tissues and bi-allelically in other tissues; thus multihormone resistance occurs only when $G\alpha_s$ mutations are inherited maternally [21]. As a result, maternal transmission of *GNAS1* mutations leads to the classical AHO phenotype with resistance to several hormones (e.g. parathyroid hormone) that activate $G\alpha_s$ in their target tissues (pseudohypoparathyroidism type IA), while paternal transmission leads only to AHO (pseudo pseudohypoparathyroidism). To date, a diagnosis of AHO is generally only considered when the classical clinical features are present. Clinical series of patients with features of AHO often include a large number of patients with deletions or frameshift mutations [22]. However, recent studies in large cohorts of children with severe obesity have shown that some patients with *GNAS1* mutations present with obesity without some of the classical features of AHO including short stature (although learning difficulties are present in over 60%) [23]. Molecular studies of these variants, which are often missense mutations that impair signalling by some but not all $G\alpha_s$-coupled GPCRs, suggest that the spectrum of clinical phenotypes associated with *GNAS1* mutations may in part be explained by the molecular consequences of specific mutations. In view of this heterogeneity, this genetic disorder should be considered in all patients with severe obesity and developmental delay even in the absence of the classical clinical features of AHO.

SIM1 Deficiency

Chromosomal deletions and heterozygous loss-of-function mutations that disrupt single-minded 1 (SIM1) result in severe dominantly inherited obesity [24]. SIM1 is a basic helix-loop-helix transcription factor involved in the development and function of the paraventricular nucleus (PVN) of the hypothalamus and may be involved in mediating melanocortin and oxytocin signalling. Prevalent features in SIM1 deficient patients are speech and language delay and neurobehavioural abnormalities including autistic type behaviours. Patients are hyperphagic with evidence of autonomic dysfunction (characterized by low systolic blood pressure). As the hyperphagia of *sim1* haplo-insufficient mice is partly ameliorated by the central administration of oxytocin [25, 26], a neurotransmitter involved in the modulation of emotion and social interaction, impaired oxytocinergic signalling is one possible mechanism implicated in the obesity and behavioural phenotype seen in SIM1 deficiency. A small number of patients have been reported with heterozygous mutations in Orthopedia (Otp) [26], a gene that interacts with SIM1 to regulate the differentiation of neurons in the PVN. These patients also exhibit speech and language delay, autistic type behaviours, and anxiety.

Ciliary Disorders

Bardet-Biedl syndrome (BBS) is a rare autosomal recessive disease characterized by obesity, learning disability, dysmorphic extremities (syndactyly, brachydactyly or polydactyly), retinal dystrophy or pigmentary retinopathy, hypogonadism, and structural abnormalities of the kidney or functional renal impairment [27]. BBS is a genetically heterogeneous disorder caused by mutations in at least 16 genes with mutations in more than one gene sometimes required for complete expression of the phenotype (tri-allelic inheritance). These genes all affect proteins localized to the basal body of the primary cilium. While hypothalamic neurons express cilia, their precise function is unknown. Other recessive disorders of ciliary function include Alström syndrome (retinal dystrophy, severe insulin resistance, deafness) and *TUB* mutations [28]. The presence of cilia on leptin-responsive neurons involved in energy balance and animal data suggesting that their disruption causes obesity suggest

that there is a link between ciliary function, leptin signalling and energy homeostasis that remains to be explored [29]. Rare homozygous/compound heterozygous mutations in *ADCY3* (encoding adenylate cyclase 3) have also been identified in children with severe obesity [30] with some evidence that this gene is localized to the cilium [31].

Obesity Without Developmental Delay

Neural circuits within the hypothalamus regulate energy balance in response to peripheral nutrient-related cues [32, 33]. Leptin-responsive Agouti-related protein (Agrp)-expressing neurons in the arcuate nucleus of the hypothalamus are activated during fasting/caloric deficit to drive an increase in food intake, while in the nutritionally replete/fed state, Pro-opiomelanocortin (Pomc) neurons are activated to reduce food intake [34, 35]. Loss-of-function mutations that disrupt the function of these neural circuits result in severe obesity, demonstrating their pivotal role in human energy homeostasis [2, 36].

Leptin and Leptin Receptor Deficiency

Homozygous mutations in the genes encoding leptin and the leptin receptor cause recessively inherited severe obesity [37]. Leptin receptor mutations have been found in some non-consanguineous families, where both parents are unrelated but carry rare heterozygous variants [8]. Patients exhibit rapid weight gain in the first few months of life resulting in severe obesity by age 2 years (mean body mass index standard deviation score (BMI SDS): +5.8–7.8). Early development is usually normal. The most notable feature is intense hyperphagia with food-seeking behaviour and aggressive behaviour when food is denied. Children with leptin deficiency have impaired T cell-mediated immunity [38] consistent with high rates of childhood infection and of mortality from infection [39]. In those that survive, obesity continues into adult life with hepatic steatosis [40], hyperinsulinemia and type 2 diabetes [41]. Leptin and leptin receptor deficiency are associated with hypothalamic hypothyroidism and hypogonadotropic hypogonadism but there is some evidence for the delayed but spontaneous onset of menses in some leptin and leptin receptor deficient adults. Leptin and leptin receptor deficient children have normal linear growth in childhood and normal IGF-1 concentrations but final adult height is usually reduced due to the absence of a pubertal growth spurt.

Although leptin deficiency is very rare, it is treatable with daily subcutaneous injections of recombinant leptin [38] which is currently available to patients on a named patient basis in a small number of centres around the world. The major effect of leptin treatment is on food intake, with normalization of hyperphagia and enhanced satiety [38, 42]. Importantly, leptin also has permissive effects on the development of puberty and if given in early childhood permits appropriate linear growth. A recent trial of Setmelanotide, a MC4R agonist, which would be predicted to improve signalling downstream of LEPR, suggests that patients with loss-of-function leptin receptor mutations may potentially be treated [43].

Serum leptin is a useful test in patients with severe early-onset obesity as an undetectable concentration is highly suggestive of congenital leptin deficiency. Very rare mutations in the *LEP* gene that result in a bio-inactive form of the hormone but appropriate leptin concentrations have also been reported, however [44]. Serum leptin concentrations are appropriate for the degree of obesity in leptin receptor deficient patients; elevated serum leptin is not necessarily a predictor of leptin receptor deficiency [8]. However, a subset of *LEPR* mutations resulting in abnormal cleavage of the extracellular domain of LEPR, which acts as a leptin binding protein, are associated with markedly elevated leptin concentrations [45].

Pro-Opiomelanocortin (POMC) Deficiency

Children who have homozygous or compound heterozygous mutations in the *POMC* gene present in neonatal life with adrenal crisis due to adrenocorticotrophic hormone (ACTH) deficiency and require long-term corticosteroid replacement [46]. Such children have pale skin and white Caucasians often have red hair due to the lack of melanocortin signalling at melanocortin 1 receptors in the skin. Children from different ethnic backgrounds may present with a less obvious pigmentary phenotype. POMC deficiency results in hyperphagia and early-onset obesity due to loss of melanocortin signalling at the melanocortin 4 receptor, MC4R. In a clinical trial, Setmelanotide led to dramatic weight loss in patients with complete POMC deficiency [47]. Heterozygous mutations affecting POMC-derived peptides are more prevalent [48]; trials in these patient groups are currently taking place.

PC1/3 (PCSK1) Deficiency

Proprotein Convertase Subtilisin/Kexin type 1 2 (PCSK1) is expressed in neuroendocrine tissues and acts to cleave prohormones including pro-opiomelanocortin (POMC), pro-thyrotropin-releasing hormone (TRH), proinsulin, proglucagon, and pro-gonadotrophin-releasing hormone (GnRH) to release biologically active peptides. Compound heterozygous or homozygous mutations in *PCSK1*, which encodes PC1/3, cause small bowel enteropathy and patients may present in neonatal life or early infancy with persistent diarrhoea requiring parenteral feeding. Other important clinical features include hypoglycaemia and neuroendocrine effects (including diabetes insipidus) due to a failure to process a number of prohormones. Hyperphagia and severe obesity tend to become apparent by 2–3 years of age [49]. Measurement of the ratio of immature proinsulin to mature insulin is a useful diagnostic test for this disorder and potentially for homozygous mutations in Carboxypeptidase E, which have been reported in one adult female with overlapping clinical features [50].

MC4R Deficiency

Heterozygous loss of function mutations in *MC4R* represent the most common genetic form of obesity and assessment of the sequence of the MC4R is increasingly seen as a necessary part of the clinical evaluation of the severely obese child [7]. The prevalence of pathogenic *MC4R* mutations ranges from 0.5% to 2.5% of people with a BMI >30 kg/m^2 in UK and European populations to 5% in patients with severe childhood obesity [7, 51]. Over 300 different heterozygous mutation in MC4R have been identified to date (https://www.mc4r.org.uk). Codominance, with modulation of expressivity and penetrance of the phenotype, is the most appropriate descriptor for the mode of inheritance. Homozygous mutations in *MC4R* have been identified in children from consanguineous families.

The clinical features of MC4R deficiency include hyperphagia in early childhood and accelerated linear growth, which may be

a consequence of disproportionate early hyperinsulinemia [52]. Reduced sympathetic nervous system activity in MC4R-deficient patients is likely to explain the lower prevalence of hypertension and lower systolic and diastolic blood pressures seen in adults [53]. Thus, leptin-melanocortin signalling appears to play an important role in the regulation of blood pressure and its coupling to changes in weight. Several studies have shown that adolescents and adults with heterozygous *MC4R* mutations respond to Roux-en-Y-bypass surgery [54]. Moreover, the GLP-1 agonist, Liraglutide, can also be effective at inducing weight loss in some patients with MC4R deficiency [55]. As most patients are heterozygotes with one functional allele intact, it is possible that small molecule MC4R agonists [56] or pharmacological chaperones which improve receptor trafficking to the cell surface might be appropriate treatments for this disorder in the future.

SH2B1 Deficiency

Severe obesity without developmental delay is associated with a significantly increased burden of rare, typically singleton copy number variants (deletions/duplications) [57]. Deletion of a 220-kb segment of 16p11.2 is associated with highly penetrant familial severe early-onset obesity and severe insulin resistance [6]. This deletion includes a small number of genes, one of which is *SH2B1*, known to be involved in leptin, insulin, and BDNF signalling. These patients gain weight in the first years of life, with hyperphagia and fasting plasma insulin concentrations that are disproportionately elevated compared to age- and obesity-matched controls. Loss-of-function mutations in the *SH2B1* gene have been reported in association with early-onset obesity, severe insulin resistance, and behavioural abnormalities including aggressive behaviour [58, 59].

KSR2 Deficiency

To date, most of the genetic obesity syndromes are characterized by hyperphagia as a major driver of the obesity. Heterozygous mutations in *KSR2* (Kinase Suppressor of Ras2) are associated with increased food-seeking behaviour in childhood and low basal metabolic rate (BMR) in the presence of normal thyroid function [60]. Clinical reports suggested that some carriers of *KSR2* mutations experienced marked weight loss in childhood when prescribed the antidiabetic drug metformin (for severe insulin resistance) although the cellular mechanisms underlying these effects have not been fully explored to date.

Rare Genetic Variants That Contribute to Severe Childhood Obesity

There is increasing evidence that rare genetic variants that do not follow Mendelian patterns of inheritance may also contribute to severe obesity in some patients. The class 3 semaphorins (SEMA3A-G) direct the development of gonadotropin-releasing hormone (GnRH) neurons into the hypothalamus [61, 62] and rare variants that disrupt SEMA3 signalling are associated with hypogonadotropic hypogonadism [62, 63]. In a recent study, rare loss-of-function variants in the genes encoding SEMA3 ligands, receptors and coreceptors were found to be enriched in severely obese cases compared to healthy controls [64]. These variants altered the function of these proteins through multiple molecular mechanisms which led to aberrant development of hypothalamic melanocortin circuits that modulate energy homeostasis. However, variants were not fully penetrant and did not segregate with severe obesity in families demonstrating non-Mendelian inheritance.

There are parallels in this emerging pattern of genetic association with hypogonadotropic hypogonadism, where incomplete penetrance and variable expressivity within and across families has been observed [65]. As the number of genes implicated in hypogonadotropic hypogonadism has increased, it has become clear that oligogenic inheritance (i.e. more than one gene mutated in the same individual) can in part explain these observations. Indeed, some of the obese adolescents with rare variants in these genes had hypogonadotropic hypogonadism. Additionally, rare heterozygous loss-of-function variants in *SEMA3C* and *SEMA3D* have been associated with Hirschsprung's disease [66], a disorder characterized by failure of development of parasympathetic ganglion cells in the large intestine. Several obese patients carrying these variants had severe medication-resistant constipation in childhood with multiple hospital admissions, suggesting that these variants contribute to the spectrum of phenotypes seen.

Another example of rare variants that do not follow Mendelian inheritance but seem to contribute to a distinct clinical phenotype which includes severe obesity, involves Steroid receptor coactivator (SRC)-1, which belongs to a family of coactivators (SRC-1, -2, and -3) that regulate gene expression downstream of nuclear hormone receptors and other transcription factors. SRC-1 is abundantly expressed in neurons in the arcuate nucleus of the hypothalamus which respond to leptin. In a recent study, SRC-1 was shown to modulate leptin's ability to suppress appetite by stimulating POMC expression by interacting with phosphorylated STAT3 [67]. Rare heterozygous variants in *SRC-1* identified in severe childhood-onset obesity impaired this signalling pathway in cells, where they had a dominant negative effect. In this study, a mouse model of a human loss-of-function variant developed obesity indicating the physiological relevance of this mechanism.

Another gene in which rare variants have been found in severely obese patients and where strong biological data supports their physiological relevance is *MRAP2* (Melanocortin Receptor Accessory Protein 2), a chaperone that alters expression of MC4R [68]. Collectively, these studies indicate some of the challenges associated with establishing the contribution of rare genetic variants to clinical phenotypes. However, such mechanistic studies can lead to new understanding and can pave the way for mechanism-based approaches to therapy for patients.

Conclusions

The diagnosis of a genetic obesity syndrome can provide information that has diagnostic value for the family to whom genetic counselling can be provided. There is particular value in a genetic diagnosis in severe obesity which, unlike other clinical disorders, may not be recognized as a medical condition by some healthcare professionals and educators. A genetic diagnosis can help children and their families deal with the social stigma that comes with severe obesity and where severe obesity has been considered a reason to invoke parental neglect, the making of a genetic diagnosis can prevent

children from being taken into care. A genetic diagnosis can inform management (many such patients are relatively refractory to weight loss through changes in diet and exercise) and can inform clinical decision making regarding the use of bariatric surgery. Importantly, some genetic obesity syndromes are now treatable.

REFERENCES

1. Wardle J, Carnell S, Haworth CM, Plomin R . Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment . *Am J Clin Nutr*, 2008; **87**: 398–404.
2. van der Klaauw AA, Farooqi IS . The hunger genes: pathways to obesity. *Cell*, 2015; **161**: 119–32.
3. Bray GA. From farm to fat cell: why aren't we all fat? *Metab Clin Exp*, 2015; **64**: 349–53.
4. Bochukova E, Schoenmakers N, Agostini M, *et al*. A mutation in the thyroid hormone receptor alpha gene. *N Engl J Med*, 2012; **366**: 243–9.
5. Styne DM, Arslanian SA, Connor EL, *et al*. Pediatric obesity-assessment, treatment, and prevention: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*, 2017; **102**: 709–57.
6. Bochukova EG, Huang N, Keogh J, *et al*. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, 2010; **463**: 666–70.
7. Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, O'Rahilly S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene . *N Engl J Med*, 2003; **348**: 1085–95.
8. Farooqi IS, Wangensteen T, Collins S, *et al*. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. *N Engl J Med*, 2007; **356**: 237–47.
9. Cassidy SB, Schwartz S, Miller JL, Driscoll DJ . Prader–Willi syndrome . *Genet Med,* 2012; **14**: 10–26.
10. Goldstone AP. Prader–Willi syndrome: advances in genetics, pathophysiology and treatment . *Trends Endocrinol Metab*, 2004; **15**: 12–20.
11. Tauber M, Diene G, Molinas C . Sequelae of GH treatment in children with PWS . *Pediatr Endocrinol Rev*, 2016; **14**: 138–46.
12. Sahoo T, del Gaudio D, German JR, *et al*. Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet*, 2008; **40**: 719–21.
13. de Smith AJ, Purmann C, Walters RG, *et al*. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet*, 2009; **18**: 3257–65.
14. Bochukova EG, Lawler K, Croizier S, *et al*. A transcriptomic signature of the hypothalamic response to fasting and BDNF deficiency in Prader–Willi syndrome. *Cell Rep*, 2018; **22**: 3401–8.
15. Manning KE, Holland AJ . Puzzle pieces: neural structure and function in Prader–Willi syndrome . *Diseases*, 2015; **3**: 382–415.
16. Swaab DF, Purba JS, Hofman MA . Alterations in the hypothalamic paraventricular nucleus and its oxytocin neurons (putative satiety cells) in Prader–Willi syndrome: a study of five cases . *J Clin Endocrinol Metab*, 1995; **80**: 573–9.
17. Tauber M, Boulanouar K, Diene G, *et al*. The use of oxytocin to improve feeding and social skills in infants with Prader–Willi syndrome. *Pediatrics*, 2017; 139: pii: e20162976.
18. Cummings DE, Purnell JQ, Vaisse C, *et al*. Elevated plasma ghrelin levels in Prader Willi syndrome. *Nat Med*, 2002; **8**: 643–4.
19. Yeo GS, Connie Hung CC, Rochford J, *et al*. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci*, 2004; **7**: 1187–9.
20. Gray J, Yeo GS, Cox J J, *et al*. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*, 2006; **55**: 3366–71.
21. Weinstein LS, Chen M, Liu J . Gs(alpha) mutations and imprinting defects in human disease . *Ann N Y Acad Sci*, 2002; **968**: 173–97.
22. Tafaj O, Juppner H . Pseudohypoparathyroidism: one gene, several syndromes . *J Endocrinol Invest*, 2017; **40**: 347–56.
23. Hendricks AE, Bochukova EG, Marenne G, *et al*. Rare variant analysis of human and rodent obesity genes in individuals with severe childhood obesity. *Sci Rep*, 2017; **7**: 4394.
24. Ramachandrappa S, Raimondo A, Cali AM, *et al*. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest*, 2013; **123**: 3042–50.
25. Moir L, Bochukova EG, Dumbell R, *et al*. Disruption of the homeodomain transcription factor orthopedia homeobox (Otp) is associated with obesity and anxiety. *Mol Metab*, 2017; **6**: 1419–28.
26. Kublaoui BM, Gemelli T, Tolson KP, Wang Y, Zinn AR . Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice. *Mol Endocrinol*, 2008; **22**: 1723–34.
27. Forsythe E, Beales PL . Bardet-Biedl syndrome . *Eur J Hum Genet*, 2013; **21**: 8–13.
28. Borman AD, Pearce LR, Mackay DS, *et al*. A homozygous mutation in the TUB gene associated with retinal dystrophy and obesity. *Hum Mutat*, 2014; **35**: 289–93.
29. Seo S, Guo DF, Bugge K, Morgan DA, Rahmouni K, Sheffield VC . Requirement of Bardet-Biedl syndrome proteins for leptin receptor signaling . *Hum Mol Genet*, 2009; **18**: 1323–31.
30. Saeed S, Bonnefond A, Tamanini F, *et al*. Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet*, 2018; **50**: 175–9.
31. Siljee JE, Wang Y, Bernard AA, *et al*. Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat Genet*, 2018; **50**: 180–5.
32. Andermann ML, Lowell BB . Toward a wiring diagram understanding of appetite control . *Neuron*, 2017; **95**: 757–78.
33. Gautron L, Elmquist JK, Williams KW . Neural control of energy balance: translating circuits to therapies . *Cell*, 2015; **161**: 133–45.
34. Cowley MA, Smart JL, Rubinstein M, *et al*. Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature*, 2001; **411**: 480–4.
35. Cowley MA, Pronchuk N, Fan W, Dinulescu DM, Colmers WF, Cone RD . Integration of NPY, AGRP, and melanocortin signals in the hypothalamic paraventricular nucleus: evidence of a cellular basis for the adipostat . *Neuron*, 1999; **24**: 155–63.
36. O'Rahilly S, Farooqi IS . Human obesity as a heritable disorder of the central control of energy balance . *Int J Obes (Lond)*, 2008; **32 Suppl 7**: S55–61.
37. Montague CT, Farooqi IS, Whitehead JP, *et al*. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*, 1997; **387**: 903–8.
38. Farooqi IS, Matarese G, Lord GM, *et al*. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest*, 2002; **110**: 1093–103.

39. Ozata M, Ozdemir IC, Licinio J . Human leptin deficiency caused by a missense mutation: multiple endocrine defects, decreased sympathetic tone, and immune system dysfunction indicate new targets for leptin action, greater central than peripheral resistance to the effects of leptin, and spontaneous correction of leptin-mediated defects . *J Clin Endocrinol Metab*, 1999; **84**: 3686–95.
40. von Schnurbein J, Heni M, Moss A, *et al*. Rapid improvement of hepatic steatosis after initiation of leptin substitution in a leptin-deficient girl. *Horm Res Paediatr*, 2013; **79**: 310–17.
41. Strobel A, Issad T, Camoin L, Ozata M, Strosberg AD . A leptin missense mutation associated with hypogonadism and morbid obesity . *Nat Genet*, 1998; **18**: 213–15.
42. Farooqi IS, Jebb SA, Langmack G, *et al*. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med*, 1999; **341**: 879–84.
43. Clement K, Biebermann H, Farooqi IS, *et al*. MC4R agonism promotes durable weight loss in patients with leptin receptor deficiency. *Nat Med*, 2018; **24**: 551–5.
44. Wabitsch M, Funcke JB, Lennerz B, *et al*. Biologically inactive leptin and early-onset extreme obesity. *N Engl J Med*, 2015; **372**: 48–54.
45. Clement K, Vaisse C, Lahlou N, *et al*. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*, 1998; **392**: 398–401.
46. Krude H, Biebermann H, Schnabel D, *et al*. Obesity due to proopiomelanocortin deficiency: three new cases and treatment trials with thyroid hormone and ACTH4-10. *J Clin Endocrinol Metab*, 2003; **88**: 4633–40.
47. Kuhnen P, Clement K, Wiegand S, *et al*. proopiomelanocortin deficiency treated with a melanocortin-4 receptor agonist. *N Engl J Med*, 2016; **375**: 240–6.
48. Lee YS, Challis BG, Thompson DA, *et al*. A POMC variant implicates beta-melanocyte-stimulating hormone in the control of human energy balance. *Cell Metab*, 2006; **3**: 135–40.
49. Jackson RS, Creemers JW, Ohagi S, *et al*. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*, 1997; **16**: 303–6.
50. Alsters SI, Goldstone AP, Buxton JL, *et al*. Truncating homozygous mutation of carboxypeptidase E (CPE) in a morbidly obese female with type 2 diabetes mellitus, intellectual disability and hypogonadotrophic hypogonadism. *PLoS One*, 2015; **10**: e0131417.
51. Stutzmann F, Tan K, Vatin V, *et al*. Prevalence of melanocortin-4 receptor deficiency in Europeans and their age-dependent penetrance in multigenerational pedigrees. *Diabetes*, 2008; **57**: 2511–18.
52. Martinelli CE, Keogh JM, Greenfield JR, *et al*. Obesity due to melanocortin 4 receptor (MC4R) deficiency is associated with increased linear growth and final height, fasting hyperinsulinemia, and incompletely suppressed growth hormone secretion. *J Clin Endocrinol Metab*, 2011; **96**: E181–8.
53. Greenfield JR, Miller JW, Keogh JM, *et al*. Modulation of blood pressure by central melanocortinergic pathways. *N Engl J Med*, 2009; **360**: 44–52.
54. Hatoum IJ, Stylopoulos N, Vanhoose AM, *et al*. Melanocortin-4 receptor signaling is required for weight loss after gastric bypass surgery. *J Clin Endocrinol Metab*, 2012; **97**: E1023–31.
55. Iepsen EW, Zhang J, Thomsen HS, *et al*. Patients with obesity caused by melanocortin-4 receptor mutations can be treated with a glucagon-like peptide-1 receptor agonist. *Cell Metab*, 2018; **28**: 23–32 e3.
56. Collet TH, Dubern B, Mokrosinski J, *et al*. Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency. *Mol Metab*, 2017; *6*: 1321–9.
57. Wheeler E, Huang N, Bochukova EG, *et al*. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet*, 2013; **45**: 513–17.
58. Doche ME, Bochukova EG, Su HW, *et al*. Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest*, 2012; **122**: 4732–6.
59. Jiang L, Su H, Keogh JM, *et al*. Neural deletion of Sh2b1 results in brain growth retardation and reactive aggression. *FASEB J*, 2018; **32**: 1830–40.
60. Pearce LR, Atanassova N, Banton MC, *et al*. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell*, 2013; **155**: 765–77.
61. Hanchate NK, Giacobini P, Lhuillier P, *et al*. SEMA3A, a gene involved in axonal pathfinding, is mutated in patients with Kallmann syndrome. *PLoS Genet*, 2012; **8**: e1002896.
62. Cariboni A, Andre V, Chauvet S, *et al*. Dysfunctional SEMA3E signaling underlies gonadotropin-releasing hormone neuron deficiency in Kallmann syndrome. *J Clin Invest*, 2015; **125**: 2413–28.
63. Young J, Metay C, Bouligand J, *et al*. SEMA3A deletion in a family with Kallmann syndrome validates the role of semaphorin 3A in human puberty and olfactory system development. *Hum Reprod*, 2012; **27**: 1460–5.
64. van der Klaauw AA, Croizier S, Mendes de Oliveira E, *et al*. Human semaphorin 3 variants link melanocortin circuit development and energy balance. *Cell*, 2019; **176**: 729–42.e18.
65. Cassatella D, Howard SR, Acierno JS, *et al*. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*, 2018; **178**: 377–88.
66. Jiang Q, Arnold S, Heanue T, *et al*. Functional loss of semaphorin 3C and/or semaphorin 3D and their epistatic interaction with ret are critical to Hirschsprung disease liability. *Am J Hum Genet*, 2015; **96**: 581–96.
67. Yang A, van der Klaauw A A, Zhu L, *et al*. Steroid receptor coactivator-1 modulates the function of pomc neurons and energy homeostasis. *Nat Commun*, 2019; **10**: 1718.
68. Asai M, Ramachandrappa S, Joachim M, *et al*. Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science*, 2013; **341**: 275–8.

14.1.7 Management of Obesity in Children and Young People

Billy White and Russell M. Viner

Introduction

The role of medical practitioners assessing children and young people (CYP) with obesity is not limited to managing endocrine causes, comorbidities, and complications of obesity. These are relatively rare in childhood, whereas the wider medical, psychological, and social needs of those with obesity can be significant and often unmet. Medical professionals thus need to consider the holistic needs of those with obesity, including managing cardiovascular risk and the full range of obesity-related complications.

Body mass index (BMI, kg/m^2) is the best measure of adiposity for routine clinical use and should be calculated for all children and plotted on appropriate national or international (e.g. International Obesity Task Force) BMI growth charts. Country-specific BMI thresholds are based on normative population data, with 95th or 98th centiles of data obtained prior to the recent obesity epidemic commonly used as thresholds. There is less consensus on definition of severe obesity, but 120% of the 95th centile and a standard deviation score of 3.33 are two proposed BMI-based thresholds.

Obesity-related complications include type 2 diabetes, dyslipidaemia, hypertension, polycystic ovary syndrome, sleep apnoea, non-alcohol fatty liver disease and some cancers, including endometrial, liver, oesophageal adenocarcinoma, colorectal carcinoma, renal carcinoma and breast cancer. The metabolic syndrome, which is variably defined but encompasses central obesity, insulin resistance, dyslipidaemia, and hypertension, plays an important role in several of these, which are discussed in more detail in Chapter 14.1.3, 'Medical Complications of Obesity'.

Assessing Cardiometabolic Risk in Obesity

Early identification and modification of cardiovascular risk factors can reduce later morbidity and mortality and should be a key component of obesity management. Ischaemic heart disease (IHD) is very rare in children whereas IHD risk-factors (hypertension, dyslipidaemia, and type 2 diabetes) are common, particularly in those with severe obesity. In the Bogolusa Heart Study 84%, 59%, and 18% had 1, 2, or 3 of dyslipidaemia, hypertension, and hyperinsulinaemia [1]. Unlike adults, there are no good risk-predictor tools available that predict overall cardiovascular risk based on BMI, blood pressure, glycaemia, and lipids (e.g. qrisk2 in adults). Obesity-related IHD risk is associated with, but not fully explained by, mass of excess adipose tissue. Distribution of fat should be noted, with central adiposity conferring greater IHD risk however measurement of waist circumference is not recommended in children due to limited normative data.

IHD risk-factors frequently cluster in families and history of IHD and IHD risk-factors should be ascertained from first- and second-degree relatives. Early familial IHD and IHD risk-factors increase risk in the child, with doubling of risk where there is IHD affecting first-degree males under 55 years and females under 65.

There is no international consensus on screening for IHD risk-factors. The US 2011 Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents recommends the most proactive screening schedule [2] whereas the UK National Institute for Health and Care Excellence recommends 'consideration' of comorbidity screening but does not suggest which children should be screened nor how.

Hypertension

Hypertension in childhood increases risk of both hypertension and IHD in adult life. Those with severe obesity >99th centile are fourfold more likely than normal-weight children to have hypertension [3]. There are insufficient data to directly stratify cardiovascular disease (CVD) risk by absolute BP in childhood and statistical cut-offs based on normative data of 50 000 CYP are currently recommended [4]. The 95th centile for age and height, or adult hypertension threshold of 135/80 (whichever is lowest) on three separate occasions is used to define stage 1 hypertension, with readings 5 mm greater than the 95th centile defined as stage 2 hypertension. Readings between the 90th and 95th centiles are defined as high-normal.

US guidelines recommend blood pressure measurement at all clinical encounters [5]. Clinic blood pressure measurement is usually performed by the automated oscillometry method; this imputes systolic and diastolic blood pressure based on mean arterial pressure. It can be used to screen for, but not diagnose, hypertension. Oscillatory monitors read higher than manual auscultatory and cannot be used to compare to normative data which has been obtained by the auscultatory method. Therefore, BP measurement should be undertaken after 3–5-minute rest using the correct cuff size (often not available in primary care) and ideally the auscultatory method. It should be repeated three times and the average of last two measurements used. Both arms should be used at the first visit, with the higher value used as reference.

Ambulatory 24-hour blood pressure monitoring provides better real-life monitoring, higher reproducibility, and higher correlation with end-organ damage. American Academy of Pediatrics (AAP) but not European guidelines recommend ambulatory blood pressure monitoring (ABPM) to confirm diagnosis of hypertension, screen for hypertension in high-risk groups who may be normotensive in clinic (masked hypertension) or those with high blood pressure in clinic who are likely normotensive (white coat hypertension) [6]. Home blood pressure monitoring is controversial for diagnosis, but can be used for ongoing monitoring.

Hypertension in those with obesity is more likely to be primary than secondary to another condition, especially in older children. Major guidelines thus recommend only limited screening for rare causes of hypertension. Differentiation between primary and secondary causes of hypertension can usually be achieved by history and examination. All CYP with hypertension should have renal ultrasound, urine microscopy, full blood count and U+E measurement to rule out secondary renal causes, echocardiography (not electrocardiogram (ECG)) and quantification of albuminuria to screen for end-organ damage together with lipid and glucose measurement for concurrent risk stratification.

Dyslipidaemia

Fasting determination provides the best assessment of the lipid profile, particularly for triglyceride concentration, however non-fasting measurement is acceptable, pragmatic, and provides similar readings to fasted levels for total cholesterol and subfractions. Fasting lipid concentrations should be checked at least once where raised non-fasting triglyceride is found, and twice prior to initiation of pharmacotherapy. Low-density lipoprotein (LDL) levels are usually

calculated from total and non-LDL cholesterol values rather than direct assay. Non-high-density lipoprotein (non-HDL) cholesterol concentration is the best single predictor of atherosclerosis, with low levels protective against CVD. LDL above 4.9 mmol/L, or 4.1 mmol/L where there is either family history of hypercholesterolaemia and/or a strong early familial CVD risk, should prompt suspicion of familial hypercholesterolaemia.

Secondary causes of dyslipidaemia should also be considered including liver disease, nephrotic syndrome, pregnancy, drugs (e.g. isotretinoin) and hypothyroidism. Lipid management in CYP is controversial given limited long-term safety data for statins and should be supervised by experts. The pathophysiology and management of dyslipidaemia is discussed in further detail in Chapter 14.2.1, 'Lipoprotein Metabolism' and Chapter 14.2.2, 'Genetic Forms of Dyslipidaemia'.

Insulin Resistance

Puberty and excess adiposity can both cause insulin resistance, which is intimately associated with type 2 diabetes, fatty liver disease, and polycystic ovarian syndrome. It is however practically hard to quantify, and normative data are insufficient [7]. Measures such as fasting insulin concentration and HOMA-IR should not be routinely measured to assess insulin resistance. Acanthosis nigricans is suggestive of insulin resistance but not pathognomonic, and its absence rules out neither insulin resistance nor diabetes.

Type 2 Diabetes

Type 2 diabetes risk is increased by maternal diabetes in pregnancy (gestational or type 2), high-risk ethnicity (e.g. Asian, Latino, African) and first/second-degree family history [8]. Based on this, the American Diabetes Association recommends screening where a child is overweight (BMI >85th centile) with one of these risk factors or signs of insulin resistance [9]. The International Society for Paediatric and Adolescent Diabetes (ISPAD) recommends consideration of type 2 diabetes in those with comorbidities associated with diabetes, particularly obesity, non-alcohol fatty liver disease, elevated triglycerides, and elevated blood pressure.

Type 2 diabetes is best screened for using the oral glucose tolerance test (OGTT) or HbA_{1c}. While each of these methods measure different aspects of glucose control, HbA_{1c} is most closely correlated with diabetes-related complications, is pragmatic in that it can be performed at the time of seeing the patient, and is cheaper than OGTT. An HbA_{1c} of 6–6.4% (38.8–46.4 mmol/mol) is defined as 'pre-diabetes' and >=6.5% as diabetes. The term 'pre-diabetes' is controversial but can be less confusing to patients than the terms impaired glucose tolerance or impaired fasting glucose. Effective diabetes control is considered imperative for minimizing cardiovascular health and any young person diagnosed with type 2 diabetes should be managed by an expert.

Optimizing Weight and Cardiometabolic Health

Diet, physical activity, and physical inactivity influence adiposity and cardiometabolic health yet there is significant interindividual variation [10]. Changes in lifestyle may not correlate with changes in weight and health and weight alone should not be used as a marker of successful lifestyle change.

The conscious behaviour of any individual has a limited influence on bodyweight, with genes, environment, and social psychology all having a greater impact [11]. Weight loss success has often inappropriately been attributed to 'will power', motivation and education but this oversimplifies a complex construct [10]. Homeostatic mechanisms antagonise weight loss through changes in appetite, muscle efficiency and basal metabolic rate [12].

Weight loss targets of 6–8% in adults have been shown to improve health; there are no similar data in children. Weight-loss targets in children vary between national guidelines ranging from no recommended weight loss (Australia) to stratified targets based on BMI centile (USA) and some guidelines avoiding targets altogether. This variation reflects lack of evidence. Cessation of weight gain with ongoing vertical growth will cause reduction in adiposity.

Unrealistic weight-loss targets can put undue pressure on individuals and damage patient-provider relationships. Some are able to lose clinically significant amounts of weight though lifestyle change, yet it is not clear why these individuals are able to lose weight while many are not. An inability to achieve weight control does not necessarily indicate lack of adherence to recommended lifestyle changes. A step-wise approach should be considered to optimise weight and cardiovascular health once comorbidities have been assessed and managed, featuring (1) lifestyle optimisation, (2) drug therapy, and (3) bariatric surgery.

Lifestyle

A. Physical activity: Physical activity improves cardiometabolic fitness, reduces IHD risk, and should be promoted regardless of effect on weight. There is strong evidence that increased moderate-vigorous physical activity is associated with lower blood pressure, BMI, and increased fitness, as well as lowered total and LDL cholesterol, and triglyceride concentrations and insulin resistance, with increased HDL cholesterol. Conversely, there is no evidence for harm associated with increased physical activity. Daily physical activity should be recommended for all, including walking as a pragmatic form of exercise, with 60 minutes a day recommended by most major guidelines.

Back, hip, knee, and ankle pain can limit exercise; this is frequently caused by deconditioning due to inactivity and should lead to assessment and rehabilitation by physiotherapists. Rarely, obesity can cause sciatica, slipped upper femoral epiphysis, or bowed legs (Blount disease). These need effective identification and management.

Breathlessness is common during exercise. This is caused largely by lack of fitness and a supervised fitness training with an exercise professional may be needed to improve capacity to exercise. Exercise-induced asthma and inadequate asthma control can be difficult to differentiate with poor fitness and should be considered as a cause of exercise intolerance.

B. Diet: In the controlled laboratory environment, total caloric intake reduction has the greatest impact on weight. Different approaches exist to reduce overall caloric intake including portion control, reduction of energy-dense foods and macronutrient balance but there is little clear evidence (including insufficient evidence to guide the low-carbohydrate versus low-fat diet debate) [2]. The evidence that does exist is summarised in Box 14.1.7.1. The role of a diet history is unclear. Recall is often poor and the use of food diaries to monitor intake can promote behaviour change and aid discussions with health professionals.

Box 14.1.7.1 Dietary interventions for which some evidence of benefit exists

Strongest evidence

- Limit fat to 30%, saturated fat to 7–10% of total calories, and cholesterol to 300 mg/day (lowers total—and LDL—cholesterol, obesity, and insulin resistance)
- Consume fat-free milk
- Consume plant-based foods (low-calorie sources of nutrients including vitamins and fibre). Increasing access to fruit and vegetables has been shown to increase their intake

Moderate evidence

- Reduce sugar-sweetened beverages (reduces obesity)
- Increase dietary fibre (intake is inversely associated with energy density and increased body fat, and positively associated with nutrient density)
- DASH (Dietary Approaches to Stop Hypertension; a promising approach to improve nutrition and decrease cardiovascular risk)
- Reduce simple carbohydrates (associated with decreased weight and triglyceride concentration)

C. Lifestyle interventions: Evidence-based group lifestyle interventions can support behaviour change [13]. They are mostly community-based and have been trialled largely in children rather than adolescents and in academic centres. These studies are usually of short duration (<6 months) and the long-term evidence for these interventions is non-existent. Group lifestyle interventions are not widely available, with limited funding by either insurance or state-funded health systems. Some adult commercial programmes, such as Weight Watchers and Slimming World, allow adolescents to enrol with an accompanying adult and are more widely available. Some find the regular supervision and education that these programmes offer to be a helpful way of controlling weight, however, many families do not wish to take part in such interventions.

There is strong evidence for smoking avoidance (and cessation) in improving health. There is less evidence on how to effectively prevent smoking initiation or support cessation.

Pharmacotherapy

Use of weight-loss medications in CYP is rare [14]. Only the lipase-inhibitor orlistat is licenced for CYP in the EU. Within clinical trials it can result in modest BMI loss and is recommended in the UK for use in those over 12 years of age with severe obesity. Outside research trials, few tolerate the side-effects associated with fat ingestion (abdominal pain, urgency, nausea, incontinence) and the vast majority do not request a second prescription. It can be used as a feedback tool to identify fat intake in those willing to cut energy consumption by reducing fat intake.

Metformin is used off-licence, largely by endocrinologists and gynaecologists in those with insulin resistance and/or polycystic ovarian syndrome [15]. A mean change of 1.42 kg/m² was detected in a meta-analysis of small studies. It is a first line drug for those with polycystic ovary syndrome (PCOS) and can improve menstrual irregularity. Its role in diabetes prevention is less clear and it should not routinely be used for this indication.

The newest generation of weight control medications have all been licenced for other indications prior to being used for weight-reduction. Their weight beneficial effects have led to subsequent licencing for weight control, namely the GLP-1 agonist liraglutide and the combinations of naltrexone/bupropion and phentermine/topiramate. They have not been licenced to date for CYP.

Bariatric Surgery

Bariatric surgery is the only current evidence-based intervention that results in large magnitudes of weight loss (mean weight loss 13.5 kg/m² at 1 year) [16]. Furthermore, bariatric surgery appears to fundamentally change gut physiology to reduce appetite, increase satiety, and enable weight loss with concurrent control of comorbidities such as type 2 diabetes, hypertension and obstructive sleep apnoea (OSA). Bariatric surgery improves many of these comorbidities prior to significant weight loss suggesting that both the physiological changes achieved by surgery and the acute calorie reduction that it allows ameliorate these comorbidities. Physiological changes include increased and faster incretin production after food ingestion (including GLP-1) and changes in gut microbiota, vagal tone, and bile acid circulation.

Three main procedures have been used in adolescents: the adjustable gastric band (AGB), sleeve gastrectomy (SG) and gastric bypass (GB). All three result in a smaller functioning stomach. In the AGB, a silicon inflatable band is placed around the stomach and inflated via a subcutaneous port. In the SG, the greater curvature is removed to reduce stomach size by up to 80%. In the GB, the stomach is divided to form a 30 ml functioning stomach which is connected directly to the jejunum. The distal end of the 'defunctioned' proximal gut (known as the Roux limb) is anastomosed to the jejunum distally to the stomach-jejunum anastomosis to allow bile acids and stomach products to re-join the functioning gut.

These procedures have not been compared in head to head trials in this age group. Only the AGB has undergone a randomized control trial which showed greater weight loss and improved CV health compared to a group lifestyle intervention. Multiple cohort studies have published outcome data for the three procedures, with greatest weight loss at one year seen in those undergoing GB and least loss after AGB. Two high quality cohorts have published outcome data for the SG and GB at 3–5 years with outcomes similar to those seen in adults and between the two procedures [17].

Weight loss is greatest in the first 6 months and plateaus after 1–2 years with a normal distribution of weight change after surgery which includes total weight gain. The evidence underpinning selection of patients for surgery is very limited. Risk alleles including FTO appear to influence outcome but are not routinely used to select patients or match patients with procedure. There is minimal evidence for psycho-social predictors of BMI outcomes with higher levels of loss of control eating and family conflict conferring worse outcomes in one study [18]. Cessation of weight gain, appropriate attendance at appointments, prior engagement with a lifestyle intervention, appropriate mental health support, and good social support are common selection criteria by clinical teams.

Most national guidelines limit surgery to those with near completion of growth due to theoretical concerns about bone density loss after surgery. BMI thresholds of 40 kg/m², or 35 kg/m² with obesity-related comorbidities are common with the largest cohorts routinely operating on young people from around 12 years of age.

The AGB is the easiest procedure to perform however long-term risk of spontaneous gastric perforation, need for ongoing adjustment of band pressure and lesser weight loss compared to other procedures mean it is now rarely advocated. Choice of procedure

between the SG and GB is controversial with the SG increasingly being seen as the procedure of choice in this age group, given it is a less invasive and quicker procedure than the GB and has lower risk of micronutrient deficiency and stenosis. Both the SG and GB are deemed safe, with no surgery-related mortality reported in adolescents. Complications include strictures requiring endoscopic dilatation, symptomatic gallstones requiring cholecystectomy and vitamin deficiencies (particularly fat-soluble vitamins).

Weight loss after surgery can result in excess skin which can be both physically and emotionally distressing. Funding for removal of excess skin (apronectomy) is usually more limited than bariatric surgery, both in insurance- and state-funded health systems and plastic surgery itself can be associated with greater risk than bariatric surgery.

Mental health outcomes have been less widely reported than those of physical health with limited data beyond anxiety and depressive disorders [18]. Both appear to follow a similar trajectory to weight with improvement in the first 6 months and subsequent plateauing and some deterioration thereafter. Many patients believe that weight loss will dramatically improve their lives: weight loss can have significant improvement in quality of life, with greatest improvement in physical functioning domains. However, evidence for improvements in other domains is less clear and many need ongoing mental health support after surgery, particularly those whose weight loss is less than they wished.

Optimizing Other Aspects of Health Related to Obesity

Health professionals often inappropriately focus on weight loss to control obesity-related poor health while avoiding effective treatments. While weight loss can make many of these conditions better, it is often difficult to achieve, and poor health can further exacerbate this. Obesity-related poor health should be proactively medically managed concurrently with lifestyle change.

Polycystic Ovarian Syndrome

Diagnosis of polycystic ovary syndrome relies on the combination of clinical hyperandrogenism (hirsutism, severe acne), menstrual irregularity, and polycystic ovaries. Use of these criteria in adolescence is controversial as these features are frequently part of normal puberty [19]. Some recommend use of the adult Rotterdam criteria but this may lead to overdiagnosis and inappropriate worry related to future infertility. Other causes of androgen excess should be considered where there is severe hyperandrogenism or menstrual dysfunction, including congenital adrenal hyperplasia or steroid excess.

Treatment is largely symptomatic, focusing on control of clinical hyperandrogenism and menstrual dysfunction. Lifestyle change and weight loss can improve insulin resistance and menstrual irregularity (see next). Hormonal contraceptives are recommended first line treatments for cutaneous symptoms (acne, hirsutism) with careful consideration of venous thrombotic risk (combined oestrogen/progestogen contraceptives and obesity cumulatively increase risk factors for venous thromboembolism (VTE)). Improvement is usually slow, with acne not improving for at least 1 month and hirsutism for around 6 months. Metformin is additionally recommended where there are signs of metabolic syndrome. Hirsutism control can be challenging and treatments include topical ornithine decarboxylase inhibitors (Vaniqa), spironolactone, antiandrogens, and laser treatment.

Non-Alcohol Fatty Liver Disease (NAFLD)

NAFLD is usually asymptomatic and classic clinical features of liver disease are rare. Accumulation of fat in the liver (steatosis) can in some result in the progressive spectrum of fibrosis, cirrhosis, and hepatocellular carcinoma, collectively known as non-alcohol steatohepatitis (NASH). This is not directly correlated to volume of steatosis and presence of fibrosis is the best predictor of liver-related morbidity and mortality. There are limited treatment options with the exception of liver transplant in those with end stage disease which is rare in childhood.

Screening in childhood is problematic given inadequately validated screening tools and very low prevalence of those who would benefit from liver transplant. Liver ultrasound and liver enzymes have both been advocated as first line screening tools, yet both have significant issues with sensitivity. Ultrasound does not detect fibrosis or inflammation but can identify steatosis and may identify signs of significant fat-related liver disease such as nodularity or splenomegaly and other causes of deranged liver function tests. Accepted tools in adult practice (elastography using fibroscan and MRI) have not yet been validated in CYP and fibrosis markers such as ELF (enhanced fibrosis score) again have limited evidence base. Only liver biopsy can reliably differentiate between steatosis, hepatosteatosis, and fibrosis.

NAFLD is highly correlated with insulin resistance and metabolic syndrome and European guidelines suggest screening only this group [20]. In the UK, NICE additionally recommends screening in those with type 2 diabetes [21]. Those from Hispanic backgrounds have the highest risk with black ethnicity being protective [22].

The discovery of either steatosis or raised liver enzymes, either incidentally or by screening, should lead to comprehensive evaluation of secondary causes rather than assumed NAFLD, including hepatitis B and C, excess alcohol consumption, autoimmunity, Wilson, and a1-antitripysin [20].

Sleep

There is increasing evidence that inadequate or poor-quality sleep can both cause and complicate obesity, sometimes contributing to enuresis, learning problems, headaches, inattention, daytime somnolence, preference for high-calorie foods, exercise avoidance, ventricular dysfunction, and/or hypertension [23]. Obesity and adeno-tonsillar hypertrophy are the major risk factors for OSA, particularly in adolescence and with increasing adiposity [24]. Snoring is increasingly considered to be a negative predictor of good health. Some patients with OSA have the typical features of snoring with difficulty breathing and apnoeas, but many do not, making diagnosis difficult. Severe OSA can exist without snoring and a high degree of clinical suspicion is needed including consideration of both day and night symptoms, physical signs, and attention to comorbidities. Factors increasing the likelihood of OSA are shown in Box 14.1.7.2.

Inpatient polysomnography is the gold-standard investigation for the diagnosis of OSA. Overnight ambulatory saturation monitoring is poor at excluding OSA (low negative predictor value) [25]. Where OSA is suspected or diagnosed, treatments include

Box 14.1.7.2 Clinical factors indicating increased probability of obstructive sleep apnoea

Night symptoms: snoring, apnoea, difficulty breathing, enuresis

Day symptoms: morning headache, difficulty in rousing, inattention, poor behaviour, somnolence, mouth breathing (including dry mouth)

Comorbidities: sinusitis, poorly controlled asthma

Signs: anterior neck adiposity, airway abnormalities (e.g. trisomy 21, craniofacial anomalies), mouth breathing, adeno-tonsillar hypertrophy

Box 14.1.7.3 Commonly used drugs that promote weight gain

Mental health: amitriptyline, sertraline, paroxetine, haloperidol, risperidone, olanzapine, quetiapine, lithium

Diabetes: insulin, gliclazide, pioglitazone

Anticonvulsants: sodium valproate, carbamazepine, gabapentin

Steroid hormones: prednisolone, high dose inhaled steroids, some combined contraceptives

Cardiac: β-blockers

adeno-tonsillectomy, weight loss (including bariatric surgery), treatment of airway disease, and nocturnal ventilation. Secondary sequalae such as left ventricular hypertrophy should be excluded by echocardiogram where OSA is severe.

Mental Health

Effective mental health support can be critical in weight control, for example binge eating or affective disorders. Mood disturbances are more common in obesity and may be associated with obesity-related inflammation. Conversely, pharmacotherapy for mental health problems can cause weight gain, particularly if involving older generation antipsychotics.

Rapid weight gain is more likely to be caused by adverse mental health than by a medical cause, especially where normal growth is maintained. Symptoms suggestive of medical causes of obesity including nausea and headaches are also common in mental health and can render diagnosis difficult.

Obtaining a mental health history can be challenging. It should usually be undertaken in part with the young person alone. Obtaining effective mental health support can also be difficult, with families often not wanting to engage with mental health professionals, while mental health resources are often insufficient to support those who do wish to. Use of validated questionnaires can help bypass the often-stringent triage processes employed by many mental health services.

Binge eating disorders are relatively rare and hard to diagnose. Questionnaires such as the Eating Disorder Diagnostic Scale (EDDS) or the Eating Disorder Examination Questionnaire (EDE-Q) can help identify key features of a binge eating disorder, including eating large volumes of food with loss of control, associated guilt, and compensatory behaviours such as purging, diuretic use, fasting, and excessive exercise. Depressive and anxiety disorders can also drive excessive eating and can be screened using the Strengths and Difficulties Questionnaire (SDQ) and Revised Child Anxiety and Depression Scale (RCADS) questionnaire. Referral to specialist mental health services should be initiated where any questionnaires suggest a disorder.

Low self-esteem and confidence and lack of motivation to engage in behaviour change are common. Support for this is more challenging and may not require weight-specific interventions. Generic interventions such as the Duke of Edinburgh award, cadets, or dramatic arts can be effective at promoting confidence, socialisation, and behaviour change. Medical conditions reducing self-esteem should be proactively addressed, such as acne and gynaecomastia. Apparent small penis can be distressing for boys practically and psychologically, and should be differentiated from true micropenis and hypogonadism.

Low self-esteem and mental health disorders can limit exercise, particularly if exercise involved changing clothes in public (such as school PE classes), reduced clothing (such as swimming) or mixing with very fit people (such as going to the gym). Strategies to minimise these barriers should be considered including private changing facilities, single sex swimming sessions, or encouraging exercises that require minimal adaptations, such as walking.

Drugs

Certain drugs are weight-promoting and avoidance of these drugs, their replacement with weight-neutral or even weight-sparing drugs, or other treatment options should be considered. Common weight-promoting drugs are shown in Box 14.1.7.3

Activities of Daily Living

The ability to undertake activities of daily living should be checked in those with severe obesity. This includes the ability to reach and tie shoelaces in those BMI over 40 kg/m^2. Many cannot adequately toilet or wash in standard facilities nor comfortably sit/sleep in standard chairs and beds. This information will not usually be proactively shared by families. Many will be sitting/sleeping in broken chairs/beds which will be exacerbating any existing musculoskeletal and sleep abnormalities and specialist beds and chairs can be accessed by occupational therapists. Home assessments and adaptations by occupational therapists should be initiated.

REFERENCES

1. Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH . Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa Heart Study . *J Pediatr*, 2007; **150**: 12–17.e2.
2. Expert Panel on Integrated Guidelines for Cardiovascular Health and Risk Reduction in Children and Adolescents; National Heart, Lung, and Blood Institute . Expert panel on integrated guidelines for cardiovascular health and risk reduction in children and adolescents: summary report. *Pediatrics*, 2011; **128** (Suppl 5): S213–56.
3. Parker ED, Sinaiko AR, Kharbanda EO, *et al*. Change in weight status and development of hypertension. *Pediatrics*, 2016; **137**: e20151662.
4. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents . *Pediatrics*, 2004; **114**(Suppl. 2): 555.
5. Flynn JT, Kaelber DC, Baker-Smith CM, *et al*. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. *Pediatrics*, 2017; **140**: pii: e20171904.
6. Lurbe E, Agabiti-Rosei E, Cruickshank JK, *et al*. 2016 European Society of Hypertension guidelines for the management of high blood pressure in children and adolescents. *J Hypertens*, 2016; **34**: 1887–920.

7. Levy-Marchal C, Arslanian S, Cutfield W, *et al*. Insulin resistance in children: consensus, perspective, and future directions. *J Clin Endocrinol Metab*, 2010; **95**: 5189–98.
8. Viner R, White B, Christie D . Type 2 diabetes in adolescents: a severe phenotype posing major clinical challenges and public health burden . *Lancet,* 2017; **389**: 2252–60.
9. American Diabetes Association. 2 . Classification and diagnosis of diabetes: standards of medical care in diabetes—2018 . *Diabetes Care*, 2018; **41**: S13–S27.
10. Casazza K, Fontaine KR, Astrup A, *et al*. Myths, presumptions, and facts about obesity. *N Engl J Med*, 2013; **368**: 446–54.
11. Finegood DT, Merth TD, Rutter H . Implications of the foresight obesity system map for solutions to childhood obesity . *Obesity (Silver Spring)*, 2010; **18** Suppl 1: S13–16.
12. Richard D. Cognitive and autonomic determinants of energy homeostasis in obesity . *Nat Rev Endocrinol*, 2015; **11**: 489–501.
13. Ho M, Garnett SP, Baur LA, *et al*. Impact of dietary and exercise interventions on weight change and metabolic outcomes in obese children and adolescents: a systematic review and meta-analysis of randomized trials. *JAMA Pediatr*, 2013; **167**: 759–68.
14. Viner RM, Hsia Y, Tomsic T, Wong ICK . Efficacy and safety of anti-obesity drugs in children and adolescents: systematic review and meta-analysis . *Obes Rev*, 2010; **11**: 593–602.
15. Park MH, Kinra S, Ward KJ, White B, Viner RM . Metformin for obesity in children and adolescents: a systematic review . *Diabetes Care*, 2009; **32**: 1743–5.
16. Black J, White B, Viner R, Simmons R . Bariatric surgery for obese children and adolescents: a systematic review and meta-analysis . *Obes Rev*, 2013; **14**: 634–44.
17. Olbers T, Beamish AJ, Gronowitz E, *et al*. Laparoscopic Roux-en-Y gastric bypass in adolescents with severe obesity (AMOS): a prospective, 5-year, Swedish nationwide study. *Lancet Diabetes Endocrinol*, 2017; **5**: 174–83.
18. White B, Doyle J, Colville S, Nicholls D, Viner R, Christie D . Systematic review of psychological and social outcomes of adolescents undergoing bariatric surgery, and predictors of success . *Clin Obes*, 2015; **5**: 312–24.
19. Rosenfield RL. The diagnosis of polycystic ovary syndrome in adolescents . *Pediatrics,* 2015; **136**: 1154–65.
20. European Association for the Study of the Liver (EASL)1; European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease . *J Hepatol*, 2016; **64**: 1388–402.
21. National Institute for Health and Care Excellence. *Non-Alcoholic Fatty Liver Disease (NAFLD): Assessment and Management*. London: National Institute for Health and Care Excellence, 2016.
22. Chalasani N, Younossi Z, Lavine JE, *et al*. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*, 2018; **67**: 328–57.
23. Medic G, Wille M, Hemels MEH . Short—and long-term health consequences of sleep disruption . Nature and Science of Sleep, 2017 ; **9**: 151–61.
24. Marcus CL, Brooks LJ, Draper KA, *et al*. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*, 2012; **130**: 576–84.
25. Kirk V, Baughn J, D'Andrea L, *et al*. American Academy of Sleep Medicine position paper for the use of a home sleep apnea test for the diagnosis of OSA in Children. *J Clin Sleep Med*, 2017; **13**: 1199–203.

14.1.8 Planning Obesity Care Pathways

Nicholas Finer

Introduction

The prevalence of obesity, its protean and relapsing nature, and the need for multidisciplinary involvement in effective care makes planning and organizing care pathways particularly challenging. Public health perspectives have tended to focus on prevention, or short-term interventions, rather than applying chronic care models to obesity which fall under the umbrella of disease management. Despite many authoritative national and international reports, systematic reviews, official recommendations, and guidelines [1–5] the need for people with overweight and obesity to receive medical care remains largely unmet. This chapter considers, mainly from an English perspective, how clinical care pathways for the management of obesity have been developed, and how they could, and should, be implemented, as well as some of the barriers to success. The development of pathways in the United Kingdom has differed from that in the USA [6], being viewed as a way of achieving a continuum of care across care settings rather than as a framework for balancing costs and quality [7].

Obesity as a Chronic Disease

Obesity is a disease of multifactorial origin driven by innate biological factors (genetic predisposition, a physiology favouring energy accumulation and protecting against weight loss), societal influences on food supply (composition, cost, availability, hedonic power) and energy expenditure (declining physical activity at work and at home), psychological factors (beliefs, executive function) and their interactions. Societal changes over the past 40 years have driven the obesity pandemic, with human physiology now maladapted to the environment. The mechanism of weight gain can be considered straightforward (sustained positive energy balance), but unless care pathways address the complex causal network management is likely to be ineffective. Since obesity is a chronic disease, increasingly starting in childhood, a life-course approach is needed [8] and long-term, even lifelong, weight management is likely required [9–11]. Effective treatment of obesity should aim to improve health and wellbeing, rather than simply aiming for weight loss; Comprehensive care of people with obesity also needs to focus on ensuring that the many complications of obesity are diagnosed and

appropriately managed. The components of comprehensive care may include behavioural, nutritional, and physical activity advice and support, as well as pharmacological and surgical interventions. However little high-quality evidence exists on how to combine, either simultaneously or sequentially, interventions or on how to personalize treatment to the individual patient.

Obesity Care Pathways

With this background there is both a clear rationale, and need, to establish obesity care pathways. An (integrated) clinical pathway is a multidisciplinary management tool based on evidence-based best practice for a specific group of patients to aid the coordination and delivery of high-quality care [12]. A pathway will embed guidelines, protocols, and locally agreed patient centred care into everyday use for the individual patient. The pathway will define the roles of professionals involved in patient care. For obesity this will include physicians, dietitians, nurse specialists, surgeons, physiotherapists, exercise specialists, psychologists, and even psychiatrists. Clinical pathways aim to standardize treatment regimens, but record variations from planned care aiming to improve outcomes and allow continual service improvement through evaluation and audit [13]. There is increasing high level evidence evaluating the efficacy (or otherwise) of care pathways [14], but much of this evidence comes from outcomes of trials of pathways focused on relatively predictable trajectories of care where efficacy in supporting proactive care management ensured that relevant clinical interventions and/or assessments were deployed in a timely manner [15]. Clinical care pathways are thus not synonymous with guidelines, which are often more generalized and do not detail the individual patient journey. Since a comprehensive clinical care pathway for the patient with obesity is likely to span levels of care from primary to specialist and involve multiple specialities (from both medical and healthcare professionals), they are most likely to be developed within comprehensive healthcare systems (e.g. the UK National Health Service).

The 5As framework, originally developed as a model for smoking cessation [16] has been adapted by the Canadian Obesity Network to form a model of initiating discussion and behaviour change based upon the principles of motivational interviewing [17, 18]. This considers a process characterized as Ask, Assess, Advise, Agree and Assist (or Arrange) (Figure 14.1.8.1). The impact of using this approach was evaluated in primary care in a 'before and after' study and showed that implementing the 5As of Obesity Management resulted in a twofold increase in the initiation of obesity management (from 19 to 39%), and a significant increase in perceived follow-up/coordination efforts [19]. Thematic network analysis of a subsequent study involving 24 chronic disease teams in Alberta, Canada concluded that the 5As Team intervention had multiple impacts on providers and teams to improve obesity management, as well as provider confidence and capability, both preconditions for developing effective patient interventions [20].

Care Pathway Development in England

In England, there are many examples of obesity care pathways, although many are confined to bariatric surgery for severe obesity. The most overarching is that developed by the National Institute for Health and Care Excellence (NICE) [21]. An overarching design principle was developed for adults within the National Health Service in which pathways are organized under a 4-tiered model (Figure 14.1.8.2) [22, 23]. Tier 1 addresses whole population prevention activity, focusing on primary prevention and includes health promotion, where primary prevention is preventing the onset of a disease. Tier 2 consists of community weight management services, that should include lifestyle advice on diet and physical activity and focus on secondary prevention of conditions associated with overweight and obesity through encouraging behaviour change, where

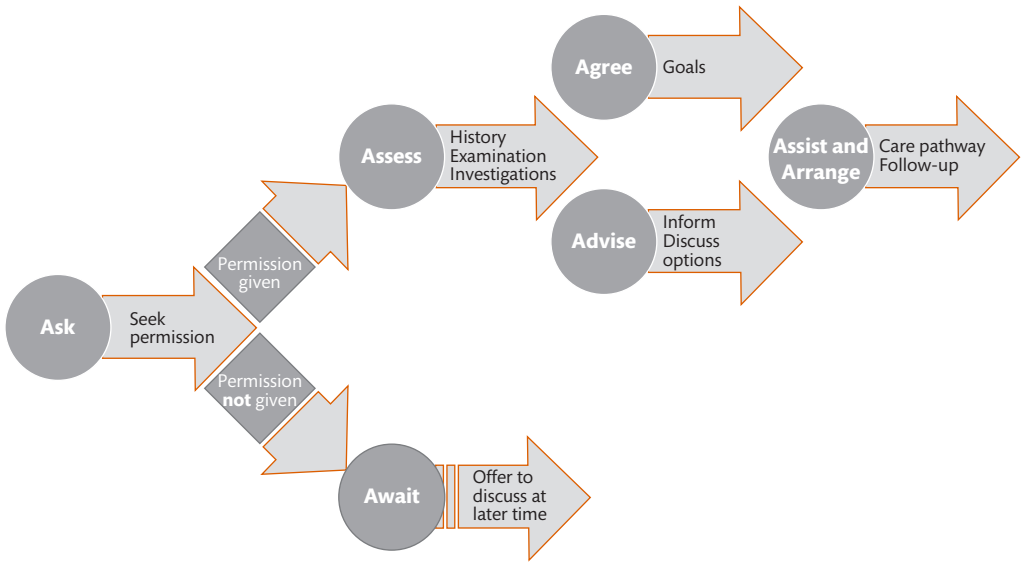


Figure 14.1.8.1 The five 'A's as a template for obesity care. In this modified schema, a sixth A—'Await' reflects that not all patients will at the point of assessment wish to proceed down the care pathway.

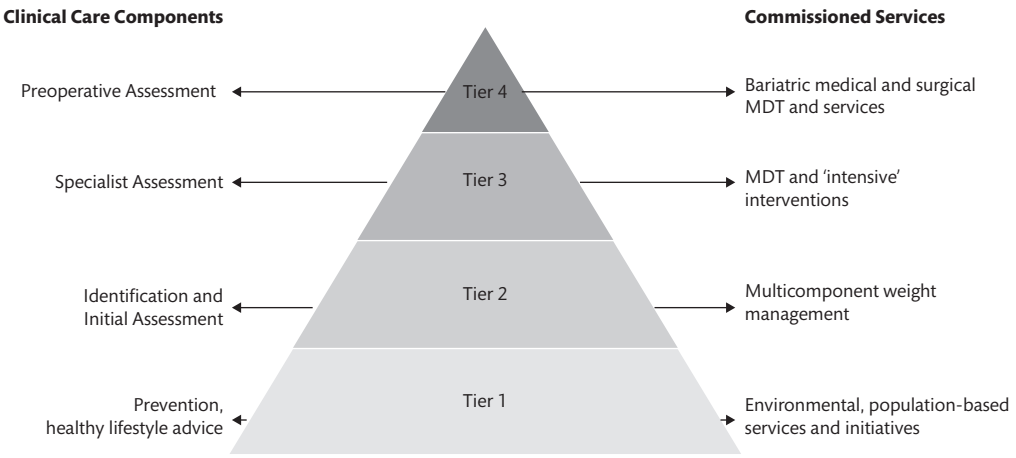


Figure 14.1.8.2 The UK tiered approach to defining the progression of obesity services. MDT, multidisciplinary team.
Adapted with permission from Blackshaw J, Montel S, King S, *et al.* Report of the working group into: Joined up clinical pathways for obesity, 2014.

secondary prevention is preventing progression of early disease. The intention is that both these tiers are funded, organized, and provided by public health experts and from the public healthcare budget. Specialist, medicalized, services are represented by Tiers 3 and 4, commissioned and funded from the health service budget. Tier 3 services are specialist multidisciplinary obesity and weight management services. Access for patients would be through specialist referral routes and engagement, with Tier 3 services a requirement for consideration for surgical services. For people with obesity, this is an example of tertiary prevention, where activities prevent complications from established disease. Tier 4 services include bariatric surgery, but also care for those with genetic or complex disorders.

A practical example of an overall adult weight management obesity pathway and service model is that given in a 2017 report from Public Health England to the East of England Clinical Senate Assembly (one of twelve similar bodies across England established to provide independent, strategic advice and guidance to commissioners and other stakeholders to assist them in healthcare planning for the populations they represent) (Figure 14.1.8.3) [24]. This pathway diverges considerably from current guidelines, including those developed by NICE. This may reflect the economic tool used to estimate the return on investment for weight management services. This calculated that for Tier 2 interventions producing an average weight loss of 2.5 kg in 74% of patients completing a 12-month programme, an upfront cost of £50 per patient enrolled would yield a healthcare saving of £36 per annum over 5 years [25, 26]. The report also highlights the gap between public health perceptions of obesity and its reality as a chronic disease, with a conclusion stating that 'The challenge (of obesity) is not to find new medicines or provide more surgery to patients. It is the population that requires treatment rather than individuals' [24]. From this viewpoint it is not surprising that there are barriers in developing, funding, and

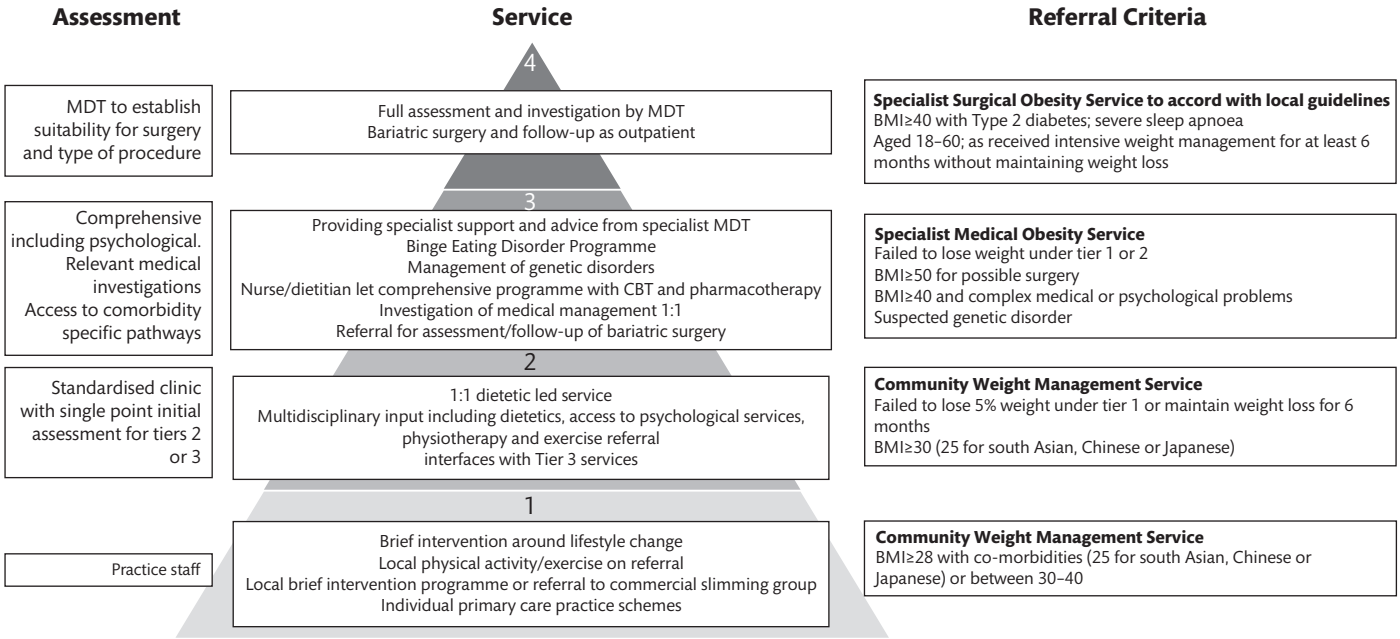


Figure 14.1.8.3 An example from the east of England of delivery of tiered obesity services. MDT, multidisciplinary team.
Adapted with permission from Brett B, Ahmed A. Addressing the Obesity Challenge in the East of England—A report to the East of England Clinical Senate Assembly Public Health England, 2017. Contains public sector information licensed under the Open Government Licence v3.0. [24]

implementing care pathways which by their very definition address individual health needs. Further evidence for such a bias and for underinvestment in obesity care was reported by the recent UK All Party Parliamentary Group on Obesity, who found, following a freedom of information request, that only 52% of local authorities commission Tier 1 services and 82% Tier 2, while 57% of NHS clinical commissioning groups commission Tier 3 services [27].

Recommendations for care pathways for those with severe obesity requiring Tier 3 and 4 services are better developed. Using a National Institute of Health Care Excellence (NICE)-accredited process, a Guidance Development Group representing and endorsed by a wide range of professional stakeholders met to establish the evidence base for best practice in Tier 2 or Tier 3 weight management services, including all interventions provided as a precursor to Tier 4, bariatric surgery. It also issued guidance for the structure and working of weight assessment and management clinics (WAMC) to provide specialist support for primary care physicians in the management of complex patients, including a pathway into bariatric surgery for suitable patients [28].

The resulting report recognized that the guidance was aspirational, and that although the final consensus was tailored to the English NHS, the basic principles should be applicable to every healthcare setting. A key element of the report was recognition that, since the facilities and resources for obesity treatment (especially specialist Tiers 3 and 4) were so poorly developed, capacity building would be required to develop such clinics. The potential overlap with diabetes clinics and sleep medicine clinics meant that some of the infrastructure necessary for WAMCs already existed and potentially could be 're-labelled'. Depending on the availability of bariatric surgical units and the preference of commissioners, WAMCs could be located either in primary or secondary care and work as a hub and spoke arrangement with the surgery service. If colocated, the whole team would in effect work together as one clinic with local arrangements for patient flow through the system for non-surgical and surgical patients (Figure 14.1.8.4). A regional WAMC would also work as a hub and spoke with different referring practices. The report also highlighted the need for hospitals to have a designated 'obesity champion' (along the lines, for example, of a lead clinician for malnutrition) as had first been proposed in a report from the Royal College of Physicians [29].

A subsequent paper updated the evidence base and expanded recommendations to include more detailed guidance on commissioning care pathways for weight assessment and management in adults and children with severe obesity [30]. It represented the views of 22 organizations including nine royal colleges, with the intention of providing a template for developing obesity services within primary and secondary care, for staffing infrastructure, and for referral practices.

Examples of Obesity Care Pathways

There are several examples in England of excellence in the development and provision of obesity care pathways within primary care. These include the Aintree LOSS scheme [31], Rotherham Institute for Obesity (RIO) services [22] and obesity services in Fakenham [32].

Aintree LOSS

Aintree LOSS is a community-based, Tier 3, multidisciplinary weight management programme for patients living in Liverpool with severe and complex obesity (a body mass index >40 kg/m² or >35 kg/m² with one or more comorbidities), focusing on a flexible and individualized service with follow-up for up to 2 years. The team initially consisted of general practitioners, a physician with a special interest in obesity, dieticians, and physiotherapists but was

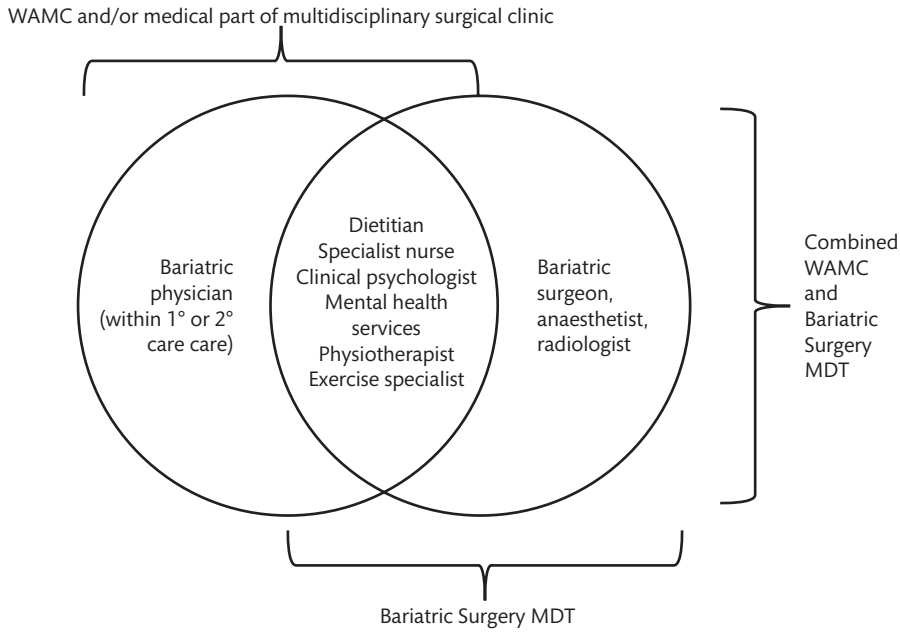


Figure 14.1.8.4 The role of the multidisciplinary team in weight management assessment clinics. MDT, multidisciplinary team; WAMC, weight management assessment clinic.

Adapted with permission from Welbourn R, Dixon J, Barth JH, *et al.* NICE-Accredited Commissioning Guidance for Weight Assessment and Management Clinics: a Model for a Specialist Multidisciplinary Team Approach for People with Severe Obesity. *Obes Surg* 2016;26(3):649–59. Copyright © 2016, Springer [28].

later expanded to include psychologists and occupational therapists. Interventions are delivered at locations across Liverpool, including a hospital clinic, general practice surgeries, community centres, and a sports centre. Following referral, the patient is contacted by telephone and booked to their preferred site. Initial assessment conducted by a GP with a special interest in obesity includes full medical history, examination, and blood tests. A personalized management plan is agreed upon from a pick list of dietetics, physiotherapy, and occupational therapy as well as group sessions (joint physiotherapy and dietetics and hydrotherapy). Individual reviews take place every 1–3 months depending on the intensity of intervention required. Contact with leisure services via swimming and aquafit session referral is also included. In appropriate cases, bariatric education is provided in a group setting, and there are monthly multidisciplinary team meetings to assess suitability for bariatric surgery onward referral. Evaluation of 2472 patients referred to the service between October 2009 and 2013 found a mean body mass index of 45.6 kg/m² at baseline with most patients coming from areas in the most deprived decile nationally. Of 2315 appropriate referrals, 1249 (55.1%) attended >2 visits; mean final weight loss was 3.50 ± 8.55 kg, and 24.1% achieved ≥5% weight loss. Of patients attending, 754 (33.3%) attended for over 6 months; mean final weight loss was 4.94 ± 10 kg, and 34% achieved 5% weight loss [31].

Rotherham Institute for Obesity

In 2008 NHS Rotherham agreed to make £3.5m available to fund their NHS obesity strategy, based upon the four-tier model, for a 3-year pilot period. Reshape Rotherham provided Tier 2 service for adults between 2009 and 2014 overseen by the local dietetics department, while for children this was delivered by Places for People leisure. Patients could self-refer to these services or be referred by an appropriate healthcare professional after an assessment of motivation. For those who did not meet their healthy weight targets, or for those considered to be more at risk of the cardiometabolic consequences associated with obesity, referral was into Tier 3 specialist services delivered by the Rotherham Institute for Obesity (RIO). After 2014 the integrated framework of services was referred to as Weigh Up Rotherham, and it had a single point of access. Although there are no peer-reviewed published outcome data, RIO has been regarded as a model care pathway [33], reporting a full analysis of data for 2010 and 2011 of 49% of 3325 referrals completing the 6-month intervention. 72% achieved a ≥5% and 18.6% a ≥10% weight loss. Despite this, in 2017 funding for the service was withdrawn.

Fakenham

In Fakenham, a rural area in the east of England, a Tier 3 service aimed to deliver evidence-based interventions including medical assessment, motivational interviewing to support behaviour change, dietary and activity advice, psychological therapies, drug therapy with orlistat, medically supervised meal replacement diets and assessment for suitability for bariatric surgery using the NHS East of England criteria (aged 18–60 years, BMI ≥40 kg/m², with either diabetes or severe obstructive sleep apnoea, and having undergone a 6-month intensive weight-loss programme). The service aimed to facilitate weight loss by implementing progressive and sustainable lifestyle changes, based on individually agreed goals over a 1-year programme. This was delivered by individual monthly appointments with obesity specialist nurses. Participants were considered for pharmacotherapy, low energy liquid diets, or bariatric surgery if clinically appropriate. Psychological therapies offered included cognitive behaviour therapy, neurolinguistic programming, emotional freedom therapy, clinical hypnosis, solution focused therapy, transactional analysis, psychodynamic therapy, and cognitive analytical therapy (Figure 14.1.8.5) [32]. An evaluation of the service

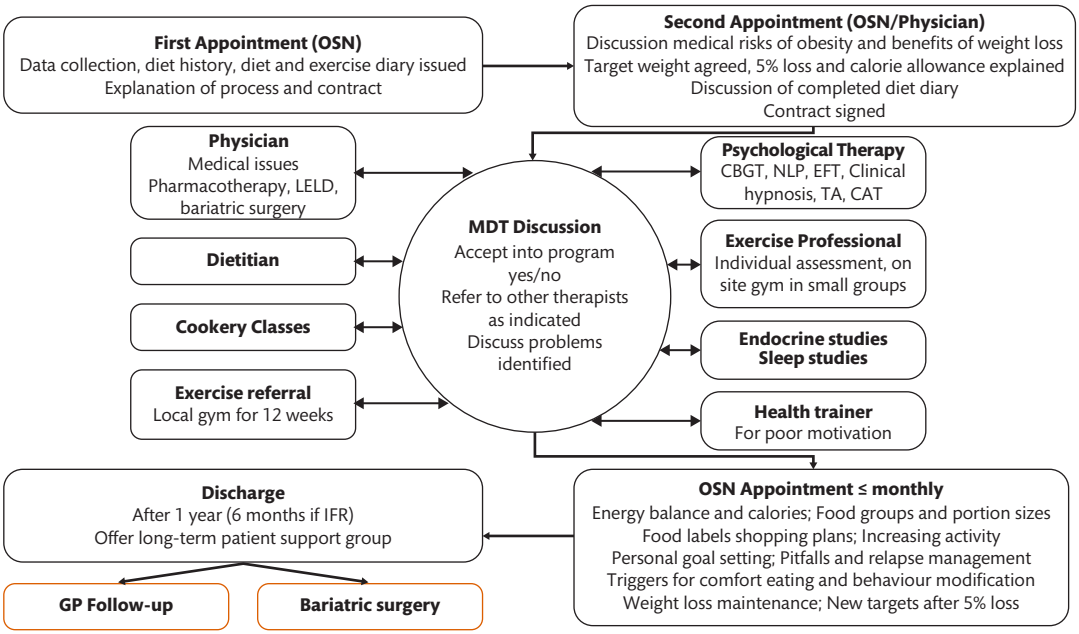


Figure 14.1.8.5 The Fakenham Clinical Pathway. MDT, multidisciplinary team; OSN, obesity specialist nurse; CBT, cognitive behavioural therapy; NLP, neurolinguistic programming; EFT, emotional freedom therapy; TA, transactional analysis; CAT, cognitive analytical therapy; LELD, low energy liquid diet.

Adapted with permission from Jennings A, Hughes CA, Kumaravel B, *et al.* Evaluation of a multidisciplinary Tier 3 weight management service for adults with morbid obesity, or obesity and comorbidities, based in primary care. *Clin Obes* 2014;4(5):254–66. © 2014 The Authors. [32]

using the National Observatory Standard Evaluation Framework [34, 35] followed a cohort of 230 patients (out of 367 referrals) sequentially enrolled over a 1-year period from August 2011. Of these, 213 entered a 1-year programme, and 17 a more intensive 6-month programme; two-thirds had a BMI above 40 kg/m². The service identified additional pathology in 48 patients: hypothyroidism, hypercholesterolaemia, abnormal liver function, diabetes mellitus, obstructive sleep apnoea, and hypertension. More than half completed their programmes and attended at 1 year. In these 117 'completers', weight losses >5% were 34.2, 53.8, 65.8, and 72.6% at 3, 6, 9, and 12 months. The mean percentage weight loss at 1 year was −6.4% (SD 6.1) in those with weights recorded, and −8% (SD 6.0) in 'completers'. There were significant improvements in quality of life measures.

Conclusions

The generic process for developing and managing care pathways is well established, but has been poorly developed for comprehensive care of people with obesity. National guidelines and systematic literature reviews have been generated and, in some cases (particularly for more severe obesity) translated into advice for commissioning integrated pathways of care, but these have hardly been implemented. This may reflect prejudice of healthcare providers against obesity as a disease worth treating, underestimation of the clinical and economic benefits of effective obesity care, or the complexities of designing pathways for its management.

REFERENCES

1. Network SIG. *Management of Obesity. A National Clinical Guideline*. Edinburgh, 2010. Available at: https://www.sign.ac.uk/assets/sign115.pdf (accessed: 16 April 2020).
2. Garvey WT, Mechanick JI, Brett EM, *et al*. American Association of Clinical Endocrinologists and American College of Endocrinology Comprehensive clinical practice guidelines for medical care of patients with obesity: executive summary. *Endocr Pract*, 2016; **22**: 842–84.
3. Fried M, Hainer V, Basdevant A, *et al*. Inter-disciplinary European guidelines on surgery of severe obesity. *Int J Obes (Lond)*, 2007; **31**: 569–77.
4. Uerlich MF, Yumuk V, Finer N, *et al*. Obesity management in europe: current status and objectives for the future. *Obes Facts*, 2016; **9**: 273–83.
5. National Health and Medical Research Council. *Clinical Practice Guidelines for the Management of Overweight and Obesity in Adults, Adolescents and Children in Australia*. Melbourne: National Health and Medical Research Council, 2013.
6. Currie L, Harvey G . *The Origins & Use of Care Pathways in the USA, Australia & the United Kingdom*. Oxford: Royal College of Nursing Institute 1997: 1–44.
7. De Bleser L, Depreitere R, De Waele K, *et al* . Defining pathways. *J Nurs Manag*, 2006; **14**: 553–63.
8. World Health Organization, Europe. *The Minsk Declaration: The Life-Course Approach in the Context of Health 2020*. Geneva: WHO, 2015.
9. Tsigos C, Hainer V, Basdevant A, *et al*. Management of obesity in adults: European clinical practice guidelines. *Obes Facts*, 2008; **1**: 106–16.
10. Fried M, Hainer V, Basdevant A, *et al*. Interdisciplinary European guidelines on surgery of severe obesity. *Obes Facts*, 2008; **1**: 52–9.
11. Mathus-Vliegen EMH, Basdevant A, Finer N, *et al*. Prevalence, pathophysiology, health consequences and treatment options of obesity in the elderly: a guideline. *Obesity Facts*, 2012; **5**: 460–83.
12. De Luc K, Kitchiner D, Layton A, *et al*. *Developing Care Pathways the Handbook*. Oxon: Radcliffe Medical Press Ltd, 2001.
13. Davis N. *Integrated Care Pathways: A Guide to Good Practice*. Wales: National Leadership and Innovation Agency for Healthcare, 2005.
14. Rotter T, Kinsman L, James E, *et al*. Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database Syst Rev,* 2010; **3**: CD006632.
15. Allen D, Gillen E, Rixson L . Systematic review of the effectiveness of integrated care pathways: what works, for whom, in which circumstances? *Int J Evid Based Healthc*, 2009; **7**: 61–74.
16. Tobacco Use and Dependence Guideline Panel . *Treating Tobacco Use and Dependence: 2008 Update*. Rockville, MD: US Department of Health and Human Services, 2008.
17. Campbell-Scherer DL, Asselin J, Osunlana AM, *et al*. Implementation and evaluation of the 5As framework of obesity management in primary care: design of the 5As Team (5AsT) randomized control trial. *Implement Sci*, 2014; **9**: 78.
18. Vallis M, Piccinini–Vallis H, Sharma AM, *et al*. Modified 5 as: minimal intervention for obesity counseling in primary care. *Canadian Family Physician,* 2013; **59**: 27–31.
19. Rueda-Clausen CF, Benterud E, Bond T, *et al*. Effect of implementing the 5As of obesity management framework on provider-patient interactions in primary care. *Clin Obes,* 2014; **4**: 39–44.
20. Asselin J, Salami E, Osunlana AM, *et al*. Impact of the 5As Team study on clinical practice in primary care obesity management: a qualitative study. *CMAJ Open*, 2017; **5**: E322–9.
21. National Institute for Health and Care Excellence. Obesity—everything NICE says in an interactive flowchart, 2018. Available at: https://pathways.nice.org.uk/pathways/obesity (accessed: 1 October 2018).
22. Capehorn MS, Haslam DW, Welbourn R . Obesity treatment in the UK health system . *Curr Obes Rep,* 2016; **5**: 320–6.
23. Blackshaw J, Montel S, King S, *et al*. Report of the working group into: joined up clinical pathways for obesity, 2014. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/03/owg-join-clinc-path.pdf (accessed: 16 April 2020).
24. Brett B, Ahmed A . *Addressing the Obesity Challenge in the East of England*—A report to the East of England Clinical Senate Assembly, Public Health England, 2017.
25. Public Health England. Weight management economic tool 2014. Available at: https://www.gov.uk/government/news/phe-launch-weight-management-economic-assessment-tool (accessed: 16 April 2020).
26. Public Health England. Local health and care planning: menu of preventative interventions, 2016. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/683016/Local_health_and_care_planning_menu_of_preventative_interventions_DM_NICE_amends_14.02.18__2_.pdf (accessed: 1 October 2018).

27. All Party Parliamentary Group on Obesity. The current landscape of obesity services: a report from the All Party Parliamentary Group on Obesity, 2018. Available at: http://www.bomss.org.uk/wp-content/uploads/2018/05/APPG-Obesity-2018.pdf (accessed: 1 October 2018).
28. Welbourn R, Dixon J, Barth JH, *et al*. NICE-accredited commissioning guidance for weight assessment and management clinics: a model for a specialist multidisciplinary team approach for people with severe obesity. *Obes Surg*, 2016; **26**: 649–59.
29. Wass J, Finer N . Action on obesity: comprehensive care for all . *Clin Med (London, England)*, 2013; **13**: 4–5.
30. Welbourn R, Hopkins J, Dixon JB, *et al*. Commissioning guidance for weight assessment and management in adults and children with severe complex obesity. *Obes Rev,* 2018; **19**: 14–27.
31. Steele T, Narayanan RP, James M, *et al*. Evaluation of Aintree LOSS, a community-based, multidisciplinary weight management service: outcomes and predictors of engagement. *Clin Obes*, 2017; **7**: 368–76.
32. Jennings A, Hughes CA, Kumaravel B, *et al*. Evaluation of a multidisciplinary Tier 3 weight management service for adults with morbid obesity, or obesity and comorbidities, based in primary care. *Clin Obes*, 2014; **4**: 254–66.
33. Aldridge S. Losing it in RIO: Diabetes UK, 2014. Available at: https://diabetes-resources-production.s3-eu-west-1.amazonaws.com/diabetes-storage/migration/pdf/Losing-it-in-Rio.pdf (accessed 1 October 2018).
34. Roberts K, Cavill N, Rutter H . National obesity observatory standard evaluation framework, 2009. Available at: https://www.noo.org.uk/core/frameworks/SEF (accessed: 1 October 2018).
35. Ells LJ, Cavill N, Roberts K, *et al*. Development of a standard evaluation framework for weight management interventions. *Public Health*, 2013; **127**: 345–7.

14.2

Lipoprotein Metabolism and Dyslipidaemia

14.2.1 Lipoprotein Metabolism

Bo Angelin and Paolo Parini

Introduction

Lipids are a heterogeneous group of substances with a myriad of structural and regulatory functions. Phospholipids and cholesterol are essential components of cell membranes, and cholesterol is also the precursor of steroid hormones, bile acids, and, in the skin, of vitamin D. Some fatty acids serve as precursors of bioactive compounds such as prostaglandins, thromboxanes, and leukotrienes; several lipids may also serve as signalling molecules in their own right. Furthermore, lipid complexes are necessary for the transport of lipid-soluble vitamins, and may have a protective role in the defence against toxins and infectious agents. From an overall perspective, however, the bulk function of plasma lipid metabolism relates to exchange of fats as energy substrates.

Plasma Lipids and Energy Transport

Fat as triglyceride is the major form of energy storage in the body [1]: about 12–15 kg of triglycerides (corresponding to ~500 000 kJ) are stored as an energy depot in the adipose tissue. Fat supplies 60–70% of cellular energy consumption, with a considerable portion of that formed through *de novo* synthesis from carbohydrate (lipogenesis) in the liver. Plasma lipids such as triglycerides and non-esterified fatty acids (NEFA) are quantitatively the most important system of energy exchange among organs (**Figure 14.2.1.1**). Each day, 6000–9000 kJ are transported in the circulation in triglycerides, and almost the same amount in NEFA. The transport of lipids is regulated very efficiently, and integrated with carbohydrate metabolism. Through the hormonal and metabolic control of enzymes (lipases) catalysing the uptake of fat in various organs and the mobilization of fat from adipose depots, energy from fat can be distributed and channelled in response to rapidly fluctuating demands. Although many hormones influence the activity of these enzymes, the predominant regulation is through insulin and catecholamines.

The inflow of dietary lipids into the circulation as chylomicron triglycerides corresponds to about 100 g/day. Another 20–50 g is secreted in very low-density lipoprotein (VLDL) triglycerides by the liver. The plasma pool of triglycerides is small, but since the elimination process normally is far from saturated, large amounts of triglycerides can be assimilated without more than a twofold to threefold rise in plasma concentration after a fat-rich meal.

Triglycerides routed to metabolically active tissues, such as skeletal muscle and heart, are rapidly utilized for oxidation. In the adipose tissue, fatty acids and monoglycerides are re-esterified to triglycerides for storage, and later mobilized rapidly on demand by lipases including hormone-sensitive lipase. The fractional uptake of free fatty acids from the circulation is relatively constant, so that the distribution to different tissues mainly follows that of blood flow. Roughly, one-third of NEFA is taken up by the liver and one-third by muscular tissue. The energy reaching the liver in this way exceeds demand, and a major fraction of fatty acids is re-esterified, returning to the circulation in VLDL triglycerides. Thus, NEFA and VLDL contribute to an energy cycle mainly between adipose tissue and the liver (**Figure 14.2.1.1**).

After food intake, when plasma concentration of insulin is high and those of catecholamines low to normal, lipoprotein lipase activity in adipose tissue is high at the same time as fatty acid mobilization is reduced. Thus, lipids are channelled to adipose tissue storage. Any surplus of carbohydrate energy can be utilized in

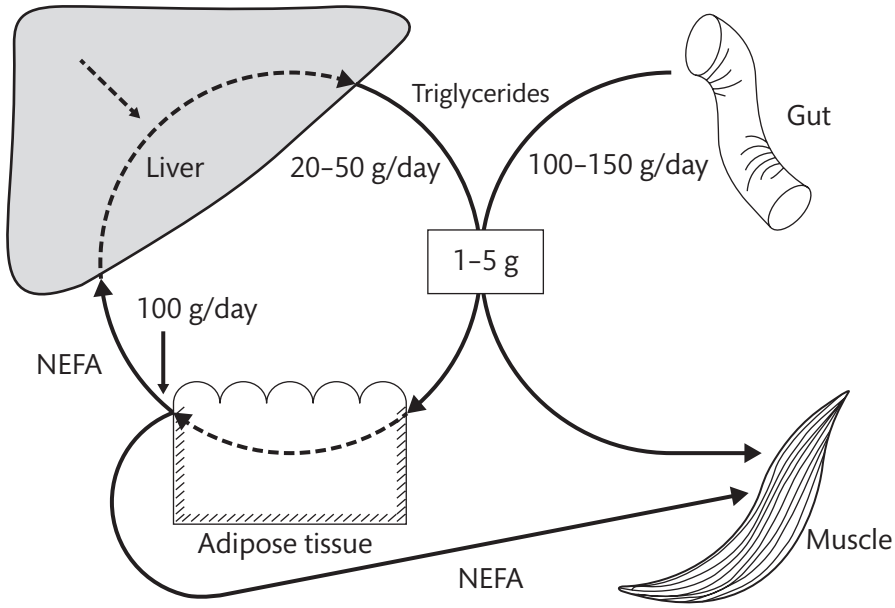


Figure 14.2.1.1 Schematic representation of the exchange of lipids between different organs.

hepatic *de novo* lipogenesis (DNL) which through VLDL transport can be redistributed to adipose tissue.

In response to starvation or stress, with low insulin and increased catecholamine levels, the lipid flux is redistributed to ensure appropriate energy delivery to metabolically active tissues, particularly heart and skeletal muscle. Lipoprotein lipase in adipose tissue is reduced, whereas it is increased in heart and muscle. Simultaneously, NEFA are mobilised from the adipose tissue, and the amount taken up by the liver can be recirculated in VLDL triglycerides. At a very high inflow of NEFA to the liver, as occurs after prolonged fasting, the surplus will be converted into ketone bodies that can be preferentially utilized in cellular metabolism.

Plasma Lipoproteins and Apolipoproteins

The requirement for transport of lipids in plasma poses a basic biophysical challenge arising from the insolubility of most lipids in water: virtually all plasma lipids thus have to be solubilized by association with specific proteins. NEFA are bound to albumin with a molar ration of 2:1, while more complex lipids are transported in hydrophilic lipoproteins [2]. These are spherical emulsion particles, in which unipolar lipids—mainly triglycerides and cholesteryl esters—are covered by a polar phospholipid membrane harbouring free cholesterol and proteins called apolipoproteins. These apolipoproteins maintain the structure of the lipoproteins, and may also serve as ligands for specific receptors or as enzyme activators. Triglycerides, cholesteryl esters, free cholesterol, and phospholipids are the main lipids present in lipoproteins, which also carries many other water-insoluble endogenous or exogenous substances.

Lipoproteins are generally classified on the basis of their hydrate density [3] or electrophoretic mobility (Table 14.2.1.1). Such classifications are obviously operational, and it is important to recognize that lipoproteins represent populations of constantly interchanging particles. Differences in density is commonly used to separate lipoproteins by ultracentrifugation. The density of lipoprotein particles is inversely related to their size, reflecting the relative amount of non-polar low-density core lipid and high-density surface protein. The two largest classes of lipoproteins thus contain mainly triglycerides: the chylomicrons (containing apoB-48) which are secreted from the intestine, and the VLDL (containing apoB-100) which are secreted from the liver. The smaller lipoprotein classes, intermediate density lipoproteins (IDL), low-density lipoproteins (LDL) and high-density lipoproteins (HDL_2 and HDL_3), mainly have cholesteryl esters in their cores, and represent products formed during the processing of the triglyceride-rich chylomicrons and VLDL. High-density lipoprotein (HDL) can also be produced by the liver and small intestine, and in their nascent form can be

Table 14.2.1.1 Normally occurring human plasma lipoproteins

Lipo-protein	Density (g/mL)	Electro-phoretic mobility	Diameter (nm)	Particle composition (weight %)				Major apolipo-proteins
				Trigly-ceride	Chole-sterol	Phos-pholipid	Protein	
Chylo-micron	0.93	Origin	80–1200	85–95	2–5	3–8	1–2	B-48, A-I, A-II, A-IV, (C, E)
VLDL IDL LDL	0.93–1.006 1.006–1.019 1.019–1.063	Pre-β β β	30–80 23–35 18–25	50 20 11	22 38 47	19 23 22	8 19 21	B-100, A-I, C, E B-100, C, E B-100
HDL_2 HDL_3	1.063–1.125 1.125–1.21	α α	9–12 5–9	6 6	22 15	30 23	41 55	A-I, A-II, C, E A-I, A-II, C, E

seen as bilayered discs on electron microscopy. The hydrolysis of triglyceride-rich lipoproteins also contributes to the formation of mature HDL particles in circulation. Lipoprotein(a) [Lp(a)] is larger but denser than LDL and exhibits slow pre-β mobility on electrophoresis. Essentially, Lp(a) consists of an LDL particle with apo(a) bound to the apoB-100 molecule.

A large number of apolipoproteins have been identified (Table 14.2.1.2). Several of the smaller ones (apoAs, apoCs, and apoE) belong to the same gene family [4]. They contain characteristic α-helical structures, which give them detergent-like properties. Like free cholesterol, these molecules can readily exchange between lipoprotein particles and with other lipid surfaces, whereas phospholipids and non-polar core lipids require specific transfer proteins for exchange. ApoB exists in two forms: apoB-100 is found mainly in VLDLs and LDLs, whereas apoB-48 is found only in chylomicrons [5]. Apo B-48 represents the amino-terminal half (48%) of apo B-100, and the synthesis of both proteins emerges from one gene. In the intestine, but not in the liver, a stop codon is introduced into the mRNA after transcription of the apo B gene (posttranscriptional editing), resulting in the formation of the truncated apo B-48 protein [6]. Thus, in humans—but not in rodents—triglyceride-rich lipoproteins containing apo B-48 and apo B-100 can be identified as being of intestinal and hepatic origin, respectively. Apo B-100, but not apo B-48, contains the ligand-binding site for the LDL receptor. Despite this the half-life in circulation of the apo B-48 containing particles is much shorter than that of the apo B-100 containing particles because the former can load a higher amount of apoE on their surface. Apo E is a strong ligand for the LDL receptor, which is also known as the apoB/apo E receptor, and for other receptors.

Table 14.2.1.2 Human apolipoproteins

Apolipoprotein	Molecular weight (kDa)*	Function/related to	Chromosome location
A-I	29	LCAT cofactor	11
A-II	9		1
A-IV	44	(LCAT activator)	11
A-V	39	VLDL synthesis and secretion	11
B-100	512	VLDL synthesis, LDL receptor ligand	2**
B-48	241	Chylomicron synthesis	2**
C-I	7		19
C-II	9	Lipoprotein lipase activation	19
C-III	9	Lipase inactivation inhibits receptor binding	11
D	19	(Cholesterol transport)	3
E	34	Ligand for LDL receptor and chylomicron remnant binding Hepatic lipase activation	19
M	25	HDL metabolism	6
Apo (a)	280–800		6

* From amino acid composition.
** Same gene, posttranscriptional editing of mRNA.

The genes of all the common apolipoproteins have been cloned, and there is evidence of a considerable number of genetic polymorphisms, some of which may result in altered functional properties of the protein. One important example is genetic variation at the apoE locus, where the presence of three variant alleles results in six major phenotypes of the apoE protein (E2/2, E2/3, E2/4, E3/3, E3/4, and E4/4). ApoE2 has lower binding affinity to the lipoprotein receptors, and individuals carrying this isoform have a slower clearance of chylomicron and VLDL remnants.

Apo (a) differs from the other apolipoproteins: it has a plasminogen-like structure, with a variable number of so-called kringle IV repeats [7]. Apo (a) is linked to apo B-100 by a disulphide bond, changing the properties of the LDL particle. There is a pronounced, genetically determined size variation based on the number of repeats, and there is an inverse correlation between plasma level and apo (a) isoform size. How elevated Lp(a) levels contribute to an increased risk for cardiovascular disease is still not understood, but mechanisms related to decreased plasmin formation as well as increased endothelial cell permeability and inflammation or oxidation have been discussed [7].

During their metabolism, lipoproteins may be modified by a number of mechanisms. Among those are oxidation, glycation, carbamoylation, and enzymatic degradation [8]. Such changes may occur in the circulation, but they are probably of particular importance when lipoproteins become retained within the subintimal layer of the vessel wall [9, 10]. Such modifications, most notably mild oxidation, stimulate the uptake of lipoproteins in tissue macrophages by so-called scavenger receptors leading to an uncontrolled expansion of cellular cholesteryl ester content. The result is the formation of foam cells, initiating the development of early atherosclerotic plaques. Changes in the lipid composition, particularly triglyceride enrichment of LDL, make the lipoprotein particles more susceptible to oxidation which may explain the increased propensity of atherosclerosis in conditions such as diabetes, insulin resistance, and obesity. Modified lipoproteins and crystals of cholesterol precipitated in the atherogenic plaque activates the inflammasome in macrophages and initiates inflammatory responses, which in turn accelerate atherogenesis [11].

Lipoprotein Receptors

The delivery of lipoprotein lipids to various tissues in the body is regulated by specific receptors and enzymes that interact with the individual apolipoproteins. The *LDL receptor* is a cell surface receptor present on all cells, capable of binding and internalizing lipoproteins containing apo B-100 or apo E [12]. It belongs to a gene family containing ligand-binding 40-amino acid cysteine-rich repeats, growth factor repeats, and spacer sequences that show homology with the epidermal growth factor precursor. The LDL receptor is synthesized in the endoplasmic reticulum and transported to the Golgi apparatus for glycosylation. After transport to the cell surface, the receptors cluster in specific regions rich in clathrin, the so-called coated pits. In these receptor-enriched regions, internalization of the LDL receptor (with or without bound LDL) takes place continuously through endocytosis. Whereas LDL is degraded in the lysosomal compartment, the receptors are recycled to the cell surface several times. The

expression of the LDL receptor is closely entrained to cellular demand for cholesterol.

Among other members of this gene family are the **LDL receptor-related proteins (LRPs)**, the **VLDL receptors**, and **megalin** [13]. LRP-1 is a large structure that is mainly expressed in the liver and seems to be involved in the binding and uptake of chylomicron remnants through apoE, but it can also bind several non-lipoprotein ligands. A number of *scavenger receptors* with a wide spectrum of ligand-binding properties are expressed on macrophages, and some of these are probably involved in the uptake of atherogenic lipoproteins to form foam cells in the arterial wall. *Scavenger receptor type B class 1 (SCARB1)* is a receptor highly expressed in the liver and adrenal, and seems to be involved in reverse cholesterol transport [14].

Enzymes and Transfer Proteins

Three enzymes have a major role in plasma lipoprotein lipid transport: lipoprotein lipase, hepatic lipase, and lecithin:cholesterol acyl transferase. The first two belong to a large gene family also including pancreatic lipase [15]. Cholesteryl ester transfer protein and phospholipid transfer protein are required to promote the exchange of core lipids between lipoprotein particles during their metabolism in plasma [16].

Lipoprotein lipase (LPL) is synthesized in the endoplasmic reticulum (ER) of adipocytes and muscle cells. In the ER of these cells a chaperone protein called lipase maturation factor 1 (LMF1) is responsible for LPL assembly and folding. Another protein called GPIHBP1 (GPI-anchored endothelial cell protein) binds LPL in the interstitial space and shuttles it across endothelial cells to its site of action in the capillary lumen [17] where it is tethered to the luminal surface of the capillary endothelium via heparan sulphate proteoglycans [15]. This enzyme, which is active as a dimer, hydrolyses triglyceride in the lipoprotein core, converting it into fatty acids and monoglycerides. APOC2 and APOA5 are cofactors for the enzymatic reaction [18]. The activity of LPL is regulated by nutritional demand via hormonal control. Enzyme activity is increased in adipose tissue during carbohydrate feeding and reduced during fasting; the regulation is reverse in skeletal muscle and heart. Loss of function mutations of all these genes (LPL, LMF1, GPIHBP1, APOA5, and APOC2) cause a rare monogenic dyslipidaemia characterized by very high levels of circulating triglyceride (see Chapter 14.2.2).

Hepatic lipase is synthesized in hepatocytes, secreted and bound to proteoglycans on the capillary endothelium of the liver. It is also present on endothelial cells in the adrenals and gonads. This enzyme has a broad activity and is believed to be involved in the metabolism of HDL and IDL [19]. Thus, by hydrolysing triglycerides and phospholipids in HDL_2, hepatic lipase is thought to promote its conversion to HDL_3. The enzyme activity is stimulated by thyroid hormone, insulin, and androgens, and decreased by oestrogen and glucocorticoids.

Lecithin: cholesterol acyl transferase (LCAT) is synthesized in the liver [20]. It binds to HDL, and catalyses the conversion of lecithin and free cholesterol to lysolecithin and cholesteryl esters; apo A-I and also apo A-IV act as cofactors for the enzyme. Through the action of LCAT, free cholesterol on the surface of nascent HDL is converted to more lipophilic esters, leading to expansion of the HDL particle.

Cholesteryl ester transfer protein (CETP) is predominantly synthesized in the liver and intestine, but also in adipocytes and the spleen [16]. It mediates transfer of cholesteryl esters from HDL and LDL to VLDL and chylomicrons, and the reciprocal exchange of triglycerides.

Phospholipid transfer protein (PLTP) is a member of the same gene family as CETP [16]. It mediates the transfer of phospholipids from triglyceride-rich lipoproteins to nascent HDL during lipoprotein lipase-mediated lipolysis.

Lipoprotein Metabolism

Lipoproteins serve three main functions: (1) transport of exogenous dietary lipids from the gut to peripheral tissues and the liver, carried out by chylomicrons and their remnant particles; (2) transport of endogenous lipids (triglyceride and cholesterol) from the liver to peripheral tissues, by the VLDL-IDL-LDL-pathway; and (3) reverse cholesterol transport from peripheral tissues to the liver, where HDL plays a central role (Figure 14.2.1.2). Apo B-100 and apo B-48 thus

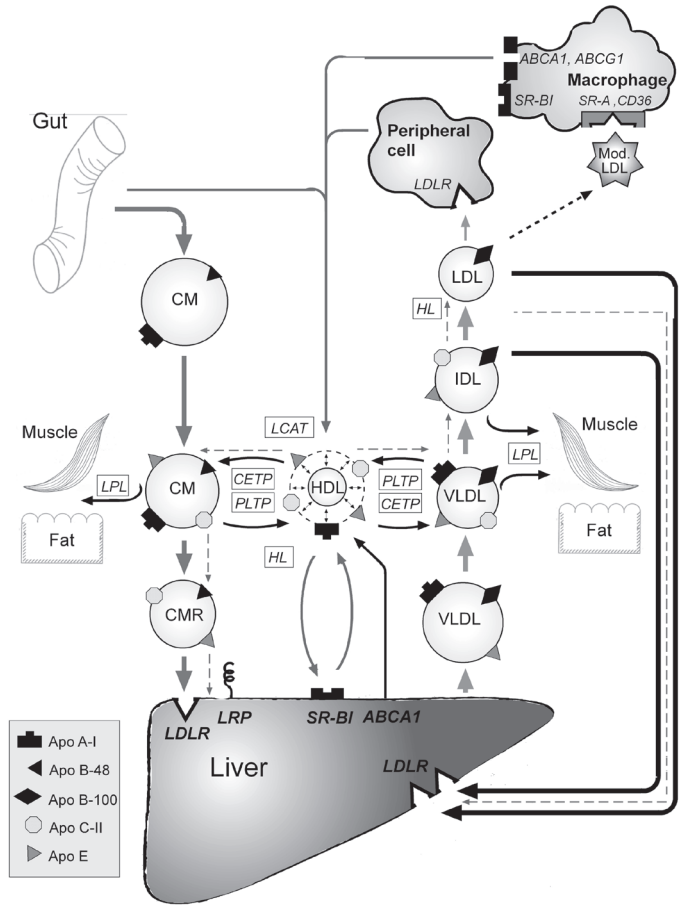


Figure 14.2.1.2 Simplified scheme of normal lipoprotein metabolism in humans. CM, chylomicrons; CMR, chylomicron remnants; VLDL, very low-density lipoprotein; IDL intermediate density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein; LPL, lipoprotein lipase; HL, hepatic lipase; CETP, cholesteryl ester transfer protein; PLTP, phospholipid transfer protein; LCAT, lecitin:cholesterol acyl transferase; ABC-1, ATP-binding cassette transporter 1; LDLR, LDL receptor; LRP, LDL receptor-related protein; SR-B1; scavenger receptor class B type 1; SR-A, scavenger receptor class A.

constitute the transport chains for chylomicron–chylomicron remnants (exogenous) and VLDL-IDL-LDL (endogenous), respectively, whereas apo A-I is closely related to reverse cholesterol transport.

Exogenous Lipid Transport

Ingested triglycerides (100–150 g/day) are hydrolysed by pancreatic lipase and almost completely absorbed in the upper small intestine [21, 22]. The solubilizing actions of biliary bile acids are vital in this process. Cholesterol of both biliary (600–800 mg/day) and dietary (300–400 mg/day) origin is absorbed to a much lesser degree (30–50%). Due to the presence of active influx via the Niemann-Pick C1-like 1 protein (NPC1L1) and active efflux promoted by the ATP-binding cassette transporters G5 and G8 (ABCG5/G8), enterocytes absorb unesterified cholesterol at a faster net rate than plant sterols. Intracellular fatty acid binding proteins facilitate fatty acid uptake in the enterocytes. After re-esterification, triglycerides are incorporated into the very large chylomicrons together with esterified cholesterol produced by the enzyme acyl CoA: cholesterol acyl transferase-2 (ACAT-2), encoded by the sterol O-acyltransferase 2 (*SOAT2*) gene; this core is surrounded by a monolayer of phospholipids, free cholesterol, and APOB48. APOB48 is partially lipidated during its translocation across the membrane of the ER, whereafter it fuses with the bulk of lipids. The fully lipidated particles are transported to the Golgi apparatus, whence they are transported to the basolateral plasma membrane and secreted into the extracellular space. Chylomicrons pass into the intestinal lacteals and enter the circulation via the thoracic duct. In addition to apo B-48, nascent chylomicrons contain apo A-I, A-II, and A-IV that have been synthesized in the enterocytes. Upon exchange with HDL in lymph and blood, chylomicrons rapidly acquire apo Cs and apo E.

In the circulation, chylomicrons are trapped on the capillary walls of adipose tissue and muscle, probably involving a complex interaction between the lipoprotein, lipoprotein lipase, and endothelial proteoglycans. Apo C-II is an obligate activator of lipoprotein lipase that attacks the triglyceride core of the chylomicron, resulting in intravascular hydrolysis of triglycerides liberating NEFA and glycerol. Most of the NEFA are taken up by adipose or muscular tissue, but some are also transported with albumin to the liver and fuelling the increased VLDL secretion that occurs in later phases of the postprandial period [23]. In the peripheral tissues, NEFA are either re-esterified to depot triglyceride or oxidized for energy production. The whole process of lipolysis is very efficient and only takes about 10–15 minutes in normal individuals.

The excess surface lipids and apolipoproteins resulting from lipase-induced reduction of the chylomicron core volume are transferred to HDL. APOB48, together with APOE, remains on a much smaller particle containing esterified cholesterol and small amounts of triglyceride in its core. The liver rapidly clears this **chylomicron remnant** via APOE-recognizing receptors on the surface of the hepatocytes [24]. Chylomicron remnants bind with high affinity to the LDL receptor in the liver, but there is ample evidence that other lipoprotein receptors including LRP-1, and surface heparan sulphate proteoglycans are also involved in the hepatic uptake. Normally, remnants have a very short half-life, which is why such particles are normally absent in plasma from fasting individuals, except those carrying the APOE2 phenotype. After endocytosis, released fatty acids and cholesterol can be used in VLDL assembly, or can be stored as triglycerides and cholesteryl esters. Cholesterol can also be excreted into the bile as free cholesterol or after conversion to bile acids.

Endogenous Lipoprotein Transport

VLDLs are similar to chylomicrons in structure and composition, but are much smaller and contain relatively less triglyceride (Table 14.2.1.1). The assembly of VLDL in the liver is essentially analogous to that of chylomicrons in the intestine. Each VLDL particle contains a single molecule of apo B-100 which is synthesized on ribosomes attached to the ER and lipidated cotranslationally. Increases of triglyceride secretion in VLDL determines increases in particle volume. Hormones such as oestrogens and growth hormone, and nutrients such as alcohol and excess carbohydrate stimulate VLDL triglyceride synthesis, whereas insulin retards secretion of VLDL. Fully lipidated VLDL are transported to the Golgi, where glycosylation occurs before trafficking and release into the space of Disse. Nascent VLDLs contain some apo A-I, apo A-II, apo E and apo Cs, and their exchange with HDL probably occurs very rapidly when in plasma. While VLDLs of varying size are released from the liver, it is still not clear whether LDL-like particles may be secreted directly.

In the circulation, VLDLs are subject to lipolysis by lipoprotein lipase, and in the postprandial state there is competition between VLDLs and chylomicrons for the lipoprotein lipase. The estimated half-life of VLDLs is considerably longer than that of chylomicrons, 1–2 hours. A major difference from the exogenous pathway is that the resulting VLDL 'remnants', referred to as IDLs, can have alternative fates. More than half, presumably those enriched in APOE(24), undergo rapid endocytosis in the liver, whereas the remaining fraction is metabolized further via hepatic lipase and transfer of lipids and apolipoproteins to HDL. The final products of this processing of circulating IDL are mature LDLs, only carrying the apo B-100 protein. The continued presence of apo C-III on IDL particles inhibits a rapid removal by the LDL receptor and leads to the formation of denser LDL particles that are very poor in triglycerides, also known as small-dense LDL [25]. Recently, the composition of LDL in respect to their phospholipid content (phosphocholine vs. sphingomyelin) and in respect of their lipid core content (triglycerides vs. esterified cholesterol) determines the susceptibility of these particles to aggregate to each other, a process that promote their atherogenic potential [26].

The half-life of LDL is 2–3 days in normal humans, which explains why the number of LDL particles is much larger than that of VLDLs. Their prolonged residence in the circulation also explains why these particles may be the major site of modification (e.g. oxidation, glycation, and carbamylation). Although passive uptake of LDL (proportional to plasma concentration) takes place in all tissues, the predominant route for elimination of LDL is endocytosis by the LDL receptor in the liver [27].

While all cells contain the enzymatic machinery to synthesize cholesterol from acetate, receptor-mediated endocytosis of LDL appears to be the preferred pathway for acquiring the cholesterol necessary for maintaining cellular homeostasis in peripheral cells. By uptake of LDLs, the cells are also supplied with tocopherols, which are present in these lipoproteins. The number of LDL receptors is increased during cell division and under other circumstances when the demand for cholesterol is increased. High numbers of LDL receptors are expressed on steroidogenic cells, and also in some

tumour cells [27]. The absolute mass of LDL receptor is greatest in the liver, and at least 50% of LDL catabolism occurs in this organ [27]. The great importance of the expression of LDL receptors in normal lipoprotein metabolism is evident from the phenotype observed in patients with familial hypercholesterolemia, where the LDL receptor is functionally deficient.

Each cell has the capacity to regulate its cholesterol balance by sensing its cellular membrane cholesterol level [28]. This influences the activity of transcription factors which are bind sterol responsive elements (SREs) in the promoter regions of the LDL receptor gene. Such SREs are also present in the genes coding for two major enzymes regulating cellular cholesterol synthesis, 3-hydroxy-3-methylglutaryl (HMG) CoA reductase and HMG Co A synthase, and of the posttranscriptional modulator of the LDL receptor, proprotein convertase subtilisin/kexin type 9 (PCSK9). The transcription factors, the so-called SRE-binding proteins (SREBPs), and are normally attached to the cellular membrane, but undergo controlled proteolytic cleavage by specific enzymes which are activated when sterols are depleted from the membrane. The resulting SREBP fragment then migrates into the nucleus and activates gene transcription of the LDL receptor, cholesterol-synthesizing enzymes, and PCSK9. In addition, the enzymatic activity of HMG-CoA reductase is regulated by sterol-induced degradation of the enzyme. PCSK9 acts by increasing the degradation of the LDL receptor during its cell cycle (Figure 14.2.1.3), adding to the complexity of the regulation of cellular cholesterol uptake [29]. Of particular interest, genetic variation in PCSK9 expression is important for variation in LDL cholesterol levels and incidence of coronary heart disease [30]. Emerging evidence suggests moreover that several hormones influence hepatic LDL receptor number through modulation of PCSK9. LDL receptor cycling is also influenced by cellular proteins including LDLRAPs and IDOL, in which genetic variants may also be important.

HDL Metabolism and Reverse Cholesterol Transport

HDL denotes a heterogeneous, metabolically active lipoprotein fraction, in which the particles are continuously remodelled, changing their composition accordingly [16]. The two major classes of HDL, HDL_2, and HDL_3, differ in their density and their protein content (Table 14.2.1.1). By serving as an apolipoprotein reservoir which provides enzyme cofactors and lipid transfer factors, HDL may stimulate both the exogenous and endogenous lipid transport pathways (Figure 14.2.1.2).

Apo A-I is the common structural element of HDL particles. There are HDL particles that contain both apo A-I and apo A-II, and particles with only apo A-I. Whereas apo A-I is synthesized both in the intestine and the liver, apo A-II is made exclusively in the liver. A major part of the apolipoproteins and phospholipids that eventually contribute to plasma HDL originates from chylomicrons and VLDL secreted into the circulation. The transmembrane transporter ABCA1 secretes phospholipids and free cholesterol to poorly lipidated apo A-I to form a nascent HDL (pre β1-HDL) (Figure 14.2.1.4). The liver seems to contribute more than the intestine to this process [31]. The various lipid and protein components of HDL have different pathways of metabolism, resulting in considerably differing turnover rates. The catabolic half-life of plasma apo A-I and apo A-II in normal humans averages 4–5 days; cholesteryl esters may turn over 10–40 times more rapidly.

Overall, there is a cycle of enlargement of HDL from influx of lipids, apolipoproteins, and LCAT activity, followed by cholesterol

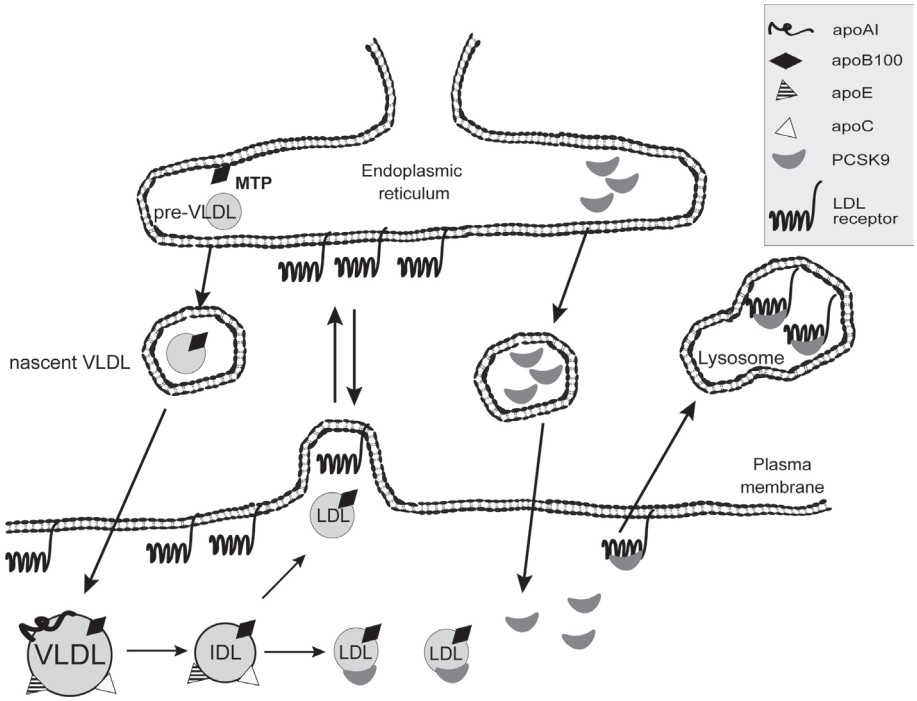


Figure 14.2.1.3 Secretion of VLDL and interaction between PCSK9 and LDL receptor. PCSK9 is transported from endoplasmic reticulum to cell surface via COP II (coat protein complex II) vesicles. Circulating PCSK9 prevents the normal recirculation of LDL receptors to the cell surface reducing their numbers at cell.

ester—triglyceride exchange and then shrinkage of HDL and loss of lipid and protein following hepatic lipase activity [16]. The cholesterol esters formed in HDLs following LCAT action can be transferred to triglyceride-rich lipoproteins by CETP, they may be transferred to cells by selective uptake without degradation of the HDL particle, and they may be catabolized with intact HDL.

HDL is currently believed to be active in the removal of cholesterol from peripheral tissues, promoting transport of cholesterol to the liver, the only organ that excretes significant amounts of cholesterol from the body. Cellular cholesterol removal occurs by several mechanisms, relying upon different HDL subclasses that play specific roles in distinct mechanisms, namely (a) aqueous diffusion (which involves unknown transporters) to mature HDL; (b) SR-BI-mediated efflux to mature HDL; (c) ABCG1-mediated efflux to mature HDL and preβ-HDL; and (d) ABCA1-mediated efflux to apoAI, preβ-HDL and small HDL particles [32–35] (Figure 14.2.1.4).

In humans ABCA1 is of major importance, as evident from the lack of HDL in patients with homozygous ABCA1 deficiency (Tangier disease). HDL can transport cholesterol directly to the liver, probably by the interaction of apo A-I with SCARB1, the human homologue of SR-BI [25], but also via exchange of core triglycerides in the triglyceride-rich lipoproteins for core cholesterol esters in HDL. By such mechanisms, cholesterol elimination from peripheral tissues may occur also utilizing the exogenous and endogenous pathways (Figures 14.2.1.2 and 14.2.1.4).

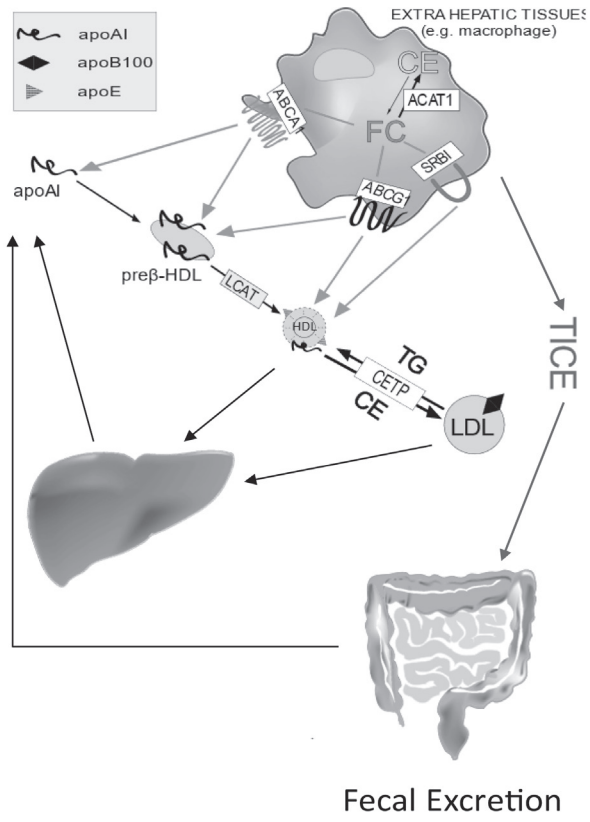


Figure 14.2.1.4 Reverse cholesterol transport and transintestinal cholesterol efflux pathways. CE, cholesteryl esters; FC, free cholesterol; TG, triglycerides; VLDL, very low-density lipoprotein; LDL, low-density lipoprotein; preβ—HDL, discoidal nascent high-density lipoprotein: HDL, high-density lipoprotein; apoAI, apolipoprotein AI; apoB100, apolipoprotein B100; apoE, apolipoprotein E; CETP, cholesterol ester transfer protein; LCAT, lecithin cholesterol:acyltransferase; ABCA1, ATP-binding cassette transporter A1; ABCG1, ATP-binding cassette transporter G1; SRBI, scavenger receptor type BI; ACAT1, Acyl-conzyme A:cholesterol acyltransferase 1.

Cholesterol Elimination

The liver has a central role in maintaining body cholesterol homeostasis [36, 37]. Several pathways may deliver cholesterol to this organ, as discussed (Figure 14.2.1.5): chylomicron remnants, IDL and LDL, and via HDL. In addition, cholesterol can be synthesized de novo in the liver, a process under tight transcriptional and posttranscriptional feedback control. Hepatic cholesterol may be stored as cholesterol esters, incorporated in secreted VLDLs, or excreted in the bile, either directly or after conversion to bile acids.

Cholesterol esterification is catalysed by the microsomal enzyme ACAT. In humans, two genes exist, *SOAT1* and *SOAT2,* encoding two different enzymes: ACAT1 and ACAT2, respectively. While ACAT1 is present in most tissues, ACAT2 appears to be specific to hepatocytes and enterocytes, and has been proposed to be the enzyme responsible for synthesizing cholesterol esters destined for VLDL secretion [38].

While minor amounts of cholesterol are converted into steroid hormones (50 mg/day) or lost through skin exfoliation (80 mg/day), faecal excretion is the major pathway for net excretion of cholesterol from the body. This process takes place either by direct secretion of free cholesterol in the bile, or by conversion of cholesterol to bile acids (Figure 14.2.1.5). Bile acids additionally promote hepatic secretion of cholesterol by stimulation of bile flow. Simultaneously, bile acids also contribute to cholesterol input by facilitating cholesterol absorption in the small intestine. A non-biliary route for cholesterol elimination from the body has also been described, named transintestinal cholesterol excretion (TICE) (Figure 14.2.1.4). TICE is responsible for direct transport of cholesterol from blood across the enterocytes into the intestinal lumen [39] and its relevance for humans is under debate [40].

Whereas only 30–50% of the cholesterol present in the intestinal lumen is absorbed, bile acids are almost completely reabsorbed via passive diffusion along the small intestine and via active transport in the distal ileum. The latter process is critically reliant on a sodium-dependent ileal bile acid transporter (IBAT) [41]. After reabsorption, bile acids are returned to the liver via the portal vein, and subsequently resecreted into the bile. By this efficient recycling, the preponderance of bile acids is conserved within the enterohepatic region, and only about 300–500 mg are lost in the faeces each day.

The capacity to synthesize bile acids from cholesterol is a unique property of the hepatocytes. In a normal adult human, about 300–500 mg of bile acids are produced daily, to compensate for faecal loss [36, 37]. In humans, the first and rate-limiting step in bile acid synthesis is the conversion of cholesterol into 7α-hydroxycholesterol, a reaction catalysed by the microsomal enzyme cholesterol 7α-hydroxylase [42]. This enzyme is mainly regulated by the flux of bile acids in the enterohepatic circulation, sensed through the bile acid-binding nuclear receptor farnesoid X-receptor (FXR) [43] which is expressed both in the ileum and in the liver. Details of this regulation are still under study, but it appears as if humans have a low

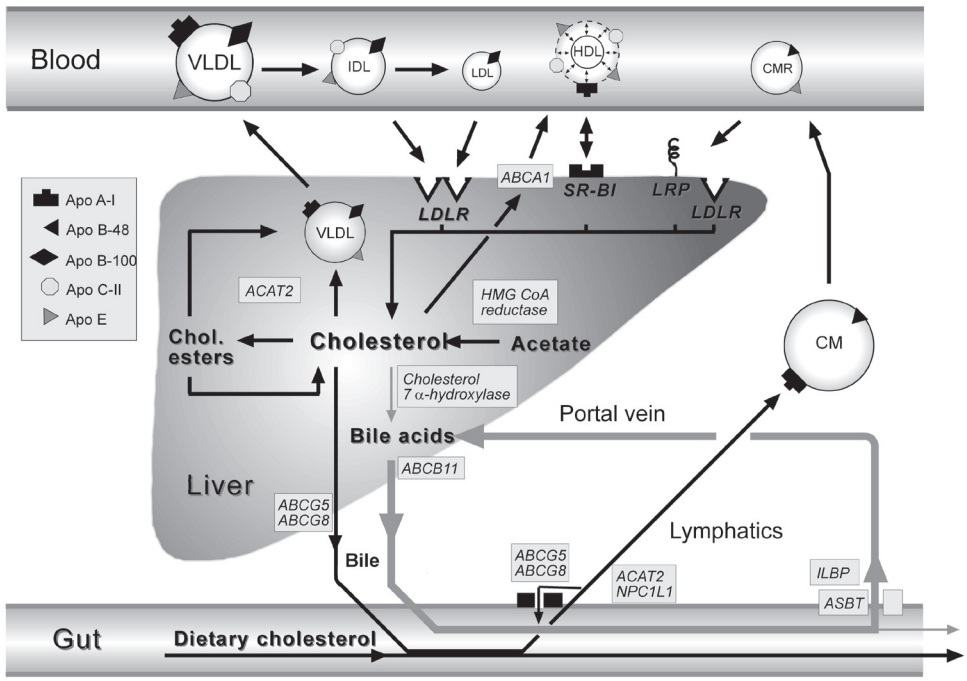


Figure 14.2.1.5 Schematic representation of hepatic cholesterol metabolism and the enterohepatic circulation. CM, chylomicrons; CMR, chylomicron remnants; VLDL, very low-density lipoprotein; IDL intermediate density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein; LDLR, LDL receptor; LRP, LDL receptor-related protein; SR-B1; scavenger receptor class B type 1; ACAT, acyl CoA: cholesteryl acyl transferase; HMG-CoA reductase, 3-hydroxy-3-methylglutaryl CoA reductase; IBAT, ileal bile acid transporter.

basal capacity for bile acid synthesis compared to most other species. Instead they show marked stimulation following interruption of the enterohepatic circuit (e.g. by ileal resection or treatment with bile acid binding resins), resulting in marked adaptive responses mediated via the SREBP-system: both HMG-CoA reductase activity and LDL receptor expression being stimulated [44]. This is the rationale for therapy with resins, and particularly explains the pronounced effects when such drugs are administered together with inhibitors of cholesterol synthesis. Inhibition of intestinal absorption and biliary reabsorption of cholesterol by ezetimibe, which block both NPC1L1 in the intestine and in the liver, does not affect the synthesis of bile acid in humans [45].

Stressors of Lipoprotein Metabolism Across the Lifespan

As discussed in Chapter 14.2.2, many severe monogenic disorders of lipoprotein metabolism are now known. Common genetic variants may also produce dyslipidaemia; however, this commonly only becomes overt in the presence of physiological or other stressors. Disturbances of lipoprotein metabolism thus represent excellent examples of interaction between genetic background and environmental influence. Selected conditions and circumstances in which clinically significant perturbation of lipoprotein metabolism may occur across the lifespan are discussed next.

Insulin has an antilipolytic effect on adipose tissue, inhibits VLDL release from the liver, and is necessary for lipoprotein lipase function. Thus, in insulin deficiency, as in newly diagnosed or severely undertreated type 1 diabetes, the absence of insulin increases NEFA and VLDL triglyceride levels, and fasting chylomicrons are frequently present, resulting in massive hypertriglyceridaemia [46]. HDL levels are reduced due to the inefficient triglyceride clearance. When type 1 diabetes is well controlled with insulin, in contrast, lipoprotein abnormalities are minimal, and if hyperlipidaemia remains it may reflect a coexisting (often primary) disorder of lipoprotein metabolism. In type 2 diabetes, insulin resistance is probably the main factor driving dyslipidaemia characterized by elevated VLDL, small-dense LDL, and low HDL [47]. Metabolic dyslipidaemia is discussed further in Chapter 15.13.4, 'Diabetic Dyslipidaemia'.

Hypercaloric intake enhances production of VLDL triglyceride [48], and, particularly together with insulin resistance, this may produce hypertriglyceridaemia. Increased VLDL triglyceride production also appears to be a major mechanism for the hyperlipoproteinaemia sometimes caused by **alcohol** consumption [49]. If VLDL clearance mechanisms (such as lipoprotein lipase) are intact and able to compensate for this increased production, enhanced flux through the VLDL pathway may explain the increase in HDL concentration seen after moderate alcohol intake. Stimulation of VLDL catabolism may also explain the positive influence of **physical exercise** in elevating plasma HDL cholesterol.

Thyroid hormone stimulates expression of hepatic LDL receptors [50]. Whereas LDL levels are low in hyperthyroidism, increased lipolysis and triglyceride production may lead to hypertriglyceridaemia. Raised LDL cholesterol is a frequent finding in hypothyroidism, where hypertriglyceridaemia and increased Lp(a) are also common findings. *Growth hormone* seems to be essential for the normal expression of hepatic LDL receptors [51], and also stimulates VLDL triglyceride flux. Replacement therapy

in growth hormone deficient subjects normalizes the elevated LDL cholesterol levels and increases HDL cholesterol; interestingly Lp(a) is also increased by growth hormone treatment.

Plasma LDL cholesterol concentration typically increases with **ageing** [52], apparently due to a reduced capacity for LDL clearance, reflecting diminished expression of hepatic LDL receptors. It has been suggested that these may at least partly relate to a relative deficiency of growth hormone with ageing.

Gender differences in plasma lipoproteins appear at puberty [53]. In comparison to males of similar age, females have lower VLDL and LDL levels and higher HDL_2 levels. Although gender-related differences in plasma lipoproteins are contributed to by differences in diet, alcohol, and smoking habits, the direct action of sex hormones on plasma lipoprotein metabolism is likely to be the dominant explanation. By the sixth decade of age, the gender differences in plasma lipoproteins change: women present increased LDL levels, associated with a decrease in LDL particle size. This relatively marked increase in LDL cholesterol levels, observed after menopause, is probably related to loss of oestrogen stimulation of LDL receptors. A fall in HDL cholesterol, due to a reduction of both HDL_2 and HDL_3 is also observed in menopausal women underlining the importance that oestrogen has for apo A-I metabolism [54]. Lp(a) levels are also influenced by oestrogen.

Understanding of the role of disturbances in lipoprotein metabolism in the development and progression of coronary heart disease is constantly becoming deeper. Because dyslipidaemias are the consequence of interactions between genetic background and environmental influences, therapeutic strategies should not only try to target contributory genetic defects, but should also consider hormonal and nutritional status. Therefore, a pharmacological approach needs to be individualized and associated with appropriate hormonal substitution, if relevant, and with changes in lifestyle by modification of diet, increased physical activity, and discontinuation of smoking and overconsumption of alcohol.

REFERENCES

1. Appling D, Anthony-Cahill SJ, Mathews CK. *Biochemistry: Concepts and Connections*. Harlow: Pearson, 2016.
2. Gotto AMJ, Pownall HR, Havel RJ . Introduction to the plasma lipoproteins. *Methods Enzymol*, 1986; **128**: 3–41.
3. Chapman MJ . Animal lipoproteins: chemistry, structure, and comparative aspects . *J Lipid Res*, 1980; **21**: 789–853.
4. Luo CC, Li WH, Moore MN, Chan L . Structure and evolution of the apolipoprotein multigene family . *J Mol Biol*, 1986; **187**: 325–40.
5. Fazio S, Linton MF . Regulation and clearance of apolipoprotein B-containing lipoproteins. In: Ballantyne CM, ed. *Clinical Lipidology*. Philadelphia, PA: Saunders Elsevier, 2009: 11–25.
6. Davidson NO, Anant S, MacGinnitie AJ. Apolipoprotein B messenger RNA editing: insights into the molecular regulation of post-transcriptional cytidine deamination . *Curr Opin Lipidol*, 1995; **6**: 70–4.
7. Ellis KL, Boffa MB, Sahebkar A, Koschinsky ML, Watts GF . The renaissance of lipoprotein(a): brave new world for preventive cardiology? *Prog Lipid Res*, 2017; **68**: 57–82.
8. Chistiakov DA, Melnichenko AA, Orekhov AN, Bobryshev YV . How do macrophages sense modified low-density lipoproteins? *Int J Cardiol*, 2017; **230**: 232–40.
9. Boren J, Williams KJ . The central role of arterial retention of cholesterol-rich apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a triumph of simplicity . *Curr Opin Lipidol*, 2016; **27**: 473–83.
10. Apro J, Parini P, Broijersen A, Angelin B, Rudling M . Levels of atherogenic lipoproteins are unexpectedly reduced in interstitial fluid from type 2 diabetes patients . *J Lipid Res*, 2015; **56**: 1633–9.
11. Karasawa T, Takahashi M . Role of NLRP3 Inflammasomes in atherosclerosis . *J Atheroscler Thromb*, 2017; **24**: 443–51.
12. Brown MS, Goldstein JL . A receptor-mediated pathway for cholesterol homeostasis . *Science*, 1986; **232**: 34–47.
13. Gliemann J . Receptors of the low density lipoprotein (LDL) receptor family in man. Multiple functions of the large family members via interaction with complex ligands . *Biol Chem,* 1998; **379**: 951–64.
14. Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M . Identification of scavenger receptor SR-BI as a high density lipoprotein receptor . *Science*, 1996; **271**: 518–20.
15. Zechner R, Zimmermann R, Eichmann TO, *et al.* Fat signals—lipases and lipolysis in lipid metabolism and signaling. *Cell Metab*, 2012; **15**: 279–91.
16. Rye KA, Clay MA, Barter PJ . Remodelling of high density lipoproteins by plasma factors . *Atherosclerosis*, 1999; **145**: 227–38.
17. Doolittle MH, Ehrhardt N, Peterfy M . Lipase maturation factor 1: structure and role in lipase folding and assembly . *Curr Opin Lipidol*, 2010; **21**:198–203.
18. Stroes E, Moulin P, Parhofer KG, Rebours V, Lohr JM, Averna M . Diagnostic algorithm for familial chylomicronemia syndrome . *Atheroscler Suppl*, 2017; **23**: 1–7.
19. Cohen JC, Vega GL, Grundy SM . Hepatic lipase: new insights from genetic and metabolic studies . *Curr Opin Lipidol*, 1999; **10**: 259–67.
20. Kuivenhoven JA, Pritchard H, Hill J, Frohlich J, Assmann G, Kastelein J . The molecular pathology of lecithin:cholesterol acyltransferase (LCAT) deficiency syndromes . *J Lipid Res*, 1997; **38**: 191–205.
21. Wang DQ, Cohen DE . Absorption and excretion of cholesterol and other sterols . In: Ballantyne CM, ed. *Clinical Lipidology*. Philadelphia, PA: Saunders Elsevier, 2009: 26–44.
22. Iqbal J, Hussain MM . Intestinal lipid absorption . *Am J Physiol*, 2009; **296**: E1183–94.
23. Heath RB, Karpe F, Milne RW, Burdge GC, Wootton SA, Frayn KN . Dietary fatty acids make a rapid and substantial contribution to VLDL-triacylglycerol in the fed state . *Am J Physiol*, 2007; **292**: E732–9.
24. Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology . *Science*, 1988; **240**: 622–30.
25. Shin MJ, Krauss RM . Apolipoprotein CIII bound to apoB-containing lipoproteins is associated with small, dense LDL independent of plasma triglyceride levels in healthy men . *Atherosclerosis*, 2010; **211**: 337–41.
26. Ruuth M, Nguyen SD, Vihervaara T, *et al.* Susceptibility of low-density lipoprotein particles to aggregate depends on particle lipidome, is modifiable, and associates with future cardiovascular deaths. *Eur Heart J*, 2018; **39**: 2562–73.
27. Rudling MJ, Reihner E, Einarsson K, Ewerth S, Angelin B. Low density lipoprotein receptor-binding activity in human tissues: quantitative importance of hepatic receptors and evidence for regulation of their expression *in vivo*. *Proc Natl Acad Sci U S A*, 1990; **87**: 3469–73.

28. Brown MS, Goldstein JL . A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood . *Proc Natl Acad Sci U S A*, 1999; **96**: 11041–8.
29. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH . Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9 . *Nat Genet*, 2005; **37**: 161–5.
30. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH . Sequence variations in PCSK9, low LDL, and protection against coronary heart disease . *N Engl J Med*, 2006; **354**: 1264–72.
31. Singaraja RR, Van Eck M, Bissada N, *et al*. Both hepatic and extrahepatic ABCA1 have discrete and essential functions in the maintenance of plasma high-density lipoprotein cholesterol levels *in vivo*. *Circulation*, 2006; **114**: 1301–9.
32. Annema W, Dikkers A, de Boer JF, *et al*. HDL cholesterol efflux predicts graft failure in renal transplant recipients. *J Am Soc Nephrol*, 2016; **27**: 595–603.
33. Du XM, Kim MJ, Hou L, *et al*. HDL particle size is a critical determinant of ABCA1-mediated macrophage cellular cholesterol export. *Circ Res*, 2015; **116**: 1133–42.
34. Favari E, Calabresi L, Adorni MP, *et al*. Small discoidal pre-beta1 HDL particles are efficient acceptors of cell cholesterol via ABCA1 and ABCG1. *Biochemistry*, 2009; **48**: 11067–74.
35. Yancey PG, Bortnick AE, Kellner-Weibel G, de la Llera-Moya M, Phillips MC, Rothblat GH. Importance of different pathways of cellular cholesterol efflux. *Arterioscler Thromb Vasc Biol*, 2003; **23**: 712–19.
36. Dietschy JM, Turley SD, Spady DK . Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species, including humans . *J Lipid Res*, 1993; **34**: 1637–59.
37. Angelin B . 1994 Mack-Forster Award Lecture. Review. Studies on the regulation of hepatic cholesterol metabolism in humans . *Eur J Clin Invest*, 1995; **25**: 215–24.
38. Parini P, Davis M, Lada AT, *et al*. ACAT2 is localized to hepatocytes and is the major cholesterol esterifying enzyme in human liver. *Circulation*, 2004; **110**: 2017–23.
39. van der Velde AE, Vrins CL, van den Oever K, *et al*. Direct intestinal cholesterol secretion contributes significantly to total fecal neutral sterol excretion in mice. *Gastroenterology*, 2007; **133**: 967–75.
40. Moreau F, Blanchard C, Perret C, *et al*. *In vivo* evidence for transintestinal cholesterol efflux in patients with complete common bile duct obstruction. *J Clin Lipidol*, 2019; **13**: 213–17.e1.
41. Love MW, Dawson PA . New insights into bile acid transport . *Curr Opin Lipidol*, 1998; **9**: 225–9.
42. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis . *Annu Rev Biochem*, 2003; **72**: 137–74.
43. Chiang JY. Regulation of bile acid synthesis: pathways, nuclear receptors, and mechanisms . *J Hepatol*, 2004; **40**: 539–51.
44. Reihner E, Angelin B, Rudling M, Ewerth S, Bjorkhem I, Einarsson K . Regulation of hepatic cholesterol metabolism in humans: stimulatory effects of cholestyramine on HMG-CoA reductase activity and low density lipoprotein receptor expression in gallstone patients . *J Lipid Res*, 1990; **31**: 2219–26.
45. Ahmed O, Littmann K, Gustafsson U, *et al*. Ezetimibe in combination with simvastatin reduces remnant cholesterol without affecting biliary lipid concentrations in gallstone patients. *J Am Heart Assoc*, 2018; **7**: e009876.
46. Ginsberg HN. Lipoprotein physiology in nondiabetic and diabetic states. Relationship to atherogenesis . *Diabetes Care*, 1991; **14**: 839–55.
47. Howard BV, Howard WJ . Dyslipidemia in non-insulin-dependent diabetes mellitus . *Endocr Rev*, 1994; **15**: 263–74.
48. Grundy S. Lipids, nutrition, and coronary heart disease. In: Fuster V, Ross R, Topol E, eds. *Atherosclerosis and Coronary Artery Disease*. Philadelphia, PA: Lippincott Raven, 1996: 45–68.
49. Baraona E, Lieber CS . Effects of ethanol on lipid metabolism . *J Lipid Res*, 1979; **20**: 289–315.
50. Angelin B, Rudling M . Lipid lowering with thyroid hormone and thyromimetics . *Curr Opin Lipidol*, 2010; **21**: 499–506.
51. Angelin B, Rudling M . Growth hormone and hepatic lipoprotein metabolism . *Curr Opin Lipidol*, 1994; **5**: 160–5.
52. Heiss G, Tamir I, Davis CE, *et al*. Lipoprotein-cholesterol distributions in selected North American populations: the lipid research clinics program prevalence study. *Circulation*, 1980; **61**: 302–15.
53. Crook D, Seed M . Endocrine control of plasma lipoprotein metabolism: effects of gonadal steroids . *Baillieres Clin Endocrinol Metab*, 1990; **4**: 851–75.
54. Matthews KA, Meilahn E, Kuller LH, Kelsey SF, Caggiula AW, Wing RR . Menopause and risk factors for coronary heart disease . *N Engl J Med*, 1989; **321**: 641–6.

14.2.2 Genetic Forms of Dyslipidaemia

Stefano Romeo, Bo Angelin, and Paolo Parini

Introduction

Disturbances of blood lipids are frequent and may contribute to premature atherosclerosis and its complications, including coronary heart disease, angina pectoris, stroke, and claudication [1]. Some forms of dyslipidaemia are also linked to pancreatitis, lipid deposition in skin and tendons, and corneal arcus. Dyslipidaemias may be 'primary' and genetic, in which severe dyslipidaemia is the inevitable result of an underlying genetic mutation, and these will be the main focus of this chapter. For many dyslipidaemias, however, a genetic predisposition interacts with exacerbating nutritional, pharmacological, or disease factors, for example in conditions such as diabetes, insulin resistance, hypothyroidism, hypopituitarism, hypo/hypercortisolism, chronic kidney disease, and polycystic ovary syndrome. Suggested investigations helpful in screening for secondary dyslipidaemia are shown in Box 14.2.2.1.

Pathogenic Aspects of Genetic Dyslipidaemias

Highly penetrant genetic dyslipidaemias are often severe and relatively easy to recognize. However, the most common heterozygous forms of genetic dyslipidaemias may go unrecognized for many years until experienced doctors identify them. If left untreated, genetic dyslipidaemias may lead to reduced lifespan or poor quality of life, depending on the severity of the phenotype. Since the body is generally exposed to such negative effects from birth, it is important to recognize genetic dyslipidaemias early in life, in order to initiate preventive treatment.

The study of genetic dyslipidaemias has yielded critical insights into the physiology and pathophysiology of human lipoprotein metabolism, also allowing identification of metabolic networks and endo-phenotypes that are common with chronic non-communicable diseases. Considering the vast number of proteins involved in lipid metabolism, it is not surprising that a relatively large number of monogenic diseases have been identified (Box 14.2.2.2). With the development of modern sequencing technology, we are rapidly approaching a more detailed understanding of interindividual variation. Accurate molecular diagnoses and targeted therapies are becoming a reality in an increasing proportion of patients with genetic dyslipidaemias.

Classification of lipoprotein disorders has historically been based on serum lipoprotein distribution, assessed by electrophoresis as originally outlined by Fredrickson in 1965 (Box 14.2.2.3) [2]. Currently, the classification of dyslipidaemias is based on the descriptions of more specific phenotypes (Box 14.2.2.4). These genetic forms of dyslipidaemias will be the focus of the rest of this chapter.

Box 14.2.2.1 Suggested laboratory parameters to be screened for to exclude secondary dyslipidaemia

Total cholesterol
HDL-cholesterol
LDL-cholesterol
Triglycerides
Lp(a)
APOA1
APOB
TSH, fT_4
Glucose, insulin
Full blood count
EVF
Creatinine, Na, K
ALT, AST, ALP
CK

Box 14.2.2.2 Terminology in genetics

Homozygous: Harbouring the same mutation in both maternal and paternal gene copies.

Compound heterozygous: Harbouring two different mutations, one in the maternal and one in the paternal gene copy.

Double heterozygous: Harbouring two different heterozygous mutations in two different genes (e.g. a mutation in the LDL-receptor and a mutation in the APOB gene)

(Autosomal) Recessive: Disease become manifest in the presence of mutations in both maternally and the paternally inherited copies of a gene. Affected individuals can be homozygous, compound heterozygous, or double heterozygous.

(Autosomal) Dominant: The disease become manifest in the presence of only one mutation, usually in either the maternal or paternal gene copy.

Familial Hypercholesterolaemia

Heterozygous Familial Hypercholesterolaemia (FH)

Familial hypercholesterolaemia (FH) is an autosomal codominant disease caused by mutations in the genes encoding the LDL-receptor (LDLR), apolipoprotein B (APOB) or proprotein convertase subtilisin/kexin type 9 (PCSK9). While mutations in the LDLR or APOB causing loss-of-function determine FH, those in PCSK9 cause gain-of-function. Described as early as in the 1930s, FH is the most common monogenic disorder of lipid metabolism with a prevalence of ~1/250 in most populations [3]. Some regions are known to have clusters with a higher incidence of FH, due to population isolation permitting expansion of rare mutations—the so-called founder effect—as in the French Canadians and Boers in South Africa, for example. The most striking feature of individuals with FH is elevated levels of LDL-C (usually > 5 mmol/L) resulting in 10–15-fold increased risk in cardiovascular disease. If untreated, heterozygous FH carriers have an estimated 10–15 years shorter expected lifespan.

Box 14.2.2.3 Fredrickson's classification of hyperlipidaemias

Type I fasting chylomicronaemia
Type II LDL elevation A isolated
B Combined with VLDL elevation
Type III Beta-VLDL (dysbetalipoproteinamia)
Type IV VLDL elevation
Type V Chylomicronaemia combined with VLDL elevation

Box 14.2.2.4 Monogenic dyslipidaemias

High LDL-C (Xanthomata)
Autosomal dominant hypercholesterolaemia (familial hypercholesterolaemia)
Elevated lipoprotein (a)
Autosomal recessive hypercholesterolaemia (ARH)
Polygenic hypercholesterolaemia
Sitosterolaemia
Cerebrotendineous xanthomatosis (CTX)

High circulating triglycerides
Familial chylomicronaemia syndrome
Familial hypertriglyceridaemia

Combined hypercholesterolaemia and hypertriglyceridaemia
Familial combined hyperlipidaemia
Familial dysbetalipoproteinaemia (type III)

Low cholesterol
Abetalipoproteinaemia
Hypobetalipoproteinaemia
Chylomicron retention disease (Anderson's disease)
Combined hypolipidaemia
PCSK9 loss-of-function mutations
APOE2/2

Disorders of HDL-C
Hypoalphalipoproteinaemia (low HDL)
Tangier disease (ABCA1); apo A-1 deficiency; LCAT deficiency
Hyperalphalipoproteinaemia (high HDL)
CETP or SCARB1 deficiency

Box 14.2.2.5 Dutch Lipid Clinic Network Criteria for making a diagnosis of familial hypercholesterolaemia in adults

Family history score
First-degree relative with known premature coronary and/or vascular disease (men aged <55 years and women aged <60 years) or
First-degree relative with known low-density lipoprotein-cholesterol (LDL-C) above the 95th percentile for age and sex 1
First-degree relative with tendinous xanthomata and/or arcus cornealis or
Children aged <18 years with LDL-C above the 95th percentile for age and sex 2

Clinical history
Patient with premature coronary artery disease (ages as above) 2
Patient with premature cerebral or peripheral vascular disease (as above) 1

Physical examination
Tendinous xanthomata 6
Arcus cornealis prior to 45 years of age 4

LDL-C (mmol/L)
LDL-C ≥8.5 8
LDL-C 6.5–8.4 5
LDL-C 5.0–6.4 3
LDL-C 4.0–4.9 1

DNA analysis
Functional mutation in the low-density lipoprotein receptor (LDLR), apolipoprotein B
(APOB) or proprotein convertase subtilisin/kexin type 9 (PCSK9) gene 8

Stratification Total score
Definite familial hypercholesterolaemia (FH) ≥8
Probable FH 6–7
Possible FH 3–5
Unlikely FH <3

Data derived from Mach F, Ray KK, Wiklund O, *et al.* Adverse effects of statin therapy: perception vs. the evidence – focus on glucose homeostasis, cognitive, renal and hepatic function, haemorrhagic stroke and cataract. *European Heart Journal*, 2018; 39: 2526–39.

Diagnosis

Clinical diagnosis of FH is based on a dominant pattern of inheritance in the family pedigree. Total cholesterol levels are ≥8 mmol/L (≥6 mmol/L in children), triglyceride, and HDL-cholesterol levels are generally unremarkable. Scores to evaluate the probability of FH based on clinical criteria have been developed, and although far from perfect may be useful for decisions about family screening. The most commonly used is the Dutch Lipid Clinic Network (DLCN) score (Box 14.2.2.5) [3]; others are the Simon Broome and the Make Early Diagnosis to Prevent Early Deaths (MEDPED) diagnostic criteria. Among the clinical signs of cholesterol accumulation, spheroidal xanthomata of the extensor tendons of the hands, elbows, knees, or the Achilles tendons are a strong support for the diagnosis of FH (Figure 14.2.2.1). Arcus senilis (a white circle of cholesterol deposit in the outer edge of the iris appearing before the age of 40) or xanthelasmata (a demarcated yellowish area of cholesterol under the skin around the eyelids) are also common but are not pathognomonic for FH (Figure 14.2.2.1).

Lifestyle Intervention and Assessing the Risk

It is recommended that FH patients adhere to a healthy lifestyle and diet [3]. It is imperative that smokers quit smoking, and are referred to a specialized tobacco unit/programme when necessary. Advice to children and young adults not to start smoking is very important. Individuals with this condition generally do not achieve clinically relevant reduction of cholesterol in response to dietary intervention or weight loss alone. The absence of reduction of LDL-C after changing lifestyle indeed increases the suspicion of FH. The cardiovascular risk in FH patients cannot be inferred by using the traditional scores (e.g. Framingham risk score); these individuals are at high risk for cardiovascular disease and should be treated to reduce LDL-C levels early in life.

Treatment

LDL-cholesterol is the primary target of therapy and the reduction in both cardiovascular and total mortality is proportional to the degree of LDL-cholesterol reduction [3]. Treatment in FH should be started during teenage years with a goal of 3.5 mmol/L. In FH adults, the LDL-C to achieve should be 2.5 mmol/L while in individuals with additional risk factors for cardiovascular disease (diabetes mellitus, hypertension, smoking, obesity, insulin resistance) or in secondary prevention LDL-C levels should be maintained below 1.8 mmol/L. All untreated individuals with FH above age 40 should be considered to be at very high cardiovascular risk, as they have been exposed to elevated LDL-cholesterol levels since birth.

Statins are the first line treatment because they reduce circulating LDL-C by increasing removal of LDL particles through upregulation of hepatic LDLRs. Treatment with the maximal dose of a potent statin should ideally be used: either atorvastatin 80 mg or rosuvastatin 40 mg daily (or pitavastatin 4 mg/daily). Alternative statins are simvastatin or pravastatin. Simvastatin 80 mg should not be used because it is associated with elevated risk of myositis and

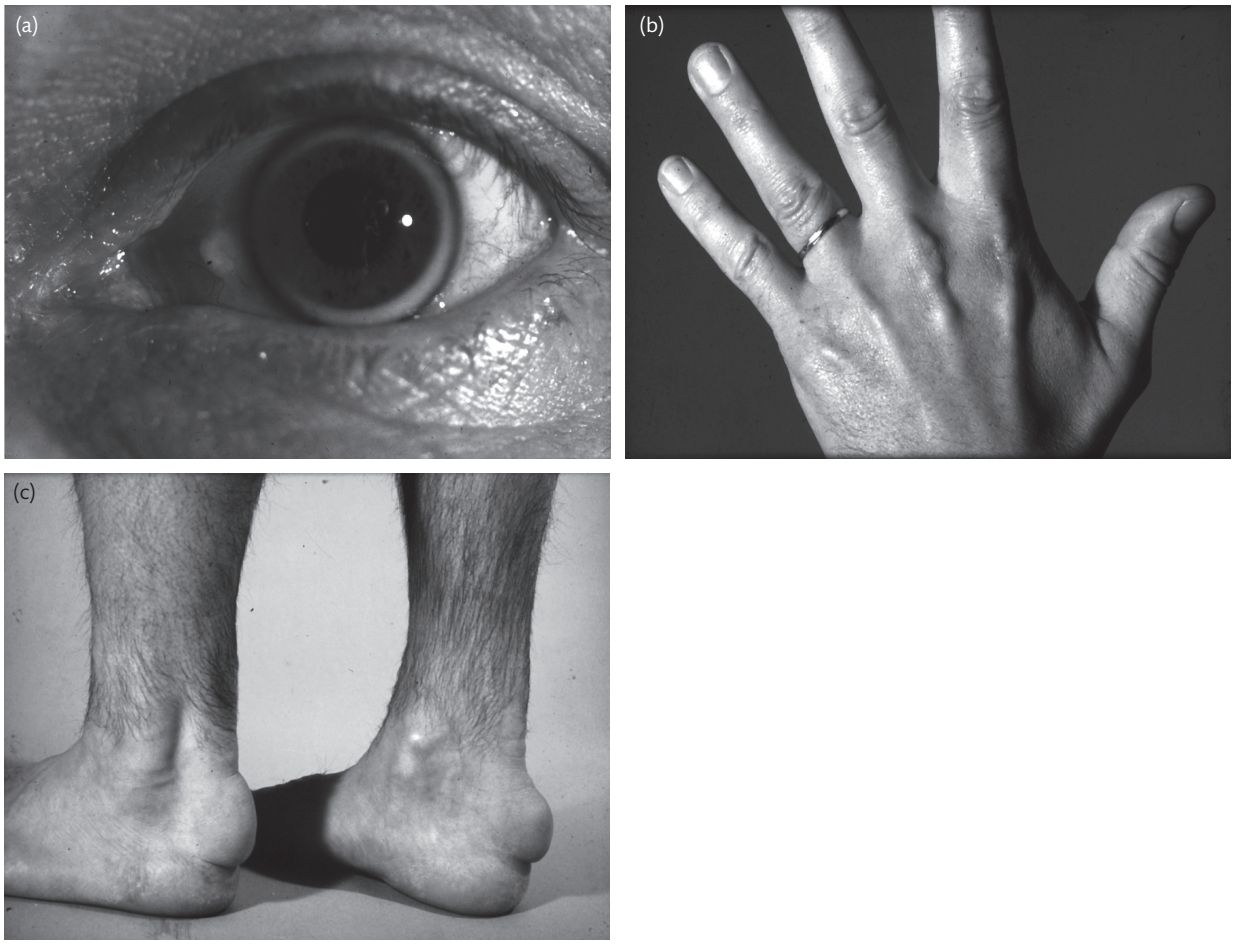


Figure 14.2.2.1 Characteristic clinical signs in a 40-year-old male patient with heterozygous FH (left to right): *Arcus senilis*; xanthomas on flexor tendons; xanthoma on Achilles tendon.

rhabdomyolysis. Statins are generally well tolerated [4, 5]. The most severe statin side effect is rhabdomyolysis occurring in approximately 1 of every 100 000 prescriptions. The major symptom of this is severe pain in large proximal muscles not explained by fever or severe physical exercise. The risk of rhabdomyolysis increases when a statin is combined with some drugs, including gemfibrozil, antifungals, macrolides, calcium channel blockers, protease inhibitors, and cyclosporin. Transient increases in liver enzymes occur in 0.5–2% of patients taking statins but are not clinically relevant. Routine monitoring of creatine kinase or transaminases is not recommended, and an increase of up to threefold the reference levels of these enzymes should not generally contraindicate continued treatment. Statin therapy is associated with a modest increase in the risk of new-onset diabetes mellitus (ca. 1/1000 patient-years). The risk is significantly higher in patients with metabolic syndrome or prediabetes. Statins do not adversely affect cognitive function.

Second-line treatment, in addition or alone, is with ezetimibe (10 mg daily), which achieves an average further reduction of 20% by reducing uptake of cholesterol from the intestine. Bile acid binding resins such as colesevelam (up to 3750 mg a day) generally further effectively reduce LDL-C. Treatment with the recently available inhibitors of PCSK9, such as the subcutaneous injection of monoclonal antibodies alirocumab (75–150 mg every two weeks) or evolocumab (140 mg every 2 weeks), can reduce LDL-C up to 70%, opening up the possibility to reach even the very stringent treatment goals for LDL-C in heterozygous FH [6, 7].

Cascade Screening in Families with FH

Considering the high risk for premature cardiovascular disease, the dominant pattern of inheritance and the well-established positive effects from lipid-lowering therapy, it is reasonable to offer first-degree relatives of patients with a diagnosis of heterozygous FH screening for this disorder [3]. This can be achieved through screening of lipid levels (Box 14.2.2.1), and/or through DNA analysis. There is ample evidence for clinical benefit from such procedures, and a number of programmes have implemented them, with offspring of known carriers actively sought through their healthcare providers. Although this may be relatively straightforward in a family with established hypercholesterolaemia and a known disease-causing mutation in the LDLR, APOB, or PCSK9, there are some caveats that should be considered when engaging in such an approach [8, 9]. First of all, the spectrum of mutations in the LDLR is large, with thousands of different mutations described, complicating genetic screening in families where the primary defect has not yet been identified. Furthermore, a remaining fraction of patients with clinical FH may have mutations in additional genes yet to be implicated in FH. In addition, there are relatively frequent disease-modifying mutations in other genes (such as variants of

apolipoprotein E (APOE) or loss-of-function mutations in PCSK9) which prevent the hypercholesterolaemic phenotype being expressed in carriers of known FH-causing mutations, thus leading to underestimation of their potential to transmit the trait. Increased availability of next generation sequencing techniques will make diagnosis more exact but will also require a more profound understanding from managing physicians.

Homozygous FH

Individuals homozygous, or compound heterozygous, for mutations in FH-causing genes generally present with a much more severe phenotype [10]. The prevalence of this disorder is approximately 1 in 500 000–1 000 000 except for in isolated ethnic groups, particularly when intermarriage between cousins is frequent. In the classical form, these individuals generally have LDL-C levels > 9 mmol/L and may present with cardiovascular disease even before the age of 10. Xanthomata are present at an early age, and frequently are the reason for clinical presentation. Double heterozygous FH (i.e. FH caused by heterozygosity for disease-causing mutations in two different genes; **Box 14.2.2.2**) has an intermediate phenotype. In addition to marked tendon xanthomas, planar cutaneous xanthomata are a distinct sign of these more severe forms (**Figure 14.2.2.2**). FH homozygotes with no residual LDLR expression are essentially resistant to treatments aimed at increasing LDLR function, whereas some lipid-lowering can be achieved in those with some remaining function. The most efficient treatment available today is removal of plasma LDL by selective LDL-apheresis, generally at weekly intervals [10]. Inhibition of hepatic very low-density lipoprotein (VLDL) secretion by oral treatment with lomitapide (10–40 mg a day) or subcutaneous injection of mipomersen (200 mg/ml per week) also has effect and may reduce the frequency of LDL-apheresis therapy. A side effect of such treatment is development of fatty liver disease, which needs to be monitored. For resistant cases of homozygous FH, more drastic treatments such as liver transplantation has been tried, and initial attempts at gene therapy have been undertaken, with modest success to date.

Autosomal Recessive Hypercholesterolaemia (ARH)

In rare cases, a phenotype similar to homozygous FH but with a recessive inheritance, with early extensive xanthomata but with normal parental LDL levels may be observed (**Figure 14.2.2.3**). This is likely to be due to homozygosity for mutations in the LDLR adapter protein-1 (*LDLRAP1*) gene, which influences recycling of the LDLR, leading to marked LDL-C elevations and premature cardiovascular disease [11, 12]. These patients generally respond well to treatment with statins in combination with ezetimibe, which may even produce evidence of regression of manifest atherosclerosis.

Sitosterolaemia

Another very rare form of recessive dyslipidaemia is sitosterolaemia, where large xanthomata accumulate due to the deposition of plant sterols such as sitosterol and campesterol. This is generally due to homozygous or compound heterozygous mutations in the ABCG5 and ABCG8 genes, which code for an intestinal and hepatic transmembrane transporter complex responsible for the elimination of plant sterols and cholesterol [13]. The patients have very high levels of sitosterol and campesterol, while plasma LDL-C may or

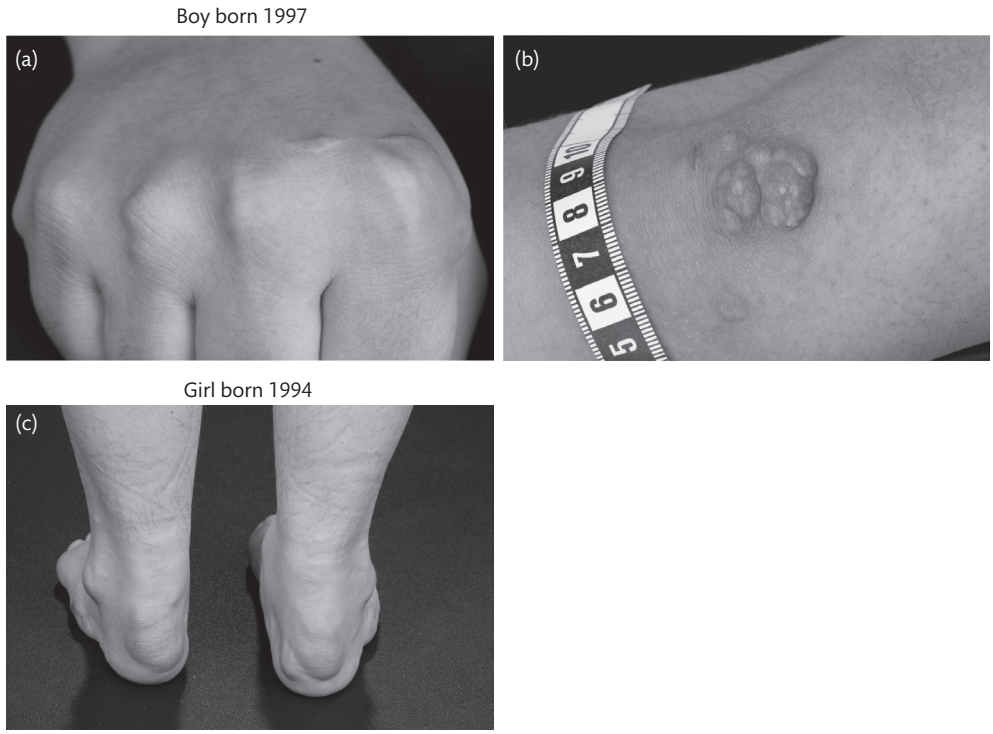


Figure 14.2.2.2 Characteristic clinical signs in a 15yr-old male patient with homozygous FH (left to right): Xanthomas on flexor tendons; Xanthomas on Achilles tendons; planar xanthomas on elbows.

Autosomal recessive hypercholesterolaemia: normalization of plasma LDL cholesterol by ezetimibe in combination with statin treatment

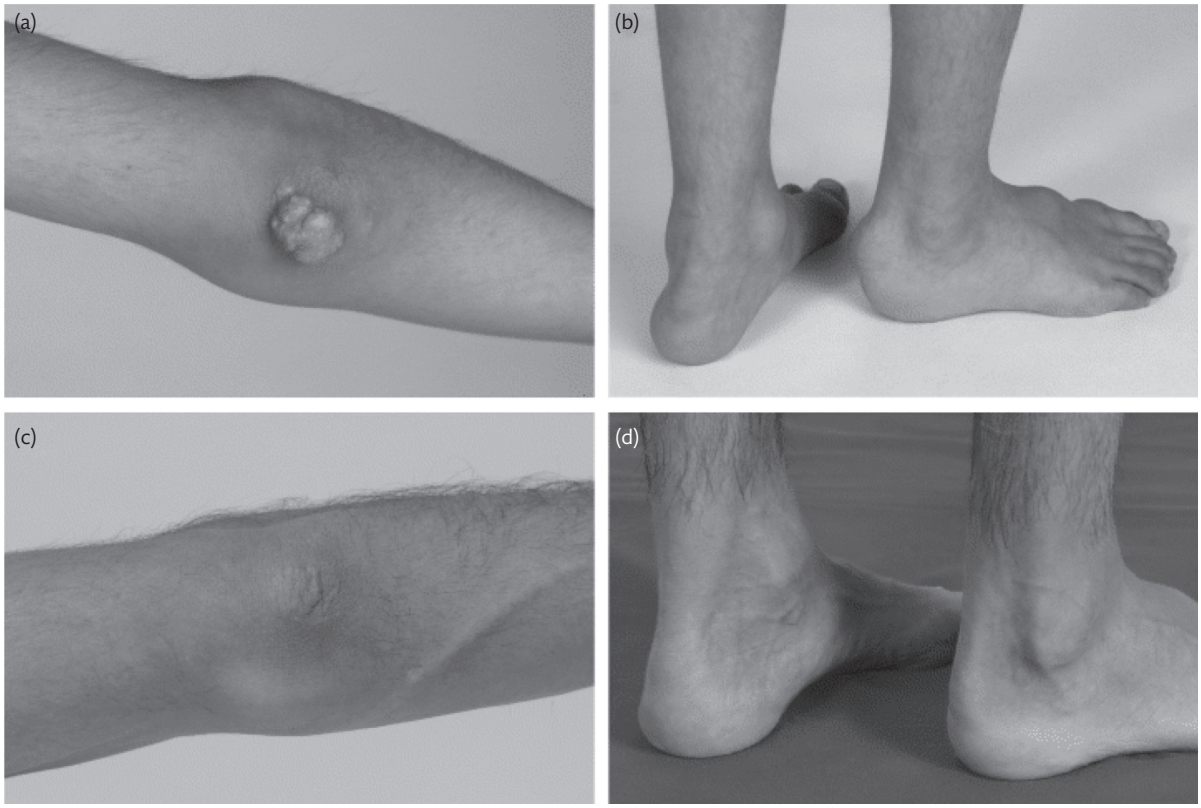


Figure 14.2.2.3 Xanthomas in a patient with ARH before (a, b) and following treatment with ezetimibe and statin (c, d).
Reproduced with permission from Lind S, Olsson AG, Eriksson M, Rudling M, Eggertsen G, Angelin B. Autosomal recessive hypercholesterolaemia: normalization of plasma LDL-cholesterol by ezetimibe in combination with statin treatment. *J Intern Med*. 2004; 256: 406–12. Copyright © 2004, John Wiley and Sons. (Ref 10)

may not be elevated. In these cases, plant sterols are a major component of the xanthomata. Patients with sitosterolaemia show extreme phenotypic heterogeneity, ranging from almost asymptomatic individuals to those with severe hypercholesterolaemia leading to accelerated atherosclerosis and premature cardiac death. Haematologic manifestations include haemolytic anaemia with stomatocytosis, macrothrombocytopaenia, splenomegaly, and abnormal bleeding. The therapy includes dietary restriction of both cholesterol and plant sterols and the use of ezetimibe, an intestinal sterol absorption inhibitor. Vegetable oils, wheat germs, nuts, seeds, avocado, shortening, margarine, and chocolate should be restricted. If present, hypercholesterolaemia in these patients is extremely responsive to low cholesterol diet and bile acid sequestrants.

Cerebrotendinous Xanthomatosis (CTX)

In this very rare recessive disease, the underlying molecular defect is in the sterol 27-hydroxylase (*CYP27A1*) gene, resulting in an inability to synthesize normal bile acids from cholesterol [14]. Loss of feedback inhibition of bile acid synthesis results in a very high hepatic production of bile acids and cholestanol. The increased bile acid synthesis leads to increased *LDLR* expression. Hence, patients with CTX generally have low levels of LDL-C. Characteristic clinical manifestations of CTX include chronic diarrhoea, bilateral cataracts, tendon xanthomas, and neurologic dysfunction. CTX is associated with abnormally high levels of cholestanol in the blood and accumulation of cholestanol and cholesterol in the brain, tendon xanthomas, and bile. There is marked variability in signs and symptoms, severity, and age of onset between patients. CTX is progressive and potentially debilitating or fatal, particularly with respect to neurologic presentations that can include intellectual disability, autism, behavioural and psychiatric problems, and dementia. Chenodeoxycholic acid treatment helps to restore normal bile acid synthesis and reduces the levels of cholestanol. Since chenodeoxycholic acid also appears to be generally effective in preventing adverse clinical manifestations of CTX from occurring or progressing if administered early enough, an early diagnosis is of great importance.

Elevated Lp(a)

Lipoprotein Lp(a) is an LDL-like particle where the plasminogen-like APO(a) is covalently linked to the structural APOB molecule. Studies using genetic variations whose primary effect is to increase circulating Lp(a) levels have shown that elevated Lp(a) increases the risk for cardiovascular disease independently of LDL-C or non-HDL-C levels and of the presence of other cardiovascular risk factors [15]. The level of Lp(a) is strongly influenced by a genetically determined size variation, with an inverse relation between the plasma level and isoform size. Individuals with FH tend to have higher Lp(a) than predicted by their APO(a) genotype, and the

degree of Lp(a) increase predicts risk in this condition. Clinically, elevated levels of Lp(a) should also be suspected in the context of progressive and recurrent myocardial infarction despite effective statin therapy. The role of Lp(a) in the pathogenesis of cardiovascular disease seems to be due to its atherogenic and prothrombotic properties and the European Atherosclerosis Society has suggested that the desirable level of Lp(a) is less than 125 nmol/L (50 mg/dl) [15]. There is no effective treatment to reduce Lp(a), although sex hormones and thyroid hormone are known to decrease its concentration. Traditional lipid-lowering therapies (e.g. statins, fibrates, and ezetimibe) have no or very minor effects on Lp(a). Treatment with niacin, lipoprotein apheresis, or PCSK9-inhibitors can decrease Lp(a) levels. Niacin and PCSK9-inhibition decrease the Lp(a) levels approximately 30%, whereas lipoprotein apheresis decreases them by 25–40% [16]. Since these treatments also have a positive effect on plasma lipid profiles; it has been difficult to discern the specific effects due to Lp(a) lowering.

Familial Chylomicronaemia Syndrome (FCS)

FCS, previously referred to as type I hyperlipoproteinemia according to Fredrickson's classification, is a recessive disorder characterized by high circulating levels of triglycerides, typically above 10 mmol/L, in the absence of secondary causes [17, 18]. This disease results from reduced catabolism of chylomicrons. Clinically, patients may have abdominal pain and recurrent pancreatitis that may be associated with eruptive xanthomata, hepatomegaly, and lipemia retinalis. Moreover, the hypertriglyceridemia is usually resistant to treatment with fibrates. The prevalence of this disease is low at ~1/1 000 000, except in isolated populations or in those in which consanguineous marriage is prevalent. The most common genetic defects responsible for FCS are loss-of-function mutations in lipoprotein lipase (LPL), APOC2, apolipoprotein A5 (APOA5), lipase maturating factor 1 (LMF1) and glycosylphosphatidylinositol anchored high-density lipoprotein binding protein 1 (GPIHBP1) genes [19]. Double heterozygotes for mutations in pairs of these genes generally have a milder phenotype and respond to fibrate treatment. There are no specific drugs to treat FCS, but promising therapies are currently under investigation, including inhibition of APOC3 and of angiopoietin like-3 (ANGPTL3). Also gene therapy for LPL deficiency has been tested but it is not currently available for a lack efficacy. FCS generally responds well to a diet with a fat content of 15–20 g per day (10–15% of total daily energy intake). The diet should meet recommendations for essential fatty acids: alpha-linolenic acid and linoleic acid. Complex carbohydrate foods should be preferred while limiting simple and refined carbohydrates. Supplementation with fat-soluble vitamins, minerals, and medium-chain triglyceride oil, is given as needed. Recommended foods include vegetables, whole grains, legumes, lean protein foods, fruits in limited amounts, and fat-free milk products without added sugars. Foods to avoid include alcohol and products high in sugar.

Familial Hypertriglyceridaemia (FHTG)

Apparently autosomal dominant elevated triglyceride concentration due to increased VLDL (type IV hyperlipoproteinaemia), is relatively common and is probably mechanistically heterogeneous, being attributable to various underlying molecular defects [17, 20]. It may be associated with the presence of chylomicrons in the fasting state (type V hyperlipoproteinaemia). Although the risk of cardiovascular disease is probably increased, it is not as marked as for increased LDL levels [20, 21]. Eruptive xanthomas may occur, and at very high triglyceride levels there is an enhanced risk of pancreatitis. Overproduction of VLDL-triglycerides is frequent, and there are links to enhanced cholesterol and bile acid production in some cases [22]. Optimizing weight, low caloric intake, and exercise are important, and a family history of early cardiovascular disease strengthens the introduction of drug therapy, with statins or fibrate drugs as first choice.

Familial Combined Hyperlipidaemia (FCHL)

This clustering of hyperlipidaemia of varying hyperlipidaemic phenotypes (elevated VLDL, LDL or their combination) with apparent autosomal dominant inheritance was first described in 1973 [23, 24], and still remains a pathogenic mystery, lacking solid diagnostic markers. The pattern is common (perhaps 1%), and it is clearly linked to an increased risk for early cardiovascular risk [14], being seen in at least 10% of individuals with myocardial infarction before the age of 60. The lipid phenotype appears to be more clearly expressed with increasing age, and there is a large overlap with abdominal obesity, insulin resistance, hypertension, and type 2 diabetes. 'FCHL' is most probably due to a combination of several genetic variants and aggravated by lifestyle influences [25]. In carefully selected patients, overproduction of APOB-containing lipoproteins and deficient lipolysis seem to be involved [26–28]. Importantly, the presence of a combined lipid phenotype in a patient with a family history of mixed hyperlipidaemia and early vascular disease, particularly in combination with high APOB levels, could motivate active screening in first-degree relatives. In addition to lifestyle adjustment, treatment with statins and/or fibrates generally has very good effect.

Type III Dyslipoproteinaemia

This is a very rare form of highly atherogenic dyslipidaemia which is also important to identify in general practice, since its treatment frequently is very successful. It is generally caused by the combination of two abnormalities: homozygosity for the APOE2/2 genotype (which is present in ~1% of the population), and overproduction of lipoproteins such as seen in FCHL or type 2 diabetes [29, 30]. Hypothyroidism in a subject with the APOE2/2 genotype frequently results in a type III lipoprotein pattern. It should be recognized that subjects with the APOE2/2 genotype generally have low lipoprotein levels, but when subject to an increased lipoprotein flux they accumulate chylomicron and VLDL remnants due to the reduced interaction of these with hepatic LDLRs. This results in elevated cholesterol and triglyceride levels, generally in the range of 8–10 mmol/L; LDL-cholesterol (and APOB) is generally low. The remnant particles make the plasma opalescent, and the uptake of those so-called beta-VLDL in macrophages results in a very typical (almost pathognomic)

lipid deposition in the palmar creases as well as tuberous xanthomata, particularly of the elbows and knees. This is probably also the reason for the high propensity to develop atherosclerotic deposits, often most pronounced in peripheral arteries. Following treatment with drugs such as fibrates or statins, there is generally a very marked response, not only in plasma lipids but also both in xanthomas and arteriopathy. Clinically, it is also of note that type III hyperlipidaemia carries a high risk of atherosclerosis despite generally having a low APOB level and a low APOB/APOA1 ratio.

Familial Hypobetalipoproteinaemia (FHBL) and Abetalipoproteinaemia (ABL)

These are disorders belonging to the same spectrum, characterized by low production of lipoproteins [13, 31]. HBL can occur in homozygous or heterozygous form, and is caused by mutations in APOB, generally leading to truncation of the protein and deficient or decreased secretion of chylomicrons and VLDL, respectively. While the heterozygotes have low LDL, they generally have normal fat absorption and seem to have a normal or even increased life span. The homozygotes have fat malabsorption that results in failure to thrive and deficiency of fat-soluble vitamins. In ABL, there is generally a homozygote or compound heterozygote mutation in microsomal triglyceride transfer protein (MTTP) which gives essentially a similar phenotype. Both conditions also predispose to fatty liver disease due to the retention of VLDL, with progressive fibrosis and risk of developing hepatocellular carcinoma. Treatment is generally based on dietary treatment with vitamin substitution and medium chain length triglyceride/MCT).

Chylomicron Retention Disease (CMRD)

Also known as Anderson's disease, CMRD is a very rare autosomal recessive disease characterized by lipid malabsorption due to mutations in the SAR1B gene, a protein involved in intracellular trafficking and chylomicron assembly [13, 32]. This causes the chylomicrons and VLDL synthesized in the small intestine and liver to be retained in the respective cells of origin, resulting in fat malabsorption and failure to thrive. Following a fatty meal, there is no postprandial chylomicron secretion, and a 'fatty gut' can be seen in intestinal biopsies. Treatment is similar to that in FHBL and ABL, and the clinical course seems somewhat milder.

Low LDL Levels

Loss-of-function mutations in the *PCSK9* gene are associated with low LDL-C levels and have been shown to protect against cardiovascular disease in several populations [33]. The APOE2/2 variant is also linked to lower LDL levels, and in both situations this probably reflects increased clearance of LDL from the circulation. Subjects with heterozygous FHBL, which may be as frequent as 0.1% in the population, generally have extreme low LDL levels, probably due to the combination of deficient production and increased clearance of APOB [13]. The risk for development of fatty liver is increased and carriers of these sequence variations have higher risk of hepatocellular carcinoma [34].

Familial Combined Hypolipidaemia

This is a condition characterized by very low/undetectable levels of all lipoprotein species, caused by mutations in the angiopoietin like protein 3 (ANGPTL3) gene [35]. This gene encodes a protein which inhibits the activity of LPL explaining the very low levels of circulating triglycerides; the mechanisms resulting in the reductions in HDL-C and LDL-C are incompletely understood. Individuals with this condition have been shown to be more sensitive to insulin, and are probably protected from cardiovascular disease and type 2 diabetes. These findings have stimulated the development of treatment of hyperlipidaemic conditions by administration of antibodies or antisense oligonucleotides targeting ANGPTL3.

Hypoalphalipoproteinaemia (Low HDL)

Low HDL-cholesterol (<1 mmol/L for males, <1.3 mmol/L for females) is relatively common, and has been associated with an enhanced risk for cardiovascular disease. In most cases, this finding probably reflects inefficient lipid transport as seen in the 'metabolic syndrome', with slow VLDL-clearance, moderate triglyceride increase, abdominal obesity, and insulin resistance [36]. This should serve as an indicator for interventions to promote metabolic health, including lifestyle changes. There is no clearly beneficial pharmacological treatment directed at HDL levels, although oestrogen substitution after menopause probably works at least partly via this pathway. There are some rare monogenic disorders that may cause substantial lowering of HDL-C.

Tangier Disease

This is a classical recessive dyslipidaemia which is caused by mutations in the *ABCA1* gene [37]. Clinically, there is relatively modest evidence for an increased risk of cardiovascular disease, despite substantial lipid deposition in macrophages, seen as hypersplenism and large, orange-coloured tonsils. The lack of functional *ABCA1* results in deficient uptake of cholesterol from peripheral cells, and APOA1 protein is rapidly lost from the circulation.

Lecithin–Cholesterol Acyltransferase (LCAT) Deficiency

This is a recessive disorder in which esterification of cholesterol and its transfer from HDL to VLDL and chylomicrons are absent, leading to very low levels of esterified cholesterol in plasma, and accumulation of small discoid HDL_3 particles [37]. Clinically, the most relevant finding is gradual deterioration of kidney function which frequently leads to end stage kidney disease which may require dialysis and kidney transplantation. There is also a high frequency of early cataract formation, particularly seen in so-called fish-eye disease. The evidence for an increased risk of cardiovascular disease is so far not fully convincing.

APOA1 Deficiency

Although there are structural variants in APOA1 that seem to result in deficient reverse cholesterol transport and early cardiovascular

disease, they are relatively rare [37]. Most of those described are inherited as a dominant trait.

Hyperalphalipoproteinaemia (High HDL)

Based on epidemiological evidence of an inverse correlation between HDL-C and cardiovascular risk, there was for a long time a conviction that genetic variants resulting in high HDL or APOA1 levels would be beneficial, a concept supported by reports of a reduced frequency of atherosclerosis in Japanese families with cholesterol ester transfer protein (CETP) deficiency. However, essentially negative results from trials of treatment with CETP inhibitors, together with accumulating genetic evidence of a more complex risk pattern, have led to a more balanced view. The mechanism responsible for the HDL-C elevation (e.g. through a reduced flux of cholesterol removal or by an increased production of APOA1) is probably more important for the development of atherosclerosis than changes in static HDL-C concentration, and explains observations of a U-shaped risk curve for HDL-C [38].

Cholesterol Ester Transfer Protein (CETP) Deficiency

CETP transfers and exchanges cholesteryl esters and triglycerides between HDL and triglyceride-rich lipoproteins. Homozygous carriers, identified mainly in Japan, show 3 to 6-fold elevated HDL-C levels, with concomitant high levels of APOAI, APOAII, APOCIII and APOE [39]. The levels of LDL-C and APOB are usually slightly reduced. The HDL particles in CETP-deficient patients are very large, rich in cholesterol esters, and poor in triglycerides. The increase in HDL-C is accounted for by a rise of HDL_2, whereas HDL_3-C is normal.

Scavenger Receptor Class B Type I (SCARB1) Mutations

SCARB1 mediates delivery of HDL-C to the liver and steroidogenic organs. Families with a missense mutation (P297S) in this gene have been identified [40]. The carriers have increased HDL-C and a reduced capacity for efflux of cholesterol from macrophages, but the carotid artery intima-media thickness was similar in carriers and non-carriers. Platelets from carriers had an increased content of unesterified cholesterol and were functionally impaired. In carriers, adrenal steroidogenesis was attenuated, as evidenced by decreased urinary excretion of corticosteroids metabolites, decreased response to corticotropin stimulation, and in some cases symptoms of diminished adrenal function.

Concluding Remarks

Atherosclerosis is a chronic process started by accumulation of cholesterol in the subintimal space of arteries. The inflammatory component is driven by cholesterol crystals and sustained by some fatty acids present in triglycerides and reduced by HDL particles. Hence, dyslipidaemia represents a highly important risk factor for cardiovascular disease by driving the atherosclerosis process. The phenotypes of the different dyslipidaemias are the consequence of an interplay between genetic and environmental influences. Abnormalities of lipid metabolism are also commonly due to or precipitated by underlying endocrine diseases. The molecular understanding of how lipid metabolism is regulated has resulted in an explosive development of diagnostic tools for individual diagnosis and in the rapid emergence of effective and specific treatment options. The knowledge that the degree of lifetime exposure to cholesterol carried in LDL particles (and remnant particles) in circulation determines the occurrence of clinical relevant cardiovascular disease dictates an early diagnosis and prevention in order to lead to large positive health effects. Hence, the continuous dissemination of new knowledge to practising physicians is a pillar of prevention strategies. Nevertheless, the recognition of characteristic physical signs in patient care, and the value of obtaining a relevant family history in patients with primary dyslipidaemia will continue to be of major importance also in the future.

REFERENCES

1. Durrington P. Dyslipidaemia. *Lancet, 2003;* **362**: 717–31.
2. Fredrickson DS, Levy RI, Lees RS . Fat transport in lipoproteins--an integrated approach to mechanisms and disorders . *N Engl J Med*, 1967; **276**: 34–42, 94–103, 148–56, 215–25, 273–81.
3. Nordestgaard BG, Chapman MJ, Humphries SE, *et al.* Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur Heart J*, 2013; **34**: 3478–90.
4. Mach F, Ray KK, Wiklund O, *et al.* Adverse effects of statin therapy: perception vs. the evidence—focus on glucose homeostasis, cognitive, renal and hepatic function, haemorrhagic stroke and cataract. *Eur Heart J*, 2018; **39**: 2526–39.
5. Stroes ES, Thompson PD, Corsini A, *et al.*; European Atherosclerosis Society Consensus Panel. Statin-associated muscle symptoms: impact on statin therapy-European Atherosclerosis Society Consensus Panel Statement on Assessment, Aetiology and Management. *Eur Heart J*, 2015; **36**: 1012–22.
6. Sharpiro MD, Tavori H, Fazio S. PCSK9: from basic science discoveries to clinical trials. *Circ Res*, 2018; **122**: 1420–38.
7. Larsen LE, Stoekenbroek RM, Kastelein JJP, Holleboom AG. Moving targets: recent advances in lipid-lowering therapies. *Arterioscler Thromb Vasc Biol*, 2019; **39**: 349–59.
8. Sturm AC, Knowles JW, Gidding SS, *et al.* Clinical genetic testing for familial hypercholesterolemia: JACC Scientific Expert Panel. *J Am Coll Cardiol*, 2018; **72**: 662–80.
9. Berberich AJ, Hegele RA. The complex molecular genetics of familial hypercholesterolemia. *Nat Rev Cardiol*, 2019; **16**: 9–20.
10. Cuchel M, Bruckert E, Ginsberg HN, *et al.* Homozygous familial hypercholesterolaemia: new insights and guidance for clinicians to improve detection and clinical management. A position paper from the Consensus Panel on Familial Hypercholesterolaemia of the European Atherosclerosis Society. *Eur Heart J*, 2014; **35**: 2146–57.
11. Lind S, Olsson AG, Eriksson M, Rudling M, Eggertsen G, Angelin B . Autosomal recessive hypercholesterolaemia: normalization of plasma LDL cholesterol by ezetimibe in combination with statin treatment . *J Intern Med*, 2004; **256**: 406–12.
12. Wilund KR, Yi M, Campagna F, *et al.* Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet*, 2002; **11**: 3019–30.
13. Calandra S, Taruqi P, Speedy HE, Dean AF, Bertolini S, Shoulders CC. Mechanisms and genetic determinants regulating sterol

absorption, circulating LDL levels, and sterol elimination: implications for classification and disease risk. *J Lipid Res*, 2011; **52**: 1885–926.
14. Salen G, Steiner RD . Epidemiology, diagnosis, and treatment of cerebrotendinous xanthomatosis (CTX). *J Inherit Metab Dis,* 2017; **40**: 771–81.
15. Nordestgaard BG, Chapman MJ, Ray K, *et al.* Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur Heart J*, 2010; **31**: 2844–53.
16. Waldmann E, Parhofer KG. Lipoprotein apheresis to treat elevated lipoprotein (a). *J Lipid Res*, 2016; **57**: 1751–7.
17. Hegele RA, Ginsberg HN, Chapman MJ, *et al.* The polygenic nature of hypertriglyceridemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol*, 2014; **8**: 655–66.
18. Brahm AJ, Hegele RA . Chylomicronaemia—current diagnosis and future therapies . *Nat Rev Endocrinol*, 2015; **11**: 352–62.
19. Surendran RP, Visser ME, Heemelaar S, *et al.* Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe hypertriglyceridaemia. *J Intern Med*, 2012; **272**: 185–96.
20. Berglund L, Brunzell JD, Goldberg AC, Godberg IJ, Sacks F, Murad MH, Stalenhoef AF, Endocrine Society . Evaluation and treatment of hypertriglyceridemia: an Endocrine Society practice guideline . *J Clin Endocrinol Metab*, 2012; **97**: 2969–89.
21. Nordestgaard BG, Varbo A . Triglycerides and cardiovascular disease . *Lancet,* 2014; **384**: 626–35.
22. Angelin B, Hershon K, Brunzell JD . Bile acid metabolism in hereditary forms of hypertriglyceridemia: evidence for an increased synthesis rate in monogenic familial hypertriglyceridemia . *Proc Natl Acad Sci U S A*, 1987; **84**: 5434–8.
23. Nikkilä EA, Aro A. Family study of serum lipids and lipoproteins in coronary heart-disease. *Lancet*, 1973; **1**: 954–9.
24. Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J Clin Invest*, 1973; **52**: 1544–68.
25. Brahm AJ, Hegele RA . Combined hyperlipidemia: familial but not (usually) monogenic . *Curr Opin Lipidol*, 2016; **27**: 131–40.
26. Castro Cabezas M, de Bruin TW, de Valk HW, Shoulders CC, Jansen H, Erkelens DW. Impaired fatty acid metabolism in familial combined hyperlipidemia. A mechanism associating hepatic apolipoprotein B overproduction and insulin resistance . *J Clin Invest*, 1993; **92**: 160–8.
27. Reynisdottir S, Eriksson M, Angelin B, Arner P . Impaired activation of adipocyte lipolysis in familial combined hyperlipidemia . *J Clin Invest*, 1995; **95**: 2161–9.
28. Veerkamp MJ, de Graaf J, Hendriks JC, Demacker PN, Stalenhoef AF . Nomogram to diagnose familial combined hyperlipidemia on the basis of results of a 5-year follow-up study. *Circulation*, 2001; **109**: 2980–5.
29. Mahley RW, Angelin B. Type III hyperlipoproteinemia: recent insights into the genetic defect of familial dysbetalipoproteinemia . *Adv Intern Med*, 1984; **29**: 385–411.
30. Hopkins PN, Brinton EA, Nanjee MN. Hyperlipoproteinemia type 3: the forgotten phenotype. *Curr Atheroscler Rep,* 2014; **16**: 440.
31. Levy E. Insights from human congenital disorders of intestinal lipid metabolism. *J Lipid Res,* 2015; **56**: 945–62.
32. Jones B, Jones EL, Bonney SA, *et al.* Mutations in a Sar1 GTPase of COPII vesicles are associated with lipid absorption disorders. *Nat Genet*, 2003; **34**: 29–31.
33. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against heart disease. *N Engl J Med*, 2001; **354**: 1264–72.
34. Pelusi S, Baselli G, Pietrelli A, *et al.* Rare pathogenic variants predispose to hepatocellular carcinoma in nonalcoholic fatty liver disease. *Sci Rep*, 2019; **9**: 3682.
35. Dewey FE, Gusarova V, Dunbar RL, *et al.* Genetic and pharmacologic inactivation of ANGPTL3 and cardiovascular disease. *N Engl J Med*, 2017; **377**: 211–21.
36. Xiang AS, Kingwell BA. Rethinking good cholesterol: a clinicians' guide to understanding HDL. *Lancet Diabetes Endocrinol*, 2019; **7**: 575–82.
37. Schaefer EJ, Anthanonta P, Diffender MR, Poliseckib E, Asztalosa BF. Diagnosis and treatment of high density lipoprotein deficiency. *Prog Cardiovasc Dis,* 2016; **59**: 97–106.
38. Madsen CM, Varbo A, Nordestgaard BG. Extreme high high-density lipoprotein cholesterol is paradoxically associated with high mortality in men and women: two prospective cohort studies. *Eur Heart J,* 2017; **38**: 2478–86.
39. Yokoyama S. Unique features of high-density lipoproteins in the Japanese: in population and in genetic factors. *Nutrients,* 2015; **7**: 2359–8.
40. Ljunggren SA, Levels JH, Hovingh K, *et al.* Lipoprotein profiles in human heterozygote carriers of a functional mutation P297S in scavenger receptor class B1. *Biochim Biophys Acta*, 2015; **1851**: 1587v95.

14.3

Other Metabolic Disorders

14.3.1 Hyperinsulinaemic Hypoglycaemia

Khalid Hussain and Sonya Galcheva

Introduction

Hyperinsulinaemic hypoglycaemia (HH) is a condition in which insulin secretion from the pancreatic β-cells is dysregulated so that insulin continues to be secreted even in the presence of hypoglycaemia [1]. HH is a major cause of persistent and recurrent hypoglycaemia in neonates and children. In HH, plasma glucose levels provide the main source of fuel for the brain as the main alternative substrates (ketone bodies) are suppressed by hyperinsulinaemia. It is this biochemical milieu (hyperinsulinaemic hypoketotic hypoglycaemia) which increases the risk of brain injury (leading to neurological sequelae such as developmental delay, cerebral palsy and epilepsy) in children with HH. In order to prevent or reduce the extent of hypoglycaemic brain injury the diagnosis of HH should be made as soon as possible and appropriate treatment started early [2].

HH can be secondary to various factors such as maternal diabetes, intrauterine growth retardation (IUGR) and perinatal asphyxia, in which case it is usually self-limiting, or congenital due to defects in genetic pathways regulating insulin secretion from pancreatic β-cells [3]. The congenital forms of HH can be classified histologically into diffuse and focal forms. The presurgical differentiation of these two subtypes is important for management [4].

Classification of HH Subtypes

Table 14.3.1.1 lists all the different causes of HH.

Transient HH

Transient HH occurs in at-risk neonates, caused by neonatal factors (prematurity, IUGR, large size for gestational age, perinatal stress and birth asphyxia, erythroblastosis fetalis and Rh isoimmunization) or maternal factors (diabetes mellitus (gestational or type 1)), medication use (sulphonylureas) or intravenous glucose infusions during labour) [5]. In some at-risk infants HH can be prolonged and require treatment with diazoxide to maintain euglycaemia until β-cell insulin secretion is normalized within several weeks or months [6].

The physiological and genetic mechanisms underlying transient HH are not known but could relate to failure of postnatal metabolic adaptation, depletion of hepatic glycogen stores, abnormal gluconeogenesis, transient alteration in the β-cell insulin secretion and/or increased insulin sensitivity, or a combination of these [7].

Monogenic Forms of Congenital Hyperinsulinism (CHI)

Monogenic CHI is relatively rare (incidence 1:40 000–50 000 in the general population) but in some populations the incidence can be as high as 1:2500 due to increased consanguinity [5]. These monogenic forms of CHI are caused by defects in key genes involved in regulation of β-cell insulin secretion. Mutations in at least 14 different

Table 14.3.1.1 Causes of hyperinsulinaemic hypoglycaemia

Causes of Transient HH	
Neonatal factors	Prematurity, IUGR, large for gestational age (LGA) infants, perinatal stress/birth asphyxia, Rhesus isoimmunization
Maternal factors	Diabetes mellitus (gestational or type 1) Medication use (sulphonylureas) Intravenous glucose infusion during labour
Monogenic forms of HH	
Mutations in genes encoding channels or ion transporters	K_{ATP} channel: *ABCC8* or *KCNJ11* Lactate transporter: *SLC16A1* Calcium channel: *CACNA1D*
Mutations in genes involved in regulation of insulin secretion	*GLUD1, GCK, HK1, HADH, UCP2, PGM1, PMM2*
Mutations in genes encoding transcription factors	*HNF1A, HNF4A, FOXA2*
Syndromic causes of HH	
Prenatal or postnatal overgrowth syndromes	Beckwith–Wiedemann, Sotos, Simpson–Golabi–Behmel, Perlman syndrome
Chromosomal abnormality syndromes	Trisomy 13, Mosaic Turner syndrome
Postnatal growth failure syndromes	Kabuki syndrome, Costello syndrome
Contiguous gene deletion including ABCC8	Usher syndrome
Syndromes with abnormal calcium homoeostasis	Timothy syndrome
Other syndromes	Congenital central hypoventilation syndrome (Ondine's curse), Donohue syndrome (Insulin resistance syndrome*), Poland syndrome, CHARGE syndrome
Metabolic causes of HH	
Congenital disorders of glycosylation	CDG Ia, CDG Ib, CDG Id
Tyrosinemia type 1	
Miscellaneous causes of HH	
Postprandial HH	Dumping syndrome, insulin receptor gene mutation*, non-insulinoma pancreatogenous hypoglycaemia, insulin autoimmune syndrome
Other causes	Insulinoma, Factitious hypoglycaemia, Drug-induced

IUGR, Intrauterine growth retardation; LGA, Large for gestational age; GDH, Glutamate dehydrogenase; GCK, Glucokinase; HK1, Hexokinase 1; HADH, Hydroxyacyl-CoA dehydrogenase; HNF1α, Hepatocyte nuclear factor 1α; HNF4α, Hepatocyte nuclear factor 4α; Kir6.2, inward rectifier potassium channel 6.2; MCT1, Monocarboxylate transporter 1; SLC16A1, solute carrier family 16, member 1; SUR1, Sulfonylurea receptor 1; UCP2, Mitochondrial uncoupling protein 2; CDG, Congenital disorder of glycosylation; PGM1, Phosphoglucomutase 1; PMM2, Phosphomannomutase 2; CACNA1D Calcium voltage-gated channel subunit alpha1 D; FOXA2, Forkhead Box A2. * In extreme insulin resistance

genes have been described to date leading to non-physiological insulin hypersecretion [4]. Monogenic CHI can broadly be classified into types due to (a) defects in genes regulating channel and ion transporter proteins; (b) gain and loss-of-function of genes encoding enzymes involved in regulating insulin secretion; and (c) transcription factor defects. These subtypes of HH will be described briefly.

Defects in Genes Regulating Channel and Ion Transport Proteins

K_{ATP} channel: *ABCC8* and *KCNJ11* mutations

Insulin secretion from pancreatic β-cell is mainly under the control adenosine triphosphate (ATP)-sensitive potassium (K_{ATP}) channels located in the pancreatic β-cell membrane. These channels play a key role in glucose homeostasis by entraining the electrical excitability of the β-cell membrane to glucose metabolism, and thus to subsequent insulin secretion [3].

The major monogenic causes of CHI are defects in *ABCC8* and *KCNJ11* genes, encoding the SUR1 and Kir6.2 proteins of the K_{ATP} channel, respectively [5, 8]. Homozygous or compound heterozygous inactivating mutations cause the most severe forms of CHI, which are typically unresponsive to diazoxide treatment, often requiring pancreatectomy to control hypoglycaemia [8]. These mutations lead to defects in K_{ATP} channel biogenesis and turnover, defects in channel trafficking from the endoplasmic reticulum and Golgi apparatus to the plasma membrane, and/or alteration of channel regulation by nucleotides or open-state frequency [5].

Dominant inactivating mutations of *ABCC8* or *KCNJ11* usually cause a milder phenotype of CHI with a later age at presentation than the recessive condition. However, medically unresponsive dominant forms have also been reported [9]. A single, paternally inherited K_{ATP} channel mutation with a concomitant postzygotic loss of the maternal 11p15.1-11p15.5 chromosomal region, encoding the K_{ATP} channel and tumour suppressor genes, usually results in a focal adenomatous hyperplasia [10].

Lactate Transporter and Calcium Channel Defects

The *SLC16A1* gene encodes the monocarboxylate transporter 1 (MCT1), which transports monocarboxylates such as pyruvate and lactate into cells. In normal physiological conditions, pancreatic β-cells have low MCT1 expression, so intracellular pyruvate and lactate concentrations are too low to induce insulin secretion. However, dominantly inherited promoter-activating mutations in *SLC16A* lead to increased β-cell MCT1 expression, resulting in pyruvate uptake and metabolism by cells, driving increased ATP

production and thus insulin exocytosis [11]. CHI caused by these mutations is known as exercise-induced hyperinsulinism (EIHI). Affected individuals (older children and adolescents) become hypoglycaemic within 30–45 min after active anaerobic exercise due to pyruvate accumulation [12].

CACNA1D gene encodes the α-subunit of the main L-type calcium voltage-gated channel that has a key role in insulin secretion from the pancreatic β-cell. A gain-of-function mutation in this gene (p.G403D) has been reported to cause diazoxide-responsive HH in a macrosomic neonate [13].

Defects in Key Enzymes Involved in Controlling Insulin Secretion

Activating mutations in the enzymes glutamate dehydrogenase (GDH), CCK, and Hexokinase 1.

The *GLUD1* gene encodes the mitochondrial matrix enzyme glutamate dehydrogenase (GDH). It catalyses reversible oxidation of glutamate to α-ketoglutarate and ammonia. De novo or dominantly inherited missense mutations in *GLUD1* result in gain of enzyme function by reducing its sensitivity to allosteric inhibition by guanosine triphosphate (GTP). This allows activation of insulin secretion by the amino acid leucine. These heterozygous activating mutations cause a protein (leucine)-sensitive form of CHI—hyperinsulinism/hyperammonaemia syndrome (HI/HA), characterized by inappropriate pancreatic insulin secretion and excessive ammonia production. This form of HH is associated with asymptomatic hyperammonaemia and postprandial hypoglycaemia after protein/leucine-rich meals [1, 4, 14].

GCK is a glycolytic enzyme that functions as a glucose sensor in the pancreatic β-cell, regulating the glucose-stimulated insulin secretion [15]. Heterozygous activating mutations in the *GCK* gene lead to an increased glucose affinity of the enzyme, resulting in an increased ATP:ADP ratio in the β-cells, closure of the K_{ATP} channel and inappropriate insulin secretion at low glucose concentrations. Patients often have a dominant family history of hypoglycaemia, with variable severity of symptoms within families [16].

The *HK1* gene encodes the glucose-phosphorylating enzyme hexokinase1 (HK1), which catalyses the first step of the glucose metabolism to glucose-6-phosphate. The absence of HK1 prevents normal β-cells from secreting insulin at the time of hypoglycaemia. Recently a dominantly inherited form of diazoxide-responsive HH has been reported, linked to a gain-of-function mutation in the *HK1* locus [17]. Patients present with mild hypoglycaemia in infancy after prolonged fasting or following a glucose load.

Loss-of-Function Mutations in the Enzymes HADH, UCP2, PGM1, PMM2

HADH is a mitochondrial enzyme that catalyses the NAD^+ dependent dehydrogenation of 3-hydroxyacyl-CoA to 3-ketoacyl-CoA in the β-oxidation of fatty acids [18]. It is highly expressed in pancreatic β-cells. Recessively inherited mutations in the *HADH* gene lead to HH [19]. In HADH deficiency there is a loss of the inhibitory protein-protein interaction between GDH and HADH, which causes overstimulation of GDH, an increase in cellular ATP and upregulated insulin secretion [18]. Unlike in HI/HA syndrome, activation of GDH in HADH deficiency is largely limited to the pancreatic β-cells, so the enzyme deficiency does not lead to hyperammonaemia. The clinical presentation is heterogeneous varying from mild fasting or protein/leucine-induced late-onset hypoglycaemia to severe neonatal hypoglycaemia. Some patients have abnormal acylcarnitine metabolites (raised plasma 3-hydroxybutyrylcarnitine and urinary 3-hydroxyglutarate levels) [3].

The *UCP2* gene encodes a mitochondrial carrier protein largely expressed in pancreatic α- and β-cells. This gene plays a role in protection against oxidative stress during fatty acid metabolism and has been implicated in negative regulation of glucose-stimulated insulin secretion [20]. UCP2 loss-of-function heterozygous mutations have been reported in children with congenital HH, however, another study has shown that variants in *UCP2* may be not a monogenic cause of CHI [21].

Phosphoglucomutase 1 (PGM1) catalyses reversible interconversion of glucose-6-phosphate and glucose-1-phosphate and is essential for glycogen formation, glycogenolysis, and protein glycosylation. PGM1 activity within the pancreatic islets modulates glucose-stimulated insulin secretion. Affected children have both fasting ketotic hypoglycaemia and symptomatic postprandial hypoketotic HH due to a defect in SUR1 glycosylation and exaggerated glucose-stimulated insulin secretion [22].

Phosphomannomutase 2 (PMM2) gene encodes for a key enzyme in glycosylation, and deglycosylation in the pancreatic β-cells has been shown to alter insulin secretion. A recessive mutation in the promoter region of the *PMM2* gene (c.-167G.T) causes a newly described form of CHI, associated with the development of congenital polycystic kidney disease [23].

Transcription Factor Defects

Mutations in *HNF4/1α* and *FOXA2* Genes

HNF4α gene encodes the transcription factor hepatocyte nuclear factor 4α (HNF4α), which belongs to the nuclear hormone receptor superfamily and controls the expression of genes involved in the glucose-stimulated insulin secretion. Heterozygous *HNF4α* mutations have been reported to cause mild transient or persistent congenital HH. Mutations in *HNF4α* cause maturity-onset diabetes of the young 1 (MODY1), and such patients are thus at increased risk of diabetes in later life, and commonly have a family history of diabetes [24]. The *HNF1α* gene encodes a transcription factor, which enhance expression of liver and pancreatic specific genes. Heterozygous, loss-of-function mutations in this gene cause *HNF1α* CHI. Patients present with mild macrosomia at birth and usually develop MODY3 in later life [25]. Two cases of a mutation in *FOXA2* with hypopituitarism, hyperinsulinism and endoderm-derived organ abnormalities have been reported [26, 27]. However, the direct link between HH and the mutation in *FOXA2* has not been proven yet.

Syndromes and Metabolic Causes Associated with HH

Many syndromes have been reported to feature HH either in the neonatal period or at a later age [28]. The most common syndrome associated with HH in childhood is Beckwith–Wiedemann

syndrome (BWS) [29]. Almost half of the patients with BWS develop hypoglycaemia but in most of these children, HH is transient and resolves spontaneously in a few days. However fewer than 5% of the children have HH beyond the neonatal period. In rare cases there is no response to medical therapy and near-total pancreatectomy may be required. Kabuki syndrome is the second most common syndromic form of HH. It is caused by inactivating mutations in two genes: autosomal recessive mutations in *KMT2D* (70–75% of cases) and X-linked mutations in *KDM6A*, which account for 1–9% of cases [30]. Up to 70% of children with Kabuki syndrome have HH and most of these are diazoxide-responsive. Congenital disorders of glycosylation (CDG) are metabolic multisystem diseases caused by defects in the biosynthesis or transfer of lipid-linked oligosaccharides to the nascent protein chain [31].

Postprandial HH (PPHH)

Postprandial hyperinsulinaemic hypoglycaemia (PPHH) occurs within a few hours of food consumption due to inappropriate insulin secretion in response to a meal. It is rare in childhood and the oral glucose tolerance test (OGTT) and mixed-meal tolerance (MMT) test are used for its diagnosis. The most common cause of PPHH in children is '*dumping syndrome*' after gastroesophageal surgery (Nissen's fundoplication). In these patients the secretion of glucagon-like peptide-1 (GLP-1) after ingestion of hyperosmolar carbohydrate-containing solutions is exaggerated, contributing to an inappropriate insulin release and consequent hypoglycaemia within 1 to 4 hours [32]. Another rare cause of PPHH is extreme insulin resistance caused by insulin receptor (*INSR*) gene mutations [33]. PPHH occurs a few hours after meal ingestion and may be related to decreased insulin clearance.

As discussed in more detail in Chapter 14.3.2, 'Autoimmune Hypoglycaemia', PPHH can also be caused by **insulin autoimmune syndrome**, an autoimmune condition characterized by the presence of high levels of insulin autoantibodies in patients who have not previously been exposed to exogenous insulin [34].

Insulinoma

Insulinoma is a rare neuroendocrine tumour, which causes fasting or postprandial HH in older children and adolescents. It can occur either in isolation, or may be a part of the familial dominant syndrome, multiple endocrine neoplasia type 1 (MEN1), so the family history is an important clue to the diagnosis of familial cases [35]. The diagnosis of insulinoma is confirmed by histological examination of the excised pancreatic tissue, showing well-demarcated nodules of endocrine tissue without hyperplasia of the exocrine components.

Munchausen by Proxy Syndrome

Factitious HH is an important consideration in cases with no clear explanation for hypoglycaemic episodes. It is caused by covert exogenous administration of insulin or oral antidiabetic drugs (sulphonylureas). In some cases, this has led to misdiagnosis and consequent pancreatectomy [1].

Histological Subtypes of HH

There are two main histological subtypes of congenital HH, namely diffuse and focal. The diffuse form affects all pancreatic β-cells with variable involvement of the islets. It features enlarged hyperfunctioning cells with very abundant cytoplasm and nuclei that are 3–4 times larger than normal [4, 36]. The most common genetic causes of diffuse HH are recessive and dominant mutations in *ABCC8* and *KCNJ11* genes, as well as all the other genes described earlier. This form of HH accounts for 60–70% of all diagnosed cases.

Focal HH is associated with small (*c.* 2–10 mm diameter) and poorly delineated areas of the pancreas. Lesions consist of large endocrine cells with a large cytoplasm and dispersed abnormal nuclei of irregular and angular shape. The area of abnormal pancreatic development is multilobular and can have satellites in nearby pancreatic tissue that necessitate intraoperative margins analysis to ensure complete excision and avoid recurrence [4, 37].

Focal disease is sporadic and has a distinctive genetic aetiology involving two independent events—inheritance of a paternal mutation in *ABCC8* or *KCNJ11* genes and somatic loss of the maternal 11p allele (11p15.1-15.5) involving the *ABCC8* and *KCNJ11* region within the focal lesion [38]. This paternal uniparental isodisomy unmasks the paternally inherited K_{ATP} channel mutation, which leads to β-cell proliferation due to the altered expression of a number of imprinted genes (maternally expressed tumour suppressor genes *H19* and *CDKN1C* and paternally expressed growth factor IGF2). This accounts for almost 30–40% of all congenital HH cases.

Clinical Presentation

HH most commonly presents with recurrent and persistent hypoglycaemia, which can be fasting or precipitated by the ingestion of protein meals or strenuous physical exercise. The most severe forms of HH typically present in the neonatal period whereas in older children the clinical presentation may be more subtle. Affected newborns may be macrosomic due to fetal hyperinsulinaemia, with hepatomegaly and hypertrophic cardiomyopathy. The clinical manifestations vary between patients, ranging from mild non-specific symptoms (such as poor feeding in the neonatal period) to severe, medically unresponsive symptoms such as apnoea, seizures or even coma [1, 3, 4].

Diagnosis of HH

The diagnosis of HH should be made as soon as possible and appropriate treatment started early to avoid hypoglycaemic brain injury. The diagnosis of HH relies on the combination of clinical findings and biochemical markers consistent with inappropriate insulin secretion at the time of hypoglycaemia.

The medical history should focus on the presence of maternal risk factors, eventful delivery, birth weight, and gestational age, family history of diabetes mellitus or infantile seizures and unexplained infant deaths, parental consanguinity, the time and clinical characteristics of the hypoglycaemic episode, and its relationship to the most recent meal or exercise. On physical examination, there might be evidence of macrosomia or IUGR, syndromic features, hepatomegaly, and cardiomegaly [39].

Laboratory (blood and urine) findings are informative and diagnostic if a 'critical sample' has been taken at the time of hypoglycaemia (usually with a BG level below 3.0 mmol/L) before treatment. When hypoglycaemic samples are not available, a controlled fast should be performed under strict monitoring conditions to induce hypoglycaemia. CHI is the only condition in which hypoglycaemia can persist despite the continuous administration of glucose. The maintenance of normoglycaemia with an intravenous glucose infusion rate higher than 8 mg/kg/min is a powerful diagnostic criterion for HH [39].

Inappropriate insulin secretion at the time of hypoglycaemia, whether spontaneous or provoked, is pathologic and consistent with the presence of HH. Since insulin secretion is pulsatile and since it is rapidly cleared by hepatic metabolism serum insulin levels may be falsely low and there is no correlation between the severity of hypoglycaemia and serum insulin levels. At the same time, the C-peptide concentration, reflecting endogenous insulin secretion, is almost always elevated during hypoglycaemia and its measurement could be more helpful for the diagnosis of dysregulated insulin secretion compared to insulin measurement. In addition, inappropriately low or suppressed serum ketone (predominately 3-β-hydroxybutyrate) and fatty acid concentrations during hypoglycaemia and absence of ketonuria are highly suggestive of HH. Increased serum ammonia concentrations in a patient suspected to have CHI suggest HI/HA syndrome, while increased plasma hydroxybutyrylcarnitine and urinary 3-hydroxyglutarate may indicate *HADH* deficiency.

In some settings, specific stimulation tests can be performed to confirm the diagnosis of HH. A positive glycaemic response of more than 1.5 mmol/L, following a glucagon bolus at the time of hypoglycaemia, is confirmatory of HH [39]. Improvement in hypoglycaemia after a subcutaneous dose of octreotide is also supportive evidence of HH.

Some types of HH may be demonstrated by provocation tests. In patients with HI/HA syndrome, having both fasting and protein-induced hypoglycaemic episodes, a protein/leucine load test will be required for the diagnosis [14]. Patients with EIHI will require a formal exercise test and/or a pyruvate load to demonstrate hypoglycaemia [12]. Patients with a history of PPHH should undergo an MMT test, which is superior to OGTT.

Currently genetic testing is a well-established method to define the best approach to treatment of CHI. Finding two pathogenic variants (homozygous or compound heterozygous mutations in *ABCC8* and *KCNJ11*) is usually diagnostic of medically unresponsive diffuse disease. Patients with a single recessive pathogenic variant on the paternal allele of *ABCC8* and *KCNJ11* genes or without identified mutations potentially have curable focal disease and require further imaging to localize the lesion(s) precisely [3, 4].

Management of Congenital HH

Management of HH can be very challenging. Patients should ideally be managed in centres of excellence with the necessary experience and expertise, as they will require multidisciplinary care. Patients with HH require frequent blood glucose monitoring and often need a central venous catheter to deliver concentrated glucose infusions. The key aims in meticulous management are to maintain normoglycaemia (blood glucose levels >3.5mmol/L), to encourage oral feeding, to assess the response to medical therapy, and to decide if urgent genetic testing is indicated, undertaking pancreatic imaging with 18F-DOPA PET/CT if required and surgical intervention again if required [1, 39]. The management cascade and medications used are shown in Figure 14.3.1.1 and Table 14.3.1.2.

Surgery in HH

The type of surgical procedure will depend on the underlying histology of the HH. Before the patient undergoes any type of surgery it will be important to know the underlying histological basis of the HH. Patients with autosomal recessive mutations in *ABCC8/KCNJ11* will have histologically diffuse disease. However, those with a paternal ABCC8/KCNJ11 mutation are likely to have focal disease and will need imaging with 18F-DOPA PET/CT to localize the focal lesion(s).

Focal Lesions

Focal forms of CHI can be completely cured with partial pancreatectomy or limited lesionectomy with a complete excision of the lesion or enucleation of the tumour confirmed by intraoperative biopsies. In large focal lesions in the pancreatic head or neck, especially when they are close to the common bile or pancreatic duct, open laparotomy with a resection and Roux-en-Y pancreaticojejunostomy is performed. Fluorine-18-L-dihydroxyphenylalanine positron emission tomography (18F-DOPA PET/CT) is the current gold standard technique for precise preoperative localization of focal lesions [40]. As a result, the extent of the pancreatic surgical resection can be guided and children with focal HH can be cured by limited surgery. This technique is based on the ability of the β-cells to uptake L-dihydroxyphenylalanine (L-DOPA) and to convert it to dopamine by the enzyme DOPA decarboxylase. The positron emitting tracer 18F-DOPA is an analogue of DOPA and the former can be used to detect lesions of hyperfunctional β-cells visualized as bright areas over a darker background [40].

Diffuse Forms

Children with diffuse forms of medically unresponsive HH requires near-total pancreatectomy (removal of 95–98% of the pancreas) leaving a small amount of pancreatic tissue around the duodenum and the common bile duct [41]. The immediate post-surgical outcome is unpredictable—from persistent hypoglycaemia to insulin-requiring diabetes and exocrine pancreatic insufficiency. Postoperatively some children still have episodes of hypoglycaemia and further postoperative therapy with medications and/or frequent feedings is required or re-operation may be performed [42]. Near-total pancreatomy is also associated with a high incidence of

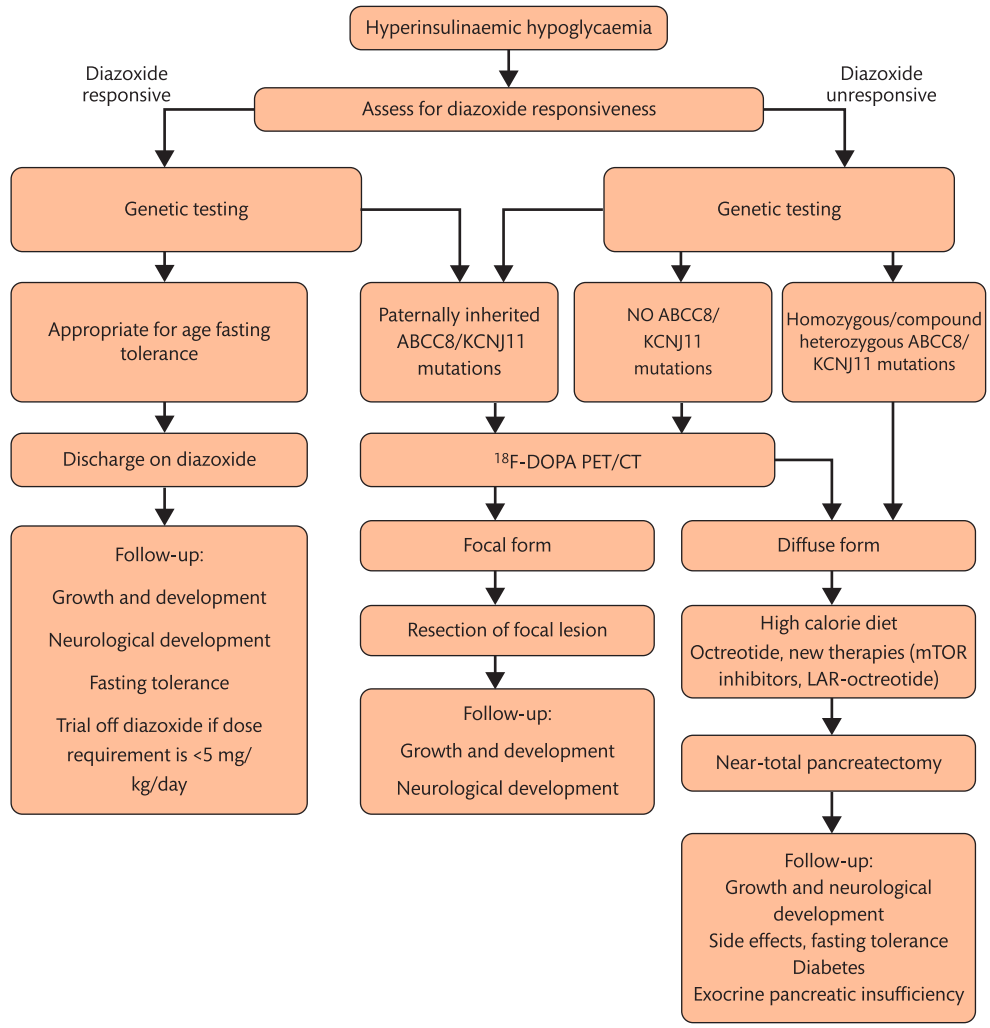


Figure 14.3.1.1 Diagnostic and management algorithm in patients with hyperinsulinaemic hypoglycaemia.

Table 14.3.1.2 Drugs used in the management of hyperinsulinaemic hypoglycaemia

Drug	Dose	Mechanism of action	Side effects
Older conventional medications			
Diazoxide	5–20 mg/kg/day in 3 oral doses	K_{ATP} channel agonist	**Common:** fluid retention, hypertrichosis, loss of appetite. **Rare:** cardiac failure, hyperuricemia, leukopenia, eosinophilia, paradoxical hypoglycaemia, pulmonary hypertension
Chlorothiazide	7–10 mg/kg/day in 2 oral doses	Synergistic effects with diazoxide	Hyponatraemia, hypokalaemia
Nifedipide	0.25–2.5 mg/kg/day in 2–3 oral doses	Calcium channel blocker	Hypotension
Glucagon	0.02 mg/kg s.c/i.m. bolus or 1–10 µg/ kg/h of s.c/i.v. infusion	Induces glycogenolysis/ gluconeogenesis	Nausea, vomiting, skin rash, rebound hypoglycaemia (at higher doses)
Octreotide	5–35 µg/kg/day, divided in 3–4 s.c. doses or continuous s.c. infusion	Somatostatin analogue; reduces insulin biosynthesis and secretion	**Acute:** nausea, abdominal distension, hepatitis, long QT, tachyphylaxis, necrotizing enterocolitis. **Long term:** decreased intestinal motility, bile sludge, cholelithiasis, pituitary suppression
Newer medications			
Lanreotide/ LAR-octreotide	Total dose of octreotide as above *or* 15–60 mg given 4 weekly as i.m./deep s.c. injection	Long-acting somatostatin analogues	Similar to daily multiple injection octreotide; long-term follow-up not reported
Rapamycin (Sirolimus)	Initially 0.5 mg/m^2/day in 1–2 oral doses, then titrated to blood levels of 5–15 ng/ml	mTOR inhibitor; inhibits β-cell proliferation and insulin secretion	Mucositis, immunosuppression, hyperlipidaemia, transient elevation of liver enzymes, thrombocytosis, impaired immune response to Bacillus Calmette–Guérin (BCG) vaccine

diabetes mellitus and exocrine pancreatic insufficiency as the risk for these complications increases with age.

Conclusion

Infants and children with HH should be diagnosed as soon as possible and treatment commenced early. HH remains a challenging condition for paediatric endocrinologists and it is recommended that children with congenital HH should be referred to centres of excellence if possible. Although there have been significant advances in understanding the genetic basis of HH, in imaging and in medical therapy there are still a large number of (diazoxide-responsive) children in whom no genetic aetiology is identified. The diffuse form of the disease remains a challenge as some children will not respond to any of the current treatments and therefore new treatment options are required in these patients. New and readily available imaging techniques (as ^{18}F-DOPA is available in only scattered centres) are required for children suspected of having focal HH.

REFERENCES

1. Shah P, Rahman SA, Demirbilek H, *et al*. Hyperinsulinaemic hypoglycaemia in children and adults. *Lancet Diabetes Endocrinol*, 2017; **5**: 729–42.
2. Helleskov A, Melikyan M, Globa E, *et al*. Both low blood glucose and insufficient treatment confer risk of neurodevelopmental impairment in congenital hyperinsulinism: a multinational cohort study. *Front Endocrinol*, 2017; **8**: 156.
3. Arya VB, Mohammed Z, Blankenstein O, De Lonlay P, Hussain K . Hyperinsulinaemic hypoglycaemia. *Horm Metab Res*, 2014; **46**: 157–70.
4. Demirbilek H, Hussain K . Congenital hyperinsulinism: diagnosis and treatment update . *J Clin Res Pediatr Endocrinol*, 2017; **9**(Suppl 2): 69–87.
5. Senniappan S, Arya VB, Hussain K . The molecular mechanisms, diagnosis and management of congenital hyperinsulinism. *Indian J Endocrinol Metab,* 2013; **17**: 19–30.
6. Fafoula O, Alkhayyat H, Hussain K . Prolonged hyperinsulinaemic hypoglycaemia in newborns with intrauterine growth retardation . *Arch Dis Child Fetal Neonatal Ed*, 2006; **91**: F467.
7. Mizumoto H, Uchio H, Yamashita S, Hata D . Transient neonatal hyperinsulinism with adaptation disorders: a report of three cases. *J Pediatr Endocrinol Metab*, 2015; **28**: 337–40.
8. Flanagan SE, Kapoor RR, Hussain K . Genetics of congenital hyperinsulinemic hypoglycemia . *Semin Pediatr Surg,* 2011; **20**: 13–17.
9. Nessa A, Aziz QH, Thomas AM, *et al*. Molecular mechanisms of congenital hyperinsulinism due to autosomal dominant mutations in ABCC8. *Hum Mol Genet*, 2015; **24**: 5142–53.
10. Arya VB, Guemes M, Nessa A, *et al*. Clinical and histological heterogeneity of congenital hyperinsulinism due to paternally inherited heterozygous ABCC8/KCNJ11 mutations. *Eur J Endocrinol*, 2014; **171**: 685–95.
11. Pullen TJ, Sylow L, Sun G, *et al*. Overexpression of monocarboxylate transporter-1 (SLC16A1) in mouse pancreatic beta-cells leads to relative hyperinsulinism during exercise. *Diabetes*, 2012; **61**: 1719–25.
12. Meissner T, Friedmann B, Okun JG, *et al*. Massive insulin secretion in response to anaerobic exercise in exercise induced hyperinsulinism. *Horm Metab Res*, 2005; **37**: 690–4.
13. Flanagan SE, Vairo F, Johnson MB, *et al*. A CACNA1D mutation in a patient with persistent hyperinsulinaemic hypoglycaemia, heart defects, and severe hypotonia. *Pediatr Diabetes*, 2017; **18**: 320–3.
14. Kapoor RR, Flanagan SE, Fulton P, *et al*. Hyperinsulinism-hyperammonaemia syndrome: novel mutations in the GLUD1 gene and genotype-phenotype correlations. *Eur J Endocrinol*, 2009; **161**: 731–35.
15. Matschinsky FM. Regulation of pancreatic beta-cell glucokinase: from basics to therapeutics . *Diabetes*, 2002; **51** Suppl 3: S394–404.
16. Sayed S, Langdon DR, Odili S, *et al*. Extremes of clinical and enzymatic phenotypes in children with hyperinsulinism caused by glucokinase activating mutations. *Diabetes*, 2009; **58**: 1419–27.
17. Pinney SE, Ganapathy K, Bradfield J, *et al*. Dominant form of congenital hyperinsulinism maps to HK1 region on 10q. *Horm Res Paediatr*, 2013; **80**: 18–27.
18. Li C, Chen P, Palladino A, *et al*. Mechanism of hyperinsulinism in short chain 3 hydroxyacyl CoA dehydrogenase deficiency involves activation of glutamate dehydrogenase. *J Biol Chem*, 2010; **285**: 31806–18.
19. Clayton PT, Eaton S, Aynsley Green A, *et al*. Hyperinsulinism in short chain L 3 hydroxyacyl CoA dehydrogenase deficiency reveals the importance of beta oxidation in insulin secretion. *J Clin Invest*, 2001; **108**: 457–65.
20. Chan CB, De Leo D, Joseph JW, *et al*. Increased uncoupling protein 2 levels in beta cells are associated with impaired glucose stimulated insulin secretion: mechanism of action. *Diabetes*, 2001; **50**: 1302–10.
21. Laver TW, Weedon MN, Caswell R, *et al*. Analysis of large-scale sequencing cohorts does not support the role of variants in UCP2 as a cause of hyperinsulinaemic hypoglycaemia. *Hum Mutat*, 2017; **38**: 1442–4.
22. Ackermann AM, Li C, Freeze HH, *et al*. Hypoglycemia due to lower threshold of glucose-stimulated insulin secretion in phosphoglucomutase 1 deficiency. Platform Presentation at: Annual Meeting of the Pediatric Academic Societies, 25–28 April 2015, San Diego, CA.
23. Cabezas OR, Flanagan SE, Stanescu H, *et al*. Polycystic kidney disease with hyperinsulinemic hypoglycemia caused by a promoter mutation in phosphomannomutase 2. *J Am Soc Nephrol*, 2017; **28**: 2529–39.
24. Kapoor RR, Locke J, Colclough K, *et al*. Persistent hyperinsulinemic hypoglycemia and maturity onset diabetes of the young due to heterozygous HNF4A mutations. *Diabetes*, 2008; **57**: 1659–63.
25. Dusatkova P, Pruhova S, Sumnik Z, *et al*. HNF1A mutation presenting with fetal macrosomia and hypoglycemia in childhood prior to onset of overt diabetes. *J Pediatr Endocrinol Metab*, 2011; **24**: 187–9.
26. Giri D, Vignola ML, Gualtieri A, *et al*. Novel FOXA2 mutation causes hyperinsulinism, hypopituitarism with craniofacial and endoderm-derived organ abnormalities. *Hum Mol Genet*, 2017; **26**: 4315–26.
27. Vajravelu ME, Chai J, Krock B, *et al*. Congenital hyperinsulinism and hypopituitarism attributable to a mutation in FOXA2. *J Clin Endocrinol Metab*, 2018; **103**: 1042–7.
28. Kapoor RR, James C, Hussain K. Hyperinsulinism in developmental syndromes. *Endocr Dev*, 2009; **14**: 95–113.

29. Kalish JM, Boodhansingh KE, Bhatti TR, *et al*. Congenital hyperinsulinism in children with paternal 11p uniparental isodisomy and Beckwith–Wiedemann syndrome. *J Med Genet*, 2016; **53**: 53–61.
30. Bogershausen N, Gatinois V, Riehmer V, *et al*. Mutation update for Kabuki syndrome genes KMT2D and KDM6A and further delineation of X-linked Kabuki syndrome subtype 2. *Hum Mutat,* 2016; **37**: 847–64.
31. Jaeken J. Komrower Lecture . Congenital disorders of glycosylation (CDG): it's all in it! *J Inherit Metab Dis*, 2003; **26**: 99–118.
32. Palladino AA, Sayed S, Levitt Katz LE, *et al*. Increased glucagon-like peptide-1 secretion and postprandial hypoglycemia in children after Nissen fundoplication. *J Clin Endocrinol Metab*, 2009; **94**: 39–44.
33. Højlund K, Hansen T, Lajer M, *et al*. A novel syndrome of autosomal-dominant hyperinsulinemic hypoglycemia linked to a mutation in the human insulin receptor gene. *Diabetes*, 2004; **53**: 1592–8.
34. Alves C, Constança J, De León DD, *et al*. A novel atypical presentation of insulin autoimmune syndrome (Hirata's disease) in a child. *J Pediatr Endocrinol Metab*, 2013; **26**: 1163–6.
35. Kwon EB, Jeong HR, Shim YS, *et al*. Multiple endocrine neoplasia type 1 presenting as hypoglycemia due to insulinoma. *J Korean Med Sci*, 2016; **31**: 1003–6.
36. Han B, Newbould M, Batra G, *et al*. Enhanced islet cell nucleomegaly defines diffuse congenital hyperinsulinism in infancy but not other forms of the disease. *Am J Clin Pathol*, 2016; **145**: 757–68.
37. Rahier J, Guiot Y, Sempoux C. Morphologic analysis of focal and diffuse forms of congenital hyperinsulinism . *Semin Pediatr Surg*, 2011; **20**: 3–12.
38. Verkarre V, Fournet JC, de Lonlay P, *et al*. Paternal mutation of the sulfonylurea receptor (SUR1) gene and maternal loss of 11p15 imprinted genes lead to persistent hyperinsulinism in focal adenomatous hyperplasia. *J Clin Invest*, 1998; **102**: 1286–291.
39. Hussain K. Investigations for neonatal hypoglycaemia . *Clin Biochem*, 2011; **44**: 465–66.
40. Blomberg BA, Moghbel MC, Saboury B, *et al*. The value of radiologic interventions and 18F-DOPA PET in diagnosing and localizing focal congenital hyperinsulinism: systematic review and meta-analysis. *Mol Imaging Biol*, 2012; **15**: 97–105.
41. Pierro A, Nah SA . Surgical management of congenital hyperinsulinism of infancy . *Semin Pediatr Surg*, 2011; **20**: 50–3.
42. Beltrand J, Caquard M, Arnoux JB, *et al*. Glucose metabolism in 105 children and adolescents after pancreatectomy for congenital hyperinsulinism. *Diabetes Care*, 2012; **35**: 198–203.

14.3.2 Autoimmune Hypoglycaemia

Phillip Gorden and Noemi Malandrino

Introduction

Two rare causes of hypoglycaemia have an autoimmune aetiology. The first, insulin autoimmune syndrome, also known as Hirata disease, is characterized by hyperinsulinaemic hypoglycaemia and a high titre of antibodies against endogenous insulin in the absence of both pancreatic islet disease and prior exposure to exogenous insulin [1]. The second condition, type B insulin resistance, presents with extreme insulin resistance, and hyperglycaemia and/or hypoglycaemia, due to autoantibodies against the insulin receptor [2].

Both disorders are mediated by antibodies to self-antigens. This is distinct from antibodies against exogenous recombinant human insulin, which are commonly detected in patients with insulin-treated diabetes, but which, except in rare cases of labile glycaemic control, do not have clinical significance. Differentiating non-self-antigen-induced antibodies from autoantibodies is difficult in patients treated with exogenous insulin, making the biochemical diagnosis of insulin autoimmune syndrome problematic. The risk of overdiagnosis of pathogenic antibodies is illustrated by the low prevalence of insulin resistance in patients who received animal-derived insulins and developed high titres of insulin-binding antibodies [3, 4].

This chapter reviews the clinical characteristics, pathophysiology, diagnosis, and management of autoimmune hypoglycaemia syndromes. Additionally, the complexity of diagnosis of insulin autoimmune syndrome in patients treated with exogenous insulin and showing positive insulin antibodies is discussed, together with the challenges in discrimination of those antibodies with clinically significant effects on insulin kinetics.

Insulin Autoimmune Syndrome

Clinical Characteristics in Asian Populations

The insulin autoimmune syndrome was first described in Japan by Hirata in 1970 [5]. Most of the cases have been reported in Asia, including at least 380 cases in Japan and more than 70 cases in China [5, 6]. In Japan, the insulin autoimmune syndrome is the third leading cause of spontaneous hypoglycaemia, after insulinoma and extrapancreatic neoplasms [1]. The peak age of onset is 60–69 years [1]. A similar age of onset has been described in male Chinese patients, while the disease most commonly occurs at 30–39 years in Chinese females [6]. The reasons for this sex difference in Chinese patients are not fully understood. Males and females are equally affected, except for the age groups 20–29 and 30–39 in Japanese and Chinese patients, where a higher percentage of females (75–85%) is reported [6].

Asian patients show a high prevalence of the Cw4/B62/DR4 allelic combination [1]. Most of the DR4-positive patients are DRB1*0406 positive, although some patients are DRB1*0403 or DRB1*0407

positive. The DRB1*0406 allele shows the highest risk for susceptibility to the insulin autoimmune syndrome and seems to mediate the presentation of insulin peptides to T cells [5].

In about 42% of Japanese patients, prior exposure to medications containing sulfhydryl groups has been reported, including methimazole, alpha-mercaptopropionyl glycine, glutathione, gold thioglucose, captopril, penicillamine, and alpha lipoic acid [5]. Sulfhydryl compounds may cleave the disulphide bonds linking the A and B-chains of insulin, allowing the interaction between antigen-presenting cells and insulin-derived peptides and the activation of self-insulin T-helper cells. Non-sulfhydryl compounds have also been associated with insulin autoimmune syndrome, including aceglatone, steroids, interferon-alpha and non-steroidal anti-inflammatory drugs [5].

Clinical Characteristics in Non-Asian Populations

A smaller number of patients with the insulin autoimmune syndrome has been reported in non-Asian countries [7]. The disease in these populations develops more frequently in individuals over 40 years of age, with similar prevalence in males and females [7]. While the DRB1*0406 allele is 10–20 times more common in Asian patients, in non-Asian patients the insulin autoimmune syndrome is more frequently associated with the DRB1*0403 allele. The different prevalence of DRB1*04 alleles in these populations may account for the higher frequency of insulin autoimmune syndrome in Asia, given the higher risk associated with DRB1*0406 [8].

In non-Asian patients, frequent association of insulin autoimmune syndrome is noted with rheumatologic diseases, including systemic lupus erythematosus, rheumatoid arthritis and systemic sclerosis, or with haematologic diseases, including benign monoclonal gammopathy and multiple myeloma. Patients may have positive antinuclear antibodies, antidouble-stranded DNA and rheumatoid factor, and an abnormal blood count, urine and serum protein electrophoresis and bone marrow examination [7].

As for Asian patients, exposure to various medications before development of the disease has been reported, including captopril, penicillamine, pyritinol, carbimazole, and methimazole, imipenem, propylthiouracil, hydralazine, procainamide, isoniazid, penicillin G, and alpha lipoic acid. However, the relationship between sulfhydryl-containing medications and the occurrence of the syndrome is not as clear as in Japanese patients [7].

Pathophysiology and Mechanism of Hypoglycaemia

Hypoglycaemic episodes in the insulin autoimmune syndrome are caused by the production of antibodies against endogenous insulin in patients with no exposure to exogenous insulin. Hypoglycaemia most often occurs in the postprandial state, although fasting hypoglycaemia and hypoglycaemia exacerbated by exercise have been described. Paradoxically, hyperglycaemia may occur immediately following a meal or oral glucose challenge, and may explain the elevated glycated haemoglobin seen in some patients, despite hypoglycaemic episodes [7].

Both hyper and hypoglycaemia are thought to be related to the binding of endogenous insulin by antibodies, resulting in physiologically inappropriate levels of bioavailable insulin. After a meal or oral glucose load, insulin is secreted in response to rising blood glucose. In the insulin autoimmune syndrome, the binding of secreted insulin to the antibodies reduces insulin availability to receptors in the liver and peripheral tissues, which may result in hyperglycaemia and further insulin release. As the blood glucose concentration decreases and insulin secretion declines, the antibody-bound insulin is released, resulting in inappropriately high free insulin concentrations and hypoglycaemia (see Figure 14.3.2.1) [7].

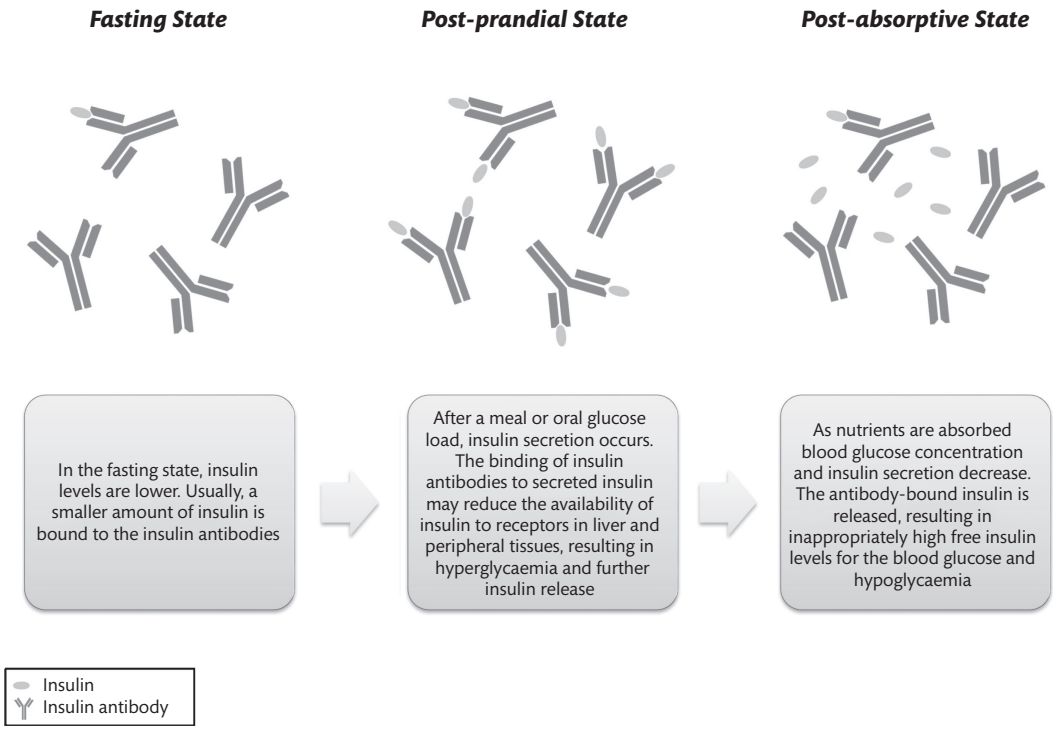


Figure 14.3.2.1 Mechanism of hypoglycaemia in the insulin autoimmune syndrome.

In the insulin autoimmune syndrome, insulin levels may be markedly elevated, due to interference of endogenous autoantibodies with the insulin immunoassay. In this situation the insulin assay is unreliable. Alternatively, the insulin values may be low depending on the specific immunoassay technique. Thus, in these situations the insulin immunoassay may be misleading and the most accurate technique to determine the presence of an insulin-binding protein, such as an antibody is polyethylene glycol or gel filtration chromatography [9]. Insulin autoantibodies may sometimes interfere with C-peptide and proinsulin immunoassays, leading to erroneously elevated concentrations of these hormones too [7].

Most insulin antibodies are immunoglobulin G (IgG) with various ratios of kappa to lambda light chains. However, other immunoglobulin classes have also been reported. Based on affinity curves for binding to human insulin and the presence of solitary light chains, insulin antibodies are classified as monoclonal or polyclonal. The clonality has been related to the type of amino acid at position 74 of the HLA-DR4 beta 1 chain. Most of the Asian patients develop polyclonal IgG insulin antibodies with two classes of insulin-binding sites: high affinity/low capacity and low affinity/high capacity. In contrast, non-Asian patients show a higher prevalence of monoclonal low affinity/high capacity antibodies [7]. High affinity/low capacity antibodies seem to be frequently associated with hyperglycaemia and insulin resistance, whereas low affinity/high capacity antibodies are more likely to induce hypoglycaemia due to the dissociation of insulin from the insulin-antibody complex [10].

Insulin antibodies can be measured by assessing binding by serum of ^{125}I-labelled insulin, by enzyme-linked immunosorbent assay, or indirectly by determining recovery of immunoreactive insulin after polyethylene glycol precipitation of plasma [7]. Radioligand binding assays are also used to assess the clonality and affinity of insulin antibodies. After incubation of a fixed concentration of serum insulin antibody with a fixed amount of ^{125}I-insulin and increasing concentrations of unlabelled insulin, Scatchard analysis is performed by plotting the bound/free versus bound insulin concentration, and binding capacities and affinities of antibodies are determined. While Scatchard analysis has been frequently used, its interpretation is problematic in the context of polyclonal antibodies [4].

Complexity in the Diagnosis of Insulin Autoimmune Syndrome in Patients Treated with Exogenous Insulin and with Positive Insulin Antibodies

Patients with diabetes treated with recombinant human insulin or insulin analogues commonly develop antibodies against exogenous insulin, which are polyclonal IgGs with a wide range of affinities and capacities [4, 11]. In patients with type 1 diabetes autoantibodies may be present prior to the development of hyperglycaemia. These autoantibodies are markers of the immune process and do not have biologic significance. Episodes of insulin resistance and/or severe hypoglycaemia have been described in patients with positive insulin antibodies and treated with exogenous animal-derived insulin [3, 4]. With the introduction of recombinant human insulin, the incidence and severity of immunological response to insulin treatment has significantly decreased. While the low titres of antibodies induced by recombinant human insulin do not usually affect glycaemic control or interfere with therapy, insulin-treated patients with high levels of insulin antibodies have rarely developed hyperglycaemic and/or hypoglycaemic episodes [4, 11, 12]. These patients most commonly have type 1 diabetes, and the poor glycaemic control is attributed to circulating antibodies against exogenous insulin [11, 12]. In a few cases, coexistence of insulin autoimmune syndrome has been proposed as the cause of hypoglycaemia [11, 13]. An additional uncommon condition proposed as a possible cause of insulin resistance and/or hypoglycaemia in type 1 diabetes is subcutaneous insulin resistance, defined according to three criteria: (1) resistance to the action of subcutaneous insulin, with normal sensitivity to intravenous insulin; (2) lack of increased circulating insulin after the administration of subcutaneous insulin; (3) increased degradation of insulin in the subcutaneous tissue. The existence of this condition is highly questionable and difficult to document [14].

Patients with type 1 diabetes treated with exogenous insulin may present with unstable glycaemic control, hypoglycaemia, and insulin antibodies. However, the role of the antibodies in causing hypoglycaemia is not always clear. Thus, several factors may affect blood glucose stability in these patients. The type and dose of insulin, and the site and technique of injection can affect the rate of insulin absorption. Slower absorption of insulin may be associated with longer time to peak and longer duration of the effect, resulting in initial hyperglycaemia, followed by hypoglycaemia a few hours after a meal. Alterations at the injection site (e.g. lipohypertrophy), and factors influencing the subcutaneous blood flow, including exercise, massage at the injection site and temperature, may also affect insulin absorption. Finally, psychosocial factors and poor compliance to insulin treatment may be associated with decreased glycaemic control [15–17]. Identification and correction of these factors is extremely important and often results in significantly improved glycaemic control. Therefore, administration of antibody-depleting therapies in individuals treated with exogenous insulin and with positive insulin antibodies should be considered with extreme caution and after careful consideration of alternative factors associated with poor glycaemic control. Furthermore, in those cases where the coexistence of insulin autoimmune syndrome is suspected, it is difficult to distinguish antibodies induced by exogenous insulin from autoantibodies. As discussed earlier, Scatchard analysis may be used to study the affinity and binding capacity of antibodies, to determine their possible role in the development of insulin resistance and/or hypoglycaemic events [4, 11]. In addition, a gel filtration chromatography method coupled to *ex vivo* insulin binding/exchange has been developed to improve identification of insulin autoantibodies that are likely to alter insulin pharmacokinetics and pharmacodynamics (see Figure 14.3.2.2) [18].

Type B Insulin Resistance

Clinical Characteristics

Type B insulin resistance is a rare disease, although the exact prevalence is unknown. The largest cohort includes 48 patients followed at the National Institutes of Health (NIH). The typical patient is a middle-aged African-American woman (~85% of patients). In the NIH cohort, the average age of onset is 41 years in women, and 57 years in men, although the disease can occur in children and the elderly [7, 19].

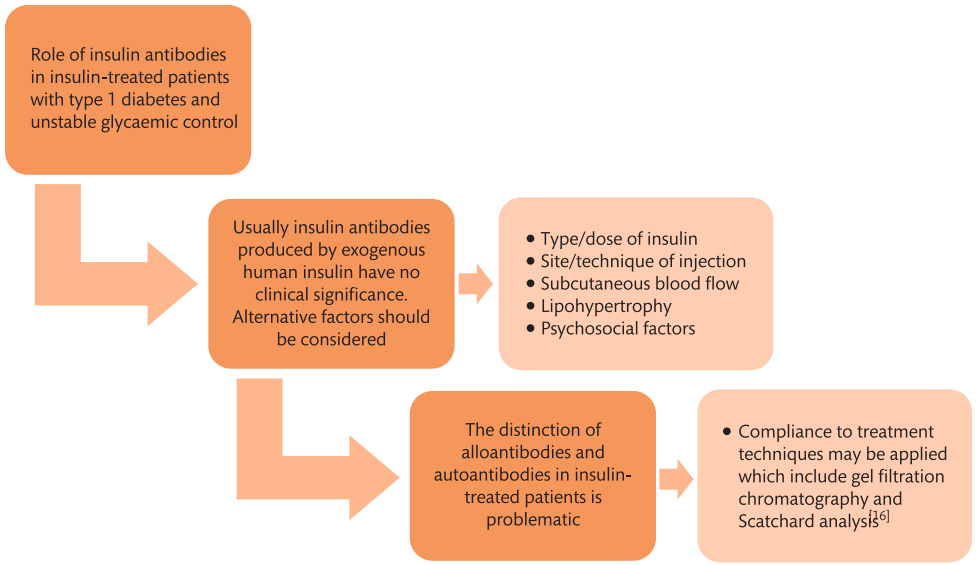


Figure 14.3.2.2 Factors affecting the stability of glycaemic control in type 1 diabetes.

Most patients present with extreme insulin resistance, relatively acute onset of moderate to severe hyperglycaemia and profound glycosuria, polyuria, and weight loss. Endogenous insulin levels are elevated prior to administration of exogenous insulin. Some patients may develop a hyperglycaemic phase followed by hypoglycaemia, or hypoglycaemia alone. The hypoglycaemia may develop in the fasting or postprandial state. Several patients presenting with hypoglycaemia show higher body mass index (BMI) than those with hyperglycaemia. In a subgroup of the NIH cohort, 62% of the hyperglycaemic patients had a BMI <25 kg/m² and 19% had a BMI >30 kg/m², whereas 33% of the hypoglycaemic patients had a BMI <25 kg/m² and 67% had a BMI >30 kg/m² [19].

In type B insulin resistance, acanthosis nigricans is a striking clinical feature. While it most commonly affects the axillae, groin, and neck, patients with type B insulin resistance may show more diffuse anatomical extent involving the periocular, perioral, and labial regions. Some patients also develop acanthosis on the trunk, buttocks, lips, and vulva [19]. Polycystic ovarian enlargement and hyperandrogenism are commonly seen in premenopausal women. Reduction in testosterone levels occurs when the autoantibody syndrome remits [7, 19].

Low fasting triglycerides and high plasma concentrations of the fat cell-derived protein adiponectin are additional biochemical characteristics associated with this disease. These may help distinguish type B insulin resistance from other syndromes of insulin resistance, which are typically associated with high fasting triglycerides and hypoadiponectinemia. Of note, low triglycerides and high adiponectin levels are also observed in patients with insulin-receptor mutations, therefore these metabolic characteristics appear to be a unique feature of insulin-receptor dysfunction as opposed to non-receptor-mediated insulin resistance [20]. In addition, in type B insulin resistance, the reduction of insulin-receptor autoantibodies and the resolution of insulin resistance result in normalization of adiponectin levels [7].

Type B insulin resistance is frequently associated with systemic lupus erythematous and/or other autoimmune diseases [19]. The metabolic derangements of type B insulin resistance usually appear after the onset of the underlying systemic autoimmune disease. Type B insulin resistance may also develop as a paraneoplastic manifestation of conditions such as multiple myeloma and Hodgkin's disease [19]. Proteinuria has been reported in ~50% of the patients. Despite uncontrolled hyperglycaemia, proteinuria appears to be related to different forms of lupus nephritis, such as mesangial and proliferative lupus nephritis, rather than diabetic nephropathy [7, 19]. Haematologic abnormalities, including leucopoenia, thrombocytopenia, and anaemia, may also be related to the underlying autoimmune disease [19].

Pathophysiology and Mechanism of Hypoglycaemia

Type B insulin resistance is caused by circulating autoantibodies against the insulin receptor which are primarily polyclonal IgGs. Patients may present with either insulin resistance or hypoglycaemia. These dual effects are related to the partial agonist activity of the receptor autoantibodies. These autoantibodies can in fact mimic insulin action in cell cultures, by causing autophosphorylation of the insulin receptor and stimulation of 2-deoxyglucose uptake or glucose oxidation, which in turn result in hypoglycaemia. On the other hand, extended exposure of the cell cultures to the autoantibodies induces increased insulin-receptor degradation, decreased number of cell-surface receptors and insulin resistance. Similarly, when administered acutely to fasting rats, the autoantibodies induce hypoglycaemia in a dose-dependent manner, whereas chronic administration of the receptor autoantibodies to fed rats results in persistent hyperglycaemia and insulin resistance. The partial agonist activity of the receptor autoantibodies is consistent with the observation that low antibody titres are associated with hypoglycaemia and high antibody titres are associated with hyperglycaemia (see Figure 14.3.2.3). A fall in the autoantibody titre is observed in some patients shifting from the insulin resistant to the hypoglycaemic state [7, 19].

Receptor autoantibodies appear to be specific for the insulin receptor in its native conformation, as binding does not occur with denatured receptors. Additionally, the autoantibodies bind

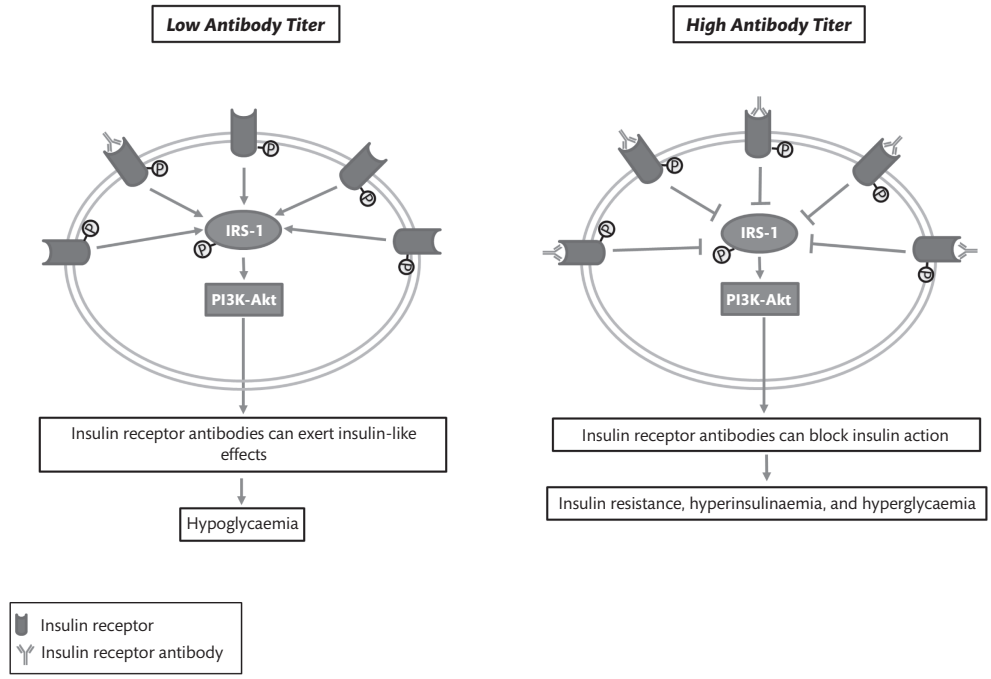


Figure 14.3.2.3 Partial agonist activity of the insulin-receptor antibodies in type B insulin resistance.

to a region of the insulin receptor in the carboxy terminal half of the alpha subunit, which is different from the domain containing the insulin-binding site in the amino terminus of the alpha subunit. Therefore, it has been suggested that, without binding to the same site as insulin, the receptor antibodies can inhibit insulin binding to the receptor through alternative mechanisms including allosteric inhibition, steric hindrance, or misalignment of the insulin-binding site. Furthermore, the binding of the autoantibodies to the insulin receptor may interfere with the receptor-mediated insulin degradation, leading to the hyperinsulinaemia [19].

To date, assays for the measurement of the autoantibodies to the insulin receptor are available only in research laboratories. The immunoprecipitation assay, which is the most sensitive and specific method, measures the ability of patients' sera to immunoprecipitate recombinant human insulin receptors [21]. Additional assays include those based on inhibition of ^{125}I-labelled insulin binding or the stimulation of glucose oxidation/lipogenesis in cells incubated with patients' sera [7].

In patients affected by type B insulin resistance, circulating insulin antibodies may be detected in addition to the insulin-receptor antibodies. Although a few cases have been reported in patients with no history of prior exposure to exogenous insulin, insulin antibody production is most commonly related to the administration of insulin [22–24].

Clinical Comparison Between Insulin Autoimmune Syndrome and Type B Insulin Resistance

The insulin autoimmune syndrome and type B insulin resistance can both present with hypoglycaemia during fasting and/or after a meal. Clinical and biochemical characteristics may help differentiate between the two syndromes (see Table 14.3.2.1). The insulin autoimmune syndrome occurs frequently between ages 40 and 80 years, equally affecting men and women. Additionally, most non-Asian patients are Caucasian. Type B insulin resistance typically occurs in African-American women between ages 40 and 50 years, though it is also seen in other ethnic groups.

In the insulin autoimmune syndrome, hypoglycaemia may develop during the fasting state and/or more commonly during the postprandial state. Hyperglycaemia may occur immediately after a meal because binding of insulin to the autoantibodies reduces insulin availability to the peripheral tissues. In type B insulin resistance, most patients present with severe hyperglycaemia. Some patients may develop hypoglycaemia after a period of hyperglycaemia, or present only with severe fasting hypoglycaemia.

Acanthosis nigricans and signs of hyperandrogenism may develop in type B insulin resistance, but not in the insulin autoimmune syndrome. While both types of autoimmune hypoglycaemia may be associated with rheumatologic or haematologic diseases, the insulin autoimmune syndrome is also associated with exposure to medications, especially sulfhydryl-containing compounds. In these cases, the hypoglycaemia often resolves with medication cessation.

In the insulin autoimmune syndrome, insulin, C-peptide, and proinsulin concentrations determined by immunoassay may be unreliable, and although values returned are usually extremely high, most of the insulin is bound to antibody and not bioavailable. In the type B syndrome, insulin values are usually high in the basal state and even more exaggerated after glucose. A remarkable feature of the type B syndrome is an extremely high insulin/C-peptide ratio, due to impaired receptor-mediated insulin degradation and unaffected non-receptor-mediated C-peptide degradation.

Table 14.3.2.1 Clinical features in patients with insulin autoimmune syndrome and type B insulin resistance

Feature	Insulin autoimmune syndrome (Asians)	Insulin autoimmune syndrome (non-Asians)	Type B insulin resistance
Type of hypoglycaemia	Reactive (most common) and/or fasting	Reactive (most common) and/or fasting	Fasting; reactive uncommon
Hyperglycaemia	Common	Variable	Common
Insulin level	Very high*	Very high*	High
C-Peptide level	Very high*	Variable*	High
Proinsulin	Very high*	Variable*	High
Sex	Men and women	Men and women	Mostly women
Age of highest incidence (yrs)	40–80	40–80	40–50
Race	–	Mostly Caucasian	Mostly African-American
Acanthosis nigricans	Rare	Rare	Usually severe
Hyperandrogenism	No	No	Common
Associated Rheumatologic disease	Less common	Common	Common
Associated Haematologic disease	No	Uncommon	Common
Medication induced	Yes, most commonly sulfhydryl compounds	Yes, role of sulfhydryl compounds less clear	No
Insulin autoantibodies	Universal, frequently polyclonal	Universal,frequently monoclonal	Usually no[§¶]
Insulin receptor Antibodies	Usually no[§]	Usually no[§]	Universal
Response to therapy[#]	Variable	Variable	Usually good

* Artefactual; [§] Few cases with no prior exposure to exogenous insulin and having both insulin and insulin-receptor antibodies have been reported; [¶] Treatment with exogenous insulin may be associated with both insulin and insulin-receptor antibodies; [#] For those cases with no spontaneous remission.

Comparison Between Autoimmune Hypoglycaemia and Other Causes of Hyperinsulinaemic Hypoglycaemia

Hyperinsulinaemic hypoglycaemia in individuals not receiving exogenous insulin is most commonly caused by insulinoma. However, other causes, including autoimmune hypoglycaemia, should be investigated. In the presence of hypoglycaemic episodes, factitious hypoglycaemia should always be considered. The surreptitious administration of insulin is typically associated with high insulin and low C-peptide and proinsulin levels.

Treatment of Autoimmune Hypoglycaemia

For many autoimmune diseases, autoantibodies are valuable quantifiable disease markers but do not mediate the underlying pathologic tissue destruction, as in, for example, type 1 diabetes. In autoimmune hypoglycaemia syndromes, in contrast, the autoantibody itself drives the metabolic derangement and there is no associated tissue destruction unless there is another concomitant autoimmune disease. Thus, the goal of therapy is to deplete and ideally eliminate the autoantibody.

The insulin autoimmune syndrome and type B insulin resistance have a different clinical course and response to treatment. Approximately 80% of patients with insulin autoimmune syndrome show transient hypoglycaemia with spontaneous remission [1]. In all patients, medications possibly triggering the onset should be discontinued, and dietary modification to include small low-carbohydrate meals (low-glycaemic index diet) at regular intervals would help maintain normal blood glucose levels. Some patients with severe or prolonged symptoms may require pharmacologic treatment. Several different immunomodulatory treatments have been described, including plasmapheresis, glucocorticoids, azathioprine, 6-mercaptopurine, methotrexate and rituximab [1, 9, 11, 25, 26], but no standardized protocol has emerged to date. Treatments aimed at limiting endogenous insulin secretion, including acarbose, diazoxide, octreotide, and partial pancreatectomy, have also been employed [1, 7]. These therapies show variable efficacy in inducing remission.

The approach at the NIH has been to use high-dose steroids for short time periods (60 mg of prednisone/day for not more than 3 weeks with a rapid taper). This aims for immediate relief from hypoglycaemia and an initial reduction of antibody titre. This is followed by azathioprine for variable periods to sustain remission. This course may be repeated if necessary [27].

Type B insulin resistance is associated with high mortality, mostly related to the severity of the underlying systemic autoimmune disease. Other causes of death include hypoglycaemia, cardiovascular or cerebrovascular disease, or malignancy [19]. Treatment of type B insulin resistance has two main goals. The first goal is to reverse the severe catabolic state and hyperglycaemia seen in most patients. To achieve this goal, massive amounts of insulin, ranging from several hundred to several thousand units per day, are usually needed. Use of concentrated insulin U-500 is thus preferred. In patients showing hypoglycaemia, treatment with high-dose corticosteroids results in prompt resolution of the hypoglycaemia, usually within 24 hours. This effect occurs before any detectable change in the antibody titre and is likely mediated by stimulation of gluconeogenesis [19].

The second goal of treatment is to eliminate autoantibodies to the insulin receptor. About one-third of the patients with type B insulin resistance undergo spontaneous remission, but the time course is unpredictable and variable. Immunosuppressive treatment is usually required to correct the metabolic derangements. Several therapeutic approaches have been tried with no clear benefit, including plasmapheresis, glucocorticoid regimens, azathioprine, mycophenolate mofetil and cyclophosphamide [28, 29]. More

recently, the NIH has developed and tested a standardized therapeutic regimen combining rituximab, pulse steroids, and cyclophosphamide. Rituximab is a monoclonal antibody against CD-20, a molecule expressed on B-cell progenitors' surface, and is used to deplete antibody-producing B-lymphocytes and reduce the production of new autoantibodies. Pulses of high-dose steroids suppress the activity of pre-existing plasma cells. Four-day administration of oral dexamethasone 40 mg/day, or 2-day administration of intravenous methylprednisolone 1 g/day are associated with better tolerance than continuous low-dose steroid therapy. Cyclophosphamide is used as an adjunctive immunomodulatory agent to further suppress B—and T-cell function. If severe neutropenia occurs, cyclosporine can be chosen instead of cyclophosphamide. The administration of the NIH standardized therapeutic regimen is associated with an average time to remission of 5 months, with most patients showing reduction in their insulin requirements and improvement of their symptoms after the first cycle of treatment. Once remission is achieved, with improvement of the hyperglycaemia and discontinuation of insulin therapy, maintenance immunotherapy with azathioprine is indicated for at least 1 year [28, 29]. Beside the necessity of monitoring the patient's haematologic status during this therapy, there is a special precaution of ovarian toxicity in reproductive age women receiving cyclophosphamide and this drug should not be used longer than 6 months.

In type B insulin resistance, patients initially presenting with hypoglycaemia are rare, and currently it is not known whether the NIH combination therapy would be beneficial. Two case reports have recently described complete and partial remission with rituximab or bortezomib respectively, in patients with type B insulin resistance and hypoglycaemia [30, 31]. However, to date, available data are too limited to recommend a specific treatment.

Conclusion

Autoimmune forms of hypoglycaemia are rare. However, they should be considered in the differential diagnosis of hyperinsulinaemic hypoglycaemia. Patients may develop hypoglycaemia during the fasting and/or postprandial state. Insulin, C-peptide, and proinsulin levels are extremely elevated, with variable patterns of derangement of the insulin to C-peptide molar ratio depending on the underlying condition. The measurement of insulin or insulin-receptor antibodies can confirm the diagnosis and help differentiate autoimmune hypoglycaemia from insulinoma, therefore sparing patients from unnecessary invasive imaging or pancreatic surgical procedures. Administration of combined immunomodulatory therapies is beneficial for those patients who do not undergo spontaneous remission.

In rare cases, antibodies against endogenous or exogenous insulin may play a role in causing unstable glycaemia in patients with type 1 diabetes treated with exogenous insulin. However, the production of insulin antibodies is often induced by insulin therapy and usually these antibodies do not affect glycaemic control. In these patients, factors influencing blood glucose stability, including the site and technique of injection, altered absorption of insulin or compliance to the prescribed treatment need to be taken into account before considering the use of immunosuppressive therapies, which are associated with significant side effects and toxicity.

ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

REFERENCES

1. Hirata Y, Uchigata Y . Insulin autoimmune syndrome in Japan . *Diabetes Res Clin Pract*, 1994; **24** Suppl: S153–7.
2. Flier JS, Bar RS, Muggeo M, Kahn CR, Roth J, Gorden P . The evolving clinical course of patients with insulin receptor autoantibodies: spontaneous remission or receptor proliferation with hypoglycemia . *J Clin Endocrinol Metab*, 1978; **47**: 985–95.
3. Kahn CR, Rosenthal AS . Immunologic reactions to insulin: insulin allergy, insulin resistance, and the autoimmune insulin syndrome . *Diabetes Care*, 1979; **2**: 283–95.
4. Fineberg SE, Kawabata TT, Finco-Kent D, Fountaine RJ, Finch GL, Krasner AS . Immunological responses to exogenous insulin. *Endocr Rev*, 2007; **28**: 625–52.
5. Uchigata Y, Hirata Y, Iwamoto Y . Drug-induced insulin autoimmune syndrome . *Diabetes Res Clin Pract*, 2009; **83**: e19–20.
6. Wang YL, Yao PW, Zhang XT, Luo ZZ, Wu PQ, Xiao F . Insulin autoimmune syndrome: 73 cases of clinical analysis . *Chin Med J (Engl)*, 2015; **128**: 2408–9.
7. Lupsa BC, Chong AY, Cochran EK, Soos MA, Semple RK, Gorden P . Autoimmune forms of hypoglycemia . *Medicine (Baltimore)*, 2009; **88**: 141–53.
8. Uchigata Y, Hirata Y, Omori Y, Iwamoto Y, Tokunaga K . Worldwide differences in the incidence of insulin autoimmune syndrome (Hirata disease) with respect to the evolution of HLA-DR4 alleles . *Hum Immunol*, 2000; **61**: 154–7.
9. Church D, Cardoso L, Kay RG, *et al.* Assessment and management of anti-insulin autoantibodies in varying presentations of insulin autoimmune syndrome. *J Clin Endocrinol Metab*, 2018; **103**: 3845–55.
10. Kure M, Katsura Y, Kosano H, *et al.* A trial to assess the amount of insulin antibodies in diabetic patients by surface plasmon resonance. *Intern Med*, 2005; **44**: 100–6.
11. Jassam N, Amin N, Holland P, *et al.* Analytical and clinical challenges in a patient with concurrent type 1 diabetes, subcutaneous insulin resistance and insulin autoimmune syndrome. *Endocrinol Diabetes Metab Case Rep*, 2014; **2014**: 130086.
12. Trabucchi A, Iacono RF, Guerra LL, *et al.* Characterization of insulin antibodies by Surface Plasmon Resonance in two clinical cases: brittle diabetes and insulin autoimmune syndrome. *PLoS One*, 2013; **8**: e84099.
13. Ishizuka T, Ogawa S, Mori T, *et al.* Characteristics of the antibodies of two patients who developed daytime hyperglycemia and morning hypoglycemia because of insulin antibodies. *Diabetes Res Clin Pract*, 2009; **84**: e21–3.
14. Schade DS, Duckworth WC . In search of the subcutaneous-insulin-resistance syndrome . *N Engl J Med*, 1986; **315**: 147–53.
15. Rave K, Heise T, Heinemann L, Boss AH . Inhaled Technosphere insulin in comparison to subcutaneous regular human insulin: time action profile and variability in subjects with type 2 diabetes. *J Diabetes Sci Technol*, 2008; **2**: 205–12.
16. Hildebrandt P . Subcutaneous absorption of insulin in insulin-dependent diabetic patients. Influence of species, physicochemical properties of insulin and physiological factors . *Dan Med Bull*, 1991; **38**: 337–46.
17. Galloway JA, Spradlin CT, Nelson RL, Wentworth SM, Davidson JA, Swarner JL . Factors influencing the absorption, serum

insulin concentration, and blood glucose responses after injections of regular insulin and various insulin mixtures . *Diabetes Care*, 1981; **4**: 366–76.
18. Church D, Cardoso L, Bradbury S, *et al*. Diagnosis of insulin autoimmune syndrome using polyethylene glycol precipitation and gel filtration chromatography with ex vivo insulin exchange. *Clin Endocrinol*, 2017; **86**: 347–53.
19. Arioglu E, Andewelt A, Diabo C, Bell M, Taylor SI, Gorden P . Clinical course of the syndrome of autoantibodies to the insulin receptor (type B insulin resistance): a 28-year perspective . *Medicine (Baltimore)*, 2002; **81**: 87–100.
20. Semple RK, Halberg NH, Burling K, *et al*. Paradoxical elevation of high-molecular weight adiponectin in acquired extreme insulin resistance due to insulin receptor antibodies. *Diabetes*, 2007; **56**: 1712–17.
21. Brown RJ, Joseph J, Cochran E, Gewert C, Semple R, Gorden P . Type B Insulin Resistance Masquerading as Ovarian Hyperthecosis . *J Clin Endocrinol Metab*, 2017; **102**: 1789–91.
22. Bortolotti D, Mothe-Satney I, Ferrari P, *et al*. Spontaneous hypoglycaemia in the presence of both anti-insulin antibody and anti-insulin receptor antibody. *Diabetes Metab*, 2006; **32**: 598–603.
23. Kato T, Itoh M, Hanashita J, *et al*. Severe hypoglycaemia in a person with insulin autoimmune syndrome accompanied by insulin receptor anomaly type B. *Diabet Med,* 2007; **24**: 1279–81.
24. Kiyokawa H, Kono N, Hamaguchi T, *et al*. Hyperinsulinemia due to impaired insulin clearance associated with fasting hypoglycemia and postprandial hyperglycemia: an analysis of a patient with antiinsulin receptor antibodies. *J Clin Endocrinol Metab*, 1989; **69**: 616–21.
25. Church D, Hunter RW, Lyall M, *et al*. Resolution of hypoglycemia and cardiovascular dysfunction after rituximab treatment of insulin autoimmune syndrome. *Diabetes Care*, 2017; **40**: e80–2.
26. Saxon DR, McDermott MT, Michels AW . Novel management of insulin autoimmune syndrome with rituximab and continuous glucose monitoring . *J Clin Endocrinol Metab*, 2016; **101**: 1931–4.
27. Cochran E, Brown RJ, Gorden P . Other antibodies resulting in diabetes mellitus: type b insulin resistance and insulin autoimmune syndrome . *AACE Clin Case Rep*, 2016; **2**: e272–5.
28. Malek R, Chong AY, Lupsa BC, *et al*. Treatment of type B insulin resistance: a novel approach to reduce insulin receptor autoantibodies. *J Clin Endocrinol Metab*, 2010; **95**: 3641–7.
29. Klubo-Gwiezdzinska J, Lange M, Cochran E, *et al*. Combined immunosuppressive therapy induces remission in patients with severe type B insulin resistance: a prospective cohort study. *Diabetes Care*, 2018; **41**: 2353–60.
30. Bourron O, Caron-Debarle M, Hie M, *et al*. Type B Insulin-resistance syndrome: a cause of reversible autoimmune hypoglycaemia. *Lancet*, 2014; **384**: 1548.
31. Shanker K, Daley T, Semple R, Rouster-Stevens K, Ham JN . Intractable hypoglycemia in the setting of autoimmune overlap syndrome . *Pediatrics*, 2017; **139**.

14.3.3 Disorders of Carbohydrate Metabolism

Robin H. Lachmann

Introduction

Sugar molecules play many roles in metabolism. Glucose is an essential source of energy in the body, but carbohydrates also have important structural and signalling functions as constituents of glycoproteins, glycolipids, and glycosaminoglycans. In this chapter, we will describe a number of monogenic diseases which involve the monosaccharides glucose, galactose, and fructose and their roles in intermediary metabolism. The many other inherited metabolic diseases which affect the formation of glycosylated macromolecules (the congenital disorders of glycosylation) or their breakdown (lysosomal storage disorders) [1] will not be discussed.

Disorders of carbohydrate metabolism, although caused be defects in individual enzymes, are best viewed as disorders of metabolic *pathways*. Their tissue pathology can be due to deficiency of a product of metabolism, but just as often it is due to accumulation of toxic molecules which cannot be metabolized. Thus hypoglycaemia, which occurs in a number of these disorders, may be due to disabling mutations in enzymes of the glycogenolytic or gluconeogenic pathways, or to the inhibition of these enzymes by other accumulating molecules. In addition, hypoglycaemia can be caused by disorders in metabolic pathways affecting use of other fuels, such as those producing fatty acids and ketone bodies, which are both important alternative sources of energy. The interrelation between these glucose generating pathways is shown in **Figure 14.3.3.1**.

Glycogen Storage Diseases

The first process which maintains blood glucose during fasting is glycogenolysis: that is, the release of glucose from carbohydrate stores. The body stores carbohydrate in the form of glycogen, a branched polymer of glucose. Glycogen stores in the liver are used to maintain normoglycaemia, but muscle also stores glycogen for its own use as an energy source during exercise. Glycogen storage disorders (GSDs) are the inherited metabolic diseases which affect the synthesis and breakdown of glycogen.

Glycogen

Glycogen allows for the compact storage of glucose in a form that has a minimal osmotic effect but which is readily accessible and metabolically active. The core of a glycogen molecule is a small protein, glycogenin. Branched chains of polymerized α-d-glucose units are covalently attached to this at their reducing termini (**Figure 14.3.3.2**).

The glucose molecules in glycogen chains are linked to each other by α-1,4 glycosidic bonds with α-1,6 bonds at the branch

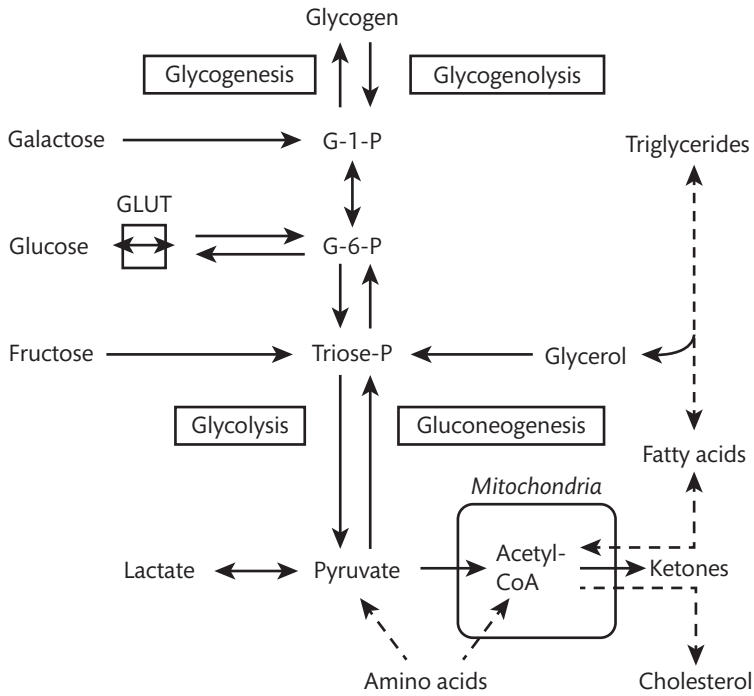


Figure 14.3.3.1 Main pathways of carbohydrate metabolism and relation with lipid metabolism. Acetyl-CoA, acetyl-coenzyme A; G-1-P, glucose-1-phosphate; G-6-P, glucose-6-phosphate; GLUT, glucose transporter; triose-P, triose phosphate.

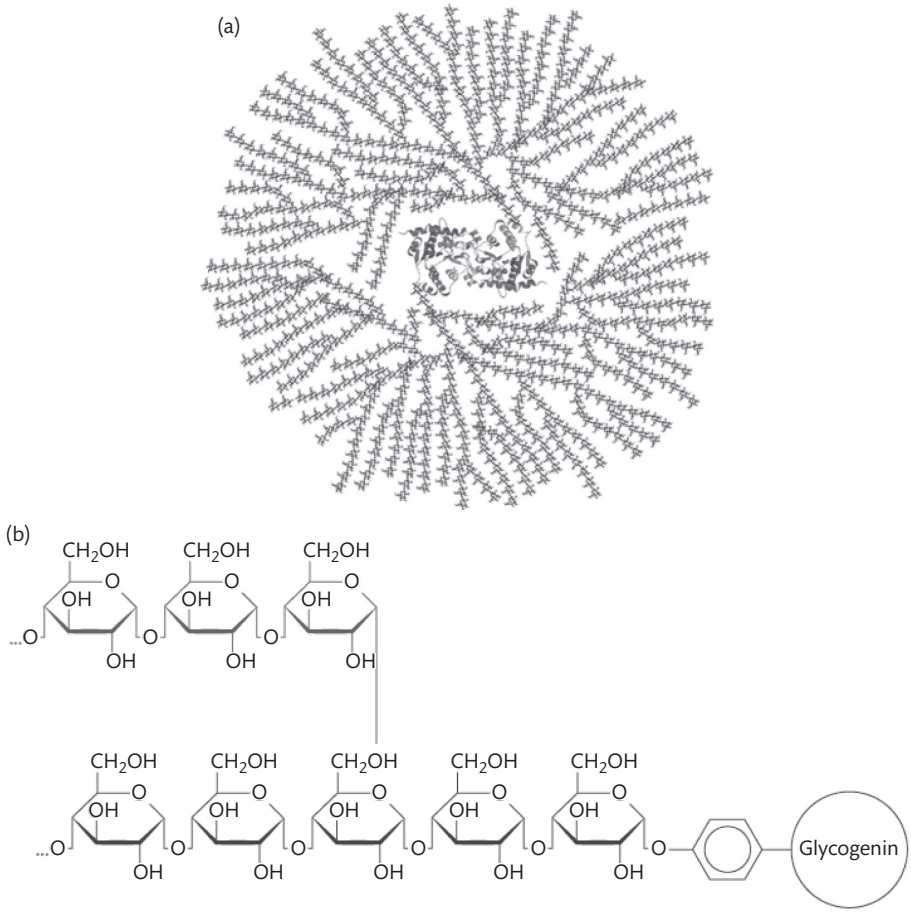


Figure 14.3.3.2 (a) A cross-sectional view of glycogen, showing the core glycogen protein surrounded by branches of glucose units, up to 60 000 of which can be contained within a glycogen granule. (b) Linear chains of glucose are formed by a-1,4 glycosidic bonds, with a-1,6 bonds at the branch points.

points. Glycogen molecules can contain up to 60,000 glucose molecules, have a molecular weight of several million daltons and are visible to the electron microscope. The liver and muscle contain between 200 and 300 g of glycogen. The 'arborization', or extensive branching, of the molecule, with large numbers of long outer chains that terminate in non-reducing glucose residues means that the enzymes of glycogen degradation can rapidly release large quantities of glucose.

Glycogen storage diseases can be caused by defects in glycogen synthesis or glycogen breakdown. The stored glycogen may have a normal or an aberrant structure. Depending on the enzymatic defect, glycogen metabolism in the liver, muscle, or both tissues may be affected. In this chapter we will focus on the diseases which affect glycogen breakdown in the liver and which can cause hypoglycaemia (Figure 14.3.3.3).

Glycogen Breakdown

Two enzymes are involved in the breakdown of glycogen in the cytoplasm: phosphorylase and debranching enzyme. Other enzymes are required subsequently to produce free glucose.

Phosphorylase sequentially removes glucose 1-phosphate units from the α-1,4-linked chains of glycogen. Debranching enzyme possesses transferase and α-1,6-glucosidase activities. When phosphorylase has degraded glycogen chains to within four α-1,4-glucosyl units of an α-1,6 linkage, three glucose residues are transferred to the end of another chain by the glycosyltransferase activity. Debranching enzyme then hydrolyses the remaining α-1,6 bond to release free glucose using its amylo-1,6-glucosidase activity. Debranching enzyme also cleaves the unique glucosyl–tyrosine linkage that anchors the terminal reducing glucose unit to glycogenin.

The main product of glycogen breakdown in muscle and liver is glucose 1-phosphate. Glucose 1-phosphate is a key intermediate of glycolysis, gluconeogenesis, glycogenolysis, and the pentose-phosphate pathway (Figure 14.3.3.1), but cannot be transferred outside the cell. However, after conversion to glucose 6-phosphate by phosphoglucomutase, free glucose is formed by the action of glucose 6-phosphatase. Glucose 6-phosphatase exists as a multicomponent complex in the endoplasmic reticulum of hepatocytes and, to a lesser extent, in renal tubular cells—it is not found in muscle. This complex contains glucose 6-phosphatase, several proteins that facilitate the transport of glucose, glucose 6-phosphate, and phosphate, as well as other stabilizing and regulatory units. Hepatic activity of glucose 6-phosphatase is the predominant metabolic source of blood glucose.

A summary of hepatic GSDs, their enzymology, and principal features is given in Table 14.3.3.1.

The commoner disorders are described next.

Glycogen Storage Disease Type I (Von Gierke's Disease)

Biochemistry

GSD I is due to Glucose-6-phosphatase deficiency. GSD Ia is caused by defects in subunits of the endoplasmic reticular enzyme complex that enable production of glucose from glucose 6-phosphate. In

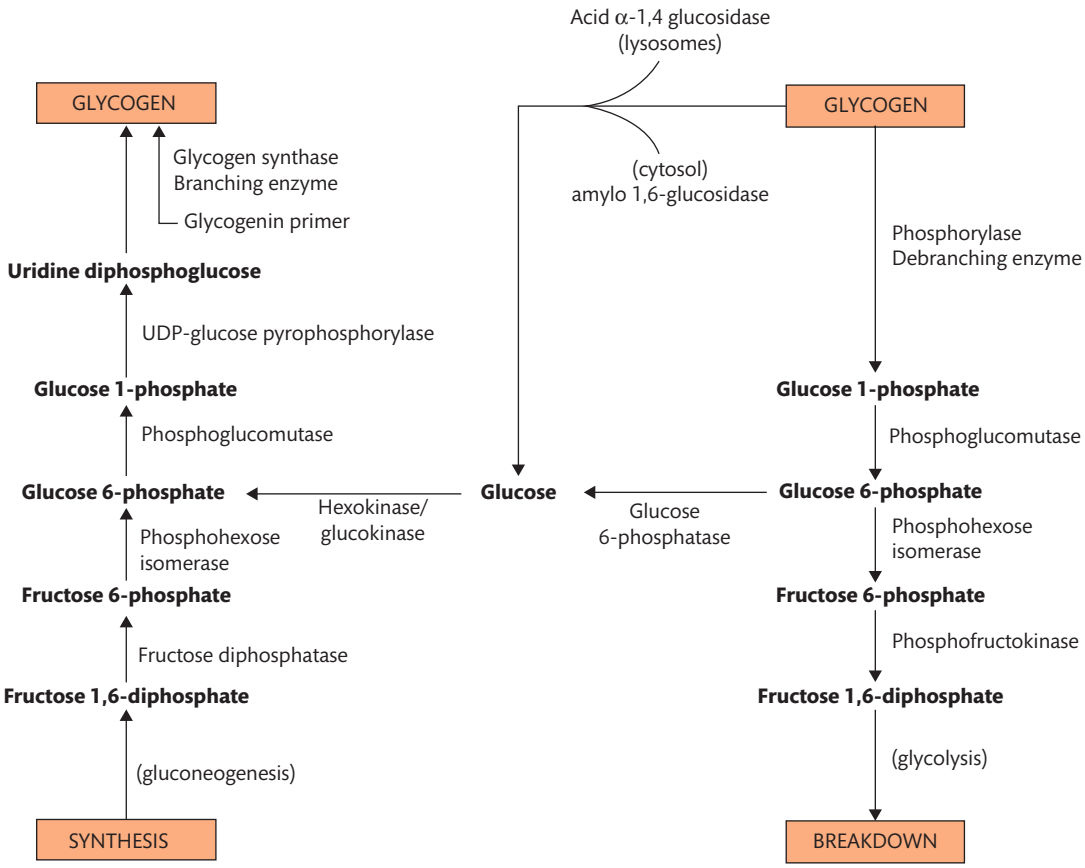


Figure 14.3.3.3 Glycogen breakdown.
Source data from Lee P, Bhattacharya K. In: Warrell D, Cox T, Firth J, eds. *Oxford Textbook of Medicine*. Oxford: Oxford University Press (2010).

Table 14.3.3.1 The glycogen storage disorders: genetic and enzymatic defects, and principal clinical features

GSD designation	Gene	Locus	Common term/Implicated protein	Supplementary	Clinical features
I (A) I (B)	*G6P6* *SLC37A4*	17q 21.31 11q 23.3	Von Gierke disease	Glucose 6-phosphatase G-6-P translocase	Hypoglycaemia, hyperlacticaemia, hyperuricaemia, hypertriglyceridaemia. Hepatomegaly, hepatic adenomata. Renal failure. As GSD Ia with neutropenia and colitis
III	*AGL*	1p21.2	Cori-Forbes disease (Limit dextrinosis)	Debranching enzyme	Hypoglycaemia, hepatomegaly, cardiomyopathy, myopathy
VI	*PYGL*	14q22.1	Hers disease	Glycogen phosphorylase Liver (very rare—Mennonite)	Hypoglycaemia, hepatomegaly
IX a2	*PHKA2*	Xp22.13	Phosphorylase kinase α2 subunit	Liver (regulatory)	Hypoglycaemia, hepatomegaly

GSD Ib the endoplasmic reticular transmembrane protein glucose-6-phosphate translocase is deficient. In both forms, the production of glucose from both glycogenolytic and gluconeogenic pathways is blocked, resulting in profound fasting hypoglycaemia (although this may be somewhat mitigated by the fact that small quantities of free glucose can be liberated by the α-1,6-glucosidase activity of the secondary action of debranching enzyme). Accompanying this, there is a build-up of glucose 6-phosphate. This is then metabolized by the pentose-phosphate shunt, or transferred back into glycogen which is stored in the liver and, to a lesser extent, the kidney, causing organomegaly. The products of glucose 6-phosphate metabolism have an important role to play in the other metabolic consequences of GSD I: hyperlacticaemia, hyperuricaemia, and hypertriglyceridaemia.

Lactic acidaemia results from stimulation of glycolysis at the level of phosphofructokinase by high concentrations of glucose 6-phosphate. Failure to dephosphorylate glucose 6-phosphate stimulates substrate cycling and increases the activity of the pentose-phosphate pathway, with enhanced production of purines. Degradation of purine nucleotides leads to overproduction of uric acid in the liver and hyperuricaemia. Enhanced flux through glycolysis and underutilization of gluconeogenic precursors leads to a cascade of metabolic changes that hypertriglyceridaemia.

Clinical Presentation

Patients typically present in infancy with symptomatic hypoglycaemia and failure to thrive, accompanied by a swollen abdomen due to hepatomegaly. Hypoglycaemic encephalopathy is often accompanied by seizures and can be fatal: recurrent episodes lead to permanent neurodisability. Children have impaired growth and increased subcutaneous fat deposition leading to a 'doll's face' appearance.

With aggressive dietary management (see next), the immediate life-threatening complications can be avoided. With improved survival, chronic, multisystem complications of GSD I have emerged (Figure 14.3.3.4).

There is persistent hepatomegaly, with glycogen storage accompanied by gross infiltration with fat. Cirrhosis and portal hypertension are, however, rare. Short stature, often combined with obesity, is common. The kidneys are enlarged by glycogen deposition. Progressive focal glomerulosclerosis and proximal tubular failure with a secondary Fanconi syndrome may also occur. Short periods of fasting, or other metabolic stressors such as infection provoke hypoglycaemia and lactic acidosis. In the longer term, poor metabolic control causes: growth arrest; hyperuricaemia and gout; marked hypertriglyceridaemia (which can lead to acute pancreatitis) and hypercholesterolaemia with raised very low-density lipoprotein (VLDL) and normal low-density lipoprotein (LDL) cholesterol concentrations in the plasma; and prolonged bleeding time related to an acquired von Willebrand-like defect affecting the platelet. Hepatic adenomata are seen in adults.

Patients with defects of the glucose 6-phosphate translocase system (type 1B) also have a neutropenia with impaired neutrophil migration and chemotaxis and are prone to recurrent bacterial infections. These patients are also at risk of developing granulomatous colitis with clinical features resembling ulcerative colitis.

Management

Historically, GSD I (and the other GSDs presenting with hypoglycaemia in infancy) were associated with a very poor outcome. This has been transformed by the introduction of aggressive dietary management aimed at maintaining a constant exogenous supply of glucose to meet basal requirements. Regular oral carbohydrate during the day and continuous overnight pump feeding with glucose, delivered either by nasogastric or gastrostomy tube, clearly improved clinical and biochemical parameters. Subsequently, fasting tolerance has been improved with the use of uncooked corn starch (obtained from the supermarket and suspended in water): this acts as a 'slow-release' form of glucose and, particularly in older patients, allows for more time between meals during the day and for some patients to discontinue overnight feeds. Modified corn starches have now been produced with the aim of increasing fasting tolerance further, although it is not yet clear that they offer a significant benefit over shop bought cornflour.

Maintaining normoglycaemia requires a diet with about 65% of dietary energy as carbohydrate (2). Continuous glucose monitoring can be useful in adjusting doses of uncooked corn starch and concentrations of overnight feeds. Regular dietetic review is important to minimize excessive weight gain, insulin resistance, and ensure the diet is nutritionally complete.

Intercurrent illness can rapidly provoke hypoglycaemia and patients with GSD I are given an 'emergency regime' to use in times of metabolic stress. This consists of frequent oral glucose polymer. If for any reason patients can't tolerate oral intake, glucose should be given intravenously at a rate of 0.2 g/kg/h.

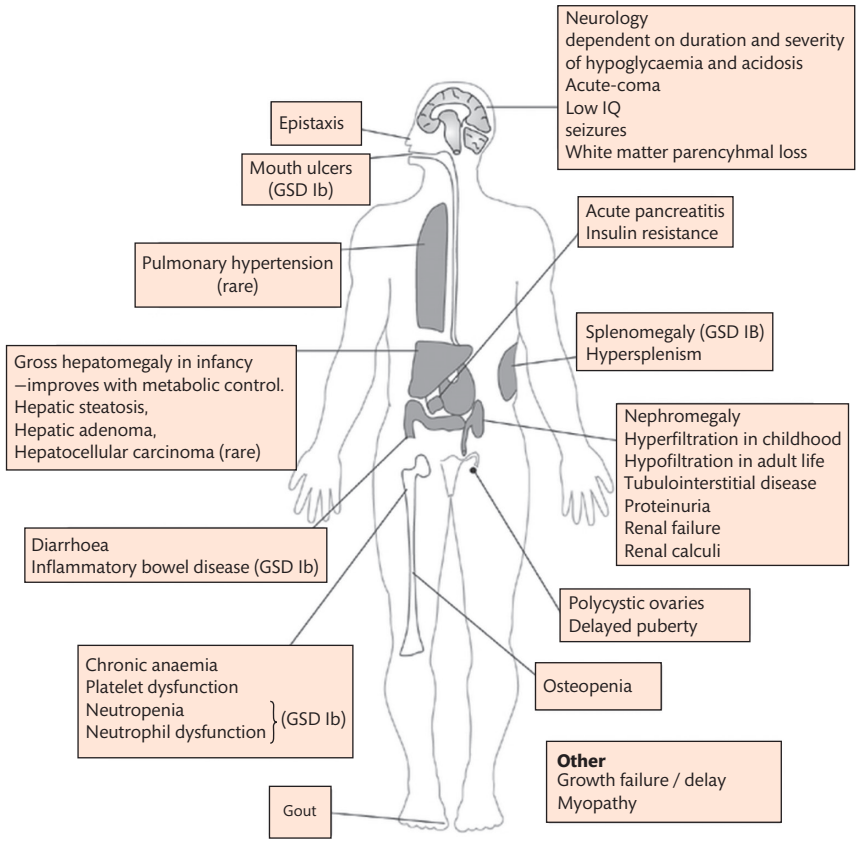


Figure 14.3.3.4 Complications of GSD I.

Adapted with permission from Lee P, Bhattacharya K. In: Warrell D, Cox T, Firth J, eds. *Oxford Textbook of Medicine*. Oxford: Oxford University Press (2010). with permission from Oxford University Press

Hyperlipidaemia and hyperuricaemia need to be treated. Hyperfiltration or albuminuria indicate renal involvement and ACE inhibitors or ARBs should be introduced. Hypocitraturia may contribute to the increased incidence of nephrolithiasis and citrate supplementation may be useful. Iron supplementation is often needed. Osteopenia is common and calcium and vitamin D supplementation should be considered.

Surveillance for hepatic adenomata is important. MRI with the use of iv contrast is the preferred method. Around 70–80% of adult patients have been reported to have at least one lesion, and these progress in size or number in 50% of cases. The occurrence of adenomata seems to be related to metabolic control and in some cases improving biochemical parameters can lead to adenoma regression. Spontaneous regression is also seen.

The occurrence of hepatic adenomata is concerning because they can progress to hepatocellular carcinoma: predicting this progression is difficult [2]. Blood markers such as α-fetopotein have not proved useful. A rapid increase in size or number of adenomata, changes in vascularization and bleeding should lead to a multidisciplinary team (MDT) review to discuss surgical intervention, including liver transplantation.

Human granulocyte colony-stimulating factor is often required in patients with GSD Ib to increase the neutrophil count and control mouth ulcers, recurrent infections, and inflammatory bowel disease. Long-term use of granulocyte colony-stimulating factor is associated with a number of complications and should be supervised by a haematologist.

Due to the dangers of fasting and the bleeding tendency associated with GSD I, surgery must be managed carefully. Patients should be admitted the day before so that fasting can be covered with intravenous glucose. Platelets should be available in case of postoperative haemorrhage.

Pregnancy in women with GSD I is now relatively routine. With careful planning, close attention to glycaemic control and increased carbohydrate requirements, especially in the second half of pregnancy, and a well-managed labour, outcomes are good.

With optimal medical management patients with GSD I now lead relatively normal lives, but some patients remain in whom good metabolic control is never obtained. For these patients, liver transplantation offers a long-term 'cure' for many features of the disease [3]. Where there is also end stage renal failure, combined liver and renal transplantation can be performed.

Glycogen Storage Disease Type III (Forbes–Cori Disease)

Biochemistry

GSD III is due to deficiency of debranching enzyme. This results in the storage of structurally abnormal glycogen, with short outer chains, called limit dextrin, in both liver and muscle. Although glycogenolysis is blocked, gluconeogenesis is unaffected and fasting hypoglycaemia is milder than that seen in GSD I and accompanied

by ketosis rather than lactic acidosis. The secondary metabolic consequences are mostly confined to mild hyperlipidaemia.

Clinical Presentation

GSD III affects both liver and muscle. Hypoglycaemia and the hepatic consequences of storage dominate the clinical picture in children, with fasting hypoglycaemia and poor growth. The condition is less severe than GSD I and even in historic cohorts, most patients survived to adulthood.

In adults, fasting tolerance improves and on the whole hypoglycaemia can be prevented with dietary management. Hepatic adenomata have only rarely been reported, although patients can occasionally develop cirrhosis, and the kidneys are not affected.

Patients do, however, develop muscle symptoms and complain of exercise intolerance [4], although rhabdomyolysis is not a recognized feature. Some patients develop a progressive, disabling myopathy with pronounced distal weakness and myopathic facies. Cardiac muscle is also involved and hypertrophic cardiomyopathy can result in arrhythmias or heart failure [5].

Management

The management of hypoglycaemia in childhood is as in GSD I [6]. In adult patients it is important not to overtreat as this can lead to insulin resistance: with home glucose monitoring it is often possible to reduce the dietary content of complex carbohydrate.

Although left ventricular hypertrophy occurs in many patients, its clinical significance is not clear [7]. To date there are very few case reports of heart failure or significant arrhythmia in adults. This may change as patients age and periodic echocardiography and electrocardiographic (ECG) monitoring is probably prudent.

The incidence of clinically significant hepatic fibrosis and cirrhosis may also increase with age and liver imaging can be used to monitor this as well as the occurrence of hepatic adenomata.

GSD IX

Biochemistry

GSD IX is due to deficiency of phosphorylase kinase. Phosphorylase kinase consists of four subunits, two of which have tissue specific isoforms. The commonest form of GSD IX, and the commonest GSD, is X-linked, and due to mutations in *PHKA2*.

Clinical Presentation

GSD IXa is a hepatic GSD presenting early in life with hepatomegaly and fasting hypoglycaemia and ketosis. It is milder than GSD I and symptoms generally resolve in adulthood. Liver fibrosis has, however, been reported as a long-term complication.

Management

Management of hypoglycaemia is as for GSD I, but adult patients have normal fasting tolerance and do not need uncooked corn starch.

Diagnosis of GSD

Most patients with hepatic GSDs present with hypoglycaemia in early life. Historically, definitive diagnosis relied on demonstrating glycogen storage and assaying enzyme activity in the affected tissue: many adults with GSD I still bear the scars of liver biopsies performed in infancy. This approach to diagnosis was not only invasive but also technically difficult, and has to a large extent been superseded by new techniques.

In infants with suggestive symptoms, biochemical profiling, with measurements of glucose, lactate, and ketones, can suggest the correct diagnosis. In some cases (e.g. GSD III) this can be confirmed by enzymological study of blood-derived leucocytes, but in GSD I molecular genetic analysis is required as the enzymes are only expressed in liver.

Some patients present with hepatomegaly without biochemical features suggesting GSD. In these cases, histological examination reveals glycogen storage. If frozen tissue has been kept, enzymology may then confirm the diagnosis of a GSD. Electron microscopy can also be helpful if structurally abnormal glycogen is present. A diagnostic fast, with measurement of glucose, lactate, and ketones can also provide useful information. Hepatomegaly with glycogen storage is not, however, always due to a GSD: hepatic glycogenosis is a well-recognized complication of poorly controlled diabetes mellitus [8].

Increasingly, molecular genetics is becoming the diagnostic test of choice. Next generation sequencing allows large arrays of genes to be sequenced at the same time, and genetic panels to screen for all known GSDs, are now available [9].

Disorders of Galactose Metabolism

Galactosaemia

Biochemistry

The disaccharide lactose, which is the most important carbohydrate in both human and cow's milk, is formed from glucose and galactose. Galactose therefore forms a large part of the energy intake of infants. Galactose metabolism is shown in Figure 14.3.3.5.

Galactosemia is caused by deficiency of galactose-1-phosphate uridyltransferase (transferase, EC 2.7.7.10) (Figure 14.3.3.5, enzyme 2). Galactose-1-phosphate, which accumulates due to the metabolic block, is toxic to many organs and tissues, including the liver, kidney, and brain. It suppresses the activities of some enzymes of glycogenolysis and gluconeogenesis, which may lead to hypoglycaemia. Galactitol, produced from galactose excess, leads to cataract formation and is excreted in the urine. As well as being derived from exogenous sources of galactose, galactose-1-phosphate is also produced endogenously from glucose-1-phosphate, by a reversal of the pyrophosphorylase-epimerase pathway (Figure 14.3.3.5, enzymes 3 and 4). This 'self-intoxication' may contribute to late complications, such as ovarian failure and neurological disease, which can develop despite a strict lactose-free diet.

Clinical Presentation

Infants with classic galactosaemia (MIM 230400), caused by severe deficiency of galactose-1-phosphate uridyltransferase (GALT) (transferase, EC 2.7.7.10) (Figure 14.3.3.5, enzyme 2) appear normal at birth but rapidly become very unwell. Refusal to feed and vomiting are accompanied by signs of liver disease: jaundice, hepatomegaly, oedema, and ascites. Cataracts appear within a few days

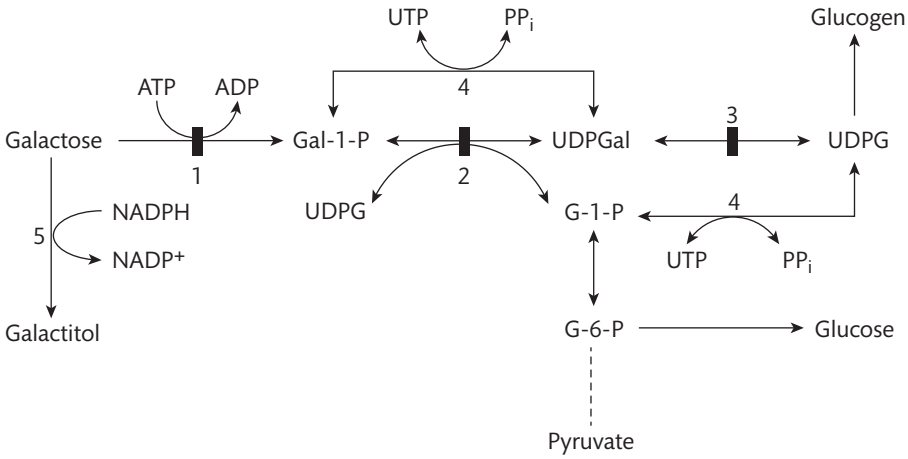


Figure 14.3.3.5 Galactose metabolism. 1, galactokinase; 2, galactose-1-phosphate uridyltransferase; 3, UDP galactose 4′-epimerase; 4, UDP glucose (UDP galactose) pyrophosphorylase; 5, aldose reductase. The three enzyme defects are depicted by solid bars across the arrows. Gal-1-P, galactose-1-phosphate; G-1-P, glucose-1-phosphate; G-6-P, glucose-6-phosphate; NADP, nicotinamide adenine dinucleotide phosphate; NADPH, reduced NADP; UDPG, uridine diphosphoglucose; UDPGal, uridine diphosphogalactose.

Reproduced with permission from Gitzelmann R. Disorders of galactose metabolism. In: Fernandes J, Saudubray J-M, van den Berghe G, eds. *Inborn Metabolic Diseases*. 3rd edn. Heidelberg, Springer-Verlag, 2000.

or weeks. Left untreated, liver and kidney failure develop and, along with sepsis are rapidly fatal [10].

Diagnosis

Many countries include galactosemia in their newborn screening programmes but, as classical galactosaemia presents very early in life, the newborn bloodspot needs to be taken early and processed quickly if the result is to be available before the neonate presents clinically: for this reason it is not included in newborn screening in the United Kingdom. Diagnosis otherwise rests on screening for total blood galactose, red blood cell (RBC) Gal-1-P, and/or urinary galactitol and confirmation by the measurement of GALT activity in RBC with or without genetic confirmation [11].

Treatment

Rapid introduction of a galactose-free diet rapidly reverses the acute presentation [11]. This involves withdrawal of breast milk or whey-based infant formula and the introduction of a soy-based or elemental formula. At weaning, the child should be started on a diet which is essentially dairy-free. Although there are small amounts of galactose in some fruits and vegetables, these are negligible when compared to endogenous galactose production, which is at least 8.4 mg/kg/day in adults, and so non-dairy sources of galactose are allowed. Many mature cheeses are also allowed as bacterial fermentation has removed all sugars and these provide a valuable source of calcium which is otherwise lacking in this diet. Low bone density is well recognized in adults with galactosemia and calcium and vitamin D intake should be carefully monitored throughout life and supplemented if required.

Long-Term Complications

Although acute intoxication with liver and kidney failure is only seen in infants, the long-term outcomes in galactosemia, even with early institution of dietary treatment, are not always good [12]. Average IQ is below the normal range and many patients have verbal dyspraxia. Adult patients often find it difficult to function in society and form relationships [13]. Neurological complications can also occur, and a few patients develop a progressive leukodystrophy [14]. The vast majority of women develop either primary or premature ovarian failure and require hormone replacement [11, 12]. This also contributes to the high risk of osteoporosis.

Currently it is recommended that a lactose-free diet is maintained for life, but there is little evidence to support this. There are reports of patients with galactosaemia who have relaxed their diets and not experienced any ill effects. Equally, there are patients who remain on a strict diet but go on to develop neurological disease.

Long-term problems might be due to continuous intoxication with galactose-1-phosphate, either produced endogenously or derived from complex sugars such as raffinose and stachyose by bacterial fermentation in the gut. In fact, it is possible that these effects are not related to the build-up of toxic metabolites at all, but actually reflect a generalized defect in glycosylation of proteins. Patients with congenital glycosylation defects demonstrate similar clinical features including hypogonadotropic premature ovarian failure, which is thought to relate to lack of glycosylation of follicle-stimulating hormone. Certainly, outcomes for patients with galactosemia on current treatment are not good and research is needed to find ways to improve them.

Disorders of Fructose Metabolism

Fructose is found in fruits, vegetables, and honey. With glucose, it forms the disaccharide sucrose, which is an important carbohydrate in many foods and beverages. Sucrose is hydrolysed into its two monosaccharides by the enzyme sucrase (EC 3.2.1.48) on the small intestinal mucosa. Another source of fructose is sorbitol, which is widely distributed in fruits and vegetables. It is converted in the liver into fructose by the enzyme sorbitol dehydrogenase (EC 1.1.1.14). The three inborn errors of fructose metabolism are shown in Figure 14.3.3.6.

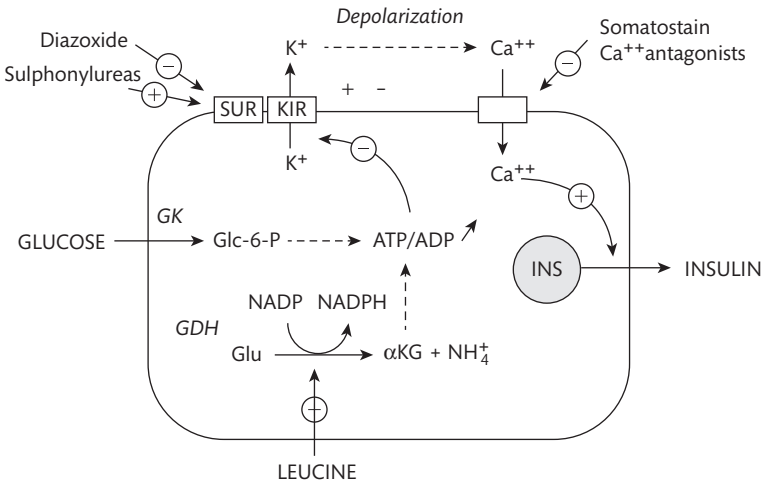


Figure 14.3.3.6 Fructose metabolism. 1, fructokinase; 2, aldolase-B; 3, fructose-1,6-bisphosphatase; 4, phosphofructokinase; 5, sorbitol dehydrogenase. The enzyme defects are depicted by solid bars across the arrows. ADP, adenosine diphosphate; ATP, adenosine triphosphate; DHA-P, dihydroxyacetone phosphate; F-1-P, fructose-1-phosphate; F-6-P, fructose-6-phosphate; F-1,6-P_2, fructose-1,6-bisphosphate; G-6-P, glucose-6-phosphate; Pi, inorganic phosphate; GAH-3-P, glycer-aldehyde-3-phosphate.

Essential Fructosuria

This is a rare 'non-disease', which does not show any clinical symptoms. It is caused by a deficiency of fructokinase (EC 2.7.1.4) (Figure 14.3.3.6, enzyme 1), which is normally found in liver, kidney, and small intestinal mucosa. Thus, fructose cannot be phosphorylated into fructose-1-phosphate. Instead, it is slowly phosphorylated into fructose-6-phosphate by the enzyme hexokinase in adipose tissue and muscle with the excess being excreted in the urine. The resultant fructosuria is the only finding. A discrepancy between a positive test for reducing sugars and a negative reaction with glucose oxidase should allow the identification of fructose as a non-glucose-reducing sugar. Dietary treatment is unnecessary.

Hereditary Fructose Intolerance

Clinical Presentation

Deficiency of aldolase-B (EC 4.1.2.13, MIM 229600), the second enzyme of the fructose pathway (Figure 14.3.3.6, enzyme 2) causes fructose-1-phosphate to accumulate after consumption of fructose containing foods. The mechanism of toxicity of fructose-1-phosphate is like that of galactose-1-phosphate in classical galactosaemia and there are some clinical similarities between the two disorders. When fructose/sucrose containing foods are introduced into the diet at weaning, the infant starts to vomit and refuse food, and develops failure to thrive with jaundice, hepatomegaly, oedema, ascites, and a bleeding tendency, reflecting liver dysfunction. Urinary findings are mellituria, proteinuria, and aminoaciduria, reflecting renal proximal tubular dysfunction. Diarrhoea and malabsorption reflect small intestinal involvement. Lethargy, tremor, and convulsions are due to hypoglycaemia (see next). The larger the fructose load and the younger the infant, the more acute the symptoms of intolerance. Older children may selectively refuse fructose containing products and never present in acute crisis. In these cases the diagnosis may be made by their dentist due to a complete freedom from dental caries.

Biochemistry

Fructose-1-phosphate, which accumulates due to the aldolase-B defect, is toxic. It causes hypoglycaemia by inhibiting enzymes of both glycogenolysis and gluconeogenesis: not only is the splitting of fructose-1-phosphate into three-carbon sugars impaired (Figure 14.3.3.6, enzyme 2), but also the condensation of the three—carbon sugars into fructose-1,6-bisphosphate (Figure 14.3.3.6, enzyme 3). The accumulation of fructose-1-phosphate also leads to the sequestration of inorganic phosphate, which is then not available for the regeneration of adenosine triphosphate (ATP) from ADP, causing a generalized energy defect in the cell. This provokes the catabolism of adenine nucleotides, which leads to the overproduction of uric acid.

Diagnosis

The clinical picture, combined with the finding of a combination of fructosuria and disturbed liver function tests, should lead to the suspicion of hereditary fructose intolerance Traditionally an intravenous fructose tolerance test was performed, but assay of aldolase-B activity in a biopsy of liver, jejunal mucosa or kidney cortex, or DNA analysis is simpler and safer.

Treatment and Prognosis

All sources of fructose, sucrose, and sorbitol (which can be present in many products such as medicines) should be excluded from the diet. They should be replaced by glucose, maltose, and starch. This elimination diet rapidly corrects all abnormalities except the hepatomegaly, which is slower to resolve. If small amounts of fructose remain in the diet growth may remain slow, but will catch up after further adjustment of the diet and overall, with treatment, the prognosis is excellent [15].

Fructose-1,6-Bisphosphatase Deficiency (MIM 229700)

Clinical Presentation

Fructose-1,6-bisphosphatase (EC 3.1.3.11) has a role both in the conversion of fructose to glucose (Figure 14.3.3.6, enzyme 3) and in gluconeogenesis, in which fructose-1,6-bisphosphatase is the third unidirectional enzyme (Figure 14.3.3.3, enzyme 3). Therefore, deficiency of the enzyme leads to abnormalities due to the impairment of both pathways, though those of failing gluconeogenesis are more serious than those of impaired fructose conversion. Hypoglycaemia, associated with lactic acidosis, occurs in the neonatal period and can recur in later childhood. It usually develops after prolonged fasting or with an intercurrent febrile illness. The clinical symptoms of hypoglycaemia (lethargy, irritability, apnoea, coma, and convulsions) are accompanied by hyperpnoea, somnolence, and vomiting due to the lactic acidosis. Attacks may also occur after ingestion of fructose or sucrose. The frequency of attacks decreases with increasing age. Mild hyperlactacidaemia may persist between episodes. Growth and psychomotor development are usually normal.

Biochemistry

As gluconeogenesis is blocked, glucose cannot be synthesized from lactate, pyruvate, alanine, glycerol, or fructose. The patient depends on exogenous glucose and galactose and endogenous glycogen for their glucose requirements. On fasting, hypoglycaemia and lactic acidosis develop, sometimes accompanied by hyperketonaemia. Lactate, pyruvate, alanine, glycerol, and glycerol-3-phosphate accumulate in blood and urine.

Diagnosis

Enzymatic assay of fructose-1,6-bisphosphatase in a biopsy of the liver, jejunal mucosa, or kidney cortex is the only reliable means for the diagnosis.

Treatment and Prognosis

The acute, life-threatening attack is treated with an intravenous glucose drip: an intravenous glucose bolus (200 mg glucose/kg bodyweight over 5 min) followed by a continuous infusion (*c.* 12 mg glucose/kg bodyweight per min). Sodium bicarbonate may be given to treat the lactic acidosis. In order to prevent further attacks, it is important to avoid prolonged fasting. An emergency regimen consisting of frequent carbohydrate is given during intercurrent infection. In small children restriction (not elimination) of fructose, sucrose, and sorbitol is recommended. The tolerance for fasting improves with age and the prognosis is good if adequate treatment is introduced in infancy.

REFERENCES

1. Hollak CEM, Lachmann R . *Inherited Metabolic Disease in Adults: A Clinical Guide*. Oxford; New York: Oxford University Press, 2016: 648.
2. Kishnani PS, Austin SL, Abdenur JE, *et al*. Diagnosis and management of glycogen storage disease type I: a practice guideline of the American College of Medical Genetics and Genomics. *Genet Med*, 2014; **16**: e1.
3. Boers SJB, Visser G, Smit PGPA, Fuchs SA . Liver transplantation in glycogen storage disease type I. *Orphanet J Rare Dis*, 2014; **9**: 47.
4. Preisler N, Laforêt P, Madsen KL, *et al*. Skeletal muscle metabolism is impaired during exercise in glycogen storage disease type III. *Neurology*, 2015; **84**: 1767–71.
5. Dagli A, Sentner CP, Weinstein DA . Glycogen Storage Disease Type III. In: Pagon RA, Adam MP, Ardinger HH, *et al*., eds. GeneReviews(®) [Internet]. Seattle, WA: University of Washington, 1993. Available at: http://www.ncbi.nlm.nih.gov/books/NBK26372/ (accessed: 2 Oct 2015).
6. Derks TGJ, Smit GPA . Dietary management in glycogen storage disease type III: what is the evidence? *J Inherit Metab Dis*, 2015; **38**: 545–50.
7. Kishnani PS, Austin SL, Arn P, *et al*. Glycogen storage disease type III diagnosis and management guidelines. *Genet Med*, 2010; **12**: 446–63.
8. Julián MT, Alonso N, Ojanguren I, Pizarro E, Ballestar E, Puig-Domingo M . Hepatic glycogenosis: an underdiagnosed complication of diabetes mellitus? *World J Diabetes,* 2015; **6**: 321–5.
9. Olpin SE, Murphy E, Kirk RJ, Taylor RW, Quinlivan R . The investigation and management of metabolic myopathies . *J Clin Pathol*, 2015; **68**: 410–17.
10. Bosch AM. Classical galactosaemia revisited . *J Inherit Metab Dis*, 2006; **29**: 516–25.
11. Welling L, Bernstein LE, Berry GT, *et al*. International clinical guideline for the management of classical galactosemia: diagnosis, treatment, and follow-up. *J Inherit Metab Dis*, 2017; **40**: 171–6.
12. Waisbren SE, Potter NL, Gordon CM, *et al*. The adult galactosemic phenotype. *J Inherit Metab Dis*, 2012; **35**: 279–86.
13. Hoffmann B, Dragano N, Schweitzer-Krantz S . Living situation, occupation and health-related quality of life in adult patients with classic galactosemia . *J Inherit Metab Dis*, 2012; **35**: 1051–8.
14. Rubio-Agusti I, Carecchio M, Bhatia KP, *et al*. Movement disorders in adult patients with classical galactosemia. *Mov Disord*, 2013; **28**: 804–10.
15. Ali M, Rellos P, Cox TM . Hereditary fructose intolerance . *J Med Genet*, 1998; **35**: 353–65.

14.3.4 Haemochromatosis and Other Inherited Diseases of Iron Metabolism

Yves Deugnier and Edouard Bardou-Jacquet

Introduction

Iron is a vital element for many biological processes, but can also be deleterious if present in excess. Because we lack a physiological pathway to remove iron absorption is tightly regulated to maintain a balanced iron store. Iron overload disease has been recognized centuries ago, but its understanding drastically improved in the recent decades. Identification of the main genes involved in iron metabolism, and association of mutations in some of these genes with different iron overload syndromes over the past 25 years, have streamlined precise diagnosis of these heterogeneous disorders, and have made it possible to construct a mechanism-based classification [1]. Genetic disorders are classified as haemochromatotic and non-haemochromatotic, and contrast with acquired iron excess, which is associated mainly with haematological conditions, cirrhosis, and metabolic syndrome. However even in these disorders clinical expression is often modulated by genetic as well as environmental factors

Hereditary Disorders

Genetic Haemochromatosis

Genetic haemochromatosis is heterogeneous, accounted for by several gene defects that share certain characteristics, namely (i) dysregulation of hepcidin synthesis or impairment of hepcidin action resulting in intestinal hyperabsorption and increased macrophage egress of iron; (ii) elevated serum iron with increased transferrin saturation; (iii) variable clinical phenotypes, ranging from early and severe disease in juvenile haemochromatosis related to the hemojuvelin (*HJV*) or the hepcidin (*HAMP*) genes to late and less severe disease in adult haemochromatosis related to the *HFE* or the transferrin 2 receptor (*TfR2*) gene (Figure 14.3.4.1); and (iv) recessive transmission, except in the exceptional case of ferroportin disease, caused by underlying gain-of-function mutations.

HFE Haemochromatosis

This condition alone accounts for almost all cases of haemochromatosis in populations of European descent [2].

Genetics and Pathophysiology

A major causal mutation, C282Y, and many minor mutations have been described on the *HFE* gene [3]. The C282Y mutation (Replacement of tyrosine by cysteine in position 282 of the protein) immobilizes the HFE protein intracellularly and impairs regulation of hepcidin synthesis. This leads to intestinal hyperabsorption of iron and to release of iron from macrophages with the following consequences: (i) hypersaturation of transferrin, the plasma iron transport protein; (ii) appearance of non-transferrin-bound iron (NTBI) [4] which, penetrating easily into parenchyma, is responsible for tissue iron loading (Figure 14.3.4.1). Among *HFE* genotypes, only C282Y homozygosity—whose prevalence ranges between 1 and 9% in Europe [5]—is associated with overt haemochromatosis. However, its clinical penetrance is rather low [6, 7], estimated at 1% in females and 27% in males [8]. Double heterozygosity for C282Y and a minor mutation, including the H63D and the S65C mutations, has no significant effect on intracellular migration of the HFE protein and its physiological role. However, compound C282Y—H63D heterozygosity may mildly increase body iron stores, but with no clinical expression in the absence of genetic or environmental cofactors [6, 9, 10].

Natural History

Initially, the disease is quiescent. An increase in serum iron and in transferrin saturation (the earliest and most sensitive marker) then occurs, followed by a gradual rise in serum ferritin concentration. Later, most often between 35 and 45 years of age in men and 45 and 55 years of age in women, clinical symptoms and then visceral complications may occur. This results in the division of disease progression into five stages (Figure 14.3.4.2), however passage from one stage to the next is never guaranteed given the low penetrance of the disease.

Clinical Expression

In daily practice, a-, mono—and pauci-symptomatic forms have become the rule due to (i) the highly variable expressiveness of the disease, particularly in women in whom asthenia and joint symptoms dominate; (ii) the frequent involvement of cofactors, which may determine selective visceral expression, such as hepatic forms in men; and (iii) the systematization of biochemical assessment and family screening in healthy people, leading to early diagnosis.

Endocrine Complications

Dominated by diabetes and pituitary disease, endocrine complications of HFE-related iron overload indicate an advanced form of the disease. Overt diabetes [11], which was part of the historical picture of hemochromatosis, is currently present in less than 10% of cases at the time of diagnosis. Moreover, systematic searches for haemochromatosis in newly developed diabetes reveal it in less than 2% of cases. Conversely, the incidence of insulin-requiring diabetes in haemochromatotic cirrhosis is significantly higher than in non-haemochromatotic cirrhosis, which supports a specific role of iron. The underlying mechanism is dual, involving insulin resistance in liver and, to a lesser degree, muscle, as well as insulinopenia secondary to destruction of pancreatic islet cells by iron accumulation. Haemochromatotic subjects exhibit insulin resistance/hyperinsulinemia while glucose tolerance is still normal, and function of both β cells (normal C peptide blood concentration) and α cells (normal glucagonemia and secretion of GIP—gastric inhibitory polypeptide) is intact. Subsequently there is parallel progression of hyperglycaemia towards diabetes due to iron-mediated loss of β cells, and of liver disease towards cirrhosis, while secretion of other pancreatic hormones remains normal. Involvement of additional factors, many genetic, has been advanced on the basis of early family studies, but has not been

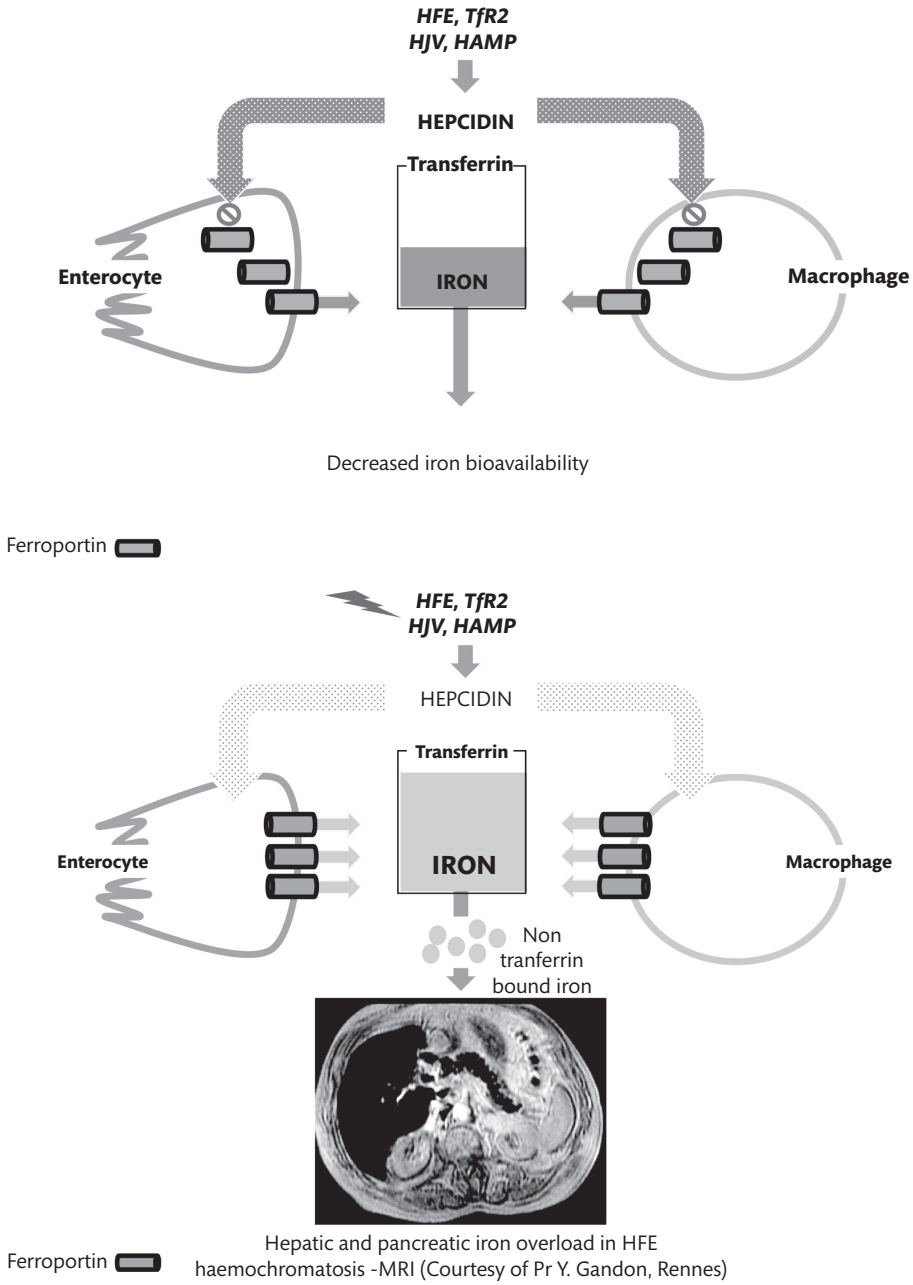


Figure 14.3.4.1 (a) Physiological role of hepcidin. The iron stock regulates hepcidin synthesis via a cascade including the HFE, hemojuvelin and transferrin 2 receptor (TfR 2) proteins. Hepcidin controls the release of iron by ferroportin, especially in enterocytes and macrophages. It acts by reducing the cellular egress of iron, which results in decreasing both serum iron and transferrin saturation and, then, the bioavailability of iron. (b) Pathophysiology of iron loading in haemochromatosis. Alteration of the HFE, hemojuvelin, transferrin 2 receptor, or hepcidin gene is responsible for the dysregulation of hepcidin production. This results in lifting of the inhibition of iron egress from enterocytes and macrophages which is responsible for increased serum iron and transferrin saturation. The subsequent production of non-transferrin-bound iron (NTBI) leads to iron loading, as NTBI is avidly taken up by parenchymas.

confirmed by the most recent studies that failed to find an increased frequency of hyperglycaemia in relatives of people with haemochromatotic diabetes. The role of C282Y heterozygosity in development of diabetes also remains debated but, if it exists, it is minimal.

Diabetes in haemochromatosis does not differ from other forms of diabetes with respect to the frequency of onset and presentation of ophthalmologic, cardiovascular, neurologic, and renal complications. However, it is a major independent prognostic factor: survival of haemochromatotic patients is reduced in those with diabetes (93% at 5 years, 77% at 10 years, 62% at 15 years and 55% at 20 years) while that of non-diabetic haemochromatotic patients is almost identical to that of the general population. Diabetes is rarely the direct cause of death, however. It rather intervenes indirectly by promoting the occurrence of coronary heart disease, infections, and even hepatocellular carcinoma.

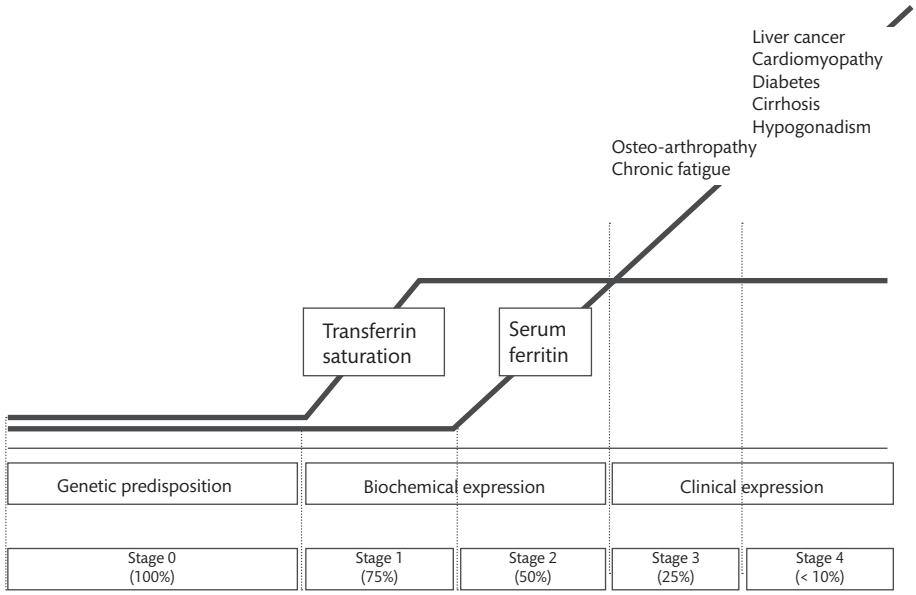


Figure 14.3.4.2 Phenotypic stages in HFE haemochromatosis (C282Y homozygosity). Not all patients progress to overt disease. According to the recommendations of the Haute Autorité de Santé, http://www.anaes.fr.

Pituitary damage in haemochromatosis [12, 13] has long been known from autopsy studies that showed major iron loading in pituitary cells, particularly gonadotrophs. Its main clinical consequence is hypogonadism [14], seen in 12% to 38% of men at the time of diagnosis and manifest as decreased libido, testicular atrophy, and hair loss and, biochemically, as decreased testosteronemia [14]. Lowering of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) concentrations, the lack of response to gonadotropin-releasing hormone (GnRH) that is general seen, and the improvement of hypogonadism on administration of hCG all support a predominantly pituitary origin of the disorder. Cirrhosis also contributes to the decline of testosteronemia through a poorly understood mechanism, and diabetes can also be involved through neurological and vascular damage. Panhypopituitarism, central hypothyroidism [15], GH hyposecretion and adrenal insufficiency [16] due to haemochromatosis are all exceptional and anecdotal.

Extra-Endocrine Disorders

Chronic fatigue is one of the first symptoms of the disease, and biochemical workup frequently has the initial aim of seeking iron deficiency. The classically described greyish-brown skin pigmentation, predominating in exposed areas, nipples, and external genitalia, is a late sign, sometimes associated with ichthyosis of the arms and legs, leukonychia, and koilonychia.

Joint symptoms are common initial presentations, especially in women [17, 18] and are not correlated with the extent of iron excess. Nevertheless, they can be seriously debilitating [19]. Small distal joints are worst affected, especially the metacarpophalangeal and proximal interphalangeal joints of the second and third fingers, resulting in pain on shaking hands. Other joints may also be affected including wrists, ankles and, more rarely, hips, knees, and shoulders. Symptoms are arthralgias with a rather inflammatory rhythm, with a relapsing and remitting pattern, and acute flares in 5–10% of cases. In advanced disease joints are deformed and functional impairment can become major. Radiologically, reduction of joint space, cystic subchondral osteopathy, sclerosis, and osteophytosis are seen in different combinations. Chondrocalcinosis is present in a quarter of cases. The pathophysiology of joint damage remains poorly understood.

Decreased bone mineral density [20] is detected in one-third of patients and is associated in 15% to 18% of cases with spinal fractures. It is somewhat correlated with iron excess, but is also driven by hypogonadism, cirrhosis and, possibly hyperparathormonaemia [21].

Hepatomegaly can be major in advanced disease and is characterized by (i) classical left lobe predominance; (ii) undisturbed or almost normal liver functions (the most frequent abnormality is moderately (<3 fold) increased serum alanine amino transaminase (ALT) concentration); and (iii) rarity of hepatocellular failure and portal hypertension, even when cirrhosis is present, in the absence of cofactors. A major risk, however, is the development of primary liver cancer—especially hepatocellular carcinoma—as soon as cirrhosis has developed [22, 23]. Although reduced, the risk persists even if the patients is appropriately treated and fibrosis has regressed [24].

When present, heart damage is usually limited to echocardiographic abnormalities including cardiac hyperechogenicity, increased left ventricular volume without parietal thickening, and decreased left ventricular ejection fraction. In advanced disease, rhythm disorders may be seen, including atrial fibrillation, paroxysmal supraventricular tachycardia, atrial flutter or atrioventricular block and/or congestive heart failure associated with cardiomegaly and systolic dysfunction on echocardiography and cardiac hyposignal on magnetic resonance imaging (MRI).

A causal relationship between iron overload (and/or chelator therapy) and certain infections, particularly those due to *Yersinia enterolitica*, *Vibrio vulnificus*, and *Mucor*, has been suggested. Asymptomatic deficiency in vitamins A and C is common.

Diagnosis

Haemochromatosis is diagnosed based on a combination of abnormal serum iron tests, and presence of the HFE C282Y mutation.

Serum iron concentration greater than 30 mmol/L and, more specifically, transferrin saturation exceeding 60%, are common, together with slight hypotransferrinaemia of around 1.9 g/L. It is essential, however, to consider and exclude (i) false negatives where there is associated systemic inflammation or metabolic syndrome; and (ii) false positives connected with haemolysis in the sampling tube, cell lysis (e.g. hepatic necrosis, haemolysis, rhabdomyolysis), metabolic syndrome, and/or hepatic disease (increased transferrin saturation due to deficient transferrin synthesis by the liver). Transferrin saturation remains the best index for both diagnosis of and disease monitoring in haemochromatosis, and indeed in the absence of systemic inflammation, metabolic syndrome, and outside the context of systematic screening, normal transferrin saturation rules out the disease.

Unlike serum iron concentration, serum ferritin concentration provides information on the degree of iron excess and therefore varies greatly from one subject to another. However, although a normal concentration allows exclusion of the diagnosis of iron overload, normal serum ferritin concentration does not exclude HFE C282Y homozygosity. Conversely, there are frequent false positive cases of hyperferritinaemia in the absence of excess body iron due to numerous conditions including excessive alcohol consumption, metabolic syndrome, cell lysis, inflammatory, and macrophage activation syndromes, tumours, Gaucher disease, and hyperferritinaemia-cataract syndrome.

HFE genotyping is indicated if, and only if, transferrin saturation is elevated and requires prior consent and subsequent genetic counselling. A diagnosis of HFE haemochromatosis is certain in people with HFE C282Y homozygosity. C282Y-H63D compound heterozygosity may be responsible for mild, clinically asymptomatic iron excess but not overt haemochromatosis. The diagnosis must be ruled out in the case of any other HFE genotypes **except** when the three following conditions are met: (1) authentication of significant iron overload by MRI or liver biopsy; (2) exclusion of all other causes of iron overload (see next); and (3) demonstration of two major mutations on the HFE gene, which generally corresponds to compound heterozygote states combining C282Y with a very rare or 'private' mutation [25].

Further Clinical Evaluation

No special evaluation is needed in patients with stage 0/1 haemochromatosis. In stages 2, 3, and four clinical examination should seek the various manifestations of the disease, and blood count, determination of serum transaminases and fasting glucose, and electrocardiography should be undertaken. Additional assessments may be helpful depending on the context. Liver biopsy may be necessary to assess hepatic fibrosis as cirrhosis confers an increased risk of liver cancer and thus modifies subsequent management. It is not indicated in C282Y homozygotes whose the liver is not enlarged AND serum ferritin levels are lower than 1000 ng/ml AND serum alanine aminotransaminase (ALAT) levels are normal, since these patients have been shown never to have bridging fibrosis or cirrhosis [26]. It remains indicated in those who do not meet these three criteria. Elastometry may, in the near future, further restrict use of biopsy [27]. Hepatic MRI replaces biopsy in quantification of iron excess [28] when it is likely to be overestimated by serum ferritin alone. It also detects (pre)neoplastic lesions [29] which exhibit relative hyperintensity of signal (by absence of iron at their level) [30]. Joint radiography may be indicated in case of painful joints as well as bone densitometry if risk factors for low bone mineral density are present (e.g. peri or postmenopause, cirrhosis, and/or hypogonadism). Determination of serum testosterone may be useful in men with sexual disturbances, while echocardiography is indicated in case of heart-related symptoms or systematically when serum ferritin levels exceed 1000 ng/ml.

Treatment

Phlebotomy remains the main approach to depletion of excess iron. This involves removal of 5–7 ml blood/kg per session, which are weekly until serum ferritin concentration falls below 50 µg/L. This first phase can last from a few weeks to more than two years, depending on the extent of the initial iron excess. It is followed by a lifelong maintenance phase consisting of regular phlebotomy whose rate and volume are adjusted to maintain serum ferritin concentration below 50–100 µg/L and, if possible, transferrin saturation below 50% [31]. One phlebotomy every 1–4 months is then sufficient, except for (i) during pregnancy, when iron removal is suspended; (ii) the time of menopause, when the rhythm and/or volume of phlebotomy must be adapted to the new physiological conditions; and (iii) in the elderly, when it is customary to reduce the unit volume of phlebotomy to 150–250ml or to stop, according to the patient's general condition and tolerance of the procedure.

Tolerance of phlebotomy is excellent as long as puncture points are varied and fluid is replenished before and after each session. Blood pressure measured before and after each blood removal, and blood count monthly initially or at each bleeding during the maintenance phase are used to monitor for side effects of therapy. Its effectiveness is monitored by serum ferritin concentration monthly at the beginning of treatment then every two bleeds once serum ferritin has dropped below 300 ng/ml (200 ng/ml in women).

Survival outcomes are excellent. Indeed, when initial serum ferritin concentration lies between the upper limit of normal and 1000 µg/L, survival is longer than that of the general population due to lower mortality from cardiovascular disease and extrahepatic cancer [32]. Only people with initial serum ferritin greater than 2000 µg/L have reduced life expectancy, mainly due to liver cancer and diabetes [32]. Phlebotomy normalizes prevalence of asthenia, melanoderma, heart damage, hepatomegaly, and increased serum transaminase levels, but does not lead to full resolution of joint manifestations, diabetes, and hypogonadism. Glucose intolerance can be improved, but not insulin-requiring diabetes, although insulin requirement is reduced in 40% of cases [11]. A few cases of regression of hypogonadism on depletive treatment have been reported [14, 33, 34], but hypogonadism generally requires additional specific management. The risk of liver cancer is reduced but remains above background in 'de-ironed' cirrhotic patients, mandating continued biannual ultrasound surveillance.

Although phlebotomy is convenient, cheap, and efficient, some patients are reluctant to undergo this treatment or are unable to tolerate it because of associated disorders (e.g. anaemia, heart failure). In this setting iron chelators may be used to deplete the excess iron burden. Deferoxamine has been used for decades in iron-overloading anaemias, and some reports showed good results in haemochromatosis. However continuous subcutaneous infusion is required, hampering adherence to treatment. The oral iron chelator deferasirox showed good efficacy but poor tolerability in a phase 1/2 clinical trial [35] and cannot be used to maintain low body iron

stores. In the near future, hepcidin agonists currently under development and clinical evaluation could prove valuable options, as they would fully address the pathological hepcidin deficiency that causes haemochromatosis [36].

There is no evidence that a low iron diet is beneficial in the long term. Reduction of intestinal absorption of iron by tea consumption [37] or by proton pump inhibitor has been advocated in small controlled studies [38, 39]. Alcohol consumption should be discouraged during the initial depletion period and should then be limited, except in cirrhosis where it should be prohibited. Vitamin C supplementation should be avoided at the beginning of depletive treatment as it may precipitate cardiac decompensation. The finding of macrocytosis may motivate the use of folic acid. Visceral complications do not require specific measures. Treatment of hypogonadism may include use of androgens administered transcutaneously at physiological doses, other forms of androgen therapy being suspected of increasing risk of liver cancer.

Screening and Prevention

European [5] and American [40] recommendations do not advocate systematic screening for the disease but strongly recommend family screening of first-degree relatives of any C282Y homozygote. Collection of both phenotypic and genotypic data is recommended in siblings, but screening of parents may be phenotypic only. For children it is suggested either to wait until the age of 18 or to offer genotyping to the other parent to clarify their risk. In C282Y homozygotes detected at stage 0 or 1, simple biological monitoring is required every 3 years (stage 0) or annually (stage 1). For those who are screened at stages 2 to 4, depletive therapy should be implemented after appropriate pretherapeutic assessment. C282Y heterozygotes can be reassured and released from any surveillance, except in the event of an abnormality of their iron balance, which requires the search for compound heterozygosity or any other cause of disturbance of iron metabolism. Genotyping of their partner is suitable in order to assess any possible risk of homozygosity in any offspring.

Non-HFE Haemochromatosis

These are extremely rare conditions. Type 2 haemochromatosis or juvenile haemochromatosis is linked to mutation of the hemojuvelin gene (*HJV*) [41] or, even more exceptionally, in the hepcidin gene (*HAMP*) [42]. It is expressed in adolescents or young adults by hypogonadotropic hypogonadism, heart failure, and cirrhosis which, as a rule, respond well to vigorous treatment with phlebotomy. Type 3 haemochromatosis has been described in some Mediterranean and other European families [43]. It is linked to mutation of the transferrin receptor 2 gene (*TFR2*) and produces a picture superimposable upon that of HFE hemochromatosis.

Non-Haemochromatotic Iron Overload Syndromes

These are genetic conditions that also affect iron metabolism, but that have a disease phenotype distinct from that of genetic haemochromatosis.

Ferroportin Disease

Ferroportin is a protein that, under the control of hepcidin, ensures egress of iron from enterocytes and macrophages to plasma [44]. Mutations of the ferroportin gene are rare and lead to type 4 haemochromatosis, also known as 'ferroportin disease'. The most frequent form (Type 4A) is clearly distinguished from HFE haemochromatosis by (i) its dominant autosomal mode of transmission; (i) its very low clinical expressiveness; (iii) its biochemical phenotype, which is characterized by elevated serum ferritin concentration but normal or only slightly increased transferrin saturation; and (iv) a predominantly mesenchymal pattern of iron deposition [45]. The disease is usually diagnosed in adulthood because of its long clinical latency in the absence of environmental cofactors. Rare mutations rendering ferroportin insensitive to hepcidin lead to phenotype similar to that of HFE haemochromatosis (Type 4B).

Aceruloplasminemia

This is an autosomal recessive disease of adulthood featuring parenchymal iron overload affecting the liver, pancreas (diabetes) and central nervous system (retinitis pigmentosa, extrapyramidal syndrome and disorders of cognitive functions up to dementia) [46]. Biochemically, it is characterized by hyperferritinaemia with low transferrin saturation, but its hallmark is the extremely low rate of serum ceruloplasmin. Phlebotomy rapidly induces anaemia. Iron chelation is then the treatment of choice.

Acquired Disorders

Excessive iron therapy of anaemias, excessive iron intake, and liver cirrhosis all usually induce significant iron overload. The metabolic syndrome is responsible for only mild iron excess, in contrast.

Haematologic Disorders

Dyserythropoiesis, whatever its cause (e.g. thalassemia, hereditary or acquired sideroblastic anaemias, congenital dyserythropoietic anaemia) can induce iron overload secondary to iron hyperabsorption even before blood transfusion. The resulting clinical picture is similar to that of haemochromatosis. Most often, the haematological disorder is known and the patient is already treated by regular blood transfusions when the problem of iron overload is unmasked, however. In thalassemia major [47], which remains the most frequent cause of secondary iron overload, iron hyperabsorption and high transfusion requirements (on average 200–300 ml/kg/day, or 0.25–0.40 mg/kg/day of iron) contribute to early iron overload. In the first decade, this manifest as hepatomegaly with fibrosis and growth retardation. It is later responsible for delayed sexual maturation and cardiomyopathy, the latter being the main cause of death in young patients. All serum iron load markers are increased. Treatment is based on dietary measures to limit intake and absorption of iron (e.g. tea consumption) and use of iron chelators. Deferasirox, an orally active chelator that is generally well tolerated by young patients has revolutionized the management of this type of overload.

Excessive Iron Intake

Whatever its route of administration, regular and prolonged iron intake can lead to iron overload, as seen sometimes in high performance athletes unduly supplemented with iron [48], for example.

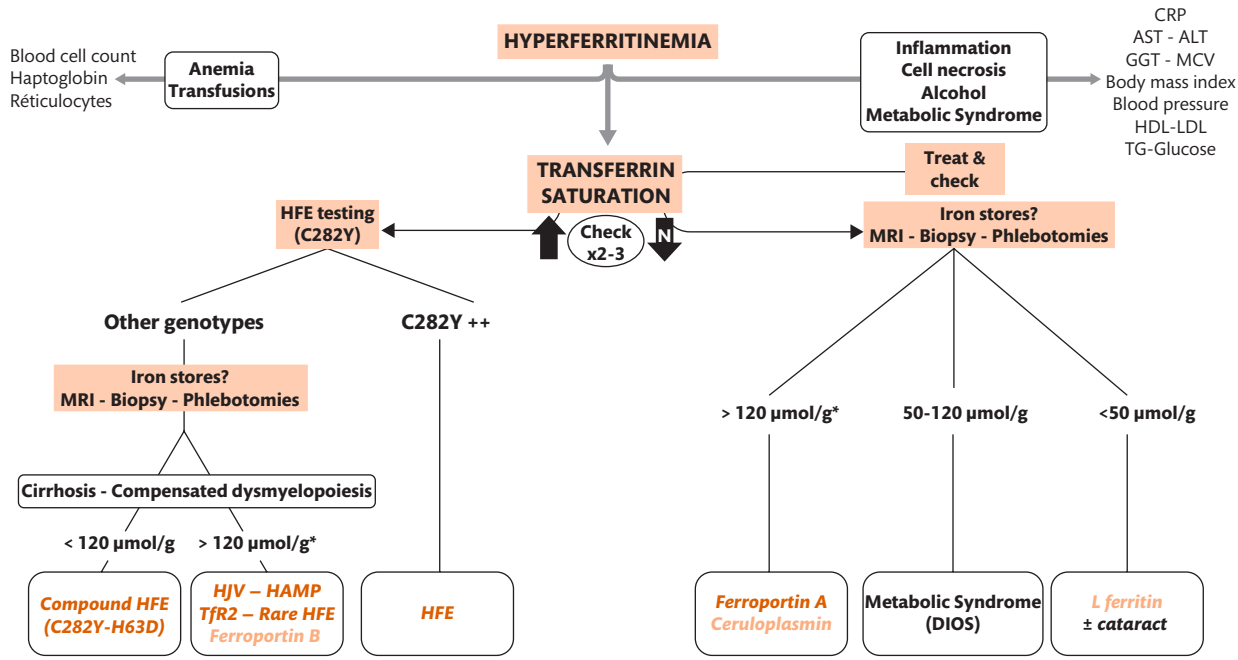


Figure 14.3.4.3 Diagnostic algorithm for hyperferritinaemia. DIOS, dysmetabolic iron overload syndrome; HAMP, hepcidin; HJV, hemojuvelin; TfR2, transferrin receptor 2. *Italics: genetic disorders (recessive/dominant).* * Or amount of iron removed by weekly/bimonthly phlebotomies: 3 g in males and 2 g in females (1 L of blood = 0.5 g of iron).

Cirrhosis of the Liver

In any cirrhosis, a pseudo-haemochromatotic picture including melanoderma, hypogonadism, diabetes, hyperferritinaemia, increased transferrin saturation (by hypotransferrinaemia and hypohepcidinaemia secondary to hepatocellular insufficiency [49–51]) and hepatic iron overload (by hepatocytic uptake of NTBI). In these cases, it is often necessary to look for the underlying mutation to distinguish confidently between genetic haemochromatosis and iron overload secondary to cirrhosis.

Metabolic Syndrome

Hyperferritinaemia is present in almost half of patients with metabolic syndrome [52, 53]. It is related to subclinical inflammation and steatohepatitis associated with insulin resistance, but may indicate true excess of body iron, as part of what has been coined as the dysmetabolic iron overload syndrome (DIOS) [54]. DIOS denotes a combination of unexplained hepatic iron overload and one or more elements of the metabolic syndrome. Its diagnosis is based on (i) this association; (ii) the absence of any other identifiable cause of hepatic iron overload; and (iii) certain biochemical patterns of abnormality, including absence of increased transferrin saturation despite hyperferritinaemia up to 1500 ng/ml [55], a mixed pattern of hepatic iron deposits (i.e. in hepatocytes and macrophages) and a moderate increase of hepatic iron concentration, rarely exceeding 120 µmol/g (N <35). In half of such patients, steatosis, or steatohepatitis coexists, with severe fibrosis or cirrhosis in 15% of cases. This explains the fairly broad indications for liver biopsy in these individuals. Alteration of iron metabolism in DIOS likely results from a multifactorial and dynamic process driven by chronic positive energy balance, and featuring cross-talk between the liver and overloaded adipose tissue. Successful treatment of metabolic abnormalities to current targets is mandatory but is insufficient to normalize serum ferritin. Two recent controlled studies [56, 57], however, have failed to support phlebotomy as a useful option in DIOS patients. Sustained modification of diet and lifestyle thus remains the first therapeutic intervention, together with pharmacological control of blood pressure, blood glucose, and dyslipidaemia when necessary.

As shown in Figure 14.3.4.3, diagnosis of disorders of iron metabolism has benefited greatly from advances in molecular genetics and MRI. Nevertheless, it remains firmly rooted upon a strict clinical approach recognizing that acquired conditions are the rule and genetic causes the exception. Targeting hepcidin with drugs represents the most promising prospect for novel therapeutics in future.

ACKNOWLEDGEMENTS

The authors thank Stuart Byrom (DRCI, Pontchaillou University Hospital, 35033 Rennes cedex, France) for reviewing the English language.

REFERENCES

1. Powell LW, Seckington RC, Deugnier Y . Haemochromatosis . *Lancet*, 2016; **388**: 706–16.
2. Brissot P, Moirand R, Jouanolle AM, *et al.* A genotypic study of 217 unrelated probands diagnosed as 'genetic hemochromatosis' on 'classical' phenotypic criteria. *J Hepatol*, 1999; **30**: 588–93.
3. Hollerer I, Bachmann A, Muckenthaler MU . Pathophysiological consequences and benefits of HFE mutations: 20 years of research. *Haematologica*, 2017; **102**: 809–17.
4. Brissot P, Ropert M, Le Lan C, Loreal O . Non-transferrin bound iron: a key role in iron overload and iron toxicity . *Biochimica et Biophysica Acta*, 2012; **1820**: 403–10.

5. Pietrangelo A, Deugnier Y, Dooley J, Erhardt A, Zoller H . EASL clinical practice guidelines for HFE hemochromatosis . *J Hepatol*, 2010; **53**: 3–22.
6. Moirand R, Jouanolle AM, Brissot P, Le Gall JY, David V, Deugnier Y. Phenotypic expression of HFE mutations: a French study of 1110 unrelated iron-overloaded patients and relatives . *Gastroenterology*, 1999; **116**: 372–7.
7. Beutler E, Felitti VJ, Koziol JA, Ho NJ, Gelbart T . Penetrance of 845G--> A (C282Y) HFE hereditary haemochromatosis mutation in the USA . *Lancet,* 2002; **359**: 211–18.
8. Allen KJ, Gurrin LC, Constantine CC, *et al.* Iron-overload-related disease in HFE hereditary hemochromatosis. *N Engl J Med*, 2008; **358**: 221–30.
9. Gurrin LC, Bertalli NA, Dalton GW, *et al.* HFE C282Y/H63D compound heterozygotes are at low risk of hemochromatosis-related morbidity. *Hepatology*, 2009; **50**: 94–101.
10. Walsh A, Dixon JL, Ramm GA, *et al.* The clinical relevance of compound heterozygosity for the C282Y and H63D substitutions in hemochromatosis. *Clin Gastroenterol Hepatol*, 2006; **4**: 1403–10.
11. Strohmeyer G, Niederau C . Diabetes mellitus and hemochromatosis. In: Barton JC, Edwards CQ, eds. *Hemochromatosis Genetics: Pathophysiology, Diagnosis, and Treatment*. Cambridge: Cambridge University Press, 2000: 268–77.
12. Kontogeorgos G, Handy S, Kovacs K, Horvath E, Scheithauer BW. The Anterior Pituitary in Hemochromatosis . *Endocr Pathol*, 1996; **7**: 159–64.
13. Walsh CH . Non-diabetic endocrinopathy in hemochromatosis . In: Barton JC, Edwards CQ, eds. *Hemochromatosis Genetics: Pathophysiology, Diagnosis, and Treatment*. Cambridge: Cambridge University Press; 2000: 278–89.
14. McDermott JH, Walsh CH . Hypogonadism in hereditary hemochromatosis . *J Clin Endocrinol Metab*, 2005; **90**: 2451–5.
15. Murphy MS, Walsh CH . Thyroid function in haemochromatosis . *Ir J Med Sci*, 2004; **173**: 27–9.
16. Hempenius LM, Van Dam PS, Marx JJ, Koppeschaar HP . Mineralocorticoid status and endocrine dysfunction in severe hemochromatosis . *J Endocrinol Invest*, 1999; **22**: 369–76.
17. Ines LS, da Silva JA, Malcata AB, Porto AL . Arthropathy of genetic hemochromatosis: a major and distinctive manifestation of the disease. *Clin Exp Rheumatol*, 2001; **19**: 98–102.
18. Rull M, Schumacher Jr R . The arthropathy of hemochromatosis . In: Barton JC, Edwards CQ, eds. *Hemochromatosis Genetics: Pathophysiology, Diagnosis, and Treatment*. Cambridge: Cambridge University Press, 2000: 258–67.
19. Adams PC, Speechley M . The effect of arthritis on the quality of life in hereditary hemochromatosis . *J Rheumatol*, 1996; **23**: 707–10.
20. Guggenbuhl P, Deugnier Y, Boisdet JF, *et al.* Bone mineral density in men with genetic hemochromatosis and HFE gene mutation. *Osteoporos Int*, 2005; **16**: 1809–14.
21. Pawlotsky Y, Le Dantec P, Moirand R, *et al.* Elevated parathyroid hormone 44–68 and osteoarticular changes in patients with genetic hemochromatosis. *Arthritis Rheum*, 1999; **42**: 799–806.
22. Bradbear RA, Bain C, Siskind V, *et al.* Cohort study of internal malignancy in genetic hemochromatosis and other chronic nonalcoholic liver diseases. *J Natl Cancer Inst*, 1985; **75**: 81–4.
23. Deugnier YM, Guyader D, Crantock L, *et al.* Primary liver cancer in genetic hemochromatosis: a clinical, pathological, and pathogenetic study of 54 cases. *Gastroenterology*, 1993; **104**: 228–34.
24. Falize L, Guillygomarc'h A, Perrin M, *et al.* Reversibility of hepatic fibrosis in treated genetic hemochromatosis: a study of 36 cases. *Hepatology*, 2006; **44**: 472–7.
25. Hamdi-Roze H, Beaumont-Epinette MP, Ben Ali Z, *et al.* Rare HFE variants are the most frequent cause of hemochromatosis in non-c282y homozygous patients with hemochromatosis. *Am J Hematol*, 2016; **91**: 1202–5.
26. Guyader D, Jacquelinet C, Moirand R, *et al.* Noninvasive prediction of fibrosis in C282Y homozygous hemochromatosis. *Gastroenterology*, 1998; **115**: 929–36.
27. Legros L, Bardou-Jacquet E, Latournerie M, *et al.* Non-invasive assessment of liver fibrosis in C282Y homozygous HFE hemochromatosis. *Liver Int*, 2015; **35**: 1731–8.
28. Gandon Y, Olivie D, Guyader D, *et al.* Non-invasive assessment of hepatic iron stores by MRI. *Lancet*, 2004; **363**: 357–62.
29. Deugnier YM, Charalambous P, Le Quilleuc D, *et al.* Preneoplastic significance of hepatic iron-free foci in genetic hemochromatosis: a study of 185 patients. *Hepatology*, 1993; **18**: 1363–9.
30. Guyader D, Gandon Y, Sapey T, *et al.* Magnetic resonance iron-free nodules in genetic hemochromatosis. *Am J Gastroenterol*, 1999; **94**: 1083–6.
31. Bardou-Jacquet E, Laine F, Guggenbuhl P, *et al.* Worse outcomes of patients with HFE hemochromatosis with persistent increases in transferrin saturation during maintenance therapy. *Clin Gastroenterol Hepatol*, 2017; **15**: 1620–7.
32. Bardou-Jacquet E, Morcet J, Manet G, *et al.* Decreased cardiovascular and extrahepatic cancer-related mortality in treated patients with mild HFE hemochromatosis. *J Hepatol*, 2015; **62**: 682–9.
33. Angelopoulos NG, Goula A, Dimitriou E, Tolis G . Reversibility of hypogonadotropic hypogonadism in a patient with the juvenile form of hemochromatosis . *Fertil Steril*, 2005; **84**: 1744.
34. Hamer OW, Gnad M, Scholmerich J, Palitzsch KD . Successful treatment of erectile dysfunction and infertility by venesection in a patient with primary haemochromatosis . *Eur J Gastroenterol Hepatol*, 2001; **13**: 985–8.
35. Phatak P, Brissot P, Wurster M, *et al.* A phase 1/2, dose-escalation trial of deferasirox for the treatment of iron overload in HFE-related hereditary hemochromatosis. *Hepatology*, 2010; **52**: 1671–779.
36. Casu C, Nemeth E, Rivella S . Hepcidin agonists as therapeutic tools . *Blood*, 2018; **131**: 1790–4.
37. Kaltwasser JP, Werner E, Schalk K, Hansen C, Gottschalk R, Seidl C . Clinical trial on the effect of regular tea drinking on iron accumulation in genetic haemochromatosis . *Gut*, 1998; **43**: 699–704.
38. Hutchinson C, Geissler CA, Powell JJ, Bomford A . Proton pump inhibitors suppress absorption of dietary non-haem iron in hereditary haemochromatosis . *Gut*, 2007; **56**: 1291–5.
39. Vanclooster A, van Deursen C, Jaspers R, Cassiman D, Koek G . proton pump inhibitors decrease phlebotomy need in HFE hemochromatosis: double-blind randomized placebo-controlled trial . *Gastroenterology*, 2017; **153**: 678–80.e2.
40. Bacon BR, Adams PC, Kowdley KV, Powell LW, Tavill AS, American Association for the Study of Liver Disease. Diagnosis and management of hemochromatosis: 2011 practice guideline by the American Association for the Study of Liver Diseases. *Hepatology*, 2011; **54**: 328–43.
41. Papanikolaou G, Samuels ME, Ludwig EH, *et al.* Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet*, 2004; **36**: 77–82.
42. Roetto A, Papanikolaou G, Politou M, *et al.* Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat Genet*, 2003; **33**: 21–2.

43. Camaschella C, Roetto A, Cali A, *et al*. The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. *Nat Genet*, 2000; **25**: 14–15.
44. Pietrangelo A. Ferroportin disease: pathogenesis, diagnosis and treatment . *Haematologica*, 2017; **102**: 1972–84.
45. Le Lan C, Mosser A, Ropert M, *et al*. Sex and acquired cofactors determine phenotypes of ferroportin disease. *Gastroenterology*, 2011; **140**: 1199–207.e1–2.
46. Kono S . Aceruloplasminemia: an update . *Int Rev Neurobiol*, 2013; **110**: 125–51.
47. Viprakasit V, Ekwattanakit S . Clinical classification, screening and diagnosis for thalassemia . *Hematol Oncol Clin North Am*, 2018; **32**: 193–211.
48. Deugnier Y, Loreal O, Carre F, *et al*. Increased body iron stores in elite road cyclists. *Med Sci Sports Exerc*, 2002; **34**: 876–80.
49. Deugnier Y, Turlin B, le Quilleuc D, *et al*. A reappraisal of hepatic siderosis in patients with end-stage cirrhosis: practical implications for the diagnosis of hemochromatosis. *Am J Surg Pathol*, 1997; **21**: 669–75.
50. Ludwig J, Hashimoto E, Porayko MK, Moyer TP, Baldus WP. Hemosiderosis in cirrhosis: a study of 447 native livers. *Gastroenterology*, 1997; **112**: 882–8.
51. Detivaud L, Nemeth E, Boudjema K, *et al*. Hepcidin levels in humans are correlated with hepatic iron stores, hemoglobin levels, and hepatic function. *Blood*, 2005; **106**: 746–8.
52. Deugnier Y, Bardou-Jacquet E, Laine F . Dysmetabolic iron overload syndrome (DIOS). *Presse medicale*, 2017; **46**(12 Pt 2): e306-e11.
53. Bozzini C, Girelli D, Olivieri O, *et al*. Prevalence of body iron excess in the metabolic syndrome. *Diabetes Care*, 2005; **28**: 2061–3.
54. Mendler MH, Turlin B, Moirand R, *et al*. Insulin resistance-associated hepatic iron overload. *Gastroenterology*, 1999; **117**: 1155–63.
55. Moirand R, Mortaji AM, Loreal O, Paillard F, Brissot P, Deugnier Y . A new syndrome of liver iron overload with normal transferrin saturation . *Lancet*, 1997; **349**: 95–7.
56. Adams LA, Crawford DH, Stuart K, *et al*. The impact of phlebotomy in nonalcoholic fatty liver disease: a prospective, randomized, controlled trial. *Hepatology*, 2015; **61**: 1555–64.
57. Laine F, Ruivard M, Loustaud-Ratti V, *et al*. Metabolic and hepatic effects of bloodletting in dysmetabolic iron overload syndrome: a randomized controlled study in 274 patients. *Hepatology*, 2017; **65**: 465–74.

14.3.5 The Porphyrias

Michael N. Badminton and Danja Schulenburg-Brand

Introduction

The porphyrias are metabolic diseases resulting from deficiency, or in one disease increased activity, of specific enzymes of haem biosynthesis [1]. Each of the eight types of porphyria is defined by association of characteristic clinical features with a specific pattern of overproduction of pathway intermediates which identifies the underlying enzymatic abnormality (Figure 14.3.5.1).

The porphyrias are subdivided clinically into those that are acute, characterized by acute neurovisceral attacks associated with the overproduction of the porphyrin precursor, 5-aminolaevulinic acid (ALA) and porphobilinogen (PBG), and those that are non-acute, in which attacks are absent and photocutaneous symptoms due to excess porphyrins are the main clinical manifestation. An

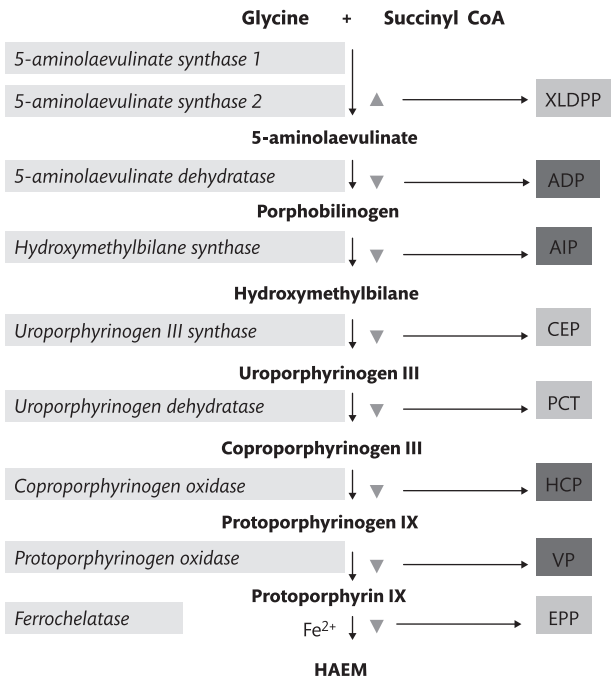


Figure 14.3.5.1 Haem biosynthesis pathway and enzyme abnormalities in the porphyrias. Enzymes (and genes) denoted in italics; ▲, increased enzyme activity; ▼, decreased enzyme activity. 5-Aminolaevulinate synthase 2 is expressed only in erythroid cells; inherited abnormal function of ubiquitous 5-aminolevulinate synthase 1, has not yet been identified in any disease. Dark box denotes acute porphyria. XLDPP, X-linked erythropoietic protoporphyria; AIP, acute intermittent porphyria; ADP, 5—aminolaevulinate dehydratase deficiency porphyria; CEP, congenital erythropoietic porphyria; PCT, porphyria cutanea tarda; HCP, hereditary coproporphyria; VP, variegate porphyria; EPP, erythropoietic protoporphyria.

alternative classification of erythropoietic or hepatic porphyria is based on the principal site of metabolite overproduction.

Acute Porphyrias–Autosomal Dominant

Clinically identical acute neurovisceral attacks occur in three AD porphyrias: acute intermittent porphyria (AIP; OMIM176000), hereditary coproporphyria (HCP; OMIM 121300), and variegate porphyria (VP; OMIM600923) [2]. Acute attacks can be life threatening if not recognized and appropriately treated. In most countries, AIP affects about 1 in 170 000 of the population, VP about 1 in 300 000, and HCP about 1 in 10^6. Acute neurovisceral attacks are the main clinical feature of AIP; photocutaneous symptoms do not occur. In VP, 40% of patients present with acute attacks, of whom half also have skin lesions, but 60% present with skin lesions alone. HCP presents with acute attacks that are accompanied by skin lesions in about 30% of patients; photocutaneous symptoms alone are rare and usually provoked by intercurrent liver disease. The skin lesions of VP and HCP are identical to those of bullous non-acute porphyrias.

Biochemical and Genetic Basis

The underlying defect in each of the AD acute porphyrias is a mutation leading to inactivation of one copy of the gene encoding the enzyme, and enzyme activities are thus half–normal in all tissues. Haem supply is maintained in the liver and other non-erythroid tissues by up-regulation of the rate controlling enzyme 5-aminolaevulinate synthase (ALAS1), with increases in substrate concentration of the affected enzyme. Compensatory change varies between tissues and between individuals with some showing no evidence of heme precursor overproduction while others have biochemically manifest disease with or without clinical symptoms.

Low clinical penetrance is a feature of AIP, HCP, and VP and many affected individuals remain asymptomatic throughout life. About 25% of patients with overt acute porphyria therefore have no family history. *De novo* mutations are uncommon. All three diseases show extensive allelic heterogeneity except in countries where founder effects occur. No clear genotype-phenotype correlation has been identified.

Pathogenesis of Acute Attacks

The symptomatology of acute attacks relates to dysfunction of autonomic, motor, and central nervous system neurons. The exact pathogenesis is not fully understood although ALA toxicity is currently the leading hypothesis [2].

Precipitating Factors

Factors implicated either individually or in combination include: hormonal fluctuations (particularly menstrual), certain prescribed and illicit drugs, excessive alcohol intake, calorie restriction, systemic illness, and stress. Attacks are more common in women, rare before puberty, and unusual after menopause. Attacks complicate less than 10% of pregnancies. A clear precipitant is not always identified [3].

Clinical Features

Pain is virtually always the initial symptom of an acute attack (95–100%). It is usually abdominal, but can affect the lower back, buttocks, and thighs [3]. It is frequently associated with nausea, vomiting, and constipation. Abdominal pain can mimic an acute abdomen, and may lead to inappropriate surgery. Diminishing pain in the absence of treatment can indicate worsening neuropathy. Hypertension and tachycardia due to autonomic dysfunction are common.

Hyponatraemia is common and can be severe, leading to seizures. Seizures may also be secondary to porphyric encephalopathy. Despite inappropriate urine sodium excretion patients are usually dehydrated and fluid restriction is not usually effective, implying renal sodium wasting rather than inappropriate antidiuretic hormone (ADH) secretion.

Symmetrical, distal motor neuropathy can occur, particularly where the attack is severe or treatment delayed. This can progress rapidly to complete motor paralysis mimicking Guillain–Barré syndrome. Other neurological signs include cranial nerve deficits and sensory changes in the same distribution as the motor neuropathy.

CNS involvement may cause confusion, anxiety, hallucinations, and paranoia. Chronic psychiatric illness is not a feature of acute porphyria.

Urine may appear 'port-wine red' due to the high content of porphyrin metabolites.

A small minority of patients, usually female, have repeated acute attacks. These may be premenstrual, occur as frequently as every month and are more likely in AIP and HCP than VP. Chronic complications include hypertension, impaired renal function, and hepatocellular carcinoma. Patients with active porphyria appear most at risk.

Laboratory Diagnosis

Measurement of urinary PBG excretion is the essential diagnostic investigation for an acute attack. Excretion of PBG and, to a lesser extent ALA is always increased during an attack, with PBG concentrations at least ten times the upper reference limit. PBG excretion decreases during remission, however, falling to normal over a period of years in AIP and a few weeks in HCP or VP. PBG excretion may be increased in adults with AIP who have never had symptoms. Although increased urinary PBG excretion makes it likely that symptoms are due to porphyria, the final diagnosis should be made on clinical grounds. Once the diagnosis of an attack of acute porphyria has been established, further investigation by analysis of plasma and faecal porphyrin identifies the type of porphyria (Table 14.3.5.1) [4]. If urine, plasma, and faecal porphyrins are normal, an acute attack is excluded as the cause of current or recent symptoms. Neither enzyme measurement nor mutational analysis are helpful for diagnosis of active porphyria in patients with suggestive symptoms [4].

Managing Acute Attacks

Unsafe drugs should immediately be withdrawn, and prompt treatment of possible precipitants with medication known to be safe begun [5]. Opiate analgesia is invariably needed and very large doses are usually required. Nausea and vomiting may be treated with an antiemetic (ondansetron, prochlorperazine) and convulsions with clonazepam, lorazepam, or magnesium sulphate. Hypertension and tachycardia should be treated with β-blockers. Intravenous fluids should include saline plus dextrose to provide calories without

Table 14.3.5.1 Inheritance and main clinical and diagnostic features of the porphyrias

Porphyria	Gene	Inheritance	Symptoms	Diagnosis
Acute intermittent porphyria	*HMBS*	autosomal dominant	acute attacks	urine PBG, ALA increased; normal faecal copro III[b]
Hereditary coproporphyria	*CPOX*	autosomal dominant	acute attacks and/or skin fragility, bullae	urine PBG, ALA increased[c]; faecal copro III increased[b]
Variegate porphyria	*PPOX*	autosomal dominant	skin fragility, bullae, and/or acute attacks	urine PBG, ALA increased[c]; plasma porphyrin peak at 624–628 nm; faecal proto increased
ALA dehydratase deficiency porphyria	*ALAD*	autosomal recessive	acute attacks	urine ALA and copro III increased; erythrocyte Zn-proto increased
Porphyria cutanea tarda	*UROD*	complex[a]	skin fragility, bullae	urine PBG, ALA normal with increased uro and hepta; faecal hepta, isocopro increased[b]
Congenital erythropoietic porphyria	*UROS*	autosomal recessive	skin fragility, bullae; haemolytic anaemia	urine PBG, ALA normal with increased uro I and copro I; faecal copro I increased; erythrocyte porphyrin increased[b]
Erythropoietic protoporphyria	*FECH*	autosomal recessive	acute painful photosensitivity	urine PBG, ALA, porphyrins normal; erythrocyte protoporphyrin increased; plasma porphyrin peak at 626–634 nm
X-linked erythropoietic protoporphyria	*ALAS2*	X-linked	acute painful photosensitivity	urine PBG, ALA, porphyrins normal; erythrocyte proto and Zn-proto increased; plasma porphyrin peak at 626–634 nm

[a] Acquired or, in about 20%, autosomal dominant; [b] plasma porphyrin peak at 615–622 nm may be present in AIP and HCP and is always present in PCT and CEP; [c] urine PBG may be normal in patients with skin lesions alone.

Copro, hepta, isocopro, proto, uro indicate coproporphyrin, heptacarboxylate porphyrin, isocoproporphyrin, protoporphyrin, uroporphyrin.

HMBS, hydroxymethylbilane synthase; CPOX, coproporphyrinogen oxidase; PPOX, protoporphyrinogen oxidase; ALAD, ALA dehydratase; UROD, uroporphyrinogen decarboxylase; UROS, uroporphyrinogen synthase; FECH, ferrochelatase; ALAS, ALA synthase

inducing hyponatraemia. Intravenous haematin is the recommended first line treatment [2, 4, 5], and is available as haem arginate (Normosang, Orphan Europe) in Europe. It is administered as a 30-minute infusion of 3 mg/kg daily on four consecutive days.

Recurrent acute attacks should prompt the seeking of advice from an expert porphyria centre [2, 5]. Options for treatment include suppression of ovulation using gonadorelin analogues [6] or regular administration of haematin. For the most severely affected patients where acute attacks become life threatening, venous access for ongoing treatment with haematin is inadequate and/or quality of life is severely reduced, liver transplantation may be considered [7]. Givosiran, a gene silencing therapy which targets hepatic ALAS1, was licensed for treating recurrent acute attacks by the US Food and Drug Administration Drugs (FDA) in November 2019 [8].

Preventing Acute Attacks

Patients should be advised about known precipitating factors and strategies to reduce the risk of acute attacks through patient information leaflets (e.g. available in 10 languages from the European Porphyria Network website, http://porphyria.eu/). Safe drug lists are available from several sources (e.g. the Welsh Medicines Information Centre; http://www.wmic.wales.nhs.uk). Patient support groups include the British Porphyria Association (http://www.porphyria.org.uk), and warning jewellery in case of emergencies should also be considered (e.g. MedicAlert). Finally, family members should be offered genetic testing so that presymptomatic relatives can also be advised on reducing the risk of acute attacks.

Acute Porphyrias—5-Aminolaevulinate Dehydratase Porphyria (ADP)

ADP (OMIM125270) is an exceedingly rare autosomal recessive porphyria that results from almost complete deficiency of 5-aminolaevulinate dehydratase (ALAD) activity with consequent overproduction of ALA. Only six cases have been reported [9]. Clinically it is indistinguishable from AIP and requires the same treatment. Diagnosis depends on demonstration of markedly increased urinary ALA excretion without an increase in PBG, increased erythrocyte zinc-protoporphyrin concentration, and very low erythrocyte ALAD activity.

Non-Acute Porphyrias

In the non-acute porphyrias, symptoms result from photosensitization of the skin by porphyrins activated by sunlight (UVA-visible range ~410 nm). The skin reacts to photo-damage in two different ways. First, accumulation of protoporphyrin in erythropoietic protoporphyria (EPP) and X-linked erythropoietic protoporphyria (XLEPP) causes acute painful photosensitivity. Second, accumulation of less hydrophobic porphyrins in porphyria cutanea tarda (PCT), congenital erythropoietic porphyria (CEP) VP and HCP, causes fragility of sun-exposed skin with erosions and subepidermal bullae. Painful photosensitivity is usually absent.

Porphyria Cutanea Tarda (PCT)

PCT is the most common type of porphyria with an annual incidence in the United Kingdom of between 2 and 5 per million.

Biochemical and Genetic Basis

The primary enzyme defect in PCT is decreased activity of liver uroporphyrinogen decarboxylase (UROD) leading to overproduction of uroporphyrin and other intermediate porphyrins. About 75% of patients have an acquired PCT (OMIM176090) usually associated with liver disease. The *UROD* gene is normal

and there is usually no family history of PCT. The remainder have familial PCT (OMIM176100) in which half–normal UROD activity is caused by mutations in the *UROD* gene, inherited in an autosomal dominant pattern with low penetrance. Further inactivation of UROD by the same process involved in sporadic PCT is required. Hepatic UROD is inactivated by an inhibitor produced by iron-dependent oxidation of uroporphyrinogen, the substrate of the UROD reaction [10].

Clinical Features

PCT occurs at all ages in both sexes with onset usually during the 5th–6th decades. Familial PCT tends to occur at younger ages. Patients have fragile sun-exposed skin. Minor trauma results in erosions as well as vesicle and bullae formation, particularly on backs of the hands, forearms, and face. Lesions are slow to heal, leaving atrophic scars, milia, and patchy depigmentation. Hyperpigmentation and melanosis may also develop and facial hypertrichosis is common. Alopecia may develop at sites of repeated trauma or bullae.

Skin lesions are often the first sign of underlying hepatocyte damage. Overt liver disease is uncommon, but minor alterations in biochemical liver function tests are common. The majority have hepatic iron overload with increased transferrin saturation and ferritin. Cirrhosis is present in <15% of patients, but confers high risk of hepatocellular carcinoma.

Associated Conditions

This combination of skin lesions and liver damage is strongly associated with excessive alcohol intake, oestrogen use, hepatitis C, or HIV infection, or mutations in the hemochromatosis (*HFE*) gene [11]. PCT may also occur in association with other disorders, notably chronic renal failure, systemic lupus erythematosus, and haematological malignancies.

Treatment

Avoidable risk factors such as alcohol or oestrogens should be identified, and underlying associated disorders sought and treated [12]. Exposure to light should be reduced by use of suitable clothing and reflectant sunscreens protecting against UVA/visible light. Either of two specific treatments produce clinical and biochemical remission in most patients in 6 to 12 months, and can be monitored by urinary total porphyrin measurement. The first is reduction of iron stores by phlebotomy (450 ml, 1 to 2 weekly) until borderline iron deficient is achieved (assessed by haemoglobin, plasma transferrin saturation, and ferritin measurement). This is preferred in patient with genetic haemochromatosis. The second option is low dose intermittent oral chloroquine (125 mg) or hydroxychloroquine (100 mg) taken twice weekly. Antimalarial doses can provoke a severe hepatotoxic reaction.

Congenital Erythropoietic Porphyria (CEP)

CEP or Gunther disease (OMIM236700) is an autosomal recessive severe bullous porphyria that normally presents in infancy [13, 14]. It affects <1:1 000 000 of the UK population.

Biochemical and Genetic Basis

Mutations in the uroporphyrinogen synthase (*UROS)* gene, or, rarely, the *GATA1* gene, decreases UROS activity resulting in massive overproduction of predominantly bone marrow uroporphyrin-I. Porphyrin-laden erythroid cells have a shortened life span leading to haemolytic anaemia, porphyrin release into plasma, and ineffective erythropoiesis. There is some correlation between genotype and severity of disease.

Clinical Features

Clinical severity varies from non-immune hydrops fetalis to a mild PCT-like syndrome in young adults, but the majority of patients present in infancy with red urine, skin lesions, and haemolytic anaemia. Skin lesions are similar to those of PCT but more persistent and severe. Progression to severe scarring with photo-mutilation is common. Haemolytic anaemia may worsen porphyrin overproduction and photosensitivity; splenomegaly is common. Porphyrin accumulates in the bones and is visible in the teeth as erythrodontia (brown pigmentation which fluoresces under ultraviolet light). Expansion of hyperactive bone marrow may result in pathological fractures, vertebral compression, or collapse, shortness of stature, and rarely osteolytic and sclerotic lesions in the skeleton.

Treatment

Protection against sunlight and prevention of skin infections are essential. In addition to reflectant sunscreen ointments, rigorous physical avoidance of sunlight is usually necessary. Haemolytic anaemia may require repeated transfusion and iron chelation to prevent iron overload, but splenectomy is rarely effective. Bone involvement may require bisphosphonate treatment and vitamin D supplementation. Allogeneic haematopoietic stem cell transplantation is curative, while gene therapy for those without donors or otherwise unsuitable for transplantation is under development.

Erythropoietic Protoporphyria (EPP)

EPP (OMIM 177000) is the more common of two porphyrias characterized by acute painful photosensitivity resulting from accumulation of protoporphyrin IX in the skin. The prevalence of EPP in Western Europe is one per 75 000 to 130 000.

Biochemical and Genetic Basis

Deficiency of ferrochelatase (FECH) activity leads to accumulation of protoporphyrin IX in skin, liver, and erythroid cells. In >90% of families, patients are compound heterozygotes for a *FECH* mutation that markedly decreases activity on one allele and a hypomorphic variant (*FECH* IVS3-48C) [15]. These reduce FECH activity below the threshold at which protoporphyrin accumulates. The population prevalence of EPP relates directly to the frequency of the hypomorphic allele which ranges from 45% in Japan to less than 1% in West Africa. The frequency in Europe is 7–11%, sufficiently high for some families to show pseudodominant inheritance. About 4% of EPP families have deleterious FECH mutations on both alleles.

Clinical Features

Symptoms normally start between birth and 6 years in both sexes, with median age of onset 1 year [16]. Exclusively photocutaneous symptoms occur in light-exposed areas such as the face and hands. Within 1 hour of sun exposure, stinging or painful burning sensations occur in skin, usually followed several hours later by erythema and oedema although in some there may be no objective signs of phototoxicity. Petechiae, purpura, vesicles, and crusting

may develop, and persist for several days. Recurrent episodes lead to chronic changes; typically, shallow linear scars over the nose and face with thickened waxy skin, especially over the knuckles. Symptoms tend to be more severe during spring and summer and may improve during pregnancy.

Protoporphyrin is hepatotoxic. About 20% of patients have abnormal biochemical liver function tests and 2% to 5% develop liver failure [17]. EPP may also increase the risk of cholelithiasis. Erythropoiesis is impaired in all patients with a lower haemoglobin concentration resulting in a mild microcytic anaemia. Fifty per cent (50%) of patients are vitamin D deficient and at risk of osteoporosis [18].

Treatment

Acute photosensitivity can be controlled by avoidance of sunlight, suitable clothing, and reflectant sunscreens. A subcutaneous implant, afamelanotide (α-melanocyte-stimulating hormone analogue) has recently been licensed. It increases the time during which patients are pain free in direct sunlight with improvement in quality of life reported [19]. Other options include production of a photoprotectant tan by narrowband ultraviolet B (UVB) phototherapy or oral β-carotene, which acts as a singlet oxygen quencher. Liver function and vitamin D status should be monitoring at least annually, and orthotopic liver transplantation is the only treatment for acute liver failure. Patients and families may request genetic counselling to assess the risk that offspring will be affected. Testing the unaffected partner for the presence of the hypomorphic *FECH* IVS3-48C allele is helpful.

X-Linked Erythropoietic Protoporphyria (XLEPP)

About 2% of protoporphyria families have XLEPP (OMIM 300752) [20]. Protoporphyrin accumulation is secondary to increased erythroid 5-aminolaevulinate synthase (ALAS2) activity caused by gain-of-function mutations in the *ALAS2* gene. This X-linked condition is expressed in males and most females and may carry a higher risk of severe liver disease than EPP.

Laboratory Diagnosis of Non-Acute Porphyrias

Skin biopsy is not required to diagnose cutaneous porphyria (Table 14.3.5.1). Active skin lesions are always accompanied by increased circulating plasma porphyrins. The first line test is therefore a plasma porphyrin screen, which if negative, excludes active cutaneous porphyria. Specialist porphyrin biochemistry can readily distinguish between the different non-acute porphyrias provided appropriate samples are analysed (Table 14.3.5.1). Mutational analysis of the *UROD* gene to distinguish familial from sporadic PCT is not necessary for clinical management. Identification of *UROS* mutations may help to assess prognosis in CEP.

REFERENCES

1. Badminton MN, Elder GH . The porphyrias: inherited disorders of haem synthesis. In: Marshall W, Lapsley M, Day AP, Ayling RM, eds. *Clinical Biochemistry: Metabolic and Clinical Aspects,* 3rd edn. Edinburgh: Elsevier Ltd, 2014: 533–49.
2. Stein PE, Badminton MN, Rees DC. Update review of the acute porphyrias. *Br J Haematol,* 2017; **176**: 527–38.
3. Hift RJ, Meissner PN. An analysis of 112 acute porphyric attacks in Cape Town, South Africa. *Medicine*, 2005; **84**: 48–60.
4. Whatley SD, Mason NG, Woolf JR, Newcombe RG, Elder GH, Badminton MN . Diagnostic strategies for autosomal dominant acute porphyrias: retrospective analysis of 467 unrelated patients referred for mutational analysis of the *HMBS, CPOX,* or *PPOX* gene . *Clin Chem*, 2009; **55**: 1406–14.
5. Stein, P, Badminton, M, Barth J, Rees D, Stewart MF. Best practice guidelines on clinical management of acute attacks of porphyria and their complications . *Ann Clin Biochem,* 2013; **50**: 217–23.
6. Schulenburg-Brand D, Gardiner T, Guppy S, *et al*. An audit of the use of gonadorelin analogues to prevent recurrent acute symptoms in patients with acute porphyria in the United Kingdom. *JIMD Rep*, 2017; **36**: 99–107
7. Soonawalla ZF, Orug T, Badminton MN, *et al*. Liver transplantation as a cure for acute intermittent porphyria. *Lancet*, 2004; **363**: 705–6.
8. Balwani M, Sardh E, Ventura P, *et al*. Phase 3 trial of RNAi therapeutic givosiran for acute intermittent porphyria. *N Engl J Med*, 2020; **382**: 2289–301.
9. Doss MO, Stauch T, Gross U, *et al*. The third case of Doss porphyria (delta-amino-levulinic acid dehydratase deficiency) in Germany. *J Inherit Metab Dis*, 2004; **27**: 529–36.
10. Phillips JD, Bergonia HA, Reilly CA, Franklin MR, Kushner JP. A porphomethene inhibitor of uroporphyrinogen decarboxylase causes porphyria cutanea tarda . *Proc Natl Acad Sci U S A*, 2007; **104**: 5079–84.
11. Ryan Caballes F1, Sendi H, Bonkovsky HL . Hepatitis C, porphyria cutanea tarda and liver iron: an update. *Liver Int*, 2012; **32**: 880–93.
12. Sarkany, RE . The management of porphyria cutanea tarda . *Clin Exp Dermatol*, 2001; **26**: 225–32.
13. Fritsch C, Bolsen K, Ruzicka T, Goerz G . Congenital erythropoietic porphyria . *J Am Acad Dermatol*, 1997; **36**: 594–610.
14. Katugampola RP, Anstey AV, Finlay AY, *et al*. A management algorithm for congenital erythropoietic porphyria derived from a study of 29 cases. *Brit J Dermatol*, 2012; **167**: 888–900.
15. Gouya L, Martin-Schmitt C, Robreau A-M, *et al*. Contribution of a single-nucleotide polymorphism to the genetic predisposition for erythropoietic protoporphyria. *Am J Hum Genet*, 2006; **78**: 2–14.
16. Holme SA, Anstey AV, Finlay AY, Elder GH, Badminton MN . Erythropoietic protoporphyria in the United Kingdom: clinical features and effect on quality of life . *Br J Dermatol*, 2006; **155**: 574–81.
17. Anstey AV, Hift RJ . Liver disease in erythropoietic protoporphyria: insights and implications for management. *Gut*, 2007; **56**: 1009–18.
18. Biewenga M, Matawlie RHS, Friesema ECH, *et al*. Osteoporosis in patients with erythropoietic protoporphyria. *Br J Dermatol*, 2017; **177**: 1693–8.
19. Langendonk JG, Balwani M, Anderson KE *et al*. Afamelanotide for erythropoietic protoporphyria. *N Engl J Med*, 2015; **373**: 48–59.
20. Whatley SD, Ducamp S, Gouya L, *et al*. C-Terminal deletions in the ALAS2 gene lead to gain of function and cause a previously undefined type of human porphyria, X-linked dominant protoporphyria, without anemia or iron overload. *Am J Human Genet*, 2008; **83**: 408–14.

SECTION 15

Diabetes Mellitus

15.1

Introduction to Diabetes Mellitus

15.1.1 Physiology of Glucose Homeostasis

Shanta J. Persaud and Peter M. Jones

Islets of Langerhans

The pancreas possesses both exocrine and endocrine functions. The acinar cells of the exocrine pancreas produce digestive enzymes and account for over 97% of the pancreatic mass. The endocrine function is performed by the islets of Langerhans, which are spherical clusters of 1000–3000 cells scattered throughout the pancreas. A human pancreas contains up to 1 million individual islets, each approximately 0.2 mm in diameter and composed of three main cell types: insulin-producing beta-cells, glucagon-secreting alpha-cells, and somatostatin-producing delta-cells, which constitute ~60%, 35%, and 5%, respectively, of a typical human islet. The islets are penetrated by blood capillaries into which the hormones are secreted.

Insulin Biosynthesis and Storage

Human insulin consists of 51 amino acids that are arranged into two polypeptide chains. The A chain contains 21 amino acids and is linked to the 30 amino acid B chain by two disulphide bonds. The biosynthesis of insulin takes place in the beta-cell rough endoplasmic reticulum and occurs via two intermediate single chain polypeptides: preproinsulin, which is cleaved to generate proinsulin. Proinsulin contains the A and B chains of insulin linked by a connecting peptide (C-peptide) that ensures that the A and B chains are correctly orientated within the proinsulin molecule and that the disulphide bonds are aligned. Proinsulin is transported from the endoplasmic reticulum to the Golgi apparatus and packaged into secretory granules, where C-peptide is removed by enzymatic cleavage at specific amino acid residues. It is in these granules that the newly synthesized insulin forms insoluble crystalline hexamers by binding of insulin B-chain histidine residues to granule zinc ions (see **Figure 15.1.1.1**). C-peptide is also stored in the secretory granules, and it is released in equimolar amounts with insulin when the appropriate signal from the extracellular environment causes fusion of the granules with the beta-cell plasma membrane. C-peptide has been traditionally regarded as an inert molecule, useful only as an indicator of residual β-cell function in patients receiving insulin injections. However, it has been suggested that C-peptide exerts significant physiological effects such as increased glucose disposal, improved metabolic control and amelioration of diabetic nephropathy [1], which are thought to be mediated by activation of the G-protein-coupled receptor GPR146 [2].

The amino acid sequence of insulin is very similar among a range of mammalian species. Thus, porcine and human insulin differ by only one amino acid, and bovine insulin by three, which meant that insulin extracted from pig and cow pancreases was used to treat

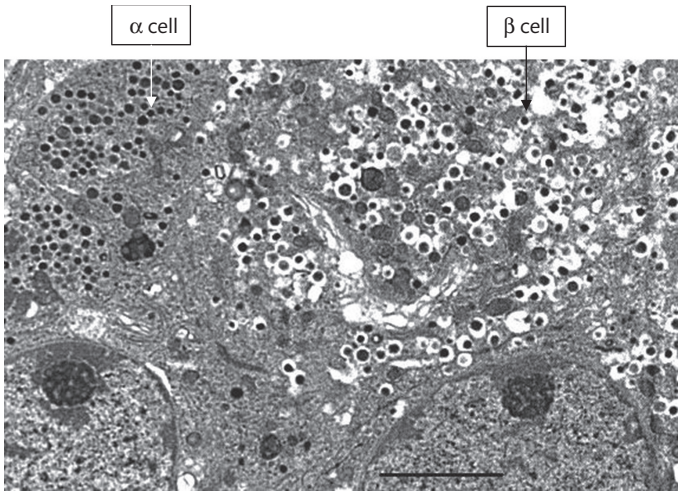


Figure 15.1.1.1 Islet ultrastructure. Transmission electron micrograph showing an islet beta-cell with typical dense core secretory granules and electrotranslucent halos. The secretory granules in the neighbouring alpha-cell have very dense cores and no halos. The scale bar is 2 μm.

diabetes until production of human insulin by recombinant DNA technology in the 1980s.

Nutrient-Induced Insulin Secretion

Glucose

Glucose is the major physiological regulator of insulin secretion in mammals. In humans beta-cells are adapted to secrete insulin only when circulating glucose concentrations are elevated above 4 mM so that glucose is tightly regulated in the range of ~4–7 mM [3]. The beta-cell responses are both very sensitive and rapid, usually within 30 seconds of the arrival of a glucose stimulus. Insulin is secreted by a process known as exocytosis, whereby the insulin-containing secretory granules fuse with the beta-cell plasma membrane, thus releasing the granule contents into the circulation. It occurs in a biphasic manner in response to glucose, with an initial spike of insulin release within 1–2 min, followed by a plateau phase that persists for as long as the stimulus is maintained. Islet beta-cells are well-equipped for immediate and sustained insulin release. Insulin is approximately 10% of the total beta-cell protein and there are in excess of 10 000 secretory granules in each beta-cell, sufficient for a 24-hour supply of insulin in the absence of any additional synthesis.

Glucose is taken up into beta-cells by glucose transporters (GLUTs), and the first stage in its metabolism is phosphorylation by a high specificity glucokinase. Subsequent metabolic events result in an increase in the cellular ATP/ADP ratio, which leads to closure of ATP-sensitive K^+-channels (K_{ATP}-channels) in the beta-cell plasma membrane (see Figure 15.1.1.2). The ensuing inhibition of K^+ efflux from the beta-cells results in changes in cellular electrical activity (depolarization) that leads to opening of specialized voltage-gated Ca^{2+} channels (VGCCs) that allow Ca^{2+} entry down a steep concentration gradient. The increase in beta-cell Ca^{2+} levels in response to glucose leads to activation of Ca^{2+}-sensitive proteins that are involved in the exocytotic fusion of insulin secretory granules. In addition, Ca^{2+}-dependent enzymes, such as the Ca^{2+}/calmodulin-dependent protein kinases (CaMKs), are activated when elevations in intracellular Ca^{2+} ($[Ca^{2+}]_i$) are detected by the Ca^{2+}-sensing protein calmodulin (CaM). This leads to the formation of Ca^{2+}/CaM complexes and activation of CaMKs to regulate Ca^{2+}-initiated cellular responses. CaMK II, in particular, is expressed by beta-cells and its activation has been implicated in glucose-induced insulin secretion [4].

The Ca^{2+}-sensitive enzyme phospholipase C (PLC) is also activated by glucose, resulting in hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) to generate two second messenger molecules, diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). DAG activates protein kinase C to phosphorylate proteins, including a subunit of the K_{ATP}-channel and VGCCs, both of which are important in beta-cell secretory responses to glucose. However, glucose-stimulated insulin release is not dependent on PKC activation, but DAG-sensitive PKC isoforms are required for potentiation of glucose-stimulated insulin secretion by some agonists such as the cholinergic neurotransmitter acetylcholine (ACh). The other second messenger produced by PLC-stimulated PIP_2 hydrolysis, IP_3, binds to specific receptors on the endoplasmic reticulum to stimulate release of stored Ca^{2+}.

Glucose also activates adenylate cyclase in beta-cells, most likely as a consequence of elevations in Ca^{2+}, to increase cyclic AMP generation from ATP. Cyclic AMP potentiates both phases of glucose-induced

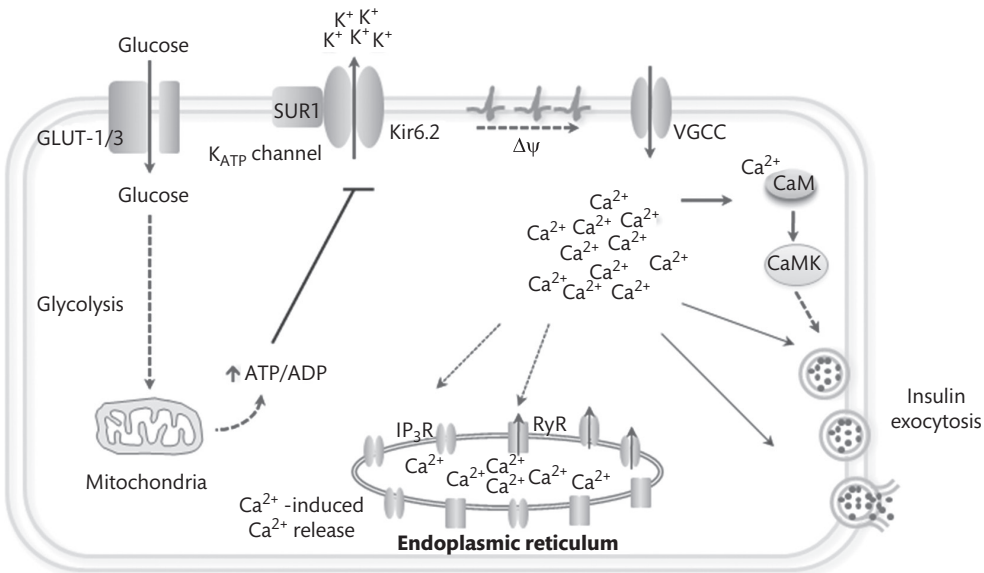


Figure 15.1.1.2 Islet beta-cell stimulus-secretion coupling pathways. Glucose is taken up by human beta-cells via GLUT-1 and GLUT-3 transporters and it is phosphorylated to glucose-6-phosphate by glucokinase. The end product of glycolysis, pyruvate, is transported to the mitochondria where ATP is produced upon pyruvate oxidation through the TCA cycle. The increase in ATP generation inhibits activity of K_{ATP}-channels, which consist of an ATP-sensitive sulphonylurea receptor (SUR1) and an inward rectifier K^+ channel (Kir6.2). ATP-induced closure of K_{ATP}-channels results in decreased K^+ efflux, beta-cell depolarization, and the opening of voltage-gated Ca^{2+} channels (VGCCs). Influx of Ca^{2+} through the VGCCs leads to a rapid increase in $[Ca^{2+}]_i$ which is further enhanced by Ca^{2+}-induced Ca^{2+} release through the opening of inositol 1,4,5-trisphosphate receptors (IP_3Rs) and ryanodine receptors (RyRs) on the endoplasmic reticulum. Cytosolic Ca^{2+} activates Ca^{2+}-sensing kinases (such as CaMKs) and Ca^{2+}-sensitive proteins of the exocytotic machinery, which facilitate release of insulin from secretory granules.

insulin secretion, exerting its effects via protein kinase (PKA) and cyclic AMP-activated GTP-exchange factors (Epacs). Once activated by cyclic AMP, PKA phosphorylates beta-cell proteins, including VGCCs and IP_3 receptors, on serine and threonine residues, which causes sustained increases in $[Ca^{2+}]_i$ during glucose-induced insulin secretion. Cyclic AMP activation of Epacs can inhibit beta-cell K_{ATP}-channels, resulting in decreased K^+ efflux, increased $[Ca^{2+}]_i$ and insulin secretion. One of the Epac proteins (Epac2) associates with two other proteins, Rim and Piccolo, which allow the docking and fusion steps of insulin exocytosis. Agonists that selectively activate Epac2 have been developed, and they are able to potentiate glucose-induced insulin secretion from human islets [5], confirming the importance of this cyclic AMP-activated signalling cascade in physiological human beta-cell function.

The precise mechanisms responsible for the first and second phases of glucose-stimulated insulin secretion have not been fully identified. However, it is thought that the first phase involves fusion of 'primed' insulin secretory granules that are present in a readily releasable pool close to the plasma membrane site of Ca^{2+} influx [6]. In contrast, the second phase of insulin release is a more sustained process that occurs from the reserve granule pool and depends on granule maturation and translocation along directed pathways involving microtubule and actin filaments.

Amino Acids

Most amino acids only stimulate insulin secretion in the presence of stimulatory concentrations of glucose, but leucine, lysine, and arginine act as insulin secretagogues in the absence of glucose. Leucine is transported into beta-cells by a sodium-independent transporter and stimulates biphasic insulin release, similar to that induced by glucose. Its metabolism leads to closure of K_{ATP}-channels with the consequent series of events that triggers insulin secretion in response to glucose, as described for glucose. Lysine and arginine are charged amino acids that enter β-cells via a cationic amino acid transporter and stimulate Ca^{2+} influx via VGCCs following membrane depolarization. Alanine, a neutral amino acid, has weak stimulatory effects on insulin secretion. It enters β-cells mainly through a sodium-dependent transport system, to elicit a small depolarization of the β-cell plasma membrane and elevation in $[Ca^{2+}]_i$.

Fatty Acids

Dietary free fatty acids (FFAs) stimulate insulin secretion in the absence of glucose and also potentiate the stimulatory effect of glucose on insulin release. A family of cell surface G-protein coupled FFAR receptors are known to be important in mediating FFA effects in β-cells. In particular, FFAR2 has been implicated in the secretagogue effects of the short chain fatty acid propionate in humans [7] while agonists that target the long chain fatty acid receptor FFAR1 are candidates for novel therapies for type 2 diabetes [8].

Regulation of Insulin Secretion by Non-Nutrient Secretagogues

Parasympathetic and Sympathetic Innervation

Parasympathetic stimulation leads to the release of acetylcholine (ACh) from cholinergic nerve terminals found close to all islet cell types, both at the periphery and deep within the islet. ACh binds to and activates cholinergic muscarinic receptors on β-cells, leading to potentiation of glucose-stimulated insulin release. This occurs via activation of PLC, which generates IP_3 and DAG, and PKC activation by DAG is important in a sustained potentiation of insulin release by cholinergic agonists. The main function of islet cholinergic innervations is potentiation of insulin secretion in response to the cephalic phase of food intake, which is triggered by the sight, smell, or taste of food.

Sympathetic neural activity has a direct effect through the release of the neurotransmitter noradrenaline (NA) from nerve terminals within islets. NA inhibits glucose-induced insulin secretion by activating α_2-adrenergic receptors, which are coupled to inhibition of adenylate cyclase activity resulting in reductions in cyclic AMP levels. NA also promotes activation of beta-cell K_{ATP}-channels and inhibition of VGCCs, resulting in decreased $[Ca^{2+}]_i$ that contributes to the reduction in insulin secretion. The related catecholamine, adrenaline, which is secreted from the adrenal medulla under conditions of stress, also inhibits insulin secretion and this contributes to the increased availability of glucose as an energy source for 'fight or flight' activities.

Islet Hormones

Hormones released from neighbouring alpha- or delta-cells, or from the beta-cells themselves, can bind to specific receptors on beta-cells to modify beta-cell function in a paracrine or autocrine fashion.

Glucagon is synthesized in islet alpha-cells as a precursor peptide, proglucagon, which is enzymatically cleaved to functional glucagon. Like insulin, glucagon is packaged into secretory granules within alpha-cells, and its secretion is regulated by changes in circulating glucose concentrations. However, while increased blood glucose stimulates exocytotic release of insulin, it inhibits glucagon secretion. The most important stimuli for glucagon release are hypoglycaemia and sympathetic nervous input, and when it is released from alpha-cells glucagon stimulates glycogenolysis and gluconeogenesis, thus restoring blood glucose levels. It also acts locally at neighbouring β-cells to stimulate insulin secretion via Gs-coupled receptors, which activate adenylate cyclase to increase cyclic AMP and PKA activation.

Somatostatin (SST) is synthesized and secreted by islet delta-cells in response to elevations in blood glucose levels. It binds to G-protein-coupled SST receptors on β-cells to act as a local inhibitor of insulin release via Gi-coupled inhibition of adenylate cyclase. SST also hyperpolarizes beta-cells through the opening of K_{ATP}-channels, which decreases $[Ca^{2+}]_i$ and contributes to its inhibitory effects. As well as being released from islets, SST is also secreted from gastrointestinal cells after food intake, but the circulating levels are insufficient to significantly affect insulin secretion. Thus, SST appears to inhibit insulin release mainly by a paracrine rather than a circulatory route.

Insulin receptors and downstream signalling elements are expressed by human islet beta-cells, suggesting that when insulin is released it may interact with these receptors to modify beta-cell function in an autocrine or paracrine manner [9]. There have been reports of insulin exerting positive or negative feedback on further insulin output from beta-cells, but the general consensus is that insulin does not regulate its own secretion. The most likely role for

insulin on β-cells is the regulation of gene expression and maintenance of beta-cell mass through stimulation of proliferation and/or inhibition of apoptosis.

Incretins

It has been known for over 50 years that islets are more responsive to oral rather than intravenous administration of glucose [10]—the incretin effect. This is due to the release of the gastrointestinal peptides GLP-1 and glucose-dependent insulinotropic peptide (GIP), which enhance the insulin secretory response to glucose by up to 60%. In the fasting state plasma levels of GIP and GLP-1 are low and they are released from specialized gut endocrine cells into the bloodstream after food intake.

GLP-1 is a 30 amino acid peptide that is synthesized by alternative splicing of proglucagon in L-cells of the ileum and colon and released in response to nutrient absorption in the gastrointestinal tract [11]. The major biologically active form of GLP-1 is GLP-1 (7–36) amide, which is the predominant form in plasma. GLP-1 binds to its Gs-coupled receptors on islet beta-cells to stimulate insulin release through elevations in intracellular cyclic AMP. In addition, GLP-1 also regulates glucose homeostasis through effects at α-cells to inhibit glucagon secretion and it reduces food intake by direct anorexigenic central effects. In addition, it is reported to stimulate beta-cell mass in rodents, but similar beneficial effects have not yet been identified in humans.

These properties have made GLP-1 an attractive therapy for type 2 diabetes, but its actions *in vivo* are rapidly curtailed through its degradation by dipeptidyl protease 4 (DPP-4), resulting in a half-life in plasma of only approximately two minutes. Pharmaceutical agents that improve the longevity of GLP-1 action through modification of its sequence have therefore been developed and are now in clinical use, as are therapies that inhibit DPP4-mediated GLP-1 degradation.

GIP, originally known as gastric inhibitory polypeptide due to its inhibitory effects on gastric acid secretion, is a 42-amino acid peptide that is synthesized from a precursor peptide (proGIP) in K-cells of the duodenum and jejunum. Circulating levels of GIP increase in response to nutrient absorption and it is carried in the circulation to islets where it potentiates glucose-induced insulin secretion through action at beta-cell GIP receptors to elevate cyclic AMP. As for GLP-1, the activity of GIP is terminated by DPP-4-mediated degradation. In contrast to GLP-1, however, GIP is unlikely to be developed as an antidiabetes agent because in type 2 diabetes GIP levels are normal, but it has a reduced effect to stimulate insulin secretion. In addition, GIP does not have the additional beneficial effects of GLP-1 such as inhibition of glucagon secretion. Nonetheless, DPP-4 inhibitor therapies that have been developed to increase the half-life of GLP-1 will also maintain plasma GIP levels.

Regulation of Plasma Glucose Levels

Insulin stimulates storage of fuels in adipose tissue, liver, and skeletal muscle, thus providing readily available depots of glycogen and fat for energy requirements during periods where there is no food intake. Thus, in healthy individuals blood glucose levels are maintained in the range of 3–4 mM under fasting conditions and 6–7 mM postprandially. Following food intake the elevation in blood glucose levels leads to increased glucose transport into beta-cells with a resulting increase in insulin output, as indicated in Figure 15.1.1.2. Insulin is transported in the blood to target tissues where it binds to cell surface insulin receptors. Skeletal muscle, one of the major sites for glucose storage in the form of glycogen, is responsible for up to 75% of glucose uptake, and insulin stimulates this by promoting translocation of GLUT4 glucose transporters to the myocyte cell surface. In addition, insulin also stimulates glucose uptake via GLUT4 in adipocytes, while glucose transport into hepatocytes is mediated by insulin-independent GLUT2 transporters. Insulin exerts its effects by binding to a heterotetrameric receptor complex that consists of two extracellular α subunits and two transmembrane beta-subunits. When insulin binds to the alpha-subunits it causes a conformational change that results in activation of the beta-subunit tyrosine kinase domain and receptor autophosphorylation. This fully activates the insulin receptor tyrosine kinase, which then phosphorylates intracellular substrates on multiple tyrosine residues [12].

The main proteins responsible for signalling downstream of insulin binding to its receptors are the insulin receptor substrate (IRS) family members, IRS-1 and IRS-2, which act as adapter proteins following insulin-induced tyrosine phosphorylation. Global deletion of IRS-1 in mice indicated its importance in intrauterine and postnatal growth, but insulin metabolic signalling was only impaired rather than abolished, and this led to the discovery of IRS-2. IRS-2 deletion in mice results in type 2 diabetes because it plays important roles in maintenance of β-cell mass as well transducing insulin action in peripheral tissues [13].

The pathways downstream of IRS phosphorylation are now reasonably well understood [12]. Thus, effectors containing src homology (SH2) domains bind to phosphotyrosine residues on IRS and are recruited into insulin receptor/IRS signalling complexes. Following its interaction with phosphorylated IRS the 85 kDa regulatory subunit of class I phosphatidylinositol 3 kinases (PI3K) binds to and activates the 110kDa PI3K catalytic subunit leading to phosphorylation of phosphatidylinositols on position 3 of the inositol ring. The resulting generation of phosphatidylinositol-3,4,5-trisphosphate (PI-3,4,5-P_3) recruits phosphoinositide-dependent kinases (PDKs) and protein kinase B (PKB, also known as Akt) to the plasma membrane site of PI-3,4,5-P_3 synthesis via their Pleckstrin homology domains. PDKs phosphorylate and activate PKB, which then phosphorylates substrates on serine and/or threonine residues to bring about the metabolic effects of insulin. In this way, GLUT4 transporters are translocated to the plasma membrane of adipocytes and myocytes, glycogen and fatty acid synthesis are stimulated, and gluconeogenesis and fatty acid oxidation are inhibited. The growth-promoting effects of insulin are largely mediated by mitogen-activated protein kinases (MAPKs), which, like PI3K, are downstream of IRS tyrosine phosphorylation. In this pathway phosphorylated IRS proteins bind to the SH2 domain of growth factor receptor binding protein 2 (Grb2), which forms a complex with son-of-sevenless to promote GTP binding to Ras and activation of the MAPK cascade. When MAPK is phosphorylated on tyrosine and threonine residues it translocates into the nucleus and phosphorylates transcription factors that stimulate cell proliferation.

The anabolic actions of insulin are counteracted by glucagon, which is released from islet alpha-cells when blood glucose levels fall. Glucagon acts at specific G-protein-coupled receptors on

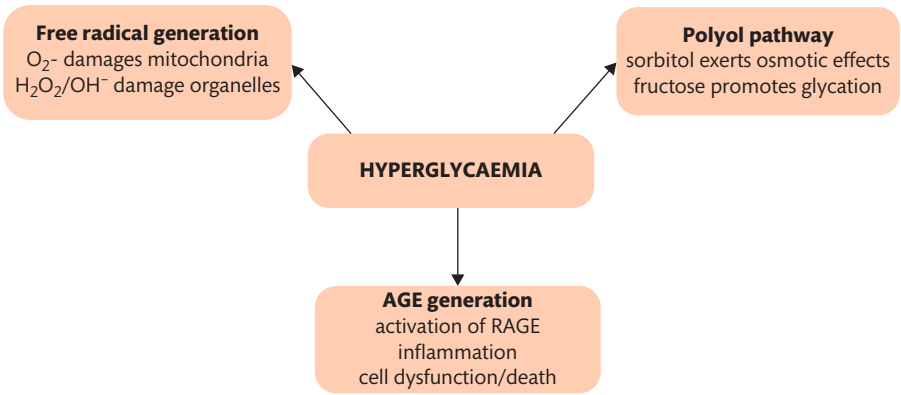


Figure 15.1.1.3 Consequences of chronic hyperglycaemia. In untreated or poorly controlled diabetes glucose levels are high in tissues that do not require insulin for glucose uptake. Deleterious effects arise in these tissues through enhanced free radical generation, overactivation of the polyol pathway, and increased production of advanced glycation end-products, all of which contribute to cell dysfunction and death.

hepatocytes to stimulate gluconeogenesis and glycogenolysis, to restore glycaemia. The glucagon receptors are coupled to adenylate cyclase via Gs, resulting in increased cyclic AMP generation, PKA activation, and phosphorylation of effectors such as phosphorylase kinase, which promotes glycogen breakdown. Thus, under normal conditions the relative contributions of insulin in the absorptive state (after food intake) and glucagon in the post-absorptive state (when the gastrointestinal tract is empty) ensure that blood glucose levels are tightly regulated.

Consequences of Chronic Hyperglycaemia in Diabetes

The importance of maintaining good blood glucose regulation to prevent or slow the onset of diabetic complications was defined in two landmark studies: the Diabetes Control and Complications Trial for type 1 diabetes [14], and the UK Prospective Diabetes Study for type 2 diabetes [15]. It is clear that prolonged exposure to elevated glucose concentrations damages tissues through metabolic changes that are mostly related to increased reactive oxygen species (ROS) generation, enhanced polyol pathway activity, and protein glycation, as summarized in Figure 15.1.1.3.

Hyperglycaemia in untreated or poorly controlled diabetes raises intracellular glucose levels in tissues such as neurons, renal glomerulus, lens, and retina that do not require insulin for glucose uptake. It increases cellular mitochondrial oxidative phosphorylation, resulting in enhanced production of superoxide ($O_2^{\cdot-}$) radicals through incomplete reduction of molecular oxygen. These charged $O_2^{\cdot-}$ radicals are the major ROS generated in mitochondria. Mitochondrial superoxide dismutase converts $O_2^{\cdot-}$ to hydrogen peroxide, another ROS, which is normally removed by catalase. However, when glucose levels are high radical production exceeds detoxification by cellular antioxidants. Hydrogen peroxide generates highly reactive hydroxyl radicals that damage organelles and the intramitochondrial accumulation of $O_2^{\cdot-}$ is associated with damage to mitochondrial membranes and DNA.

Elevated blood glucose levels are also associated with increased reduction of glucose to sorbitol by aldose reductase, the rate-limiting enzyme of the polyol pathway. Sorbitol does not readily cross cell membranes so it accumulates intracellularly, and elevated sorbitol levels correlate with the onset of diabetic complications. It causes damage through its osmotic effects, by depleting levels of intracellular myoinositol, a precursor of the membrane phospholipid phosphatidylinositol and by altering the redox state of pyridine nucleotides. Epalrestat, an aldose reductase inhibitor, has been shown to delay progression of diabetic neuropathy [16], most likely as a consequence of reductions in production of both sorbitol and fructose, the product of sorbitol oxidation, which exerts deleterious effects by promoting glycation.

Glycation is the non-enzymatic attachment of monosaccharides (glucose or fructose) to intracellular proteins, lipids, and nucleic acids and chronic hyperglycaemia causes gradual, cumulative, and irreversible changes in these long-lived molecules. Early glycation products combine to form complex cross-linked structures termed advanced glycation end-products (AGE) and impaired renal function in diabetes leads to their accumulation. They bind to a cell surface receptor known as RAGE, whose activation leads to signal transduction cascades that generate more ROS, increase pro-inflammatory cytokines, and culminate in cell dysfunction and apoptosis. RAGE activation induces permanent changes, which may explain why restoration of normoglycaemia does not always reverse tissue damage inflicted by previous exposure to hyperglycaemia. However, reduction in glycation with the use of AGE cross-link breakers such as alagebrium may be clinically useful to prevent hyperglycaemia-induced progression to microvascular complications [17].

REFERENCES

1. Brunskill NJ. C-peptide and diabetic kidney disease. *J Intern Med*, 2017; **281**: 41–51.
2. Yosten GLC, Kolar GR, Redlinger LJ, Samson WK. Evidence for an interaction between proinsulin C-peptide and GPR146. *J Endocrinol*, 2013; **218**: B1–B8.
3. Cryer PE. Regulation of glucose metabolism in man. *J Intern Med Suppl*, 1991; **735**: 31–9.
4. Jones PM, Persaud SJ. Islet function and insulin secretion. In: Holt RIG, Cockram CS, Flyvberg A, Goldstein BJ. *Textbook of Diabetes*, 5th edn. Oxford: Blackwell Scientific Press, 2017: 87–102.
5. Schwede F, Bertinetti B, Langerijs CN, *et al.* Structure-guided design of selective Epac1 and Epac2 agonists. *PLoS Biol*, 2015; **13**: e1002038.

6. Rorsman P, Braun M. Regulation of insulin secretion in human pancreatic islets. *Annu Rev Physiol*, 2013; **75**: 155–79.
7. Pingitore A, Chambers ES, Hill T, *et al.* The diet-derived short chain fatty acid propionate improves beta-cell function in humans and stimulates insulin secretion from human islets *in vitro*. *Diabetes Obes Metab*, 2017; **19**: 257–65.
8. Ghislain J, Poitout V. The role and future of FFA1 as a therapeutic target. *Handb Exp Pharmacol*, 2017; **236**: 159–80.
9. Muller D, Huang GC, Amiel S, *et al.* Identification of insulin signalling elements in human β-cells: autocrine regulation of insulin gene expression. *Diabetes*, 2006; **55**: 2835–42.
10. McIntyre N, Turner DS, Holdsworth CD. New interpretation of oral glucose tolerance. *Lancet*, 1964; **2**: 20–1.
11. Drucker DJ, Habener JF, Holst JJ. Discovery, characterization, and clinical development of the glucagon-like peptides. *J Clin Invest*, 2017; **127**: 4217–27.
12. White MF. Mechanism of insulin action. In: Holt RIG, Cockram CS, Flyvberg A, Goldstein BJ, eds. *Textbook of Diabetes*, 5th edn. Oxford: Blackwell Scientific Press, 2017: 114–32.
13. Withers DJ, Gutierrez JS, Towery H *et al.* Disruption of IRS-2 causes type 2 diabetes in mice. *Nature*, 1998; **391**: 900–4.
14. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*, 1993; **329**: 977–86.
15. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33) *Lancet*, 1998; **352**: 837–53.
16. Hotta N, Kawamori R, Fukuda M, Shigeta Y. Long-term clinical effects of epalrestat, an aldose reductase inhibitor, on progression of diabetic neuropathy and other microvascular complications: multivariate epidemiological analysis based on patient background factors and severity of diabetic neuropathy. *Diabet Med*, 2012; **29**: 1529–33.
17. Borg DJ, Forbes JM. Targeting advanced glycation with pharmaceutical agents: where are we now? *Glycoconj J*, 2016; **33**: 653–70.

15.1.2 Classification and Diagnosis of Diabetes Mellitus

Stephen Colagiuri and Crystal Man Ying Lee

Introduction

Diabetes is a major contributor to the global disease burden. The first International Diabetes Federation (IDF) global estimate in 2000 reported that 151 million people aged 20–79 years had diabetes [1]. Since then the numbers have increased substantially with the latest estimate that 425 million people had diabetes in 2017 [2]. Independently, the NCD Risk Factor Collaboration (NCD-RisC) estimated that 422 million people aged ≥18 years had diabetes in 2014 [3]. Accurate estimations of diabetes prevalence rely on sources and quality of data, modelling assumptions, and methods and criteria used to diagnose diabetes. The IDF selects data based on method of diagnosis, sample size, representation of the sample, age of data source, and type of publication [4]. NCD-RisC included data that measured at least one glycaemic biomarker and adjusted for differences in diabetes prevalence between these biomarkers [3, 5]. Despite different methodologies, the two groups generated similar estimations.

There have been several modifications to diagnostic criteria and classification of diabetes since the first World Health Organization (WHO) report was published over half a century ago [6]. These changes impact comparability of epidemiological data over time, medical care, and non-health related issues associated with labelling of a person as having diabetes.

Diagnostic Criteria for Diabetes

A diagnosis of diabetes can be made when an individual has classic symptoms of diabetes and a clearly elevated plasma glucose. However, many people with diabetes are asymptomatic for many years and blood tests are required for diagnosis. The WHO and American Diabetes Association (ADA) use the same diagnostic criteria for diabetes based on fasting plasma glucose (FPG) ≥7.0 mmol/L (126 mg/dl), 2-hour post-load plasma glucose during an oral glucose tolerance test (2hPG) ≥11.1 mmol/L (200 mg/dl), and glycated haemoglobin (HbA1c) ≥6.5% (48 mmol/mol) [7–9] (Table 15.1.2.1). Repeat testing in asymptomatic individuals is recommended to confirm the diagnosis.

Diagnostic Criteria for Intermediate Hyperglycaemia

Intermediate hyperglycaemia (IH) denotes an elevated blood glucose but lower than levels considered diagnostic of diabetes and is associated with increased risk of developing diabetes and cardiovascular disease. Criteria for IH differ between the WHO and ADA. WHO separates IH into impaired glucose tolerance (IGT) and impaired fasting glucose (IFG) based on FPG and 2hPG levels [7–8] (Table 15.1.2.1). The ADA refers to IH as 'prediabetes', a debated term as the progression to diabetes is variable with a recent report suggesting that two out of three individuals with IFG never develop diabetes [10]. The ADA FPG cut-point is lower than that recommended by WHO and the ADA uses HbA1c as a criterion for prediabetes [9] (Table 15.1.2.1). The ADA criterion of 5.7–6.4% (39–47 mmol/mol) has been questioned as identifying too many people with prediabetes with one large study reporting 35% with

Table 15.1.2.1 Current World Health Organization and American Diabetes Association diagnostic criteria for diabetes and intermediate hyperglycaemia

	Diabetes (WHO)	Intermediate hyperglycaemia (WHO)		Diabetes (ADA)	Prediabetes (ADA)
		Impaired glucose tolerance	Impaired fasting glucose		
Fasting plasma glucose	≥7.0 mmol/L (126 mg/dl)	<7.0 mmol/L (126 mg/dl)	6.1–6.9 mmol/L (110–125 mg/dl)	≥7.0 mmol/L (126 mg/dl)	5.6–6.9 mmol/L (100–125 mg/dl)
	AND/OR	AND	AND	OR	OR
2 hour plasma glucose during a 75 g OGTT	≥11.1 mmol/L (200 mg/dl)	7.8–11.0 mmol/L (140–199 mg/dl)	<7.8 mmol/L (140 mg/dl) (if measured)	≥11.1 mmol/L (200 mg/dl)	7.8–11.0 mmol/L (140–199 mg/dl)
	AND/OR			OR	OR
Glycated haemoglobin	≥6.5% (48 mmol/mol)			≥6.5% (48 mmol/mol)	5.7–6.4% (39–47 mmol/mol)
Random plasma glucose	≥11.1 mmol/L (200 mg/dl) in the presence of diabetes symptoms			≥11.1 mmol/L (200 mg/dl) in the presence of diabetes symptoms	

OGTT, oral glucose tolerance test.

prediabetes based on HbA1c compared with 8% based on 2h PG [11]. WHO is currently considering the use of HbA1c to identify IH but has yet to issue a statement about its use or level.

Major Changes in Diagnostic Criteria Over Time

The WHO initially recommended a single diagnostic criterion for diabetes (2hPG ≥7.2 mmol/L [130 mg/dl]) and referred to 2hPG 6.1–7.1 mmol/L (110–128 mg/dl) as a borderline state [6]. The current 2hPG cut-offs of ≥11.1 mmol/L (200 mg/dl) for diabetes and 7.8–11.0 mmol/L (140–199 mg/dl) for IGT were introduced in 1985 [12]. Glucose doses of 50 g and 100 g were used in the oral glucose tolerance test in the early years but was standardized to 75 g in 1980 [13]. FPG was included in the WHO diagnostic criteria in 1980 [13]. The FPG diagnostic level for diabetes was initially ≥8.0 mmol/L (145 mg/dl) but was revised to ≥7.8 mmol/L (140 mg/dl) in 1985 and further revised down to ≥7.0 mmol/L (126 mg/dl) by the ADA in 1997 and WHO in 1999 [11–14]. IFG was introduced to define FPG levels equivalent to IGT in 1997 by the ADA [14] and in 1999 by WHO [15]. In 2003, the ADA lowered the FPG cut-point for IFG from 6.1 mmol/L (110 mg/dl) to 5.6 mmol/L (100 mg/dl). However, the WHO decided in 2006 to continue with a FPG of 6.1 mmol/L (110 mg/dl) [7]; hence the difference in FPG criterion for IH between these two groups.

HbA1c measurement has a number of potential advantages compared with FPG and 2hPG testing. Although HbA1c had been used to monitor diabetes control for decades, there was concern about considering it as a diagnostic criterion for diabetes until international standardization of the HbA1c assay improved. In 2009, an international expert committee recommended the use of HbA1c for the diagnosis of diabetes [16]. The ADA adopted HbA1c for the diagnosis of diabetes in 2010 [17] and WHO in 2011 [8].

Methods for Determining Diagnostic Cut-Points

A 1971 Pima Indian study reported that the distribution of 2hPG was represented by two overlapping curves [18]. Later studies in populations with high prevalence of diabetes reported a similar bimodal glucose distribution [19, 20]. This bimodal distribution was assumed to comprise two distinct but overlapping groups (normal glucose versus abnormal glucose) and the point where the two curves intersect was used to define the 2hPG cut-point to diagnose diabetes. However, the DETECT-2 Collaboration showed that cut-points derived from bimodal distribution varied considerably between countries and that not all populations demonstrated a bimodal distribution [21]. Moreover, cut-points obtained from this method may not align with those derived from increased risk of diabetes complications.

The 1997 decision of the ADA to lower the diabetes diagnostic cut-point for FPG to the current level was based largely on three studies on the relationship between diabetic retinopathy and FPG, 2hPG, and HbA1c [14]. The prevalence of retinopathy was examined by deciles (ten equally sized groups) of the distribution for each glycaemic measure. The distribution graphs showed that the prevalence of retinopathy was initially low and then increased substantially from the eighth decile (FPG 7.2 mmol/L [130 mg/dl]; 2hPG 12.1 mmol/L [218 mg/dl]; HbA1c 6.9% [52 mmol/mol]) in the Egyptian population, ninth decile (FPG 7.5 mmol/L [135 mg/dl]; 2hPG 13.5 mmol/L [243 mg/dl]; HbA1c 6.7% [50 mmol/mol]) in the Pima Indian population, and tenth decile (FPG 6.7 mmol/L [121 mg/dl]; 2hPG 10.8 mmol/L [195 mg/dl]; HbA1c 6.2% [44 mmol/mol]) in the US population. By consensus, the new FPG cut-point was established and the 2hPG cut-point was not changed. Using data from 45 000 participants, the DETECT-2 Collaboration studied the relationship by vigintiles (20 equally sized groups) and by 0.5 unit intervals of glycaemic measures and was able to focus on moderate to severe retinopathy cases which are specific to diabetes [22]. The various analyses performed in this study indicated that an HbA1c of 6.5% (48 mmol/mol) was a suitable alternative diagnostic criterion for diabetes and has now been universally adopted [8, 16, 17].

Effect of the Different Diagnostic Criteria on Prevalence

The diagnostic cut-points of the three glycaemic measures are individually derived and relate to their association with diabetic retinopathy rather than attempting to identify the same individuals with

hyperglycaemia by each of these measures. Therefore, the diagnostic result of one test may not correspond to that of another test. Data from the 16 000 participants without known diabetes in the DETECT-2 Collaboration who had all three glycaemic measures showed that the proportion with newly diagnosed diabetes varied between the three measures (7.7% for FPG, 13.9% for 2hPG, and 5.7% for HbA1c) [22]. More recently, the NCD-RisC reported 2–6% higher diabetes prevalence based on FPG or 2hPG than that based solely on FPG [5]. More importantly, the results suggested 47% of participants with newly diagnosed diabetes based on the FPG would not have had diabetes based on the HbA1c measurement. This incongruence remains a concern both for population estimates of diabetes but more importantly for the diagnosis of diabetes in an individual and will need to be addressed in future deliberations on diagnostic criteria for diabetes.

Classification of Diabetes

Diabetes is a heterogenous group of diseases which share the common feature of hyperglycaemia. Different types of diabetes are recognized based on natural history, aetiology, pathophysiology, risk of developing complications, and treatment. Classification systems serve several purposes including informing clinical care decisions and individualizing treatment, describing aetiopathology, and for epidemiological studies. While types of diabetes should be defined by categorical features specific and exclusive to that type of diabetes [3 New WHO report], our current state of knowledge does not allow a single classification system to address all these objectives for the majority of people with diabetes.

The first WHO report in 1965 recommended classification of diabetes by age of diabetes onset but also recognized other clinical types, including juvenile-onset diabetes, brittle diabetes, insulin resistant diabetes, gestational diabetes, pancreatic diabetes, endocrine diabetes, and iatrogenic diabetes [6]. The 1979 National Diabetes Data Group [23] and the 1980 WHO [13] classification described four classes of diabetes namely insulin dependent diabetes mellitus, non-insulin dependent diabetes mellitus, gestational diabetes, and another type of diabetes with six subtypes (pancreatic, hormonal, drug or chemical-induced, insulin-receptor abnormalities, genetic syndromes, and others). Malnutrition-related diabetes mellitus was added as a fifth class in the 1985 WHO report [12].

The most recent WHO classification was published in 1999 and returned to four basic types—type 1 diabetes, type 2 diabetes, other specific types, and gestational diabetes [15] (**Box 15.1.2.1**). The subtypes for other specific types included genetic defects of β-cell function which included maturity onset diabetes of the young (MODY), genetic defects in insulin action, diseases of the exocrine pancreas, endocrinopathies, drug- or chemical-induced diabetes, infections, uncommon forms of immune-mediated diabetes, and other genetic syndromes sometimes associated with diabetes [15]. In addition, the classification acknowledged different stages in the natural history of diabetes—normoglycaemia, intermediate hyperglycaemia, diabetes not requiring insulin, diabetes requiring insulin for control, and diabetes requiring insulin for survival [15].

Few studies have reported population prevalence by diabetes types. The 2016 US National Health Interview Survey classified diabetes types based on self-report and insulin use and estimated that of all diagnosed cases, 5.8% were type 1 diabetes, 90.9% were type 2 diabetes and the remaining 3.3% were other types of diabetes [24]. A Swedish study of 13 720 adults reported 1.5% with type 1 diabetes, 88.3% with type 2 diabetes, 5.3% with latent autoimmune diabetes of adults (LADA), 1.2% with diabetes secondary to pancreatic disease, and 3.8% as unclassifiable [25]. A Norwegian study in antibody-negative children with diabetes aged under 15 years showed up to 6.5% had MODY, one in three of whom had not been previously recognized [26].

Type 1 and type 2 diabetes are becoming increasingly difficult to distinguish due to obesity in the young and the occurrence of type 2 diabetes in young people [27]. This and the increasing recognition of intermediate forms of diabetes such as LADA and ketosis-prone type 2 diabetes, a syndrome of episodic diabetic ketoacidosis without immunologic markers characterized by variable insulin dependence [28], have led to calls to review the classification system for diabetes to better assist clinicians and reflect advances in knowledge of pathophysiological pathways [29, 30]. One proposal suggested classification centred on the abnormal β-cell function common to all forms of diabetes and 11 distinct pathways contributing to β-cell stress, dysfunction, or loss [31].

A recent study reported five distinct subtypes of type 2 diabetes based on a cluster analysis of six variables (glutamate decarboxylase antibodies, age at diagnosis, BMI, HbA1c, and homoeostatic model assessment of β-cell function and insulin resistance) which compared time to medication, time to reach treatment goal, risk of diabetic complications and genetic associations [25]. Whether this clustering is applicable to other populations and ethnicities remains to be demonstrated. In addition, adopting such a system in most parts of the world would be challenging due to lack of the necessary laboratory tests. Therefore, it seems that for the immediate future, classification systems will rely on phenotypic characteristics, clinical judgement, and careful monitoring of clinical course and responses to treatment. This is especially important in those individuals in whom diabetes type is difficult to assign at presentation.

Box 15.1.2.1 Aetiological classification of diabetes

Type 1—β-cell destruction, usually leading to absolute insulin deficiency
- Autoimmune
- Idiopathic

Type 2—may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with or without insulin resistance

Other specific types
- Genetic defects of β-cell function
- Genetic defects of insulin action
- Diseases of the endocrine pancreas
- Endocrinopathies
- Drug- or chemical-induced
- Infections
- Uncommon forms of immune-related diabetes
- Other genetic syndromes sometimes associated with diabetes

Gestational diabetes

REFERENCES

1. International Diabetes Federation. *Diabetes Atlas 2000*. Brussels: International Diabetes Federation, 2000.
2. International Diabetes Federation. *IDF Diabetes Atlas*, 8th edn. Brussels: International Diabetes Federation, 2017.
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*, 2016; **387**: 1513–30.
4. Guariguata L, Whiting D, Weil C, Unwin N. The International Diabetes Federation Diabtes Atlas methodology for estimating global and national prevalence of diabetes in adults. *Diabetes Res Clin Pract*, 2011; **94**: 322–32.
5. NCD Risk Factor Collaboration (NCD-RisC). Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: a pooled analysis of 96 population-based studies with 331288 participants. *Lancet Diabetes Endocrinol*, 2015; **3**: 624–37.
6. World Health Organization. *Diabetes Mellitus: Report of a WHO Expert Committee* (Technical report series no. 310). Geneva: World Health Organization, 1965.
7. World Health Organization, International Diabetes Federation. *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia* (Report of a WHO/IDF consultation). Geneva: World Health Organization, 2006.
8. Report of a World Health Organization Consultation. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. *Diabetes Res Clin Pract*, 2011; **93**: 299–309.
9. American Diabetes Association. Classification and diagnosis of diabetes. *Diabetes Care*, 2015; **38**: S8–16.
10. Echouffo-Tcheugui JB, Niiranen TJ, McCabe EL, *et al.* Lifetime prevalence and prognosis of prediabetes without progression to diabetes. *Diabetes Care*, 2018; **41**: e117–18.
11. Xu Y, Wang L, He J, *et al.*, for the 2010 China Noncommunicable Disease Surveillance Group. Prevalence and control of diabetes in Chinese adults. *JAMA*, 2013; **310**: 948–58.
12. World Health Organization. *Diabetes Mellitus: Report of a WHO Study Group* (Technical report series no. 727). Geneva: World Health Organization, 1985.
13. World Health Organization. *Expert Committee on Diabetes Mellitus* (Technical report series no. 646). Geneva: World Health Organization, 1980.
14. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, 1997; **20**: 1183–97.
15. World Health Organization. *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications*. Report of a WHO consultation. Part 1: Diagnosis and classification of diabetes mellitus. Geneva: World Health Organization, 1999.
16. The International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*, 2009; **32**: 1327–34.
17. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 2010; **33**: S62–9.
18. Rushforth NB, Bennett PH, Steinberg AG, Burch TA, Miller M. Diabetes in the Pima Indians. Evidence of bimodality in glucose tolerance distribution. *Diabetes*, 1971; **20**: 756–65.
19. Zimmet P, Whitehouse S. Bimodality of fasting and two-hour glucose tolerance distributions in a Micronesian population. *Diabetes*, 1978; **27**: 793–800.
20. Lim TO, Bakri R, Morad Z, Hamid MA. Bimodality in blood glucose distribution. Is it universal? *Diabetes Care*, 2002; **25**: 2212–17.
21. Vistisen D, Colagiuri S, Borch-Johnsen K; the DETECT-2 Collaboration. Bimodal distribution of glucose is not universally useful for diagnosing diabetes. *Diabetes Care*, 2009; **32**: 397–403.
22. Colagiuri S, Lee CM, Wong TY, Balkau B, Shaw JE, Borch-Johnsen K; the DETECT-2 Collaboration writing group. Glycemic thresholds for diabetes-specific retinopathy. Implications for diagnostic criteria for diabetes. *Diabetes Care*, 2011; **34**: 145–50.
23. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes*, 1979; **28**: 1039–57.
24. Bullard KM, Cowie CC, Lessem SE, *et al.* Prevalence of diagnosed diabetes in adults by diabetes type—United States, 2016. *MMWR Morb Mortal Wkly Rep*, 2018; **67**: 359–61.
25. Ahlqvist E, Storm P, Käräjämäki A, *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*, 2018; **6**: 361–9.
26. Johansson BB, Irgens HU, Molnes J, *et al.* Targeted next-generation sequencing reveals MODY in up to 6.5% of antibody-negative diabetes cases listed in the Norwegian Childhood Diabetes Registry. *Diabetologia*, 2017; **60**: 625–35.
27. Lesile RD, Palmer J, Schloot NC, Lernmark A. Diabetes at the crossroads: relevance of disease classification to pathophysiology and treatment. *Diabetologia*, 2016; **59**: 13–20.
28. Sobngwi E, Gautier JF. Adult-onset idiopathic Type I or ketosis-prone Type II diabetes: evidence to revisit diabetes classification. *Diabetologia*, 2002; **45**: 283–5.
29. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. the many faces of diabetes: a diabetes with increasing heterogeneity. *Lancet*, 2014; **383**: 1084–94.
30. Skyler JS, Bakris GL, Bonifacio E, *et al.* Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes*, 2017; **66**: 241–55.
31. Schwartz SS, Epstein S, Corkey BE, Grant SF, Gavin III JR, Aguilar RB. The time is right for a new classification system for diabetes: rationale and implications of the β-cell-centric classification schema. *Diabetes Care*, 2016; **39**: 179–86.

15.2

Type 1 Diabetes

15.2.1 Epidemiology and Public Health

Elizabeth J. Mayer-Davis and Daria Igudesman

Introduction

Diabetes is broadly classified into type 1 diabetes mellitus (T1D), type 2 diabetes mellitus (T2D), gestational diabetes mellitus (GDM), and specific types due to other causes [1]. In T1D, a combination of genetic and environmental factors results in autoimmune destruction of pancreatic β-cells and insufficient insulin production. Meanwhile, insulin resistance is the hallmark of T2D—the most common diabetes type—alongside insufficient insulin production, resulting in hyperglycaemia. GDM usually occurs in the second or third trimesters of pregnancy and resolves post-delivery. Specific diabetes types due to other causes include those resulting from a genetic mutation, diseases of the exocrine pancreas, and drug- or chemical-induced diabetes.

While T1D develops most commonly in childhood and adolescence, T2D is more frequent (albeit not exclusively) among adults and is associated with obesity. Genetic risk scores for T1D have been established through genome-wide association studies (GWAS) that have identified several alleles involved in immunity [2]. GWAS have also linked T2D and obesity-related genetic loci, including body mass index (BMI) and waist-to-hip ratio [3]. Strong evidence exists for lifestyle prevention of T2D, which in long-term studies has been more effective than metformin [4]. Because acute complications are typically less severe than in T1D, T2D can go undetected for long periods, and is often diagnosed due to complications of prolonged hyperglycaemia.

Diagnosis of Diabetes

According to the American Diabetes Association (ADA), a fasting plasma glucose (FPG) of >126 mg/dl indicates diabetes, while impaired fasting glucose (IFG) is between 100 and 125 mg/dl. The World Health Organization (WHO) recommends performing an oral glucose tolerance test (OGTT), by which diabetes is defined as 2-hour plasma glucose (PG) ≥ 200 mg/dl. Haemoglobin A1c (HbA1c) may also be used for diagnosis (prediabetes between 5.7% and 6.4%; diabetes >6.5%), although the assay is not universally standardized, can be affected by other diseases such as malaria, and is not feasible in some low-resource settings (see Chapter 15.1.2).

Diabetes Prevalence

Globally, an estimated 8.8% (95% CI 7.2–11.3) of adults, or 425 million people, have diagnosed diabetes. Of these, 79% live in low- and middle-income countries. This is expected to increase to 629 million adults by 2045, owing largely to countries whose economic status will shift from low to middle income with increasing exposure to risk factors for development of (primarily) T2D. Prevalence is lower among women compared to men (8.4% compared to 9.1%), and greater in urban compared to rural environments (10.2% vs. 6.9%). Meanwhile, undiagnosed diabetes accounts for half of all diabetes cases worldwide, with 84.5% occurring in low- and middle-income countries. The proportion of undiagnosed diabetes is as high as 69.2% in Africa, likely due to low screening priority and resource availability [1].

US age-standardized estimates of total (diagnosed and undiagnosed) diabetes prevalence increased from 10.8% (9.5–12.0) in 2001–2002 to 12.4% (10.8–14.2) in 2011–2012 (P <0.001) [5]. This was attributable to increases in diagnosed diabetes, as undiagnosed prevalence only changed among Mexican Americans (5.6% and 5.9% in 1988–1994 and 2011–2012, respectively). Estimates

have increased significantly in every stratum of age, sex, race/ethnicity, education level, and poverty income ratio tertiles, although the only increase among BMI strata occurred in the 30+ stratum (18% (15.9–20.0) in 1988–1994 vs. 20.1% (17.8–22.4) in 2011–2012, $P = 0.003$). 2011–2012 prevalence was lower among non-Hispanic whites (NHW) (11.3% (9.0–14.1)) than non-Hispanic blacks (21.8% (17.7–26.7)), non-Hispanic Asians (20.6% (15.0–27.6)), and Hispanics (22.6% (18.4–27.5)). Meanwhile, prevalence of diagnosed diabetes among adults was 9.1% (7.8–10.6), and 5.2% (4.0–6.9) for undiagnosed diabetes; prediabetes prevalence was 38% (34.7–41.3), of which 36.4% (30.5–42.7) cases were undiagnosed.

T1D Prevalence and Incidence

In high income countries, between 7% and 12% of all individuals with diabetes have T1D [1, 6]. In the United States, three-quarters of all T1D diagnoses are made to individuals under the age of 18 [7]. Global annual incidence has ranged from 0.5 to over 60 cases per 100 000 children under the age of 15 [8].

Incidence rates of T1D appear to be increasing in the United States, where age, sex, and race/ethnicity adjusted estimates among youths aged 0 to 19 have increased by 1.8% annually between 2002–2003 and 2011–2012 [9]. This equates to 15 900 cases diagnosed in 2002–2003, compared to 17 900 new cases in 2011–2012. The estimated annual increase in incidence was highest among Hispanic youth. Trends in international incidence rates have varied, however; while in Australia, annual incidence among children has followed a sinusoidal trend [10], it seems to have stabilized in Scandinavian nations [11], and increased in other European paediatric populations [12].

T2D Prevalence and Incidence

US estimates of diagnosed T2D prevalence increased from 4.96% to 7.78% between 1999 and 2010 among adults over age 30, while undiagnosed T2D prevalence increased from 3.45% to 4.36%. Adjusted incidence rates among US youths with T2D increased by 4.8% annually from the 2002–2003 to the 2011–2012 period ($P < 0.001$). Incidence rates increased in all age, sex, and race/ethnicity categories, except among NHW; greater relative increases were observed in non-Hispanic blacks, Asians, and Native Americans combined compared to NHW (8.9% vs. 3.1%, $P < 0.01$).

Between 2001 and 2009, prevalence of diagnosed T2D increased among youths aged 10–19 from 0.34 per 1000 (0.31–0.37) to 0.46 per 1000 (0.43–0.49) [13]. Reliable international estimates for T2D among youths are lacking, as diagnostic procedures vary by region. However, a systematic review suggests that both incidence and prevalence rates are increasing worldwide [14].

GDM Prevalence and Complications

GDM occurs in 8.3% of pregnancies worldwide [15], and is thought to be increasing due to increased prevalence of overweight and obesity [16]. Fetal complications include high and low birth weight, congenital defects, and perinatal death. Maternal complications include eclampsia, preeclampsia, and Caesarean section [17]. Both women with GDM and children born to mothers with GDM are at increased risk for developing T2D [18, 19].

Complications of Diabetes

Acute Complications

Although severe hypoglycaemia has been associated with intensive diabetes therapy, long-term studies suggest that rates are becoming more similar among intensively treated versus conventionally treated T1D patients [20]. Incidence of hospitalization for severe hypoglycaemia among adults with T1D has increased by 3.74% (1.70–5.83) annually between 1998 and 2013, and by 8.59% (5.76–11.50) among elderly adults with T2D between 1998 and 2009. Although incidence decreased by 8.05% (–14.48–1.13) annually among this population between 2009 and 2013, 2013 incidence remained higher than in 1998. This reduction may be attributed to reduced use of sulfonylurea medications, which can induce hypoglycaemia [21].

Diabetic ketoacidosis (DKA) can be life-threatening, and often results in hospitalization. Annual incidence of hospitalizations for DKA in T1D remained stable in an English population between 1998 and 2004, but increased by 14.10% annually between 2004 and 2007, after which it plateaued. Incidence among adults with T2D increased annually by 4.24% between 1998 and 2013 [22].

Chronic Complications

Macrovascular complications lead to cardiovascular disease (CVD), and can be prevented through management of blood glucose (BG), hypertension, and dyslipidaemia. Individuals with diabetes are two to four times more likely to have CVD than non-diabetic individuals [23]; this estimate increases with age, and in low- and middle-income compared to high-income countries [1]. While intensive glucose-lowering therapy was shown to reduce the hazard of myocardial infarction among individuals with T1D and T2D compared to standard therapy [24, 25], intensive insulin treatment of T1D can also result in weight gain, which itself may increase risk for CVD [25]. US estimates approximate a 6.3% 10-year CVD risk among participants aged 30–39 [26]. Approximately 14–45% of children with T1D have two or more CVD risk factors, with girls at higher risk than boys.

Increased diabetes duration and poor glycaemic control have been associated with microvascular complications. Of these, diabetic kidney disease (DKD) is a major cause of end-stage renal disease, morbidity, and mortality among young adults with T1D. Because DKD is associated with CVD [27], controlling blood pressure and BG through lifestyle and pharmacology is key for prevention of progression to dialysis or need for kidney transplant [28].

Diabetic retinopathy (DR) is a microvascular complication that is the primary cause of vision loss among 20–65 year olds, and occurs to some degree in 1 of every 3 individuals with diabetes (see Chapter 15.12.2). It is more prevalent in T2D than T1D, and adolescents are at greater risk of progression to vision loss than adults. Prevalence is as high as 41% in South-East Asia, while the lowest documented prevalence of 12% is in Africa [1]. Risk of DR progression may be reduced by managing lipids [29], and severity may

be reduced through intensive BG control [30]. Prevention requires regular retinal screening, as DR is often asymptomatic in early stages, although screening is often unavailable in low- and middle-income countries.

Uncontrolled BG can also lead to neuropathy—microvascular-mediated damage of nerve cells—particularly in peripheral tissues such as the feet (see Chapter 15.4.1). The resulting loss of sensation can lead to unattended injuries and wounds. Thus, ulceration, infection, and amputation occur at higher rates among individuals with diabetes compared with non-diabetic controls [31]. Global diabetic foot ulceration was estimated to be 6.4% (4.6–8.1) among individuals with T2D—higher than those with T1D (5.5%, 3.2–7.7).

Cerebrovascular risk, or damage to the microvasculature of the brain through poorly managed BG, has been associated with cognitive impairment among individuals with diabetes [32]. Complex tasks of executive function have been subtly reduced among young adults with T1D compared with controls, although no differences in IQ or memory were observed between groups [33]. Likewise, academic performance among children with T1D has been negatively associated with severe hypoglycaemia [34]. Reduced performance in verbal memory, attention tasks, and processing speed were detected in a meta-analysis of individuals with T2D compared with controls [35].

A recent WHO report attributes 1.6 million annual deaths to diabetes globally. This number is greater among women (2.1 million, 1.7–2.7) than men (1.0 million, 1.5–2.3) in all global regions, except in the North American and Caribbean region. Among individuals with T1D, standardized mortality ratios (SMR) for total mortality and CVD are elevated among women compared to men [36]. Among middle-aged individuals with T2D in high- and middle-income countries, 27 out of every 1000 are estimated to experience CVD-related deaths annually, primarily due to stroke and coronary artery disease [1]. Other diabetes-related complications result in 4.0 (3.2–5.0) million deaths worldwide among adults.

Additional complications of diabetes include depression, dementia, tuberculosis, and non-alcoholic fatty liver disease (NAFLD). Quality of life is also impacted heavily. For instance, DR has been associated with deteriorating physical state among individuals with diabetes, with 58% of those surveyed globally reporting limitations in daily activities of life, compared with 37% without DR [1].

Economic Impact

Diabetes care expenditures were recently estimated to be 727 billion USD annually, or one-eighth of all global income spent on healthcare (see Chapter 15.15.2). This is expected to increase by 104 billion USD by 2045 [1]. Another study estimated global diabetes cost to be 1.3 trillion USD in 2015 [37], while diabetes-related CVD resulted in a loss of 84 billion USD globally between 2005 and 2015. Direct healthcare costs to those living with diagnosed diabetes tend to be twice as high as those of non-diabetic individuals, while work absenteeism in the United States results in annual loss of 5 billion USD to individuals; reduced productivity while working and inability to work lead to annual losses of 20.8 and 21.6 billion USD, respectively.

Summary

The epidemiology of diabetes mellitus is a critical element in developing public health policy and advancing healthcare systems. Subgroup analyses of incidence and prevalence estimates can help identify high-risk subpopulations that would benefit from tailored interventions. However, data are limited in some settings, which complicates efforts to obtain valid estimates and to differentiate between T1D and T2D. Nonetheless, updates in epidemiologic data are necessary for improving diabetes prevention and treatment both globally and regionally.

REFERENCES

1. International Diabetes Federation. *IDF Diabetes Atlas*, 8th edn. Brussels: IDF, 2017.
2. Redondo MJ, Oram RA, Steck AK. Genetic risk scores for type 1 diabetes prediction and diagnosis. *Curr Diabetes Rep*, 2017; **17**: 129.
3. Karaderi T, Drong AW, Lindgren CM. Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of obesity-related traits. *Curr Diabetes Rep*, 2015; **15**: 83.
4. Group DPPR. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol*, 2015; **3**: 866–75.
5. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA*, 2015; **314**: 1021–9.
6. Holman N, Young B, Gadsby R. Current prevalence of type 1 and type 2 diabetes in adults and children in the UK. *Diabetic Med*, 2015; **32**: 1119–20.
7. American Diabetes Association. 4. Lifestyle management: standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41**(Suppl. 1): S38–50.
8. World Health Organization. *Global Report on Diabetes*. Geneva: World Health Organization, 2016.
9. Mayer-Davis EJ, Lawrence JM, Dabelea D, *et al.* Incidence trends of type 1 and type 2 diabetes among youths, 2002–2012. *N Engl J Med*, 2017; **376**: 1419–29.
10. Haynes A, Bulsara MK, Bower C, Jones TW, Davis EA. Regular peaks and troughs in the Australian incidence of childhood type 1 diabetes mellitus (2000–2011). *Diabetologia*, 2015; **58**: 2513–16.
11. Berhan Y, Waernbaum I, Lind T, Möllsten A, Dahlquist G, Group SCDS. Thirty years of prospective nationwide incidence of childhood type 1 diabetes: the accelerating increase by time tends to level off in Sweden. *Diabetes*, 2011; **60**: 577–81.
12. Patterson C, Gyürüs E, Rosenbauer J, *et al.* Trends in childhood type 1 diabetes incidence in Europe during 1989–2008: evidence of non-uniformity over time in rates of increase. *Diabetologia*, 2012; **55**: 2142–7.
13. Dabelea D, Mayer-Davis EJ, Saydah S, *et al.* Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *JAMA*, 2014; **311**: 1778–86.
14. Farsani SF, Van Der Aa M, Van Der Vorst M, Knibbe C, De Boer A. Global trends in the incidence and prevalence of type 2 diabetes in children and adolescents: a systematic review and

evaluation of methodological approaches. *Diabetologia*, 2013; **56**: 1471–88.
15. Brussels: International Diabetes Federation. *IDF Diabetes Atlas*, 6th edn. Brussels: IDF, 2013.
16. Pu J, Zhao B, Wang EJ, *et al.* Racial/ethnic differences in gestational diabetes prevalence and contribution of common risk factors. *Paediatr Perinat Epidemiol*, 2015; **29**: 436–43.
17. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study Cooperative Research Group. Hyperglycemia and adverse pregnancy outcome (HAPO) study: preeclampsia. *Am J Obstet Gynecol*, 2010; **202**: 255.e1–7.
18. Bellamy L, Casas J-P, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet*, 2009; **373**: 1773–9.
19. Clausen TD, Mathiesen ER, Hansen T, *et al.* High prevalence of type 2 diabetes and pre-diabetes in adult offspring of women with gestational diabetes mellitus or type 1 diabetes: the role of intrauterine hyperglycemia. *Diabetes Care*, 2008; **31**: 340–6.
20. Gubitosi-Klug RA, Braffett BH, White NH, *et al.* Risk of severe hypoglycemia in type 1 diabetes over 30 years of follow-up in the DCCT/EDIC study. *Diabetes Care*, 2017; **40**: 1010–16.
21. Zhong VW, Juhaeri J, Cole SR, *et al.* Incidence and trends in hypoglycemia hospitalization in adults with type 1 and type 2 diabetes in England, 1998–2013: a retrospective cohort study. *Diabetes Care*, 2017; **40**: 1651–60.
22. Zhong VW, Juhaeri J, Mayer-Davis EJ. Trends in hospital admission for diabetic ketoacidosis in adults with type 1 and type 2 diabetes in England, 1998–2013: a retrospective cohort study. *Diabetes Care*, 2018; **41**: 1870–7.
23. Mozaffarian D, Benjamin EJ, Go AS, *et al.* Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation*, 2015; **131**: e29–322.
24. Gerstein HC, Miller ME, Ismail-Beigi F, *et al.* Effects of intensive glycaemic control on ischaemic heart disease: analysis of data from the randomised, controlled ACCORD trial. *Lancet*, 2014; **384**: 1936–41.
25. Purnell JQ, Braffett BH, Zinman B, *et al.* Impact of excessive weight gain on cardiovascular outcomes in type 1 diabetes: results from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study. *Diabetes Care*, 2017; **40**: 1756–62.
26. Miller RG, Mahajan HD, Costacou T, Sekikawa A, Anderson SJ, Orchard TJ. A contemporary estimate of total mortality and cardiovascular disease risk in young adults with type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complications Study. *Diabetes Care*, 2016; **39**: 2296–303.
27. Fox CS, Matsushita K, Woodward M, *et al.* Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet*, 2012; **380**: 1662–73.
28. Agrawal L, Azad N, Bahn GD, *et al.* Long-term follow-up of intensive glycaemic control on renal outcomes in the Veterans Affairs Diabetes Trial (VADT). *Diabetologia*, 2018; **61**: 295–9.
29. Azad N, Bahn GD, Emanuele NV, *et al.* Association of blood glucose control and lipids with diabetic retinopathy in the Veterans Affairs Diabetes Trial (VADT). *Diabetes Care*, 2016; **39**: 816–22.
30. Zhu C-H, Zhang S-S, Kong Y, Bi Y-F, Wang L, Zhang Q. Effects of intensive control of blood glucose and blood pressure on microvascular complications in patients with type II diabetes mellitus. *Int J Ophthalmol*, 2013; **6**: 141.
31. Moxey P, Gogalniceanu P, Hinchliffe R, *et al.* Lower extremity amputations—a review of global variability in incidence. *Diabet Med*, 2011; **28**: 1144–53.
32. Jacobson A, Ryan C, Cleary P, *et al.* Biomedical risk factors for decreased cognitive functioning in type 1 diabetes: an 18-year follow-up of the Diabetes Control and Complications Trial (DCCT) cohort. *Diabetologia*, 2011; **54**: 245–55.
33. Ly TT, Anderson M, McNamara KA, Davis EA, Jones TW. Neurocognitive outcomes in young adults with early-onset type 1 diabetes: a prospective follow-up study. *Diabetes Care*, 2011; **34**: 2192–7.
34. Hannonen R, Komulainen J, Riikonen R, *et al.* Academic skills in children with early-onset type 1 diabetes: the effects of diabetes-related risk factors. *Dev Med Child Neurol*, 2012; **54**: 457–63.
35. Palta P, Schneider AL, Biessels GJ, Touradji P, Hill-Briggs F. Magnitude of cognitive dysfunction in adults with type 2 diabetes: a meta-analysis of six cognitive domains and the most frequently reported neuropsychological tests within domains. *J Int Neuropsychol Soc*, 2014; **20**: 278–91.
36. Huxley RR, Peters SA, Mishra GD, Woodward M. Risk of all-cause mortality and vascular events in women versus men with type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2015; **3**: 198–206.
37. Bommer C, Sagalova V, Heesemann E, *et al.* Global economic burden of diabetes in adults: projections from 2015 to 2030. *Diabetes Care*, 2018; **41**: 963–70.

15.2.2 Presentation and Natural History of Type 1 Diabetes

Augustin Brooks

Historical Perspective and Type 1 Diabetes Subtypes

Physician Aretaeus of Cappadocia first used the term 'diabetes' to summarize the presentation and natural history of sufferers in the second century AD:

> Diabetes is a dreadful affliction, not very frequent among men, being a melting down of the flesh and limbs into urine. The patients never stop making water and the flow is incessant, like the opening of aqueducts. Life is short, unpleasant and painful, thirst

> unquenchable, drinking excessive, and disproportionate to the large quantity of urine … the viscera seem scorched up, the patients are affected by nausea, restlessness and a burning thirst, and within a short time, they expire.

The term 'diabetes' means siphon, and it is likely Aretaeus' patients had type 1A diabetes, the classical hallmark of which is autoimmune destruction of pancreatic β-cells (seen in 70–90% of type 1 diabetes), producing an insulin deficient state which results in spontaneous hyperglycaemia and subsequent ketone formation. Rate of β-cell loss is highly variable, and onset in prepubertal childhood may reflect a more aggressive destructive process, while presentation later in life may be associated with high residual β-cell mass (see Chapter 15.2.3, 'Pathogenesis'). Latent autoimmune diabetes of adulthood (LADA) is slowly progressive but ultimately requires insulin therapy; autoimmune antibodies are detectable at presentation and genetic studies suggest that it is not a distinct entity from type 1A diabetes.

A minority of those presenting with ketoacidosis due to insulin insufficiency show no evidence of autoimmune activation, and can later achieve good glycaemic control for many years with adjustment of lifestyle, with or without insulin therapy. This form of type 1B idiopathic diabetes is described in individuals of African and Caribbean origin, termed 'J type' or 'Flatbush' diabetes and best summarized as periodic insulin insufficiency, although diabetic ketoacidosis can recur many years after initial presentation [1].

Incidence

Classically described as a disease of childhood, onset of type 1 diabetes can occur at any age beyond the first 6 months of life, with two peak incidences at 5–7 years old and around commencement of puberty, with a slight male preponderance.

Incidence varies markedly between countries and is positively related to distance away from the equator: the Chinese province of Guizhou has a low incidence of 0.1/100 000/year, while highest rates are found in European countries peaking in Finland at 60/100 000/year. Temporal trends suggest that incidence is increasing at 3–4% per calendar year, most notably in younger age groups. A seasonal winter peak supports a theoretical role for an environmental agent driving pathogenesis.

Natural History Preceding Clinical Onset

Genetic Predisposition

Aetiology of type 1 diabetes remains unclear but a genetic predisposition is apparent: 4–6% of siblings of type 1 diabetes probands develop the condition. Risk is 5–8% in offspring from diabetic fathers, and 2–5% from diabetic mothers, although cumulative incidence in a monozygotic twin is less than 50%.

Over 40 genetic loci contributing to type 1 diabetes risk have been identified [2]. The Human Leukocyte Antigen (HLA) complex on chromosome 6p21 provides the greatest contribution to the overall genetic susceptibility (~60%). HLA genes encode molecules that participate in antigen presentation to CD4 cells, and the HLA-DQ locus has a greater association with type 1 diabetes than any other HLA subtype. HLA class II proteins associated with diabetes might promote antigen presentations that generate T-helper cells responsible for an immune response to specific islet cell autoantigens, activating cytotoxic CD8 T cells that destroy insulin-producing β cells in the islets of Langerhans.

Environmental Trigger

Twin studies demonstrate that not all individuals with genetic susceptibility to type 1 diabetes develop the condition and rapidly changing incidence of diabetes within a genetically stable population implies non-genetic factors must have a causative role.

Leading hypotheses have proposed early rubella and enterovirus infections as environmental triggers; destruction of infected β cells by virus-induced cytolysis could release antigens that trigger a pathogenic autoreactive T cell response. However, no direct evidence for a causative viral strain has been identified. Nutritional influences have been considered, with shortened period of breastfeeding, early exposure to cow's milk or wheat gluten and low vitamin D levels all proposed as potential causative factors (see Chapter 15.2.3, 'Pathogenesis').

Prodrome and Autoantibody Formation

The β-cell decline hypothesis remains a widely referenced benchmark model for type 1 diabetes [3]. This model proposes that genetically susceptible individuals encounter an environmental trigger that initiates an autoimmune process, which subsequently leads to autoantibody formation, activation of autoreactive T cells capable of destroying β-cells and a linear decay in β-cell mass. In this model, type 1 diabetes presents when 80–90% of β-cells have been destroyed, although β-cell mass between individuals varies 3–5-fold and timeline to clinical onset might be determined by initial β-cell mass rather than aggressiveness of the autoimmune attack.

Antibodies formed against prominent antigens are important markers of β-cell autoimmunity, detectable in 85% of patients at time of disease presentation. More than two dozen autoantibodies have been identified, the most significant being autoantibodies to insulin (micro IAA), glutamic acid decarboxylase (GAD65), insulinoma-associated antigen-2 (IA-2A), and zinc transporter 8 (ZnT8). Current generation assays demonstrate high diagnostic sensitivity and specificity for type 1 diabetes.

Autoantibodies are surrogate measures for β-cell autoimmunity and type 1 diabetes risk, with 5-year incidence 20–25% for subjects with one autoantibody, 50–60% with two autoantibodies, 70% with three antibodies, and 80% with four autoantibodies [4]. IAA autoantibodies are usually the earliest detectable, and a higher titre is consistent with a more aggressive disease course, associated with presentation before the age of 5 years. Autoantibody titres typically drop to low levels in the years after diagnosis. Loss of endogenous insulin production is more rapid in the first year following diagnosis than the second year, with a greater loss in children and adolescents than adults. Some residual β cell function may persist for many years, with detectable C-peptide production reported decades after diagnosis in small cohorts [5].

Box 15.2.2.1 Physiological effects of insulin insufficiency

Consequences of insulin insufficiency

Hepatic
- Glycogenolysis
- Reduced glycogen stores
- Reduced glucose oxidation
- Ketone body formation

Muscle
- Reduced glucose uptake
- Reduced glucose oxidation
- Reduced glycogen synthesis
- Increased proteolysis
- Loss of muscle mass

Fat tissue
- Increased lipolysis
- Reduced triglyceride synthesis
- Reduced adipocyte maturation
- Fat tissue breakdown

Other consequences
- Reduced cell and tissue growth
- Impaired immune modulation

Clinical Presentation

There is a wide gap between subclinical initiation of autoimmune damage and symptom onset; loss of first phase insulin response is usually followed by a period of glucose intolerance and patients do not present with pancreatic inflammation. Eventually, reduction in endogenous insulin production is so severe that a catabolic state of unrestrained glycogenolysis and hepatic gluconeogenesis develops, alongside failure of glucose uptake into peripheral tissues (see Box 15.2.2.1). Chronic hyperglycaemia ensues with clinical onset of symptoms.

Classic Symptom Triad

Osmotic symptoms of polyuria and polydipsia occur in 96% of individuals presenting with type 1 diabetes, with weight loss reported in 85% [6]. Symptom prevalence is increased in autoantibody positive individuals, while increasing age has been independently associated with weight loss and fatigue. Median symptom duration prior to presentation is typically 4 weeks in females and 3 weeks in males. Symptoms and signs of type 1 diabetes are listed in Table 15.2.2.1; their presence in combination with venous plasma glucose ≥11.1 mmol/L is diagnostic of diabetes.

Diabetic Ketoacidosis

Diabetic ketoacidosis (DKA) was identified in 42% of individuals presenting with newly diagnosed type 1 diabetes in a recent large cohort study [6]. Higher rates can occur in the very young, with an associated mortality risk as a result of cerebral oedema.

Insulin insufficiency causes unrestrained lipolysis, increasing non-esterified fatty acids that the liver converts to ketones providing an alternative energy source to glucose. Ketonaemia leads to metabolic acidosis, however, associated with nausea and eventually vomiting, with ketone excretion into the urine increasing osmotic diuresis. A vicious cycle of increasing ketones, vomiting, and acute renal impairment contribute to a worsening metabolic acidosis. Bicarbonate falls as carbon dioxide excretion is increased from respiratory compensation, while acid accumulation from ketones produces an increased anion gap.

The inability of insulin-deprived cells to take up glucose and potassium intracellularly leads to worsening hyperglycaemia and hyperkalaemia, with loss in urine and gastric secretions. Although hyperkalaemia occurs at DKA presentation, it must be remembered that total body potassium is depleted and a rapid reduction in serum potassium levels can occur if fluid and insulin are administered without potassium replacement.

Prolonged solute loss prior to DKA presentation may deplete total body sodium, and hyperglycaemia causes a further artefactual fall in measured plasma sodium exacerbated by hypertriglyceridaemia: each 3 mmol/L elevation in glucose accounts for approximately a 1 mmol/L drop in measured plasma sodium. Increased serum sodium at DKA presentation therefore reflects profound dehydration.

Natural History Following Diagnosis

In 1913, Harvard Medical School described the near uniform fatality of childhood onset diabetes, with anticipated survival 18–24 months if diagnosed under the age of 7 years [7]. The discovery of exogenous insulin therapy in 1921 was a monumental achievement, transforming type 1 diabetes from a terminal disease to a manageable medical condition. A contemporary picture of mortality between 1970 and 2007 in Finland reported an overall cumulative mortality of 17.9% after 35 years' duration of type 1 diabetes [8]. Standardized mortality ratio (SMR) 20 years post-diagnosis decreased with time from 3.5 for patients diagnosed in 1970–1974 to 1.9 for those diagnosed in 1985–1989. Mortality due to chronic complications reduced from 10.0 to 2.2 per 10 000 person-years for patients diagnosed between the ages of 0–15 years.

Loss in life expectancy from type 1 diabetes has been estimated at 11.1 years in men and 12.9 years in women, with greatest loss in expectancy from ischaemic heart disease [9]. Diabetic coma and ketoacidosis remain the greatest contributors to mortality before the age of 50 years, and even today childhood diabetes remains a rapidly progressive and ultimately fatal condition in parts of sub-Saharan Africa. Addressing global disparity in access to healthcare is a priority for present day management.

The Diabetes Control and Complications Trial (DCCT) confirmed that tight blood glucose control can reduce the incidence of complications but regulation of blood glucose levels within the normal physiological range remains challenging despite recent advances in medical therapy. Exogenous insulins are unable to mimic the natural regulatory ability of pancreatic β cells and nearly all individuals with type 1 diabetes remain prone to microvascular and macrovascular complications.

Macrovascular Complications Influence Mortality

Improvements in life expectancy have unmasked higher incidence of macrovascular disease in type 1 diabetes populations. Atheroma develops faster and more extensively in individuals with diabetes, with coronary lesions more prone to plaque ulceration and rupture. The risk of a cardiac event is ten times that of age-matched non-diabetic populations [10], although intensive diabetes treatment

Table 15.2.2.1 Symptoms and signs of type 1 diabetes

Symptoms	Signs	Causation
Urinary symptoms Polyuria Nocturia Polydipsia Dysuria	Dehydration	Hyperglycaemia exceeds the capacity of the renal tubule to reabsorb glucose in the distal convoluted tubule and increased urinary glucose creates an osmotic diuresis, producing increased urine production. Urine volume and frequency are increased, and affected individuals can report nocturia. Increase in urine volume differentiates new onset diabetes from infection or causes of bladder outflow obstruction, although high glucose levels do predispose to increased risk of infection.
Polydipsia Thirst Dizziness on standing	Dehydration Resting tachycardia Postural hypotension	Increased plasma osmolality and fluid losses from recurrent osmotic diuresis cause thirst and increased fluid intake. Intracellular dehydration occurs as water moves out of cells into the hyperosmolar interstitial fluid and circulation. In the later stages of undiagnosed diabetes preceding diabetic ketoacidosis (see below), the patient's intake may be insufficient to replace fluid losses, causing dehydration.
Weight loss	Reduced body mass index Muscle loss Loss of subcutaneous and visceral fat	Common in weeks preceding diagnosis, due to dehydration and worsening catabolic state. Lipolysis and proteolysis occur with insulin deficiency. Although weight loss is common, average body weight at presentation is usually within the normal range.
Fatigue	Excessive somnolence	Increased urinary frequency, nocturia, fluid and muscle loss lead to increased lassitude.
Visual disturbance Blurred vision	Altered visual acuity (Retinal and macular changes should be excluded by examination)	Hyperglycaemia causes osmotic changes in the lens and chambers of the eye that alter refraction; reduction in ambient glucose that occurs with treatment restores fluid levels in the dehydrated eye chamber, correcting the refraction capacity of the lens.
Musculoskeletal Leg pains, cramps	Muscle cramps, especially lower extremities	Excessive electrolyte loss with diuresis, altered carbohydrate, fat and protein metabolism, and subsequent ketosis formation can lead to cramping.
Infection Skin infections Dysuria Genital soreness	Cellulitis Abscesses Otitis externa Conjunctivitis Vaginal thrush Vulvovaginitis Balanitis Urinary tract/Genital infection	Increased glucose levels in body secretions encourage culture of bacteria, fungi, and viruses. Glucose levels should be tested in any individual presenting with significant skin or urogenital infection. Insulin deficiency may also have adverse effects on neutrophil function, and it is not uncommon for type 1 diabetes to present at times of infection as the stress response increases sympathetic activation, cortisol, and growth hormone release, all of which produce further hyperglycaemia.
Gastrointestinal symptoms Nausea Constipation (paediatric cases)	Oral and oesophageal candidiasis may complicate newly presenting type 1 diabetes.	Nausea and bloating result in physiological delay of gastric emptying with concurrent hyperglycaemia. Constipation noted in paediatric presentations, arising as a consequence of constipation from dehydration. Diarrhoea is rare, unless associated with coeliac disease at presentation.
Diabetic ketoacidosis		
Extreme thirst Excessive polyuria Vomiting Hyperventilation Reduced consciousness	Dehydration Increased skin turgor Resting tachycardia Postural hypotension Sunken fontanelle (infants) Kussmaul breathing Coma	Fluid losses excessive due to osmotic effects of profound hyperglycaemia. Acute renal failure may be present, and hyperkalaemia occurs as potassium cannot be moved into cells with insulin insufficiency. Unrestrained lipolysis and ketone formation cause nausea and subsequently vomiting. Ketones are acidic, reducing plasma pH. Metabolic acidosis drives respiratory compensation and hyperventilation can ensue, producing a deep sighing respiration. The breath smells of ketones, likened to pear drops. Rising osmolality and reduction in plasma pH reduce consciousness; often a late sign of established ketoacidosis necessitating intensive care admission.

might reduce the risk of cardiovascular events by more than 40%. Hazard ratio for ischaemic stroke is three times that of control subjects, increasing incrementally with worsening glycaemic control [11]. Amputation risk is 30 times that of the general population [12], with 20% of peripheral vascular disease patients dying within 2 years of symptom onset. Incidence of hypertension is increased in patients with type 1 diabetes, and attenuation of cardiovascular risk factors (blood pressure and cholesterol) remains paramount to reduce the risk of fatal coronary heart disease, although guidance is often based on outcomes from type 2 diabetes intervention studies [13].

Microvascular Complications Influence Morbidity

Prolonged exposure to hyperglycaemia particularly affects tissues that cannot downregulate glucose uptake, leading to raised intracellular concentrations and mitochondrial overproduction of superoxides that cause vascular damage.

Retinopathy

Each 5-year increase in the duration of diabetes at baseline is associated with a 10% increase in the risk of diabetic retinopathy and a 26% increase in the risk of sight-threatening severe retinopathy. Childhood diabetes exposure, puberty, and pregnancy contribute to added risk. Complication rates are improving; incidence of severe diabetic retinopathy in the United Kingdom decreased non-significantly from 48 events per 10 000 person-years in 2004 to 25.4 events per 10 000 years in 2014 [14]. Treatment options such as pan-retinal laser photocoagulation can control neovascularization and prevent blindness, although 25-year cumulative incidence of severe visual impairment (defined as visual acuity poorer than 6/60

in the better eye) has been reported at 3% [15]. In addition, lifetime risk of developing cataracts is increased threefold and glaucoma risk by nearly 50% in individuals with type 1 diabetes.

Nephropathy

Incidence of diabetic nephropathy increases linearly with duration of diabetes, with genetic propensity peaking at 10–14 years. Microalbuminuria develops in 30–40% of individuals with type 1 diabetes, although a significant proportion of patients can revert to normoalbuminuria with improvements in glycaemic control and blood pressure management. A third of microalbuminuria patients progress to macroalbuminuria over a decade [16], of which two-thirds later develop end-stage renal disease. Kidney disease accounts for a fifth of deaths in type 1 diabetes and all stages of nephropathy are associated with increased risk of death from cardiovascular disease (2–3-fold with microalbuminuria, 10-fold with overt proteinuria).

Neuropathy

Chronic distal symmetrical neuropathy with reduced sensation in the lower limbs and feet occurs in approximately a third of patients with 20–25 years' type 1 diabetes, although encouragingly incidence has decreased with improvements in glycaemic control. Distal neuropathy increases the likelihood of ulceration and amputation in diabetes, and in Westernized countries 7% of patients with diabetes are managed for foot ulcers. Abnormalities of the autonomic nervous system can manifest as gustatory sweating, altered bowel habit, postural hypotension, impotence, and blunting of physiological heart-rate variations; overt gastroparesis with vomiting and bladder dysfunction are rare.

Hypoglycaemia

Insulin therapy is life-sustaining and enables glycaemic control. However, hypoglycaemia is a potentially lethal side effect, and repeated exposure can result in impaired awareness of low blood glucose levels. Impaired awareness affects an estimated 20–30% of individuals with type 1 diabetes, and is associated with increased risk of severe hypoglycaemia requiring third party assistance. One episode has potential implications regarding lifestyle choices and driving but ultimately it can be truly life-threatening: not waking from sleep during nocturnal hypoglycaemia increases the risk of 'dead-in-bed' syndrome and sudden death [17]. The frequency of impaired awareness of hypoglycaemia has not declined in the past two decades, and despite major advances in conventional insulin therapy, there remains a group of patients in whom recurrent severe hypoglycaemia is associated with increased mortality (see Chapter 15.5.5 'Hypoglycaemia in the Treatment of Diabetes Mellitus').

Autoimmune Associations

Clustering of autoimmune disorders due to shared genetic susceptibility increases the lifetime risk of other autoimmune conditions developing in individuals with type 1 diabetes. A prevalence of 6% was recently reported [6], and the most common autoimmune conditions associated with type 1 diabetes are listed in Table 15.2.2.2.

Table 15.2.2.2 Autoimmune associations of type 1 diabetes

Condition	Autoantibody prevalence in type 1 diabetes	Screening test
Thyroid disease (Hyper and hypothyroidism)	15–30% express thyroid autoantibodies	Thyroid function tests (British Thyroid Association and American Diabetes Association suggest annual screening)
Addison's disease	1–2% express adrenal antibodies	9 am cortisol Short synacthen test
Coeliac disease	5–10% express coeliac antibodies	Tissue transglutaminase antibodies Duodenal biopsy if positive
Gastritis and pernicious anaemia	Antiparietal cell antibodies in 5–15%	Full blood count Vitamin B_{12} levels Antiparietal cell antibodies Iron studies
Vitiligo Alopecia areata	Rarely measured Reported frequency 2–10%	Clinical diagnosis Skin biopsy rarely used to confirm diagnosis

Future Management Strategies Might Influence Natural History

Redefining type 1 diabetes as having a silent or asymptomatic phase defined by autoantibody status and genetic risk could allow therapeutic interventions to be given earlier in the natural history of the disease, with the goal of maintaining β-cell mass and endogenous insulin production. Organized trial networks now facilitate testing of immunotherapy agents which might be capable of preventing disease onset. However, the current risk-benefit ratio is such that interventions have not yet provided sufficient evidence for routine clinical use to be indicated. Given that exogenous insulin therapy has considerably increased life expectancy for type 1 diabetes, a high safety bar has been set for immune-based interventions, and they must achieve long-lasting outcomes with minimal side effects before they can replace insulin therapy.

REFERENCES

1. Mbanya JC, Motala AA, Sobngwi E, Assah FK, Enoru ST. Diabetes in sub-Saharan Africa. *Lancet*, 2010; **375**: 2254–66.
2. Barrett JC, Clayton DG, Concannon P, *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*, 2009; **41**: 703–7.
3. Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med*, 1986; **314**: 1360–8.
4. Winter WE, Schatz DA. Autoimmune markers in diabetes. *Clin Chem*, 2011; **57**: 168–75.
5. Oram RA, McDonald TJ, Shields BM, *et al.* Most people with long-duration type 1 diabetes in a large population-based study are insulin microsecretors. *Diabetes Care*, 2015; **38**: 323–8.
6. Bravis V, Kaur A, Walkey HC, *et al.* Relationship between islet autoantibody status and the clinical characteristics of children and adults with incident type 1 diabetes in a UK cohort. *BMJ Open*, 2018; **8**: e020904.
7. Morse J. Diabetes in infancy and childhood. *Boston Med Surg J*, 1913; **168**: 530–5.
8. Harjutsalo V, Forsblom C, Groop PH. Time trends in mortality in patients with type 1 diabetes: nationwide population based cohort study. *BMJ*, 2011; **343**: d5364.

9. Livingstone SJ, Levin D, Looker HC, *et al.* Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008–2010. *JAMA*, 2015; **313**: 37–44.
10. Orchard TJ, Costacou T, Kretowski A, Nesto RW. Type 1 diabetes and coronary artery disease. *Diabetes Care*, 2006; **29**: 2528–38.
11. Stahl CH, Lind M, Svensson AM, Gudbjornsdottir S, Martensson A, Rosengren A. Glycaemic control and excess risk of ischaemic and haemorrhagic stroke in patients with type 1 diabetes: a cohort study of 33 453 patients. *J Int Med*, 2017; **281**: 261–72.
12. Khanolkar MP, Bain SC, Stephens JW. The diabetic foot. *QJM*, 2008; **101**: 685–95.
13. de Boer I, Kestenbaum B, Rue TC, *et al.* for the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group. Insulin therapy, hyperglycemia, and hypertension in type 1 diabetes mellitus. *Arch Intern Med*, 2008; **168**: 1867–73.
14. Mathur R, Bhaskaran K, Edwards E, *et al.* Population trends in the 10-year incidence and prevalence of diabetic retinopathy in the UK: a cohort study in the Clinical Practice Research Datalink 2004–2014. *BMJ Open*, 2017; **7**: e014444.
15. Klein R, Lee KE, Gangnon RE, Klein BE. The 25-year incidence of visual impairment in type 1 diabetes mellitus the wisconsin epidemiologic study of diabetic retinopathy. *Ophthalmology*, 2010; **117**: 63–70.
16. Rossing P, Hougaard P, Parving HH. Progression of microalbuminuria in type 1 diabetes: ten-year prospective observational study. *Kidney Int*, 2005; **68**: 1446–50.
17. Gill GV, Woodward A, Casson IF, Weston PJ. Cardiac arrhythmia and nocturnal hypoglycaemia in type 1 diabetes—the 'dead in bed' syndrome revisited. *Diabetologia*, 2009; **52**: 42–5.

15.2.3 Pathogenesis

Ayat Bashir, Richard A. Oram, and F. Susan Wong

Introduction

The autoimmune basis for human type 1 diabetes is well established. It is recognized that there may be considerable heterogeneity in the features of disease, and that type 1 diabetes is not a disease only of children, but also presents in adulthood. There are several lines of evidence supporting the autoimmune nature of type 1 diabetes, and these are summarized in Figure 15.2.3.1; each of these will be discussed in turn in the sections next.

Adding to the heterogeneity of disease, the period of time from initiation of the disease process to clinical presentation may range from weeks (in those presenting before the age of 5 years) to years in those presenting at an older age. A current model setting out the early stages of disease pathogenesis is shown in Figure 15.2.3.2. There are genetic and environmental risk factors, which may or may not lead to the development of autoimmunity. In stage 1, following initiation of the immune process, there is autoimmunity against beta-cells but normal blood glucose. As more beta-cells are destroyed there is asymptomatic dysglycaemia, finally leading to presentation of symptomatic diabetes.

Genetics—Association with Immune System Genes

Type 1 diabetes is a polygenic disease with a significant heritable component stemming from multiple risk genes [1, 2]. Monozygotic twin concordance for type 1 diabetes is 30–70%, compared to dizygotic twin concordance of 6–10%. Risk for people who have a first degree relative affected varies from 3% with an affected mother, to 5% with an affected father, and 8% with an affected sibling. Type 1 diabetes is most common in white Caucasian populations, but can occur in all races and ethnicities. Genetic associations of type 1 diabetes highlight the critical role of the immune system in type 1 diabetes, with the human leucocyte antigen (HLA) class II genes having the strongest association. Large differences in incidence of type 1 diabetes in genetically similar populations [3], and the increasing incidence of type 1 diabetes in low genetic risk individuals, highlight the importance of the interaction between genetic and environmental risk factors in type 1 diabetes development.

The Main Susceptibility Locus: Human Histocompatibility Complex (HLA)

The association of T1D with the HLA region on chromosome 6p21 was determined over 40 years ago [4, 5]. This region is now known to contribute around 50% of genetic susceptibility to type 1 diabetes [6]. The strongest association is with highly polymorphic HLA DR and DQ genes, which are commonly inherited together, due to tight genetic linkage between DR and DQ alleles. These genes encode proteins that display self and non-self-peptides on the surface of antigen-presenting immune cells. The highest risk haplotypes include DR4-DQ8 (DRB1*0401 DQA1*0301-DQB1*0302) and DR3-DQ2 (DRB1*0301-DQA1*0501-DQB1*0201) and at least one of these alleles is found in 85% of white Caucasians with T1D. The primary genetic drivers of risk seem to be amino acid and protein structural changes in the antigen binding pockets of HLA class II molecules at 3 key amino acid positions in the DRB1 and DQB1

Figure 15.2.3.1 Indicators of the autoimmune pathogenesis of type 1 diabetes.

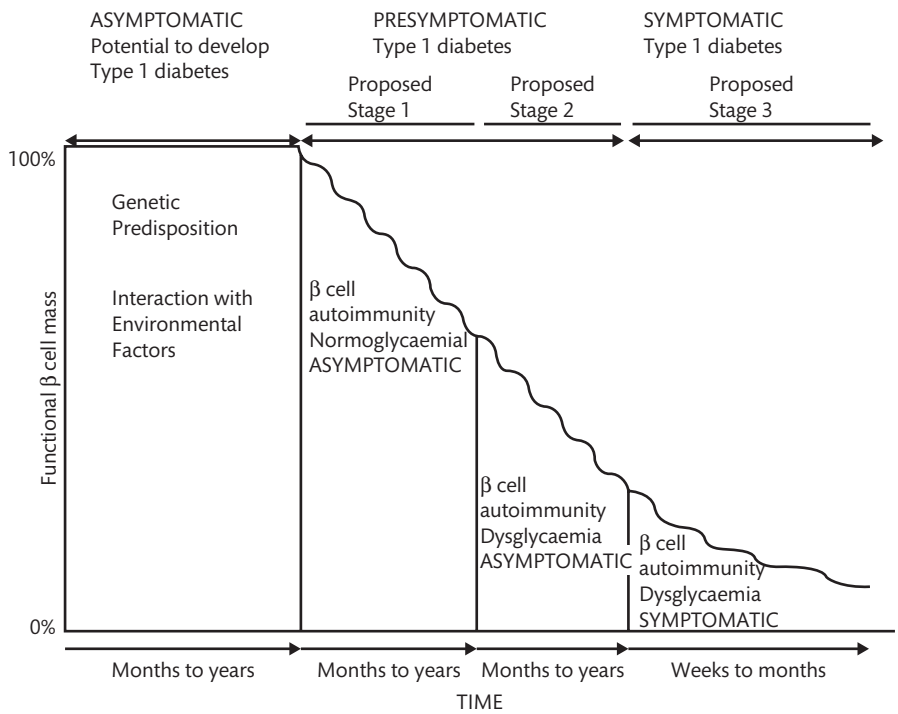


Figure 15.2.3.2 Early stages of type 1 diabetes.

Source data from Insel RA, Dunne JL, Atkinson MA, Chiang JL, Dabelea D, Gottlieb PA, *et al*. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes care*. 2015;38(10):1964–74. (1)

genes [7]. This presumably impacts the binding affinity, the repertoire of peptides that can be presented and their impact on the T cells they bind. Certain DR-DQ alleles (such as DRB1*1501-DQA1*0102-DQB1*0602) confer strong dominant protection from type 1 diabetes [8]. The mechanisms for this are not fully understood but may offer valuable insights into how to reduce or prevent islet-specific autoimmunity.

Additional associations with the HLA region include the class II DPB1 locus, and HLA class I A, B, and C loci. Alleles with particularly strong associations include the protective B*57:01 allele, and B*39:06 and A24 risk alleles [9]. HLA class I/peptide antigen complexes are involved in in the development of the $CD8^+$ T-cell repertoire in the thymus and antigen-specific T-cell-mediated cytotoxicity.

Non-HLA Loci are Important Contributors to Genetic Risk of Type 1 Diabetes and Reveal Further Immune Insights

Linkage and genome wide association studies have demonstrated over 50 additional non-HLA genetic associations with type 1 diabetes risk [2]. The majority of these associations are with genes in key immune pathways and further highlight the importance of the immune system function and regulation in the pathogenesis of type 1 diabetes [10] (Figure 15.2.3.3).

The strongest non-HLA genetic association, initially described as the IDDM2 locus, was discovered in the early 1980s, and is a variable number of tandem repeats (VNTR) polymorphism next to the insulin gene (INS) on chromosome 11p15 [11]. The class III protective VNTR is associated with increased insulin expression in the thymus and may increase negative selection of insulin-specific autoreactive T cells.

The PTPN22 gene on chromosome 1p13 encodes the lymphoid-specific phosphatase LYP, which normally suppresses T-cell activation [12]. A gain of function variant in PTPN22 may lead to a failure of deletion of autoreactive T cells during thymic selection, and altered function of effector and regulatory T (Treg) lymphocytes, and B lymphocytes in the periphery. Cytotoxic T-lymphocyte–associated protein 4 (CTLA4), encoded on chromosome 2q33, plays a role in T-cell development and it is a negative regulator of T-cell activation. Its expression in both regulatory and effector T cells is essential for suppression of autoreactive T-cell-mediated cytotoxicity. The IL-2 receptor alpha (IL-2RA) gene encodes the IL-2 receptor alpha chain, CD25. The T1D-susceptibility *IL2RA* association is associated with decreased signalling via the IL-2 pathway in both memory T cells and Tregs and this is linked to diminished Treg function. These and other genetic immune associations (reviewed in [2, 13]) have led to better understanding of mechanisms of islet-specific autoimmunity, and identified pathways that are being targeted for therapeutic intervention.

Not all genetic associations of type 1 diabetes point to adaptive immunity. Both common and rare deleterious variants in IFIH1 are associated with type 1 diabetes [14]. IFH 1 encodes melanoma differentiated associated protein 5 (MDA5) which binds to double-stranded RNA viruses and thus mediates the innate immune system's interferon response to certain viruses. A recent study demonstrated evidence for a causal relationship between the common IFIH variant (rs1990760) and both autoimmunity and viral resistance [15]. Additionally there is emerging evidence that several type 1 diabetes genes are expressed in insulin-producing beta-cells and abnormalities within the beta-cell may contribute to type 1 diabetes pathogenesis [10].

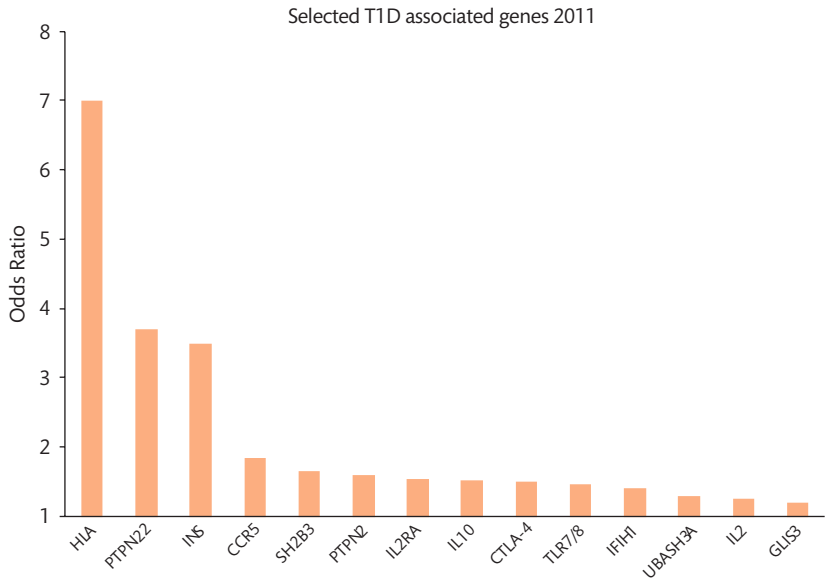


Figure 15.2.3.3 Genetic associations with type 1 diabetes. Genome-wide association studies have identified a number of genes associated with type 1 diabetes. Colour-coding designates year of discovery of these candidate genes. The *y*-axis indicates the best estimate of the odds ratio (OR) for risk alleles at each of the indicated loci. For each genomic region where convincing association with type 1 diabetes has been reported, the gene of interest, or containing the most associated SNP is indicated on the x-axis. The majority of these genes are implicated in the immune response, but several of the non-HLA genes are expressed in human pancreatic islets (marked with *).

Reproduced with permission from Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G, *et al.* Genetics of type 1 diabetes: what's next? *Diabetes.* 2010;59(7):1561–71.

The majority of genetic linkage and association studies have been in groups of type 1 diabetes cases compared to control individuals. However, there is increasing awareness of heterogeneity of aetiopathology, immune phenotype, age of presentation, and beta-cell loss in type 1 diabetes. To elucidate the genetic contribution to this heterogeneity will require large association studies of these features within cohorts of people who progress to type 1 diabetes and well-phenotyped individuals with type 1 diabetes. It is possible, as is the case with some other autoimmune diseases, that genetic associations and predictors of disease progression may be independent of risk loci. Initial analyses have highlighted independent genetic associations with age of diagnosis [16], beta-cell loss [17], and incidence of islet autoantibody development [18].

Thus, there have been considerable advances in determining genetic factors that lead to increased susceptibility to type 1 diabetes. However, it is clear that genetic susceptibility alone is not sufficient for diabetes to develop, and environmental factors have been implicated in triggering or perpetuating the immune responses that lead to islet beta-cell damage, discussed in the next section.

What is the Role of the Environment in the Loss of Immune Tolerance?

Several studies have been designed to characterize the progression over time to islet autoimmunity and clinical disease in children at risk of type 1 diabetes, including The Environmental Determinants of Diabetes in the Young (TEDDY) [19], and the Diabetes Autoimmunity Study in the Young (DAISY) [20]. Evidence emerging from these and other studies is consistent with diverse genetic and environmental factors driving initiation of islet autoimmunity and progression to type 1 diabetes.

Infections, particularly viral, have long been implicated as a potential cause of type 1 diabetes. Evidence of persistent enteroviral infection has come from detection of enteroviral VP1 protein, and RNA in the beta-cells of individuals with type 1 diabetes [21, 22]. However, whether this is a causative factor or an additional factor causing stress to beta-cells is unknown. Recently, considerable interest has been focused on the intestinal microbiome. The gut bacteria are acquired at the time of birth, and the microbiota increase in diversity, particularly over the first 3 years of life [23]. Children who develop type 1 diabetes have reduced diversity [24] of the gut microbiota. These commensal bacterial influence host metabolism, and have an important role in maturation of the immune system and development of immune tolerance. In animal models, gut microbes could activate autoreactive T cells and contribute to development of diabetes by molecular mimicry [25], but this has yet to be proven in humans. Other influences in early life, such as dietary factors including early introduction of cows' milk, cereals, polyunsaturated fatty acids, and vitamin D have all been considered as possible determinants, but as yet, no firm conclusions can be drawn [26, 27]. Vaccines and antibiotic use have not been shown to be a factor in diabetes development. Large prospective studies will be important to help throw light on this, although heterogeneity of type 1 diabetes may continue to make this a very difficult task.

Thus, either one or more of these environmental factors may be important in both triggering and perpetuating the immune responses that damage the islet beta-cells. A proposed scheme for the components involved in the immune damage is shown in Figure 15.2.3.4, and these will be discussed further in the sections that follow.

Insulitis—Beta-Cell Antigen-Specific T Cells Found in Human Islets

The earliest evidence for immune cell involvement was infiltration of mononuclear cells in the pancreatic islets [28, 29]. Autopsies of young patients who died shortly after diagnosis of diabetes showed variable degrees of lymphocytic infiltration, resembling that seen in autoimmune thyroiditis. In a large series of post-mortem examinations, infiltration was found in insulin-containing islets, with few cells found in insulin-deficient islets [30]. More recently, the insulitis has been further characterized in post-mortem samples where $CD8^+$ T cells and B cells are found more commonly than $CD4^+$ cells, with other cell types including $CD68^+$ macrophages [31]. In addition, biopsies of pancreas in newly diagnosed individuals confirmed $CD8^+$ T cells to be the most abundant cells present [32]. Furthermore, infiltrating $CD8^+$ T cells have been shown to be islet autoantigen-reactive to proinsulin, glutamic acid decarboxylase, and islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) [33, 34]. Nevertheless, it is perhaps somewhat surprising that relatively few cells are found in the islets of the pancreatic samples that have been studied post-mortem, and this may reflect the fact that the immune processes may have commenced a considerable time before diagnosis.

Humoral Immunity—The Presence of Anti-Islet Autoantibodies in the Circulation

The identification of islet cell antibodies (ICA) by immunofluorescence in 1974 was a key step in the recognition of type 1 diabetes as an immune-mediated disease. ICA have now been characterized at the molecular level and are summarized in Table 15.2.3.1. Circulating islet autoantibodies are present in sera from ~95% of children with new-onset diabetes, and provide evidence of an active and disease-specific B lymphocytic response. Furthermore, the observation that ICA preceded the development of diabetes in high-risk relatives led to the understanding that onset of autoimmunity commonly occurs years before clinical presentation with diabetes. Although individuals with single autoantibody positivity do not necessarily progress to a decline in beta-cell function, the presence of two or more high-titre autoantibodies predicts development of type 1 diabetes over the following 10–15 years [35]. The risk of developing autoantibodies in these relatives is high in the youngest children and reduces with age. Models have been developed to estimate the risk of developing type 1 diabetes at different ages from birth through childhood, and was greatest in younger islet autoantibody positive children [88].

While anti-islet autoantibodies may appear many years before clinical onset of diabetes and allow prediction and classification of type 1 diabetes, the autoantibodies are not thought to be effectors of the immune attack; rather they are a sign of autoreactive B cell

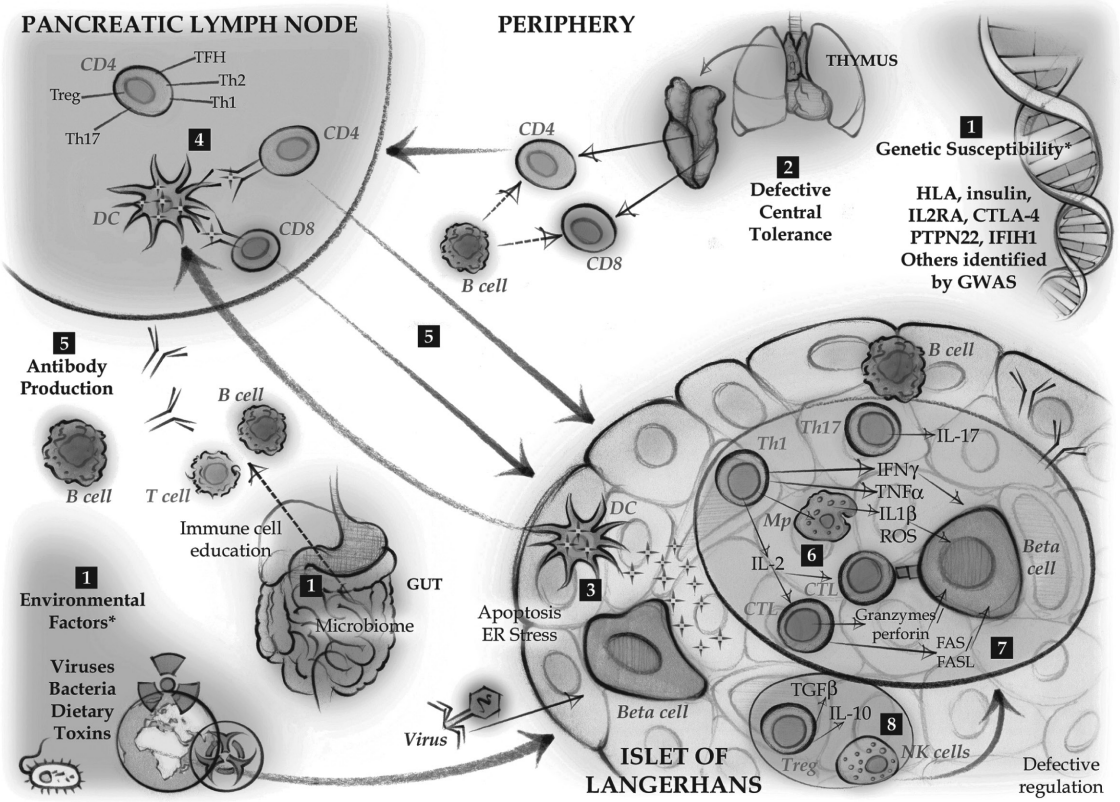


Figure 15.2.3.4 Proposed scheme for pathogenesis of type 1 diabetes. (1) Genetic and environmental factors, such as viral infections, are likely to impact on loss of peripheral tolerance, a major contributing factor in the pathogenesis of type 1 diabetes. The intestinal microbiome has also been implicated as an environmental factor in pathogenesis. (2) Defective central tolerance allows naïve islet-reactive $CD4^+$ and $CD8^+$ T cells to escape the thymus and migrate to the pancreatic lymph node. (3) Several possible mechanisms, including remodelling by apoptosis, viral infection, endoplasmic reticulum (ER) stress could promote beta-cell death and release of beta-cell antigens. Antigens can be taken up by dendritic cells (DC), which migrate to the pancreatic lymph node. (4) Dendritic cells present beta-cell antigens to naïve CD4 T cells, leading to activation of several possible subsets including: Th1, Th2, TFH, Th17, Treg. Dendritic cells also cross-present antigens to $CD8^+$ T cells. (5) $CD4^+$ cells can help B lymphocytes to produce autoantibodies that target beta-cell proteins. They may also assist in activation and migration of $CD8^+$ T cells. Activated T cells traffic to the islets. (6) $CD4^+$ Th1 secrete pro-inflammatory cytokines IFNγ and TNFα, which could induce beta-cell death and stimulate macrophages to produce reactive oxygen species (ROS), TNFα and IL1β. This may augment beta-cell death. (7) Furthermore CTL infiltration can induce direct killing of beta-cell by lysing beta-cells presenting self-antigen, via secretion of granzymes/perforin and through the Fas–FasL pathway. (8) Treg cells function to reduce autoimmune responses through interaction with dendritic cells and secretion of TGFβ and IL-10. In diabetes, regulatory T-cell function is impaired.
Adapted with permission from van Belle TL, Coppieters KT, von Herrath MG. Type 1 diabetes: etiology, immunology, and therapeutic strategies. *Physiological reviews*. 2011;91(1):79–118. Artwork courtesy of Juliet Percival.

activation. As many of these autoantibodies are IgG, suggesting that they are T-cell-dependent antibodies, they are also likely to indicate autoreactive T-cell activity. It is interesting to note that many of the islet autoantibodies are directed towards components of the islet secretory granule, and some of these are also targets of T lymphocytes (see next).

Cellular Immunity–Pancreas Transplantation, Adoptive Transfer of Diabetes and Islet Beta-Cell Antigen-Responsive T Cells

Understanding of immune cell subsets has increased considerably over the last few years, and many of these subsets, shown in Figure 15.2.3.4, participate in autoimmune reactivity. As illustrated, the interactions are complex and the diversity of cells involved is likely to make immune intervention potentially difficult. Much information relating to cellular immunity has initially been derived from rodent models that develop spontaneous autoimmune diabetes. The most studied of these models is the non-obese diabetic (NOD) mouse, which has given us early insights and signposted cellular immune processes. Like humans, the most important genetic susceptibility factor is in the MHC class II region, and the structure of the MHC class II molecule most closely associated with diabetes, is very similar in both humans and NOD mice. As in humans, environmental factors also play a role, which contributes to the variation in incidence of diabetes in different mouse colonies across the world, as well as over time within the same colony [49]. Some of the target T-cell antigens in NOD mice, which are also targeted in humans, are very similar to the autoantibody targets (Table 15.2.3.2). Early studies in the NOD mouse indicated that both $CD8^+$ and $CD4^+$ T cells as well as B cells themselves (rather than antibodies) are likely to be important in the developing cellular responses against islet beta-cells (reviewed in [50]). This has been

Table 15.2.3.1 Autoantibodies in type 1 diabetes

Target antigen	Key information	References
Insulin (IAA)	• Presence demonstrated using sensitive radiobinding assay. • Appear very early in the natural history of type 1 prodrome. • Overall prevalence in childhood diabetes (<15 years) has been reported as 70%, with higher prevalence in younger children. • High affinity IAA strongly associated with *HLA DRB1*04*, subsequent progression to multiple autoantibody positivity, binding to human insulin A chain residues 8–13, and binding to proinsulin.	[36–38]
Glutamate decarboxylase 65 kD molecular weight isoform (GAD65)	• Originally discovered by immunoprecipitation of a 64kDa protein from 35S methionine-labelled rat islets by sera from patients with type 1 diabetes. • Found in >70% of people with type 1 diabetes at all ages. GAD is not confined to beta-cells, and is also found in other cell types. • GAD antibodies show a positive correlation with age and can persist longer than other islet cell autoantibodies. • The affinity and epitope specificity of the antibody response predicts progression of the disease process.	[37, 39, 40]
Insulinoma-associated antigen-2 (IA2)	• Identified by screening patient sera with an islet cDNA expression library. • Found in 65–75% patients at diagnosis. • Almost always detected with other islet autoantibodies and are very specific for type 1 diabetes. • More likely to appear closer to clinical onset. • Critical epitope regions/residues for IA-2A and IA-2β antibodies have been defined, with a hierarchy of risk for future development of type 1 diabetes. • Subreactivity to IA-2β is strongly associated with progression to diabetes within 5 years.	[41–44]
Zinc-transporter isoform 8 (ZnT8)	• Discovered by screening highly-expressed islet beta-cell specific molecules. • Found in 60–80% patients at diagnosis. • Typically disappear very early after disease onset. • Autoimmunity against the COOH-terminal region of ZnT8 is of particular prognostic significance. • Greatest risk of disease progression occurs in ZnT8A-positive children who are homozygous for either arginine or tryptophan at position 325.	[45, 46]
Tetraspanin-7	• Identified by mass spectrometry. • Found in approximately 19–38% of patients. • Significantly higher prevalence in children. • In addition to other antibodies, Tetraspanin-7 provides only minor further value in identifying beta-cell autoimmunity. • Not known to influence disease progression.	[47, 48]

facilitated by the ability to clone islet-reactive T cells, and to study their targets and mechanism of action. Furthermore, of note in considering the immune role in destruction of islet beta-cells, T cells, both polyclonal populations from diabetic NOD mice as well as individual cloned T cells, are able to transfer diabetes to non-diabetic immunocompromised mice.

The strongest MHC susceptibility factor associated with the immune response in type 1 diabetes is in the MHC class II region, as noted in Section 1, coding for antigen-presenting molecules that are required for $CD4^+$ T cells to recognize their antigen. The autoantibodies indicate that the B cells that produce them are likely to have interacted with CD4+ T cells. However, the presence of hyperexpression of HLA class I in insulitis, genetic HLA class I associations, and $CD8^+$ T-cell transfer experiments in mice highlight that $CD8^+$ T cells are also critical to the autoimmune process. A major barrier to progress in understanding type 1 diabetes has been the difficulty in identifying and studying islet-specific autoreactive lymphocytes from insulitis lesions in humans, with a reliance on studying the very small proportion of islet autoreactive lymphocytes circulating in peripheral blood [51].

The importance of $CD8^+$ T cells in the final effector phases of type 1 diabetes was emphasized in early identical twin-to-twin pancreas transplants. Unaffected monozygotic twins in long-term discordant pairs who were thought not to require immunosuppression because they were HLA-identical, donated the tail of their pancreas to their affected co-twin. However, the graft was rapidly destroyed by recrudescence of the type 1 diabetes disease process, with a preponderance of $CD8^+$ T cells in the insulitis that developed in the transplanted organ within weeks of transplantation and unfortunately, rapid recurrence of diabetes occurred [52].

Evidence for transfer of diabetes by T cells in humans comes from bone marrow transplantation, from donors with type 1 diabetes to non-diabetic recipients. In nine cases in which the donor had type 1 diabetes, 2 of 3 long-term survivors developed type 1 diabetes [53]. This transfer of disease can be prevented by depletion of mature T cells prior to transplantation, which is now routine practice.

What do the T cells recognize? Many, but not all, of the autoantigenic targets are those that are also targets of the humoral response, shown in Table 15.2.3.2. $CD8^+$ T cells recognizing islet autoantigens can be found within the islets, as highlighted earlier, as well as in the peripheral blood [54, 55] and some of these also demonstrate the ability to damage islets *in vitro* [56]. Under normal circumstances, autoreactive T cells are removed in the thymus, by the process of central tolerance. T cells that have sufficient affinity for peptides of self-antigens, presented on medullary thymic epithelial cells within the thymus, will be deleted. It is recognized that a number of the islet autoantigens recognized, are either not presented in the thymus or the peptide recognized by the islet antigen-specific T cells are of very low affinity and do not trigger thymic deletion. Furthermore, some of the genetic susceptibility factors associated with type 1 diabetes encode alleles that reduce thymic negative selection, as discussed in the earlier genetics section, and reviewed in [57]. In addition, recent studies have indicated that some autoantigenic targets recognized by

Table 15.2.3.2 Major autoantigens targeted by T cells in humans

Autoantigen	Responding T cell		References
	CD4	CD8	
Insulin/Proinsulin	✓	✓	[60, 61]
GAD65	✓	✓	[62, 63]
Islet-specific glucose-6-phosphase catalytic subunit-related protein (IGRP)		✓	[63–65]
Insulinoma-associated antigen-2	✓	✓	[64, 66, 67]
Insulin antigen-2-Phogrin	✓		[68]
ZnT8	✓	✓	[69, 70]
Insulin hybrid peptides	✓		[58, 59, 71]
Defective ribosomal insulin gene product		✓	[72]

$CD4^+$ T cells may be hybrid peptides of two autoantigens, generating unique antigens within the beta-cells or local environment (58, 59). This may further explain why some autoreactive T cells are not eliminated by central tolerance mechanisms in the thymus during development.

The role of pathogenic $CD4^+$ T cells is less clear, as fewer of these are found within the islet infiltrate, but it is likely that they may help activation of both $CD8^+$ T cells as well as B cells to produce autoantibodies, although this has not been shown directly in man. In recent years, newer subsets of $CD4^+$ T cells including T follicular helper (T_{FH}) cells were shown to be increased in the circulation [73]. These T_{FH}, which are found in germinal centres and aid B cells in class-switching of antibodies, are important in autoantibody production. Furthermore, among the many subsets of $CD4^+$ T cells, the Treg cells, involved in terminating immune responses, are important in preventing autoimmunity under normal circumstances. While, overall, these cells are not altered in number, there is evidence that their function is defective in type 1 diabetes, being less able to control cytokine production and proliferation of effector $CD4^+$ T cells [74]. This may be related to a variety of mechanisms that are genetically encoded and include reduced sensitivity to IL-2, increased apoptosis, alteration in Treg function and instability of the expression of forkhead box P3 (FOXP3), a central transcription factor for a major subset of Tregs, and alteration of other genes involved in T reg function reviewed in [75]. In addition, not only have defects in these cells been shown in peripheral blood, but also in pancreatic lymph nodes, the draining nodes of the pancreas [76].

How are the Beta-Cells Damaged or Destroyed?

Type 1 diabetes is a disease of progressive, severe insulin deficiency caused by autoimmune beta-cell destruction. Despite historical dogma, it is increasingly clear that progression of beta-cell destruction is variable and does not always lead to complete beta-cell loss. Post-mortem histological studies of individuals who have had diabetes for more than 50 years, have indicated that some pancreatic beta-cells remain insulin-positive [77]. Almost all people with type 1 diabetes have measurable beta-cell function at diagnosis, and recently, highly sensitive C-peptide assays have shown that even years after diagnosis, many people with type 1 diabetes, particularly those diagnosed as adults, have low levels of detectable beta-cell function [77]. The source of remaining beta-cells is unclear and could relate to a "burning out" of autoimmunity, some beta-cells being more resistant to immune attack, or regeneration of beta-cells.

The islet beta-cells are likely to be damaged by a combination of both necrosis and apoptosis. Cytotoxic CD8+ T cells play a role in immune damage, and release of perforin and granzymes from the cytotoxic granules damage the cell membrane. Islet beta-cells are also particularly susceptible to damage by cytokines [78] and reactive oxygen species (ROS) [79]. The cytokines and ROS cause mitochondrial damage, and lead to increased permeability, disrupting oxygen phosphorylation and production of ATP. However, apoptosis may also play a role, and this may also be initiated by cytokine damage by TNF, as well as activation through the Fas/Fas ligand pathway.

Recent advances in imaging of beta-cells will enhance understanding of the rate of loss, and will allow the loss of beta-cell mass to be quantified. The insulin-containing beta-cell granules express vesicular monoamine transporter2 (VMAT2) and the density of this can be visualized and quantified by PET imaging [80]. Further advances in this and other MRI imaging techniques, may considerably aid the understanding of the processes of islet cell damage. In addition to the imaging techniques, turnover of islet beta-cells in the context of beta-cell death has also been studied by measurement of unmethylated insulin gene (*INS*) DNA that originates in beta-cells where the insulin gene is transcribed, compared with other tissues where no transcription takes place. Thus, these epigenetic changes can be estimated, by unmethylated to methylated *INS* gene measurements in serum by polymerase chain reaction (PCR), to provide an indication of beta-cell death. While there are limitations to this method, reviewed in [81], refinements may allow functional measurements of beta-cell mass and indicate ongoing beta-cell destruction.

Immune Intervention

As evidence points to autoimmunity as a major contributor to the pathogenesis of type 1 diabetes, this should be amenable to therapy targeting the immune system. Early trials of cyclosporin A showed that there was an apparent delay in loss of beta-cells using a non-specific immunosuppressive approach [82, 83]. However, despite much optimism that type 1 diabetes can be halted or reversed by immune intervention, in the human studies to date, this goal has not yet been achieved [84]. This may, in part, be because of the complexity and heterogeneity of the immune process in type 1 diabetes, or the fact that too many beta-cells are lost by the time of presentation. A few recent studies have shown some slowing of loss of beta-cells, as detected by preservation of C-peptide, within the first year with anti-CD3 (teplizumab, otelixizumab) [85, 86] as well as anti-CD20 targeting B cells [87]. Further promise has also been shown with agents targeting costimulatory molecules. Of importance would be the ability to more specifically target either pathogenic cells with antigen-specific tolerance strategies or boosting regulatory T cells, and clearly there is much scope for further research. These will be discussed in more detail in Chapter 15.6.3, 'Immunotherapy in Type 1 Diabetes'.

Summary

There have been considerable advances in our understanding of the genetic susceptibility and the immune pathogenesis of type 1 diabetes. However, there are still unanswered questions about the pathogenesis of type 1 diabetes, particularly in relation to the environmental factors that may trigger the immune attack, and disease heterogeneity. The refinements in modern technology that will allow more sophisticated imaging, as well as advances in our ability to interrogate the immune processes at a single cell level, and understand the interaction with islet biology should allow further progress to be made in this area.

REFERENCES

1. Insel RA, Dunne JL, Atkinson MA, *et al.* Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*, 2015; **38**: 1964–74.
2. Pociot F, Lernmark A. Genetic risk factors for type 1 diabetes. *Lancet*, 2016; **387**: 2331–9.
3. Kondrashova A, Reunanen A, Romanov A, *et al.* A six-fold gradient in the incidence of type 1 diabetes at the eastern border of Finland. *Ann Med*, 2005; **37**: 67–72.
4. Nerup J, Platz P, Andersen OO, *et al.* HL-A antigens and diabetes mellitus. *Lancet*, 1974; **2**: 864–6.
5. Cudworth AG, Woodrow JC. Evidence for HL-A-linked genes in 'juvenile' diabetes mellitus. *BMJ*, 1975; **3**: 133–5.
6. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet*, 1996; **59**: 1134–48.
7. Hu X, Deutsch AJ, Lenz TL, *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet*, 2015; **47**: 898–905.
8. Noble JA. Immunogenetics of type 1 diabetes: a comprehensive review. *J Autoimmun*, 2015; **64**: 101–12.
9. Noble JA, Valdes AM, Varney MD, *et al.* HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. *Diabetes*, 2010; **59**: 2972–9.
10. Pociot F, Akolkar B, Concannon P, *et al.* Genetics of type 1 diabetes: what's next? *Diabetes*, 2010; **59**: 1561–71.
11. Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes*, 1984; **33**: 176–83.
12. Vang T, Miletic AV, Bottini N, Mustelin T. Protein tyrosine phosphatase PTPN22 in human autoimmunity. *Autoimmunity*, 2007; **40**: 453–61.
13. Robertson CC, Rich SS. Genetics of type 1 diabetes. *Curr Opin Genet Dev*, 2018; **50**: 7–16.
14. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science (New York, NY)*, 2009; **324**: 387–9.
15. Gorman JA, Hundhausen C, Errett JS, *et al.* The A946T variant of the RNA sensor IFIH1 mediates an interferon program that limits viral infection but increases the risk for autoimmunity. *Nat Immunol*, 2017; **18**: 744–52.
16. Inshaw JRJ, Walker NM, Wallace C, Bottolo L, Todd JA. The chromosome 6q22.33 region is associated with age at diagnosis of type 1 diabetes and disease risk in those diagnosed under 5 years of age. *Diabetologia*, 2018; **61**: 147–57.
17. Roshandel D, Gubitosi-Klug R, Bull SB, *et al.* Meta-genome-wide association studies identify a locus on chromosome 1 and multiple variants in the MHC region for serum C-peptide in type 1 diabetes. *Diabetologia*, 2018; **61**: 1098–111.
18. Torn C, Hadley D, Lee HS, *et al.* Role of type 1 diabetes-associated SNPs on risk of autoantibody positivity in the TEDDY study. *Diabetes*, 2015; **64**: 1818–29.
19. Krischer JP, Lynch KF, Lernmark A, *et al.* Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. *Diabetes Care*, 2017; **40**: 1194–202.
20. Frederiksen BN, Kroehl M, Baron A, *et al.* Assessing age-related etiologic heterogeneity in the onset of islet autoimmunity. *BioMed Res Int*, 2015; **2015**: 708289.
21. Krogvold L, Edwin B, Buanes T, *et al.* Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. *Diabetes*, 2015; **64**: 1682–7.
22. Richardson SJ, Willcox A, Bone AJ, Foulis AK, Morgan NG. The prevalence of enteroviral capsid protein vp1 immunostaining in pancreatic islets in human type 1 diabetes. *Diabetologia*, 2009; **52**: 1143–51.
23. Yassour M, Vatanen T, Siljander H, *et al.* Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Trans Med*, 2016; **8**: 343ra81.
24. Kostic AD, Gevers D, Siljander H, *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*, 2015; **17**: 260–73.
25. Tai N, Peng J, Liu F, *et al.* Microbial antigen mimics activate diabetogenic CD8 T cells in NOD mice. *J Exp Med*, 2016; **213**: 2129–46.
26. Rewers M, Ludvigsson J. Environmental risk factors for type 1 diabetes. *Lancet*, 2016; **387**: 2340–8.
27. van Belle TL, Coppieters KT, von Herrath MG. Type 1 diabetes: etiology, immunology, and therapeutic strategies. *Physiol Rev*, 2011; **91**: 79–118.
28. Gepts W. Pathologic anatomy of the pancreas in juvenile diabetes mellitus. *Diabetes*, 1965; **14**: 619–33.
29. Gepts W, Lecompte PM. The pancreatic islets in diabetes. *Am J Med*, 1981; **70**: 105–15.
30. Foulis AK, Liddle CN, Farquharson MA, Richmond JA, Weir RS. The histopathology of the pancreas in type 1 (insulin-dependent) diabetes mellitus: a 25-year review of deaths in patients under 20 years of age in the United Kingdom. *Diabetologia*, 1986; **29**: 267–74.
31. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG. Analysis of islet inflammation in human type 1 diabetes. *Clin Exp Immunol*, 2009; **155**: 173–81.
32. Itoh N, Hanafusa T, Miyazaki A, *et al.* Mononuclear cell infiltration and its relation to the expression of major histocompatibility complex antigens and adhesion molecules in pancreas biopsy specimens from newly diagnosed insulin-dependent diabetes mellitus patients. *J Clin Invest*, 1993; **92**: 2313–22.
33. Babon JA, DeNicola ME, Blodgett DM, *et al.* Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. *Nat Med*, 2016; **22**: 1482–7.
34. Coppieters KT, Dotta F, Amirian N, *et al.* Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent

onset and long-term type 1 diabetes patients. *J Exp Med*, 2012; **209**: 51–60.
35. Ziegler AG, Rewers M, Simell O, *et al.* Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*, 2013; **309**: 2473–9.
36. Palmer JP, Asplin CM, Clemons P, *et al.* Insulin antibodies in insulin-dependent diabetics before insulin treatment. *Science (New York, NY)*, 1983; **222**: 1337–9.
37. Williams AJ, Norcross AJ, Dix RJ, Gillespie KM, Gale EA, Bingley PJ. The prevalence of insulin autoantibodies at the onset of Type 1 diabetes is higher in males than females during adolescence. *Diabetologia*, 2003; **46**: 1354–6.
38. Achenbach P, Koczwara K, Knopff A, Naserke H, Ziegler A-G, Bonifacio E. Mature high-affinity immune responses to (pro) insulin anticipate the autoimmune cascade that leads to type 1 diabetes. *J Clin Invest*, 2004; **114**: 589–97.
39. Baekkeskov S, Aanstoot HJ, Christgau S, *et al.* Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature*, 1990; **347**: 151–6.
40. Mayr A, Schlosser M, Grober N, *et al.* GAD autoantibody affinity and epitope specificity identify distinct immunization profiles in children at risk for type 1 diabetes. *Diabetes*, 2007; **56**: 1527–33.
41. Rabin DU, Pleasic SM, Shapiro JA, *et al.* Islet cell antigen 512 is a diabetes-specific islet autoantigen related to protein tyrosine phosphatases. *J Immunol (Baltimore, MD: 1950)*, 1994; **152**: 3183–8.
42. Tsirogianni A, Pipi E, Soufleros K. Specificity of islet cell autoantibodies and coexistence with other organ specific autoantibodies in type 1 diabetes mellitus. *Autoimmun Rev*, 2009; **8**: 687–91.
43. Bonifacio E, Lampasona V, Bingley PJ. IA-2 (islet cell antigen 512) is the primary target of humoral autoimmunity against type 1 diabetes-associated tyrosine phosphatase autoantigens. *J Immunol*, 1998; **161**: 2648–54.
44. Achenbach P, Warncke K, Reiter J, *et al.* Stratification of type 1 diabetes risk on the basis of islet autoantibody characteristics. *Diabetes*, 2004; **53**: 384–92.
45. Wenzlau JM, Liu Y, Yu L, *et al.* A common nonsynonymous single nucleotide polymorphism in the SLC30A8 gene determines ZnT8 autoantibody specificity in type 1 diabetes. *Diabetes*, 2008; **57**: 2693–7.
46. Achenbach P, Lampasona V, Landherr U, *et al.* Autoantibodies to zinc transporter 8 and SLC30A8 genotype stratify type 1 diabetes risk. *Diabetologia*, 2009; **52**: 1881–8.
47. McLaughlin KA, Richardson CC, Ravishankar A, *et al.* Identification of tetraspanin-7 as a target of autoantibodies in type 1 diabetes. *Diabetes*, 2016; **65**: 1690–8.
48. Walther D, Eugster A, Jergens S, *et al.* Tetraspanin 7 autoantibodies in type 1 diabetes. *Diabetologia*, 2016; **59**: 1973–6.
49. Yurkovetskiy L, Burrows M, Khan AA, *et al.* Gender bias in autoimmunity is influenced by microbiota. *Immunity*, 2013; **39**: 400–12.
50. Pearson JA, Wong FS, Wen L. The importance of the Non Obese Diabetic (NOD) mouse model in autoimmune diabetes. *J Autoimmun*, 2016; **66**: 76–88.
51. Battaglia M, Atkinson MA. The streetlight effect in type 1 diabetes. *Diabetes*, 2015; **64**: 1081–90.
52. Sutherland DE, Goetz FC, Sibley RK. Recurrence of disease in pancreas transplants. *Diabetes*, 1989; **38** Suppl 1: 85–7.
53. Lampeter EF, McCann SR, Kolb H. Transfer of diabetes type 1 by bone-marrow transplantation. *Lancet*, 1998; **351**: 568–9.
54. Velthuis JH, Unger WW, Abreu JRF, *et al.* simultaneous detection of circulating autoreactive CD8(+) T-cells specific for different islet cell-associated epitopes using combinatorial MHC multimers. *Diabetes*, 2010; **59**: 1721–30.
55. James EA, Abreu JRF, McGinty JW, *et al.* Combinatorial detection of autoreactive CD8(+) T cells with HLA-A2 multimers: a multi-centre study by the Immunology of Diabetes Society T Cell Workshop. *Diabetologia*, 2018; **61**: 658–70.
56. Skowera A, Ellis RJ, Varela-Calvino R, *et al.* CTLs are targeted to kill beta cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. *J Clin Invest*, 2008; **118**: 3390–402.
57. Clark M, Kroger CJ, Tisch RM. Type 1 diabetes: a chronic anti-self-inflammatory response. *Front Immunol*, 2017; **8**: 1898.
58. Delong T, Wiles TA, Baker RL, *et al.* Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science (New York, NY)*, 2016; **351**: 711–14.
59. Kent SC, Mannering SI, Michels AW, Babon JAB. Deciphering the pathogenesis of human type 1 diabetes (T1D) by interrogating T cells from the 'scene of the crime'. *Curr Diabetes Rep*, 2017; **17**: 95.
60. Michels AW, Landry LG, McDaniel KA, *et al.* Islet-derived CD4 T cells targeting proinsulin in human autoimmune diabetes. *Diabetes*, 2017; **66**: 722–34.
61. Zhang L, Nakayama M, Eisenbarth GS. Insulin as an Autoantigen in NOD/Human Diabetes. *Curr Opin Immunol*, 2008; **20**: 111–18.
62. Reijonen H, Novak EJ, Kochik S, *et al.* Detection of GAD65-specific T-cells by major histocompatibility complex class II tetramers in type 1 diabetic patients and at-risk subjects. *Diabetes*, 2002; **51**: 1375–82.
63. Mallone R, Martinuzzi E, Blancou P, *et al.* CD8+ T-cell responses identify beta-cell autoimmunity in human type 1 diabetes. Diabetes, 2007; **56**:613–21.
64. Martinuzzi E, Novelli G, Scotto M, *et al.* The frequency and immunodominance of islet-specific CD8+ T-cell responses change after type 1 diabetes diagnosis and treatment. *Diabetes*, 2008; **57**: 1312–20.
65. Yang J, Danke NA, Berger D, *et al.* Islet-specific glucose-6-phosphatase catalytic subunit-related protein-reactive CD4+ T cells in human subjects. *J Immunol (Baltimore, MD: 1950)*, 2006; **176**: 2781–9.
66. Arif S, Tree TI, Astill TP, *et al.* Autoreactive T cell responses show proinflammatory polarization in diabetes but a regulatory phenotype in health. *J Clin Invest*, 2004; **113**: 451–63.
67. Peakman M, Stevens EJ, Lohmann T, *et al.* Naturally processed and presented epitopes of the islet cell autoantigen IA-2 eluted from HLA-DR4. *J Clin Invest*, 1999; **104**: 1449–57.
68. Kelemen K, Gottlieb PA, Putnam AL, Davidson HW, Wegmann DR, Hutton JC. HLA-DQ8-associated T cell responses to the diabetes autoantigen phogrin (IA-2 beta) in human prediabetes. *J Immunol (Baltimore, MD: 1950)*, 2004; **172**: 3955–62.
69. Enee E, Kratzer R, Arnoux JB, *et al.* ZnT8 is a major CD8+ T cell-recognized autoantigen in pediatric type 1 diabetes. Diabetes, 2012; **61**:1779–84.
70. Dang M, Rockell J, Wagner R, *et al.* Human type 1 diabetes is associated with T cell autoimmunity to zinc transporter 8 (ZnT8). *J Immunol (Baltimore, MD: 1950)*, 2011; **186**: 6056–63.
71. Harbige J, Eichmann M, Peakman M. New insights into non-conventional epitopes as T cell targets: the missing link for breaking immune tolerance in autoimmune disease? *J Autoimmun*, 2017; **84**: 12–20.

72. Kracht MJ, van Lummel M, Nikolic T, *et al.* Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nat Med*, 2017; **23**: 501–7.
73. Kenefeck R, Wang CJ, Kapadi T, *et al.* Follicular helper T cell signature in type 1 diabetes. *J Clin Investig*, 2015; **125**: 292–303.
74. Buckner JH. Mechanisms of impaired regulation by CD4(+) CD25(+)FOXP3(+) regulatory T cells in human autoimmune diseases. *Nat Rev Immunol*, 2010; **10**: 849–59.
75. Hull CM, Peakman M, Tree TIM. Regulatory T cell dysfunction in type 1 diabetes: what's broken and how can we fix it? *Diabetologia*, 2017; **60**: 1839–50.
76. Ferraro A, Socci C, Stabilini A, *et al.* Expansion of Th17 cells and functional defects in T regulatory cells are key features of the pancreatic lymph nodes in patients with type 1 diabetes. *Diabetes*, 2011; **60**: 2903–13.
77. Keenan HA, Sun JK, Levine J, *et al.* Residual insulin production and pancreatic ss-cell turnover after 50 years of diabetes: Joslin Medalist study. *Diabetes*, 2010; **59**: 2846–53.
78. Brozzi F, Nardelli TR, Lopes M, *et al.* Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms. *Diabetologia*, 2015; **58**: 2307–16.
79. Padgett LE, Broniowska KA, Hansen PA, Corbett JA, Tse HM. The role of reactive oxygen species and proinflammatory cytokines in type 1 diabetes pathogenesis. *Ann N Y Acad Sci*, 2013; **1281**: 16–35.
80. Naganawa M, Lim K, Nabulsi NB, *et al.* Evaluation of pancreatic VMAT2 binding with active and inactive enantiomers of [(18)F] FP-DTBZ in healthy subjects and patients with type 1 diabetes. *Mol Imaging Biol*, 2018; **20**: 835–45.
81. Wilcox NS, Rui J, Hebrok M, Herold KC. Life and death of beta cells in Type 1 diabetes: a comprehensive review. *J Autoimmun*, 2016; **71**: 51–8.
82. Cyclosporin-induced remission of IDDM after early intervention. Association of 1 yr of cyclosporin treatment with enhanced insulin secretion. The Canadian-European Randomized Control Trial Group. *Diabetes*, 1988; **37**: 1574–82.
83. Feutren G, Papoz L, Assan R, *et al.* Cyclosporin increases the rate and length of remissions in insulin-dependent diabetes of recent onset. Results of a multicentre double-blind trial. *Lancet*, 1986; **2**: 119–24.
84. Skyler JS. Hope vs hype: where are we in type 1 diabetes? *Diabetologia*, 2018; **61**: 509–16.
85. Aronson R, Gottlieb PA, Christiansen JS, *et al.* Low-dose otelixizumab anti-CD3 monoclonal antibody DEFEND-1 study: results of the randomized phase III study in recent-onset human type 1 diabetes. *Diabetes Care*, 2014; **37**: 2746–54.
86. Herold KC, Gitelman SE, Ehlers MR, *et al.* Teplizumab (anti-CD3 mAb) treatment preserves C-peptide responses in patients with new-onset type 1 diabetes in a randomized controlled trial: metabolic and immunologic features at baseline identify a subgroup of responders. *Diabetes*, 2013; **62**: 3766–74.
87. Pescovitz MD, Greenbaum CJ, Bundy B, *et al.* B-lymphocyte depletion with rituximab and β-cell function: two-year results. *Diabetes Care*, 2014; **37**: 453–9.
88. Hoffman VS, Weiß A, Winkler C, *et al.* Landmark models to define the age-adjusted risk of developing stage 1 type 1 diabetes across childhood and adolescence. *BMC Med*, 2019; **17**: 125.

15.3

Type 2 Diabetes

15.3.1 Epidemiology and Public Health

Sarah Wild and Jackie Price

Definition of Type 2 Diabetes

The 2006 World Health Organization (WHO) definition of diabetes is based on a fasting plasma glucose concentration of 7 mmol/L or above and/or a 2 hour venous plasma glucose value of 11.1 mmol/L or above after a 75 g glucose load during an oral glucose tolerance test (OGTT) [1]. The American Diabetes Association (ADA) definition uses a fasting glucose of 7 mmol/L or above alone to identify diabetes. Both WHO and ADA now recognize the use of non-fasting glycosylated haemoglobin (HbA1c) levels >6.5% (48 mmol/mol) to diagnose diabetes. WHO has concluded that 'HbA1c can be used as a diagnostic test for diabetes, provided that stringent quality assurance tests are in place and assays are standardized to criteria aligned to the international reference values, and there are no conditions present which preclude its accurate measurement' [2]. Diabetes may still be diagnosed using glucose criteria even if HbA1c is less than 6.5% (48 mmol/mol). In people who are asymptomatic two abnormal tests are required to diagnose diabetes. Recent research has identified clinically important subtypes of type 2 diabetes that can be distinguished by measurement of six variables: glutamate decarboxylase antibodies, age at diagnosis, body mass index (BMI), HbA1c, and homoeostatic model assessment two estimates of β-cell function and insulin resistance although several of these are not measured in routine clinical practice [3].

Estimates of diabetes prevalence (that is the proportion of people within a population with diabetes) may be based on self-reported diabetes from population surveys, medical records/diabetes registers or on models of diabetes prevalence which extrapolate data from other populations. Such estimates, based on known diabetes, have been shown to underestimate the prevalence of diabetes identified by biochemical testing by anywhere between 30% and 70% depending on the population and the year of data collection. Undiagnosed diabetes can be identified by screening.

Prevalence of Type 2 Diabetes

The prevalence of diabetes recorded in any given population is affected by the definition and data sources used. The risk of type 2 diabetes increases with age and therefore the crude prevalence of diabetes will tend to be higher in older than younger populations. As a consequence, comparisons of prevalence are usually age-standardized to enable fair comparisons. Ageing of populations and improving survival of people with diabetes contributes to increasing prevalence of diabetes over time [4] even if incidence (that is the number of people developing new diabetes each year) remains stable or declines.

Many epidemiological studies of diabetes prevalence do not distinguish between type 1 and type 2 diabetes. However, the proportion of all diabetes accounted for by type 2 diabetes is likely to be a minimum of 85% (e.g. in Scandinavian countries in which type 1 diabetes is relatively common) and may be higher than 95% in countries in which type 1 diabetes is less common. The International Diabetes Federation (IDF) produces an atlas of diabetes prevalence (https://www.diabetesatlas.org/en/). In the estimates for 2019, the world age-standardized prevalence of combined known and undiagnosed diabetes among 20–79-year-olds was estimated to be 8.3% overall and varied from 4.7% in Africa to 12.2% in the Middle East and North Africa (https://www.diabetesatlas.org/en/).

The IDF also produces projections of estimated number of adults with diabetes and have estimated that numbers of people with diabetes of 20-79 years of age in the world will increase from 463 million 2019 to 700 million in 2045. These projections of future numbers of people with diabetes take into account changes in population distribution by age and urban/rural residence for less developed countries. Increasing trends in prevalence of obesity other than those associated with urbanization in less developed countries and improvements in survival of people with diabetes

mean that these figures are likely to underestimate future diabetes prevalence.

Risk Factors for Type 2 Diabetes

Non-modifiable risk factors for type 2 diabetes include age, sex, genetic background, and ancestry. Incidence and prevalence in most countries increase with increasing age up to late middle age and fall in the oldest age groups. The incidence of type 2 diabetes among northern European populations increases with age up to 70–79 years of age. The sex ratio differs between populations with, for example, higher prevalence in European men than women but higher prevalence among Middle Eastern women than men. This probably reflects sex differences in the prevalence of obesity. A family history of type 2 diabetes increases the risk of type 2 diabetes, arising from a combination of shared genetic and environmental factors. Multiple genetic polymorphisms have been found to be associated with increased risk of diabetes, but at present these only explain a fraction of the heritability of diabetes. However, it is likely that interaction between genetic and lifestyle factors contribute to the risk of diabetes. Considerably higher prevalence of type 2 diabetes has been found among people of South Asian, African, Middle Eastern, and Hispanic descent and among indigenous populations of America, Australasia, and Pacific Islands than among European populations. The important ethnic differences in the prevalence of diabetes are discussed in detail in Chapter 15.7.3. Ethnic differences in the prevalence of diabetes may be at least partly explained by genetic adaptation to famine during evolution, resulting in a differential maladaptation to the energy imbalance that has developed in modern life, a situation termed the 'thrifty genotype'. There is evidence that fetal adaptation to the maternal environment, or programming, results in the thrifty phenotype—babies who are relatively malnourished at birth are at higher risk of developing diabetes in later life than normal birth weight babies, suggesting an important developmental component to diabetes risk.

Modifiable Risk Factors for Diabetes

Lifestyle factors associated with urbanization, particularly increases in energy intake and decreases in physical activity, lead to an increase in the prevalence of obesity and an associated increase in the risk of type 2 diabetes. Increasing prevalence of obesity is likely to explain much of the increasing prevalence and incidence of type 2 diabetes observed in recent years. The prevalence of type 2 diabetes tends to be higher in lower socioeconomic groups in developed countries and in higher socioeconomic groups in developing countries, reflecting similar patterns in the prevalence of obesity. The importance of obesity as a risk factor for diabetes is demonstrated by associations observed in numerous cross-sectional and prospective studies and in the prevention of diabetes observed with interventions resulting in weight loss in several populations [5–7]. Differences in fat distribution between populations may explain some of the ethnic differences in diabetes risk—people with central fat distribution are at higher risk of diabetes than those with peripheral fat distribution and ethnic specific cut points of BMI should be used to define obesity. Regular physical activity is associated with reduced risk of developing diabetes even after adjusting for BMI (for review see [8]). The effect of cigarette smoking is complicated by its association with BMI and fat distribution but a meta-analysis suggests that current smoking is associated with increased risk of developing diabetes [9].

Epidemiological studies investigating the role of diet have found that increased consumption of fibre and coffee is associated with a decreased risk of diabetes. High trans fatty acid intake and both absent and excess alcohol consumption have been associated with a higher risk of type 2 diabetes. However, diet is difficult to measure accurately and it is not clear how dietary components influence the risk of diabetes independent of BMI and physical activity levels. A striking inverse dose-response association has been shown between plasma vitamin C (as a marker of fruit/vegetable intake) and risk of incident diabetes, independent of lifestyle and anthropometric factors suggesting that biomarkers of dietary intake may be more reliable than self-reported dietary intake [10]. In contrast positive associations have been reported between consumption of sugar-sweetened beverages and incidence of diabetes [11].

The incidence of diabetes appears to be unchanged or increased by use of thiazide diuretics and β-blockers and unchanged or decreased by ACE inhibitors, calcium channel blockers, and angiotensin receptor blockers [12]. Treatment with antiretroviral agents is associated with an approximate doubling of the risk of type 2 diabetes [13]. Treatment with corticosteroids also increases the risk of diabetes. Diabetes is more common among people with schizophrenia, bipolar disorder, and schizoaffective disorder than among the general population and this may be related to side effects of antipsychotic drugs and increased prevalence of obesity.

Prevention of Type 2 Diabetes

Risk Scores for Diabetes

Several risk scores for diabetes have been developed either to identify people at high risk of developing future diabetes who would be eligible for primary prevention or to help identify people with undiagnosed diabetes who could be screened in order to start appropriate secondary prevention. The variables included in risk scores differ and include readily available factors from self-report on age, BMI, and family history, through those available routinely in medical records to those requiring biochemical testing of glucose and lipids. Risk scores tend to perform reasonably well in the populations in which they are developed but perform less well in other populations, particularly those of different ethnicity. Examples of risk scores include FINDRISC [14], a modified UK version of FINDRISC [15], the Cambridge Risk Score [16], the simplified Indian Diabetes Risk Score [17], and QDScore [18].

Primary Prevention

Several trials have now shown that diabetes can be prevented in a variety of populations with lifestyle intervention with long-term benefits [5, 6, 19, 20]. Metformin has also been used in some diabetes prevention studies but in general this does not appear to confer additional benefit over lifestyle intervention [21]. It remains

challenging to introduce effective diabetes prevention programmes outside the context of trials but the National Health Service in England is attempting to do this (see https://www.england.nhs.uk/diabetes/diabetes-prevention/ for more details).

Secondary Prevention and Screening

The role of screening for diabetes remains controversial with no evidence available at present to show cost-effectiveness of screening for diabetes alone. However screening for both type 2 diabetes and impaired glucose tolerance in relatively high risk populations if an intervention is offered to people with impaired glucose tolerance appears to be cost effective [22]. The Anglo-Danish-Dutch Study of Intensive Treatment In People with Screen Detected Diabetes in Primary Care (ADDITION) trial found that intensive management of patients with screen-detected type 2 diabetes was associated with a small, non-significant reduction in the incidence of cardiovascular events and death [23].

Complications of Diabetes and Tertiary Prevention

Complications of diabetes make the major contribution to healthcare costs of treating people with diabetes. Many people will have macrovascular (coronary, cerebrovascular, or peripheral arterial disease) and microvascular (retinopathy, nephropathy, and neuropathy) complications of diabetes present at diagnosis of diabetes. Increasing duration of diabetes and poor glycaemic control are important risk factors for microvascular disease and hypertension and dyslipidaemia, which are common among people with type 2 diabetes, are important risk factors for macrovascular disease. Diabetes is a common cause of blindness and end-stage renal disease and is the leading cause of non-traumatic leg amputation. It is important that evidence-based approaches such as those used in national guidelines are used to screen for and manage complications of diabetes and their risk factors.

Depression is approximately twice as common among people with diabetes compared with the general population. Type 2 diabetes is also associated with an increased risk of cognitive impairment and dementia. Collated results from a large number of epidemiological studies indicate that overall, people with diabetes have a 1.2–1.7 times greater decline in cognitive performance than those without diabetes and are 1.6 times more likely to develop dementia. Perhaps predictably, given the association between diabetes and cardiovascular disease, diabetes is associated with a 2.2–3.4 times greater risk of vascular dementia.

Non-alcoholic fatty liver disease (NAFLD, which covers a spectrum ranging from fat accumulation alone (steatosis), through steatohepatitis (NASH) and advanced fibrosis to end-stage liver disease with cirrhosis and hepatocellular carcinoma), occurs more commonly in people with type 2 diabetes than those without diabetes. Prevalence of NAFLD among people with diabetes ranges from 34% to 74% in studies of participants unselected by BMI but has been reported to be up to 100% in those with coexistent obesity. The severity of NAFLD is also increased in type 2 diabetes and several studies have shown that type 2 diabetes is a risk factor for progression of NAFLD from steatosis to fibrosis and cirrhosis. The presence of NAFLD among people with diabetes confers increased risk of cardiovascular disease and mortality from all causes [24].

Mortality

Estimating the effect of diabetes on mortality is difficult from routine data because diabetes is poorly recorded on death certificates. Life expectancy for people with diabetes in Scotland is between 0 and 6 years shorter on average compared to people without diabetes, depending on age, sex, and socioeconomic status [25]. The global mortality attributable to diabetes and its complications for 20-79- year-olds in the year 2019 was estimated to be 4.2 million deaths, equivalent to 11.3% of all death. Older studies suggest that relative risks of mortality for people with diabetes are approximately double those compared to those without diabetes but relative risks of mortality associated with type 2 diabetes have declined in developed countries in recent years. Relative risks of mortality associated with diabetes are generally higher for younger than older people, for women than for men, for type 1 diabetes than type 2 diabetes and for cardiovascular disease than for all-cause mortality. Cardiovascular disease accounts for a large proportion of deaths among people with diabetes in developed countries with larger proportions of deaths due to infectious diseases and renal disease reported in less developed than more developed countries.

REFERENCES

1. World Health Organization, International Diabetes Federation. *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia*. Geneva: World Health Organization, 2006.
2. World Health Organization. *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus*. Abbreviated report of a WHO consultation. Geneva: World Health Organization, 2011.
3. Ahlqvist E, Storm P, Käräjämäki A, *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*, 2018; **6**: 361–9.
4. Read SH, Kerssens JJ, McAllister DA, *et al.* Trends in type 2 diabetes incidence and mortality in Scotland between 2004 and 2013. *Diabetologia*, 2016; **59**: 2106–13.
5. Pan XR, Li GW, Hu YH, *et al.* Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care*, 1997; **20**: 537–44.
6. Knowler WC, Barrett-Connor E, Fowler SE, *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*, 2002; **346**: 393–403.
7. Lindstrom J, Louheranta A, Mannelin M, *et al.* The Finnish Diabetes Prevention Study (DPS): lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care*, 2003; **26**: 3230–36.
8. Gill JM, Cooper AR. Physical activity and prevention of type 2 diabetes mellitus. *Sports Med*, 2008; **38**: 807–24.
9. Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2015; **3**: 958–67.
10. Harding AH, Wareham NJ, Bingham SA, *et al.* Plasma vitamin C level, fruit and vegetable consumption, and the risk of new-onset type 2 diabetes mellitus: the European prospective investigation of cancer—Norfolk prospective study. *Arch Intern Med*, 2008; **168**: 1493–9.

11. Imamura F, O'Connor L, Ye Z, *et al.* Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. *BMJ*, 2015; **351**: h3576.
12. Padwal R, Laupacis A. Antihypertensive therapy and incidence of type 2 diabetes: a systematic review. *Diabetes Care*, 2004; **27**: 247–55.
13. Ledergerber B, Furrer H, Rickenbach M, *et al.* Factors associated with the incidence of type 2 diabetes mellitus in HIV-infected participants in the Swiss HIV Cohort Study. *Clin Infect Dis*, 2007; **45**: 111–19.
14. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*, 2003; **26**: 725–31.
15. Gray LJ, Taub NA, Khunti K, *et al.* The Leicester Risk Assessment score for detecting undiagnosed Type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabet Med*, 2010; **27**: 887–95.
16. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes Metab Res Rev*, 2000; **16**: 164–71.
17. Mohan V, Deepa R, Deepa M, Somannavar S, Datta M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. *J Assoc Physicians India*, 2005; **53**: 759–63.
18. Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ*, 2009; **338**: b880.
19. Tuomilehto J, Lindstrom J, Eriksson JG, *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*, 2001; **344**: 1343–50.
20. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V. The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia*, 2006; **49**: 289–97.
21. Crandall J, Schade D, Ma Y, *et al.* The influence of age on the effects of lifestyle modification and metformin in prevention of diabetes. *J Gerontol A Biol Sci Med Sci*, 2006; **61**: 1075–81.
22. Gillies CL, Lambert PC, Abrams KR, *et al.* Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ*, 2008; **336**: 1180–5.
23. Griffin SJ, Borch-Johnsen K, Davies MJ, *et al.* Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. *Lancet*, 2011; **378**: 156–67.
24. Wild SH, Walker JJ, Morling JR, *et al.* Cardiovascular disease, cancer, and mortality among people with type 2 diabetes and alcoholic or nonalcoholic fatty liver disease hospital admission. *Diabetes Care*, 2018; **41**: 341–7.
25. Walker J, Colhoun H, Livingstone S, *et al.* Type 2 diabetes, socioeconomic status and life expectancy in Scotland (2012–2014): a population-based observational study. *Diabetologia*, 2018; **61**: 108–16.

15.3.2 Presentation and Natural History of Type 2 Diabetes

Roy Taylor

Clinical Presentations of Type 2 Diabetes

Screening

Presence of risk factors for type 2 diabetes such as a family history of the disease or obesity allow targeted screening. Given the typically long phase of prediabetes, this is of benefit. In an era when screening was much less common, tissue damage was present in around 50% of people newly diagnosed [1].

Thirst, Polydipsia, and Polyuria

These symptoms result from an osmotic diuresis as a consequence of hyperglycaemia. For glucose to escape into the urine, plasma concentration must exceed the renal threshold for tubular reabsorption of glucose. The renal threshold averages 11 mmol/L but displays a wide individual variation of around 6–14 mmol/L [2]. The maximum absorptive capacity rises with age such that older people exhibit glycosuria at higher plasma glucose levels [3]. They also have a higher threshold for thirst [4].

Hyperosmolar Hyperglycaemic Syndrome

The principal clinical features of this syndrome are profound dehydration with confusion, and associated focal neurological symptoms such as hemi-paresis or hemi-sensory abnormalities may easily be confused with stroke. Mild ketosis may be present at diagnosis simply because of antecedent decrease in food intake, and calculation of plasma osmolarity ((sodium + potassium) × 2 + urea + glucose) is important to distinguish this from ketoacidosis.

Infections

Leucocytes exhibit paralysis of phagocytic and other functions above glucose concentration of 11 mmol/L [5]. This explains the impaired ability to fight off bacterial and fungal infections. Recurrent or refractory yeast infections, vaginal candidiasis in women, or balanitis in men, may draw attention to previously undiagnosed diabetes.

Very rare but serious infective presentations of diabetes must be considered. Necrotizing fasciitis is commoner in people with diagnosed or undiagnosed diabetes [6]. Fournier's gangrene (gangrene of the perineum and genitalia) is associated with diabetes in almost 50% of instances [7]. The rare and facial/maxillary sinus fungal infection mucormycosis is most often associated with diabetes [8].

Complications of Diabetes

Type 2 diabetes frequently presents for the first time with acute myocardial infarction or stroke. Symptomatic loss of vision due to macula oedema, vitreous haemorrhage or central/ branch retinal vein occlusion may occasionally be the presenting feature. Very early ophthalmological treatment is essential as the initial treatment of the diabetes will decrease blood glucose levels, cause retinal blood flow to return acutely to normal levels, and may result in marked worsening of the retinopathy. As loss of pain sensation in the feet, affecting up to 60% in people with newly presenting diabetes [9, 10], is usually unnoticed, foot ulcer or black toe are common presentations.

Acute diabetic mononeuropathy and diabetic amyotrophy are much less common presenting features.

Pregnancy

Presentation of type 2 diabetes during pregnancy is not uncommon due to the physiological insulin resistance which increases through the second and early third trimester. Usually, true type 2 diabetes can be distinguished from gestational diabetes if HbA1c is raised (indicating pronged hyperglycaemia) or if it is detected before 24 weeks. Screening is wise for all women at higher risk of type 2 diabetes as early in pregnancy as healthcare resources allow.

Natural History of Prediabetes

The earliest predictor of development of type 2 diabetes is low insulin sensitivity in skeletal muscle [11–14]. This is associated with raised plasma insulin levels, and lack of ability to store meal-derived glucose as glycogen [15–17]. Meal-derived glucose must either be stored as glycogen, oxidized, or converted to triglyceride for storage. Especially as the process of *de novo* lipogenesis is stimulated by higher insulin levels [17, 18], hepatic fat accumulation is enhanced. Excess fat deposition in the liver is a strong predictor of development of diabetes [19–24]. In individuals with normal glucose tolerance, there is a very low 8-year incidence of type 2 diabetes if fatty liver is excluded, whereas if present the risk of developing diabetes rises fivefold [19]. In support of this, a temporal progression from weight gain to raised liver enzymes, onwards to hypertriglyceridemia and then glucose intolerance has been demonstrated [23, 25].

In the Nurses' Health study, women who put on weight during adult life and achieved body mass index (BMI) of between 23 and 25 exhibited a fourfold increase risk of diabetes compared with those whose BMI remained below 22 kg/m^2 [26]. This is a striking level of risk associated with very modest accumulation of excess total body weight, although usually overshadowed by the 39-fold increase in prevalence if BMI increases to over 35 kg/m^2 [26].

The 2008 Twin Cycle Hypothesis postulated that type 2 diabetes only occurred after a prolonged period of positive calorie balance [27]. Liver fat would gradually accumulate, cause increased export of triglyceride from the liver to all tissues, and result in increased fat exposure of the pancreatic islets. Consequently, a slow decline in beta-cell ability to respond to meals was postulated to pass a threshold after which a relatively rapid rise in plasma glucose levels would occur. The postulated time course was subsequently found to be consistent with the data from the Whitehall II study [28]. The Whitehall II study demonstrated that in an initially normoglycaemic population followed prospectively, those people developing diabetes exhibited a slight and gradual rise in fasting glucose over at least 12 years [28]. This was followed by a marked rise over an average of 18 months immediately prior to diagnosis. This insight has replaced the previous assumption that there would be a slow linear rise in fasting plasma glucose and gradual beta-cell decompensation [29].

The rate of decrease in glucose tolerance in first-degree relatives of type 2 diabetic individuals is determined by decline in beta-cell function, while the degree of muscle insulin sensitivity changes little [30, 31]. This mirrors observations on populations showing that the transition from hyperinsulinaemic normal glucose tolerance to overt diabetes involves a rapid rise in glucose levels as a result of a relatively small further loss of acute β-cell competence [32, 33]. The potential to halt or at least slow the progression to type 2 diabetes has been demonstrated by several population-based intervention studies. The Finnish Diabetes Prevention Programme demonstrated a 48% decrease in conversion from impaired glucose tolerance to type 2 diabetes after 4 years [34]. In that study, the central role of weight loss is demonstrated by the observation that change in weight of 5kg or more was associated with similar decrease in diabetes incidence in both intervention and control groups. After 15 years of follow-up in the Diabetes Prevention Programme, the initial weight loss of 5 kg (falling to 2 kg) was associated with a 9% decrease in risk of onset of type 2 diabetes [35]. The Da Qing study reported the longest follow-up, with a 19% decrease in diabetes incidence after 23 years [36].

Natural History of Diabetes

The Belfast diet study provided the first clear information on the apparently progressive nature of type 2 diabetes [37]. Fasting plasma glucose levels rose steadily over 6 years with increasing requirement for oral hypoglycaemic agents. In the UK Prospective Diabetes Study individuals were managed according to randomized group, with tighter glycaemic control in the intervention group [29, 38]. There was a striking deterioration in blood glucose control over years, and the rise in average blood glucose levels flattened out only when a majority had been transferred to insulin therapy. There was no difference in the rate of deterioration between intervention and control groups. 50% of people required insulin therapy 10 years after diagnosis. More detailed analysis showed that the deterioration was due to steady decline in beta-cell competence rather than any change in insulin resistance [39].

These data appeared to confirm the clinical impression that type 2 diabetes was inevitably progressive, requiring more and more oral agents and inexorably moving towards insulin requirement. However, these observations were made in the context of year-on-year weight gain. The observations relate only to a population of people with type 2 diabetes who follow the common trajectory of steady weight gain after diagnosis. Under these circumstances, the supranormal liver fat present at the time of diagnosis would continue to drive the underlying pathogenic processes [40–42]. However, current clinical guidelines reinforce the inevitably progressive view of natural history of the disease, emphasizing the expectation that hypoglycaemic agents must be added sequentially [43].

The possibility that the natural history could be influenced by weight loss was first raised over 40 years ago [44]. In 2005, Petersen and colleagues defined the detailed effects of moderate weight loss in type 2 diabetes [45]. They demonstrated that 8 kg weight loss over 2 months brought about normalization of liver fat content and normalization of hepatic insulin sensitivity. This allowed a clinically useful decrease in fasting plasma glucose levels. The rapidity of return to normal fasting plasma glucose levels in type 2 diabetes was observed following bariatric surgery [46], and subsequently in the Counterpoint study of calorie restriction [27]. The latter study was conducted to test the Twin Cycle Hypothesis of the aetiology of type 2 diabetes. It observed complete normalization of liver insulin sensitivity within 7 days of commencing a diet of 600 kcal/day and recovery of normal beta-cell function within 8 weeks. Subsequent studies found that this normalization of liver and pancreas function is durable over 6–12 months of weight stability during avoidance of weight regain while eating normally [41, 42]. The natural history of type 2 diabetes can now be seen to depend upon body weight. If the personal level of weight which allowed type 2 diabetes to develop is maintained or even allowed to increase, then the disease is steadily progressive. If weight loss sufficient to normalize intraorgan fat is achieved and maintained, non-diabetic glucose control can be re-established.

Insulin resistance in muscle has long been regarded as being responsible for the continued hyperglycaemia. Certainly, the muscle insulin resistance determines poor glycogen storage after meals [15] and drives on hepatic *de novo* lipogenesis [17] as the natural history of type 2 diabetes unfolds. However, reversing the processes in liver and pancreas which directly cause the syndrome of type 2 diabetes has minimal effect upon muscle insulin resistance [40, 41]. Insulin resistance in muscle appears to be permissive, leading to increase in liver fat levels in the circumstance of excess calorie intake, but not determinant for the natural history of type 2 diabetes.

Natural History of Post-Diabetes

Following demonstration of durable remission of type 2 diabetes after both dietary and surgically induced weight loss, it is important to describe the clinical state of post-diabetes. This is characterized by relatively low risk of both micro- and macro-vascular complications. It is defined as having achieved a non-diabetic level of HbA1c or fasting plasma glucose, after weight loss and off all hypoglycaemic agents. Confirmation by a second test at least two months later is required [47]. Even though HbA1c may be in the range associated with prediabetes, the clinical implications are very different. Prediabetes is a state of high risk for cardiovascular events whereas post-diabetes is not. This is illustrated in Figure 15.3.2.1, showing the change in major components of cardiovascular risk before and after weight loss plus 6 months of weight stability [41]. During the transition from diabetes to post-diabetes, the QRISK score falls on average from 23% to 7%. In the study group, with an average chronological age of 55 years, calculated heart age fell from 71 to 56 years.

The first hint that type 2 diabetes was a fully reversible syndrome came from bariatric surgery [48]. The phenomenon was examined in a randomized prospective study comparing gastric banding with intensive medical therapy for type 2 diabetes [49]. Mean fasting plasma glucose normalized in many, and this was related to the degree of weight loss rather than allocation to either surgical or medical group [49]. The rapidity of the fall in fasting plasma glucose after bariatric surgery led to the widespread belief that surgery itself brought about specific changes mediated via incretin hormone secretion [46, 50, 51]. This assumption overlooked the major change which follows bariatric surgery: an acute, profound decrease in calorie intake with sudden increase in carbon flow from stored triglyceride depots. Several more recent studies have confirmed the dramatic effect of calorie restriction *per se* with no role of incretins in achieving rapid glucose control [52, 53].

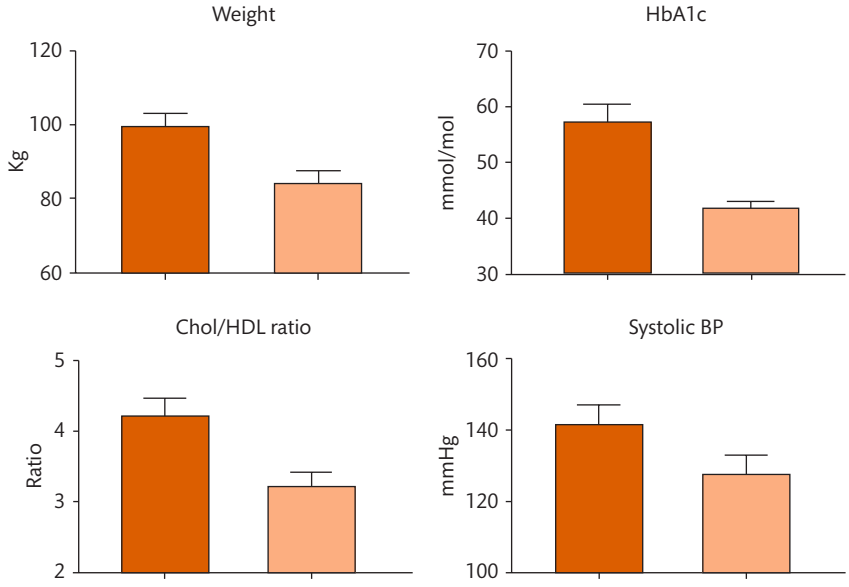


Figure 15.3.2.1 Effect of weight loss in type 2 diabetes on components of cardiovascular risk before (darker bars) and after 6 months of weight maintenance (lighter bars). The average QRISK 10-year cardiovascular risk score fell from 23.2% to 6.9%.
Data are plotted from Steven *et al.* with permission from American Diabetes Association.

The Counterpoint study demonstrated that in short duration type 2 diabetes, decreasing calorie intake sharply allowed return to normal fasting plasma glucose within 7 days [40]. The rapid return to normal was associated with a 30% decrease in liver fat content and normalization of insulin sensitivity in regulating hepatic glucose production. It also demonstrated that over the 8-week study period, the level of fat in the pancreas gradually decreased and in step with this, first phase insulin response returned to within the normal range [40].

The durability of post-diabetes was first tested in the Counterbalance study. Weight loss of around 15 kg followed by 6 months of weight stability achieved stable intrahepatic and intrapancreatic fat levels with both liver and pancreas function remaining normal [41].

Whether these changes could be reproduced by existing nursing or dietetic staff in Primary Care was examined in the Diabetes REmission Controlled Trial (DiRECT) [54]. This demonstrated that weight loss of 15 kg or more could be achieved in 25% with remission of type 2 diabetes maintained over 12 months in 46%, respectively (intention to treat analysis). To bring about this change in natural history of type 2 diabetes, the intervention consisted of 8

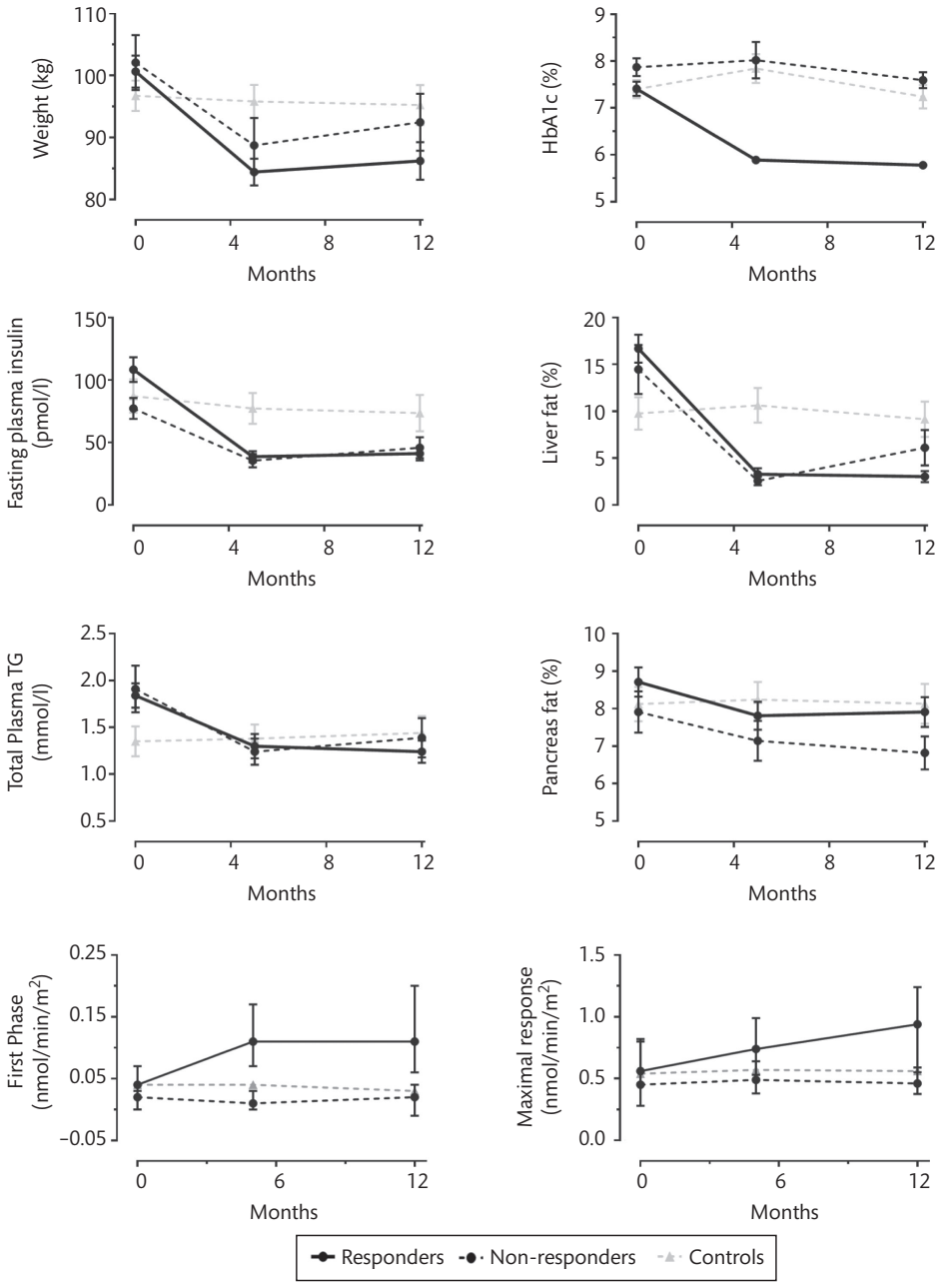


Figure 15.3.2.2 Change in metabolic parameters from baseline after weight loss on an isocaloric diet (5 months) and after the weight maintenance phase of DiRECT (12 months). Participants randomized to the control group continued best management according to guidelines. Those randomized to intervention group stopped all antihyperglycaemic therapy on day 1 of the weight loss diet, and data are shown separately for those who achieved non-diabetic levels of HbA1c (<48 mmol/mol) and fasting plasma glucose (<7.0 mmol/L).

hours of structured training for a primary care nurse or dietitian. Currently, data on durability of remission are sparse, but case histories suggest that if weight regain is avoided, the natural history of type 2 diabetes can be changed to long-term non-diabetic blood glucose control [55–57].

It is important to recognize the lack of relevance of the absolute level of BMI in maintaining the post-diabetic state. After weight loss in both Counterbalance and DiRECT, half of those who returned to non-diabetic glucose control still had a BMI greater than 30, but this had no impact upon their ability to avoid subsequent intraorgan fat re-accumulation during weight stability. This illustrates the important concept of the Personal Fat Threshold [58]. Each individual had dropped below a personal threshold permitting safe storage of fat in the subcutaneous compartment, and this was equally so whether BMI dropped from 40 to 36 or 28 to 24 kg/m². The ability to store fat safely in adipose tissue varies considerably between otherwise normal individuals [59, 60], and once that capacity has been exceeded the body has to use other storage sites including liver and pancreas. If pancreas fat exposure is too high there is clearly variable susceptibility to beta-cell dysfunction. Individuals can exceed their personal fat threshold at BMI's well within the 'normal' range. In the United Kingdom Prospective Diabetes Study [1], at the time of diagnosis of diabetes, 36% had a BMI less than 25 kg/m². The BMI distribution of participants was right-shifted compared with that of the UK population of that time, when 64% had a BMI less than 25 kg/m² [61].

The natural history of type 2 diabetes appears to depend upon more fat supply than could be tolerated by the beta-cells. Excess saturated fat produces endoplasmic reticulum stress, and that this is associated with dedifferentiation of the beta-cell [62–65]. In this survival mode its specialized functions are shut down. Removal of the fat allows re-expression of the insulin gene and resumption of acute glucose mediated insulin secretion [66]. When an individual achieves remission of type 2 diabetes after weight loss, that post-diabetic state is durable providing that weight regain is avoided [42, 67]. The changes in HbA1c and important physiological parameters during the natural history of post-diabetes is shown in **Figure 15.3.2.2**. The first phase insulin response is stable from the end of the weight loss phase to 12 months, and the maximal capacity of the beta-cells increases over the whole 12-month period.

Given that muscle insulin resistance remains essentially unchanged during reversal to the non-diabetic state [40, 67], individuals can be advised that weight regain of sufficient degree will inevitably precipitate return to the diabetic state.

Subsequent studies have found that the normalization of liver and pancreas function is durable over 6–12 months of weight stability during avoidance of weight regain while eating normally [41, 42]. Case reports indicate durability of the post-diabetic state for up to 5 years [55–57]. However, regular follow-up remains essential to monitor and minimize risk of weight regain.

REFERENCES

1. UKPDS. UK Prospective Diabetes Study (UKPDS). VIII. Study design, progress and performance. *Diabetologia*, 1991; **34**: 877–90.
2. Johansen K, Svendsen PA, Lorup B. Variations in renal threshold for glucose in Type 1 (insulin-dependent) diabetes mellitus. *Diabetologia*, 1984; **26**: 180–2.
3. Butterfield WJ, Keen H, Whichelow MJ. Renal glucose threshold variations with age. *Br Med J*, 1967; **4**: 505–7.
4. Baylis PH, Thompson CJ. Osmoregulation of vasopressin secretion and thirst in health and disease. *Clin Endocrinol (Oxf)*, 1988; **29**: 549–76.
5. Turina M, Fry DE, Polk HC, Jr. Acute hyperglycemia and the innate immune system: clinical, cellular, and molecular aspects. *Crit Care Med*, 2005; **33**: 1624–33.
6. Dworkin MS, Westercamp MD, Park L, McIntyre A. The epidemiology of necrotizing fasciitis including factors associated with death and amputation. *Epidemiol Infect*, 2009; **137**: 1609–14.
7. Yanar H, Taviloglu K, Ertekin C, *et al.* Fournier's gangrene: risk factors and strategies for management. *World J Surg*, 2006; **30**: 1750–4.
8. Haliloglu NU, Yesilirmak Z, Erden A, Erden I. Rhino-orbito-cerebral mucormycosis: report of two cases and review of the literature. *Dentomaxillofac Radiol*, 2008; **37**: 161–6.
9. Young MJ, Boulton AJ, MacLeod AF, Williams DR, Sonksen PH. A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population. *Diabetologia*, 1993; **36**: 150–4.
10. Tapp RJ, Shaw JE, de Courten MP, Dunstan DW, Welborn TA, Zimmet PZ. Foot complications in Type 2 diabetes: an Australian population-based study. *Diabet Med*, 2003; **20**: 105–13.
11. Lillioja S, Mott DM, Spraul M, *et al.* Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. *N Engl J Med*, 1993; **329**: 1988–92.
12. Bock G, Chittilapilly EG, Basu R, *et al.* Contribution of hepatic and extrahepatic insulin resistance to the pathogenesis of impaired fasting glucose. Role of increased rates of gluconeogenesis. *Diabetes*, 2007; **56**: 1703–11.
13. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*, 1988; **37**: 1595–607.
14. Taylor R. Insulin resistance and type 2 diabetes. *Diabetes*, 2012; **61**: 778–9.
15. Carey PE, Halliday J, Snaar JEM, Morris PG, Taylor R. Direct assessment of muscle glycogen storage after mixed meals in normal and type 2 diabetic subjects. *Am J Physiol*, 2003; **284**: E286–94.
16. Macauley M, Smith FE, Thelwall PE, Hollingsworth KG, Taylor R. Diurnal variation in skeletal muscle and liver glycogen in humans with normal health and type 2 diabetes. *Clin Sci (Lond)*, 2015; **128**: 707–13.
17. Petersen KF, Dufour S, Savage DB, *et al.* The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc Natl Acad Sci U S A*, 2007; **104**: 12587–94.
18. Schwarz JM, Linfoot P, Dare D, Aghajanian K. Hepatic de novo lipogenesis in normoinsulinemic and hyperinsulinemic subjects consuming high-fat, low-carbohydrate and low-fat, high-carbohydrate isoenergetic diets. *Am J Clin Nutr*, 2003; **77**: 43–50.
19. Shibata M, Kihara Y, Taguchi M, Tashiro M, Otsuki M. Nonalcoholic fatty liver disease is a risk factor for type 2 diabetes in middle-aged Japanese men. *Diabetes Care*, 2007; **30**: 2940–4.
20. Perseghin G, Bonfanti R, Magni S, *et al.* Insulin resistance and whole body energy homeostasis in obese adolescents with

fatty liver disease. *AJP Endocrinology and Metabolism*, 2006; **291**: E697–703.
21. Perseghin G, Lattuada G, De Cobelli F, *et al.* Habitual physical activity is associated with intrahepatic fat content in humans. *Diabetes Care*, 2007; **30**: 683–8.
22. Cali AM, Zern TL, Taksali SE, *et al.* Intrahepatic fat accumulation and alterations in lipoprotein composition in obese adolescents: a perfect proatherogenic state. *Diabetes Care*, 2007; **30**: 3093–8.
23. Sattar N, McConnachie A, Ford I, *et al.* Serial metabolic measurements and conversion to type 2 diabetes in the west of Scotland coronary prevention study: specific elevations in alanine aminotransferase and triglycerides suggest hepatic fat accumulation as a potential contributing factor. *Diabetes*, 2007; **56**: 984–91.
24. Targher G, Bertolini L, Padovani R, *et al.* Prevalence of nonalcoholic fatty liver disease and its association with cardiovascular disease among type 2 diabetic patients. *Diabetes Care*, 2007; **30**: 1212–18.
25. Suzuki A, Angulo P, Lymp J, *et al.* Chronological development of elevated aminotransferases in a nonalcoholic population. *Hepatology (Baltimore, MD)*, 2005; **41**: 64–71.
26. Hu FB, Manson JE, Stampfer MJ, *et al.* Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med*, 2001; **345**: 790–7.
27. Taylor R. Pathogenesis of Type 2 diabetes: tracing the reverse route from cure to cause. *Diabetologia*, 2008; **51**: 1781–9.
28. Tabak AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimaki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet*, 2009; **373**: 2215–21.
29. U.K. prospective diabetes study 16. Overview of 6 years' therapy of type II diabetes: a progressive disease. U.K. Prospective Diabetes Study Group. *Diabetes*, 1995; **44**: 1249–58.
30. Cnop M, Vidal J, Hull RL, *et al.* Progressive loss of beta-cell function leads to worsening glucose tolerance in first-degree relatives of subjects with type 2 diabetes. *Diabetes Care*, 2007; **30**: 677–82.
31. Festa A, Williams K, D'Agostino R, Jr., Wagenknecht LE, Haffner SM. The natural course of beta-cell function in nondiabetic and diabetic individuals: the Insulin Resistance Atherosclerosis Study. *Diabetes*, 2006; **55**: 1114–20.
32. Ferrannini E, Nannipieri M, Williams K, Gonzales C, Haffner SM, Stern MP. Mode of onset of type 2 diabetes from normal or impaired glucose tolerance. *Diabetes*, 2004; **53**: 160–5.
33. Weyer C, Bogardus C, Mott DM, Pratley RE. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest*, 1999; **104**: 787–94.
34. Tuomilehto J, Lindstrom J, Eriksson JG, *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*, 2001; **344**: 1343–50.
35. Diabetes Prevention Program Research G. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol*, 2015; **3**: 866–75.
36. Li G, Zhang P, Wang J, *et al.* Cardiovascular mortality, all-cause mortality, and diabetes incidence after lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: a 23-year follow-up study. *Lancet Diabetes Endocrinol*, 2014; **2**: 474–80.
37. Hadden DR, Blair AL, Wilson EA, *et al.* Natural history of diabetes presenting age 40–69 years: a prospective study of the influence of intensive dietary therapy. *Q J Med*, 1986; **59**: 579–98.
38. UKPDS. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, 1998; **352**: 854–65.
39. Rudenski AS, Hadden DR, Atkinson AB, *et al.* Natural history of pancreatic islet B-cell function in type 2 diabetes mellitus studied over six years by homeostasis model assessment. *Diabetic Med*, 1988; **5**: 36–41.
40. Lim EL, Hollingsworth KG, Aribisala BS, Chen MJ, Mathers JC, Taylor R. Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia*, 2011; **54**: 2506–14.
41. Steven S, Hollingsworth KG, Al-Mrabeh A, *et al.* Very low calorie diet and 6 months of weight stability in type 2 diabetes: pathophysiological changes in responders and nonresponders. *Diabetes Care*, 2016; **39**: 158–65.
42. Taylor R, Al-Mrabeh A. Remission of human type 2 diabetes requires decrease in liver and pancreas fat content but is dependent upon capacity for beta cell recovery. *Cell Metab*, 2018; **28**: 547–56.e3.
43. Inzucchi SE, Bergenstal RM, Buse JB, *et al.* Management of hyperglycaemia in type 2 diabetes, 2015: a patient-centred approach. Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia*, 2015; **58**: 429–42.
44. Henry RR, Wallace P, Olefsky JM. Effects of weight loss on mechanisms of hyperglycaemia in obese non-insulin dependent diabetes mellitus. *Diabetes*, 1986; **35**: 990–8.
45. Petersen KF, Dufour S, Befroy D, Lehrke M, Hendler RE, Shulman GI. Reversal of nonalcoholic hepatic steatosis, hepatic insulin resistance, and hyperglycemia by moderate weight reduction in patients with type 2 diabetes. *Diabetes*, 2005; **54**: 603–8.
46. Guidone C, Manco M, Valera-Mora E, *et al.* Mechanisms of recovery from type 2 diabetes after malabsorptive bariatric surgery. *Diabetes*, 2006; **55**: 2025–31.
47. McCombie L, Leslie W, Taylor R, Kennon B, Sattar N, Lean MEJ. Beating type 2 diabetes into remission. *BMJ*, 2017; **358**: j4030.
48. Pories WJ, MacDonald KG, Jr., Morgan EJ, *et al.* Surgical treatment of obesity and its effect on diabetes: 10-y follow-up. *Am J Clin Nutr*, 1992; **55**(2 Suppl): 582S–5S.
49. Dixon JB, O'Brien PE, Playfair J, *et al.* Adjustable gastric banding and conventional therapy for type 2 diabetes: a randomized controlled trial. *JAMA*, 2008; **299**: 316–23.
50. Rubino F, Forgione A, Cummings DE, *et al.* The mechanism of diabetes control after gastrointestinal bypass surgery reveals a role of the proximal small intestine in the pathophysiology of type 2 diabetes. *Ann Surg*, 2006; **244**: 741–9.
51. Holst JJ. Postprandial insulin secretion after gastric bypass surgery: the role of glucagon-like peptide 1. *Diabetes*, 2011; **60**: 2203–5.
52. Lingvay I, Guth E, Islam A, Livingston E. Rapid improvement in diabetes after gastric bypass surgery: is it the diet or surgery? *Diabetes Care*, 2013; **36**: 2741–7.
53. Steven S, Hollingsworth KG, Small PK, *et al.* Calorie restriction and not glucagon-like peptide-1 explains the acute improvement in glucose control after gastric bypass in Type 2 diabetes. *Diabetic Med*, 2016; **33**: 1723–31.
54. Lean ME, Leslie WS, Barnes AC, *et al.* Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet*, 2017; **391**: 541–51.

55. Taylor R. Calorie restriction and reversal of type 2 diabetes. *Exp Rev Endocrinol Metab*, 2016; **11**: 521–8.
56. Paisey RB, Frost J, Harvey P, *et al.* Five year results of a prospective very low calorie diet or conventional weight loss programme in type 2 diabetes. *J Hum Nutr Diet*, 2002; **15**: 121–7.
57. LookAhead. Eight-year weight losses with an intensive lifestyle intervention: the look AHEAD study. *Obesity*, 2014; **22**: 5–13.
58. Taylor R, Holman R. Normal weight individuals who develop type 2 diabetes: the personal fat threshold. *Clin Sci (Lond)*, 2015; **128**: 405–10.
59. Robbins AL, Savage DB. The genetics of lipid storage and human lipodystrophies. *Trends Mol Med*, 2015; **21**: 433–8.
60. Virtue S, Vidal-Puig A. Adipose tissue expandability, lipotoxicity and the metabolic syndrome—an allostatic perspective. *Biochim Biophys Acta*, 2010; **1801**: 338–49.
61. Rosenbaum S, Skinner RK, Knight IB, Garrow JS. A survey of heights and weights of adults in Great Britain, 1980. *Ann Hum Biol*, 1985; **12**: 115–27.
62. Pinnick K, Neville M, Clark A, Fielding B. Reversibility of metabolic and morphological changes associated with chronic exposure of pancreatic islet beta-cells to fatty acids. *J Cell Biochem*, 2010; **109**: 683–92.
63. Pinnick KE, Collins SC, Londos C, Gauguier D, Clark A, Fielding BA. Pancreatic ectopic fat is characterized by adipocyte infiltration and altered lipid composition. *Obesity*, 2008; **16**: 522–30.
64. Talchai C, Xuan S, Lin HV, Sussel L, Accili D. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. *Cell*, 2012; **150**: 1223–34.
65. White MG, Marshall HL, Rigby R, *et al.* Expression of mesenchymal and alpha-cell phenotypic markers in islet beta-cells in recently diagnosed diabetes. *Diabetes Care*, 2013; **36**: 3818–20.
66. White MG, Shaw JAM, Taylor R. Type 2 diabetes: the pathologic basis of reversible beta-cell dysfunction. *Diabetes Care*, 2016; **39**: 2080–88.
67. Steven S, Hollingsworth KG, Small P, *et al.* Weight loss decreases excess pancreatic triacylglycerol specifically in type 2 diabetes. *Diabetes Care*, 2016; **39**: 158–65.

15.3.3 Pathogenesis

Mark Walker, Xuefei Yu, and Amalia Gastaldelli

Introduction

Hyperglycaemia is the hallmark of type 2 diabetes. It results from a combination of decreased insulin secretion and insulin resistance (also referred to as decreased insulin sensitivity) as stated in the WHO definition of type 2 diabetes. These two characteristics are physiologically linked as demonstrated by previous studies of subjects with different glycaemic states. As shown in **Figure 15.3.3.1**, there exists an inverse hyperbolic relationship between insulin secretory response and insulin sensitivity for subjects with normal glucose tolerance. The relationship persists in patients with diabetes, but the curve is shifted downwards such that less insulin is secreted for a given level of whole body insulin sensitivity [1]. Progression from normal glucose tolerance to diabetes results from the failure of insulin secretion to compensate for the level of insulin resistance. Even within the range of normal glucose tolerance, there is evidence of progressive decrease in the insulin secretory response once adjustment has been made for the degree of insulin resistance [2] (**Figure 15.3.3.2**).

Over the last few decades, a large number of detailed physiological studies have been conducted to understand glucose metabolism in healthy, non-diabetic subjects and how it is perturbed in type 2 diabetes [3]. In the fasting state in non-diabetic subjects, circulating insulin acts primarily at the liver to regulate hepatic glucose release. In type 2 diabetes, fasting hepatic glucose release is increased and closely correlated with the fasting glucose concentration. The increased hepatic glucose release reflects the interaction of a number of factors including decreased hepatic insulin sensitivity and raised fasting glucagon levels. In the postprandial state in non-diabetic subjects, the increased circulating insulin levels serve to suppress hepatic glucose release and promote glucose uptake and metabolism in the peripheral insulin sensitive tissues, in particular skeletal muscle [3]. These processes are altered in type 2 diabetes and contribute to the postprandial hyperglycaemia. Specifically, the combination of impaired insulin secretion and decreased insulin sensitivity at the liver and peripheral tissues result, respectively, in impaired suppression of hepatic glucose release and impaired peripheral glucose uptake and metabolism [3]. The purpose of this chapter is to review the mechanisms underlying these physiological abnormalities that contribute to the hyperglycaemic state in type 2 diabetes.

Genetic Architecture

Type 2 diabetes is a complex trait with genetic and non-genetic factors contributing to the pathogenesis. The non-genetic factors are well documented as described in Chapter 15.3.1, 'Epidemiology and Public Health', with increased adiposity and physical activity being key factors. Our understanding of the genetic basis of type 2 diabetes has evolved over the last 12–15 years, driven by mapping of the human genome and technological advances in genotyping and sequencing.

Genome-wide association studies (GWAS) in populations of North European extraction have found that the genetic structure is primarily based upon common variants (minor allele frequency >5%) across many susceptibility loci [4]. At the present time, around 200 susceptibility loci have been identified. While rare variants in

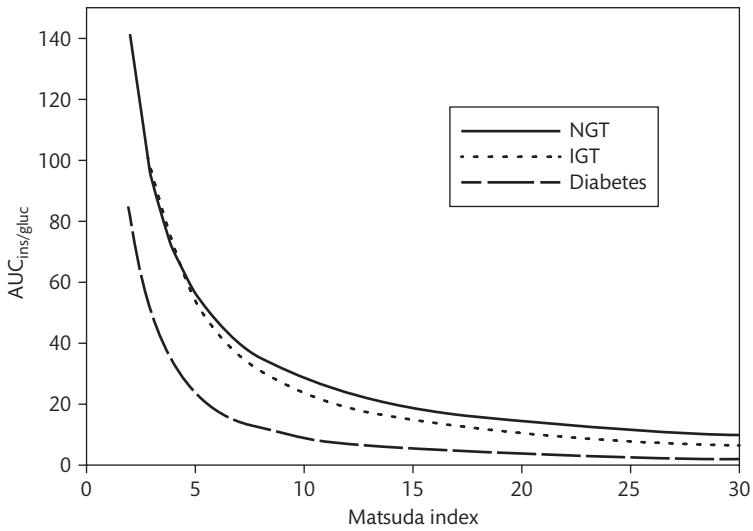


Figure 15.3.3.1 Inverse hyperbolic relationship between insulin secretory response (AUC ins/gluc) and whole body insulin sensitivity (Matsuda index) measures derived during an oral glucose tolerance test. For any given level of insulin sensitivity, the insulin secretory response is less in the diabetic versus normal glucose tolerant subjects.

Reproduced with permission from Retnakaran R, Shen S, Hanley AJ, Vuksan V, Hamilton JK, Zinman B. Hyperbolic Relationship Between Insulin Secretion and Sensitivity on Oral Glucose Tolerance Test. *Obesity*. 2008;16(8): 1901–7. doi:10.1038/oby.2008.307. Copyright © 2012, John Wiley and Sons (Ref 1).

known monogenic genes can contribute to diabetes segregation within certain families, there is little evidence for the role of rare variants in other loci [5]. This may not apply to genetically isolated populations such as the Greenland Inuits, in which a variant in the *TBC1D4* gene specific to this population was found to predispose to diabetes [6].

The common risk variants identified by GWAS are functionally weak, so an individual's genetic type 2 diabetes risk is a function of the accumulated number of common risk variants that they carry [7]. The level of risk is predicted to be modulated further by potential interactions between specific variants, and other genetic variants and non-genetic factors.

The majority of the susceptibility loci have been linked to impaired pancreatic beta-cell function [8]. This is not surprising as decreased insulin secretion is a prerequisite for hyperglycaemia, the defining feature of all type 2 diabetes. Variation at the *TCF7L2* locus provides the strongest signal for type 2 diabetes risk. It is involved in the WNT signalling pathway, and early physiological studies pointed to a role in modulating the incretin-insulin secretory axis. However, the exact mechanisms by which variation in other susceptibility gene loci such as *CDKAL1* and *CNKN2A/2B* remain to be defined. An important lesson arose from the study of *FTO*. Variation in this gene was linked to increased BMI. However, subsequent analysis demonstrated that the key risk variant did not alter *FTO*

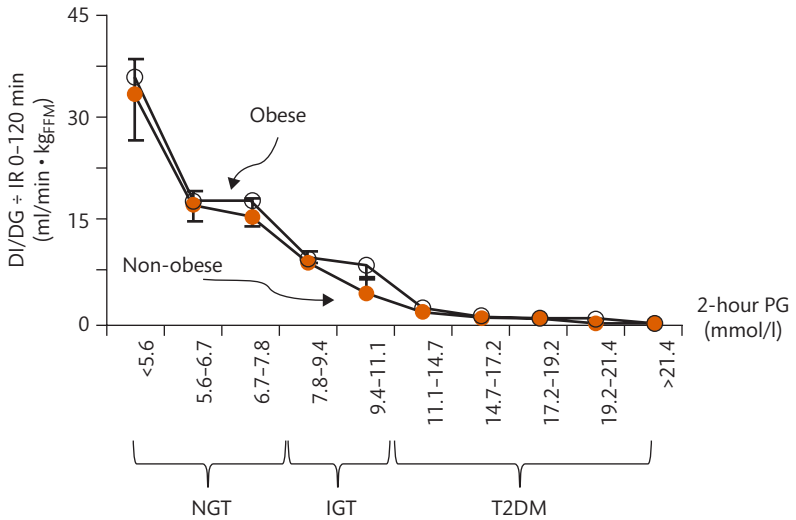


Figure 15.3.3.2 Disposition index (insulin secretion/insulin resistance); calculated as ΔAUC-Insulin/ΔAUC-Glucose ÷ clamp derived IR) plotted against 2-hour plasma glucose from an oral glucose tolerance test. Even in the normal glucose tolerance range, the disposition index markedly falls as the 2 hour plasma glucose increases in both obese and non-obese subject groups.

Gastaldelli A, Ferrannini E, Miyazaki Y, Matsuda M, DeFronzo RA. Beta- cell dysfunction and glucose intolerance; results from the San Antonio Metabolism (SAM) study. *Diabetologia*, 2004; **47**: 31–9. (Ref 2)

expression, but influenced expression of two distant genes, *IRX3* and *IRX5*, and these were involved in the process of thermogenesis [9]. In other words, while the diabetes risk variant was located at the *FTO* locus, the increased risk was mediated via regulation of genes located some physical distance away. This could complicate the delineation of the underlying mechanisms of diabetes susceptibility if it applies to other risk loci.

Adipose Tissue

Increased adiposity is a key risk factor for the development of type 2 diabetes. The rapid rise in the worldwide prevalence of type 2 diabetes is primarily driven by a concordant increase in the prevalence of obesity as described in Chapter 14.1.2. Our understanding of how and why obesity is linked to the pathogenesis of type 2 diabetes has evolved dramatically over the last few decades leading to the concept of lipotoxicity. A key question is what factors contribute to adipose tissue dysfunction that in turn impact on pancreatic beta-cell function and target tissue insulin sensitivity.

Adipose Tissue Dysfunction and Inflammation

Adipocyte number and distribution are key determinants of their function in man. Adipocyte size has been found to be negatively correlated with whole body insulin sensitivity in man [10]. An elegant study found that the number of adipocytes increases during childhood, but then remains stable throughout adulthood [11]. This implies that increased lipid storage in adulthood is achieved by adipocyte hypertrophy of the existing complement of cells. This is important as hypertrophy is linked to adipocyte stress, possibly through cellular hypoxia that is the trigger for a number of adverse processes including adipose tissue inflammation [12]. This involves the expression of pro-inflammatory cytokines by adipocytes and the attraction and infiltration of immune cells such as macrophages that serve to amplify the inflammatory process within the adipose tissue stores. A positive correlation was found between macrophage abundance and both BMI and adipocyte size in human adipose tissue [13]. The release of pro-inflammatory cytokines, for example TNF-alpha, is predicted to interfere with insulin action within the adipocytes and other local tissues such as muscle. Expansion of the adipose tissue stores leads to increased circulating fatty acid levels [14] and a relative decrease in circulating adiponectin [15]. Adiponectin is exclusively secreted by adipose tissue and exerts major effects on glucose metabolism in liver and muscle.

Adipose Tissue Insulin Resistance

Adipose tissue lipogenesis is regulated by insulin that promotes free fatty acid (FFA) re-esterification and triglyceride synthesis. Since adipose tissue lacks the enzyme glycerol kinase, it relies upon glucose and pyruvate to make glycerol-3-phosphate that is used to synthesize triglyceride. Insulin not only promotes fatty acid re-esterification but also glucose uptake in the adipose tissue. Moreover, insulin is the most potent antilipolytic hormone, since it inhibits triglyceride hydrolysis and fatty acid release. In type 2 diabetic patients, the antilipolytic effect of insulin is compromised leading to increased fasting and postprandial fatty acid concentrations, despite normal or raised circulating insulin levels [16]. This is recognized as adipose tissue insulin resistance that promotes the overflow of FFAs that are then taken up by other tissues and contribute to ectopic fat accumulation (see next).

Adipose Tissue Distribution

It is well recognized that adipose tissue distribution affects type 2 diabetes risk. Familial partial lipodystrophy is a monogenic condition characterized by decreased peripheral subcutaneous fat stores and compensatory increase in central fat deposition with increased visceral adiposity. This extreme form of fat redistribution is associated with decreased insulin sensitivity and increased risk of type 2 diabetes, fatty infiltration of the liver and cardiovascular disease. Genetic cluster analysis has been applied in two recent studies to identify overlapping clusters comprising 11 [17] and 53 [18] common genetic variants associated with decreased insulin sensitivity. Both studies found that the clusters were associated with increased central and decreased peripheral adiposity, but an increased risk of type 2 diabetes and coronary artery disease. This highlights the importance of altered adipose tissue distribution and diabetes risk.

Ectopic fat deposition describes the situation where lipid is stored in tissues other than adipose tissue, such as the liver, muscle, and pancreas. It arises when an individual's adipose storage capacity is overwhelmed, and is a particular problem in subjects with limited peripheral subcutaneous adipose tissue stores. Ectopic fat deposition *per se* is not a problem so long as the recipient tissues have the capability to fully metabolize the stored lipid. This capability appears to be impaired in type 2 diabetes and contributes to the situation of incomplete lipid metabolism and lipotoxicity.

Pancreatic Beta-Cell

Insulin secretion in non-diabetic subjects in response to either an oral or intravenous glucose load is characterized by an initial rapid response over 10 to 15 minutes (first phase) followed by slower more prolonged rise called the second phase [19]. It is well established that this insulin response is altered in type 2 diabetes, with a blunted first phase response and a delayed and lower second phase. These changes contribute to the hyperglycaemia of type 2 diabetes. It has been demonstrated that the first phase response plays a crucial role in regulating hepatic glucose output in non-diabetic subjects [20] and that the impaired first phase response leads to impaired suppression of hepatic glucose output and postprandial hyperglycaemia in type 2 diabetes [19].

By administering a progressive stepped glucose infusion, a dose response curve can be generated for insulin secretion rates against plasma glucose levels [21]. The slope of this curve decreases as glucose intolerance progresses to frank type 2 diabetes (**Figure 15.3.3.3**). In other words, there is a progressive decrease in the insulin secretion rate for a given plasma glucose concentration. Type 2 diabetes is characterized by abnormalities of pancreatic beta-cell function in the fasting state. Basal insulin secretion shows regular periodic oscillations in non-diabetic individuals which is lost in patients with type 2 diabetes [22]. Furthermore, the fasting proinsulin to insulin ratio is increased in type 2 diabetic patients compared to non-diabetic controls which is thought to reflect incomplete proinsulin processing in the diabetic state.

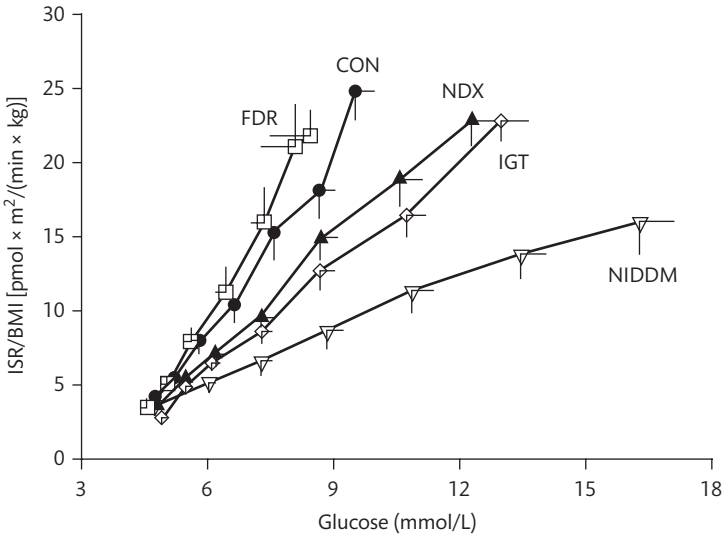


Figure 15.3.3.3 Dose response curves for insulin secretory rates versus plasma glucose concentration curves derived using a stepped glucose infusion. Curves are shown for subjects with different degrees of glucose tolerance, including those with normal glucose tolerance with (FDR) and without (CON) a family history of diabetes, and those with established type 2 diabetes (NIDDM).

While the pancreatic beta-cell phenotype is well characterized in type 2 diabetes, the underlying mechanisms remain to be determined. There is evidence that the impaired insulin secretion results from both decreased beta-cell function and mass.

Beta-Cell Genetic Susceptibility

As specified, the majority of genetic susceptibility loci identified to date are associated with impaired pancreatic beta-cell function. Some of the risk variants are within genes already associated with monogenic diabetes such as Wolfram Syndrome (*WFS1*) and different forms of Maturity Onset Diabetes of the Young (*GCK* and *HNF1A*). Others were within genes known to be involved in the insulin secretory process such *SLC30A8*, and *KCNJ11* that encodes part of the potassium channel. Physiological studies have tried to dissect the relationships between gene variants and the beta-cell phenotype. Such an approach showed that *TCF7L2, SLC30A8*, and a variant in the gastric inhibitory peptide receptor gene (*GIPR*) were all associated with raised circulating proinsulin levels and a decreased early insulin response, a pattern thought to reflect decreased insulin secretion on a background of decreased beta-cell mass/processing [8]. A further study examined the relationship between known risk variants and the insulin response to an intravenous glucose challenge to try to identify genes that affect the intrinsic beta-cell response [23]. Variants in *MTNR1B* and *CDKAL1* showed the strongest association, while variation in *TCF7L2* was comparatively weak supporting the concept that variation at this gene locus affects insulin secretion by mechanisms beyond the beta cell.

Beta-Cell Dysfunction

A series of mechanisms have been implicated in the development of pancreatic beta-cell dysfunction and type 2 diabetes. However, these have invariably been investigated using pancreatic islets or beta-cell lines derived from rodent models of diabetes. It well known that there are differences between rodent and human islets, for example, in terms of cellular composition and hyperplastic capacity. While the molecular mechanisms are expected to be the same across species, their relative importance may differ between rodents and man. With this in mind, emphasis has been given to those mechanisms where there is evidence for a role in type 2 diabetes in man.

The endoplasmic reticulum (ER) is a membranous tubular system involved in protein synthesis, and the removal of protein products. A number of conditions can lead to ER stress, including excess pro-inflammatory cytokines, and raised glucose and lipid levels that are known to impair beta-cell function [24]. ER stress elicits an adaptive reaction called the unfolded protein response (UPR) that temporarily decreases the ER synthetic output and increases the removal of protein products. In this way the stress is relieved and ER homeostasis is restored. A facet of the UPR in beta-cells is a decrease in insulin secretion that under chronic conditions will predispose to diabetes. Furthermore, failure of the UPR to restore ER homeostasis can with time lead to activation of the pro-apoptotic pathway and a loss of beta-cells [24]. There is evidence that ER dysfunction contributes to type 2 diabetes in man. First, mutations of the *WFS1* gene lead to altered ER function, and common low penetrant *WFS1* variants increase the susceptibility to type 2 diabetes [25]. ER distension is a marker of ER stress and has been identified in isolated pancreatic islets from type 2 diabetic patients [26].

There is strong evidence that sustained metabolic stress can impair pancreatic beta-cell function. Chronic elevated glucose levels decrease the expression of PDX-1, a key regulator of insulin gene transcription [27]. This appears to be mediated in part by the increased generation of reactive oxygen species (ROS) and underlies the concept of glucotoxicity. Similarly, a sustained increase in saturated fatty acid levels, such as palmitate, has been shown to impair beta-cell function and insulin secretion [28]. The mechanism appears to involve increased ROS production by the mitochondria and peroxisomes [29] and the induction of ER stress. The combination

of raised glucose and saturated fatty acid levels (glucolipotoxicity) is thought to be particularly harmful to beta-cell function.

Insulin secretion is highly energy dependent. This is reflected in the high mitochondrial content of pancreatic beta-cells that generates the required ATP through oxidative phosphorylation. Mutations of the mitochondrial DNA that impair mitochondrial function and ATP generation can lead to diabetes in man [30]. The role of mitochondrial dysfunction in type 2 diabetes remains to be fully defined. However, it is recognized that decreased mitochondrial DNA content can decrease insulin secretion, and that mitochondrial DNA content has been shown to decline with age in isolated human islets [31]. Uncoupling protein 2 (UCP2) is expressed by pancreatic beta-cells [32]. Increased UCP2 expression increases proton leak by the respiratory chain, leading to decreased ATP synthesis and decreased insulin secretion. Hyperglycaemia has been shown to increase UCP2 expression which could contribute to the mechanism of beta-cell glucotoxicity [32].

Immune cell infiltration and amyloid deposition have been identified in pancreatic islets in a proportion of patients with type 2 diabetes [33]. The macrophage is the predominant immune cell, with evidence for a pro-inflammatory subtype that predominates in type 2 diabetes islets and mediates the inflammatory process through cytokine release, in particular IL-1B. As a consequence, anti-IL-1B immune therapies have been developed and have been found to improve pancreatic beta-cell function and blood glucose levels in patients with type 2 diabetes [34]. There is extensive evidence from animal studies that elevated saturated fatty acids, including palmitate, trigger the pro-inflammatory process in pancreatic islets. The accumulation of islet amyloid is another activator [33]. Pancreatic islet amyloid deposition has been long been recognized as a feature of type 2 diabetes [35]. Islet amyloid polypeptide (IAPP) is cosecreted with insulin and is non-toxic in its monomeric form. However, there is recent evidence that oligomerization of human IAPP to form aggregates within islets can activate the pro-inflammatory response and impair insulin secretion in type 2 diabetes [33].

The mechanisms and their interactions contributing to pancreatic beta-cell dysfunction are summarized in Figure 15.3.3.4.

Beta-Cell Mass

There is consistent evidence for decreased beta-cell mass in type 2 diabetes, with estimates of 40–60% decrease in the number of beta cells [36]. A number of mechanisms may contribute to the decreased beta-cell mass in type 2 diabetes. Histological analyses of human islets have found an increase frequency of apoptotic markers in samples from type 2 diabetic compared with non-diabetic control subjects [37]. The increased beta-cell apoptosis could be driven by a number of factors through the common pathway of ER stress. Another process that has gained traction is beta-cell dedifferentiation. In a mouse model of diabetes, knockout of the transcription factor Fox01 resulted in a decrease in beta-cell mass and the development of diabetes [38]. Lineage tracing showed evidence of beta-cell dedifferentiation with a loss of expression of beta-cell markers, such as PDX1, and the expression of markers of other endocrine cell type including glucagon and somatostatin. Analysis of human pancreas samples reported an increase in the number of cells expressing markers of dedifferentiation and of other endocrine cell types, including glucagon [39], implying that dedifferentiation may be relevant to the pathogenesis of type 2 diabetes in man.

Glucagon and Incretins

Glucagon and glucagon-like peptide-1 (GLP-1) are hormones that are important for glucose metabolism and insulin release. Both are derived by the enzymatic cleavage of proglucagon, by tissue specific prohormone convertase (PC) enzymes. In the pancreatic α cells proglucagon is converted to glucagon, while in the intestinal L cells it is converted to GLP-1 [40].

Glucagon and insulin have opposite effects and counterregulate each other. Glucagon stimulates hepatic glucose production mainly by stimulating glycogenolysis [41]. Glucagon release is modulated by glucose (high glucose inhibits glucagon release while hypoglycaemia stimulates glucagon release) and by GLP-1 that suppresses glucagon secretion [42]. Glucagon concentrations are higher in

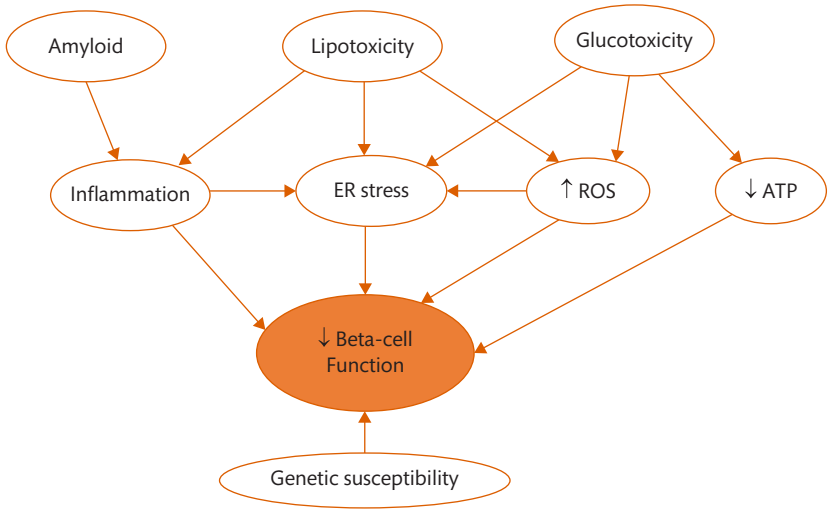


Figure 15.3.3.4 Schematic diagram summarizing the key mechanisms, and how they interact, that impair pancreatic beta-cell function in type 2 diabetes. ER = endoplasmic reticulum and ROS = reactive oxygen species.

T2D patients compared with non-diabetic subjects [41] and reduction of the plasma glucagon concentration is a potential target for treatment of patients with type 2 diabetes [43].

GLP-1 is an incretin hormone whose main action is the potentiation of the glucose stimulated insulin secretion [44]. GLP-1 is released by the intestinal L cells in response to nutrients (carbohydrates and lipids and its effects are mediated by activation of G-protein-coupled receptors) [45].

GLP-1 has a very short half-life (1–2 minutes) since it is rapidly degraded by dipeptidyl peptidase 4 (DPP-4). GLP-1 receptor agonists (GLP-1RA) and DPP-4 inhibitors are currently used as approved treatment for diabetic hyperglycaemia. The treatment with GLP-1RA has other beneficial effects that include the regulation of appetite and food intake [46].

Enzymatic cleavage of proglucagon also produces oxyntomodulin (OXM) that is a glucagon/GLP-1 receptor agonist [40]. OXM has a profound effect on appetite (reduced) and energy expenditure (increased) that improves glucose tolerance and insulin sensitivity [47].

Skeletal Muscle

Using a combination of hyperinsulinaemic clamp and limb catheterization techniques, it has been shown that skeletal muscle insulin stimulated glucose uptake is decreased in patients with type 2 diabetes [3]. Subsequent studies using magnetic resonance spectroscopy showed that the decreased insulin stimulated glucose uptake primarily resulted in decreased glycogen storage [48]. This combination of decreased insulin stimulated glucose and glycogen synthesis was also seen in the non-diabetic first-degree relatives of type 2 diabetes patients [49]. These subjects are at increased risk of progressing to diabetes, suggesting that impaired insulin action at skeletal muscle is an early defect in the pathogenesis of type 2 diabetes.

Insulin-Mediated Glucose Uptake

The insulin signalling pathway in relation to glucose metabolism is well defined [50]. The initial step involves insulin receptor binding and tyrosine kinase activation that leads to phosphorylation of insulin receptor substrate-1 (IRS1) that initiates a complex signalling cascade. For glucose uptake, the end point of the cascade is the activation of AS160 (previously known as TBC1D4) that stimulates translocation of GLUT4 to the cell membrane and glucose uptake. For glycogen synthesis, the principle end point is activation of glycogen synthase through the inhibition of the regulator, GSK-3.

A small proportion of the type 2 diabetes susceptibility variants identified by GWAS are associated with decreased insulin sensitivity [51]. These include *IGF1* and *IRS1* variants that are involved in signal transduction. No common variants that specifically and exclusively affect insulin sensitivity in skeletal muscle have been identified to date by GWAS. However, the retention of defects on insulin action in cultured skeletal muscle cells from type 2 diabetic patients [52] and their non-diabetic first-degree relatives [53] provides support for the role of genetic/epigenetic factors in the development of the peripheral insulin resistance in type 2 diabetes. Gene expression studies of skeletal muscle biopsy samples from type 2 diabetic patients found decreased expression of PGC1alpha, a transcriptional regulator of mitochondrial gene expression [54, 55]. It would seem, however, that these changes may not contribute directly to the peripheral insulin resistance in type 2 diabetes, as targeted *PGC1A* knockdown did not alter insulin action in skeletal muscle [56].

Metabolic and Inflammatory Mechanisms

Patients with type 2 diabetes have raised circulating FFA levels compared with weight matched non-diabetic control subjects [14]. In addition, patients with type 2 diabetes are prone to ectopic fat deposition in skeletal muscle. Interestingly, it has been shown that the intramuscular lipid content is increased to a comparable degree in type 2 diabetic patients and normal glucose tolerant athletes [57]. However, it would seem that the athletes but not the type 2 diabetic patients have the capacity for the efficient metabolism of the intramuscular lipid to generate energy.

The elevation of circulating FFA levels has been shown to decrease insulin signalling and impaired glucose uptake into skeletal muscle. Through a series of animal and human studies [58], it has been shown that elevated fatty acid levels increase intracellular lipid intermediaries that include diacylglycerol (DAG). DAG activates protein kinase C theta which serine phosphorylates IRS1 and inhibits downstream insulin signalling. As a consequence, there is decreased insulin stimulated glucose uptake and glycogen synthesis in the skeletal muscle. The same DAG-mediated mechanism is thought be involved in the decreased insulin action that is associated with ectopic lipid accumulation in skeletal muscle in type 2 diabetes.

Adipose tissue related inflammation is predicted to exacerbate skeletal muscle insulin sensitivity through the effect of locally released cytokines, in particular TNF-alpha, on skeletal muscle metabolism. TNF-alpha has been shown to inhibit insulin signalling through serine phosphorylation of IRS1 [59]. However, it is recognized that skeletal muscle *per se* can release a number of molecules, so called myokines, and includes IL-6. IL-6 release from contracting muscle increases as part of the normal physiological response to exercise in healthy man [60]. But there is emerging evidence that inflammation is a feature of skeletal muscle in type 2 diabetes, and that it mirrors the inflammatory process seen in other tissues in type 2 diabetes. Skeletal muscle cells from type 2 diabetic patients express a pro-inflammatory profile that persists in culture [61, 62], and there is evidence for skeletal muscle macrophage infiltration in type 2 diabetic patients [63].

Non-Insulin-Mediated Glucose Disposal

Insulin independent pathways also promote glucose uptake into skeletal muscle and clearance from the circulation. A major pathway involves AMPK activation, GLUT4 translocation, and increased glucose uptake. AMPK is the principle energy sensor of the cell and is activated by processes that diminish the intracellular ATP/AMP ratio [64]. A key activator of AMPK is muscle contraction, and underlines the potential role of exercise in controlling glycaemia in type 2 diabetes.

Adiponectin is another activator of AMPK and can exist as multiple complexes, with the high molecular form being the most active [65]. Adiponectin mediates its actions by binding to two receptors types, AdipoR1 and Adipo2. AdipoR1 is the predominant form in skeletal muscle, and receptor binding leads to activation of AMPK as well as other regulators of cellular energy provision such as

PGC1A. Through these mechanisms adiponectin increases GLUT4 translocation and glucose uptake into skeletal muscle [66]. There is evidence that the beneficial effects of adiponectin may be limited in type 2 diabetes. First, increased adiposity is associated with decreased circulating adiponectin levels and, second, adiponectin receptor expression has been reported to be decreased in skeletal muscle from type 2 diabetic patients [65].

Liver

The liver is the main organ that produces glucose during the fasting state. Glucose is produced either from the hydrolysis of the glycogen stored during the fed state or from gluconeogenesis (GNG). This involves the synthesis of glucose from gluconeogenic precursors such as lactate/pyruvate, gluconeogenic amino acids, and glycerol. While GNG can also occur in the kidney and small intestine, their contribution to the generation of glucose is minor relative to the liver.

In non-diabetic subjects, endogenous glucose production (EGP) by the liver is relatively similar among subjects when whole body fluxes are normalized by lean body mass [67]. However, in patients with type 2 diabetes the EGP is increased and closely linked to the fasting hyperglycaemia [68].

The application of stable isotope-based techniques has allowed the investigation of GNG in humans. After an overnight fast, GNG comprises approximately 50% of EGP in non-diabetic subjects and after 42h of fasting the great majority of glucose is produced as GNG as the hepatic glycogen stores are depleted [69]. The contribution of lactate/pyruvate, glycerol, or amino acids to GNG can vary. It has been shown that the liver utilizes these gluconeogenic precursors in proportion to their availability, a mechanism called 'hepatic autoregulation', although the overall rate of GNG remains steady [70]. However, in patients with type 2 diabetes, hepatic autoregulation is lost. As a consequence, an increase in gluconeogenic precursor supply leads to a direct increase GNG that in turn drives the increase in EGP [68].

Hepatic Insulin Resistance and Abnormal Hormonal Control

Hepatic glucose metabolism is tightly regulated by several hormones, in particular insulin and glucagon [71]. During the fasting state the most important hormone that regulates EGP is insulin by decreasing of hepatic glucose production. After a meal this is achieved through the inhibition of glycogenolysis and the stimulation of glycogen synthesis. The effect of insulin on gluconeogenesis is mainly indirect, since insulin also decreases peripheral lipolysis and protein catabolism thus diminishing the supply of gluconeogenic precursors to the liver.

In patients with type 2 diabetes, the ability of physiological insulin levels to suppress EGP is impaired indicating hepatic insulin resistance [3]. Using tracer techniques, it has been shown that EGP is raised under basal and elevated insulin levels in patients with type 2 diabetes due to the impaired ability of insulin to regulate hepatic gluconeogenesis [72]. Conversely, the effect of insulin to suppress hepatic glycogenolysis in type 2 diabetes is normal. Thus in type 2 diabetes hepatic insulin resistance is mediated by increased rates of gluconeogenesis under both fasting and insulinized states, thereby contributing to glucose overproduction under both conditions. The hepatic insulin resistance in type 2 diabetes is closely associated with fatty infiltration of the liver and decreased hepatic insulin clearance suggesting a potential mechanistic link [73].

The other important hormone that regulates hepatic metabolism is glucagon whose action is to stimulate gluconeogenesis, glycogenolysis, and net hepatic glucose output [71]. In type 2 diabetes, fasting plasma glucagon levels are increased and promote ongoing EGP.

Non-Alcoholic Fatty Liver Disease

Non-alcoholic fatty liver disease (NAFLD) affects almost 25% of the world population and the prevalence is much higher in patients with type 2 diabetes [74]. NAFLD is diagnosed when fat accumulation in the liver exceeds 5% in absence of excess alcohol intake [75].

Simple fat accumulation is considered a benign effect of excess fat overload to the liver. However, approximately 15–20% of subjects with NAFLD can progress to non-alcoholic steatohepatitis (NASH), defined by liver biopsy when together with steatosis hepatocytes show inflammation and ballooning with or without fibrosis [75]. These patients are at high risk of cirrhosis and hepatic cancer. International guidelines emphasize the importance of diagnosing NAFLD at its early stages in order to avoid or delay comorbidities, not only severe liver disease, but also cardiovascular disease and cancer [75, 76].

The pathophysiology of NAFLD is incompletely understood, although it is established that increased adiposity and diabetes are major risk factors. Several genotypes have been found to increase the risk of NAFLD and its progression to NASH, that include variants in *PNPLA3, TM6SF2, DGAT*, and *MBOAT7* [77]. Intriguingly, these genotypes are associated only with NAFLD and not insulin resistance [78].

Hepatic insulin resistance is increased in patients with type 2 diabetes and NAFLD, possibly related to an increased GNG [79]. It has been proposed that increased hepatic fat decreases insulin signalling by the same lipotoxicity mechanisms that operate in muscle [58]. Subjects with NAFLD have also increased peripheral insulin resistance in muscle and adipose tissue [79]. This peripheral insulin resistance results in unrestrained lipolysis and increased glycerol and FFA release that fuel hepatic lipogenesis and gluconeogenesis as summarized in Figure 15.3.3.5.

Conclusions

The principal pathophysiological changes of decreased pancreatic beta-cell function and impaired insulin action that characterize type 2 diabetes are well defined. Headway is now being made in defining the mechanisms that lead to these pathophysiological changes. As described in this chapter, a number of mechanisms are emerging that are common across the different tissues involved in the development of type 2 diabetes. In particular, ectopic fat deposition, tissue lipotoxicity, and the activation of pro-inflammatory pathways are emerging as common mechanisms. This presents the possibility that interventions targeted to these mechanisms could act across multiple tissues leading to an amplified therapeutic impact.

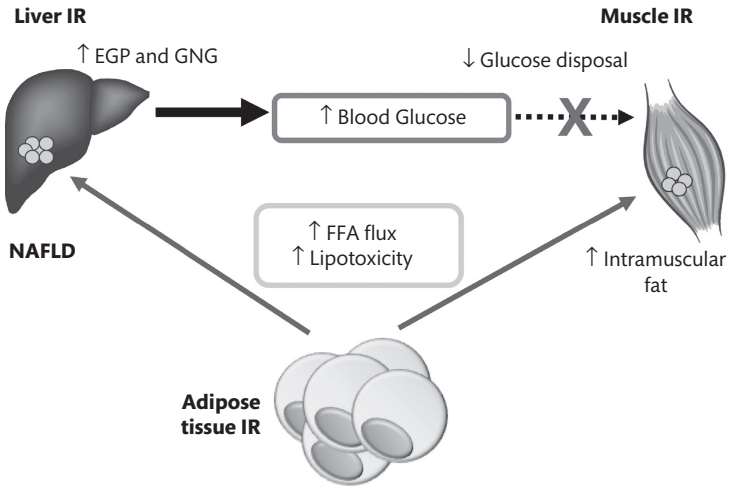


Figure 15.3.3.5 Insulin resistance (IR) at the liver leads to increased gluconeogenesis (GNG) and endogenous glucose production (EGP), and at the muscle it leads to decreased glucose clearance. Adipose tissue IR results in impaired suppression of lipolysis and increased FFA flux to increase hepatic lipogenesis and gluconeogenesis, and to impair insulin action in muscle.

REFERENCES

1. Retnakaran R, Shen S, Hanley AJ, Vuksan V, Hamilton JK, Zinman B. Hyperbolic relationship between insulin secretion and sensitivity on oral glucose tolerance test. *Obesity*, 2008; **16**: 1901–7.
2. Gastaldelli A, Ferrannini E, Miyazaki Y, Matsuda M, DeFronzo RA. Beta-cell dysfunction and glucose intolerance; results from the San Antonio Metabolism (SAM) study. *Diabetologia*, 2004; **47**: 31–9.
3. DeFronzo, RD. Pathogenesis of type 2 diabetes mellitus. *Med Clin N Am*, 2004; **88**: 787–835.
4. The Wellcome Trust Case Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007; **447**: 661–78.
5. Fuchsberger C, Flannick J, Teslovich TM, *et al.* The genetic architecture of type 2 diabetes. *Nature*, 2016; **563**: 41–7.
6. Moltke I, Grarup N, Jørgensen ME, *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*, 2014; **512**: 190–204.
7. Weedon MN, McCarthy MI, Hitman G, *et al.* Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med*, 2006; **3**: e374.
8. Ingelsson E, Langenberg C, Hivert M-F, *et al.* Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. *Diabetes*, 2010; **59**: 1266–75.
9. Claussnitzer M, Dankel SN, Kim K-H, *et al.* FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*, 2015; **373**: 895–905.
10. Arner P, Arner E, Hammarstedt A, Smith U. Genetic predisposition for type 2 diabetes, but not for overweight/obesity, is associated with a restricted adipogenesis. *PLoS One*, 2011; **6**: e18284.
11. Spalding KL, Arner E, Westermark PO, *et al.* Dynamics of fat cell turnover in humans. *Nature*, 2008; **453**: 783–7.
12. Schenk S, Saberi M, Olefsky JM. Insulin sensitivity: modulation by nutrients and inflammation. *J Clin Investig*, 2008; **118**: 2992–3002.
13. Weisberg SP, McCann D, Desai M, Rosenbaum M, Leibel R, Ferrante ML. Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest*, 2003; **112**: 1796–808.
14. Boden G. Role of fatty acids in the pathogenesis of insulin resistance and NIDDM. *Perspect Diabetes*, 1997; **46**: 1–10.
15. Arita Y, Kihara S, Ouchi N, *et al.* Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. *Biochem Biophys Res Commun*, 1999; **257**: 79–83.
16. Gastaldelli A, Gaggini M, DeFronzo RA. Role of adipose tissue insulin resistance in the natural history of type 2 diabetes: results from the San Antonio metabolism study. *Diabetes*, 2017; **66**: 815–22.
17. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*, 2017; **49**: 17–26.
18. Yaghootkar H, Scott RA, White CC, *et al.* Genetic evidence for a normal-weight 'metabolically obese' phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes*, 2014; **63**: 4369–77.
19. Del Prato, S. Loss of early insulin secretion leads to postprandial hyperglycaemia. *Diabetologia*, 2003; **46**(Suppl1): M2–M8.
20. Luzi L, Defronzo RA. Effect of loss of first-phase insulin secretion on hepatic glucose production and tissue glucose disposal in humans. *Am J Physiol Endocrinol Metab*, 1989; **257**: E241–6.
21. Byrne, MM, Sturis J, Sobel RJ, Polonsky KS. Elevated plasma glucose 2 h postchallenge predicts defects in β-cell function. *Am J Physiol Endocrinol Metab*, 1996; **276**: E572–9.
22. Pørksen N, Hollingdal M, Juhl C, Butler P, Veldhuis JD, Schmitz O. Pulsatile insulin secretion: detection, regulation, and role in diabetes. *Diabetes*, 2002; **51**(Suppl1): S245–54.
23. Wood AR, Jonsson A, Jackson AU, *et al.* A genome-wide association study of IVGTT-based measures of first-phase insulin secretion refines the underlying physiology of type 2 diabetes variants. *Diabetes*, 2017; **66**: 2296–309.
24. Herbert TP, Ross Laybutt D. A reevaluation of the role of the unfolded protein response in islet dysfunction: maladaptation or a failure to adapt? *Diabetes*, 2016; **65**: 1472–80.
25. Sandhu MS, Weedon MN, Fawcett KA, *et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet*, 2007; **39**: 951–3.
26. Marchetti P, Bugliani M, Lupi R, *et al.* The endoplasmic reticulum in pancreatic beta cells of type 2 diabetes patients. *Diabetologia*, 2007; **50**: 2486–94.

27. Paul Robertson R. Chronic oxidative stress as a central mechanism for glucose toxicity in pancreatic islet beta cells in diabetes. *J Biol Chem*, 2004; **279**: 42351–4.
28. Ly LD, Xu S, Choi SK, *et al.* Oxidative stress and calcium dysregulation by palmitate in type 2 diabetes. *Exp Mol Med*, 2017; **49**: e291.
29. Elsner, M, Gehrmann W, Lenzen S. Peroxisome-generated hydrogen peroxide as important mediator of lipotoxicity in insulin-producing cells. *Diabetes*, 2011; **60**: 200–8.
30. Schaefer AM, Walker M, Turnbull DM, Taylor RW. Endocrine disorders in mitochondrial disease. *Mol Cell Endocrinol*, 2013; **379**: 2–11.
31. Cree LM, Patel SK, Pyle A, *et al.* Age-related decline in mitochondrial DNA copy number in isolated human pancreatic islets. *Diabetologia*, 2008; **51**: 1440–3.
32. Lowell BB, Shulman GI. Mitochondrial dysfunction and type 2 diabetes. *Science*, 2005; **307**: 384–7.
33. Eguchi K, Nagai R. Islet inflammation in type 2 diabetes and physiology. *J Clin Invest*, 2017; **127**: 14–23.
34. Larsen CM, Faulenbach M, Vaag A, *et al.* Interleukin-1–receptor antagonist in type 2 diabetes mellitus. *N Engl J Med*, 2007; **356**: 1517–26.
35. Clark A, Nilsson MR. Islet amyloid: a complication of islet dysfunction or an aetiological factor in Type 2 diabetes? *Diabetologia*, 2004; **47**: 157–69.
36. Chen C, Cohrs CM, Stertmann J, Bozsak R, Speier S. Human beta cell mass and function in diabetes: recent advances in knowledge and technologies to understand disease pathogenesis. *Mol Metab*, 2017; **6**: 943–7.
37. Marchetti P, Del Guerra S, Marselli L, *et al.* Pancreatic islets from type 2 diabetic patients have functional defects and increased apoptosis that are ameliorated by metformin. *J Clin Endocrinol Metab*, 2004; **89**: 5535–41.
38. Talchai C, Xuan S, Lin HV, Sussel L, Accili D. Pancreatic β Cell Dedifferentiation as a Mechanism of Diabetic β Cell Failure. *Cell*, 2012; **150**: 1223–34.
39. Cinti F, Bouchi R, Kim-Muller JY, *et al.* Evidence of β-Cell dedifferentiation in human type 2 diabetes. *J Clin Endocrinol Metab*, 2016, **101**: 1044–54.
40. Gastaldelli A, Gaggini M, DeFronzo R. Glucose kinetics: an update and novel insights into its regulation by glucagon and GLP-1. *Curr Opin Clin Nutr Metab Care*, 2017; **20**: 300–9.
41. Raskin P, Unger RH. Hyperglucagonemia and its suppression. Importance in the metabolic control of diabetes. *N Engl J Med*, 1978; **299**: 433–36.
42. Zander M, Madsbad S, Madsen JL, Holst JJ. Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and beta-cell function in type 2 diabetes: a parallel-group study. *Lancet*, 2002; **359**: 824–30.
43. Pearson MJ, Unger RH, Holland WL. Clinical trials, triumphs, and tribulations of glucagon receptor antagonists. *Diabetes Care*, 2016; **39**: 1075–7.
44. Campbell JE, Drucker DJ. Pharmacology, physiology, and mechanisms of incretin hormone action Cell. *Metab*, 2013; **17**: 819–37.
45. Holst JJ, Seino Y. GLP-1 receptor agonists: targeting both hyperglycaemia and disease processes in diabetes. *Diabetes Res Clin Pract*, 2009; **85**: 1–3.
46. Flint A, Raben A, Astrup A, Holst JJ. Glucagon-like peptide 1 promotes satiety and suppresses energy intake in humans. *J Clin Invest*, 1998; **101**: 515–20.
47. Pocai A. Action and therapeutic potential of oxyntomodulin. *Mol Metab*, 2014; **3**: 241–51.
48. Shulman GI, Rothman DL, Jue T, Stein P, DeFronzo RA, Shulman RG. Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by ^{13}C nuclear magnetic resonance spectroscopy. *N Engl J Med*, 1990; **322**: 223–8.
49. Rothman DL, Magnusson I, Cline G, *et al.* Decreased muscle glucose transport/phosphorylation is an early defect in the pathogenesis of non-insulin-dependent diabetes mellitus. *Proc Natl Acad Sci*, 1995; **92**: 983–7.
50. Saltiel AR, Ronald Kahn C. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 2001; **414**: 799–806.
51. Brown AE, Walker M. Genetics of insulin resistance and the metabolic syndrome. *Curr Cardiol Rep*, 2016; **18**: 75.
52. Henry RR, Abrams L, Nikoulina S, Ciaraldi TP. Insulin action and glucose metabolism in nondiabetic control and NIDDM subjects. Comparison using human skeletal muscle cell cultures. *Diabetes*, 1995; **44**: 936–46.
53. Jackson S, Bagstaff SM, Lynn S, Yeaman SJ, Turnbull DM, Walker M. Decreased insulin responsiveness for glucose uptake in cultured human skeletal muscle cells from insulin resistant non-diabetic relatives of type 2 diabetic families. *Diabetes*, 2000; **49**: 1169–77.
54. Patti ME, Butte AJ, Crunkhorn S, *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: potential role of PGC1 and NRF1. *PNAS*, 2003; **100**: 8466–71.
55. Mootha VK, Lindgren CM, Eriksson K-F, *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 2003; **34**: 267–73.
56. Handschin C, Choi CS, Chin S, *et al.* Abnormal glucose homeostasis in skeletal muscle–specific PGC-1α knockout mice reveals skeletal muscle–pancreatic β cell crosstalk. *J Clin Invest*, 2007; **117**: 3463–74.
57. Goodpaster BH, He J, Watkins S, Kelley DE. Skeletal muscle lipid content and insulin resistance: evidence for a paradox in endurance-trained athletes. *J Clin Endocrinol Metab*, 2001; **86**: 5755–61.
58. Shulman GI. Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. *N Engl J Med*, 2014; **371**: 1131–41.
59. Gökhan S, Hotamisligil S, Murray DL, Choy LN, Spiegelman BM. Tumor necrosis factor A inhibits signaling from the insulin receptor. *Proc Natl Acad Sci U S A*, 1994; **91**: 4854–8.
60. Pedersen BK, Steensberg A, Schjerling P. Muscle-derived interleukin-6: possible biological effects. *J Physiol*, 2001; **536**: 329–37.
61. Brown AE, Palsgaard J, Borup R, *et al.* p38 MAPK activation upregulates pro-inflammatory pathways in skeletal muscle cells from insulin resistant type 2 diabetic patients. *AJP Endocrinol Metab*, 2015; **308**, E63–E70.
62. Green CJ, Pedersen M, Pedersen BK, Scheele C. Elevated NF-kappaB activation is conserved in human myocytes cultured from obese type 2 diabetic patients and attenuated by AMP-activated protein kinase. *Diabetes*, 2011; **60**: 2810–19.
63. Torres SH, De Sanctis JB, de LBM, Hernandez N, Finol HJ. Inflammation and nitric oxide production in skeletal muscle of type 2 diabetic patients. *J Endocrinol*, 2004; **181**: 419–27.
64. Hardie DG, Carling D. The AMP-activated protein kinase. Fuel gauge of the mammalian cell? *Eur J Biochem*, 1997; **246**: 259–73.

65. Kadowaki T, Yamauchi T, Kubota N, Hara K, Ueki K, Tobe K. Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. *J Clin Invest*, 2006; **116**: 1784–92.
66. Liu Y, Sweeney G. Adiponectin action in skeletal muscle. *Best Pract Res Clin Endocrinol Metab*, 2014; **28**: 33–41.
67. Natali A, Toschi E, Camastra S, Gastaldelli A, Groop L, Ferrannini E. Determinants of postabsorptive endogenous glucose output in non-diabetic subjects. European Group for the Study of Insulin Resistance (EGIR). *Diabetologia*, 2000; **43**: 1266–72.
68. Gastaldelli A, Baldi S, Pettiti M, *et al.* Influence of obesity and type 2 diabetes on gluconeogenesis and glucose output in humans: a quantitative study. *Diabetes*, 2000; **49**: 1367–73.
69. Landau BR, Wahren J, Chandramouli V, Schumann WC, Ekberg K, Kalhan SC. Contributions of gluconeogenesis to glucose production in the fasted state. *J Clin Invest*, 1996; **98**: 378–85.
70. Moore MC, Connolly CC, Cherrington AD. Autoregulation of hepatic glucose production. *Eur J Endocrinol*, 1998; **138**: 240–8.
71. Roden M, Perseghin G, Petersen KF, *et al.* The roles of insulin and glucagon in the regulation of hepatic glycogen synthesis and turnover in humans. *J Clin Invest*, 1996; **97**: 642–8.
72. Gastaldelli AE, Toschi M, Pettiti S, *et al.* Effect of physiological hyperinsulinemia on gluconeogenesis in nondiabetic subjects and in type 2 diabetic patients. *Diabetes*, 2001; **50**: 1807–12.
73. Kotronen A, Juurinen L, Tikkainen M, Vehkavaara S, Yki-Järvinen H. Increased liver fat, impaired insulin clearance, and hepatic and adipose tissue insulin resistance in type 2 diabetes. *Gastroenterology*, 2008; **135**: 122–30.
74. Younossi Z, Anstee QM, Marietti M, *et al.* Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol*, 2018; **15**: 11–20.
75. EASL, Marchesini G, Day CP, *et al.* EASL-EASD-EASO clinical practice guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol*, 2016; **64**: 1388–402.
76. Chalasani N, Younossi Z, Lavine JE, *et al.* The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology*, 2012; **142**: 1592–609.
77. Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol*, 2017; **23**: 1–12.
78. Petaja EM, Yki-Järvinen H. Definitions of normal liver fat and the association of insulin sensitivity with acquired and genetic NAFLD—a systematic review. *Int J Mol Sci*, 2016; **17**: pii: E633.
79. Gastaldelli A, Cusi K, Pettiti M, *et al.* Relationship between hepatic/visceral fat and hepatic insulin resistance in nondiabetic and type 2 diabetic subjects. *Gastroenterology*, 2007; **133**: 496–506.

15.4

Non Type 1, Non Type 2 Diabetes

15.4.1 Diagnosis of Non Type 1, Non Type 2 Forms of Diabetes

Katharine R. Owen

Introduction

Diabetes is an umbrella term for a raised blood glucose caused by a wide variety of different aetiologies. It is defined by exceeding a blood glucose threshold, and for many years was divided by treatment modality—insulin or not. The latter delineation has tended to rather stubbornly persist, with the labels 'NIDDM' (non-insulin dependent diabetes mellitus) and 'IDDM' (insulin dependent diabetes mellitus) still creeping into notes and referral letters. All too often the difference in aetiology and management between even type 1 and type 2 diabetes is not well appreciated, and the lack of a specialist input to assess aetiology around the time of diagnosis means that many individuals are miscoded [1]. This may lead to the use of inappropriate patient pathways, suboptimal treatment, and missed educational opportunities.

Although type 1 and type 2 diabetes together account for about 95% of people with diabetes, the remaining few percent reflect a large number of individuals with an eclectic group of aetiologies. In this chapter we will discuss how the correct use of clinical assessment along with a few basic investigations can differentiate aetiology at diagnosis (and beyond) and guide when more specialized investigations are required.

The wide range of differential diagnosis is particularly striking in diabetes arising in the young adult age range, up to around 40 years of age. Most monogenic causes of diabetes arise in this age range, as well as both classic and more slow-onset autoimmune diabetes, and an increasing prevalence of young-onset type 2 diabetes. There is often considerable clinical overlap between the subtypes so a careful assessment is required.

It is important to correctly diagnose the aetiology for a number of reasons. This will influence the first line treatment e.g. insulin for type 1 diabetes, low dose sulphonylureas for HNF1A- or HNF4A-MODY or metformin for unambiguous type 2 diabetes. Similarly, early interventions aimed at altering progression of diabetes, such as immunotherapy or closed loop glucose-regulated insulin delivery in type 1 diabetes; or weight reduction programmes to induce remission in type 2 diabetes rely on accurate assignment of aetiology. Finally, for the single gene causes of diabetes, family members require screening for diabetes and/or the gene variant.

Causes of Diabetes

Box 15.4.1.2 shows a classification incorporating some examples of the wide range of 'Non-type 1, non-type 2' causes of diabetes [2]. These include monogenic causes of beta-cell function and insulin resistance, primary pancreas pathologies (Chapter 15.10.3), endocrinopathies affecting glucose metabolism (Chapter 15.10.3), other rare syndromes, and drug-induced diabetes. Diabetes may be the only feature, or may present before or after other features of an associated condition. It is important to consider whether other clinical findings are coincident, or part of a unifying diagnosis. Thus the full clinical picture should be taken into account when assessing aetiology of diabetes, and the diagnosis revisited if additional features develop or information emerges (e.g. see Box 15.4.1.2). The finding of a monogenic or syndromic cause for diabetes should also trigger screening for other associated features of the syndrome—for example, investigation of the renal tract in those with diabetes and an *HNF1B* mutation, or cardiomyopathy screening in mitochondrial diabetes.

Table 15.4.1.1 contrasts the clinical features of type 1 and type 2 diabetes with the commoner forms of monogenic diabetes.

Monogenic Beta-Cell Dysfunction

Monogenic beta-cell dysfunction comprises neonatal diabetes (diagnosed before 6 months of age); Maturity Onset Diabetes of the Young (MODY); and mitochondrial diabetes. Some monogenic syndromes (e.g. Wolfram syndrome) also cause diabetes largely through beta-cell dysfunction. It is increasingly appreciated that there is an overlap between the genes causing neonatal diabetes and those causing MODY and it is important to specify the gene involved when describing cases. In general, MODY arises in adolescents and young adults, and the majority of children diagnosed

Box 15.4.1.1 Aetiology of diabetes

Autoimmune diabetes
- Classic type 1 diabetes
- Latent autoimmune diabetes of adulthood
- Monogenic autoimmunity

Type 2 diabetes

An insulin secretory defect with varying levels of insulin resistance; other specific diagnoses excluded

Genetic defects of β-cell function
- MODY
- Neonatal diabetes
- Mitochondrial diabetes

Genetic defects of insulin action
- Lipodystrophy
- Insulin receptor mutations
- Post-receptor insulin signalling defects
- Genetic severe obesity

Diseases of the exocrine pancreas
- Pancreatitis
- Haemochromatosis
- Cystic fibrosis-related diabetes
- Neoplasia

Endocrinopathies
- Cushing's syndrome
- Acromegaly
- Phaeochromocytoma

Syndromes associated with diabetes
- Chromosome abnormalities: Down's syndrome, Turner syndrome
- Myotonic dystrophy
- Friedreich ataxia
- Werner and other progeria syndromes
- Alström syndrome

Drug-induced diabetes
- Corticosteroids
- Diazoxide
- Thiazides
- Atypical antipsychotics
- Checkpoint inhibitors

between 6 months and 10 years have type 1 diabetes. One exception is glucokinase-MODY which can be diagnosed at any age.

MODY is the commonest form of monogenic diabetes seen in clinical practice. This is usually characterized by (Table 15.4.1.1):

- Diabetes onset in the 2nd–4th decade
- Absence of beta-cell antibodies/high risk HLA associated with type 1 diabetes
- Absence of obesity/insulin resistance associated with type 2 diabetes
- Non-acute presentation/DKA rare
- Frequent parental and other family history of diabetes

The most frequent subtypes seen in the clinic are due to mutations in transcription factors (Hepatocyte Nuclear Factor (HNF) 1A, HNF4A, HNF1B) and the glycolytic enzyme, glucokinase (GCK). See Chapter 15.10.1 for a more detailed description of the MODY subtypes and details of management. MODY is found in 1–6% of those with childhood or young-adult onset diabetes [3, 4] and has important treatment implications for those diagnosed with the commoner subtypes. Diagnostic genetic testing is now widely available, although uptake of this tends to be variable, dictated more by clinician experience and interest rather than formal guidelines.

It is very important to take a careful history of age of onset in individuals who were diagnosed in young childhood. A monogenic cause can be identified in more than 80% of those with a clinical diagnosis of neonatal diabetes (diabetes arising in the first 6 months of life) [5], often with an associated syndrome (see Chapter 15.10.1). Genetic testing is mandated in all those diagnosed under the age of 1 year, and (similarly to MODY), making a molecular diagnosis may have a substantial impact on treatment. Increasingly the early molecular diagnosis of syndromes presenting with diabetes in neonatal life means that timely counselling and screening for other features of the condition is possible.

Mitochondrial diabetes is due to mutations in the small mitochondrial genome [6]. Classically these individuals have diabetes and sensorineural deafness (see Box 15.4.1.2), but other neurological features are common and patients often complain of fatigue due to myopathy. Management of individuals with mitochondrial diabetes requires a multidisciplinary approach, including audiology, cardiology, and neurology review.

Genetic Defects of Insulin Action

Monogenic severe insulin resistance consists of disorders of body fat distribution (lipodystrophy), insulin signalling defects including insulin receptor mutations and syndromes associated with severe obesity [7]. More detailed descriptions of the subtypes can be found in Chapter 15.10.2. Severe insulin resistance is less common than MODY, but may still be encountered in a general diabetes clinic. Cases may also present to the endocrine clinic as polycystic ovarian syndrome in young women.

One clue to a rare monogenic form of severe insulin resistance is that features of metabolic syndrome may occur without significant obesity, or with evidence of central obesity only. Reactive hypoglycaemia may be a prominent early feature.

Diagnostic genetic testing is available for the commoner forms of severe insulin resistance including the lipodystrophy genes Lamin A/C (*LMNA*) and Peroxisome proliferator-activated gamma (*PPARG*) and for the insulin receptor.

Box 15.4.1.2 Case study—reassessing diagnosis

A 23-year-old woman presented to her GP with tinnitus. She also mentioned thirst and a blood glucose was requested. A fasting glucose of 7.2 mmol/L confirmed the diagnosis of diabetes and she was referred to the hospital clinic. There were no acute symptoms, weight loss, or ketonaemia. She was lean with no features of insulin resistance and 2 beta-cell antibodies were negative, so MODY was suspected. Genetic sequencing of the MODY genes *HNF1A*, *HNF4A*, and *GCK* was performed without any abnormal findings. She was commenced on metformin and reviewed after 3 months.

At review it was noted that she was wearing bilateral hearing aids. In the meantime, the original presentation of tinnitus had been investigated and she had been shown to have sensorineural hearing loss. This led to a suspicion of mitochondrial disease and the common A3243G mitochondrial mutation was confirmed on further genetic testing. Metformin was discontinued and she was managed on gliclazide, progressing to require insulin after a further 2 years.

Table 15.4.1.1 Clinical features associated with common forms of monogenic diabetes in comparison to type 1 and type 2 diabetes

	Age of onset	Features of insulin resistance	Family history of diabetes	Beta-cell antibodies	C-peptide	Neurological features	Exocrine pancreas dysfunction	Other features
Type 1 Diabetes	Peak incidence 1st–2nd decades, diagnosis often missed in older adults	No	10% of cases	80–90%	Classically C-peptide negative, but usually low detectable level at diagnosis with maintained low level in up to 10%	No	No	
Type 2 diabetes	Peak incidence after 4th decade, still relatively rare in paediatric age range	Typically found	50% of cases	No, by definition	C-peptide typically high at diagnosis, declines over time	No	No	
HNF1A/HNF4A MODY	2nd–4th decade	No	60–90% report parental diabetes	Not usually, some mixed MODY/T1 reported	Generally low normal range	No	No	HNF1A: Low renal glucose threshold, Low CRP HNF4A: fetal hyperinsulinaemia and macrosomia
GCK-MODY	Fasting hyperglycaemia present throughout life, may be diagnosed at any age	No	Less common that in other forms of MODY as often asymptomatic	Not usually, some mixed MODY/T1 reported	Normal range	No	No	Treatment not usually required Low risk of microvascular complications
HNF1B-MODY	Diabetes presents in 2nd–4th decade, other features may present earlier	Not typical	Less common than in other forms of MODY as spontaneous whole gene deletions are a frequent cause	No	Low normal range Usually progress to insulin quickly	No	Yes, common	Structural abnormalities of renal and genitourinary (GU) system, renal failure
Mitochondrial diabetes	Typically 3rd–4th decade	Not typical	Typically maternally transmitted, but clinical phenotype very variable due to heteroplasmy	No	Low normal range Usually progress to insulin quickly	Yes, typical	No	Myopathy, cardiomyopathy, pigmented retinopathy
Partial lipodystrophy	2nd-4th decade	Yes, often disproportionately to BMI	Autosomal dominant inheritance, but parents may be asymptomatic	No	High reflecting insulin resistance	Not typically	No	Abnormal body fat distribution from puberty with central adiposity and decreased fat on limbs. Muscular hypertrophy
Insulin receptor (IR) defects	Frequently presents in females with PCOS during puberty, rare severe cases in childhood. Males may be relatively asymptomatic	Partial features of metabolic syndrome: acanthosis nigricans and PCOS but not dyslipidaemia or hepatic steatosis	Parental diabetes may be seen in autosomal dominantly inherited cases, parents may be asymptomatic	No	High reflecting insulin resistance	Not typical	No	Postprandial hypoglycaemia Failure to thrive in childhood forms

Tools for Diagnosis

Clinical Assessment: Features from History, Examination, and Routine Biochemistry

Table 15.4.1.1 compares the features of the commoner types of monogenic diabetes in young adults. The most obvious clinical discriminator favouring type 2 diabetes, as opposed to type 1 diabetes or MODY, is obesity with associated features of metabolic syndrome such as dyslipidaemia, hypertension, hepatic steatosis, and PCOS. Clinical examination at diagnosis is important to detect acanthosis nigricans and lipodystrophy. Features of insulin resistance in the absence of obesity are strongly suggestive of genetically determined severe insulin resistance (see Chapter 15.10.2) warranting further evaluation.

As just discussed, coexistence of other clinical features with diabetes may give clues to the aetiology.

Beta-Cell Antibodies

At diagnosis 80–90% of those with type 1 diabetes will have at least one positive beta-cell antibody. In general younger individuals (under 45 years old) with positive antibodies are considered to have autoimmune diabetes. They may have a classic onset of type 1 diabetes, or a slower progression without initial insulin requirement consistent with a diagnosis of latent autoimmune diabetes of adulthood (seen in 10–20% of apparent young-onset type 2 diabetes [8]). Beta-cell antibodies are a good test for helping to confirm a diagnosis of type 1 diabetes in children and young adults, while absence of antibodies may indicate that an alternative diagnosis should be considered. However, there are some limitations to the tests which it is important to understand [9].

Firstly older tests such as Islet cell antibodies (ICA) are still commonly measured, but are difficult to standardize due to the qualitative immunofluorescence method used and have a low sensitivity. Glutamic acid decarboxylase (GAD), tyrosine phosphatase islet antigen 2 (IA2) and Zinc transporter (ZnT8) antibodies are measured using immunoassay which is standardized in an international laboratory programme. An appropriate reference range from the population being tested should be used. GAD is most often offered in diagnostic labs. Where only one or two antibodies are offered, there is a risk of missing a positive result, so that a negative antibody test does not exclude a diagnosis of type 1 diabetes.

Interpreting a positive test may also be difficult. A small percentage of people with non-autoimmune diabetes (as well as non-diabetic individuals) will have false-positive beta-cell antibodies depending on the threshold of the test. In older individuals where the proportion of people with type 2 diabetes is higher, the finding of a single positive antibody has a higher probability of being a false-positive test. This presents a conundrum—we know that type 1 diabetes in older adults is often missed, but the main diagnostic test we have at our disposal becomes less useful.

C-peptide

C-peptide is a marker for endogenous insulin secretion and is thus a useful indicator of the severe insulin deficiency seen in type 1 diabetes. C-peptide tends to be preserved in the low normal range in MODY and long-duration type 2 diabetes and is generally high as a marker of insulin resistance in more recent onset type 2, severe insulin resistance, or prediabetes. C-peptide testing can thus be used in those with insulin treated diabetes to confirm absent or very low insulin secretion (usually means type 1 diabetes), and where it is present to suggest an alternative diagnosis. C-peptide should ideally also be measured if discontinuation of insulin treatment is planned, particularly if insulin has been used since diagnosis. This gives confidence to both patients and healthcare professionals that this is a safe thing to do.

Interpreting C-peptide levels is not always straightforward, as both ambient glucose level and insulin sensitivity affect the insulin response. Measuring a paired glucose with any measure of insulin or C-peptide gives context for interpretation of the value, as a higher glucose provides greater beta-cell stimulation. C-peptide test should be repeated if the paired glucose is below 4 mmol/L.

In most studies of type 1 diabetes a stimulated C-peptide or C-peptide increment on stimulation of ≥200 pmol/L is considered to represent residual endogenous pancreatic insulin secretion [10]. A fasting C-peptide in an insulin-sensitive person could be lower than this and still represent useful insulin secretion. However in the context of a high glucose, a C-peptide of 200 pmol/L would be a poor response. There are no definitive thresholds for making treatment decisions, so the clinical scenario should be taken into account and gaining experience in using C-peptide values is important. For that reason C-peptide is probably best used in clinical settings where expertise exists to interpret results.

Research settings use standardized tests of beta-cell function such as the Homeostatic model assessment (HOMA) for the fasting state and glucagon, oral glucose tolerance test, or mixed meal tests for stimulated C-peptide, however formal testing is not usually needed to make clinical decisions.

Duration of diabetes is clearly one of the main factors when assessing residual insulin secretion. Most people with type 1 diabetes have measurable insulin secretion at diagnosis and this declines at a variable rate, tending to be a slower decline in those diagnosed as adults and in those with a higher C-peptide at diagnosis. One suggestion is that those with T1D should have a C-peptide measured after the 'honeymoon' period, when a negative C-peptide will confirm T1D. However if it is desirable to avoid a type 1 pathway (intensive insulin treatment, carb counting, etc.) for those who do not need it, then waiting months or years to make that assessment feels too long and other factors such as antibody status and clinical scenario will be the main determinants of such decisions.

Use of Novel Biomarkers and Other Tools for Diagnosing Monogenic Diabetes

Extrapancreatic effects of monogenic diabetes genes may result in measurable alterations in serum or urine constituents, which could then form the basis for a screening test for a specific subtype of diabetes. This has been explored most for *HNF1A*, which regulates many hepatic genes. One of the best potential candidates for clinical use is C-reactive protein (CRP), which was found to be low in individuals with HNF1A-MODY compared to other forms of diabetes [11]. The best discrimination, with a ROC curve C-statistic of >0.8, is from those with young-onset type 2 diabetes, partly because a higher CRP is seen in type 2 diabetes. Because HNF1A-MODY is uncommon compared to type 2 diabetes, using a low CRP of <0.5 mg/L as a screening test to select individuals for genetic testing for *HNF1A* would lead to sequencing of about 20 individuals with type 2 diabetes in order to diagnose one person with HNF1A-MODY.

To attempt to overcome the shortcomings of interpreting a number of clinical features and biomarkers, a MODY prediction model has been devised [12] which compares the clinical features of

the patient being considered to well-defined groups of MODY, type 1, and type 2 diabetes. This produces a probability that the patient has MODY, and the clinician can set their own acceptable probability threshold for requesting genetic testing (e.g. 20–30% likelihood of MODY). The probability calculator has some limitations in that it compares cases to a narrow test group and it is only currently recommended for use in those aged <35 years at diagnosis and in the white European population. Clinically useful biomarkers including antibodies, C-peptide, and CRP are not currently included in the model.

Genetic Testing in Diabetes

Diagnostic genetic testing for monogenic diabetes has been available in some countries since the early 2000s, allowing a definitive diagnosis of monogenic diabetes to be made. This is important to inform treatment decisions and to allow a proper evaluation of risk of diabetes in family members. The individual benefits of making a molecular diagnosis are easy to see and there is potential for cost saving if insulin can be discontinued in the common forms of MODY. To date the health economics of introducing widespread genetic testing in diabetes has not been proven although some modelling has been performed [13]. Availability and uptake of genetic testing remains rather variable in different parts of the world and even within the same country [14]. Educational initiatives such as the Genetic Diabetes Nurse network in the United Kingdom can help raise awareness of genetic testing and increase uptake [15].

Genetic testing for diabetes is now usually performed using targeted next generation sequencing for a panel of genes. The genes incorporated in the panel vary from lab to lab, but include at a minimum the common MODY genes. The panel may also include the common mitochondrial variant and lipodystrophy genes. All genes in the panel are sequenced simultaneously, rather than the older method of making a clinical diagnosis first and requesting the gene most likely to be responsible. Potentially a number of gene variants will be identified for each individual, and the challenge is to assess whether any of these variants may be responsible for the diabetes. Monogenic causes represent only a small proportion of diabetes diagnosed in adults and children after the neonatal period, so a careful assessment is required so that the correct advice can be given to patients. The American College of Medical Genetics and Genomics has produced guidelines to assist diagnostic laboratories and clinicians in the interpretation and reporting of sequence variants in suspected Mendelian disease [16] and it is important that any diagnostic service follows these recommendations. Even with genes such as *HNF1A*, where thousands of people have been sequenced and there is much clinical experience, it can be difficult to be confident that a variant is disease-causing. Exome and whole genome sequencing in populations have demonstrated that what may appear to be loss of function variants in genes can be associated with complete absence of clinical phenotype [17].

A polygenic genetic risk score for type 1 diabetes can also be calculated using combinations of known risk variants for type 1 [18]. The score captures the effect of the risk and protective HLA alleles which have traditionally been complex and expensive to measure using classic HLA typing. This can assist in determining likelihood of a type 1 diagnosis in cases where clinical features are ambiguous and antibodies are negative.

Who Should Be Seen in Secondary Care?

There is a body of evidence for misclassification of diabetes subtype in adults [19]. This could be due to assumptions based on age of onset, or because rarer subtypes of diabetes are not considered. Younger adults are more likely to be labelled as having type 1

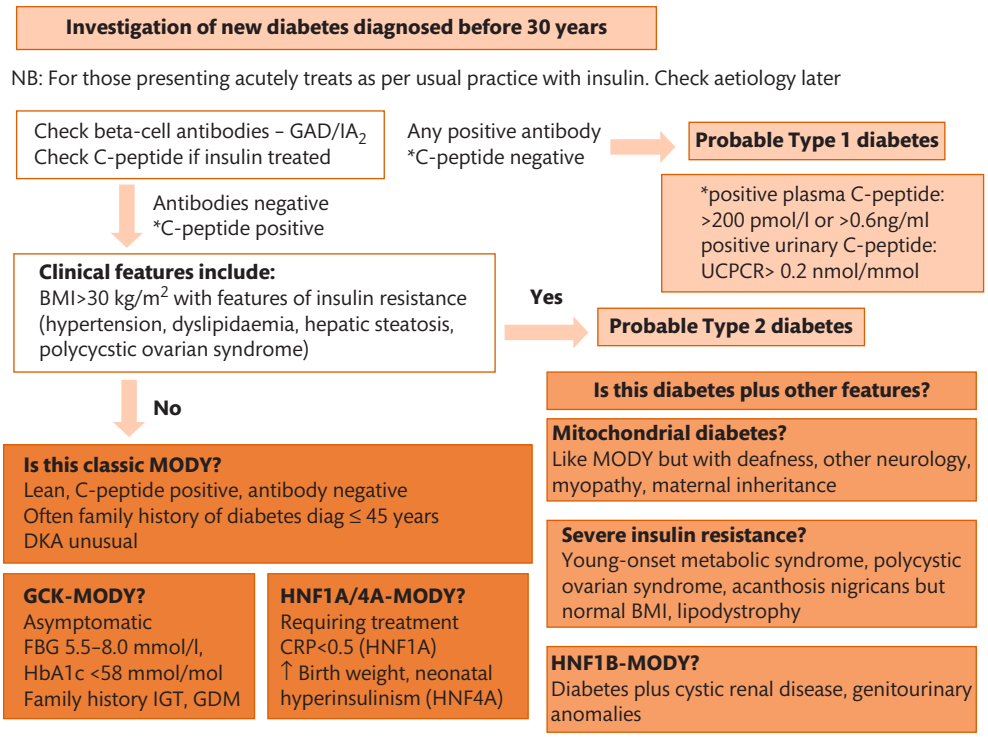


Figure 15.4.1.1 How to assess a newly diagnosed young adult.

diabetes and treated with insulin—this may be the safest option at diagnosis when uncertain, but should be revisited if there is any suggestion of an ambiguous clinical presentation or if beta-cell antibodies are negative. Conversely older adults may be routinely treated as type 2 diabetes and not have appropriate treatment escalation even if they respond poorly to oral hypoglycaemic agents. Therefore, there is much to be gained by developing pathways for assessing aetiology that can direct suitable individuals to specialist clinics—an example of such a pathway for newly diagnosed young adults is given in **Figure 15.4.1.1**.

Most UK care pathways direct type 1 diabetes into secondary care, and this is appropriate for the rare monogenic forms of diabetes too, with the exception of GCK-MODY where individuals can be discharged following genetic diagnosis. As young type 2 diabetes represents an equally high risk for poor outcomes as type 1 diagnosed at a similar age [20], one could argue that they would also benefit from intensive input.

Conclusions

Accurate assessment of aetiology is important for management of diabetes, both for choice of immediate pharmacological agent and for ensuring individuals with diabetes follow the most appropriate care pathway. Diagnosing non-type 1, non-type 2 diabetes can be challenging due to clinical overlap with the common forms of diabetes and relies on a variety of clinical skills and specialist investigation. It is important to understand the limitations of investigations as well as the most appropriate use. Monogenic diabetes may present with diabetes alone, or as part of a syndrome and family members will need to be counselled and investigated.

While at the moment stratified medicine approaches can be applied mainly to monogenic subtypes, there is much research aimed at understanding the clinical molecular features which define treatment response and complication development in polygenic type 2 diabetes. Assessment of aetiology will become both more complex and more commonplace in the future.

REFERENCES

1. de Lusignan S, Sadek N, Mulnier H, Tahir A, Russell-Jones D, Khunti K. Miscoding, misclassification and misdiagnosis of diabetes in primary care. *Diabet Med*, 2012; **29**: 181–9.
2. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. *Diabetes Care*, 2018. **41**(Suppl 1): S13–S27.
3. Johansson BB, Irgens HU, Molnes J, *et al.* Targeted next-generation sequencing reveals MODY in up to 6.5% of antibody-negative diabetes cases listed in the Norwegian Childhood Diabetes Registry. *Diabetologia*, 2017; **60**: 625–35.
4. Shields BM, Shepherd M, Hudson M, *et al.* Population-based assessment of a biomarker-based screening pathway to aid diagnosis of monogenic diabetes in young-onset patients. *Diabetes Care*, 2017; **40**: 1017–25.
5. De Franco E, Flanagan SE, Houghton JA, *et al.* The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet*, 2015; **386**: 957–63.
6. Murphy R, Turnbull DM, Walker M, Hattersley AT. Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A>G mitochondrial point mutation. *Diabet Med*, 2008; **25**: 383–99.
7. Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic syndromes of severe insulin resistance. *Endocr Rev*, 2011; **32**: 498–514.
8. Thanabalasingham G, Pal A, Selwood MP, *et al.* Systematic assessment of etiology in adults with a clinical diagnosis of young-onset type 2 diabetes is a successful strategy for identifying maturity-onset diabetes of the young. *Diabetes Care*, 2012; **35**: 1206–12.
9. Misra S. Pancreatic autoantibodies: who to test and how to interpret the results. *Practical Diabetes*, 2017; **34**: 221–3.
10. Jones AG, Hattersley AT. The clinical utility of C-peptide measurement in the care of patients with diabetes. *Diabet Med*, 2013; **30**: 803–17.
11. Thanabalasingham G, Shah N, Vaxillaire M, *et al.* A large multicentre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes. *Diabetologia*, 2011; **54**: 2801–10.
12. Shields BM, McDonald TJ, Ellard S, Campbell MJ, Hyde C, Hattersley AT. The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia*, 2012; **55**: 1265–72.
13. Naylor RN, John PM, Winn AN, *et al.* Cost-effectiveness of MODY genetic testing: translating genomic advances into practical health applications. *Diabetes Care*, 2014; **37**: 202–9.
14. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia*, 2010; **53**: 2504–8.
15. Shepherd M, Colclough K, Ellard S, Hattersley AT. Ten years of the national genetic diabetes nurse network: a model for the translation of genetic information into clinical care. *Clin Med (Lond)*, 2014; **14**: 117–21.
16. Richards S, Aziz N, Bale S, *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 2015; **17**: 405–24.
17. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 2016; **536**: 285–91.
18. Oram RA, Patel K, Hill A, *et al.* A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care*, 2016; **39**: 337–44.
19. Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: a systemic review. *Diabet Med*, 2010; **27**: 491–7.
20. Constantino MI, Molyneaux L, Limacher-Gisler F, *et al.* Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. *Diabetes Care*, 2013; **36**: 3863–9.

15.5

Principles of Management of Diabetes

15.5.1 Structured Education

Simon Heller and Jackie Elliott

Introduction

Diabetes mellitus is almost unique, even among chronic diseases, in that long-term outcomes are determined not by the knowledge and skill of the healthcare professional but by how successfully a patient can contribute to his/her own management. A crucial responsibility of their professional carers is to ensure that those with the condition are provided with knowledge and help to self-manage their condition effectively. There is now considerable evidence that acquisition of complex self-management skills requires high quality educational support and that this is best provided by attending structured educational courses. Thus, education provision is a fundamental component of the management of both type 1 and type 2 diabetes. In this chapter we describe the essential features of structured education and review the evidence for its effectiveness.

Background

The aim of treatment in both type 1 and type 2 diabetes is maintaining blood glucose levels close to normal both to prevent acute complications and in the long-term reduce micro- and macrovascular disease. In type 2 diabetes there is additional focus on controlling both blood pressure and lipids to reduce macrovascular events. Management frequently requires medication but there is strong evidence that major changes in lifestyle can achieve many of these aims without the need for medication. Significant weight loss achieved by a very low calorie diet or bariatric surgery can lead to complete and prolonged remission of diabetes while regular physical activity reduces macrovascular disease over the medium to long-term.

Despite the evidence for the effectiveness of lifestyle change, interventions specifically aimed at altering dietary habits or increasing physical activity often have only modest effects. The challenges faced by those affected are considerable. They include changing habits of a lifetime among individuals who have often been socialized into adopting an unhealthy lifestyle. Since these choices are one of the main reasons why individuals develop diabetes it is unsurprising that people find it so difficult people to alter their behaviour. It is increasingly realized that traditional approaches to diabetes management in which patients are encouraged to achieve healthcare outcomes framed from the healthcare professional prospective are less likely to be effective [1]. The recognition that regular 1:1 consultations with healthcare professionals who make the major decisions on management, particularly medication, are not conducive to successful treatment has prompted a move to a more patient-centred model of care. Diabetes associations in the United States, United Kingdom, and elsewhere now propose that newly diagnosed patients receive structured education from diagnosis and beyond. This has been reinforced in the United Kingdom by recent guidance from the National Institute for Health and Clinical Excellence (NICE) regarding structured education in both type 1 and type 2 diabetes as a fundamental component of treatment [2, 3].

How is Structured Education Defined?

A recent position statement from the United States has defined structured education as the process of facilitating the knowledge, skill, and ability necessary for diabetes self-care [4]. In 2003, the National Institute for Health and Clinical Excellence (NICE) proposed that structured education was 'a planned and graded programme that is comprehensive in scope, flexible in content, responsive to an individual's clinical and psychological needs, and adaptable to his or her educational and cultural background'[5].

The essential components of a structured education package have been expanded in recent NICE guidance [3].

These state that any structured education programme for adults with diabetes should include the following components:

- Outcomes are audited regularly.
- It is evidence-based, and suits the needs of the person.
- It has specific aims and learning objectives, supporting the person and their family members and carers in developing attitudes, beliefs, knowledge, and skills to self-manage diabetes.
- It has a structured curriculum that is theory-driven, evidence-based, and resource-effective, has supporting materials, and is written down.
- It is delivered by trained educators who have an understanding of educational theory appropriate to the age and needs of the person, and who are trained and competent to deliver the principles and content of the programme.
- It is quality assured, and reviewed by trained, competent, independent assessors who measure it against criteria that ensure consistency.

Structured Education Programmes Which Have Changed Practice

Type 1 Diabetes

The development of structured education in type 1 diabetes owes much to the vision of Berger and colleagues working in Düsseldorf. The basic principles of their insulin treatment and training programme (ITTP) included separation of basal and soluble insulin, given as pre-meal boluses based on pre-meal, blood glucose testing. Doses were calculated according to the carbohydrate content of the meal and the resulting dietary freedom allowed those participating to eat freely and flexibly, comparable to that of people without diabetes.

Their education course, provided as an in-patient, 5-day programme to groups of 6–8 adults, was evaluated in a series of clinical trials in the 1980s. In a large controlled trial involving 300 adults with type 1 diabetes, the course resulted in a 1.5% fall in HbA1c at 1 year (compared with control participants taught basic information alone) and was accompanied by a marked reduction in ketoacidosis [6].

A large observational study from the same group subsequently presented findings from a 6-year follow-up in a cohort of individuals with type 1 diabetes who had undergone the course [7]. They reported a reduction in HbA1c of 0.7%, while the incidence of severe hypoglycaemia had fallen from 0.28 episodes per person per year in the year at baseline to 0.17 episodes per person per year at follow-up (P <0.05). Observational studies provide a weaker level of evidence compared with clinical trials, but the large numbers and the relatively long follow-up indicate that structured education not only improves HbA1c but also reduces rates of severe hypoglycaemia by approximately 50%.

The success of these programmes has led to their adoption in other countries. In the United Kingdom, a multicentre randomized controlled trial (RCT), utilizing a waiting list design in three centres measured the impact of a structured education course (Dose Adjustment for Normal Eating [DAFNE]), based on the ITTP [8]. HbA1c improved by between 0.7 to 1% at 6 months, while no change was observed in the control group. Of equal importance was the marked improvement in quality of life, not included in previous evaluations. Flexible intensive insulin therapy demands more blood tests and injections but increased dietary freedom and the ability to adopt a more flexible lifestyle was highly valued by the participants. The approach has also been the subject of a health economics analysis, which concluded that even if HbA1c improvements were maintained only partially, the course would be highly cost-effective, due to the reduced incidence of microvascular complications [9].

Longer-term observational studies involving larger numbers of participants have demonstrated reductions in the incidence of severe hypoglycaemia that are comparable with those seen in the original ITTP studies from Germany, albeit at higher HbA1c levels. One reported reductions in HbA1c (mean difference pre- and post-DAFNE: 0.27%; P <0.001) and around a 50% reduction in rates of severe hypoglycaemia (mean ± SD: pre-DAFNE 1.7 ± 8.5 vs. post-DAFNE 0.6 ± 3.7 episodes per person per year; P <0.05), together with improved awareness of hypoglycaemia in up to 43% of participants at 1 year follow-up [10]. The overall rate of impaired awareness fell from 39.9% to 33%, with reductions in psychological distress and improved well-being up to 1 year following programme attendance [10]. A more recent study has also reported that DAFNE reduces the risk of diabetic ketoacidosis by 61% and combined emergency costs for DKA and severe hypoglycaemia are decreased by 64% [11].

DAFNE courses are now provided in more than 70 centres across the United Kingdom and Ireland (having been delivered to over 40 000 participants) and have been rolled out to other countries, including Australia, New Zealand, and Kuwait. In the United Kingdom, structured education is now recognized as a standard of care. Many centres who don't offer DAFNE training provide structured education in some form but, although some are based on the Düsseldorf course, the standard is inconsistent. Some centres provide shorter courses of a few days, despite evidence that these are less effective [12]. Few include formal training for educators, peer review, quality assurance, or regular audits as recommended by the Department of Health. This has been recognized by NHS England who have provided funding to try and ensure access to high quality structured education across the country.

Type 2 Diabetes

Structured education courses in type 1 diabetes primarily teach participants the skills of flexible intensive insulin therapy, focus on insulin delivery and maintaining glucose targets. In contrast, structured education in type 2 diabetes has a broader remit. Programmes need to address lifestyle issues of healthy eating, weight control, and the role of physical activity, particularly in those newly diagnosed. Individuals with more advanced type 2 diabetes require sufficient information to understand more advanced therapy, including the different medications used to lower glucose as well as those required to lower cholesterol, blood pressure, and reduce the risk of cardiovascular disease. A successful programme will also engage people in the management of their condition and help them understand how important their own contribution is to the success of treatment. This also explains why programmes often incorporate psychological approaches such as motivational interviewing or cognitive behavioural

therapy techniques. Indeed, the term 'psycho-educational' is often used to acknowledge the combined nature of these interventions.

In many respects this makes structured education programmes in type 2 diabetes more challenging to evaluate. In type 1 diabetes, appropriate outcomes include HbA1c, rates of severe hypoglycaemia, and DKA as well as psychosocial measures. While these are relevant in type 2 diabetes, it is necessary to include weight, lipids, blood pressure, and smoking. It is also important to understand the clinical context when interpreting the results. For example, a course designed for newly diagnosed individuals is unlikely to demonstrate major differences in HbA1c or glucose lowering medication whereas weight loss or levels of physical activity are more useful.

Other factors which add to the demands of comparing different interventions include different methods of delivery, varied contact time, group size, inclusion of technology, and healthcare setting. However, some generalizations can be made. A systematic review of trials in type 2 diabetes showed that on average, reductions in HbA1c of >0.4% compared to control groups were achieved, and improvements in glycaemic control were correlated with courses with delivery time of 11 hours or more [13]. Importantly in terms of resource there was no advantage in individual as opposed to group interventions. Other useful predictors of success appeared to relate to the demographics of the cohort. Younger participants under 65 years of age, and those of non-white ethnic backgrounds often appeared to gain greater benefit. Unsurprisingly those with higher baseline HbA1c exhibited greater improvement in glycaemic control. Measures of quality of life generally improved, but this was not universal.

The Diabetes Education and Self-Management for Ongoing and Newly Diagnosed (DESMOND) study was a randomized controlled trial of a group structured education intervention designed for adults with newly diagnosed diabetes [14]. The intervention consisted of 6 hours of education delivered in either one day, or two half-day sessions by two professional educators. Participants in both the intervention and the control arms of the study received routine care, with the intervention additionally receiving the DESMOND programme.

HbA1c fell in both intervention and control groups by around 1.5% with no significant difference in fall in HbA1c between the groups. Differences were observed in some secondary outcomes. Those in the intervention group lost more weight (3 kg vs. 1.8 kg), had lower levels of reported smoking, reduced depression scores, and reduced cardiovascular risk scores, at 12 months. The intervention group also demonstrated significant positive changes in their health beliefs about their diabetes, sustained at 12 months.

The lack of difference in HbA1c between the groups has been interpreted by some as indicating no additional benefit from DESMOND training. However, it is important that at diagnosis, HbA1c levels will fall in most individuals with type 2 diabetes. Somewhat disappointingly, a follow-up of the original cohort at 3 years, showed no difference in any biomedical outcome (including weight and HbA1c). The DESMOND intervention has been widely implemented in primary care in the United Kingdom, supported by an infrastructure which has provided training, quality assurance, and professional development.

The X-PERT study was conducted in a single centre primary care trust in the North West of England [15]. The 314 participants from a multiethnic population had a mean of duration of type 2 diabetes of 6.7 years. The intervention was delivered by single dietitian and consisted of six 2-hourly group sessions held once a week. The control group received routine care and additional individual education.

Significant reductions in HbA1c, total cholesterol, weight, and waist measurements were seen in the intervention group at 14 months compared to controls. In addition, the intervention group had significantly improved diabetes knowledge, treatment satisfaction, and feelings of empowerment. Intervention participants reported improved dietary habits and greater levels of physical activity. There were no differences in other clinical measures, nor in quality of life. A limitation of this study is that it was undertaken at a single centre. Nevertheless, the programme has been taken up in a number of UK centres.

The ROMEO trial also studied people with established type 2 diabetes and reported 4-year results in 2010 [16]. The intervention was partly structured education but an essential component was that the education was delivered within regular consultations to a group of patients. The trial was multicentre, delivered across Italy in 13 specialist clinics in which 815 participants received group education in sessions 40–50 minutes long, delivered every 3–4 months. The control group received standard routine care. At 4 years, the intervention group showed significantly improved HbA1c and other clinical outcomes, including blood pressure, lipids, weight, and body mass index (BMI) compared to controls. Quality of life and diabetes knowledge also improved. Despite the overall success, of the original 815 participants, only 592 completed the trial at 4 years and two of the participating centres were unsuccessful in implementing the education programme. This demonstrates the challenges inherent in integrating a successful intervention into routine care.

Technology in Structured Education

Technological innovations are making an increasingly important contribution to many aspects of care in diabetes and offer considerable potential to contribute to patient education. These range from continuous glucose monitoring to delivering education online. Innovations such as telehealth may provide education more effectively but there is a danger that the technology might merely be used to reduce costs. It is therefore important that technological development is accompanied by rigorous evaluation to measure its effectiveness.

Web-based strategies are potentially attractive in that they allow the user to access educational material at a time that suits them. They also increase flexibility by enabling users to learn at their own pace and decide how much to cover at each sitting. On the other hand, participants may lose the benefits of contextualized social learning, when participating individually rather than in a group. One systematic review examining online behaviour change interventions highlighted some benefits in type 2 diabetes [17] including improvements in both clinical and psychological outcomes. Relatively few behaviour change techniques were used and these were not always related to the theories underlying their intervention. A recent Cochrane review [18] assessed the effects on health status and health-related quality of life of computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus in a large number of clinical trials. Effects on HbA1c

were modest (falls of around 0.2% compared to controls) although mobile phone applications appeared to have a greater effect. This review found no other biomedical improvements or psychological or behavioural benefits.

The Need to Improve the Uptake of Structured Education

Despite increasing evidence confirming the benefit of structured education, uptake is still generally poor. In the United Kingdom, many people with diabetes are now offered structured education at some point (~50% in type 1 and ~90% in type 2 diabetes) but rates of uptake for adults with both types of diabetes remain below 10% [19]. Low uptake is also seen in many other countries, where self-reported rates of receiving any formal educational input were between 23% and 84% in one study [20].

A recent systematic review examined reasons cited by patients for not attending structured education across several countries and cultures [21]. It found that the barriers to uptake are sometimes patient-orientated, for example, location, timing, duration, loss of income if in working hours, an unawareness that education can improve outcomes, and an unwillingness to return to a classroom type experience. However, healthcare professionals are frequently responsible; a perceived lack of effectiveness preventing them from encouraging patients to attend. This factor was also highlighted in a recent study examining the reason for poor take-up of structured education in a population of young people with type 1 diabetes in South London [22].

Conclusion

Structured education is increasingly considered to be a crucial component of diabetes care with an essential role to play in enabling people with diabetes to initiate and sustain successful management of their own condition. This requires stakeholders to commission high quality programmes which meet national and international standards. Such a policy need not be expensive. Spending on diabetes is largely directed at trying to contain devastating, but largely preventable long-term complications. If a small fraction of this spending were directed at trying to prevent complications occurring by encouraging better self-management this would be highly cost-effective.

Nevertheless, more work still needs to be done. Healthcare professionals should explain to their patients and their families that structured education is a fundamental part of diabetes management and insist that they participate from diagnosis. We need to understand the factors which underlie an individual's decision to engage with their diabetes and the perceived barriers to putting knowledge into practice. We must work closely with experts in the area of behaviour change if we are to develop more effective interventions.

Developing and implementing structured education programmes is complex and challenging. Recent achievements and the positive feedback from participants, demonstrate the potential of these programmes in enabling people with diabetes to exercise control over their experience of living with diabetes, rather than being controlled by it.

REFERENCES

1. Wolpert HA, Anderson BJ. Management of diabetes: are doctors framing the benefits from the wrong perspective? *Brit Med J*, 2001; **323**: 994–6.
2. NICE. Type 1 diabetes in adults: diagnosis and management. 2015. Available at: https://www.nice.org.uk/guidance/ng17 (accessed: 21 April 2020).
3. NICE. Type 2 diabetes in adults: management. 2015. Available at: https://www.nice.org.uk/guidance/ng28 (accessed: 21 April 2020).
4. Powers MA, Bardsley J, Cypress M, *et al.* Diabetes self-management education and support in type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. *Clin Diabetes*, 2016; **34**: 70–80.
5. NICE. Guidance on the use of patient-education models for diabetes. *Technology Appraisals (TA60)* 2003. Available at: https://www.nice.org.uk/guidance/ta60 (accessed: 21 April 2020).
6. Muhlhauser I, Jorgens V, Berger M, *et al.* Bicentric evaluation of a teaching and treatment programme for type 1 (insulin-dependent) diabetic patients: improvement of metabolic control and other measures of diabetes care for up to 22 months. *Diabetologia*, 1983; **25**: 470–6.
7. Samann A, Muhlhauser I, Bender R, *et al.* Glycaemic control and severe hypoglycaemia following training in flexible, intensive insulin therapy to enable dietary freedom in people with type 1 diabetes: a prospective implementation study. *Diabetologia*, 2005; **48**: 1965–70.
8. DAFNE Study Group. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: Dose Adjustment for Normal Eating (DAFNE) randomised controlled trial. *BMJ*, 2002; **325**: 746–51.
9. Kruger J, Brennan A, Thokala P, *et al.* The cost-effectiveness of the Dose Adjustment for Normal Eating (DAFNE) structured education programme: an update using the Sheffield Type 1 Diabetes Policy Model. *Diabet Med*, 2013; **30**: 1236–44.
10. Hopkins D, Lawrence I, Mansell P, *et al.* Improved biomedical and psychological outcomes 1 year after structured education in flexible insulin therapy for people with type 1 diabetes: the U.K. DAFNE experience. *Diabetes Care*, 2012; **35**: 1638–42.
11. Elliott J, Jacques RM, Kruger J, *et al.* Substantial reductions in the number of diabetic ketoacidosis and severe hypoglycaemia episodes requiring emergency treatment lead to reduced costs after structured education in adults with type 1 diabetes. *Diabet Med*, 2014; **31**: 847–53.
12. George JT, Valdovinos AP, Russell I, *et al.* Clinical effectiveness of a brief educational intervention in Type 1 diabetes: results from the BITES (Brief Intervention in Type 1 diabetes, Education for Self-efficacy) trial. *Diabet Med*, 2008; **25**: 1447–53.
13. Pillay J, Armstrong MJ, Butalia S, *et al.* Behavioral programs for type 1 diabetes mellitus: a systematic review and meta-analysis. *Ann Intern Med*, 2015; **163**: 836–47.
14. Davies MJ, Heller S, Skinner TC, *et al.* Effectiveness of the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes: cluster randomised controlled trial. *Brit Med J*, 2008; **336**: 491–5.
15. Deakin TA, Cade JE, Williams R, *et al.* Structured patient education: the diabetes X-PERT Programme makes a difference. *Diabet Med*, 2006; **23**: 944–54.

16. Trento M, Gamba S, Gentile L, *et al.* Rethink Organization to iMprove Education and Outcomes (ROMEO): a multicenter randomized trial of lifestyle intervention by group care to manage type 2 diabetes. *Diabetes Care*, 2010; **33**: 745–7.
17. van Vugt M, de Wit M, Cleijne WH, *et al.* Use of behavioral change techniques in web-based self-management programs for type 2 diabetes patients: systematic review. *J Med Internet Res*, 2013; **15**: e279.
18. Pal K, Eastwood SV, Michie S, *et al.* Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. *Cochrane Database Syst Rev*, 2013; **3**: CD008776.
19. HSCIC. *Health and Social Care Information Centre: National Diabetes Audit, 2016–17 Report 1: Care Processes and Treatment Targets*. 2018.
20. Nicolucci A, Kovacs Burns K, *et al.* Diabetes Attitudes, Wishes and Needs second study (DAWN2): cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes. *Diabet Med*, 2013; **30**: 767–77.
21. Horigan G, Davies M, Findlay-White F, *et al.* Reasons why patients referred to diabetes education programmes choose not to attend: a systematic review. *Diabet Med*, 2017; **34**: 14–26.
22. Harris SM, Joyce H, Miller A, *et al.* The attitude of healthcare professionals plays an important role in the uptake of diabetes self-management education: analysis of the Barriers to Uptake of Type 1 Diabetes Education (BUD1E) study survey. *Diabet Med*, 2018. doi: 10.1111/dme.13704.

15.5.2 Glucose Monitoring and Sensing

John Pickup and Nick Oliver

Introduction

Blood glucose concentrations are measured to identify hyper- and hypoglycaemia. Healthcare teams use glucose data to diagnose diabetes, or impaired glucose regulation, adjust therapies, and to inform hypo- and hyperglycaemia correction. People with diabetes self-monitor blood glucose concentrations to inform decision making; enabling food, insulin, and activity to be titrated according to glucose levels, empowering people to achieve optimal glucose concentrations. Self-monitoring of blood glucose (SMBG) also supports activities such as driving and can provide reassurance to the person with diabetes, and their family and carers, especially for children and frail people.

Glucose monitoring has traditionally been performed by intermittent sampling of capillary blood glucose concentrations using finger-prick samples applied to reagent strips inserted into portable glucose meters. Since around 2000, continuous glucose monitoring (CGM) has entered clinical practice as a supplement to SMBG, and more recently flash glucose monitoring (FGM) has been available. CGM and FGM are based on the implantation of needle-type sensors into the subcutaneous tissue for measurement of interstitial glucose concentrations which are then calibrated to an estimate of blood glucose, either by SMBG or factory calibration.

Self-Monitoring of Blood Glucose

The History and Technology of SMBG

The first point of care blood glucose testing system was introduced in 1965; requiring application of blood to a strip containing immobilized glucose oxidase, peroxidase, and a dye. A colour change in the dye was compared semi-quantitatively with a colour chart to estimate the glucose concentration [1]. Automated and more quantitative reading of the strips using a portable reflectance meter was introduced in the early 1970s. SMBG was introduced into clinical practice in 1978 [2, 3], and has become an integral part of the self-management of type 1 diabetes.

Most SMBG technologies are now based on the electrochemical detection of glucose and a current response, which is converted into a digital read-out of the glucose concentration [4]. A small molecular weight mediator (e.g. ferrocene, hexacyanoferrate, or a quinone) is used to shuttle electrons from glucose to an electrode, thereby producing a current, catalysed by an enzyme such as glucose oxidase or glucose dehydrogenase. Very small volumes of blood are now required, typically less than 1 μl, with the blood taken up into the strip by capillary forces. A reading can usually be obtained in around 5 seconds.

Modern meters are equipped with a solid-state memory and are able to store test results for later display, and reporting simple statistics, such as mean glucose over preceding periods and at defined times of day. Data can be downloaded to a personal computer or in some cases meters can connect to a smartphone for display of data and more complete statistical analysis, including blood glucose mean, modal day display, standard deviation, and percentage of values within, over, or under target.

The Evidence Base for SMBG

The popularity of SMBG accelerated following the Diabetes Control and Complications Trial (DCCT) [5], where SMBG was used as a component of intensified therapy by both the multiple dose injection (MDI) and continuous subcutaneous insulin infusion (CSII) arms. The study evidence for the effectiveness of SMBG, however, is far from conclusive, and its best use is not always agreed by healthcare professionals [6].

In type 2 diabetes, the majority of trials indicate either marginal or no effectiveness of SMBG [7, 8] over no monitoring, although trials are difficult to interpret. Studies have failed to differentiate between types of oral agent; many have low statistical power; many did not assess hypoglycaemia, a major reason for

testing; and many did not give instructions on SMBG frequency or interpretation.

The reasons for, and likely benefits of, SMBG in type 1 diabetes seem more obvious than in type 2 diabetes: hypoglycaemia is more common and glucose more variable so frequent monitoring will allow adjustments in insulin dosage, avoidance of hypoglycaemia, and greater time in target. Surprisingly, the evidence for its effectiveness from formal randomized studies is weak, mostly because of poor trial design and reporting. Cross-sectional studies generally favour lower HbA_{1c} levels in those testing frequently in type 1 diabetes [9, 10] with a high negative correlation in type 1 diabetes between strip prescription and HbA_{1c}.

A Rational Approach to SMBG in Type 1 and 2 Diabetes

With conflicting evidence for effectiveness, how should SMBG be supported for people with diabetes? The major determinants of blood glucose testing frequency are the lability of the blood glucose concentration; the impact of the data on decision making, and personal preference [11]. With greater within-, and between-day glucose variability, as seen with absolute insulin deficiency, more glucose values are required to reveal glucose changes [12]. In type 2 diabetes, glucose is more stable and, although concentrations are elevated, may be more predictable. A single blood glucose measurement in type 2 diabetes, therefore, relates more closely to overall control and correlates well with HbA_{1c} [13].

The treatment of people with diabetes also determines the frequency of SMBG. Those with type 2 diabetes managed with diet alone, metformin, glitazones, or dipeptidyl peptidase-IV inhibitors have little or no hypoglycaemia; while those with insulin or sulphonylurea-treated diabetes are more exposed to hypoglycaemia.

In type 1 diabetes, SMBG is usually advised before and after meals, and at bedtime, with extra tests at certain times. National Institute for Health and Care Excellence (NICE) guidelines for self-management of type 1 diabetes support SMBG at least four times a day, and up to ten times a day if the desired target for blood glucose control is not achieved; if the frequency of hypoglycaemic episodes increases; if there is a legal requirement to do so (such as before driving); during periods of illness; before, during, and after sport and when planning pregnancy, during pregnancy and while breastfeeding [14].

The NICE guidance for self-management of type 2 diabetes recommends SMBG for adults with type 2 diabetes if the person is on insulin; there is evidence of hypoglycaemic episodes; the person is on oral medication that may increase their risk of hypoglycaemia while driving or operating machinery; or the person is pregnant, or is planning to become pregnant [15].

There is increasing interest in not only the mean blood glucose levels, but also in glycaemic variability and its clinical importance. While there is continued debate and conflicting evidence on whether glycaemic variability is a risk factor for complications in diabetes [16], the clearest consequence of excessive variability is that it restricts the HbA_{1c} level which can be achieved. The advent of continuous data from CGM has added further measures of glycaemic control such as 'time-in-range' (per cent glucose readings or hours per day between 3.9 and 10 mmol/L [70–180 mg/dl]) but the best indices have not been agreed [17].

CGM

The Technology and Performance of CGM

Most CGM used in clinical practice is based on needle-type electrodes [18] that are inserted subcutaneously. It is therefore minimally invasive. These amperometric (current-measuring) enzyme electrodes consist of glucose oxidase immobilized at a charged base electrode; oxidation of glucose to gluconic acid and hydrogen peroxide is monitored by the electrochemical detection of the hydrogen peroxide (Medtronic, Dexcom). One glucose sensor on the market (FreeStyle Navigator [CGM] and FreeStyle Libre [FGM], Abbott) uses a modification of this system, whereby an osmium mediator covalently bound to a polymer matrix 'wires' the enzyme to the electrode.

Glucose sensors are implanted in the subcutaneous tissue and record interstitial fluid glucose concentrations every 1–5 min over an operating period of 7–10 days. A transmitter attached to the sensor relays the data wirelessly to a storage and display device, either a portable meter, a modified insulin pump, or a smartphone. Most sensors require calibration using capillary blood glucose samples but calibration-free CGM is now available. CGM data can be either downloaded to a computer and reviewed by the multidisciplinary team and user retrospectively in order to identify patterns and aid treatment changes ('professional CGM'), or can be viewed in real-time with alerts and alarms for impending and established high and low blood glucose ('personal CGM'). Trend arrows give an indication of the direction and rate of change of blood glucose levels.

Recently a novel type of CGM technology has entered clinical practice based on a fully subcutaneously implanted capsule containing a non-enzymatic glucose-indicating fluorescence hydrogel polymer with a transmitter worn over the sensor site and transmission to a hand-held display device (Eversense, Senseonics). This technology has been successfully tested for up to 180 days' implantation, but has not yet received extensive clinical evaluation [19].

The Evidence Base for the Effectiveness of Real-Time CGM

CGM has the potential to improve overall glycaemic control (e.g. measured by HbA_{1c} and time-in-range), reduce exposure to hypoglycaemia, and positively impact on fear of hypoglycaemia. A landmark Juvenile Diabetes Research Foundation International (JDRF) study in 2008 demonstrated that CGM, when used for 6 days out of 7, reduces HbA_{1c} in children, young people and adults with type 1 diabetes compared with SMBG [20]. A subsequent meta-analysis of this and other randomized controlled trials of CGM vs. SMBG in people with type 1 diabetes treated by CSII or MDI reinforced this finding and showed that the HbA_{1c} reduction is dependent on both the frequency of sensor use and the baseline HbA_{1c} [21]. A reduction in median exposure to hypoglycaemia of 23% (CGM vs. SMBG) was also shown.

More recently studies have assessed the impact of CGM in populations most likely to use CGM in clinical practice: those with high baseline HbA_{1c} values or disabling hypoglycaemia during SMBG, as well as those using MDI regimens rather than CSII. These studies have reproducibly shown a mean reduction in HbA_{1c} of around 0.6% (6 mmol/mol) and clinically, and statistically, significant reductions in exposure to hypoglycaemia at all thresholds compared with SMBG [22, 23]. High risk populations have also been studied

with reductions in mild-to-moderate and severe hypoglycaemia (requiring the assistance of a third party) demonstrated compared to SMBG in people with type 1 diabetes and impaired awareness of hypoglycaemia or a recent history of severe hypoglycaemia [24]. Similar findings have been reported for CGM compared with FGM, suggesting that CGM is the preferred monitoring modality for people at high risk of hypoglycaemia [25]. In studies of people at highest risk of hypoglycaemia, hypoglycaemia fear is also reduced with CGM compared to SMBG and flash monitoring. Real-world data collected in a reimbursed health system suggest that CGM in people with type 1 diabetes managed in specialist centres improves HbA_{1c}, lessens fear of hypoglycaemia, and improves quality of life, while work absenteeism and admissions for acute diabetes complications, including severe hypoglycaemia, decreased [26].

Though the use of CGM in type 2 diabetes has been little investigated there is emerging RCT evidence that those using MDI and CGM have lowered HbA_{1c} compared to MDI and SMBG [27]

Clinical Recommendations for CGM

The NICE guidance for adults with type 1 diabetes supports the use of CGM devices where people have severe hypoglycaemia, unawareness of hypoglycaemia, frequent asymptomatic hypoglycaemia, extreme fear of hypoglycaemia, or where SMBG has failed, despite an intensified insulin regimen. CGM is supported by NICE for children with type 1 diabetes who are very young, unable to communicate the symptoms of hypoglycaemia, or who are at high risk of hypoglycaemia due to physical activity [14].

Flash Glucose Monitoring (FGM)

FGM or intermittently scanned CGM (isCGM) (FreeStyle Libre, Abbott) involves a subcutaneously implanted, factory-calibrated glucose sensor with up to 14 days lifetime. It provides a contemporaneous estimation of blood glucose, a trend arrow, and a graph of the preceding 8 hours' glycaemic data when a reader or smartphone is swiped ('flashed') over the sensor. There are no alarms on the original version but the recently introduced Libre 2 has customisable alarms for low and high blood glucose and is being selectively launched in several countries. It is less expensive than CGM and potentially enables users to reduce or avoid capillary SMBG. There is a limited evidence base for FGM, with few randomized controlled trials in type 1 and type 2 diabetes. First results have shown a reduction in time spent below 3.9 mmol/L (70 mg/dl) compared with SMBG, but no impact on HbA_{1c} [28, 29]. isCGM may therefore be best implemented as a replacement for SMBG where driving regulations allow, and where the budget impact is neutral or better.

Conclusion

Blood glucose monitoring in diabetes is in an exciting phase, undergoing a transition from SMBG to CGM and flash (isCGM). CGM provides information about the direction of change in glycaemic control and allows alarms when hypo- and hyperglycaemic thresholds are exceeded. Coupling sensor information to an insulin pump via suitable algorithms allows closed-loop insulin delivery—an artificial pancreas. Initially this was achieved via automatic threshold or predictive low glucose suspension of the basal insulin infusion rate of an insulin pump, and such devices are effective at reducing hypoglycaemia and are in routine clinical use [30]. Moment-to-moment, fully closed-loop systems remain a challenge but hybrid closed-loop devices (feedback control in the basal state but patient-activated meal insulin) are entering clinical practice.

REFERENCES

1. Owens DR. History and vision: what is important for patients with diabetes? *Diabet Technol Therapeut*, 2008; **10** (Suppl 1): S5–9.
2. Sönksen PH, Judd SL, Lowy C. Home monitoring of blood glucose. Method for improving control. *Lancet*, 1978; **8**: 729–32.
3. Walford S, Gale EA, Allison SP, Tattersall RB. Self monitoring of blood glucose. Improvement of diabetic control. *Lancet*, 1978; **8**: 732–5.
4. Hönes J, Műller P, Surridge N. The technology behind glucose meters: test strips. *Diabet Technol Therapeut*, 2008; **10**(Suppl 1): S10–26.
5. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long term complications of insulin dependent diabetes mellitus. *N Engl J Med*, 1993; **329**: 977–86.
6. O'Kane MJ, Pickup JC. Self monitoring of blood glucose in diabetes: is it worth it? *Annal Clin Biochem*, 2009; **46**: 273–82.
7. Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R. Self monitoring in type 2 diabetes mellitus: a meta-analysis. *Diabet Med*, 2000; **17**: 755–61
8. Welschen LMC, Bloemendal E, Nijpels G, *et al*. Self-monitoring of blood glucose in patients with type 2 diabetes who are not using insulin. *Diabetes Care*, 2005; **28**: 1510–17.
9. Karter AJ, Parker MM, Moffett HH, *et al*. Longitudinal study of new and prevalent use of self-monitoring of blood glucose. *Diabetes Care*, 2006; **29**: 1757–63.
10. Evans JMM, Ruta DA, MacDonald TM, Stevenson RJ, Morris AD. Frequency of blood glucose monitoring in relation to glycaemic control: observational study with diabetes database. *BMJ*, 1999; **319**: 83–6.
11. Owens D, Barnett AH, Pickup JC, *et al*. Blood glucose self-monitoring in type 1 and type 2 diabetes: reaching a multidisciplinary consensus. *Diabetes Primary Care*, 2004; **6**: 8–16.
12. Pickup JC. Diabetic control and its measurement. In: Pickup JC, Williams G, eds. *Textbook of Diabetes*. Oxford: Blackwell Science, 2003: 34.1–34.17.
13. McCance DR, Ritchie CM, Kennedy L. Is HbA1 measurement superfluous in NIDDM? *Diabetes Care*, 1988; **11**: 512–14.
14. NICE. Guideline NG17: Type 1 diabetes in adults: diagnosis and management. Available at: https://www.nice.org.uk/guidance/ng17 (accessed: 15 July 2018).
15. NICE Guideline NG28: Type 2 diabetes in adults: management. Available at: https://www.nice.org.uk/guidance/ng28 (accessed: 15 July 2018).
16. Service FJ. Glucose variability. *Diabetes*, 2013; **62**: 1398–404.
17. Danne T, Nimri R, Battelino T, *et al.* International consensus on use of continuous glucose monitoring. *Diabetes Care*, 2017; **40**: 1631–40
18. Mastrototaro J. The MiniMed Continuous Glucose Monitoring System [CGMS]. *J Paediatr Endocrinol Metab*, 1992; **12**(Suppl 3): 751–8.
19. Kropff J, Choudhary P, Neupane S, *et al.*: Accuracy and longevity of an implantable continuous glucose sensor in the PRECISE study: a 180-day, prospective, multicenter, pivotal trial. *Diabetes Care*, 2017; **40**: 63–8.

20. Tamborlane WV, Beck RW, Bode BW, *et al.* Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med*, 2008; **359**: 1464–76.
21. Pickup, JC, Freeman S C., Sutton, A J. Glycaemic control in type 1 diabetes during real time continuous glucose monitoring compared with self monitoring of blood glucose: meta-analysis of randomised controlled trials using individual patient data. *BMJ*, 2011; **343**: d3805.
22. Lind M, Polonsky W, Hirsch IB, *et al.* Continuous glucose monitoring vs conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections. The GOLD Randomized Clinical Trial. *JAMA*, 2017; **317**: 379–87.
23. Beck RW, Riddlesworth T, Ruedy K, *et al.* Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections. The DIAMOND Randomized Clinical Trial. *JAMA*, 2017; **317**: 371–8.
24. Heinemann L, Freckmann G, Ehrmann D, *et al.* Real-time continuous glucose monitoring in adults with type 1 diabetes and impaired hypoglycaemia awareness or severe hypoglycaemia treated with multiple daily insulin injections (HypoDE): a multicentre, randomised controlled trial. *Lancet*, 2018; **6736**: 1–11.
25. Reddy M, Jugnee N, El Laboudi A, Spanudakis E, Anantharaja S, Oliver N. A randomized controlled pilot study of continuous glucose monitoring and flash glucose monitoring in people with Type 1 diabetes and impaired awareness of hypoglycaemia. *Diabet Med*, 2018; **35**: 483–90.
26. Charleer S, Mathieu C, Nobels F, *et al.* Effect of continuous glucose monitoring on glycemic control, acute admissions, and quality of life: a real-world study. *J Clin Endocr Metab*, 2018; **103**: 1224–32.
27. Beck RW, Riddlesworth TD, Ruedy K, *et al.* Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections. A randomized trial. *Ann Intern Med*, 2017; **167**: 365–74.
28. Bolinder J, Antuna R, Geelhoed-Duijvestijn P, Kröger J, Weitgasser R. Novel glucose-sensing technology and hypoglycaemia in type 1 diabetes: a multicentre, non-masked, randomised controlled trial. *Lancet*, 2016; **388**: 2254–63.
29. Haak T, Hanaire H, Ajjan R, Hermanns N, Riveline JP, Rayman G. Flash glucose-sensing technology as a replacement for blood glucose monitoring for the management of insulin-treated type 2 diabetes: a multicenter, open-label randomized controlled trial. *Diabetes Ther*, 2017; **8**: 55–73.
30. Battelino T, Nimri R, Dovc K, Phillip M, Bratina N. Prevention of hypoglycemia with predictive low glucose insulin suspension in children with type 1 diabetes: a randomized controlled trial. *Diabetes Care*, 2017; **40**: 764–70.

15.5.3 Insulins and Insulin Delivery Devices

Pratik Choudhary and Peter Jacob

Introduction

Insulin is the life-saving treatment in type 1 diabetes, and increasingly used to treat advanced type 2 diabetes. Over the years there have been improvements and alterations in insulin treatment, moving from animal-derived insulin, through recombinant human insulin to genetically modified analogue insulins to help support people with diabetes to achieve better glucose control. Improved insulin delivery through insulin pens and more complex devices such as insulin pumps have helped improve quality of life and biomedical outcomes such as lower HbA1c and reduced hypoglycaemia. In this chapter we will describe currently available insulins with a focus on clinical trials that demonstrate differences relevant to individual users. We go on to discuss the wide variety of devices used to administer these insulins together with a discussion of their relative advantages and disadvantages.

Insulin

The insulin molecule is produced from post-translational modification of pre-proinsulin and then proinsulin. The final active hormone is a peptide comprising two amino acid chains linked by bi-sulphide bonds (Figure 15.5.3.1). In solution, insulin aggregates into hexamers around a zinc core. When injected subcutaneously, these hexamers dissociate to form monomers which can then be taken up into capillaries and enter the blood stream to exert their physiological action.

In health, insulin is secreted in a glucose-dependent manner into the portal circulation. After first pass metabolism, where more than 50% of insulin is metabolized by the liver, it acts on other organs, predominantly adipose tissue, and skeletal muscle. Ultimately, remaining insulin is cleared via kidneys [1]. Approximately 50% of insulin secretion is in response to meals (prandial insulin) and 50% is 'basal' insulin secreted at a background level even when the person is not eating [2].

Development of Insulin

The first insulins were prepared from animal pancreases and were typically either bovine or porcine. However, these only lasted a few hours and often patients needed frequent injections every day. The first modification was development of protamine or zinc-treated insulins [3]. This stabilized the insulin hexamers and prolonged the action of insulin. Human insulin was the first protein to be fully synthesized using recombinant DNA technology and this led to the development of recombinant human insulin generated within genetically modified *E. coli* [4]. These insulins were introduced in the 1980s. More recently, alterations to the amino acid sequence or addition of other molecules have

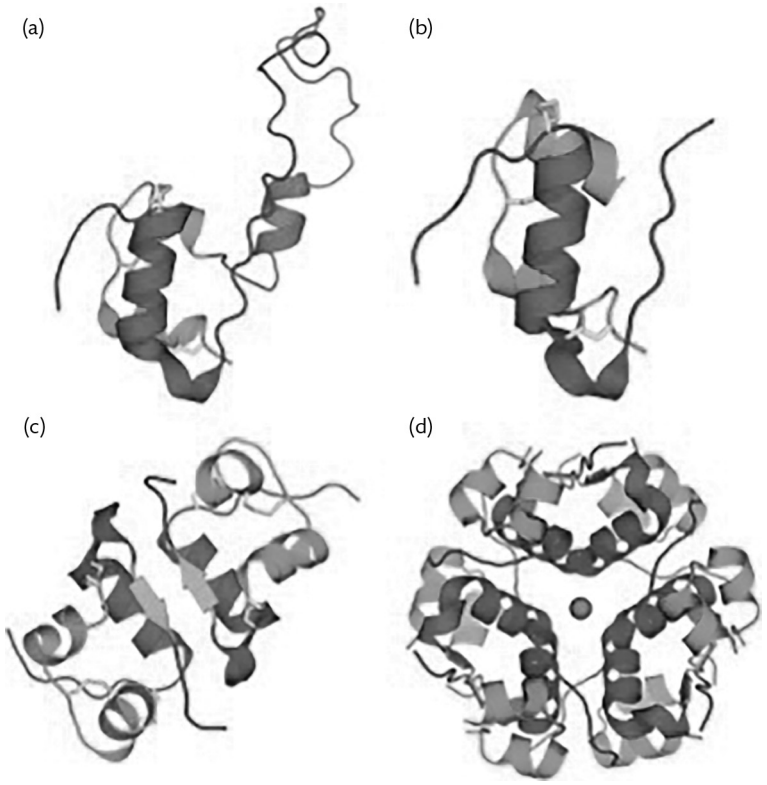


Figure 15.5.3.1 Structure of insulin. (a) Proinsulin. (b) Insulin monomer. (c) Insulin dimer. (d) Insulin hexamer.
Reproduced with permission from Hilgenfeld R, Seipke G, Berchtold H *et al*. The evolution of insulin glargine and its continuing contribution to diabetes care. *Drugs* 2014;74:911–27. Copyright © 2014, The Authors [54].

been used to create rapid-acting or long-acting insulin analogues which allow the individual more flexibility with the treatment regimen (Figure 15.5.3.2).

Today, we consider insulins in two main groups. The soluble or rapid-acting insulins are designed to cover the rapid rise in insulin requirements after ingestion of carbohydrates, or to correct acutely raised blood glucose levels, and are called meal-time or bolus insulins. Other longer acting insulins provide background or 'basal' insulin replacement to balance endogenous glucose production towards maintenance of stable blood glucose levels stable in the fasted or basal state.

Key goals of insulin development have been to increase the speed of onset of 'rapid-acting insulin' to better cover meal-time requirements; and to slow down and attain a longer and flatter profile for 'basal' insulins (Figure 15.5.3.3). Further development has been towards delivering the insulin via the inhaled or oral routes. On the horizon are numerous other insulins which have faster onset of action or the ability to reduce or switch off action in response

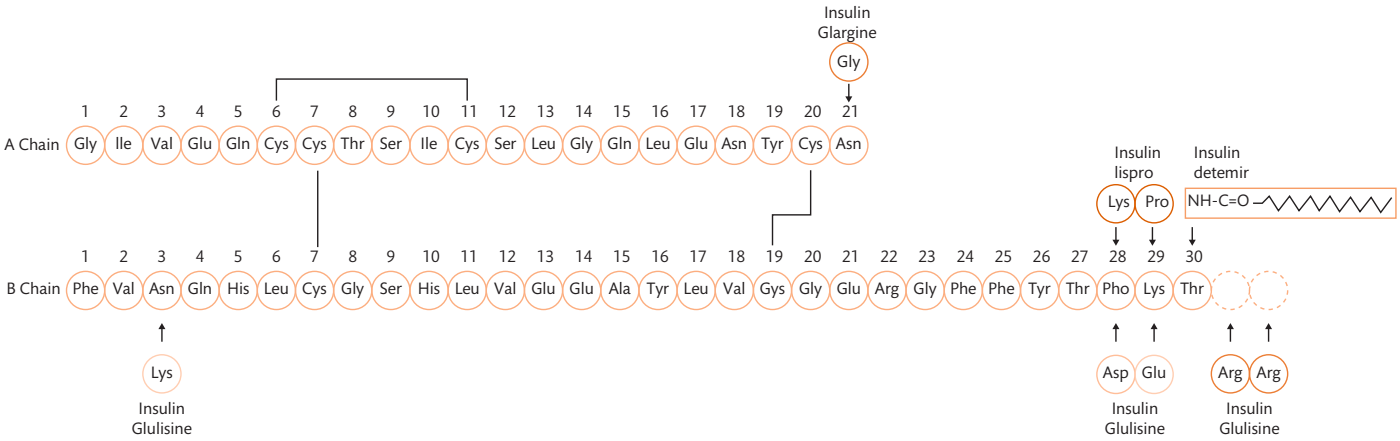


Figure 15.5.3.2 Amino acid structure of insulin. Insulin analogues are created by substitution of amino acids, or insertion of a new group on either the A or B chain at specific sites. Aspart–aspartate replaces B28 proline. Glulisine–lysine replaces B3 asparagine and glutamate replaces B29 lysine. Lispro–lysine replaces B28 proline and proline replaces B29 lysine. Detemir–myristic acid moiety replaces B30 threonine binding directly to B29 lysine. Glargine–glycine replaces A21 and two arginines bound to B30 threonine.

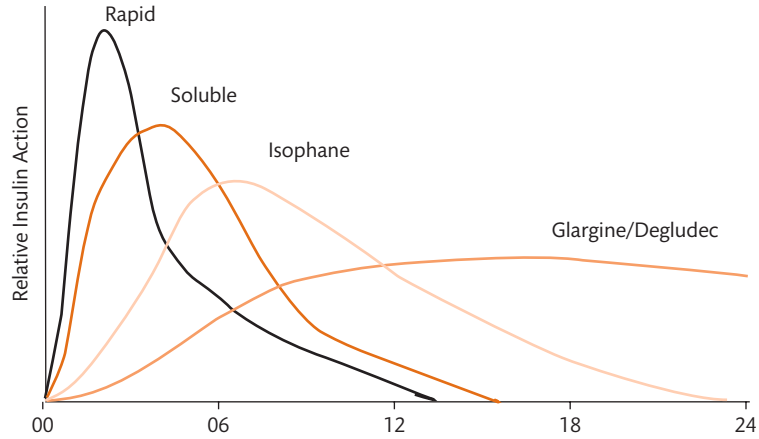


Figure 15.5.3.3 Relative insulin activity following a single subcutaneous injection of each insulin.

to low blood glucose (hypoglycaemia). A very small number of patients continue treatment with bovine or porcine insulins, but these treatments have been superseded and will not be discussed further here.

Prandial Insulins

Soluble human insulin is a synthetic insulin manufactured using recombinant DNA technology and has the same amino acid structure as native human insulin. It aggregates into hexamers at higher concentrations and thus when injected subcutaneously has a delayed onset of 30–60 minutes, peak effect after 2–4 hours and prolonged duration of action (6–8 hours) (Table 15.5.3.1). Evidence suggests that soluble insulin should be injected up to 30 minutes before eating to control postprandial excursions and reduce the risk of late hypoglycaemia [5]. Its long duration of action continues to predispose to a risk of late hypoglycaemia with between meal carbohydrate-containing snacks often required. There is significant interdose variability in absorption and duration of action may be prolonged with larger injection volumes impacting on reliability.

Table 15.5.3.1 Comparison of insulin action profiles for commonly used insulins

Insulin	Time to onset	Peak activity	Duration of action
Prandial insulin			
Soluble human insulin	30–60 min	2–3 min	6–8 hr
Rapid-acting analogues	15 min	30–90 min	4–6 hr
Faster aspart	5–15 min	30–90 min	3–5 hr
Basal Insulin			
Isophane	2–4 hr	4–10 hr	12–14 hr
Detemir	1–2 hr	6–12 hr	Up to 24 hr
Glargine	2–4 hr	Peakless	Up to 24 hr
Degludec	2–4 hr	Peakless	Up to 42 hr

Rapid-acting insulin analogues were designed to minimize the need for early preprandial injections and to reduce the risk of hypoglycaemia late after the meal. Insulins lispro, aspart and glulisine were created by substituting amino acids at key residues at the end of the insulin B chain to rapidly reverse hexamerization following injection. With lispro, the amino acids proline and lysine at positions B28 and B29 of the insulin β-chain are reversed. To create aspart the B28 proline is exchanged for aspartic acid. Finally, with glulisine, the asparagine at B3 is altered to lysine and the B29 lysine replaced with glutamic acid. The resulting insulins all have similar pharmacokinetic profiles with onset time of 5–15 minutes and a duration of action of between 4 and 6 hours [6]. Crucially, the peak insulin concentration and the peak insulin effect may still be 40–60 minutes or 120 minutes post-dose, respectively [5]. With these data, patients should still be advised to inject 15–20 minutes before a meal when it is safe to do so. Despite clear advantages for the user, the published data suggest only modest HbA1c improvements of –0.1% compared to soluble insulin [7]. Reductions in total number of hypoglycaemic episodes, number of severe hypoglycaemia events and nocturnal hypoglycaemia were seen in the analogue group but due to differences in trial design clear comparative analysis was not possible. A more recent RCT looked specifically at hypoglycaemia rates in high risk patients with more than two severe hypoglycaemic events per year. This found a significant reduction in the rate of hypoglycaemia when using rapid-acting insulin analogues in comparison to soluble insulin [8]. Benefit of rapid-acting analogue insulins in T2D is less clear [7].

There are a some 'ultra-fast' acting insulins in development, with faster acting insulin aspart (FiAsp®) now clinically available. In this preparation, nicotinamide (Vitamin B3) has been added, intended to enhance absorption by encouraging insulin monomerization, together with L-arginine which is purported to enhance insulin stability. The aspart amino acid structure is unchanged, so that once in the bloodstream the effect is identical to regular aspart. These modifications shift the insulin action curve 5–10 minutes to the left, providing greater insulin action within the first 30 minutes [9]. In clinical trials, better postprandial control was associated with lower overall HbA1c and similar hypoglycaemia rates in type 1 diabetes [10]. Although there was a reduction in postprandial glucose, there was no reduction in HbA1c in type 2 diabetes [11].

Basal Insulins

The treatment of insulin with protamine (isophane insulins) or its co-injection with zinc (Lente insulin series), provided basal insulin treatment for many years. Isophane insulins have persisted as a treatment option, predominantly because of lower acquisition cost in comparison to newer basal analogues. Due to a duration of action of 12–14 hours, isophane is best used twice daily in T1D, but may be used once daily as part of T2D management. There are two major limitations with the use of these 'cloudy' suspensions however, variability in both peak action and duration. Studies have shown variable rate of absorption of isophane insulin on a day-to-day basis of 25–35% which may lead to variable control and risk of hypoglycaemia when aiming for good control [12]. Additionally, the pharmacokinetic profiles of isophane insulins show a peak in action at 4–6 hours which may predispose to hypoglycaemia, particularly when administered before bed, necessitating a suppertime long-acting carbohydrate snack. Duration of action may be injection volume dependent and variable with the same dose on different days.

Analogue Basal Insulins

Subsequently, analogue basal insulins were introduced in which modifications to the insulin chain lengthened the duration of action. The first basal analogue was insulin glargine which was created by adding two basic arginine amino acid residues onto the B chain and substituting a glycine for an asparagine residue on the A-chain (Figure 15.5.3.2). This molecule is soluble at acidic pH, but once exposed to the neutral subcutaneous environment it crystallizes. Its duration of action is long at 18–24 hours and in many people can be used as a once daily insulin, however others may experience pre-dose hyperglycaemia and require twice daily administration. It is described as 'peakless' (Figure 15.5.3.3) and has an onset of action 2–4 hours after injection. Insulin detemir has a fatty acid chain (myristic acid) bound to a lysine amino acid on the B chain of the insulin molecule. This slows down its absorption and allows greater binding to subcutaneous and serum albumin where it is chemically inert. Detemir similarly has onset of action 2–4 hours after dose but duration of action of 16–18 hours and is thus commonly administered twice daily in T1D.

A direct, head to head, non-inferiority study comparing detemir and glargine showed comparable results between the two [13]. Thus, each insulin could legitimately be chosen according to user preference. Those with more variable lifestyles could choose detemir allowing dose down-titrations on days when exercise or alcohol might increase the risk of nocturnal hypoglycaemia whereas glargine might be preferred in those who desire stability and once daily injection. A meta-analysis has suggested that long-acting insulin analogues are superior to intermediate acting insulin but at greater acquisition cost [14].

More recently new insulin analogues have been introduced to the market. A biosimilar insulin glargine is available and has been demonstrated to be non-inferior to standard insulin glargine in both type 1 and type 2 diabetes [15, 16]. Insulin degludec is another modified insulin where hexadecanedioic acid is conjugated to the lysine amino acid at position B29. This modification reduces absorption and enhances protein binding in a similar fashion to detemir but gives insulin degludec a very long duration of action of 42 hours. It therefore allows once daily administration with some variability in daily timing of dose. In T1D and T2D statistically significant reductions in the rate of hypoglycaemia are seen in comparison to glargine [17, 18].

Premixed Insulins

Premixed insulins contain rapid-acting and basal insulin in a predefined ratio. Historically, most patients would 'free-mix' a variable amount of soluble and isophane insulin twice a day. For example, a mixture of soluble and intermediate insulin in the morning would allow a child to get through the school day without needing an injection at school. Certainly, premixed insulins as part of a biphasic regime may be useful in patients where adherence or patient preference for fewer injections dictates treatment decisions such as in children or adolescents. Evidence for their use is stronger in T2D where a meta-analysis has shown that greater than 50% of people can achieve good glycaemic control on a biphasic regime although patients who can manage a basal-bolus had better control with a similar hypoglycaemia rate [19].

Numerous premixed insulins are available and there are important differences between them. Until recently the combinations included either soluble insulin or rapid-acting analogue with isophane. The ratio of short to long-acting insulin varies between preparations but is usually between 25% to 50% short/rapid acting. The 30/70 mixtures (with 30% soluble and 70% isophane insulin) are the most widely used, but in cultures with high carbohydrate intake, 50:50 mixtures have been shown to provide better glucose control [20]. Nomenclature varies around the world and prescribers should be familiar with the terminology used in their country. Due to the pH dependent crystallization of insulin glargine, premixed insulins with analogue basals have not been available. This has changed with licensing of a combination of aspart with degludec.

The limitation of the premixed insulin is the in-flexibility of dosing schedule. The peak of the soluble component is designed to cover the immediate meal, and the peak of the isophane component that arrives 4–6 hours later is designed to cover a second meal. However, if that meal is delayed or of a different carbohydrate content, this could predispose to hypo or hyperglycaemia. For this reason, these fixed combination insulins are best used in those with fixed daily habits.

Different Strengths of Insulins

In some markets, different strengths of insulin are available. The most commonly available insulin is U100 that contains 100 IU of insulin/ml. In some markets, more dilute insulin is available (U40–40 IU/ml) that can be used especially in very young children. It is key to know which insulin is being used and use the appropriate delivery device that is calibrated to the strength of insulin being delivered. In the 1980s it was recognized that the numerous different strengths of insulins contributed to hypoglycaemia [21]. Poor understanding of the different strengths among doctors, pharmacists, and patients together with unclear labelling of insulin syringes compounded the problem. Insulin production and sale in most western markets was consequently restricted to production at a fixed strength (U100).

The obesity epidemic has caused increased rates of insulin resistance and a new need for higher-strength insulins. Higher insulin concentration and hence, smaller dose volumes have lower

variability in dose absorption, duration of action, and have the advantages of single rather than multiple injection with higher doses. U500 insulin is available comprising 500 IU of insulin/ml and can be used in extremely insulin resistant patients [22]. A concentrated insulin lispro U200 (200 units/ml) was introduced to the market in 2015. Both are soluble or prandial insulins.

There are currently two concentrated basal insulins available. Insulin degludec is available as a U200 formulation (200 units/ml) and insulin glargine is available as U300 (300 units/ml). The pharmacokinetic profiles indicate that glargine U300 has a longer duration of action with lower peak activity [23]. In T2D there was a small but statistically significant reduction in hypoglycaemia (particularly at night) when using U300 [24] and it may be considered in patients where this is a particular concern. There was no difference between Glargine-U100 and Glargine-U300 in T1D trials [25].

Prescribers and pharmacists must have a good awareness of the differences between these insulins and patients must be educated appropriately to prevent dose errors. If we can overcome these barriers higher-strength insulins may play a significant role in optimized diabetes treatment [26].

Insulin Delivery Devices

Historically, insulin was drawn up from a vial and injected via a specific insulin syringe with marked graduations allowing the correct volume and dose of insulin delivery. In the early days, these were glass syringes that required regular cleaning and sterilization. Devices such as the Palmer Injector were developed to make things easier, although these were not widely used. In the 1980s, manufacturers developed insulin pens that come as pre-filled disposable pens, or as re-usable pens in which cartridges can be changed [27]. Insulin pens remove the need to draw up insulin which can reduce error, making them more convenient, more discreet, and easier to handle. They require disposable needles that can be of different lengths. Insulin users describe pens as more acceptable and greater satisfaction can be achieved with them while achieving improved HbA1c and reduced hypoglycaemia rates [28].

Key differentiators between pens are differences in the manual dexterity and force required to use them and ease of delivery. Newer pen devices offer the ability to synchronize with a mobile application featuring a bolus advisor, enabling convenient insulin dose tracking and more accurate bolus advice [29].

There has long been an interest in needle-free injection devices, as the use of a needle is a major barrier for many patients who need insulin therapy. There have been some needle-free 'jet' injector devices available on the market, but these are not widely used as they often cause pain or bruising and have less predictable delivery [30].

It is clear that, whatever the preferred method of insulin delivery, high quality education and support from the multidisciplinary diabetes team is crucial in delivering best outcomes [31].

Inhaled Insulins

The possibility of absorption of insulin via inhaled devices offers a clear advantage to some patients (particularly those with phobia of injections and extreme variability in the action of insulin administered subcutaneously). A key benefit is the rapid onset of action with associated reductions in postprandial spikes, but this is offset by reduced and variable bioavailability together with short duration of action.

Initial attempts at development were beset with problems including variable bioavailability and concerns around effects of insulin on lung function. This ultimately led to withdrawal of all drugs from development for some time [32]. More recently, insulin adsorbed to porous Technosphere particles has been developed. Technosphere insulin (TI) has been shown to have more rapid absorption and onset of action (10 minutes) compared to rapid-acting analogues. It has peak action after 30–60 minutes and short duration of action (2–3 hours). Despite this, a meta-analysis has shown greater reductions in HbA1c with subcutaneous insulin compared to TI, albeit with a greater propensity to severe hypoglycaemia and weight gain [33]. Given this, TI may be considered for use when subcutaneous prandial insulin injection is not possible.

Continuous Subcutaneous Insulin Infusion (CSII)

Insulin pumps can be used to deliver rapid-acting insulin continuously into the subcutaneous compartment and were first used in clinical studies in the 1970s [34]. The pump has an electromechanical motor and a reservoir of insulin, usually rapid-acting analogue insulin. The patient and clinician preprogramme different basal rate profiles into the pump that allow basal insulin delivery that can be varied through the day in response to the known circadian profile of insulin sensitivity [2]. In health, basal insulin delivery is lower overnight and rises immediately before the patient wakes up, often called the 'dawn phenomenon' [35]. Pump users can program the pump with different basal rate profiles to accommodate changes in insulin requirements with work, weekends, or illness (Figure 15.5.3.4). Crucially, this basal insulin delivery can also be temporarily increased or decreased using a 'temporary basal rate' in response to exercise, illness, or alcohol.

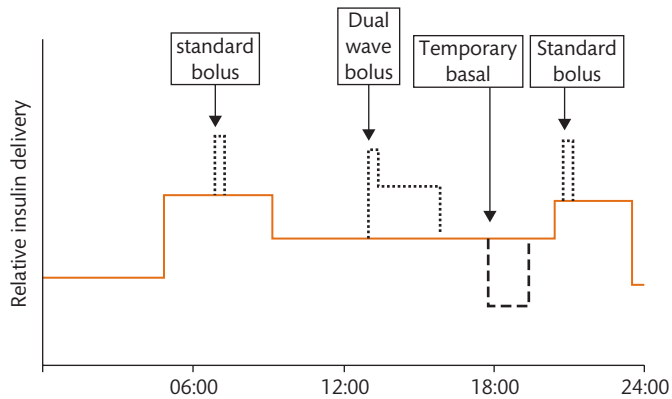


Figure 15.5.3.4 Demonstration of several features of insulin pump insulin delivery.

The flat line indicates continuous insulin delivery. An example alternative basal profile is superimposed with low rate of insulin delivery overnight to prevent nocturnal hypoglycaemia and greater delivery in the morning and evening to combat hyperglycaemia at times of physiological insulin resistance. Finally at meal times, boluses are delivered 15 minutes before meals to cover breakfast and the evening meal as well as a dual wave bolus at lunchtime to cover a high protein/fat meal.

When a bolus of insulin is required to cover carbohydrate intake or to correct hyperglycaemia, the user presses a button to deliver the desired bolus. Boluses can be delivered immediately, or over a prolonged duration (square or dual wave boluses) to cover high fat meals that produce sustained hyperglycaemia. Most pumps have software (bolus calculators) that perform the calculations required for boluses using preset insulin to carbohydrate ratio (ICR) and insulin sensitivity factor (correction factor). One of the key benefits of insulin pumps is the ability to easily deliver multiple boluses to cover all carbohydrate intakes, and take small corrections when needed, as well as the ability to rapidly adapt the basal profile as required. These features allow better matching of insulin requirement to insulin delivery, facilitating improved diabetes control in terms of lower HbA1c as well as less hypoglycaemia.

Observational data show that CSII therapy is associated with significant improvements in HbA1c and hypoglycaemia. A meta-analysis of CSII studies demonstrated a 4-fold reduction in rates of severe hypoglycaemia requiring third party assistance, and a reduction in HbA1c [36]. Benefits in both hypoglycaemia and in glucose control were dependant on baseline. Most studies have shown significant improvements in quality of life, which is an important factor for a lifelong, chronic condition. Initially developed in the United Kingdom and the United States, the uptake of insulin pump therapy is markedly different in different countries, varying from above 40% of people with T1D in the United States to about 20–30% in many European countries and under 15% in the United Kingdom. This can be explained by the cost of these devices, and the perceived complexity, and increased education required for safe and effective use. Cost aside, the potential benefit in motivated and well-educated people with T1D with either sustained hyperglycaemia or troublesome hypoglycaemia is clear. The best results with CSII are demonstrated in specialized centres with access to a highly trained staff group for education and support of insulin pump users.

The REPOSE RCT enrolled patients who did not have a clear preference or indication for pump therapy and compared the outcomes of a structured education programme with or without pump therapy. Both arms showed modest improvements in HbA1c with no differences in hypoglycaemia rates, but with clear improved satisfaction in those with pump therapy. Interestingly, a post-hoc per-protocol analysis showed that when patients who crossed over from pump to MDI or vice versa were analysed according to the treatment they actually received, there was a significant benefit of pump therapy on HbA1c. Similarly, the HypoCOMPaSS study in patients with impaired awareness of hypoglycaemia found that when provided intensive support, there was no incremental benefit of pump therapy over the control arm using injections, although again, there was a clear benefit in quality of life [37, 38].

A recent randomized controlled trial of insulin pumps in patents with type 2 diabetes and high insulin requirements demonstrated improved diabetes control with lower insulin requirements when using insulin pumps [39].

Types of CSII Insulin Pump

Broadly speaking, there are two common insulin pump types that are in use today: tethered pumps and patch pumps. The key difference is the presence or absence of tubing between the pump and the skin.

Tethered pumps deliver insulin via tubing to a cannula inserted subcutaneously. The cannula itself usually looks like a small disc and is placed 4–8 mm deep using an insertion needle. The pump is usually powered by domestically available batteries and can be carried in a comfortable, convenient position in pockets or pouches as decided by the user. The principle control of the pump is generally via an interface on the pump itself, although some may be controlled via a remote device. This allows the user to control the pump unobtrusively. However, some people do not like to be attached to the tubing, finding that the tubing can catch on things and get ripped out, or have deeper-seated body image issues decreasing acceptability of pump therapy.

Patch pumps are directly attached to the skin for the duration of use and have the convenience of no tubing but have the disadvantage of being dislodged if knocked. Control of the pump is generally achieved by a wirelessly connected remote device which usually also functions as a blood glucose meter. There is no particular advantage of one type of pump over the other, apart from patient preference [40].

Most currently available pumps allow data including basic settings, blood glucose values, and insulin doses delivered to be stored. These can then be downloaded by users or healthcare professionals to allow periodic review of control.

Advanced Features of Pumps

The delivery of basal insulin can be set up in numerous ways. Some teams prefer to start with a 'flat' basal that is usually 40–50% of the total daily insulin (Figure 15.5.3.4). Other groups use a circadian profile, or a modified circadian profile that can be programmed using a template, allowing the basal insulin delivery to match the normal circadian profile of insulin requirement. It is important to recognize that overnight insulin requirements vary by up to 200% in the same person between nights [41], and so patients still need to adjust their insulin levels based on review of their blood glucose data.

The basal insulin delivery can be temporarily increased or decreased by either a percentage of the current basal rate or by a fixed amount for a fixed duration, termed a 'temporary basal rate'. This is particularly helpful when the patient needs to reduce basal insulin delivery in response to exercise or alcohol, or increase it in response to illness. Changes in basal rates can take time to impact on plasma insulin, and so these features are most useful when the increase or decrease is required for a longer period.

Different Types of Bolus

High fat or protein content in meals can cause a carbohydrate independent rise in glucose levels. This may be due to temporary insulin resistance caused by fat, or gluconeogenesis stimulated by protein intake. High fat meals can lead to late hyperglycaemia that can last up to eight hours after a meal. Insulin pump users can program square or dual wave insulin boluses that can cover this delayed increase in insulin requirement.

Bolus Advisors

Most insulin pumps currently available have integrated bolus advisors. These are calculators that can perform simple calculations required to deliver precise amounts of insulin taking into consideration factors such as ICR, insulin sensitivity factor, and insulin

action time. The ICR is the amount of carbohydrate that is covered by a single unit of insulin. This can also be expressed as the amount of insulin required to cover a fixed amount of carbohydrate (usually a carbohydrate exchange which can be 10 or 12 grams). This usually varies between 8 and 12 grams per unit of insulin, although for insulin sensitive patients this can be higher and for insulin resistant patients, this can be as low as 2 or 3 grams per unit of insulin.

The use of bolus advisors offers greater accuracy in dosing and some studies have shown that they help reduce glucose variability and improved glucose control [42]. Additionally users report this technology saved time, improving confidence in the dose given [43].

Sensor Augmented Pumps (SAP)

With the advent of real time continuous glucose monitoring systems, some pumps can connect with integrated continuous glucose monitoring (CGM) systems allowing display of glucose values on the pump (see Chapter 15.5.2, 'Glucose Monitoring and Sensing'). The user can utilize these data to make frequent small adjustments in insulin dosing via the pump. Ultimately SAPs could provide complete 'closed loop' control that functions like an 'artificial pancreas'.

The first step along this pathway was systems that used 'threshold suspend' of insulin. These systems can automatically suspend insulin delivery for a fixed duration (usually for up to 2 hours). For example, if the 'low threshold' is set at 3.6 mmol/L (65 mg/dl), the system alerts the user when the glucose reaches this threshold, and if there is no response from the user, it will automatically suspend insulin delivery for up to 2 hours. These types of system do not actually prevent hypoglycaemia, but can minimize the impact of prolonged low glucose and in randomized controlled trials have been shown to significantly reduce duration of nocturnal hypoglycaemia [44].

The next generation of SAP systems proactively aim to prevent hypoglycaemia, by automatically suspending insulin delivery when hypoglycaemia is predicted. Often these systems will suspend insulin delivery while glucose is still in the normal or desirable range, based on the trajectory of glucose. They can also restart insulin once the glucose is back above a safe level and rising. These systems offer progressive improvements in protection against hypoglycaemia [45]. Similarly, the devices can alarm if glucose levels are too high, prompting the user to take appropriate corrective action.

Closed Loop 'Artificial Pancreas' Systems

Over the last 10 years, various closed-loop systems have been developed and tested in different environments (see Chapter 15.8.2, "Closed Loop" Insulin Delivery). A closed-loop system includes three main components. The first is a continuous glucose sensor that detects glucose levels including direction of change. The second is an algorithm that can be housed on a portable computer, a mobile phone, or even the insulin pump itself. This controller calculates the insulin required at any given time point to maintain glucose levels as close as possible to the target glucose. Often the calculation is overseen by safety modules that govern the maximum amount of insulin that can be given and may adjust the calculation if food or exercise are detected. The third component, the insulin pump, delivers the insulin as advised by the algorithm.

Closed-loop systems have been shown to be relatively safe in clinical trials, providing improved glucose control with reduced hypoglycaemia compared to standard treatments. A major challenge for closed-loop systems is the delay in glucose sensing (changes in interstitial glucose often occur 5–10 minutes after blood glucose), which means there is a delay before a rise due to food or glucose fall due to exercise is sensed. The other issue is the relative slowness of currently available 'fast' acting analogues that have a delayed peak and prolonged duration of action, especially when compared to portally secreted endogenous insulin.

A key issue for closed-loop systems is their ability to deal with unannounced meals or exercise. Current insulin kinetics require rapid-acting insulin to be dosed 10–15 minutes before a meal. This is obviously a problem for a truly closed-loop system which requires no input from the user. One solution is a 'hybrid' closed-loop system that requires 'meal announcement' from the user. The more precise this is, the greater the degree of post-meal glucose control.

Another problem is unannounced exercise. Even though systems can be programmed to reduce insulin delivery when exercise is detected, the prolonged action of current insulin preparations and delivery through the subcutaneous route mean that even after complete suspension of insulin there is considerable risk of hypoglycaemia with unannounced exercise.

One hybrid closed-loop system is currently available, the Medtronic Minimed™ 670 g Insulin Pump System with others in development.

Continuous Intraperitoneal Insulin Infusion (CIPII)

A very small number of people use an insulin pump where insulin is continuously delivered into the peritoneum and absorbed preferentially into the portal blood via peritoneal blood vessels [46]. There are two methods for CIPII: fully implanted pumps or external pumps with intraperitoneal delivery.

Implantable pumps have been used in a small group of patients, predominantly in France for many years. An insulin pump is inserted into a pocket in the abdominal wall and a catheter tunnelled to the peritoneal cavity [47]. CIPII has been shown to improve HbA1c and reduce number of hospital admissions in a series of complex patients [48]. It is limited by the need for pump refills and system flushes in addition to potential for 'pocket' infection so cannot be considered a mainstay of treatment [46].

An alternative technology is the DiaPort system in which an external insulin pump is connected to an intraperitoneal delivery system. This has been shown to reduce severe hypoglycaemia in a complex group of people with T1D [49] albeit with problems of port site infections. This has been addressed in a more recent clinical trial without fully published results, but in which many of the infections were managed conservatively [50]. Both implantable technologies require insulin specifically manufactured to prevent aggregation within the pump [51].

Treatments on the Horizon

Glucose-responsive 'smart' insulins are being developed in which the insulin activity depends upon the serum or tissue glucose level [52]. Several strategies for achieving this are being investigated. They include incorporating insulin within a polymer or matrix which has the capacity to release insulin according to glucose concentration. Another strategy is to modify the insulin molecule to allow glucose-dependent action or conformational change which inhibits binding to the insulin receptor during hypoglycaemia.

Micro needles in a patch may offer a method of delivering insulin intradermally with faster absorption due to the greater vascularity of the dermal space. Systems for combining this mode of therapy with glucose-responsive elements are in development [53, 54].

As well as offering a more acceptable route of administration, insulin absorbed through the gut offers more physiological delivery to the vasculature. Absorption into the portal veins allows highly concentrated action on the liver, with relatively little acting systemically providing potentially lower hypoglycaemia rates. With ongoing research into this field, oral insulins may come to the market as a viable treatment option in the coming years.

REFERENCES

1. Duckworth WC, Kitabchi AE. Insulin metabolism and degradation. *Endocr Rev*, 1981; **2**: 210–33.
2. Polonsky KS, Given BD, Vancauter E. 24-hour profiles and pulsatile patterns of insulin-secretion in normal and obese subjects. *J Clin Invest*, 1988; **81**: 442–8.
3. Bliss M. The history of insulin. *Diabetes Care*, 1993; **16**: 4–7.
4. Goeddel DV, Kleid DG, Bolivar F, *et al.* Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A*, 1979; **76**: 106–10.
5. Slattery D, Amiel SA, Choudhary P. Optimal prandial timing of bolus insulin in diabetes management: a review. *Diabet Med*, 2018; **35**: 306–16.
6. Torlone E, Fanelli C, Rambotti AM, *et al.* Pharmacokinetics, pharmacodynamics and glucose counterregulation following subcutaneous injection of the monomeric insulin analogue [Lys(B28), Pro(B29)] in IDDM. *Diabetologia*, 1994; **37**: 713–20.
7. Siebenhofer A, Plank J, Berghold A, *et al.* Short acting insulin analogues versus regular human insulin in patients with diabetes mellitus. *Cochrane Database Syst Rev*, 2006; **2**: CD003287.
8. Pedersen-Bjergaard U, Kristensen PL, Beck-Nielsen H, *et al.* Effect of insulin analogues on risk of severe hypoglycaemia in patients with type 1 diabetes prone to recurrent severe hypoglycaemia (HypoAna trial): a prospective, randomised, open-label, blinded-endpoint crossover trial. *Lancet Diabetes Endocrinol*, 2014; **2**: 553–61.
9. Russell-Jones D, Bode BW, De Block C, *et al.* Fast-acting insulin aspart improves glycemic control in basal-bolus treatment for type 1 diabetes: results of a 26-week multicenter, active-controlled, treat-to-target, randomized, parallel-group trial (onset 1). *Diabetes Care*, 2017; **40**: 943–50.
10. Russell-Jones D, Bode BW, De Block C, *et al.* Fast-acting insulin aspart improves glycemic control in basal-bolus treatment for type 1 diabetes: results of a 26-week multicenter, active-controlled, treat-to-target, randomized, parallel-group trial (onset 1). *Diabetes Care*, 2017; **40**: 943–50.
11. Bowering K, Case C, Harvey J, *et al.* Faster aspart versus insulin aspart as part of a basal-bolus regimen in inadequately controlled type 2 diabetes: the onset 2 trial. *Diabetes Care*, 2017; **40**: 951–57.
12. Heinemann L. Variability of insulin absorption and insulin action. *Diabetes Technol Ther*, 2002; **4**: 673–82.
13. Heller S, Koenen C, Bode B. Comparison of insulin detemir and insulin glargine in a basal-bolus regimen, with insulin aspart as the mealtime insulin, in patients with type 1 diabetes: a 52-week, multinational, randomized, open-label, parallel-group, Treat-to-Target noninferiority tr. *Clin Ther*, 2009; **31**: 2086–97.
14. Tricco AC, Ashoor HM, Antony J, *et al.* Safety, effectiveness, and cost effectiveness of long acting versus intermediate acting insulin for patients with type 1 diabetes: systematic review and network meta-analysis. *BMJ*, 2014; **349**: g5459–g5459.
15. Blevins TC, Dahl D, Rosenstock J, *et al.* Efficacy and safety of LY2963016 insulin glargine compared with insulin glargine (Lantus®) in patients with type 1 diabetes in a randomized controlled trial: the ELEMENT 1 study. *Diabetes Obes Metab*, 2015; **17**: 726–33.
16. Rosenstock J, Hollander P, Bhargava A, *et al.* Similar efficacy and safety of LY2963016 insulin glargine and insulin glargine (Lantus®) in patients with type 2 diabetes who were insulin-naïve or previously treated with insulin glargine: a randomized, double-blind controlled trial (the ELEMENT 2 study). *Diabetes, Obes Metab*, 2015; **17**: 734–41.
17. Lane W, Bailey TS, Gerety G, *et al.* Effect of insulin degludec vs insulin glargine u100 on hypoglycemia in patients with type 1 diabetes. *JAMA*, 2017; **18**: 33–44.
18. Wysham C, Bhargava A, Chaykin L, *et al.* Effect of insulin degludec vs insulin glargine u100 on hypoglycemia in patients with type 2 diabetes. *JAMA*, 2017; **318**: 45–56.
19. Giugliano D, Maiorino MI, Bellastella G, *et al.* Efficacy of insulin analogs in achieving the hemoglobin A1c target of <7% in type 2 diabetes: meta-analysis of randomized controlled trials. *Diabetes Care*, 2011; **34**: 510–17.
20. Deed G, Kilov G, Dunning T, *et al.* Use of 50/50 premixed insulin analogs in type 2 diabetes: systematic review and clinical recommendations. *Diabetes Ther*, 2017; **8**: 1265–96.
21. Bloom A, Keen H, Watkins PJ. A change to 100-unit insulin dosage will reduce errors. *Br Med J (Clin Res Ed)*, 1981; **283**: 33.
22. Cochran E, Musso C, Gorden P. The use of U-500 in patients with extreme insulin resistance. *Diabetes Care*, 2005; **28**: 1240–4.
23. Hurren KM, O'Neill JL. Pharmacodynamic and pharmacokinetic evaluation of insulin glargine U300 for the treatment of type 1 diabetes. *Expert Opin Drug Metab Toxicol*, 2016; **12**: 1521–6.
24. Ritzel R, Roussel R, Bolli GB, *et al.* Patient-level meta-analysis of the EDITION 1, 2 and 3 studies: glycaemic control and hypoglycaemia with new insulin glargine 300 U/ml versus glargine 100 U/ml in people with type 2 diabetes. *Diabetes, Obes Metab*, 2015; **17**: 859–67.
25. Home PD, Bergenstal RM, Bolli GB, *et al.* New insulin glargine 300Units/mL versus glargine 100 Units/mL in people with type 1 diabetes: a randomized, phase 3a, open-label clinical trial (EDITION 4). *Diabetes Care*, 2015; **38**: 2217–25.
26. Dashora U, Castro E. Insulin U100, 200, 300 or 500? *Br J Diabetes*, 2016; **16**: 10.
27. Selam J-L. Evolution of diabetes insulin delivery devices. *J Diabetes Sci Technol*, 2010; **4**: 505–13.
28. Lasalvia P, Barahona-Correa JE, Romero-Alvernia DM, *et al.* Pen devices for insulin self-administration compared with needle and vial: systematic review of the literature and meta-analysis. *J Diabetes Sci Technol*, 2016; **10**: 959–66.
29. Bailey TS, Stone JY. A novel pen-based Bluetooth-enabled insulin delivery system with insulin dose tracking and advice. *Expert Opin Drug Deliv*, 2017; **14**: 697–703.
30. Engwerda EEC, Tack CJ, De Galan BE. Pharmacokinetic and pharmacodynamic variability of insulin when administered by jet injection. *J Diabetes Sci Technol*, 2017; **11**: 947–52.
31. Frid AH, Kreugel G, Grassi G, *et al.* New insulin delivery recommendations. *Mayo Clin Proc*, 2016; **91**: 1231–55.
32. Setji TL, Hong BD, Feinglos MN. Technosphere insulin: inhaled prandial insulin. *Expert Opin Biol Ther*, 2016; **16**: 111–17.
33. Pittas AG, Westcott GP, Balk EM. Efficacy, safety, and patient acceptability of Technosphere inhaled insulin for people with

diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2015; **3**: 886–94.
34. Pickup J, Keen H. Continuous subcutaneous insulin infusion at 25 years. *Diabetes Care*, 2002; **25**: 593–8.
35. Bachran R, Beyer P, Klinkert C, *et al.* Basal rates and circadian profiles in continuous subcutaneous insulin infusion (CSII) differ for preschool children, prepubertal children, adolescents and young adults. *Pediatr Diabetes*, 2012; **13**: 1–5.
36. Pickup JC, Sutton AJ. Severe hypoglycaemia and glycaemic control in Type 1 diabetes: meta-analysis of multiple daily insulin injections compared with continuous subcutaneous insulin infusion. *Diabet Med*, 2008; **25**: 765–74.
37. Heller S, Amiel S, Campbell M, *et al.* Relative effectiveness of insulin pump treatment over multiple daily injections and structured education during flexible intensive insulin treatment for type 1 diabetes: cluster randomised trial (REPOSE). *BMJ*, 2017; **356**: j1285.
38. Little SA, Leelarathna L, Walkinshaw E, *et al.* Recovery of hypoglycemia awareness in long-standing type 1 diabetes: a multicenter 2 × 2 factorial randomized controlled trial comparing insulin pump with multiple daily injections and continuous with conventional glucose self-monitoring (HypoCOMPaSS). *Diabetes Care*, 2014; **37**: 2114–22.
39. Reznik Y, Cohen O, Aronson R, *et al.* Insulin pump treatment compared with multiple daily injections for treatment of type 2 diabetes (OpT2mise): a randomised open-label controlled trial. *Lancet*, 2014; **384**: 1265–72.
40. Leelarathna L, Roberts SA, Hindle A, *et al.* Comparison of different insulin pump makes under routine care conditions in adults with type 1 diabetes. *Diabet Med*, 2017; **34**: 1372–9.
41. Ruan Y, Thabit H, Leelarathna L, *et al.* Variability of insulin requirements over 12 weeks of closed-loop insulin delivery in adults with type 1 diabetes. *Diabetes Care*, 2016; **39**: 830–2.
42. Ziegler R, Rees C, Jacobs N, *et al.* Frequent use of an automated bolus advisor improves glycemic control in pediatric patients treated with insulin pump therapy: results of the Bolus Advisor Benefit Evaluation (BABE) study. *Pediatr Diabetes*, 2016; **17**: 311–18.
43. Lawton J, Kirkham J, Rankin D, *et al.* Perceptions and experiences of using automated bolus advisors amongst people with type 1 diabetes: a longitudinal qualitative investigation. *Diabetes Res Clin Pract*, 2014; **106**: 443–50.
44. Weiss R, Garg SK, Bode BW, *et al.* Hypoglycemia reduction and changes in hemoglobin A1c in the ASPIRE in-home study. *Diabetes Technol Ther*, 2015; **17**: 542–7.
45. Abraham MB, Nicholas JA, Smith GJ, *et al.* Reduction in hypoglycemia with the predictive low-glucose management system: a long-term randomized controlled trial in adolescents with type 1 diabetes. *Diabetes Care*, 2018; **41**: 303–10.
46. Spaan NA, Teplova AE, Renard E, *et al.* Implantable insulin pumps: an effective option with restricted dissemination. *Lancet Diabetes Endocrinol*, 2014; **2**: 358–60.
47. Renard E, Schaepelynck-Bélicar P. Implantable insulin pumps. A position statement about their clinical use. *Diabetes Metab*, 2007; **33**: 158–66.
48. DeVries JH, Eskes SA, Snoek FJ, *et al.* Continuous intraperitioneal insulin infusion in patients with `brittle' diabetes: favourable effects on glycaemic control and hospital stay. *Diabet Med*, 2002; **19**: 496–501.
49. Liebl A, Hoogma R, Renard E, *et al.* A reduction in severe hypoglycaemia in type 1 diabetes in a randomized crossover study of continuous intraperitoneal compared with subcutaneous insulin infusion. *Diabetes Obes Metab*, 2009; **11**: 1001–8.
50. Garcia-Verdugo R, Erbach M, Schnell O. A new optimized percutaneous access system for CIPII. *J Diabetes Sci Technol*, 2017; **11**: 814–21.
51. Schaepelynck P, Riveline J-P, Renard E, *et al.* Assessment of a new insulin preparation for implanted pumps used in the treatment of type 1 diabetes. *Diabetes Technol Ther*, 2014; **16**: 582–9.
52. Rege NK, Phillips NFB, Weiss MA. Development of glucose-responsive 'smart' insulin systems. *Curr Opin Endocrinol Diabetes Obes*, 2017; **24**: 267–78.
53. Chen G, Yu J, Gu Z. Glucose-responsive microneedle patches for diabetes treatment. *J Diabetes Sci Technol*, 2019; **13**: 41–8.
54. Hilgenfeld R, Seipke G, Berchtold H, *et al.* The evolution of insulin glargine and its continuing contribution to diabetes care. *Drugs*, 2014; **74**: 911–27.

15.5.4 Non-Insulin Glucose-Lowering Agents

Clifford J. Bailey and Melanie J. Davies

Introduction

This chapter will focus on non-insulin therapies to treat hyperglycaemia in type 2 diabetes. Glycaemic control is an important part of the management of type 2 diabetes, which is a complex process due to the diverse, variable, and progressive nature of the disease pathogenesis, clinical complications, and societal impact (**Box 15.5.4.1**). Care should be individualized, flexible, and realistic, with provision for patient education and empowerment to enable optimal benefit from the guidance and interventions offered by healthcare professionals. Patient preference should be considered with a focus on

Box 15.5.4.1 Multiple features of type 2 diabetes mellitus that make this a complex disease to treat

- Polygenic susceptibility
- Multiple environmental impositions
- Variable endocrine defects (e.g. insulin resistance and islet dysfunction)
- Diverse metabolic disturbances
- Progressive natural history
- Extensive complications and comorbidities (e.g. cardiovascular and renal diseases)
- Many concurrent therapies usually involved
- Substantial patient education, empowerment, and commitment required
- Impacts family, friends, and work colleagues
- Many different healthcare disciplines implicated

choices which optimize adherence to and persistence with therapy. Relief of acute symptoms and attention to long-term complications and comorbidities often preoccupy and sometimes overwhelm the treatment process. However, early and sustained remediation of endocrine and metabolic disturbances, in combination with containment of modifiable cardiovascular risk factors, prevents the onset, and limits the severity of chronic pathology.

Glycaemic Management

Glycaemic control serves as a valuable indicator of metabolic status, and we re-emphasize that good glycaemic control can reduce both micro- and macrovascular complications. Reductions in microvascular disease have been confirmed in almost all trials of intensive glycaemic control, whereas improved cardiovascular outcomes may only emerge after several years, being more evident in younger patients with better control from the time of diagnosis [1]. Moreover, the closer that long-term glycaemic control approaches to normoglycaemia, the fewer the complications (Figure 15.5.4.1), excepting that overintensification of treatment to cause persistent or severe episodes of hypoglycaemia must be avoided [2]. There is also evidence that some glucose-lowering agents offer independent benefits against the major cardiovascular and renal complications of type 2 diabetes.

The post-trial follow-up of the United Kingdom Prospective Diabetes Study (UKPDS) has provided clinical confirmation of the so-called 'glycaemic memory' or 'legacy effect'. This effect, which probably reflects the accumulated damage from glucotoxicity, renders individuals with poor glycaemic control during earlier stages of the disease at much higher risk of complications in later life, even if their control is subsequently improved [3]. Thus, available evidence supports the early use of blood glucose-lowering therapy to attain and maintain a level of glycaemic control as close to normal as safely practicable and commensurate with the circumstances of the patient.

Guidelines

The many guidelines and consensus statements regarding pharmacological therapies to control blood glucose in type 2 diabetes show a considerable degree of conformity [4–6] (Figure 15.5.4.2).

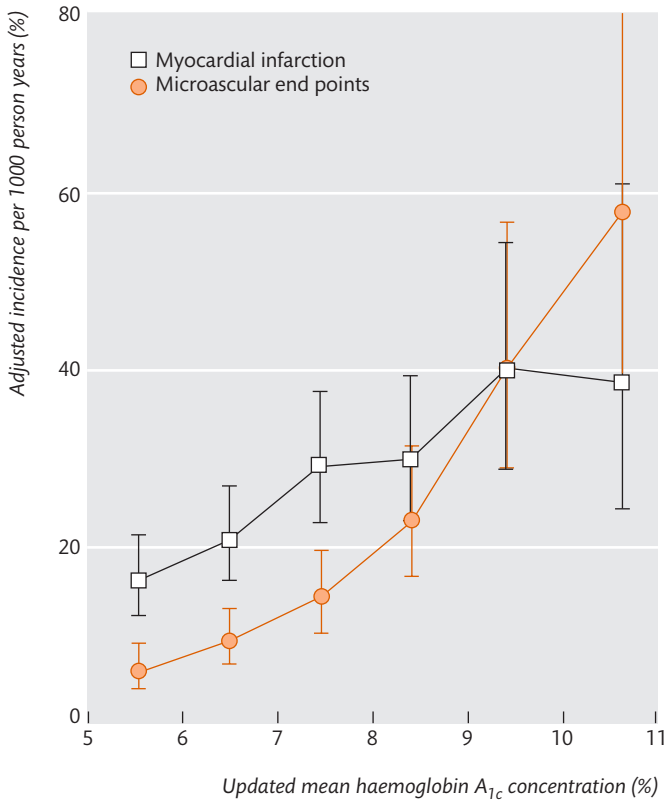
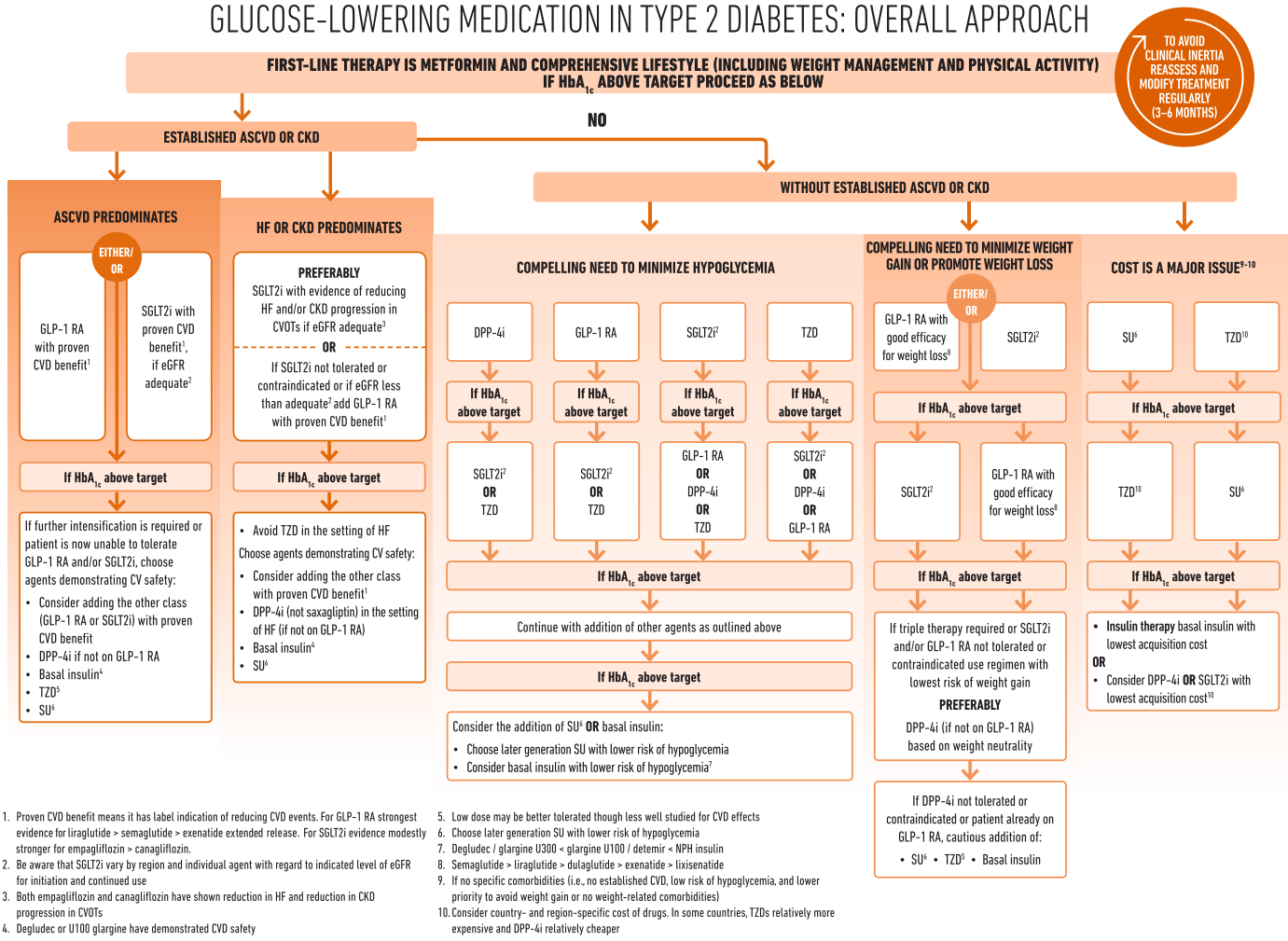


Figure 15.5.4.1 Improved glycaemic control as measured by a lowering of average HbA1c over 10 years during the UKPDS was associated with reductions in microvascular complications and myocardial infarction.
Reproduced with permission from Stratton et al, *Br Med J* 2000; **321**: 405–12, ref 2.

Lifestyle measures invariably form the foundation for management of type 2 diabetes, especially diet and exercise for weight control, cardiovascular wellbeing, and psychosocial health [4–8]. Pharmacological therapies are added to achieve and sustain the desired glycaemic target, generally starting with one oral agent. If the target is not achieved or maintained by increasing the doses, a differently acting oral agent or injectable agent is added. If this combination is unable to prevent disease progression, a third agent is added. The selection of agents at each stage will be strongly influenced by the presence of comorbidities, including CVD, obesity, renal or hepatic impairment, age, and personal circumstances. Submaximal doses of two or more agents may be preferred to enhance efficacy and reduce side effects that are sometimes encountered with a large dose of one agent. Insulin is customarily reserved for patients in whom a combination of other agents does not provide continuing control, but insulin may be introduced earlier if patients are severely hyperglycaemic and substantially symptomatic with comorbidities that deter the use of other therapies.

Synopsis of Therapies

The main orally administered blood glucose-lowering therapies are metformin, sulphonylureas, meglitinides, dipeptidyl peptidase-4 (DPP-4) inhibitors (gliptins), sodium/glucose cotransporter-2 (SGLT-2) inhibitors, thiazolidinediones (TZDs), and α -glucosidase inhibitors (AGIs). Bromocriptine and colesevelam also have an indication



1. Proven CVD benefit means it has label indication of reducing CVD events. For GLP-1 RA strongest evidence for liraglutide > semaglutide > exenatide extended release. For SGLT2i evidence modestly stronger for empagliflozin > canagliflozin.
2. Be aware that SGLT2i vary by region and individual agent with regard to indicated level of eGFR for initiation and continued use
3. Both empagliflozin and canagliflozin have shown reduction in HF and reduction in CKD progression in CVOTs
4. Degludec or U100 glargine have demonstrated CVD safety
5. Low dose may be better tolerated though less well studied for CVD effects
6. Choose later generation SU with lower risk of hypoglycemia
7. Degludec / glargine U300 < glargine U100 / detemir < NPH insulin
8. Semaglutide > liraglutide > dulaglutide > exenatide > lixisenatide
9. If no specific comorbidities (i.e., no established CVD, low risk of hypoglycemia, and lower priority to avoid weight gain or no weight-related comorbidities)
10. Consider country- and region-specific cost of drugs. In some countries, TZDs relatively more expensive and DPP-4i relatively cheaper

Figure 15.5.4.2 Algorithm for the management of hyperglycaemia in type 2 diabetes as set out in the 2020 consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) [4]. This algorithm takes account of emerging evidence of the properties of individual agents in relation to particular patient groups, particularly those with renal impairment or cardiovascular diseases. A decision to treat high-risk individuals with a GLP-1 receptor agonist or SGLT2 inhibitor to reduce major adverse cardiovascular events, hospitalisation for heart failure, cardiovascular death or chronic kidney disease (CKD) progression should be considered independently of baseline HbA1c or the individualised HbA1c target. Also, GLP-1 receptor agonists can be considered in patients with type 2 diabetes without established cardiovascular disease but with the presence of specific indicators of high risk. SGLT2 inhibitors are recommended in patients with type 2 diabetes and heart failure, particularly those with heart failure with reduced ejection fraction, as well as patients with CKD (eGFR 30 to ≤60 ml/min/1.73 m^2 or urinary albumin-to-creatinine ratio >30 mg/g. HbA1c target is usually <6.5%-7.5% (<48-58 mmol/mol), but will vary with patient circumstance.
ASCVD, atherosclerotic cardiovascular disease; CKD, chronic kidney disease; CVD, cardiovascular disease; DPP-4i, dipeptidyl peptidase-4 inhibitor; GLP-1RA, glucagon-like peptide-1 receptor agonist; HbA1c, glycated haemoglobin; SGLT-2i, sodium/glucose co-transporter-2 inhibitor; SU, sulphonylurea; TZD, thiazolidinedione.

for use as glucose-lowering agents in some countries (not Europe) (9). The parenterally administered agents are glucagon-like peptide-1 (GLP-1) receptor agonists and insulin: pramlintide injections are indicated for glucose lowering with insulin in some countries (not Europe). The main actions, typical efficacy, and important cautions for agents other than insulin are summarized in Table 15.5.4.1.

Starting Pharmacological Therapy

If hyperglycaemia is severe, or initial lifestyle measures do not achieve desired glycaemic control after one to two months, pharmacological therapy is indicated. Many guidelines advise immediate prescription of metformin at the same time as instituting lifestyle change. In theory, a pharmacological agent might be chosen to address a predominant underlying pathogenic factor, such as insulin resistance or beta-cell dysfunction. In practice, however, it is often difficult to determine the relative impact of different pathogenic factors, and any safe and effective agent to improve glycaemic control may need to be considered. Indeed, therapies may be prescribed on the basis of local protocols, patient preference, or what is left when preferred options are excluded by contraindications or tolerability [4–9]. The suitability of a treatment for an individual requires consideration of the risks of the disease, risks of other agents, and risk of the proposed agent compared with the potential benefits for the current and likely future circumstances of that individual.

Table 15.5.4.1 Synopsis of the main actions, typical reductions in HbA_{1c} effects on body weight, key adverse effects, exclusions, and precautions for glucose-lowering agents other than insulin

Class *with examples*	Main mode of action	HbA_{1c}	Body wt	Main adverse events	Main exclusions and precautions
Oral					
Biguanide Metformin	Counters insulin resistance[a] ↓ hepatic glucose output, ↑ peripheral glucose utilization	↓ ~ 1–2% (~ 11–22 mmol/mol)	–/↓ Mostly weight neutral	GI intolerance Lactic acidosis (rare)	Renal impairment Any hypoxic condition
Sulphonylureas Glibenclamide (glyburide) Gliclazide Glimepiride Glipizide	Increase insulin secretion[b] Direct effect on pancreatic β cells	↓ ~ 1–2% (~ 11–22 mmol/mol)	↑ ~1–4 kg	Hypoglycaemia	Choice restricted by severe liver or renal disease, or porphyria
Meglitinides Nateglinide Repaglinide	Increase insulin secretion[b, c] (Prandial administration: rapid onset, short duration of action). Direct effect on pancreatic β cells	↓ ~ 0.5–1.5% (~ 5–16 mmol/mol)	↑/– ~1–2 kg	Hypoglycaemia (fewer and less severe than sulphonylureas)	Liver or severe renal disease
DPP-4 inhibitors Alogliptin Linagliptin Saxagliptin Sitagliptin Vildagliptin	Potentiate nutrient-induced insulin secretion[b] Inhibit DPP-4 to increase endogenous incretin activity	↓ ~ 0.5–1.5% (~ 5–16 mmol/mol)	– Mostly weight neutral	Possible risk of pancreatitis	Severe renal or liver disease Caution if history of pancreatitis
SGLT-2 inhibitors Canagliflozin Dapagliflozin Empagliflozin Ertugliflozin	Increase renal glucose elimination	↓ ~ 0.5–1.5% (~ 5–16 mmol/mol)	↓ ~2–3 kg	Dehydration, urino-genital tract infections. Atypical ketoacidosis and fractures[d]	Renal disease
Thiazolidinediones Pioglitazone Rosiglitazone	Increase insulin action[a] Stimulate PPARγ , to increase adipogenesis and adjust the glucose-fatty acid cycle	↓ ~ 1–1.5% (~ 11–16 mmol/mol)	↑ ~2–3 kg	Oedema and heart failure, anaemia, Fractures	Cardiac disease, Fluid retention, Severe liver or renal disease
α -glucosidase inhibitors Acarbose Miglitol	Slow carbohydrate digestion[e]	↓ ~ 0.5–1.0% (~ 5–11 mmol/mol)	– Mostly weight neutral	–	Intestinal diseases, Severe kidney disease
Dopamine D2 agonist Bromocriptine	Neurally mediated ↓ hepatic glucose output, other effects not clear[a]	↓ ~ 0.5–0.8% (~ 5–9 mmol/mol)	– Mostly weight neutral	Fibrotic reactions, hypotension	Psychotic disorders
Bile acid sequestrant Colesevelam	Increase incretin secretion and ↓ hepatic glucose output	↓ ~ 0.4–0.8% (~ 4–9 mmol/mol)	– Mostly weight neutral	May increase triglyceride	Intestinal obstruction
Subcutaneous injection					
GLP-1 receptor agonists Dulaglutide Exenatide BD QW Liraglutide Lixisenatide Semaglutide	Potentiate nutrient–induced insulin secretion[b], Suppress glucagon secretion, Satiety effect	↓ ~ 0.5–2% (~ 5–22 mmol/mol)	↓ ~3 kg	Possible risk of pancreatitis	Caution if history of pancreatitis Severe liver, renal or intestinal disease
Amylin analogue Pramlintide[f]	Slow gastric emptying, Suppress glucagon secretion, Satiety effect	↓ ~ 0.3–0.6% (~ 3–6 mmol/mol)	↓ ~1–2 kg	Risk of hypoglycaemia when used with insulin	Gastroparesis, Other medications affecting GI motility Hypoglycaemia unawareness

↑ increase; ↓ decrease; ~ approximately;—no change

Most agents have rarely caused sensitivity reactions.

[a] Efficacy requires presence of circulating insulin.
[b] Efficacy requires presence of a functional β cell mass.
[c] Taken with meals, fewer and less severe hypoglycaemia than sulfonylureas.
[d] Possible increased risk of amputation with canagliflozin.
[e] Taken with meals rich in complex carbohydrate.
[f] Usually used in conjunction with insulin.

Not all blood glucose-lowering agents are available in all countries, and they do not always carry the same indications for use, e.g. pramlintide and rosiglitazone are not available in Europe and gliclazide is not available in the USA, while bromocriptine and colesevelam are not indicated for glucose-lowering in Europe. The same drug may be named differently in different countries, e.g. glibenclamide (Europe) is glyburide in the USA, although the formulation is different. Similarly named drugs may also have different formulations in different countries, e.g. glipizide. The exclusions, precautions, and monitoring requirements may also vary, e.g. pioglitazone is excluded for New York Heart Association (NYHA) categories I–V in Europe, but III-IV in the USA. Maximum doses are sometimes different between countries, e.g. maximum daily dose for metformin is 3000 mg in Europe and 2550 mg in the USA. Many countries introduce local restrictions, e.g. in England the National Institute for Health and Clinical Excellence (NICE) recommends that a GLP-1 receptor agonist should only be continued if its use is associated with a reduction in HbA1c of ≥ 11 mmol/mol [1.0%] and weight loss of ≥3% of initial body weight in 6 months. Also, individual agents within a class can vary in their glucose-lowering efficacy and effects on body weight which affects their suitability for individual patients. For these reasons, it is not possible to be fully prescriptive or inclusive in this chapter, and prescribers are urged to check national and local formulary restrictions before administering pharmacological therapies.

Starting pharmacotherapy assumes that a chosen agent is not precluded by comorbidity, potential interactions with other medications, or incompatibility with lifestyle, and that appropriate information and support are offered to the patient. Note baseline glycaemia and begin therapy with a low dose. A summary of individual agents with pharmacokinetic information is given in Table 15.5.4.2. Titrate up at intervals of one to two weeks for metformin, sulphonylureas, meglitinides, and AGIs. Some GLP-1 receptor agonists (administered by subcutaneous injection) have two or three dosage strengths through which titration can be made, usually two-weekly to monthly. Intervals of at least one month are usually required for a thiazolidinedione. DPP-4 inhibitors and SGLT-2 inhibitors do not usually require titration. Continue to

Table 15.5.4.2 Summary of dose and pharmacokinetic information for individual glucose-lowering agents other than insulins[a]

Class/agent	Dose range[b] mg (except where stated)	Starting dose[b] mg (except where stated)	Duration of action hr[c]	Plasma protein bound	Metabolites[c]	Elimination
Oral						
Biguanide						
Metformin Meformin XR/SR	500–2550/3000	500–850 od 500–1000 od	6–18 12–24[d]	<12%	–	u ~ 100%
Sulphonylureas						
Glibenclamide	2.5–15/20	2.5 od	18–24	> 98%	Active	B > 50%
Gliclazide Gliclazide MR	40–320 30–120	40 od 30 od	12–24 18–24	> 85%	Inactive	u ~ 65%
Glimepimide	1–6	1 od	18–24	> 99%	Active	u ~ 60%
Glipizide	2.5–20	2.5 od/bd	6–18	~ 98%	Inactive	u ~ 70%
Tolbutamide[e]	500–2000	500 od/bd	6–12	> 95%	Inactive	u ~ 100%
Meglitinides						
Nateglinide	180–540	60 tds, ac	1–4	> 97%	Inactive[f]	u ~ 80%
Repaglinide	1–16	0.5 bd/tds, ac	1–6	> 98%	Inactive	B ~ 90%
DPP-4 inhibitors						
Alogliptin	6.25–25	25 od	~ 24	20–30%	Inactive	u ~76%
Linagliptin	5	5 od	~ 24	70–99%	Inactive	B ~95%
Saxagliptin	5	5 od	~ 24	negligible	Active	u ~ 75%
Sitagliptin	100[f]	100 od	~ 24	~ 38%	Inactive	u ~ 79%
Vildagliptin	100	50 bd	~ 24	~ 9%	Inactive[e]	u ~ 85%
SGLT-2 inhibitors						
Canagliflozin	100–300	100 od	~24	~99%	Inactive	u ~33%
Dapagliflozin	5–10	10 od	~24	~91%	Inactive	u ~75%
Empagliflozin	10–25	10 od	~24	~86%	Inactive	u ~54%
Ertugliflozin	5–15	5 od	~24	~95%	Inactive	u ~50%
Thiazolidinediones						
Pioglitazone	15–45	15 od/bd	~ 24	> 99%	Active	B > 60%
Rosiglitazone[g]	2–8	2 od/bd	~ 24	> 99%	Inactive[f]	u ~ 64%
α-glucosidase inhibitors						
Acarbose	150–300	50 bd/tds, ac	~ 6	–%	Inactive	I ~ 50%
Miglitol	75–300	25 bd/tds, ac	~ 6	<4%	–	u > 95%
Dopamine agonist						
Bromocriptine[h]	0.8–4.8	0.8 od	12–24	> 90%	–	B > 94%
Bile acid sequestrant						
Colesevelam[h]	625–4375	1875 bd, ac	12–24	Not absorbed	–	–
Subcutaneous						
GLP-1 receptor agonists						
Dulaglutide	0.75–1.5 mg	0.75 or 1.5 mg QW	>7 days	>99% linked to IgG4 fragment	–	–
Exenatide bd Exenatide QW	5–10 µg 2 mg	5 µg bd, ac 2 mg QW	4–6 >7 days	Negligible	–	–

Table 15.5.4.2 *Continued*

Class/agent	Dose range[b] mg (except where stated)	Starting dose[b] mg (except where stated)	Duration of action hr[c]	Plasma protein bound	Metabolites[c]	Elimination
Liraglutide	0.6–1.8 mg	0.6 mg od	18–>24	>98%	Inactive	–
Lixisenatide	10–20 ug	10 ug od[i]	~18–24	~55%	–	–
Semaglutide	0.25–1 mg	0.25 mg QW	>7 days	>99%	–	–
Amylin analogue						
Pramlintide[j]	60–480 µg	60–120 tds, ac	2–4	~50%	Active	–

~ approximately; u, urine; B, bile; ac, before meals; bd, twice daily; QW, once weekly; tds, three times daily

[a] Availability of some agents, dose range, indications for use, exclusions, and contraindications vary between countries. Prescribers are urged to check national and local formulary restrictions before administering pharmacological therapies.
[b] All doses in mg except exenatide bd, lixisenatide and pramlintide, in µg.
[c] Different formulations affect pharmacokinetics, timing of doses, and duration of action. In this table metabolites are considered to be 'active' if they make a significant contribution to the overall pharmacodynamic effect of the agent.
[d] Slow-release formulations of metformin (XR/SR) have a duration of action up to 24h; recommended maximum dose 2000 mg od.
[e] Not usually used to initiate sulphonylurea therapy.
[f] Some slightly active metabolites.
[g] Discontinued in Europe in 2010.
[h] Not indicated for glucose-lowering in Europe.
[i] Taken before the main meal of the day.
[j] Not available in Europe.

Many single-tablet fixed-dose combinations of two differently acting glucose-lowering agents are available. These may be used as a convenient way to reduce the pill burden. Fixed-ratio combinations of a basal insulin with a GLP-1 receptor agonist in the same injection are also available. These are titrated in the same way as insulin but the presence of the GLP-1 receptor agonist increases the glucose-lowering effect at a lower dose of insulin, off-sets the weight gain associated with insulin therapy and reduces the risk of serious hypoglycaemia.

titrate until the desired level of glycaemic control is achieved. If intolerance supervenes, reduce a dose level and attempt titration again later. If a titration step does not provide any additional benefit, return to the previous dosage and, if significant adverse events are experienced, consider discontinuation and switching to another class of agent. Switching within class is rarely helpful, except when contraindications develop that can be circumvented by different pharmacokinetics. Appropriate monitoring, which may extend beyond glycaemic parameters, and reinforcement of lifestyle compliance, should be undertaken as required.

Many single-tablet fixed-dose combinations of two differently acting oral agents are available. These include combinations of metformin with each of a sulphonylurea, meglitinide, thiazolidinedione, DPP-4 inhibitor, or SGLT-2 inhibitor. There are also fixed-dose combinations of a DPP-4 inhibitor with an SGLT-2 inhibitor. Single-tablet fixed-dose combinations may be used as a convenient way to reduce the pill burden. Bearing in mind the progressive nature of type 2 diabetes, additional therapy to address deteriorating control, or switching therapy to accommodate emerging comorbidity, is an expected part of the treatment process. Rapidly advancing hyperglycaemia in patients with long-standing diabetes, often with unintentional weight loss, is generally an indication of substantial beta-cell failure, signalling the need for insulin replacement therapy. Reassessment of diabetes therapy is required when renal or liver function deteriorates, or patients experience cardiovascular events. Glucose-lowering therapy may require alteration during certain investigations, such as the use of contrast media, and during cultural and religious events during which dietary intake is substantially changed. Fixed-ratio combinations of a basal insulin with a GLP-1 receptor agonist in the same injection are now available: these are titrated in the same way as insulin but the presence of the GLP-1 receptor agonist increases the glucose-lowering effect at a lower dose of insulin, off-sets the weight gain associated with insulin therapy and reduces the risk of serious hypoglycaemia.

Metformin

Three biguanide drugs (metformin, phenformin, buformin) were introduced in the late 1950s. Their origins relate to the glucose-lowering effect of guanidine that was identified in *Galega officinalis*, a plant used to treat diabetes in traditional herbal medicine [10]. Due to a high incidence of lactic acidosis, phenformin and buformin were withdrawn by the late 1970s. Metformin (Figure 15.5.4.3) carries negligible risk of lactic acidosis, if appropriately prescribed, and has since become the preferred first-line agent in most guidelines and the most used oral glucose-lowering agent worldwide.

Actions

Metformin lowers blood glucose concentrations without risk of serious hypoglycaemia and without weight gain. This involves several different actions, mostly serving to counter insulin resistance (Table 15.5.4.3). Some of these actions are achieved through enhanced insulin sensitivity, while others are independent of insulin, but the glucose-lowering efficacy of metformin requires the presence of at least some circulating insulin [9, 11].

Hepatic glucose output is reduced by metformin, particularly the suppression of gluconeogenesis, but with low potency such that the counterregulatory response is not impeded when glucose levels fall below the normal range [12]. Metformin also modestly enhances insulin-stimulated glucose uptake and glycogenesis by skeletal muscle, associated with increased deployment of glucose transporter type 4 protein (GLUT4) in the cell membrane

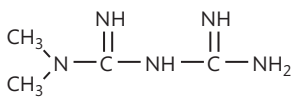


Figure 15.5.4.3 Structure of metformin (dimethyl biguanide).

Table 15.5.4.3 Key insulin-dependent and independent effects of metformin that contribute to its glucose-lowering efficacy

	Insulin dependent	Insulin independent
Liver	↓ Gluconeogenesis	↓ Gluconeogenesis
	↑ IR-TKA	↓ Respiratory chain complex 1
	↓ Glucagon action	↓ mitochondrial GPD
	↑ AMPK	↑ AMPK
	↓ Glycogenolysis	↓ Lipogenesis
	↓ G-6-Pase	
Muscle	↑ Glucose uptake	
	↑ Glycogenesis	
	↑ Glucose oxidation	
	↑ IR-TKA	
	↑ GLUT4 translocation and activation	
Gut		↑ Anaerobic glycolysis
		↓ Respiratory chain complex 1

↑, increase; ↓, decrease; AMPK, adenosine monophosphate-activated protein kinase; IR-TKA, insulin receptor tyrosine kinase activity, G-6-Pase, glucose-6-phosphatase; GLUT4, glucose transporter isoform 4. GPD, glycerophosphate dehydrogenase.

(Figure 15.5.4.4). Anaerobic glucose metabolism in the walls of the intestine is increased by the presence of high concentrations of metformin, probably due to suppression of respiratory chain activity at complex 1. This increases glucose–lactate turnover, which may contribute to futile cycling and increased energy dissipation that helps to prevent weight gain.

Efficacy

The glucose-lowering efficacy of metformin has been affirmed in many studies, typically reducing HbA1c by 1–2% (11–22 mmol/mol). The lack of weight gain and low risk of hypoglycaemia have contributed to the general preference for this agent as initial monotherapy, especially in patients who are overweight or more vulnerable to hypoglycaemia. Metformin does not stimulate insulin secretion and generally reduces basal insulin concentrations in hyperinsulinaemic patients. A small improvement in the lipid profile is often seen and reductions in plasma triglyceride and free fatty acids (FFA) are not uncommon. Evidence for a vasoprotective effect of metformin has also contributed to its positioning as initial therapy. Use of metformin reduced the risk of myocardial infarction (by 39% over 10 years) in the UKPDS independently of the glucose-lowering effect, and this benefit persisted during the post-trial follow-up for more than eight years [3]. Metformin has been shown to improve a range of vascular risk markers (e.g. reducing plasminogen activator inhibitor 1 (PAI-1) and increasing fibrinolysis) and surrogate measures such as reducing carotid intima-media thickness and increasing vasoreactivity. Metformin does not appear to affect blood pressure, although a lowering of blood pressure may coincide with reduced body weight [12, 13].

Metformin provides additive efficacy when combined with most other glucose-lowering agents. When used in conjunction with insulin therapy, metformin can reduce insulin dose requirement, improve glucose profile, and reduce hypoglycaemic episodes and weight gain. Hence, metformin is frequently continued when patients with type 2 diabetes start insulin therapy.

Cautions

The main tolerability issue with metformin is gastrointestinal symptoms. Diarrhoea may limit dose titration in some patients, although it is usually transient, reduced by temporary dose reduction and by taking after meals. Symptoms can also be reduced by switching to a slow-release formulation, but around 5–15% of patients do not tolerate titration of metformin.

Since metformin is eliminated unchanged in the urine, it is important to check renal function before and at intervals during therapy to avoid drug accumulation, as this may predispose to

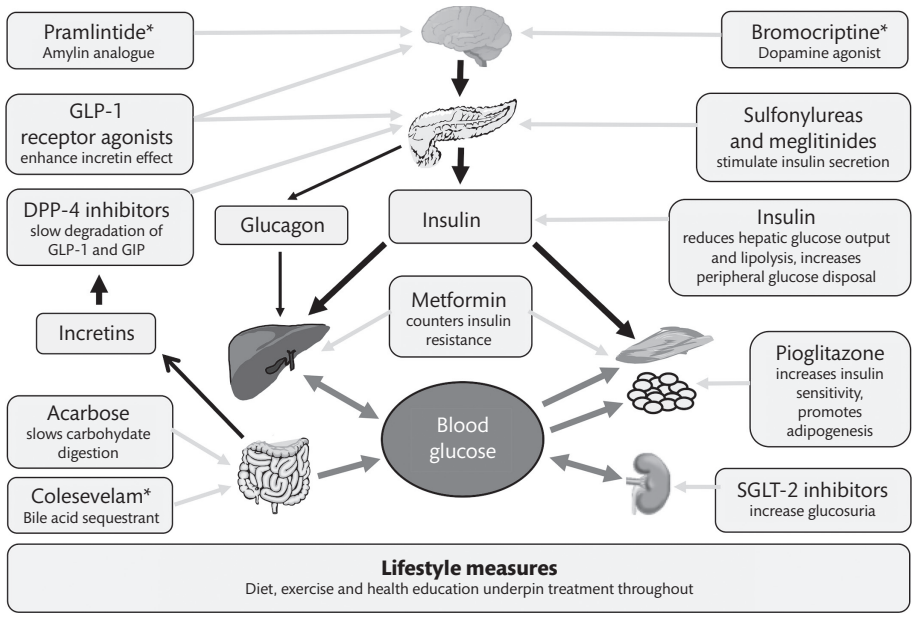


Figure 15.5.4.4 Main sites of action of blood glucose-lowering agents. *Pramlintide, bromocriptine, and colesevelam are not indicated for glucose-lowering in Europe. DPP-4, dipeptidyl peptidase-4; GLP-1, glucagon-like peptide-1; SGLT-2, sodium/glucose cotransporter-2.

lactate accumulation [14]. The renal exclusion criteria, which vary slightly between countries, typically recommend reviewing the dose if the estimated glomerular filtration rate (eGFR) is <60 ml/min/1.73 m², preferably limiting the dose to 1000 mg/day if the eGFR is <60 ml/min/1.73 m², not initiating metformin therapy if the eGFR is <45 ml/min/1.73 m², and stopping metformin if the eGFR is <30 ml/min/1.73 m². Most cases of metformin-associated lactic acidosis are due to failure to respect or recognize deteriorating renal function. Although such cases are rare (variously estimated at 0.03–0.09 per 1000 patient years of treatment), 30–50% are fatal. Treatment should be started immediately, usually with intravenous bicarbonate, and haemodialysis may be helpful to remove excess metformin.

While metformin therapy assists in the prevention of cardiovascular events, it is noted that metformin is contraindicated in conditions of hypoxaemia, which include cardiac or respiratory insufficiency, septicaemia, or hypotension. Also, metformin is contraindicated by alcohol abuse, previous history of metabolic acidosis, or severe cirrhotic liver disease. Nevertheless, with suitable caution, metformin may benefit patients with non-alcoholic fatty liver disease (NAFLD). Metformin is sometimes used to treat the insulin resistance in polycystic ovary syndrome, where it can assist ovulation and conception. Metformin has not been shown to have adverse effects on embryonic or fetal development, and may indeed reduce spontaneous abortion and the risk of maternal gestational diabetes. Since metformin may reduce vitamin B_{12} absorption, it is advised to check haemoglobin occasionally. Metformin should be stopped temporarily during use of contrast media until normal urine flow returns.

Sulphonylureas

Sulphonylureas were developed in the 1950s following an observation that sulfonamide drugs could cause hypoglycaemia. Since the introduction of sulphonylureas, many members of this class have received extensive use worldwide, and they remain the second most used oral glucose-lowering agent (Figure 15.5.4.5).

Actions

The glucose-lowering effect of sulphonylureas is attributed almost entirely to increased insulin secretion resulting from a direct action on the pancreatic beta-cells. Sulphonylureas bind to the so-called sulphonylurea receptor 1 (SUR1), which is part of a transmembrane protein complex with the ATP-sensitive potassium (K_{ATP}) channel [15]. Binding of a sulphonylurea to SUR1 closes the K_{ATP} channel. This prevents K^+ efflux, depolarizes the membrane, and opens adjacent voltage-dependent L-type calcium channels. The resulting increase in cytosolic calcium activates calcium-dependent signalling proteins that regulate the secretion of insulin from pre-formed granules (Figure 15.5.4.6).

Sulphonylureas initiate insulin secretion independently of the glucose concentration. Hence, sulphonylureas can increase insulin secretion at all times, including periods when glucose concentrations are low. They therefore predispose to hypoglycaemia. Sulphonylureas may produce a small decrease of glucagon secretion, and there have been reports of minor glucose-lowering activity that is independent of effects on the pancreas.

Efficacy

The variety of sulphonylureas (see Table 15.5.4.2) with different pharmacokinetic properties enables choice for the onset and duration of action, as well as the mode of metabolism and elimination [16, 17]. Typically, sulphonylureas reduce HbA1c by 1–2% (11–22 mmol/mol) although desensitization may occur along with disease progression to reduce efficacy during prolonged use. Sulphonylureas are effective as monotherapy, and in combination with most other glucose-lowering agents (excepting

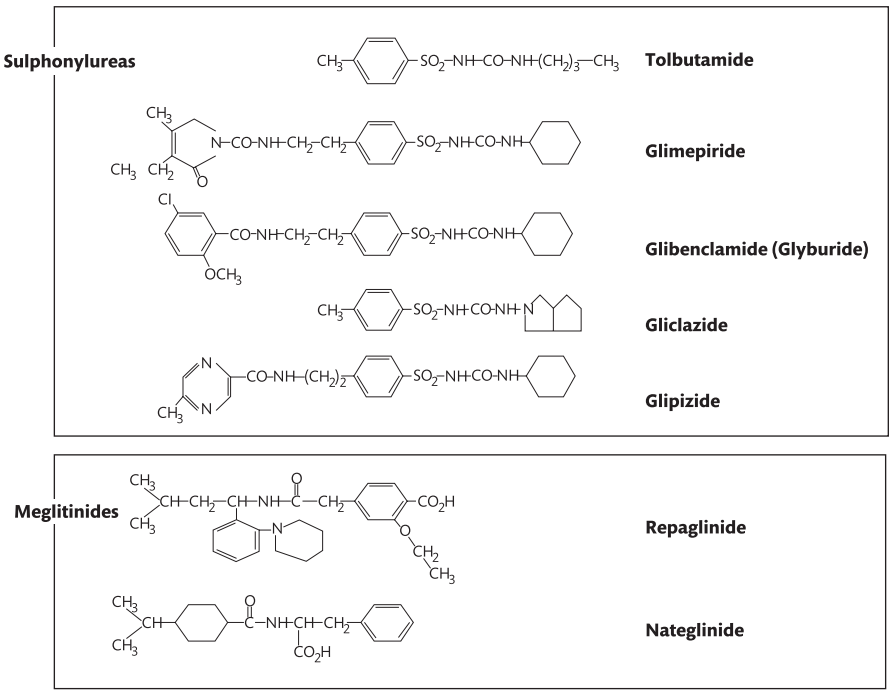


Figure 15.5.4.5 Structures of sulphonylureas and meglitinides.

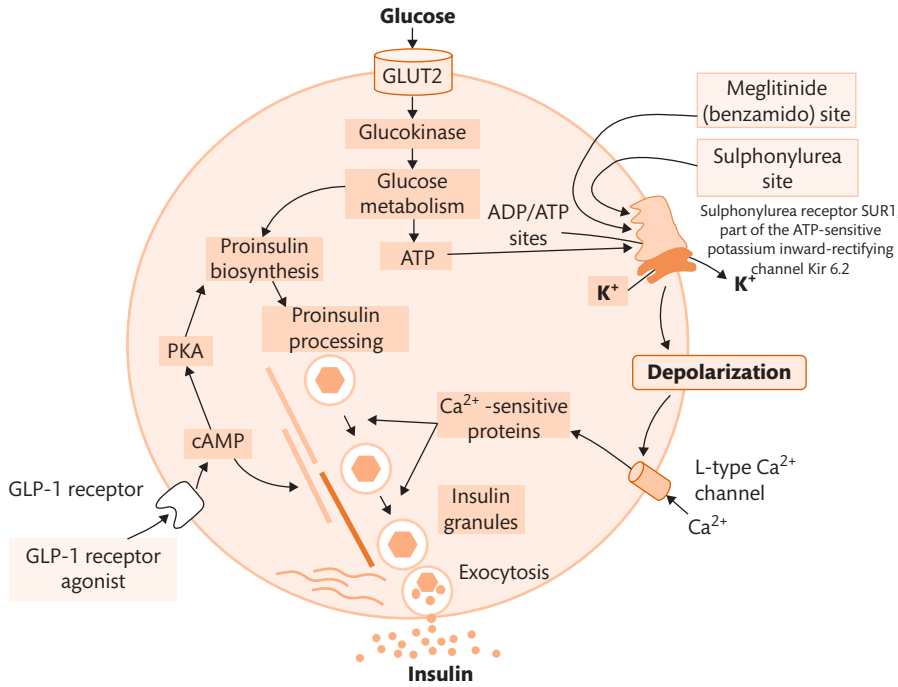


Figure 15.5.4.6 Schematic representation of the control of insulin secretion by a pancreatic beta-cell.

meglitinides, and there is little gain when used in combination with a DPP-4 inhibitor). While there is residual beta-cell function, sulphonylureas can also usefully supplement insulin treatment in type 2 diabetes. By increasing the portal delivery of endogenous insulin, sulphonylureas help to reduce hepatic glucose output, especially during meals: this complements the predominantly peripheral effect of subcutaneously administered insulin. Sulphonylureas have not significantly affected lipids or blood pressure in most studies.

Cautions

Sulphonylureas are prone to cause weight gain, usually 1–4 kg that stabilizes by about 6 months [16, 17]. This is probably due to the anabolic effect of extra insulin and the reduced loss of glucose in the urine. Tolerability is generally good and facilitated by the choice of sulphonylureas available.

Hypoglycaemia is the main serious adverse event associated with sulphonylurea therapy, requiring vigilant glucose monitoring by patients who drive and especial consideration of vulnerability in the elderly and frail. It is important to start with a low dose and titrate in concert with blood glucose monitoring, harmonizing drug therapy with other aspects of lifestyle. It is also necessary to educate patients to recognize and respond to early warning symptoms of hypoglycaemia, and to prevent hypoglycaemia wherever possible. One or more episodes of hypoglycaemia are likely to occur in up to 20% of sulphonylurea-treated patients annually, although severe episodes occur in only about 1% of patients and the mortality risk has been reported at 0.014–0.033 per 1000 patient years. Severe sulphonylurea-associated hypoglycaemia may initially require intravenous glucose, and glucose supplementation should be continued with appropriate monitoring for more than 24 hours to prevent a recurrence, which can occur if longer-acting or more slowly metabolized agents are involved. Treatment of hypoglycaemia with glucagon is discouraged in type 2 diabetes patients, as this is itself an insulin secretagogue.

Meglitinides

Meglitinides are short-acting (prandial) insulin releasers developed after the observation that the benzamido compound meglitinide, which is a component of some sulphonylurea molecules, can stimulate insulin secretion. Two meglitinide agents (repaglinide and nateglinide) were introduced in the late 1990s/early 2000s (see Figure 15.5.4.5).

Actions

Meglitinides bind with the benzamido site on the beta-cell SUR1 receptor [15], setting in motion the same sequence of events described for stimulation of insulin secretion by sulphonylureas (see Figure 15.5.4.6). The main differences are pharmacokinetic: meglitinides are rapidly absorbed, rapidly acting, but with a shorter duration of action than sulphonylureas, making them suitable for administration to coincide with the period of meal digestion.

Efficacy

The main application of meglitinides is to increase prandial insulin secretion to coincide with demand and reduce postprandial glucose excursions [18]. There is typically a carry-over effect to reduce basal glycaemia, but the reduction in HbA1c is usually slightly less than with sulphonylureas. Meglitinides are appropriate for individuals with irregular lifestyles with unpredictable or missed meals. Meglitinides are conveniently used with metformin or a thiazolidinedione, and to supplement insulin therapy in type 2 diabetes.

Cautions

Although meglitinides can precipitate hypoglycaemia, such episodes are fewer and less severe than with sulphonylureas. Weight gain is generally less than with sulphonylureas, although there is

little noticeable effect when switching between a sulphonylurea and meglitinide, or when combined with metformin. A drug interaction between repaglinide and gemfibrozil should be noted. Also multiple daily dosing means that meglitinides are not a practical option for many patients.

GLP-1 Receptor Agonists

GLP-1 receptor agonists enhance the incretin effect which is the extra insulin response to nutrients generated during meal digestion beyond that caused by the increase in circulating nutrients alone. The incretin effect is produced by the hormone GLP-1 and other endocrine (e.g. glucose-dependent insulinotropic polypeptide; GIP) and neural pathways from the intestinal tract [19]. The incretin effect is reduced in type 2 diabetes, but the capacity of GLP-1 to potentiate nutrient-induced insulin secretion (and suppress excess glucagon secretion) is largely retained, hence the therapeutic application of GLP-1 to improve glycaemic control in type 2 diabetes. However, GLP-1 is rapidly degraded in the circulation (half-life, <2 minutes) by the enzyme DPP-4 which breaks peptides with an Ala or Pro residue at position N2 (GLP-1 has an Ala residue in this position). Accordingly, several analogues and formulations of GLP-1 have been developed to avoid or reduce degradation by DPP-4 in order to sustain the incretin effect (Figure 15.5.4.7).

Exenatide, introduced in 2005, was the first GLP-1 receptor agonist. It is synthetic exendin-4, a peptide identified in the saliva of the Gila monster lizard *Heloderma suspectum*. Exenatide has 53% homology with human GLP-1, and lixisenatide is an analogue of exendin. Other GLP-1 receptor agonists have a much closer structure to human GLP-1 with only one to three amino acid residues altered (notably alterations in the N2 position) and/or a fatty acid side chain added for attachment to albumin: these are designed to prevent degradation by DPP-4 and enable formulations with different durations of action.

Actions

GLP-1 is secreted by L cells located throughout the intestine but mostly in the ileum and colon. GLP-1 interacts with G-protein-coupled receptors on the pancreatic beta-cells, potentiating distal steps in the nutrient-induced insulin secretion pathway as well as activating protein kinase A to promote insulin biosynthesis (Figure 15.5.4.6). Additionally, GLP-1 interacts with receptors in the portal system to trigger neural reflexes that affect other pancreatic, gastrointestinal, and metabolic functions. These include the suppression of excess glucagon secretion by pancreatic alpha-cells, as well as slowing gastric emptying, and augmenting meal-induced satiety. GLP-1 can increase pancreatic beta-cell mass in animal models, but an effect on beta-cell mass in human type 2 diabetes has yet to be confirmed, although the durable efficacy of GLP-1 receptor agonists has been encouraging. Many further physiological effects of GLP-1 have been reported which impact cardiovascular, renal, neural and inflammatory activity [19–21].

The pharmacokinetic properties of the various GLP-1 receptor agonists include preparations with different affinities for the GLP-1 receptor and different durations of action which afford different administration schedules (Table 15.5.4.4). As all GLP-1 receptor agonists bind to the GLP-1 receptors to produce the glucose-lowering effect, it is likely that each will exert additional GLP-1 receptor-mediated effects.

Efficacy

GLP-1 receptor agonists typically reduce HbA1c by 1–2% (11–22 mmol/mol), associated with substantial reductions (by ~4 mmol/l or 72 mg/dl) in postprandial glycaemia. Durability of the glucose-lowering effect has been observed over several years, mostly as add-on to metformin. Adding a GLP-1 receptor agonist to other oral glucose-lowering agents as dual or triple therapy is similarly effective provided there is adequate residual beta-cell capacity. Use of a GLP-1 receptor agonist is usually associated with a reduction in body weight of about 3 kg over 6–12 months. However there is considerable heterogeneity in the glucose-lowering and

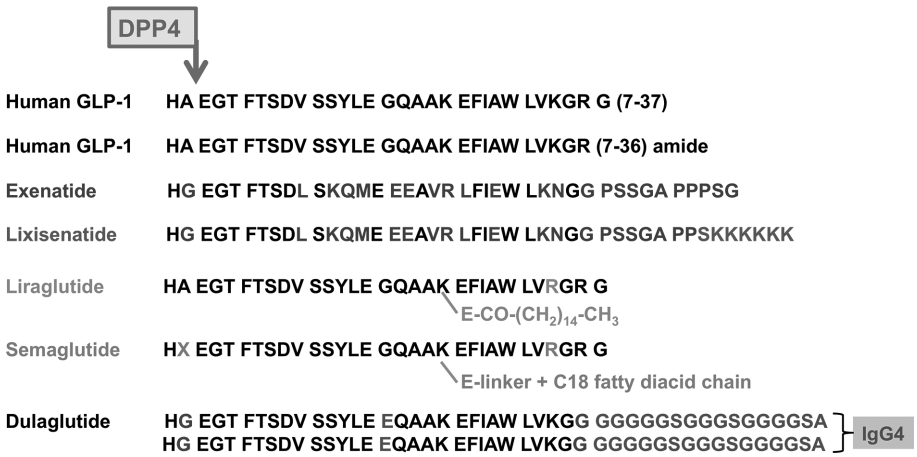


Figure 15.5.4.7 Structures of the incretin hormone GLP-1 (glucagon-like peptide-1) and the GLP-1 receptor agonists exenatide, liraglutide, lixisenatide, dulaglutide, and semaglutide. Dulaglutide comprises two GLP-1 analogue molecules covalently linked to a modified human immunoglobulin G4 (IgG4) heavy chain fragment (Fc) by a small peptide linker.

[Semaglutide is additionally available as an oral formulation.] X, alpha-aminoisobutyric acid. Note that the N2 position of the biologically active GLP-1 peptides 7–37 and 7–36 amide is the same as the N8 position of the fully translated sequence.

Table 15.5.4.4 Characteristics of GLP-1 receptor agonists

Drug	Brand	Structure (sequence homology)	IC_{50} (nM)	Dose	Admin	C_{max}	T_{max}	Half-life
Exenatide *twice daily*	Byetta	Exendin-4 (53%)	0.55	5 µg, 10 µg	BD	~160–250 pg/ml	2–3 h	~3.5 h
Liraglutide *once daily*	Victoza	GLP-1 (97%)	0.11	0.6 mg, 1.2 mg, 1.8 mg	OD	~34 nmol/L (1.8 mg dose)	10–14 h	11.6–13 h
Exenatide QW *once weekly*	Bydureon	Exendin-4 (53%)	0.55	2 mg	QW	Steady state ~300 pg/ml	2–6 weeks at steady state	Unspecified
Lixisenatide *once daily*	Lyxumia	Exendin-4 plus extra Lys residues	1.4	20 µg	OD	~190	1.2–2.5 h	2–4 h
Dulaglutide *once weekly*	Trulicity	GLP-1 (91%)	?	0.75 mg, 1.5 mg	QW	114 ng/ml (1.5 mg dose)	2–4 weeks at steady state	~4.7 days
Semaglutide *once weekly*	Ozempic	GLP-1 (94%)	?	0.5 mg, 1.0 mg	QW	~30 nmol/L (1.0 mg dose)	1–3 days	~6.2 days

Admin, administration; BD, twice daily; C_{max}, peak serum concentration; GLP-1, glucagon-like peptide-1; GLP-1RA, GLP-1 receptor agonist; OD, once daily; QW, once weekly; T_{max}, time to peak serum concentration.

weight-lowering efficacy of individual agents and in the responsiveness of individuals. There is also debate regarding the effectiveness of twice-daily versus once-daily versus once-weekly administration, but head-to-head studies have reported generally similar effects over 6–12 months [9]. For reasons unknown, a minority of individuals do not respond to GLP-1 receptor agonists, even when there is evident beta-cell capacity remaining. GLP-1 receptor agonists can be used effectively in combination with basal (long-acting) insulin, and there are fixed-ratio combinations of a GLP-1 receptor agonist with a basal insulin in the same subcutaneous injection which can be titrated in the same way as insulin alone. A fixed-ratio combination of liraglutide with insulin degludec (IDegLira) confers greater blood glucose-lowering at a lower dose of insulin and with less weight gain and no increased risk of hypoglycaemia compared with degludec alone. A fixed-ratio combination of lixisenatide with insulin glargine offers similar efficacy.

Most studies with GLP-1 receptor agonists have noted a small reduction of systolic blood pressure, although there is often a small increase in heart rate. Large prospective outcome studies have reported either no significant changes or beneficial reductions in a composite of major adverse cardiac events (CV-related deaths, non-fatal MI and stroke) during use of a GLP-1 receptor agonist [22]. Accordingly, some guidelines have offered recommendations for the use of particular GLP-1 receptor agonists in patients with cardiovascular diseases or at high risk of cardiovascular disease [4]. Lower rates of development and progression of albuminuria have also been noted in some long-term studies with GLP-1 receptor agonists.

Cautions

A limiting factor for the use of GLP-1 receptor agonists is initial nausea, presumed to reflect a reduced rate of gastric emptying which may affect the absorption of concomitant oral medications. The nausea is usually transient and ameliorated by introducing therapy at a low dose for several weeks. Administration to patients with severe gastrointestinal disease, including gastroparesis, should be avoided. GLP-1 receptor agonists carry little risk of severe hypoglycaemia when used as monotherapy or combination with agents that do not raise insulin concentrations, but the risk of hypoglycaemia should be appreciated when used in combination with insulin and insulin-releasing agents. Dose reduction or avoidance should be considered in patients with moderate to severe renal disease, although liraglutide, semaglutide, and dulaglutide can be used in renal patients (except end-stage renal disease) without dose adjustment. It is advised to discontinue these agents in pregnancy. Although some patients develop antibodies to exenatide, these do not usually have a noticeable effect on efficacy, and reactions at the injection site are seldom problematic. A possible increased risk of pancreatitis in patients receiving GLP-1 receptor agonists remains unclear, but it is advised that these agents are discontinued if pancreatitis is suspected or diagnosed. Caution is also suggested in patients with thyroid disease.

DPP-4 Inhibitors

Several dipeptidyl peptidase-4 (DPP-4) inhibitors (also terms gliptins) have become available since 2007, namely sitagliptin, vildagliptin, saxagliptin, linagliptin and alogliptin which are taken once daily (or twice-daily for vildagliptin) (**Figure 15.5.4.8**). In Japan there are long-acting once weekly DPP-4 inhibitors (e.g. omarigliptin and trelagliptin). DPP-4 inhibitors act by increasing the endogenous incretin effect.

Actions

Each of the DPP-4 inhibitors produces almost complete inhibition of DPP-4 activity for 12–18 hours, and >80% inhibition for the remainder of a 24h period. This slows the proteolytic degradation of the two main incretin hormones GLP-1 and GIP. The rise in active GLP-1 concentrations is about 2–3-fold compared with >10 fold increased concentration after administration of a GLP-1 receptor agonist. Thus, DPP-4 inhibitors exert similar effects to GLP-1 receptor agonists but to a lesser extent. Adequate residual beta-cell capacity is required for DPP-4 inhibitors to potentiate

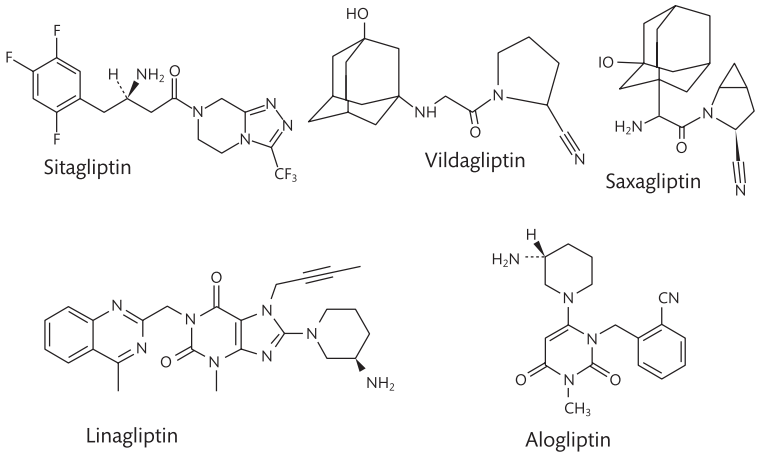


Figure 15.5.4.8 Structures of the dipeptidyl peptidase-4 (DPP-4) inhibitors sitagliptin, vildagliptin, saxagliptin, linagliptin, and alogliptin.

nutrient-induced insulin release and they may also reduce excess glucagon, although GIP tends to raise glucagon which diminishes the glucagon-lowering effect of GLP-1. Neural and gastrointestinal effects of DPP-4 inhibitors are also less marked than with GLP-1 receptor agonists [19, 23, 24].

Efficacy

As monotherapy, DPP-4 inhibitors reduce postprandial glucose concentrations (by about 3 mmol/l or 54 mg/dl) and basal glycaemia (by about 1–1.5 mmol/l or 18–27 mg/dl). This typically achieves a reduction in HbA1c of 0.7–1% (8–11 mmol/mol). Because the increase in insulin secretion is glucose dependent, the release of extra insulin does not occur when glucose levels fall to normal basal values, thereby reducing the risk of interprandial hypoglycaemia. DPP-4 inhibitors do not cause weight gain and they may assist a small amount of weight loss: they seldom reduce the rate of gastric emptying to a clinically significant extent (so nausea is not a problem), and they do not produce a clinically demonstrable satiety effect. DPP-4 inhibitors can give added efficacy when used in combination with agents that improve insulin sensitivity and with SGLT-2 inhibitors. They are approved for use with other glucose-lowering agents such as sulphonylureas, meglitinides, GLP-1 receptor agonists, and insulin, although the extra effects are generally modest with these combinations.

Although DPP-4 inhibitors might theoretically slow the degradation of some vaso-active peptides, large prospective trials with saxagliptin, alogliptin and sitagliptin have confirmed the cardiovascular safety of these agents in type 2 diabetes patients who are at high cardiovascular risk. However these trials did not identify cardiovascular benefits, and a small increase in heart failure hospitalization was noted with saxagliptin. Use in renal disease requires dose reduction of DPP-4 inhibitors except linagliptin which is eliminated mostly via the bile. A reduction in some markers of kidney function suggests a possible slower progression of diabetic renal disease with some DPP-4 inhibitors [23].

Cautions

DPP-4 inhibitors have shown a commendable safety profile: several large studies have indicated a small increased risk of pancreatitis, and DPP-4 inhibitors should be discontinued if pancreatitis is suspected or diagnosed. The DPP-4 enzyme is also the CD26 T-cell activator, but inhibition of the dipeptidase activity does not interfere with the immune function of the molecule, and DPP-4 inhibitors have not been reported to exert any immunological effects.

Sodium/Glucose Transporter-2 Inhibitors

Sodium/glucose transporter-2 (SGLT-2) inhibitors were introduced in 2013: they lower blood glucose by increased elimination of glucose in the urine (glucosuria). The development of SGLT-2 inhibitors arose from the glucosuric effect of phlorizin, a natural phenolic glycoside isolated from apple tree bark [25]. Current members of the class in Europe and North America are canagliflozin, dapagliflozin, empagliflozin and ertugliflozin (Figure 15.5.4.9). Further SGLT-2 inhibitors (ipragliflozin, luseogliflozin, and tofogliflozin) are available in Japan. A dual SGLT-1/SGLT-2 inhibitor (sotagliflozin) is advanced in development.

Actions

SGLT-2 in the first segment of renal proximal tubules is responsible for reabsorption of most of the glucose filtered by the kidney. SGLT-1 in the third segment of renal proximal tubules also contributes to glucose reabsorption, and SGLT-1 in the brush border of enterocytes is responsible for the intestinal absorption of glucose (Figure 15.5.4.10). Competitive inhibition of SGLT-2 reduces the renal threshold for glucose, and in individuals with substantial hyperglycaemia (e.g. blood glucose concentrations >10 mmol/L) SGLT-2 inhibitors can eliminate 80–100 g of glucose/day into the urine. Depending on its specificity and kinetics an SGLT-2 inhibitor may slightly reduce the activity of renal SGLT-1 transporters. The glucosuric effect is self-limiting such that at low glucose concentrations entering the renal filtrate (e.g. <4 mmol/l), 'unblocked' transporters can reabsorb most of the glucose. Also SGLT-2 inhibitors do not compromise the counterregulatory mechanism, making the risk of hypoglycaemia unlikely when these agents are used alone or in combination with

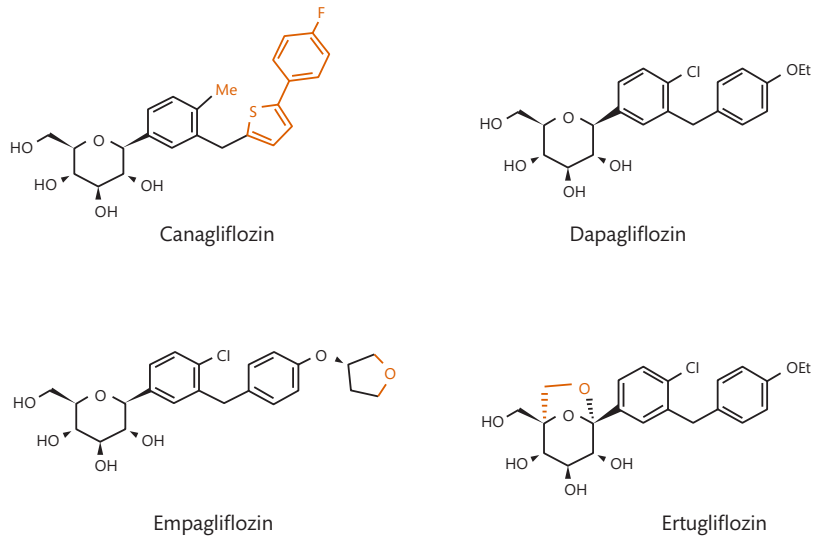


Figure 15.5.4.9 Structures of the SGLT-2 inhibitors canagliflozin, dapagliflozin, empagliflozin, and ertugliflozin.

other low risk agents. The effects of SGLT-2 inhibitors are independent of insulin: hence their efficacy is not affected by the extent of insulin resistance or insulinopenia, but it is appreciated that sufficient insulin must be present to address the underlying endocrine pathogenesis. The caloric loss associated with the glucosuria assists weight reduction, and the mild osmotic diuresis generated by the glucosuria is likely to contribute to a reduction of systolic blood pressure with SGLT-2 inhibitors. Lowering of blood glucose with reduced glucotoxicity during use of an SGLT-2 inhibitor can improve insulin sensitivity and glucose sensitivity of pancreatic beta-cells, assisting the extent and durability of the glucose-lowering effect. The unique mode of action of SGLT-2 inhibitors enables their use in combination with any other class of glucose-lowering agent.

The inhibition of SGLT-2 in the proximal tubules is anticipated to allow more sodium into the loop of Henle and to deliver more sodium to the macula densa. Through the tubular-glomerular feedback mechanism this would constrict the afferent glomerular arterioles, reducing intraglomerular pressure and reducing the risk of hyperfiltration. This could account for emerging evidence

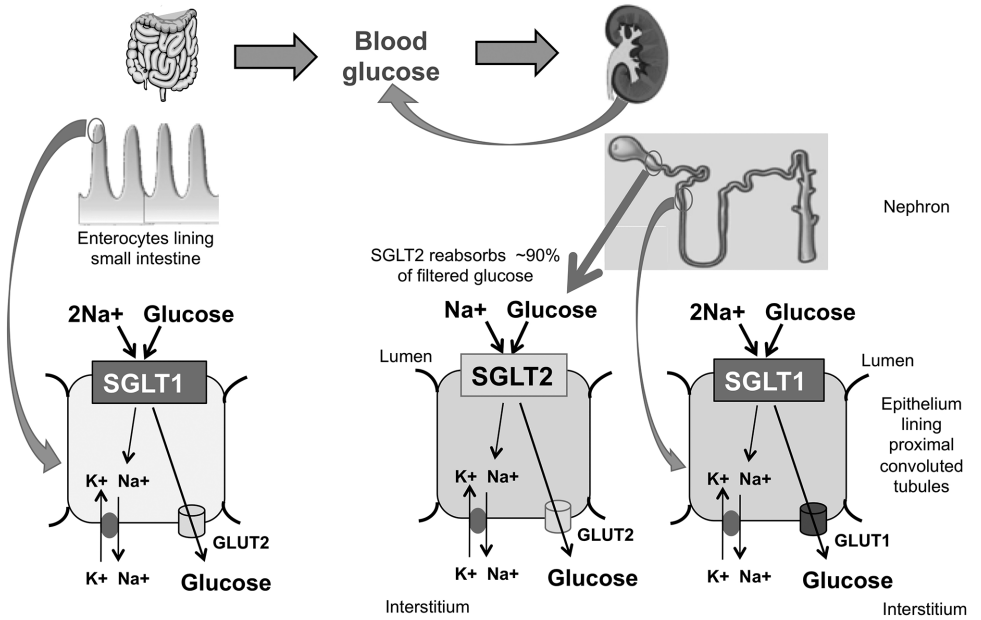


Figure 15.5.4.10 SGLT-2 inhibitors act mainly by competitive inhibition of SGLT-2 transporters in the renal proximal tubules and possibly by slight inhibition of renal SGLT-1 transporters. SGLT-2 transporters are low-affinity, high-capacity glucose transporters that normally reabsorb most of the filtered glucose.

Reproduced with permission from Bailey CJ, Day C. *Br J Diabetes Vasc Dis* 2010; 10 193–9.

that SGLT-2 inhibitors exert a renoprotective effect in diabetes by slowing the rate of long-term decline in glomerular filtration and slowing the onset and progression of albuminuria.

Efficacy

The prompt glucosuric effect of SGLT-2 inhibitors causes a rapid reduction of blood glucose that is approximately proportional to the extent of hyperglycaemia provided there is an adequate glomerular filtration rate (GFR). Reductions in HbA1c are similar whether an SGLT-2 inhibitor is used as monotherapy or add-on to any other glucose-lowering agent(s) as a double or triple combination including insulin. From a baseline HbA1c of around 8% (64 mmol/mol) the reduction in HbA1c is typically ~0.7–1.0% (8–11 mmol/mol) and durability has been observed over several years provided kidney function is maintained [9, 25–27]. Risk of hypoglycaemia is negligible unless used with an agent that increases circulating insulin. Body weight reductions with SGLT-2 inhibitors are usually in the order of 2–3 kg after 6–12 months, which is less than might be predicted in the long term from the amount of glucosuria. This is possibly due to an increase in metabolic efficiency and to a compensatory increase in feeding. In 'real world' observational studies the reductions in HbA1c have been similar to the prospective controlled trials, although the weight reductions have generally been >3 kg. Reduction in systolic blood pressure are typically 2–5 mmHg and there are reductions in uric acid.

Cardiovascular outcome trials to-date have noted a reduced composite of major adverse cardiac events (CV-related deaths, non-fatal MI, and stroke) and consistent demonstration of fewer hospitalizations for heart failure during use of SGLT-2 inhibitors. The rapid emergence of these effects suggests that they are unlikely to reflect weight loss or reduced atherogenesis and may be due in part to reduced plasma volume and blood pressure. It is also possible that SGLT-2 inhibitors can act directly on the myocardium to improve energy metabolism and reduce activity of the Na^+/H^+ exchanger.

Cautions

Though labels vary between region and compounds it is generally recommended that SGLT-2 inhibitors should not be initiated in patients with eGFR <60 ml/min/1.73 m^2, although patients who are already receiving and tolerating an SGLT-2 inhibitor can be continued down to an eGFR of 45 ml/min/1.73 m^2. An initial drop in glomerular filtration often occurs after introduction of an SGLT-2 inhibitor, but this will usually return to pretreatment levels by 3–6 months. Indeed trials have consistently shown reductions in a renal composite end point with SGLT-2 inhibitors [28–30]. The main tolerability concern with SGLT-2 inhibitors is genital infection, particularly women with a history of these infections, and there is a small increased risk of urinary tract infections. Potential for volume depletion, dehydration, and hypotension should be appreciated and patients are usefully reminded to ensure that fluid intake is adequate, especially in hot climates. Those who regularly self-monitor blood glucose, particularly if insulin-treated, may equate a decrease in blood glucose with a lesser need for other glucose-lowering therapies. Any reduction of insulin should be cautious to avoid underinsulinization which can lead to increased lipolysis, increased ketone production, and risk of an atypical form of ketoacidosis that presents without severe hyperglycaemia and sometimes with euglycaemia. Slightly increased risks of fractures and possibly of amputation in individuals with severe peripheral vascular disease have not been excluded.

Thiazolidinediones

TZDs were introduced at the turn of the century. Of three original class members (troglitazone, rosiglitazone and pioglitazone), troglitazone was withdrawn globally due to idiosyncratic hepatotoxicity, and rosiglitazone was withdrawn in Europe and given restricted use in some countries due to reports of increased cardiovascular risk. Pioglitazone remains in use in most countries but was withdrawn in some European countries due to reports of a possible association with bladder cancer (Figure 15.5.4.11) [9].

Actions

TZDs produce most of their antidiabetic activity by activation of a nuclear receptor, peroxisome proliferator-activated receptor-gamma (PPARγ). This receptor forms a heterodimeric complex with the retinoid X receptor (RXR), and, when a TZD and retinoic acid are bound to the complex, repressors are shed and coactivators recruited [31]. The activated receptor locates a nucleotide sequence termed the 'peroxisome proliferator response element' within the promoter regions of a range of genes. Many of these genes are insulin sensitive, and others promote complementary effects on glucose and lipid metabolism that improve insulin sensitivity. PPARγ is highly expressed in adipose tissue, and modestly expressed in other key tissues involved in nutrient homeostasis. Stimulation of PPARγ promotes adipogenesis through the differentiation of

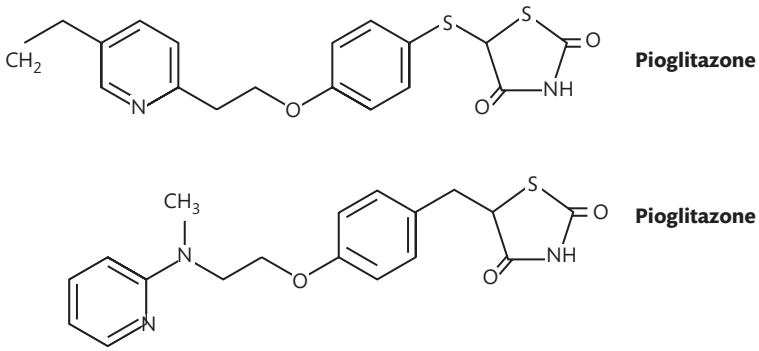


Figure 15.5.4.11 Structure of the thiazolidinediones pioglitazone and rosiglitazone.

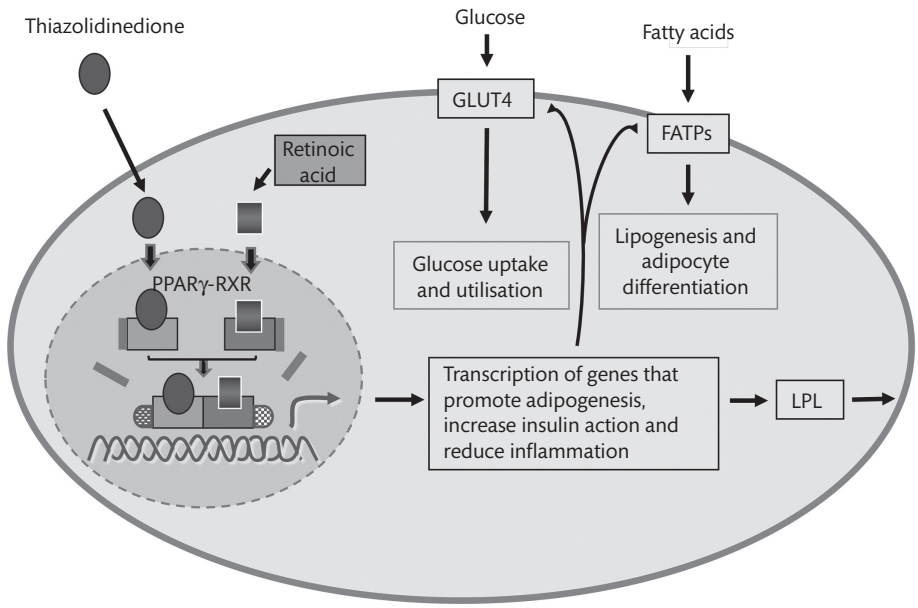


Figure 15.5.4.12 Cellular mode of action of thiazolidinediones via activation of peroxisome proliferator-activated receptor-gamma (PPARγ). FATPs, fatty acid transport proteins; GLUT4, glucose transporter isoform 4; LPL, lipoprotein lipase; RXR, retinoid X receptor.

preadipocytes into new, small, insulin-sensitive adipocytes, mainly in subcutaneous adipose depots (Figure 15.5.4.12). These take up fatty acids, decrease circulating fatty acid concentrations and rebalance the glucose-fatty acid Randle cycle to favour glucose utilization, reducing fatty acid availability as an energy source for hepatic gluconeogenesis. TZDs also reduce the accumulation of lipids in muscle and liver, and increase glucose uptake into skeletal muscle and adipose tissue through increased availability of GLUT4.

Additional and diverse ('pleiotropic') effects of TZDs include reduced production of several pro-inflammatory cytokines by adipose tissue, notably tumour necrosis factor-alpha (TNFα). TZDs also increase adiponectin production, improve vasoreactivity, and tend to reduce blood pressure, with beneficial effects on a range of cardiovascular risk factors and markers.

Efficacy

TZDs exert a slowly generated glucose-lowering effect, typically decreasing HbA1c by ~1% (11 mmol/mol). This generally requires 2–4 months to achieve full efficacy, reflecting the predominantly genomic mechanism of action: hence dose titration may be a prolonged process [32, 33]. However, durability of action has been longer than with sulfonylureas. TZDs do not stimulate insulin secretion, and they do not cause hypoglycaemia. They can be used as monotherapy, or in combination with most other types of glucose-lowering agents. Pioglitazone exerts some PPARα agonism and generally reduces triglycerides and often increases HDL. It also increases the proportion of larger and more buoyant (less atherogenic) LDL particles, and long-term trials have noted a reduction in risk of MI and recurrence of stroke.

Cautions

Use of TZDs is typically associated with weight gain, often more than with sulphonylureas. This is mainly an increase in subcutaneous adipose tissue, but rapid and marked weight gain may be due to fluid retention, and development of oedema after initiation of TZD therapy is usually accompanied by a modest dilutional reduction in the circulating haemoglobin. Although TZDs reduce several markers of cardiovascular disease, they are contraindicated in individuals with manifest cardiovascular disease. TZDs carry a risk of increased heart failure and their use is contraindicated by New York Heart Association I-IV (Europe) or III-IV (USA). Impaired liver function is also a contraindication, although pioglitazone has been used cautiously to reduce hepatic triglycerides in NAFLD. Improved insulin sensitivity with TZDs can restore ovulation in polycystic ovary syndrome (PCOS), but TZDs should be discontinued in pregnancy. Risk of bone fracture is increased with a TZD, especially in postmenopausal women, and individuals with low bone density or osteoporotic disease are not recommended for a TZD. Pioglitazone is not recommended in patients with current or previous bladder cancer, pending ongoing debate over a possible link.

α-Glucosidase Inhibitors

Following evidence that metabolites from cultures of actinomycete fungi could inhibit cell surface glucosidase enzymes, specific AGIs were developed as glucose-lowering drugs. The first, acarbose, was introduced in the early 1990s, followed by miglitol. Another, voglibose, is available in some countries (Figure 15.5.4.13).

Actions

AGIs bind to α-glucosidase enzymes with high affinity, acting competitively to prevent the binding and cleavage of disaccharides and oligosaccharides into monosaccharides [34]. This impedes the final step of carbohydrate digestion. AGIs do not specifically affect the glucose absorption process, but the delay in carbohydrate digestion can defer and slow the appearance of glucose in the circulation, significantly reducing the postprandial glycaemic excursion when there is substantial complex carbohydrate in the diet. The

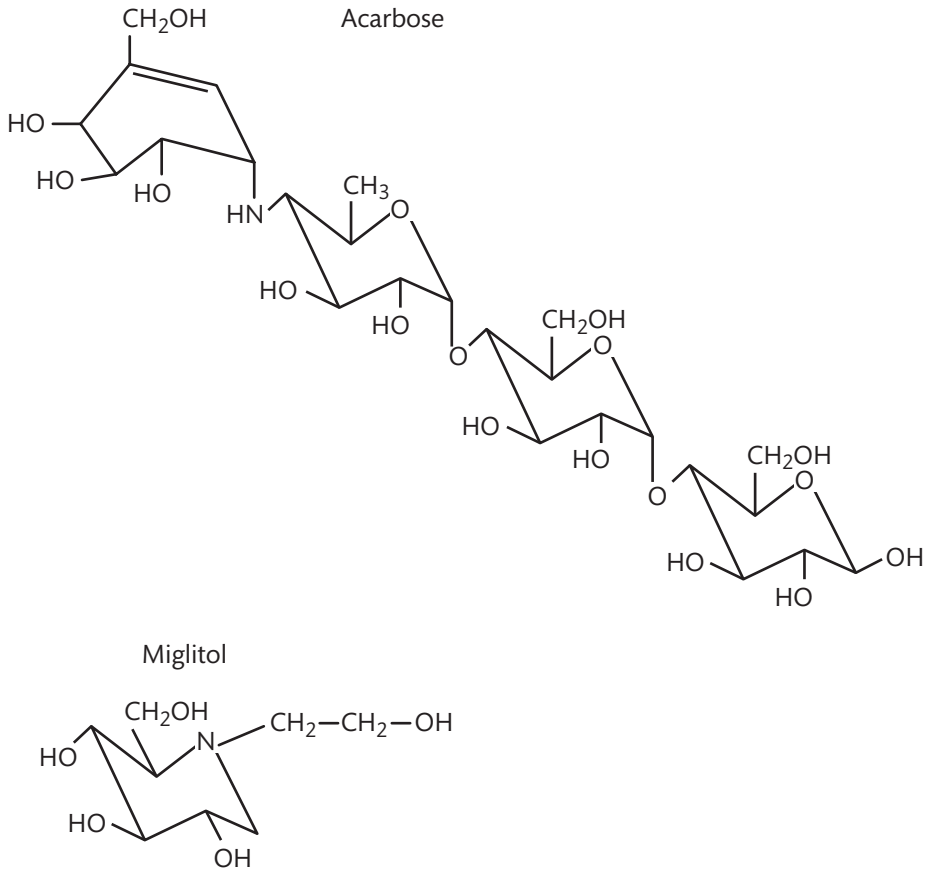


Figure 15.5.4.13 Structures of the α -glucosidase inhibitors acarbose, miglitol, and voglibose.

activity profiles of acarbose and miglitol vary slightly: acarbose has a greater affinity for glucoamylase than sucrase, whereas miglitol has a stronger inhibitory effect on sucrase.

Efficacy

AGIs act mainly to reduce postprandial hyperglycaemia, and their effects are generally modest, reducing HbA1c by about 0.5% (5 mmol/mol), although this can be greater in individuals consuming mainly a carbohydrate-rich diet. Usefully, an AGI can be added in to any other therapy, and this is not usually associated with risk of hypoglycaemia. Indeed, by extending the duration of meal digestion, AGIs can reduce the risk of interprandial hypoglycaemia in individuals receiving insulin or an insulin secretion initiator. Also, AGIs do not cause weight gain, and some individuals may show a reduced postprandial triglyceride profile. It has been suggested that, by extending carbohydrate digestion to more distal regions of the intestine, AGIs might increase GLP-1 secretion. However, postprandial insulin concentrations are commonly reduced by an AGI, commensurate with the lowering of postprandial glycaemia. There have been reports of fewer cardiovascular events in patients receiving an AGI, but a large long-term prospective trial in people with impaired glucose tolerance found no significant effect on major adverse cardiovascular events [35].

Cautions

AGIs are prone to cause some carbohydrate malabsorption. Undigested carbohydrate passing into the large bowel is fermented and can create considerable flatulence. Thus, AGIs should be given with appropriate meals and titrated slowly to minimize this effect. AGIs are contraindicated in individuals with gastrointestinal conditions: caution is needed with any agents affecting gut motility, and liver function should be checked in individuals receiving a high dose of acarbose, since increased alanine transaminase levels have been noted very occasionally.

Pramlintide

Amylin (islet amyloid polypeptide, IAPP) is synthesized and cosecreted with insulin from the pancreatic beta-cells. Precipitates of IAPP within the islets in type 2 diabetes have been ascribed a pathogenic role in beta-cell demise, although the extent of involvement remains uncertain. Amylin exerts central effects that independently affect nutrient metabolism: to retain these effects without detrimental effects on the islets, a non-precipitating amylin analogue (pramlintide) was developed and introduced in the USA in 2005, but is not available in Europe [36].

Actions

Pramlintide (Figure 15.5.4.14) acts centrally to complement the effects of insulin in the control of postprandial glucose homeostasis. It acts predominantly via the area postrema in the brain stem, which communicates with the hypothalamus and activates neural pathways to suppress glucagon secretion by pancreatic alpha-cells.

Amylin KCNTAT CATQRL ANFLVH SSNNF GAILSS TNVGSNTY-(NH_2)

Pramlintide KCNTAT CATQRL ANFLVH SSNNF GPILPP TNVGSNTY-(NH_2)

Figure 15.5.4.14 Structures of amylin (islet amyloid polypeptide, IAPP) and pramlintide.

By reducing glucagon, pramlintide reduces hepatic glucose output. Additionally, via a central route, pramlintide decreases the rate of gastric emptying and reduces the secretion of gastric juice, which slows the rate of digestion with resultant slowing of nutrient absorption. Pramlintide also acts centrally to reduce food intake, which may, in the long term, assist weight control.

Efficacy

Pramlintide is typically used to reduce the insulin dose and prevent weight gain associated with higher doses of insulin, while improving glycaemic control. It is injected before meals in patients with type 1 or type 2 diabetes who are already receiving insulin therapy, and has been shown to improve the glycaemic profile with reductions in HbA1c of about 0.3–0.6% (3–7 mmol/mol) and reductions in body weight of 1–2 kg. Since these effects of pramlintide are usually achieved with a reduction in the insulin dose, it is generally advised to reduce the pre-meal short-acting insulin dose by about half when initiating pramlintide therapy, to avoid risk of hypoglycaemia. The pH difference between insulin and pramlintide precludes the combination of these agents in the same syringe.

Cautions

The most common side effect of pramlintide is nausea, which is usually transient and minimized by gradual dose titration. Since pramlintide is used with insulin, it increases the risk of hypoglycaemia unless appropriately titrated in conjunction with suitable meals and a reduced insulin dose. A drawback to the use of pramlintide is the need for an injection before each main meal, which is additional to the injection required for insulin therapy. Antibodies to pramlintide have been detected in some patients, but this does not appear to have interfered with efficacy.

Bromocriptine

Bromocriptine received an indication for glucose-lowering in the USA in 2009 but is not indicated for this purpose in Europe. Its antihyperglycaemic potential had been appreciated through studies decades earlier, and through experience during use in the treatment of Parkinsonism and pituitary tumours.

Actions

Bromocriptine is a dopamine D2 receptor agonist that lowers glucose concentrations without stimulating insulin secretion [37]. Taken in a low dose early in the morning it appears to assist the normal diurnal periodicity of glucoregulation, acting centrally to stimulate neural pathways that reduce hepatic glucose output.

Efficacy

Low-dose bromocriptine reduces HbA1c by about 0.5–0.6% (5–7 mmol/mol). It can be used as monotherapy or in combination with other oral agents, is unlikely to cause severe hypoglycaemia and is not associated with weight gain.

Cautions

Experience with bromocriptine during use for other indications suggests that risk of respiratory and pericardial fibrosis, hypotension, and exacerbation of psychotic disorders should be borne in mind, and appropriate monitoring should be undertaken. Also, interactions can occur with dopamine antagonist therapy, drugs that are highly protein bound, and other drugs that are metabolized by or induce the P450 isoform CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4). Use in pregnancy is not recommended.

Colesevelam

The bile acid sequestrant colesevelam, used in the treatment of dyslipidaemia, received an indication for glucose-lowering in the USA in 2008, but is not indicated for this purpose in Europe.

Actions

Colesevelam is a non-absorbed allylamine polymer. Its glucose-lowering mechanism is unclear, but by transferring bile acids distally along the intestine colesevelam may facilitate bile acid stimulation of GLP-1 secretion by L cells. Also, by reducing bile acid recirculation to the liver, colesevelam might decrease gluconeogenesis by preventing activation of hepatic farnesoid receptors.

Efficacy

In trials, colesevelam reduced HbA1c by about 0.3–0.5% as add-on to metformin, sulphonylurea, or insulin, without changing body weight or increasing the risk of hypoglycaemia [38].

Cautions

Colesevelam should be taken with plenty of water, and avoided in patients with gastrointestinal motility disorders. Use is associated with risk of increased triglyceride concentrations, possible decreased absorption of fat-soluble vitamins and delayed absorption of other medications.

Conclusions

Managing the progressive and variable natural history of type 2 diabetes with its emergent complications and comorbidities continues to present a formidable therapeutic challenge. Diet and other lifestyle measures are fundamental throughout, supplemented with a series of oral and injectable glucose-lowering agents, as appropriate, eventually requiring insulin in many patients. Individualization and flexibility of therapy to suit the particular needs and circumstances of the patient are important to the

management process, and must be complemented with adequate education and empowerment.

REFERENCES

1. Ray KK, Seshasar SRK, Wijesuriya S, *et al.* Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. *Lancet*, 2009; **373**: 1765–72.
2. Stratton UM, Adler AI, Neil AW, *et al.* Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ*, 2000; **321**: 405–12.
3. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*, 2008; **359**: 1577–89.
4. Buse JB, Wexler DJ, Tsapas A, et al. 2019 update to: management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia*, 2020; **63**: 221-8.
5. Garber AJ, Abrahamson MJ, Barzilay JI, *et al.* Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm—2018 executive summary. *Endocr Pract*, 2018; **24**: 91–120.
6. International Diabetes Federation. IDF Clinical Practice Recommendations for managing Type 2 Diabetes in Primary Care. International Diabetes Federation, Brussels, 2017. Available at: https://www.idf.org/e-library/guidelines/128-idf-clinical-practice-recommendations-for-managing-type-2-diabetes-in-primary-care.html (accessed: February 2018).
7. American Diabetes Association. Standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41** (Suppl 1); S1–158.
8. National Institute for Health and Clinical Excellence. Type 2 diabetes in adults: management NICE guideline [NG28] Published date: December 2015, updated May 2017. Available at: https://www.nice.org.uk/guidance/ng28 (accessed: February 2018)
9. Tahrani AA, Barnett AH, Bailey CJ. Pharmacology and therapeutic implications of current drugs for type 2 diabetes mellitus. *Nat Rev Endocrinol*, 2016; **12**: 566–92.
10. Bailey CJ. Metformin: historical overview. *Diabetologia*, 2017; **60**: 1566–76.
11. Rena G, Hardie DG, Pearson ER. The mechanism of action of metformin. *Diabetologia*, 2017; **60**: 1577–85.
12. Bailey CJ, Turner RC. Metformin. *N Engl J Med*, 1996; **334**: 574–9.
13. Bailey CJ. Metformin: effects on micro and macrovascular complications in type 2 diabetes. *Cardiovasc Drugs Ther*, 2008; **22**: 215–24.
14. Inzucchi SE, Lipska KJ, Mayo H, Bailey CJ, McGuire DK. Metformin in patients with type 2 diabetes and kidney disease: a systematic review. *JAMA*, 2014; **312**: 2668–75.
15. de Wet H, Proks P. Molecular action of sulphonylureas on KATP channels: a real partnership between drugs and nucleotides. *Biochem Soc Trans*, 2015; **43**: 901–7.
16. Rendell M. The role of sulfonylureas in the management of type 2 diabetes. *Drugs*, 2004; **64**: 1339–58.
17. Deacon CF, Lebovitz HE. Comparative review of dipeptidyl peptidase-4 inhibitors and sulphonylureas. *Diabetes Obes Metab*, 2016; **18**: 333–47.
18. Guardado-Mendoza R, Prioletta A, Jiménez-Ceja LM, Sosale A, Folli F. The role of nateglinide and repaglinide, derivatives of meglitinide, in the treatment of type 2 diabetes mellitus. *Arch Med Sci*, 2013; **9**: 936–43.
19. Nauck M. Incretin therapies: highlighting common features and differences in the modes of action of glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. *Diabetes Obes Metab*, 2016; **18**: 203–16.
20. Campbell JE, Drucker DJ. Pharmacology, physiology, and mechanisms of incretin hormone action. *Cell Metab*, 2013; **17**: 819–37.
21. Drucker D. The cardiovascular biology of glucagon-like peptide-1. *Cell Metab*, 2016; **24**: 15–30.
22. Hernandez AF, Green JB, Janmohamed S, *et al.* Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet*, 2018; **392**: 1519–29.
23. Rosenstock J, Perkovic V, Johansen OE, *et al.* Effect of linagliptin vs placebo on major cardiovascular events in adults with type 2 diabetes and high cardiovascular and renal risk: the CARMELINA randomized clinical trial. *JAMA*, 2019; **321**: 69–79.
24. Deacon CF. A review of dipeptidyl peptidase-4 inhibitors. Hot topics from randomized controlled trials. *Diabetes Obes Metab*, 2018; **20** Suppl 1: 34–46.
25. Tahrani AA, Barnett AH, Bailey CJ. SGLT inhibitors in management of diabetes. *Lancet Diabetes Endocrinol*, 2013; **1**: 140–51.
26. Nauck MA. Update on developments with SGLT2 inhibitors in the management of type 2 diabetes. *Drug Des Devel Ther*, 2014; **8**: 1335–80.
27. Zhang X-L, Zhu Q-Q, Chen Y-H, Li X-L, Chen F, Huang J-A, Xu B. Cardiovascular safety, long-term noncardiovascular safety, and efficacy of sodium-glucose cotransporter 2 inhibitors in patients with type 2 diabetes mellitus: a systemic review and meta-analysis with trial sequential analysis. *J Am Heart Assoc*, 2018; **7**: pii: e007165.
28. Neal B, Perkovic V, Mahaffey KW, *et al.* Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*, 2017; **377**: 644–57.
29. Wiviott SD, Raz I, Bonaca MP, *et al.* Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2019; **380**: 347–57.
30. Wanner C, Inzucchi SE, Zinman B. Empagliflozin and progression of kidney disease in type 2 diabetes. *N Engl J Med*, 2016; **375**: 323–34.
31. Grygiel-Gorniak B. Peroxisome proliferator-activated receptors and their ligands: nutritional and clinical implications—a review. *Nutr J*, 2014, **13**: 17.
32. Liao H, Saver JL, Wu Y, *et al.* Pioglitazone and cardiovascular outcomes in patients with insulin resistance, pre-diabetes and type 2 diabetes: a systematic review and meta-analysis. *BMJ Open*, 2017; **7**: e013927.
33. Davidson MA, Mattison DR, Azoulay L, Krewski D. Thiazolidinedione drugs in the treatment of type 2 diabetes mellitus: past, present and future. *Crit Rev Toxicol*, 2018; **48**: 52–108.
34. Lebovitz HE. Alpha-glucosidase inhibitors as agents in the treatment of diabetes. *Diabetes Rev*, 1998; **6**: 132–45.
35. Holman RR, Coleman RL, Chan JCN, *et al.* Effects of acarbose on cardiovascular and diabetes outcomes in patients with coronary heart disease and impaired glucose tolerance (ACE): a randomised, double-blind, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 877–86.

36. Edelman S, Maier H, Wilhelm K. Pramlintide in the treatment of diabetes mellitus. *BioDrugs*, 2008; **22**: 375–86.
37. Holt RI, Barnett AH, Bailey CJ. Bromocriptine: old drug, new formulation and new indication. *Diabetes Obes Metab*, 2010, **12**: 1048–57.
38. Fonseca VA, Handelsman Y, Staels B. Colesevelam lowers glucose and lipid levels in type 2 diabetes: the clinical evidence. *Diabetes Obes Metab*, 2010; **12**: 384–92.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

15.5.5 Hypoglycaemia in the Treatment of Diabetes Mellitus

Stephanie A. Amiel

Introduction

Hypoglycaemia (low blood glucose concentration) is the most important acute complication of the pharmacological treatment of diabetes mellitus. Low blood glucose impairs brain function, may precipitate cardiac arrhythmias, and is associated with an inflammatory stress response. The brain has minimal endogenous stores of energy, with small amounts of glycogen in astroglial cells, and is therefore largely dependent on circulating glucose as the substrate to fuel cerebral metabolism and support cognitive performance. If blood glucose concentrations fall sufficiently, cognitive dysfunction occurs. In health, efficient glucose sensing and counterregulatory mechanisms exist to prevent clinically significant hypoglycaemia. These are impaired by diabetes and by its therapies. Patients with diabetes rank fear of hypoglycaemia alongside fear of chronic complications such as nephropathy or retinopathy [1]. Fear of hypoglycaemia, hypoglycaemia itself and attempts to avoid hypoglycaemia limit the degree to which glycaemic control can be intensified to reduce the risk of chronic complications of diabetes, both for type 1 and type 2 diabetes.

Definition and Classification of Hypoglycaemia

Truly good metabolic control in diabetes can only be defined as maintenance of near-normal blood glucose concentrations, reflected by measures of medium-term glycaemic control, such as glycated haemoglobin (HbA_{1c}), associated with least risk of microvascular complications, *plus* absence of troublesome hypoglycaemia. This two-item definition matters not only in optimizing diabetes management for people with diabetes, but also in the assessment of the relative efficacy of new treatments for diabetes coming to market, where superiority of performance may be claimed on the basis of lesser risk of hypoglycaemia.

Hypoglycaemia may be defined clinically as an episode in which the low blood glucose results in a characteristic symptom complex or in the presentation of signs. Whipple's triad, in which an event requires symptoms consistent with hypoglycaemia, a measured low plasma glucose concentration and relief of symptoms by administration of glucose to be defined as hypoglycaemia, was devised originally for surgeons to diagnose spontaneous hypoglycaemia in pathological hyperinsulinism [2]. It remains valid for diabetes management today, if we add 'or signs' to the 'symptoms' to allow inclusion of episodes of asymptomatic hypoglycaemia, and can agree on a biochemical value for hypoglycaemia. In theory, the latter should be a simple statement of the plasma glucose concentration that is lower than values seen in health, traditionally defined as under 3.5 mmol/L, but this has been controversial, as some counterregulatory responses can be detected at higher glucose concentrations, and there is no evidence for actual harm until lower values are reached.

Expressing hypoglycaemia prevalence rates or an individual's hypoglycaemia experience using symptoms alone to define an episode is obviously dependent on subjective awareness of hypoglycaemia and the affected person's memory. Many symptomatic mild episodes will not be recalled, while normoglycaemic episodes with symptoms mimicking hypoglycaemia may be recorded. Nevertheless, symptom-based definitions are useful clinically. They categorize episodes by degree:

- severe hypoglycaemia, in which third-party intervention is required, because cognitive dysfunction is so impaired that the patient cannot either perceive or make logical response to the hypoglycaemia; useful subgroups of this may be episodes requiring parenteral therapy (intravenous glucose or intramuscular glucagon) and those resulting in coma or seizure (which may include episodes for which third-party intervention was not available), or for health economic assessments those requiring ambulance call out, hospital attendance, or admission. It is worth noting that some people with diabetes, particularly those with type 1, count only this category as 'hypoglycaemia', which will lead to underestimation of their experience and missed opportunities for prevention unless more specific questions are asked. It is also worth noting that in paediatric practice, children will often need assistance to understand and treat their hypoglycaemia, so the need for assistance is a poor marker of the severity of the hypoglycaemia itself. However, by stressing the presence of cognitive impairment as the reason for the requirement for help, the paediatric definition of severe hypoglycaemia has been harmonized and is given as having 'severe cognitive impairment (including coma and convulsions) requiring external assistance by another

person to actively administer carbohydrates, glucagon, or take other corrective actions'. Again, events associated with a seizure or loss of consciousness are recognized as a subgroup [3].

- non-severe hypoglycaemia is then any episode which is self-recognized and self-treated. The term 'non-severe' is recommended in preference to the term 'mild', as such episodes, while medically safe, can be experienced by the patient as unpleasant precisely because they may be very symptomatic. Some authorities include an intermediate category of 'moderate', where the patient can perceive and self-treat the episode, but where the symptoms and/or the treatment cause significant disruption of normal activity, but there are obvious problems with consistency of this definition, leaving division into 'non-severe' and 'severe' solely by ability to self-treat the favoured option.

This simple and widely accepted categorization assumes that every documented episode is true hypoglycaemia and is either symptomatic or otherwise clinically obvious. A 2013 consensus statement from the American Diabetes Association (4) classifies hypoglycaemia as follows:

- severe, as just defined (although without explicitly including episodes for which help was not available);
- documented symptomatic in which typical symptoms are experienced in the presence of a low measured plasma glucose and, by distinction from severe, the event is self-treated;
- asymptomatic, a biochemical episode not accompanied by typical symptoms, describing episodes in which the hypoglycaemia was only detected because someone other than the patient recognized it and/or where it was detected because of a coincidental blood test, i.e. not being done because the patient was aware of being low;
- probable symptomatic, in which the patient experiences symptoms but does not have a confirmatory biochemical measurement;
- pseudo-hypoglycaemia, formerly defined as relative hypoglycaemia, in which the patient experiences typical symptoms of hypoglycaemia, but with a measured blood glucose that is not in the hypoglycaemic range. This is likely to occur in patients accustomed to running high blood glucose concentrations because endogenous hypoglycaemia sensing is programmed by recent glycaemic experience and so may trigger responses at normal or even high values. Furthermore, there is a still a strong clinical impression, not entirely supported by experimental evidence [5, 6], that rate of glucose fall can be important in determining symptomatic hypoglycaemia.

The ADA classification depends on defining a biochemical cut-off for hypoglycaemia, quoted in the 2013 document as 3.9 mmol/L (70 mg/dl) or less, and certainly hypoglycaemia should be definable solely biochemically. Frequency will then be dependent upon frequency of monitoring. As just intimated, there is however no universally agreed threshold glucose concentration for defining biochemical hypoglycaemia. Use of capillary, venous, venous arterialized (sampling from a distal venous cannula in a heated hand) or arterial samples will introduce variability, as will measurement of concentration in either whole blood or plasma. However, international consensus has recently been achieved on the biochemical definitions of important low blood glucose [3, 7] (**Figure 15.5.5.1**), in which a plasma glucose reading of 3.9 mmol/L (less than 70 mg/dl) is recognized, not as hypoglycaemia *per se*, but as the lower limit of a desirable target range for diabetes therapies. Consequently, plasma glucose values below this are defined as a 'hypoglycaemia alert', a value that should be treated to prevent hypoglycaemia; while values <3 mmol/L (<54 mg/dl) are recognized as being associated with negative medical outcomes, including cognitive impairment, cardiac arrhythmia and generation of impaired awareness of hypoglycaemia (IAH), and therefore identify a serious, clinically important hypoglycaemia to be avoided. The term 'severe hypoglycaemia' is retained to denote an episode including severe cognitive impairment requiring external assistance for recovery, or in which there is loss of consciousness or seizure, not requiring a biochemical value. These definitions, proposed in 2017, have been accepted by the American Diabetes Association and the European Association for the Study of Diabetes with simultaneous publications in their respective journals, although it should be noted that that original document used 3.9 mmol/L or less as the hypoglycaemia alert, based

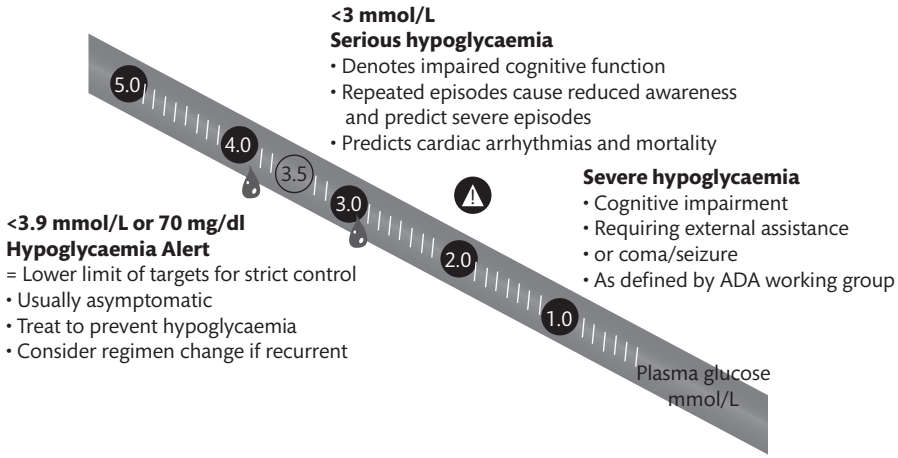


Figure 15.5.5.1 Schematic of plasma glucose readings for significant hypoglycaemia. For description, refer to text.

Adapted with permission from the International Hypoglycaemia Study Group (IHSG). Diabetic Hypoglycemia: Questions and Controversies. Presented at the 9th World Congress on Prevention of Diabetes and its Complications. (2016). Available at: https://ihsgonline.com/wp-content/uploads/2017/05/Diabetic_Hypoglycemia_Questions_and_Controversies_v1.0.pdf Accessed 19 June 2019. See also reference number 7.

on the earlier recommendation from the ADA, and subsequent discussion has accepted <3.9 mmol/L (<70 mg/dl) for consistency. This is because 3.9 mmol/L (70 mg/dl) is the quoted lower limit of the recommended target range for diabetes therapies intended to minimize risk of future complications and so using <3.9 mmol/L for the hypoglycaemia alert avoids defining a value of precisely 3.9 mmol/L in two ways. The scheme (Figure 15.5.5.1) has been endorsed by a multidisciplinary group including patient advocates and people with diabetes [8]. Important for clinical practice, there is evidence that those who perceive their own hypoglycaemia at plasma glucose concentrations of 3 mmol/L (54 mg/dl) or higher are at significantly lower risk for severe hypoglycaemia than those who perceive symptoms of hypoglycaemia at plasma glucose concentrations below 3 mmol/L (54 mg/dl) or who do not perceive their hypoglycaemia at all [9] (see 'Impaired Awareness of Hypoglycaemia', to follow).

Definitions for hypoglycaemia in continuous glucose monitoring data are also uncertain, with no clear clinical correlates for different values, given that the readings are made in interstitial fluid, not plasma, although equilibrated to plasma glucose concentrations, and there is an inevitable time lag between interstitial and plasma concentrations [10]. An early definition of 'low interstitial glucose' or 'LIG' described low glucose readings lasting at least 20 minutes at values of less than 3 mmol/L (54 mg/dl, LIG_3) or less than 2.2 mmol/L ($LIG_{2.2}$) [11]. More recently, an international consensus came to an agreement in line with the consensus on plasma glucose readings, and defined a CGM reading of <3 mmol/L (54 mg/dl) for at least 15 minutes as clinically significant hypoglycaemia, retaining the range from 3 to <3.9 mmol/L (54 to <70 mg/dl) as a hypoglycaemia alert [12]. Most episodes of asymptomatic, usually nocturnal, hypoglycaemia detected by CGM with the wearer of the device unaware of the event in fact last much longer than this and the clinical significance of the shorter duration episodes remains uncertain. That such episodes are seen in non-diabetic individuals is not necessarily helpful, as hypoglycaemia in the presence of an intact feedback loop including endogenous insulin and glucagon may be very different from hypoglycaemia during which the plasma insulin concentrations are artificially maintained. Nevertheless, there has been recent validation for a nocturnal value of <3 mmol/L (54 mg/dl) as having clinical consequences, with no indication yet for problems for any specific values between 3.0 and <3.9 mmol/L (54 to <70 mg/dl) [13].

Incidence of Hypoglycaemia

Even allowing for the differences in definition, reported incidence of hypoglycaemia—even severe hypoglycaemia—varies considerably. The first major considerations are the type and duration of the diabetes and the diabetes therapy being used. Tables 15.5.5.1 and 15.5.5.2 give some estimates from the literature for incidence of severe hypoglycaemia in type 1 and type 2 diabetes respectively. Lifestyle and insulin sensitizing or glycosuric agents used alone carry virtually no risk of severe hypoglycaemia and the nature of the hypoglycaemia reported in such treatments (e.g. [42]) is not

Table 15.5.5.1 Frequency of severe hypoglycaemia in the literature: studies in type 1 diabetes

First author, year, and reference number	Number of subjects	Mean HbA_{1c} (%) (mmol/mol)	% using intensive therapy	Severe hypoglycaemia	
				% patients	Events/person per year
Observational (retrospective) studies					
Casparie (1985) [14]	200	8.1 (65 mmol/mol)	n/a	13	–
MacLeod (1993) [15]	600	10.7 (HbA_1)	10.8	29.2	1.7
Bali (1997) [16]	458	n/a			
Mulhhauser (1998) [17]	669	8.0 (64 mmol/mol)	79 (9% CSII)	19	0.82
ter Braak (2000) [18]	195	7.8 (62 mmol/mol)	82	40.5 (20% coma)	1.5
Linkeschova (2002) [99]	103	7.7 (61 mmol/mol)	100	n/a	1.2
Johnson (2002)[a] [20]	1113	–	–	12	0.05
Holstein (2002)[a] [21]	600	–	–	15.3	0.04
Leese (2003)[a] [22]	977	7.8 (62 mmol/mol)	–	7	0.012
Weinstock (2003) [23]	4973	7.7 (61 mmol/mol)	52% on CSII for ≥12 months	11.8	Not given
Pedersen-Bjergaard (2004) [24]	1076	8.6 (70 mmol/mol)	71.6	29.2	1.04
Zammitt (2007) [25]	300	8.2 (66 mmol/mol)	–	31	0.93
Paediatric registry studies					
Haynes (2017) [26] USA Germany Australia					 0.71 0.33 0.67
Prospective studies					
Janssen (2000) [27]	31	7.2 (55 mmol/mol)	100	50	–
Pederson-Bjergaard (2003) [28]	171	8.4 (68 mmol/mol)	87	39	1.10

Table 15.5.5.1 *Continued*

First author, year, and reference number	Number of subjects	Mean HbA_{1c} (%) (mmol/mol)	% using intensive therapy	Severe hypoglycaemia	
				% patients	Events/person per year
Leckie (2005) [29]	243	9.1 (76 mmol/mol)	80	34	0.98
Khunti (2018) [30]	8022, observed over 4 weeks	7.9 (62.8 mmol/mol)	No data	14.4	4.90
Interventional studies					
DCCT (1997) [31]	711	7.2 (55 mmol/mol)	100	65	0.62
Bott (1997) [32]	636	7.6 (60 mmol/mol)	100	12	0.17
DAFNE (2002) [33]	169	8.4 (68 mmol/mol)	100	18	–
Hoogma (2006) [34]	129 127	7.7 (61 mmol/mol) 7.5 (58 mmol/mol)	MDI CSII	–	0.50 0.20
Hopkins (2012) [9]	359	8.5 (69 mmol/mol) 8.2 (66 mmol/mol)	Pre-education 12 months post education	25 Not stated	1.93 0.61

Table 15.5.5.2 Frequency of severe hypoglycaemia in the literature: studies in type 2 diabetes

First author, year	Number of subjects	Mean HbA_{1c} (%) (mmol/mol)	Mean age	Treatment modality	% patients with mild hypo	Severe hypoglycaemia	
						% patients	Events per patient per year
Observational studies							
Shorr (1997)[b] [35]	19932	–	>65	SU INS			0.01 SU 0.03 INS[a]
Miller (2001) [36]	1055	7.6 (60 mmol/mol)	60.9	SU	16 OHA 30 INS	0.5	
Johnson (2002) [20]	1113	–					0.05
Holstein (2003) [21]	*c.*9000	–		INS All therapies			0.015 0.004
Henderson (2003) [37]	215	8.6 (70 mmol/mol)	68	INS	64	15	0.28
Leese (2003) [22]	2823	7.6 (60 mmol/mol)	65.2	SU			0.09
	901	8.1 (65 mmol/mol)	64.9	INS		12	
Donnelly (2005) [38]	173	8.9 (74 mmol/mol)	66	INS	45	3	0.35
Murata (2005) [39]	344	–	65.5	INS	51.2		0.20
Akram (2006) [40]	401	8.3 (67 mmol/mol)	66	INS		16.5	0.44
Khunti (2018) [30]	19563 observed over 4 weeks	8 (64.2 mmol/mol)	60.8	INS	46.5	8.9	0.25
Interventional studies							
Abraira (VA CSDM, 1995) [41]	153	9.3 (78 mmol/mol)	60 (6)		56 (conv) 93 (intensive)	–	2/pt/yr in both groups
UKPDS (1998) [42]	3935	6.2 (44 mmol/mol)	54 (8)	SU INS	13	11 OHA 37 insulin	
4-T (2007) [43]	235 239	7.3 (56 mmol/mol) 7.2 (55 mmol/mol)	61.7 61.6	INS: bd Prandial	10 20	11 16	
ACCORD (2008) [44]	5123 5128	7.5 (58 mmol/mol) 6.5 (48 mmol/mol)	62.2 62.2	Standard intensive		8.6 26.7	1.0 3.1
ADVANCE (2008) [45]	5569 5571	7.3 (56 mmol/mol) 6.5 (48 mmol/mol)	66 66	Standard intensive	*c.*38 *c.*53	1.5 2.7	0.5 0.7
VADT (2008) [46]	899 892	8.4 (68 mmol/mol) 6.9 (52 mmol/mol)	60.3 60.5	Standard Intensive		17.6 24.1	4.0 12.0

[a] These population-based studies only recorded events that were reported to the emergency services or required hospital admission.
[b] This study defined serious hypoglycaemia as that requiring hospitalization/emergency department admission or bringing about death associated with a blood glucose of <2.8 mmol/l. bd, twice daily pre-mixed insulin (given before breakfast and evening meal); CSII, continuous subcutaneous insulin infusion; INS, insulin therapy; MDI, multiple daily insulin injection therapy; Prandial, fast-acting analogue given before meals only; OHA, oval hypoglycaemic agent; SU, sulphonylurea therapy.

clear. The nature of the population will affect the prevalence—a specialist clinic population may have particular characteristics not shared by people not attending that clinic, while questionnaire surveys of large populations, which may be thought to be more representative of the population at large, may suffer from bias in that people interested in the topic are more likely to respond. Large database surveys can be helpful, especially for health economic analyses, but the data collected may be imperfect. Another important consideration is whether the data are collected retrospectively by questionnaire from patient recall, or prospectively, and the duration of time over which the data are collected. Rates also vary by country. In one recent large global study of people using insulin therapy and estimating rates from a 4-week prospective recording, the Hypoglycaemia Assessment Tool (HAT) study, rates of severe hypoglycaemia ranged from 2.1 to 10.8 events per person per year (global mean 4.9) for type 1 diabetes and 1.3–3.7 (global mean 2.5) for type 2, and noted that the percentage of people reporting at least one episode of any hypoglycaemia ranged from 54% to 87.4% and from 30.1% to 53.8% for types 1 and 2, respectively [30]. The study importantly showed no link with HbA1c, in contrast to randomized controlled trials of intensified insulin therapy, where higher rates of severe hypoglycaemia are associated with the lower HbA1c seen in groups allocated to intensive therapy [31, 42, 44–46]. The rates reported by the HAT study are high. A much smaller, focused study of a clinic population of adults with type 1 diabetes reported an annual severe hypoglycaemia rate of 1.3 episodes per patient per year, but very importantly showed that only 40% of people experience a severe hypoglycaemia during a year and 10% of the population contribute nearly 70% of all episodes [24].

People with type 2 diabetes are at much lower risk for hypoglycaemia than those with type 1. However, the type 2 population is heterogeneous. As the disease progresses, insulin deficiency becomes more profound. The insulin-deficient late type 2 patient behaves more as a patient with type 1 diabetes, including in terms of hypoglycaemia risk. The UK Hypoglycaemia Study Group reported both patient-reported episodes and episodes in which circulating glucose, monitored as interstitial tissue glucose calibrated to capillary plasma continuously over 96 h, fell below 2.2 mmol/L for 20 min or more. This prospective study found low rates of hypoglycaemia in participants with type 2 diabetes on sulphonylurea therapy and in those who had been on insulin for less than 2 years, but the rates in participants who had been on insulin for more than 5 years were higher and not different from rates in participants with type 1 diabetes of less than 5 years' duration (**Figure 15.5.5.2**) [11]. People with type 1 diabetes of more than 15 years' duration had much higher rates than the other groups. One contributor to the differences is duration of diabetes *per se* and increasing insulin deficiency over time. It is likely that the patients with type 2 diabetes who are treated with insulin secretagogues and those in the early stages of insulin therapy, are relatively protected from hypoglycaemia by residual ability to suppress endogenous insulin secretion and therefore, also respond to hypoglycaemia with endogenous glucagon. As insulin deficiency progresses, these primary counterregulatory mechanisms are lost and, as already suggested, one hypoglycaemic episode diminishes the defences against and therefore, increases the risk for another in the near future. Importantly, despite the lower rates of severe hypoglycaemia compared to rates in people with type 1, patients

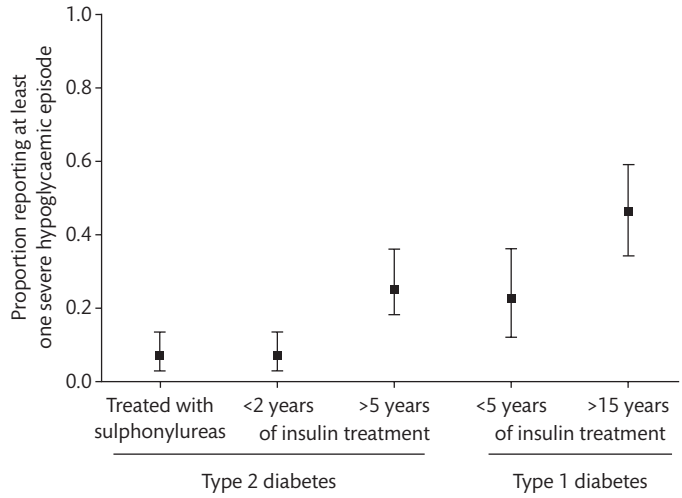


Figure 15.5.5.2 Graph showing relative frequency of severe hypoglycaemia in different groups of people with diabetes.
Reproduced with permission from UK Hypoglycaemia Study Group. Risk of hypoglycaemia in types 1 and 2 diabetes: effects of treatment modalities and their duration. *Diabetologia*. 2007 Jun;50(6):1140–7. Copyright © 2007, Springer-Verlag (Ref 11).

with type 2 diabetes may account for almost the same number of presentations to emergency services, because of the much greater prevalence of the disease [22]. Nevertheless, a meta-analysis of studies of people on sulphonylureas conducted in 2014 suggested only about 5% of people on such therapies experienced severe hypoglycaemia [47].

Physiological Responses to Hypoglycaemia

The response to hypoglycaemia is a generic stress response. As with other stressors, it involves activation of the hypothalamic–pituitary–adrenal axis and the autonomic nervous system. It is primarily driven centrally, although there are important local responses in the pancreas and liver (see next three paragraph) and modulating influences from peripheral sources, such as hepatic portal glucose sensors [48]. The brain's glucose-sensing apparatus is part of a complex system for control of energy balance. Despite vast fluctuations in energy intake and expenditure in normal daily living, in health, these mechanisms preserve plasma glucose concentrations within an extremely narrow physiological range. In health, hypoglycaemia sufficient to cause significant impairment of cognitive function does not occur.

The physiological responses to hypoglycaemia have been examined experimentally in human and in animal models by creating hypoglycaemia with insulin. This can be done by unopposed insulin injection, as in the insulin-tolerance test, or infusion, but, in human studies, a commonly used experimental technique is to infuse relatively high doses of insulin and control the hypoglycaemia with a simultaneous glucose infusion. This applies a very controlled hypoglycaemic challenge that allows comparisons of the responses to be made between different populations. It should be recognized that the counterregulatory responses to hypoglycaemia will normally be stimulated in a different physiological state but much useful information has been achieved with the technique.

Any change in plasma glucose concentration is detected by neurones in the brain stem and hypothalamus, which respond to changes

in blood glucose by altering their firing rate. The cellular and molecular mechanisms whereby these glucose responsive neurones are activated have recently been reviewed [49, 50]. The glucose-sensing neurones use many of the same mechanisms as the insulin-secreting cells of the pancreas, which release insulin in proportion to intracellular glucose metabolism and thus directly in response to the glucose supply to the cell. In the neurones, the likely response is a change in firing rate and neurotransmitter release. These eventually drive changes in endogenous glucose production from the liver and kidney and in peripheral glucose uptake by muscle and fat to sustain glucose concentrations in the circulation.

The first hormonal responses to a falling blood glucose concentration in the non-diabetic are cessation of pancreatic insulin release and stimulation of glucagon release (**Figure 15.5.5.3**). Hepatic glucose output rises as a result of glycogenolysis (which can sustain circulating plasma glucose for about 6–8 hours after the last food intake) and, later, gluconeogenesis. The role of hepatic autoregulation, in which hepatic glucose production rises in response to a falling blood glucose level independently of hormonal stimulation, is unclear. It is likely that changes in insulin and glucagon alone can protect against hypoglycaemia during normal daily living, with the balance between these processes maintaining circulating glucose concentrations despite wide changes in glucose availability (fed vs. fasting) and utilization (rest vs. exercise). If however glucose concentrations do fall, a stress response, including adrenaline secretion and sympathetic nervous system activation [51], further stimulates hepatic glucose output both directly and indirectly by peripheral effects on lipolysis and proteolysis to increase hepatic delivery of gluconeogenic precursors, such as alanine, glycerol, and lactate. Stimulation of the autonomic nervous system is reflected by changes in circulating noradrenaline and pancreatic polypeptide. Sympathetic activation directly stimulates endogenous hepatic glucose production, shuts down insulin secretion, and stimulates glucagon release. Growth hormone and cortisol also rise in hypoglycaemia, later or at slightly lower glucose concentrations. They are probably of secondary importance in acute counterregulation, but are important in sustaining euglycaemia, acting predominantly to reduce peripheral glucose utilization and provide precursors for gluconeogenesis. Other endocrine changes during hypoglycaemia include increased plasma renin activity and rises in prolactin, vasopressin, oxytocin, and β-endorphin. The role of these hormones in glucose regulation is uncertain.

The stress response to hypoglycaemia just described is associated with a characteristic symptom complex. This occurs at a lower glucose concentration than the stimulation of pancreatic glucagon and cessation of insulin secretion, and is not usually important in counterregulation in health, although subjective awareness has been reported in healthy volunteers during experimental hypoglycaemic induction studies at arterialized plasma glucose concentrations ranging from 2.8 to 3.6 mmol/L [51, 52]. The classical symptoms of hypoglycaemia are traditionally divided into autonomic (pounding heart/palpitations, shaking/tremor, hunger, sweating) and neuroglycopenic (drowsiness, difficulty speaking, incoordination, difficulty concentrating/confusion), either on the basis of the known aetiology of the symptom (e.g. sweating due to cholinergic sympathetic outflow) or by using statistical techniques. Groups of symptoms that tend to cluster together can be identified by principal component analysis, either during experimental hypoglycaemia, or from population-based studies in people with diabetes (**Table 15.5.5.3**) [53]. Hunger is a particularly useful symptom, since it both warns of hypoglycaemia and promotes eating and restoration of glucose levels. Symptoms of early hypoglycaemia become key in defence against more profound falls in circulating glucose concentrations in people with diabetes, *qv.*

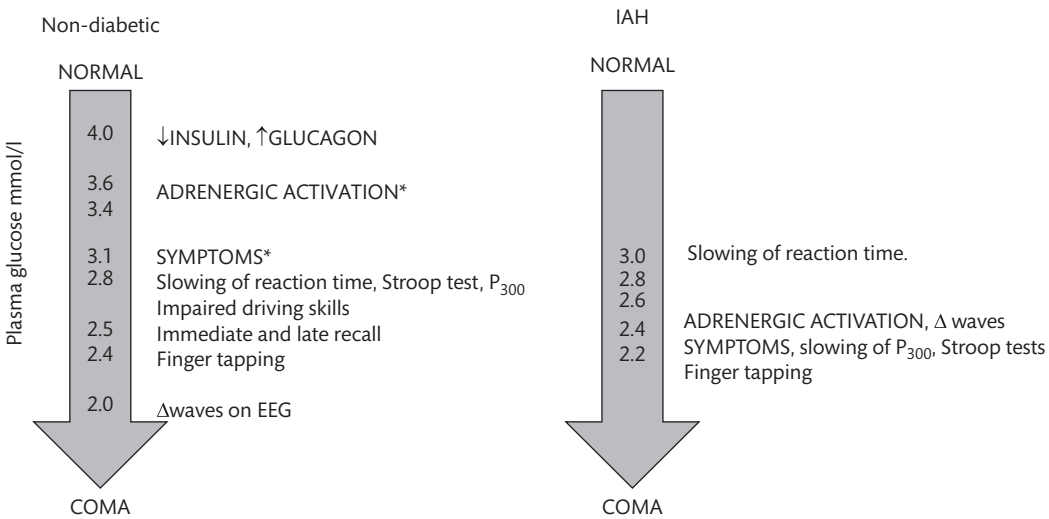


Figure 15.5.5.3 Counterregulation to hypoglycaemia. The arrow on the left shows the normal protective hierarchy of responses against progressive hypoglycaemia in with autonomic and humoral responses (upper case) occur before any detectable evidence of cognitive dysfunction (lower case). Of these, the Stroop test measures speed of reading lists of names of colours in appropriately, and, later, inappropriately coloured ink; P300 is an evoked potential recorded in response to a stimulus by scalp electrodes. * = In some studies of people with poorly-controlled diabetes, these parameters have been found to change at even higher glucose concentrations. Arterialized venous plasma glucose concentrations are given in mmol/L. On the right, the responses of a person with diabetes and impaired awareness are shown. The ability to regulate insulin and glucagon is lost and the glucose concentration for the onset of the rest of humoral response has fallen below the glucose concentration at which cognitive function is detectably deteriorated.

Table 15.5.5.3 Symptoms of acute hypoglycaemia

Autonomic	Neuroglycopenic	Other	Miscellaneous
• Sweating • Tremor • Warmth • Anxiety • Nausea	• Dizziness • Confusion • Tiredness • Difficulty in speaking • Headache	• Hunger • Blurred vision • Drowsiness • Tiredness	• Palpitations • Shivering

Reproduced with permission from McAulay V, Deary IJ, Frier BM. Symptoms of hypoglycaemia in people with diabetes. *Diabet Med*, 2001; 18: 690–705. Copyright © Diabetes UK.

Impaired cognitive function, particularly changes in reaction time tasks, is noted in healthy volunteers undergoing experimental hypoglycaemia at glucose concentrations of between 2.6 and 3.2 mmol/L [52, 54], depending on the circumstances of the study and the age of the participants. Impairment of multiple memory functions has been shown at around 2.5 mmol/L [55], although the impact of lesser degrees of hypoglycaemia on memory in non-diabetic subjects has not been studied. Older people show evidence of cognitive decline at slightly higher glucose concentrations compared to young adults, and a later onset of symptoms [52]. In theory, if glucose levels are lowered far enough, brain function becomes further impaired, so that drowsiness and even seizures or coma may eventually develop. This cannot, of course, be done in experimental settings, but it has been well described in insulin shock therapy in treatments for severe depression in the nineteenth century. Outside this, hypoglycaemia sufficient to cause significant cognitive dysfunction does not occur in health, except in extreme circumstances such as prolonged vigorous exercise. Starvation, for example, does not cause loss of consciousness, especially as the brain can use non-glucose fuels, such as ketones and lactate, to support its function.

Responses to Hypoglycaemia in Diabetes

In diabetes, endogenous protection against hypoglycaemia is impaired. With loss of endogenous insulin secretory capacity comes loss of the ability to reduce insulin release during hypoglycaemia, a problem compounded by artificial elevation of the circulating insulin from injected insulin, or as a result of the action of insulin secretagogues or meglitinides. The glucagon response to hypoglycaemia is also lost, as the alpha-cell response is apparently driven by signals from adjacent β-cells. The loss of the glucagon response to acute hypoglycaemia is seen within the first five years of type 1 diabetes [56]. The primary neurohumoral defenWEWce against hypoglycaemia is then adrenaline and autonomic activation. Defects in the former have been shown in otherwise healthy subjects with diabetes, associated with longer disease duration [56]. People with diabetes are, therefore, peculiarly dependent upon subjective recognition of early degrees of hypoglycaemia and voluntary ingestion of carbohydrate to compensate for the failure of endogenous responses. As gastric emptying is accelerated during hypoglycaemia, oral carbohydrate ingestion results in a rapid rise in circulating glucose (see 'Treatment', to follow).

The symptoms of hypoglycaemia early in diabetes are similar to those reported by healthy volunteers made hypoglycaemic in experimental protocols but are influenced by many factors, including the age group of the person. Studies have been made in children with type 1 diabetes, who may experience symptoms that they cannot report, and manifest signs, often behavioural, that are detected by their parents [57]. By the age of nine, children become able to detect and describe their own hypoglycaemia [57]. Symptoms of hypoglycaemia also wane with ageing and elderly patients with diabetes and treated with insulin report more lightheadedness. Unsteadiness incoordination/disarticulation symptoms become important [58]. Diabetes duration is also relevant, with many people reporting a weakening of particularly autonomic symptoms over time [59]. It is possible that treatment modality affects the glucose concentration at which symptoms of hypoglycaemia are reported—one study showed earlier symptom perception in people with type 2 diabetes using sulphonylureas, vs. insulin [60], and although an impact of the different degree of insulin deficiency is likely between patients on the different treatment modalities, a direct effect of the sulphonylureas cannot be ruled out. Agents that impact on sulphonylurea receptors have been investigated for their ability to enhance subjective awareness of hypoglycaemia. Antecedent glycaemic experience has a major impact, with evidence for the ability of antecedent hypoglycaemia to reduce the glucose concentration for symptom and neurohumoral responses to subsequent hypoglycaemia demonstrated experimentally in both non-diabetic people [61, 62] and in people with diabetes [63–65] and evidence also for prior hyperglycaemia, as seen in type 2 diabetes, to elevate the glucose concentrations associated with onset of counterregulatory responses [66].

Glucose thresholds for cognitive impairment during acute hypoglycaemia are not so obviously impacted upon by the presence of diabetes. Deterioration in task performance during experimentally induced controlled slow reduction in plasma glucose has been seen at a mean of between 3.2 and 2.6 mmol/L in all studies, irrespective of type of diabetes or its treatment [54, 60, 67, 68]. Performance of simpler tasks may be slightly better preserved during hypoglycaemia and glucose concentrations for some tasks may show some adaptation to antecedent hypoglycaemia, but this effect if present is small [63]. Importantly, one study in young people with type 1 diabetes who performed brief cognitive tasks prior to measuring plasma glucose demonstrated preservation of ability at glucose concentrations of 3–3.9 mmol/L but significant impairment at lower values, as well as at values of over 22.2 mmol/L [69].

Impaired Awareness of Hypoglycaemia

Because of their lost ability to make an appropriate insulin or glucagon response to a falling plasma glucose, people with insulin-deficient diabetes are dependent on subjective recognition of the start of a fall in order to arrest it by ingestion of rapidly absorbed carbohydrate and avoid severe hypoglycaemia with significant cognitive impairment. Between 20% and 40% of people with type 1 diabetes have impaired awareness of hypoglycaemia (IAH), in which they fail to be able to recognize their own hypoglycaemia in timely fashion [9, 70]. This increases their risk of severe hypoglycaemia sixfold [70]. The prevalence of IAH in type 2 diabetes is less well studied but there is evidence that 10% of people using insulin will have IAH, with a 17-fold increased risk of severe episodes [71].

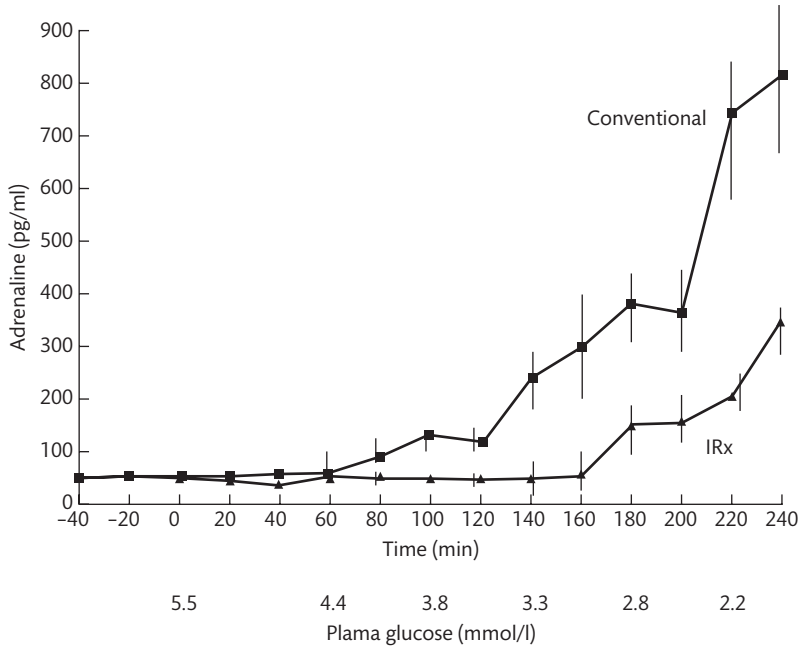


Figure 15.5.5.4 Delayed and diminished adrenaline responses to hypoglycaemia after institution of intensified insulin therapy in the mid-1980s. The samples were collected during controlled hypoglycaemia established by the insulin clamp technique, so the adrenaline responses are to the same hypoglycaemic challenge, applied before and after intensification of therapy in a small group of people with type 1 diabetes.

IAH is associated with defects in the normal counterregulatory responses to acute hypoglycaemia, additional to the inability to regulate insulin secretion or mount a glucagon response. The neurohumoral responses are delayed in onset to lower glucose concentrations (Figure 15.5.5.3) and diminished in magnitude at any given concentration (Figure 15.5.5.4) [72]. The glucose concentration at which cognitive dysfunction is first detectable is, however, not lower [67]. Performance in some cognitive tasks may be preserved at slightly lower plasma glucose concentrations, but the change is small and much less than the change seen in glucose concentrations associated with symptoms or catecholamine responses [73]. For some tasks, there is no change at all and the protective window between the glucose concentration for subjective awareness and hormonal responses and the onset of cognitive function is lost [67, 74]. By the time the stress response starts, the patient may have too much cognitive deficit to be able to recognize or respond to it. The major contributor to this situation is prior exposure to low blood glucose concentrations, with the impact on symptomatic and hormonal responses to subsequent hypoglycaemia alluded to above [63, 65, 68]. It is not clear whether some people are genetically more resistant to this phenomenon, or to the cognitive decline of hypoglycaemia, but, in general, people frequently exposed to plasma glucose readings of <3 mmol/L will develop impaired awareness of hypoglycaemia and experience more severe episodes [9]. Equally, avoidance of such exposure can restore awareness [75–77].

The mechanisms of IAH induction are not certain. Postulated mechanisms include increased glucose uptake by neuronal cells in response to hypoglycaemia exposure, and alterations in the brain's ability to utilize non-glucose fuels (e.g. [78]). Neuroimaging studies have failed to confirm that loss of response to hypoglycaemia is associated with increased brain glucose uptake, a phenomenon that has been well described in animal models and in early human studies, but have shown that hypoglycaemia is associated with altered responses in brain regions, including in the brain stem and hypothalamic and pituitary regions, involved in energy balance, glucose sensing, and stress responses, balance and coordination, and also in brain regions that are involved in desire to eat, emotional salience and recall [79, 80]. These differences in brain response to hypoglycaemia may be relevant not just to the reduced perception of symptoms, but also to the difficulties some patients experience in avoiding hypoglycaemia in the future, a strategy that can restore awareness to the unaware.

Nocturnal Hypoglycaemia

Hypoglycaemia at night has important specific features and is the topic of a recent review [81]. The counterregulatory response is greatly attenuated during deep sleep and people with diabetes experiencing hypoglycaemia during such sleep will not wake up, or be aware, and may also show no external signs. They may also sleep through alarms generated by continuous glucose monitors. Symptomatic hypoglycaemia at night may only occur if the hypoglycaemia is still present or occurs during lighter sleep. Nocturnal hypoglycaemia may thus continue for several hours. Risk of nocturnal hypoglycaemia is increased because the early hours of the morning are the time of maximum insulin sensitivity (because of diurnal variations in counterregulatory hormone secretion) and may also coincide with peak action of insulins injected to provide overnight cover. Hypoglycaemia in the night does not appear to impair cognitive performance next day but can be associated with low mood. It confounds the laying down of memory [82] and may be enough to reduce the stress response to hypoglycaemia the next

day [63]. Symptoms of an unsuspected hypoglycaemia in the night range from none at all to temporary neurological deficit (Todd's paresis), although a headache in the morning may be a result.

Contributory Factors to Hypoglycaemia

Hypoglycaemia in patients with diabetes is driven by exogenous insulin and insulin secretagogue therapy, with the loss of endogenous control of insulin secretion. Such therapies are inevitably associated with times of insulin excess and loss of glucagon responses to hypoglycaemia consequent upon the loss of β-cell function increases the risk. Additional risk occurs with defective responses in other elements of the normal stress response, most notably with the development of IAH. Risk of severe hypoglycaemia rises as insulin deficiency, marked by C-peptide deficiency, advances; with increasing age and diabetes duration and, most robustly, with a history of severe hypoglycaemia in the past [24]. A relationship between increased hypoglycaemia and desirable HbA1c is not inevitable. Most recent studies fail to find a higher incidence of severe hypoglycaemia with a lower HbA_{1c}, at least in type 1 populations [23, 26], although one recent study has shown increased risk in type 2 patients attending primary care with HbA1c below 53 mmol/mol [83]. Rates of severe hypoglycaemia are always higher in the intensive treatment arms of research trials [31, 44–46] but in type 1 diabetes, improving patients' ability to manage their own insulin regimen flexibly through structured education lowers HbA1c and risk of hypoglycaemia (e.g. [9]). In type 2 diabetes, there is no risk for severe episodes with lifestyle therapies or insulin sensitizing therapies (during which the patient is using endogenous insulin only) and minimal risk with incretin-based therapies, or glycosuric agents, when used alone. The risk is restricted to insulin secretagogue therapies and highest with longer-acting agents, such as first- and second-generation sulphonylureas [83]. This is not to suggest that such agents should not be used, as the absolute risk remains small and they are well-tested, effective, and familiar agents, but care should be taken in their use, especially in elderly patients and in those with other comorbidities, such as renal impairment or coexisting cognitive dysfunction, now that newer agents with lower hypoglycaemia risk are available.

Exogenous insulin can cause hypoglycaemia because it is not under endogenous control; it is delivered systemically, rather than into the portal circulation, creating a relatively high peripheral-to-portal gradient; its absorption is unpredictable, depending on factors such as skin blood flow (and, therefore, site of injection, local lipohypertrophy, exercise of underlying muscle, or physical disruption of the injection site by rubbing, ambient temperature) and because of intrinsic variability in the absorption, which affects time to peak action and duration of the effects of an injection. Whether the absence of C-peptide leaves the patient lacking in a specific defensive action, or is simply a reflection of the complete inability to achieve any lowering of insulin effect by shutting off an endogenous supply or influencing the glucagon response, is not known.

Often, a single episode of hypoglycaemia can be attributed to a likely precipitant. Exercise, missed meals/snacks, treatment errors, and alcohol are the most commonly reported (Box 15.5.5.1). Exercise causes immediate hypoglycaemia as it drives insulin and insulin independent muscle glucose uptake and glucose oxidation. Unaccustomed, vigorous, or prolonged exercise (all relative to the exerciser's accustomed exercise level) will continue to lower blood glucose over the next 12–24 hours as expended muscle and liver glycogen stores are replaced. The possibility that exercise may cause hypoglycaemia many hours later is often not appreciated. Extreme exercise, by producing a stress response, may alleviate this and one paper has suggested a sprint at the end of a period of less intense exercise may provide some protection in the immediate post-exercise phase [84], but not the delayed hypoglycaemia [85] which needs to be addressed by reduced insulin administration.

Alcohol increases the risk of delayed severe hypoglycaemia, by inhibiting the gluconeogenesis needed to maintain blood glucose levels after fasting. The hypoglycaemia typically occurs in the early hours of the morning after drinking in the evening; or even after breakfast the next day. Alcohol also causes neglect of self-care of diabetes (affecting compliance with medical treatment and dietary regimens), decreases perception of symptoms of early hypoglycaemia and may impair the ability to take appropriate action to recover from a low glucose.

Box 15.5.5.1 Risk factors for hypoglycaemia and problematic hypoglycaemia

Common contributors to an individual episode of hypoglycaemia

- Missed or inadequate meals/snacks
- Exercise
- Drug or insulin dose error
- Alcohol
- Change in absorption from injection, e.g. hot bath, sauna, change in injection site
- Overcorrection of a high glucose reading, e.g. postprandial, during action of earlier correction dose

Risk factors for experiencing severe hypoglycaemia:

- History of severe hypoglycaemia
- Hypoglycaemia unawareness/deficient counterregulation
- Long duration of diabetes
- C-peptide deficiency
- Under 5 years of age
- Nocturnal hypoglycaemia
- Intensified glycaemic control (insulin or drug treatment)
- Inappropriate insulin/drug regimen, e.g. long-acting sulphonylurea in elderly patients
- Comorbidities and polypharmacy

Other factors increasing risk for recurrent, problematic hypoglycaemia

- Failure of glucose counterregulation
  - Cortisol and/or growth hormone deficiency
- Failure of insulin or sulphonylurea clearance
  - Renal failure
  - Liver failure
  - Hypothyroidism
  - Insulin-binding antibodies
- Failure of endogenous glucose production
  - Liver failure
  - Glycogen storage diseases
- Failure of exogenous glucose supply
  - Malabsorption, including coeliac disease
  - Gastroparesis
  - Eating disorders, especially dieting or anorexia
- Postpartum/breast feeding
- Manipulation of therapy (includes suicide/parasuicide and obsessive fear of hyperglycaemia)

Presence of insulin-binding antibodies may interfere with the pharmacodynamics of any given insulin, and has been associated with increased risk of severe hypoglycaemia, especially when present in high concentrations in patients using older, less pure insulin preparations [56]. Insulin antibody levels are higher with genetically modified insulin analogues and with inhaled insulin, but, so far, have not been shown to have the same impact. In autoimmune disease, both insulin-binding and insulin-receptor antibodies have been known to cause hypoglycaemia in the absence of insulin therapy. Similarly, there are reports of hypoglycaemia occurring in the prodrome to type 2 diabetes, thought to be due to an exaggerated second-phase insulin response, brought on by early hyperglycaemia associated with loss of the first-phase insulin response.

Comorbidities, perhaps particularly in the elderly patient with type 2 diabetes, are well established to increase risk of severe hypoglycaemia. Renal failure and hypothyroidism (by decreasing insulin clearance) and hypoadrenalism from Addison's disease or hypopituitarism and growth hormone deficiency (by reducing insulin antagonism) are rare causes of recurrent hypoglycaemia. Anecdotally, insulin-treated patients on adrenal replacement therapy may be more safely managed on longer-acting agents, e.g. prednisolone rather than hydrocortisone, to avoid periods of very low corticosteroid levels at night, although this should be the subject of research. Drugs that have been reported to increase hypoglycaemia risk in treated diabetic patients include angiotensin-converting enzyme (ACE) inhibitors in patients with type 1 diabetes, any hypertensive therapy in patients with type 2 diabetes (possibly reflecting more active medical management, rather than a direct effect of a particular drug), antithyroid treatment, and cessation of steroid treatment. Other drugs that can cause hypoglycaemia include quinolones, pentamidine, quinine, β-blockers, and insulin-like growth factor [86]. Recurrent hypoglycaemia, or mixed hypoglycaemia/ketoacidosis, occur in some subjects labelled as having 'brittle diabetes'. Psychosocial factors may contribute to this presentation. Some patients may choose to run their blood glucose at near-hypoglycaemic levels, overcorrecting minor or transient high glucose concentrations.

Anecdotal reports of loss of hypoglycaemia awareness and increased risk of severe hypoglycaemia associated with conversion from one insulin species to another, most particularly after switching from animal to human insulin, were not confirmed in randomized trials. However, insulins do differ in terms of pharmacokinetics and lipid solubility. The randomized trials cannot exclude idiosyncratic or very small differences that might be clinically relevant to susceptible patients. Caution should always be exercised in converting patients from one insulin regimen to another and patient preference should be respected.

Hypoglycaemia may become a particular problem in pregnancy, where women with diabetes strive for very tight control, often introduced urgently. The therapeutic targets are tighter. Insulin sensitivity may be increased in the first 12 weeks of pregnancy, and, importantly, the insulin resistance of pregnancy stops immediately placental function ceases, so that protocols for insulin administration during delivery must anticipate and account for this (see Chapter 15.10.4, 'Diabetes in Pregnancy'). Loss of hypoglycaemia awareness and severe hypoglycaemia are common in pregnancy and can create a significant problem for the mother, although with no apparently negative impact on the fetus. Because of the attention to very tight glucose targets, including postprandially in pregnancy, review of eating patterns and, perhaps, introduction of routine between-meal snacking, may be necessary. The insulin resistance of late pregnancy resolves with delivery of the placenta, and labour ward protocols for blood glucose control should reflect this, allowing reduction in insulin administration immediately postpartum, to avoid maternal hypoglycaemia. Because of the imperfections in our ability to control blood glucose in the diabetic mother, their babies are at risk of hypoglycaemia, probably secondary to fetal hyperinsulinaemia before and during labour. Breast feeding is an energy-consuming activity and maternal insulin doses may need further reduction during lactation.

There is some controversial evidence for a genetic predisposition to IAH and severe hypoglycaemia, with the deletion (D-) allele of the ACE insertion/deletion (I/D) polymorphism and polymorphisms in or near the genes for the β1 and β2 adrenergic receptor implicated in adults with diabetes (e.g. [87–89]), although other studies, including one in young people have been negative [90, 91].

Consequences of Hypoglycaemia

Acute Effects

Acute episodes of hypoglycaemia resulting in cognitive impairment may result in embarrassing incidents, errors of judgement or performance, and accidents. Altered behaviour may manifest as aggression, or may mimic alcohol intoxication. If glucose levels fall low enough, drowsiness, coma, and/or epileptic seizures may result in injury and/or hospital admission.

Hypoglycaemia is a particular risk when undertaking activities that require concentration and fast reactions. Driving is a particular issue, usefully reviewed in [92]. In many countries, risk of hypoglycaemia is taken into account by driving licensing authorities and this has led to underreporting [93]. In the United Kingdom, hypoglycaemia is not accepted by courts as a valid medicolegal explanation for driving offences and the onus is on the individual to ensure that hypoglycaemia does not occur behind the steering wheel. Cox has found a relationship between experience of hypoglycaemia-related road traffic accidents and risk-taking behaviours in general [94]. Patients with a current history of severe hypoglycaemia while awake, or asymptomatic hypoglycaemia with high risk of severe hypoglycaemia, must be advised not to drive at all, and steps should be taken immediately to attempt to restore hypoglycaemia awareness and protection. While the UK licensing authorities suggest any normal blood glucose (5 mmol/L or more) is acceptable before driving, we advise patients with any degree of impaired awareness not to drive with a blood glucose less than 7 mmol/L, as there is evidence that blood glucose falls during driving. If a driver does suspect that he or she is hypoglycaemic during driving, he/she should stop the car, get out of the driver seat, and treat the hypoglycaemia. Because of evidence of delayed restoration of cognitive function after hypoglycaemia, driving should not be resumed for at least 45 minutes after an event.

Acute hypoglycaemia impairs formation of memory for events occurring at the time of the hypoglycaemia [95], and may impair consolidation of memory during sleep [82]. Prolonged hypoglycaemia may result in seizure [96]. Nocturnal hypoglycaemia may

also impair awareness of hypoglycaemia the next day [62, 63]. It may result in transient neurological sequelae, such as hemiplegia, which may mimic a cerebrovascular event (hemiplegic hypoglycaemia). Rarely, hypoglycaemia may be followed by cerebral oedema. Persistent neurological damage may follow very severe and prolonged hypoglycaemia, usually after a major insulin overdose.

Cardiac arrythmias are detected at plasma glucose concentrations of under 3 mmol/L (54 mg/dl) and may be associated with the autonomic activation, adrenaline responses and/or hypokalaemia that accompany insulin-induced hypoglycaemia [97, 98]. Prolongation of the qQTc interval is detected. The suggestion that some people may be predisposed to certain arrhythmias in hypoglycaemia raises the possibility of screening for risk in future [99].

Mortality resulting from acute hypoglycaemia is difficult to quantify accurately. Each year, 25–30 deaths are recorded in the United Kingdom as being directly related to hypoglycaemia. Population studies suggest that the actual rate may be higher. While diabetic ketoacidosis remains the main cause of acute diabetes-related death in young people, Cryer has quoted data reporting between 4% and 10% of deaths in people with type 1 diabetes are due to hypoglycaemia, but it is important to recognize that all the studies were in young people, with low rates of death from other causes [100]. A history of severe hypoglycaemia is associated with a significant increased mortality but not necessarily from hypoglycaemia itself in type 1 or type 2 diabetes, such that hypoglycaemia may be a marker of frailty and high risk of death from other causes [101, 102]. Nevertheless, recent studies show that death from hypoglycaemia, at least in people with type 1 diabetes, continues to occur probably at a stable rate. 'Dead in bed' syndrome, the event of a sudden, presumed cardiac, death occurring during sleep without evidence of seizure, was recently reported as accounting for 5% of deaths in young people with type 1 diabetes, very similar to the 4% originally described in 1991 [103].

In type 2 diabetes, and/or with elderly patients, presentations with severe hypoglycaemia have been associated with stroke and heart attack, although it cannot be established which event came first.

The costs of hypoglycaemia are difficult to calculate, as many episodes even of severe hypoglycaemia, are managed without involvement of emergency services. Furthermore the hidden costs in terms of loss of working time, for either the person with diabetes or his/her carers, and additional self-monitoring or contact with healthcare professionals are not easily recorded. One recent estimate of the direct and indirect costs of episodes of severe hypoglycaemia that led to hospitalization was approximately £1180 (US $1505) medical costs per person per episode, and £485 (US $620) additional indirect costs in terms of lost work [103]. There are no significant costs associated with non-severe episodes [104].

Chronic Effects of Hypoglycaemia

Neurological

The chronic effects of recurrent hypoglycaemia, from which apparently full recovery is made at the time, are unknown. Other repeated brain insults (e.g. trauma in boxers) may lead on to irreversible brain damage and it is known that neonatal hypoglycaemia (not usually diabetic) is associated with impaired brain function later. The developing brain is more susceptible to the effects of hypoglycaemia than the adult brain, and severe hypoglycaemia before the age of seven does seem to have impact on later performance of cognitive tasks [105, 106]. In contrast, in adolescents, as well as in adults, prospective studies of intensified therapy, with its associated increased risk of severe hypoglycaemia, have failed to show any cognitive deficit in those with recurrent hypoglycaemia [107], which is important information when deciding appropriate therapies to offer to minimize all complications of diabetes and its therapies, including cognitive performance. At the other end of the age range, one recent study has reported a high rate of impaired cognitive performance in older adults with type 1 diabetes, and impaired performance was associated with recent severe hypoglycaemia and impaired awareness of hypoglycaemia [108] and a study in older people with type 2 diabetes also links impaired performance and cognitive decline with hypoglycaemia [109]. However, the type 1 data also confirm earlier studies showing associations with chronic hyperglycaemia and microvascular complications, and the direction of any link between impaired cognitive function and hypoglycaemia remains unknown—certainly people with dementia struggle to adjust insulin regimens effectively and poor cognition may be a cause of problematic hypoglycaemia. A meta-analysis of seven studies in adults, with diabetes, comparing those with and without a history of severe hypoglycaemia, indicates impaired memory functions and impaired speed of processing in type 2 diabetes only, while other cognitive functions remain intact [110]. Anecdotally, people with hypoglycaemia often complain of poor memory, and animal studies of profound hypoglycaemia show damage in the hippocampus, but it is possible that the failure to make or consolidate memory mentioned earlier may be as much to blame as permanent damage. Depression and anxiety deteriorate performance of cognitive function tests.

Cardiovascular

Hypoglycaemia is a risk marker for cardiovascular mortality. Apart from acute effects, the stress response of hypoglycaemia is accompanied by a coagulopathy, with elevated circulating inflammatory markers, adhesion molecules, and altered platelet and endothelial function. These changes do not resolve immediately, but persist and it is hypothesized that they may contribute to vascular damage and increase the risk of macrovascular events [111]. The hypothesis is very plausible but remains to be proven.

Societal and Psychological

Experience of hypoglycaemia can have long-term impact on quality of life, both of the person with diabetes and their families. There is evidence for absenteeism after severe hypoglycaemia, with some evidence for reduced work attendance after non-severe hypoglycaemia, although those data are currently soft. As already discussed, recurrent exposure to plasma glucose values of <3 mmol/L (54 mg/dl) reduces awareness of subsequent hypoglycaemia, increasing risk of severe episodes. Both IAH and severe hypoglycaemia can cause embarrassment at work, or worse, with accidents and trauma, and may lead to loss of driving privileges and in severe cases access to children and dependents. Fear of hypoglycaemia is higher in people with experience of severe episodes but may be high even in people not at particularly high risk [112] and can of itself impair quality of life [113]. Healthcare professionals may have higher concern about

hypoglycaemia than people with diabetes and such fears may lead to reduced enthusiasm for intensifying diabetes treatments to minimize long-term diabetes complications. The impact of IAH and severe hypoglycaemia on family members can be devastating as they thrust the family into an extreme carer role [114].

Management of Hypoglycaemia

Any person taking insulin or an insulin secretagogue is at risk.

Management

Most acute episodes are self-managed with oral carbohydrate; 15 g is usually adequate and rapidly absorbed carbohydrate must be used. Glucose tablets, such as dextrosol (3.1 g/tablet), or 200 ml fresh fruit juice, are widely recommended, but any palatable sources of concentrated glucose can be used. A glycaemic response should occur within 10–15 minutes and treatment should be repeated after 15 min if glucose concentrations are still below 3.9 mmol/L. If refined glucose is used, most authorities recommend a starchy snack should also be ingested to avoid glucose levels dropping again after rapid absorption of available glucose from the stomach. This may be achieved by expediting the next meal or by deliberately ingesting 20 g complex carbohydrate on recovery. In children who may be reluctant to eat or drink, concentrated glucose preparations, such as Glucogel (32 g/100 ml), or honey can be squeezed inside the cheek. The absorption of this is probably by the concentrated glucose trickling back down the oropharynx, inducing reflex swallowing. Such therapies should not be attempted in the unconscious, when parenteral glucose or glucagon should be used.

Glucagon can be given by intravenous, subcutaneous, or intramuscular injection, with nasal glucagon being developed as an effective substitute. Glucagon mobilizes hepatic glycogen. It may be ineffective in liver disease, malnutrition, after repeated episodes of hypoglycaemia and extreme exercise, where there may be inadequate hepatic glycogen stores. It is available as a 1 mg injection pack for emergency intramuscular injection by a third party at home. The effect will be short-lived, so the recovered patient should take oral carbohydrate to prevent glucose levels from falling again. Intravenous glucose injection (70 ml of 20% glucose, or 125 ml of 10%) will rapidly elevate blood glucose levels and is the standard emergency department treatment for hypoglycaemia that cannot be managed by oral intake. 25 ml 50% glucose may be used if a central venous line is in place and fluid load is a concern. Having treated the acute episode, an attempt should be made to identify the underlying cause (missed meals, exercise) and to give advice/education as appropriate.

Patients who become hypoglycaemic on sulphonylurea therapy will require monitoring, and, probably, additional glucose support for up to 48 hours, as the hypoglycaemia is recurrent and prolonged. Advice for self-treating hypoglycaemia with ingestion of simple sugar should be reiterated, perhaps especially for patients also using alpha glucosidase inhibitors, in whom the ability of which to break down and absorb glucose from complex carbohydrate will be retarded. If there is a history of repeated hypoglycaemia and/or unawareness of hypoglycaemia, the patient should be formally reviewed by a specialist diabetes team.

Avoidance of Hypoglycaemia

The first step towards minimizing hypoglycaemia risk is to identify and quantify it. Documenting hypoglycaemia experience and hypoglycaemia awareness should be part of every annual review for anyone using exogenous insulin or an insulin secretagogue. The International Hypoglycaemia Study Group has produced a Red-Amber-Green rating for risk for use by healthcare professionals and one for use by people with diabetes (ihsgonline.com) The healthcare professional should establish how often the patient experiences severe, non-severe, and asymptomatic (from relatives and friends and from home records showing readings <3 mmol/L [54 mg/dl]) episodes and if there are times of particular risk. Nocturnal hypoglycaemia should be sought by asking the patient to check blood glucose occasionally at around 03.00 h (3 a.m.), the time of maximum insulin sensitivity and, in regimens using conventional background insulins (NPH and isophane), the time of peak insulin action. Hypoglycaemia awareness should be formally measured using one of the published validated questionnaires identifying people at high risk of severe hypoglycaemia [9, 70, 71, 115, 116]. Of these, the Gold score, asking the person to rate their awareness on a numeric scale from 1 to 7, and the Dose Adjustment For Normal Eating (DAFNE) tool asking if symptoms are usually perceived at or above a plasma glucose of 3 mmol/L (54 mg/dl) vs. below this value or not at all, are single item questionnaires, the latter being less subjective. Impaired awareness with either (defined as a score of 4 or more on the Gold score, or usual detection being at under 3 mmol/L or not occurring at all) predict significantly increased risk of severe hypoglycaemia in both type 1 (Figures 15.5.5.5 and 15.5.5.6) and in type 2 diabetes [71].

Inspection of home blood glucose monitoring records can help identify times of particular risk of hypoglycaemia, and it is worth going over recent records with the patient to confirm impressions and to ascertain awareness status. Using continuous glucose monitoring or intermittently monitored, retrospective ('Flash') glucose monitoring can also help identify hypoglycaemia patterns, especially overnight. Short-term use of blinded CGM has been used diagnostically, although not of proven utility.

Identification of risk factors may be made through a careful history and examination. Duration of diabetes, duration of insulin therapy, and degree of insulin secretory failure reflected by C-peptide concentrations are not amenable to treatment, but indicate someone at increased risk for severe hypoglycaemia, as do comorbidity and frailness. Discussing with the patient the timing of seminal activities, such as rising, insulin administration, eating, exercise, and alcohol ingestion, can identify possible contributors to non-physiological insulin replacement strategies. Intercurrent illnesses that might enhance hypoglycaemia risk, such as other hormone deficiencies, other drugs, malabsorption, etc., should be eliminated or treated.

Thereafter, there is a clear pathway for reducing hypoglycaemia risk [117]. In patients with type 2 diabetes, not dependent on insulin, review of medication, with consideration given to replacing insulin secretagogues with insulin sensitizers, incretin-based, or glycosuric therapies. For patients requiring insulin, particularly those with type 1 diabetes, structured education around flexible insulin use and self-adjustment of doses to achieve pre-meal

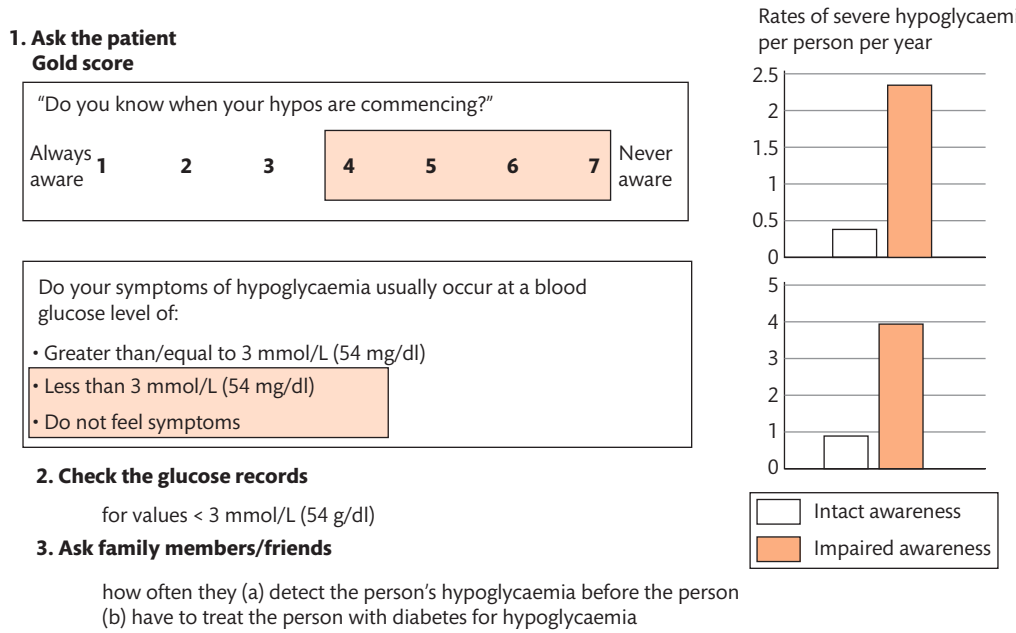


Figure 15.5.5.5 Two tools for assessing awareness of hypoglycaemia in people with diabetes by asking the patient. In the top panel, the Gold score is illustrated, below it the DAFNE tool. With each tool, the patient answers that define impaired awareness are shaded in grey. On the right the grey bars in each graph show the rates of severe hypoglycaemia in people scoring for impaired awareness with each tool compared to the rate in people who do not score for impaired awareness in white.
Data taken from references 70 and 9.

target blood glucose in the near-normoglycaemic range (most particularly stressing the lower limit of the desired range, normally 4 mmol/L (72 mg/dl)) has been reliably shown to reduce severe hypoglycaemia, with evidence for many such programmes, some specifically focused around hypoglycaemia avoidance, others more generally targeting improved glycaemic control and flexibility of lifestyle, demonstrated in a systematic review [118]. Such programmes depend on teaching about insulin pharmacodynamics, how to administer insulin flexibly around meals to be eaten, and common precipitants of hypoglycaemia such as exercise or alcohol. Patients may increase their carbohydrate intake (e.g. 30 g before and at 30-minute intervals during moderate exercise) and/or adjust insulin doses before and after exercise to avoid low blood glucose. During the exercise, continuation of some, perhaps reduced, basal insulin replacement is usually needed, but meal doses can be reduced considerably. Unaccustomed, vigorous, or prolonged exercise necessitates a further reduction in the overnight insulin dose (e.g. by between 15% and 50% in a well-controlled patient). Alcohol, with its ability to cause delayed hypoglycaemia, taken with exercise (e.g. dancing and drinking at parties) may be particularly

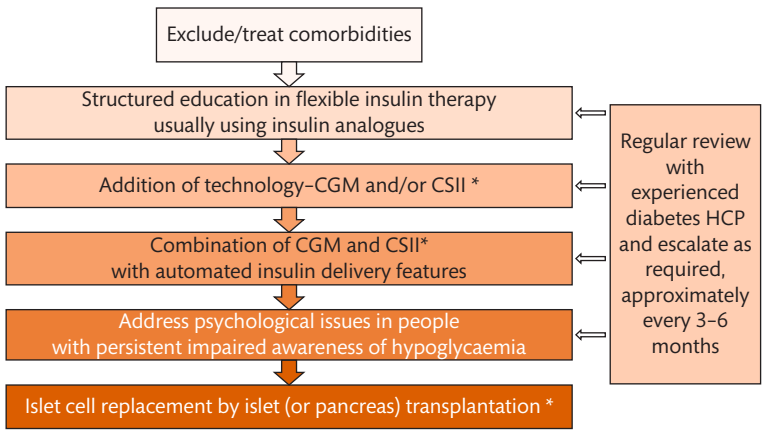


Figure 15.5.5.6 Treatment pathway for people with type 1 diabetes experiencing recurrent severe hypoglycaemia and with impaired awareness of hypoglycaemia. The patient should be carefully assessed and progressed along the pathway, with the option to increase a stage in between reviews if there is clearly no benefit. CSII = continuous subcutaneous insulin infusion; MDI = multiple daily insulin injection; SMBG = self-monitored blood glucose (finger pricking); RT-CGM = real-time continuous glucose monitoring with alarms. * psychological issues may play a role and psychological assessment and/or treatment may be appropriate at any stage when clinically indicated.
Adapted with permission from Choudhary P, Rickels MR, Senior PA, Vantyghem MC, Maffi P, Kay TW, Keymeulen B, Inagaki N, Saudek F, Lehmann R, Hering BJ. Evidence-informed clinical practice recommendations for treatment of type 1 diabetes complicated by problematic hypoglycemia. *Diabetes Care*. 2015 Jun;38(6):1016–29.

dangerous. In insulin-treated patients, consideration of the insulin kinetics will identify times of high risk of hypoglycaemia (e.g. 2–3 hours after meals and in the early hours of the night). Use of the short-acting insulin analogues for meals may help achieve good postprandial glucose levels without hypoglycaemia before the next meal, and, for some, reintroduction of between-meal snacking may need to be considered. Avoidance of postprandial administration of insulin is key—even with the fastest currently available insulins for subcutaneous use, a 15-minute gap between insulin administration and food ingestion gives optimal results [119, 120]. For nocturnal hypoglycaemia, use of fast-acting analogues for evening meals, 'peakless' analogues for background control, or, where conventional NPH or isophane insulins are used, taking the bedtime dose as late as possible or resuming a bedtime snack (including, perhaps, uncooked cornstarch) may all help. Use of the ultralong acting insulin degludec is associated with reduced hypoglycaemia, compared to other insulins, and although this lacks the flexibility of twice daily insulin detemir, many patients do not use this facility and for them, degludec may effectively reduce even severe hypoglycaemia risk (see Chapter 15.5.3, 'Insulins and Insulin Delivery Devices').

Adoption of technology is the next step, if the hypoglycaemia persists after proper education. The prior use of education is key, as this alone can be very beneficial and may account for some of the benefits originally ascribed to technology [121]. Use of automated insulin dose advisors linked to home glucose monitoring and in receipt of information about earlier insulin doses can reduce behaviours such as stacking (repeat 'correction' doses of fast-acting insulin in quick succession, or shortly after carbohydrate ingestion, without allowing the first correction dose to have worked—a minimum interval of two hours is recommended) but the strongest evidence, in people with type 1 diabetes, is for either continuous glucose monitoring (see Chapter 15.5.2, 'Glucose Monitoring and Sensing'), or for infusing background insulin using continuous subcutaneous insulin therapy (for more on insulin pump therapy, see Chapter 15.5.3), followed by sensor augmented pump therapy, in which the pump's insulin delivery is suspended automatically, when the sensor predicts a hypoglycaemic event. Recent studies indicate that use of real-time continuous glucose monitoring alone reduces severe hypoglycaemia risk, and such systems can be added to multiple daily injection therapy with benefit [122]. Interestingly, there is no evidence that such systems result in restored hypoglycaemia awareness, so the protection is only effective when the devices are being worn [123]. If all else fails, islet or whole organ pancreas transplantation provide protection from severe hypoglycaemia while islet function persists (see Chapter 15.6.4, 'Transplantation'). Such therapies cannot be undertaken lightly, but are available for patients with intractable problems.

The described pathway can be very successful in reducing risk of severe hypoglycaemia but there remain a few patients who fail to benefit. These are people with apparently treatment resistant IAH and recurrent severe hypoglycaemia and there is evidence that they have abnormal cognitive responses to hypoglycaemic events, as just described, that create barriers to their engaging successfully with the treatment pathway. Fears of hyperglycaemia and failure to prioritize treatment of hypoglycaemia are key cognitions which may need to be addressed through targeted psychological support before this group of people can successfully reduce their hypoglycaemia experience [124].

Summary

Hypoglycaemia is the most important acute complication of pharmacological therapies for diabetes that either replace or artificially drive endogenous insulin secretion. Awareness of all hypoglycaemia provides the patient using such therapies with essential protection against severe hypoglycaemia, in which the circulating glucose concentrations fall too low to support full, normal, cognitive function. Defects occur in the normal defences against hypoglycaemia as soon as therapies are started and progress over time. Exposure to mild hypoglycaemia can cause defective responses to and subjective recognition of, subsequent hypoglycaemia. Severe hypoglycaemia has associated adverse outcomes, which range from the embarrassing to the disastrous. Healthcare professionals should regularly check each patient's hypoglycaemia experience and adjust therapies to maximize the patient's ability to defend themselves. Avoidance of hypoglycaemia is, therefore, an important goal for diabetes therapies and can be achieved without deterioration of overall glycaemic control. To achieve this, however, requires good systems for patient education and engagement of the patient in these.

REFERENCES

1. Stuckey HL, Mullan-Jensen CB, Reach G, *et al*. Personal accounts of the negative and adaptive psychosocial experiences of people with diabetes in the second Diabetes Attitudes, Wishes and Needs (DAWN2) study. *Diabetes Care*, 2014; **37**: 2466–74.
2. Whipple AO. The surgical therapy of hyperinsulinism. *J Int Chir*, 1938; **3**: 237–276.
3. Abraham MB, Jones TW, Naranjo D, *et al*. Assessment and management of hypoglycemia in children and adolescents with diabetes. *Pediatr Diabetes*, 2018; **19** Suppl 27: 178–92.
4. Seaquist ER, Anderson J, Childs B, *et al*. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care*, 2013; **36**: 1384–95.
5. Amiel SA, Simonson DC, Tamborlane WV, DeFronzo RA, Sherwin RS. Rate of glucose fall does not affect counterregulatory hormone responses to hypoglycemia in normal and diabetic humans. *Diabetes*, 1987; **36**: 518–22.
6. Fanelli CG, Pampanelli S, Porcellati F, *et al*. Rate of fall of blood glucose and physiological responses of counterregulatory hormones, clinical symptoms and cognitive function to hypoglycaemia in Type I diabetes mellitus in the postprandial state. *Diabetologia*, 2003; **46**: 53–64.
7. International Hypoglycaemia Study Group. Glucose concentrations of less than 3.0 mmol/L (54 mg/dl) should be reported in clinical trials: a joint position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, 2017; **40**: 155–7; and *Diabetologia*, 2017; **60**: 3–6.
8. Beyond A1C Writing Group. Need for regulatory change to incorporate beyond A1C glycemic metrics. *Diabetes Care*, 2018; **41**: e92–4.
9. Hopkins D, Lawrence I, Mansell P, *et al* Improved biomedical and psychological outcomes 1 year after structured education in flexible insulin therapy for people with type 1 diabetes: the U.K. DAFNE experience. *Diabetes Care*, 2012; **35**: 1638–4.
10. Choudhary P, Lonnen K, Emery CJ, Freeman JV, McLeod KM, Heller SR. Relationship between interstitial and blood glucose

during hypoglycemia in subjects with type 2 diabetes. *Diabetes Technol Ther*, 2011; **13**: 1121–7.
11. UK Hypoglycaemia Study Group. Risk of hypoglycaemia in types 1 and 2 diabetes: effects of treatment modalities and their duration. *Diabetologia*, 2007; **50**: 1140–7.
12. Danne T, Nimri R, Battelino T, *et al*. International consensus on use of continuous glucose monitoring. *Diabetes Care*, 2017; **40**: 1631–40.
13. Henriksen MM, Andersen HU, Thorsteinsson B, Pedersen-Bjergaard U. Hypoglycemic exposure and risk of asymptomatic hypoglycemia in type 1 diabetes assessed by continuous glucose monitoring. *J Clin Endocrinol Metab*, 2018; **103**: 2329–35.
14. Casparie AF, Elving LD. Severe hypoglycemia in diabetic patients: frequency, causes, prevention. *Diabetes Care*, 1985; **8**: 141–5.
15. MacLeod KM, Hepburn DA, Frier BM. Frequency and morbidity of severe hypoglycaemia in insulin-treated diabetic patients. *Diabet Med*, 1993; **10**: 238–45.
16. Bali C, Gurdet C, Irsigler K. [Retrospective analysis of the incidence of severe hypoglycemia in 458 type 1 diabetic patients] [article in German]. *Acta Med Austriaca*, 1997; **24**: 165–9.
17. Mühlhauser I, Overmann H, Bender R, Bott U, Berger M. Risk factors of severe hypoglycaemia in adult patients with type I diabetes—a prospective population based study. *Diabetologia*, 1998; **41**: 1274–82.
18. ter Braak EW, Appelman AM, van de Laak M, Stolk RP, van Haeften TW, Erkelens DW. Clinical characteristics of type 1 diabetic patients with and without severe hypoglycemia. *Diabetes Care*, 2000; **23**: 1467–71.
19. Linkeschova R, Raoul M, Bott U, Berger M, Spraul M. Less severe hypoglycaemia, better metabolic control, and improved quality of life in Type 1 diabetes mellitus with continuous subcutaneous insulin infusion (CSII) therapy; an observational study of 100 consecutive patients followed for a mean of 2 years. *Diabet Med*, 2002; **19**: 746–51.
20. Johnson ES, Koepsell TD, Reiber G, Stergachis A, Platt R. Increasing incidence of serious hypoglycemia in insulin users. *J Clin Epidemiol*, 2002; **55**: 253–9.
21. Holstein A, Plaschke A, Egberts EH. Incidence and costs of severe hypoglycemia. *Diabetes Care*, 2002; **25**: 2109–10; and Holstein A, Plaschke A, Egberts EH. Clinical characterisation of severe hypoglycaemia—a prospective population-based study. *Exp Clin Endocrinol Diabetes*, 2003; **111**: 364–9.
22. Leese GP, Wang J, Broomhall J, *et al*. Frequency of severe hypoglycemia requiring emergency treatment in type 1 and type 2 diabetes: a population-based study of health service resource use. *Diabetes Care*, 2003; **26**: 1176–80.
23. Weinstock RS, Xing D, Maahs DM, *et al*. Severe hypoglycemia and diabetic ketoacidosis in adults with type 1 diabetes: results from the T1D Exchange clinic registry. *J Clin Endocrinol Metab*, 2013; **98**: 3411–19.
24. Pedersen-Bjergaard U, Pramming S, Heller SR, *et al*. Severe hypoglycaemia in 1076 adult patients with type 1 diabetes: influence of risk markers and selection. *Diabetes Metab Res Rev*, 2004; **20**: 479–86.
25. Zammitt NN, Geddes J, Warren RE, Marioni R, Ashby JP, Frier BM. Serum angiotensin-converting enzyme and frequency of severe hypoglycaemia in Type 1 diabetes: does a relationship exist?. *Diabet Med*, 2007; **24**: 1449–54.
26. Haynes A, Hermann JM, Miller KM, *et al*. Severe hypoglycemia rates are not associated with HbA1c: a cross-sectional analysis of 3 contemporary pediatric diabetes registry databases. *Pediatr Diabetes*, 2017; **18**: 643–50.
27. Janssen MM, Snoek FJ, de Jongh RT, Casteleijn S, Devillé W, Heine RJ. Biological and behavioural determinants of the frequency of mild, biochemical hypoglycaemia in patients with type 1 diabetes on multiple insulin injection therapy. *Diabetes Metab Res Rev*, 2000; **16**: 157–63.
28. Pedersen-Bjergaard U, Pramming S, Thorsteinsson B. Recall of severe hypoglycaemia and self-estimated state of awareness in type 1 diabetes. *Diabetes Metab Res Rev*, 2003; **19**: 232–40.
29. Leckie AM, Graham MK, Grant JB, Ritchie PJ, Frier BM. Frequency, severity, and morbidity of hypoglycemia occurring in the workplace in people with insulin-treated diabetes. *Diabetes Care*, 2005; **28**: 1333–8.
30. Khunti K, Alsifri S, Aronson R, *et al*. Rates and predictors of hypoglycaemia in 27 585 people from 24 countries with insulin-treated type 1 and type 2 diabetes: the global HAT study. *Diabetes Obes Metab*, 2016; **18**: 907–15.
31. The Diabetes Control and Complications Trial (DCCT) Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*, 1993; **329**: 977–86.
32. Bott S, Bott U, Berger M, Mühlhauser I. Intensified insulin therapy and the risk of severe hypoglycaemia. *Diabetologia*, 1997; **40**: 926–32.
33. DAFNE Study Group. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: Dose Adjustment For Normal Eating (DAFNE) randomised controlled trial. *BMJ*, 2002; **325**: 746.
34. Hoogma RP, Hammond PJ, Gomis R, Kerr D, Bruttomesso D, Bouter KP. Comparison of the effects of continuous subcutaneous insulin infusion (CSII) and NPH-based multiple daily insulin injections (MDI) on glycaemic control and quality of life: results of the 5-nations trial. *Diabet Med*, 2006; **23**: 141–7.
35. Shorr RI, Ray WA, Daugherty JR, Griffin MR. Incidence and risk factors for serious hypoglycemia in older persons using insulin or sulfonylureas. *Arch Intern Med*, 1997; **157**: 1681–6.
36. Miller CD, Phillips LS, Ziemer DC, Gallina DL, Cook CB, El-Kebbi IM. Hypoglycemia in patients with type 2 diabetes mellitus. *Arch Intern Med*, 2001; **161**: 1653–9.
37. Henderson JN, Allen KV, Deary IJ, Frier BM. Hypoglycaemia in insulin-treated Type 2 diabetes: frequency, symptoms and impaired awareness. *Diabet Med*, 2003; **20**: 1016–21.
38. Donnelly LA, Morris AD, Frier BM, *et al*. Frequency and predictors of hypoglycaemia in Type 1 and insulin-treated Type 2 diabetes: a population-based study. *Diabet Med*, 2005; **22**: 749–55.
39. Murata GH, Duckworth WC, Shah JH, Wendel CS, Mohler MJ, Hoffman RM. Hypoglycemia in stable, insulin-treated veterans with type 2 diabetes: a prospective study of 1662 episodes. *J Diabetes Complications*, 2005; **19**: 10–17.
40. Akram K, Pedersen-Bjergaard U, Carstensen B, Borch-Johnsen K, Thorsteinsson B. Frequency and risk factors of severe hypoglycaemia in insulin-treated Type 2 diabetes: a cross-sectional survey. *Diabet Med*, 2006; **23**: 750–6.
41. Abraira C, Colwell JA, Nuttall FQ, *et al*. Veterans Affairs Cooperative Study on glycemic control and complications in type II diabetes (VA CSDM). Results of the feasibility trial. Veterans Affairs Cooperative Study in Type II Diabetes. *Diabetes Care*, 1995; **18**: 1113–23.

42. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with 2 diabetes (UKPDS 33). *Lancet*, 1998; **352**: 837–53.
43. Holman RR, Thorne KI, Farmer AJ, *et al.* Addition of biphasic, prandial, or basal insulin to oral therapy in type 2 diabetes. *N Engl J Med*, 2007; **357**: 1716–30.
44. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, *et al.* Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*, 2008; **358**: 2545–59.
45. ADVANCE Collaborative Group, Patel A, MacMahon S, *et al.* Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*, 2008; **358**: 2560–72.
46. Duckworth W, Abraira C, Moritz T, *et al.* Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med*, 2009; **360**: 129–39.
47. Edridge CL, Dunkley AJ, Bodicoat DH, *et al.* Prevalence and incidence of hypoglycaemia in 532,542 people with type 2 diabetes on oral therapies and insulin: a systematic review and meta-analysis of population based studies. *PLoS One*, 2015; **10**: e0126427.
48. Donovan CM, Bohland M. Hypoglycemic detection at the portal vein: absent in humans or yet to be elucidated? *Diabetes*, 2009; **58**: 21–3.
49. Fioramonti X, Chrétien C, Leloup C, Pénicaud L. Recent advances in the cellular and molecular mechanisms of hypothalamic neuronal glucose detection. *Front Physiol*, 2017; **8**: 875.
50. Ogunnowo-Bada EO, Heeley N, Brochard L, Evans ML. Brain glucose sensing, glucokinase and neural control of metabolism and islet function. *Diabetes Obes Metab*, 2014; **16** Suppl 1: 26–32.
51. Mitrakou A, Ryan C, Veneman T, *et al.* Hierarchy of glycemic thresholds for counterregulatory hormone secretion, symptoms, and cerebral dysfunction. *Am J Physiol*, 1991; **260**(1 Pt 1): E67–74.
52. Matyka K, Evans M, Lomas J, Cranston I, Macdonald I, Amiel SA. Altered hierarchy of protective responses against severe hypoglycemia in normal aging in healthy men. *Diabetes Care*, 1997; **20**: 135–41.
53. McAulay V, Deary IJ, Frier BM. Symptoms of hypoglycaemia in people with diabetes. *Diabet Med,* 2001; **18**: 690–705.
54. Heller SR, Macdonald IA, Herbert M, Tattersall RB. Influence of sympathetic nervous system on hypoglycaemic warning symptoms. *Lancet*, 1987; **2**: 359–63.
55. Sommerfield AJ, Deary IJ, McAulay V, Frier BM. Moderate hypoglycemia impairs multiple memory functions in healthy adults. *Neuropsychology*, 2003; **17**: 125–32.
56. Bolli G, de Feo P, Compagnucci P, *et al.* Abnormal glucose counterregulation in insulin-dependent diabetes mellitus. Interaction of anti-insulin antibodies and impaired glucagon secretion. *Diabetes*, 1983; **32**: 134–41.
57. Graveling AJ, Noyes KJ, Allerhand MH, *et al* Prevalence of impaired awareness of hypoglycemia and identification of predictive symptoms in children and adolescents with type 1 diabetes. *Pediatr Diabetes*, 2014; **15**: 206–13.
58. Jaap AJ, Jones GC, McCrimmon RJ, Deary IJ, Frier BM. Perceived symptoms of hypoglycaemia in elderly type 2 diabetic patients treated with insulin. *Diabet Med*, 1998; **15**: 398–401.
59. Olsen SE, Asvold BO, Frier BM, Aune SE, Hansen LI, Bjørgaas MR. Hypoglycaemia symptoms and impaired awareness of hypoglycaemia in adults with Type 1 diabetes: the association with diabetes duration. *Diabet Med,* 2014; **31**: 1210–17.
60. Choudhary P, Lonnen K, Emery CJ, *et al.* Comparing hormonal and symptomatic responses to experimental hypoglycaemia in insulin- and sulphonylurea-treated type 2 diabetes. *Diabet Med*, 2009; **26**: 665–72.
61. Heller SR, Cryer PE. Reduced neuroendocrine and symptomatic responses to subsequent hypoglycemia after 1 episode of hypoglycemia in nondiabetic humans. *Diabetes*, 1991; **40**: 223–6.
62. Veneman T, Mitrakou A, Mokan M, Cryer P, Gerich J. Induction of hypoglycemia unawareness by asymptomatic nocturnal hypoglycemia. *Diabetes*, 1993; **42**: 1233–7.
63. Fanelli CG, Paramore DS, Hershey T, *et al.* Impact of nocturnal hypoglycemia on hypoglycemic cognitive dysfunction in type 1 diabetes. *Diabetes*, 1998; **47**: 1920–7.
64. Ovalle F, Fanelli CG, Paramore DS, Hershey T, Craft S, Cryer PE. Brief twice-weekly episodes of hypoglycemia reduce detection of clinical hypoglycemia in type 1 diabetes mellitus. *Diabetes*, 1998; **47**: 1472–9.
65. Davis SN, Mann S, Briscoe VJ, Ertl AC, Tate DB. Effects of intensive therapy and antecedent hypoglycemia on counterregulatory responses to hypoglycemia in type 2 diabetes. *Diabetes*, 2009; **58**: 701–9.
66. Spyer G, Hattersley AT, MacDonald IA, Amiel S, MacLeod KM. Hypoglycaemic counter-regulation at normal blood glucose concentrations in patients with well controlled type-2 diabetes. *Lancet*, 2000; **365**: 1970–4.
67. Maran A, Lomas J, Macdonald IA, Amiel SA. Lack of preservation of higher brain function during hypoglycaemia in patients with intensively-treated IDDM. *Diabetologia*, 1995; **38**: 1412–18.
68. Korzon-Burakowska A, Hopkins D, Matyka K, *et al.* Effects of glycemic control on protective responses against hypoglycemia in type 2 diabetes. *Diabetes Care*, 1998; **21**: 283–90.
69. Gonder-Frederick LA, Zrebiec JF, Bauchowitz AU, *et al.* Cognitive function is disrupted by both hypo- and hyperglycemia in school-aged children with type 1 diabetes: a field study. *Diabetes Care*, 2009; **32**: 1001–6.
70. Geddes J, Schopman JE, Zammitt NN, Frier BM. Prevalence of impaired awareness of hypoglycaemia in adults with Type 1 diabetes. *Diabet Med*, 2008; **25**: 501–4.
71. Schopman JE, Geddes J, Frier BM. Prevalence of impaired awareness of hypoglycaemia and frequency of hypoglycaemia in insulin-treated type 2 diabetes. *Diabetes Res Clin Pract*, 2010; **87**: 64–8.
72. Amiel SA, Sherwin RS, Simonson DC, Tamborlane WV. Effect of intensive insulin therapy on glycemic thresholds for counterregulatory hormone release. *Diabetes*, 1988; **37**: 901–7.
73. Fanelli CG, Pampanelli S, Porcellati F, Bolli GB. Shift of glycaemic thresholds for cognitive function in hypoglycaemia unawareness in humans. *Diabetologia*, 1998; **41**: 720–3.
74. Korytkowski MT, Mokan M, Veneman TF, Mitrakou A, Cryer PE, Gerich JE. Reduced beta-adrenergic sensitivity in patients with type 1 diabetes and hypoglycemia unawareness. *Diabetes Care*, 1998; **21**: 1939–43.
75. Cranston I, Lomas J, Maran A, Macdonald I, Amiel SA. Restoration of hypoglycaemia awareness in patients with long-duration insulin-dependent diabetes. *Lancet*, 1994; **344**: 283–7.
76. Fanelli CG, Epifano L, Rambotti AM, *et al.* Meticulous prevention of hypoglycemia normalizes the glycemic thresholds and magnitude of most of neuroendocrine responses to, symptoms of, and cognitive function during hypoglycemia in intensively treated patients with short-term IDDM. *Diabetes*, 1993; **42**: 1683–9.

77. Dagogo-Jack S, Rattarasarn C, Cryer PE. Reversal of hypoglycemia unawareness, but not defective glucose counterregulation, in IDDM. *Diabetes*, 1994; **43**: 1426–34.
78. Wiegers EC, Rooijackers HM, Tack CJ, Heerschap A, de Galan BE, van der Graaf M. Brain lactate concentration falls in response to hypoglycemia in patients with type 1 diabetes and impaired awareness of hypoglycemia. *Diabetes*, 2016; **65**: 1601–5.
79. Dunn JT, Choudhary P, Teh MM, *et al*. The impact of hypoglycaemia awareness status on regional brain responses to acute hypoglycaemia in men with type 1 diabetes. *Diabetologia*, 2018; **61**: 1676–87.
80. Nwokolo M, Amiel SA, O'Daly O, *et al*. Impaired awareness of hypoglycemia disrupts blood flow to brain regions involved in arousal and decision making in type 1 diabetes. *Diabetes Care*, 2019; **42:** 2127–35.
81. Graveling AJ, Frier BM. The risks of nocturnal hypoglycaemia in insulin-treated diabetes. *Diabetes Res Clin Pract*, 2017; **133**: 30–9.
82. Jauch-Chara K, Hallschmid M, Gais S, *et al*. Hypoglycemia during sleep impairs consolidation of declarative memory in type 1 diabetic and healthy humans. *Diabetes Care*, 2007; **30**: 2040–5.
83. Malawana M, Kerry S, Mathur R, Robson J. HbA1c and hypoglycaemia in intensively treated type 2 diabetes: a retrospective cohort study in primary care. *JRSM Open*, 2018; **9**: 2054270418773669.
84. Bussau VA, Ferreira LD, Jones TW, Fournier PA. A 10-s sprint performed prior to moderate-intensity exercise prevents early post-exercise fall in glycaemia in individuals with type 1 diabetes. *Diabetologia*, 2007; **50**: 1815–18.
85. Davey RJ, Bussau VA, Paramalingam N, *et al*. A 10-s sprint performed after moderate-intensity exercise neither increases nor decreases the glucose requirement to prevent late-onset hypoglycemia in individuals with type 1 diabetes. *Diabetes Care*, 2013; **36**: 4163–5.
86. Murad MH, Coto-Yglesias F, Wang AT, *et al*. Clinical review: drug-induced hypoglycemia: a systematic review. *J Clin Endocrinol Metab*, 2009; **94**: 741–5.
87. Davis WA, Brown SG, Jacobs IG, *et al*. Angiotensin-converting enzyme insertion/deletion polymorphism and severe hypoglycemia complicating type 2 diabetes: the Fremantle Diabetes Study. *J Clin Endocrinol Metab*, 2011; **96**: E696–700.
88. Pedersen-Bjergaard U, Nielsen SL, Akram K, *et al*. Angiotensin-converting enzyme and angiotensin II receptor subtype 2 genotypes in type 1 diabetes and severe hypoglycaemia requiring emergency treatment: a case cohort study. *Pharmacogenet Genomics*, 2009; **19**: 864–8.
89. Schouwenberg BJ, Coenen MJ, Paterson AD, *et al*. Pharmacogenetic determinants of impaired awareness of hypoglycemia in type 1 diabetes. *Genomics*, 2017; **27**: 323–8.
90. Freathy RM, Lonnen KF, Steele AM, *et al*. The impact of the angiotensin-converting enzyme insertion/deletion polymorphism on severe hypoglycemia in Type 2 diabetes. *Rev Diabet Stud*, 2006; **3**: 76–81.
91. Bulsara MK, Holman CD, van Bockxmeer FM, *et al*. The relationship between ACE genotype and risk of severe hypoglycaemia in a large population-based cohort of children and adolescents with type 1 diabetes. *Diabetologia*, 2007; **50**: 965–71.
92. Graveling AJ, Frier BM. Driving and diabetes: problems, licensing restrictions and recommendations for safe driving. *Clin Diabetes Endocrinol*, 2015; **1**: 8.
93. Pedersen-Bjergaard U, Færch L, Allingbjerg ML, Agesen R, Thorsteinsson B. The influence of new European Union driver's license legislation on reporting of severe hypoglycemia by patients with type 1 diabetes. *Diabetes Care*, 2015; **38**: 29–33.
94. Cox DJ, Penberthy JK, Zrebiec J, *et al*. Diabetes and driving mishaps: frequency and correlations from a multinational survey. *Diabetes Care*, 2003; **26**: 2329–3.
95. Warren RE, Zammitt NN, Deary IJ, Frier BM. The effects of acute hypoglycaemia on memory acquisition and recall and prospective memory in type 1 diabetes. *Diabetologia*, 2007; **50**: 178–85.
96. Buckingham B, Wilson DM, Lecher T, Hanas R, Kaiserman K, Cameron F. Duration of nocturnal hypoglycemia before seizures. *Diabetes Care*, 2008; **31**: 2110–12.
97. Novodvorsky P, Bernjak A, Chow E, *et al*. Diurnal differences in risk of cardiac arrhythmias during spontaneous hypoglycemia in young people with type 1 diabetes. *Diabetes Care*, 2017; **40**: 655–62.
98. Chow E, Bernjak A, Williams S, *et al*. Risk of cardiac arrhythmias during hypoglycemia in patients with type 2 diabetes and cardiovascular risk. *Diabetes*, 2014; **63**: 1738–47.
99. Campbell M, Heller SR, Jacques RM. Response to comment on Novodvorsky *et al*. Diurnal differences in risk of cardiac arrhythmias during spontaneous hypoglycemia in young people with type 1 diabetes. *Diabetes Care*, 2017; **40**: 655–62. *Diabetes Care*, 2018; **41**: e65–6.
100. Cryer P. Severe hypoglycaemia predicts mortality in type 1 diabetes. *Diabetes Care*, 2012; **35**: 1814–16.
101. McCoy RG, Van Houten HK, Ziegenfuss JY, Shah ND, Wermers RA, Smith SA Increased mortality of patients with diabetes reporting severe hypoglycemia. *Diabetes Care*, 2012; **35**: 1897–901.
102. Gagnum V, Stene LC, Jenssen TG, *et al*. Causes of death in childhood-onset Type 1 diabetes: long-term follow-up. *Diabet Med*, 2017; **34**: 56–63.
103. O'Reilly DJ, Burke N, Tarride JE, Hahn J, Nurkanovic L. Direct health-care costs and productivity costs associated with hypoglycemia in adults with type 1 and type 2 diabetes mellitus that participated in the Canadian hypoglycemia assessment tool program. *Can J Diabetes*, 2018; **42**: 659–63.
104. de Groot S, Enters-Weijnen CF, Geelhoed-Duijvestijn PH, Kanters TA. A cost of illness study of hypoglycaemic events in insulin-treated diabetes in the Netherlands. *BMJ Open*, 2018; **8**: e019864.
105. Northam EA, Rankins D, Lin A, *et al*. Central nervous system function in youth with type 1 diabetes 12 years after disease onset. *Diabetes Care*, 2009; **32**: 445–50.
106. He J, Li S, Liu F, *et al*. Glycemic control is related to cognitive dysfunction in Chinese children with type 1 diabetes mellitus. *J Diabetes*, 2018; **10**: 948–57.
107. Musen G, Jacobson AM, Ryan CM, *et al*.; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Impact of diabetes and its treatment on cognitive function among adolescents who participated in the Diabetes Control and Complications Trial. *Diabetes Care*, 2008; **31**: 1933–8.
108. Chaytor NS, Barbosa-Leiker C, Ryan CM, Germine LT, Hirsch IB, Weinstock RS. Clinically significant cognitive impairment in older adults with type 1 diabetes. *J Diabetes Complications*, 2019; **33**: 91–7.

109. Feinkohl I, Aung PP, Keller M, *et al.*; Edinburgh Type 2 Diabetes Study Investigators. Severe hypoglycemia and cognitive decline in older people with type 2 diabetes: the Edinburgh type 2 diabetes study. *Diabetes Care*, 2014; **37**: 507–15.
110. Chen YX, Liu ZR, Yu Y, Yao ES, Liu XH, Liu L. Effect of recurrent severe hypoglycemia on cognitive performance in adult patients with diabetes: a meta-analysis. *J Huazhong Univ Sci Technolog Med Sci*, 2017; **37**: 642–8.
111. International Hypoglycaemia Study Group. Hypoglycaemia, cardiovascular disease, and mortality in diabetes: epidemiology, pathogenesis, and management [published correction appears in *Lancet Diabetes Endocrinol* 2019 Jun; 7(6): e18]. *Lancet Diabetes Endocrinol,* 2019;7: 385–96.
112. Anderbro T1, Gonder-Frederick L, Bolinder J, *et al.* Fear of hypoglycemia: relationship to hypoglycemic risk and psychological factors. *Acta Diabetol*, 2015; **52**: 581–9.
113. Polonsky WH, Fisher L, Hessler D, Edelman SV. Identifying the worries and concerns about hypoglycemia in adults with type 2 diabetes. *J Diabetes Complications*, 2015; **29**: 1171–6.
114. Lawton J, Rankin D, Elliott J, Heller SR, Rogers HA, De Zoysa N, Amiel S. Experiences, views, and support needs of family members of people with hypoglycemia unawareness: interview study. U.K. NIHR DAFNE Study Group. *Diabetes Care*, 2014; **37**: 109–15.
115. Gold AE, MacLeod KM, Frier BM. Frequency of severe hypoglycemia in patients with type I diabetes with impaired awareness of hypoglycemia. *Diabetes Care*, 1994; **17**: 697–703.
116. Clarke WL, Cox DJ, Gonder-Frederick LA, Julian D, Schlundt D, Polonsky W. Reduced awareness of hypoglycemia in adults with IDDM. A prospective study of hypoglycemic frequency and associated symptoms. *Diabetes Care*, 1995; **18**: 517–22.
117. Choudhary P, Rickels MR, Senior PA, *et al.* Evidence-informed clinical practice recommendations for treatment of type 1 diabetes complicated by problematic hypoglycemia. *Diabetes Care*, 2015; **38**: 1016–29.
118. Yeoh E, Choudhary P, Nwokolo M, Ayis S, Amiel SA. Interventions that restore awareness of hypoglycemia in adults with type 1 diabetes: a systematic review and meta-analysis. *Diabetes Care*, 2015; **38**: 1592–609.
119. Slattery D, Amiel SA, Choudhary P. Optimal prandial timing of bolus insulin in diabetes management: a review. *Diabet Med*, 2018; **35**: 306–16.
120. Peters A, Van Name MA, Thorsted BL, Piltoft JS, Tamborlane WV. Postprandial dosing of bolus insulin in patients with type 1 diabetes: a cross-sectional study using data from the t1d exchange registry. *Endocr Pract*, 2017; **23**: 1201–9.
121. Little SA, Speight J, Leelarathna L, *et al.* Sustained reduction in severe hypoglycemia in adults with type 1 diabetes complicated by impaired awareness of hypoglycemia: two-year follow-up in the HypoCOMPaSS randomized clinical trial. *Diabetes Care*, 2018; **41**: 1600–7.
122. Heinemann L, Freckmann G, Ehrmann D, *et al.* Real-time continuous glucose monitoring in adults with type 1 diabetes and impaired hypoglycaemia awareness or severe hypoglycaemia treated with multiple daily insulin injections (HypoDE): a multicentre, randomised controlled trial. *Lancet*, 2018; **391**: 1367–77.
123. Choudhary P, Amiel SA. Hypoglycaemia in type 1 diabetes: technological treatments, their limitations and the place of psychology. *Diabetologia*, 2018; **61**: 761–9.
124. de Zoysa N, Rogers H, Stadler M, *et al.* A psychoeducational program to restore hypoglycemia awareness: the DAFNE-HART pilot study. *Diabetes Care*, 2014; **37**: 863–6.

15.6

Evidence-Based Management of Type 1 Diabetes

15.6.1 Strategies for the Management of Type 1 Diabetes

Peter Hammond and Fiona Campbell

Introduction

The person living with type 1 diabetes has to meet the challenge of integrating glucose self-management with everyday life 24/7 in order to ensure optimal glycaemic control, which not only reduces their risks of long-term complications but enhances their quality of life and allows them to maximize their potential.

The components of this self-management strategy are use of an appropriate insulin regimen, timely and effective glucose monitoring, and training to allow the person with diabetes to adjust their management in response to daily events which impact on blood glucose levels, such as diet, alcohol, exercise, stress, and illness. This self-management training will also need to be provided to parents and carers for children and adults who require support.

Supporting the person with diabetes to implement their self-management strategy requires the input of all members of the specialist multidisciplinary team, ideally comprising a diabetes consultant, diabetes specialist nurse, specialist dietitian, and psychologist with training in diabetes. Similar but separate teams will be needed to support adults and children with diabetes and appropriate local arrangements will be needed to ensure seamless transition for the person with diabetes between children's and adult services. Appropriate benchmarking can assess the performance of individual services. Regional, national, and international benchmarking schemes have been shown to improve overall biomedical outcomes.

Insulin Regimens

The cornerstone of effective blood glucose self-management is intensive insulin therapy or flexible insulin therapy with either multiple daily injections or insulin pump therapy. The aim of treatment is to allow the person with diabetes to target an HbA1c of 6.5% while minimizing the risk of hypoglycaemia.

A multiple daily injection, or basal-bolus, insulin regimen is the favoured starting injection regimen for the newly diagnosed adult or child with diabetes, although in some cases the newly diagnosed child may be started immediately on insulin pump therapy. The multiple daily injection regimen is designed to mimic as closely as possible, given the limitations of the subcutaneous route of administration, a physiological insulin profile, with a long-acting insulin providing the basal, or background, circulating insulin, and a rapid-acting insulin being used for bolus insulin to cover glucose excursions following meals or snacks, and to correct high glucose levels at any time.

Long-acting analogue insulins are preferred for people with type 1 diabetes as they deliver a more reliable diurnal insulin profile with less intraindividual variation in circulating insulin levels when compared to human insulin preparations. A once daily dosing regimen for the long-acting analogue insulin is often preferred by the person with diabetes to minimize the number of daily injections, but a twice-daily injection regimen potentially provides greater flexibility for those with a more varied lifestyle. This is reflected in NICE guidance for adults with type 1 diabetes which advises favouring a regimen using twice-daily insulin detemir over once daily insulin glargine. The twice-daily insulin detemir approach is being increasingly used in children. Since publication of this guidance there is more evidence to support the use of once daily insulin degludec, particularly in reducing the risk of nocturnal hypoglycaemia.

For an adult with type 1 diabetes the starting basal insulin dose can be estimated as:

- 16–24 units depending on the size of the individual, or
- 0.3 units/kg body weight.

A lower dose should be considered in the elderly and where intravenous insulin has been used prior to starting an injection regimen and the individual appears to be very sensitive to insulin. In the latter scenario ½ the units of intravenous insulin administered over the previous 24 hours can be used as a starting basal insulin dose.

In children, the dose and frequency of the long-acting analogue insulin varies depending on the child's age and weight. Smaller children are more sensitive to insulin. Children who are pubertal require increasing amounts of insulin. However, insulin dosages are very individual and two children of the same age or same weight often need quite different amounts. For children under 5 years, if insulin pump therapy is not considered appropriate, then a basal insulin dose of 0.25 units/kg/day, administered once daily, could be considered as a starting dose of insulin detemir.

Children between 5 and 10 years of age could be started on 0.3 units/kg/day. Children aged 11 years and above are often prescribed 0.4 units/kg/day insulin detemir, at this age given in two divided doses.

Rapid-acting insulin analogues are preferred as the bolus insulin for people with type 1 diabetes, the more reliable circulating insulin profile and shorter duration of action compared to soluble human insulin resulting in a lower postprandial glucose peak and less risk of late post-meal hypoglycaemia.

If the newly diagnosed person receives training, and subsequently develops an understanding of carbohydrate counting, then they can be trained to give rapid-acting insulin according to the amount of carbohydrate ingested. To do this they will need to know an insulin: carbohydrate ratio (ICR), the amount of insulin to be administered per 10 g carbohydrate or 1 carbohydrate portion (CP). In most adults, a reasonable starting point is an ICR of 1 unit: 10 g, with 1 unit: 5 g for those who are insulin resistant or are obese, and 1 unit: 15 g for those who are very insulin sensitive. Where this training takes place at a later date, fixed mealtime insulin doses can be used, with the total daily bolus dose equating to the total daily basal dose and divided according to the reported relative carbohydrate content of each meal. In children, an ICR of between 1unit: 5 g to 1 unit: 50 g may be required depending on the child's age, weight, and pubertal status.

Rapid-acting insulin can also be used to correct high blood glucose levels and the individual will need to determine by how much 1 unit insulin reduces blood glucose levels. This is called the insulin sensitivity factor (ISF). In adults, a starting point of 1 unit to bring blood glucose by 2.5–3 mmol/L is typical. In children, an ISF of between 1:2 and 1: 25 may be required depending upon the child's age, weight, and pubertal status,

Established rapid-acting insulin analogues represent modifications of the structure of human insulin. Recently faster-acting insulin aspart has become available, using the adjuvant nicotinamide to accelerate the absorption of insulin from the subcutaneous depot. There is no definitive evidence to indicate superiority of this insulin over other rapid-acting analogue insulins in a multiple daily injection regimen, but it may be considered for those people with diabetes who, despite giving bolus insulin 15 to 20 minutes pre-meals, are struggling with postprandial blood glucose spikes or late post-meal hypoglycaemia.

In certain circumstances a simpler insulin regimen may be considered:

- In the months after diagnosis the person with diabetes may go through a honeymoon phase where insulin requirements fall significantly, and on occasion, for a short time, insulin may not be needed at all. This typically lasts for 6 to 12 months, although rarely can last longer than this. During this honeymoon phase there may be a need for only basal or bolus insulin. However, it is important that the person with diabetes is aware that this is a transient phase and that their maintenance regimen is with multiple daily injections.
- The use of a twice-daily regimen using either a fixed pre-mixture of short and long-acting insulin or a free mixture of these insulins performed by the person with diabetes has been favoured in the past as a starting regimen for people with diabetes before improvements in insulins and glucose monitoring allowed for more effective and widespread use of multiple daily injection regimens. Now, use of a twice-daily regimen using pre-mixed insulin is reserved for those who would struggle to manage a multiple daily injection regimen, particularly those dependent on carers, such as people with learning difficulties, the frail elderly, or young people who are having significant difficulty adhering to this intensive insulin regimen. Such regimens are likely to achieve satisfactory blood glucose control only if the individual maintains a relatively unvarying lifestyle with consistent mealtime carbohydrate intake.

The alternative to a multiple daily injection regimen is continuous subcutaneous insulin infusion or insulin pump therapy. This therapy offers potential advantages over multiple daily injections as a method of delivering intensive insulin therapy. Insulin can be delivered more flexibly via the pump. In particular this includes a variable rate of basal insulin delivery, different types of bolus insulin profiles and an ease of giving bolus doses which favours more frequent use of correction boluses. Insulin pump therapy has been shown to be potentially superior to multiple daily injection therapy in reducing HbA1c and at the same time reducing the frequency of hypoglycaemia in both adults and children, but its use is limited by cost considerations and to a lesser extent personal preference. In one meta-analysis [1], that included 22 studies comparing insulin pump therapy with multiple daily injections, severe hypoglycaemia was reduced on pump therapy compared to injection therapy, with a rate ratio of 2.89 for randomized controlled trials and 4.34 for before/after studies. The mean difference in HbA_{1c} between treatments was 0.21% for randomized controlled trials and 0.72% for before/after studies. The greatest reductions in severe hypoglycaemia were seen in those with the highest baseline incidence and the higher the starting HbA1c the greater the reported reduction. Studies reporting long-term outcomes of pump therapy [2–3] indicate that sustained reductions in HbA1c averaging around 1% for at least 10 years are seen in those who start HbA1c because of inability to reach target HbA1c on multiple injection therapy.

In England and Wales NICE guidance for insulin pump therapy [4] recommends it should be used:

- For children younger than 12 years of age where multiple daily injection therapy is considered to be impractical or inappropriate

- For adults and children 12 years and older where:
  - Attempts to achieve target HbA1c levels with multiple daily injections result in the person experiencing disabling hypoglycaemia, the repeated and unpredictable occurrence of hypoglycaemia that results in persistent anxiety about recurrence and is associated with a significant adverse effect on quality of life, or
  - HbA1c levels have remained high at 8.5% (69 mmol/mol) or above on multiple daily injection therapy despite a high level of care

On the basis of the recommended criteria for using insulin pump therapy the expectation is that in time over 20% of adults and up to 50% of children with type 1 diabetes should be using this technology.

Safe and effective use of insulin pump therapy requires a significant commitment to self-management either by the person with diabetes or their carer and so there should be regular review to ensure that the insulin pump user is demonstrating adequate engagement with the therapy. While the user may choose to discontinue pump therapy themselves or it may be withdrawn due to an absence of clinical benefit, consideration should be given to discontinuing pump therapy if the user is not monitoring blood glucose levels adequately as this will prevent them identifying unrecognized pump failure. If pump therapy is withdrawn due to safety concerns, this should be done in a supportive way, with educational and psychological support provided and an option to restore pump therapy again if appropriate. In reality insulin pump therapy is rarely discontinued with reported continuation rates in excess of 95%.

Glucose Monitoring

Self-monitoring blood glucose is essential for ensuring that insulin dosing is optimized and the risk of hypoglycaemia is mitigated for the person with diabetes. A minimum of 4 capillary blood glucose tests per day is recommended, based on 3 pre-meal and 1 pre-bed test. In reality a greater number of tests is likely to be needed, for example to ensure safety when driving or exercising and to inform the person with diabetes or their carer about the impact of external factors such as alcohol and stress on blood glucose levels.

The NICE guidance for adults with type 1 diabetes [5] recognized this need for additional testing with a recommendation to support testing up to ten times a day if any of the following apply:

- The desired target for blood glucose control, measured by HbA1c, is not achieved
- The frequency of hypoglycaemic episodes increases
- There is a legal requirement to do so, such as before driving, in line with the Driver and Vehicle Standards Agency guidance
- During periods of illness
- Before, during, and after sport
- When planning pregnancy, during pregnancy, and while breastfeeding
- If there is a need to know blood glucose levels more than four times a day for other reasons, such as impaired awareness of hypoglycaemia or taking part in high-risk activities

In certain circumstances more than ten blood glucose tests per day may be required because of the person's lifestyle, for example driving for a long period of time, undertaking high-risk occupations, during long-distance travel, or if the person has impaired awareness of hypoglycaemia. There is good evidence that a greater frequency of blood glucose testing is associated with a lower HbA1c [6].

The NICE guidance for children with type 1 diabetes [7] recommends that an explanation should be given to children and young people with type 1 diabetes and their family members or carers (as appropriate) that an HbA1c target level of 48 mmol/mol (6.5%) or lower is ideal to minimize the risk of long-term complications. They should also be advised to routinely perform at least five capillary blood glucose tests per day [8].

Target blood glucose levels recommended by NICE [5, 7], based on achieving target HbA1c of 48 mmol/mol (6.5%) are:

- Fasting plasma glucose level (on waking) of 4–7 mmol/L for children, 5–7 mmol/L for adults
- Plasma glucose level of 4–7 mmol/L before meals at other times of the day
- Plasma glucose level of 5–9 mmol/L post-meals if checked after at least 90 minutes
- Plasma glucose level of at least 5 mmol/L when driving

The use of subcutaneous interstitial fluid glucose sensing allows an alternative to conventional self-monitoring of capillary blood glucose. These sensors allow glucose levels to be conveniently checked at any time and provide information as to the direction of change in blood glucose that is not evident when taking isolated capillary glucose readings. Two options for interstitial fluid glucose testing are available:

- Flash glucose monitoring—glucose levels are available when the sensing device is scanned
- Continuous glucose monitoring—glucose levels are continuously relayed to a display device and alarms/alerts can be set to warn of existing or impending hypo- or hyperglycaemia

Flash glucose monitoring is relatively inexpensive compared to continuous glucose monitoring and therefore can be regarded as a substitute for self-monitoring of blood glucose. It should be considered for all children with diabetes or for adults with diabetes needing to test eight times a day or more, or whose HbA1c remains high despite adequate blood glucose self-monitoring or who cannot perform capillary testing due to physical or psychological issues.

The indications for adding continuous glucose monitoring into the management strategy are similar to the indications for pump therapy. A meta-analysis of individual patient data from six randomized controlled trials [9] showed a mean difference in HbA1c of 0.3% in favour of those using continuous glucose monitoring rather than capillary glucose self-monitoring, and as for pump therapy the higher the baseline HbA1c the greater the reduction in HbA1c when using continuous glucose monitoring. Clinical trials indicate that in order to get the desired impact from continuous glucose monitoring it should be used at least 70% of the time.

In England and Wales NICE recommend continuous glucose monitoring [5, 7] should be:

- Offered to children and young people with diabetes who have:
  - Frequent severe hypoglycaemia, or
  - Impaired awareness of hypoglycaemia associated with adverse consequences (e.g. seizures or anxiety), or

  - An inability to recognize, or communicate about, symptoms of hypoglycaemia (for example, because of cognitive or neurological disabilities)
- Considered for children and young people with diabetes in the following situations:
  - Neonates, infants, and preschool children
  - Children and young people who undertake high levels of physical activity (e.g. sport at a regional, national, or international level)
  - Children and young people who have comorbidities (e.g. anorexia nervosa) or who are receiving treatments (e.g. corticosteroids) that can make blood glucose control difficult
- Considered for adults with diabetes who despite optimal use of insulin therapy and conventional blood glucose monitoring have:
  - More than one episode a year of severe hypoglycaemia with no obviously preventable precipitating cause
  - Complete loss of awareness of hypoglycaemia
  - Frequent (more than two episodes a week) asymptomatic hypoglycaemia that is causing problems with daily activities
  - Extreme fear of hypoglycaemia
  - HbA1c of 9% (75 mmol/mol) or higher that persists despite testing at least ten times a day. For this indication real-time continuous glucose monitoring should only be considered if HbA1c can be sustained at or below 7% (53 mmol/mol) and/or there has been a fall in HbA1c of 2.5% (27 mmol/mol)

The final recommendation is based on a cost-effectiveness analysis so cheaper continuous glucose monitoring would allow it to be used for lesser degrees of hyperglycaemia and require less stringent glycaemic targets for ongoing usage.

In addition to blood glucose monitoring the person with type 1 diabetes needs a way of monitoring for ketones in the event of high blood glucose levels. While urine ketone testing is feasible, the preferred option is capillary blood glucose testing for its accuracy and predictive value. Generally capillary blood ketones should be checked when blood glucose levels are greater than 15 mmol/l, with a level of 1.5–3 mmol/L mandating action such as administering an insulin correction bolus and increasing fluid intake, and a level of >3 mmol/L the threshold for considering hospital admission.

Combining Technologies for Optimizing Control

A strategy for introducing and combining technologies for optimizing blood glucose control is illustrated in Figures 15.6.1.1 and 15.6.1.2.

The use of sensor-augmented pump therapy, combining an insulin pump and continuous glucose monitoring, is of greatest benefit to the person with problematic hypoglycaemia, particularly impaired awareness of hypoglycaemia. The diabetes team should ensure that all people with type 1 diabetes are assessed at least annually for hypoglycaemic problems by recording the frequency of severe hypoglycaemia and identifying impaired hypoglycaemia awareness using the Clarke or Gold score.

Continuous glucose monitoring alone can alert the user to the existence or imminence of hypoglycaemia but automation of insulin delivery can protect against hypoglycaemia. A sensor-augmented pump system using a predictive low glucose suspend algorithm to temporarily stop insulin delivery when the sensor glucose trend indicates there is a likelihood of hypoglycaemia developing within 30 minutes, has been shown to prevent over 80% of probable hypoglycaemic episodes [9]. A hybrid closed-loop insulin delivery system, using the same technology, with a different algorithm which adjusts basal insulin delivery according to sensor glucose values, is now available in the United Kingdom and may be an option for the person with type 1 diabetes who has a persistently high HbA1c despite optimal use of insulin pump therapy or multiple daily injections supported by continuous glucose monitoring.

Structured Education

The person with type 1 diabetes needs training to effectively use intensive insulin therapy with appropriate blood glucose monitoring.

There is mixed evidence for the value of structured educations for adults with type 1 diabetes. While meta-analysis has failed to show that structured education based around carbohydrate counting has a consistently positive impact on HbA1c [10], the DAFNE (Dose Adjustment For Normal Eating) course is of proven benefit with a reduction in HbA1c of 1% (11 mmol/mol) 6 months after attending the course [11]. Longer-term follow-up shows HbA1c remains lower but to a lesser extent, but quality of life improvements are maintained [12]. It is possible that refresher education courses may help maintain the benefit from the initial course.

The NICE guideline for adults with type 1 diabetes [5] concluded that a structured education course based on the principles of the DAFNE course should be offered to all people with type 1 diabetes between 6 and 12 months after diagnosis. Carbohydrate counting training alone could be offered while awaiting more comprehensive structured education. The NICE guideline for children with type 1 diabetes [7] recommends that children and young people with type 1 diabetes and their family members or carers (as appropriate) are offered a continuing programme of education from diagnosis.

The NICE guideline for adults [5] gives specific criteria for an acceptable structured education course (described in greater detail in Chapter 15.5.1). The structured education course should cover at a minimum the following aspects of self-management:

- Nutrition, including carbohydrate counting, alcohol, healthy eating, and weight management
- Insulin dosing for meals, including insulin to carbohydrate ratios and correction doses
- Hypoglycaemia/hyperglycaemia management
- Exercise and blood glucose control
- Sick day rules
- Complications of diabetes
- Pregnancy and contraception
- Driving and travel

The group approach to structured education appears to have significant benefits over one-to-one training both in terms of efficiency for the diabetes educator and learning for the person with diabetes.

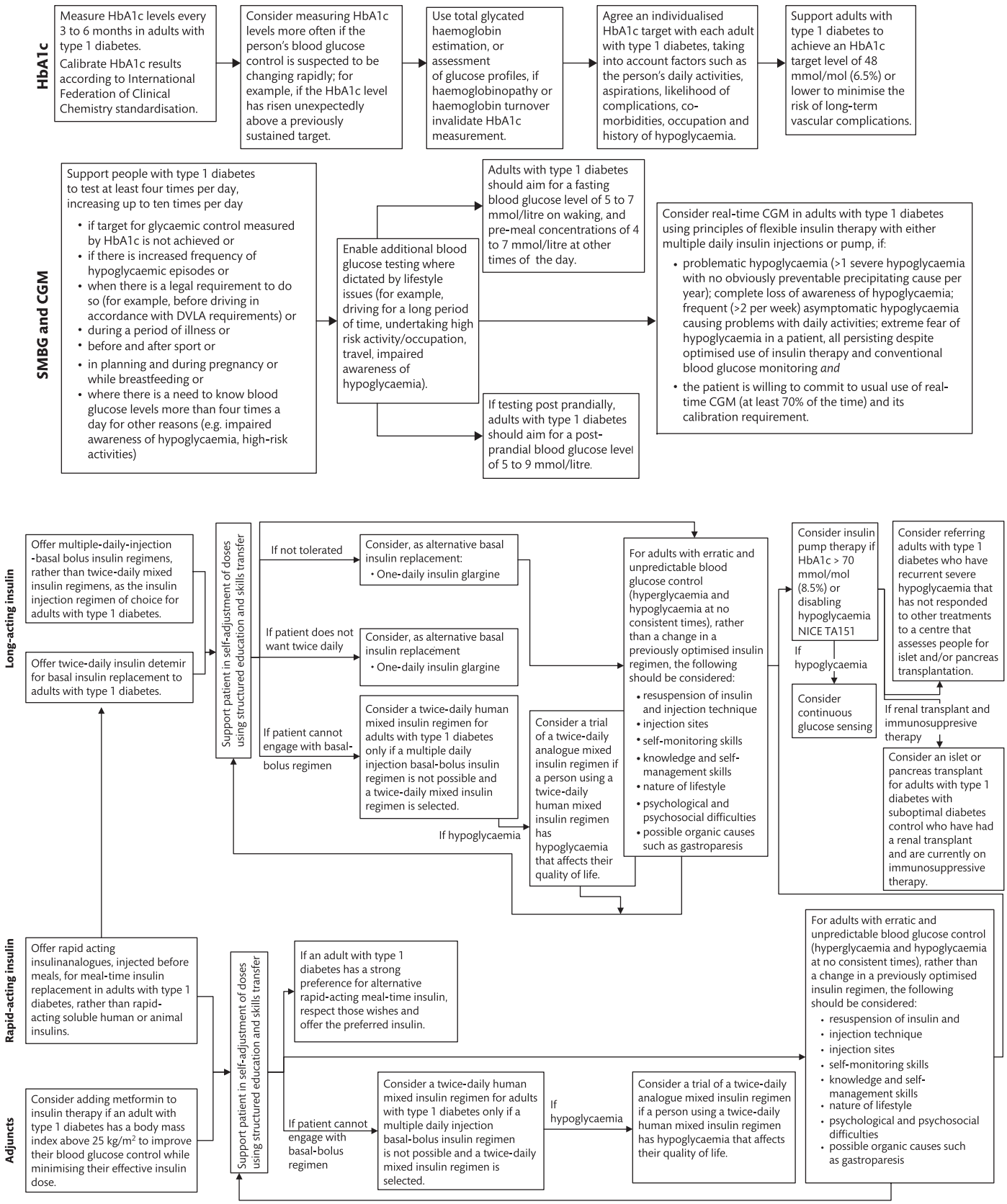


Figure 15.6.1.1 NICE guidance (NG17) for adults with type 1 diabetes (2015): (a) monitoring glycaemia; (b) choice of insulin regimen.

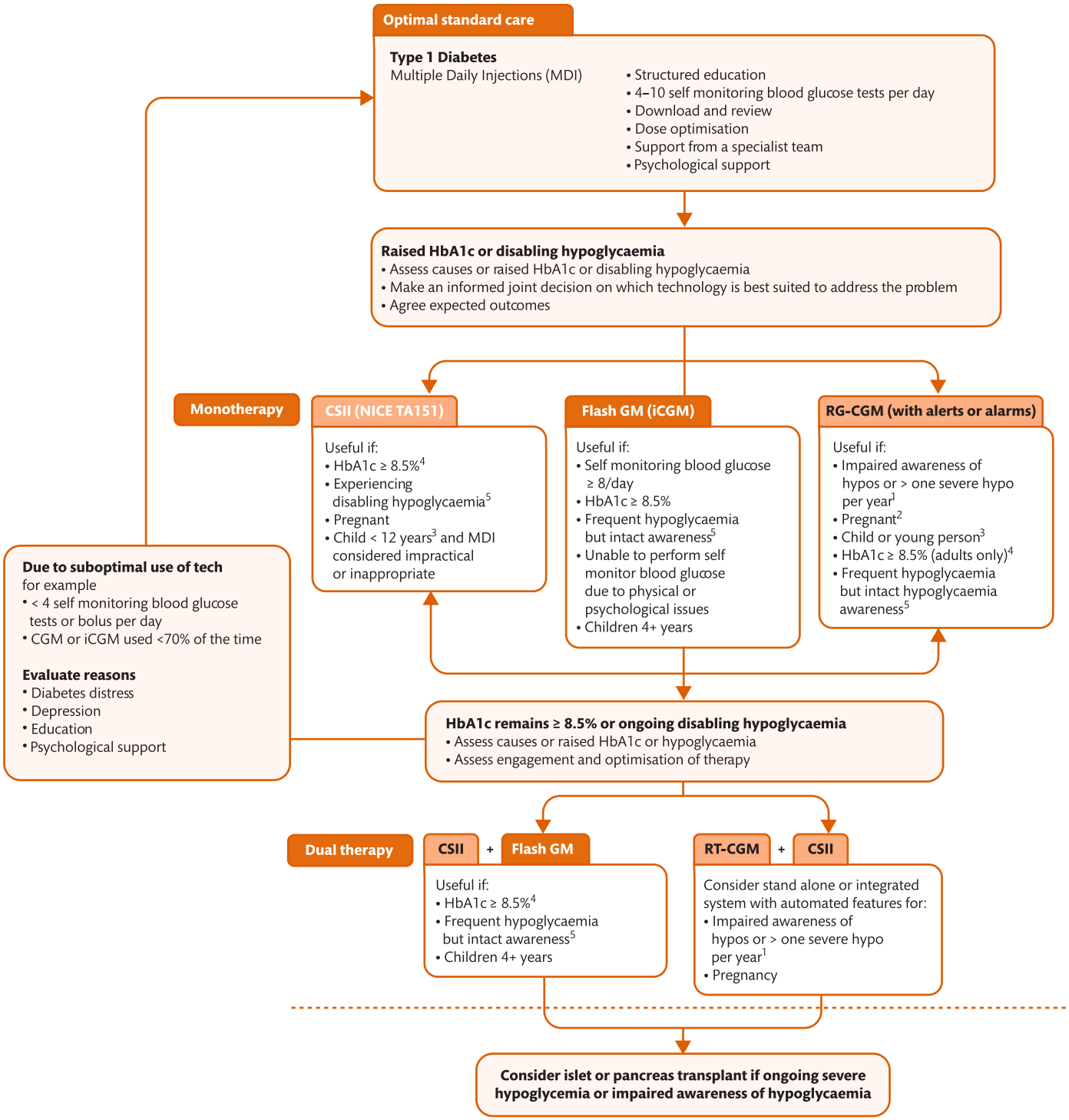


Figure 15.6.1.2 Diabetes UK algorithm for the use of technology for type 1 diabetes (2018)—includes flash glucose monitoring which was not considered in the NICE guidance (NG17).

Reproduced with permission from Type 1 Diabetes Technology: A Consensus Guideline. June 2018. https://www.diabetes.org.uk/resources-s3/2018-06/Diabetes%20UK%20consensus%20guideline.pdf

Certain subgroups may benefit from further education in specific areas:

- Preconceptional counselling has been shown to benefit women with type 1 diabetes preparing for pregnancy
- The person with diabetes who is adept at carbohydrate counting may wish to fine tune insulin dosing with meals by gaining an understanding of how fat and protein content additionally impact on post-meal blood glucose levels
- More detailed advice may be helpful for the person with type 1 diabetes who undertakes serious exercise as the impact of factors such as duration of activity, whether activity is aerobic or anaerobic or mixed, and whether competition is involved, will influence

the strategy for managing glucose levels and the balance between adjusting insulin doses and taking additional carbohydrate
- A DAFNE-pump course has been developed for adults so that pump users can be taught the same self-management principles as applied to insulin pump therapy
- Users of insulin pump therapy and/or flash/continuous glucose monitoring will benefit from training in uploading data from their device and retrospectively analysing the information obtained in order to make adjustments to insulin therapy and consider the impact of external factors, such as nutrition, exercise, and stress on blood glucose levels

People with type 1 diabetes suffering from problematic hypoglycaemia may benefit from specific behavioural training to help avoid hypoglycaemia

Trained educators supported by the wider diabetes multidisciplinary team should deliver these components of structured education and advanced training in diabetes self-management.

The Multidisciplinary Team

The multidisciplinary diabetes team is central to providing the necessary support for the individual person with diabetes to implement their self-management strategy.

The key individuals in this team are:

- Consultant trained in management of type 1 diabetes
- Diabetes specialist nurse
- Diabetes specialist dietitian
- Clinical psychologist with expertise in the management of long-term conditions, and preferably specialist expertise in the management of diabetes
- A diabetes educator, who is likely to be either a diabetes specialist nurse or dietitian

These team members will be the same for adult and paediatric services but the expertise will be either in adult or paediatric diabetes care, with the only overlap occurring at the transition between paediatric and adult services, which usually covers age 16 to 19 years.

Other non-core team members may include:

- Diabetes service administrator
- Diabetes pharmacist
- Diabetes podiatrist
- Diabetes technician, for supporting technology use
- Diabetes youth worker
- Diabetes family support worker
- Diabetes social worker

The multidisciplinary team should meet regularly, probably weekly, in order to discuss complex decisions about individual people with type 1 diabetes. This could include for example:

- People being considered for insulin pump therapy or continuous glucose monitoring
- People who are suffering from problematic hypoglycaemia
- People who are pregnant or planning to conceive
- People who have a persistently high HbA1c
- People who are repeatedly failing to attend the service
- People on insulin pump therapy who fulfil criteria for pump withdrawal
- People who are new to the diabetes service
- People who have been in-patients with diabetes-related complications
- People who have safeguarding related concerns

The multidisciplinary team should oversee improvements in service delivery, provision of appropriate structured education and audit/benchmarking activity.

The amount of contact that the individual with type 1 diabetes has with the multidisciplinary team will vary depending on their age, their degree of education and experience, the effectiveness of their blood glucose self-management, and whether they are using specific technologies.

Most guidance supports all people with type 1 diabetes being under the care of the multidisciplinary specialist team, even if they only make contact once a year. The paediatric best practice, year of care tariff in England mandates at least quarterly face-to-face contact with the team and a minimum of an additional eight contacts with the young person between clinic visits. There is no such mandated contact for those aged over 19, however NICE guidance clearly states standards of care that should be met for young people preparing to transition from paediatric to adult care:

- Allow sufficient time for young people with diabetes to familiarize themselves with the practicalities of the transition from paediatric to adult services because this improves clinic attendance.
- Agree specific local protocols for transferring young people with diabetes from paediatric to adult services.
- Base the decision about the age of transfer to the adult service on the young person's physical development and emotional maturity, and local circumstances.
- Ensure that transition from the paediatric service occurs at a time of relative stability in the young person's health and is coordinated with other life transitions.
- Explain to young people with type 1 diabetes who are preparing for transition to adult services that some aspects of diabetes care will change at transition.

In adult care, following pump therapy initiation, there are best practice guidelines [13] that provide a template for the first year's follow-up and NICE guidance for the pregnant woman [14] with diabetes advises an appropriate schedule of clinic follow-ups.

Contact with the multidisciplinary team is likely to evolve with greater use of technology to support self-management for the person with type 1 diabetes. The ability to upload insulin delivery and blood glucose monitoring data to the cloud, using platforms such as Glooko/Diasend, Tidepool, and Carelink and to communicate remotely can allow for more timely intervention to support changing insulin regimens when problems with blood glucose control occur. There is likely to be rise in remote advice giving, including use of virtual clinics, as the uptake of these technologies becomes more widespread, and healthcare professionals and users become more skilled in interpreting the information obtained from uploaded data.

Benchmarking and Audit

There is good evidence that benchmarking networks and regional, national and international audit programmes can help improve performance of services supporting the person with diabetes in implementing their self-management strategy.

The National Paediatric Diabetes Audit (NPDA) [15] is now in its fourteenth year and is used to compare the care and outcomes of all children and young people with diabetes receiving care from Paediatric Diabetes Units (PDUs) in England and Wales. There is 100% participation in this audit. It has been commissioned by the Health Quality Improvement Partnership (HQIP), funded by NHS England and the Welsh Government, and is managed by the Royal College of Paediatrics and Child Health (RCPCH).

The audit's aims are to monitor the incidence and prevalence of all types of diabetes among children and young people receiving care from a PDU in England and Wales and to establish which key care processes are being received by children and young people with diabetes. The audit enables benchmarking of performance against standards of care specified by the National Institute for Health and Care Excellence (NICE) guidance at PDU and national level and determines the prevalence and incidence of diabetes-related complications among children and young people with diabetes.

Although the NPDA has demonstrated marked improvements in paediatric diabetes outcomes in England and Wales over the last 8 years, it is felt by those working in the field of paediatric diabetes that in order to maintain the momentum towards the best international results and to meet current NICE Standards, new initiatives are required. As a result, the RCPCH has recently commenced the delivery of a high quality 3-year renewable programme of Peer Review and Quality Improvement involving all the diabetes units in England and Wales. The programme is a partnership between the RCPCH and the National Children and Young People's Diabetes Network, which has been supported by HQIP and NHS England and it complies with the Peer Review Framework from NHS Wales.

The programme comprises four elements:

- **An annual programme of self-assessment** against nationally-agreed measures which, alongside existing NPDA data provide a complete picture of a unit's compliance with national standards and good practice
- **A process of external verification** which provides assurance as to consistency and credibility of the assessments
- **A three-year cycle of peer review** visits providing consistent, authoritative assessment of every unit's achievements and practical challenges in improving the care for children with diabetes
- **A cascading managed programme of Quality Improvement** transforming whole diabetes teams and through them their network colleagues

It is hoped that by introducing this Quality Programme more rapid improvement in the variation of standards of care and outcomes will be achieved.

In adult practice the National Diabetes Audit has in recent years been extended to specialist secondary care diabetes services and has started to allow comparison between centres in terms of outcome for people with type 1 diabetes and the uptake of technology in these centres. There is considerable variation across England and Wales, for example the uptake of insulin pump therapy ranging between centres from 3% to over 30% of adults with type 1 diabetes [16] while evidence from the audit demonstrates that greater uptake of technology is associated with better outcomes for the person with diabetes. Given the potential improvements in care that could be facilitated by action to improve a centre's performance in the National Diabetes Audit, participation in such national audit programmes should be mandated for all centres providing specialist type 1 diabetes care.

REFERENCES

1. Pickup JC, Sutton AJ. Severe hypoglycaemia and glycaemic control in insulin dependant diabetes: meta-analysis of multiple daily insulin injections compared with continuous subcutaneous insulin infusion. *Diabet Med*, 2008; **25**: 765–74.
2. Beato-Vibora P, Yeoh E, Rogers H, Hopkins D, Amiel SA, Choudhary P. Sustained benefit of continuous subcutaneous insulin infusion and hypoglycaemia in adults with type 1 diabetes. *Diabet Med*, 2015; **32**: 1453–9.
3. Anyanwagu U, Olaoye H, Jennings P, *et al.* Continuous subcutaneous insulin infusion at Derby Teaching Hospitals: sustained benefits in glucose control. *Diabet Med*, 2017; **34**: 1154.
4. NICE. Continuous subcutaneous insulin infusion for the treatment of diabetes mellitus. NICE, London, 2008. Available at: https://www.nice.org.uk/Guidance/TA151 (accessed: 22 April 2020).
5. NICE guideline [NG17] Type 1 diabetes in adults: diagnosis and management. Last updated: July 2016. Available at: https://www.nice.org.uk/guidance/ng17 (accessed: 22 April 2020).
6. Miller KM, Beck RW, Bergenstal RM, *et al.* Evidence of a strong association between frequency of self-monitoring of blood glucose and hemoglobin A_{1c} levels in T1D exchange clinic registry participants. *Diabetes Care*, 2013; **36**: 2009–14.
7. NICE guideline [NG18] Diabetes (type 1 and type 2) in children and young people: diagnosis and management. Last updated: November 2016. Available at: https://www.nice.org.uk/guidance/ng18 (accessed: 22 April 2020).
8. Pickup JC, Freeman SC, Sutton AJ. Glycaemic control in type 1 diabetes during real time continuous glucose monitoring compared with self monitoring of blood glucose: meta-analysis of randomised controlled trials using individual patient data. *BMJ*, 2011; **343**: d3805.
9. Choudhary P, Olsen BS, Conget I, Welsh JB, Vorrink L, Shin JL. Hypoglycemia prevention and user acceptance of an insulin pump system with predictive low glucose management *Diabetes Technol Ther*, 2016; **18**: 288–91.
10. Bell KJ, Barclay AW, Petocz P, Colagiuri S, Brand-Miller JC. Efficacy of carbohydrate counting in type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2014; **2**: 133–40.
11. DAFNE Study Group. Training in flexible, intensive insulin management to enable dietary freedom in people with type 1 diabetes: dose adjustment for normal eating (DAFNE) randomised controlled trial. *BMJ*, 2002; **325**: 746–52.
12. Speight J, Amiel SA, Bradley C, *et al.* Long-term biomedical and psychosocial outcomes following DAFNE (dose adjustment for normal eating) structured education to promote intensive

insulin therapy in adults with sub-optimally controlled Type 1 diabetes. *Diabetes Res Clin Pract*, 2010; **89**: 22–9.
13. ABCD Diabetes Technology Network CSII Best Practice Guide. Continuous subcutaneous insulin infusion (CSII). A clinical guide for adult diabetes services. Available at: https://abcd.care/dtn-uk-best-practice-guides (accessed: 22 April 2020).
14. NICE guideline [NG3] Diabetes in pregnancy: management from preconception to the postnatal period. Last updated: August 2015. Available at: https://www.nice.org.uk/guidance/ng3 (accessed: 22 April 2020).
15. National Paediatric Diabetes Audit Report 2015–16. Available at: https://www.npda@rcpch.ac.uk (accessed: 22 April 2020).
16. National Diabetes Insulin Pump Audit, 2015–2016. NHS Digital. Available at: http://www.digital.nhs.uk/pubs/ndapump1516 (accessed: 22 April 2020).

15.6.2 Psychological and Behavioural Aspects of Type 1 Diabetes Management

Christel Hendrieckx and Jane Speight

Introduction

Type 1 diabetes is the most common chronic condition of childhood—but it can develop at any age, and around half of cases are diagnosed in adulthood. Age of onset can affect the individual trajectory in adapting to life with diabetes [1]. An adult who developed diabetes in childhood may not have memory of the diagnosis and its initial impact on life. While in adulthood, a diagnosis has been shown to have an immediate disruptive effect on physical and psychosocial aspects of life [2]. At any age, type 1 diabetes and its management is challenging, but there are also particular transitions in life that can impact on diabetes management and outcomes and vice versa. For example, puberty/adolescence [3] and pregnancy [4] often coincide with unstable glucose levels and elevated emotional distress, and conversely diabetes impacts negatively on the experience of these key life stages.

This chapter focuses on the behavioural implications and diabetes-related emotional problems that are commonly experienced by people with type 1 diabetes. Adequate assessment of the psychological impact informs the most effective treatment strategies. Also, these are best addressed by diabetes health professionals, as they are associated with diabetes self-care and outcomes.

Managing Type 1 Diabetes is Complex

The aim of type 1 diabetes self-care is to maintain glycated haemoglobin (HbA1c) within an individualized target range (e.g. <53 mmol/mol; <7%). This requires proactive self-management, including adjusting insulin according to glucose levels (checked several times a day) and the nutritional content of meals or snacks. Planning and adjustments are needed for physical activity (e.g. sports, walking, gardening) to avoid extreme glucose levels. However, many other factors affect glucose levels, including several beyond the person's control—e.g. hormones, weather, stress, illness—and the limitations of current treatments. The individual needs to develop advanced self-care knowledge, skills, and confidence, as well as acceptance of the need for relentless decision-making and lifelong perseverance in performing multiple daily self-care tasks. Therefore, managing glycaemic excursions is relentless and places a considerable behavioural burden on the individual.

Diabetes technologies (e.g. insulin pumps and continuous glucose monitoring) may reduce this burden and can offer greater flexibility of lifestyle and choice in treatment modalities. But such benefits are associated with the person's level of engagement and optimal use of the devices [5]. Also, technologies are not an option for all. This is partly due to their financial implications (both initial outlay and ongoing consumables, either for the individual or their healthcare service) and to personal preferences. These can relate to body image concerns (not wanting to be attached to a device), and that the devices currently require significant interaction to maintain 'control' and do not negate the need for human input. The 'human factor' is one reason why not all people with type 1 diabetes benefit from these novel technologies, which is evidenced by the substantial number not achieving HbA1c at target [6]. Despite significant advances in treatments, technologies, and screening procedures, type 1 diabetes can still lead to serious acute complications (e.g. hypoglycaemia, ketoacidosis), long-term complications (e.g. amputation, blindness), remains associated with a shorter life expectancy, and can have a significant psychological impact.

The relentless efforts and frustrations of managing type 1 diabetes are often underestimated by health professionals. Often their approach is still largely driven by the acute biomedical model of care in which the health professional's role is to prescribe treatment and the patient's role is to 'comply' with the prescriptions. This is a wholly inappropriate model of care for type 1 diabetes, which at the least needs to be replaced with a bio-psycho-social model of care [7]. Although health professionals are becoming more aware of the psychological impact of diabetes management and outcomes, there is still a gap between the holistic support people with diabetes expect and what they receive from their health professionals [8, 9]. If an analogy is needed, perhaps that of a mountain climber, supported by an expert team is appropriate—as health professionals can advise but, ultimately, the daily efforts of self-care to achieve individualized targets rest with the person.

Type 1 Diabetes Affects Emotional Well-Being

Emotional responses to the diagnosis, daily burden of self-care, and progress of the condition can be a barrier to achieve optimal

self-care, health, and quality of life. The individual's emotional reactions may be related directly to diabetes and its management (e.g. diabetes-specific distress or fear of hypoglycaemia). People with type 1 diabetes may also experience more general mental health problems such as depression, anxiety, or eating disorders. While these may or may not be caused by type 1 diabetes, these psychological problems can certainly impact on how the person manages their diabetes.

Diabetes Distress

Diabetes distress is a common emotional reaction to living with diabetes [10]. It is not a diagnosable mental health condition. If left unchecked, diabetes distress can become severe and/or place the person at risk of developing depression. Although there is some similarity between diabetes distress and clinical depression, they are distinct constructs and require different treatment strategies [11]. Due to the overlap between symptoms of depression and signs of suboptimal diabetes management (e.g. tiredness and lack of energy), it could well be that people with diabetes are diagnosed with depression, while they actually experience hyperglycaemia or diabetes distress [10]. With appropriate psychological assessment tools, this distinction can now be established [11].

Diabetes distress is defined by its content and severity, which can fluctuate from mild to severe. Around 20–30% of adults with type 1 diabetes experience severe diabetes distress [12]. Sources of diabetes distress in people with type 1 diabetes include, but are not limited to, feeling powerless, food/eating, hypoglycaemia, negative social reactions [13]. The impact of these stressors should not be underestimated because they affect the emotional and physical well-being of the person with type 1 diabetes. Greater diabetes distress is associated with suboptimal diabetes management (e.g. reduced physical activity, unhealthy eating, no or limited self-monitoring of glucose levels) and, as a consequence, elevated HbA1c and severe hypoglycaemia.

Few interventions have been developed specifically to reduce diabetes distress among people with type 1 diabetes—most were designed for people with type 2 diabetes or to reduce diabetes distress as a secondary outcome [14]. There is evidence that structured type 1 diabetes education programmes [15, 16] and psychological interventions [17, 18] are effective in reducing diabetes distress, particularly in those with greater baseline distress. Diabetes distress is reduced after completing the Dose Adjustment For Normal Eating (DAFNE) programme, and further analysis shows this to be driven by reductions in regimen-related distress [16]. In mixed samples of people with diabetes, monitoring diabetes distress routinely with a validated questionnaire and acting on the outcomes in clinical consultations reduced severe diabetes distress and improved general emotional well-being [19]. Finally, positive psychological factors such as optimism, self-esteem, self-efficacy, have been identified as protective against distress and elevated HbA1c [20].

Fear of Hypoglycaemia

Hypoglycaemia is an acute complication of some blood glucose lowering medications, particularly insulin, and is a real concern for people with type 1 diabetes and their families [8, 21]. It can occur unexpectedly, often overnight, with debilitating physical and cognitive symptoms and social embarrassment. *Fear* of hypoglycaemia is the specific and extreme fear arising from the risk and/or experience of hypoglycaemia. This differs from *concerns* about low glucose levels, which are constructive in keeping the person attentive to the early warning signs of hypoglycaemia.

Fear of hypoglycaemia can be triggered by impaired awareness of hypoglycaemia or experiencing a severe episode, which may have caused a trauma (e.g. a car accident or injury). Sometimes, a person's fear is disproportionate to their actual risk [22]. This may be due to limited understanding, skills, and confidence in preventing, recognizing, and treating low glucose levels. It may also coincide with personality traits, such as neuroticism or elevated trait anxiety [23].

To manage their fear, affected individuals may develop compensatory behaviours, such as reducing or omitting insulin doses, or continually snacking. These strategies are effective in reducing fear but when they become a habit, the consequence is persistent hyperglycaemia and increased risk of long-term complications. Other strategies include avoiding being alone or restricting activities that could lead to hypoglycaemia (e.g. refraining from exercise or travelling), which may be detrimental to their quality of life. Underlying reasons may include safety concerns and the fear of embarrassment in public. Furthermore, it is not uncommon that a severe hypoglycaemic event is mistaken for alcohol or drug abuse by those who are not familiar with diabetes, and this results in perceived or experienced stigma [24].

Fear of hypoglycaemia is often not recognized and addressed by health professionals. People with diabetes may not report experiencing hypoglycaemia for valid reasons [25]. They may blame themselves for not being able to prevent hypoglycaemia or fear the potential implications for their driving licence or work. Pathways to address hypoglycaemia and the associated fear include treatment optimization [26] and fear reduction strategies [27]. Psychoeducational intervention—addressing unhelpful beliefs and upskilling the person in their understanding of, and response to, early and personal symptoms of low glucose—has also been shown to be effective [28, 29].

Fear of Hyperglycaemia

At the other end of the spectrum is extreme concern about elevated glucose levels. Affected individuals may strive to keep their glucose levels in a very narrow range, by correcting even slightly elevated glucose levels. Clinical observation suggests that this strategy may be associated with 'perfectionist' tendencies. Such fear may also be a reaction to experiencing unpleasant symptoms of elevated glucose or extreme concerns about developing diabetic ketoacidosis or long-term complications. Irrespective of the cause, without an appropriate level of concern about low blood glucose, hyperglycaemia avoidance appears to be a risk factor for recurrent severe hypoglycaemia [30]. Fear of hyperglycaemia [31] and lack of concerns about hypoglycaemia or impaired awareness [30], have only recently been investigated and underlying mechanisms are not yet fully understood.

Clinical Implications

The psychological burdens just described are real, common and, moreover, a natural consequence of living with this lifelong, self-managed condition. They are barriers to optimal self-care and, as such, are likely to impact on diabetes outcomes. For these reasons,

these are best addressed, in first instance, by the diabetes health professionals as part of holistic individualized care [32].

Health professionals often believe that talking about emotions will divert the conversation from the medical agenda and that it will take up much more time. However, it is crucial to understand how a person is translating professional advice into their daily self-management plan, the barriers they are facing, their beliefs about what does and does not work for them, their preferences and satisfaction with treatment modalities (including use of technologies), and how they feel about their life with diabetes. Only in this way can a mutually agreed, personalized care plan be established and implemented with maximum benefit to the person with type 1 diabetes. Often, simply acknowledging these issues and showing empathy can ease diabetes distress [13] and strengthen the individual's engagement with their self-management.

Diabetes Management or Psychological Approach?

As noted earlier, negative emotions may result from a limited understanding of diabetes and its management, self-care skills, and/or confidence. In this case, the first-step approach should focus on diabetes education and training (preferably a structured programme with a person-centred philosophy and proven effectiveness, such as DAFNE), to enhance these skills and in turn reduce diabetes distress and/or fear of hypo/hyperglycaemia.

However, if the emotional problem (e.g. low/high fear unrelated to actual risk of hypoglycaemia) is due to inappropriate health beliefs or a personality trait, a psychological approach is recommended in parallel with continuous support from the diabetes team. Also, when diabetes-specific emotional problems coincide with other complex issues (e.g. anxiety disorder, eating disorder and/or social problems), it will require an integrated multidisciplinary approach including diabetes health professionals and mental health professionals.

Conclusion

This chapter has highlighted the relentless '24/7' behavioural burden and common negative emotional reactions experienced by people living with type 1 diabetes, which have significant implications for both their health and quality of life. Consequently, such problems need to be acknowledged and addressed, preferably in the first instance by the diabetes healthcare team supporting the individual. For those who experience negative emotional reactions, the intensity can vary over time. However, attentive observation by a concerned clinician is recommended, to enable early detection and addressing of psychological problems when these occur. Finally, it is important to note that the majority of individuals living with type 1 diabetes are coping well and enjoying life, albeit with the added complexity of also managing their condition.

REFERENCES

1. Lašaite L, Ostrauskas R, Žalinkevičius R, *et al.* Diabetes distress in adult type 1 diabetes mellitus men and women with disease onset in childhood and in adulthood. *J Diabetes Complications*, 2016; **30**: 133–7.
2. Due-Christensen M, Zoffmann V, Willaing I, *et al.* The process of adaptation following a new diagnosis of type 1 diabetes in adulthood: a meta-synthesis. *Qual Health Res*, 2018; **28**: 245–58.
3. Hagger V, Hendrieckx C, Sturt J, *et al.* Diabetes distress among adolescents with type 1 diabetes: a systematic review. *Curr Diab Rep*, 2016; **16**: 9.
4. Singh H, Murphy HR, Hendrieckx C, *et al.* The challenges and future considerations regarding pregnancy-related outcomes in women with pre-existing diabetes. *Curr Diab Repo*, 2013; **13**: 869–76.
5. Gonder-Frederick LA, Shepard JA, Grabman JH, Ritterband LM. Psychology, technology, and diabetes management. *Am Psychol*, 2016; **71**: 577.
6. Beck RW, Tamborlane WV, Bergenstal RM, *et al.* The T1D exchange clinic registry. *J Clin Endocrinol Metab*, 2012; **97**: 4383–9.
7. Speight J. Managing diabetes and preventing complications: what makes the difference. *Med J Aust*, 2013; **198**: 16–17.
8. Nicolucci A, Kovacs Burns K, Holt R, *et al.* Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national benchmarking of diabetes-related psychosocial outcomes for people with diabetes. *Diabet Med*, 2013; **30**: 767–77.
9. Holt RI, Nicolucci A, Kovacs Burns K, *et al.* Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national comparisons on barriers and resources for optimal care—healthcare professional perspective. *Diabet Med*, 2013; **30**: 789–98.
10. Gonzalez JS, Fisher L, Polonsky WH. Depression in diabetes: have we been missing something important? *Diabetes Care*, 2011; **34**: 236–9.
11. Snoek FJ, Bremmer MA, Hermanns N. Constructs of depression and distress in diabetes: time for an appraisal. *Lancet Diabetes Endocrinol*, 2015; **3**: 450–60.
12. Sturt J, Dennick K, Due-Christensen M, McCarthy K. The detection and management of diabetes distress in people with type 1 diabetes. *Curr Diabetes Rep*, 2015; **15**: 101.
13. Fisher L, Polonsky WH, Hessler DM, *et al.* Understanding the sources of diabetes distress in adults with type 1 diabetes. *J Diabetes Complications*, 2015; **29**: 572–7.
14. Sturt J, Dennick K, Hessler D, *et al.* Effective interventions for reducing diabetes distress: systematic review and meta-analysis. *International Diabetes Nursing*, 2015; **12**: 40–55.
15. Hopkins D, Lawrence I, Mansell P, *et al.* Improved biomedical and psychological outcomes 1 year after structured education in flexible insulin therapy for people with type 1 diabetes: the U.K. DAFNE experience. *Diabetes Care*, 2012; **35**: 1638–42.
16. Speight J, Holmes-Truscott E, Harvey DM, *et al.* Structured type 1 diabetes education delivered in routine care in Australia reduces diabetes-related emergencies and severe diabetes-related distress: the OzDAFNE program. *Diabetes Res Clin Pract*, 2016; **112**: 65–72.
17. Schmidt C, van Loon BJP, Vergouwen A, *et al.* Systematic review and meta-analysis of psychological interventions in people with diabetes and elevated diabetes-distress. *Diabet Med.* June 2018; **35**: 1157–72.
18. Fisher L, Hessler D, Polonsky WH, *et al.* T1-REDEEM: a randomized controlled trial to reduce diabetes distress among adults with type 1 diabetes. *Diabetes Care.* July 2018; **41**: 1862–9.
19. Snoek FJ, Kersch NYA, Eldrup E, *et al.* Monitoring of individual needs in diabetes (MIND)-2: follow-up data from the cross-national Diabetes Attitudes, Wishes, and Needs (DAWN) MIND study. *Diabetes Care*, 2012; **35**: 2128–32.

20. Yi JP, Vitaliano PP, Smith RE, *et al.* The role of resilience on psychological adjustment and physical health in patients with diabetes. *British Journal Health Psychology*, 2008; **13**: 311–25.
21. Kovacs Burns K, Nicolucci A, Holt RI, *et al.* Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national benchmarking indicators for family members living with people with diabetes. *Diabet Med*, 2013; **30**: 778–88.
22. Anderbro T, Gonder-Frederick L, Bolinder J, *et al.* Fear of hypoglycemia: relationship to hypoglycemic risk and psychological factors. *Acta Diabetologica*, 2015; **52**: 581–9.
23. Wild D, von Maltzahn R, Brohan E, *et al.* A critical review of the literature on fear of hypoglycemia in diabetes: implications for diabetes management and patient education. *Patient Educ Couns*, 2007; **68**: 10–15.
24. Browne JL, Ventura A, Mosely K, Speight J. 'I'm not a druggie, I'm just a diabetic': a qualitative study of stigma from the perspective of adults with type 1 diabetes. *BMJ Open*, 2014; **4**: e005625.
25. Östenson C, Geelhoed-Duijvestijn P, Lahtela J, *et al.* Self-reported non-severe hypoglycaemic events in Europe. *Diabet Med*, 2014; **31**: 92–101.
26. Choudhary P, Rickels MR, Senior PA, *et al.* Evidence-informed clinical practice recommendations for treatment of type 1 diabetes complicated by problematic hypoglycemia. *Diabetes Care*, 2015; **38**: 1016–29.
27. Vallis M, Jones A, Pouwer F. Managing hypoglycemia in diabetes may be more fear management than glucose management: a practical guide for diabetes care providers. *Curr Diabetes Rev*, 2014; **10**: 364–70.
28. De Zoysa N, Rogers H, Stadler M, *et al.* A psychoeducational program to restore hypoglycaemia awareness: the DAFNE-HART Pilot study. *Diabetes Care*, 2014; **37**: 863–6.
29. Cox DJ, Gonder-Frederick L, Ritterband L, *et al.* Blood glucose awareness training: what is it, where is it, and where is it going? *Diabetes Spectrum*, 2006; **19**: 43–9.
30. Rogers H, De Zoysa N, Amiel S. Patient experience of hypoglycaemia unawareness in type 1 diabetes: are patients appropriately concerned? *Diabet Med*, 2012; **29**: 321–7.
31. Singh H, Gonder-Frederick L, Schmidt K, *et al.* Assessing hyperglycemia avoidance in people with types 1 diabetes. *Diabetes Management*, 2014; **4**: 263–71.
32. Hendrieckx C, Halliday JA, Beeney LJ, Speight J. *Diabetes and Emotional Health: A Handbook for Health Professionals Supporting Adults with Type 1 and Type 2 Diabetes*. Canberra: National Diabetes Services Scheme; 2016.

15.6.3 Immunotherapy for Type 1 Diabetes

Colin Dayan and Danijela Tatovic

Introduction

It is now well-recognized that the appearance of multiple antibodies to beta-cell antigens in otherwise normoglycaemic individuals marks the start of the process that will eventually lead to clinical type 1 diabetes (T1D) [1, 2]—see Figure 15.6.3.1. This period of several years represents a window of opportunity for immunological intervention in order to halt the autoimmune process and preserve beta-cell function (see online only Figure 15.6.3.2). The most rapid loss of beta-cell function happens in the first 12 months after diagnosis [3], which represents a further opportunity for interventions leading to effective beta-cell preservation.

The Risk–Benefit Balance of Immunotherapy in T1D

Many diabetologists have been unwilling to consider immunotherapy as they believe the risks outweigh the benefits versus insulin therapy. However, with the new era of therapies that are very well-tolerated, the risks remain largely theoretical and need to be set against the shortcomings of even advanced insulin therapy. Even with optimal management, less than 30% of children and adults with T1D achieve a level of HbA1c which protects them from long-term complications (<7% (53 mmol/mol)) and for certain groups, such as teenagers, compliance and metabolic control is even worse. And even these levels of control are only achieved by acceptance of a level of hypoglycaemic events that impairs quality of life.

Most individuals have 10–20% of beta-cell function remaining at the time of diagnosis of T1D. Preservation of even 5% of beta-cell function has been shown to lower HbA1c by 1%, permit over 50% of people to reach target glycaemic levels, reduce hypoglycaemic risk by >50% and reduce long-term complications by 50% [4, 5]. Immunotherapy has the potential to preserve endogenous beta-cell function and thereby improve metabolic control even in poorly compliant individuals, with less hypoglycaemia and the possibility of intermittent rather than daily therapy [6–8].

Mechanisms of Immunomodulation

Non-Antigen-Specific Interventions

Generalized immunosuppression of the kind used in transplantation provides the evidence that immunotherapy can be effective in preserving beta-cell function [9, 10], but is rarely used in diabetes immunotherapy nowadays because the risk profile cannot generally be justified versus treatment with insulin alone. Selective immunosuppression (non-antigen-specific) disables selected immune pathways relevant to progression of T1D (see Figure 15.6.3.3). Using modern-day low-risk biologic therapies that target specific cell subsets or cytokines, this can now be achieved with limited side effects and little evidence of generalized immunosuppression. As a result, these agents are now widely used to treat other autoimmune conditions such as rheumatoid arthritis, iritis, psoriasis, and multiple sclerosis (see Table 15.6.3.1).

At the time of writing, six therapeutic interventions have shown positive results when given over a short-term period in recent-onset

Stage	Health	Stage 1	Stage 2	Stage 3
Risk factors	YES	YES	YES	YES
Antoimmunity	NO	YES	YES	YES
Metabolic status	Normal	Normal	Dysglycaemia	Diabetes
Symptomatic	NO	NO	NO	YES

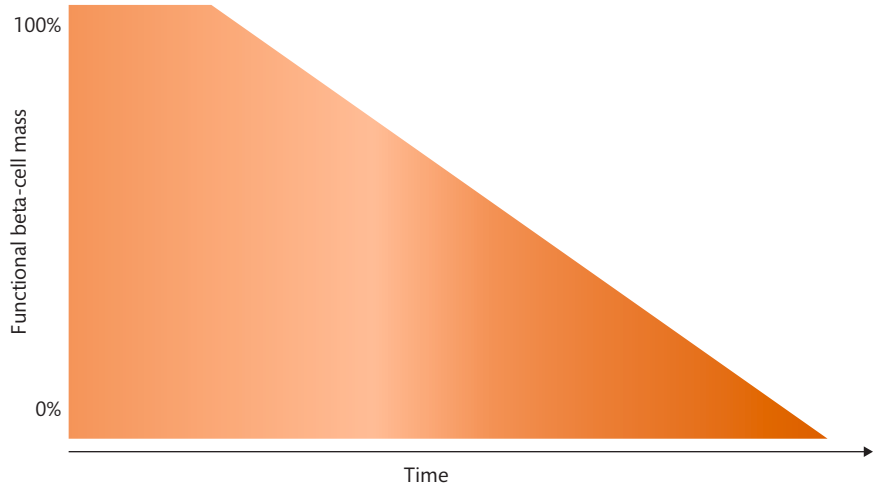


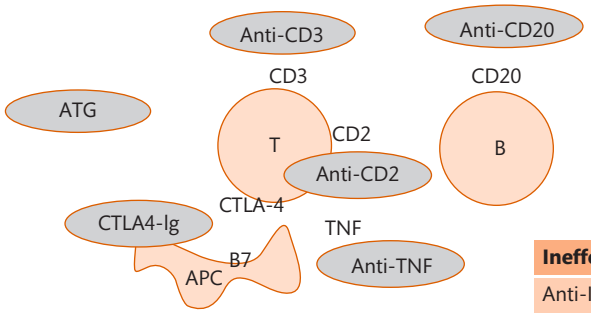
Figure 15.6.3.1 Immunotherapy of type 1 diabetes.
Source data from Insel RA, *et al.* Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care.* 2015 Oct;38(10):1964–74.

T1D, while others were ineffective (Figure 15.6.3.3). They target various molecules involved in the immune process—Figure 15.6.3.3 [11]—delaying C-peptide loss by 8.2–15.9 months in new-onset T1D [1]. The earlier cell depleting therapies—anti-CD3, anti-CD20 and antithymocyte globulin (ATG) are associated with infusion reactions and cytokine release, that are not apparent with anti-CD2, CTLA4-Ig, and anti-TNF.

Targeting T Cells—Anti-CD3, Antithymocyte Globulin (ATG)

Anti-CD3 monoclonal antibodies target CD3 molecules expressed on the surface of T cells as part of the T-cell receptor (TCR) complex. These antibodies have been successfully used to prevent transplant rejection, but had the significant side effect of causing cytokine storm [12]. Their modification to reduce Fc receptor binding affinity led to the reduction in side effects, but retained immunosuppression [13]. Two of these antibodies have been developed for use in immune interventions: teplizumab and otelixizumab.

Early phase studies of teplizumab showed promising results in slowing the loss of beta-cell function, in both children and adults, that extended up to 2 years of follow-up with most of the effect observed in patients with recent T1D onset [14–16]. The same effect was replicated in a larger study, although the trial did not meet its endpoints, which were focused on diabetes control (HbA1c level and insulin requirement) that is, unlike C-peptide level, dependent on manipulation by physicians and patients [17]. However, younger



Ineffective
Anti-Il-1
MMF+Daclizumab
IL-2 + Rapamycn
Interferon alpha

Figure 15.6.3.3 Agents with evidence of efficacy in T1D.

Table 15.6.3.1 Different types of immunotherapy

Treatment	Example	Risk of side effects
General immunosuppression	Drugs use for organ transplants	High
Selective immunosuppression/ immunoregulation	Includes newer drugs used for example in arthritis, skin diseases	Low
Antigen-specific immunotherapy	'Vaccines', depletion of autoantigen-specific T cells or boosting specific Treg	Very low

Source data from Rigby MR, Harris KM, Pinckney A, DiMeglio LA, Rendell MS, Felner EI, *et al.* Alefacept provides sustained clinical and immunological effects in new-onset type 1 diabetes patients. *The Journal of clinical investigation*. 2015;125(8):3285–96.

patients with good metabolic control and higher C-peptide levels had better therapeutic benefit [18].

Low-dose otelixizumab failed to show benefit in beta-cell preservation in a larger study [19], despite promising results at a higher dose in a smaller cohort, extending up to 18 months of follow-up after a single treatment, probably because the low dose was subtherapeutic [20, 21].

Low-dose ATG depletes T cells and has also been shown to preserve C-peptide in two separate studies (+/– G-CSF) [22, 23] although administration is associated with infusion reactions and transient serum sickness.

Targeting Effector Memory T Cells—Anti-CD2

Alefacept targets CD2 cell surface protein expressed on effector memory cells that are likely responsible for the autoimmune targeting of the pancreatic beta-cells. Indeed, Alefacept preserved C-peptide production (see online only **Figure 15.6.3.4**), lowered insulin use, and reduced hypoglycaemic events by 50% (see online only **Figure 15.6.3.5**) in patients with newly diagnosed T1D [24]. The effect extended over the 12 months after cessation of therapy [25].

Targeting T-Cell Costimulatory Signals—Anti-CTLA4

This approach is based on the observation that stimulation of T cells through the TCR complex without costimulation (interaction between CD28 on T cells and B7 molecules on antigen-presenting cells (APC)) leads to T cell inactivation. Modulating costimulation would, in theory, affect only T cells that have their antigen-specific receptors engaged. This is the basis of trials studying the CD28 homologue, cytotoxic T-lymphocyte antigen 4 immunoglobulin (CTLA4-Ig), also known as Abatacept. It selectively binds the B7 molecules on the APC, interfering with APC–T-cell interaction and inhibiting T cells.

Indeed, when Abatacept was trialled in patients with recent-onset T1D it slowed reduction in beta-cell function over 2 years [26]. The effect of this costimulation modulation was sustained for at least 1 year after cessation of treatment (i.e. 3 years after diagnosis of T1D [27]).

Targeting B Cells—Anti-CD20

The success of Rituximab, a monoclonal anti-CD20 antibody directed against B cells, in rheumatoid arthritis indicated the possibility of its use in diabetes. The rationale for this approach is to target a large population of cells with antigen-presenting properties, such as B cells. It slowed decline in C-peptide over 12 months in subjects with recent-onset T1D [28]. The rate of C-peptide decline shifted by 8.2 months in Rituximab-treated subjects [29].

Anti-TNF—TNF Receptor (Etanercept)

Etanercept is widely used to treat rheumatoid arthritis and other diseases and in a small study showed benefit in T1D [30]. Larger confirmatory studies are awaited.

Antigen-Specific Interventions

Antigen-specific immunotherapy (ASI), in which autoantigen molecules, or peptide epitopes derived from them, are delivered to patients to promote immune tolerance, is an attractive approach to treatment of antigen-specific autoimmune diseases as it offers the possibility of slowing the disease process without systemic immunosuppression The aim is to boost immunoregulation, by expanding antigen-specific regulator T cells, although some ASI appears to act predominantly by deleting or anergizing antigen-specific effector cells [31–33].

One of the first trials in the field of ASI interventions was a glutamic acid decarboxylase (GAD)–alum trial, in which the GAD65, one of the major autoantigens in T1D, was combined with an aluminium hydroxide adjuvant called alum, known to preferentially induce Th2 responses [34] and promote Treg expansion [11], with potential benefit in T1D reversal. Despite positive effects shown in early phase trials in patients treated within 6 months of diagnosis [35], the effect was not confirmed in subsequent larger studies [36–38]. A more recent study of intralymphatic administration appears promising [39].

Studies based on the autoantigen proinsulin—using either proinsulin DNA or T-cell target peptides derived from proinsulin have recently shown some promise but require replication [40, 41].

Currently, no immunotherapies are licensed for use in T1D, but results of many ongoing trials [42, 43] and networks have been established to increase recruitment of newly diagnosed and at risk individuals into trials (https://www.trialnet.org, http://www.address2.org, https://www.innodia.eu, https://www.type1diabetesresearch.org.uk).

For almost a hundred years, treatment for T1D has not advanced beyond insulin replacement. We are now on the brink of fundamentally changing the management of this devastating disease by using immunotherapy to preserve endogenous beta-cell function and make metabolic control substantially easier. It seems likely that non-antigen-specific therapies will be licensed first, but antigen-specific therapy may follow, offering the possibility of treating T1D in the preclinical phase and delaying or preventing the need for insulin therapy.

REFERENCES

1. Greenbaum C, Lord S, VanBuecken D. Emerging concepts on disease-modifying therapies in type 1 diabetes. *Curr Diab Rep*, 2017; **17**: 119.
2. Insel RA, Dunne JL, Atkinson MA, *et al.* Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*, 2015; **38**: 1964–74.

3. Greenbaum CJ, Beam CA, Boulware D, *et al.* Fall in C-peptide during first 2 years from diagnosis: evidence of at least two distinct phases from composite type 1 diabetes TrialNet data. *Diabetes*, 2012; **61**: 2066–73.
4. Lachin JM, McGee P, Palmer JP, Group DER. Impact of C-peptide preservation on metabolic and clinical outcomes in the Diabetes Control and Complications Trial. *Diabetes*, 2014; **63**: 739–48.
5. Sorensen JS, Johannesen J, Pociot F, *et al.* Residual beta-cell function 3–6 years after onset of type 1 diabetes reduces risk of severe hypoglycemia in children and adolescents. *Diabetes Care*, 2013; **36**: 3454–9.
6. Effect of intensive therapy on residual beta-cell function in patients with type 1 diabetes in the Diabetes Control and Complications Trial. A randomized, controlled trial. The Diabetes Control and Complications Trial Research Group. *Ann Intern Med*, 1998; **128**: 517–23.
7. Nathan DM, Cleary PA, Backlund JY, *et al.* Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*, 2005; **353**: 2643–53.
8. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA*, 2003; **290**: 2159–67.
9. Bougneres PF, Landais P, Boisson C, *et al.* Limited duration of remission of insulin dependency in children with recent overt type I diabetes treated with low-dose cyclosporin. *Diabetes*, 1990; **39**: 1264–72.
10. Jenner M, Bradish G, Stiller C, Atkison P. Cyclosporin A treatment of young children with newly-diagnosed type 1 (insulin-dependent) diabetes mellitus. London Diabetes Study Group. *Diabetologia*, 1992; **35**: 884–8.
11. Ryden AK, Wesley JD, Coppieters KT, Von Herrath MG. Non-antigenic and antigenic interventions in type 1 diabetes. *Hum Vaccin Immunother*, 2013; **10**: 838–46.
12. Chatenoud L, Legendre C, Ferran C, Bach JF, Kreis H. Corticosteroid inhibition of the OKT3-induced cytokine-related syndrome—dosage and kinetics prerequisites. *Transplantation*, 1991; **51**: 334–8.
13. Li B, Wang H, Dai J, Ji J, Qian W, Zhang D, *et al.* Construction and characterization of a humanized anti-human CD3 monoclonal antibody 12F6 with effective immunoregulation functions. *Immunology*, 2005; **116**: 487–98.
14. Herold KC, Hagopian W, Auger JA, *et al.* Anti-CD3 monoclonal antibody in new-onset type 1 diabetes mellitus. *N Engl J Med*, 2002; **346**: 1692–8.
15. Herold KC, Gitelman SE, Masharani U, *et al.* A single course of anti-CD3 monoclonal antibody hOKT3gamma1(Ala-Ala) results in improvement in C-peptide responses and clinical parameters for at least 2 years after onset of type 1 diabetes. *Diabetes*, 2005; **54**: 1763–9.
16. Herold KC, Gitelman SE, Ehlers MR, *et al.* Teplizumab (Anti-CD3 mAb) treatment preserves C-peptide responses in patients with new-onset type 1 diabetes in a randomized controlled trial: metabolic and immunologic features at baseline identify a subgroup of responders. *Diabetes*, 2013; **62**: 3766–74.
17. Sherry N, Hagopian W, Ludvigsson J, *et al.* Teplizumab for treatment of type 1 diabetes (Protege study): 1-year results from a randomised, placebo-controlled trial. *Lancet*, 2011; **378**: 487–97.
18. Hagopian W, Ferry RJ, Jr., Sherry N, *et al.* Teplizumab preserves C-peptide in recent-onset type 1 diabetes: two-year results from the randomized, placebo-controlled Protege trial. *Diabetes*, 2013; **62**: 3901–8.
19. Ambery P, Donner TW, Biswas N, Donaldson J, Parkin J, Dayan CM. Efficacy and safety of low-dose otelixizumab anti-CD3 monoclonal antibody in preserving C-peptide secretion in adolescent type 1 diabetes: DEFEND-2, a randomized, placebo-controlled, double-blind, multi-centre study. *Diabet Med*, 2014; **31**: 399–402.
20. Keymeulen B, Vandemeulebroucke E, Ziegler AG, *et al.* Insulin needs after CD3-antibody therapy in new-onset type 1 diabetes. *N Engl J Med*, 2005; **352**: 2598–608.
21. Keymeulen B, Walter M, Mathieu C, *et al.* Four-year metabolic outcome of a randomised controlled CD3-antibody trial in recent-onset type 1 diabetic patients depends on their age and baseline residual beta cell mass. *Diabetologia*, 2010; **53**: 614–23.
22. Haller MJ, Schatz DA, Skyler JS, *et al.* Low-dose anti-thymocyte globulin (ATG) preserves beta-cell function and improves HbA1c in new-onset type 1 diabetes. *Diabetes Care*, 2018; **41**: 1917–25.
23. Haller MJ, Gitelman SE, Gottlieb PA, *et al.* Anti-thymocyte globulin/G-CSF treatment preserves beta cell function in patients with established type 1 diabetes. *J Clin Invest*, 2015; **125**: 448–55.
24. Rigby MR, DiMeglio LA, Rendell MS, *et al.* Targeting of memory T cells with alefacept in new-onset type 1 diabetes (T1DAL study): 12 month results of a randomised, double-blind, placebo-controlled phase 2 trial. *Lancet Diabetes Endocrinol*, 2013; **1**: 284–94.
25. Rigby MR, Harris KM, Pinckney A, *et al.* Alefacept provides sustained clinical and immunological effects in new-onset type 1 diabetes patients. *J Clin Invest*, 2015; **125**: 3285–96.
26. Orban T, Bundy B, Becker DJ, *et al.* Co-stimulation modulation with abatacept in patients with recent-onset type 1 diabetes: a randomised, double-blind, placebo-controlled trial. *Lancet*, 2011; **378**: 412–19.
27. Orban T, Bundy B, Becker DJ, *et al.* Costimulation modulation with abatacept in patients with recent-onset type 1 diabetes: follow-up 1 year after cessation of treatment. *Diabetes Care*, 2014; **37**: 1069–75.
28. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H, *et al.* Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. *N Engl J Med*, 2009; **361**: 2143–52.
29. Pescovitz MD, Greenbaum CJ, Bundy B, *et al.* B-lymphocyte depletion with rituximab and beta-cell function: two-year results. *Diabetes Care*, 2014; **37**: 453–9.
30. Mastrandrea L, Yu J, Behrens T, *et al.* Etanercept treatment in children with new-onset type 1 diabetes: pilot randomized, placebo-controlled, double-blind study. *Diabetes Care*, 2009; **32**: 1244–9.
31. Mallone R, Brezar V, Boitard C. T cell recognition of autoantigens in human type 1 diabetes: clinical perspectives. *Clin Dev Immunol*, 2011; **2011**: 513210.
32. Tian J, Kaufman DL. Antigen-based therapy for the treatment of type 1 diabetes. *Diabetes*, 2009; **58**: 1939–46.
33. von Herrath M, Peakman M, Roep B. Progress in immune-based therapies for type 1 diabetes. *Clin Exp Immunol*, 2013; **172**: 186–202.
34. Brewer JM, Conacher M, Satoskar A, Bluethmann H, Alexander J. In interleukin-4-deficient mice, alum not only generates T helper 1 responses equivalent to freund's complete adjuvant, but continues to induce T helper 2 cytokine production. *Eur J Immunol*, 1996; **26**: 2062–6.
35. Ludvigsson J, Faresjo M, Hjorth M, *et al.* GAD treatment and insulin secretion in recent-onset type 1 diabetes. *N Engl J Med*, 2008; **359**: 1909–20.

36. Ludvigsson J, Krisky D, Casas R, *et al.* GAD65 antigen therapy in recently diagnosed type 1 diabetes mellitus. *N Engl J Med*, 2012; **366**: 433–42.
37. Elding Larsson H, Lundgren M, Jonsdottir B, Cuthbertson D, Krischer J, Di A-ITSG. Safety and efficacy of autoantigen-specific therapy with 2 doses of alum-formulated glutamate decarboxylase in children with multiple islet autoantibodies and risk for type 1 diabetes: a randomized clinical trial. *Pediatr Diabetes*, 2017; **19**: 410–19.
38. Wherrett DK, Bundy B, Becker DJ, *et al.* Antigen-based therapy with glutamic acid decarboxylase (GAD) vaccine in patients with recent-onset type 1 diabetes: a randomised double-blind trial. *Lancet*, 2011; **378**: 319–27.
39. Ludvigsson J, Wahlberg J, Casas R. Intralymphatic injection of autoantigen in type 1 diabetes. *N Engl J Med*, 2017; **376**: 697–9.
40. Roep BO, Solvason N, Gottlieb PA, *et al.* Plasmid-encoded proinsulin preserves C-peptide while specifically reducing proinsulin-specific CD8(+) T cells in type 1 diabetes. *Sci Transl Med*, 2013; **5**: 191ra82.
41. Alhadj Ali M, Liu YF, Arif S, *et al.* Metabolic and immune effects of immunotherapy with proinsulin peptide in human new-onset type 1 diabetes. *Sci Transl Med*, 2017; **9**: pii: eaaf7779.
42. Atkinson MA, Roep BO, Posgai A, Wheeler DCS, Peakman M. The challenge of modulating beta-cell autoimmunity in type 1 diabetes. *Lancet Diabetes Endocrinol*, 2019; **7**: 52–64.
43. Roep BO, Wheeler DCS, Peakman M. Antigen-based immune modulation therapy for type 1 diabetes: the era of precision medicine. *Lancet Diabetes Endocrinol*, 2019; **7**: 65–74.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

15.6.4 Transplantation (Islet and Solid Organ)

Anneliese Flatt, Martin Drage, Chris Callaghan, and Peter Senior

Historical Perspectives and Context

The first reports of unsuccessful attempts at transplantation in the treatment of diabetes, initially with fragments of pancreas from sheep and later from human cadavers, predate the discovery of insulin when death from ketoacidosis was inevitable (reviewed by Shapiro, 2007 [1]). After the discovery of insulin, type 1 diabetes was no longer fatal but chronic complications and hypoglycaemia emerged as, and remain, key care gaps and areas of focus for beta-cell replacement therapy.

In the 1960s pancreas transplantation evolved with progress in solid organ transplantation including advances in effective immunosuppression. The first deceased donor vascularized pancreas transplant was performed in combination with a kidney transplant at the University of Minnesota in 1966 with the goal of protecting the donor kidney in a patient with type 1 diabetes and diabetic nephropathy. Technical advances have addressed the high morbidity and mortality due to graft thrombosis, sepsis and anastomotic leaks which were common after early procedures [1].

Successful clinical translation of islet transplantation took longer. The use of enzymatic digestion to isolate islets effectively and identification of the intrahepatic portal site as optimal for transplantation by Moskalewski and Lacy were key milestones paving the way for the first clinical islet transplant performed by Sutherland in Minnesota in 1974. Refinement of isolation techniques including Ricordi's semi-automatic digestion chamber improved islet yield (potency) and purity (safety). Early results of islet allotransplantation for type 1 diabetes were far surpassed by outcomes of islet autotransplantation following pancreatectomy, highlighting the dual challenges of overcoming auto- and alloimmunity [1]. Prior to 2000, insulin independence rates at one year following pancreatic islet allotransplantation were only 8% [2]. The 'Edmonton Protocol' demonstrated improved outcomes through transplantation of sufficient islet mass from 2 to 4 donor pancreases over several procedures, using steroid-free immunosuppression to achieve one-year insulin independence in the first seven consecutive subjects [3]. Longer term results were less impressive with 5-year insulin independence rates of only 10%, although persistent graft function (C-peptide positivity) afforded ongoing clinical benefit through improved glycaemic control, reduced variability and liberation from previously disabling hypoglycaemia [4].

Restoration of endogenous insulin secretion through islet or pancreas transplantation can achieve glycemia superior to that shown to protect against both micro- and macrovascular complications of diabetes without any risk of significant hypoglycaemia [5]. The current rationale for beta-cell replacement is restoration of normoglycaemia and the prevention of hypoglycaemia and progressive complications of diabetes when best medical approaches have failed [6, 7].

Transplant Options in Type 1 Diabetes

Although both pancreas and islet transplantation have similar objectives, they are often considered for distinct populations, albeit with some overlap. Because donor supply is limited and major abdominal surgery is required, pancreas transplants are often restricted to younger, fitter recipients. Islet transplantation is generally indicated to reduce risk of hypoglycaemia and/or improve glycaemic control and most are performed in subjects with preserved renal function or after kidney transplant [8]. Some of the similarities and differences are described in Table 15.6.4.1 [8–13].

Table 15.6.4.1 Comparison of pancreas and islet transplantation

	Pancreas transplant	Islet transplant
Indications	Suboptimal glycaemic control or metabolic instability End stage renal disease (SPK) or functioning renal transplant (PAK)	Recurrent severe hypoglycaemia with impaired awareness of hypoglycaemia Severe glycaemic lability Progressive microvascular complications
Cautions	Current smoking Alcohol/drug use Non-adherence with medical therapy	Alcohol/drug use Non-adherence with medical therapy High body mass index/insulin requirement
Contraindications	Untreated infection or malignancy High cardiovascular risk	Untreated infections or malignancy Liver disease Portal hypertension/portal vein thrombosis
Recipient selection	Good cardiovascular health, younger recipients	Less stringent requirements for cardiovascular health
Renal status		
CKD 4–5	Simultaneous pancreas-kidney transplant indicated	Islet transplant deferred until after kidney transplant (IAK) or less commonly as simultaneous islet kidney
CKD 1–3	Pancreas alone transplants have inferior outcomes and are being performed less frequently	Most islet transplants performed as Islet Transplant Alone in individuals with preserved kidney function
Organ availability	Highly selected donors, with preference for young, lean donors	Less stringent donor selection (older, higher BMI). Islet yields often higher in heavier, middle aged donors
Procurement	Adequate preservation of vessels and an intact pancreatic capsule	Intact pancreatic capsule
Number reported to international registry	48 000 procedures between 1966 and 2014 [9]	2150 infusions between 1999 and 2015 [10]
Surgical procedure	Major abdominal procedure, general anaesthesia	Minimally invasive, local anaesthesia
Length of hospital admission	1–2 weeks	≤1 week
Procedural risks	Bleeding, pancreatitis, anastomotic leak, re-laparotomy, pancreatic thrombosis	Bleeding, portal vein thrombosis, abdominal pain
Immunosuppression		
Steroids	Often used short term Dose minimized/avoided in longer term when possible	Avoided
Maintenance	Required long-term	Required long-term
Beta-cell Mass/secretory capacity	Robust	Minimal/marginal

SPK, simultaneous pancreas kidney transplant; PAK, pancreas after kidney; IAK, islet after kidney transplant.

Pancreas Transplantation

Surgical Procedure

Advances in organ procurement and surgical technique have improved pancreas transplant outcomes. Pancreas grafts are often placed in the right lower quadrant, with a side-to-side anastomosis of the duodenal stump to a loop of ileum permitting exocrine drainage (not usually to the bladder which was used historically). Vascular anastomoses are often with the iliac vessels although portal venous drainage is used by some surgeons. Technical and anatomical considerations may be most important since no clear advantages of portal drainage have been found. Anticoagulation seems to reduce the risk of pancreatic thrombosis which has been a leading cause of early graft loss [14].

Types of Pancreas Transplant

The vast majority of pancreas transplants have been performed in association with renal transplants, with 90% of cases as a simultaneous pancreas kidney transplant (SPK) [9].

Demand for renal transplantation exceeds supply and suitable pancreatic donors are even more limited. Renal transplant recipients have a survival advantage over dialysis with best outcomes achieved with living donor kidneys [15]. Pancreas after kidney transplantation (PAK) may attain the benefits of improved glycaemic control while reducing time on dialysis [11]. SPK with maintained pancreas graft function reduces cardiovascular mortality risk in comparison to living donor kidney transplant alone [15].

Pancreas transplantation alone (PTA) performed in individuals with well-preserved renal function is reserved for metabolic instability in type 1 diabetes, defined by recurrent severe hypoglycaemia, hyperglycaemia, or diabetic ketoacidosis. PTA represents a small minority of pancreas transplants and the benefits of improved glycaemic control and insulin independence must be weighed against the procedural risk, with islet transplantation providing a lower risk alternative [11].

Islet Transplantation

Islet transplantation is a minimally invasive procedure performed with local anaesthesia with/without sedation which generally employs a percutaneous, transhepatic approach to access the portal vein using fluoroscopic guidance. The major procedural risks are related to either bleeding (intrahepatic or intraperitoneal) as a result of the hepatic puncture; or thrombosis of branches of the portal vein. The use of highly purified islet preparations with low tissue volume (below 5 ml)

minimizes elevations of portal vein pressure, and together with use of heparin, reduces the risk for thrombosis. Skilled operators can rapidly access the portal vein with a single pass, and the tract can be ablated to minimize the risk of bleeding, using a paste made from collagen flour and contrast medium or a haemostatic sponge [16, 17].

The success of the Edmonton Protocol was attributed initially to a sufficient mass (a total of >10 000 islet equivalents per kg body weight (IEQ/kg)) of high quality, freshly isolated islets [3]. Most centres now perform islet transplantation after brief periods of islet culture (24–48 hours). This has a number of practical advantages including the opportunity to administer conditioning immunosuppression and schedule transplants during regular hours with experienced staff. While some islets are lost during culture, islet purity generally increases (and tissue volume decreases) as exocrine tissue is preferentially lost [10, 16]. Unfortunately, large numbers of islets are lost in the first few hours after transplantation as a result of the instant blood mediated inflammatory response [13]. Insulin independence can be achieved occasionally with single infusions of <10 000 IEQ/kg, but may not be achieved in all cases despite exceeding this threshold after several infusions [18]. Suffice to say that in the majority of islet recipients, the mass of engrafted islets is substantially lower than in a healthy individual [13, 19].

Successful islet transplant programmes are highly dependent on the skill and expertise of their islet isolation teams to consistently provide high quality preparations. Islet transplant outcomes can vary substantially between sites, with better outcomes achieved by more experienced, high volume centres—emphasizing the importance of effective multidisciplinary teams for achieving optimal outcomes [18].

While insulin independence was the focus of early reports and (with sufficient islet mass) can be achieved in almost all cases, long-term insulin independence may not be maintained [10]. Protection from severe hypoglycaemia can be maintained in the long term such that the major indication for islet transplantation alone (ITA) is recurrent severe hypoglycaemia and impaired awareness of hypoglycaemia and/or glycaemic lability [18]. Islet after kidney transplantation (IAK) is performed less frequently to improve metabolic control with the expectation that this will protect the donor kidney. Simultaneous islet-kidney transplantation has been reported relatively infrequently [10] but has recently been made available within the UK integrated programme as an alternative to SPK in those with insufficient cardiorespiratory fitness. The TRIMECO multicentre randomized trial demonstrated the benefits of islet transplantation (ITA or IAK) over medical therapy for improved metabolic control and prevention of hypoglycaemia [20].

Immunosuppression

Induction with lymphodepleting therapies (thymoglobulin, alemtuzumab) is commonly used in both islet and pancreas transplant [9, 10]. Use of lymphodepletion has been an important factor associated with improved islet transplant outcomes [10]. Maintenance immunosuppression commonly uses combinations of tacrolimus or sirolimus with mycophenolate. While prednisone use is widespread early after pancreas transplant [9], it is avoided in islet transplantation because of adverse metabolic effects believed to negatively impact islet engraftment [3].

While short-term side effects of immunosuppression are common, these can generally be mitigated by dose adjustment. The major concern is the risks of long-term immunosuppression, particularly infection and neoplasia [10]. With careful monitoring, these risks can be minimized, but cases of opportunistic infection do occur. Recently, cases of post-transplant lymphoproliferative disease have been reported in islet transplant recipients [21]. While solid organ tumours are somewhat increased in transplant recipients, skin cancers are significantly more common and seem to progress more rapidly in those on immunosuppression [22].

Outcomes

Patient and Graft Survival

Patient and graft survival reported by international registries are shown in Table 15.6.4.2. Pancreas graft survival is defined by insulin independence with islet graft survival defined by the presence of C-peptide secretion [23–25].

Metabolic Function and Protection from Hypoglycaemia

Both pancreas and islet transplantation can restore counterregulatory response to hypoglycaemia [19]. However, while marginal graft function protects against severe hypoglycaemia [26], more substantial beta-cell mass, routinely provided by successful pancreas transplantation, is required for insulin independence. Despite insulin independence, pancreas transplant recipients may have dysglycaemia [27], although detailed metabolic assessments are not reported routinely. Registry data for islet and pancreas transplant metabolic outcomes are summarized in Table 15.6.4.3 [28].

Table 15.6.4.2 Islet patient survival data from CIT-07 trial. Islet graft failure reported from NHSBT annual report on pancreas and islet transplantation 2017/2018 with 1-year survival data for islet transplants between 2013 and 2017, and 5-year survival for transplants between 2008 and 2017

	Patient survival (%)			Beta-cell (pancreas or islet) graft survival (%)		
	1 year	5 years	10 years	1 year	5 years	10 years
SPK	97.4†	90††	76††	91.3†	73††	56††
PAK	97.9†	87††	71††	86.0†	64††	38††
PTA	97.0†	90††	82††	85.7†	53††	36††
ITA	100%	100%		84%	48%	

†Analysis of transplant outcomes from UNOS and IPTR report for US transplants performed between 2010 and 2014[9].
††Analysis of IPTR outcomes for US transplants with reported 5-year survival outcomes for transplants 2008/2009 and 10-year outcomes for transplants between 2004/2005.

Table 15.6.4.3 Metabolic outcomes after pancreas and islet transplant. Islet data obtained from 10th CITR Annual report with assessment of outcomes 1- and 5-years post last infusion. Pancreas data from UNOS[9]. In pancreas transplants, a return to exogenous insulin has generally been viewed as graft failure, and it is uncommon for HbA1c and severe hypoglycaemia to be assessed routinely

	Pancreas transplant		Islet transplant	
	Short term (1 year)	Long-term (5 years)	Short term (1 year)	Long term (5 years)
Insulin independence	>90%	53–73%	50%	~30%
Reduced insulin dose	Insulin independence expected	Expected, unless explanted or recipient develops insulin resistance/PTDM	Expected, unless C-peptide negative	Sustained, if C-peptide positive
Absence of severe hypoglycaemia	Expected	Expected	95%	>90%
HbA1c <7% (53 mmol/mol) and no severe hypoglycaemia	Expected	Expected	~70%	~50%

PTDM, post-transplant diabetes mellitus.

Impact on Diabetes Complications

Data from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications study (DCCT/EDIC) imply that improved glycaemic control should result in improved microvascular outcomes [5]. Compared with DCCT participants, transplant recipients are older, have much longer diabetes duration and a substantially higher burden of complications [10]. Rodent data suggest that reversal of complications is harder with longer diabetes duration. Some immunosuppressant drugs (calcineurin inhibitors) are nephrotoxic [17]. Thus long-term studies to evaluate the net effect of transplantation are required.

Ten years of normoglycemia after pancreas transplant has shown reversal of histological changes of diabetic nephropathy (without preservation of GFR) and stabilization of diabetic retinopathy [29]. Small studies support a reduced rate of microvascular disease progression following islet transplantation versus medical therapy [30]. Macrovascular disease may also be positively impacted as islet transplantation has been associated with reduced progression of carotid intima-medial thickening [31] and cardiovascular mortality is lower with sustained pancreas graft function after SPK compared with living donor kidney transplant alone [15].

Quality of Life

Improved quality of life is expected with successful pancreas and islet transplantation [32]. Diabetes distress and hypoglycaemia fear are substantially reduced [33].

Can We Meaningfully Compare Beta-Cell Replacement Outcomes, Risks, and Benefits?

A major challenge is the paucity of comparable data regarding outcomes of pancreas or islet transplantation. Most reports in pancreas transplantation are derived from registry data and focus on patient and graft survival. Graft function is defined as no insulin use (often without any assessment of glycaemic control) [9]. In contrast, long-term patient survival data are not yet available in islet transplantation and reports have tended to focus on defined metabolic outcomes and hypoglycaemia (Table 15.6.4.3).

To address this the International and European Pancreas & Islet Transplant Associations have proposed the Igls criteria (Table 15.6.4.4) to standardize assessment of outcomes for beta-cell replacement [6].

Patient Selection and Indications for Transplantation

Informed choices around transplant options necessitate weighing short- and long-term risks and benefits, including the consideration of renal reserve and other comorbidities. Most contraindications are relative and depend on the strength of the indications. Robust early function appears to predict long-term function [34], although achieving insulin independence with islet transplantation may be challenging in heavier recipients [10], or with higher pre-transplant insulin requirement and/or insulin resistance. It is not clear whether strategies prioritizing achievement of insulin independence (several

Table 15.6.4.4 Outcomes of beta-cell replacement therapy: Igls criteria

Beta-cell functional status	HbA1c (mmol/mol)	Severe hypoglycaemia events per year	Insulin requirements unit/kg/day	C-peptide	Treatment success
Optimal	≤48 (6.5%)	None	None	>Baseline	Yes
Good	<53 (7%)	None	<50% baseline use of other OHA	>Baseline	Yes
Marginal	Baseline	<Baseline	≥50% baseline	>Baseline	No
Failure	Baseline	Baseline	Baseline	Baseline	No

OHA: oral hypoglycaemic agents.

Adapted with permission from Rickels MR, Stock PG, de Koning EJP, *et al.* Defining outcomes for beta-cell replacement therapy in the treatment of diabetes: a consensus report on the Igls Criteria from the IPITA/EPITA opinion leaders workshop. *Transplantation*. 2018;102(9):1479–86. Copyright 2018 © Steunstichting ESOT.

islet donors per recipient—Canada) or amelioration of hypoglycaemia in more recipients (maximum two infusions—UK) will prove more effective in the long-term but supplemental islet infusions have been shown to be effective in re-establishing insulin independence in individuals with islet graft dysfunction [35].

Pancreas Allocation and Availability

Because of the limited availability of donor organs efforts should be made to ensure they are used most appropriately and effectively. In some parts of the world (e.g. Canada) organ allocation is determined by local policies, with pancreata first being offered for whole pancreas transplantation (usually SPK), with declined organs then being offered for islet transplantation. This may reflect historical precedence with many established pancreas transplant programmes compared with a small number of islet transplant programmes.

In the United Kingdom, nationally retrieved organs are allocated to pancreas and islet transplant recipients on a shared waiting list. Priority for islet or pancreas transplant is assessed by the weighting of factors including: recipient waiting time; dialysis status; HLA sensitization/donor-recipient mis-match; donor body mass index (BMI); donor-recipient age matching; and cold ischaemic time. The quality of the donor organ has been shown to directly influence the success of islet isolation and graft outcomes. In the UK programme, pancreata from higher BMI donors are preferentially offered for islet transplant. Additional priority is also awarded following first islet transplant towards rapid access to a second transplant when indicated [36].

Clinical Monitoring and Management

Interval monitoring is important to ensure adequate, but not excessive, exposure to immunosuppression with metabolic monitoring to assess graft function. Dysglycaemia can predict islet and pancreas graft failure [27] and may be due to allo-immune rejection, recurrence of autoimmunity, diabetogenic effects of immunosuppression or beta-cell exhaustion [18]. Pancreas rejection can be confirmed by biopsy and may respond to steroids [12] or antithymocyte globulin Currently there are no reliable markers for rejection in islet transplant, and dysglycaemia is most commonly addressed by early use of insulin to reduce metabolic demand and avoid glucose toxicity. DPP4 inhibitors and GLP-1 receptor agonists appear to be safe glucose-lowering agents in islet transplantation. Other non-insulin glucose-lowering agents have not been tested, but consensus guidelines suggest avoidance of insulin secretagogues [8].

Other Indications and Future Directions

Long-term safety and limited organ supply are the two major barriers to broader application of transplantation in type 1 diabetes. Clinical trials of a potentially unlimited supply of beta-cells derived from embryonic stem cells began in 2016 [37]. Advances in immunotherapy and immunobarrier encapsulation may improve the effectiveness and safety of, or potentially obviate the need for, chronic immunosuppression.

No immunosuppression is required in islet autotransplantation following total pancreatectomy for chronic pancreatitis complicated by severe, refractory pain to prevent labile diabetes and may avoid the need for insulin in up to 41% of patients for up to 3 years [38].

Reports of case series of pancreas transplantation for type 2 diabetes have shown that insulin independence can be achieved [39], certainly in the absence of BMI >30, but the efficacy of the marginal mass achieved by islet transplantation in the face of insulin resistance in type 2 diabetes seems questionable. However, given the growing prevalence of type 2 diabetes with end stage renal disease, transplantation (SPK) may be an appropriate and effective therapy [40]. Transplant physicians will continue to carefully balance the risks and benefits for an individual with the optimal use of limited donor pancreata.

REFERENCES

1. Shapiro AMJ. A historical perspective on experimental and clinical islet transplantation. In: Shaprio AMJ, Shaw JAM, ed. *Islet Transplantation and Beta-Cell Replacement Therapy*, 1st edn. New York: Informa Healthcare USA Inc., 2007: 1–27.
2. Brendel M, Hering B, Schulz A, *et al. International Islet Transplant Registry Report*. Germany: University of Giessen, 1999: 1–20.
3. Shapiro AMJ, Lakey JR, Ryan EA, *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppression regimen. *N Engl J Med*, 2000; **343**: 230–8.
4. Ryan EA, Paty BW, Senior PA, *et al.* Five-year follow-up after clinical islet transplantation. *Diabetes*, 2005; **54**: 2060–9.
5. Nathan DM, Cleary PA, Backlund JY, *et al.* Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*, 2005; **353**: 2643–53.
6. Rickels MR, Stock PG, de Koning EJP, *et al.* Defining outcomes for beta-cell replacement therapy in the treatment of diabetes: a consensus report on the Igls Criteria from the IPITA/EPITA opinion leaders workshop. *Transplantation*, 2018; **102**: 1479–86.
7. Choudhary P, Rickels MR, Senior PA. Evidence-informed clinical practice recommendations for treatment of type 1 diabetes complicated by problematic hypoglycaemia. *Diabetes Care*, 2015; **38**: 1016–29.
8. Diabetes Canada Clinical Practice Guidelines Expert Committee, Senior PA, AlMehthel M, *et al.* Diabetes and transplantation. *Can J Diabetes*, 2018; **42** Suppl 1: S145–9.
9. Gruessner AC, Gruessner RW. Pancreas transplantation of US and non-US cases from 2005 to 2014 as reported to the United Network for Organ Sharing (UNOS) and the International Pancreas Transplant Registry (IPTR). *Rev Diabet Stud*, 2016; **13**: 35–58.
10. CITR Coordinating Center. Collaborative Islet Transplant Registry tenth annual report, 2017. Available at: https://citregistry.org/system/files/10th_AR.pdf (accessed: 10 August 2018).
11. Dean PG, Kukla A, Stegall MD, *et al.* Pancreas transplantation. *BMJ*, 2017; **357**: j1321.
12. White SA, Shaw JA, Sutherland DE. Pancreas transplantation. *Lancet*, 2009; **373**: 1808–17.
13. Rickels MR, Mueller R, Teff KL, *et al.* Beta-cell secretory capacity and demand in recipients of islet, pancreas, and kidney transplants. *J Clin Endocrinol Metab*, 2010; **95**: 1238–46.
14. Rogers J, Farney AC, Orlando G. Pancreas transplantation with portal venous drainage with an emphasis on technical aspects. *Clin Transplant*, 2014; **28**: 16–26.

15. Lindahl JP, Hartmann A, Aakhus S, *et al.* Long-term cardiovascular outcomes in type 1 diabetic patients after simultaneous pancreas and kidney transplantation compared with living donor kidney transplantation. *Diabetologia*, 2016; **59**: 844–52.
16. Ricordi C, Goldstein JS, Balamurugan AN, *et al.* National Institutes of Health-Sponsored clinical islet transplantation consortium phase 3 trial: manufacture of a complex cellular product at eight processing facilities. *Diabetes*, 2016; **65**: 3418–28.
17. Venturini M, Sallemi C, Marra P, *et al.* Allo- and autopercutaneous intra-portal pancreatic islet transplantation (PIPIT) for diabetes cure and prevention: the role of imaging and interventional radiology. *Gland Surg*, 2018; **7**: 117–31.
18. Shapiro AM, Ricordi C, Hering BJ, *et al.* International trial of the Edmonton protocol for islet transplantation. *N Engl J Med*, 2006; **355**: 1318–30.
19. Rickels MR. Recovery of endocrine function after islet and pancreas transplantation. *Curr Diab Rep*, 2012; **12**: 587–96.
20. Lablanche S, Vantyghem MC, Kessler L, *et al.* Islet transplantation versus insulin therapy in patients with type 1 diabetes with severe hypoglycaemia or poorly controlled glycaemia after kidney transplantation (TRIMECO): a multicentre, randomised controlled trial. *Lancet Diabetes Endocrinol*, 2018; **6**: 527–37.
21. Peters A, Olateju T, Deschenes J, *et al.* Posttransplant lymphoproliferative disorder after clinical islet transplantation: report of the first two cases. *Am J Transplant*, 2017; **17**: 2474–80.
22. Ulrich C, Schmook T, Sachse MM, *et al.* Comparative epidemiology and pathogenic factors for nonmelanoma skin cancer in organ transplant patients. *Dermatol Surg*, 2004; **30**: 622–7.
23. Gruessner AC, Gruessner RWG. Long-term outcome after pancreas transplantation: a registry analysis. *Curr Opin Organ Transplant*, 2016; **21**: 377–85.
24. Hering BJ, Clarke WR, Bridges ND, *et al.* Phase 3 trial of transplantation of human islets in type 1 diabetes complicated by severe hypoglycaemia. *Diabetes Care*, 2016; **39**: 1230–40.
25. NHS Blood and Transplant. Annual report of pancreas and islet transplantation: report for 2017/2018. 2018. Available at: https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/12251/nhsbt-pancreas-and-islet-transplantation-annual-report-2017-2018.pdf (accessed: 24 November 2018).
26. Vantyghem MC, Raverdy V, Balavoine AS. Continuous glucose monitoring after islet transplantation in type 1 diabetes: an excellent graft function (beta score greater than 7) is required to abrogate hyperglycaemia, whereas a minimal function is necessary to suppress severe hypoglycaemia (beta score greater than 3). *J Clin Endocrinol Metab*, 2012; **97**: E2078–83.
27. Mittal S, Nagendran M, Franklin RH, *et al.* Postoperative impaired glucose tolerance is an early predictor of pancreas graft failure. *Diabetologia*, 2014; **57**: 2076–80.
28. Huang C, Zhang L, Shi Y, *et al.* The KCa3.1 blocker TRAM34 reverses renal damage in a mouse model of established diabetic nephropathy. *PLoS One*, 2018; **13**: e0192800.
29. Fioretto, P, Mauer SM, Bilous RW, *et al.* Effects of pancreas transplantation on glomerular structure in insulin-dependent diabetic patients with their own kidneys. *Lancet*, 1993; **342**: 1193–6.
30. Thompson DM, Meloche M, Ao Z, *et al.* Reduced progression of diabetic microvascular complications with islet cell transplantation compared with intensive medical therapy. *Transplantation*, 2011; **91**: 373–8.
31. Danielson KK, Hatipoglu B, Kinzer K, *et al.* Reduction in carotid intima-media thickness after pancreatic islet transplantation in patients with type 1 diabetes. *Diabetes Care*, 2013; **36**: 450–6.
32. Speight J, Reaney MD, Woodcock AJ, *et al.* Patient-reported outcomes following islet cell or pancreas transplantation (alone or after kidney) in type 1 diabetes: a systematic review. *Diabetic Med*, 2010; **27**: 812–22.
33. Foster ED, Bridges ND, Feurer ID, *et al.* Improved health-related quality of life in a phase 3 islet transplantation trial in type 1 diabetes complicated by severe hypoglycaemia. *Diabetes Care*, 2018; **41**: 1001–8.
34. Vantyghem MC, Kerr-Conte K, Arnalsteen L, *et al.* Primary graft function, metabolic control and graft survival after islet transplantation. *Diabetes Care*, 2009; **32**: 1473–8.
35. Koh A, Imes S, Kin T, *et al.* Supplemental islet infusions restore insulin independence after graft dysfunction in islet transplant recipients. *Transplantation*, 2010; **89**: 361–5.
36. Hudson A, Bradbury L, Johnson R, *et al.* The UK pancreas allocation scheme for whole organ and islet transplantation. *Am J Transplant*, 2015; **15**: 2443–55.
37. Clinical trials.gov. A safety, tolerability and efficacy study of VC-01™ combination product in subjects with type 1 diabetes mellitus. NCT02239354, 2014. Available at: https://clinicaltrials.gov/ct2/show/NCT02239354 (accessed: 24 November 2018).
38. Chinnakotla S, Bellin MD, Schwarzenberg SJ, *et al.* Total pancreatectomy and islet autotransplantation in children for chronic pancreatitis: indication, surgical techniques, postoperative management, and long term outcomes. *Ann Surg*, 2014; **260**: 56–64.
39. Nath DS, Gruessner AC, Kandaswamy R, *et al.* Outcomes of pancreas transplants for patients with type 2 diabetes mellitus. *Clin Transplant*, 2005; **19**: 792–7.
40. Whiting DR, Guariguata L, Weil C, *et al.* IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract*, 2011; **94**: 311–21.

15.7

Evidence-based Prevention and Management of Type 2 Diabetes

15.7.1 Strategies for the Management of Type 2 Diabetes

Peter Winocour and Sagen Zac-Varghese

Introduction

Type 2 diabetes is characterized by insulin resistance and inadequate insulin secretion leading to hyperglycaemia. The overarching objectives of antihyperglycaemic therapy are: symptom relief, reduction of microvascular complications, and reduction of macrovascular complications over the longer term. Therapies may reduce insulin resistance, stimulate insulin secretion, replace insulin, reduce glucose production, delay glucose absorption, or increase glucose elimination.

The increasing availability of new antihyperglycaemic agents, while beneficial, has led to increased uncertainty with regards to choice of agent, sequence, suitability, safety, and appropriate combination. Guidelines state that care plans should be individualized. However, opinions on how this is to be achieved are varied. In the future, subcategorization of type 2 diabetes may have an impact on therapy [1].

National guidelines are broadly similar in recommending stepwise increments in therapy. Ultimately a successful treatment regimen would enable a patient to achieve glycaemic control in the absence of weight gain and hypoglycaemia. Some guidelines have reflected cardiovascular and renal benefits of certain agents, and have incorporated these into treatment algorithms.

This chapter looks at the approach to therapy selection and discusses some of the background, theoretical framework, and evidence behind these algorithms.

Socioeconomic Determinants of Health

The socioeconomic determinants of health and cost of therapy are important considerations. Where patients have to pay for their treatments this can lead to non-compliance [2]. Nationally funded healthcare may restrict access to certain agents. Globally, food insecurity could lead to the risk of hypoglycaemia with certain medications. Conversely, the widespread availability of highly processed, refined carbohydrate snacks, and beverages has contributed to the global obesity epidemic.

Lifestyle Measures and Obesity Management

Guidelines advocate healthy lifestyle interventions (diet and exercise) for type 2 diabetes [3, 4]. Weight loss can delay the progression of prediabetes to type 2 diabetes and there is increasing evidence that weight loss has important effects on improving glycaemic control both initially and at later stages of diabetes [5]. The demonstration that remission from type 2 diabetes can be achieved after more extreme weight loss (through an approximately 800 calories/day diet) may preclude the need for medical therapies or at least enable dose reduction or withdrawal of certain hypoglycaemic agents

[6, 7]. However, it is recognized that such an approach is limited by weight regain (after diet cessation) unless a long-term maintenance programme is followed.

American, Canadian and European guidelines reflect the importance of both good glycaemic control and weight management and advocate the use of antihyperglycaemic agents which either promote weight loss or which are weight neutral [3, 4, 8]. Current National Institute for Health and Care (NICE) guidelines do not specifically prioritize the use of these agents [9]. However, the fundamental importance of individualized care is a central theme of that guidance.

Antihyperglycaemic agents that contribute to weight loss include: glucagon-like peptide-1 (GLP-1) receptor agonists, sodium and glucose cotransporter (SGLT) 2 inhibitors and amylin (only available in the United States). Medications with possible benefits or weight neutral drugs include: metformin, dipeptidyl peptidase IV (DPP-IV) inhibitors and α-glucosidase inhibitors (acarbose). Sulphonylureas (SU), meglitinides, and particularly insulin all lead to weight gain (increased adiposity). Weight gain is inevitable with insulin and this is more notable in insulin resistant patients who require higher insulin doses. Thiazolidinediones (TZD) are associated with weight gain due to fluid retention.

A recent meta-analysis concluded that oral antihyperglycaemic agents were equally beneficial irrespective of baseline body mass index (BMI) [10]. However, SU were less effective in obese patients and insulin was not studied. In the UKPDS 49 study, obese patients were more likely to require multiple therapies [11].

The American Diabetes Association (ADA) guidelines recommend specific pharmacotherapy for obesity (either alone or as an adjunct to standard diabetes therapy) [8]. Efficacy may be limited by adverse effects and variable compliance. In contrast, metabolic bariatric surgery leads to marked early improvement in glycaemic control and weight loss. Following bariatric surgery, type 2 diabetes remission is between 30% and 80% depending on the type of surgery performed [12].

Principles of Antihyperglycaemic Treatment for Type 2 Diabetes

Patient empowerment in understanding diabetes and the associated treatments is essential. People with diabetes should be actively involved in therapy selection and in monitoring the response to treatment. The choice of antihyperglycaemic agent is based on a number of factors including:

- individualized glycaemic targets,
- effectiveness of the therapy versus tolerability and side effects,
- comorbidities including renal impairment,
- cardiovascular disease, and obesity,
- the impact of the agent on weight,
- the impact of the agent on renal disease and cardiovascular disease, patient preference,
- risk from polypharmacy,
- appropriateness of tight glycaemic control (risk of hypoglycaemia).

Patient preferences may dictate treatment based on their job (shift work or work involving heavy machinery), driving restrictions and fasting (e.g. Ramadan). This may preclude the use of agents which predispose to hypoglycaemia. Patient comorbidity and concomitant medications, in particular atypical antipsychotics, steroids, and antiepileptic agents may also be associated with unwanted weight gain and hyperglycaemia.

Changes during a lifetime, such as during puberty, during pregnancy and breastfeeding, in older patients, in dementia and at the end of life, can lead to changes in diabetes management with either intensification or relaxation of glycaemic targets and treatment regimes. Type 2 diabetes is increasingly prevalent in older adults who are at increased risk of polypharmacy. It is important to simplify regimes and implement strategies with low risk of hypoglycaemia as elderly patients may not recognize hypoglycaemic symptoms or seek appropriate help. In these patients, the focus should be on symptom control and agents that encourage satiety and weight loss may be undesirable as malnutrition may be an issue. At the end of life, in patients with type 2 diabetes, many treatments can be discontinued [13] (see Table 15.7.1.1).

Initiation of Antihyperglycaemic Agents for Type 2 Diabetes

A patient with type 2 diabetes who is asymptomatic and diagnosed through a screening programme will require a different treatment approach compared to a patient diagnosed due to symptoms, metabolic decompensation, or due to the complications of diabetes. All antihyperglycaemic agents (including insulin) have side effects and the importance of sick day guidance should be reinforced. In general, patients with undiagnosed type 2 diabetes presenting with complications will be less amenable to tight control, especially if associated with obesity, compared to asymptomatic patients diagnosed early in the disease through screening programmes.

Patients with Type 2 Diabetes Presenting With Metabolic Decompensation

In newly presenting patients with symptomatic hyperglycaemia and metabolic decompensation, the type of diabetes may be initially unclear and glycaemic control may be best achieved with increasing doses of either SU or insulin. Insulin would be the treatment of choice where severe osmotic symptoms, ketosis, or weight loss (regardless of HbA1c) raise the possibility of type 1 diabetes, latent autoimmune diabetes of adults (LADA), or insulin deficiency.

Guidelines employ an initial stratification scheme based on the level of glucose and HbA1c (regardless of symptoms). If the HbA1c is <75 mmol/mol initial oral monotherapy is recommended, if the HbA1c is ≥75 mmol/mol dual therapy is recommended (e.g. metformin plus insulin) and at levels ≥86 mmol/mol combination injectable therapy is recommended (e.g. a basal bolus insulin regime) [3, 8, 9]. The Canadian guidelines use individualized glycaemic targets and an algorithm which references percentage deviance from the target [3].

The initial intensive treatment of hyperglycaemia may have some benefit as studies have demonstrated the effects of glycaemic memory which may reflect a reduction in accumulated glucotoxicity [14]. There is also some evidence that early insulin therapy leads to

Table 15.7.1.1 Summarized benefits, side effects, and other considerations for antihyperglycaemic agents

Agent	Weight	Hypoglycaemia	Cardiovascular effects	Renal effects	Side effects	Additional contraindication	Additional benefits
Metformin	Neutral/ loss	No	Benefit	Contraindicated if eGFR < 30	Gastrointestinal, B12 deficiency, lactic acidosis (rare)	Liver failure, hypoxia (acute heart failure, sepsis)	Benefit in NAFLD and PCOS
Sulphonylureas	Gain	Yes	Association with heart failure, RCT—no statistical difference	Risk of hypoglycaemia		Liver failure	
Meglitinides	Gain	Yes	Neutral	Risk of hypoglycaemia		Liver failure	Rapid onset and offset of action. Suitable for erratic meal patterns
Thiazolidinedione	Gain (fluid)	No	Increased risk of heart failure. Benefit in PROactive study	Not recommended in renal failure due to oedema	Fluid, maculopathy, osteoporotic fracture peripheral (and hip) in postmenopausal women	Heart failure, liver failure, bladder cancer (EASD—no significant association)	Benefit in NAFLD/NASH
SGLT-2 inhibitors	Loss	No	Benefit –class effect reduction in hospitalization for heart failure. Reduction in MACE—empagliflozin, canagliflozin	Benefit – dapagliflozin, empagliflozin, canagliflozin eGFR restrictions—variation within class	Urinary tract and genital infections, DKA risk, possible risk of amputation and increased fracture risk (canagliflozin), association with Fournier gangrene, small increase in LDL-C	Liver failure	Reduced blood pressure
DPP-IV inhibitors	Neutral	No	Potential risk—saxagliptin, alogliptin	Renal dose adjustment, variation within class		Previous pancreatitis, angioedema, urticaria	
GLP-1 receptor agonist	Loss	No	Benefit—liraglutide, semaglutide, dulaglutide, albiglutide	Benefit—liraglutide eGFR restrictions—variation within class	Rare cases of acute gallstone disease, gastrointestinal	Previous pancreatitis Black box—risk of thyroid C cell tumours (seen in animal studies only), semaglutide—caution with severe retinopathy	Benefit in NAFLD (liraglutide)
Insulin	Gain	Yes	Neutral	Risk of hypoglycaemia			
Acarbose	Neutral	No	Possible benefit	Avoid in dialysis	Gastrointestinal	Liver failure, gut dysmotility e.g. gastroparesis	

partial beta-cell recovery [15]. Unfortunately, the agents associated with rapid improvement in glycaemic control (insulin and SU) are associated with risk of weight gain and hypoglycaemia.

While high glucose and HbA1c values may prompt insulin initiation, it is important to take a clear dietary history. Intake of high glucose containing drinks to address the osmotic symptoms of type 2 diabetes could lead to hyperglycaemia which could readily improve by their withdrawal and institution of standard dietary advice.

It is also important to reflect on comparative studies and the baseline HbA1c. All therapies will have greater efficacy the higher the starting level of glycaemia. Some have flatter dose response curves and insulin is the most potent glucose lowering therapy on a dose dependent basis.

While several studies have demonstrated reduced microvascular complications with improved glycaemic control, the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study found significantly higher mortality in older patients with prevalent macrovascular disease treated with intensive insulin therapy [16].

It is also recognized that overly rapid correction of HbA1c can lead to deteriorating retinopathy and worsening neuropathy [17, 18]. Currently, there are no recommendations with regards to rate of HbA1c improvement, other than increased retinal screening in patients considered to be at increased risk, i.e. those with baseline moderate to severe diabetic retinopathy and longer duration of diabetes [18]. In the majority of patients, the worsened retinopathy and/or neuropathy are reversible and the benefit of improved glycaemic control outweighs the risk [17, 18]. However, there is still a case for gradual HbA1c reduction in at risk patient groups but the evidence that this will reduce the risk of deterioration is lacking.

Patients with Type 2 Diabetes Presenting Without Metabolic Decompensation

Across all guidelines, metformin is considered the standard first line therapy for type 2 diabetes and there is a long-term accumulation of experience of over 60 years. The benefits of metformin include its efficacy, low risk of hypoglycaemia, weight neutrality (or slight weight loss), potential cardiovascular benefits, ability to combine metformin with most other agents and its cost [14].

The main contraindication for metformin is renal impairment with eGFR <30 ml/min/1.73 m^2 (see 'Renal Impairment and Choice of Antihyperglycaemic Agent'). There is a potential risk of lactic acidosis in situations of metabolic decompensation. Thus metformin is contraindicated in sepsis, acute coronary disease, hepatic and renal failure, and in the event of concurrent illness (sick day guidance). This guidance is perceived by many as overly cautious and it is not yet clear if metformin is a bystander or active participant contributing to lactic acidosis in these situations. A Cochrane systematic review of metformin found no fatal or non-fatal cases of lactic acidosis with the use of metformin compared to placebo or other antihyperglycaemic agents [19]. The latest ABCD-Renal Association guidelines support the use of metformin for some patients with chronic kidney disease (CKD) stage 4 with dose reduction [20].

In situations where metformin is contraindicated due to significant renal impairment, several other classes of antihyperglycaemic agent may be similarly cautioned (notably SU, TZD, and SGLT-2 inhibitors). If metformin is contraindicated for other reasons, for example, due to gastrointestinal intolerance (including intolerance of modified released metformin), then the European Association for the Study of Diabetes (EASD), Canadian guidelines and ADA suggests initiating any one of the six treatment options: SU, TZD, DPP-IV inhibitors, SGLT-2 inhibitors, GLP-1 receptor agonists or basal insulin [3, 4, 8]. In the ADA and Canadian guidelines these options are given equivalence. The Scottish Intercollegiate Guidelines Network (SIGN) gives equivalence to four treatment options: SU, TZD, DPP-IV inhibitors and SGLT-2 inhibitors [21]. NICE guidelines advocate the use of SU, TZD, or DPP-IV inhibitors as next step therapy. SGLT-2 inhibitors are considered appropriate if these three are not suitable or are contraindicated.

Intensification of Therapy for Type 2 Diabetes

Due to the progressive nature of type 2 diabetes, patients may require additional therapy. All guidelines follow a similar stepwise approach to treatment intensification. An individualized target HbA1c is decided upon and following lifestyle or monotherapy, treatments are sequentially added every 3–6 months until the target HbA1c is achieved. However, this may lead to a degree of therapeutic inertia. In ADA, NICE, and SIGN guidelines, the target HbA1c at diagnosis is initially set at 48 mmol/mol and 53 mmol/mol thereafter if therapy is associated with hypoglycaemia [8, 9, 21]. Higher targets have been suggested for patients with moderate to severe renal failure or reduced life expectancy [8, 20, 22].

Following metformin, the sequence of therapy selection differs in the guidance from the various agencies. NICE guidelines advise adding SU, TZD, or DPP-IV inhibitors. SIGN suggest adding SU, TZD, DPP-IV inhibitors, or SGLT-2 inhibitors. ADA, EASD, and Canadian guidelines base further treatment choice on the presence of atherothrombotic disease, heart failure or renal disease (see 'Cardiovascular Disease and Antihyperglycaemic Agents'), or the effect of the agent on body weight (see **Figure 15.7.1.1**). Individual patient and antihyperglycaemic agent characteristics influence the subsequent choice of antihyperglycaemic agent (see **Table 15.7.1.1**).

Within the NICE and SIGN guidelines, GLP-1 receptor agonists are only recommended in the third intensification of treatment if the BMI is >35 (lower in patients of Asian ethnicity) with obesity-related complications. GLP-1 receptor agonists are also recommended in patients with BMI <35 where insulin use would be restricted or where weight loss would directly benefit obesity-related complications [9, 21].

Insulin is initiated in patients failing to achieve glycaemic targets. Most guidelines advise starting with a basal insulin [4, 8, 9, 21]. If this is ineffective, short acting prandial insulin is added or the patient is switched to a biphasic regime. The 4T study looked at the use of various insulin regimes and found that basal insulin is least likely to induce weight gain and hypoglycaemia [23]. In addition, the AT-LANTUS study comparing clinician-led to patient-led insulin titration algorithms found that the patient-led protocol was associated with a reduced HbA1c without increased hypoglycaemia [24]. This highlights the importance of self-management and patient education. In general, hypoglycaemia can be mitigated with carbohydrate awareness, exercise guidance, and the use of analogue insulins with improved pharmacokinetic profiles. In insulin resistant patients, the concentrated insulins (e.g. U-500) may play a role. However, care needs to be taken with regards to safety of insulin prescription.

ADA guidelines highlight a role for α-glucosidase inhibitors (acarbose), bile acid sequestrants, bromocriptine and pramlintide (an injectable amylin analogue only available for use in the United States). However, the placement of these agents within the algorithm is not clear. These drugs are limited due to their side effect profile and are less commonly used in clinical practice [4].

Combination of Antihyperglycaemic Agents

In contrast to using maximum doses of single agents, there is some evidence that combinations of submaximal doses may be more effective [25]. When combining antihyperglycaemic agents, agents targeting different pathways should be used together. For example, the combination of an insulin secretagogue such as an SU with an insulin sensitizer would be appropriate. However, the combination of two insulin secretagogues (e.g. an SU with a meglitinide) or the combination of two agents that act on the incretin system (e.g. a DPP-IV inhibitor with a GLP-1 receptor agonist) are less rational with limited benefit.

In addition, specific issues arise with drug combinations with similar side effect profiles. For example, canagliflozin and pioglitazone both potentially increase the risk of fracture; thus this combination should be avoided in patients with metabolic bone disease.

Many patients eventually end up with a combination of oral antihyperglycaemic agents and insulin. When metformin is combined with insulin, this is associated with reduced HbA1c and

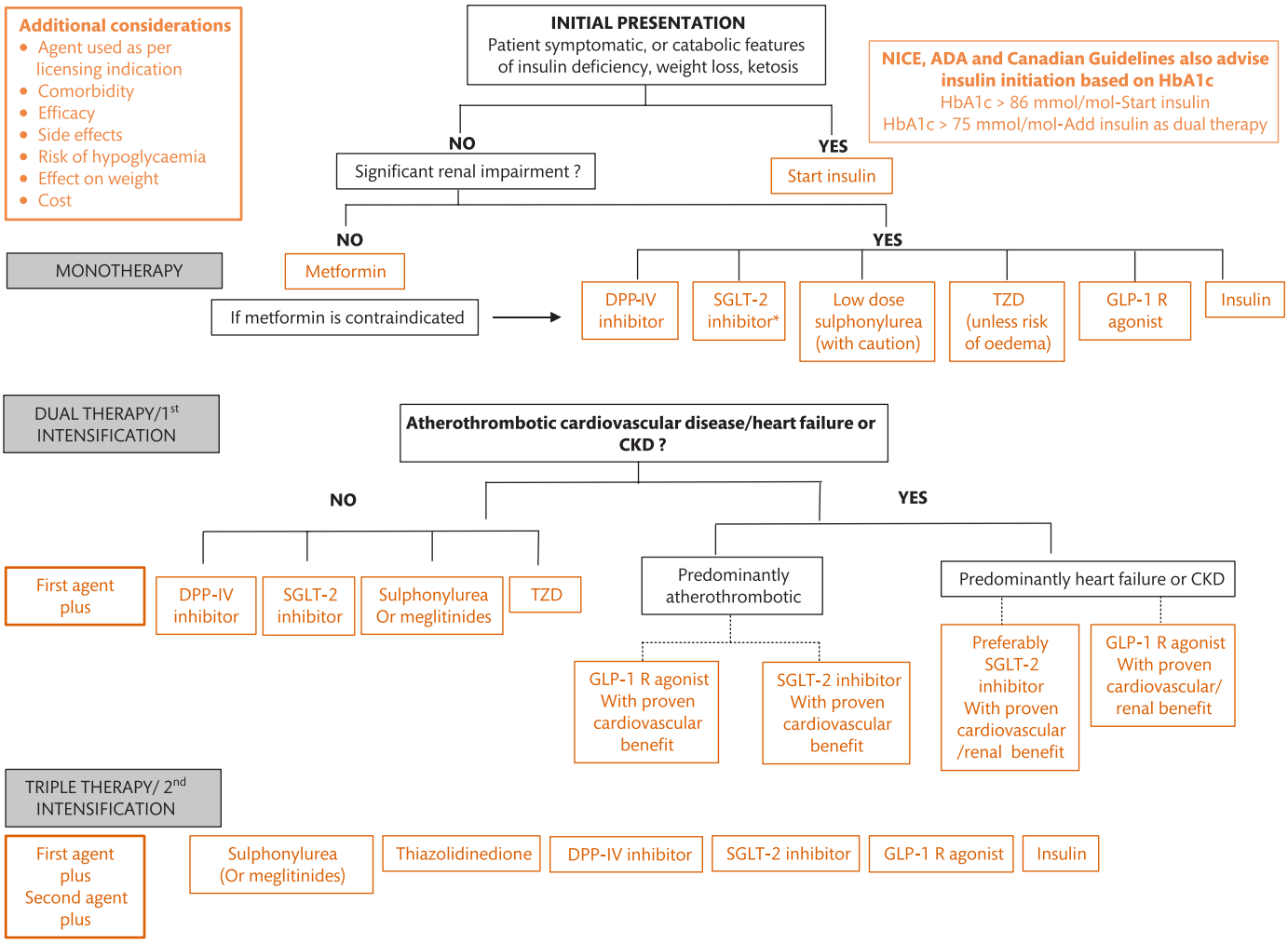


Figure 15.7.1.1 Strategy for choice of antihyperglycaemic agent in type 2 diabetes. *Dependent on eGFR and licensing.

decreased weight gain. However, SU in combination with insulin could increase the risk of hypoglycaemia. It would be reasonable to reduce or discontinue SU after insulin is started (especially if a basal bolus or biphasic insulin regime is used). TZD use alongside high insulin doses can lead to significant weight gain through fluid retention. SGLT-2 inhibitors can be used with insulin and can also lead to weight loss with insulin dose reduction; however, this combination can be associated with increased risk of hypoglycaemia. In general, GLP-1 agonists do not need to be discontinued with insulin.

Cardiovascular Disease and Antihyperglycaemic Agents

In 2008, following cardiovascular concerns regarding rosiglitazone, the Food and Drugs Administration (FDA) advised that cardiovascular outcome trials (CVOT) be performed for all new antihyperglycaemic agents. Since 2008, restrictions on the use of rosiglitazone have been removed and it is currently available in the United States (not licensed in Europe). Pioglitazone (available in United States and Europe) is associated with heart failure; however, paradoxically, pioglitazone was found to improve all-cause mortality, non-fatal MI and stroke in the PROactive study [26, 27].

The Thiazolidinediones Or Sulfonylureas Cardiovascular Accidents Intervention Trial (TOSCA.IT) study looked at the risk of TZD versus SU in patients on metformin monotherapy [28]. This study did not find differences in the primary end point (cardiovascular death, non-fatal MI, non-fatal stroke) between these two agents. However, heart failure was not included in the primary or secondary outcomes and had a relatively high incidence in both arms in the expanded composite outcomes. It is not yet known if pioglitazone has a beneficial role in patients where heart failure is controlled with diuretics.

The UK GP database (observational study) identified an 18–30% increased risk of heart failure with SU [29]. However, randomized control trials demonstrated no statistical difference between SU and metformin [21, 30]. It is possible that SU increase the risk of heart failure, but not of atherothrombotic heart disease.

In addition to TZD and SU, select DPP-IV inhibitors (alogliptin and saxagliptin) have been associated with an increased risk of heart failure. Cardiovascular safety has been established with sitagliptin and more recently for linagliptin in the CArdiovascular safety and Renal Microvascular outcomE study with LINAgliptin (CARMELINA) study [31].

The cardiovascular benefit of metformin was previously suggested in UKPDS [32]. A recent systematic review and meta-analysis of the

effects of metformin has moderated this view. This demonstrated that metformin reduced the risk of all-cause mortality, cardiovascular disease (fatal and non-fatal) and peripheral vascular disease but potentially increased the risk of stroke (although the absolute risk is very small) [33].

Cardiovascular benefit has been demonstrated for empagliflozin, canagliflozin, and dapagliflozin with all three associated with a reduction in hospitalization for heart failure. The EMPA-REG OUTCOME trial demonstrated a reduction in cardiovascular outcome and death with empagliflozin compared to placebo [34]. For canagliflozin, the CANVAS trial showed a reduced risk of cardiovascular death and a reduction in hospitalization for heart failure [35]. This cardiovascular benefit was confirmed in the canagliflozin renal outcome study, CREDENCE, in a high-risk cohort with CKD [36]. CREDENCE demonstrated a reduction in the risk of cardiovascular death, myocardial infarction or stroke and a reduction in hospitalization for heart failure [36]. For dapagliflozin, the Dapagliflozin Effect on CardiovascuLAR Events (DECLARE-TIMI 58) study demonstrated a reduction in hospitalization due to heart failure, a lower rate of cardiovascular death and a neutral effect on major adverse cardiovascular events [37, 38]. Of note, the study populations for these three CVOTs differed and it is of interest that the SGLT-2 inhibitor that demonstrated the largest effect was studied in the highest cardiovascular risk population.

Cardiovascular benefit has been demonstrated in established cardiovascular disease for liraglutide, semaglutide, albiglutide, and dulaglutide. The Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results—A Long Term Evaluation (LEADER) trial demonstrated a reduction in cardiovascular death and death from any cause [39]. The SUSTAIN-6 trial demonstrated that semaglutide, a once-weekly GLP-1 receptor agonist, reduced cardiovascular death, non-fatal MI, and non-fatal stroke [40]. The HARMONY outcome trial comparing albiglutide to placebo found a reduction in a composite primary end point of major adverse cardiovascular events and a significant reduction in myocardial infarction [41]. The REWIND trial looking at dulaglutide demonstrated a reduction in the primary composite cardiovascular outcome compared to placebo [42]. While cardiovascular safety has been established for GLP-1 receptor agonists not all agents have shown superiority. Both the Evaluation of Lixisenatide in Acute Coronary Syndrome (ELIXA) study and the Exenatide Study of Cardiovascular Event Lowering (EXSCEL) study met the primary objective of non-inferiority but without added benefit [43].

The CVOTs for the newer antihyperglycaemic agents (GLP-1 receptor agonists and SGLT-2 inhibitors) have influenced European, American and Canadian guidelines which currently recommend adding one of these agents with proven benefit following lifestyle intervention and metformin in patents with atherothrombotic cardiovascular disease or in patients with clinical heart failure [3, 4].

Renal Impairment and Choice of Antihyperglycaemic Agent

Several antihyperglycaemic agents are untested in patients with moderate to severe CKD and therefore are not licensed or recommended below specific eGFR levels. Uncertainty rather than safety precludes their use. As study data and real-world data accumulates, this may influence licensing and prescribing of these newer agents.

There are several considerations with regards to antihyperglycaemic agent pharmacokinetics and pharmacodynamics in the context of renal impairment and renal failure [20]. This influences the appropriateness and dosage of therapy [44]. It should also be recognized that eGFR is inaccurate in various circumstance such as in obesity. In this group of patients the Cockcroft-Gault formula may be used.

Insulin requirements rise in the early stages of diabetic kidney disease and then fall [45]. Therefore, agents which promote hypoglycaemia (SU, meglitinides, and insulin) should be used with caution. SU and meglitinides are currently not recommended for use in dialysis [22]. Basal bolus insulin regimes are recommended in haemodialysis as these allow greater flexibility [22].

There are a number of other agent specific considerations with regards to the complications related to renal failure. Where renal failure is associated with fluid overload, agents associated with oedema, such as TZD, would be contraindicated. Where oedema is not a concern TZD can be continued without dose adjustment in renal impairment. However, TZD are not licensed for use in haemodialysis and currently not recommended [22].

The Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation (CREDENCE) trial evaluated the effects of 100 mg canagliflozin in patients with established renal disease and reported a reduction in the primary renal endpoint of 30% in the canagliflozin group [36]. There was a significant reduction in end stage renal failure at a median follow-up of 2.62 years [36]. This was the first study looking at primary renal endpoints in a population with established macroalbuminuria and diabetic kidney disease. In the CANVAS study, the renal outcomes were not considered statistically significant, however, this study complements the CREDENCE study. This showed reduced frequency of progression of albuminuria and reduced frequency of the composite renal outcome (reduction in eGFR, need for renal replacement, or death from renal causes) [35]. The DECLARE-TIMI study suggested similar renal benefit with a reduced frequency of renal events (reduction in eGFR, new endstage renal disease, or death from renal or cardiovascular causes) occurring in the dapagliflozin group [38]. The EMPA-REG OUTCOME trial demonstrated that empagliflozin reduced renal progression (microalbuminuria, doubling of creatinine and initiation of renal replacement) as a secondary end point [34, 46]. Studies looking specifically at the renal outcomes of empagliflozin (EMPA-KIDNEY) and dapagliflozin (DAPA-CKD) are awaited.

SGLT-2 inhibitors are currently licensed for initiation when eGFR >60 ml/min/1.73 m^2 and can be maintained in lower doses when eGFR remains >45 ml/min. These cut-off values are being re-evaluated with a likely reduction in the threshold eGFR for use with increased safety data from the cardiovascular and renal outcomes trials.

GLP-1 receptor agonists have a less evident effect on renal function. In acute illness they have been anecdotally linked with acute kidney injury. However, sustained reductions in albuminuria have been noted with dulaglutide, liraglutide, and semaglutide [40, 47, 48]. Definitive renal outcome studies with hard renal endpoints are awaited.

Hepatic Impairment and Choice of Antihyperglycaemic Agent

Elevation of liver enzymes is common in patients with type 2 diabetes, affecting up to 24% of diabetic patients screened [49]. This is often associated with non-alcoholic fatty liver disease (NAFLD). Hepatic impairment leads to several challenges. Firstly, most oral antihyperglycaemic agents are metabolized in the liver. Secondly, patients with hepatic impairment are at increased risk of hypoglycaemia [50]. Thirdly, hepatic impairment, unlike renal impairment, is less clearly defined and guidance can be misinterpreted.

National guidelines regarding the safety of drugs in hepatic impairment may reflect the paucity of data in these study populations rather than actual risk of harm. Most oral antihyperglycaemic agents can be used in mild chronic liver disease (i.e. where there is no evidence of decompensation, e.g. ascites or encephalopathy). However, caution is advised in moderate disease and most agents are contraindicated in severe disease [50, 51].

NAFLD is characterized by insulin resistance; thus agents, such as metformin, which improve insulin sensitivity may be beneficial in reversing disease pathology. This has also been demonstrated with pioglitazone in patients with biopsy proven non-alcoholic steatohepatitis (NASH) [52]. However, if the ALT levels increase further during therapy (>3× upper limit of normal) then it may be necessary to stop pioglitazone while assessing the aetiology.

SGLT-2 inhibitors undergo extensive hepatic metabolism. They are largely untested in severe liver disease and so while they can be used in mild hepatic impairment, they should be used cautiously and avoided in moderate to severe hepatic disease. However, the SGLT-2 inhibitor dapagliflozin has a documented benefit in fatty liver disease. In addition, in NASH, studies have found liraglutide to be safe with a beneficial effect on liver enzymes [53]. See Table 15.7.1.1.

Conclusion

The strategy for successful management of hyperglycaemia in type 2 diabetes depends on a patient-centred approach with careful consideration of the risks and benefits of the various agents. Currently, metformin is the first line recommended agent. Following this, different classes may be given equal consideration (where cost is not prohibitive). It is useful to consider the impact of the agent on weight given the increasing global prevalence of obesity. The main change in recommendations is due to cardiovascular and renal outcome trials demonstrating cardiovascular and renal benefits of specific agents. The results of further renal outcome trials may also have an impact on recommendations. In the future, long-term effectiveness trials of the newer agents will influence treatment algorithms.

REFERENCES

1. Ahlqvist E, Storm P, Käräjämäki A, *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*, 2018; **6**: 361–9.
2. Piette JD, Heisler M, Wagner TH. Cost-related medication underuse among chronically ill adults: the treatments people forgo, how often, and who is at risk. *Am J Public Health*, 2004; **94**: 1782–7.
3. Lipscombe L, Booth G, Butalia S, *et al.* Pharmacologic glycemic management of type 2 diabetes in adults. *Can J Diabetes*, 2018; **42** Suppl 1: S88–103.
4. Buse JB, Wexler DJ, Tsapas A, *et al.* 2019 update to: Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia*, 2020; **63**: 221–8.
5. Herman WH, Pan Q, Edelstein SL, *et al.* Impact of lifestyle and metformin interventions on the risk of progression to diabetes and regression to normal glucose regulation in overweight or obese people with impaired glucose regulation. *Diabetes Care*, 2017; **40**: 1668–77.
6. Lean ME, Leslie WS, Barnes AC, *et al.* Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet*, 2018; **391**: 541–51.
7. Steven S, Hollingsworth KG, Al-Mrabeh A, *et al.* Very low-calorie diet and 6 months of weight stability in type 2 diabetes: pathophysiological changes in responders and nonresponders. *Diabetes Care*, 2016; **39**: 808–15.
8. 9. Pharmacologic Approaches to glycemic treatment: standards of medical care in diabetes—2020. *Diabetes Care*, 2020; **43**(Suppl 1): S98–S110.
9. NICE. Type 2 diabetes in adults: management. Available at: https://www.nice.org.uk/guidance/ng28 (accessed: 23 April 2020).
10. Cai X, Yang W, Gao X, Zhou L, Han X, Ji L. Baseline body mass index and the efficacy of hypoglycemic treatment in type 2 diabetes: a meta-analysis. *PloS One*, 2016; **11**: e0166625.
11. Turner RC, Cull CA, Frighi V, Holman RR. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group. *JAMA*, 1999; **281**: 2005–12.
12. Adams TD, Davidson LE, Litwin SE, *et al.* Weight and metabolic outcomes 12 years after gastric bypass. *N Engl J Med*, 2017; **377**: 1143–55.
13. Rowles S, Kilvert A, Sinclair A. ABCD position statement on diabetes and end of life care. *Pract Diabetes Int*, 2011; **28**: 26–7.
14. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*, 2008; **359**: 1577–89.
15. Kramer CK, Zinman B, Retnakaran R. Short-term intensive insulin therapy in type 2 diabetes mellitus: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*, 2013; **1**: 28–34.
16. Gerstein HC, Miller ME, Byington RP, *et al.* Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*, 2008; **358**: 2545–59.
17. Gibbons CH, Freeman R. Treatment-induced diabetic neuropathy: a reversible painful autonomic neuropathy. *Ann Neurol*, 2010; **67**: 534–41.
18. Feldman-Billard S, Larger É, Massin P. Early worsening of diabetic retinopathy after rapid improvement of blood glucose control in patients with diabetes. *Diabetes Metab*, 2018; **44**: 4–14.
19. Salpeter S, Greyber E, Pasternak G, Salpeter E. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database Syst Rev*, 2006; **1**: CD002967.
20. Winocour P, Bain SC, Chowdhury TA, *et al.* Managing hyperglycaemia in patients with diabetes and diabetic

nephropathy–chronic kidney disease. Summary of recommendations 2018. *Br J Diabetes*, 2018; **18**: 78–89.
21. SIGN 154. Pharmacological management of glycaemic control in people with type 2 diabetes: a national clinical guideline. November 2017. Available at: https://www.sign.ac.uk/assets/sign154.pdf (accessed: 23 April 2020).
22. Frankel AH, Kazempour-Ardebili S, Bedi R, *et al.* Management of adults with diabetes on the haemodialysis unit: summary of guidance from the Joint British Diabetes Societies and the Renal Association. *Diabet Med*, 2018; **35**: 1018–26.
23. Holman RR, Farmer AJ, Davies MJ, *et al.* Three-year efficacy of complex insulin regimens in type 2 diabetes. *N Engl J Med*, 2009; **361**: 1736–47.
24. Davies M, Lavalle-González F, Storms F, Gomis R. Initiation of insulin glargine therapy in type 2 diabetes subjects suboptimally controlled on oral antidiabetic agents: results from the AT.LANTUS trial. *Diabetes Obes Metab*, 2008; **10**: 387–99.
25. Garber AJ, Larsen J, Schneider SH, Piper BA, Henry D. Simultaneous glyburide/metformin therapy is superior to component monotherapy as an initial pharmacological treatment for type 2 diabetes. *Diabetes Obes Metab*, 2002; **4**: 201–8.
26. Lincoff AM, Wolski K, Nicholls SJ, Nissen SE. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: a meta-analysis of randomized trials. *JAMA*, 2007; **298**: 1180–8.
27. Dormandy JA, Charbonnel B, Eckland DJA, *et al.* Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. *Lancet,* 2005; **366**: 1279–89.
28. Vaccaro O, Masulli M, Nicolucci A, *et al.* Effects on the incidence of cardiovascular events of the addition of pioglitazone versus sulfonylureas in patients with type 2 diabetes inadequately controlled with metformin (TOSCA.IT): a randomised, multicentre trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 887–97.
29. Tzoulaki I, Molokhia M, Curcin V, *et al.* Risk of cardiovascular disease and all cause mortality among patients with type 2 diabetes prescribed oral antidiabetes drugs: retrospective cohort study using UK general practice research database. *BMJ*, 2009; **339**: b4731.
30. Bolen S, Tseng E, Hutfless S, *et al. Diabetes Medications for Adults With Type 2 Diabetes: An Update*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2016.
31. Rosenstock J, Perkovic V, Johansen OE, *et al.* Effect of linagliptin vs placebo on major cardiovascular events in adults with type 2 diabetes and high cardiovascular and renal risk: the CARMELINA Randomized Clinical Trial. *JAMA*, 2019; **321** : 69–79.
32. [No authors]. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, 1998; **352**: 854–65.
33. Griffin SJ, Leaver JK, Irving GJ. Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. *Diabetologia*, 2017; **60**: 1620–9.
34. Zinman B, Wanner C, Lachin JM, *et al.* Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*, 2015; **373**: 2117–28.
35. Neal B, Perkovic V, Matthews DR. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*, 2017; **377**: 2099.
36. Perkovic V, Jardine MJ, Neal B, *et al.* Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*, 2019; **380**: 2295–306.
37. Cefalu WT, Leiter LA, de Bruin TW, *et al.* Dapagliflozin's effects on glycemia and cardiovascular risk factors in high-risk patients with type 2 diabetes: a 24-week, multicenter, randomized, double-blind, placebo-controlled study with a 28-week extension. *Diabetes Care*, 2015; **38**: 1218–27.
38. Wiviott SD, Raz I, Bonaca MP, *et al.* Dapagliflozin and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2019; **380**: 347–57.
39. Marso SP, Daniels GH, Brown-Frandsen K, *et al.* Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2016; **375**: 311–22.
40. Marso SP, Bain SC, Consoli A, *et al.* Semaglutide and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*, 2016; **375**: 1834–44.
41. Hernandez AF, Green JB, Janmohamed S, *et al.* Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet*, 2018; **392**: 1519–29.
42. Gerstein HC, Colhoun HM, Dagenais GR, *et al.* Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet*, 2019; **394**: 121–30.
43. Holman RR, Bethel MA, Mentz RJ, *et al.* Effects of once-weekly exenatide on cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2017; **377**: 1228–39.
44. Zac-Varghese S, Winocour P. Managing diabetic kidney disease. *Br Med Bull*, 2018; **125**: 55–66.
45. Gerich JE, Meyer C, Woerle HJ, Stumvoll M. Renal gluconeogenesis: its importance in human glucose homeostasis. *Diabetes Care*, 2001; **24**: 382–91.
46. Wanner C, Inzucchi SE, Lachin JM, *et al.* Empagliflozin and progression of kidney disease in type 2 diabetes. *N Engl J Med*, 2016; **375**: 323–34.
47. Tuttle KR, McKinney TD, Davidson JA, Anglin G, Harper KD, Botros FT. Effects of once-weekly dulaglutide on kidney function in patients with type 2 diabetes in phase II and III clinical trials. *Diabetes Obes Metab*, 2017; **19**: 436–41.
48. Mann JFE, Ørsted DD, Buse JB. Liraglutide and renal outcomes in type 2 diabetes. *N Engl J Med*, 2017; **377**: 2197–8.
49. Belcher G, Schernthaner G. Changes in liver tests during 1-year treatment of patients with Type 2 diabetes with pioglitazone, metformin or gliclazide. *Diabet Med*, 2005; **22**: 973–9.
50. Tolman KG, Fonseca V, Dalpiaz A, Tan MH. Spectrum of liver disease in type 2 diabetes and management of patients with diabetes and liver disease. *Diabetes Care*, 2007; **30**: 734–43.
51. Gangopadhyay KK, Singh P. Consensus statement on dose modifications of antidiabetic agents in patients with hepatic impairment. *Indian J Endocrinol Metab*, 2017; **21**: 341–54.
52. Cusi K, Orsak B, Bril F, *et al.* Long-term pioglitazone treatment for patients with nonalcoholic steatohepatitis and prediabetes or type 2 diabetes mellitus: a randomized trial. *Ann Int Med*, 2016; **165**: 305–15.
53. Armstrong MJ, Gaunt P, Aithal GP, *et al.* Liraglutide safety and efficacy in patients with non-alcoholic steatohepatitis (LEAN): a multicentre, double-blind, randomised, placebo-controlled phase 2 study. *Lancet*, 2016; **387**: 679–90.

15.7.2 Psychological and Behavioural Aspects of Type 2 Diabetes Management

Timothy C. Skinner and Jane Speight

Introduction

The challenge of type 2 diabetes is that, if individuals wish to minimize their risk of developing long-term complications, then they need to maintain their blood glucose, blood pressure, cholesterol, and their emotional well-being as close as possible to recommended levels [1]. This can be achieved, but only through significant and sustainable changes, including dietary choices, physical activity, taking medication, and monitoring glucose levels; for many, there is the additional challenge of needing to sustain some weight loss (1). All of these are independent risk factors for heart disease, stroke, retinopathy, nephropathy and neuropathy and mortality [2]. Unfortunately, managing these risks is complex and most people do not achieve recommended targets [3]. An important consideration is that biomedical risk factors (blood glucose, blood pressure, cholesterol) are largely asymptomatic, and a lack of biofeedback means little positive reward for optimal diabetes self-management. In the long term, at best, individuals are rewarded with an absence of complications some 10–20 years into the future (i.e. they maintain the status quo). It is little wonder that many people with type 2 diabetes find it challenging to maintain motivation for sustained optimal diabetes self-management.

In the past 20 years, substantial innovations in medical technologies, pharmaceutical agents, and digital technologies have not been matched with comparable improvements in type 2 diabetes outcomes [4]. The reality remains that diabetes is a self-managed condition, and outcomes will not improve until this is acknowledged fully by clinicians, researchers and policymakers [5]. Regardless of healthcare professional advice, people with diabetes make the day-to-day decisions about their food choices and activity levels, taking medications, and monitoring their health, and are in charge of managing *their* diabetes.

Given this, it is unsurprising that many people with type 2 diabetes experience high levels of distress. A recent meta-analysis (55 studies from 16 countries) indicated that 36% of people with type 2 diabetes experience clinically meaningful levels of diabetes-specific distress [6]. This diabetes distress is possibly the main barrier to effective self-management of diabetes. Further, some intervention studies have demonstrated that a reduction in diabetes distress is associated with improved glucose levels [7, 8].

Psychometric instruments have been developed to assess diabetes distress, and there is substantial overlap between scales [9]. For clinical purposes, it can be useful to think of five facets or 'SHADEs' of diabetes distress, and recognize that many people with diabetes feel: Scared, Helpless, Alone, Deprived, and Exhausted.

Scared

Individuals with type 2 diabetes report feeling worried about the future and the possibility of complications, and this is consistently one of the most commonly endorsed problems in diabetes [10]. This can surprise healthcare professionals, who may assume that individuals do not follow 'clear' advice and recommendations because they do not consider diabetes a serious condition—although, underestimation of risk may be the case for a minority.

Many healthcare professionals struggle to communicate balanced and effective messages regarding the seriousness of diabetes while conveying the message that it is possible to prevent complications. When they see that individuals are not making the self-care changes that they advised, they may use fear of complications as the 'stick' with which to motivate behaviour change. However, a review found that people usually overestimate their risk [11], with one study reporting individuals' perceptions of risk for coronary heart disease (CHD) and stroke about 3.5 and 5.5 times, respectively, greater than actual risk [12].

However, fear alone does not motivate action. A recent meta-analysis showed that increasing perceptions of threat (i.e. likelihood and severity of diabetes complications) does not result in behaviour change [13]. More importantly, when individuals have low confidence in their capability to address a health threat, increasing the perceived threat can have a negative impact on their health behaviours and self-care.

Increasing evidence suggests that providing individuals with accurate, personalized information about their complication risk can have positive effects on both emotional well-being and self-care. Randomized controlled trials (RCTs) have demonstrated the efficacy of providing people with type 2 diabetes with their actual cardiovascular risk and a small RCT has used personalized retinal imaging effectively [14–16]. People experience the provision of their actual risk as engaging and constructive. If this is used to support the individual to develop a specific action plan for managing their risk, it may improve blood glucose, pressure, and cholesterol, and reduce diabetes distress.

Helpless

Feeling scared is likely to be accompanied by feeling helpless, and a number of factors contribute to this. First, people with type 2 diabetes are usually given a long list of recommendations that they are told they need to follow. For example: *'eat less sugar, less fat, more fibre, more fruit and vegetables (but not too much fruit), be more physically active, take (up to 10 different) medications at the right time every day, monitor glucose levels, give up smoking, and limit alcohol intake'*.

People with type 2 diabetes usually know all these things—many are common public health messages—but few feel capable to achieve them all. There is little nuance about which are most important/effective actions, or any emphasis on *how* to develop these new healthy habits. So, if the person cannot do all these things, what is their response? Typically, to do none of them. And what is the healthcare professional's response? Typically, they assume the problem to be lack of knowledge—so, they give more advice/recommendations; or to be lack of motivation—so, they scare them a bit more.

Compounding this 'all or nothing' problem, people with diabetes receive different advice from different people, and struggle to know what advice to follow. Healthcare professionals in the same team often do not agree on basic principles of diabetes management [17, 18], and so different professionals in different roles can give conflicting information. This is without considering the advice from well-meaning friends and family, the stigmatizing consequences of well-intended media messages [19], and that written information is frequently incomprehensible to the people for whom it is intended [20, 21].

So, we need to help people appreciate the value of any, small change in reducing risk of complications. If they are even aware of targets, most are fixated on, and demotivated by, unattainable numbers rather than targeting incremental risk reduction. Again, this is where the freely available risk engines can be valuable, to demonstrate the impact of reducing just one risk factor. Critically important, though, is taking the time to appreciate what the individual understands, and what sense they make of all the disparate information to which they are exposed. Without this insight, we simply add to confusion and challenges.

Alone

Many individuals report feeling alone, or unsupported in managing diabetes [22]. In this context, there are several issues to consider: the 'diabetes police', lack of social support, perceived/experienced social stigma and that our words have power.

Nearly everyone living with diabetes has experienced the 'diabetes police'—the family, friends, colleagues, and acquaintances, who may be well-meaning with their comments but are often controlling or undermining [23]. Added to this, studies have now identified social stigma surrounding diabetes, often blaming and shaming people for bringing their condition upon themselves, leading to unwillingness to disclose their condition to others [24, 25], which can have serious implications for self-managing in public.

Added to these challenges, it is important to acknowledge the role that healthcare professionals (and their language) play in leaving people with diabetes feeling unsupported and stigmatized. In recent years, patient advocacy organizations, professional bodies and healthcare organizations have raised awareness that the words we use in diabetes care are often inappropriate, inaccurate, and harmful [26, 27] (**Box 15.7.2.1**).

Many of the words we use, such as 'adherence' or 'compliance', are not only inaccurate and inappropriate [29], they shape and inform a negative relationship with the person with diabetes. Thus, if we wish to reduce the distress that people with diabetes experience, and want them to feel supported in their self-management efforts, then we need to change our own attitudes, approach, and language [26, 29].

Box 15.7.2.1 The inappropriate, inaccurate, and harmful nature of the words we use in diabetes

The NHS England 'Language Matters' guide [28] indicates much of our language can be:

- Stigmatizing, e.g. *'you're in denial'*
- Shaming/blaming, e.g. *'it's your weight that is causing all these problems'*
- Authoritarian, e.g. *'you must take your medications properly in future'*
- Demanding, e.g. *'before you come to see me next, I want you to do 4 blood tests a day for 3 days, so I can check what's going wrong'*
- Disapproving, e.g. *'you aren't meant to take your insulin like that'*
- Discriminating, e.g. *'I don't think they'd get much from a diabetes education class'*
- Stereotyping, e.g. *'people from xx background often dislike the idea of injections'*
- Assumptive, e.g. *'I think you'd cope best with once a day insulin, as it's simpler'*
- Pre-judging, e.g. *'no-one in that family has ever taken much notice of their diabetes, they will be the same'*
- Judgemental, e.g. *'I think you're making the wrong decision'*
- Threatening, e.g. *'If you don't improve your control you will end up on insulin'*

Deprived

While self-managing diabetes is a significant burden, it is made harder by much of the language we use being focused on *not* doing things. *Not* eating sweet foods, *not* eating fat, *not* drinking sugary drinks. The language of reducing, stopping, and not doing creates a sense of deprivation. It is estimated that we make around 200 food/drink-related decisions every day [30]. If the majority of those decisions end up as 'no I can't', 'no I shouldn't' etc., then that is a lot of negativity every day. Therefore, it is unsurprising that many people with diabetes struggle with a sense of feeling deprived of the things that make life enjoyable for them.

This is why, when working with people with diabetes, the goal is to focus on what people *can* do. Many healthcare professionals have heard of SMART goal setting, but few report that A is for Actions, i.e. to focus on the thing someone *will* do, not on the thing they will *not* do, or negative goals such as weight loss, or reducing blood glucose. So, when talking about reducing dietary fat/sugar, the key is to help individuals reframe this, into a positive, i.e. what they are going replace/swap or do instead. Helping people focus on something they *can* achieve (i.e. keeping a focus on R for Realistic) reduces this sense of deprivation.

It is also worth considering some of the behavioural lessons derived from 'mindless eating' [31]. This focuses on how our environment shapes our thoughts, feelings, and behaviour around food. For example, the same amount of food, presented on a smaller plate, is perceived to be a larger portion size, and reported as more satisfying, than the same amount on a larger plate [32]. A wealth of factors shape our eating without our awareness, but these same factors can improve our food choices, if we become aware, and use them to our advantage.

Exhausted

Self-managing diabetes is a '24/7' task—it never goes away, so the effort to effectively manage glucose, blood pressure, and cholesterol, is unrelenting and exhausting. Then there is the challenge of a lack of tangible reward. The best possible outcome from all the person's efforts is that nothing bad happens; they just grow older like someone without diabetes. This means there is no sense of progress, and nothing to re-energize the individual for the continuous effort required to change and maintain that change. Moreover, let's not forget that the individual is trying to maintain healthy behaviours in the context of an 'obesogenic' social and cultural environment [33], which have shaped the person's lifestyle, potentially contributing to their development of diabetes in the first place. Individuals with diabetes are also at increased risk of having a number of sleep problems (see Chapters 1.11, 'Endocrinology, Sleep, and Circadian Rhythms' and 15.11.2, 'Type 2 Diabetes and Psychiatry'), and diabetes itself is associated with feeling fatigued and tired.

Therefore, it is critical to create mechanisms whereby individuals can connect to progress with their diabetes management. This is where self-monitoring plays a major role. A powerful tool to deal with the emotional exhaustion of diabetes management is to support individuals to find creative and informative ways to visualize both the problems and their progress toward a solution, ensuring positive rewards for their efforts. While many studies of self-monitoring of blood glucose appear to show it to be ineffective and burdensome for people with type 2 diabetes who are not using insulin, behaviourally-informed trials have demonstrated that this structured biofeedback improves engagement in medication taking, confidence in self-care, well-being and glucose outcomes [34, 35].

Conclusion

Living with type 2 diabetes (and being the recipient of a biomedical model, rather than a biopsychosocial model, of care) often results in high levels of diabetes distress. Many feel Scared, Helpless, Alone, Deprived, and Exhausted. By providing people with their actual risk of diabetes complications, helping them focus on realistic actions that target their concerns, using constructive and supportive language, helping them use their environment for change, sleep better, and find creative ways to monitor their health and visualize what they have achieved, we can reduce distress and improve both the physical and emotional health of people with type 2 diabetes.

REFERENCES

1. de Groot M, Anderson R, Freedland KE, *et al.* Association of depression and diabetes complications: a meta-analysis. *Psychosom Med*, 2001; **63**: 619–30.
2. van Dooren FE, Nefs G, Schram MT, *et al.* Depression and risk of mortality in people with diabetes mellitus: a systematic review and meta-analysis. *PLoS One*, 2013; **8**: e57058.
3. Khunti K, Ceriello A, Cos X, *et al.* Achievement of guideline targets for blood pressure, lipid, and glycaemic control in type 2 diabetes: a meta-analysis. *Diabetes Res Clin Pract*, 2018; **137**: 137–48.
4. Edelman SV, Polonsky WH. Type 2 diabetes in the real world: the elusive nature of glycemic control. *Diabetes Care*, 2017; **40**: 1425–32.
5. Speight J. Behavioural innovation is key to improving the health of one million Australians living with type 2 diabetes. *Med J Aust*, 2016; **205**: 149–51.
6. Perrin NE, Davies MJ, Robertson N, *et al.* The prevalence of diabetes-specific emotional distress in people with Type 2 diabetes: a systematic review and meta-analysis. *Diabet Med*, 2017; **34**: 1508–20.
7. Schmidt CB, van Loon BJP, Vergouwen ACM, *et al.* Systematic review and meta-analysis of psychological interventions in people with diabetes and elevated diabetes-distress. *Diabet Med*, 2019; **35**: 1157–72.
8. Sturt J, Dennick K, Hessler D, *et al.* Effective interventions for reducing diabetes distress: systematic review and meta-analysis. *International Diabetes Nursing*, 2015; **12**: 40–55.
9. Dennick K, Sturt J, Speight J. What is diabetes distress and how can we measure it? A narrative review and conceptual model. *J Diabetes Complications*, 2017; **31**: 898–911.
10. McGuire BE, Morrison TG, Hermanns N, *et al.* Short-form measures of diabetes-related emotional distress: the Problem Areas in Diabetes Scale (PAID)-5 and PAID-1. *Diabetologia*, 2010; **53**: 66–9.
11. Rouyard T, Kent S, Baskerville R, *et al.* Perceptions of risks for diabetes-related complications in type 2 diabetes populations: a systematic review. *Diabet Med*, 2017; **34**: 467–77.
12. Asimakopoulou KG, Skinner TC, Spimpolo J, *et al.* Unrealistic pessimism about risk of coronary heart disease and stroke in patients with type 2 diabetes. *Patient Educ Couns*, 2008; **71**: 95–101.
13. Peters GJ, Ruiter RA, Kok G. Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychol Rev*, 2013; **7**(Suppl 1): S8–31.
14. Benner JS, Erhardt L, Flammer M, *et al.* A novel programme to evaluate and communicate 10-year risk of CHD reduces predicted risk and improves patients' modifiable risk factor profile. *Int J Clin Pract*, 2008; **62**: 1484–98.
15. Tawfik MY, Mohamed RA. The impact of communicating cardiovascular risk in type 2 diabetics on patient risk perception, diabetes self-care, glycosylated hemoglobin, and cardiovascular risk. *J Public Health*, 2016; **24**: 153–64.
16. Rees G, Lamoureux EL, Nicolaou TE, *et al.* Feedback of personal retinal images appears to have a motivational impact in people with non-proliferative diabetic retinopathy and suboptimal HbA1c: findings of a pilot study. *Diabet Med*, 2013; **30**: 1122–5.
17. Skinner TC, Lange KS, Hoey H, *et al.* Targets and teamwork: understanding differences in pediatric diabetes centers treatment outcomes. *Pediatr Diabetes*, 2018; **19**: 559–65.
18. Swift PG, Skinner TC, de Beaufort CE, *et al.* Target setting in intensive insulin management is associated with metabolic control: the Hvidoere childhood diabetes study group centre differences study 2005. *Pediatr Diabetes*, 2010; **11**: 271–8.
19. Browne JL, Zimmet P, Speight J. Individual responsibility for reducing obesity: the unintended consequences of well intended messages. *Med J Aust*, 2011; **195**: 386.
20. Dorcely B, Agarwal N, Raghuwanshi M. Quality assessment of diabetes online patient education materials from academic institutions. *Health Education Journal*, 2015; **74**: 568–77.
21. Hutchinson N, Baird GL, Garg M. Examining the reading level of internet medical information for common internal medicine diagnoses. *Am J Med*, 2016; **129**: 637–9.
22. Pouwer F, Skinner TC, Pibernik-Okanovic M, *et al.* Serious diabetes-specific emotional problems and depression in a

Croatian-Dutch-English Survey from the European Depression in Diabetes [EDID] Research Consortium. *Diabetes Res Clin Pract*, 2005; **70**: 166–73.
23. Newton-John TR, Ventura AD, Mosely K, *et al.* 'Are you sure you're going to have another one of those?': a qualitative analysis of the social control and social support models in type 2 diabetes. *J Health Psychol*, 2017; **22**: 1819–29.
24. Browne JL, Ventura A, Mosely K, *et al.* 'I call it the blame and shame disease': a qualitative study about perceptions of social stigma surrounding type 2 diabetes. *BMJ Open*, 2013; **3**: e003384.
25. Nicolucci A, Kovacs Burns K, Holt RI, *et al.* Correlates of psychological outcomes in people with diabetes: results from the second Diabetes Attitudes, Wishes and Needs (DAWN2) study. *Diabet Med*, 2016; **33**: 1194–203.
26. Speight J, Conn J, Dunning T, *et al.* Diabetes Australia position statement. A new language for diabetes: improving communications with and about people with diabetes. *Diabetes Res Clin Pract*, 2012; **97**: 425–31.
27. Dickinson JK, Guzman SJ, Maryniuk MD, *et al.* the use of language in diabetes care and education. *Diabetes Care*, 2017; Oct: dci170041.
28. NHS England. Language Matters, 2018. Available at: https://www.england.nhs.uk/wp-content/uploads/2018/06/language-matters.pdf (accessed: 23 April 2020).
29. Anderson RM, Funnell MM. Compliance and adherence are dysfunctional concepts in diabetes care. *Diabetes Educ*, 2000; **26**: 597–604.
30. Wansink B, Sobal J. Mindless eating: the 200 daily food decisions we overlook. *Environ Behav*, 2007; **39**: 106–23.
31. Wansink B. *Mindless Eating: Why We Eat More Than We Think*. New York: Bantam, 2007.
32. Van Ittersum K, Wansink B. Plate size and color suggestibility: the delboeuf illusion's bias on serving and eating behavior. *JCR*, 2012; **39**: 215–28.
33. Egger G, Swinburn B. An 'ecological' approach to the obesity pandemic. *BMJ*, 1997; **315**: 477–80.
34. Polonsky WH, Fisher L, Schikman CH, *et al.* Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. *Diabetes Care*, 2011; **34**: 262–7.
35. Speight J, Browne JL, Furler J. Challenging evidence and assumptions: is there a role for self-monitoring of blood glucose in people with type 2 diabetes not using insulin? *Curr Med Res Opin*, 2013; **29**: 161–8.

15.7.3 Type 2 Diabetes in Different Ethnic Groups

Nitin Narayan Gholap and Kamlesh Khunt

Introduction

The prevalence of type 2 diabetes mellitus (T2DM) and its associated complications varies significantly between populations, with non-white ethnic populations, indigenous groups, and developing countries bearing a disproportionately higher burden of the disease. High-quality studies examining various aspects of T2DM and its management among diverse non-white ethnic groups remain limited. Furthermore, ethnic disparities in health may vary between countries due to various factors. However, existing evidence has identified common themes across these groups as being key contributors to their greater burden of disease. In the increasingly globalized world, healthcare professionals, and policymakers need to understand ethnic disparities in relation to T2DM and their determinants to provide effective care.

Ethnicity and Health

An ethnic group can be defined as a group of people with shared ancestral, social, cultural, religious, lingual, ideological, *or* homeland background and biological features which are distinctive and maintained between generations. The study of ethnicity in health is important because it allows us to identify patterns of given diseases among people with shared ethnic characteristics, thereby helping us to assess health needs, formulate health policies, develop treatments, plan service delivery strategies, and identify areas for further research.

The Epidemiology of Type 2 Diabetes

Latest estimates from the International Diabetes Federation (IDF) suggest that there are currently 451 million (8.8%) adults aged 20–79 years with diabetes worldwide, compared with 151 million (4.6%) in the year 2000. This figure is expected to increase to 629 million (9.9%) by 2045 [1]. While not differentiating between types of diabetes, these estimates are very much representative of T2DM. However, there is a significant geographical and ethnic variation in prevalence, with the global burden of the disease rising much faster among non-white ethnic groups (see **Figure 15.7.3.1** and **Table 15.7.3.1**) [1–5]. This is reflected in the variation seen in the rates of diabetes across the seven IDF geographical regions (**Figure 15.7.3.1**) and among multiethnic populations in Western countries. The estimated age-adjusted prevalence of diabetes is over 12% in Mexico and some of the Middle East/North African countries, and 9–12% in India, China, and North America, which is higher than in Europe and Australia (**Figure 15.7.3.1**). It is further predicted that differences between the countries will become more pronounced in the future. Between 2017 and 2045, prevalence of diabetes is set to rise by 156% in sub-Saharan Africa, 110%

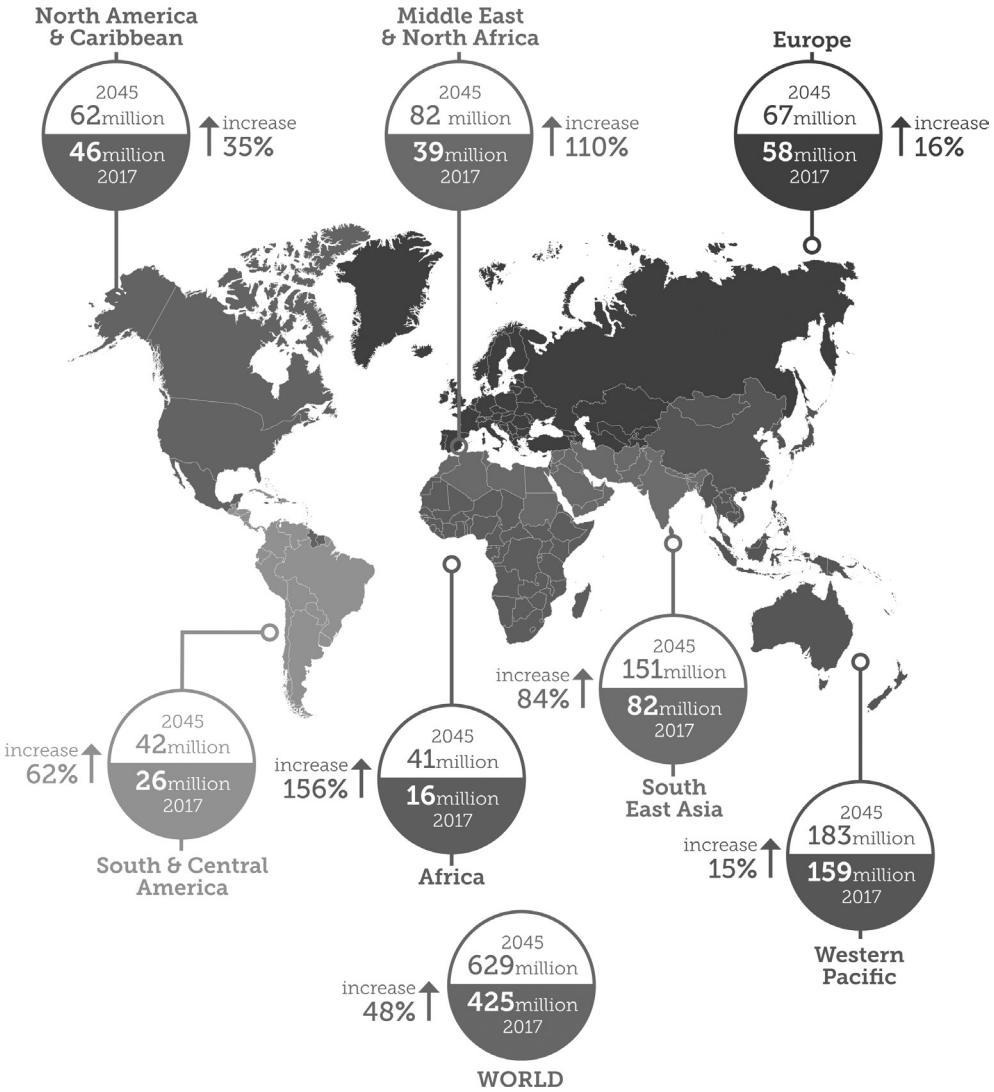


Figure 15.7.3.1 Increase in number of people with diabetes per region between 2017 and 2045.
Adapted with permission from International Diabetes Federation. *IDF diabetes atlas*, 8th ed. International Diabetes Federation; 2017. Available at: http://www.diabetesatlas.org1

Table 15.7.3.1 Top ten countries with highest number of people with diabetes (age 20–79 years) in 2017 and 2045

Rank	Country ranking 2017	Number of people with diabetes (millions), 2017	Country ranking 2045	Number of people with diabetes (millions), 2045
1	China	114.4	India	134.3
2	India	72.9	China	119.8
3	United States	30.2	United States	35.6
4	Brazil	12.5	Mexico	21.8
5	Mexico	12.0	Brazil	20.3
6	Indonesia	10.3	Egypt	16.7
7	Russia	8.5	Indonesia	16.7
8	Egypt	8.2	Pakistan	16.1
9	Germany	7.5	Bangladesh	13.7
10	Pakistan	7.5	Turkey	11.2

Adapted with permission from International Diabetes Federation. IDF diabetes atlas, 8th ed. International Diabetes Federation; 2017. Available at: http://www.diabetesatlas.org

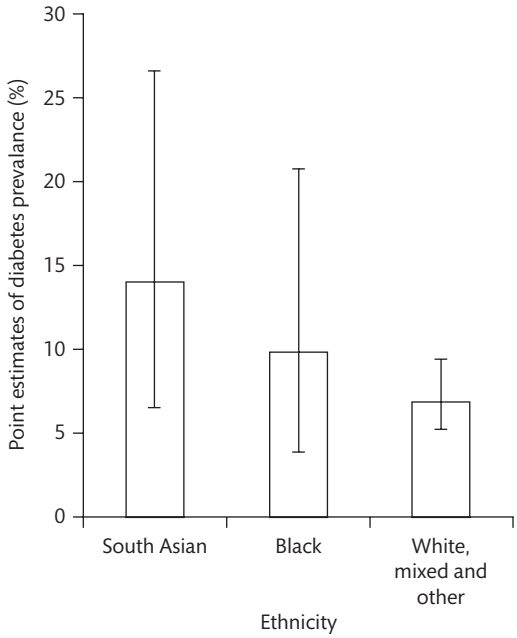


Figure 15.7.3.2 Summary estimates of diabetes prevalence by ethnicity for England in 2010. Error bars show the uncertainty range.

Adapted with permission from Holman N, Forouhi NG, Goyder E, *et al.* The Association of Public Health Observatories (APHO) diabetes prevalence model: estimates of total diabetes prevalence for England, 2010–2030. *Diabet Med* 2011;28:575–82; with data from Hanif W, Khunti K, Bellary S, *et al.* Type 2 diabetes in the UK South Asian population. An update from the South Asian Health Foundation; 2014. Available at: https://www.sahf.org.uk/publications.

in North Africa, and the Middle East, 84% in South East Asia and 62% in South and Central America compared with 16% in Europe (Figure 15.7.3.1).

The UK population comprises 13% of people from ethnic minority groups, with South Asian (people originating from India, Pakistan, Bangladesh, or Sri Lanka) and black African populations forming the two largest groups. The prevalence of diabetes (known and undiagnosed) in the South Asian population in England is 2–4 times higher than the general population, at 14.0% compared with 6.9% in the general population. The corresponding figure for the black African population is about 10% (Figure 15.7.3.2) [2]. Similarly, the incidence of T2DM is about three-times higher in first generation British South Asians (34%) and two-times higher in British black Africans (30%) compared with the native white (14%) population, as seen in the SABRE (Southall and Brent Revisited) study [3]. Higher incidence and prevalence of T2DM have also been reported among different ethnic minority groups in the United States [4], Canada, Australia, and European countries. Disproportionately high rates of T2DM are also seen in certain indigenous populations living in developed countries; for example, Pima Indians in the United States [5].

Among non-white ethnic groups, T2DM is more common in young and middle-aged individuals [5]. In a recent UK study, South Asian, and black African individuals developed T2DM at a mean age of 46 and 48 years respectively, with 31% and 23% developing it below the age of 40 [6]. In contrast, the mean age in white individuals was 58 years and only 9% developed T2DM below 40 years of age [6]. In recent years, there has also been an alarming rise in T2DM and obesity among children and adolescents from indigenous groups in the United States and Australia, and in black African, Hispanic, and Asian populations [7].

Table 15.7.3.2 Top 10 countries with highest number of people with undiagnosed diabetes in 2017

Rank	Country ranking 2017	Number of people with undiagnosed diabetes (millions), 2017	Proportion undiagnosed
1	China	61.3	53.6%
2	India	42.2	57.9%
3	United States	11.5	38.2%
4	Indonesia	7.6	73.7%
5	Brazil	5.7	46.0%
6	Pakistan	4.6	61.5%
7	Russia	4.5	53.7%
8	Mexico	4.5	37.4%
9	Egypt	4.4	53.1%
10	Bangladesh	3.9	56.0%

Adapted with permission from International Diabetes Federation. IDF diabetes atlas, 8th ed. International Diabetes Federation; 2017. Available at: http://www.diabetesatlas.org.

Diabetes remains undiagnosed in about 50% of the population globally; this proportion is much higher in low-to-middle income countries where the proportion of undiagnosed diabetes is 85% (Table 15.7.3.2) [1]. While this can be partly explained by inadequate resources and lack of diabetes screening programmes, undiagnosed diabetes remains high among South Asian and black African populations in countries with free access to care, such as the United Kingdom.

Pathogenesis and Risk Factors

The precise mechanisms behind the excess risk of T2DM in non-white populations remain unclear. However, it is indisputably driven by the heightened adverse impact of rapid economic development and urbanization, especially of physical inactivity, energy-rich diet and obesity, in the background of genetic predisposition [8–14]. Research has not yet identified distinct genetic variants that fully explain the excess risk, and genetic susceptibility accounts for less than 10% of total diabetes risk in any ethnic group [14, 15]. In contrast, epigenetic changes may be more pertinent to excess risk of T2DM in non-white groups as described next [16–19].

Development Origins of Health and Diseases (DOHaD)

The thrifty phenotype [18] and Development Origins of Health and Diseases (DOHaD) [16, 19] postulate that nutritional mismatch between periconception/intrauterine/infancy and adult life significantly increases risk of obesity and T2DM in later life, possibly through epigenetic changes. These epigenetic alterations are mitotically heritable changes in gene expression caused by distinct mechanisms of histone modification and DNA methylation, without involving changes in the DNA sequence *per se* [5, 17]. According to the thrifty phenotype hypothesis, maternal

malnutrition, and low birth weight lead to a programmed tendency towards nutritional thrift, impaired fetal pancreatic development, and beta-cell function, impaired peripheral glucose uptake and insulin resistance [18]. When such individuals undergo rapid weight gain during the postnatal period, especially due to abundance of food and physical inactivity, this can result in disturbances in glucose metabolism and adverse fat deposition [16, 18, 19]. A recent meta-analysis highlighted the inverse relationship between low birth weight and excess risk of T2DM and further found that each 1 kg gain in birth weight reduced the risk by 25% [20]. Apart from undernutrition, intrauterine overnutrition from maternal obesity, hyperglycaemia, or T2DM and resultant high birth weight is also associated with risk of T2DM in offspring at a younger age, mediated again possibly via epigenetic effect [21]. Such phenomena related to early-life nutritional status and intergenerational effect of hyperglycaemia are pertinent to the rapid emergence of T2DM in fast-developing countries, especially in South Asia, China, and the Middle East [1, 5, 22].

Obesity, Insulin Resistance, and Beta-Cell Dysfunction

Non-white ethnic groups tend to develop glucometabolic abnormalities at lower thresholds of obesity [6, 23]. In a study of 490 288 UK Biobank participants, for an equivalent prevalence of T2DM, body mass index—BMI (and waist circumference) was significantly lower in non-white participants (South Asian 22.0 kg/m^2; black African 26.0 kg/m^2; Chinese women 24.0 kg/m^2; Chinese men 26.0 kg/m^2) compared with white participants (30 kg/m^2) [23]. A possible explanation for this is that non-white populations, particularly Asians, have a high total body fat percentage ('thin-fat' phenotype) that is distributed mainly in the abdomen and ectopic areas of the liver and muscles, which could lead to insulin resistance at lower BMI [10, 22–25]. Such ethnic discrepancies have led to recommendations to use ethnic-specific cut-offs for defining obesity or metabolic syndrome. According to emerging evidence, early beta-cell impairment is also a major determinant of increased risk of T2DM in Asian groups [11–13]. This may reflect early beta-cell 'exhaustion' following years of insulin hypersecretion in response to higher degrees of insulin resistance, or reduced beta-cell mass/capacity/function as a consequence of adverse early-life factors or genetic predisposition [11–13].

Lifestyle Habits and Social and Cultural Determinants of Health

The rapid rise in obesity and T2DM in non-white ethnic groups is largely attributed to changes in lifestyle habits. There has been a major shift in nutrition towards consumption of diets rich in calories, sugar (including sweetened beverages), saturated fat and processed meat, and low in fibre and fruit and vegetables. In addition, multiple studies report significantly low levels of physical activity in non-white ethnic groups [9–12, 14]. Unawareness of the benefits of physical activity, lack of time and space, and, for women from Asia and the Middle East, cultural norms are major barriers to engaging in leisure-time exercise [9–11, 22, 24–26]. Additionally, health illiteracy, religious beliefs that attribute disease to being an inevitable 'destiny', and poor socioeconomic [27], educational, and psychological status all contribute to ethnic disparities in T2DM risk and its outcomes.

Mortality and Complications

High-risk ethnic populations and geographical regions are generally at a greater risk of developing associated adverse outcomes. In 2017, diabetes accounted for an estimated 10.7% of total mortality worldwide, with premature and higher mortality rates seen in South Asia, the Middle East, and North Africa, Central and South American regions [1]. In an analysis of pooled mortality data from different time periods from 1991 to 2007 in seven European countries, adjusted mortality rate ratios were about twofold higher in migrants from developing countries compared with the native population, with the highest mortality observed in people from the Caribbean and South Asia [28]. Likewise, in the United States, non-Hispanic black, native Americans and Alaskans, and Hispanic Americans have been found to be 2.1, 1.8, and 1.4 times respectively more likely to die from diabetes compared with non-Hispanic whites [14, 29].

Macrovascular Complications

The prevalence of cardiovascular disease (CVD) and associated mortality among people with T2DM is disproportionately high in developing countries such as India, China, and Russia [1]. Certain ethnic populations, particularly South Asians, suffer from excess overall CVD mortality, likely due to higher burden of insulin resistance, T2DM and related atherosclerotic risk factors [9, 10]. It is also possible that T2DM is more adversely associated with CVD risk in some non-white ethnic groups. In the SABRE study, during a median 20.5-year follow-up, diabetes was associated with a 3.0-fold and 2.5-fold increase in age-adjusted incidence of stroke in British black African and South Asian individuals respectively, compared with 1.3-fold excess risk in the native white group [30]. In the UKPDS (UK Prospective Diabetes Study), at the median 18-year follow-up, the adjusted relative risk was non-significantly high for stroke (1.18 (0.97–1.43)) in black African/Caribbean individuals and for myocardial infarction (1.11 (0.96–1.28)) in South Asian individuals compared with white individuals with T2DM [31]. In a recent retrospective case–control study, the risk of major cardiovascular events (MACE) among patients with T2DM compared with those without T2DM was significantly higher in overweight and obese South Asian patients in comparison with their white counterparts [32]. The risk of coronary heart disease in Chinese and black African individuals with T2DM is generally low compared with white individuals. Similarly, the relationship between ethnicity and peripheral artery disease in patients with T2DM shows rather mixed results, with a higher risk observed in black Africans and a lower risk in Asian populations.

Microvascular Complications

Individuals with T2DM of South Asian, African American, Middle Eastern, Hispanic, Polynesian and from indigenous populations are at a particularly high risk of developing chronic kidney disease (CKD) and its progression to end-stage renal disease (ESRD) [33]. In the UK and elsewhere, Asian and black African patients suffer from increased risk of albuminuria or proteinuria, rapid CKD progression, and higher incidence of stage 4 or 5 CKD [34]. Similarly, a recent systematic review reported a generally higher prevalence of diabetic retinopathy among ethnic minority groups, indigenous populations and developing countries [35]. In a cross-sectional analysis of the DRIVE UK (Diabetic Retinopathy in Various Ethnic

Groups) study, the age-standardized prevalence of any diabetic retinopathy, significant macular oedema and sight-threatening retinopathy was significantly higher in black African and South Asian compared with white patients [36]. The prevalence of diabetic peripheral neuropathy is generally similar among white, black African, and Hispanic populations but low in the Asian population.

Implications for Management of T2DM

Apart from possible genetic predisposition, much of the excess risk of adverse outcomes in non-white ethnic populations can be attributed to poorer diabetes management and control of conventional risk factors [37]. Individuals from non-white groups within and between countries are generally less likely to: adhere to medications and diabetes self-management; engage in physical activity; receive insulin therapy despite poor glycaemia control; and achieve treatment targets for HbA1c, blood pressure, and cholesterol [14, 37, 38]. The major barriers to effective diabetes care in these groups include issues with access and quality of care, lack of cultural adaptation to diabetes management, limited knowledge, and misconceptions about diabetes management including the role of preventative care, cultural health beliefs and behaviours and social determinants of health [14, 27, 37, 39]. In addition, the low participation of non-white ethnic groups in clinical trials has resulted in lack of knowledge on ethnic-specific treatment targets and efficacy of various pharmacological and non-pharmacological interventions. Nonetheless, the available evidence suggests that intensive control of risk factors as per the current targets derived predominantly from the white population are beneficial in non-white ethnic groups [31]. However, challenges remain in the delivery of care. To achieve desired clinical outcomes, a greater emphasis needs to be placed on a comprehensive diabetes management plan that combines pharmacotherapy with culturally adapted diabetes education and advice on self-management, regular exercise, and dietary changes [39, 40]. Additionally, in middle- and low-income countries, more emphasis is needed on developing and implementing national policies and programmes to raise awareness about T2DM and its complications, deliver screening, facilitate early diagnosis, develop prevention strategies for diabetes complications and, to improve organization of overall diabetes care [41]. Such strategies have been successful in significantly reducing rates of diabetes complications such as CVD and all-cause mortality in high-income countries [42]. In addition, wide-scale screening, early detection, and prevention of T2DM through national programmes are equally important in middle and low-income countries to reduce future burden of the disease and associated resources [42]. Any diabetes screening programme needs to use ethnic-specific strategies, for example, screening at a younger age, using lower BMI cut-points, and adopting a more aggressive and culturally adapted approach to diabetes prevention.

Summary

Non-white ethnic populations across the globe are facing an extremely worrying health, economic, and societal challenge from the rapidly growing burden of T2DM. Compared with white populations they have a 2–4-fold higher incidence and prevalence of T2DM. They also suffer from higher rates of certain diabetes-related complications. However, much of the excess burden of T2DM in non-white populations can be attributed to poor control of modifiable risk factors and suboptimum diabetes care. To improve the situation, a culturally adapted and innovative approach to the delivery of diabetes care is needed. This approach should focus on the proactive management of risk factors, address misconceptions about T2DM, and encourage lifestyle behaviour change. Further consideration should be given to early detection, control, and prevention of complications of T2DM and targeting high-risk individuals for prevention of T2DM.

REFERENCES

1. International Diabetes Federation. *IDF Diabetes Atlas*, 8th edn. International Diabetes Federation, 2017. Available at: http://www.diabetesatlas.org (accessed: 23 April 2020).
2. Holman N, Forouhi NG, Goyder E, *et al.* The Association of Public Health Observatories (APHO) diabetes prevalence model: estimates of total diabetes prevalence for England, 2010–2030. *Diabet Med*, 2011; **28**: 575–82.
3. Tillin T, Hughes AD, Godsland IF, *et al.* Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. *Diabetes Care*, 2012; **36**: 383–93.
4. Menke A, Casagrande S, Geiss L, *et al.* Prevalence of and trends in diabetes among adults in the United States, 1988–2012. *JAMA*, 2015; **314**: 1021–9.
5. Zimmet PZ, Magliano DJ, Herman WH, *et al.* Diabetes: a 21st century challenge. *Lancet Diabetes Endocrinol*, 2014; **2**: 56–64.
6. Paul SK, Owusu Adjah ES, Samanta M, *et al.* Comparison of body mass index at diagnosis of diabetes in a multi-ethnic population: a case-control study with matched non-diabetic controls. *Diabetes Obes Metab*, 2017; **19**: 1014–23.
7. Viner R, White B, Christie D. Type 2 diabetes in adolescents: a severe phenotype posing major clinical challenges and public health burden. *Lancet*, 2017; **389**: 2252–60.
8. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care*, 2011; **34**: 1249–57.
9. Gholap N, Davies M, Patel K, *et al.* Type 2 diabetes and cardiovascular disease in South Asians. *Prim Care Diabetes*, 2011; **5**: 45–56.
10. Sattar N, Gill JM. Type 2 diabetes in migrant south Asians: mechanisms, mitigation, and management. *Lancet Diabetes Endocrinol*, 2015; **3**: 1004–16.
11. Hanif W, Khunti K, Bellary S, *et al.* Type 2 diabetes in the UK South Asian population. An update from the South Asian Health Foundation; 2014. Available at: https://www.sahf.org.uk/publications (accessed: 24 April 2020).
12. Ma RC, Lin X, Jia W. Causes of type 2 diabetes in China. *Lancet Diabetes Endocrinol*, 2014; **2**: 980–91.
13. Kodama K, Tojjar D, Yamada S, *et al.* Ethnic differences in the relationship between insulin sensitivity and insulin response: a systematic review and meta-analysis. *Diabetes Care*, 2013; **36**: 1789–96.
14. Golden SH, Brown A, Cauley JA, *et al.* Health disparities in endocrine disorders: biological, clinical, and nonclinical factors—an Endocrine Society scientific statement. *J Clin Endocrinol Metab*, 2012; **97**: E1579–639.

15. Haiman CA, Fesinmeyer MD, Spencer KL, *et al.* Consistent directions of effect for established type 2 diabetes risk variants across populations: the Population Architecture using Genomics and Epidemiology (PAGE) Consortium. *Diabetes*, 2012; **61**: 1642–8.
16. Gluckman PD, Hanson MA, Cooper C, *et al.* Effect of in-utero and early-life conditions on adult health and disease. *N Engl J Med*, 2008; **359**: 61–73.
17. Pinney SE, Simmons RA. Epigenetic mechanisms in the development of type 2 diabetes. *Trends Endocrinol Metab*, 2010; **21**: 223–9.
18. Hales CN, Barker DJ Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia*, 1992; **35**: 595–601.
19. Fleming TP, Watkins AJ, Velazquez MA, *et al.* Origins of lifetime health around the time of conception: causes and consequences. *Lancet*, 2018; **391**: 1842–52.
20. Whincup PH, Kaye SJ, Owen CG, *et al.* Birth weight and risk of type 2 diabetes: a systematic review. *JAMA*, 2008; **300**: 2886–97.
21. Ma RC, Chan JC, Tam WH, *et al.* Gestational diabetes, maternal obesity, and the NCD burden. *Clin Obstet Gynecol*, 2013; **56**: 633–41.
22. Nanditha A, Ma RC, Ramachandran A, *et al.* Diabetes in Asia and the Pacific: implications for the global epidemic. *Diabetes Care*, 2016; **39**: 472–85.
23. Ntuk UE, Gill JM, Mackay DF *et al.* Ethnic specific obesity cutoffs for diabetes risk: cross-sectional study of 490,288 UK biobank participants. *Diabetes Care*, 2014; **37**: 2500–7.
24. Kengne AP, Echouffo-Tcheugui JB, Sobngwi E, *et al.* New insights on diabetes mellitus and obesity in Africa–Part 1: prevalence, pathogenesis and comorbidities. *Heart*, 2013; **99**: 979–83.
25. Aschner P, Aguilar-Salinas C, Aguirre L, *et al.* Diabetes in South and Central America: an update. *Diabetes Res Clin Pract*, 2014; **103**: 238–43.
26. Ye J. Mechanisms of insulin resistance in obesity. *Front Med*, 2013; **7**: 14–24.
27. Walker RJ, Smalls BL, Campbell JA, *et al.* Impact of social determinants of health on outcomes for type 2 diabetes. *Endocrine*, 2014; **47**: 29–48.
28. Vandenheede H, Deboosere P, Stirbu I, *et al.* Migrant mortality from diabetes mellitus across Europe: the importance of socioeconomic change. *Eur J Epidemiol*, 2012; **27**: 109–17.
29. Office of Minority Health 2012. Minority population profiles. Available at: https://www.minorityhealth.hhs.gov/omh/browse.aspx?lvl=2&lvlID=26 (accessed: 23 April 2020).
30. Tillin T, Hughes AD, Mayet J, *et al.* The relationship between metabolic risk factors and incident cardiovascular disease in Europeans, South Asians, and African Caribbeans: SABRE (Southall and Brent Revisited)—a prospective population-based study. *J Am Coll Cardiol*, 2013; **61**: 1777–86.
31. Davis TM, Coleman RL, Holman RR, UKPDS Group. Ethnicity and long-term vascular outcomes in Type 2 diabetes: a prospective observational study (UKPDS 83). *Diabet Med*, 2014; **31**: 200–27.
32. Adjah ES, Bellary S, Hanif W, *et al.* Prevalence and incidence of complications at diagnosis of T2DM and during follow-up by BMI and ethnicity: a matched case–control analysis. *Cardiovasc Diabetol*, 2018; **17**: 70.
33. Thomas MC, Cooper ME, Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. *Nat Rev Nephrol*, 2016; **12**: 73–81.
34. Dreyer G, Hull S, Mathur R, *et al.* Progression of chronic kidney disease in a multi-ethnic community cohort of patients with diabetes mellitus. *Diabet Med*, 2013; **30**: 956–63.
35. Ruta LM, Magliano DJ, Lemesurier R, *et al.* Prevalence of diabetic retinopathy in type 2 diabetes in developing and developed countries. *Diabet Med*, 2013; **30**: 387–98.
36. Sivaprasad S, Gupta B, Gulliford MC, *et al.* Ethnic variations in the prevalence of diabetic retinopathy in people with diabetes attending screening in the United Kingdom (DRIVE UK). *PLoS One*, 2012; **7**: e32182.
37. Lanting LC, Joung IM, Mackenbach JP, *et al.* Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. *Diabetes Care*, 2005; **28**: 2280–8.
38. Verma A, Birger R, Bhatt H, *et al.* Ethnic disparities in diabetes management: a 10-year population-based repeated cross-sectional study in UK primary care. *J Public Health*, 2010; **32**: 250–8.
39. Sohal T, Sohal P, King-Shier KM, *et al.* Barriers and facilitators for type-2 diabetes management in South Asians: a systematic review. *PloS One*, 2015; **10**: e0136202.
40. Hawthorne K, Robles Y, Cannings-John R, *et al.* Culturally appropriate health education for Type 2 diabetes in ethnic minority groups: a systematic and narrative review of randomized controlled trials. *Diabet Med*, 2010; **27**: 613–23.
41. Khunti K, Gavin JR, Boulton AJ, *et al.* The Berlin Declaration: a call to improve early actions related to type 2 diabetes. Why is primary care important?. *Prim Care Diabetes*, 2018; **12**: 383–92.
42. Gregg EW, Sattar N, Ali MK. The changing face of diabetes complications. *Lancet Diabetes Endocrinol*, 2016; **4**: 537–47.

15.7.4 Prevention of Type 2 Diabetes

Nicholas J. Wareham

Introduction

There is a growing body of literature which has unequivocally established that type 2 diabetes is preventable. This chapter reviews that evidence and then proceeds to consider how that evidence should be translated into preventive action. Since this textbook is focused on clinical management, the perspective that will be taken

is predominantly a clinical rather than a public health one. However it is impossible to thoroughly describe the topic of diabetes prevention without considering public health approaches and thus these are described here not least because clinicians are important advocates for this important complementary strategy.

Is Type 2 Diabetes Preventable by Behavioural Interventions and Weight Loss?

A series of randomized controlled trials (RCT) have shown than behavioural interventions aimed at dietary change, physical activity, and weight loss are effective in halving the risk of progression from impaired glucose tolerance to diabetes. The first major RCT was conducted in Da Qing in the North Eastern province of Heilongjiang in China. In this trial, 577 people who had impaired glucose tolerance (IGT) (2-hour plasma glucose 7.8—< 11.1 mmol/L) were cluster randomized to a diet intervention alone, physical activity alone, both interventions together or standard care [1]. The cumulative incidence of diabetes in the control group at 6 years was very high (67.7%) and in an analysis adjusting for baseline differences in body mass index (BMI) and fasting glucose, the diet intervention was associated with a 31% reduction in risk (P <0.03). In the physical activity only group, the risk reduction was 46% (P <0.0005) but there was no evidence of an additional effect of both interventions, as the risk reduction in the combined group was 42% (P <0.005).

The largest prevention trial was the U.S Diabetes Prevention Program (DPP) [2] which randomized 3234 middle-aged adults with IGT and fasting hyperglycaemia (fasting plasma glucose 5.3—<7.0 mmol/L). In addition to these two glucose-based parameters, participants were also selected to be high risk by virtue of other characteristics including age, obesity, and ethnic group. The mean BMI at baseline was 34 kg/m^2 and 45% of those randomized were of African, Hispanic, and Asian or American-Indian origin. Participants were originally randomized to four arms; an intensive lifestyle intervention, metformin (850 mg twice a day), troglitazone, or placebo. The troglitazone arm was discontinued in 1998 because of evidence of hepatotoxicity [3]. The intensive lifestyle intervention which was delivered by individual counselling sessions was focused on achieving a loss of 7% in body weight over 24 weeks through achieving 150 minutes per week of moderately intense physical activity and consumption of a low calorie, low fat diet. The incidence rate of diabetes, assessed by annual oral glucose tolerance tests and measurement of fasting plasma glucose at 6 monthly intervals, was reduced by 58% (95% confidence interval 48–66%) compared to the placebo group. The risk reduction in the metformin arm was less at 31% (95% CI 17–43%) and adherence to this treatment was high, with 72% of the participants in the metformin arm being compliant with therapy as defined by taking >80% of the prescribed medication.

At approximately the same time as the DPP, the Finnish Diabetes Prevention Study (DPS) reported the results of a randomized trial of intensive lifestyle intervention compared to placebo in middle-aged people (mean age 55 years) with IGT [4]. As with the DPP, participants in the DPS were largely overweight or obese (mean BMI 31 kg/m^2) and the intervention aimed to reduce body weight by at least 5% by reduction of consumption of fat, particularly saturated fat, increased consumption of fibre, and increased moderately intense physical activity to 30 minutes per day. As in the DPP, the DPS showed that the reduction in rate of progression to diabetes was 58% (95% confidence interval 30–70%). Subsequent similar intervention trials have demonstrated similar efficacy of lifestyle interventions in other populations including Japanese men with IGT [5] (n = 458) and Asian Indians with IGT [6] (n = 531). In a recent Cochrane systematic review, Hemmingsen *et al.* [7] meta-analysed these trials with a further seven studies [8–14] and reported a summary relative risk of progression to diabetes of 0.57 (95% CI 0.50–0.64).

Are the Benefits of Lifestyle Intervention the Same in All People at Risk?

The question of differential response to preventive interventions can only realistically be addressed in the DPP since only a large study has the power to examine heterogeneity between subgroups. In that study, there were no significant differences in relative risk reduction between subgroups defined on the basis of age, sex, and ethnicity, baseline level of obesity, or fasting glucose. A significantly larger relative risk reduction was, however, observed in those with lower levels of 2-hour glucose. While differences between subgroups in relative risk are interesting from an aetiological perspective, it is differences in absolute risk that matter from a clinical perspective. In general high-risk prevention approaches have their biggest impact when targeted on people at high absolute risk, because in those individuals, for a given relative risk reduction, the absolute risk reduction is greatest. The inverse of the absolute risk reduction is the number needed to treat, which is the parameter that makes sense to individual people who are offered a preventive intervention since it addresses the question 'How many people like me would need to have this intervention for one person to avoid getting diabetes over a particular time period?'. Thus logically, the individual interventions that have been shown to be efficacious in clinical trials ought to be first offered to people who have the highest absolute risk of developing type 2 diabetes. Whether they are offered to a wider set of people at moderate risk is an affordability question. However, the total number of people at moderate risk is considerably greater than those at high risk, since risk is roughly normally distributed, and since those people stand to gain less from a preventive intervention than those at high risk, it is very easy for a prevention programme to become overwhelmed by people who are not those who stand to benefit most. This is one of the challenges of translating trial results in diabetes prevention into real-world programmes.

Do the Benefits of Behavioural Interventions Persist Once the Active Intervention is Discontinued?

Several lifestyle diabetes prevention trials have examined the impact of interventions over the longer term and critically beyond the duration of the active phase of supported behaviour change. In the DPS the hazard ratio in the intervention group compared to placebo during a 9-year post-trial follow-up period was 0.67

(95% CI 0.48–0.95), suggesting that the benefits of lifestyle intervention persist beyond the point at which people were being provided with active lifestyle change interventions [15]. A similar follow-up in the Da Qing study also suggested persistence of the impact on lifestyle intervention with those in a combined lifestyle intervention groups having a 43% lower incidence over 20 years of follow-up [16]. In the follow-up study to the DPP [17], the intervention groups became more similar after the active period of the trial was ended and became unblinded, with the reduction in risk in the lifestyle intervention group compared to the original placebo group in the 10 years since randomization being 34% (95% CI 24–42%). Together these data suggest that there are persistent benefits of lifestyle intervention that go beyond the period of active intervention.

Are There Long-Term Health Benefits of Diabetes Prevention?

Although these landmark trials have shown that behavioural interventions and weight loss reduce progression to diabetes, it has proven very difficult to show that by reducing diabetes incidence there are long-term health benefits, since the trials are underpowered to examine these effects. The Da Qing study showed a 47% reduction in risk of severe retinopathy in the intervention group compared to control but no significant differences in risk of nephropathy or neuropathy [18]. In the DPP after 15 years of follow-up, there was no significant difference in risk of an aggregate microvascular outcome between treatment groups [19]. The only trial that has reported the impact on cardiovascular outcomes is the Da Qing trial which reported a relative risk for cardiovascular disease mortality of 0.59 (95% CI 0.36–0.96) in the intervention group over 23 years of follow-up [20]. From a patient perspective, it may be important not only to consider these long-term outcomes, but also the impact of diabetes preventive interventions on the more immediate outcome of quality of life. In the DPP, the lifestyle intervention improved patients' perception of the functional capability and quality of life compared to placebo, whereas there was no difference between these outcomes for the metformin and placebo groups [21].

Is Type 2 Diabetes Preventable by Pharmacological Intervention?

Beyond metformin, which was reported in the DPP, a whole series of trials have examined the effect of pharmacological therapy on diabetes prevention. Given that diabetes is defined by glucose levels, it is self-evident and effectively tautological to suggest that prescribing glucose lowering therapies to people with pre-diabetes should reduce progression to diabetes. Indeed, the level of relative risk reduction can be predicted by the degree of glucose reduction that an agent causes [22]. Thus thiazolidinediones including troglitazone (before it was withdrawn) [3], pioglitazone [23], rosiglitazone alone [24], or in combination with metformin [25], α-glucosidase inhibitors (acarbose [26, 27] and voglibose [28]), glucagon-like peptide-1 analogues (liraglutide [29]), and insulin secretagogues (Nateglinide [30]) have all been shown to reduce risk of progression to diabetes from IGT. Whether reduction of progression from IGT to diabetes is sufficient evidence to justify pharmacological intervention in people who don't consider themselves to be ill and who often have not even sought help, but may have been offered it as a consequence of being screened, is debatable. Unlike lifestyle intervention, the benefits for pharmacological diabetes prevention don't persist beyond the duration of active treatment, nor do patients being treated pharmacologically feel better or have better functional capability. The evaluation of the benefits of pharmacological diabetes prevention on hard clinical outcomes is the ultimate test of clinical utility of this approach but is a high level of evidence to aspire to and will inevitably require large and long-term clinical trials. In the absence of such evidence, it may be logical to restrict use of glucose lowering treatments in people without diabetes to population groups who have tried lifestyle interventions but without success. This stepwise approach using metformin, for example, has been shown to reduce diabetes incidence by a third in the D-CLIP trial [31].

Translating the Results of Trials into Real-World Interventions

As this chapter has summarized, the trial evidence underpinning individual-level diabetes prevention is extremely strong, showing that interventions are efficacious when assessed under the ideal circumstances of a randomized controlled trial. However, a major question is whether those interventions are as effective when used under the less ideal situations that prevail in the real world. An analysis by Ali and colleagues of 28 studies of real-world interventions in the United States suggested that these interventions achieved an average weight loss at 12 months of 4% of their baseline weight [32]. Since the DPP had established that weight loss was the most important determinant of reduction of diabetes incidence, with that incidence being reduced by 16% for every kilogramme of weight lost, this magnitude of average weight loss is not inconsiderable. Alongside the challenge of average effect size, is the issue of retention. In the MOVE-IT trial, a lifestyle intervention programme was effective in the group of individuals who completed the programme in an intense and sustained manner, but these participants represented less than 1% of the population for whom the intervention was designed [33], creating a massive gap between efficacy and effectiveness [34]. Many agencies in developed and developing countries are working to design programmes that can combine reach, effectiveness, and affordability since an additional challenge of the DPP approach was that it cost $1399 per participant in the first year of intervention. Cheap and scalable interventions may be possible but as yet, there is limited evidence of effectiveness of any programmes [35, 36]. Since the real clinical benefits of a preventive intervention lie not in the relative risk but in the absolute risk reduction, it may be preferable to focus such individual-level approaches to prevention on subgroups of the population who are at demonstrably high absolute risk because for those participants it is possible to get a large absolute risk reduction and a relatively low number needed to treat, the inverse of the absolute risk reduction.

Finding Complementary Strategies for Diabetes Prevention

Striving to deal with the problem of diabetes prevention only through an individual-level strategy is unlikely to result in a major impact on population-level diabetes incidence even though it could be of benefit to participants who are compliant and engaged. Thus an integrated diabetes prevention strategy also needs to take a whole population approach and to reflect the need to deal with the public health problem of type 2 diabetes through public health means. There is no single population-level intervention that is likely to have a major impact by itself on diabetes risk, nor is there likely to be strong evidence of effectiveness before policy action is implemented. Thus multiple policy level interventions have to be encouraged on the basis of observational evidence and modelled long-term health impact, with the effectiveness of approaches evaluated after the event through quasi-experimental approaches. This is the basis, for example, of the evidence that has driven the implementation of sugar-sweetened beverage taxes in multiple countries [37–39]. It is not within the remit of a medical textbook to provide a full discussion of the full range of public health interventions, but perhaps the critical thing for clinicians to appreciate is that the population-level approaches are a complement to more clinical approaches to prevention rather than alternative, that no single intervention is ever likely to have a large effect alone, but that considerable impact can be accumulated by intervening in multiple ways, and above all that the evidence base that underpins such interventions in inevitably different to that that is possible for individual-level interventions.

REFERENCES

1. Pan XR, Li GW, Hu YH, *et al.* Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care*, 1997; **20**: 537–44.
2. Knowler WC, Barrett-Connor E, Fowler SE, *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*, 2002; **346**: 393–403.
3. Knowler WC, Hamman RF, Edelstein SL, *et al.* Prevention of type 2 diabetes with troglitazone in the Diabetes Prevention Program. *Diabetes*, 2005; **54**: 1150–6.
4. Tuomilehto J, Lindstrom J, Eriksson JG, *et al.* Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*, 2001; **344**: 1343–50.
5. Kosaka K, Noda M, Kuzuya T. Prevention of type 2 diabetes by lifestyle intervention: a Japanese trial in IGT males. *Diabetes Res Clin Pract*, 2005; **67**: 152–62.
6. Ramachandran A, Snehalatha C, Mary S, *et al.* The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia*, 2006; **49**: 289–97.
7. Hemmingsen B, Gimenez-Perez G, Mauricio D, *et al.* Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. *Cochrane Database Syst Rev*, 2017; **12**: CD003054.
8. Penn L, White M, Lindstrom J, *et al.* Importance of weight loss maintenance and risk prediction in the prevention of type 2 diabetes: analysis of European Diabetes Prevention Study RCT. *PloS One*, 2013; **8**: e57143.
9. Hellgren MI, Jansson PA, Wedel H, Lindblad U. A lifestyle intervention in primary care prevents deterioration of insulin resistance in patients with impaired glucose tolerance: a randomised controlled trial. *Scand J Public Health*, 2016; **44**: 718–25.
10. Katula JA, Vitolins MZ, Rosenberger EL, *et al.* One-year results of a community-based translation of the Diabetes Prevention Program: Healthy-Living Partnerships to Prevent Diabetes (HELP PD) Project. *Diabetes Care*, 2011; **34**: 1451–7.
11. Sakane N, Sato J, Tsushita K, *et al.* Effect of baseline HbA1c level on the development of diabetes by lifestyle intervention in primary healthcare settings: insights from subanalysis of the Japan Diabetes Prevention Program. *BMJ Open Diabetes Res Care*, 2014; **23**: e000003.
12. Oldroyd JC, Unwin NC, White M, Mathers JC, Alberti KGMM. Randomised controlled trial evaluating lifestyle interventions in people with impaired glucose tolerance. *Diabetes Res Clin Pract*, 2006; **72**: 117–27.
13. Bhopal RS, Douglas A, Wallia S, *et al.* Effect of a lifestyle intervention on weight change in south Asian individuals in the UK at high risk of type 2 diabetes: a family-cluster randomised controlled trial. *Lancet Diabetes Endocrinol*, 2014; **2**: 218–27.
14. Mensink M, Blaak EE, Corpeleijn E, Saris WH, De Bruin TW, Feskens EJ. Lifestyle intervention according to general recommendation improves glucose tolerance. *Obes Res*, 2003; **11**: 1588–96.
15. Lindstrom J, Ilanne-Parikka P, Peltonen M, *et al.* Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet*, 2006; **368**: 1673–9.
16. Li G, Zhang P, Wang J, *et al.* The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet*, 2008; **371**: 1783–9.
17. DPP Research Group, Knowler WC, Fowler SE, *et al.* 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol*, 2009; **374**: 1677–86.
18. Gong Q, Gregg EW, Wang J, *et al.* Long-term effects of a randomised trial of a 6-year lifestyle intervention in impaired glucose tolerance on diabetes-related microvascular complications: the China Da Qing Diabetes Prevention Outcome Study. *Diabetologia*, 2011; **54**: 300–7.
19. DPP Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol*, 2015; **3**: 866–75.
20. Li G, Zhang P, Wang J, *et al.* Cardiovascular mortality, all-cause mortality, and diabetes incidence after lifestyle intervention for people with impaired glucose tolerance in the Da Qing Diabetes Prevention Study: a 23-year follow-up study. *Lancet Diabetes Endocrinol*, 2014; **2**: 474–80.
21. Florez H, Pan Q, Ackermann RT, *et al.* Impact of lifestyle intervention and metformin on health-related quality of life: the diabetes prevention program randomized trial. *J Gen Intern Med*, 2012; **27**: 1594–601.
22. Tuomilehto J, Wareham N. Glucose lowering and diabetes prevention: are they the same? *Lancet*, 2006; **368**: 1218–19.

23. DeFronzo RA, Tripathy D, Schwenke DC, *et al.* ACT NOW study: pioglitazone for diabetes prevention in impaired glucose tolerance. *N Engl J Med*, 2011; **364**: 1104–15.
24. DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) Trial Investigator Investigators, Gerstein HC, Yusuf S, *et al.* Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial. *Lancet*, 2006; **368**: 1096–105.
25. Zinman B, Harris SB, Neuman J, *et al.* Low-dose combination therapy with rosiglitazone and metformin to prevent type 2 diabetes mellitus (CANOE trial): a double-blind randomised controlled study. *Lancet*, 2010; **376**: 103–11.
26. Chiasson JL, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *Lancet*, 2002; **359**: 2072–7.
27. Holman RR, Coleman RL, Chan JCN, *et al.* Effects of acarbose on cardiovascular and diabetes outcomes in patients with coronary heart disease and impaired glucose tolerance (ACE): a randomised, double-blind, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 877–86.
28. Kawamori R, Tajima N, Iwamoto Y, *et al.* Voglibose for prevention of type 2 diabetes mellitus: a randomised, double-blind trial in Japanese individuals with impaired glucose tolerance. *Lancet*, 2009; **373**: 1607–14.
29. le Roux CW, Astrup A, Fujioka K, Greenway F, Lau DCW, Van Gaal L. 3 years of liraglutide versus placebo for type 2 diabetes risk reduction and weight management in individuals with prediabetes: a randomised, double-blind trial. *Lancet*, 2017; **389**: 1399–409.
30. Group. NS, Holman RR, Haffner SM, *et al*. Effect of nateglinide on the incidence of diabetes and cardiovascular events. *N Engl J Med*, 2010; **362**: 1463–76.
31. Weber MB, Ranjani H, Staimez LR, *et al.* The stepwise approach to diabetes prevention: results from the D-CLIP randomized controlled trial. *Diabetes Care*, 2016; **39**: 1760–7.
32. Ali MK, Echouffo-Tcheugui J, Williamson DF. How effective were lifestyle interventions in real-world settings that were modeled on the Diabetes Prevention Program? *Health Aff (Millwood)*, 2012; **31**: 67–75.
33. Jackson SL, Long Q, Rhee MK, *et al.* Weight loss and incidence of diabetes with the Veterans Health Administration MOVE! lifestyle change programme: an observational study. *Lancet Diabetes Endocrinol*, 2015; **3**: 173–80.
34. Wareham NJ. Mind the gap: efficacy versus effectiveness of lifestyle interventions to prevent diabetes. *Lancet Diabetes Endocrinol*, 2015; **3**: 160–1.
35. Ramachandran A, Snehalatha C, Ram J, *et al.* Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. *Lancet Diabetes Endocrinol*, 2013; **1**: 191–8.
36. Thankappan KR, Sathish T, Tapp RJ, *et al.* A peer-support lifestyle intervention for preventing type 2 diabetes in India: a cluster-randomized controlled trial of the Kerala Diabetes Prevention Program. *PLoS Med*, 2018; **15**: e1002575.
37. Imamura F, O'Connor L, Ye Z, *et al.* Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. *BMJ*, 2015; **351**: h3576.
38. Colchero MA, Popkin BM, Rivera JA, Ng SW. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. *BMJ*, 2016; **352**: h6704.
39. Sánchez-Romero LM, Penko J, Coxson PG, *et al.* Projected impact of Mexico's sugar-sweetened beverage tax policy on diabetes and cardiovascular disease: a modeling study. *PLoS Med*, 2016; **13**: e1002158.

15.8

Emerging Approaches to Restoring Euglycaemia in Diabetes

15.8.1 Regenerative Medicine for Diabetes

Michael G. White, Cara E. Ellis, and Timothy J. Kieffer

Introduction

Both type 1 (T1D) and type 2 (T2D) diabetes are characterized by insulin insufficiency and ultimately beta-cell loss. While T2D is multifactorial, T1D principally results from the selective loss of beta-cells involving autoimmune processes and requiring lifelong insulin replacement therapy to prevent clinical complications including ketoacidosis. Despite substantial improvements in insulin delivery systems and glucose monitoring, insulin therapy remains inadequate to maintain euglycaemia. Moreover, while intensive insulin therapy decreases average glycaemia and thereby reduces long-term complications of T1D, exogenous insulin delivery can also result in episodes of severe hypoglycaemia [1]. Therefore, there is an unmet need for better approaches to insulin replacement.

Beta-cell replacement through islet transplantation can provide restoration of more physiological glycaemic control, almost entirely preventing significant hypoglycaemia [2]. While islet transplantation offers unquestionably superior glycaemic control compared to insulin therapy, widespread implementation is severely limited by the availability of donor tissue. This has fuelled efforts to develop alternative beta-cell replenishment strategies including: (1) derivation of beta-cells from human embryonic stem cells (ESCs)/induced pluripotent stem cells (iPSCs); and (2) induction of endogenous beta-cell regeneration. Since both T1D and T2D are associated with beta-cell mass deficits, these therapeutic strategies could be effective for both forms of the disease.

Generating Beta-Cells from Pluripotent Stem Cells

ESCs and iPSCs are attractive sources of unlimited numbers of beta-cells to treat diabetes because they can self-renew indefinitely and differentiate into all cell types. To coax stem cells to beta-cells, differentiation protocols typically aim to mimic normal developmental processes, with the added challenge of using protocols that can be adapted for industrial scale. The pluripotency of stem cells makes it extremely challenging to achieve homogenous differentiation and undifferentiated cells can form teratomas. Several strategies are being explored to enhance the safety of potential stem-cell products, including cell containment devices and genetic modifications.

Beta-Cells from Embryonic Stem Cells

Building upon knowledge of pancreatic development in animal models and humans, protocols have been developed to differentiate human ESCs to pancreatic endoderm cells [3]. Efforts have focused on the production of cells expressing the key marker proteins PDX1 and NKX6.1, transcription factors that in combination specify pancreatic progenitor cells. Notably, cells at this stage are sufficiently committed to the pancreatic lineage and are fully capable of finishing differentiation and maturation post-implant to produce mixed populations of islet endocrine cells that can regulate glycaemia in rodents [3, 4]. Clinical trials implanting human ESC-derived pancreatic progenitor cells are now underway to assess safety and efficacy in patients with T1D (ClinicalTrials.gov Identifiers: NCT02239354, NCT03163511). Cell differentiation protocols continue to be refined and extended towards preimplant generation of mature beta-cells [5–8]. While not yet fully optimized, such cells can more rapidly reverse diabetes following implantation in mice compared to progenitor cells [5, 6].

Beta-Cells from Induced Pluripotent Stem Cells

The development of technologies to produce iPSCs through reprogramming of patient-derived cells such as fibroblasts [9] offers an alternative pluripotent cell type that addresses some of the ethical concerns related to the destruction of an embryo to generate

ESCs. Directed differentiation of iPSCs towards islet endocrine cells is being pursued using similar approaches as with ESCs [5, 6] but with additional challenges due to variability (including epigenetic) in the donor cells [10] and the need for protocol optimization for individual lines. Also, iPSCs may carry increased risk of teratoma formation compared to ESCs due to the chance of the reprogrammed donor cells acquiring a pro-oncogenic mutation that could be selected for during culture. However, earlier fears that the reprogramming process itself may introduce mutations appear to be unfounded [11]. Patient-derived cells would theoretically not be recognized as foreign by the immune system, but in T1D there is still beta-cell targeted autoimmunity to deal with, and generation of personalized tissues for millions of people with T1D will be prohibitively expensive for the foreseeable future. Alternatives to both ESCs and iPSC sources include creating HLA-matched banks from a subset of donors or creating universal donor cells through genetic editing [12]. A genetic 'safety switch' could also be added, allowing drug-mediated destruction of the implanted cells should the need arise, including in the case of off-target differentiation or teratoma formation.

Graft Encapsulation

Another challenge for the therapeutic application of stem-cell derived beta-cells is a safe and functional implant site. Graft retrievability and immune protection have been the goal stem cell-derived product development and thus clinical trials have not yet been conducted whereby differentiated stem cells are infused into the liver to mimic islet transplant using the Edmonton protocol. PEC-Encap™ is under development by ViaCyte [13, 14] and a phase I/IIa clinical trial began in 2014 (NCT02239354). This and other macroencapsulation devices are designed to contain the cells and physically prevent immune rejection while allowing nutrient exchange and insulin delivery via diffusion. However, results from other macroencapsulation devices implanted in humans suggest that a foreign body response to implanted devices contributing to lack of graft oxygenation may preclude optimal graft function [15]. As a substantial subset of patients with T1D are already on immunosuppression for other transplanted organs, use of a fully immunoprotective barrier may not always be necessary. ViaCyte is now testing PEC-Direct™, a device with holes that are designed to permit direct graft vascularization. A new multicentre phase I/IIa clinical trial began in 2017 (NCT03163511) that will offer important insights for the future direction of encapsulation technology.

An alternative to centimetre-scale immunoprotective devices is to encapsulate cell clusters, producing micrometre-to-millimetre-scale spheres which prevent physical interactions between the recipient immune system and the implanted cells. Moderately promising results from clinical trials (for example [16]) and advances in alginate formulas [17], the most common microencapsulation material, suggest that this technology could be a potential alternative to immunosuppression, although challenges with this approach include retrievability, achieving effective nutrient exchange, implant site, and foreign body responses.

Remaining Challenges for Stem Cell-Derived Beta-Cells

Regardless of the stem cell source, safety, and efficacy of the implanted cells are vital. Having completed maturation, whether before or after implantation, these cells must sense glucose and not only increase insulin secretion when blood glucose levels rise to minimize hyperglycaemia, but rapidly decrease insulin secretion when blood glucose levels drop to prevent hypoglycaemia. Challenges relating to immunosuppression or immune protection have not been completely solved, although many patients would exchange the heavy burden of multiple daily insulin injections for the side effects of immunosuppression if restoration of normal glucose homeostasis could be achieved. Recent encouraging progress suggests that these challenges, though difficult, are not insurmountable, and stem cell-derived beta-cells may be a functional cure for diabetes in the future.

Expanding Endogenous Beta-Cell Mass

Physiological Mechanisms Regulating Beta-Cell Mass

During embryogenesis beta-cells are formed from a population of pancreatic progenitor cells. While beta-cell mass in adults is determined, at least partially, by the size of this progenitor pool [18], it is also heavily influenced by the rapid beta-cell replication that occurs during infancy [19]. Physiological turnover of β-cells in both adult rodents and humans is comparatively slow; however, during pregnancy [20, 21] and excessive weight gain [22, 23], beta-cell mass expands to compensate for the increased metabolic demand. While the mechanisms governing such compensatory responses remain relatively poorly understood, and indeed may differ considerably between species, new beta-cell formation is thought to occur through both beta-cell replication [24] and differentiation from non-beta-cell sources—a process termed neogenesis [25, 26]. It may be possible to target replication and/or neogenesis to restore beta-cell mass and function in patients with diabetes.

Replication of Beta-Cells

Experiments using approaches that label beta-cells and their descendants provided compelling evidence for replication of pre-existing beta-cells as the major source of new beta-cells during adulthood in rodents [24, 27]. During pregnancy in rodents, beta-cell replication has been described as the main compensatory mechanism, predominately mediated by a prolactin-dependent inhibition of the tumour and replication suppressor, menin [20]. Similarly, increases in beta-cell mass observed in obese mice have been attributed to beta-cell replication [22]. Human beta-cells may have less capacity to replicate. While it has been shown that replication-mediated human beta-cell expansion occurs during infancy [19], there is a significant decline in this capacity with no evidence of beta-cell replication in donors aged over 30 [28], perhaps as a result of age-related increase in expression of p16INK4a, an inhibitor of cell–cycle associated activators, including cyclin D3 [29]. While increases in human beta-cell numbers during pregnancy have been reported, there was no evidence for these resulting from beta-cell replication [21]. Rather, increased numbers of small islets and duct associated insulin-positive cells suggest neogenesis as a major source of new β-cells [21], but this conclusion relies on circumstantial evidence from observations in autopsy material. Although it is increasingly clear that adaptive β-cell responses differ between species, evidence for beta-cell replication following human islet transplants into mice supports the hypothesis that human beta-cells have

the capacity, albeit perhaps limited, for proliferation [30, 31], and thus strategies to exploit this potential should be explored.

Inducing Beta-Cell Replication

While seeming an obvious therapeutic approach, inducing human beta-cell replication has proved difficult. Loss- or gain-of-function studies in rodents have yielded critical insights into the importance of cell–cycle molecules, particularly the D cyclins, which act predominantly at the G1/S transition [32]. However, species differences in the expression and localization of these proteins have hindered potential translation to humans [33], and there is evidence of potential beta-cell apoptosis following forced entry into the cell cycle [34]. However, high-throughput screens have identified small molecules that induce human beta-cell proliferation, including harmine [35], 5-iodotubercidin [36], and a series of aminopyrazines [37], all of which target a common pathway affecting DYRK1a and NFAT activity. Further, exendin-4, a clinically used GLP-1 receptor agonist, induced proliferation of juvenile human beta-cells [30]. While these studies demonstrate the feasibility of inducing beta-cell replication, the clinical utility of such compounds will require development of methods to ensure beta-cell specificity, particularly given the potential health implications of unchecked cell proliferation.

Beta-Cell Neogenesis/Transdifferentiation

Despite all pancreatic cell types arising from a common progenitor, evidence for a true adult pancreatic stem cell remains unconvincing, and the existence of facultative progenitors after birth continues to be debated. It is increasingly accepted, however, that pancreatic cells can display significant plasticity. Pancreatic cell types can switch phenotype, either directly (transdifferentiation) [26] or indirectly via a dedifferentiated intermediate [38]. Understanding of these events has been gained through rodent models of pancreatic regeneration including partial pancreatectomy [38], partial duct ligation [39], and targeted β-cell ablation [26]. Lineage tracing within these models reveals endocrine cell neogenesis, demonstrating several potential sources of new beta-cells, including pancreatic ducts and non-beta-cell endocrine cells.

Duct-Derived Neogenesis

Responsible for channelling digestive enzymes to the duodenum, the pancreatic ductal epithelium has long been proposed to harbour a post-natal progenitor cell population for islet cells, with several reports detailing observations of duct-mediated beta-cell neogenesis in response to pancreatic injury. Specifically, lineage tracing of ductal cells and their derivatives demonstrated that ductal cells can give rise to both new islet and acinar cells following partial duct ligation [40]. Characterization of rodent pancreas following 90% pancreatectomy identified several stages to this process, including initial ductal cell dedifferentiation and rapid proliferation, followed by activation of the endocrine progenitor marker neurogenin-3 (Ngn3) and subsequent formation of new beta-cells [38]. The gut hormone gastrin can enhance this series of neogenesis events and increase beta-cell numbers in a pancreatectomy model [41]. Lineage tracing following partial duct ligation also demonstrated the contribution of Ngn3-positive ductal cells to new beta-cell formation [39]. Recently it was reported that subpopulations of cells within the duct differ in their ability to differentiate towards endocrine cells [42], which may explain conflicting reports that refute duct-mediated beta-cell neogenesis in the adult [43, 44]. Analyses of human autopsy samples have provided circumstantial evidence for progenitor cells in the ductal regions, most notably a CD133-positive population [45]. Although this cell population has demonstrated some β-cell differentiation potential *in vitro* [46], overall the existence of such events in humans remains poorly understood and additional research is needed to better understand the translational potential.

Endocrine Cell Transdifferentiation

Forming functional micro-organs within the pancreas, islets consist of four main cell types: beta-cells (insulin), alpha-cells (glucagon), delta-cells (somatostatin), and PP cells (pancreatic polypeptide). While characterized by discrete gene expression profiles at full maturity, each of these endocrine cell types is derived from a common Ngn3-positive progenitor population. Diabetes reversal can occur in rodents following near total beta-cell ablation, mediated by alpha-to-beta-cell transdifferentiation [26]. This specific conversion appears to be age restricted, with recovery of diabetes in young mice (before puberty) resulting from delta-to-beta-cell conversion, without a contribution from alpha cells. The potential occurrence of similar plasticity events in human T1D has been described [47]. Further, it was demonstrated that transdifferentiation was dependent on activation of the G-protein coupled receptor, PAR2, and could be induced by the agonist 2-furoyl-LIGRLO-amide trifluoroacetate salt (2fLI) in both rodent models [48] and human islets [49]. However, there have yet been no lineage tracing studies to define cell transitions in human islets, and the continual autoimmunity associated with T1D must be considered. Despite these caveats, further work in this area may eventually allow the therapeutic induction of new beta-cells harnessing islet endocrine cell transdifferentiation potential.

Interplay Between Islet- and Duct-Cell Plasticity in Beta-Cell Neogenesis

Alpha-to-beta-cell conversion can be induced through genetic manipulation of key transcription factors. Ectopic expression of Pax4 and inhibition of Arx in alpha cells leads to their conversion to functional, neo-generated beta-cells that are capable of rescuing streptozotocin-induced diabetes in mice [50, 51]. Lineage tracing studies revealed activation of Ngn3-positive cells in the pancreatic ducts, which give rise to new α-cells and subsequently beta-cells following α-cell transdifferentiation [50, 51]. These studies may reconcile observations from the varying animal models of beta-cell regeneration, demonstrating a role for ductal and islet cell plasticity in beta-cell neogenesis in rodents. A study by the same group identified the neurotransmitter GABA as an inducer of the same series of events that lead to beta-cell replacement following Pax4 and Arx manipulation, including in human islets both *in vitro* and following transplantation into rodents [52]. In a parallel study, the antimalarial drug, artemisinin, was reported to induce alpha-to-beta-cell conversion in islets from multiple species through $GABA_A$ receptor signalling [53]. However, recent reports contest the impact of GABA and artemisinin on alpha-cell mediated beta-cell neogenesis [54, 55]. More investigation will be required to elucidate and potentially harness the putative plasticity of human islet cells.

Conclusions

Transforming lives of many patients with T1D, the success of islet transplantation has paved the way for development of novel cellular therapies for diabetes. Although in its infancy, substantial advancements have been made in our understanding and application of beta-cell replacement strategies, particularly in the generation and implant of differentiated hESCs. While significant challenges remain, including the development of technology to protect implanted cells from immune attack, emerging and ongoing clinical trials highlight the potential. Although a renewable source of *ex vivo* generated beta-cells is highly attractive, the ultimate regenerative goal in diabetes is to stimulate beta-cell formation *in situ*. The finding that the major pancreatic cell types appear to display a degree of plasticity which may be harnessed to generate new beta-cells has been met with considerable excitement. However, despite encouraging data, the existence of beta-cell proliferation and neogenesis/transdifferentiation in humans remains unclear, and thus whether strategies to enhance them are therapeutically relevant is unknown. Despite this, animal models have facilitated the identification of FDA-approved molecules that may target these potential regenerative pathways. This has led to an increasing number of clinical trials involving patients with T1D, from which the results are eagerly anticipated.

REFERENCES

1. Nathan DM. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*, 2014; **37**: 9–16.
2. Shapiro AMJ, Lakey JRT, Ryan EA, *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med*, 2000; **343**: 230–8.
3. Kroon E, Martinson LA, Kadoya K, *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nat Biotechnol*, 2008; **26**: 443–52.
4. Rezania A, Bruin JE, Riedel MJ, *et al.* Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. *Diabetes*, 2012; **61**: 2016–29.
5. Rezania A, Bruin JE, Arora P, *et al.* Reversal of diabetes with insulin-producing cells derived *in vitro* from human pluripotent stem cells. *Nat Biotechnol*, 2014; **32**: 1121–33.
6. Pagliuca FW, Millman JR, Gurtler M, *et al.* Generation of functional human pancreatic beta cells *in vitro*. *Cell*, 2014; **159**: 428–39.
7. Russ HA, Parent AV, Ringler JJ, *et al.* Controlled induction of human pancreatic progenitors produces functional beta-like cells *in vitro*. *EMBO J*, 2015; **34**: 1759–72.
8. Nostro MC, Sarangi F, Yang C, *et al.* Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines. *Stem Cell Reports*, 2015; **4**: 591–604.
9. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 2006; **126**: 663–76.
10. Liang G, Zhang Y. Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell*, 2013; **13**: 149–59.
11. Kwon EM, Connelly JP, Hansen NF, *et al.* iPSCs and fibroblast subclones from the same fibroblast population contain comparable levels of sequence variations. *Proc Natl Acad Sci U S A*, 2017; **114**: 1964–9.
12. Figueiredo C, Blasczyk R. A future with less HLA: potential clinical applications of HLA-universal cells. *Tissue Antigens*, 2015; **85**: 443–9.
13. Motte E, Szepessy E, Suenens K, *et al.* Composition and function of macroencapsulated human embryonic stem cell-derived implants: comparison with clinical human islet cell grafts. *Am J Physiol Endocrinol Metab*, 2014; **307**: E838–46.
14. Agulnick AD, Ambruzs DM, Moorman MA, *et al.* Insulin-producing endocrine cells differentiated *in vitro* from human embryonic stem cells function in macroencapsulation devices *in vivo*. *Stem Cells Transl Med*, 2015; **4**: 1214–22.
15. Ryan AJ, O'Neill HS, Duffy GP, O'Brien FJ. Advances in polymeric islet cell encapsulation technologies to limit the foreign body response and provide immunoisolation. *Curr Opin Pharmacol*, 2017; **36**: 66–71.
16. Basta G, Montanucci P, Luca G, *et al.* Long-term metabolic and immunological follow-up of nonimmunosuppressed patients with type 1 diabetes treated with microencapsulated islet allografts: four cases. *Diabetes Care*, 2011; **34**: 2406–9.
17. Vegas AJ, Veiseh O, Gurtler M, *et al.* Long-term glycemic control using polymer-encapsulated human stem cell-derived beta cells in immune-competent mice. *Nat Med*, 2016; **22**: 306–11.
18. Stanger BZ, Tanaka AJ, Melton DA. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature*, 2007; **445**: 886.
19. Meier JJ, Butler AE, Saisho Y, *et al.* β-cell replication is the primary mechanism subserving the postnatal expansion of β-cell mass in humans. *Diabetes*, 2008; **57**: 1584–94.
20. Karnik SK, Chen H, McLean GW, *et al.* Menin controls growth of pancreatic β-cells in pregnant mice and promotes gestational diabetes mellitus. *Science*, 2007; **318**: 806–9.
21. Butler AE, Cao-Minh L, Galasso R, *et al.* Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. *Diabetologia*, 2010; **53**: 2167–76.
22. Cox AR, Lam CJ, Rankin MM, *et al.* Extreme obesity induces massive beta cell expansion in mice through self-renewal and does not alter the beta cell lineage. *Diabetologia*, 2016; **59**: 1231–41.
23. Saisho Y, Butler AE, Manesso E, Elashoff D, Rizza RA, Butler PC. β-Cell mass and turnover in humans: effects of obesity and aging. *Diabetes Care*, 2013; **36**: 111–17.
24. Dor Y, Brown J, Martinez OI, Melton DA. Adult pancreatic β-cells are formed by self-duplication rather than stem-cell differentiation. *Nature*, 2004; **429**: 41.
25. Inada A, Nienaber C, Katsuta H, *et al.* Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth. *Proc Natl Acad Sci USA*, 2008; **105**: 19915–19.
26. Thorel F, Népote V, Avril I, *et al.* Conversion of adult pancreatic α-cells to β-cells after extreme β-cell loss. *Nature*, 2010; **464**: 1149–54.
27. Teta M, Rankin MM, Long SY, Stein GM, Kushner JA. Growth and regeneration of adult β cells does not involve specialized progenitors. *Dev Cell*, 2007; **12**: 817–26.
28. Perl S, Kushner JA, Buchholz BA, *et al.* Significant human β-cell turnover is limited to the first three decades of life as determined by *in vivo* thymidine analog incorporation and radiocarbon dating. *J Clin Endocrinol Metab*, 2010; **95**: E234–E9.

29. Krishnamurthy J, Ramsey MR, Ligon KL, *et al.* p16INK4a induces an age-dependent decline in islet regenerative potential. *Nature*, 2006; **443**: 453.
30. Dai C, Hang Y, Shostak A, Poffenberger G, Hart N, Prasad N, *et al.* Age-dependent human β cell proliferation induced by glucagon-like peptide 1 and calcineurin signaling. *J Clin Invest*, 2017; **127**: 3835–44.
31. Tyrberg B, Ustinov J, Otonkoski T, Andersson A. Stimulated endocrine cell proliferation and differentiation in transplanted human pancreatic islets. *Diabetes*, 2001; **50**: 301.
32. Georgia S, Hinault C, Kawamori D, *et al.* Cyclin D2 is essential for the compensatory β-cell hyperplastic response to insulin resistance in rodents. *Diabetes*, 2010; **59**: 987–96.
33. Fiaschi-Taesch NM, Kleinberger JW, Salim FG, *et al.* Cytoplasmic-nuclear trafficking of G1/S cell cycle molecules and adult human β-cell replication: a revised model of human β-cell G1/S control. *Diabetes*, 2013; **62**: 2460–70.
34. Rieck S, Zhang J, Li Z, *et al.* Overexpression of hepatocyte nuclear factor-4α initiates cell cycle entry, but is not sufficient to promote β-cell expansion in human islets. *Mol Endocrinol*, 2012; **26**: 1590–602.
35. Wang P, Alvarez-Perez J-C, Felsenfeld DP, *et al.* Induction of human pancreatic beta cell replication by inhibitors of dual specificity tyrosine regulated kinase. *Nat Med*, 2015; **21**: 383–8.
36. Dirice E, Walpita D, Vetere A, *et al.* Inhibition of DYRK1A stimulates human β-cell proliferation. *Diabetes*, 2016; **65**: 1660–71.
37. Shen W, Taylor B, Jin Q, *et al.* Inhibition of DYRK1A and GSK3B induces human β-cell proliferation. *Nat Commun*, 2015; **6**: 8372.
38. Li W-C, Rukstalis JM, Nishimura W, *et al.* Activation of pancreatic-duct-derived progenitor cells during pancreas regeneration in adult rats. *J Cell Sci*, 2010; **123**: 2792–802.
39. Van de Casteele M, Leuckx G, Baeyens L, *et al.* Neurogenin 3+ cells contribute to β-cell neogenesis and proliferation in injured adult mouse pancreas. *Cell Death Dis*, 2013; **4**: e523.
40. Inada A, Nienaber C, Katsuta H, *et al.* Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth. *Proc Natl Acad Sci U S A*, 2008; **105**: 19915–19.
41. Téllez N, Montanya E. Gastrin induces ductal cell dedifferentiation and β-cell neogenesis after 90% pancreatectomy. *J Endocrinol*, 2014; **223**: 67–78.
42. Rezanejad H, Ouziel-Yahalom L, Keyzer CA, *et al.* Heterogeneity of SOX9 and HNF1β in pancreatic ducts is dynamic. *Stem Cell Reports*, 2018; **10**: 725–38.
43. Kopp JL, Dubois CL, Schaffer AE, *et al.* Sox9(+) ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. *Development*, 2011; **138**: 653–65.
44. Solar M, Cardalda C, Houbracken I, *et al.* Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Dev Cell*, 2009; **17**: 849–60.
45. Gomez DL, O'Driscoll M, Sheets TP, *et al.* Neurogenin 3 expressing cells in the human exocrine pancreas have the capacity for endocrine cell fate. *PLoS One*, 2015; **10**: e0133862.
46. Lee J, Sugiyama T, Liu Y, *et al.* Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. *eLife*, 2013; **2**: e00940.
47. Piran R, Lee SH, Li CR, Charbono A, Bradley LM, Levine F. Pharmacological induction of pancreatic islet cell transdifferentiation: relevance to type I diabetes. *Cell Death Dis*, 2014; **5**: e1357.
48. Piran R, Lee S-H, Kuss P, *et al.* PAR2 regulates regeneration, transdifferentiation, and death. *Cell Death Dis*, 2016; **7**: e2452.
49. Lee S-H, Hao E, Scharp D, Levine F. Insulin acts as a repressive factor to inhibit the ability of PAR2 to induce islet cell transdifferentiation. *Islets*, 2018; May 3: 1–24.
50. Collombat P, Xu X, Ravassard P, *et al.* The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into α- and subsequently β-cells. *Cell*, 2009; **138**: 449–62.
51. Courtney M, Gjernes E, Druelle N, *et al.* The inactivation of Arx in pancreatic α-cells triggers their neogenesis and conversion into functional β-like cells. *PLoS Genet*, 2013; **9**: e1003934.
52. Ben-Othman N, Vieira A, Courtney M, *et al.* Long-term GABA administration induces alpha cell-mediated beta-like cell neogenesis. *Cell*, 2017; **168**: 73–85.e11.
53. Li J, Casteels T, Frogne T, *et al.* Artemisinins target GABA(A) receptor signaling and impair α cell identity. *Cell*, 2017; **168**: 86–100.e15.
54. Ackermann AM, Moss NG, Kaestner KH. GABA and artesunate do not induce pancreatic α-to-β cell transdifferentiation *in vivo*. *Cell Metab*, 2018; **28**: 787–92.e3.
55. van der Meulen T, Lee S, Noordeloos E, *et al.* Artemether does not turn α cells into β cells. *Cell Metab*, 2018; **27**: 218–25.e4.

15.8.2 "Closed Loop" Insulin Delivery

Roman Hovorka and Charlotte Boughton

Introduction

Following the Diabetes Control and Complications Trial, focus on intensive insulin therapy has led to increased uptake of insulin pump therapy (continuous subcutaneous insulin infusion, CSII) as it is recognized that CSII delivery can achieve improved glucose control, reduce the risk of hypoglycaemia and improve quality of life in people with T1D (see Chapter 15.5.3, 'Insulins and Insulin Delivery Devices') [1].

Advances in diabetes technology have led to the development of continuous glucose monitoring (CGM) systems which measure real-time interstitial glucose concentration (see Chapter 15.5.2, 'Glucose Monitoring and Sensing'). There has been steady improvement in the reliability and accuracy of CGM systems with the best having an accuracy of below 10% mean absolute relative difference [2]. There remains a known issue with lag between blood and interstitial

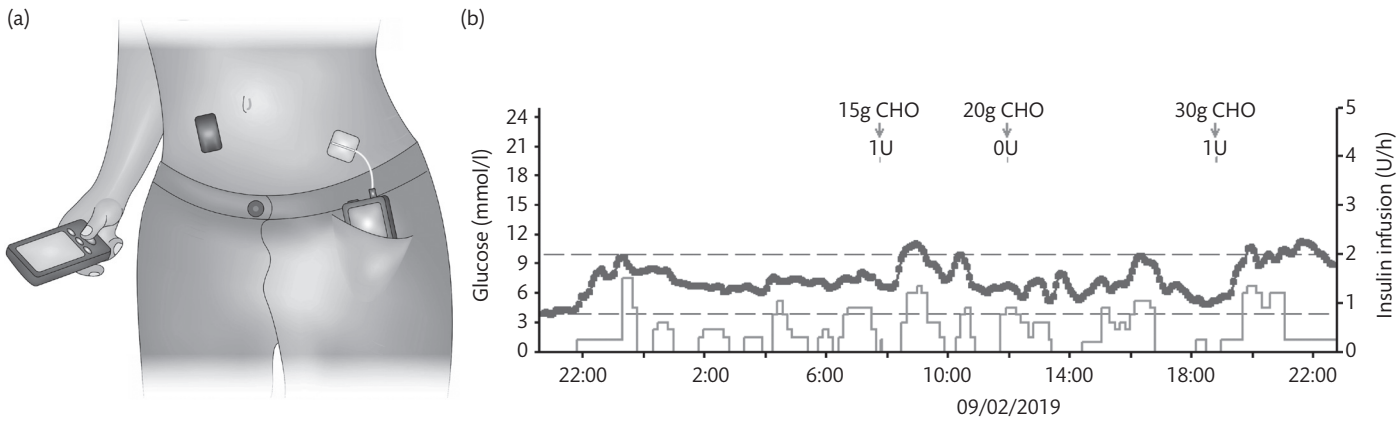


Figure 15.8.2.1 (a) A sensor (black rectangle) transmits information about interstitial glucose levels to a mobile-phone-sized controller (red box located in the hand), which runs a control algorithm and interacts with the user. An insulin pump (blue box) delivers a rapid-acting insulin analogue subcutaneously. Insulin delivery is modulated in real-time by the control algorithm. The communication among system components is wireless. (b) Glucose levels and insulin delivery profiles during a hybrid single-hormone closed-loop study. The bold red line shows the continuous subcutaneous glucose trace. Control algorithm-directed insulin infusion rates during closed-loop are denoted by the thick blue line. Vertical arrows indicate meals and snacks (orange arrows) and blue vertical lines illustrate insulin boluses administered by the user at mealtimes. Horizontal dashed red lines indicate the target glucose range. *For a colour version of this figure, please see colour plate section.*

glucose levels; however, use of CGM is associated with improvements in glucose control and reductions in hypoglycaemia [3].

Insulin pumps can be used in conjunction with real-time CGM, allowing users to manually modify the insulin infusion rate according to CGM values (sensor augmented pump therapy, SAP). Introduction of a low glucose suspend (LGS) feature allowed for automatic insulin infusion suspension when a pre-programmed CGM threshold value is reached or when sensor glucose is predicted to cross the pre-programmed CGM threshold value (predictive LGS). These features are effective in reducing the frequency and duration of hypoglycaemia [3].

Closed-loop insulin delivery (the artificial pancreas) is more sophisticated and combines insulin pump and CGM with *a control algorithm* to deliver insulin in a glucose-responsive manner (single-hormone closed-loop system) by automatically adjusting the insulin infusion rate based on sensor glucose levels. Glucagon or other hormones can also be delivered in a similar glucose-responsive fashion within dual-hormone closed-loop systems. The control algorithm may be embedded in the pump or reside in a separate device (e.g. a smartphone (Figure 15.8.2.1) and these components communicate wirelessly).

Closed-Loop Algorithms

The key component of a closed-loop system is the control algorithm, which directs insulin delivery according to real-time sensor glucose levels while accounting for inter- and intrasubject variability, inherent sensor, and insulin delivery errors as well as kinetic delays. Various algorithms have been developed with the most commonly used being the model-predictive control approach, the proportional–integral–derivative control, and fuzzy logic [4].

Model-predictive control (MPC) algorithms employ an individualized mathematical model of glucose regulation to predict glucose excursions based on input such as subcutaneous insulin. Insulin delivery is calculated by minimizing the difference between model-predicted glucose-concentrations and the target glucose levels over a pre-specified prediction time horizon.

Proportional–integral–derivative (PID) control algorithms continuously adjust the insulin infusion rate by assessing glucose excursions from three perspectives, the deviation from the target glucose (proportional component), the area-under-the-curve between the measured and target glucose level (integral component), and the rate of change in the measured glucose levels (derivative component).

The fuzzy logic control approach modulates insulin delivery on the basis of approximate rules to express the empirical knowledge of diabetes practitioners.

Meta-analysis data suggests that PID algorithms may be associated with less time in target compared with MPC and fuzzy logic algorithms, although the analysis of subgroup differences was not significant [5]. Direct comparison of the performance of comparable PID and MPC algorithms in a cross-over study support this finding [6].

Most control algorithms include additional safety modules to constrain insulin delivery, limiting the amount of insulin on board or the maximum rate of insulin delivery, or suspending insulin delivery when glucose levels are low or decreasing rapidly.

User specific parameters (body weight, insulin requirements) are usually required for initialization. Adaptation of the algorithm to changes in physiological parameters with real-time adjustment of closed-loop control parameters is essential for optimal performance.

Types of Closed-Loop Systems

Hybrid and Fully Closed-Loop Systems

Hybrid closed-loop systems require meal announcements while fully closed-loop systems are entirely automated offering the

advantage of reduced user intervention. There have been relatively few, small in-patient studies of fully automated closed-loop systems in people with T1D. Postprandial hyperglycaemic excursions and late postprandial hypoglycaemia are significant challenges with this approach and glucose control is significantly compromised compared to hybrid closed-loop at present. Ultra-rapid insulin analogues are needed to enable safe and efficacious fully closed-loop systems.

Single and Dual-Hormone Closed-Loop Systems

Impaired glucagon response to hypoglycaemia is characteristic of T1D. Addition of glucagon to closed-loop systems (dual-hormone) is an attractive option to counteract the relatively long duration of action of subcutaneous insulin but also increases system complexity and cost. The potential role of glucagon is further reduction in the risk of hypoglycaemia while optimizing insulin delivery, or to compensate with higher glucagon doses, buffering 'aggressive' insulin delivery.

Addition of glucagon to closed-loop systems is currently limited by the instability of available glucagon formulations at room temperature. These require reconstitution prior to use and exchange of the glucagon depot every 24 hours. Novel glucagon analogues are under development but the full pharmacokinetic and safety profile is to be established. Long-term low-dose glucagon administration has the potential for pleiotropic effects and long-term follow-up safety studies are necessary. The main side-effects of high-dose rescue glucagon include nausea, vomiting, and headaches, but adverse effects of smaller doses of glucagon used in trials of dual-hormone systems were limited to mild gastrointestinal symptoms [7].

Adjuncts

One of the main challenges with closed-loop insulin delivery is postprandial glucose control. Studies have demonstrated short-term efficacy of adjuncts such as synthetic amylin (pramlintide), glucagon-like peptide-1 agonists, and dipeptidyl peptidase-4 inhibitors in combination with closed-loop insulin delivery to reduce postprandial hyperglycaemia, although long-term benefits have not been evaluated.

Clinical Evidence of Closed-Loop Systems

Clinical studies have progressed from those undertaken in highly supervised clinical research facilities to studies lasting several months in unsupervised, free-living conditions. Clinical studies have shown that hybrid closed-loop systems, as compared with CSII or SAP, improve glycaemic control and reduce hypoglycaemia. Closed-loop insulin delivery has demonstrated efficacy and safety in children, adolescents, adults, and pregnant women with T1D and in hospitalized patients with hyperglycaemia.

Closed-Loop in Adults and Children with Type 1 Diabetes

Two recent meta-analyses have evaluated the efficacy and safety of closed-loop therapy in non-pregnant outpatients with type 1 diabetes [5, 8]. The larger of these included 41 studies (1042 participants).

Time in Range

Meta-analysis data from 32 studies has shown that closed-loop systems increase time spent in near normoglycaemia (3.9–10.0 mmol/L) throughout 24 hours by 9.6% compared with controls (140 additional minutes/24 h period). The beneficial effect of closed-loop systems is more evident on the percentage of time spent in near normoglycaemia overnight, which is 15% greater than controls [8]. Increased time in range with closed-loop systems is due to reduced time in hyperglycaemia (>10 mmol/L) compared to controls.

While dual-hormone systems have been associated with a greater improvement in time in target range compared with single-hormone systems, almost all dual-hormone system studies have been relatively short and have been compared to pump therapy alone, whereas almost all single-hormone systems have used SAP as a comparator. Studies directly comparing single- and dual-hormone closed-loop systems have been under supervised conditions and of short duration but observed no difference in the time spent in target glucose range over 24 hours [9].

Hypoglycaemia

Closed-loop systems reduce time spent in hypoglycaemia (<3.9 mmol/L) by 1.5% (approximately 20 minutes/24 h period) compared to control [8]. The incidence of severe hypoglycaemia is very low in clinical studies comparing closed-loop with control so there is insufficient evidence to determine any benefit of closed-loop insulin delivery on severe hypoglycaemia. However, use of closed-loop systems was associated with a decrease in low blood glucose index overnight compared to controls, a measure of the risk of severe hypoglycaemia.

Clinical studies contrasting dual-hormone with single-hormone systems report less time spent in hypoglycaemia and fewer hypoglycaemic events with dual-hormone systems [7]. The reduction in hypoglycaemia was observed with comparable mean glucose levels.

Mean Glucose and HbA1c

Compared to control, use of closed-loop systems has a favourable effect on the mean sensor glucose level over 24h, which is reduced by 0.48 mmol/L in meta-analysis data [8]. These findings are consistent with a reduction in HbA1c of 0.26% observed with closed-loop systems compared with control in studies with a duration per intervention of more than eight weeks.

Glycaemic Variability

In almost all closed-loop studies, glucose variability, measured as the standard deviation and coefficient of variation of the sensor glucose level between days, is lower with closed-loop systems than with controls [10].

Insulin Requirements

Individual studies have conflicting results with respect to the effect of closed-loop systems on daily insulin dose. Meta-analysis data suggests there is no difference between closed-loop and control systems in the mean daily insulin requirement [5, 8].

Closed Loop in Pregnancy

Closed-loop insulin delivery has been shown to be safe and effective in women with T1D during pregnancy. In a randomized, cross-over

study in pregnant women with T1D, overnight closed-loop therapy for 4 weeks resulted in increased time in target range (3.5–7.8 mmol/L) by 15.2 percentage points (74.7% vs. 59.5%) compared to SAP therapy [11]. There was no difference between closed-loop and SAP therapy in the percentage of time in hypoglycaemia. During the continuation phase (up to 14.6 additional weeks including antenatal hospitalizations, labour, and delivery), of day-and-night closed loop until delivery, glucose levels were in the target range 68.7% of the time. The impact of closed-loop insulin delivery on perinatal outcomes has not been established.

Closed Loop in Inpatients

In-patient hyperglycaemia is a marker of poor prognosis associated with increased morbidity, mortality, length of stay, and healthcare costs. In-patient studies have focused on fully closed-loop insulin delivery (without meal announcements) both in the intensive care unit (ICU) and the general wards.

Most closed-loop insulin delivery studies in the ICU have utilized intravenous insulin delivery and venous glucose sampling. These studies have demonstrated efficacy and safety of automated insulin delivery although the requirement for large blood sample volumes has resulted in limited uptake [12]. Use of fully automated closed-loop therapy with continuous subcutaneous glucose monitoring led to increased time in target glucose range (6.0–8.0 mmol/L) compared with control intravenous variable rate insulin (54.3% vs. 18.5%) over a 48 h period in ICU without increasing the risk of hypoglycaemia [13].

In inpatients with type 2 diabetes in the general wards, closed-loop insulin delivery significantly increased time spent within target glucose range (5.6–10.0 mmol/L) compared to conventional subcutaneous insulin delivery (65.8% vs. 41.5%) without any increase in the risk of hypoglycaemia [14].

Psychosocial Aspects of Closed Loop

Long-term usage of closed-loop technology is likely to be influenced by user expectations and experience. Expectations include stable glucose regulation, reduced need for glucose self-monitoring, relief of daily concerns, and timesaving, however, trust in automated insulin delivery is perceived as a potential barrier [15].

Reported benefits, in addition to improved glycaemic control, include reduced fear of hypoglycaemia, reassurance for users and family members, reduced anxiety, improved sleep, confidence, 'time off' from diabetes demands, greater freedom to engage in activity, excitement, and empowerment. Reported burdens included technical difficulties, alarm intrusiveness and interrupted sleep, increased time spent thinking about diabetes, size, and appearance of the equipment, limitations on exercise, and perceptions of deskilling and data obsession.

OpenAPS (Open Artificial Pancreas System) and LOOP lay communities suggest high parental and end-user interests in the technology [16]. The OpenAPS movement includes individuals building alternative non-commercial do-it-yourself (DIY) closed-loop systems from commercially available insulin pumps (sometimes out of warranty), CGM devices, and an open source algorithm.

Individual users have varied levels of engagement (a 'hands-off' approach with minimal input to high levels of engagement and adjustment of the control algorithm to their individual needs) and differing preferences with potential implications for effective long-term usage, associated glycaemic outcomes, and training needs. Clinician attitudes to closed-loop systems have yet to be considered.

Figure 15.8.2.2 Hybrid closed-loop system comprising fourth generation Enlite 3 glucose sensor, MiniMed® 670G insulin pump, with an embedded proportional–integral–derivative algorithm with insulin feedback (Medtronic, Northridge, CA).
Reproduced with permission from Medtronic.

Commercially Available Closed-Loop Systems

The MiniMed® 670G pump (Medtronic, Northridge, CA), approved in 2016 by the U.S. Food and Drug Administration (FDA) for use by people with T1D over 14 years of age, is a single-hormone hybrid closed-loop system with the control algorithm integrated in the insulin pump (Figure 15.8.2.2). The pump basal rate is automated based on a proportional–integral–derivative algorithm with insulin on board feedback. Clinical trial evaluation to assess safety was non-randomized and lacked a control arm therefore statements pertaining to its efficacy are limited [17]. The closed-loop system was used day and night for 3 months by 94 adults and 30 adolescents. No episodes of severe hypoglycaemia or ketoacidosis were observed. Several other hybrid closed-loop systems are being developed for commercial use. Post-marketing studies are ongoing.

Outlook

Future closed-loop systems will utilize improved system components. Sensors with improved accuracy, reduced or factory calibrations, and increased wear time will improve both performance and acceptability. Faster insulin analogues may further improve closed-loop system performance. At present, meal announcement with accurate carbohydrate counting is recommended to achieve optimal glucose control. Algorithms capable of integrating multiple signals may more accurately reflect the rapidly changing insulin

requirements of a person with diabetes than CGM alone. Device interoperability, cloud upload, and remote monitoring capabilities may further increase appeal and acceptability.

Assessment of cost-effectiveness for reimbursement by healthcare systems, and development of adequate infrastructure to support closed-loop users is essential for wide adoption of this technology.

REFERENCES

1. Pickup JC. Insulin-pump therapy for type 1 diabetes mellitus. *N Engl J Med*, 2012; **366**: 1616–24.
2. Rodbard D. Characterizing accuracy and precision of glucose sensors and meters. *J Diabetes Sci Technol*, 2014. **8**: 980–5.
3. Rodbard D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. *Diabetes Technol Ther*, 2017; **19**(Suppl 3): S25–37.
4. Hovorka R. Closed-loop insulin delivery: from bench to clinical practice. *Nat Rev Endocrinol*, 2011; **7**: 385–95.
5. Weisman A, Bai JW, Cardinez M, Kramer CK, Perkins BA. Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. *Lancet Diabetes Endocrinol*, 2017; **5**: 501–12.
6. Pinsker JE, Lee JB, Dassau E, *et al.* Randomized crossover comparison of personalized MPC and PID control algorithms for the artificial pancreas. *Diabetes Care*, 2016; **39**: 1135–42.
7. Haidar A, Messier V, Legault L, *et al.* Outpatient 60-hour day-and-night glucose control with dual-hormone artificial pancreas, single-hormone artificial pancreas, or sensor-augmented pump therapy in adults with type 1 diabetes: an open-label, randomised, crossover, controlled trial. *Diabetes Obes Metab*, 2017; **19**: 713–20.
8. Bekiari E, Kitsios K, Thabit H, *et al.* Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis. *BMJ*, 2018; **361**: k1310.
9. Haidar A, Legault L, Messier V, *et al.* Comparison of dual-hormone artificial pancreas, single-hormone artificial pancreas, and conventional insulin pump therapy for glycaemic control in patients with type 1 diabetes: an open-label randomised controlled crossover trial. *Lancet Diabetes Endocrinol*, 2015; **3**: 17–26.
10. Thabit H, Tauschmann M, Allen JM, *et al.* Home use of an artificial beta cell in type 1 diabetes. *N Engl J Med*, 2015; **373**: 2129–40.
11. Stewart ZA, Wilinska ME, Hartnell S, *et al.* Closed-loop insulin delivery during pregnancy in women with type 1 diabetes. *N Engl J Med*, 2016; **375**: 644–54.
12. Okabayashi T, Shima Y, Sumiyoshi T, *et al.* Intensive versus intermediate glucose control in surgical intensive care unit patients. *Diabetes Care*, 2014; **37**: 1516–24.
13. Leelarathna L, English SW, Thabit H, *et al.* Feasibility of fully automated closed-loop glucose control using continuous subcutaneous glucose measurements in critical illness: a randomized controlled trial. *Crit Care*, 2013; **17**: R159.
14. Bally L, Thabit H, Hartnell S, *et al.* Closed-loop insulin delivery for glycemic control in noncritical care. *N Engl J Med,* 2018; **379:** 547–56.
15. Farrington C. Psychosocial impacts of hybrid closed-loop systems in the management of diabetes: a review. *Diabet Med*, 2018; **35**: 436–49.
16. Lewis D, Leibrand S. Real-world use of open source artificial pancreas systems. *J Diabetes Sci Technol*, 2016; **10**: 1411.
17. Bergenstal RM, Garg S, Weinzimer SA, *et al.* Safety of a hybrid closed-loop insulin delivery system in patients with type 1 diabetes. *JAMA*, 2016; **316**: 1407–8.

15.9

Emergency and Hospital Management of Diabetes

15.9.1 Hyperglycaemic Emergencies

Ketan Dhatariya

Diabetic Ketoacidosis (DKA): Introduction

Diabetic ketoacidosis (DKA) is a frequently encountered acute metabolic emergency most frequently occurring in people with type 1 diabetes mellitus, but increasingly in those with type 2 as well. DKA is a condition, that if not managed correctly in a timely manner it is still associated with an appreciable mortality. The Centers for Disease Control and Prevention in the United States reported that in 2014 there were 168 000 admissions to hospital for this condition, a prevalence of 7.7 per 1000 people with diabetes [1]. These data are broadly unchanged from the late 1970s and early 1980s [2, 3]. In 2014, admissions for DKA represented ~2.3% of all admissions where diabetes was listed as a discharge diagnosis. In 1993, the mortality rate for DKA was reported as 3.9% in the United Kingdom [4], but data from the United States suggests that more recently this has fallen from ~8% to less than 1% [5, 6]. Among children and young adults the long-term standardized mortality ratio for children and young people has also come down from 9.3 to 5.6 between 1965 and 1979 [7, 8]. However, in this age group, DKA remains the most common cause of death with type 1 diabetes.

Pathophysiology

DKA occurs due to absolute insulin deficiency or when the concentrations of counterregulatory hormones (glucagon, catecholamines, cortisol and, to a lesser extent, growth hormone) are high. In these circumstances there is almost no insulin mediated cellular glucose uptake meaning that cells need an alternative energy substrate.

Insulin deficiency results in the generation of free fatty acids due to the increase in hormone-sensitive lipase activity on triglycerides [9]. The fatty acids are β-oxidized to form acetyl coenzyme A (CoA), which usually enters the tricarboxylic acid (TCA) cycle. This excess acetyl CoA entering the TCA cycle overwhelms the enzyme systems, and is converted into ketone bodies in the liver [10]. These ketones provide the alternative energy substrate needed for cellular metabolism to continue, mainly in the form of β-hydroxybutyrate (a hydroxyl acid) and acetoacetate at an approximate ratio of 10:1 [11].

Definition of DKA

To date, there is no recognized 'standardized' definition of DKA. While it may sound obvious, to make a diagnosis of DKA, all three of these criteria must be met. In the absence of large scale clinical studies to help determine the best management of DKA in the United Kingdom, the Joint British Diabetes Societies have produced a widely used national guideline on the management of DKA, used by over 90% of all UK hospitals [12, 13]. The diagnostic criteria they use states that the individual should either have an admission glucose of >11.1 mmol/L (200 mg/dl) or be known to have diabetes—irrespective of the admission glucose (the 'D'); that the plasma β-hydroxybutyrate concentration be ≥3.0 mmol/L, or there be significant ketonuria (>2+ on standard urine ketone sticks) (the 'K'); and the plasma bicarbonate concentration be <15.0 mmol/L *and/or* a venous pH of <7.3 (the 'A') [12].

Probably the most widely adopted guidelines used elsewhere for the management of DKA come from the 2009 American Diabetes Association position statement on hyperglycaemia emergencies [14]. That document uses a slightly different set of diagnostic criteria for DKA: a glucose concentration of >13.9 mmol/L (250 mg/dl); that ketones must be present (in urine or in the blood); and a metabolic acidosis must be present, with a pH of <7.30 (measured on arterial or venous blood) *and* a serum bicarbonate of ≤18.0 mmol/L.

Conflict of interest: Prof Dhatariya has been the lead or coauthor on several of the guidelines on the management of patients with diabetes in hospital produced by the Joint British Diabetes Societies for Inpatient Care, available at https://abcd.care/joint-british-diabetes-societies-jbds-inpatient-care-group
External Funding: None. Prof Dhatariya is a full-time employee of the National Health Service

Recently there has been a call to amend the ADA guideline in the face of emerging concerns about the sodium-glucose cotransporter 2 (SGLT-2) inhibitors and their association with 'euglycaemic DKA', where glucose concentrations may be less than 10.0 mmol/L (180 mg/dl) [15]. This condition may occur in between 3.5% and 15% of all cases of DKA, but because the ADA guideline states that DKA should not be diagnosed unless the glucose is >13.9 mmol/L (250 mg/dl), there is potential room for error in diagnosis, and subsequent management [16, 17].

The ADA also grades DKA into levels of severity depending on pH, bicarbonate concentrations, anion gap and level of alertness [14]. This is important because the available evidence suggests that in patients with mild to moderate DKA, the use of weight based, low dose subcutaneous or intramuscular insulin given every one or two hours offers an acceptable alternative to intravenous insulin [18]. In addition, the use of modern technology can guide whether an individual needs to be admitted for treatment or not [19]. Thus, there is an opportunity to reduce the costs associated with treating DKA, especially where the use of intravenous insulin necessitates mandatory admission to high cost environments such as the intensive care unit.

Incidence and Prevalence of DKA

A recent systematic literature review reported a range for incidence of 0–56 per 1000 person years, and a prevalence of 0–128 per 1000 people [20]. In North America, the prevalence of 168 000 admissions per year equated to an estimated incidence of DKA in at between 1% and 5% per year [1, 3, 21]. Most of these are in people with type 1 diabetes, but in some areas with high ethnic diversity or strong family history, up to 50% of cases occur in those with type 2 diabetes [22, 23]. In other parts of the world the rates vary, with quoted rates of 10 per 100 patient years in the Western Pacific [24]. It is much lower in Northern Europe [25, 26], with rates of 4.8 per 100 patient years being reported in the United Kingdom—an incidence of 3.8% per year [27, 28]. The treatment of DKA is expensive, with costs per admission in the US estimated at ~$17 500 [14], and £2064 in the United Kingdom [29].

Precipitants

Recent data has suggested that DKA is the first presentation for type 1 diabetes in only 3–6% of cases [16, 30], far below the previously quoted figure of ~30% [3, 31–35]. While in many cases no clear cause is found, the other main causes of DKA are infections, poor adherence to treatment, and drug and alcohol related events [16, 36, 37]. As mentioned, DKA—in particular euglycaemic DKA—has been increasingly reported with the use of SGLT-2 inhibitors [38] (and in SGLT1/2 inhibitors [39]) in those with type 1 and type 2 diabetes. More recently, a systematic review has found that lower socio-economic status is also associated with an increased risk of DKA [40].

DKA is still reported as the largest cause of death in children and young adults with type 1 diabetes under 30 years old [8]. However, the mortality rate in adults and in children may be dropping [41]. In 2016 a national survey of DKA management in the United Kingdom reported no deaths due to DKA in over 280 acute admissions [16], compared to previously reported rates of between 1.7% and 3.9% in the United Kingdom and China [4, 42, 43].

Box 15.9.1.1 Factors that should prompt an urgent assessment for care in a Level 2/High Dependency Unit environment in DKA

- Severe ketoacidosis (blood ketones >6.0 mmol/L; serum bicarbonate <5.0 mmol/L; venous or arterial pH <7.1; anion gap >16)
- Hypokalaemia (<3.5 mmol/L)
- Impaired consciousness (e.g. abnormal Glasgow coma score (GCS) or AVPU (Alert, Voice, Pain, Unresponsive) score)
- Oxygen saturation <92% breathing air (if baseline respiratory function normal)
- Haemodynamic compromise (systolic BP <90 mmHg and/or heart rate >100 or <60 beats per minute)

The Management of DKA

The principles of managing DKA centre primarily on replacing the fluid deficit. In addition to enhancing organ perfusion, fluid replacement helps reduce counterregulatory hormones production. Once a fluid replacement regimen has been started, then a weight based, fixed rate intravenous insulin infusion (FRIII) should be administered to suppress lipolysis and ketone production. 50 units of human soluble insulin should be made up to 50 ml with 0.9% sodium chloride solution and infused at a starting dose of 0.1 units/kg/hour. As mentioned previously, the American Diabetes Association guideline suggests grading the severity of DKA allowing for the use of regular subcutaneous insulin should hospital admission need to be avoided [14].

A bolus dose of subcutaneous or intravenous insulin should only be administered if there is a delay in setting up an FRIII. At the same time as the fluid and insulin are being administered, precipitating factors should be sought and treated appropriately. The presence of any of the factors listed in Box 15.9.1.1 should prompt an urgent assessment for care in a level 2/high dependency unit environment

The aim of treatment should be to restore euvolaemia, euglycaemia, and a normal pH without inducing iatrogenic hypokalaemia or hypoglycaemia.

Table 15.9.1.1 gives an example of a fluid replacement regimen.

Hyperkalaemia is common at presentation in DKA, and, if severe, should be treated as an emergency. However, high serum potassium often masks a total body deficiency potassium, and because potassium is driven into cells by treatment with insulin and fluids, serum

Table 15.9.1.1 An example of a fluid replacement regimen in DKA

Fluid	Volume
0.9% sodium chloride 1 L	1000 ml over 1st hour
0.9% sodium chloride 1 L with potassium chloride	1000 ml over next 2 hours
0.9% sodium chloride 1 L with potassium chloride	1000 ml over next 2 hours
0.9% sodium chloride 1 L with potassium chloride	1000 ml over next 2 hours
0.9% sodium chloride 1 L with potassium chloride	1000 ml over next 4 hours
0.9% sodium chloride 1 L with potassium chloride	1000 ml over next 4 hours

Re-assessment of cardiovascular status at 12 hours is mandatory, further fluid may be required

Table 15.9.1.2 An example of a potassium replacement regimen in DKA

Potassium concentration in first 24 hours (mmol/L)	Potassium replacement in mmol/L of infusion solution
Over 5.5	Nil
3.5–5.5	40 mmol/L
Below 3.5	Senior review because additional potassium needs to be given

potassium may fall sharply. Hypokalaemia occurs frequently, and serum potassium should be checked regularly and replaced proactively to avoid hypokalaemia [16].

Table 15.9.1.2 gives an example of a potassium replacement regimen.

Regular clinical and biochemical assessment (ideally hourly initially) should be mandatory to ensure continued improvement after the initiation of therapy. As discussed next, the urine ketone estimation should be avoided, while the use of venous blood gases is strongly encouraged.

Resolution of DKA is defined as a venous pH >7.3, serum bicarbonate >15.0 mmol/L, and blood ketone <0.6 mmol/L.

The early involvement of the diabetes specialist team in the longer-term management of the person with DKA is crucial.

Urinary vs. Plasma Ketones

In addition to the call for changes in the glucose concentration used to define DKA, there has been an increasing focus on the potentially inappropriate use of urinary ketones. There are many reasons for this. Firstly, the predominant ketone in urine is acetoacetate [44]. The urinary concentrations of acetoacetate do not reflect the concentration of plasma β-hydroxybutyrate [11]. Furthermore, as the DKA resolves, β-hydroxybutyrate is converted into acetoacetic acid which is then renally excreted. This leads to the false impression that the DKA is taking longer to resolve than is the case [45]. With the increasing availability of handheld, point-of-care ketone meters, there is call to stop using urinary ketones and to use blood ketones to help diagnose, and guide the management of DKA [17, 45, 46]. This move towards the DKA being 'ketone-centric' rather than 'glucose-centric' is increasingly important given the recognition of the previously mentioned 'euglycaemic DKA'. While it is increasingly being reported with the use of SGLT-2 inhibitors [38] (and in SGLT1/2 inhibitors [39]) euglycaemic DKA can also occur in pregnancy, starvation, and alcoholism [47–49].

Venous vs. Arterial Blood Gases

There is increasing evidence to suggest that the difference between venous and arterial pH is very minor—in the region of 0.02–0.15 pH units, and that the difference between arterial and venous bicarbonate is ~1.88 mmol/L [50–53]. These differences are highly unlikely to change the diagnosis or alter the management of DKA. Given the potential iatrogenic harm that can be caused by an arterial blood gas sample being taken, in the United Kingdom it is felt that it is no longer necessary to use arterial blood to measure acid-base status [12]. Venous blood can be used in portable and fixed blood gas analysers and therefore venous measurements (pH, bicarbonate, chloride, and potassium) are easily obtained in most admitting units. The UK guideline suggests that an arterial line should only be inserted if its use will influence management [12].

Continuation of Subcutaneous Long-Acting Basal Insulin

The use of long-acting basal insulin is now widespread. These can be either human or analogue insulin. They should be continued in patients already taking them, or started at a dose of 0.25 units/kg subcutaneously once daily in those for whom DKA is the first presentation of type 1 diabetes. Continuation of these subcutaneous injections during the initial management of DKA provides background insulin when the intravenous insulin is eventually discontinued, and so avoids any rebound hyperglycaemia [54]. This should avoid extending the length of stay [55]. This only applies to long-acting insulin and does not obviate the need to give short acting insulin before discontinuing the intravenous insulin infusion.

Future Developments

It is likely that there will be several changes in the management of DKA over the next few years. The unexpected finding of DKA with the use of SGLT-2 inhibitors has highlighted some issues with the diagnostic criteria used to diagnose DKA. The DKA event in these trials were not adjudicated, and on closer inspection of the data, it is clear that the definitions used to diagnose the conditions in these trials varied across the world, and even within countries at different institutions. However, despite there being no clear definitions used, licencing authorities have placed warnings on the use of these drugs [56, 57] Because of this, there has recently been a call to ensure that the same diagnostic criteria are used, especially with clinical trials, to ensure that data used to make such decisions are robust [58].

In addition, the awareness that euglycaemic DKA occurs with some frequency, may change the diagnostic algorithm used by the ADA to allow a glucose of lower than 13.9 mmol/L (250 mg/dl) before a diagnosis can be made.

The recent move towards the management being more 'ketone-centric', because of the increasing availability of point-of-care ketone testing equipment, means that there is likely to be a greater call for their use—in diagnosing the disorder and also in guiding its management [45]. And while there remain differences in the management of DKA between the two most commonly used guidelines—those of the United States and the United Kingdom [59]. These differences are likely to become smaller as the evidence base increases and there are changes in the paradigm in how DKA presents [17]. This is important because of the data suggesting that fragmentation of—and by extension—differences in care—is associated with recurrent DKA as well as increased mortality [60].

Finally, recent evidence from a large national survey showed that the current management paradigm used in the United Kingdom was associated with over 27% of patients developing hypokalaemia (a potassium concentration of <4.0 mmol/L) and 55% developing mild (i.e. self-treated) hypoglycaemia (blood glucose <72 mg/dl [4.0 mmol/L]) [16]. While these biochemical abnormalities were not associated with any harm, what remains unclear is whether these were due to staff not following the guidelines, or whether the rate of insulin infusion needs to be reduced to reduce the likelihood of them occurring. Further work needs to be done to determine the best intravenous insulin infusion rates.

Hyperosmolar Hyperglycaemic State (HHS)

Introduction

HHS was initially described as a separate entity causing diabetic coma by Dreschfield [61] and Von Frerichs [62]. Dreschfield and others described three forms of 'diabetic coma'—one of which was the gradual onset of loss of consciousness that occurred most frequently in adults over the age of 40, who were overweight without the characteristic acetone breath or acetone in the urine found in other forms of diabetic coma [61, 63–65]. It was with the advent of post-war technology that the true extent of the metabolic derangement associated with HHS became apparent [66, 67].

Definition and Pathophysiology

While there is no formally accepted definition of HHS, it is a condition characterized by severe dehydration, hyperglycaemia in the absence of ketoacidosis, and hyperosmolarity. The diagnosis is made when there is hypovolaemia, a glucose of >30.0 mmol/L (>33.3 mmol/L in the United States), with a serum osmolality >320 mOsm/kg—in the absence of significant acidosis or ketonaemia [14, 68]. Osmolality should be calculated as or [2 × measured Na (mmol/L) + glucose (mmol/L) + [urea (mmol/L)] or [2 × measured Na (mEq/L) + glucose (mg/dl)/18 + blood urea nitrogen (mg/dl)]/2.8]. HHS is believed to result from the presence of sufficient insulin to suppress hepatic ketone production, but not to suppress gluconeogenesis. The slow onset of progressive hyperglycaemia causes an osmotic diuresis, which itself leads to often severe dehydration. Because the renal losses of water are in excess of sodium and potassium the osmolality rises. Fluid losses are estimated at between 100 and 220 ml/kg. Despite this, patients with HHS may not look dehydrated, because of the redistribution of body water due to blood hypertonicity and preservation of intravascular volume.

Some patients may present with a mixed picture, with a metabolic acidosis and a raised anion gap—this may be due in part to ketonaemia, but also to high lactate concentrations.

Precipitants

The most common precipitant of HHS is infection. Other common causes are omission of glucose lowering agents, other comorbidities, such as acute coronary syndrome, or the addition of high dose glucocorticoid therapy. In only a minority of cases is HHS the first presentation of diabetes [36].

Morbidity and Mortality

HHS tends to occur in individuals who are older than those with DKA. As such, they often have several other comorbidities at the time of presentation. In addition, they are at greater risk of developing other complications, with atherosclerosis, thrombosis, and foot ulceration posing particular risks. Patients with HHS should be offered heel protection—particularly to those with neuropathy, peripheral vascular disease, or lower limb deformity. The feet should be examined daily.

It is likely that the reported mortality of HHS of 5–16% (about 10-fold higher than for DKA) is partly because of these pre-existing comorbidities [69, 70].

Management of HHS

As with DKA, the principles of managing HHS centre primarily on replacing the fluid deficit and lowering glucose concentrations. The two main guidelines used in many parts of the world—the ADA guidelines [14] and the UK guidelines from the Joint British Diabetes Societies for Inpatient Care [68], both advocate the use of 0.9% sodium chloride solution given at an appropriate rate (depending on the presence of cardiac, pulmonary or renal comorbidities) to correct the dehydration and reduce the serum osmolality. The UK guideline advocated that approximately 50% of the estimated fluid loss should be replaced within the first 12 hours, and the remainder in the following 12 hours. There may be an initial rise in serum sodium concentrations but because the glucose and urea concentrations would fall faster than the sodium rise, the overall calculated serum osmolality should fall—ideally by 3–8 mOsmol/kg/hr. The use of 0.45% sodium chloride solution should be avoided initially—except when the osmolality fails to fall despite adequate fluid replacement. Despite the absence of hard data, a decline in serum sodium of 0.5 mmol/L per hour has been recommended for hypernatraemic dehydration. Potassium should be replaced as for DKA (see Table 15.9.1.2).

The only time that insulin should be started before glucose concentrations stop falling in HHS is if, at the time of presentation or anytime thereafter, ketonaemia (>1.0 mmol/L) develops. This would indicate a relative hypoinsulinaemia. Most patients with HHS are insulin sensitive, and the addition of insulin prior to the fluids alone allowing glucose concentrations to fall may lead to a precipitous drop in glucose, causing a rapid lowering of osmolality, potentially leading to central pontine myelinolysis. In addition, the rapid fluid shifts associated with the movement of glucose intracellularly out of the intravascular space, may lead to cardiovascular collapse if the rate of fluid replacement is insufficient, or not possible due to other comorbidities.

Careful and regular monitoring of fluid balance and biochemistry—ideally hourly for the first 6 hours—should be instituted, and the intravenous fluid should be continued until the osmolality stops falling. In addition, once the glucose concentration has stopped falling with fluid administration alone should a weight based, FRIII be initiated at a rate of 0.05 unit/Kg/hour. The aim should be to reduce the glucose concentration at a rate of 5.0 mmol/L/hour, but to maintain it initially between 10.0 and 15.0 mmol/L (according to the UK guideline). As the condition of the patient improves, with the addition of further intravenous insulin, or the use of oral hypoglycaemic agents, the glucose concentration may be reduced further as necessary.

While infusing intravenous fluids, clinical assessment should look for evidence of a precipitating cause, e.g. inflammation, sepsis, or myocardial infarction, or a change in medication, e.g. recently starting steroids. Once fluid resuscitation has been initiated, then the underlying cause should be sought and treated appropriately, while at the same time ensuring the prevention of complications such as arterial or venous thrombosis, cerebral oedema, central pontine myelinolysis, and foot ulceration. Foot examination is mandatory at the time of admission and daily afterwards. Obtunded or uncooperative patients should be assumed to be at particularly high risk.

Box 15.9.1.2 Factors that should prompt an urgent assessment for care in a level 2/high dependency unit environment in HHS

- Difficulty in obtaining intravenous access
- Serum osmolality >350 mOsm/kg and/or serum sodium >160 mmol/L
- Venous/arterial pH <7.1
- Severely deranged serum potassium (<3.5 mmol/L or >6.0 mmol/L)
- Impaired consciousness (e.g. GCS <12 or abnormal AVPU score)
- Oxygen saturation <92% breathing air (if baseline respiratory function normal)
- Haemodynamic compromise (systolic BP <90 mmHg and/or heart rate >100 or <60 beats per minute)
- Hypothermia
- Acute or serious comorbidity (e.g. MI, CCF, or CVA)
- Urine output <0.5 ml/kg/h or other evidence of acute kidney injury

The electrolyte and osmolality abnormalities may take up to 72 hours to normalise.

The presence of any of the factors listed in Box 15.9.1.2 should prompt an urgent assessment for care in a level 2/high dependency unit environment.

Education

For DKA and HHS the key is prevention—educating patients before they develop any complications. For those with insulin treated diabetes, revision of 'sick day rules' should form a part of the annual review.

For those who develop DKA or HHS, assessment by the inpatient diabetes specialist team should be mandatory, to educate the patient and provide a contact for patients and carers for when the patient leaves hospital.

REFERENCES

1. Centers for Disease Control and Prevention. National diabetes statistics report, 2017. Estimates of diabetes and its burden in the United States. Available at: https://dev.diabetes.org/sites/default/files/2019-06/cdc-statistics-report-2017.pdf (accessed: 24 April 2020).
2. Johnson DD, Palumbo PJ, Chu CP. Diabetic ketoacidosis in a community-based population. *Mayo Clin Proc*, 1980; **55**: 83–8.
3. Faich GA, Fishbein HA, Ellis SE. The epidemiology of diabetic acidosis: a population-based study. *Am J Epidemiol*, 1983; **117**: 551–8.
4. Basu A, Close CF, Jenkins D, *et al.* Persisting mortality in diabetic ketoacidosis. *Diabetic Med*, 1993; **10**: 282–4.
5. Lin SF, Lin JD, Huang YY. Diabetic ketoacidosis: comparisons of patient characteristics, clinical presentations and outcomes today and 20 years ago. *Chang Gung Med J*, 2005; **28**: 24–30.
6. Wang J, Williams DE, Narayan KM, Geiss LS. Declining death rates from hyperglycemic crisis among adults with diabetes, U.S., 1985–2002. *Diabetes Care*, 2006; **29**: 2018–22.
7. Secrest AM, Becker DJ, Kelsey SF, Laporte RE, Orchard TJ. All-cause mortality trends in a large population-based cohort with long-standing childhood-onset type 1 diabetes. *Diabetes Care*, 2010; **33**: 2573–9.
8. Wasag DR, Gregory JW, Dayan C, Harvey JN. Excess all-cause mortality before age 30 in childhood onset type 1 diabetes: data from the Brecon Group Cohort in Wales. *Arch Dis Child*, 2018; **103**: 44–8.
9. Foster DW, McGarry JD. The metabolic derangements and treatment of diabetic ketoacidosis. *N Eng J Med*, 1983; **309**: 159–69.
10. McGarry JD, Woeltje KF, Kuwajima M, Foster DW. Regulation of ketogenesis and the renaissance of carnitine palmitoyltransferase. *Diabetes Metab Rev*, 1989; **5**: 271–84.
11. Stentz FB, Umpierrez GE, Cuervo R, Kitabchi AE. Proinflammatory cytokines, markers of cardiovascular risks, oxidative stress, and lipid peroxidation in patients with hyperglycemic crises. *Diabetes*, 2004; **53**: 2079–86.
12. Savage MW, Dhatariya KK, Kilvert A, *et al.* Joint British Diabetes Societies guideline for the management of diabetic ketoacidosis. *Diabetic Med*, 2011; **28**: 508–15.
13. NHS Digital. National Diabetes Inpatient Audit (NaDIA)—2017. Available at: https://digital.nhs.uk/catalogue/PUB30248 (accessed: 18 March 2018).
14. Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN. Hyperglycemic crises in adult patients with diabetes. *Diabetes Care*, 2009; **32**: 1335–43.
15. Munro JF, Campbell IW, McCuish AC, Duncan JP. Euglycaemic diabetic ketoacidosis. *Br Med J*, 1973; **2**: 578–80.
16. Dhatariya KK, Nunney I, Higgins K, Sampson MJ, Iceton G. A national survey of the management of diabetic ketoacidosis in the UK in 2014. *Diabetic Med*, 2016; **33**: 252–60.
17. Dhatariya KK, Umpierrez GE. Guidelines for management of diabetic ketoacidosis: time to revise? *Lancet Diabetes Endocrinol*, 2017; **5**: 321–3.
18. Vincent M, Nobécourt E. Treatment of diabetic ketoacidosis with subcutaneous insulin lispro: a review of the current evidence from clinical studies. *Diabetes Metab*, 2016; **39**: 299–305.
19. Ullal J, McFarland R, Bachand M, Aloi J. Use of a computer-based insulin infusion algorithm to treat diabetic ketoacidosis in the emergency department. *Diabetes Technol Ther*, 2016; **18**: 100–3.
20. Farsani SF, Brodovicz K, Soleymanlou N, *et al.* Incidence and prevalence of diabetic ketoacidosis (DKA) among adults with type 1 diabetes mellitus (T1D): a systematic literature review. *BMJ Open*, 2017; **7**: e016587.
21. Ginde AA, Pelletier AJ, Camargo CA. National study of U.S. emergency department visits with diabetic ketoacidosis, 1993–2003. *Diabetes Care*, 2006; **29**: 2117–19.
22. Wang ZH, Kihl-Selstam E, Eriksson JW. Ketoacidosis occurs in both type 1 and type 2 diabetes—a population-based study from Northern Sweden. *Diabetic Med*, 2008; **25**: 867–70.
23. Umpierrez GE, Smiley D, Kitabchi AE. Narrative review: ketosis-prone type 2 diabetes mellitus. *Ann Intern Med*, 2006; **144**: 350–7.
24. Craig ME, Jones TW, Silink M, Ping YJ. Diabetes care, glycemic control, and complications in children with type 1 diabetes from Asia and the Western Pacific Region. *J Diabetes Complications*, 2016; **21**: 280–7.
25. Henriksen OM, Roder ME, Prahl JB, Svendsen OL. Diabetic ketoacidosis in Denmark. *Diabetes Res Clin Pract*, 2007; **76**: 51–6.
26. Rosilio M, Cotton JB, Wieliczko MC, *et al.* Factors associated with glycemic control: a cross-sectional nationwide study in 2,579 French children with type 1 diabetes. *Diabetes Care*, 1998; **21**: 1146–53.

27. Health and Social Care Information Centre. National Diabetes Audit 2012–2013. Report 2: complications and mortality. 2015. Available at: http://content.digital.nhs.uk/catalogue/PUB16496/nati-diab-audi-12-13-rep2.pdf (accessed: 18 March 2018).
28. Karges B, Rosenbauer J, Holterhus PM, *et al.* Hospital admission for diabetic ketoacidosis or severe hypoglycemia in 31,330 young patients with type 1 diabetes. *Eur J Endocrinol*, 2015; **173**: 341–50.
29. Dhatariya KK, Skedgel C, Fordham R. The cost of treating diabetic ketoacidosis in the UK: a national survey of hospital resource use. *Diabetic Med*, 2017; **34**: 1361–6.
30. Edge JA, Nunney I, Dhatariya KK. Diabetic ketoacidosis in an adolescent and young adult population in the UK in 2014: a national survey comparison of management in paediatric and adult settings. *Diabetic Med*, 2016; **33**: 1352–9.
31. Ellemann K, Soerensen JN, Pedersen L, Edsberg B, Andersen OO. Epidemiology and treatment of diabetic ketoacidosis in a community population. *Diabetes Care*, 1984; **7**: 528–32.
32. Bui TP, Werther GA, Cameron FJ. Trends in diabetic ketoacidosis in childhood and adolescence: a 15-yr experience. *Pediatr Diabetes*, 2002; **3**: 82–8.
33. Dunger DB, Sperling MA, Acerini CL, *et al.* European Society for Paediatric Endocrinology/Lawson Wilkins Pediatric Endocrine Society Consensus Statement on diabetic ketoacidosis in children and adolescents. *Pediatrics*, 2004; **113**: e133–40.
34. Rewers A, Dong F, Slover RH, Klingensmith G, Rewers M. Incidence of diabetic ketoacidosis at diagnosis of type 1 diabetes in Colorado youth, 1998–2012. *JAMA*, 2015; **313**: 1570–2.
35. Usher-Smith JA, Thompson M, Ercole A, Walter FM. Variation between countries in the frequency of diabetic ketoacidosis at first presentation of type 1 diabetes in children: a systematic review. *Diabetologia*, 2012; **55**: 2878–94.
36. Dhatariya KK, Glaser NS, Codner E, Umpierrez GE. Diabetic ketoacidosis. *Nat Rev Dis Primers,* 2020; **6:** 40.
37. Randall L, Begovic J, Hudson M, *et al.* Recurrent diabetic ketoacidosis in inner-city minority patients. *Diabetes Care*, 2011; **34**: 1891–6.
38. Peters AL, Buschur EO, Buse JB, *et al.* Euglycemic diabetic ketoacidosis: a potential complication of treatment with sodium-glucose cotransporter 2 inhibition. *Diabetes Care*, 2015; **38**: 1687–93.
39. Garg SK, Henry RR, Banks P, *et al.* Effects of sotagliflozin added to insulin in patients with type 1 diabetes. *N Eng J Med*, 2017; **377**: 2337–48.
40. Lindner LM, Rathmann W, Rosenbauer J. Inequalities in glycaemic control, hypoglycaemia and diabetic ketoacidosis according to socio-economic status and area-level deprivation in type 1 diabetes mellitus: a systematic review. *Diabetic Med*, 2017; **35**: 12–32.
41. Poovazhagi V. Risk factors for mortality in children with diabetic ketoacidosis from developing countries. *World J Diabetes*, 2014; **5**: 932–8.
42. Wright J, Ruck K, Rabbitts R, *et al.* Diabetic ketoacidosis (DKA) in Birmingham, UK, 2000–2009: an evaluation of risk factors for recurrence and mortality. *Br J Diabetes Vasc Dis*, 2009; **9**: 278–82.
43. Xu Y, Bai J, Wang G, *et al.* Clinical profile of diabetic ketoacidosis in tertiary hospitals in China: a multicentre, clinic-based study. *Diabetic Med*, 2016; **33**: 261–8.
44. Rothera AC. Note on the sodium nitro-prusside reaction for acetone. *J Physiol*, 1908; **37**: 491–4.
45. Dhatariya K. Blood ketones—measurement, interpretation, limitations and utility in the management of diabetic ketoacidosis. *Rev Diabet Stud*, 2017;**13**: 217–25.
46. Dhatariya K, Savage M, Kelly T, *et al.* Joint British Diabetes Societies Inpatient Care Group. The management of diabetic ketoacidosis in adults. Second edition. Update: September 2013. Available at: https://abcd.care/joint-british-diabetes-societies-jbds-inpatient-care-group (accessed: 18 March 2018).
47. Guo R-X, Yang L-Z, Li L-X, Zhao X-P. Diabetic ketoacidosis in pregnancy tends to occur at lower blood glucose levels: case–control study and a case report of euglycemic diabetic ketoacidosis in pregnancy. *J Obstet Gynaecol Res*, 2008; **34**: 324–30.
48. McGuire LC, Cruickshank AM, Munro PT. Alcoholic ketoacidosis. *Emerg Med J*, 2006; **23**: 417–20.
49. Palmer BF, Clegg DJ. Electrolyte disturbances in patients with chronic alcohol-use disorder. *N Eng J Med*, 2017; **377**: 1368–77.
50. Ma OJ, Rush MD, Godfrey MM, Gaddis G. Arterial blood gas results rarely influence emergency physician management of patients with suspected diabetic ketoacidosis. *Acad Emerg Med*, 2003; **10**: 836–41.
51. Gokel Y, Paydas S, Koseoglu Z, Alparslan N, Seydaoglu G. Comparison of blood gas and acid-base measurements in arterial and venous blood samples in patients with uremic acidosis and diabetic ketoacidosis in the emergency room. *Am J Nephrol*, 2000; **20**: 319–23.
52. Kelly A-M, Mcalpine R, Kyle E. Venous pH can safely replace arterial pH in the initial evaluation of patients in the emergency department. *Emerg Med J*, 2001; **18**: 340–2.
53. Herrington WG, Nye HJ, Hammersley MS, Watkinson PJ. Are arterial and venous samples clinically equivalent for the estimation of pH, serum bicarbonate and potassium concentration in critically ill patients? *Diabetic Med*, 2012; **29**: 32–5.
54. Hsia E, Seggelke S, Gibbs J, *et al.* Subcutaneous administration of glargine to diabetic patients receiving insulin infusion prevents rebound hyperglycemia. *J Clin Endocrinol Metab*, 2012; **97**: 3132–7.
55. Martin C, McKinney ZJ, Hoody D, Fish L. Diabetic ketoacidosis critical care pathway implimentation: incorporation into EMR significantly decreases length of stay. *Endocr Pract*, 2016; **22**: 673–8.
56. US Food and Drug Administration. FDA Drug Safety Communication: FDA warns that SGLT2 inhibitors for diabetes may result in a serious condition of too much acid in the blood. Available at: https://www.fda.gov/Drugs/DrugSafety/ucm475463.htm. 2015 (accessed: 18 March 2018).
57. Medicines and Healthcare products Regulatory Agency. Drug safety update: SGLT2 inhibitors (canagliflozin, dapagliflozin, empagliflozin): risk of diabetic ketoacidosis. 2015. Available at: https://www.gov.uk/drug-safety-update/sglt2-inhibitors-canagliflozin-dapagliflozin-empagliflozin-risk-of-diabetic-ketoacidosis (accessed: 18 March 2018).
58. Dhatariya KK. Why the definitions used to diagnose diabetic ketoacidosis should be standardised. *Diabetes Res Clin Pract*, 2017; **135**: 227–8.
59. Dhatariya KK, Vellanki P. Treatment of diabetic ketoacidosis (DKA)/hyperglycemic hyperosmolar state (HHS): novel advances in the management of hyperglycemic crises (UK versus US). *Curr Diab Rep*, 2017; **17**: 33–9.
60. Mays JA, Jackson KL, Derby TA, *et al.* An evaluation of recurrent diabetic ketoacidosis, fragmentation of care, and mortality across Chicago, Illinois. *Diabetes Care*, 2016; **39**: 1671–76.

61. Dreschfeld J. The Bradshawe lecture on diabetic coma. *Br Med J* 1886; **2**: 358–63.
62. von Frerichs F. Uber den diabetes. *August Hirschwald, Berlin*, 1884; **1884**: 113.
63. Rosenbloom J. A form of diabetic coma, not due to the acetone bodies. *N Y Med J*, 1915; **102**: 294–6.
64. Marble A, Root HF, White P. Diabetic Coma. *N Eng J Med*, 1935; **212**: 288–97.
65. Root HF, Leech R. Diabetic coma and hyperglycemic stupor compared. *Med Clin North Am*, 1946; **30**: 1115.
66. De Graeff J, Lips JB. Hypernatraemia in diabetes mellitus. *Acta Med Scand*, 2016; **1957**: 1–71.
67. Sament S, Schwartz MB. Severe diabetic stupor without ketosis. *S Afr Med J*, 1957; **31**: 893–4.
68. Scott A, on Behalf of the Joint British Diabetes Societies (JBDS) for Inpatient Care. Management of hyperosmolar hyperglycaemic state in adults with diabetes. *Diabetic Med*, 2015; **32**: 714–24.
69. Fadini GP, De Kreutzenberg SV, Rigato M, *et al.* Characteristics and outcomes of the hyperglycemic hyperosmolar non-ketotic syndrome in a cohort of 51 consecutive cases at a single center. *Diabetes Res Clin Pract*, 2011; **94**: 172–9.
70. Pinies JA, Cairo G, Gaztambide S, Vazquez JA. Course and prognosis of 132 patients with diabetic non ketotic hyperosmolar state. *Diabete et Metabolisme*, 1994; **20**: 43–8.

15.9.2 Management of the Inpatient with Diabetes Mellitus

Gerry Rayman

Introduction

The previous edition of this textbook lamented on the general lack of interest in inpatient diabetes care. Since then there has been a significant change, partly brought about by increasing recognition of the significant financial burden of inpatient diabetes, increased lengths of stay, worryingly high levels of patient harms associated with poor control (both hyperglycaemia and hypoglycaemia), and poor patient experience [1, 2]. In England and Wales, the National Diabetes Inpatient Audit (NaDIA—*vide infra*), the world's first national diabetes audit was partially instrumental in highlighting the levels of patient harms and patient dissatisfaction as well revealing the paucity of inpatient diabetes specialists [2]. Additionally, recognizing the need to improve inpatient diabetes care the Joint British Diabetes Societies (JBDS) for inpatient care was formed to produce evidence and consensus based national guidelines to promote better care and reduce variation in practice [3]. Similarly, in 2013, diabetes specialists in the United States formed a consortium for Planning Research in Inpatient Diabetes (PRIDE) to improve inpatient diabetes care [4]. Thus, driven by evidence of suboptimal management and the associated high costs, inpatient diabetes care has now become a major focus for funders, hospital management, clinicians, and patients. The goal should be to ensure that the outcomes for people with diabetes admitted to hospital are no different from those without diabetes by prevention of inpatient hyperglycaemia, hypoglycaemia, and hospital-acquired foot lesions and ensuring early and safe discharge.

Prevalence and Economic Burden

In England, accurate data on bed occupancy by people with diabetes was for the first time available following the first NaDIA in 2010. This annual bedside snapshot audit that takes place on a single day in a week in September, in over 98% of all acute hospitals, surveys of every inpatient to find those with diabetes and then evaluates the care they have received and documents any complications that have arisen during their stay. In the first NaDIA, people with diabetes were found to occupy 14.9% of inpatient beds; higher than expected from the often quoted 10% from Hospital Episode Statistics and three times the prevalence of the condition in the general population [2, 5, 6]. Furthermore, in line with the increasing prevalence in the general population there has been a year upon year increase such that in 2017, diabetes patients accounted for more than one in six inpatient beds in England and Wales (17.9%) [7]. Projecting forward it is estimated that by 2030 more than one in four inpatients will have diabetes. Indeed, in some UK hospitals the prevalence is as high as 30% [7], a prevalence already exceeded in the whole of some American states [8].

In 2011, the NHS in England was estimated to spend £2.3–2.5 billion a year on inpatient care for people with diabetes—around 11% of total inpatient care expenditure and approximately 25% of the total expenditure on diabetes care [1]. Very similar estimates have been made for the French (22%) and American (25%) healthcare systems, both confirming that inpatient care is the largest component of medical expenditure for diabetes [9, 10].

Given the inexorable increase in prevalence, the high cost and associated demands on staffing this has major implications for funding and planning inpatient diabetes care.

Patient Outcomes: Length of Stay, Re-Admissions, and Mortality

Numerous publications based on hospital activity data confirm that patients with diabetes have longer lengths of stay and emergency re-admission compared with people without diabetes [5, 11, 12]. They also have a greater risk of dying in hospital [5]. Reasons include being an older group with more comorbidities, however after correcting for these it has been shown that unstable diabetes control in the form of hypoglycaemia and hyperglycaemia is associated with increased length of stay, morbidity and mortality [5, 12–15]. Recent studies have however demonstrated that focused initiatives can reduce length of stay and re-admission rates, which if replicated on a nationwide basis would result in substantial savings [16, 17].

Bedside Diabetes Care: Lessons from the National Diabetes Inpatient Audit (NaDIA)

Hospital activity data, collected mainly for the purposes of billing, though providing important data on length of stay, re-admissions, and mortality contains no information on the actual care patients receive. The NaDIA is unique in that it provides comprehensive information on bedside care, which includes patient harms, treatment and management errors, structures of care such as inpatient diabetes specialist staffing levels, and direct feedback from patients of their inpatient experience. Uniquely it allows hospitals to benchmark against other hospitals

The key findings of the NaDIA 2017 report, the 7th iteration providing data on 16 010 inpatients with diabetes admitted to 208 hospital sites in England and Wales are summarized in **Box 15.9.2.1**. The provision of specialist diabetes staff remains a concern and has not improved since the audits began. Medication errors and insulin errors have significantly reduced however, as shown in the box remain at unacceptable levels. Inpatient hypoglycaemia and hospital acquired foot ulceration have reduced year on year however, cases of diabetic ketoacidosis and hyperosmolar hyperglycaemic state occurring in hospital remain unacceptably high. Shockingly, 1 in 25 patients with type 1 diabetes experiences diabetic ketoacidosis during their hospital stay due to due to inadequate care. Although exclusive to England and Wales, anecdotal feedback suggests that the experience in many other countries is very similar.

Box 15.9.2.1 Key findings from the National Diabetes Inpatient Audit in 2017

NaDIA 2017 Key messages summary

- 30% of hospitals have no inpatient diabetes specialist nurse
- 30% who should be seen by the diabetes specialist are NOT SEEN
- Less than half of those insulin-treated have a 'good' diabetes day (no BG<4.0 and no more than 2 >11.0 mmol/L)
- 31% experienced medication errors—more frequent on surgical wards
- Insulin errors occurred in 40% of those treated with insulin
- Inhospital DKA occurred 1 in 25 people with type 1 diabetes
- Severe hypoglycaemia occurs in 1 in 4 people with type 1 diabetes
- 1 in 80 of all inpatients with diabetes required injectable rescue treatment
- Patient satisfaction with timing of and meal choice continue to worsen

Glucose Abnormalities on Admission and During the Hospital Stay

All patients admitted to hospital should have a blood glucose measured on admission. Those known to have diabetes with blood glucose levels persistently greater than 7.8 mmol/L require alterations to their diet and medication. In these patients, an admission HbA1c is helpful in determining the diabetes control prior to admission. In those without history of diabetes, persistent glucose results greater than 7.0 mmol/L may signify stress hyperglycaemia, which will resolve after discharge or newly diagnosed diabetes, which will persist after discharge. An admission HbA1c of ≥6.5% is helpful in confirming pre-existing diabetes and <6.5% in indicating stress hyperglycaemia.

In hospital-hypoglycaemia is clinically significant at glucose levels below 3.0 mmol/L; however, clinical staff should be alerted to levels below 4.0 mmol/L as many people with diabetes are symptomatic below these levels. In the NaDIA, mild hypoglycaemia is defined as any glucose level below 4.0 mmol/L and ≥3.0 mmol/L, severe hypoglycaemia as below 3.0 mmol/L and a third category of severe hypoglycaemia requiring injectable rescue treatment (IV dextrose or IV/IM glucagon) was included to indicate life-threatening hypoglycaemia.

The NaDIA revealed concerning levels of hypoglycaemia occurring in UK hospital. In 2011, 25.7% of people with diabetes experienced a hypoglycaemic episode during their admission; 10.6% were severe. Since, there have been year on year improvements; in 2017, the respective figures were 18.4% and 7.0%. Importantly, hypoglycaemic episodes requiring injectable rescue treatment (IV glucose or IV/IM glucagon) have fallen from 2.1% to 1.3%.

Glycaemic Targets in Hospitalized Patients with Diabetes

Over the past decade, inpatient glycaemic management has focused mainly on preventing hyperglycaemia rather than avoidance of hypoglycaemia as the former has been shown to be associated with increased morbidity and mortality, particularly in cardiac and orthopaedic surgery [18–22]. The initial studies of 'tight glycaemic control' (4.4–6.1 mmol/L), were conducted in a single intensive care unit (ICU) where it was possible to very intensively monitor glucose and intravenous insulin infusion rates. The first study in surgical ICU patients demonstrated an impressive 34% reduction in mortality, 34% reduction in septicaemia and 46% reduction in renal failure requiring dialysis by 41% [23]. A subsequent study by the same group in medical ICU patients did not show a reduction in mortality but benefit in length of stay in the ICU [24]. This led to much excitement among intensivists with many adopting strategies to achieve 'tight glycaemic control'. However, subsequent 'real

world' studies across multiple ICUs have not supported these findings and indeed suggest that 'tight control' may be associated with greater mortality possibly related to increased frequency of hypoglycaemia [25–27]. As the vast majority of diabetes admissions are not to critical care units, these studies are not helpful in determining glycaemic targets in general wards where glucose monitoring is less intensive and the patient to nursing/medical staff ratio significantly lower. A meta-analysis aimed at establishing set guidelines in non-critical hospitalized patients concluded that there is no difference in the mortality or macrovascular complications for non-critical patients with diabetes receiving intensive vs. standard glycaemic control but cautioned that the strength of this conclusion is limited by a dearth of literature from the non-critical setting [28]. In the absence of definite evidence for 'tight glycaemic control' and a broad consensus that hyperglycaemia and hypoglycaemia are associated with adverse outcomes, the pragmatic approach is to avoid hypoglycaemia and significant hyperglycaemia. Thus, American Diabetes Association recommends a target glucose range of 7.8–10.0 mmol/L for the majority of critically ill and non-critically ill patients. In the United Kingdom, the JBDS recommend an ideal blood glucose target range of 6.0–10.0 mmol/L (acceptable range 4.0–12.0 mmol/L). More stringent targets may be appropriate in selected groups such as those undergoing cardiac surgery where intensive monitoring is possible and in whom there is good evidence for the benefit of 'tight glycaemic control' [29]. Conversely, higher glucose targets may be appropriate in those who are terminally ill, those with severe comorbidities, and elderly frail patients.

Barriers to Achieving Good Glycaemic Control in Hospital

There are numerous reasons why achieving good inpatient diabetes control is challenging. These include the physiological response to illness and/or surgery, reduction in carbohydrate intake from emesis, interruption of enteral feeding, periods of fasting for procedures, comorbidities such as renal, cardiac, and hepatic disease, and alteration in insulin sensitivity due use of new medications such as steroids. System issues include coordinating meals and insulin, staff knowledge, staffing levels, communication between staff and with the patient, and insufficient frequency of blood glucose monitoring. Importantly, as over 90% of patients with diabetes are admitted for reasons unrelated to diabetes (e.g. pneumonia or a fracture) and are cared for by non-diabetes specialist staff with relatively little experience in insulin dose adjustment, the need to tailor diabetes medications during illness and in the management of diabetic emergencies. Added to this new classes of diabetes medications, new combinations of existing medications and biosimilar insulins with new names has made diabetes management significantly more challenging for diabetes specialists let alone non-specialists. Indeed, one of the largest surveys of junior doctors in the United Kingdom found that a majority felt their diabetes training did not allow them to confidently manage inpatients with diabetes safely, and that only 40% would take the initiative to optimize glycaemic control for patients under their care [30]. For these reasons, systems need to be in place systems to train non-specialist staff in the basic care of people with diabetes in hospital, and they need easy access to guidelines and policies on diabetes management as well as instruction on whom when and how to refer for specialist support. Patient and system factors that increase the risk of hypoglycaemia are listed in Box 15.9.2.2.

Box 15.9.2.2 Factors contributing to hypoglycaemia

Patient factors contributing to hypoglycaemia

- Critical illness (hepatic, cardiac, and renal failure)
- Alterations in medications such as rapid steroid dose reduction
- Reduced CHO intake: vomiting, reduced oral intake, e.g. fasting for procedures: interruption of intravenous glucose infusion, enteral or parenteral nutrition
- Impaired patient awareness and response to hypoglycaemia related to altered mental status due to sedation or illness

System failures resulting in hypoglycaemia

Inappropriate insulin administration and adjustment:

- Unawareness of interaction between timing of timing in insulin administration to meal
- Lack of information of carbohydrate content of meal and corresponding adjustment of insulin dose
- Lack of knowledge on different insulin types
- Errors in insulin prescription, dosage, and administration

Infrequent/missed blood glucose monitoring:

- Suboptimal levels of staff knowledge on the importance of monitoring
- Lack of blood glucose re-testing and reviewing, following a hypoglycaemic episode

Insufficient ward staffing levels and competencies:

- Low staffing levels, especially in specialist staff, overnight and weekends
- General knowledge and hands-on experience of diabetes lacking for nurses and junior doctors, particularly on wards not commonly dealing with patients diagnosed with diabetes
- Non-specialist staff reluctant to seek specialist's help and advice

Point-of-Care Testing and Frequency of Blood Glucose Monitoring

It is recommended that patients on insulin and/or sulphonylurea therapy are monitored prior to meals and before bed. Less frequent monitoring may be appropriate in other groups such as patients who are solely diet treated if initial monitoring suggests satisfactory and stable glycaemic control. The frequency should be intensified in the deteriorating patient or when there is a change of medication known to effect glycaemia (e.g. steroid therapy).

Glycaemic control in hospitalized patients with diabetes requires accurate near-patient glucose monitoring systems. Blood glucose meters for patient use should not be used in the hospital setting as they lack the required accuracy particularly at the lower glycaemic ranges and unlike those designed for hospital may be adversely affected by hypoxia, dehydration, haematocrit, non-glucose sugars, and interfering drugs [31].

Currently, fourth-generation point-of-care BG data management systems are available that connect seamless and bidirectionally with wireless enabled point-of-care BG meters [32]. These can be configured to provide alerts of out of range glucose results to enable early review of hypoglycaemic medication to prevent recurrent hypoglycaemia. Data from such systems have been also been used to investigate patterns of hypoglycaemia and in one study identified nocturnal hypoglycaemia as a particular issue in patients receiving nocturnal basal insulin and/or sulphonylurea therapy [33].

Glycaemic Therapies in Hospitalized Patients

In the United Kingdom as in many other countries, most patients admitted to hospital remain on their usual antihyperglycaemic regimen in contrast to the United States where the American Diabetes Association recommends a basal-bolus regimen for all diabetes admissions irrespective of their pre-admission treatments [34]. This policy has not been adopted in the United Kingdom principally because nurse to patient ratios are a considerably lower, and the provision of diabetes inpatient nurses is insufficient to support the more intensive management required. Indeed, given the high levels of harm related to insulin use already seen in UK hospitals, the widespread adoption of the basal-bolus regime in inpatients with diabetes may be unwise.

On admission, there should be a review of the patient's glycaemic therapy. In certain circumstances some oral hypoglycaemic agents may need to be discontinued (e.g. metformin and sulphonylureas in the presence of acute kidney injury); metformin should be temporarily discontinued in patients with impaired renal function if intravenous contrast material is to be administered. There will also be a need to consider altering the dose of insulin and/or sulphonylurea therapy depending on patient circumstances. Thus, in many patients the evening basal insulin and or/sulphonylurea dose may need to be reduced take into account differences in feeding patterns and nutritional content of the evening hospital meals compared with those usually eaten at home. Incretin therapies can be used but should be avoided in patients who develop heart failure. Glucagon-like peptide 1 agonists may be continued provided gastrointestinal symptoms are not an issue.

SGLT-2 inhibitors, a relatively new group of oral hypoglycaemic agents currently indicated for the treatment of type 2 diabetes but also increasingly used off licence in type 1 diabetes, should be discontinued, as there is a risk of euglycaemic ketoacidosis in ill and dehydrated patients.

A summary example of the adjustments to hypoglycaemic treatments that should be considered on admission in all patients is shown in Table 15.9.2.1.

Patients started on corticosteroid therapy must be closely monitored and where high doses are required it very likely there will be a need to escalate hypoglycaemic treatment. Furthermore, when the dose of corticosteroid is tapered dose reduction guided by frequent blood glucose monitoring is necessary to avoid hypoglycaemia. Detailed guidance for monitoring and managing steroid induced hyperglycaemia in patients with and without pre-existing diabetes is available through the JBDS [3].

Continuous Subcutaneous Insulin Infusion (CSII) Therapy

An increasing number of patients are using CSII. Unless incapacitated, most will be safest remaining on CSII if admitted to hospital. They will know how best to use the device however but they should be reviewed by the diabetes team. If the patient is unable to manage the CSII, and no specialist advice is immediately available, the CSII should be discontinued and a conventional intravenous insulin infusion or S/C basal-bolus insulin regimen.

Intravenous Insulin Infusion

In critically ill patients and those unable to eat or drink, continuous intravenous insulin infusion is the choice method for achieving and maintaining glucose control. However, as intravenous insulin has a half-life of a few minutes, if interrupted

Table 15.9.2.1 Diabetes medicines management on admission card

Diabetes medicines management on admission	Preventing hypoglycaemia in the first 48 hours after admission
• SGLT2 inhibitors (e.g. empagliflozin)—RISK of DKA. Stop for major surgery or severe medical illness. • Metformin—RISK of lactic acidosis in presence of significant kidney, liver, respiratory, or cardiac failure. Stop if any of above or in those with and eGFR <60 undergoing a procedure that requires prolonged fasting, or a diagnostic test requiring radio-opaque contrast material which can impair kidney function. • NEVER stop the patient's basal insulin (e.g. Lantus or Levemir). Continue even when on an insulin infusion. • Gliclazide—RISK of nocturnal hypoglycaemia in renal impairment or where food intake is reduced. Reduce or omit (if already on low dose) evening dose. • All oral diabetes drugs should preferably be administered with food.	• Hypoglycaemia is common within 48 hours of admission; therefore unless admitted with a hyperglycaemic emergency or admission blood glucose > 15 mmol/l: – Basal-bolus regiment—reduce pre-bed basal insulin by 20% (i.e. give 80% of usual dose); – If on a pre-mix insulin (e.g. Humulin M3) reduce evening dose by 20% (i.e. give 80% of usual dose); – Halve or omit (if already on low dose) the evening dose of sulphonylurea (Gliclazide). • Reassess hypoglycaemic therapy the following day.

Note: Never omit insulin after correcting a hypoglycaemic event as rebound DKA may occur.
Source: The Ipswich Hospital NHS Trust.

ketoacidosis can quickly ensue. This can occur if the infusion line becomes blocked or dislodged or the infusion pump is inadvertently switched off, or intentionally stopped but not restarted (e.g. when transferring from the ward to an investigation area). For these reasons, it is recommended the basal insulin be continued during the infusion in those already receiving a basal insulin. It is also essential that patients placed on an intravenous infusion have hourly monitoring. In most circumstances the variable rate intravenous insulin infusion (VRIII) is used. In those already insulin-treated the initial algorithm may be determined by the previous insulin total daily dose, otherwise a mid-range algorithm is used. The blood glucose response to the insulin infusion must be reviewed regularly to determine whether the glucose/insulin algorithm should be altered. Because of the risk of hypoglycaemia an intravenous glucose infusion should be administered either as 5% dextrose with 40 mmol/L KCL at 125 ml/hr if the serum potassium is 3.5–5.5 mmol/L or preferably 0.45% NaCl and 5% glucose with 0.3% KCl (40 mmol/L) at 125 ml/hr if serum K is 3.5–5.5 mmol/L. A fixed rated intravenous insulin infusion (FRIII) is now considered the most effective means of correcting ketosis in patients with diabetic ketoacidosis. Examples of infusion scales for the VRIII as well as the insulin dose calculation for the FRIII are shown in **Table 15.9.2.2**. More detailed information on the use of VRIIIs and FRIIIs is available from the Joint British Diabetes Societies as well as in specials situations such as patients on parenteral nutrition [3].

In addition to diabetic ketoacidosis, hypoglycaemia, fluid overload, hypokalaemia and/or hyponatraemia, and cannula site infection are potential risk of intravenous insulin infusion. Daily electrolyte measurement is recommended. Given these risks, and the discomfort and intrusion of frequent monitoring especially overnight for the patient, intravenous insulin infusions should only be used where absolutely necessary. In the NaDIA 2010, 7.6% of infusions were considered unnecessary, 9.9% continued for too long and in 25.8% of patients, the transfer back to subcutaneous insulin not well managed. Fortunately, the situation has improved and in 2017, the respective figures were 6.1%, 7.1, and 16.4%.

When transferring from intravenous to subcutaneous insulin, it is important that the latter has been given prior to stopping the infusion to prevent insulinopenia and subsequent ketoacidosis. Ideally, basal insulin should have been continued during the infusion. The VRIII should only be discontinued 30 minutes after subcutaneous short acting or mixed insulin has been given and ideally at a mealtime. Transfer should not be done at bedtime where there is less observation by staff.

If the patient is new to insulin the total daily insulin dose can be calculated from the insulin requirement over the last 6 hours on the VRIII. It is prudent to reduce this by 20%. Fifty per cent (50%) is given as basal insulin and the rest divided initially equally for each of the three meals.

An exciting new development in achieving glucose levels within a 'tight' target range without the risk of hypoglycaemia is the use of automated, closed-loop insulin-delivery systems. To date, in inpatients their use has been limited to those with type 2 diabetes and at present they remain at the research stage [35, 36]. Though attractive, considerable expertise in the use of closed-loop technology and the involvement of diabetes staff expert in fitting and calibrating the sensors is essential. Although such systems hold considerable promise, further work is necessary to improve their ease of use before they can be used on busy general wards [37].

Self-Management

All clinicians who talk to people with diabetes are aware that many are deeply unhappy about the care they receive in hospital. The NaDIA reports that patients are disempowered, particularly in their involvement in their own diabetes control; this includes being permitted to being allowed to self-monitor their own blood glucose, self-administer insulin, and contribute to decisions on insulin dose adjustment. They also report significant concerns about the quality, choice and timing of meals in relation to insulin administration [7].

Self-management is appropriate in selected patients and should be encouraged provided the patient is known to successfully manage their diabetes at home and is well enough and not cognitively affected by their illness. With one in 25 people with type 1 diabetes experiencing diabetic ketosis during their hospital admission, far greater than would occur self-caring in the community, it is not surprising that many patients self-discharge in the belief that they are more capable of controlling their diabetes that the staff in the hospital.

Safe Discharge from Hospital and Preventing Re-Admissions

Re-admission rates within 28 days for people with diabetes are between 14% and 20% and are 59% higher than those without diabetes [38]. A proportion of these can be prevented by a 'safe diabetes discharge policy'. This should include a formal review of diabetes therapies before discharge, which can be informed by HbA1c measured on admission. Care should be taken to ensure that the patient or carer is able to recognize and treat hypoglycaemia, has written instruction on insulin doses and their timing and that they leave with all the necessary equipment and treatments (e.g. needles, insulin pens, blood glucose monitoring). Care should also be taken to avoid transcription errors when communicating medication dosages to the patient, carers, and the primary care physician and all should be aware of follow-up arrangements. The inpatient diabetes specialist nurse is often the key member of the team in this discharge policy.

A relatively small cohort of patients with recurrent admissions account for a significant proportion of inpatient stays in people with diabetes. In one study, 30% of patients who had two or more hospital admissions accounted for more than 50% of such hospitalizations [39]. Targeting these patients may decrease re-admissions. This would include provision of psychological support for those with recurrent ketoacidosis and hypoglycaemia related to behavioural problems. Frail elderly and patients from nursing and residential homes are at particular risk of recurrent hypoglycaemic admissions and/or ambulance call outs with over a third of these being repeated events [40]. Support from the diabetes specialist team, in particular focusing on medication de-escalation and setting less stringent blood glucose targets may reduce re-admissions in this group.

Table 15.9.2.2 Examples of variable rate and fixed rate intravenous infusion scales

Variable Rate Intravenous Insulin Infusion (Vriii)–*Scale 2 Most Commonly Used*								Fixed rate intravenous insulin infusion (friii) Fixed irrespective of prevailing glucose *See DKA/HHS guidelines on concurrent fluid management*	
Scale 1 **Daily insulin requirement less than 24 units.**		**Scale 2** **Daily insulin requirement 24–100 units.**		**Scale 3** **Daily insulin requirement over 100 units.**		**Scale 4 (Others)** **Please prescribe insulin rate in boxes below**		**DKA Scale**	**HHS Scale**
Glucose	Insulin Rate Units/hour	Glucose	Insulin Rate Units/hour	Glucose	Insulin Rate Units/hour	Glucose	Insulin Rate Units/hour	Rate = Estimate wt..... kg × 0.1 units/hr Initial starting rate = ... units / hr (if exceeds 12 units/hr, call Diabetes Team) Date/Time: Signature:	Rate = Estimate wt..... kg × 0.05 units/hr Initial starting rate = ... units / hr (if exceeds 12 units/hr, call Diabetes Team) Date/Time: Signature:
≤ 4	0*	≤ 4	0*	≤ 4	0*	≤ 4	0*		
4.1–8.0	0.5	4.1–8.0	1	4.1–8.0	2	4.1–8.0			
8.1–12.0	1	8.1–12.0	2	8.1–12.0	4	8.1–12.0			
12.1–16.0	2	12.1–16.0	4	12.1–16.0	6	12.1–16.0			
16.1–20.0	3	16.1–20.0	5	16.1–20.0	7	16.1–20.0			
20.1–24.0	4	20.1–24.0	6	20.1–24.0	8	20.1–24.0			
≥ 24.1	6	≥ 24.1	8	≥ 24.1	10	≥ 24.1		If changed, new rate = ... units / hr (See DKA guidelines on when to change) Date/Time of change: Signature:	If changed, new rate = ... units/hr (See HHS guideline on when to change) Date/Time of change: Signature:
Date / Time		Date / Time		Date / Time		Date / Time			
Signature		Signature		Signature		Signature			
* Treat hypoglycaemia and restart IV insulin within 20 minutes									

REFERENCES

1. Inpatient Care for People with Diabetes: The Economic Case for Change. Marion Kerr 2011. Available at: https://www.diabetes.org.uk/resources-s3/2017-10/Inpatient%20Care%20for%20People%20with%20Diabetes%20%20The%20Economic%20Nov%202011_1.pdf (accessed: August 2018).
2. National Diabetes Inpatient Audit 2012. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-inpatient-audit/national-diabetes-inpatient-audit-open-data-2011 (accessed: August 2018).
3. Joint British Diabetes Societies (JBDS) for Inpatient Care Group. Available at: https://abcd.care/joint-british-diabetes-societies-jbds-inpatient-care-group (accessed: August 2018).
4. Draznin B, Gilden J, Golden S, for the PRIDE investigators. Pathways to quality inpatient management of hyperglycemia and diabetes: a call to action. *Diabetes Care*, 2013; **36**: 1807–14.
5. Whitston M, Chung S, Henderson J, Young B. What can be learned about the impact of diabetes on hospital admissions from routinely recorded data? *Diabet Med*, 2012; **29**: 1199–205.
6. Ogurtsova K, da Rocha Fernandes J, Huang Y, *et al.* IDF Diabetes atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract*, 2017; **128**: 40–50.
7. National Diabetes Inpatient Audit 2017. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-inpatient-audit/national-diabetes-inpatient-audit-nadia-2017 (accessed: 24 April 2020).
8. Meng YY, Pickett M, Babey SH, Davis AC, Goldstein H. Diabetes tied to a third of California hospital stays, driving health care costs higher. Los Angeles: UCLA Center for Health Policy Research and California Center for Public Health Advocacy, 2014. Available at: http://healthpolicy.ucla.edu/publications/search/pages/detail.aspx?PubID=1278 (accessed: 27 September 2015).
9. de Lagasnerie G, Aguadé AS, Denis P, Fagot-Campagna A, Gastaldi-Menager C. The economic burden of diabetes to French national health insurance: a new cost-of-illness method based on a combined medicalized and incremental approach. *Eur J Health Econ*, 2018; **19**: 189–201.
10. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. *Diabetes Care*, 2018; **41**: 917–28.
11. Sampson MJ, Dozio N, Ferguson B, Dhatariya K. Total and excess bed occupancy by age, speciality and insulin use for nearly one million diabetes patients discharged from all English Acute Hospitals. *Diabetes Res Clin Pract*, 2007; **1**: 92–8.
12. Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia is associated with increased length of stay and mortality in people with diabetes who are hospitalized. *Diabet Med*, 2012; **29**: e445–8.
13. Umpierrez GE, Isaacs SD, Bazargan N, You X, Thaler LM, Kitabchi AE. Hyperglycaemia: an independent marker of inhospital mortality inpatients with undiagnosed diabetes. *J Clin Endocrinol Metab*, 2002; **87**: 978–82.
14. Krinsley JS, Maurer P2, Holewinski S, *et al.* Glucose control, diabetes status, and mortality in critically ill patients: the continuum from intensive care unit admission to hospital discharge. *Mayo Clin Proc*, 2017; **92**: 1019–29.
15. Turchin A, Matheny ME, Shubina M, Scanlon JV, Greenwood B, Pendergrass ML. Hypoglycemia and clinical outcomes in patients with diabetes hospitalized in the general ward. *Diabetes Care*, 2009; **32**: 1153–7.
16. Sampson MJ1, Crowle T, Dhatariya K, *et al.* Trends in bed occupancy for inpatients with diabetes before and after the introduction of a diabetes inpatient specialist nurse service. *Diabet Med*, 2006; **23**: 1008–15.
17. Garg R, Schuman B, Bader A, *et al.* Effect of preoperative diabetes management on glycemic control and clinical outcomes after elective surgery. *Ann Surg*, 2018; **267**: 858–62.
18. Furnary AP, Zerr KJ, Grunkemeier GL, Starr A. Continuous intravenous insulin infusion reduces the incidence of deep sternal wound infection in diabetic patients after cardiac surgical procedures. *Ann Thorac Surg*, 1999; **67**: 352–60.
19. Frisch A, Chandra P, Smiley Dl. Prevalence and clinical outcome of hyperglycemia in the perioperative period in noncardiac surgery. *Diabetes Care*, 2010; **33**: 1783–8.
20. Martin ET, Kaye KS, Knott C, *et al.* Diabetes and risk of surgical site infection: a systematic review and meta-analysis. *Infect Control Hosp Epidemiol*, 2016; **37**: 88–99.
21. Ramos M, Khalpey Z, Lipsitz S, *et al.* Relationship of perioperative hyperglycemia and postoperative infections in patients who undergo general and vascular surgery. *Ann Surg*, 2008; **248**: 585–91.
22. Umpierrez GE, Smiley D, Jacobs S, *et al.* Randomized study of basal-bolus insulin therapy in the inpatient management of patients with type 2 diabetes undergoing general surgery (RABBIT 2 surgery). *Diabetes Care*, 2011; **34**: 256–61.
23. Van den Berghe G, Wouters P, Weekers F, *et al.* Intensive insulin therapy in critically ill patients. *N Engl J Med*, 2001; **345**: 1359–67.
24. Van den Berghe G, Wilmer A, Hermans G, *et al.* Intensive insulin therapy in the medical ICU. *N Engl J Med*, 2006; **354**: 449–61.
25. 10. The NICE–SUGAR Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*, 2009; **360**: 1283–97.
26. Abdelmalak BB, Lansang MC. Revisiting tight glycemic control in perioperative and critically ill patients: when one size may not fit all. *J Clin Anesth*, 2013; **25**: 499–507.
27. Yamada T, Shojima N, Noma H, Yamauchi T, Kadowaki T. Glycemic control, mortality, and hypoglycemia in critically ill patients: a systematic review and network meta-analysis of randomized controlled trials. *Intens Care Med*, 2017; **43**: 1–15.
28. Murad MH, Coburn JA, Coto-Yglesias F, *et al.* Glycemic control in non-critically ill hospitalized patients: a systematic review and meta-analysis. *J Clin Endocrinol Metab*, 2012; **97**: 49–58.
29. American Diabetes Association. Diabetes care in the hospital: standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41**(Suppl 1): S144–51.
30. George JT, Warriner D, McGrane DJ, *et al.* Lack of confidence among trainee doctors in the management of diabetes: the trainees own perception of delivery of care (TOPDOC) diabetes study. *QJM*, 2011; **104**: 761–6.
31. Rajendran R, Rayman G. Point-of-care blood glucose testing for diabetes care in hospitalized patients: an evidence-based review. *J Diabetes Sci Technol*, 2014; **8**: 1081–90.
32. Kim JY, Lewandrowski K. Point-of-care testing informatics. *Clin Lab Med*, 2009; **29**: 449–61.
33. Kerry C, Mitchell S, Sharma S, Scott A, Rayman G. Diurnal temporal patterns of hypoglycaemia in hospitalized people with diabetes may reveal potentially correctable factors. *Diabetic Med*, 2013; **30**: 1403–6.

34. Moghissi ES, Korytkowski MT, DiNardo M, *et al.*; American Association of Clinical Endocrinologists; American Diabetes Association. American Association of Clinical Endocrinologists and American Diabetes Association consensus statement on inpatient glycemic control. *Diabetes Care*, 2009; **32**: 1119–31.
35. Thabit H, Hartnell S, Allen JM, *et al.* Closed-loop insulin delivery in inpatients with type 2 diabetes: a randomised, parallel-group trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 117–24.
36. Bally L, Thabit H, Hartnell S, *et al.* Closed-loop insulin delivery for glycemic control in noncritical care. *N Engl J Med*, 2018; **379**: 547–56.
37. Rayman G. Closer to closing the loop on inpatient glycaemia. *Lancet Diabetes Endocrinol*, 2017; **5**: 81–3.
38. Rubin DJ. Hospital readmission of patients with diabetes. *Curr Diab Rep*, 2015; **15**: 17.
39. Jiang HJ, Stryer D, Friedman B, Andrews R. Multiple hospitalizations for patients with diabetes. *Diabetes Care*, 2003; **26**: 1421–6.
40. Sampson M, Bailey M, Clark J, *et al.* A new integrated care pathway for ambulance attended severe hypoglycaemia in the East of England: The Eastern Academic Health Science Network (EAHSN) model. *Diabetes Res Clin Pract*, 2017; **133**: 50–9.

15.9.3 Care of Diabetes in ICU and Perisurgery

Jan Gunst and Greet Van den Berghe

Introduction

Evoked by severe physical stress, critically ill patients and patients undergoing major surgery usually develop stress hyperglycaemia, irrespective of their pre-admission diabetes status. Large observational studies have shown that the degree of stress hyperglycaemia relates to poor outcome (Figure 15.9.3.1), with the lowest risk of death being associated with normal or nearly normal age-adjusted fasting blood glucose levels [1, 2]. Remarkably, in critically ill patients with established diabetes, the relationship between hyperglycaemia and mortality is blunted and shifted to the right [1]. Since these associations are derived from observational studies, they do not imply causality. Indeed, sicker patients will develop more severe insulin resistance, with consequently higher blood glucose levels.

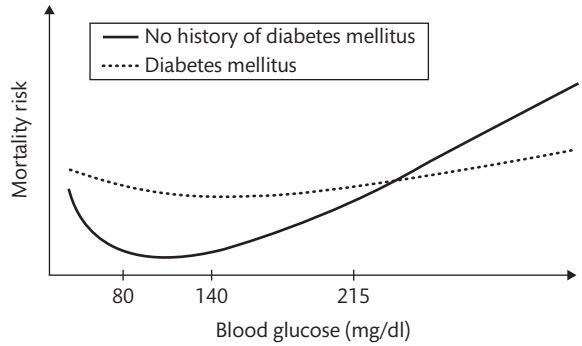


Figure 15.9.3.1 The association of blood glucose concentrations with outcome in critically ill adults. In observational studies, the statistical association between blood glucose concentrations and the risk of mortality follows a J-shaped curve. To convert mg/dl to mmol/L, divide by 18.

A causal relationship can only be derived from adequately powered randomized controlled trials (RCTs).

Tight Glucose Control in the Intensive Care Unit

Over the last two decades, several RCTs—varying in design, study population, size, blood glucose target, and outcome measures—have tested the hypothesis that tight glucose control may improve outcome of critically ill patients [3–8].

The Leuven Studies on Tight Glucose Control in the Intensive Care Unit

In three pioneer studies performed in Leuven, Belgium, tight glucose control improved morbidity and mortality of critically ill adults and children, as compared to tolerating hyperglycaemia up to 12 mmol/L (215 mg/dl) [3–5]. In the intervention group, age-adjusted normal fasting glucose levels were targeted (4.4–6.1 mmol/L for adults [80–110 mmol/L], 3.9–5.6 mmol/L [70–100 mg/dl] for children older than 1 year, and 2.8–4.4 mmol/L for infants [50-80 mg/dl]) and insulin was administered through continuous intravenous infusion (no boluses). As long as an arterial line was present, glucose measurements were performed on arterial blood, using an accurate blood gas analyser, with simultaneous measurement of potassium. All patients received early parenteral nutrition as part of the standard care at that time. The intervention protocol was highly efficient, with >70% of measurements within target and a clear separation in blood glucose levels between the randomization groups.

Importantly, the short-term clinical benefit was maintained on the long-term in both adults and children [9, 10], and the intervention was shown to reduce healthcare costs [11]. Subsequent mechanistic studies attributed the beneficial impact to the avoidance of hyperglycaemia and consequent glucose overload in vital organs, rather than to glycaemia-independent effects of insulin [12].

Subsequent Studies and Possible Contributors to Interstudy Discrepancies

After the Leuven trials, several large implementation studies and small RCTs have confirmed a positive clinical impact of implementing tight glucose control in critically ill patients [13–16].

However, the intervention remains debated, since multicentre RCTs have not reproduced the benefit observed in the pioneer studies [6–8]. The largest RCT, the NICE-SUGAR trial, even found excess mortality by tight glucose control, which was subsequently attributed to the increased risk of hypoglycaemia [7, 17]. Substantial differences in study design may explain these seemingly contradictory results, including differences in blood glucose target, the accuracy of the glucose measurements, the insulin infusion protocol, and the feeding protocol.

In contrast to the Leuven studies, in which tight glucose control was compared to tolerating severe hyperglycaemia, many subsequent studies, including NICE-SUGAR, targeted stricter blood glucose levels in the control group, explained by a shift in standard care after the Leuven studies [3–5, 7]. The outcome effect of lowering blood glucose from less severely elevated to tight blood glucose levels is presumably smaller than with a larger difference in blood glucose target. This is confirmed by a secondary analysis of the Leuven studies, which found a dose-dependent impact of blood glucose lowering on outcome [18].

Second, the accuracy of the blood glucose measurement and the insulin protocol differed among studies. Whereas in the Leuven studies, blood glucose was measured on arterial blood using an accurate blood gas analyser, in many subsequent studies including NICE-SUGAR, a variety of—at that time—inaccurate glucometers were used [19]. Also, subsequent studies allowed capillary glucose measurements, although unreliable in patients with shock [12]. Apart from that, the insulin treatment protocol differed. In Leuven, insulin was only administered through continuous intravenous infusion, in contrast to subsequent studies in which insulin boluses were allowed. Theoretically, inaccurate glucose measurements and insulin boluses increase blood glucose variability and the risk of prolonged and undetected hypoglycaemia, which may have contributed to the harm observed in NICE-SUGAR.

Finally, the feeding protocol differed among studies. In the Leuven studies, all patients received early parenteral nutrition as part of the standard treatment, whereas in NICE-SUGAR and other subsequent studies, patients received less parenteral calories in the acute phase [3–5, 7]. Early parenteral nutrition increases the risk of hyperglycaemia and was subsequently shown to be harmful, even in a context of tight glucose control [20, 21]. A meta-analysis revealed that the mortality impact of tight glucose control related to the amount of parenteral glucose given in the trial [22]. However, this meta-analysis did not adjust for other methodological differences. In contrast, a secondary analysis of the Leuven studies showed that the benefit of tight glucose control was also present in the subgroup who received the lowest amount of parenteral glucose [18]. Hence, it currently remains unknown whether tight glucose control is still beneficial in the absence of early parenteral nutrition.

Intraoperative Tight Glucose Control

Most large RCTs have investigated the effect of tight glucose control in intensive care unit (ICU) patients. However, major surgery *per se* also evokes hyperglycaemia, which has been associated with an increased risk of complications [2]. Several RCTs have investigated whether tight, intraoperative control of blood glucose also improves outcome as compared to tolerating hyperglycaemia. Altogether, these studies showed mixed results. Whereas some RCTs showed benefit in patients undergoing cardiac and vascular surgery, other trials in cardiac and non-cardiac surgery patients were neutral [23–26]. However, as the duration of the intervention was relatively small (mainly the time in the operating room), the benefit the investigators anticipated was probably too large. Hence, several of these RCTs were likely underpowered [27]. Interestingly, the largest RCT, which randomized 2383 elective cardiac surgery patients to intraoperative tight (4.4–6.1 mmol/L [80–110 mg/dl]) or liberal blood glucose control (<10 mmol/L [180 mg/dl]), followed by tight glucose control in all patients once admitted to the ICU, found less postoperative complications with the addition of intraoperative tight glucose control to only postoperative tight control [23]. Likewise, a recent RCT found reduced morbidity and mortality by a hyperinsulinaemic normoglycaemic clamp during elective cardiac surgery, as compared to standard care, which consisted of tolerating hyperglycaemia up to 8.3 mmol/L (150 mg/dl) [28]. As the study combined a hyperinsulinaemic clamp with maintaining normoglycaemia by adjusting the glucose intake, it is unclear whether the clinical benefit is explained by the lower glucose levels, by the higher insulin dose, or by a combination of both. However, as previous RCTs on glucose-insulin-potassium therapy in acute myocardial infarction revealed negative results when targeting normoglycaemia was not part of the intervention protocol [29, 30], it seems plausible that at least part of the benefit of the latter trial is explained by maintaining normal blood glucose levels.

Tight Glucose Control: Which Patients Benefit Most?

Theoretically, the mixed results across the different RCTs on tight glucose control could be explained by a different case mix. However, subgroup analyses could not identify a patient subgroup who responded differently to the intervention than others, with the possible exception of patients with pre-admission diabetes [7, 8, 18]. Indeed, in a secondary analysis of the Leuven RCTs, all patient subgroups showed benefit from tight glucose control, except patients with pre-admission diabetes in whom the intervention was neutral [18]. Likewise, in the large, intraoperative RCT, the benefit of tight glucose control was only present in the subgroup of non-diabetes patients [23]. These findings support observational studies, which revealed a flattened and right-shifted relationship of blood glucose with outcome, especially in patients with poorly controlled diabetes [1, 31] (**Figure 15.9.3.1**). Conversely, in NICE-SUGAR, all patient subgroups were harmed by the intervention, including patients without pre-admission diabetes [7]. Moreover, a large observational study in 5510 diabetic cardiac surgery patients found reduced mortality and infectious complications by implementation of a perioperative glucose control protocol [32].

Tight Glucose Control and the Risk of Hypoglycaemia

Tight glucose control inevitably increases the risk of hypoglycaemia. When severe or prolonged, hypoglycaemia can cause convulsions, coma, and irreversible brain damage, as well as cardiac

arrhythmias. The risk of hypoglycaemia is a concern in critically ill and surgical patients because early hypoglycaemic symptoms are not easily recognized in sedated patients [18]. Multiple observational studies have associated the occurrence of hypoglycaemia with an increased risk of death. However, hypoglycaemia is more prevalent in patients with more severe disease (particularly liver failure and sepsis). It remains unknown whether a short-lasting episode of iatrogenic hypoglycaemia adversely impacts outcome [18]. A secondary analysis of NICE-SUGAR attributed the increased mortality risk by tight glucose control to an increased risk of hypoglycaemia [17]. However, also prolonged hyperglycaemia may cause harm. A 4-year neurocognitive follow-up study of patients included in the paediatric Leuven study showed a better neurocognitive outcome of children randomized to tight glucose control, despite a high incidence of hypoglycaemia [10]. Moreover, a nested case-control study, matching critically ill children who effectively experienced an episode of hypoglycaemia to critically ill children without a hypoglycaemic event, also could not detect any harmful impact of hypoglycaemia on neuronal biomarkers and on 4-years neurocognitive outcome [10, 33]. The fact that overcorrection of hypoglycaemia was carefully avoided in the latter study [5], may have played a role. Indeed, animal data suggest that neuronal death after hypoglycaemia is mainly triggered by activation of neuronal NADPH oxidase, which occurs during glucose reperfusion [34]. Nevertheless, treatment protocols should aim at minimizing the risk of hypoglycaemia. The use of validated computerized treatment algorithms and of near-continuous glucose monitoring may substantially reduce this risk [35].

Summary and Conclusions

Currently, the optimal blood glucose target for critically ill and surgical patients remains unclear and may depend on the context. Tight glucose control has been shown to reduce morbidity and mortality of critically ill patients when performed with accurate glucose monitoring tools and an effective treatment algorithm that avoids insulin boluses in a context of early parenteral nutrition. Analyses of patient subgroups have not shown a differential impact of the intervention, with the possible exception of patients with pre-admission diabetes, who may benefit less from tight glucose control. The impact of tight glucose control in patients in whom early parenteral nutrition is avoided, is unknown. When the logistic context does not allow frequent, accurate glucose monitoring, tolerating higher glucose concentrations appears safer than strictly targeting the healthy fasting range. In any case, the incidence of hypoglycaemia and severe hyperglycaemia should be minimized, which can be achieved by the use of validated computerized treatment algorithms.

REFERENCES

1. Falciglia M, Freyberg RW, Almenoff PL, D'Alessio DA, Render ML. Hyperglycemia-related mortality in critically ill patients varies with admission diagnosis. *Crit Care Med*, 2009; **37**: 3001–9.
2. van den Boom W, Schroeder RA, Manning MW, Setji TL, Fiestan GO, Dunson DB. Effect of A1C and glucose on postoperative mortality in noncardiac and cardiac surgeries. *Diabetes Care*, 2018; **41**: 782–8.
3. Van den Berghe G, Wouters P, Weekers F, *et al.* Intensive insulin therapy in critically ill patients. *N Engl J Med*, 2001; **345**: 1359–67.
4. Van den Berghe G, Wilmer A, Hermans G, *et al.* Intensive insulin therapy in the medical ICU. *N Engl J Med*, 2006; **354**: 449–61.
5. Vlasselaers D, Milants I, Desmet L, *et al.* Intensive insulin therapy for patients in paediatric intensive care: a prospective, randomised controlled study. *Lancet*, 2009; **373**: 547–56.
6. Preiser JC, Devos P, Ruiz-Santana S, *et al.* A prospective randomised multi-centre controlled trial on tight glucose control by intensive insulin therapy in adult intensive care units: the Glucontrol study. *Intensive Care Med*, 2009; **35**: 1738–48.
7. Finfer S, Blair D, Bellomo R, *et al.* Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*, 2009; **360**: 1283–97.
8. Kalfon P, Giraudeau B, Ichai C, *et al.* Tight computerized versus conventional glucose control in the ICU: a randomized controlled trial. *Intensive Care Med*, 2014; **40**: 171–81.
9. Ingels C, Debaveye Y, Milants I, *et al.* Strict blood glucose control with insulin during intensive care after cardiac surgery: impact on 4-years survival, dependency on medical care, and quality-of-life. *Eur Heart J*, 2006; **27**: 2716–24.
10. Mesotten D, Gielen M, Sterken C, *et al.* Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: a randomized controlled trial. *JAMA*, 2012; **308**: 1641–50.
11. Van den Berghe G, Wouters PJ, Kesteloot K, Hilleman DE. Analysis of healthcare resource utilization with intensive insulin therapy in critically ill patients. *Crit Care Med*, 2006; **34**: 612–16.
12. Gunst J, Van den Berghe G. Blood glucose control in the ICU: don't throw out the baby with the bathwater! *Intensive Care Med*, 2016; **42**: 1478–81.
13. Krinsley JS. Effect of an intensive glucose management protocol on the mortality of critically ill adult patients. *Mayo Clin Proc*, 2004; **79**: 992–1000.
14. Lecomte P, Van Vlem B, Coddens J, *et al.* Tight perioperative glucose control is associated with a reduction in renal impairment and renal failure in non-diabetic cardiac surgical patients. *Crit Care*, 2008; **12**: R154.
15. Bilotta F, Caramia R, Paoloni FP, Delfini R, Rosa G. Safety and efficacy of intensive insulin therapy in critical neurosurgical patients. *Anesthesiology*, 2009; **110**: 611–19.
16. Okabayashi T, Shima Y, Sumiyoshi T, *et al.* Intensive versus intermediate glucose control in surgical intensive care unit patients. *Diabetes Care*, 2014; **37**: 1516–24.
17. Finfer S, Liu B, Chittock DR, *et al.* Hypoglycemia and risk of death in critically ill patients. *N Engl J Med*, 2012; **367**: 1108–18.
18. Van den Berghe G, Wilmer A, Milants I, *et al.* Intensive insulin therapy in mixed medical/surgical intensive care units—benefit versus harm. *Diabetes*, 2006; **55**: 3151–9.
19. Scott MG, Bruns DE, Boyd JC, Sacks DB. Tight glucose control in the intensive care unit: are glucose meters up to the task? *Clin Chem*, 2009; **55**: 18–20.
20. Casaer MP, Mesotten D, Hermans G, *et al.* Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*, 2011; **365**: 506–17.
21. Fivez T, Kerklaan D, Mesotten D, *et al.* Early versus late parenteral nutrition in critically ill children. *N Engl J Med*, 2016; **374**: 1111–22.

22. Marik PE, Preiser JC. Toward understanding tight glycemic control in the ICU: a systematic review and metaanalysis. *Chest*, 2010; **137**: 544–51.
23. Bláha J, Mráz M, Kopecký P, *et al.* Perioperative tight glucose control reduces postoperative adverse events in nondiabetic cardiac surgery patients. *J Clin Endocrinol Metab*, 2015; **100**: 3081–9.
24. Subramaniam B, Panzica PJ, Novack V, *et al.* Continuous perioperative insulin infusion decreases major cardiovascular events in patients undergoing vascular surgery: a prospective, randomized trial. *Anesthesiology*, 2009; **110**: 970–7.
25. Gandhi GY, Nuttall GA, Abel MD, *et al.* Intensive intraoperative insulin therapy versus conventional glucose management during cardiac surgery: a randomized trial. *Ann Intern Med*, 2007; **146**: 233–43.
26. Abdelmalak BB, Bonilla A, Mascha EJ, *et al.* Dexamethasone, light anaesthesia, and tight glucose control (DeLiT) randomized controlled trial. *Br J Anaesth*, 2013; **111**: 209–21.
27. Van den Berghe G. Does intensive insulin therapy during cardiac surgery improve postoperative outcome? *Nat Clin Pract Endocrinol Metab*, 2007; **3**: 630–1.
28. Duncan AE, Sessler DI, Sato H, *et al.* Hyperinsulinemic normoglycemia during cardiac surgery reduces a composite of 30-day mortality and serious in-hospital complications: a randomized clinical trial. *Anesthesiology*, 2018; **128**: 1125–39.
29. Mehta SR, Yusuf S, Díaz R, Zhu J, Pais P, Xavier D, *et al.* Effect of glucose-insulin-potassium infusion on mortality in patients with acute ST-segment elevation myocardial infarction: the CREATE-ECLA randomized controlled trial. *JAMA*, 2005; **293**: 437–46.
30. Malmberg K, Rydén L, Wedel H, *et al.* Intense metabolic control by means of insulin in patients with diabetes mellitus and acute myocardial infarction (DIGAMI 2): effects on mortality and morbidity. *Eur Heart J*, 2005; **26**: 650–61.
31. Plummer MP, Bellomo R, Cousins CE, *et al.* Dysglycaemia in the critically ill and the interaction of chronic and acute glycaemia with mortality. *Intensive Care Med*, 2014; **40**: 973–80.
32. Furnary AP, Wu Y. Eliminating the diabetic disadvantage: the Portland Diabetic Project. *Semin Thorac Cardiovasc Surg*, 2006; **18**: 302–8.
33. Vanhorebeek I, Gielen M, Boussemaere M, *et al.* Glucose dysregulation and neurological injury biomarkers in critically ill children. *J Clin Endocrinol Metab*, 2010; **95**: 4669–79.
34. Suh SW, Gum ET, Hamby AM, Chan PH, Swanson RA. Hypoglycemic neuronal death is triggered by glucose reperfusion and activation of neuronal NADPH oxidase. *J Clin Invest,* 2007; **117**: 910–18.
35. Dubois J, Van Herpe T, van Hooijdonk RT, *et al.* Software-guided versus nurse-directed blood glucose control in critically ill patients: the LOGIC-2 multicenter randomized controlled clinical trial. *Crit Care*, 2017; **21**: 212.

15.10

Specialized Management of Other Forms of Diabetes

15.10.1 Monogenic Forms of Diabetes Resulting from Beta-Cell Dysfunction

Andrew Hattersley, Kashyap A. Patel, and Rachel Besser

MODY

Overview

MODY is a familial form of diabetes with autosomal dominant inheritance that typically presents with non-insulin dependent diabetes diagnosed under the age of 25 years. The term 'MODY' was based on the previous definition of juvenile-onset and maturity-onset diabetes (now type 1 and type 2 diabetes). Although the term MODY is useful as it describes a broad category, cases are more appropriately described by the gene involved when this is known.

MODY is a genetically heterogeneous condition. The commonest causes are due to defects in the transcription factor genes: hepatic nuclear factor-1 alpha (*HNF1A*; ~52%), and hepatic nuclear factor-4 alpha (*HNF4A;* ~10%), as well as the Glucokinase gene (*GCK*; ~32%) [1, 2]. The diabetes phenotype is similar in HNF1A and HNF4A MODY and distinct from that seen in GCK MODY. These are summarized in Table 15.10.1.1.

HNF1A and HNF4A MODY

Clinical Phenotype

Diabetes

HNF1A and HNF4A MODY have a similar diabetes phenotype with progressive hyperglycaemia throughout life [3] (Figure 15.10.1.1). Diabetes characteristically presents in adolescence or early adulthood, in HNF1A MODY 70% develop diabetes by age 25 years, 97% by 50 years. The mean age of diagnosis is similar in HNF1A (20 years) and HNF4A MODY (23 years) [2, 4, 5]. Postprandial glucose values are raised before fasting hyperglycaemia develops and hence, during an oral glucose tolerance test (OGTT), patients typically have an elevated 2-hour glucose with a large increment above the fasting value (>3 mmol/L) [6].

Diabetes-related complications occur frequently and are related to the degree of hyperglycaemia, therefore ongoing monitoring and follow-up is important. There is an increase in cardiovascular morbidity seen in HNF1A MODY. This is despite HDL levels being relatively high, which may appear falsely reassuring [2].

Non-Diabetes Features

HNF1A MODY can sometimes be differentiated from HNF4A MODY by the extrapancreatic features.

In HNF1A MODY decreased renal glucose reabsorption results in a low renal threshold for glucose, causing glycosuria at minimally raised blood glucose levels (<8 mmol/L). Glycosuria may be an early sign of glucose intolerance offering the potential for use as a screening tool in undiagnosed children with an affected parent [2].

In HNF4A MODY apolipoprotein synthesis is altered, resulting in a reduction in HDL, lipoprotein A1, and A2 [5]. Macrosomia also occurs in ~50% (average increased birth weight 800 g) and transient neonatal hypoglycaemia may be present (~15%) with diabetes presenting later in life [2, 7]. A history of marked macrosomia (birth weight >4.4 kg) or prolonged (>48 hours) hypoglycaemia in

Table 15.10.1.1 Comparison between HNF1A/HNF4A MODY and GCK MODY and type 1 and type 2 diabetes

	HNF1A/HNF4A MODY	GCK MODY	Type 1 Diabetes	Type 2 Diabetes
Age at diagnosis	Typically adolescence/early adulthood	When tested	Typically child or young adult	Typically middle or old age
Presenting features	Polyuria, polydipsia, weight loss but may be asymptomatic	Usually incidental Asymptomatic	Polyuria, polydipsia, weight loss	Polyuria, polydipsia but may be asymptomatic
Glycaemia	Progressive deterioration, HbA1c ≥6.5% (48 mmol/mol)	Stable mild elevated fasting hyperglycaemia (5.4–8.3 mmol/L) from birth, HbA1c 5.8–7.6% (40–60 mmol/mol)	Progressive deterioration HbA1C ≥6.5% (48 mmol/mol)	Progressive deterioration HbA1C ≥6.5% (48 mmol/mol)
Low C-peptide (<200 pmol/L)	No	No	Yes outside honeymoon	No
Islet autoantibodies	Negative	Negative	Positive (~80–90%)	Negative
Complications	Frequent microvascular and macrovascular complications HNF1A: Excess cardiovascular morbidity	Very rare microvascular complications—mild if occur, no monitoring required	Frequent microvascular and macrovascular complications	Frequent microvascular and macrovascular complications
Extrapancreatic features	HNF1A: Renal-glycosuria at lower renal threshold HNF1A liver-Raised HDL, adenoma HNF4A: liver-reduced TG, LDL, reduced HDL Macrosomia (56%), neonatal hypoglycaemia (15%)	Fetal size determined by maternal/fetal genotype	None	None
Treatment	Sulphonylurea sensitive 30% ultimately require insulin	Treatment usually neither effective nor needed Consider insulin during pregnancy	Insulin	Diet, non-insulin antidiabetic drugs, Insulin

an infant of a diabetic parent (mother or father) is an indication for sequencing *HNF4A* before the more common *HNF1A* gene [2, 7].

Differentiating HNF1A/HNF4A MODY from Type 1 and Type 2 Diabetes

More than 80% of MODY is not diagnosed in routine clinical practice due to overlapping clinical features with T1D and T2D [1]. The classical criteria (young-onset, family history and non-insulin treatment) only identifies 48% of patients with MODY. However, identifying patients with a high probability of MODY can be greatly improved by using a statistical calculator that takes into account multiple simple clinical factors [8]. The 'MODY Probability calculator' is available as a smart phone application (Diabetes diagnostic) and at https://www.diabetesgenes.org.

HNF1A/HNF4A MODY may be misdiagnosed as T1D due to the young age at symptomatic presentation.

However unlike T1D, in HNF1A/HNF4A MODY, endogenous insulin production (C-peptide >200 pmol/L) is persistent outside the honeymoon period [9] and islet autoantibodies are negative [10]. Therefore, presence of islet autoantibodies and/or C-peptide <200 pmol/L strongly confirms presence of T1D and can be used as a rule-out test for selecting patients for MODY genetic testing.

T2D may be misdiagnosed in patients who are clearly non-insulin dependent. In contrast to T2D, patients with HNF1A/HNF4A MODY are rarely obese, do not have acanthosis nigricans, and the metabolic syndrome is less common. 'MODY Probability calculator' helps to discriminate T2D and MODY [8]. Sulphonylurea therapy may cause hypoglycaemia in HNF1A/HNF4A MODY at the standard doses used in T2D.

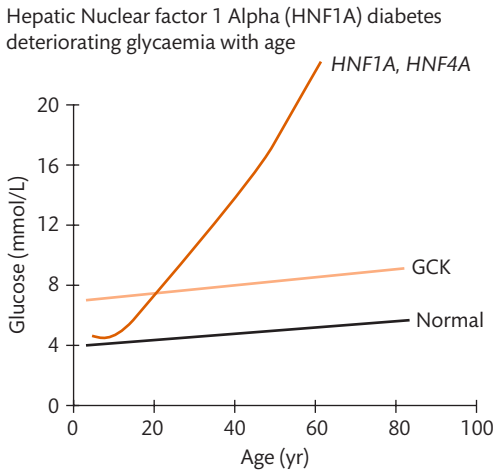


Figure 15.10.1.1 Effect of glycaemia on MODY genotype. In HNF1A/HNF4A MODY, glucose deteriorates with increasing age. In contrast, in GCK MODY, there is little change in glycaemia with age.

Genetics

Over 414 different heterozygous mutations have been reported in the *HNF1A* gene and ~103 have been reported in the *HNF4A* gene [11]. The commonest genetic abnormalities are missense mutations (55% in the *HNF1A* gene). Other mutations include nonsense, insertions, duplications, deletions, insertion/deletions, promoter, and splice site mutations.

Pathophysiology

HNF1A and *HNF4A* genes encode proteins which control gene expression during embryogenesis and in adult life. *HNF1A* and *HNF4A* are expressed in the pancreatic beta-cell and hepatocyte, regulating genes which encode proteins involved in beta-cell development and functioning. They result predominantly in progressive beta-cell failure [3].

Treatment

Patients are very sensitive to the glucose lowering effect of sulphonylurea therapy (five times more than metformin and four times more than in T2D) [3, 12]. This results from both increased insulin secretion and increased insulin sensitivity [12]. Starting doses should be low (e.g. 20–40 mg gliclazide daily) to avoid hypoglycaemia. Alternatively low-dose fast-acting novel agents like repaglinide may be used.

Increasing sulphonylurea doses are required to treat deteriorating glycaemia occurring as the patient gets older. Around 30–40% of patients progress to needing insulin treatment to control blood glucose values, so monitoring glycaemic control is essential. Patients who are misdiagnosed with T1D will be treated with insulin initially and even after several years on insulin the majority (79%) can still be successfully transferred onto sulphonylureas often with improvement in glycaemic control [13].

Due to the excess in cardiovascular morbidity, we recommend treating patients with HNF1A MODY with a statin after the age of 40 years even if HDL levels appear high [2].

GCK MODY

Clinical Phenotype

Diabetes

GCK MODY causes asymptomatic mild fasting hyperglycaemia (usually 5.4–8.3 mmol/L) from birth that deteriorates slightly with age. It is often detected incidentally and depending upon the time of detection can be misdiagnosed as gestational diabetes during routine antenatal screening, T1D or T2D [14].

In contrast to those with HNF1A/HNF4A MODY, patients always have fasting glucose >5.5 mmol/L and a small increment during an OGTT [6, 14]. This is because they have appropriate insulin secretion which is reset at a slightly higher glucose set point. This is also reflected in their HbA1c which ranges from 5.8 to 7.6% (40–60 mmol/mol) [14].

Differentiating GCK MODY from Type 1 and Type 2 Diabetes

A persistently raised mild fasting glucose without further deterioration in glycaemia strongly suggests GCK MODY especially when found in children or young adults (Figure 15.10.1.1). A dominant family history may not be identified unless fasting glucose is tested in asymptomatic family members. Diagnosis is worthwhile to avoid unnecessary treatment. Similar to HNF1A/HNF4A MODY, MODY probability calculator and C-peptide and islet autoantibodies can help to discriminate GCK MODY from T1D/T2D.

Genetics

Nearly 620 heterozygous inactivating mutations (missense, nonsense, frameshift, splice site) have been described throughout the 10 exons of the *GCK* gene [15].

Pathophysiology

GCK is a glycolytic enzyme catalysing the first step in glucose metabolism in the pancreatic beta-cell and hepatocyte (glucose to glucose-6-phosphate). Glucokinase is often referred to as the 'pancreatic glucose sensor' because this rate-determining step is key to detection of ambient glucose and maintaining euglycaemia through appropriate insulin secretion [14, 15]. A reduction in *GCK* activity decreases glucose phosphorylation, which raises the threshold concentration of glucose needed for insulin secretion.

Treatment

Treatment of hyperglycaemia is usually unnecessary outside pregnancy as microvascular complications are rare and glucose homeostasis is maintained [14, 16]. In addition as the fasting glucose level is regulated, low-dose insulin or oral agents do not alter HbA1c except when T2D is also present (HbA1c > 7.5%) [14].

Management in pregnancy depends on growth of the fetus. Growth can be predicted by parental/fetal inheritance. If the fetus and mother both have the mutation, they share a common higher glucose set point resulting in appropriate fetal growth. If there is only a maternal mutation, the fetus will increase its own insulin secretion in response to maternal hyperglycaemia and so there will be increased fetal growth. Reducing maternal hyperglycaemia may be difficult and large insulin doses (at least replacement doses) are often needed and the patient's counterregulation which is also set to the higher level will make achieving normal glucose values difficult. Early delivery may be required for a macrosomic baby. Conversely, a fetus who inherits a paternal mutation may be small for dates because of reduced fetal insulin secretion as the fetal set point is higher than its unaffected mother [17].

Less Common Subtypes of MODY

Other rarer causes of MODY include heterozygous mutations in *PDX1, NEUROD1, ABCC8, INS, KCNJ11*, and *RFX6*. Although *HNF1B* was initially described as a MODY gene and is a hepatic transcription factor it is discussed under monogenic diabetes syndromes since it typically presents with extrapancreatic (notably renal) disease [18].

Neonatal Diabetes

Overview

Neonatal diabetes is a rare (1:100 000 live births) genetically heterogeneous disorder in which diabetes is diagnosed within the first 6 months of life [19]. Around 85% of diabetes diagnosed within the first 6 months of life has a monogenic aetiology [20]. Patients present with diabetes that either persists (permanent neonatal diabetes, PNDM) or remits but frequently relapses later in life (transient

Table 15.10.1.2 Comparison of clinical features of the common genetic subgroups seen in PNDM and TNDM

	PNDM		TNDM	
Genes involved	K_{ATP} channel (~40%)	*INS* (~12%)	6q24 (70%)	K_{ATP} channel (~25%)
Age of presentation in weeks, median (range)	8 (0–40)	11 (0–1144)	Initial: 0.4 (0–4) Relapse:16 years (4–25 years)	Initial: 4 (0–16) Relapse: 5 years (3–15 years)
Birth weight in kg, median (range)	2.7 (1.5–4.2)	2.7 (1.7–3.9)	1.9 (1.6–2.7)	2.6 (1.4–3.6)
Sensitivity to sulphonylureas	85–90%	None	Usually successful on relapse	Usually successful on relapse

PNDM, permanent neonatal diabetes; TNDM, transient neonatal diabetes; K_{ATP} channel genes, *KCNJ11*, *ABCC8*.

neonatal diabetes, TNDM), in approximately equal proportions. There is some overlap in the genes involved. The underlying genetic aetiology determines both the clinical features and the treatment modalities. Those involving the ATP-sensitive potassium channel (K_{ATP} channel) should be treated with oral sulphonylureas. This makes early genetic diagnosis a priority. Diabetes may be present in isolation or associated with other features. Neonatal diabetes can also occur in babies born preterm and they should also have genetic testing. Comparison of the common subtypes of PNDM and TNDM are summarized in Table 15.10.1.2.

Permanent Neonatal Diabetes (PNDM)

Mutations in several genes have been reported giving rise to PNDM, causing abnormal pancreatic development, increased beta-cell destruction, or beta-cell dysfunction.

Common Causes

Activating (gain of function) mutations in *KCNJ11* and *ABCC8* genes account for around 40% of cases [20, 21]. They encode the subunits that make up the K_{ATP} channel of the beta-cell membrane; a 4:4 complex, with an inner pore-forming subunit Kir6.2 (*KCNJ11*) and outer regulatory subunit SUR1 (*ABCC8*). This channel is critical to maintaining glucose homeostasis through insulin secretion by the beta-cell; binding of ATP to Kir6.2 causes channel closure, membrane depolarization, and insulin secretion. The mutated K_{ATP} channel remains open and therefore unable to complete the cascade of steps culminating in insulin secretion [2, 22].

Most patients are born small (mean birthweight 2.7 kg), reflecting fetal insulin deficiency. The typical age of diagnosis is 8 weeks with marked hyperglycaemia (median glucose 33 mmol/L for those with *KCNJ11* mutations) and ketoacidosis with absent or very low C-peptide, all reflecting marked insulin deficiency. Except in a few rare cases, patients are diagnosed under the age of 6 months [21, 22].

The majority of patients have isolated diabetes, but neurological features may be seen in approximately 29% cases with *KCNJ11* mutations, and are less frequent and less marked in those with *ABCC8* mutations (20%) [20]. The neurological features result from a severely mutated K_{ATP} channel affecting the brain, muscle, and/or nerve. There are two patterns of abnormalities described. The DEND syndrome (severe Developmental delay, Epilepsy and Neonatal Diabetes) and the less severe phenotype, intermediate-DEND, in which epilepsy is less severe and less common, and the developmental delay is less marked [2, 19].

K_{ATP} channel mutations usually arise *de novo* (*KCNJ11*: ~80%; *ABCC8*: ~50%), but are typically dominantly inherited although some *ABCC8* mutations are recessive. There is a strong genotype-phenotype relationship for *KCNJ11* mutations although this is not absolute. For example, *KCNJ11* R201H mutations are typically seen in isolated diabetes, while patients with V59M typically have neurological features with the intermediate-DEND syndrome.

The critical reason to make a genetic diagnosis is that it will have a dramatic impact on treatment. Sulphonylureas close the mutated K_{ATP} channel causing membrane depolarization and insulin secretion. After stabilization on insulin, the majority of patients (85–90%) can be transferred on to sulphonylureas with improved glycaemic control, without increased hypoglycaemia that persists up to 10 years [22, 23]. Doses used are 3–4 times higher than those used in T2D. Higher doses (0.4–0.8 mg/kg/day) are typically needed in those with *KCNJ11* compared to *SUR1* mutations. (0.2–0.4 mg/kg)/day) [23]. Choice of sulphonylurea may be important in those who later develop neurological features. Glibenclamide crosses the blood–brain barrier and early treatment may reverse some neurological deficit in some cases [24].

Less Common Causes

Mutations in the *INS* gene, which encodes preproinsulin, accounts for 12% PNDM cases. They tend to be diagnosed later than those with K_{ATP} channel mutations (11 vs. 8 weeks) and require insulin treatment [21].

Rare

Rare causes of PNDM frequently involve extrapancreatic tissues and are listed in Table 15.10.1.3 [19, 20].

Transient Neonatal Diabetes (TNDM)

Commonest Causes

Approximately 70% cases are due to abnormalities in imprinting on chromosome 6q24, causing overexpression of a paternally expressed gene (*ZAC* and *HYMAI*) through different mechanisms; paternal uniparental disomy, paternal 6q24 duplication, and hypomethylation [25]. Of those with a methylation defect, some have been found due to a mutation on an upstream gene (*ZFP57*) which regulates methylation at 6q24 [25]. These patients have a global methylation problem which explains their other clinical features (congenital heart disease, developmental delay).

Irrespective of mechanism, the common phenotype is one of severe intrauterine growth retardation (IUGR) (mean birthweight

Table 15.10.1.3 Causes of neonatal diabetes

	Gene (protein) affected	PNDM/TNDM	Inheritance	Pancreatic agenesis	Common Extrapancreatic features (excluding low birth weight)
Common causes	K_{ATP} channel; *KCNJ11* (Kir6.2) and *ABCC8* (SUR1)	PNDM (50%) TNDM (25%)	AD (KCNJ11) AD/AR (ABCC8)	No	Neurological—developmental delay and epilepsy (29%)
	6q24 abnormality	TNDM (70%)	Variable	No	Macroglossia (30%) Umbilical hernia (9%)
	INS (insulin)	PNDM (12%)	AD (Coding mutation) / AR (promoter mutation)	No	Nil
Less common causes	*EIF2AK3* (eukaryotic translation initiation factor 2-alpha kinase 3)	PNDM	AR	No	Wolcott–Rallison syndrome (spondyloepiphyseal dysplasia, hepatitis, renal failure)
	FOXP3 (forkhead box protein P3)	PNDM	X-linked	No	IPEX syndrome (immunodysregulation, entero-, and endocrinopathy)
	GATA6 (GATA binding protein 6)	PNDM	AD	Yes	Congenital heart malformation, neurological defects, hypothyroidism, gut and hepato-biliary malformation, exocrine pancreatic dysfunction
	GCK (glucokinase)	PNDM	AR	No	Nil
	PTF1A (pancreas transcription factor 1 subunit alpha)	PNDM	AR	Yes	Distal enhancer mutation—isolated pancreatic agenesis; Coding mutation—cerebellar agenesis, microcephaly, exocrine pancreatic dysfunction
Rare causes	*GATA4* (GATA binding protein 4)	PNDM	AD	Yes	Congenital heart malformation, exocrine pancreatic dysfunction
	GLIS3 (GLIS family zinc finger 3)	PNDM	AR	No	Congenital hypothyroidism and glaucoma, liver fibrosis, cystic kidney disease
	HNF1B (hepatocyte nuclear factor 1-beta)	PNDM/TNDM	AD	Yes	Renal cysts, exocrine pancreatic dysfunction
	IER3IP1 (immediate early response 3-interacting protein 1)	PNDM	AR	No	Microcephaly, epilepsy
	LRBA (lipopolysaccharide-responsive and beige-like anchor protein)	PNDM	AR	No	Early onset multiple autoimmune features (e.g. hypothyroidism, autoimmune cytopenia, enteropathy), recurrent infections
	MNX1 (motor neuron and pancreas homeobox protein 1)	PNDM	AR	No	Sacral agenesis, neurological defects
	NEUROD1 (neurogenic differentiation factor 1)	PNDM	AR	No	Cerebellar hypoplasia, sensorineural deafness, visual impairment
	NEUROG3 (neurogenin-3)	PNDM	AR	No	Congenital malabsorptive diarrhoea
	NKX2-2 (homeobox protein Nkx-2.2)	PNDM	AR	No	Severe neurodevelopmental defects
	PDX1 (PANCREAS/ duodenum homeobox protein 1)	PNDM	AR	Yes	Exocrine pancreatic dysfunction
	RFX6 (regulatory factor X6)	PNDM	AR	No	Gall bladder aplasia, gut atresia
	SLC2A2 (solute carrier family 2, facilitated glucose transporter member 2)	PNDM/TNDM	AR	No	Hypergalactosaemia, liver failure
	STAT3 (signal transducer and activator of transcription 3)	PNDM	AD	No	Short stature, early onset multiple autoimmune features (e.g. hypothyroidism, autoimmune cytopenia, eczema), low IgE
	WFS1 (Wolframin)	PNDM	AD	No	Congenital sensorineural deafness, and congenital cataracts

AR, autosomal recessive; AD, autosomal dominant.

Table 15.10.1.4 Monogenic diabetes syndromes (excluding those typically presenting with neonatal diabetes)

A	Inheritance	Cause	Clinical phenotype (median age unless otherwise stated)	Distinguishing features
Wolfram syndrome	AR		DIDMOAD: cranial diabetes insipidus (73%), diabetes mellitus (6 years), optic atrophy (11years), bilateral sensorineural deafness (20 s; 62%) Other features: renal tract anomalies (30 s; 58%), cerebellar ataxia and myoclonus (40 s; 62%), gastrointestinal dysmotility (24%), primary gonadal atrophy, early death (30 years)	Childhood onset diabetes and optic atrophy
MIDD	Maternal	Mitochondrial DNA point mutations (commonly m.3243 A>G)	MIDD: Maternally Inherited Diabetes (slow onset, 20% acute, 8% ketoacidosis; 37 years; range 11–68): and deafness: sensorineural cochlear (75%; range 2–61 years; M>F) Variable phenotypes: MELAS, MERRF, CPEO, KSS, Leigh syndrome Other features: macular retinal dystrophy, myopathy, cardiac abnormalities with risk of premature death, renal disease, endocrine, gastrointestinal	Maternal inheritance Deafness frequently precedes diabetes Stroke <40 years (MELAS) Renal, ophthalmic, and neurological disease Ragged red fibres on muscle biopsy
Roger's syndrome	AR	*SLC19A2* mutations	Thiamine responsive megaloblastic anaemia, cardiac, neurological	Megaloblastic anaemia responds to thiamine, early onset deafness
RCAD	AD *de novo* 1/3	*HNF1B* gene deletions and mutations (encoding transcription factor HNF1B)	Renal developmental disorder (cysts;66%), diabetes; 58% (20 years; 15 days–61 years), genitourinary abnormalities (male 5%, female 14%), hyperuricaemia (20%), hypomagnesaemia (40%), abnormal liver function (13%)	Developmental renal disorder preceding diabetes
DPED syndrome	AD	*CEL* gene deletions	Diabetes and pancreatic exocrine function	Pancreatic exocrine dysfunction

*MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes), MERRF (myoclonic epilepsy associated with ragged red fibres), CPEO (chronic progressive external ophthalmoplegia), KSS (Kearns-Sayre syndrome), Leigh syndrome. *SLC19A2*, solute carrier family2, member 2 gene; CEL, carboxyl ester lipase.

1930 g), diabetes usually in the first weeks of life (median presentation 3 days), remission (typically at 3 months), and relapse in ~55% at around 14 years. Macroglossia (30%) and occasionally an umbilical hernia (9%) may also be present. The genetic mechanism predicts the heritability [25].

Treatment is initially with insulin but on relapse, patients are managed with sulphonylureas or other oral therapy [26].

Patients who have TNDM but do not have an abnormality at 6q24 (~25%) usually (~89%) have a gain of function mutation in one of the K_{ATP} channel genes (*KCNJ11, ABCC8*). In comparison to those with 6q24 abnormalities, patients with TNDM due to K_{ATP} channel mutations have less IUGR (mean birth weight 2.57 vs. 1.95 kg), and present later (4 vs. 0 weeks) indicating less severe insulin deficiency at birth. However, they also enter remission later (35 vs. 13 weeks), and relapse earlier (4.7 vs. 16 years). Similar to K_{ATP} channel mutations causing PNDM, neurological features may be present [27]. These TNDM patients on relapse are treated with low-dose sulphonylureas compared to high-dose sulphonylureas for PNDM patients.

Rare Causes of TNDM

Rare causes of TNDM include mutations in *INS, HNF1B, SLC2A2*, which may also cause PNDM.

Monogenic Syndromes Associated with Monogenic Beta-Cell Disorders

These are listed in Table 15.10.1.4 [2, 18, 28, 29].

Conclusions

There have been considerable advances in our understanding of the molecular genetic aetiology of beta-cell monogenic diabetes. Identifying the aetiological subtypes is crucial for appropriate patient management. Molecular genetics is now key in this area of clinical diabetes and not solely a research investigation.

REFERENCES

1. Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia*, 2010; **53**: 2504–8.
2. Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract*, 2008; **4**: 200–13.
3. Pearson ER, Velho G, Clark P, *et al.* beta-cell genes and diabetes: quantitative and qualitative differences in the pathophysiology of hepatic nuclear factor-1alpha and glucokinase mutations. *Diabetes*, 2001; **50** Suppl 1: S101–7.
4. Patel KA, Kettunen J, Laakso M, *et al.* Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance. *Nat Commun*, 2017; **8**: 888.
5. Pearson ER, Pruhova S, Tack CJ, *et al.* Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. *Diabetologia*, 2005; **48**: 878–85.
6. Stride A, Vaxillaire M, Tuomi T, *et al.* The genetic abnormality in the beta cell determines the response to an oral glucose load. *Diabetologia*, 2002; **45**: 427–35.

7. Pearson ER, Boj SF, Steele AM, *et al.* Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. *PLoS Med*, 2007; **4**: e118.
8. Shields BM, McDonald TJ, Ellard S, Campbell MJ, Hyde C, Hattersley AT. The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes. *Diabetologia*, 2012; **55**: 1265–72.
9. Besser RE, Shepherd MH, McDonald TJ, *et al.* Urinary C-peptide creatinine ratio is a practical outpatient tool for identifying hepatocyte nuclear factor 1-{alpha}/hepatocyte nuclear factor 4-{alpha} maturity-onset diabetes of the young from long-duration type 1 diabetes. *Diabetes Care*, 2011; **34**: 286–91.
10. McDonald TJ, Colclough K, Brown R, *et al.* Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from Type 1 diabetes. *Diabet Med*, 2011; **28**: 1028–33.
11. Colclough K, Bellanne-Chantelot C, Saint-Martin C, Flanagan SE, Ellard S. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia. *Hum Mutat*, 2013; **34**: 669–85.
12. Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT. Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet*, 2003; **362**: 1275–81.
13. Shepherd M, Shields B, Ellard S, Rubio-Cabezas O, Hattersley AT. A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients. *Diabet Med*, 2009; **26**: 437–41.
14. Chakera AJ, Steele AM, Gloyn AL, *et al.* Recognition and management of individuals with hyperglycemia because of a heterozygous glucokinase mutation. *Diabetes Care*, 2015; **38**: 1383–92.
15. Osbak KK, Colclough K, Saint-Martin C, *et al.* Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. *Hum Mutat*, 2009; **30**: 1512–26.
16. Steele AM, Shields BM, Wensley KJ, Colclough K, Ellard S, Hattersley AT. Prevalence of vascular complications among patients with glucokinase mutations and prolonged, mild hyperglycemia. *JAMA*, 2014; **311**: 279–86.
17. Spyer G, Macleod KM, Shepherd M, Ellard S, Hattersley AT. Pregnancy outcome in patients with raised blood glucose due to a heterozygous glucokinase gene mutation. *Diabet Med*, 2009; **26**: 14–18.
18. Clissold RL, Hamilton AJ, Hattersley AT, Ellard S, Bingham C. HNF1B-associated renal and extra-renal disease-an expanding clinical spectrum. *Nat Rev Nephrol*, 2015; **11**: 102–12.
19. Lemelman MB, Letourneau L, Greeley SAW. Neonatal diabetes mellitus: an update on diagnosis and management. *Clin Perinatol*, 2018; **45**: 41–59.
20. De Franco E, Flanagan SE, Houghton JA, Lango Allen H, Mackay DJ, Temple IK, *et al.* The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. *Lancet*, 2015; **386**: 957–63.
21. Edghill EL, Flanagan SE, Patch AM, *et al.* Insulin mutation screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood. *Diabetes*, 2008; **57**: 1034–42.
22. Pearson ER, Flechtner I, Njolstad PR, *et al.* Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med*, 2006; **355**: 467–77.
23. Rafiq M, Flanagan SE, Patch AM, Shields BM, Ellard S, Hattersley AT. Effective treatment with oral sulfonylureas in patients with diabetes due to sulfonylurea receptor 1 (SUR1) mutations. *Diabetes Care*, 2008; **31**: 204–9.
24. Beltrand J, Elie C, Busiah K, *et al.* Sulfonylurea therapy benefits neurological and psychomotor functions in patients with neonatal diabetes owing to potassium channel mutations. *Diabetes Care*, 2015; **38**: 2033–41.
25. Temple IK, Shield JP. 6q24 transient neonatal diabetes. *Rev Endocr Metab Disord*, 2010; **11**: 199–204.
26. Garcin L, Kariyawasam D, Busiah K, *et al.* Successful off-label sulfonylurea treatment of neonatal diabetes mellitus due to chromosome 6 abnormalities. *Pediatr Diabetes*, 2018; **19**: 663–9.
27. Flanagan SE, Patch AM, Mackay DJ, Edghill EL, Gloyn AL, Robinson D, *et al.* Mutations in ATP-sensitive K+ channel genes cause transient neonatal diabetes and permanent diabetes in childhood or adulthood. *Diabetes*, 2007; **56**: 1930–7.
28. Murphy R, Turnbull DM, Walker M, Hattersley AT. Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A>G mitochondrial point mutation. *Diabet Med*, 2008; **25**: 383–99.
29. Barrett TG. Differential diagnosis of type 1 diabetes: which genetic syndromes need to be considered? *Pediatr Diabetes*, 2007; **8** Suppl 6: 15–23.

15.10.2 Lipodystrophies and Severe Insulin Resistance Syndromes

Anna Stears, David B. Savage, and Stephen O'Rahilly

Introduction

The most common cause of insulin resistance, defined by impaired blood glucose lowering by insulin, is obesity. A minority of patients have severe insulin resistance without obesity. These patients, while not contributing significantly to the general prevalence of diabetes, often harbour pathogenic single gene defects affecting insulin signalling or adipose tissue function [1]. This chapter focuses on the resulting severe insulin resistance syndromes.

The normal response of the pancreatic beta-cells to insulin resistance is increased insulin secretion, and hyperglycaemia may not be present, especially earlier in life. Although clinical history and examination may offer clues to the presence of severe insulin resistance, laboratory confirmation should usually be sought. Measurement of insulin concentration is useful, although this is infrequent in clinical practice. Biochemical diagnostic thresholds for severe insulin resistance are arbitrary, and should, ideally, be interpreted in the context of BMI-adjusted normal ranges. However, one set of approximate diagnostic criteria is as follows:

- No diabetes and BMI under 30 kg/m^2 —fasting insulin above 150 pmol/L or peak insulin on oral glucose tolerance testing above 1500 pmol/L.
- Diabetes with absolute insulin deficiency and BMI under 30 kg/m^2 —exogenous insulin requirement above 3 units/kg/day.
- Diabetes with partial β cell decompensation and/or BMI over 30 kg/m^2—insulin levels are difficult to interpret in the context of obesity, while, in established diabetes, glucotoxicity, impaired islet function, and a combination of endogenous and exogenous insulin in the circulation confuse the biochemical picture. In this setting, the clinical history and features such as acanthosis nigricans are important in making a diagnosis of likely monogenic severe insulin resistance.

Causes of Severe Insulin Resistance

In addition to obesity, causes of severe insulin resistance can be categorized into:

1. Inherited or acquired defects in insulin signalling (including insulin receptoropathies)
2. Inherited or acquired impairment of adipocyte triglyceride storage (lipodystrophies)
3. Complex syndromes with associated severe insulin resistance

Insulin resistance of any cause may be exacerbated by puberty, pregnancy, weight gain, or intercurrent illness. Existing clinical nomenclature is variably applied and reflects historical descriptions of the syndromes: type A insulin resistance syndrome was named in the 1970s to discriminate it from the anti-insulin receptor antibody-mediated type B insulin resistance. It denotes acanthosis nigricans, hyperandrogenism, oligo- or amenorrhoea, and a BMI <30 kg/m^2. 'HAIR-AN' syndrome denotes 'HyperAndrogenism, Insulin Resistance, and Acanthosis Nigricans', and is identical to type A insulin resistance syndrome except that it is often used only in women with BMI >30 kg/m^2.

Prevalence of Severe Insulin Resistance

No clear prevalence figures exist for severe insulin resistance syndromes. A recent survey suggested a prevalence of all lipodystrophies across adjudicated electronic medical record databases of 3.07 cases per million (0.23 cases/million for generalized lipodystrophy and 2.84 cases/million for partial lipodystrophy) [2]. It is likely that many patients, especially males, remain undiagnosed. Experience over 5 years in the UK national specialist referral centre found that of 283 patients with confirmed severe insulin resistance and/or lipodystrophy, 14 (5%) had an insulin receptor (INSR) mutation, 1 (0.3%) had insulin-receptor antibodies, 13 (4.5%) had congenital generalized lipodystrophy, and 135 (47%) familial partial lipodystrophy (FPLD). Of the FPLD group 76 (26.5%) had FPLD1, 42 (14.6%) FPLD2 secondary to mutation of the *LMNA* gene, 17 (6%) FPLD3 secondary to mutation of the *PPARG* gene and 36 (12.5%) acquired lipodystrophy (4 generalized).

Generic Clinical Features of Severe Insulin Resistance Syndromes

Severe insulin resistance is often first noticed in patients with persistent hyperglycaemia despite very large doses of insulin. However many cases are unrecognized in the pre-diabetic phase. Indeed, a very common early feature of severe insulin resistance is spontaneous and symptomatic postprandial *hypo*glycaemia. This may dominate the clinical picture for years before hyperglycaemia supervenes, which only occurs in the face of β cell decompensation. The commonest presentation of severe insulin resistance is with acanthosis nigricans, which is nearly a *sine qua non* of all forms of severe insulin resistance. Another common presenting feature in females is hyperandrogenism (which may be severe) and oligo- or amenorrhoea. Partly for this reason, a large preponderance of presenting patients are female. Acanthosis nigricans and hyperandrogenism are common to all known forms of severe insulin resistance, but are not generally seen in insulin-deficient forms of diabetes. It is thus concluded that their pathogenesis depends upon severe hyperinsulinaemia, likely leading to ectopic activation of the insulin-like growth factor 1 (IGF-1) receptor, which is closely homologous to the insulin receptor. This mechanism, while not proven, seems plausible for components of the syndrome that feature cellular hyperproliferation, such as acanthosis nigricans and ovarian hyperthecosis [3].

Clinical Features Which Differentiate Causes of Severe Insulin Resistance

Clinical examination may reveal, in addition to acanthosis nigricans, a paucity or an unusual distribution of body fat, suggesting lipodystrophy. There may be other clinical features in patients with complex syndromes including short stature or signs of premature ageing. Age of onset of clinical features provides diagnostic clues. Fat loss is present from birth in the inherited generalized lipodystrophies, becomes most noticeable at puberty in the inherited partial lipodystrophies and tends to start with

the face and upper body in acquired lipodystrophies. Signs and symptoms often present in puberty in patients with heterozygous insulin-receptor mutations but earlier in childhood in patients with homozygous insulin-receptor mutations. The time course of onset is also important. For example, there is usually a history of rapid weight loss and onset of acanthosis nigricans and diabetes in patients with insulin receptor antibody-mediated type B insulin resistance.

Defects in Insulin Signalling Including Insulin Receptoropathies

Genetic defects in *INSR*, the gene encoding the insulin receptor, produce a spectrum of clinical syndromes: the most severe are autosomal recessive disorders with infant or childhood mortality. Although they form a continuum, the historical labels 'Donohue' or 'Rabson–Mendenhall' syndromes are still commonly used [4, 5]. Both feature characteristic dysmorphism and impaired linear growth. There is poor development of adipose tissue and muscle, which both rely on insulin-stimulated glucose uptake, contrasting with pseudoacromegaloid overgrowth of many other soft tissues. Hypertrichosis and exaggerated growth of androgen-dependent tissues may be particularly prominent. More difficult to diagnose clinically are the milder autosomal dominant insulin-receptor defects leading to type A insulin resistance, HAIR-AN (hyperandrogenism, insulin resistance, and acanthosis nigricans) syndrome, or their male equivalents. These are commonly caused by heterozygous intracellular mutations with dominant negative activity towards the coexpressed wild-type allele [6].

Traditionally, it has been difficult to discriminate patients with type A insulin resistance due to *INSR* mutations from those with other aetiologies for their insulin resistance. However, by exploiting the biochemical differences between receptoropathies and other forms of severe insulin resistance, it is now possible to undertake biochemical triage with a high degree of accuracy prior to genetic testing, expediting genetic diagnosis of affected patients and avoiding unnecessary and expensive sequencing of the large *INSR* gene. Plasma concentration of the adipose tissue-derived protein adiponectin is the most discriminating marker of receptoropathy. In receptoropathy, normal or elevated adiponectin levels are usually found in the face of extreme insulin resistance, unlike the suppressed levels seen in other insulin-resistant states including lipodystrophy [7] (**Table 15.10.2.1**). A further clinical clue to insulin receptoropathy is the absence of dyslipidaemia and hepatic steatosis despite severe insulin resistance [8]. This interesting observation has also led to the inference that this component of the common 'insulin resistance syndrome' actually reflects exposure of intact insulin signalling pathways in at least some tissues to high circulating insulin levels. A similar biochemical profile with normal or raised adiponectin levels and an absence of dyslipidaemia is seen in patients whose insulin resistance is due to mutations in the p85 alpha regulatory subunit of PI3 kinase (see SHORT syndrome, to follow). The biochemical profile of type B insulin resistance, due to anti-insulin-receptor antibodies, is also similar to that of the genetic insulin-receptor defects, but the onset is more acute, with rapid onset acanthosis nigricans and severe hyper and/or hypoglycaemia [9]. Type B insulin resistance is most often, but not exclusively, described in females of African origin.

Table 15.10.2.1 Comparison of biochemical features of insulin receptoropathy with other forms of severe insulin resistance

	Prevalent insulin resistance/ lipodystrophy	Insulin receptoropathy
Insulin	↑↑	↑↑-↑↑↑
Glucose	↑/→/↓	↑/→/↓
Triglyceride	↑-↑↑↑	→
High-density lipoprotein (HDL) cholesterol	↓	→
Adiponectin	↓	↑-↑↑↑
IGFBP-1	↓	→-↑↑
SHBG	↓	→-↑↑

Primary Lipodystrophic Syndromes

Lipodystrophies are a heterogeneous group of conditions characterized by partial or complete absence of adipose tissue [10]. They may be genetic or acquired, and are further classified according to the anatomical distribution of the lipodystrophy [10]. White adipose tissue is required for the storage of excess energy as triglyceride in lipid droplets. Failure to store triglyceride in subcutaneous adipose tissue leads to alternative storage in liver and muscle tissue causing impaired insulin action in these tissues [11]. Insulin resistance therefore is a feature of most, but not all lipodystrophies. As with all forms of insulin resistance, the clinical expression in adults is more pronounced in women.

Congenital Generalized Lipodystrophy

Congenital generalized lipodystrophy (CGL), also known as Berardinelli–Seip congenital lipodystrophy (BSCL) [12], is an autosomal recessive condition characterized by a generalized absence of adipose tissue from birth, increased appetite due to leptin deficiency, accelerated growth, and advanced bone age. Skeletal muscles, peripheral veins, and the thyroid gland are prominent due in part to paucity of subcutaneous fat. Hyperinsulinaemia from early childhood leads to organomegaly, acromegaloid features, and acanthosis nigricans. Diabetes tends to develop in the second decade. Hepatomegaly is often prominent and caused by severe non-alcoholic fatty liver disease (NAFLD), which generally progresses to non-alcoholic steatohepatitis (NASH), and, often, eventually to cirrhosis. Severe hypertriglyceridaemia, eruptive xanthomata, and pancreatitis are common. In the vast majority of cases, CGL is due to biallelic mutations in either the gene encoding 1-acylglycerol-3-phosphate O-acyltransferase 2 (AGPAT2), or the gene-encoding seipin (*BSCL2*), an endoplasmic reticulum protein. AGPAT2 is an essential enzyme in glycerophospholipid and triacylglycerol synthesis. The mechanistic link between seipin and lipodystrophy is more obscure, although seipin has been shown

to be highly expressed in white adipose tissue and to be essential for adipogenesis *in vitro* [13]. It is currently not possible to distinguish clinically between these genetic subgroups with confidence. However, adipose tissue loss in mechanical fat pads, such as the palms, soles, orbits, scalp, and periarticular regions, has been reported as a specific feature of CGL due to seipin mutations.

Familial Partial Lipodystrophies

Familial partial lipodystrophies (FPLD) are both milder and more common than CGL. Indeed, patients with these conditions may exhibit normal, or even increased, whole-body adipose stores. Consequently, indices such as the BMI have little utility in diagnosing FPLD. The abnormality instead lies in the adipose tissue topography, or fat distribution. In particular, head and neck adipose depots are often preserved or increased. Without a thorough clinical examination, it is thus possible to form the erroneous impression of normal or increased adiposity [10]. These disorders most commonly present in peri- or postpubertal women, and the loss of femorogluteal fat is particularly striking. They are difficult to detect clinically in men and in prepubertal children. Metabolic abnormalities range from asymptomatic impaired glucose tolerance and mild dyslipidaemia to severe insulin resistance with type 2 diabetes mellitus and severe dyslipidaemia, eruptive xanthomata, and pancreatitis. NAFLD/ NASH is also common.

Familial Partial Lipodystrophy Type 1 (FPLD1)

FPLD1 (Köbberling syndrome) is characterized by loss of limb fat with preserved (and, frequently, increased) truncal fat, in a pattern reminiscent of that seen in Cushing's syndrome [14]. While some of these patients do have affected family members, many do not, suggesting that not all cases are inherited. No specific genetic defects have been reported in this group and recently it has been noted that patients with FPLD1 are significantly enriched for SNPs associated with increased fasting insulin and lower gluteofemoral fat deposition from 53 distinct genomic regions, suggesting polygenic inheritance [15, 16].

Familial Partial Lipodystrophy Type 2 (FPLD2)

FPLD2 (Dunnigan syndrome) predominantly affects the limbs and gluteal fat depots with variable truncal involvement, but with normal or excess fat on the face, neck, and in the labia majora [17]. A majority of patients with FPLD2 have heterozygous loss-of-function mutations in LMNA, encoding lamin A/C, a structural component of the nuclear lamina which is nearly ubiquitously expressed. Mutations in this gene have also been convincingly linked to several different disorders, including muscular dystrophy and dilated cardiomyopathy. Detailed understanding of the mechanisms underlying the tissue-selective phenotypes associated with LMNA mutations is lacking, but proposed abnormalities include structural defects in the nuclear envelope and altered binding of the nuclear lamina to chromatin or transcription factors.

Familial Partial Lipodystrophy Type 3 (FPLD3)

FPLD3 also features a paucity of limb and gluteal fat; however, abdominal fat is generally preserved and may even be increased, and facial fat is most commonly normal. Insulin resistance and lipodystrophy have been described in prepubertal children, although peripubertal presentation is most common. The very high prevalence of early onset hypertension helps to discriminate FPLD3 from FPLD2. Loss-of-function mutations in the gene-encoding peroxisome proliferator-activated receptor γ (PPARγ)—a nuclear hormone receptor critically required for adipose tissue development and targeted by thiazolidinedione insulin sensitizers—have been described in many patients with FPLD3. All pathogenic mutations described to date have been heterozygous, located in the DNA or ligand-binding domains of the protein, and many display dominant negative activity *in vitro* [18].

Familial Partial Lipodystrophies Types 4–6

FPLD4 (*PLIN1*), FPLD5 (*CIDEC*), and FPLD6 (LIPE) are very rare subtypes of partial lipodystrophy. PLIN1 and CIDEC are involved in triglyceride storage within lipid droplets and are almost exclusively expressed in adipocytes. Rare cases associated with mutations in *AKT2* and *PCYT1A* have also been described.

Acquired Lipodystrophies

Acquired partial lipodystrophies are often associated with autoimmune disorders. The best described is Barraquer–Simons syndrome, with cephalocaudal loss of subcutaneous fat, complement C3 deficiency, C3 nephritic factor and mesangioproliferative glomerulonephritis [19]. Barraquer–Simons is not usually associated with metabolic disturbance due to the sparing of gluteofemoral fat stores. Acquired partial lipodystrophy is increasingly being described following childhood total body irradiation and/or chemotherapy and can be associated with metabolic disturbances including diabetes and NAFLD [20]. 'Late effects' screening should include metabolic screening and clinical examination for lipodystrophy and acanthosis nigricans. Acquired generalized lipodystrophies are very rare and may be associated with low C4 complement levels. Metabolic sequelae are usually severe [21].

Complex Syndromes Featuring Severe Insulin Resistance

There is group of syndromes that exhibit severe insulin resistance disproportionate to total fat mass as part of a more generalized disorder. These include Alström syndrome, primordial dwarfism, mandibuloacral dysplasia, and some forms of progeria. Often in these conditions, acanthosis nigricans is the key clinical clue, and it is important that its clinical significance is recognized and that appropriate endocrine assessment is undertaken at an early stage of evaluation. Some of these complex syndromes are associated with dyslipidaemia and hepatic steatosis but this is not the case in

SHORT syndrome, which arises due to a loss-of-function mutation in the *PIK3R1* gene. SHORT syndrome is also associated with a raised adiponectin concentration.

Principles of Management of Severe Insulin Resistance

Management of severe insulin resistance involves early, intensive use of insulin-sensitizing agents (metformin), and lifestyle modification to include a low-calorie, low-fat diet, and as much aerobic exercise as reasonably possible. Where postprandial hypoglycaemia is symptomatic, a low glycaemic index diet, corn starch/glycoside and acarbose may be efficacious. Dietary measures are particularly important in lipodystrophy, where they are crucial in preventing or improving dyslipidaemia and diabetes [22]. Some patients with lipodystrophy have benefitted from 'metabolic surgery' despite a normal body mass index, with remission of diabetes and reduction in serum triglycerides [23]. Adjunctive use of subcutaneous leptin in patients who have secondary leptin deficiency due to lack of adipose tissue has proved highly effective, especially in patients with generalized lipodystrophy [24]. Recombinant IGF1 or composite preparations have some utility in severe insulin resistance [25]. Clinical management should be introduced based on clinical and biochemical criteria, and establishment of the genetic defect should not influence therapeutic decision making significantly. The FPLDs are minor exceptions to this: it is logical to suppose that use of potent thiazolidinedione PPARγ agonists in patients with *PPARG* mutations may be particularly efficacious; however, despite some evidence for this in the case of particular mutations and novel agonists, this requires further study. Type B insulin resistance can be effectively treated with immunosuppressive regimens [9]. Many patients with severe insulin resistance require very high insulin doses and the use of concentrated insulin formulations and continuous subcutaneous insulin pump therapy is common. The incidence of early microvascular and macrovascular complications is high, especially in patients with lipodystrophy, and close attention should be paid to minimizing cardiovascular risk factors [10, 26]. The psychological burden of severe insulin resistance syndromes can be high, especially when there is severe hyperandrogenism and in patients with lipodystrophy, largely due altered body image [27], but also due to the burden of early complications.

REFERENCES

1. Melvin A, O'Rahilly S, Savage DB. Genetic syndromes of severe insulin resistance. *Curr Opin Genet Dev*, 2018; **50**: 60–7.
2. Chiquette E, Oral EA, Garg A, Araújo-Vilar D, Dhankhar P. Estimating the prevalence of generalized and partial lipodystrophy: findings and challenges. *Diabetes Metab Syndr Obes*, 2017; **10**: 375–83.
3. Cruz PD, Hud JA. Excess insulin binding to insulin-like growth factor receptors: proposed mechanism for acanthosis nigricans. *J Invest Dermatol*, 1992; **98**(6 Suppl), 82S–5S.
4. Donohue WL, Uchida I. Leprechaunism: a euphemism for a rare familial disorder. *J Pediatr*, 1954; **45**: 505–19.
5. Rabson SM, Mendenhall EN. Familial hypertrophy of pineal body, hyperplasia of adrenal cortex and diabetes mellitus; report of 3 cases. *Am J Clin Pathol*, 1956; **26**: 283–90.
6. Taylor SI, Cama A, Accili D, *et al*. Mutations in the insulin receptor gene. *Endocr Rev*, 1992; **13**: 566–95.
7. Semple RK, Cochran EK, Soos MA, *et al*. Plasma adiponectin as a marker of insulin receptor dysfunction: clinical utility in severe insulin resistance. *Diabetes Care*, **31**(5), 977–79 (2008).
8. Semple RK, Sleigh A, Murgatroyd PR, *et al*. Postreceptor insulin resistance contributes to human dyslipidemia and hepatic steatosis. *J Clin Invest*, 2009; **119**: 315–22.
9. Brown RJ, Joseph J, Cochran E, Gewert C, Semple R, Gorden P. Type B insulin resistance masquerading as ovarian hyperthecosis. *J Clin Endocrinol Metab*, 2017; **102**: 1789–91.
10. Brown RJ, Araújo-Vilar D, Cheung PT, *et al*. The diagnosis and management of lipodystrophy syndromes: a multi-society practice guideline. *J Clin Endocrinol Metab*, 2016; **101**: 4500–11.
11. Vidal-Puig A. Adipose tissue expandability, lipotoxicity and the metabolic syndrome. *Endocrinol Nutr*, 2013; **60** Suppl 1: 39–43.
12. Seip M, Trygstad O. Generalized lipodystrophy, congenital and acquired (lipoatrophy). *Acta Paediatr Suppl*, 1996; **413**: 2–28.
13. Payne VA, Grimsey N, Tuthill A, *et al*. The human lipodystrophy gene BSCL2/seipin may be essential for normal adipocyte differentiation. *Diabetes*, 2008; **57**: 2055–60.
14. Guillín-Amarelle C, Sánchez-Iglesias S, Castro-Pais A, *et al*. Type 1 familial partial lipodystrophy: understanding the Köbberling syndrome. *Endocrine*, 2016; **54**: 411–21.
15. Lotta LA, Gulati P, Day FR, *et al*. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet*, 2017; **49**: 17–26.
16. Scott RA, Fall T, Pasko D, *et al*. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes*, 2014; **63**: 4378–87.
17. Jackson SN, Howlett TA, McNally PG, O'Rahilly S, Trembath RC. Dunnigan-Köbberling syndrome: an autosomal dominant form of partial lipodystrophy. *QJM*, 1997; **90**: 27–36.
18. Semple RK, Chatterjee VK, O'Rahilly S. PPAR gamma and human metabolic disease. *J Clin Invest*, 2006; **116**: 581–9.
19. Misra A, Peethambaram A, Garg A. Clinical features and metabolic and autoimmune derangements in acquired partial lipodystrophy: report of 35 cases and review of the literature. *Medicine (Baltimore)*, 2004; **83**: 18–34.
20. Adachi M, Oto Y, Muroya K, Hanakawa J, Asakura Y, Goto H. Partial lipodystrophy in patients who have undergone hematopoietic stem cell transplantation during childhood: an institutional cross-sectional survey. *Clin Pediatr Endocrinol*, 2017; **26**: 99–108.
21. Misra A, Garg A. Clinical features and metabolic derangements in acquired generalized lipodystrophy: case reports and review of the literature. *Medicine (Baltimore)*, 2003; **82**: 129–46.
22. Stears A, Flanagan C. Diagnosis and management of lipodystrophy: a practical update. Clin Lipidol, 2014; **9**: 235–59.
23. Melvin A, Adams C, Flanagan C, *et al*. Roux-en-Y gastric bypass surgery in the management of familial partial lipodystrophy type 1. *J Clin Endocrinol Metab*, 2017; **102**: 3616–20.
24. Chan JL, Lutz K, Cochran E, *et al*. Clinical effects of long-term metreleptin treatment in patients with lipodystrophy. *Endocr Pract*, 2011; **17**: 922–32.

25. Regan FM, Williams RM, McDonald A, *et al*. Treatment with recombinant human insulin-like growth factor (rhIGF)-I/rhIGF binding protein-3 complex improves metabolic control in subjects with severe insulin resistance. *J Clin Endocrinol Metab*, (United States, 2010) 2113–22.
26. Bidault G, Garcia M, Vantyghem MC, *et al*. Lipodystrophy-linked LMNA p.R482W mutation induces clinical early atherosclerosis and *in vitro* endothelial dysfunction. *Arterioscler Thromb Vasc Biol*, 2013; **33**: 2162–71.
27. Adams C, Stears A, Savage D, Deaton C. 'We're stuck with what we've got': the impact of lipodystrophy on body image. *J Clin Nurs*, 2018; **27**: 1958–68.

15.10.3 Diabetes Secondary to Pancreatic Disease

Philip J. Weston

Cystic Fibrosis-Related Diabetes

Cystic fibrosis (CF) is an autosomal recessive disorder caused by gene mutations in the cystic fibrosis transmembrane regulator (CFTR) gene. CFTR is a chloride channel responsible for regulation of salt and water across cell walls with mutations leading to increased viscosity of secretions resulting in obstruction and fibrosis of organs including the pancreas [1].

Survival rates for CF patients have increased but this has been associated with the development of complications, the commonest of which is CF related diabetes (CFRD). CFRD occurs in 20% of adolescents and up to 50% of adults with CF. CFRD is associated with female gender, increasing age, exocrine pancreatic insufficiency, and corticosteroid usage. The median age for diagnosis of CFRD is between 18 and 21 years. CFRD has a significant impact on survival rates. Female CF patients with normal glucose tolerance survive up to 16 years longer than those with CFRD. Furthermore, CFRD is associated with reduced pulmonary function, an increase in acute exacerbations of pulmonary disease and poor nutritional status [2]. Effective treatment of CFRD improves pulmonary function [3].

CF leads to exocrine pancreatic failure in up to 85% of patients. Initially beta-cell function is preserved but there is progressive loss of beta- and alpha-cell function. Initially, hyperglycaemia may be seen only at times of acute infection or when the patient is taking corticosteroids (due to associated insulin resistance). As beta-cell function reduces further postprandial hyperglycaemia (but normal fasting glucose) and finally CFRD with fasting hyperglycaemia occurs [4].

CF patients should be screened annually from 10 years of age for CFRD [5]. The American Diabetes Association (ADA) does not recommend HbA1c for screening as it is unlikely to detect the early stages of glucose intolerance that are associated with declining pulmonary and nutritional status of the patient [5]. In addition, HbA1c can be spuriously low in patients with CF owing to increased red cell turnover. The gold standard test for CFRD remains the oral glucose tolerance test (OGTT). The diagnostic criteria for diagnosing impaired glucose regulation in CF patients are the same as in the general population but if pulmonary function is adversely affected the diagnosis of clinically significant glucose dysregulation may need to be individualized. The conventional 2-hour post glucose load OGTT may fail to detect relevant glucose rises between the glucose ingestion and the 2-hour sample. For this reason, in a person with CF who has no symptoms of diabetes and whose respiratory function is stable, an abnormal OGTT should be followed by a period of home blood glucose monitoring. Increasingly continuous glucose monitoring is used in such patients allowing for detailed examination of the abnormality in glucose regulation and tailoring of insulin therapy [4, 5].

Patients with CFRD rarely develop macrovascular disease but patients are now surviving long enough to develop microvascular complications.

The aim of treating CFRD is to maintain/improve lung function and nutritional status as well as reducing the risk of developing microvascular complications. Insulin treatment remains the mainstay and is associated with improved pulmonary function and an increase in body weight [4–6]. For this reason insulin may be initiated in patients with CF and impaired glucose tolerance who are losing weight or who have a persistent decline in their lung function.

To maintain nutrition CFRD patients need an increased calorie intake compared to the non-CF population and dietary carbohydrate is recommended to comprise 40–60% of the total energy intake [4]. While recommendations are for low glycaemic index and high fibre foods, in reality many CFRD patients, particularly those with poor nutritional status, rely on refined sugary food taken in regular quantities.

Insulin regimens need to be individualized based on the patient's home blood glucose monitoring but the majority of patients with CFRD and fasting hyperglycaemia require a multiple dose regimen including a basal insulin. Balancing exocrine replacement therapy with nutritional intake and exogenous insulin replacement is complex and patient education as well as regular review by a multidisciplinary team is essential to ensure adequate nutrition, attain diabetes control, and maintain respiratory function.

Pancreatitis

Acute pancreatitis, most commonly due to gallstones, is associated with transient hyperglycaemia in up to 50% of cases [7]. The severity of hyperglycaemia is a prognostic indicator in patients with acute pancreatitis and is included in a number of disease severity scoring systems. Persistent hyperglycaemia after the acute episode is less common and is associated with the severity of the pancreatitis, obesity, male gender, older age, the presence of chronic liver

disease, hypertension, and a history of alcohol intoxication [8]. If surgical intervention is required for the acute episode of pancreatitis then the risk of diabetes is increased. Glucose tolerance should be assessed in all patients 3–6 months after an episode of acute pancreatitis.

Chronic pancreatitis is an inflammatory disease of the pancreas initially affecting the exocrine pancreas which can, over time, impact on endocrine pancreatic function [9]. The chronic inflammation results in fibrosis and in most cases later calcification. In the United Kingdom most cases chronic pancreatitis are associated with excess alcohol or are idiopathic. It is estimated that up to 60% of patients with calcific pancreatitis develop diabetes. Chronic pancreatitis, especially hereditary pancreatitis, is linked to the risk of developing pancreatic cancer.

Fibrocalculous pancreatic diabetes is due to a form of chronic pancreatitis seen only in developing countries and is not associated with alcohol excess [10]. It affects young adults and is characterized by abdominal pain due to intraductal pancreatic calculi and the development of exocrine and later endocrine pancreatic failure. The aetiology remains uncertain, but micronutrient deficiencies or possible toxins have been proposed. Despite the degree of pancreatic fibrosis, ketoacidosis is rare in those affected. There is a strong association with pancreatic cancer.

Patients with chronic pancreatitis from any cause who develop diabetes are at risk of developing diabetes complications.

The treatment of diabetes secondary to chronic pancreatitis is complex as patients are susceptible to greater glucose variability and are more prone to iatrogenic hypoglycaemia given associated hypoglucagonaemia which results in reduced hepatic gluconeogenesis. This is often compounded by poor nutrition and malabsorption associated with failure of the exocrine pancreas and the limitations of pancreatic exocrine replacement therapy [11]. Continued alcohol excess can lead to further glycaemic instability.

Oral glucose lowering agents may be tried initially provided there is adequate beta-cell reserve but the choice of agents is limited. These patients are often malnourished and underweight so metformin is not favoured. With the continued uncertainty around the risk of pancreatitis with incretin-mimetic agents and DPP4 inhibitors they are best avoided in this group of patients. Similarly, SGLT2 inhibitors should be avoided in view of the possibility of progressive insulin deficiency and risk of developing normoglycaemic ketoacidosis. Sulphonylureas may be effective but long-acting sulphonylureas should be used with caution given the increased risk of hypoglycaemia. Insulin remains the mainstay of treatment for these patients. This may initially be with prandial fast acting insulin but as beta-cell function declines multiple daily insulin doses are needed.

Pancreatic Cancer

Pancreatic cancer is a devastating disease and the fourth commonest cause of cancer-related mortality in men in the United States [12]. The relationship between pancreatic cancer and diabetes is complex and two-way. Patients with pancreatic cancer develop diabetes, with up to 50% of patients with a new diagnosis of pancreatic cancer having associated hyperglycaemia. Conversely, patients with type 2 diabetes are at increased risk of developing pancreatic cancer [13].

The pathophysiology of developing diabetes with pancreatic cancer is far from clear. Although reduced beta-cell mass secondary to tumour-related pancreatitis and fibrosis has been proposed, islet loss may not be sufficient to fully account for pancreas cancer associated diabetes. Other possible factors include a shared association with obesity and insulin resistance, impairment of incretin secretion or earlier loss of beta-cells compared to alpha cells [14].

Haemochromatosis

Primary haemochromatosis is an autosomal recessive condition associated with the *HFE* gene. Patients with primary haemochromatosis chronically absorb iron in excess from the small bowel and the excess iron is deposited in a number of organs, including the pancreas. Identification of the HFE gene has allowed the recognition of milder forms of the disease so the incidence of associated diabetes has fallen over the years from 75% of cases to around 15% dependent on the degree of iron overload [15]. Diabetes is caused by progressive pancreatic fibrosis and other factors and is often present at the diagnosis of severe iron overload.

Diabetes in patients with haemochromatosis is associated with micro- and macrovascular complications and most patients require insulin therapy. Regular venesection to reduce the iron overload is associated with improvements in glycaemic control.

Secondary haemochromatosis occurring in patients with haematological conditions requiring frequent blood transfusions is also associated with diabetes.

Diabetes secondary to pancreatic disease is commonly seen and can present complex management problems which are best managed by a multidisciplinary team. Treating the underlying condition as well as the associated hyperglycaemia is essential.

REFERENCES

1. O'Shea D, O'Connell J. Cystic fibrosis related diabetes. *Curr Diab Rep*, 2014; **14**: 511.
2. Milla CE, Warwick WJ, Moran A. Trends in pulmonary function in patients with cystic fibrosis correlate with the degree of glucose intolerance at baseline. *Am J Respir Crit Care Med*, 2000; **162**(3 Pt 1): 891–5.
3. Kayani K, Mohammed R, Mohiaddin H. Cystic fibrosis-related diabetes. *Front Endocrinol (Lausanne)*, 2018; **9**: 20
4. Cystic Fibrosis Trust. Consensus documents. Available at: http://www.cysticfibrosis.org.uk/the-work-we-do/clinical-care/consensus-documents (accessed: 25 April 2020).
5. Moran A, Brunzell C, Katz M, *et al.* The CFRD guidelines committee. Clinical care guidelines for cystic fibrosis-related diabetes. *Diabetes Care*, 2010; **33**: 2697–708.
6. Moheet A, Moran A. Pharmacological management of cystic fibrosis related diabetes. *Exp Rev Clin Pharmacol*, 2018; **11**: 185–91.
7. Kikuta K, Masamune A, Shimosegawa T. Impaired glucose tolerance in acute pancreatitis. *World J Gastroenterol*, 2015; **21**: 7367–74.
8. Lee Y-K, Huang M-Y, Hsu C-Y, Su Y-C. Bidirectional relationship between diabetes and acute pancreatitis. *Medicine*, 2016; **95**: 1–8.
9. Braganza J, Lee S, McCloy R, McMahon M. Chronic pancreatitis. *Lancet*, 2011; **377**: 1184–97.

10. Mohan V, Premalatha G, Pitchumoni CS. Tropical chronic pancreatitis: an update. *J Clinic Gastroenterol*, 2003; **36**: 337–46.
11. National Institute for Health and Care Excellence. Pancreatitis: diagnosis and management. NICE guideline NG104. September 2018. Available at: https://www.nice.org.uk/guidance/ng104 (accessed: 24 April 2020).
12. Kamisawa T, Wood L, Itoi T, Takaori K. Pancreatic cancer. *Lancet*, 2016: **388**: 73–85.
13. Song S, Wang B, Zhang X, *et al.* Long-term diabetes mellitus is associated with an increased risk of pancreatic cancer: a meta-analysis. *PLoS One*, 2015; **10**: e0134321.
14. Andersen DK, Korc M, Petersen GM, *et al.* Diabetes, pancreatogenic diabetes and pancreatic cancer. *Diabetes*, 2017; **66**: 1103–10
15. Barton J, Acton R. Diabetes in HFE haemochromatosis. *J Diabetes Res*, 2017; **2017**: 9826930.

15.10.4 Diabetes Secondary to Endocrine Disorders

Jeremy W. Tomlinson

Acromegaly

Acromegaly is a rare condition usually caused by a pituitary somatotroph adenoma. It is estimated that up to 40% of patients with acromegaly may develop diabetes and that 20% of patients develop diabetes prior to the diagnosis of acromegaly being made. Diabetes is most likely to occur in those individuals with higher growth hormone (GH) and IGF-1 levels, older individuals, those that have had the condition untreated for longer and also where there is a family history of type 2 diabetes. The development of diabetes in the context of acromegaly may also be associated with an adverse mortality outcome. Diabetes is often diagnosed as part of the diagnostic work-up during an oral glucose tolerance test when glucose excursions can be measured alongside assessing the ability of a glucose load to suppress GH levels. Formal assessment of glucose tolerance should be made in all patients with acromegaly.

Insulin resistance, despite the relatively lean phenotype of patients with acromegaly, is thought to be the major underlying pathophysiological mechanism driving the development of diabetes. Fasting insulin levels are usually elevated and there is impaired glucose disposal. In addition, GH is able to stimulate gluconeogenesis further contributing to elevations in blood glucose. Insulin deficiency and diabetic ketoacidosis are rarely seen [1].

The aims of treating diabetes in patients with acromegaly are not different to those of treating type 2 diabetes in isolation. Normoglycaemia, with restoration of normal insulin and C-peptide levels, can be achieved with normalization of IGF-1 and GH levels following pituitary surgery and occurs in up to 60% of cases.

The use of somatostatin analogues, either as primary therapy or postoperatively, can both improve insulin sensitivity (as a result of lowering IGF-1 and GH levels) but these also impair insulin secretion. As a consequence, they can have a variable impact on glucose homeostasis, but are often reported to have no overall impact on glucose tolerance [2]. However, close monitoring of glycaemic control is advocated. Second-generation somatostatin analogues which possess broader somatostatin receptor activity have a more detrimental impact on glycaemic control. This is thought to arise as a result of a more profound inhibitory action impairing insulin release within the pancreatic islet. In patients who respond clinically to somatostatin analogues, but who have coexistent diabetes, management should be according to current type 2 diabetes consensus guidelines. Agents which stimulate insulin secretion have been used effectively as have GLP-1 analogues, insulin, and insulin analogue therapy. There are limited clinical data available, but the novel GH receptor ligand, pegvisomant, does not appear to worsen, and actually may improve glycaemic control.

Cushing's Syndrome

Cushing's syndrome is defined by circulating glucocorticoid excess due to a pituitary tumour secreting adrenocorticotropic hormone (ACTH) (Cushing's disease); secondary to ectopic ACTH secretion; from an adrenal adenoma or rarely carcinoma. However, the commonest cause of Cushing's syndrome is medically prescribed glucocorticoids; currently it is estimated that 2–3% of the population are prescribed glucocorticoid therapy.

The prevalence of altered glucose homeostasis in Cushing's syndrome is as high as 90%. Overt type 2 diabetes may be present in almost 50% of cases and glucose intolerance in up to 30% [3]. Fasting glucose is more variable, and can be normal in those with diabetes, notably in those with iatrogenic Cushing's syndrome due to prescribed glucocorticoids. Diabetes is more commonly seen in patients with ectopic ACTH secretion who often present with severe and rapidly progressive disease, although the correlations between circulating glucocorticoid levels and abnormalities of glucose homeostasis are variable, probably reflecting interindividual variability in the susceptibility to the adverse effects of glucocorticoids.

The mechanisms underpinning glucocorticoid-induced diabetes and impaired glucose tolerance are complex, but hepatic and peripheral insulin resistance are thought to be crucial. Fasting glucose and insulin levels are elevated in comparison with age-, sex-, and BMI-matched controls. Insulin resistance is thought to be driven, at least in part, through elevated circulating free fatty acid levels which inhibit skeletal muscle and hepatic insulin signalling. In addition, elevated circulating glucocorticoid levels can impair insulin secretion.

Although surgical and pharmacological treatment of Cushing's syndrome to decrease circulating glucocorticoid levels are associated with improvements in glucose homeostasis, the impact of Cushing's syndrome on glucose metabolism can persist for long after the normalization of circulating glucocorticoid levels. As with acromegaly, the second-generation somatostatin analogues with broad receptor affinity can worsen glycaemic control in patients with Cushing's disease and therefore close monitoring of glycaemic control in these patients is needed [4].

There are currently no established guidelines for the treatment of diabetes associated with glucocorticoid excess. All treatment modalities including insulin therapy, sulphonylureas, GLP-1 receptor analogues, and DPP4 inhibitors have been used. The decision to use a specific strategy will rely upon both the degree of impaired glucose metabolism as well as the urgency with which glycaemic control is needed [5].

Hyperaldosteronism

Primary hyperaldosteronism is commonly associated with features of the metabolic syndrome. Insulin sensitivity is reduced in patients with primary hyperaldosteronism in comparison to those patients with essential hypertension. Aldosterone is able to induce insulin resistance in a variety of metabolic target tissues including muscle and adipose tissue. In addition, there is also evidence to suggest that aldosterone can impair insulin secretion. Mineralocorticoid receptor antagonism and surgical removal of an aldosterone secreting adenoma have resulted in improvements in insulin sensitivity and glucose handling in some, but not all, studies [6].

Hypo- and Hyperthyroidism

Both hypo- and hyperthyroidism are associated with glucose intolerance driven by both impaired insulin secretion and insulin resistance. Furthermore, the development of hyperthyroidism may worsen established glucose intolerance or type 2 diabetes.

While the underlying autoimmune aetiology underpins the association between thyroid disease and type 1 diabetes, the association with type 2 diabetes is more complex. Thyroid hormones have been shown to regulate key elements of carbohydrate metabolism, however, variability in glucose control is also associated with increased thyroid-stimulating hormone (TSH) levels which normalize once glucose control is optimized [7].

The mechanisms underpinning these observations have largely been explored in cellular and rodent models. Muscle glucose uptake is impaired in subclinical and overt hypothyroidism and hyperthyroidism (subclinical and overt) is associated with increased gluconeogenesis and impaired glycogen synthesis [8]. Thyroid hormones have been shown to increase the expression of phosphoenolpyruvate carboxykinase (PEPCK), the rate-limiting step in gluconeogenesis. In addition, hyperthyroid-associated hepatic insulin resistance is driven *via* central mechanisms originating within the hypothalamus altering sympathetic stimulation of the liver [9].

The management of impaired glucose homeostasis and diabetes in the context of thyroid dysfunction is not different to those patients without coexistent thyroid disease.

Phaeochromocytoma

Phaeochromocytomas and paragangliomas synthesize the catecholamines norepinephrine, epinephrine, and less frequently dopamine, either in isolation or in combination. Increasingly, these tumours are being recognized as part of hereditary syndromes rather than sporadic occurrences. The prevalence of abnormal glucose handling is highly variable and may occur in up to 75% of cases, but overt type 2 diabetes is relatively rare. Catecholamines inhibit insulin secretion and this is thought to be the main factor driving the development of diabetes although insulin resistance, increased hepatic glucose output, increased glycogenolysis and increased intestinal glucose absorption may all have a role to play [10, 11]. Diabetic ketoacidosis has been reported in a very small number of cases of phaeochromocytomas. Adrenoceptor blockade can improve both insulin secretion and glucose tolerance although even after surgical intervention, abnormalities in glucose homeostasis may persist.

Other Neuroendocrine Tumours

Glucagonoma and somatostatinoma are very rare neuroendocrine tumours. Patients with glucagonoma can present with a constellation of symptoms that include diabetes (in up to 70% of cases), necrolytic migratory erythema, weight loss, stomatitis, anaemia, venous thromboembolic disease, neuropsychiatric disturbances, and diarrhoea [12].

Approximately 75% of patients with pancreatic somatostatinoma develop diabetes (contrasting with 10% of patients who have intestinally located tumours). In addition to the direct impact of local somatostatin excess and its impact upon insulin secretion, replacement of functional endocrine pancreatic tissue may contribute to the high prevalence of impaired glucose handling in this group [13]. The abnormalities in glucose homeostasis are often relatively mild and oral hypoglycaemic agents or insulin can be used therapeutically.

Diabetes is commonly associated with many other endocrine conditions. While the principles of management may not differ (and include treating the underling endocrine disease), the fundamental importance lies in making the diagnosis so that appropriate treatment can be instigated without delay.

REFERENCES

1. Moller N, Jorgensen JO. Effects of growth hormone on glucose, lipid, and protein metabolism in human subjects. *Endocr Rev*, 2009; **30**: 152–77.
2. Mazziotti G, Floriani I, Bonadonna S, Torri V, Chanson P, Giustina A. Effects of somatostatin analogs on glucose homeostasis: a metaanalysis of acromegaly studies. *J Clin Endocrinol Metab*, 2009; **94**: 1500–8.
3. Valassi E, Santos A, Yaneva M, *et al*., and the ERCUSYN Study Group. The European Registry on Cushing's syndrome: 2-year experience. Baseline demographic and clinical characteristics. *Eur J Endocrinol*, 2011; **165**: 383–92.
4. Colao A, Petersenn S, Newell-Price J, *et al*., and the Pasireotide B2305 Study Group. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med*, 2012; **366**: 914–24.

5. Munir A, Newell-Price J. Management of diabetes mellitus in Cushing's syndrome. *Neuroendocrinology*, 2010; **92**(Suppl. 1): 82–5.
6. Luther JM. Effects of aldosterone on insulin sensitivity and secretion. *Steroids*, 2014; **91**: 54–60.
7. Celani MF1, Bonati ME, Stucci N. Prevalence of abnormal thyrotropin concentrations measured by a sensitive assay in patients with type 2 diabetes mellitus. *Diabetes Res*, 1994; **27**: 15–25.
8. Iwen KA, Schröder E, Brabant G. Thyroid hormones and the metabolic syndrome. *Eur Thyroid J*, 2013; **2**: 83–92.
9. Mullur R, Liu YY, Brent GA. Thyroid hormone regulation of metabolism. *Physiol Rev*, 2014; **94**: 355–82.
10. Wiesner TD, Blüher M, Windgassen M, Paschke R. Improvement of insulin sensitivity after adrenalectomy in patients with pheochromocytoma. *J Clin Endocrinol Metab*, 2003; **88**: 3632–6.
11. Komada H, Hirota Y, So A, *et al.* Insulin secretion and insulin sensitivity before and after surgical treatment of pheochromocytoma or paraganglioma. *J Clin Endocrinol Metab*, 2017; **102**: 3400–5.
12. Song X, Zheng S, Yang G, *et al.* Glucagonoma and the glucagonoma syndrome. *Oncol Lett*, 2018; **15**: 2749–55.
13. Nesi G1, Marcucci T, Rubio CA, Brandi ML, Tonelli F. Somatostatinoma: clinico-pathological features of three cases and literature reviewed. *J Gastroenterol Hepatol*, 2008; **23**: 521–6.

15.10.5 Diabetes in Pregnancy

Helen R. Murphy and Jennifer M. Yamamoto

Epidemiology

Globally, 17% of women have hyperglycaemia in pregnancy and 16% of these women have pre-existing diabetes [1]. The prevalence of T2D and gestational diabetes (GDM) in pregnancy is increasing and varies substantially between ethnic groups and locations [1]. The increase in GDM is accompanied by rising rates of maternal overweight and obesity, and excessive gestational weight gain. Obesity may further complicate pregnancy by increasing the risk of fertility problems, excess fetal growth, and maternal hypertensive disorders.

Physiology

There is a normal physiological adaptation of maternal metabolism during pregnancy to ensure growth and development of the fetus. This involves a greater fall in plasma glucose and amino acids, and a greater rise in free fatty acids in response to overnight fasting than in the non-pregnant state ('accelerated starvation'). This is associated with increased hepatic insulin resistance. In later pregnancy, a progressive rise in postprandial glucose and its associated insulin response, associated with decreased peripheral insulin sensitivity, parallels the growth of the fetal placental unit and rapidly reverses after delivery. This 'facilitated anabolism' brings about changes in carbohydrate, amino acid, and lipid metabolism, to ensure adequate nutrients for the developing fetus.

Pathophysiology

Women who lack the necessary beta-cell reserve, either absolutely, as in T1D, or relatively, as in T2D or GDM, still have normal physiological adaptation of carbohydrate, protein, and fat metabolism. Pregnant woman with T1D require sufficient insulin to compensate for increasing caloric needs, increasing adiposity, decreasing exercise, and increasing anti-insulin hormones. The insulin dose to maintain normoglycaemia and prevent maternal ketosis may increase up to threefold during pregnancy in T1D, and women with T2D will usually require insulin treatment in pregnancy, often at high doses often related to obesity and physical inactivity. An understanding of these underlying pathophysiological mechanisms is necessary for the successful management of these women during pregnancy.

Definitions

Internationally agreed definitions of T1 and T2D apply during pregnancy. Screening for overt diabetes in early pregnancy should be performed for those at high risk. The physiologic adaptations of pregnancy, however, affect fasting and 2-hour plasma glucose values as well as HbA1c. This suggests the need for more specific diagnostic criteria pertaining to pregnancy.

GDM is defined as carbohydrate intolerance with onset or first recognition during pregnancy, though its definition remains controversial. A fundamental challenge is that the relationship between maternal hyperglycaemia and adverse pregnancy outcome appears to be continuous. Traditionally, two different schemes have been employed. Both involve an oral glucose tolerance test (OGTT), but differ in the glucose load and interpretation.

The Hyperglycaemia and Pregnancy Outcome (HAPO) trial was a multicentre, multicultural, prospective observational study that examined whether maternal hyperglycaemia, short of diabetes, is associated with adverse maternal fetal outcome [1]. The study comprised 25 000 women who underwent a 75 g OGTT at ~28 weeks'

Table 15.10.5.1 Definitions of gestational diabetes mellitus based on a 75 g oral glucose tolerance test

	Fasting plasma glucose	1 h plasma glucose	2 h plasma glucose
World Health Organization diagnostic criteria			
Any of the following are met or exceeded	5.1–6.9 mmol/L (92–125 mg/dl)	≥ 10.0 mmol/L (180 mg/dl)	8.5–11.0 mmol/L (153–199 mg/dl)
International Association of Diabetes and Pregnancy Study Group			
Any of the following are met or exceeded	5.1 mmol/L (92 mg/dl)	10.0 mmol/L (180 mg/dl)	8.5 mmol/L (153 mg/dl)
National Institute for Health and Care Excellence (NICE) UK Guidelines			
Any of the following are met or exceeded	5.6 mmol/L (100 mg/dl)	–	7.8 mmol/L (140 mg/dl)

gestation with clinicians blinded to the results, unless the fasting glucose was >5.8 mmol/L or the 2-h was >11.1 mmol/L. The results demonstrated a continuum of risk, without clear thresholds, between each of the OGTT glucose measures (fasting, 1 hour, and 2 hours post-glucose load) and each of the four primary outcomes: macrosomia, primary caesarean section, neonatal hypoglycaemia, and fetal hyperinsulinism [1].

Different GDM diagnostic criteria continue to be endorsed by various international guidelines (Table 15.10.5.1) which reflect the controversy in this area. The glycaemic threshold values chosen from the HAPO data are consistent with an odds ratio of 1.75 for large-for-gestational-age, high neonatal body fat, elevated cord blood C-peptide, or a combination of these, compared to reference glucose values. Thirteen per cent (13%) of the total group had one or more values greater than or equal to the threshold.

The UK NICE guidelines recommended a diagnosis of GDM for women with a more severe glycaemic disturbance. Their diagnostic criteria suggest a fasting blood glucose of ≥5.6 mmol/L or a 2-hour blood glucose of ≥7.8 mmol/L with a 2-hour 75 g OGTT [2, 3].

Screening for Diabetes in Pregnancy

There is again a lack of international agreement in this area. Some authorities have recommended universal or selective screening, while others advocate screening only women with risk factors for GDM. In North America, screening traditionally has been based on a 50 g 1-hour non-fasting glucose challenge test [4]. In the UK, NICE guidelines recommend screening women with a 75 g OGTT who have any of the following risk factors [4]:

- BMI >30 kg/m²
- Previous baby with macrosomia (birth weight ≥4.5 kg)
- Previous GDM
- First-degree relative with diabetes
- Family origin with high prevalence of diabetes (South Asian, black Caribbean, and Middle Eastern)

Management

Pre-Gestational Diabetes in Pregnancy

Pre-Pregnancy Care

Pre-pregnancy care includes counselling about pregnancy and safe, effective contraception as well as additional care to prepare women for pregnancy. It has been associated with improvements in both pregnancy preparation and pregnancy complications including a reduction in serious adverse outcomes [5]. Despite these data, less than half of women attend specialized pre-pregnancy care clinics. Women with T2D are less likely to receive formal pre-pregnancy care than those with T1D [6].

The education of, and discussion with, women of reproductive age about pregnancy and contraception is an essential component of every diabetes consultation. Ideally, prepregnancy care should begin at least 6 months before stopping contraception and include advice and discussion on the following:

- Continued adherence with safe effective contraception until optimization of glycaemic control
- Individualized glycaemic targets
- Importance of periconception glucose control to pregnancy outcomes including congenital malformations and stillbirth
- The risk of hypoglycaemia and instruction of family members in use of glucagon
- Assessment of microvascular complications with referral and treatment as appropriate
- Review of medications, smoking, alcohol consumption
- Folic acid supplementation (5 mg daily)
- Advice on early referral mechanisms as soon as pregnancy is confirmed

Provision of Pregnancy Care

Specialized interdisciplinary teams operating in a secondary or tertiary care setting is a commonly adopted model for the care of women with diabetes. Members of the team include a diabetologist, obstetrician, a diabetes specialist nurse, dedicated dietitian, and a diabetes-trained midwife. Review is usually fortnightly and women are often seen weekly as term approaches.

Nutrition

A healthy lifestyle consisting of a well-balanced diet and moderate physical activity should be encouraged with the aim of limiting post-meal glucose excursions and preventing hypoglycaemia between meals. Low glycaemic-index foods help to achieve this (5). Individual advice around appropriate carbohydrate intake and insulin adjustment is essential.

Obesity in pregnancies continues to be common in GDM and T2D pregnancies but has also become a substantial concern in T1D with over half of all pregnant women with type 1 diabetes entering pregnancy overweight or obese [7, 8]. Therefore, weight management may be an important part of dietary counselling both

before and during pregnancy. Appropriate weight targets should be based on the woman's pre-pregnancy weight. The following gestational weight gain targets are recommended over pregnancy: BMI <18.5 kg/m² 13–18 kg, BMI 18.5–24.9 kg/m² 11–16 kg, BMI 25–29.9 kg/m² 7–11 kg, >30.0 kg/m² 5–9 kg [9].

Glycaemic Control

In T1D, blood glucose should be measured up to eight times daily (before and after each meal, at bedtime, and, intermittently, overnight). Continuous glucose monitoring (CGM) should be offered to all pregnant women with T1D. Evidence for this is from a landmark international randomized trial which found improvement in not only maternal glycaemic control (increased time spent in-target-glucose-range) but also a halving of the odds ratio for large-for-gestational age, neonatal hypoglycaemia, neonatal intensive care unit admissions and a shorter neonatal length of hospital stay [5]. In T2D, blood glucose should generally be tested at least four times a day but may vary based on treatment strategy.

Treatment targets for women should be individualized and include an assessment of the risk of hypoglycaemia. Women with pre-gestational diabetes should generally target an HbA1c of <6.5% prior to and early in pregnancy. Self-glucose monitoring is essential and women should aim for a fasting glucose of <5.3 mmol/L and a 1 hour postprandial glucose of <7.8 mmol/L or a 2 hour postprandial glucose of <6.4 mmol/L [10]. Any strategy for intensive glycaemic control must constantly be balanced against the risk of maternal hypoglycaemia.

Intensive insulin therapy is the standard of care for women with T1D and many with T2D. This can be in the form of multiple dose injection (MDI) insulin regimen or the insulin pump. MDI usually comprises a short-acting insulin taken before meals and an intermediate- or long-acting insulin taken daily or twice a day. There are no randomized controlled trial data to suggest that insulin pump therapy is superior to MDI for most women, but its selected use in experienced centres with motivated patients may be appropriate. Insulin pump therapy can be combined with CGM (sensor augmented pump therapy) and increasing degrees of automation from low glucose suspend to hybrid closed-loop systems, whereby basal insulin delivery in fully automated but manual pre-meal boluses are still required.

Gestational Diabetes Mellitus

Women with GDM are typically diagnosed during screening or through a diagnostic OGTT around at 24–28 weeks' gestation. Treatment for GDM has been shown to decrease preeclampsia, fetal overgrowth and shoulder dystocia in large randomized controlled trials and meta-analyses [11].

Medical nutritional therapy with regular review of blood glucose results and fetal growth is successful in controlling hyperglycaemia in most women with GDM [10]. For women who are not reaching target glucose with diet alone, pharmacologic therapy should be considered. After a discussion of risks and benefits, metformin can be offered to women in whom insulin therapy may be avoided. Insulin treatment should be offered if metformin therapy is unacceptable, contraindicated, or in women with more severe glycaemic disturbances [5]. While not considered first line, NICE Guidelines suggest that glibenclamide may be considered if women decline insulin therapy and are not achieving glycaemic targets with metformin and/or cannot tolerate metformin [12]. Metformin, unlike glibenclamide, crosses the placenta and caution is necessary until safety in the first trimester has been evaluated more fully. Glibenclamide needs to be administered at least 30–60 minutes before eating for optimal postprandial glucose control which is often impractical.

Diabetic Complications in Women with Pre-Existing Diabetes

When counselling a woman with advanced diabetic complications about pregnancy, the risks both to herself and to the fetus must be considered. The development or progression of retinopathy, nephropathy, or autonomic neuropathy during pregnancy may present major management challenges and require close collaboration between multidisciplinary colleagues. Diabetic ketoacidosis is a serious concern, and every mother should be instructed to monitor ketones and seek urgent advice if needed. Early pregnancy is a high-risk time for severe hypoglycaemia (defined as hypoglycaemia requiring third party assistance) and strategies to avoid and treat hypoglycaemia must be discussed.

Obstetric Surveillance

The goal of obstetric surveillance is to identify fetuses at risk, in order to intervene in a timely and appropriate fashion to reduce perinatal morbidity and mortality. Given the limitations of the available tests and lack of rigorous randomized controlled trials, all protocols used for fetal surveillance are empirical and have limitations. The UK NICE Guideline recommends ultrasound monitoring of fetal growth and amniotic fluid volume every 4 weeks from 28 to 36 weeks, and individualized monitoring of fetal wellbeing for women at risk of intrauterine growth restriction (IUGR) such as women with macrovascular disease or nephropathy [13, 14].

Labour and Delivery

The primary objectives are to avoid stillbirth and the complications of obstructed labour or shoulder dystocia associated with fetal overgrowth. Consequently, Caesarean section rates for women with pre-existing diabetes in most parts of the world are >50% [5]. Iatrogenic prematurity has resulted in high rates of admission to neonatal intensive care particularly following T1D pregnancies (Table 15.10.5.2) [5].

An individualized approach to the timing and mode of delivery is essential. Many factors need to be taken into consideration, including glycaemic control, diabetes complications, past obstetric history, fetal growth, and the availability of healthcare resources in labour.

The management of labour should follow standard practice as for women without diabetes. The treatment of diabetes during labour should follow an established protocol in a dedicated centre with a neonatal care unit equipped to deliver the highest level of care.

Most clinical practice guidelines recommend achieving a maternal blood glucose of 4–7 mmol/L during labour and delivery

Table 15.10.5.2 National pregnancy in diabetes audit outcomes for women with type 1 and type 2 diabetes (England, Wales, and the Isle of Man)

	Type 1 diabetes n = 1563	Type 2 diabetes n = 1386
Congenital anomaly	46.2/1000	4.6/1000
Stillbirth	10.7/1000	10.5/1000
Neonatal death	8.1/1000	11.4/1000
Gestational age at delivery in weeks*	36.4 (SD 2.0)	37.1 (SD 2.0)
Preterm delivery*	39.7%	21.7%
LGA*	46.4%	23.9%

* P <0.05; SD, standard deviation; preterm delivery: defined as delivery <37 weeks' gestation; LGA, large-for-gestational-age, defined as birth weight >90th percentile.

Data derived from Martens PJ, Shafer LA, Dean HJ, Sellers EA, Yamamoto J, Ludwig S, *et al*. Breastfeeding Initiation Associated with Reduced Incidence of Diabetes in Mothers and Offspring. *Obstet Gynecol*. 2016;128(5):1095–104.

to minimize risk of neonatal hypoglycaemia [5]. Hourly capillary glucose measurements are essential for safety and to guide insulin adjustment. For women with T1D, insulin delivery is usually either via intravenous insulin infusion or subcutaneous insulin pump. Some women with insulin-treated T2D and GDM may also require intravenous insulin. For women with diet-controlled GDM, glucose should be monitored every 1–2 hours once labour is established, with the aim of keeping glucose levels <7 mmol/L. If this is not achieved, an insulin/glucose infusion may be required.

Diabetes Management Immediately Postpartum

Because of the dramatic increase in insulin sensitivity following the delivery of the placenta, all women with diabetes in pregnancy will require treatment adjustment postpartum. For women with GDM, insulin is discontinued immediately following delivery. Glucose monitoring should continue for several days to ensure a return of blood glucose values to the normal range. These women should be booked for an assessment of glucose tolerance and review approximately 6–12 weeks after delivery.

Women with pregestational diabetes should be counselled that postpartum glycaemic targets are less stringent (typically 6–10 mmol/L) and that the risk maternal hypoglycaemia during this period is high [3]. Mothers with T2D previously on oral hypoglycaemic agents can typically discontinue insulin. Metformin and glibenclamide are used, often at lower doses, while breastfeeding. However, all other oral hypoglycaemic agents should be avoided. If blood glucose levels remain consistently above target, insulin may be required. For insulin-treated women with pregestational diabetes, a large decrease (typically a 50% reduction) in insulin doses immediately postpartum is required. Dose adjustments should be individualized and based on previous insulin doses and risk of maternal hypoglycaemia.

The benefits of breastfeeding to both mother and infant are well-recognized [15]. In addition to the overall benefits of breastfeeding, women with diabetes may have additional short- and long-term advantages for both mother and offspring including a decreased risk of developing type 2 diabetes following GDM [2].

Contraception

Safe effective contraception should be discussed with all women with diabetes who may become pregnant. Important considerations when choosing contraception include: decision to breastfeed, failure rates, patient preference, and cardiovascular risk. Long-acting reversible contraception methods including progesterone implants and intrauterine contraceptive devices (IUCD) are particularly suited to women who do not wish to become pregnant within the next 1–2 years. Uncomplicated diabetes generally does not limit contraceptive choice [5]. Women with longstanding diabetes, hypertension, microvascular, or cardiovascular complications, and possibly women who smoke cigarettes and/or have a BMI of ≥35 kg/m^2, should consider progesterone-only long-acting reversible contraception methods (injections, implants, or IUCD) although all contraception options may be safer than an unplanned pregnancy in high-risk women.

Postpartum Considerations

All women should be reviewed at 6–7 weeks after delivery, with further counselling regarding contraception and possible pregnancy planning. Women with pre-gestational diabetes are referred back to their pre-pregnancy care providers. Women with GDM are offered a 75 g OGTT after 6–12 weeks.

GDM is a recognized risk factor for the future development of GDM with subsequent pregnancies, T2D, and cardiovascular disease. In addition, risk factors for the development of GDM are similar to those for T2D and metabolic syndrome. The postnatal visit, therefore, provides a unique opportunity to offer specific lifestyle advice, to screen for cardiovascular factors, and remind women of the need for early referral for future pregnancies. Those with normal or impaired glucose tolerance should have annual diabetes screening in the community.

Long-Term Outcomes

There is now increasing recognition that offspring of women with diabetes (whether T1D, T2D, GDM, or maturity-onset diabetes of the young) during pregnancy are at increased future risk of diabetes, obesity, and cardiovascular disease [5]. Since these risk factors develop early in life, they place the offspring of women with diabetes at risk throughout adulthood. The major public health challenge is to break this vicious cycle by the prevention of diabetes until after women have completed their pregnancies and, possibly, also by better control of diabetes during pregnancy.

REFERENCES

1. Guariguata L, Linnenkamp U, Beagley J, Whiting DR, Cho NH. Global estimates of the prevalence of hyperglycaemia in pregnancy. *Diabetes Res Clin Pract*, 2014; **103**: 176–85.
2. Murphy HR, Bell R, Cartwright C, *et al*. Improved pregnancy outcomes in women with type 1 and type 2 diabetes but substantial

clinic-to-clinic variations: a prospective nationwide study. *Diabetologia*, 2017; **60**: 1668–77.
3. Mackin ST, Nelson SM, Kerssens JJ, *et al.* Diabetes and pregnancy: national trends over a 15 year period. *Diabetologia*, 2018; **61**: 1081–8.
4. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, *et al.* Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*, 2008; **358**: 1991–2002.
5. National Institute of Health and Care Excellence. Diabetes in pregnancy: management from preconception to the postnatal period. 25 February 2015. Available at: https://www.nice.org.uk/guidance/ng3 (accessed: 24 April 2020).
6. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41**(Suppl 1): S13–27.
7. Murphy HR, Roland JM, Skinner TC, *et al.* Effectiveness of a regional prepregnancy care program in women with type 1 and type 2 diabetes: benefits beyond glycemic control. *Diabetes Care*, 2010; **33**: 2514–20.
8. Yamamoto JM, Hughes DJF, Evans ML, *et al.* Community-based pre-pregnancy care programme improves pregnancy preparation in women with pregestational diabetes. *Diabetologia*, 2018; **61**: 1528–37.
9. Murphy HR, Steel SA, Roland JM, *et al.* Obstetric and perinatal outcomes in pregnancies complicated by type 1 and type 2 diabetes: influences of glycaemic control, obesity and social disadvantage. *Diabet Med*, 2011; **28**: 1060–7.
10. Feig DS, Donovan LE, Corcoy R, *et al.* Continuous glucose monitoring in pregnant women with type 1 diabetes (CONCEPTT): a multicentre international randomised controlled trial. *Lancet*, 2017; **390**: 2347–59.
11. American College of Obstetricians and Gynecologists. ACOG Committee opinion no. 548: weight gain during pregnancy. *Obstet Gynecol*, 2013; **121**: 210–12.
12. Donovan L, Hartling L, Muise M, Guthrie A, Vandermeer B, Dryden DM. Screening tests for gestational diabetes: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*, 2013; **159**: 115–22.
13. Crowther CA, Hiller JE, Moss JR, *et al.* Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med*, 2005; **352**: 2477–86.
14. Landon MB, Spong CY, Thom E, *et al.* A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med*, 2009; **361**: 1339–48.
15. Martens PJ, Shafer LA, Dean HJ, *et al.* Breastfeeding initiation associated with reduced incidence of diabetes in mothers and offspring. *Obstet Gynecol*, 2016; **128**: 1095–104.

15.11

Psychiatry and Diabetes

15.11.1 Type 1 Diabetes and Psychiatry

Khalida Ismail, Chris Garrett, and Marietta Stadler

Depression

Epidemiology

Depression is a mental state of pervasive low mood with little to no enjoyment, causing a marked deterioration in function. Early studies using self-rated questionnaires in T1D populations suggested rates of depression of around 15%, and recent studies using standardized clinical interview estimate rates that are twice as frequent as the general population [1–3]. Depressive symptoms have been linked to worse glycaemic control, morbidity, including acute admissions, and premature mortality [4–7].

Assessment and Management

For some clinicians not working in mental health there is sometimes a difficulty in discerning the mental state of a chronic intractable low mood from an existential sadness through the experience of living with T1D. However, from a health resources perspective it is useful that a referring clinician is able to draw the distinction thorough clinical assessment prior to referral for treatment.

For full ICD diagnostic criteria of depression see **Table 15.11.2.1** in Chapter 15.11.2, 'Type 2 Diabetes and Psychiatry'. People with depression feel persistently low with reduced enjoyment of everyday activities. Psychological symptoms emerge from this profound sense of sadness leading to negativity in all aspects of thinking including perception of self, surroundings, and future and some patients will wish to be dead and some will make active plans to end their life.

Physical symptoms of depression include tiredness, decreased concentration and disturbed sleep, an alteration in eating habits either markedly increased or decreased, and loss of sex drive. However, these can be confused with the physical symptoms of T1D and vice versa. For example, prolonged hyperglycaemia can cause dehydration and fatigue and diabetes late complications impair quality of life, which will in turn impact on mood.

One can quickly see how depression could affect a person's ability to self-manage a condition where motivation and consistency is key to achieving glycaemic control—when there has been a marked change in a person's diabetes management, consideration of an underlying depressive illness is appropriate and necessary.

When using standardized psychological tools in screening for depression, an understanding of the co-occurrence of diabetes distress is important and therefore a tool such as the Hospital Anxiety and Depression Scale alongside the Diabetes Distress Scale is appropriate [8, 9].

Treatment of depression alongside type 1 diabetes follows the same pathway as with people without a chronic health condition according to current guidance. In milder presentations a watch, wait, and review policy is employed alongside advice on self-help literature and psychoeducation around the condition, the importance of activity and connectedness to others. Management will generally take place in primary care though a clear plan is necessary for more intensive treatment to be activated if improvement does not occur relatively quickly or diabetes management becomes clearly affected, evidenced by worsening glycaemic control or acute admissions. More intensive treatment is required in moderate to severe presentations with psychological interventions and/or antidepressants in the form of selective serotonin reuptake inhibitors (SSRIs). Where treatment is not provided within the diabetes service, liaison is required between mental health and physical health clinicians, particularly when insulin omission is prominent and where DKA has occurred.

Anxiety Disorders

Epidemiology

Anxiety disorders are a collection of mental health conditions where the most prominent aspect of the presentation is heightened anxiety coupled with physical symptoms of stress response, namely tachycardia, sweating, and increased respiratory rate. Incidence of anxiety disorders in T1D is approximately 7–9%, more than double that in the background population, though caution is again necessary in diagnosis due to the overlapping symptoms of 'diabetes distress' and

Table 15.11.1.1 Outlining the personality disorder diagnoses using ICD-10 Criteria

Personality disorder cluster	Personality disorder diagnosis	ICD-10 Criteria
A	Paranoid	ICD-10 F60.0 • Excessive sensitivity to setbacks • Unforgiving of others • Suspiciousness and jealousy • Combative • Sense of personal rights • May be excessive self-importance, and there is often excessive self-reference
	Schizoid PD	ICD-10 F60.1 • Withdrawal from affectional, social, and other contacts • Preference for fantasy, solitary activities, and introspection • Limited capacity to express feelings and to experience pleasure
B	Dissocial PD	ICD-10 F60.2 • Disregard for social obligations • Callous unconcern for others • Gross disparity between behaviour and the prevailing social norms • Behaviour is not readily modifiable by adverse experience, including punishment • Low frustration tolerance • Low threshold for discharge of aggression, including violence • Tendency to blame others
	Emotionally unstable personality disorder (PD)	ICD-10 F60.3 • Act impulsively without consideration of consequences • Unpredictable mood • Liability to outbursts of emotion and an incapacity to control the behavioural explosions • Quarrelsome • Two subtypes: (1) impulsive type, characterized predominantly by emotional instability and lack of impulse control; (2) borderline type, characterized in addition by disturbances in self-image, aims, and internal preferences, chronic feelings of emptiness, intense and unstable interpersonal relationships, and a tendency to self-destructive behaviour, including suicide gestures and attempts
	Histrionic PD	ICD-10 F60.4 • Shallow and labile affectivity • Self-dramatization and theatricality • Suggestibility • Egocentricity • Lack of consideration for others • Easily hurt feelings • Continuous seeking for appreciation, excitement, and attention
C	Anankastic PD	ICD-10 F60.5 • Feelings of doubt • Perfectionism • Excessive conscientiousness, • Checking and preoccupation with details • Stubbornness, caution, and rigidity
	Anxious	ICD-10 F60.6 • Constant tension and apprehension • Insecurity and inferiority • Continuous yearning to be liked and accepted • Hypersensitivity to rejection and criticism with restricted personal attachments, and a tendency to avoid certain activities by habitual exaggeration of the potential dangers or risks in everyday situations
	Dependent	ICD-10 F60.6 • Passive reliance on other people • Fear of abandonment • Feelings of helplessness and incompetence • Compliance with the wishes of others • Weak response to the demands of daily life

with the majority of studies largely focused on people with type 2 diabetes [10–12].

Assessment and Management

Anxiety disorders divide largely into three categories, those where anxiety is consistently present, those where there are discreet episodes of heightened anxiety without anxiety in between and those where there is background anxiety and overlapping intense episodes. Often these conditions are comorbid with depression and presence of heightened anxiety indicates the need to assess for underlying depression.

Generalized anxiety disorder is a chronic anxiety state characterized by persistent 'free-floating' anxiety, not particularly focused on a specific situation or concern, but pervasive and all-encompassing with inability to feel relaxed or at ease. Physical manifestations can

include unsteadinesss walking, trembling, and light-headedness—symptoms which can be related to T1D or its therapy. Symptomatic anxiety disorder may also be misinterpreted as symptoms of hypoglycaemia or neuropathy (e.g. numbness).

Often this constant worry will lead to avoidance behaviour to prevent anxiety, potentially resulting in functional impairment (e.g. in work, relationships, social life). This may include avoidance of diabetes self-management tasks.

Panic disorder is the experience of intermittent episodes of extremely intense anxiety. Although these episodes may be triggered by specific situations such as getting on a bus or being in a crowded place, or in the diabetes context fear that one is becoming hypoglycaemic, they often emerge 'out of the blue'. During a 'panic attack' the person feels suddenly intensely fearful, about to die, and consequent adrenal activity leads to tachycardia, sweating, and hyperventilation, symptoms also experienced in the adrenergic response to hypoglycaemia, again emphasizing the importance of a detailed history.

Anxiety disorders may impact on type 1 diabetes where many of the health behaviours required could evoke emotion. Specific phobias can occur around finger prick blood tests or self-injection of insulin and can lead to problematic self-management through avoidance. Fear of hypoglycaemia can lead to disproportionate insulin restriction and chronic hyperglycaemia.

Like depression, treatment of anxiety disorders alongside type 1 diabetes depends on severity. Milder presentations, not affecting diabetes management are usually managed by primary care physicians using 'wait and review' alongside psychoeducation and advice on avoidance behaviours and maintaining a social network as well as self-help literature. More intensive treatment is required in moderate to severe presentations with psychological interventions being particularly beneficial in simple diabetes-related phobias including fear of testing or fear of injections. More general cognitive behavioural therapy (CBT) interventions and/or antidepressants are used in more pervasive presentations and liaison is required between mental health and physical health clinicians, particularly when insulin omission is prominent.

Eating Disorders

Epidemiology

Eating disorders (ED) are twice as common in young people with T1D compared with controls without diabetes [13]. Typically, onset is during adolescence either soon after the onset of T1DM or after onset of puberty in established T1D [14]. Poor classification has resulted in broad prevalence estimates from 8% to 36% [15, 16] with a further 9–14% classified as 'subthreshold' ED. Insulin restriction in T1D, which is seen to varying degrees in ED is associated with 3-fold increased mortality compared to those who do not restrict insulin [17]. Mortality is driven by accelerated onset of diabetes late complications secondary to persistent hyperglycaemia [18, 19] and by acute complications due to almost total insulin omission leading to DKA or insulin overdose [17].

Assessment and Management

Diagnostic criteria for common EDs are detailed in Table 15.11.2.2 of Chapter 15.11.2. However, clinical presentations in T1D can vary widely and often do not fit within established ED categories. Some patients restrict by only a few units of insulin per day to keep their weight in check with a normal body mass index (BMI) so that the only clinical indicator is persistent hyperglycaemia. Other patients will restrict by much larger amounts and induce rapid weight loss within a few days (sometimes up to 5 kg) as the severe hyperglycaemia causes a rapid osmotic diuresis [20]. This may present with rDKA. Recently in the media and among diabetes communities, insulin omission for the purpose of losing weight, has been termed 'diabulimia' [21].

Screening questionnaires for ED in the T1D context are problematic due to false positive responses related to overlap between diabetes management and eating disorder behaviours such as a focus on diet, counting carbohydrates, and eating when not hungry to avoid or treat hypoglycaemia. There is only one diabetes-specific questionnaire for ED, the DEPS-R which has been tested in adolescent and adult populations [22].

From a clinical perspective, management of T1D and ED requires a multidisciplinary approach to integrate physical and mental health elements. There are few specialists in this area but treating clinicians should be cognisant of certain concepts. For example, omitting or restricting insulin to control or reduce weight is a harmful behaviour that is only possible in people with type 1 diabetes and occurrence of diabetic ketoacidosis and extent and chronicity of hyperglycaemia are key indicators, while BMI, which is usually used by psychiatrists to make decisions regarding admission in ED is less relevant [23]. In addition, physical and mental health management can be at cross-purposes, with ED therapeutic interventions actively discouraging consideration of calorie content of food, while diabetes clinicians encourage a focus on carbohydrate and fat content in order to calibrate insulin dose.

The main stay of treatment is psychotherapeutic with limited call for psychopharmacological interventions except use of SSRIs as treatment for binge eating. Therapeutic modalities employed include CBT, mentalization based treatment (MBT), and cognitive analytic therapy (CAT) but there is a very limited evidence base for outcomes and treatment generally follows what is available locally. As a general principle, therapy aims to help a person gain greater understanding of their thoughts and emotions connected to the ideas about their body, to help regulate eating patterns and to explore their relationship with diabetes. From a diabetes perspective, clinicians are actively involved in support and monitoring of insulin use as well as monitoring for complications and specialist dietician input.

Personality Disorders

Epidemiology

The personality disorders (PD) are a group of psychiatric conditions made up of enduring, extreme personality characteristics often evident in late childhood or adolescence but generally diagnosed over the age of 18, broadly dividing into three 'clusters' referred to as A (odd, eccentric), B (dramatic) and C (anxious). Table 15.11.1.1 shows ICD-10 criteria for the different PD diagnoses. A recent longitudinal study suggested that PD are twice as common in young adults with T1D, implying that the condition is affecting personality

development *per se*. A recent case-control study indicated association between personality disorder and recurrent DKA [24].

Assessment

Cluster A

The personality disorders that make up cluster A are often referred to as the 'odd' or 'eccentric' group, although an additional key element is ideas or beliefs outside reality. People with paranoid PD have pervasive distrust, struggle to maintain close relationships, bear prolonged grudges towards individuals or organizations and are often litigious. Schizoid personality disorder consists of lack of interest in others, isolation, and lack of close relationships. Typically they are non-plussed about emotions and described as 'cold'. An example in T1D may be a patient with chronic hyperglycaemia who sustains a leg ulcer, yet does not become anxious or change behaviours despite development of a long-term, life-limiting complication.

Cluster B

The cluster B PD diagnoses are described as the 'dramatic' group due to marked fluctuations in emotion regulation and difficulties in impulse control, particularly at times of interpersonal conflict. In the diabetes context, difficulty in emotion regulation is associated with diabetes distress [25]. Often T1D is used as a vehicle for expressing aggression and upset through over- and undertreatment. Such expressions can be external, creating disruptions within families, and internal, inflicting self-harm. Dissocial PD is typified by impulsiveness, recklessness, disregard for others, and externalization of difficulties leading to blaming others. In the diabetes context this can lead to a lack of responsibility for diabetes care and tendency to make others culpable for diabetes treatment and complications. Emotionally unstable PD also classically presents with emotion dysregulation and impulsivity, but recklessness tends to be more self-damaging with interpersonal conflict often acting as a trigger to self-destructive behaviour including undertreatment causing DKA or overtreatment with insulin and subsequent hypoglycaemia.

Cluster C

Cluster C is made up of the 'anxious' PDs and much of the interplay with T1D results in extreme anxiety and panic regarding self-management with subsequent avoidance of self-monitoring and insulin delivery. This may manifest as 'fear of hypoglycaemia' or 'needle phobia' and patients presenting with these common issues should be assessed by their clinician for more pervasive anxiety symptoms. Anxious PD consists of a core feature of profoundly diminished self-esteem and perceived criticism, leading to avoidance of work and social interaction. Dependent PD also has the core feature of negative self-perception. Here, rather than avoidance, feelings of inadequacy may lead to overreliance on others for emotional and physical wellbeing. This group struggle to gain independence and the T1D burden is displaced onto others. Anankastic PD is one of the few psychiatric conditions, which can interact favourably with T1D. The key feature is an obsessive drive for control, mastery over one's environment, and perfectionism, which can lead to focused attention on glycaemic control. This personality type can be associated with profound frustration with the intrinsic unpredictability of blood glucose readings with risk of superimposed depressive disorder or recurrent hypoglycaemia from overtight control.

Management

Unlike depression and anxiety disorders, which are eminently treatable and often transient, PDs are persistent and treatment remains difficult despite advances in the last two decades [26]. Eliciting the disorder early is vital to maximize treatment potential and limit impact on T1D self-management. Moreover, difficulty with relationships is a prime issue in PD and diagnosing the condition can aid understanding and the approach to therapeutic engagement for diabetes clinicians, which should include establishing trust and a consistent approach, with potential redefining of achievable treatment goals. Established psychotherapy interventions for PD focus mainly on the cluster B disorders with MBT, dialectical behaviour therapy and transference-focused psychotherapy all having an established evidence base [26].

Summary and Future Directions

This chapter has summarized the major mental health issues impacting T1D and emphasizes their bidirectional nature: physical symptoms potentiate psychopathology and mental health problems worsen prognosis through impact on self-care. Diabetes clinical care should include an element of mental health management in order to prevent early complications and mortality. Integration of mental healthcare into T1D services is therefore the gold standard. In addition, mental health should be assessed as part of the annual routine screening for diabetes complications and efforts should be made to stratify risk. Embedding mental healthcare into the physical healthcare setting will lower the threshold for accessing early diagnosis and targeted therapies. Such interventions should be delivered by clinicians with experience in managing psychopathology specific to T1D in close partnership with diabetes clinicians.

REFERENCES

1. Barnard KD, Skinner TC, Peveler R. The prevalence of co-morbid depression in adults with type 1 diabetes: systematic literature review. *Diabet Med*, 2006; **23**: 445–8.
2. Fisher L, Hessler DM, Polonsky WH, *et al.* Prevalence of depression in type 1 diabetes and the problem of over-diagnosis. *Diabet Med*, 2016; **33**: 1590–7.
3. Kampling H, Petrak F, Farin E, Kulzer B, Herpertz S, Mittag O. Trajectories of depression in adults with newly diagnosed type 1 diabetes: results from the German Multicenter Diabetes Cohort Study. *Diabetologia*, 2017; **60**: 60–8.
4. Gonzalez JS, Peyrot M, McCarl LA, *et al.* Depression and diabetes treatment nonadherence: a meta-analysis. *Diabetes Care*, 2008; **31**: 2398–403.
5. Gilsanz P, Karter AJ, Beeri MS, Quesenberry CP, Whitmer RA. The bidirectional association between depression and severe hypoglycemic and hyperglycemic events in type 1 diabetes. *Diabetes Care*, 2018; **41**: 446–52.
6. Ahola AJ, Harjutsalo V, Saraheimo M, Forsblom C, Groop P-H, Group FS. Purchase of antidepressant agents by patients with type

1 diabetes is associated with increased mortality rates in women but not in men. *Diabetologia*, 2012; **55**: 73–9.
7. Winkley K, Sallis H, Kariyawasam D, *et al.* Five-year follow-up of a cohort of people with their first diabetic foot ulcer: the persistent effect of depression on mortality. *Diabetologia*, 2012; **55**: 303–10.
8. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *J Psychosom Res*, 2002; **52**: 69–77.
9. Polonsky WH, Fisher L, Earles J, *et al.* Assessing psychosocial distress in diabetes. *Diabetes Care*, 2005; **28**: 626–31.
10. Smith KJ, Béland M, Clyde M, *et al.* Association of diabetes with anxiety: a systematic review and meta-analysis. *J Psychosom Res*, 2013; **74**: 89–99.
11. Cooper MN, Lin A, Alvares GA, De Klerk NH, Jones TW, Davis EA. Psychiatric disorders during early adulthood in those with childhood onset type 1 diabetes: rates and clinical risk factors from population-based follow-up. *Pediatr Diabetes*, 2017; **18**: 599–606.
12. Dybdal D, Tolstrup JS, Sildorf SM, *et al.* Increasing risk of psychiatric morbidity after childhood onset type 1 diabetes: a population-based cohort study. *Diabetologia*, 2018; **61**: 831–8.
13. Jacobson AM, Braffett BH, Cleary PA, Gubitosi-Klug RA, Larkin ME, Group D Research. The long-term effects of type 1 diabetes treatment and complications on health-related quality of life: a 23-year follow-up of the Diabetes Control and Complications/Epidemiology of Diabetes Interventions and Complications cohort. *Diabetes Care*, 2013; **36**: 3131–8.
14. Gonder-Frederick LA, Clarke WL, Cox DJ. The emotional, social, and behavioral implications of insulin-induced hypoglycemia. *Semin Clin Neuropsychiatry*, 1997; **2**: 57–65.
15. Jones JM, Lawson ML, Daneman D, Olmsted MP, Rodin G. Eating disorders in adolescent females with and without type 1 diabetes: cross sectional study. *BMJ*, 2000; **320**: 1563–6.
16. Colton PA, Olmsted MP, Daneman D, *et al.* Eating disorders in girls and women with type 1 diabetes: a longitudinal study of prevalence, onset, remission, and recurrence. *Diabetes Care*, 2015; **38**: 1212–17.
17. Goebel-Fabbri AE, Fikkan J, Franko DL, Pearson K, Anderson BJ, Weinger K. Insulin restriction and associated morbidity and mortality in women with type 1 diabetes. *Diabetes Care*, 2008; **31**: 415–19.
18. Stadler M, Peric S, Strohner-Kaestenbauer H, *et al.* Mortality and incidence of renal replacement therapy in people with type 1 diabetes mellitus–a three decade long prospective observational study in the Lainz T1DM cohort. *J Clin Endocrinol Metab*, 2014; **99**: 4523–30.
19. Association AD. The absence of a glycemic threshold for the development of long-term complications: the perspective of the Diabetes Control and Complications Trial. Diabetes. *Am Diabetes Assoc*, 1996; **45**: 1289–98.
20. Merwin RM, Dmitrieva NO, Honeycutt LK, *et al.* Momentary predictors of insulin restriction among adults with type 1 diabetes and eating disorder symptomatology. *Diabetes Care*, 2015; **38**: dc150753.
21. Staite E, Zaremba N, Macdonald P, *et al.* 'Diabulima' through the lens of social media: a qualitative review and analysis of online blogs by people with type 1 diabetes mellitus and eating disorders. *Diabet Med*, 2018; **35**: 1329–36.
22. Markowitz JT, Butler DA, Volkening LK, Antisdel JE, Anderson BJ, Laffel LMB. Brief screening tool for disordered eating in diabetes internal consistency and external validity in a contemporary sample of pediatric patients with type 1 diabetes. Diabetes Care. *Am Diabetes Assoc*, 2010; **33**: 495–500.
23. Group DR. Weight gain associated with intensive therapy in the diabetes control and complications trial. Diabetes Care. *Am Diabetes Assoc*, 1988; **11**: 567–73.
24. Garrett C. *Psychopathology of Recurrent Diabetic Ketoacidosis.* London: King's College, 2018.
25. Fisher L, Hessler D, Polonsky W, *et al.* Emotion regulation contributes to the development of diabetes distress among adults with type 1 diabetes. *Patient Educ Couns*, 2018; **101**: 124–31.
26. Bateman AW, Gunderson J, Mulder R. Treatment of personality disorder. *Lancet*, 2015; **385**: 735–43.

15.11.2 Type 2 Diabetes and Psychiatry

Marilia Calcia, Clare Whicher, Hermione Price, Khalida Ismail, and Calum Moulton

Depression

Epidemiology

Clinical depression is characterized by pervasive low mood, reduced enjoyment, and/or increased fatigue, accompanied by other biological and cognitive symptoms (Table 15.11.2.1). Approximately 10% of people with type 2 diabetes meet ICD-10 diagnostic criteria for clinical depression and 20–30% have significant depressive symptoms on self-report questionnaires [1, 2]. Depression has a bidirectional relationship with type 2 diabetes: depression increases the risk of incident type 2 diabetes by 60%, while type 2 diabetes increases the risk of incident depression by 15–20% [3]. Diabetes and depression comorbidity are associated with increased incidence of microvascular and macrovascular complications in addition to premature mortality [4]. Notably, depression does not appear to worsen diabetes self-care [5], and growing evidence suggests that shared biological mechanisms, such as inflammation, may link the two conditions [6].

Table 15.11.2.1 Diagnostic criteria for depression according to the International Classification of Diseases (ICD-10)

Diagnosis	Criteria
Depression	At least two of: • Low mood • Anhedonia (loss of interest and enjoyment) • Low energy Plus at least two of: • reduced concentration and attention • reduced self-esteem and self-confidence • ideas of guilt and unworthiness • bleak and pessimistic views of the future • ideas or acts of self-harm or suicide • disturbed sleep • diminished appetite Symptoms must be present for at least 2 weeks (or less if severe and of rapid onset).

Management

Guidelines recommend that clinicians working with people with chronic illnesses, such as type 2 diabetes, should actively screen for concurrent depression [7]. Although brief screening questionnaires may assist, such as the Patient Health Questionnare-9 (PHQ-9) [8], a clinical assessment should confirm the diagnosis of depression. The history should consider previous treatments for depression and social problems that might be perpetuating depression. Patients should specifically be asked about thoughts of self-harm or suicide.

Treatment of depression follows a stepped care model [7]. Patients with mild or subthreshold depressive symptoms may be referred to a group physical activity programme, group-based peer support or low-intensity psychological interventions (e.g. group cognitive-behavioural therapy (CBT)). Patients with moderate and severe depression, or with persistent mild to moderate depression despite lower-intensity treatment, need to have their treatment intensified including antidepressants and/or high-intensity psychological therapies. Selective serotonin reuptake inhibitors (SSRIs), such as sertraline and fluoxetine, are the first-line antidepressants due to their favourable side-effect profile. Prior to starting antidepressants, patients should be advised of the risk of increased agitation, nausea, and suicidal ideation in the first few days of treatment. Regular reviews are indicated, and if an antidepressant is ineffective after 2–4 weeks, the first step is to check the patients' concordance with treatment and for substance misuse. If antidepressants are ineffective or partially effective, a dose increase, or switching to a different medication should be considered.

Antidepressants should be continued for at least 6 months from the point of remission of a depressive episode to prevent early relapse or recurrence [7]. Referral to specialist mental health services is advised for patients with psychotic- or treatment-resistant depression, or those with risk of self-harm, suicide, or severe self-neglect.

Eating Disorders

Epidemiology

Disordered eating affects up to 40% of people with type 2 diabetes [9]. Binge eating disorder (BED), bulimia nervosa (BN) and night eating syndrome (NES) are eating disorders associated with excessive caloric intake and are often comorbid with overweight, obesity and type 2 diabetes [10] (see Table 15.11.2.2 for diagnostic criteria).

Table 15.11.2.2 Diagnostic criteria for eating disorders associated with type 2 diabetes

Diagnosis	Criteria
Bulimia nervosa	ICD-10 F50.2 • Persistent preoccupation with eating, and an irresistible craving for food; episodes in which large amounts of food are consumed in short periods of time • Attempts to counteract the 'fattening' effects of food by one or more of the following: – self-induced vomiting – purgative abuse – alternating periods of starvation – use of drugs such as appetite suppressants, thyroid preparations, or diuretics – neglecting insulin treatment • Morbid dread of fatness and a sharply self-defined weight threshold, well below the premorbid weight that constitutes optimum or healthy weight
Binge eating disorder (BED)	DSM 5 307.51 (American Psychiatric Association, 2013) • Recurrent episodes of binge eating, characterized by eating a large amount of food in a discrete period of time, with lack of control overeating during the episode • Episodes associated with at least 3 of the following: – Eating much more rapidly than normal – Eating until feeling uncomfortably full – Eating large amounts of food when not feeling hungry – Eating alone due to feeling embarrassed by how much one is eating – Feeling disgusted with oneself, depressed, or very guilty after the episode • Marked distress regarding binge eating • Episodes occur at least once a week for 3 months Episodes are not associated with inappropriate compensatory behaviour
Night eating syndrome	DSM 5 307.59 • Recurrent episodes of eating after awakening from sleep, or excessive food consumption after the evening meal • There is awareness and recall of the eating • Episodes are not better explained by changes in the sleep-wake cycle or by local or social norms • Episodes cause significant distress and/or functional impairment The pattern or eating is not better explained by BED, other mental or medical disorder, or the effects of a medication

Cross-sectional studies have found that patients with type 2 diabetes who have frequent episodes of binge eating tend to have higher body mass index, more depressive symptoms, poorer glycaemic control, and more frequent hospital admissions for diabetes complications [11]. Therefore, addressing these conditions is essential in supporting patients in lifestyle interventions to manage type 2 diabetes.

Management

The evidence base for treatment of BED and BN in type 2 diabetes is limited. From evidence in the general population, treatment may include guided self-help and psychological therapies such as CBT. Family therapy may help young people with BN. High-dose SSRI (e.g. fluoxetine up to 60 mg) is effective in combination with psychological therapy or guided self-help for BN [12]. Specialist referral is advised for patients with severe distress or comorbidity with other mental disorders. Those with complications associated with purging or restrictive behaviours or risk of self-harm or suicide may warrant urgent psychiatric assessment. Treatments for NES are not as well established; SSRI may be helpful in treating NES symptoms and promoting weight loss, and pilot studies have indicated that CBT may help reduce the frequency of night eating episodes [13].

Cognitive Impairment and Dementia

Epidemiology

A broad spectrum of cognitive impairment is overrepresented in people with type 2 diabetes. Most severe is dementia: a progressive deterioration in multiple cognitive domains leading to a significant deterioration in functioning [14]. The relative risks for Alzheimer disease and vascular dementia are 1.53 and 2.27, respectively, in people with diabetes, although Alzheimer's disease is more prevalent overall [15]. Mild cognitive impairment (MCI) describes cognitive impairment greater than expected based on age and education but which, unlike dementia, does not interfere with daily activities, and people with diabetes are 1.5–3 times more likely to progress from MCI to dementia [16]. There is also emerging evidence that milder subjective-only cognitive decrements are common in people with diabetes of all ages, although research is at an earlier stage [17].

Risk factors for dementia in diabetes are likely multiple. In the large-scale ACCORD-MIND trial, intensive modification of glycaemic control did not affect cognitive decline in people with type 2 diabetes [18]. By contrast, prior episodes of cardiovascular disease (CVD) and repeated hypoglycaemic episodes are consistent risk factors for dementia in people with type 2 diabetes [19]. Promising novel risk factors for dementia in type 2 diabetes, including inflammation and depression, have not been tested in clinical trials.

Management

Diabetes physicians should be mindful that there may be a bidirectional effect of cognitive impairment and diabetes control. Insight into cognitive impairment is frequently poor and dementia increases the risk of medication errors, hypoglycaemia and CVD [20]. The following steps may support non-specialists encountering suspected cognitive impairment:

1. Obtain collateral history of any cognitive decline, its time-course, and effects on function.
2. Conduct a physical examination and consider blood tests to exclude organic causes of cognitive impairment, such as hypothyroidism and vitamin B_{12} deficiency.
3. Screen for depression (e.g. PHQ-9 score ≥10), as depression can cause cognitive impairment.
4. If time permits, perform a brief supplementary cognitive assessment, such as the 6-item cognitive impairment test or Mini-Cog [14].
5. If MCI or dementia is still suspected, follow local protocols. In the United Kingdom, patients with suspected dementia should be referred to a memory clinic or community old age psychiatry service [14].

Detailed guidance is available on adjusting diabetes management in people with established dementia [20]. In particular, clinicians should be vigilant to hypoglycaemia risk and should consider mitigating this through additional support at mealtimes, relaxation of HbA1c targets to 8.5% (70 mmol/mol), and avoidance of sulphonylurea medication.

Severe Mental Illness

Epidemiology

Individuals with schizophrenia and other severe mental illness (SMI) have a 20-year reduction in life expectancy when compared to the general population, largely due to increased incidence of CVD [21]. Further compounding the risk of CVD is a 2–3-fold increased risk of developing type 2 diabetes [22]. The multiple reasons behind this include genetic susceptibility and environmental factors, such as poor diet, obesity, and lower levels of physical activity. Furthermore, antipsychotic medications, in particular olanzapine and clozapine, are associated with lipid and glucose dysregulation and increased risk of obesity [23]. These medications have sedative and appetite stimulating properties and may also have a direct toxic effect on pancreatic beta-cells [23]. In addition to long-term risks, individuals using antipsychotic medication are also at risk of developing acute complications of diabetes including diabetic ketoacidosis (DKA) [24].

Management

Deaths from acute complications of unrecognized diabetes in individuals with SMI have been reported, indicating a need for all people starting antipsychotic therapy and their carers to be made aware of the risk of developing diabetes and its acute symptoms. Effective strategies for managing obesity and glucose dysregulation in people with SMI using antipsychotic medication are lacking. In patients prescribed clozapine, coprescription of metformin, aripiprazole, or topiramate may help to reduce weight, but this evidence should be balanced against potential side effects [25]. A theory-based, group structured lifestyle education programme was unable to achieve weight loss [26]. Promisingly, glucagon-like peptide-1 receptor agonists have recently been found to achieve reversion to normal glucose tolerance in up to two-thirds of people [27].

Further compounding the challenge of managing diabetes and obesity in people living with SMI is the organization of care systems which often separate physical from mental healthcare. Moreover, innovations for people with type 2 diabetes in the general population

[28] appear to exclude those with SMI, even though these individuals may be in the greatest need.

Future Treatment Directions

In isolation, current treatments for depression are limited by failure rate of around one-third [29], poor concordance rates [30], and inconsistent effects on subsequent diabetes outcomes [31]. Novel strategies may include collaborative interventions to achieve better integration between mental and physical health [31], novel psychological therapies [32], and even the repositioning of current diabetes treatments for mental health [33].

Summary

Psychiatric disorders are overrepresented in patients with type 2 diabetes and predict poor diabetes outcomes. Clinicians in diabetes and primary care should actively screen for these comorbidities and seek to initiate at least first-line management and refer appropriately. Future treatments strategies, such as collaborative care and repositioning of diabetes treatments, may provide novel opportunities to integrate care and improve both psychological and biomedical outcomes.

REFERENCES

1. NHS Mental Health Task Force. The five year forward view mental health taskforce: public engagement findings. September 2015. Available at: https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2015/09/fyfv-mental-hlth-taskforce.pdf (accessed: 24 April 2020).
2. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*, 2001; **24**: 1069–78.
3. Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. *Diabetes Care*, 2008; **31**: 2383–90.
4. Park M, Katon WJ, Wolf FM. Depression and risk of mortality in individuals with diabetes: a meta-analysis and systematic review. *Gen Hosp Psychiatry*, 2013; **35**: 217–25.
5. Ismail K, Moulton CD, Winkley K, *et al.* The association of depressive symptoms and diabetes distress with glycaemic control and diabetes complications over 2 years in newly diagnosed type 2 diabetes: a prospective cohort study. *Diabetologia*, 2017; **60**: 2092–102.
6. Moulton CD, Pickup JC, Ismail K. The link between depression and diabetes: the search for shared mechanisms. *Lancet Diabetes Endocrinol*, 2015; **3**: 461–71.
7. National Institute for Health and Care Excellence: Clinical Guideline 91: depression in adults with a chronic physical health problem: recognition and management. Available at: https://www.nice.org.uk/guidance/cg91 (accessed: 24 April 2020).
8. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*, 2001; **16**: 606–13.
9. Garcia-Mayor R V, Garcia-Soidan FJ. Eating disorders in type 2 diabetic people: brief review. *Diabetes Metab Syndr*, 2017; **11**: 221–4.
10. Kessler RC, Berglund PA, Chiu WT, *et al.* The prevalence and correlates of binge eating disorder in the World Health Organization World Mental Health Surveys. *Biol Psychiatry*, 2013; **73**: 904–14.
11. Nicolau J, Simo R, Sanchis P, *et al.* Eating disorders are frequent among type 2 diabetic patients and are associated with worse metabolic and psychological outcomes. *Acta Diabetol*, 2015; **52**: 1037–44.
12. Treasure J, Claudino AM, Zucker N. Eating disorders. *Lancet (London, England)*. 2010; **375**: 583–93.
13. Allison KC, Tarves EP. Treatment of night eating syndrome. *Psychiatr Clin North Am*, 2011; **34**: 785–96.
14. National Institute for Health and Care Excellence. Clinical Guideline 97: dementia: assessment, management and support for people living with dementia and their carers. 2018. Available at: https://www.nice.org.uk/guidance/ng97 (accessed: 25 April 2020).
15. Gudala K, Bansal D, Schifano F, Bhansali A. Diabetes mellitus and risk of dementia: a meta-analysis of prospective observational studies. *J Diabetes Investig*, 2013; **4**: 640–50.
16. Li J, Wang YJ, Zhang M, *et al.* Vascular risk factors promote conversion from mild cognitive impairment to Alzheimer disease. *Neurology*. 2011; **76**: 1485–91.
17. Koekkoek PS, Kappelle LJ, van den Berg E, Rutten GEHM, Biessels GJ. Cognitive function in patients with diabetes mellitus: guidance for daily care. *Lancet Neurol*, 2015; **14**: 329–40.
18. Launer LJ, Miller ME, Williamson JD, *et al.* Effects of intensive glucose lowering on brain structure and function in people with type 2 diabetes (ACCORD MIND): a randomised open-label substudy. *Lancet Neurol*, 2011; **10**: 969–77.
19. Haroon NN, Austin PC, Shah BR, Wu J, Gill SS, Booth GL. Risk of dementia in seniors with newly diagnosed diabetes: a population-based study. *Diabetes Care*, 2015; **38**: 1868–75.
20. Sinclair A, Dunning T, Colagiuri S. Managing older people with type 2 diabetes. *J Diabetes Nurs*, 2014; **18**: 145.
21. Hoang U, Goldacre MJ, Stewart R. Avoidable mortality in people with schizophrenia or bipolar disorder in England. *Acta Psychiatr Scand*, 2013; **127**: 195–201.
22. Stubbs B, Vancampfort D, De Hert M, Mitchell AJ. The prevalence and predictors of type two diabetes mellitus in people with schizophrenia: a systematic review and comparative meta-analysis. *Acta Psychiatr Scand*, 2015; **132**: 144–57.
23. Miron IC, Baroana VC, Popescu F, Ionica F. Pharmacological mechanisms underlying the association of antipsychotics with metabolic disorders. *Curr Heal Sci J*, 2014; **40**: 12–17.
24. Vuk A, Kuzman MR, Baretic M, Osvatic MM. Diabetic ketoacidosis associated with antipsychotic drugs: case reports and a review of literature. *Psychiatr Danub*, 2017; **29**: 121–35.
25. Whitney Z, Procyshyn RM, Fredrikson DH, Barr AM. Treatment of clozapine-associated weight gain: a systematic review. *Eur J Clin Pharmacol*. 2015; **71**: 389–401.
26. Holt RIG, Gossage-Worrall R, Hind D, *et al.* Structured lifestyle education for people with schizophrenia, schizoaffective disorder and first-episode psychosis (STEPWISE): randomised controlled trial. *Br J Psychiatry*, 2019; **214**: 63–73.
27. Larsen JR, Vedtofte L, Jakobsen MSL, *et al.* Effect of liraglutide treatment on prediabetes and overweight or obesity in clozapine- or olanzapine-treated patients with schizophrenia spectrum disorder: a randomized clinical trial. *JAMA Psychiatry*, 2017; **74**: 719–28.

28. Lean ME, Leslie WS, Barnes AC, *et al.* Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet (London, England)*, 2018; **391**: 541–51.
29. Rush AJ, Trivedi MH, Wisniewski SR, *et al.* Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*, 2006; **354**: 1231–42.
30. Marengoni A, Onder G, Degli Esposti L, *et al.* Adherence to selective serotonin and serotonin-norepinephrine reuptake inhibitor prescriptions affects overall medication adherence in older persons. *J Clin Psychiatry*, 2016; **77**: 1712–18.
31. Katon WJ, Von Korff M, Lin EH, *et al.* The Pathways Study: a randomized trial of collaborative care in patients with diabetes and depression. *Arch Gen Psychiatry*, 2004; **61**: 1042–9.
32. Hermanns N, Schmitt A, Gahr A, *et al.* The effect of a Diabetes-Specific Cognitive Behavioral Treatment Program (DIAMOS) for patients with diabetes and subclinical depression. *Diabetes Care*, 2015; **38**: 551–60.
33. Moulton CD, Hopkins CWP, Ismail K, Stahl D. Repositioning of diabetes treatments for depressive symptoms: a systematic review and meta-analysis of clinical trials. *Psychoneuroendocrinology*, 2018; **94**: 91–103.

15.12

Microvascular Complications of Diabetes

15.12.1 Pathogenesis of Microvascular Complications

Angela Shore

Introduction

Diabetes is associated with vascular complications which affect both the large and small blood vessels. The major macrovascular complications include accelerated cardiovascular disease (CVD) and cerebrovascular disease; patients with diabetes have, for example, a 2–3-fold increase risk of myocardial infarction and CVD accounts for more than 50% of the mortality in this patient group [1].

This chapter concentrates on the complications of diabetes which involve the small blood vessels (the microvasculature comprising small arteries, arterioles, capillaries and venules) and their intimate relationship with the organ in which they reside.

By the time **clinically detectable** microvascular complications are present, an individual may have had many years during which the function of the microvasculature has been abnormal and the surrounding tissue has been challenged. Normal microvascular function is important as it maintains the delivery of oxygen, other nutrients and hormones; ensures removal of waste products; maintains the filtration barrier between the blood and tissue; and controls the local tissue blood flow and blood pressure in order to compensate for changes in metabolism, oxygenation, or perfusion pressure.

Microvascular complications are traditionally thought of as those in the eye (see Chapter 15.12.2), kidney (Chapter 15.12.3), and nerves (Chapter 15.12.4) but it is becoming increasingly clear that microvascular complications are not confined to these tissues; they occur in many other tissues, e.g. heart, brain, bone, spinal cord [2–4].

Diabetic complications are the result of the combined effects of the diabetic milieu (hyperglycaemia and altered lipid profiles) and haemodynamic factors on cells, modulated by genetic predisposition and environmental factors. Increasing evidence also suggests that the deleterious effects of hyperglycaemia are exacerbated by an imbalance between factors designed to protect tissues from a variety of toxic effects and the toxic effects of hyperglycaemia *per se*.

These complications of diabetes should not be thought of as arising from the abnormalities of the microcirculation alone but rather they are the result of the disruption of the intimate relationship between the microcirculation and abutting or nearby cells (Figure 15.12.1.1). Recently the American Diabetes Association [5] described *diabetic retinopathy as a highly specific neurovascular complication involving progressive disruption of the interdependence between multiple cell types in the retina.* These cells comprise endothelial cells, pericytes, or vascular smooth muscle cells, retinal glia including Müller cells and astrocytes, neuronal processes, and immune cells both microglia and macrophages [6]. Similarly diabetic neuropathy is due to the interactions of the direct effects of the diabetic milieu on the Schwann cell and the axon as well as the more

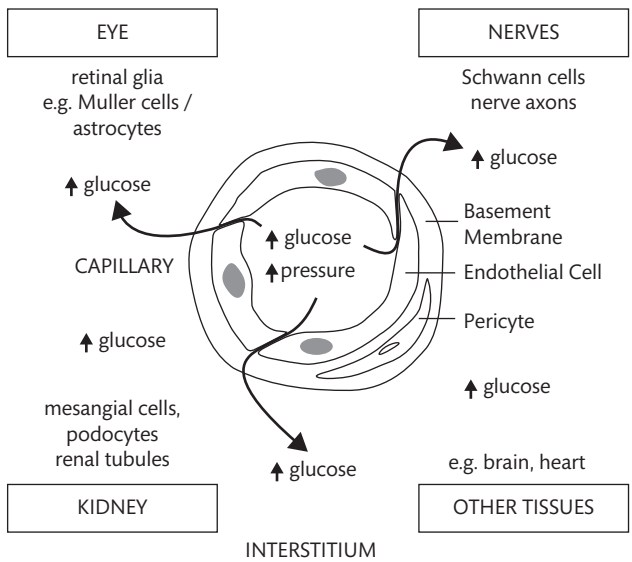


Figure 15.12.1.1 Microvascular complications are the result of disruptions of the intimate relationship between the capillaries and the cells abutting or nearby. These cells also contain the metabolic pathways which are impaired by high glucose and which are responsible for cell dysfunction.

indirect effects on these cells via altered microvascular perfusion [7]; and in the kidney, mesangial cells, podocytes and renal tubules [8] play important roles in diabetic nephropathy in addition to the cells of the vasculature.

Which of these many cell types are the initiators of the complications in diabetes is highly controversial and disputed and is likely to vary according to the tissue examined. One of the main problems in trying to establish which functional abnormality occurs first (e.g. nerve or blood vessel dysfunction) is the intertwined nature of the many components and the fact that the different techniques used to detect these earliest functional abnormalities vary in their ability to detect very small changes. It is however becoming increasingly recognized that many of these 'support' cells, for example Schwann cells or Müller cells contain the metabolic pathways which have been implicated in microvascular complications for many years, and thus they may play direct roles in the development of complications (Figure 15.12.1.1).

It is difficult to give a detailed overview of the pathogenesis of microvascular complications in diabetes as the mechanistic pathway(s) involved varies according to:-

- whether complications are just being initiated, being sustained or in a progressive phase;
- which organ (e.g. kidney, eye, etc.) is being considered and the roles of the vasculature and effects of other cells in the organ;
- whether the patient is receiving treatment which modifies the pathogenic mechanisms.

This chapter thus describes the potential factors involved in the pathogenesis of complications and some of the mechanistic pathways. Further details are available in recent reviews [9]. Subsequent chapters may add specific details about the pathogenesis of an individual complication (e.g. Chapter 15.12.3).

Factors Involved in the Pathogenesis of Microvascular Complications

Metabolic

The Effect of Hyperglycaemia

There is no doubt that the hyperglycaemia associated with diabetes is a major contributing factor but it is not, on its own, sufficient to explain all diabetes complications (see Figure 15.12.1.2). Trial evidence has shown that intensive metabolic control can reduce microvascular complications in both type 1 (DCCT) and type 2 diabetes (UKPDS). The follow-up studies of DCCT and UKPDS show clearly that early intensified glycaemic control can protect against future diabetic complications even when glycaemic control is no longer so intensive—the so-called metabolic memory [10–14].

Higher glycaemic variability, defined as greater pre- and postprandial glucose excursions or more variation of HBA1c over time, predict increased patient risk for cardiovascular complications. Higher glycaemic variability also likely increases the risk of diabetic

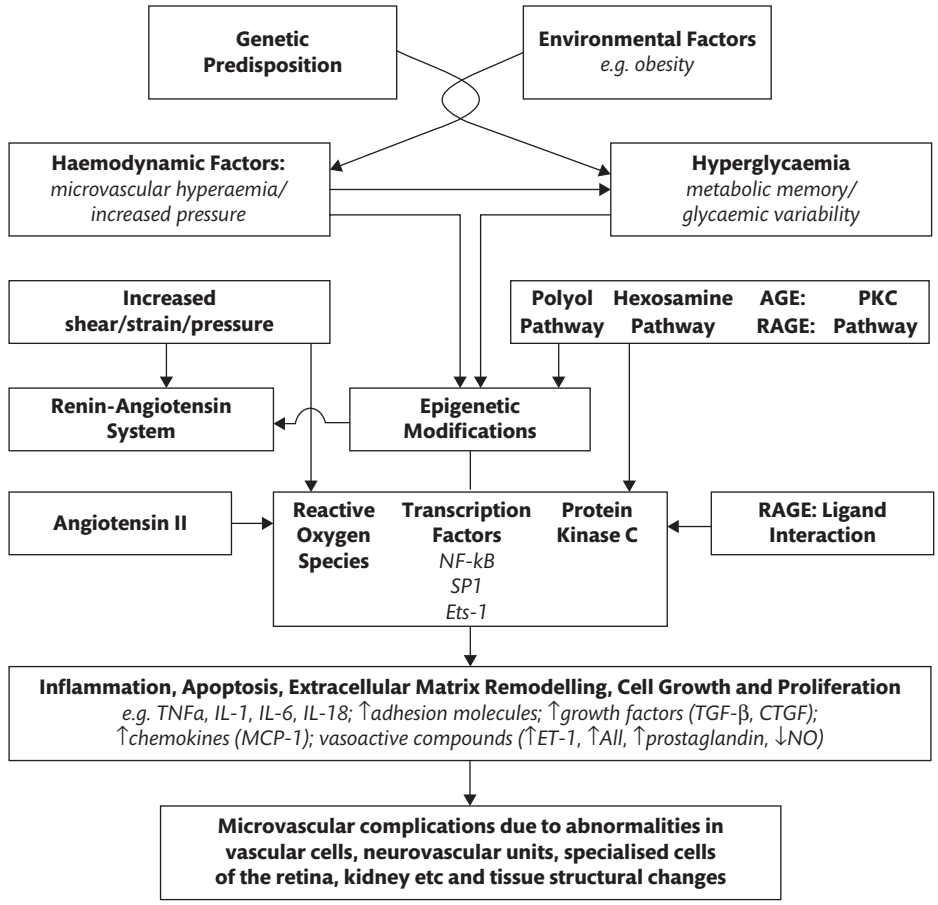


Figure 15.12.1.2 The factors involved in the pathogenesis of microvascular complications.

microvascular complications in patients with diabetes as oxidative stress is increased by acute glucose fluctuations.

Classical pathways [9] implicated in the hyperglycaemic effects include:

- An increased glucose flux through the polyol pathway;
- Increased activity of the protein kinase C (PKC) pathway;
- Increased activity of the hexosamine pathway;
- Accumulation of advanced glycation end-products.

Many of the cells involved in complications respond to an increase in blood glucose by an increase in intracellular glucose. This leads to an increased activity in the glycolysis pathway. Cell glucose is first phosphorylated then converted to fructose 6-phosphate, glyceraldehyde 3-phosphate and dihydroxyacetone phosphate, precursors of diacylglycerol (**DAG**), and at the end of glycolysis, pyruvate is increased and increases activity though the tricarboxylic acid cycle (TCA) cycle in mitochondria.

The high intracellular glucose also results in an increased glucose flux through the polyol pathway which increases the amount of glucose converted to sorbitol and fructose. This leads to a relative depletion of NADPH and activation of NADH oxidase, a reduction in the regeneration of the antioxidant glutathione and increased 3-deoxyglucosone a precursor for advanced glycation end-products (AGE) and histone glycation [15]. Activation of the polyol pathway thus increases oxidative stress and enhances AGE formation. The polyol pathway is present in many cells relevant to diabetic microangiopathy including endothelial cells, Schwann cells, Müller cells, and pericytes.

Increased activity of the PKC pathway which converts glyceraldehyde-3P, by **DAG**, **reactive oxygen species (ROS)** generated from other pathways or by hyperglycaemia causing **transcriptional upregulation of PKC** lead to:-

- Endothelial dysfunction with elevation of the vasoconstrictor endothelin-1 (ET-1), and reduction of the vasodilatory nitric oxide (NO) production, and reduced vasodilatory prostaglandins, and increased permeability via an increase in the vascular endothelial growth factor (VEGF) all important mechanisms in diabetic retinopathy and nephropathy;
- Local inflammation with increases in the transcription factor NF-κB;
- Cell apoptosis, e.g. retinal pericytes, via multiple mechanisms including NF-κB, NADPH oxidase, SHP-1 and MAP kinase;
- Activation of growth factors such as transforming growth factor-β (TGFβ) or connective tissue growth factor (CTGF).

Increased flux through the hexosamine pathway is caused by hyperglycaemia and increases gene expression, for example of plasminogen activator inhibitor-1 (PAI-1) and TGF-β. The pathway converts fructose-6-phosphate to glucosamine-6 phosphate by glutamine-fructose-6 phosphate aminotransferase. Glucosamine-6 phosphate is then converted to uridine diphosphate-N-acetyl glucosamine (UDP-GlcNAc) which, after conversion by 0-GlcNac transferase, activates transcription factors such as Sp1. Insulin vasodilatation and signalling are also impaired by N-acetyl glucosamine [16].

Hyperglycaemia accelerates the formation and **accumulation of AGEs**. AGEs are the products of non-enzymatic glycation and/or oxidation of proteins and lipids [17]. AGEs are slowly metabolized and are postulated to contribute to the metabolic memory of diabetes complications. The receptor for AGE, RAGE, is widely distributed in vascular and inflammatory cells, Müller cells, podocytes, pericytes, neurons, and microglia. RAGE is upregulated in diabetes. The interaction of AGE with their signal–transduction receptor RAGE contributes to many aspects of diabetic microangiopathy particularly inflammation, oxidative stress, and functional effects e.g. in:

- Endothelium—AGE/RAGE upregulates adhesion molecule expression and increases the adherence of leukocytes, increases oxidative stress, increases cytokines and growth factors through activation of NF-κB, increases VEGF, and thus impairs the barrier function of the vessel.
- Nerves—AGE/RAGE impairs perfusion in the vasa nervorum by increasing oxidative stress and prothrombotic signals, causes pro-inflammatory signals in Schwann cells.
- Glomerular—AGE/RAGE increases vascular permeability and proteinuria.
- Macrophages and monocytes—AGE/RAGE stimulates the generation of cytokines and migration.
- Retina—AGE/RAGE causes glial cell dysfunction leading to inflammatory cytokine and chemokine release, oxidative stress, and aberrant growth factor signalling.

AGE-RAGE interaction, via the transcription factors NF-κB and Sp1 for example, increases gene expression of pro-inflammatory cytokines and generates ROS by NADPH oxidase and mitochondrial pathways. Non-AGE ligands for RAGE are also important in diabetic microangiopathy.

Balancing Toxic Effects of Hyperglycaemia and Protective Factors

The consequences of hyperglycaemia for a particular patient depend on the balance of the toxic effects of hyperglycaemia and the patient's natural protective factors [16]. In the Joslin Clinics cohort of 1000 individuals with more than 50 years of insulin requirement [16] 12% have diabetic nephropathy, and 35% have no retinopathy, nephropathy, or neuropathy. Presence or absence of complications was not related to previous glycaemic control. Glomeruli and other cells from these individuals have revealed protective factors several of which have both protective and toxic effects.

Protective factors include:

- Insulin which, via the IRS/PI3Kinase/Akt pathway, upregulates nitric oxide synthase in endothelial cells, upregulates antioxidant activity haeme oxygenase 1 (HO-1), act as an important cell survival factor (e.g. podocytes) and counteracts leukocyte-endothelial cell adhesion;
- Increased endogenous antioxidant enzymes activity, activation of nuclear factor (erythroid derived 2)-like 2 (nrf2) upregulates HO-1 and enzymes in the glutathione pathway;
- Growth factors, e.g. platelet derived growth factor's (PDGF) important cell survival factor function, TGF-β's anti-inflammatory effect, VEGF's role as a survival factor for endothelial cells and podocytes;
- Endothelial and myeloid progenitor cells which can improve angiogenesis and show potential as protective factors.

Metabolic Memory

Metabolic memory refers to the effect of earlier periods of glycaemic control on subsequent development of microvascular complications. DCCT and 30-year follow-up, the Epidemiology of Diabetes Interventions and Complications trial (EDIC) in type 1 diabetes, and the UKPDS and its 10-year follow-up in type 2 diabetes, clearly demonstrated that intensive glycaemic control early in diabetes reduced subsequent microvascular complications compared to the normal control arm, despite the differences in glycaemic control between the treatment arms only being apparent during the trials themselves and not during considerable follow-up periods. See reviews including Cochrane systematic review [10–14].

Animal models show that glycaemic memory is established very early in diabetes causing cellular changes which are not easy to reverse. The mechanisms involved are several. The key biochemical drivers of metabolic memory appear to be oxidative stress and AGE accumulation with various epigenetic marks such as histone methylations and DNA methylation heavily involved [18, 19]. In vascular smooth muscle cells from the db/db mouse model of type 2 diabetes, an epigenetic link to a persistent pro-inflammatory phenotype associated with a reduction in chromatin histone methylation mark, H3K9me3 was reported. This pro-inflammatory phenotype persisted even after culture of the cells in normal glucose levels for several passages [19]. Monocytes of DCCT trial patients from the conventional control group demonstrated evidence of increased acetylation of H3K9 at promoters of genes related to inflammation, apoptosis, NFκB pathway and ROS which increased activation of these pathways. Epigenetic marks may also contribute to the hyperglycaemia associated endothelial dysfunction, changes to histone methylation marks in human microvascular endothelial cells incubated in high glucose, induced upregulation of NFκB p65 expression and an increase in the transcription of numerous inflammatory cytokines. A genome-wide DNA methylation analysis of blood from individuals with type 1 diabetes with or without nephropathy revealed hypermethylation of UNC13B promoter in nephropathy, a region with effects on glomerular apoptosis. DNA methylation has also been associated with glycaemic memory and progression of diabetic retinopathy in a rat model of diabetes. Thus epigenetic modifications likely have a very important effect on the formation and progression of diabetic complications as well as on glycaemic memory. An important role for these biochemical changes on endothelial progenitor cells (EPCs) has also been proposed. Given that EPCs have an extended lifetime and are important for vascular repair and homeostasis, it is likely that the metabolic memory also impairs the vasoreparative capacities of diabetic tissues.

Microvascular Haemodynamic and Structural Factors

Haemodynamic Factors

Both microvascular blood flow and blood pressure contribute to the pathogenesis of microvascular complications (see Figure 15.12.1.2) [20, 21]. Given that microvascular perfusion is not measured or reported clinically, individuals may be classified as remaining free from microvascular complications despite the fact that microvascular perfusion deficits are present. Indeed microvascular dysfunction precedes diabetes. Abnormal microvascular responses are reported in obesity and prediabetes and may contribute to insulin resistance, for example, by impaired insulin mediated vasodilation and reduced capillary recruitment leading to reduced insulin delivery and glucose uptake.

In individuals with type 2 diabetes, reductions in future microvascular complications were achieved by lowering systemic blood pressure from diagnosis in the UKDPS study, irrespective of the therapy used suggesting a beneficial effect of lowering blood pressure *per se* rather than a therapeutic specific effect.

Early in diabetes, tissue hyperperfusion is common, linked to poor glycaemic control followed later by loss of vascular autoregulation and barrier leakage. In type 1 diabetes, abnormalities of the microvascular response to stress are present within the first year of diagnosis, even in prepubertal children. In type 2 diabetes microvascular dysfunction is considerable even at presentation with disease. Endothelial dysfunction becomes increasingly abnormal with longer disease duration and is most marked in those with poor control.

Capillary hypertension, even in the absence of systemic hypertension, is a feature of poorly controlled short duration type 1 diabetes or in those with microalbuminuria. Raised capillary pressure in patients with type 2 diabetes and hypertension is likely due to impaired pressure autoregulation. Whether capillary rarefaction (reduction in number) is a feature of human diabetes is a matter of controversy. Glomerular capillary hypertension in animal models of diabetes precedes the development of glomerulosclerosis and prevention of this glomerular capillary hypertension by treatment with angiotensin converting enzyme inhibitors prevents diabetic nephropathy emphasizing the importance of capillary hypertension in the pathogenesis of diabetic microangiopathy.

Clinical observations link haemodynamic factors with diabetic microangiopathy. Patients with unilateral renal artery stenosis, for example, develop unilateral diabetic nephropathy, the kidney with the stenosed renal artery being protected both from the abnormalities of pressure and diabetic nephropathy. Similar findings are reported for diabetic retinopathy.

Early hyperaemia and elevations in capillary pressure lead to microangiopathy via the effects of pressure, strain, and shear on the vasculature (endothelial cells, vascular smooth muscle cells and pericytes) and supporting cells such as mesangial cells [20].

These haemodynamic forces:

- increase ROS
- activate PKC
- increase NF kappa Beta (NFκB)
- increase Glut 1
- increase adhesion molecules, e.g. ICAM-1, monocyte chemoattractant protein (MCP)
- activate the renin–angiotensin–aldosterone system (RAAS)
- increase growth factors, e.g. CTGF, TGF-β
- stimulate increased secretion of basement membrane components
- change DNA methylation patterns

In addition to these functional changes to the microvasculature, structural changes to the vessels and associated multiple cell types are described.

Structural Changes in the Microcirculation

In a multitude of tissues **thickening of the capillary basement membrane** is the ultrastructural hallmark of diabetic microangiopathy. It

occurs in muscle capillaries, retinal capillaries, glomerulus, and heart capillaries to name a few. The increased thickness is due to increased synthesis of fibronectin, type IV collagen, and laminin by endothelial cells and pericytes which occurs early in diabetes and is not matched by an equivalent increase in the breakdown of these components. Thickening of the basement membrane occurs prior to clinically detectable complications, it increases the rigidity of capillaries, and impairs the barrier function of the vessel wall due to its altered charge and a more open lattice work structure. In the kidney the glomerular basement membrane is twice as thick as other capillary basement membranes as it is synthesized by both the glomerular endothelial cells and the podocytes. Glomerular basement membrane thickness is increased within two years of diagnosis of diabetes at least in part due to podocyte injury, which causes increased synthesis of extracellular matrix components and dysregulation of matrix degradation. In the pathogenesis of diabetic neuropathy not only is the endoneurial capillary basement membrane thickened prior to clinically detectable neuropathy but the Schwann cell basement membrane increases in thickness and changes composition with proposed consequences for neural transmission.

The basement membrane serves as a substratum for cell attachment and regulates cell survival. Thickening of the capillary basement membrane may alter cross talk between the pericytes and endothelial cells in unspecialized capillaries or between the glial cells (Müller cells or astrocytes) and endothelial cells in the retinal neurovascular unit and impair podocyte survival in the glomerulus.

i) Pericytes are perivascular cells embedded in the basement membrane of capillaries and venules which have long processes which interact with one or many endothelial cells. They likely regulate microvascular blood flow. They contain actin, have the ability to contract and are able to respond to vasoconstrictor and vasodilator agents. The interaction between the endothelial cell and pericytes is important in the maintenance and regulation of vessel function and structure, in the control of the vascular barrier function and in the modulation of the immune response.

In diabetes, loss of pericytes from capillaries occurs early in disease. Pericyte deficient capillaries have been reported in many tissues (e.g. skeletal muscle, retina, skin, kidney, and nerves). Pericyte loss contributes to altered endothelial function, impaired barrier function, and neovascularization. Pericyte loss and subsequent weakening of the capillary may result in microaneurysms which, in the retina at least, are normally located adjacent to areas of acellular capillaries and non-perfusion. High glucose induced migration of renal pericytes from the capillary into the interstitial space and subsequent differentiation into collagen producing myofibroblasts has been implicated in tubulointerstitial fibrosis and renal dysfunction [22].

The number of capillary microaneurysms and non-perfused capillaries increases with the severity of the microangiopathy, and together with impaired vascular reactivity of larger vessels, explain the inadequate perfusion of severe microangiopathy.

Diabetic retinopathy histologically is ultimately associated with acellular non-perfused capillaries in the retina due to the death of both endothelial cells and pericytes. These non-perfused capillaries are simply tubes of basement membrane. Delivery of oxygen to such areas is compromised and results in activation of inflammatory cascades.

Diabetic nephropathy is accompanied by several main histological changes;

- extensive mesangial expansion and glomerulosclerosis;
- thickening of the glomerular basement membrane;
- broadening (effacement) of the podocyte foot processes. This reduces filtration slit area between the podocytes and increases macromolecule passage;
- abnormalities of the glycocalyx and glomerular endothelial cells.

The renal structural changes of early diabetes are heterogenous. Abnormalities to the glomerular endothelial cell surface layer (glycocalyx) have recently been reported to contribute to the development of diabetic microalbuminuria [23].

Genetic Susceptibility

Genetic factors contribute to microvascular complications: There is familial clustering of both diabetic nephropathy and severe, vision threatening forms of diabetic retinopathy as reported in the Diabetes Control and Complications Trial. The estimated heritability of albumin excretion rate is ~20–40% and diabetic siblings of patients with diabetic nephropathy have a twofold increased risk of diabetic nephropathy.

Many genetic risk variants for microvascular complications have been reported by candidate gene studies or by genome-wide association studies; however very few have shown robust replication. In the biggest study in type 1 diabetes to date, sets of alleles known to increase body mass index, and those known to increase the risk of type 2 diabetes, were associated with the risk of diabetic nephropathy [24]. Pathway analysis implicated ascorbate and aldarate metabolism, and pentose and glucoronate interconversions. In the largest T2DM study to date, the genetic analysis provided support for the importance of BMI and identified some variants associated with GFR and microalbuminuria in diabetic nephropathy [25]. Further studies are needed, accompanied by more intensive phenotyping of disease in particular using biomarkers which can identify individuals destined to develop the complications prior to any clinical evidence [26]. It is highly likely that epigenetic changes contribute to the formation of both diabetic nephropathy and retinopathy.

Environmental Factors

Environmental and behavioural factors, for example, those influencing BMI and levels of exercise, can modulate many of the pathways through which hyperglycaemia, haemodynamics, and genetic susceptibility lead to microvascular complications of diabetes and thus contribute to the findings of links between the two. Obesity is, for example, accompanied by increases in ROS, dysregulation of mitochondrial biogenesis, mitophagy, and dynamics, downregulation of both the AMP-activated protein kinase (AMPK) and sirtuins and activation of the renin–angiotensin–aldosterone system.

Pathways Involved in the Formation of Microvascular Complications

Reactive Oxygen Species (ROS)

ROS encompasses non-radical species such as hydrogen peroxide and free radical species such as hydroxyl radical and superoxide

anion (see Figure 15.12.1.2). ROS are proposed to act as a common pathway in the pathogenesis of diabetic complications, though the exact species involved has caused dispute over recent years [27]. ROS are produced physiologically in health as normal homeostatic functions. However when the balance between production and removal of ROS is altered, oxidative stress can result. At high concentrations, ROS can cause tissue damage including modifications to proteins, DNA, and lipids. Reactive nitrogen species, e.g. derivatives of NO also cause tissue damage.

In diabetic complications [27], ROS are increased in, for example, endothelial cells, vascular smooth muscle cells, cardiac myocytes, mesangial and tubular epithelial cells, retinal glial cells, Schwann cells by factors including:

- hyperglycaemia
- haemodynamic factors
- inflammatory processes.

There are many pathways which contribute to the generation of ROS in tissues in which diabetic microvascular complications arise.

These include:

- enzymes such as nicotinamide adenine dinucleotide phosphate oxidase (NADPH oxidase complex (NOX)), xanthine oxidase, 12/15 lipoxygenase, myeloperoxidase;
- altered mitochondrial ROS production and clearance due to increased oxidative phosphorylation;
- cytochrome p450;
- uncoupling of nitric oxide synthase so that superoxide anions rather than nitric oxide are produced.

Whether the mitochondrial electron transport chain is involved in diabetic complications has been at the centre of controversy recently but the consensus seems to be that changes in mitochondrial redox status by impaired electron transport are indeed important in the formation of diabetic microangiopathy even if they are subsequently reduced at later stages of the disease [27].

The antioxidant pathways which remove ROS are often also impaired in diabetes such as heme-oxygenase-1 (HO-1), Superoxide dismutase (SOD), glutathione peroxidases (GPx). Nrf2, a transcription factor, controls baseline and induced levels of antioxidants and is thus an important regulator of ROS and a target for future therapies.

The downstream redox sensitive processes which are affected by increases in ROS are numerous and include major metabolic pathways, RAGE/AGE, PKC, renin–angiotensin–aldosterone system (RAAS), cell growth, apoptosis, migration, extracellular matrix modelling, growth factors, vascular function, and permeability. The exact mechanisms and pathways involved in ROS mediated damage differs in retinopathy, nephropathy, neuropathy, and other tissues.

Inflammation

Many mechanisms, mediators and signalling cascades contribute to inflammation in diabetic complications with some variation of pathways and the cells involved in different tissues [28, 29].

Inflammation, recruitment, and activation of immune cells to clear damage to tissue or pathogens, is beneficial in the acute phase; however, chronic inflammation may cause irreversible pathological tissue changes. Pattern recognition receptors such as RAGE are involved as one of the receptors in inflammatory responses. RAGE-ligand interactions both initiate and sustain inflammation, NF-κB is activated and stimulates transcription of pro-inflammatory genes, acute phase proteins, and chemokines such as IL-6, TNF-α, monocyte chemoattractant protein (MCP-1) which all play a major role in the recruitment and activation of monocytes and leukocytes and the subsequent inflammatory response, all are implicated.

Inflammatory mechanisms have been implicated as important early pathogenic factors in retinopathy, neuropathy, and nephropathy. This is strengthened by the finding that modification of inflammation is associated with reduced complications in patients [30]. Circulating markers of inflammation are increased in both type 1 and type 2 diabetes and correlate with albuminuria and risk of progression towards End Stage Renal Disease.

Activation of macrophages, neutrophils, resident immune cells such as microglia or dendritic cells, Schwann cells, endothelial cells, or adipocytes via pattern recognition receptors [28–30] leads to:

- increased NF-κB activity, mitogen-activated protein kinases (MAPKs);
- induction of pro-inflammatory genes;
- release of pro-inflammatory cytokines/chemokines with effects locally and systemically including immune cell recruitment.

Exactly which cytokines are released, in what proportions, and the downstream effects appear to vary according to the stage of disease and the tissue examined. TNFα, IL-1, IL-6, and IL-18 are increased early in diabetic nephropathy; they increase adhesion molecule expression (e.g. ICAM-1), alter glomerular haemodynamics and increase glomerular permeability, increase prostaglandin secretion, induce mesangial cell proliferation, cause apoptosis, generate ROS, and augment the inflammatory response by stimulating transcription factors and other growth factors. Similar cytokines are also increased in retinopathy although VEGF is induced by hypoxia through HIF-1 and contributes to increased vascular permeability [31, 32].

The chemokines such as CCL-2, CCL-4, CXCL-9, and CXCL-10 are elevated in vitreous samples from patients with proliferative diabetic retinopathy as is the MCP-1. All are involved in immune stimulation and recruiting and activating microglia (resident immune cells in retina) and leukocytes. MCP-1 also contributes to fibrosis and angiogenesis [30]. Many of these chemokines are also elevated in patients with diabetic nephropathy.

The Renin–Angiotensin–System (RAS)

RAS is upregulated by both hyperglycaemia and by haemodynamic factors.

RAS is expressed locally in tissues, e.g. in the eye it is present in neuronal (Müller cells, ganglion cells) and vascular cells. The RAS components are produced locally in the tissues rather than being filtered from the circulation. Increased levels of Angiotensin II (Ang II) and other RAS components have been found in the vitreous fluid in patients with proliferative retinopathy compared to controls. Research suggests these locally produced components of RAS especially Ang II are involved in the pathogenesis of complications

and potentially that chymases rather than angiotensin converting enzymes (ACE) are the primary mechanism of intracellular Ang II generation. Chymases convert Ang I to Ang II at a much higher rate than ACE and are not inhibited by ACE inhibitors. In the eye some cells will be exposed to circulating Ang II (e.g. endothelial cells), while cells behind the vascular barrier will be primarily exposed to locally derived Ang II. Kidney Ang II is also locally generated as well as circulating.

Angiotensin II (Ang II) has haemodynamic and non-haemodynamic effects, generating its actions via ROS and stimulation of pro-inflammatory transcription factors such as NF-κB and Ets-1.

Ang II increases:

- glomerular hypertension, by preferentially vasoconstricting the efferent capillary;
- cellular apoptosis which decreases pericyte viability and impairs vascular regulation;
- vascular permeability, by stimulating expression and secretion of VEGF, leukotriene C4, PGE2, and PGE1;
- rolling and sticking leukocytes by upregulating E-selectin VCAM-1 and ICAM-1;
- recruitment of monocytes, T lymphocytes, and neutrophils by stimulating secretion of monocyte chemotactic protein MCP-1, cytokine-inducible neutrophil chemoattractant, keratinocyte-derived chemokine, and macrophage inflammatory protein (MIP)-2.

Ang II also:

- impairs tissue repair;
- mediates fibrosis via TGF-β and CTGF;
- induces apoptosis/proliferation;
- stimulates ECM deposition.

REFERENCES

1. Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in non diabetic subjects with and without prior myocardial infarction. *N Eng J Med*, 1998; **339**: 229–34.
2. Moheet A, Mangia S, Seaquist ER. Impact of diabetes on cognitive function and brian structure. *Ann N Y Acad Sci*, 2015; **1353**: 60–71.
3. Levelt E, Piechnik SK, Liu A, *et al.* Adenosine stress CMR T1-mapping detects early microvascular dysfunction in patients with type 2 diabetes mellitus without obstructive coronary artery disease. *J Cardiovasc Magn Reson*, 2017; **19**: 81.
4. Bone disease in diabetes: another manifestation of microvascular disease. Shanbhoque VV, Hansen S, Brixham K, Hermann AP. *Lancet Diabetes Endocrinol*, 2017; **5**: 827–38
5. Solomon SD, Chew E, Duh EJ, *et al.* Diabetic retinopathy: a position statement by the American Diabetes Association. *Diabetes Care*, 2017; **40**: 412–18.
6. Lecher J, O'Leary OE, Stitt AW. The pathology associated with diabetic retinopathy. *Vision Res*, 2017; **139**: 7–14.
7. Goncales NP, Voegter CB, Andersen H, Ostergaard L, Calcutt NA, Jenson TS. Schwann cell interactions with axons and microvessels in diabetic neuropathy. *Nat Rev Neurol*, 2017; **13**: 135–47.
8. Qi W, Li Q, Gordin D, King GL. Preservation of renal function in chronic diabetes by enhancing glomerular glucose metabolism. *J Molec Med*, 2018; **96**: 373–81.
9. Forbes JM, Cooper ME. Mechanisms of diabetic complications. *Physiol Rev*, 2013; **93**: 137–88.
10. The Diabetes Control and Complications Trial Research Group. The effect of the intense treatment of diabetes on the development and progression of long term complications in insulin dependent diabetes mellitus. *N Engl J Med*, 1993; **329**: 977–86.
11. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-Year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*, 2008; **359**: 1577–89.
12. The Diabetes Control and Complications trial/Epidemiology of Diabetes Interventions and Complications (EDIC) Research group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the EDIC study. *JAMA*, 2003; **290**: 2159–67.
13. Lachin JM, Grp DER. Effect of intensive diabetes therapy on the progression of diabetic retinopathy in patients with type 1 diabetes: 18 years of follow up in th DCCT/EDIC. *Diabetes*, 2015; **64**: 631–42.
14. Fullerton B, Jeitler K, Seitz M, Horvath K, Berghold A, Siebenhofer A. Intensive glucose control versus conventional glucose control for type 1 diabetes. *Cochrane Database Syst Rev*, 2014; **2**: CD009122.
15. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*, 2005; **54**: 1615–25.
16. Rask-Madson C, King GL. Vascular complications of diabetes: mechanisms of injury and protective factors. *Cell Metab*, 2013; **17**: 20–33.
17. Rhee SY, Kim YS. The role of advanced glycation end products in diabetic vascular complications. *Diabetes Metab J*, 2018; **42**: 188–95.
18. Rajasekar P, O'Neill CL, Eeles L, Stitt AW, Medina RJ. Epigenetic changes in endothelial progenitors as a possible cellular basis for glycaemic memory in diabetic vascular complications. *J Diabetes Res*, 2015; **2015**: 436879.
19. Villeneuve IM, Reddy MA, Lanting II, Wang M, Meng I, Natarajan R. Epigenetic histone H3 lysine 9 methylation in metabolic memory and inflammatory phenotype of vascular smooth muscles in diabetes. *PNAS*, 2008; **105**: 9047–52.
20. Shore AC. The microvasculature in type 1 diabetes. *Semin Vasc Med*, 2002; **2**: 9–20.
21. Stehouwer CDA. Microvascular dysfunction and hyperglycaemia: a vicious cycle with widespread consequences. *Diabetes*, 2018; **67**: 1729–41.
22. Harrell CR, Markovic BS, Fellabaum C, Arsenijevic A, Djonov V, Volarevic V. Molecular mechanisms underlying therapeutic potential of pericytes. *J Biomed Sci*, 2018; **25**: 21.
23. Satchell SC, Tooke JE. What is the mechanism of microalbuminuria in diabetes: a role for the glomerular endothelium? *Diabetologia*, 2008; **51**: 714–25.
24. Sandholm N, Zuydam NV, Ahlqvist E, *et al.*, on behalf of the SUMMIT consortium. The genetic landscape of renal complications in Type 1 diabetes. *J Am Soc Nephol*, 2017; **28**: 557–74.
25. van Zuydam NR, Ahlqvist E, Sandholm N, *et al.* A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes*, 2018; **67**: 1414–27.

26. Ahlqvist E, van Zuydam NR, Groop LC, McCarthy MI The genetics of diabetic complications. *Nat Rev Nephrol,* 2015; **11**: 277–87.
27. Pickering RJ, Rosado CJ, Sharma A, Buksh S, Tate M, de Haan JB. Recent novel approaches to limit oxidative stress and inflammation in diabetic complications. *Clin Transl Immunology*, 2018; **7**: e1016.
28. Bondeva T, Wolf G. Reactive oxygen species in diabetic nephropathy: friend or foe. *Nephrol Dial Transplant*, 2014; **29**: 1998–2003.
29. Rubsam a, Parikh S, Fort PE. Role of inflammation in diabetic retinopathy. *Int J Mol Sci*, 2018; **19**: 942.
30. Navarro JF, Mora C. Role of inflammation in diabetic complications. *Nephrol Dial Transplant*, 2005; **20**: 2601–4.
31. Altmann C, Schmidt MHH. The role of microglia in diabetic retinopathy: inflammation, microvasculature defects and neurodegeneration. Int J Mol Sci, 2018; **19**: E110.
32. Marquez E, Riera M, Pascual J, Soler J. Renin–angiotensin system within the diabetic podocyte. *Am J Physiol Renal Physiol*, 2015; **308**: F1–F10.

15.12.2 Retinopathy

Peter H. Scanlon

Definition of Diabetic Retinopathy

Diabetic retinopathy refers to the microvascular complication of diabetes with pathology of the capillaries, arterioles and venules in the retina and the subsequent effects of leakage from or occlusion of the small vessels.

The Normal Eye

The eye is an approximate globe consisting of three concentric major layers of tissue. The outermost layer is the white protective sclera except anteriorly where it becomes transparent to allow light to enter through the cornea. The middle layer is the uvea which comprises the iris, ciliary body, and choroid. The innermost layer in the posterior section is the retina which contains the photoreceptor cells. The anterior part of the globe is filled with aqueous humour and the posterior part is filled with the vitreous jelly, both of which are transparent to allow passage of light through to the retina. The majority of the retina, excepting the central foveal region, is provided with oxygen from the retinal vascular circulation. The central foveal region is thinner and has no direct retinal blood supply. Lying between the choroid and the retina are the very thin Bruch's membrane and the retinal pigment epithelium. The supply of oxygen and nutrients is provided through diffusion from the choriocapillaris, which has small gaps (fenestrations) between the cells in the walls of smaller choroidal vessels, and through contact with the retinal pigment epithelium. The blood vessels in the retina exhibit tight junctions between adjacent cells, which prevent leakage of blood components, fluids, chemicals, or drugs through the retinal blood vessel wall maintaining the *blood–retinal barrier*.

Pathology of Diabetic Retinopathy

The histological changes that have been described in diabetic retinopathy are:

1. Thickening of the basement membrane
2. Pericyte loss
3. Loss of epithelial tight junctions
4. Loss of endothelial cells
5. Smooth muscle cell death
6. Capillary weakening
7. Increased capillary permeability
8. Capillary occlusion
9. Microaneurysm formation—this has always been considered one of the earliest lesions in diabetic retinopathy and the hallmark of the condition.

The production of vascular endothelial growth factor (VEGF) during the pathological processes of diabetic retinopathy and the beneficial effects of blocking the effects with VEGF inhibitor drugs have become increasingly apparent.

Modifiable Risk Factors for Progression of DR

a) Systemic hypertension

Although there had been previous descriptions of the effect of hypertension on diabetic retinopathy progression, Joner [1] studied a nationwide cohort of 600 young people with type 1 diabetes in Norway in 1992 and found hypertension to be a significant contributor to the occurrence and progression of diabetic retinopathy in a multiple logistic regression model. Matthews [2] described the UK Prospective Diabetes Study (UKPDS) results

in type 2 diabetes that by 4.5 years after randomization, there was a highly significant difference in microaneurysm count with 23.3% in the tight BP control group and 33.5% in the less tight BP control group having 5 or more microaneurysms (relative risk [RR], 0.70; $P = 0.003$).

More recently the ACCORD [3] and the ADVANCE [4] trials did not show any benefits for retinopathy of more intensive blood pressure lowering, presumably because the blood pressure in the control arms was relatively well controlled compared to the earlier studies.

b) Glucose control.

In type 1 diabetes, the Diabetes Control and Complications Trial (DCCT) [5, 6] demonstrated the beneficial effects of glucose control. After a mean follow-up of 6.5 years in the primary-prevention cohort, intensive therapy reduced the risk for the development of DR by 76% (CI 62–85%), compared with conventional therapy. For the secondary intervention cohort, intensive therapy slowed the progression of DR by 54% (CI 39–66%) and reduced the development of proliferative diabetic retinopathy (PDR) or severe non-proliferative diabetic retinopathy (NPDR) by 47% (CI 14–67%).

In type 2 diabetes, the United Kingdom Prospective Diabetes Study UKPDS [7, 8] demonstrated that there was a 25% (CI 7–40%) risk reduction in the intensive group in development of diabetic retinopathy.

c) Blood lipids.

The role of lipids in the genesis of retinopathy is less clear than that of B/P and glucose control. In 1996 Chew [9] reported an association of elevated serum lipid levels with retinal hard exudates in diabetic patients from the Early Treatment Diabetic Retinopathy Study.

In a recent Health Technology Assessment (HTA) project [10] looking at identifying low risk individuals whose screening risk could be safely extended, the total serum cholesterol (per 1 mmol/L) was a risk factor for progression to screen positive maculopathy but not for progression to pre-proliferative or proliferative DR. The hazard ratio for progression to screen positive maculopathy was 1.14 (95% CI 1.06–1.22).

The FIELD [11] and ACCORD [3] trails suggested that fenofibrate may have beneficial effects on preventing and slowing progression of retinopathy [3, 11], although the mechanism may not be through modification of lipid levels as had previously been thought.

d) Smoking.

In type 1 diabetes Muhlhauser [12] and Karamanos [13] reported smoking to be a risk factor for progression. In type 2 diabetes the evidence is controversial and it may protect against the progression of retinopathy in some patients despite the fact that mortality from cardiovascular disease is increased.

Non-Modifiable Risk Factors for DR

These include:

a) Duration of diabetes:

Numerous studies have reported [14] on the effect of duration of diabetes. Most recently we have reported [15] on the retinopathy levels in 2125 children with diabetes screened for the first time at age 12 or 13. In those diagnosed with diabetes at 2 years of age or less the proportion with retinopathy in one eye or both eyes was 20% and 11%, respectively, decreasing to 8% and 2% in those diagnosed between 2 and 12 years ($P < 0.0001$)

b) Age:

In type 1 diabetes the relationship with age is complex as described [15] earlier in the younger age group where those diagnosed less than 2 years old had disproportionately higher levels of retinopathy at aged 12 years when they were first screened.

However, in type 2 diabetes in the UKPDS [16], in those who already had retinopathy, progression was associated with older age.

c) Genetic predisposition:

Hietala [17] found a familial clustering of proliferative retinopathy in patients with type 1 diabetes. The observation could not be accounted for by conventional risk factors, suggesting a genetic component in the pathogenesis of proliferative retinopathy in type 1 diabetes.

d) Ethnicity:

Simmons [18] compared ethnic differences in Auckland, New Zealand and found that moderate or more severe retinopathy is more common in Polynesians than Europeans and this difference could not be explained by differences in diabetes duration, insulin therapy, the extent of renal disease, blood pressure, or glycaemic control.

Other Common Conditions Affecting the Progression of DR

Pregnancy

Progression of diabetic retinopathy may occur during pregnancy. The known risk factors for progression of diabetic retinopathy in pregnancy are:

a) Pregnancy [19, 20] is independently associated with progression of diabetic retinopathy
b) Baseline severity of retinopathy [21–23]
c) Poor metabolic control at conception [21]
d) Rapid improvement of glycaemic control [20–23]
e) Poor metabolic control during pregnancy or the early postpartum period [19, 20, 23, 24]
f) Duration of diabetes [23, 25, 26]
g) Chronic hypertension and pregnancy-induced hypertension [23]

Renal Failure

Sakata [27] described four patients who exhibited the greatest decline in the estimated glomerular filtration rate (12.3–23.5 ml/min/1.73 m^2/year) had several clinical features in common, including marked retinopathy. It is common in diabetic retinopathy clinics to find deterioration in retinopathy with decline in renal function and there is often (but not always) an improvement once renal dialysis has commenced.

The Early Worsening Phenomenon

In 1998 the DCCT described the effect of early worsening of diabetic retinopathy at the 6- and/or 12-month visit in 13.1% of 711

patients assigned to intensive treatment. Early worsening led to high-risk proliferative retinopathy in two patients in the DCCT who had poorer control on entering the trial. However, the long-term benefits of intensive insulin treatment greatly outweighed the risks of early worsening.

The Early Treatment Diabetic Retinopathy Study (ETDRS)

The ETDRS [29] described the progression of diabetic retinopathy in relation to the development of the specific lesions, the most important of which were:

1. Microaneurysms (Ma)
2. Small retinal haemorrhages or dot haemorrhages
3. HMa (haemorrhage/microaneurysm)
4. Flame haemorrhages
5. Blot haemorrhages
6. Hard exudates—often just referred to as exudates (HE)
7. Cotton wool spots (referred to as soft exudates or SE in the ETDRS but this term is now rarely used)
8. Intraretinal microvascular abnormality (IRMA)
9. Venous beading (VB)
10. New vessels at the disc (NVD)
11. New vessels elsewhere (NVE)
12. Vitreous haemorrhage (VH)
13. Preretinal haemorrhage (PRH)

Table 15.12.2.1 shows the progression to PDR according to the ETDRS severity scale [30] and a comparison to English Screening Programme grades.

Screening for Diabetic Retinopathy

Screening for Diabetic Retinopathy commenced in the Four Nations of England, Scotland, Wales, and Northern Ireland between 2002

Table 15.12.2.1 Diabetic retinopathy classification of progression to proliferative DR

ETDRS final retinopathy severity scale	ETDRS (final) grade	Lesions	Risk of progression to PDR in 1 year (ETDRS interim)	English Screening Programme levels
No apparent retinopathy	10 14, 15	DR absent DR questionable		R0 Currently screen Annually
Mild NPDR	20	Micro aneurysms only		R1 Screen annually **Background** microaneurysm(s) Retinal haemorrhage(s) ± any exudate Venous loop
	35 a b c d e	One or more of the following: Venous loops ≥ definite in 1 field SE, IRMA, or VB questionable Retinal haemorrhages present HE ≥ definite in 1 field SE ≥ definite in 1 field	Level 30 = 6.2%	
Moderate NPDR	43a b	H/Ma moderate in 4–5 fields or severe in 1 field or IRMA definite in 1–3 fields	Level 41 = 11.3%	R2 Refer to ophthalmologist **Pre-proliferative** venous beading venous reduplication intraretinal microvascular abnormality (IRMA) multiple deep, round, or blot haemorrhages
Moderately severe NPDR	47 a b c d	Both level 43 characteristics – H/Ma moderate in 4–5 fields or severe in 1 field and IRMA definite in 1–3 fields **or** any one of the following: IRMA in 4–5 fields HMA severe in 2–3 fields VB definite in 1 field	Level 45 = 20.7%	
Severe NPDR	53 a b c d	One or more of the following: ≥ 2 of the 3 level 47 characteristics H/Ma severe in 4–5 fields IRMA ≥ moderate in 1 field VB ≥ definite in 2–3 fields	Level 51 = 44.2% Level 55 = 54.8%	
Mild PDR	61a b	FPD or FPE present with NVD absent or NVE = definite		**R3** Proliferative DR
Moderate PDR	65a b	1) NVE ≥ moderate in 1 field or definite NVD with VH and PRH absent or questionable or 2) VH or PRH definite and NVE < moderate in 1 field and NVD absent		
High-risk PDR	71 a b c d	Any of the following: 1) VH or PRH ≥ moderate in 1 field 2) NVE ≥ moderate in 1 field and VH or PRH definite in 1 field 3) NVD = 2 and VH or PRH definite in 1 field 4) NVD ≥ moderate		
High-risk PDR	75	NVD ≥ moderate and definite VH or PRH		
Advanced PDR	81	Retina obscured due to VH or PRH		

and 2004 with full population coverage being achieved in England by 2008. All Four Nations use digital photography as their primary screening tool with England and Wales dilating all people with diabetes, Northern Ireland dilating over the age of 50 years and Scotland using a staged approach with dilation if the screener considers the image quality to be poor. In 2014 it was reported [31] that Diabetic Retinopathy was no longer the leading cause of blindness in the working age group in England and Wales and the introduction of the screening programmes was considered to have made a significant contribution to this achievement.

The English Screening Classification for DR Progression

The English Screening Programme takes two 45-degree fields and uses the classification shown in the right-hand column of Table 15.12.2.1 for grading the images. Each eye has a retinopathy R and a maculopathy M grade, and the outcome is dependent on the grade of the worst eye.

R0M0 No Diabetic Retinopathy. No maculopathy
R1M0. Background DR. No maculopathy
R1M1. Background DR. No maculopathy
R1M1 is Background DR. Maculopathy
R2M0. Pre-proliferative DR. No maculopathy
R2M1. Pre-proliferative DR. Maculopathy
R3M0. Proliferative DR. No maculopathy
R3M1. Proliferative DR. Maculopathy
U. Unassessable images

People with diabetes who have unassessable images are referred for slit-lamp biomicroscopy investigation.

Examination Techniques Used to Assess Diabetic Retinopathy

The commonly used examination techniques used are:

1. A carefully taken history.
2. Assessment of visual acuity.
3. Dilation of the pupil with G Tropicamide 1% often with G Phenylephrine 2.5%.
4. Direct ophthalmoscopy has a limited field of view and does not have a good sensitivity and specificity for the detection of sight threatening diabetic retinopathy but is useful for ad hoc detection of diabetic retinopathy.
5. Slit-lamp biomicroscopy is the commonest method used by ophthalmologists to diagnose and monitor retinal disease using an indirect lens such as a 78D lens.
6. Contact lens biomicrosopy is used for more detailed retinal examinations and while conducting laser treatments with a contact lens placed on the eye.
7. Binocular indirect ophthalmoscopy is useful for evaluating the posterior segment and retinal periphery. A larger area can be viewed than with slit-lamp biomicroscopy but this view is less magnified.

Investigations Used to Assess Diabetic Retinopathy

1. Retinal photography—digital colour retinal photography is currently the method of choice for screening for diabetic retinopathy. In the English Screening Programme two 45-degree fields are taken, one centred on the fovea and one centred on the optic disc. In programmes that use non-mydriatic photography such as Scotland, one 45-degree field is used because of the delay in pupil dilation after the flash. All new screening methodologies are tested against a reference standard of seven-field stereophotography. See Figure 15.12.2.1.
2. Scanning laser ophthalmoscopy (SLO)—this method is increasingly being used in the eye clinic environment and they provide an artificial colour image. The SLOs use monochromatic light and provide an artificial colour image by using red, green, and blue lasers sequentially to acquire red, green, and blue reflectance images of the retina and combining these images to produce an artificial colour image. The Optos machine can take a wide 200-degree field although this width of field does limit the central resolution. Other scanning laser ophthalmoscopes such as the Heidelberg Spectralis take a smaller field (e.g. 55 degrees).
3. Scanning confocal ophthalmoscopy (not laser)—some recent machines use the same principle of confocal light but produce a colour image using LED light (e.g. the Eidon by CenterVue and the Zeiss Clarus 500).
4. Fundus fluorescein angiography is a diagnostic procedure in which fluorescein dye (typically 5 ml of 10% sodium fluorescein) is injected in a peripheral vein and a specialized fundus camera or confocal scanning laser ophthalmoscope captures high resolution, rapid sequence photographs, or digital videos to monitor the circulation of the fluorescent dye through the eye. The dye provides a yellow fluorescence with blue light and so the technique involves blue light being transmitted into the eye and yellow filters being inserted for light returning to the camera or SLO. The fluorescein dye can cause side effects in a few individuals such as nausea, vomiting, occasional syncope, skin rashes, and itching. More serious side effects, such as anaphylactic shock, are extremely rare. See Figure 15.12.2.2.
5. Optical coherence tomography (OCT) utilizes time of flight optical interferometry to resolve the different reflective layers of the retina at resolutions typically in the range of 3–8 μm for standard commercial systems. OCT images can be presented as either cross-sectional images or as topographic maps and can be combined with fluorescein angiography. See Figure 15.12.2.3.
6. Optical coherence tomography angiography—advances in OCT technology have seen the development of methods of non-invasive, dye-free angiography. This technique analyses the axial motion associated with blood flow and uses this to generate three-dimensional reconstructions of the vasculature. OCT angiography has a number of advantages over fluorescein angiography—it is non-invasive, does not require the use of fluorescein dye, and is quicker to perform. However, it presents challenges with interpretation because the absence of a vessel may mean that the blood flow is too slow for the technique to detect rather than true absence.
7. See Figure 15.12.2.4.

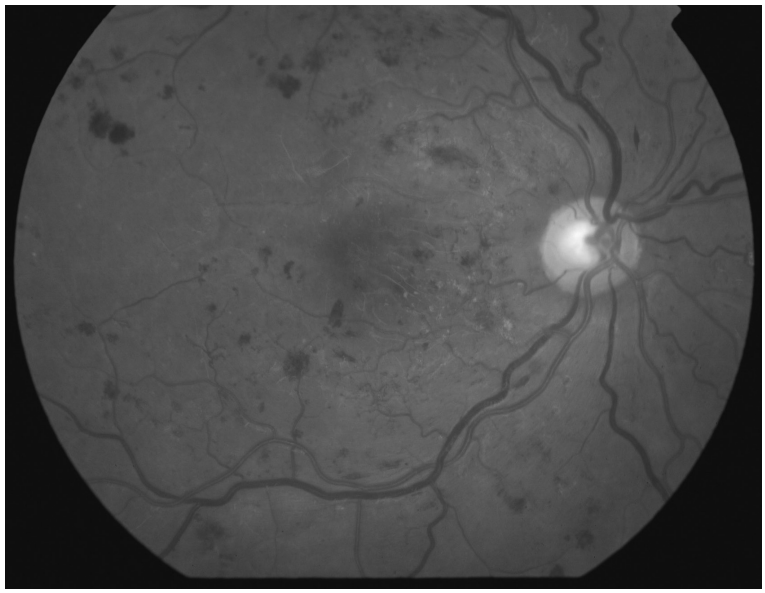


Figure 15.12.2.1 This macular centre digital photograph shows features of R2 preproliferative DR with multiple haemorrhages and IRMA and M1 maculopathy with exudates <1DD from the foveal centre.

8. Ultrasound B scan examination uses high frequency ultrasound typically in the range of 8–15 MHz, to generate images of the eye. Ocular structures are identified by their characteristic reflectivity profiles on the A-scan and B-scans provide cross-sectional images of the eye. This is useful to examine the density and extent of a VH and the presence or absence of a retinal detachment where the retinal view is obscured. This technique is particularly useful in advanced diabetic retinopathy where fibrovascular membranes, tractional retinal detachments, and VHs may coexist and assists in planning vitreoretinal surgery.
9. Visual field perimetry is the systematic measurement of differential light sensitivity in the visual field by the detection of the presence of test targets on a defined background in order to map and quantify the visual field, normally testing each eye independently but binocularly for driving field assessment.

Multidisciplinary Approach

Many patients who have severe diabetic eye disease have suboptimal control of blood glucose, blood pressure, and lipid levels. It

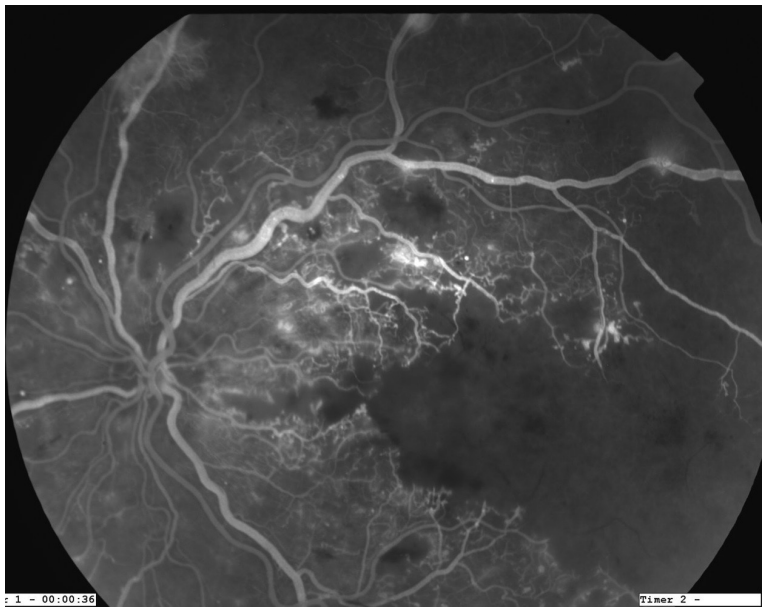


Figure 15.12.2.2 A fluorescein angiogram 36 seconds after injection of dye showing a very ischaemic retina with very poor blood supply to the macular area and leakage of dye from the new vessels elsewhere is starting and will increase over the next few minutes and will be more obvious in the later photos.

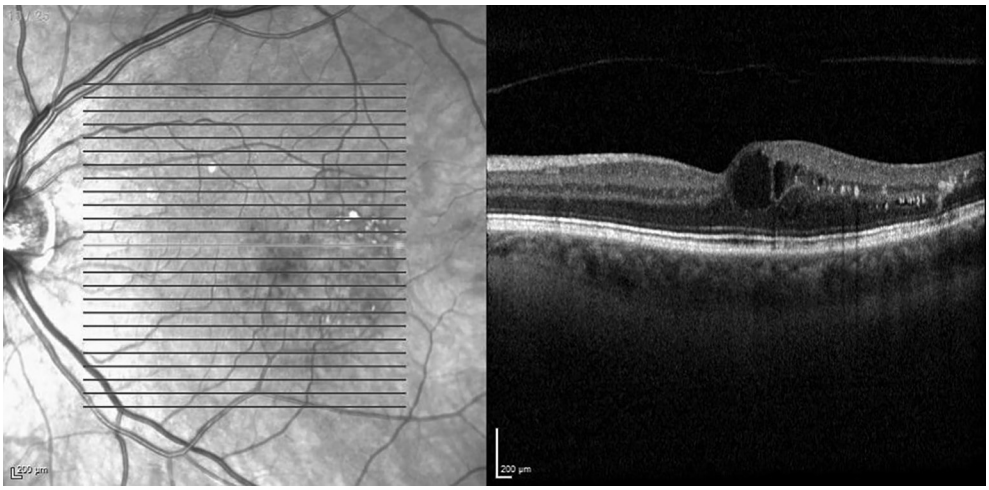


Figure 15.12.2.3 An OCT image of the left macular area showing leaking of fluid into the retinal layers on the temporal side of the macula in a patient with diabetic maculopathy.

is important to have a multidisciplinary approach to these patients so that the eye is not treated in isolation. Hence it is important that proper attention is paid to the management of:

a) Systemic hypertension
b) Glucose control
c) Blood lipids

Treatment of Diabetic Maculopathy

In the past diabetic maculopathy was treated with laser in accordance with the definition of clinically significant macular oedema which was defined by the ETDRS study as:

1. Thickening of the retina at or within 500 microns of the centre of the macula;
2. Hard exudates at or within 500 microns of the centre of the fovea, if associated with thickening of the adjacent retina (not residual hard exudates remaining after disappearance of retinal thickening);
3. A zone or zones of retinal thickening 1 disc area or larger, any part of which is within 1 disc diameter of the centre of the macula.

When treatment was carried out with laser it was recognized that one should not treat within the foveal avascular zone because of the risk of the treated individual noticing a blind spot in their vision and because laser creep over time can affect the central vision at a later date.

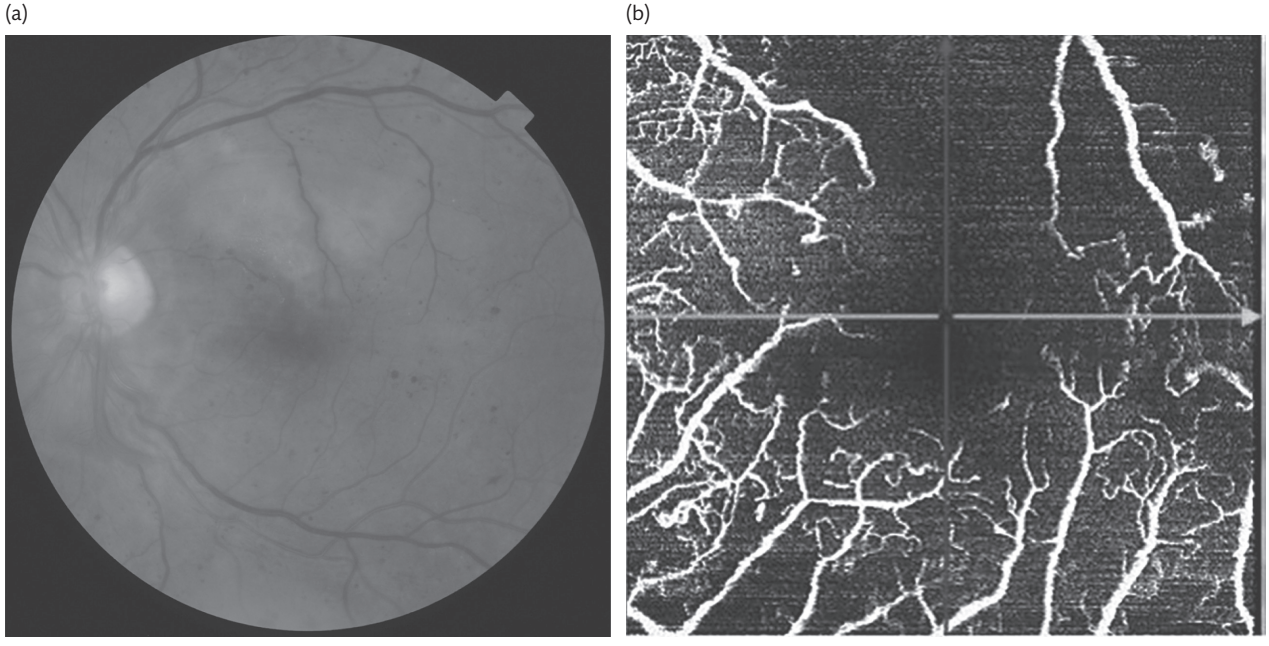


Figure 15.12.2.4 (a) Colour image of the left macula and disc showing NVD at the left disc and a superior branch arterial occlusion above the left macula. (b) An OCT angiogram of the left macular area showing the poor blood supply in the superior left macular area.

VEGF inhibitors—more recently laser has been reserved for thickening that is off centre and those people with diabetes who have diabetic maculopathy and central thickening are offered a course of intravitreal injections of VEGF inhibitors for treatment of their oedema. In England NICE guidelines restrict this treatment to those whose centre thickness are greater than or equal 400 microns.

There have been a number of studies that have shown that this treatment is beneficial in diabetic maculopathy using the two on-label preparations ranibizumab [32], aflibercept [33, 34], and the off-label bevacizumab [35]. One exciting feature of all of these studies was that there was regression of ETDRS levels and there was a significant gain in vision measured by the number of extra letters read. This was of course dependent on the starting level of vision but there was considerably more visual gain with VEGF inhibitors than with laser treatment where stabilization of vision was considered optimal and vision gain was infrequent. One study [36] suggested a higher vision gain could be achieved with aflibercept in patients whose initial vision was worse (less than 69 letters or approximately ≤6/15) than using other VEGF inhibitors. There tends to be a reduction in treatment frequency required over time with an average of 7–9 anti-VEGF injections required in the first year, 2–4 in the second year, 1–3 in the third, and approximately 1 injection per year in years 4 and 5 [37, 38].

Potential side effects of intravitreal VEGF inhibitor therapy are sustained elevation of intraocular pressure [39] (9.5% compared to 3.4% in the control group), endophthalmitis [40] (0.04–0.06/injection) and cardiovascular events [41] which, although a theoretical risk, have not been found to be higher than the control group in a meta-analysis of six randomized controlled studies. Until recently we have had a choice in England of using the on-label preparations of aflibercept and ranibizumab but NICE has recently issued guidance [42] suggesting that they found no significant differences in effectiveness and safety between the different anti-VEGF treatments in age-related macular degeneration following the subsequent statement by the GMC [43] that they would not raise any 'fitness to practice' concerns if doctors prescribed bevacizumab, we may see a wider use of this agent.

Intravitreal corticosteroids—it has been shown [44] that inflammation plays a significant role in the development of macular oedema in some patients with diabetes. Intravitreal triamcinolone acetonide is an off-label treatment that has been extensively studied [45] and has been shown to be beneficial in some patients. Side effects of all intravitreal steroid preparations are elevation of IOP requiring topical treatment or surgery, and cataracts. In addition some patients who were treated with intravitreal triamcinolone developed sterile uveitis/pseudoendophthalmitis (in up to 13% of eyes) that further limited its clinical use [46]. Dexamethasone intravitreal implant (Ozurdex, Allergan, Inc., Irvine, CA) is a sustained-release bio-gradable implant formulation of corticosteroid that has been shown [47] to improve visual acuity in a difficult-to-treat patient population with long-standing refractory DME of macular oedema due to diabetes. Dexamethasone implants only have a duration of action of approximately 3 months. In the UK, it has also been approved by NICE [48] for this indication but only in pseudophakic eyes who have not responded to non-corticosteroid treatment. Fluocinolone acetonide (Illuvien) is an intravitreal implant of fluocinolone acetonide which delivers 36 months of continuous, low-dose corticosteroid. Intravitreal Fluocinolone acetonide was recently granted approval [49] in the UK by NICE but only for treating chronic diabetic macular oedema that is insufficiently responsive to available therapies only if the implant is to be used in an eye with an intraocular (pseudophakic) lens [50].

Treatment of PDR

The Diabetic Retinopathy Study (DRS) [51, 52] provided the first treatment for PDR with panretinal photocoagulation (PRP) in 1976 and recommended prompt treatment in the presence of DRS high-risk characteristics, which reduced the two-year risk of severe visual loss by 50% or more and were defined by:

a) The presence of preretinal or VH;
b) Eyes with NVD equalling or exceeding one-quarter to one-third disc area in extent with no haemorrhage;
c) NVE equalling > one-half disc area with haemorrhage.

See **Figure 15.12.2.5**.

As the side effects of laser have reduced there has been an increasing acceptance that treatment of low risk PDR with PRP should also be undertaken. The modern lasers have reduced the duration of the burn which has reduced discomfort and enabled multiple burn spots to be produced with each depression of the foot pedal, which reduces the time taken for the procedure, although a larger number of burns appears to be required [53] than with traditional lasers.

However, for the first time since 1976 an alternative treatment is being suggested, which are regular injections of VEGF inhibitors. A study [54] by the Diabetic Retinopathy Clinical Research

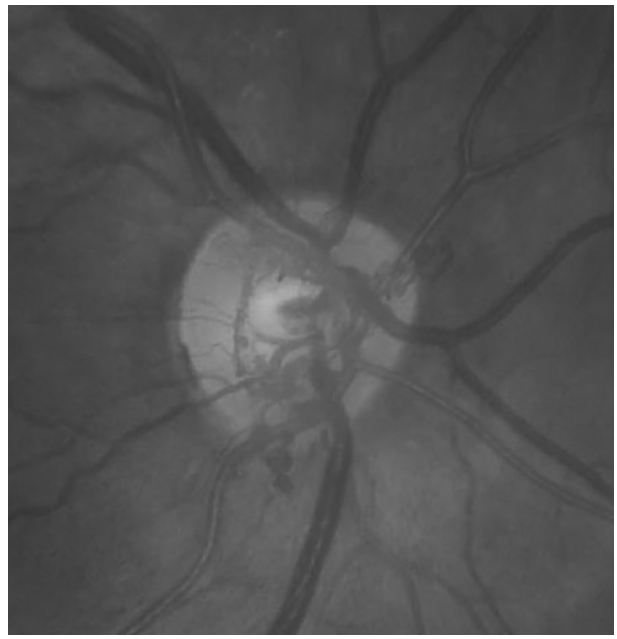


Figure 15.12.2.5 New vessels are developing at this right disc due to generalized ischaemia. As they are >1/3rd disc area they are considered high risk.

Network assessed the non-inferiority of intravitreal ranibizumab compared with PRP for visual acuity outcomes in patients with PDR. Among eyes with PDR, treatment with ranibizumab resulted in visual acuity that was not worse than PRP treatment at 2 years. A study from the UK [55] found that patients with PDR who were treated with intravitreal aflibercept had an improved outcome at 1 year compared with those treated with PRP standard care. However, the long-term requirements of number of injections required to maintain this benefit and the economic costs are currently unknown and so PRP continues to be standard care for proliferative DR.

Treatment of the Combination of Diabetic Maculopathy and PDR

Before the introduction for VEGF inhibitors, report from the DRS [56] suggested that 'reducing macular oedema by focal photocoagulation before initiating scatter treatment and dividing scatter treatment into multiple sessions with less intense burns may decrease the risk of the visual loss associated with photocoagulation'. However, the introduction of VEGF inhibitors has made VEGF inhibitors the preferred treatment if the centre thickness of the macula is ≥400 microns.

A report from the United Kingdom National Ophthalmology Database Study on diabetic retinopathy showed that of 48 250 eyes of patients with diabetes who had had a structured assessment recorded in the hospital eye service, 5066 eyes had proliferative DR. Of those with proliferative DR there were 14.2% (719 eyes) with centre involving diabetic macular oedema and 7.6% (385 eyes) with non-centre involving diabetic macular oedema. Hence, it is common to see the combination of proliferative DR and diabetic macular oedema.

Treatment of Advanced DR

Advanced DR such as non-clearing VH and tractional retinal detachment involving the macula is treated with pars plana vitrectomy. With frequent follow-up examinations and timely scatter (panretinal) photocoagulation, the 5-year cumulative rate of pars plana vitrectomy in ETDRS patients [57] was 5.3%. Surgical techniques have improved since the early reports and some surgeons [58] are now using preoperative intravitreal VEGF inhibitors to reduce intraoperative bleeding and thereby facilitating the surgery. Iris neovascularization and neovascular glaucoma are usually treated [59] with an intravitreal VEGF inhibitor followed by PRP [60] to control the retinal ischaemia. Before the availability of VEGF inhibitors it was often very difficult to treat these patients because the raised intraocular pressure produced corneal oedema that made it difficult to undertake the PRP because of the poor view. With the advent of the VEGF inhibitors the neovascularization usually responds within 24 hours and the intraocular pressure reduces and this allows a window of approximately 1 month to provide the required PRP. If there is a delay in presentation, the angle may have closed and then other surgical techniques (e.g. cyclodiode laser may be required to control the intraocular pressure).

REFERENCES

1. Joner G, Brinchmann-Hansen O, Torres CG, Hanssen KF. A nationwide cross-sectional study of retinopathy and microalbuminuria in young Norwegian type 1 (insulin-dependent) diabetic patients. *Diabetologia*, 1992; **35**: 1049–54.
2. Matthews DR, Stratton IM, Aldington SJ, Holman RR, Kohner EM. Risks of progression of retinopathy and vision loss related to tight blood pressure control in type 2 diabetes mellitus: UKPDS 69. *Arch Ophthalmol*, 2004; **122**: 1631–40.
3. Chew EY, Ambrosius WT, Davis MD, *et al.* Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*, 2010; **363**: 233–44.
4. Patel A, Group AC, MacMahon S, *et al.* Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. *Lancet*, 2007; **370**: 829–40.
5. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med*, 1993; **329**: 977–86.
6. Effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: Diabetes Control and Complications Trial. Diabetes Control and Complications Trial Research Group. *J Pediatr*, 1994; **125**: 177–88.
7. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, 1998; **352**: 837–53.
8. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. *Lancet*, 1998; **352**: 854–65.
9. Chew EY, Klein ML, Ferris FL, 3rd, *et al.* Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy. Early Treatment Diabetic Retinopathy Study (ETDRS) Report 22. *Arch Ophthalmol*, 1996; **114**: 1079–84.
10. Scanlon PH AS, Leal J, Luengo-Fernandez R, *et al.* Development of a cost-effectiveness model for optimisation of the screening interval in diabetic retinopathy screening. *Health Technol Assess*, 2015; **19**: 1–116.
11. Keech AC, Mitchell P, Summanen PA, *et al.* Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomised controlled trial. *Lancet*, 2007; **370**: 1687–97.
12. Muhlhauser I, Bender R, Bott U, *et al.* Cigarette smoking and progression of retinopathy and nephropathy in type 1 diabetes. *Diabet Med*, 1996; **13**: 536–43.
13. Karamanos B, Porta M, Songini M, *et al.* Different risk factors of microangiopathy in patients with type I diabetes mellitus of short versus long duration. The EURODIAB IDDM Complications Study. *Diabetologia*, 2000; **43**: 348–55.
14. Yau JW, Rogers SL, Kawasaki R, *et al.* Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 2012; **35**: 556–64.
15. Scanlon PH, Stratton IM, Bachmann MO, Jones C, Leese GP, Four Nations Diabetic Retinopathy Screening Study Group. Risk of diabetic retinopathy at first screen in children at 12 and 13 years of age. *Diabet Med*, 2016; **33**: 1655–8.

16. Stratton IM, Kohner EM, Aldington SJ, *et al.* UKPDS 50: risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. *Diabetologia*, 2001; **44**: 156–63.
17. Hietala K, Forsblom C, Summanen P, Groop PH. Heritability of proliferative diabetic retinopathy. *Diabetes*, 2008; **57**: 2176–80.
18. Simmons D, Clover G, Hope C. Ethnic differences in diabetic retinopathy. *Diabet Med*, 2007; **24**: 1093–8.
19. Klein BE, Moss SE, Klein R. Effect of pregnancy on progression of diabetic retinopathy. *Diabetes Care*, 1990; **13**: 34–40.
20. Effect of pregnancy on microvascular complications in the Diabetes Control and Complications Trial. The Diabetes Control and Complications Trial Research Group. *Diabetes Care*, 2000; **23**: 1084–91.
21. Chew EY, Mills JL, Metzger BE, *et al.* Metabolic control and progression of retinopathy. The Diabetes in Early Pregnancy Study. National Institute of Child Health and Human Development Diabetes in Early Pregnancy Study. *Diabetes Care*, 1995; **s**: 631–7.
22. Phelps RL, Sakol P, Metzger BE, Jampol LM, Freinkel N. Changes in diabetic retinopathy during pregnancy. Correlations with regulation of hyperglycemia. *Arch Ophthalmol*, 1986; **104**: 1806–10.
23. Rosenn B, Miodovnik M, Kranias G, *et al.* Progression of diabetic retinopathy in pregnancy: association with hypertension in pregnancy. *Am J Obstet Gynecol*, 1992; **166**: 1214–18.
24. Lauszus F, Klebe JG, Bek T. Diabetic retinopathy in pregnancy during tight metabolic control. *Acta Obstet Gynecol Scand*, 2000; **79**: 367–70.
25. Dibble CM, Kochenour NK, Worley RJ, Tyler FH, Swartz M. Effect of pregnancy on diabetic retinopathy. *Obstet Gynecol*, 1982; **59**: 699–704.
26. Temple RC, Aldridge VA, Sampson MJ, Greenwood RH, Heyburn PJ, Glenn A. Impact of pregnancy on the progression of diabetic retinopathy in Type 1 diabetes. *Diabet Med*, 2001; **18**: 573–7.
27. Sakata M, Oniki K, Kita A, *et al.* Clinical features associated with a rapid decline in renal function among Japanese patients with type 2 diabetes mellitus: microscopic hematuria coexisting with diabetic retinopathy. *Diabetes Res Clin Pract*, 2013; **100**: e39–41.
28. Early worsening of diabetic retinopathy in the Diabetes Control and Complications Trial. *Arch Ophthalmol*, 1998; **116**: 874–86.
29. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*, 1991; **98**: 786–806.
30. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*, 1991; **98**: 823–33.
31. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open*, 2014; **4**: e004015.
32. Nguyen QD, Brown DM, Marcus DM, *et al.* Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*, 2012; **119**: 789–801.
33. Korobelnik JF, Do DV, Schmidt-Erfurth U, *et al.* Intravitreal aflibercept for diabetic macular edema. *Ophthalmology*, 2014; **121**: 2247–54.
34. Brown DM, Schmidt-Erfurth U, Do DV, *et al.* Intravitreal Aflibercept for diabetic macular edema: 100-week results from the VISTA and VIVID studies. *Ophthalmology*, 2015; **122**: 2044–52.
35. Michaelides M, Kaines A, Hamilton RD, *et al.* A prospective randomized trial of intravitreal bevacizumab or laser therapy in the management of diabetic macular edema (BOLT study) 12-month data: report 2. *Ophthalmology*, 2010; **117**: 1078–86.e2.
36. Diabetic Retinopathy Clinical Research Network, Wells JA, Glassman AR, *et al.* Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema. *N Engl J Med*, 2015; **372**: 1193–203.
37. Schmidt-Erfurth U, Lang GE, Holz FG, *et al.* Three-year outcomes of individualized ranibizumab treatment in patients with diabetic macular edema: the RESTORE extension study. *Ophthalmology*, 2014; **121**: 1045–53.
38. Elman MJ, Ayala A, Bressler NM, *et al.* Intravitreal ranibizumab for diabetic macular edema with prompt versus deferred laser treatment: 5-year randomized trial results. *Ophthalmology*, 2015; **122**: 375–81.
39. Bressler SB, Almukhtar T, Bhorade A, *et al.* Repeated intravitreous ranibizumab injections for diabetic macular edema and the risk of sustained elevation of intraocular pressure or the need for ocular hypotensive treatment. *JAMA Ophthalmol*, 2015; **133**: 589–97.
40. Brown DM, Nguyen QD, Marcus DM, *et al.* Long-term outcomes of ranibizumab therapy for diabetic macular edema: the 36-month results from two phase III trials: RISE and RIDE. *Ophthalmology*, 2013; **120**: 2013–22.
41. Chen G, Li W, Tzekov R, Jiang F, Mao S, Tong Y. Ranibizumab monotherapy or combined with laser versus laser monotherapy for diabetic macular edema: a meta-analysis of randomized controlled trials. *PloS One*, 2014; **9**: e115797.
42. Age-related macular degeneration. NICE guideline [NG82]. 2018. Available at: https://www.nice.org.uk/guidance/ng82/chapter/recommendations (accessed: 29 April 2018).
43. Doctors are cleared to prescribe cheaper drug for wet AMD. 2018. Available at: https://www.bmj.com/content/360/bmj.k344 (accessed: 29 April 2018).
44. Joussen AM, Poulaki V, Le ML, *et al.* A central role for inflammation in the pathogenesis of diabetic retinopathy. *FASEB J*, 2004; **18**: 1450–2.
45. Beck RW, Edwards AR, Aiello LP, *et al.* Three-year follow-up of a randomized trial comparing focal/grid photocoagulation and intravitreal triamcinolone for diabetic macular edema. *Arch Ophthalmol*, 2009; **127**: 245–51.
46. Otsuka H, Kawano H, Sonoda S, Nakamura M, Sakamoto T. Particle-induced endophthalmitis: possible mechanisms of sterile endophthalmitis after intravitreal triamcinolone. *Invest Ophthalmol Vis Sci*, 2013; **54**: 1758–66.
47. Dutra Medeiros M, Postorino M, Navarro R, Garcia-Arumi J, Mateo C, Corcostegui B. Dexamethasone intravitreal implant for treatment of patients with persistent diabetic macular edema. *Ophthalmologica*, 2014; **231**: 141–6.
48. NICE. Dexamethasone intravitreal implant for treating diabetic macular oedema. Technology appraisal guidance [TA349]. 2015. Available at: https://www.nice.org.uk/guidance/ta349 (accessed: 29 April 2018).
49. NICE. Fluocinolone acetonide intravitreal implant for treating chronic diabetic macular oedema after an inadequate response to prior therapy. Technology appraisal guidance [TA301]. 2013. Available at: https://www.nice.org.uk/guidance/ta301 (accessed: 30 April 2018).

50. Cunha-Vaz J, Ashton P, Iezzi R, *et al.* Sustained delivery fluocinolone acetonide vitreous implants: long-term benefit in patients with chronic diabetic macular edema. *Ophthalmology*, 2014; **121**: 1892–903.
51. DRS. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical application of Diabetic Retinopathy Study (DRS) findings, DRS Report Number 8. The Diabetic Retinopathy Study Research Group. *Ophthalmology*, 1981; **88**: 583–600.
52. Indications for photocoagulation treatment of diabetic retinopathy: Diabetic Retinopathy Study Report no. 14. The Diabetic Retinopathy Study Research Group. *Int Ophthalmol Clin*, 1987; **27**: 239–53.
53. Muqit MM, Marcellino GR, Henson DB, Young LB, Turner GS, Stanga PE. Pascal panretinal laser ablation and regression analysis in proliferative diabetic retinopathy: Manchester Pascal Study Report 4. *Eye*, 2011; **25**: 1447–56.
54. Writing Committee for the Diabetic Retinopathy Clinical Research N, Gross JG, Glassman AR, *et al.* Panretinal photocoagulation vs intravitreous ranibizumab for proliferative diabetic retinopathy: a randomized clinical trial. *JAMA*, 2015; **314**: 2137–46.
55. Sivaprasad S, Prevost AT, Vasconcelos JC, *et al.* Clinical efficacy of intravitreal aflibercept versus panretinal photocoagulation for best corrected visual acuity in patients with proliferative diabetic retinopathy at 52 weeks (CLARITY): a multicentre, single-blinded, randomised, controlled, phase 2b, non-inferiority trial. *Lancet*, 2017; **389**: 2193–203.
56. Ferris FL, 3rd, Podgor MJ, Davis MD. Macular edema in Diabetic Retinopathy Study patients. Diabetic Retinopathy Study Report Number 12. *Ophthalmology*, 1987; **94**: 754–60.
57. Flynn HW, Jr., Chew EY, Simons BD, Barton FB, Remaley NA, Ferris FL, 3rd. Pars plana vitrectomy in the Early Treatment Diabetic Retinopathy Study. ETDRS report number 17. The Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology*, 1992; **99**: 1351–7.
58. Zhao LQ, Zhu H, Zhao PQ, Hu YQ. A systematic review and meta-analysis of clinical outcomes of vitrectomy with or without intravitreal bevacizumab pretreatment for severe diabetic retinopathy. *Br J Ophthalmol*, 2011; **95**: 1216–22.
59. Oshima Y, Sakaguchi H, Gomi F, Tano Y. Regression of iris neovascularization after intravitreal injection of bevacizumab in patients with proliferative diabetic retinopathy. *Am J Ophthalmol*, 2006; **142**: 155–8.
60. Shiraya T, Kato S, Shigeeda T. Optimal area of retinal photocoagulation necessary for suppressing active iris neovascularisation associated with diabetic retinopathy. *Int Ophthalmol*, 2014; **34**: 1115–17.

15.12.3 Diabetic Nephropathy

Luigi Gnudi and Sally M. Marshall

Definition

Classical Diabetic Nephropathy

The evolution of diabetic nephropathy (DN) proceeds through several distinct stages. The stages of DN have been historically depicted as progression through increasing amount of albumin in the urine (normoalbuminuria-A1, microalbuminuria-A2, and macroalbuminuria-A3), as a result of a damaged glomerular filtration barrier [1], until renal failure [2]. DN is defined as presence of albuminuria and progressive decline in glomerular filtration rate (GFR) associated with retinopathy and hypertension. Albuminuria *per se* is inadequate to stage the progression of DN; it has recently been suggested that the parallel use of albuminuria and GFR levels improves the staging of renal disease progression [2].

Non-Classical Diabetic Nephropathy

The presence of albuminuria is not always associated with a commensurate alteration/decline in renal function, especially in the early phase of renal disease in diabetes [3, 4]. GFR progressive decline has been extensively described in the absence of albuminuria [5], a frequent scenario particularly in patients with T2DM [6, 7] observed in many studies (NEFRON [8], UKPDS [9], ADVANCE [10]).

Non-albuminuric renal impairment in T2DM does not associate with poor metabolic control, retinopathy and hypertension but is linked to a higher cardiovascular risk [11]. Vascular disease (macroangiopathy) has been suggested as a potential underlying renal pathology [12].

Up to 40–50% of individuals with T2DM present with some degree of renal function decline in the absence of retinopathy, suggesting atypical renal disease is frequent and should be quickly identified with a kidney biopsy for a conclusive diagnosis [13, 14].

Changing Epidemiology?

Diabetes is the major cause of ESRD (United States Renal Data System, 2016, https://www.usrds.org/2016/view/Default.aspx). The progressive ageing of the population in the western world has contributed to an increased prevalence of T2DM and chronic kidney disease (CKD), the incidence of ESRD in elderly people being higher than in the general population [15]. Similarly, the obesity epidemic, today also observed in young adults, has been linked with an increased risk for diabetic kidney disease (DKD) [16].

The recently observed reduction in cardiovascular events in the T1DM and T2DM population [17] has not been associated with an improvement in CKD [18] which still affects ~50% of patients with T2DM [19].

Of the patients with diabetes who develop CKD, approximately one-third will present with classical DN, one-third with

non-diabetic kidney disease (mostly obesity-related focal segmental glomerulosclerosis in the absence of retinopathy) and one-third with mixed pathologies [20].

Pathology

Classic Diabetic Nephropathy

Histologically, the classical presentation of DN is characterized by specific glomerular and tubulointerstitial lesions [21]:

Glomerulus

The glomerular anatomo-ultrastructural changes comprise thickening of the glomerular basement membrane (GBM), diffuse mesangial sclerosis with accumulation of extracellular matrix and nodular formation (Kimmelstiel-Wilson lesion), as well as hyalinosis, and glomerular capillary microaneurysms [21]. There has been an attempt to classify the glomerular lesions into four classes according to the degree of extracellular matrix accumulation and mesangial expansion [22]. This is paralleled by podocyte loss in the urine [23] and progressive fall in filtration surface and decline in renal function.

Tubular Compartment

Pathological alterations within the tubular compartment occur in the advanced stage of the disease: these are characterized by tubular interstitial fibrosis with significant inflammatory infiltrates and occasionally tubular atrophy. Hyaline arteriosclerosis of renal vessels can be seen as a result of a vascular disease process that occurs from the early stages of DN [24].

Non-Classic Diabetic Nephropathy

There is no structured classification/definition for the 'non-classical' presentation of DN, mostly seen in patients with T2DM; the atypical presentation is characterized mainly by disproportionate tubulointerstitial, glomerulosclerotic, and vascular changes. The aetiology is thought to be related to ageing, atherosclerotic vascular disease in general, and hypertension.

The heterogeneity in renal ultrastructural lesions in T2DM leaves us with many unanswered questions about the natural history of albuminuria or proteinuria in patients with T2DM who may present with minimal or no renal lesions. On the other hand patients with albuminuria/proteinuria can present with mild diabetic glomerular changes but disproportionately severe vasculopathy, tubular atrophy, tubular basement membrane thickening, tubulointerstitial inflammatory infiltrate, and sclerosis.

Especially in patients with T2DM, the renal pathological presentation may have many distinct patterns with variable glomerular, tubular, and vascular involvement [25].

Pathophysiology of Microvascular Damage

Mechanisms of Glomerular Disease

A clear interaction has been proposed between haemodynamic (glomerular hypertension) and metabolic (hyperglycaemia, dyslipidaemia) perturbations as recognized drivers for renal disease in diabetes [26, 27] (see Figure 15.12.3.1 online only).

Hyperglycaemia has been implicated in the dysregulation of afferent and efferent glomerular arteriolae with dysregulation of the local tissue RAAS. Excess angiotensin-2 is a key player in glomerular hypertension [28]. Excess local angiotensin-2 is a major direct stimulus for TGFβ1, one of the main stimuli for glomerular and tubulointerstitial fibrosis [29] and a stimulus for the expression of cytokines such as VEGF-A, monocyte chemoattractant protein-1 (MCP-1), TGFβ1 itself, and connective tissue growth factor (CTGF) [30]. Of note, TGFβ1-mediated facilitative glucose transporter (GLUT)-1 upregulation has been proposed as one of the mediators of metabolic–haemodynamic interaction in diabetic glomerulopathy [27].

Both angiotensin-2 and TGFβ1 stimulate nicotinamide adenine dinucleotide phosphate (NADPH) oxidase leading to excess reactive oxygen species accumulation [31]. Excess cell glucose metabolism /oxidative stress-mediated ER-stress has been proposed as a mechanism of DN [32].

The excess cellular glucose uptake, driven by the ambient elevated circulating glucose levels, and excess oxidative stress also results in activation of different intracellular metabolic pathways such as the polyol and hexosamine pathway, increased production of advanced glycation end-products, activation of protein kinase C (PKC) and p38 mitogen activated protein kinase (MAPK), pathways linked to upregulation of mediators of glomerular damage such as angiotensin-2, VEGF-A, and TGFβ1 [33].

Oxidative stress is an important mediator of inflammation: activation of NFkB, a transcription factor that controls the expression of cytokines involved in inflammation (interleukin (IL)-1, IL-6, IL-18, and tumour necrosis factor), plays an important role in DKD [34].

Recent studies have challenged the role of oxidant species in the pathophysiology of DN, highlighting how oxidant species not only cause oxidative stress if elevated, but also function as important signalling molecules [31], implying that a balanced level of oxidant species must be maintained in physiology [31, 35].

The endothelium plays an important central role in the pathophysiology of diabetic glomerulopathy. Endothelial dysfunction precedes increased vascular permeability and albuminuria. In diabetes, glomerular endothelial cell injury with loss of glycocalyx and cell apoptosis has been proposed as one of the mechanisms of diabetic glomerulopathy [36]. Changes in glycocalyx structure are associated with changes in glomerular permeability [1]. Diabetes affects other glomerular cells, such as podocytes and mesangial cells, resulting in podocyte effacement, foot process fusion and apoptosis with podocyte loss in the urine [23], mesangial expansion, and mesangiolysis.

Diabetes affects the anatomical structure of the cells forming the glomerular filtration barrier and alters the autocrine/paracrine actions of cytokines and vascular growth factors (e.g. TGF-β1, VEGF-A, and angiopoietins) secreted by these cells [37]. Glomerular VEGF-A expression plays a role in the pathophysiology of diabetic glomerulopathy by binding to its receptors expressed in endothelial cells, mesangial cells, and podocytes [37]. Primary VEGF-A overexpression in the podocytes results in proteinuria and disruption of the GBM [38], while inhibition of VEGF-A in animal experimental models of diabetes ameliorates proteinuria and glomerular damage [21, 37].

VEGF-A and nitric oxide (NO) play important roles in endothelial dysfunction [39]. In diabetes, endothelial nitric oxide synthase (eNOS) uncoupling results in decreased NO production and increased reactive oxygen species. The role of eNOS is still debated.

More will be learned from eNOS knockout mice which develop advanced diabetic renal lesions [40]. In addition, lack of eNOS protects from diabetes-mediated hyperfiltration [41].

VEGF-A expression modulates the actions of angiopoietin (Angpt)-1 and -2 on the vasculature [37]. Angpt-1 promotes endothelial cell survival, stabilization of supporting perivascular cells, and inhibition of endothelial permeability, while Angpt-2 is considered to be a natural antagonist of Angpt-1 and its overexpression results in albuminuria [21, 37].

In diabetic glomerulopathy, Angpt-1 and Angpt-2 levels are deregulated (often with Angpt-2 > Angpt-1) and ablation of Angpt-1 in podocytes or mesangial cells results in accelerated diabetes-mediated glomerular damage, while podocyte specific Angpt-1 overexpression in diabetes is protective [37]. VEGF-A has been implicated in maintenance of the endothelial glycocalyx structure and Angpt-1 promotes a healthy endothelial glycocalyx and reduces glomerular capillary permeability [21, 37].

Mechanisms of Tubular Disease and Interstitial Fibrosis

Metabolic perturbations driving inflammation and oxidative stress have been involved in interstitial fibrosis and tubular damage in diabetes [42]; importantly renal function correlates with diabetes-mediated tubulointerstitial changes [43].

Diabetes is characterized by an upregulation of the low-affinity/high-capacity sodium-glucose cotransporter SGLT2 known to represent the major player in glucose reabsorption in the nephron. SGLT2 upregulation is paralleled by activation of the local angiotensin-2 system and prosclerotic cytokines such as CTGF and TGF-β1 which, in turn, will favour tubular cell proliferation and hypertrophy, cell senescence and increased deposition of extracellular matrix [21]. Studies in patients with T1DM have suggested that tubular hypertrophy and excess glucose tubular reabsorption could contribute to hyperfiltration and faster decline in renal function [44]. Hyperfiltration, in turn, results in glomerular proteinuria, and tubulointerstitial inflammation and fibrosis [45].

Clinical Presentation

Early Renal Abnormalities

Hyperfiltration

Whole-kidney hyperfiltration (GFR>120–140 ml/min) is common in newly-diagnosed T1DM and short-duration T2DM. In addition, as kidney disease advances and nephron mass reduces, hyperfiltration occurs in the remaining nephrons [46]. In many individuals, hyperfiltration at diagnosis of diabetes resolves as glucose control improves. In some, hyperfiltration persists and may predispose to later DN. However, this remains controversial [47].

Progression of DKD

At presentation of diabetes, albuminuria A2 or even A3 may be present. As hyperglycaemia resolves, albuminuria returns to normal in almost everyone with T1DM and the majority of those with T2DM. In classical DN, albuminuria then rises gradually through A1 to A2 and A3. Tracking changes in GFR in early disease routinely is extremely difficult because of the insensitivity of serum creatinine and creatinine-based formulae for estimating GFR to small changes in renal function at higher levels. However, it seems that GFR falls in parallel with rising albuminuria, from early persistent hyperfiltration, to normal and then low levels [48]. The fall in GFR is linear but the rate varies widely between individuals [49]. Blood pressure and cardiovascular risk rise in parallel with increasing albuminuria and declining GFR.

In non-classical DN, a substantial number of individuals with T2DM, and approximately 10–20% of those with T1DM, develop ESRD without significant progressive albuminuria. The clinical phenotype of these individuals is different to those with progressive albuminuria [9, 50]. In addition, in T2DM, histological changes are of atherosclerotic vascular disease rather than the glomerulosclerosis of DN [51]. Thus it is likely that classical and non-classical nephropathy are of different aetiology, the non-albuminuric phenotype probably developing from hypertension, atherosclerotic vascular disease, and obesity. The rate of fall of GFR is at least as great as in classical nephropathy. Frank hypertension is usually present throughout and individuals may not have significant retinopathy.

Cardiovascular Disease and the Cardiorenal Syndrome

Cardiovascular risk increases progressively DKD develops [52]. Most individuals with DKD die of cardiovascular disease before developing renal failure. Albuminuria and GFR <60 ml/min/1.73m^2 are both independently and additively associated with increased all-cause and cardiovascular mortality and cardiovascular events [47]. Individuals with both increased albuminuria and decreased GFR have the highest risk.

Non-albuminuric renal impairment is also a strong predictor of mortality, highlighting the prognostic role of renal dysfunction independently of albuminuria [53].

Potential explanations for this increased cardiovascular disease (CVD) risk include shared traditional CVD risk factors, worsening of traditional CVD risk factors as DKD progresses, novel CVD risk factors (e.g. disruption of mineral metabolism) and genetic factors. The term cardiorenal metabolic syndrome has been coined to describe the constellation of central obesity, insulin resistance, hypertension, dyslipidaemia and proteinuria and/or declining GFR [54].

Risk Factors for Disease Onset and Progression

Non-Modifiable

There is a genetic predisposition to classical DN and the associated cardiovascular disease. Different genes may influence disease initiation and progression, and separately affect the development of albuminuria and decline in GFR. Although numerous candidate genes have been suggested, none are currently clinically useful. Epigenetic modification may also play a role. The role of inheritance in the development of non-classical DKD is unknown.

Modifiable

For classical DN, the main modifiable risk factors are burden of glycaemia (increasing duration of diabetes and hyperglycaemia), blood pressure, smoking, dyslipidaemia, and obesity. These factors all probably influence initial development and progression of DN. Other features of the metabolic syndrome, such as insulin resistance, non-alcoholic fatty liver disease, and sleep apnoea are also

associated. All of these risk factors are also likely to influence the development and progression of non-classical DKD, but this is less well studied.

Diagnosis and Monitoring

Despite much work exploring novel biomarkers for DKD [55], only albuminuria and GFR are of clinical use currently. Screening relies on annual measurement of urine albumin excretion and serum creatinine, with calculation of the estimated GFR (eGFR) using the CKD-EPI equation [2] (Figure 15.12.3.2 online only). Urine samples can be spot early morning, spot random, or timed collections. For spot samples, urine creatinine is also measured and the urine albumin:creatinine ratio (ACR) calculated. Timed samples are inconvenient for the individual with diabetes and open to inaccuracies in collection, so generally are not used for screening. Spot early morning samples avoid changes in albuminuria with upright posture so are preferred to random samples.

On screening (Figure 15.12.3.3 online only), if albuminuria is A1 and eGFR>90 ml/min/1.73m^2, the patient does not have DKD and both measurements should be repeated annually. Because of the extremely high intraindividual day-to-day variation in albuminuria, if the first measurement is abnormal, further measurements should be made over 3–6 months. Albuminuria A2 and A3 should only be diagnosed if repeat measurements are abnormal. If present, measurements should be made at each clinic visit, to allow a clear picture of the trajectory of albuminuria over time, and the response to any intervention. If eGFR is 30–60 ml/min/1.73m^2, the test should be repeated 6 monthly, and if <30 ml/min/1.73m^2, every 3 months, so that the rate of change of kidney function can be assessed accurately. As eGFR declines, anaemia and bone chemistry should also be monitored.

Diagnosis of DN is usually made clinically, on the basis of a gradual rise in albuminuria and blood pressure, with GFR falling linearly, in the presence of significant retinopathy and other complications of diabetes. The absence of retinopathy suggests non-diabetic CKD. If there are concerns that other renal disease is present (e.g. sudden occurrence of albuminuria A3 in someone with previously normal ACR), then further investigations might include renal tract ultrasound (symmetrical kidneys of normal size in DN), autoantibodies and immunoglobulins (all normal in DN). If doubt remains, renal biopsy can be performed.

Prevention and Management of Diabetic Kidney Disease

Prevention of DKD is clearly of utmost importance, as once CKD stage 3 develops, its progress can only be slowed, not reversed.

Glycaemic Control

Numerous studies have demonstrated that the better the glucose control (HbA1c), the lower the risk of developing DN. In T1DM, the Diabetes Control and Complications Trial (DCCT) and its open follow-up, the Epidemiology of Diabetes Interventions and Complications Study (DCCT-EDIC) [56] demonstrated highly significant reductions in the number of individuals developing micro- and macroalbuminuria, eGFR <60 ml/min/1.73m^2, and ESRD in those originally assigned to intensive management compared to those initially receiving usual care. The difference in HbA1c during the 6.5 years randomized clinical trial (mean HbA1c 7.2% intensive group, 9.1% conventional group) was lost during open follow-up (mean HbA1c 8.0% in both groups). However, despite the loss of glycaemic separation, DCCT/EDIC showed a durable effect of initial assigned therapies, with risk reductions for new microalbuminuria of 45%, new macroalbuminuria 61% and eGFR<60 ml/min/1.73m^2 44% [56].

In T2DM, several studies have demonstrated reduction in the development of micro- and macro-albuminuria with improved glucose control [57–60], but studies have been too small and too short to demonstrate convincing effects on GFR. The exception to this are the SGLT2 studies, which are discussed later.

Once albuminuria A2 or A3 are present, there is some evidence in both T1DM and T2DM that improving glucose control may reduce albuminuria. In an observational study of T1DM individuals with persistent microalbuminuria at baseline, those who became normoalbuminuric were more likely to have HbA1c less than 8.0%, systolic BP less than 115mmHg and cholesterol less than 5.13 mmol/L [61]. In T2DM, individuals receiving intensive glucose management are less likely to have progressive albuminuria and albuminuria is more likely to regress compared to those receiving usual care [57, 59]. There is no convincing data that rate of fall of GFR is altered by improving glucose control and no head-to-head studies of different classes of oral glucose-lowering agents, so that the choice of therapy should be individualized. Most guidelines recommend a target HbA1c of 6.5–7.2% (48–52 mmol/mol), but this must be individualized. Tight glucose control is inappropriate in individuals with multiple comorbidities and limited life expectancy. In addition, as GFR declines, the risk of hypoglycaemia increases, due to declining kidney gluconeogenesis, reduced clearance of some glucose-lowering agents and declining appetite among others [62].

There is no specific trial data on the effects of improving glucose control in individuals with non-classical DKD.

Inhibition of the Renin Angiotensin Aldosterone System (RAAS)

Inhibitors of the RAAS cause dilatation of the efferent glomerular arteriole and thus reduce intraglomerular pressure and albuminuria. By doing so, we believe they reduce the rate of progression of albuminuric CKD. Experts therefore suggest that diabetic individuals with albuminuria should be prescribed a RAAS inhibitor, regardless of blood pressure, and the dose titrated to the maximum recommended or tolerated. Evidence supporting this rests on a few small studies in T1DM with BP below the definition of hypertension but clearly above that of their normoalbuminuric peers. For historical reasons, the bulk of the evidence favours angiotensin-2 converting enzyme inhibitors (ACEI) in T1DM and angiotensin-2 receptor blockers (ARB) in T2DM, although in reality there is little to choose between the agents. Blockade of RAAS using a single agent is incomplete, so that dual blockade using ACEI plus ARB or aldosterone antagonist has been explored. Dual blockade reduces blood pressure and albuminuria more than single blockade, but

the significantly higher risks of hyperkalaemia and acute kidney injury outweigh any benefit, so that this strategy is currently not recommended.

RAAS blockade does not reduce the risk of developing DN in normoalbuminuric diabetic individuals with BP <140/90 mmHg. We do not know if RAAS inhibitors have any advantages over other classes of antihypertensive medication in individuals with normoalbuminuric DKD.

All patients taking RAAS blockade should be counselled to stop these medications temporarily if they become unwell with vomiting or diarrhoea or cannot maintain a normal fluid intake, and only to restart the drugs when they recover. Women of childbearing age should be advised to stop RAAS blockade if they are considering pregnancy, because of the teratogenic effects. Other agents, including methyldopa, nifedipine, and labetalol should be substituted during the preconception phase.

Blood Pressure

Raised blood pressure is a key initiating and progressing factor in DKD. The higher the blood pressure, the higher the risk of development and progression of DN. Lowering blood pressure reduces albuminuria and slows the rate of decline of GFR and progression to renal failure. Historic data shows that without antihypertensive therapy, GFR declines by 10–12 ml/min/1.73 m^2 per year. With effective blood pressure control, this is slowed to 3–5 ml/min/1.73 m^2 per year. Thus if blood pressure is controlled adequately from early in the disease process, the time to requiring renal replacement therapy can be doubled. There are no specific trials in non-albuminuric DKD, but it seems extremely likely that good blood pressure control is also beneficial.

The target blood pressure for diabetic individuals with CKD is generally 130/80 mmHg. Too tight blood pressure control may actually be harmful, leading to renal hypoperfusion. Thus, as with glucose targets, blood pressure targets should be individualized. In fitter individuals with few comorbidities, lower targets (less than 130 mmHg systolic) can probably be achieved safely. In people with multiple comorbidities, it is probably safer to accept blood pressure in the range 130–140 mmHg. Increasingly, more weight is being put on 24-hour blood pressure monitoring or home measurements by the individual, rather than clinic blood pressure readings. Loss of normal nocturnal blood pressure dipping may occur, even though day-time blood pressure is normal.

The initial antihypertensive agent usually will be either a RAAS inhibitor or calcium-channel blocking agent. Most individuals will need two or more agents to achieve good control. Combinations used in trials include RAAS blocker plus calcium-channel blocker, or RAAS blocker plus a diuretic. There is little evidence to guide choice of a third or fourth agent, but α-blockers, β-blockers, and centrally acting agents are all used routinely. It is vital to reduce the blood pressure to a satisfactory level, using the combination of agents best suited to the individual person.

Dietary Factors

Weight Loss

Obesity may be one of the factors which drives non-albuminuric DKD. Albuminuria decreases significantly after bariatric surgery [63], but we lack robust data on its effects on GFR.

Protein Intake

A low-protein diet reduces progression of CKD [64]. Intake should not be restricted to <0.8 g/kg per day [2]. Vegetable protein may be more beneficial than animal protein.

Salt

The blood pressure lowering effect of RAAS blockade is greater in individuals whose salt intake is lower rather than higher [65]. Thus individuals should be encouraged to moderate their salt intake to <90 mmol sodium /day [2].

Potassium

Hyperkalaemia is common, particularly in those taking RAAS blockade. There is a U-shaped relationship between serum potassium and adverse outcomes [66]. Hyperkalaemia may limit the dose of RAAS inhibitor. Above a level of 5.5 mmol/L, a diet low in potassium should be advised.

Other Cardiovascular Risk Factors

It is important to ensure that cardiovascular risk factors are managed aggressively in individuals with DKD. Smoking cessation is paramount, as is lipid control. Lipid profiles change as DKD progresses, but all individuals should be offered statin therapy. Because of the very high cardiovascular risk, it has been suggested that everyone with DKD should be offered low-dose aspirin therapy as primary prevention, but there is no good evidence to support this. In addition to reducing cardiovascular risk, it is hoped that these therapies will also reduce DKD progression, but there is no trial data to support this.

Multifactorial Intervention

Addressing all renal and cardiovascular risk factors, not just individual factors, is vital, as demonstrated by the Steno-2 multifactorial intervention trial [67]. In this study, T2DM individuals with microalbuminuria were randomized to intensive risk factor management or routine care. Over the 8 years of the intervention trial, and in open follow-up thereafter, the risk of progression to proteinuria and ESRD, need for laser therapy, cardiovascular and all-cause mortality and heart failure were all significantly reduced in the intensively managed group [68, 69]. Thus, intensive management of a wide range of risk factors has tangible long-term benefits encompassing both micro- and macro-vascular complications (see Table 15.12.3.1 **online only**).

Future Treatments

Since the 1990s, blockade of RAAS, along with good metabolic control has been the key therapeutic approach for renal protection in diabetes [21].

The use of aldosterone antagonist on top of conventional RAAS blockade (ACEI or ARB therapy) has demonstrated a reduction in albuminuria and renal disease progression [21, 70], but side effects

such as hyperkalaemia, hypotension and acute kidney injury has dampened the enthusiasm for double RAAS blockade [21].

Other agents such as endothelin antagonists and activator of the antioxidant element NrF2 have failed to show any benefit in renal protection and have often raised important safety concerns [21].

In recent years inhibitors of the SGLT2 glucose transporters, introduced as oral hypoglycaemic agents, have been proposed as new cardiorenal protective treatments; their positive effects seem related to haemodynamic changes [71] and recent studies have confirmed these beneficial effects on renal protection [72].

The incretin pathway has recently been proposed to confer renoprotection; GLP-1 analogues developed as hypoglycaemic agents retain anti-inflammatory and renoprotective properties [21]. Recent human studies have suggested a significant cardiovascular-renal protection of GLP-1 analogues [73, 74]. Further work is needed to confirm and validate these findings.

REFERENCES

1. Haraldsson B, Nystrom J, Deen WM. Properties of the glomerular barrier and mechanisms of proteinuria. *Physiol Rev*, 2008; **88**: 451–87.
2. Group KCW. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl*, 2013; **3**: 1–150.
3. Jerums G, Panagiotopoulos S, Premaratne E, Power DA, MacIsaac RJ. Lowering of proteinuria in response to antihypertensive therapy predicts improved renal function in late but not in early diabetic nephropathy: a pooled analysis. *Am J Nephrol*, 2008; **28**: 614–27.
4. Karalliedde J, Viberti G. Proteinuria in diabetes: bystander or pathway to cardiorenal disease? *J Am Soc Nephrol*, 2010; **21**: 2020–7.
5. Robles NR, Villa J, Gallego RH. Non-proteinuric diabetic nephropathy. *J Clin Med*, 2015; **4**: 1761–73.
6. MacIsaac RJ, Tsalamandris C, Panagiotopoulos S, Smith TJ, McNeil KJ, Jerums G. Nonalbuminuric renal insufficiency in type 2 diabetes. *Diabetes Care*, 2004; **27**: 195–200.
7. Kramer HJ, Nguyen QD, Curhan G, Hsu CY. Renal insufficiency in the absence of albuminuria and retinopathy among adults with type 2 diabetes mellitus. *JAMA*, 2003; **289**: 3273–77.
8. Thomas MC, Macisaac RJ, Jerums G, *et al*. Nonalbuminuric renal impairment in type 2 diabetic patients and in the general population (national evaluation of the frequency of renal impairment co-existing with NIDDM [NEFRON] 11). *Diabetes Care*, 2009; **32**: 1497–502.
9. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR, Group US. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. *Diabetes*, 2006; **55**: 1832–39.
10. Ninomiya T, Perkovic V, de Galan BE, *et al*. Albuminuria and kidney function independently predict cardiovascular and renal outcomes in diabetes. *J Am Soc Nephrol*, 2009; **20**: 1813–21.
11. Penno G, Solini A, Bonora E, *et al*. Clinical significance of nonalbuminuric renal impairment in type 2 diabetes. *J Hypertens*, 2011; **29**: 1802–9.
12. MacIsaac RJ, Panagiotopoulos S, McNeil KJ, *et al*. Is nonalbuminuric renal insufficiency in type 2 diabetes related to an increase in intrarenal vascular disease? *Diabetes Care*, 2006; **29**: 1560–6.
13. Erdogmus S, Kiremitci S, Celebi ZK, *et al*. Non-diabetic kidney disease in type 2 diabetic patients: prevalence, clinical predictors and outcomes. *Kidney Blood Press Res*, 2017; **42**: 886–93.
14. Fiorentino M, Bolignano D, Tesar V, *et al*. Renal biopsy in patients with diabetes: a pooled meta-analysis of 48 studies. *Nephrol Dial Transplant*, 2017; **32**: 97–110.
15. Jungers P, Chauveau P, Descamps-Latscha B, *et al*. Age and gender-related incidence of chronic renal failure in a French urban area: a prospective epidemiologic study. *Nephrol Dial Transplant*, 1996; **11**: 1542–6.
16. Kovesdy CP, S LF, Zoccali C, World Kidney Day Steering C. Obesity and kidney disease: hidden consequences of the epidemic. *Clin Kidney J*, 2017; **10**: 1–8.
17. Rawshani A, Rawshani A, Franzén S, *et al*. Mortality and cardiovascular disease in type 1 and type 2 diabetes. *N Engl J Med*, 2017; **376**: 1407–18.
18. Neuen BL, Chadban SJ, Demaio AR, Johnson DW, Perkovic V. Chronic kidney disease and the global NCDs agenda. *BMJ Glob Health*, 2017; **2**, e000380.
19. Thomas MC, Cooper ME, Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. *Nat Rev Nephrol*, 2016; **12**, 73–81.
20. Tan J, Zwi LJ, Collins JF, Marshall MR, Cundy T. Presentation, pathology and prognosis of renal disease in type 2 diabetes. *BMJ Open Diabetes Res Care*, 2017; **5**, e000412.
21. Gnudi L, Gentile, G., Ruggenenti, P. The patient with diabetes mellitus. In: Turner N, Lamiere N, Goldsmith DJ, Wineearls CG, Himmelfarb J, Remuzzi G, eds. *Oxford Textbook of Clinical Nephrology*. Oxford: Oxford University Press, 2016: 1199–247.
22. Tervaert TW, Mooyaart AL, Amann K, *et al*. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol*, 2010; **21**: 556–63.
23. Pagtalunan ME, Miller PL, Jumping-Eagle S, *et al*. Podocyte loss and progressive glomerular injury in type II diabetes. *J Clin Invest*, 1997; **99**: 342–8.
24. Tonolo G, Cherchi S. Tubulointerstitial disease in diabetic nephropathy. *Int J Nephrol Renovasc Dis*, 2014; **7**: 107–15.
25. Fioretto P, Mauer M. Histopathology of diabetic nephropathy. *Semin Nephrol*, 2007; **27**: 195–207.
26. Cooper ME. Interaction of metabolic and haemodynamic factors in mediating experimental diabetic nephropathy. *Diabetologia*, 2001; **44**: 1957–72.
27. Gnudi L, Thomas SM, Viberti G. Mechanical forces in diabetic kidney disease: a trigger for impaired glucose metabolism. *J Am Soc Nephrol*, 2007; **18**: 2226–32.
28. Maddox DA, Brenner BM. Glomerular ultrafiltration. In: Brenner BM, Rector FC, Jr., eds. *The Kidney*. Philadelphia, PA: W.B. Saunders Company, 2000: 319–74.
29. Ziyadeh FN. Mediators of diabetic renal disease: the case for tgf-Beta as the major mediator. *J Am Soc Nephrol*, 2004; **15 Suppl 1**: S55–7.
30. Wolf G, Ziyadeh FN. Cellular and molecular mechanisms of proteinuria in diabetic nephropathy. *Nephron Physiol*, 2007; **106**: 26–31.
31. Sagoo MK, Gnudi L. Diabetic nephropathy: is there a role for oxidative stress? *Free Radic Biol Med*, 2018; **116**: 50–63.
32. Zhuang A, Forbes JM. Stress in the kidney is the road to pERdition: is endoplasmic reticulum stress a pathogenic mediator of diabetic nephropathy? *J Endocrinol*, 2014; **222**: R97–111.
33. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*, 2005; **54**: 1615–25.

34. Navarro-Gonzalez JF, Mora-Fernandez C, Muros de Fuentes M, Garcia-Perez J. Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. *Nat Rev Nephrol*, 2011; **7**: 327–40.
35. Sharma K. Mitochondrial hormesis and diabetic complications. *Diabetes*, 2015; **64**: 663–72.
36. Satchell SC, Tooke JE. What is the mechanism of microalbuminuria in diabetes: a role for the glomerular endothelium? *Diabetologia*, 2008; **51**: 714–25.
37. Gnudi L, Benedetti S, Woolf AS, Long DA. Vascular growth factors play critical roles in kidney glomeruli. *Clin Sci (Lond)*, 2015; **129**: 1225–36.
38. Veron D, Bertuccio CA, Marlier A, *et al*. Podocyte vascular endothelial growth factor (Vegf) overexpression causes severe nodular glomerulosclerosis in a mouse model of type 1 diabetes. *Diabetologia*, 2011; **54**: 1227–41.
39. Nakagawa T. Uncoupling of the VEGF-endothelial nitric oxide axis in diabetic nephropathy: an explanation for the paradoxical effects of VEGF in renal disease. *Am J Physiol Renal Physiol*, 2007; **292**: F1665–72.
40. Nakayama T, Sato W, Kosugi T, *et al*. Endothelial injury due to eNOS deficiency accelerates the progression of chronic renal disease in the mouse. *Am J Physiol Renal Physiol*, 2009; **296**: F317–27.
41. Takahashi T, Harris RC. Role of endothelial nitric oxide synthase in diabetic nephropathy: lessons from diabetic eNOS knockout mice. *J Diabetes Res*, 2014; **2014**: 590541.
42. Vallon V. The proximal tubule in the pathophysiology of the diabetic kidney. *Am J Physiol Regul Integr Comp Physiol*, 2011; **300**: R1009–22.
43. Gilbert RE, Cooper ME. The tubulointerstitium in progressive diabetic kidney disease: more than an aftermath of glomerular injury? *Kidney Int*, 1999; **56**: 1627–37.
44. Vervoort G, Veldman B, Berden JH, Smits P, Wetzels JF. Glomerular hyperfiltration in type 1 diabetes mellitus results from primary changes in proximal tubular sodium handling without changes in volume expansion. *Eur J Clin Invest*, 2005; **35**: 330–6.
45. Remuzzi G, Bertani T. Pathophysiology of progressive nephropathies. [Review] [95 refs]. *N Engl J Med*, 1998; **339**: 1448–56.
46. Tonneijck L, Muskiet MH, Smits MM, *et al.* Glomerular hyperfiltration in diabetes: mechanisms, clinical significance, and treatment. *J Am Soc Nephrol*, 2017; **28**: 1023–39.
47. Tuttle KR, Bakris GL, Bilous RW, *et al*. Diabetic kidney disease: a report from an ADA Consensus Conference. *Diabetes Care*, 2014; **37**: 2864–83.
48. Krolewski AS, Niewczas MA, Skupien J, *et al*. Early progressive renal decline precedes the onset of microalbuminuria and its progression to macroalbuminuria. *Diabetes Care*, 2014; **37**: 226–34.
49. Skupien J, Warram JH, Smiles AM, Stanton RC, Krolewski AS. Patterns of estimated glomerular filtration rate decline leading to end-stage renal disease in type 1 diabetes. *Diabetes Care*, 2016; **39**: 2262–9.
50. Penno G, Russo E, Garofolo M, *et al*. Evidence for two distinct phenotypes of chronic kidney disease in individuals with type 1 diabetes mellitus. *Diabetologia*, 2017; **60**: 1102–13.
51. Ekinci EI, Jerums G, Skene A, *et al*. Renal structure in normoalbuminuric and albuminuric patients with type 2 diabetes and impaired renal function. *Diabetes Care*, 2013; **36**: 3620–6.
52. de Ferranti SD, de Boer IH, Fonseca V, *et al*. Type 1 diabetes mellitus and cardiovascular disease: a scientific statement from the American Heart Association and American Diabetes Association. *Diabetes Care*, 2014; **37**: 2843–63.
53. Penno G, Solini A, Orsi E, *et al*. Non-albuminuric renal impairment is a strong predictor of mortality in individuals with type 2 diabetes: the Renal Insufficiency And Cardiovascular Events (RIACE) Italian multicentre study. *Diabetologia*, 2018; **61**: 2277–89.
54. Whaley-Connell A, Sowers JR. Basic science: pathophysiology: the cardiorenal metabolic syndrome. *J Am Soc Hypertens*, 2014; **8**: 604–6.
55. Colhoun HM, Marcovecchio ML. Biomarkers of diabetic kidney disease. *Diabetologia*, 2018; **61**: 996–1011.
56. DCCT/EDIC research group. Effect of intensive diabetes treatment on albuminuria in type 1 diabetes: long-term follow-up of the Diabetes Control and Complications Trial and Epidemiology of Diabetes Interventions and Complications study. *Lancet Diabetes Endocrinol*, 2014; **2**: 793–800.
57. Ismail-Beigi F, Craven T, Banerji MA, *et al*. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial. *Lancet*, 2010; **376**: 419–30.
58. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*, 1998; **352**: 837–53.
59. ADVANCE Collaborative Group, Patel A, MacMahon S, *et al*. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*, 2008; **358**: 2560–72.
60. Perkovic V, Heerspink HL, Chalmers J, *et al*. Intensive glucose control improves kidney outcomes in patients with type 2 diabetes. *Kidney Int*, 2013; **83**: 517–23.
61. Perkins BA, Ficociello LH, Ostrander BE, *et al.* Microalbuminuria and the risk for early progressive renal function decline in type 1 diabetes. *J Am Soc Nephrol*, 2007; **18**: 1353–61.
62. Alsahli M, Gerich JE. Hypoglycemia in patients with diabetes and renal disease. *J Clin Med*, 2015; **4**: 948–64.
63. Upala S, Wijarnpreecha K, Congrete S, Rattanawong P, Sanguankeo A. Bariatric surgery reduces urinary albumin excretion in diabetic nephropathy: a systematic review and meta-analysis. *Surg Obes Relat Dis*, 2016; **12**: 1037–44.
64. Nezu U, Kamiyama H, Kondo Y, Sakuma M, Morimoto T, Ueda S. Effect of low-protein diet on kidney function in diabetic nephropathy: meta-analysis of randomised controlled trials. *BMJ Open*, 2013; **3**: pii: e002934.
65. Lambers Heerspink HJ, Holtkamp FA, Parving HH, *et al.* Moderation of dietary sodium potentiates the renal and cardiovascular protective effects of angiotensin receptor blockers. *Kidney Int*, 2012; **82**: 330–7.
66. Kovesdy CP, Matsushita K, Sang Y, *et al*. Serum potassium and adverse outcomes across the range of kidney function: a CKD Prognosis Consortium meta-analysis. *Eur Heart J*, 2018; **39**: 1535–42.
67. Gaede P, Lund-Andersen H, Parving HH, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med*, 2008; **358**: 580–91.
68. Gaede P, Oellgaard J, Carstensen B, *et al*. Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial. *Diabetologia*, 2016; **59**: 2298–307.

69. Oellgaard J, Gæde P, Rossing P, *et al*. Reduced risk of heart failure with intensified multifactorial intervention in individuals with type 2 diabetes and microalbuminuria: 21 years of follow-up in the randomised Steno-2 study. *Diabetologia*, 2018; **61**: 1724–33.
70. Epstein M, Williams GH, Weinberger M, *et al*. Selective aldosterone blockade with eplerenone reduces albuminuria in patients with type 2 diabetes. *Clin J Am Soc Nephrol*, 2006; **1**, 940–51.
71. DeFronzo RA, Norton L, Abdul-Ghani M. Renal, metabolic and cardiovascular considerations of SGLT2 inhibition. *Nat Rev Nephrol*, 2017; **13**: 11–26.
72. Perkovic V, Jardine MJ, Neal B, *et al*. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*, 2019; **380**: 2295–306.
73. Marso SP, Daniels GH, Brown-Frandsen K, *et al*. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2016; **375**: 311–22.
74. Mann JFE, Ørsted DD, Brown-Frandsen K, *et al*. Liraglutide and renal outcomes in type 2 diabetes. *N Engl J Med*, 2017; **377**: 839–48.

Additional material is provided in the online version at **oxfordmedicine.com/otendocrinology3**

15.12.4 Diabetic Neuropathy

Solomon Tesfaye and Jing Wu

Introduction

Diabetic neuropathy is a major complication of diabetes and a cause of considerable morbidity and mortality [1]. Diabetic neuropathy is not a single entity but includes several neuropathic syndromes (Figure 15.12.4.1). In clinical practice, the commonest presentation of neuropathy is chronic distal symmetrical polyneuropathy (DSP) also known as diabetic peripheral neuropathy (DPN). The neuropathic syndromes depicted in Figure 15.12.4.1 have varied presentations and pathogenesis. This chapter will cover these syndromes although the main focuses will be: (1) DPN, which is the main initiating factor for foot ulceration and a cause of troublesome painful neuropathic symptoms and (2) autonomic neuropathy.

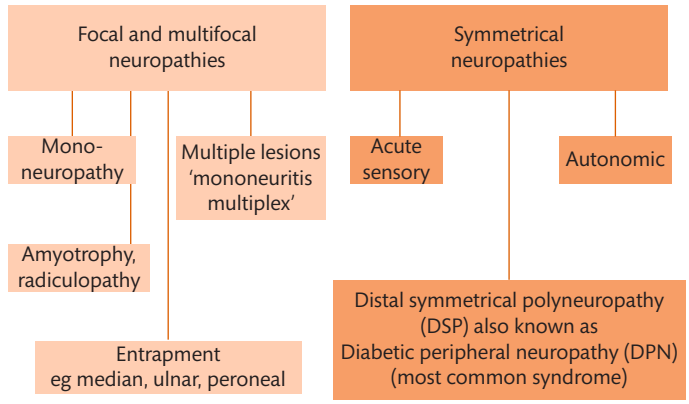


Figure 15.12.4.1 Neuropathic syndromes associated with diabetes mellitus.

Epidemiology

There is much variation in the prevalence of DPN. Where electrophysiology is employed the prevalence rates will be in excess of 50% [2], whereas when clinical parameters or QST (quantitative sensory testing) are employed both clinic- and population-based studies show similar prevalence rates at about 30% [3]. The EURODIAB Prospective Complications Study investigated 3250 type 1 patients, and found a prevalence rate of 28% for DPN [4]. The study also showed that over a 7.3 year period, about one-quarter of type 1 diabetic patients developed DPN [5]. The development of DPN was also associated with modifiable cardiovascular risk factors such as hypertension, hyperlipidaemia, obesity, and cigarette smoking (Figure 15.12.4.2) [5]. Based on recent epidemiological studies, correlates of DPN include age, duration of diabetes, poor glycaemic control, retinopathy, albuminuria, and vascular risk factors [5].

Classification of Diabetic Neuropathy

Classification of the various syndromes of diabetic neuropathy is difficult. The variations and overlap in aetiology, clinical features, natural history, and prognosis have meant that most classifications are necessarily oversimplified. Nevertheless, classification assists in the planning of clinical management.

Figure 15.12.4.1 shows a modified clinical classification of diabetic polyneuropathy [6].

Watkins and Edmonds [7] have suggested a classification for diabetic neuropathy based on the natural history of the various syndromes, which separates them into three distinct groups (Box 15.12.4.1).

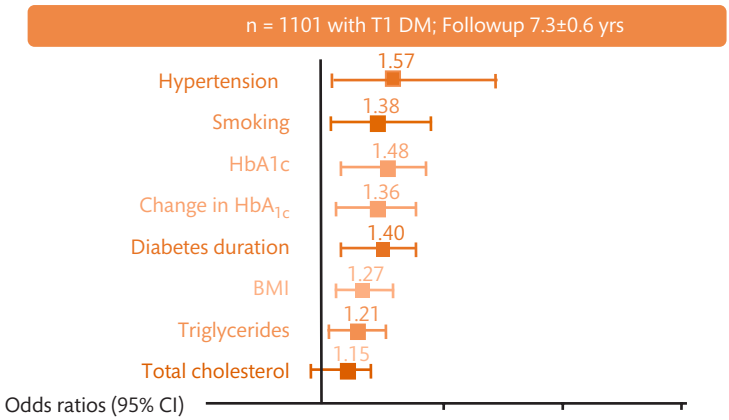


Figure 15.12.4.2 Risk factors for incident DPN in the EURODIAB prospective study.

Reproduced with permission from Tesfaye S, Chaturvedi N, Eaton SEM, Witte D, Ward JD, Fuller J. Vascular risk factors and diabetic neuropathy. *New Engl J Med* 2005; 352: 341–50. Copyright © 2005, Massachusetts Medical Society. (Ref. 5).

More recently, the American Diabetes Association Position statement on Diabetic Neuropathy [8] has suggested a classification summarized in Box 15.12.4.2.

Diabetic Peripheral Neuropathy (DPN)

This is the commonest neuropathic syndrome. The Toronto Consensus Panel recently defined DPN as 'a symmetrical, length dependent sensorimotor polyneuropathy attributable to metabolic and microvessel alterations as a result of chronic hyperglycaemia exposure (diabetes) and cardiovascular risk covariates' [9]. There is a 'length-related' pattern of sensory loss, with sensory symptoms starting in the toes and then extending to involve the feet and legs in a stocking distribution. In more severe cases, there is upper limb involvement, with a similar progression proximally. Although the nerve damage can extend over the entire body including the head and face, this is exceptional. Subclinical neuropathy detectable by autonomic function tests is usually present. However, clinical autonomic neuropathy is less common. As the disease advances, overt motor manifestations such as wasting of the small muscles of the hands and limb weakness become apparent. Subclinical motor involvement detected by magnetic resonance imaging appears to be common, and thus motor disturbance is clearly part of the functional impairment caused by DPN [10].

The main clinical presentation of DPN is sensory loss, which the patient may not be aware of or may be described as 'asleep numbness' or 'dead feeling'. However, some may experience a progressive build-up of unpleasant sensory symptoms including tingling (paraesthesiae or 'pins and needles'). Box 15.12.4.3 summarizes the 'positive' and 'negative' symptoms of DPN. There is a large spectrum of severity of these symptoms.

Diabetic neuropathic pain is characteristically more severe at night, and often prevents sleep [11]. Some patients may be in a

Box 15.12.4.1 Classification of diabetic neuropathies by natural history

1. **Progressive neuropathies.** These are associated with increasing duration of diabetes and with other microvascular complications. Sensory disturbance predominates and autonomic involvement is common. The onset is gradual and there is no recovery.
2. **Reversible neuropathies.** These have an acute onset, often occurring at the presentation of diabetes itself, and are not related to the duration of diabetes or other microvascular complications. There is spontaneous recovery of these acute neuropathies.
3. **Pressure palsies.** Although these are not specific to diabetes only, they tend to occur more frequently in diabetic patients than the general population. There is no association with duration of diabetes or other microvascular complications of diabetes.

Reproduced with permission from Watkins PJ, Edmonds ME. Clinical features of diabetic neuropathy. In: *Textbook of Diabetes* Vol.2. Pickup J, Williams G (Eds.) 1997; pp 50.1–50.20. Copyright © 1998, Springer Science Business Media New York. (Ref. 7).

Box 15.12.4.2 Classification of diabetic neuropathies

Polyneuropathies
- Distal symmetrical polyneuropathy
- Combined large and small-fibre neuropathy
- Predominantly large-fibre neuropathy
- Predominantly small-fibre neuropathy

Autonomic neuropathy
- Cardiovascular
- Resting tachycardia
- Sudden death
- Exercise intolerance
- Orthostatic hypotension
- Foot vein distension and arteriovenous shunting
- Gastrointestinal
- Gastroparesis
- Diarrhoea or constipation
- Bladder hypomotility
- Erectile dysfunction
- Gustatory sweating
- Reduced peripheral sweating

Focal neuropathies
- Mononeuropathies
- Mononeuritis multiplex
- Proximal motor neuropathy (amyotrophy)
- Thoraco-abdominal neuropathy

Box 15.12.4.3 Symptoms of DPN

'Positive' symptoms
- Persistent burning or dull pain
- Paroxysmal electric, shooting, stabbing path
- Dysesthesias (painful parasthesias)
- Evoked pain (hyperalgesia, allodynia)
- Asleep numbness

'Negative' symptoms (deficits)
- Hypoalgesia, analgesia
- Hypoesthesia, anaesthesia

constant state of tiredness because of sleep deprivation [11]. Others are unable to maintain full employment. Severe painful neuropathy can cause marked reduction in exercise threshold so as interfere with daily activities. This is particularly the case when there is an associated disabling, severe postural hypotension due to autonomic involvement. Not surprisingly therefore, mood disorders including anxiety and depressive symptoms are common [12]. It is important to appreciate that many subjects with DPN may not have any of the symptoms shown in Box 15.12.4.3, and their first presentation may be with a foot ulcer [13]. This underpins the need for carefully examining and screening the feet of all diabetic people. The insensate foot is at risk of developing mechanical and thermal injuries, and patients must therefore be warned about these and given appropriate advice with regard to foot care [13]. A curious feature of the neuropathic foot is that both numbness and pain may occur, the so-called 'painful, painless' leg. It is indeed a paradox that the patient with a large foot ulcer may also have severe neuropathic pain. In those with advanced neuropathy, there may be sensory ataxia, unsteadiness on walking, and falls.

DPN is usually easily detected by simple clinical examination (Table 15.12.4.1) [8]. Bare feet should be examined at least annually and more often if neuropathy is present. The most common presenting abnormality is a reduction or absence of vibration sense in the toes. As the disease progresses there is sensory loss in a 'stocking' and sometimes in a 'glove' distribution, involving all modalities. When there is severe sensory loss, proprioception may also be impaired, leading to a positive Romberg's sign. Ankle tendon reflexes are lost (though this may also be lost with old age in non-diabetic people) and with more advanced neuropathy, knee reflexes are often reduced or absent. The recent American Diabetes Association Position Statement made clear recommendations for the screening of DPN (Box 15.12.4.4).

Table 15.12.4.1 Clinical assessment for DPN

History	Signs
• Sensory symptoms	• Inspection (normal or distal wasting, clawing)
• Motor symptoms	• Reflexes (ankle reflex unreliable in the elderly)
• Assessment of disability due to DPN	• Sensory
• Exclude other causes of neuropathy	• Vibration • Temperature • Pinprick (good discriminator in the elderly) • 10 g mono filament
	Assess footwear

In DPN there is ↓ in: reflexes, vibration, pinprick, and pressure sensation.

Box 15.12.4.4 American Diabetes Association recommendations for the screening of DPN

- All patients should be assessed for DPN starting at diagnosis of type 2 diabetes and 5 years after the diagnosis of type 1 diabetes and at least annually thereafter.
- Consider screening patients with prediabetes who have symptoms of peripheral neuropathy.
- Assessment should include a careful history and either temperature or pinprick sensation (small-fibre function) and vibration sensation using a 128-Hz tuning fork (large-fibre function). All patients should have an annual 10-g monofilament testing to assess for feet at risk for ulceration and amputation.
- Electrophysiological testing or referral to a neurologist is rarely needed for screening, except in situations where the clinical features are atypical, the diagnosis is unclear, or a different aetiology is suspected. Atypical features include motor greater than sensory neuropathy, rapid onset, or asymmetrical presentation.

Reproduced with permission from Pop-Busui R, Boulton AJ, Feldman EL, Bril V, Freeman R, Malik RA, Sosenko JM, Ziegler D. Diabetic Neuropathy: A Position Statement by the American Diabetes Association. *Diabetes Care*. 2017; 40(1):136–54. (Ref. 8).

Muscle strength is usually normal early in the course of the disease, although mild weakness may be found in toe extensors. With progressive disease there is significant generalized muscular wasting, particularly in the small muscles of the hand and feet. Wasting of dorsal interossei is usually due to entrapment of the ulnar nerve at the elbow. Clawing of the toes is believed to be due to unopposed pulling of the long extensor and flexor tendons. This scenario results in elevated plantar pressure points at the metatarsal heads that are prone to callus formation and foot ulceration. Deformities such as a bunion can form the focus of ulceration, and with more extreme deformities such as those associated with Charcot arthropathy [14] the risk is further increased. As one of the most common precipitants to foot ulceration is inappropriate footwear, a thorough assessment should also include examination of shoes for poor fit, abnormal wear, and internal pressure areas or foreign bodies.

Autonomic neuropathy affecting the feet can cause a reduction in sweating and consequently dry skin that is likely to crack easily, predisposing the patient to the risk of infection. The 'purely' neuropathic foot is also warm due to arteriovenous shunting. This results in the distension of foot veins that fail to collapse even when the foot is elevated. It is not unusual to observe a gangrenous toe in a foot that has bounding arterial pulses, as there is impairment of the nutritive capillary circulation due to arteriovenous shunting. The oxygen tension of the blood in these veins is typically raised. The increasing blood flow brought about by autonomic neuropathy can sometimes result in neuropathic oedema.

Recent Advances in the Assessment of DPN Using Novel Point-of-Care Devices

Over the past decade there has been an advance in the development of non-invasive, objective, relatively quick and accurate

point-of-care devices (POCDs) that may be able to diagnose DPN early, before overt clinical signs are apparent. These include DPN-Check that can measure sural nerve conduction velocity and amplitude, Sudoscan that can measure sudomotor (small fibre function), Neuropad colour change that can also measure sudomotor function and corneal confocal microscopy that can assess corneal small fibres [15]. A recent study evaluated the use of POCDs in an annual, one-stop diabetes microvascular screening clinic and found that this reduces clinic visits, unmasked new diagnosis of painful DPN in 25% and the at-risk foot [16]. However, the long-term impact of this clinic on hard outcomes such as foot ulcers and amputations has not been evaluated.

Differential Diagnosis of DPN

Before attributing the neuropathy to diabetes other common causes of neuropathy must be excluded. The absence of other complications of diabetes, rapid weight loss, excessive alcohol intake and other atypical features in either the history or clinical examination should direct the physician to search for other causes of neuropathy (Box 15.12.4.5). Some clinicians routinely screen for B_{12} deficiency (particularly as metformin use is linked with B12 deficiency) and hypothyroidism in all diabetic patients.

Confirmed Diagnosis of DPN

The Toronto Consensus Panel on Diabetic Neuropathy [9] defined: *possible DPN* as the presence of typical symptoms or signs; *probable DPN* as one that requires at least two of the symptoms, distal sensory deficits, or decreased/absent ankle jerks; *confirmed DPN* as one abnormal nerve conduction attribute and the presence of symptom(s)/sign (s); *subclinical DPN* is defined as one abnormal nerve conduction attribute or validated measure of small-fibre impairment in the absence of symptom(s)/sign(s). This expert group has also proposed minimal diagnostic criteria for small-fibre neuropathy (SFN) as follows: *possible SFN*: symptoms and/or signs of small-fibre impairment; *probable SFN*: symptoms and signs of small-fibre impairment with normal nerve conduction parameters of the sural nerve; *definite SFN*: symptoms and signs of small-fibre damage, normal nerve conduction parameters of the sural nerve, with reduced intraepidermal nerve fibre density on skin biopsy and/or abnormal thermal perception thresholds.

Box 15.12.4.5 Differential diagnosis of DPN

- Metabolic
  - Diabetes
  - Amyloidosis
  - Uraemia
  - Myxoedema
  - Porphyria
- Vitamin deficiency (thiamine, B_{12}, B_6, pyridoxine)
- Drugs and chemicals
  - Alcohol
  - Cytotoxic drugs, e.g. vincristine
  - Chlorambucil
  - Nitrofurantoin
  - Isoniazid
- Neoplastic disorders
  - Bronchial or gastric carcinoma
  - Lymphoma
- Infective or inflammatory
  - HIV
  - Leprosy
  - Guillain-Barre syndrome
  - Lyme borreliosis
  - Chronic inflammatory demyelinating polyneuropathy (CIDP)
- Polyarteritis nodosa
- Genetic
  - Charcot–Marie–Tooth disease
  - Hereditary sensory neuropathies

Acute Painful Neuropathies

Acute painful neuropathies are transient neuropathic syndromes characterized by an acute onset of pain (over weeks) in the lower limbs. They are relatively rare compared to chronic DPN. There are two distinct syndromes, the first of which occurs within the context of poor glycaemic control, and the second with rapid improvement in glycaemic control.

Acute Painful Neuropathy of Poor Glycaemic Control

This occurs usually in diabetic subjects with poor glycaemic control. There is often an associated severe weight loss. Ellenberg coined the description of this condition as 'neuropathic cachexia' [17]. With improvements in the treatment of diabetes and the achievement of better standards of care, this complication is now rarely seen in developed countries such the UK. Patients typically experience persistent burning pain associated with allodynia (contact pain). The pain is most marked in the feet but often affects the whole of the lower extremities. As in chronic DPN, the pain is typically worse at night and often results in depression.

In acute painful neuropathies sensory loss is usually mild or absent. There are usually no motor signs, although ankle jerks may be absent. Nerve conduction studies are usually normal or mildly abnormal. Temperature discrimination threshold (small fibre function) is affected more commonly than vibration perception threshold (large-fibre function). There is complete resolution of symptoms within 12 months, and weight gain is usual with continued improvement in glycaemic control with the use of insulin.

Acute Painful Neuropathy of Rapid Glycaemic Control (Treatment-Induced Neuropathy of Diabetes–TIND)

Acute painful neuropathy of rapid glycaemic control, also known as treatment-induced neuropathy of diabetes (TIND) and occasionally as 'insulin neuritis' has recently been found to occur in around 10% people with DPN, more common than previously thought [18]. The term 'insulin neuritis' is a misnomer as the condition can follow rapid improvement in glycaemic control with oral hypoglycaemic agents; and is better termed 'acute painful neuropathy of rapid glycaemic control' or TIND [18]. A decrease in the glycosylated haemoglobin A1C of more than 3% in 3 months in individuals with chronic hyperglycaemia increases the risk of developing TIND. The natural history of acute painful neuropathies is an almost guaranteed improvement in contrast to chronic DPN [19]. Presentation is with burning pain, paraesthesiae, allodynia, often with a nocturnal

exacerbation of symptoms; and accompanying depression. There is no associated weight loss, sensory loss is often mild or absent, and there are may be autonomic neuropathy but there are no motor signs [19]. There is little or no abnormality on nerve conduction studies. Prognosis is good with usually complete resolution of symptoms within 12 months [19]. The management of painful neuropathic symptoms is as in chronic painful DPN.

Small-Fibre Neuropathy

The existence of 'small-fibre neuropathy' as a distinct entity has been advocated [20], usually within the context of young type 1 diabetes patients and prediabetes [21]. A dominant feature of this syndrome is neuropathic pain, which may be severe, with relative sparing of large-fibre functions (vibration and proprioception). The pain is sometimes described as burning, deep and aching. The sensation of 'pins and needles' (paraesthesiae) is often experienced. Contact hypersensitivity (allodynia) may be present. Autonomic involvement is common, and severely affected patients may be disabled by postural hypotension and/or gastrointestinal symptoms. The syndrome tends to develop within a few years of diabetes (and indeed in pre-diabetes) as a relatively early complication.

On clinical examination there is little evidence of objective signs of nerve damage, apart from a reduction in pinprick and temperature sensation, which are reduced in a 'stocking' and 'glove' distribution. There is relative sparing of vibration and position sense (due to relative sparing of the large diameter Aβ fibres). Muscle strength and reflexes are usually normal. Autonomic function tests are frequently abnormal and affected male patients usually have erectile dysfunction. Electrophysiological tests are usually normal. Controversy still exists as to whether small-fibre neuropathy is a distinct entity or an earlier manifestation of DPN [20].

Asymmetrical Neuropathies

Asymmetrical (or focal) neuropathies have a relatively rapid onset, and complete recovery is usual. This contrasts with chronic DPN, where there is usually no improvement in symptoms several years after onset. Unlike DPN their presence is not related to the presence of other diabetic complications. Asymmetrical neuropathies predominantly affect middle aged/older patients and are more common in men [22]. A high index of suspicion for a non-diabetic cause is advised.

Diabetic Amyotrophy

Proximal Motor Neuropathy, Lumbo-Sacral Polyradiculopathy

The syndrome of progressive asymmetrical proximal leg weakness and atrophy was first described by Garland [23], who coined the term 'diabetic amyotrophy'. This condition has also been named as 'proximal motor neuropathy' or 'lumbo-sacral polyradiculopathy'. The patient presents with severe pain which is felt deep in the thigh, but can sometimes be of burning quality and extend below the knee. The pain is usually continuous and often causes insomnia and depression. Both type 1 and type 2 patients over the age of 50 are affected [23]. There is an associated weight loss which can be severe, and can raise the possibility of an occult malignancy.

On examination there is profound wasting of the quadriceps with marked weakness in these muscle groups, although hip flexors and hip abductors can also be affected. Thigh adductors, glutei, and hamstring muscles may be involved. The knee jerk is usually reduced or absent. The profound weakness can lead to difficulty from getting out of a low chair or climbing stairs. Sensory loss is unusual, and if present indicates a coexistent DPN.

Other causes of quadriceps wasting such as nerve root and cauda equina lesions and occult malignancy causing proximal myopathy syndromes (e.g. polymyositis) should be excluded. MR imaging of the lumbo-sacral spine is now mandatory in order to exclude focal nerve root entrapment, neoplastic infiltrative lesions, and other pathologies. An erythrocyte sedimentation rate (ESR), an X-ray of the lumbar/sacral spine, a chest X-ray and ultrasound/CT of the abdomen may also be required. Electrophysiological studies may demonstrate increased femoral nerve latency and active denervation of affected muscles. CSF protein is often mildly elevated.

The cause of diabetic proximal motor neuropathy is not known. It tends to occur within the background of DPN. The combination of focal features superimposed on diffuse peripheral neuropathy may suggest vascular damage to the femoral nerve roots, as a cause of this condition. An immune mediated epineurial microvasculitis has been demonstrated in nerve biopsies.

There is scarcity of prospective studies that have looked at the natural history of proximal motor neuropathy. Pain usually starts to settle after about three months, and usually settles by one year, while the knee jerk is restored in 50% of the patients after two years. Recurrence is a rare event. Management is largely symptomatic and supportive. There is still controversy as to whether the use of insulin therapy influences the natural history of this syndrome. Some patients benefit from physiotherapy that including extension exercises to strengthen the quadriceps. The management of pain in diabetic amyotrophy is similar to that of painful DPN (see next).

Cranial Mononeuropathies

The commonest cranial mononeuropathy is the third cranial nerve palsy (Figure 15.12.4.3). The patient presents with pain in the orbit, or sometimes with a frontal headache [24]. There is typically ptosis and ophthalmoplegia, although the pupil is usually spared. Recovery occurs usually over 6 months. It is important to exclude any other cause of third cranial nerve palsy (aneurysm or tumour) by CT or MR scanning, where the diagnosis is in doubt. Fourth, sixth, and seventh cranial nerve palsies have also been described in diabetic subjects, but the association with diabetes is not as strong as that with third cranial nerve palsy.

Thoracoabdominal Neuropathy

Truncal Radiculopathy

Diabetic thoracoabdominal neuropathy (truncal radiculopathy) is characterized by an acute onset pain in a dermatomal

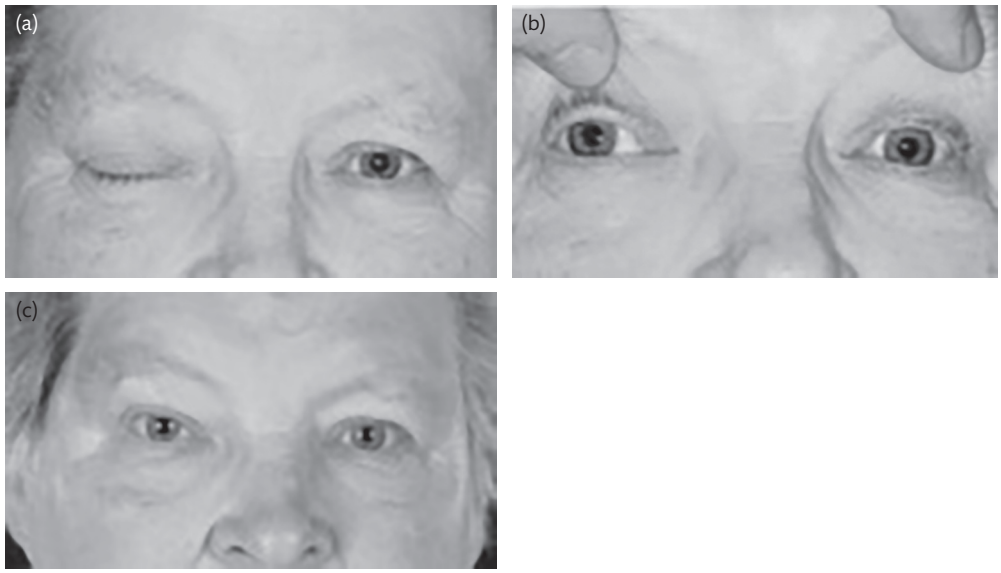


Figure 15.12.4.3 A patient with sudden onset third cranial nerve palsy (3 A) showing ptosis; ophthalmoplegia with mild dilatation of the pupil on manual opening of the eye lid (3B) and complete recovery 4 months later (3C).

distribution over the thorax or the abdomen [25] (Figure 15.12.4.4). The pain is usually asymmetrical, and can cause local bulging of the muscle due to weakness [26]. There may be patchy sensory loss over the affected area and other causes of nerve root compression should be excluded. When the patient presents with focal thoraco/abdominal pain one has to take a careful history and perform a clinical exam of the affected area assessing: pinprick sensation, light touch sensation, and dynamic mechanical allodynia. A clear difference is demonstrated when comparing with the unaffected area. Many patients have ended up with laparotomy or even worse when simple clinical examinations could have avoided this. Recovery is usually the rule within several months, although symptoms can sometimes persist for years.

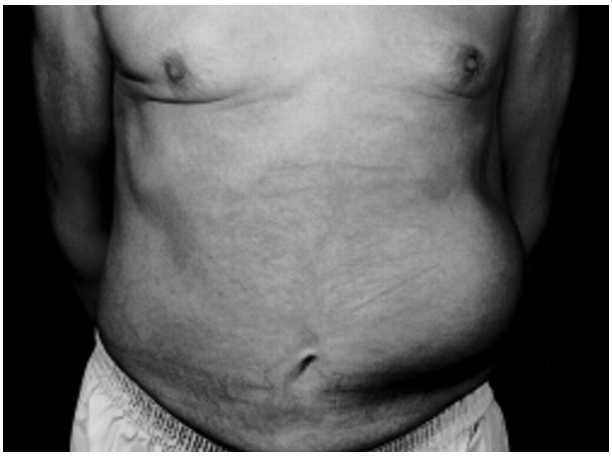


Figure 15.12.4.4 Diabetic truncal polyradiculopathy presenting as a bulge in the left abdominal wall secondary to muscle weakness.

Reproduced with permission from Weeks RA, Thomas PK, Gale AN. Abdominal pseudohernia caused by diabetic truncal radiculoneuropathy. *J Neurol Neurosurg Psychiatry*. 1999; 66(3):405. Copyright © 1999, British Medical Journal. (Ref 26).

Pressure Palsies

Carpal Tunnel Syndrome

The patient typically has pain and paraesthesia in the hands, which sometimes radiate to the forearm and are particularly marked at night. In severe cases clinical examination may reveal a reduction in sensation in the median nerve territory in the hands, and wasting of the muscle bulk in the thenar eminence. The clinical diagnosis is confirmed by median nerve conduction studies and treatment involves steroid injection and/or surgical decompression. There is generally a good response to surgery, although painful symptoms may relapse more commonly than in the non-diabetic population.

Ulnar Nerve and Other Isolated Nerve Entrapments

The ulnar nerve is also vulnerable to pressure damage at the elbow resulting in wasting of the dorsal interossei, particularly the first dorsal interosseous. This is confirmed by ulnar electrophysiological studies.

Rarely, the patients may present with wrist drop due to radial nerve palsy after prolonged sitting or while unconscious during hypoglycaemia.

In the lower limbs the common peroneal (lateral popliteal) is the most commonly affected nerve resulting in foot drop. Unfortunately, complete recovery is not usual. The lateral cutaneous nerve of the thigh is occasionally also affected with entrapment neuropathy in diabetes. Phrenic nerve involvement in association with diabetes has been described.

Pathogenesis of DPN

Despite considerable research, the specific mechanisms contributing to DPN are not completely understood [8]. There is experimental evidence for hyperglycaemia, dyslipidaemia, impaired insulin signalling and oxidative-nitrosative stress acting in concert

with other risk factors to activate several biochemical pathways, in addition to vascular factors that contribute to neuronal injury (Figures 15.12.4.5 and 15.12.4.6). These alterations promote the pathological changes of DPN including: (1) segmental demyelination and remyelination without axonal degeneration [27]; (2) a loss of large and small myelinated and unmyelinated nerve fibres [28]; (3) focal areas of demyelination on teased fibre preparations [29]. Nerve regenerative activity may also be seen with the emergence of 'regenerative, clusters' [30], containing groups of myelinated axons and non-myelinated axons sprouts. Until recently, few studies had assessed small fibres, even though these had been shown to have comparable sensitivity/specificity for detecting neuropathy [31] and constitute around 80% of all peripheral nerve fibres. Figure 15.12.4.5 shows the multifactorial pathogenesis of DPN [32].

Metabolic Pathways

Prolonged hyperglycaemia is a key risk factor for DPN and can lead to cellular damage in several ways. Hyperglycaemia induces increased formation of advanced glycation end-products (AGEs) [33]. AGEs cross-link essential proteins, lipids, or nucleic acids, altering their function and causing cellular damage [34]. Extracellular AGEs also bind to the receptor RAGE, initiating inflammatory signalling cascades, activating NADPH oxidases and generating oxidative stress [35]. The polyol pathway is the most studied pathway thought to play an important role in the pathogenesis DPN. Excess glucose is converted to sorbitol by aldose reductase resulting in increase of cellular osmolarity, reduction of NADPH levels and oxidative stress [36]. Increased glycolysis in response to excess glucose can overload the electron transport chain and initiate mitochondrial and cytosolic oxidative stress, which promote neuronal injury [37]. Moreover, the glycolysis intermediate fructose-6-phosphate also enters the hexosamine pathway which is associated with inflammatory injury [38].

Dyslipidaemia has been found to be associated with DPN [5, 39, 40]. Plasma lipoproteins, particularly low-density lipoproteins (LDLs), can be modified by oxidation (oxLDL) and/or glycation,

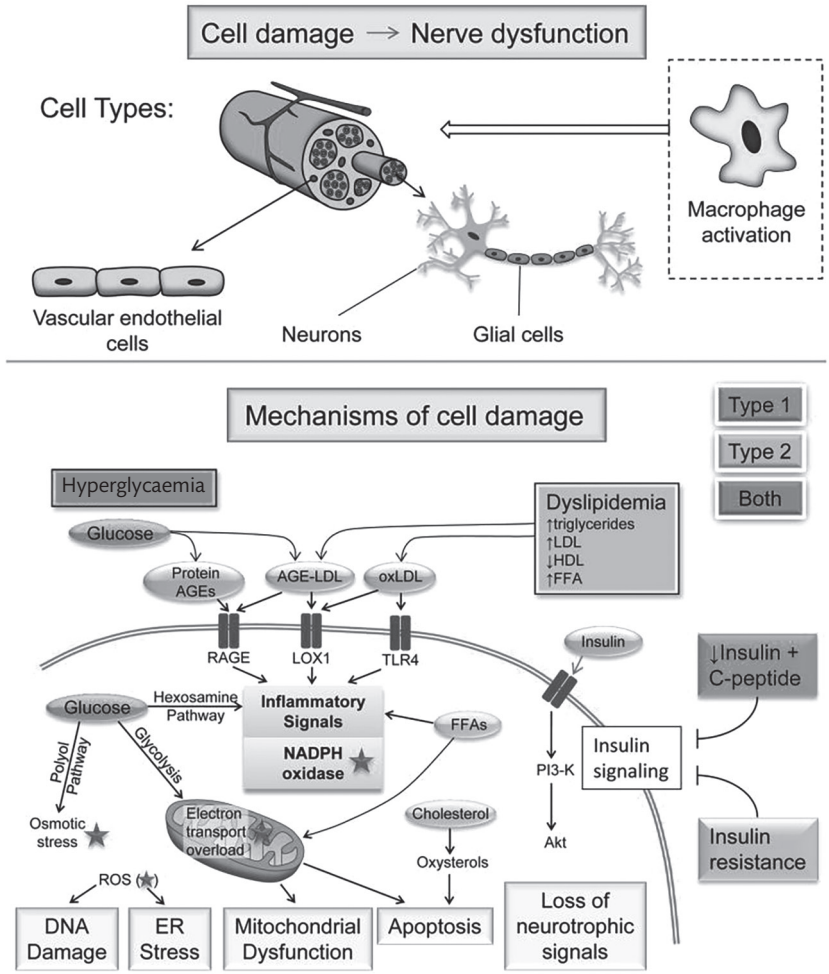


Figure 15.12.4.5 Pathological mechanisms involved in DPN. Factors linked to type 1 diabetes (yellow), type 2 diabetes (blue) and both (green) cause DNA damage, ER stress, mitochondrial dysfunction, apoptosis, and loss of neurotrophic signalling. The relative importance of the pathways in this network will vary with cell type, disease profile, and time. Abbreviations: AGE, advanced glycation end-products; LDL, low-density lipoprotein; HDL, high-density lipoprotein; FFA, free fatty acids; ROS, reactive oxygen species (red star); ER, endoplasmic reticulum; PI3K, phosphatidylinositol 3-kinase. *For a colour version of this figure, please see colour plate section.*

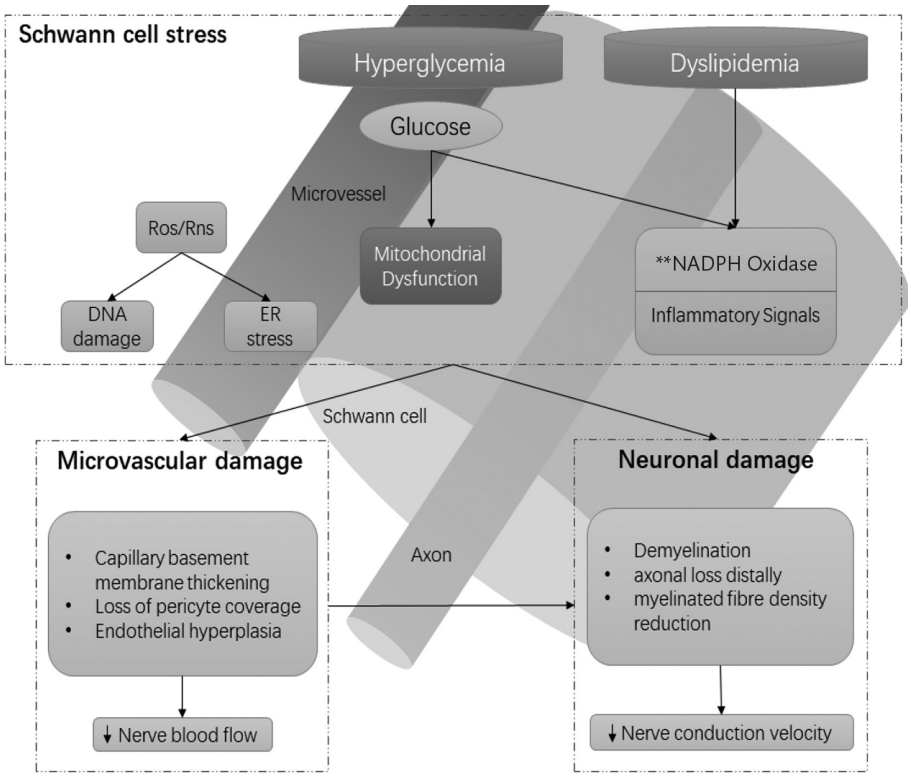


Figure 15.12.4.6 Hyperglycaemia-driven Schwann cells stress and neuronal damage. Hyperglycaemia and dyslipidaemia ultimately lead to reduction of neuronal support from Schwann cells and microvessels. Disruption of neuronal support by Schwann cells and the vascular system contributes to neuropathy, in conjunction with the direct effects of diabetes on neurons themselves.

and these modified LDLs can bind to extracellular receptors (including the oxLDL receptor LOX [41], Toll-like receptor 4 [42], and RAGE [35], triggering signalling cascades that activate NADPH oxidase and subsequent oxidative stress [41]. Free fatty acids (FFAs) result in cellular dysfunction and cell apoptosis that involves activation of inflammatory signalling, mitochondrial dysfunction and an augmented state of cellular oxidative stress [43]. Additionally, cholesterol may be oxidized to oxysterols, which have been suggested to induce neuronal apoptosis [44].

Insulin and insulin signalling dysregulation may also contribute to neuropathic damage. Although insulin does not appear to directly control glucose transport into the PNS, it is a potent neurotrophic factor capable of supporting axonal growth and survival. [45, 46]. Reduced neurotrophic signalling due to insulin deficiency (type 1 diabetes) or insulin resistance (IR; type 2 diabetes) may lead to neuronal injury [47]. In neurons, IR occurs by inhibition of the PI3K/Akt signalling pathway [48]. Disruption of this pathway may result in mitochondrial dysfunction, oxidative stress, and the development of neuropathy [47].

Oxidative Stress, Nitrosative Stress, and Inflammation

Excess glucose overloads the electron transport chain, leading to the production of superoxides and subsequent mitochondrial generation of reactive oxygen species (ROS) [48]. Hyperglycaemia also induces nitrogen generating pathways (nitrosative stress) [49]. These processes impair the capacity of the vascular endothelium to produce biologically active nitric oxide (NO), which adversely affects vascular relaxation. Superoxide combined with NO generated by the endothelial cells then leads to the formation of peroxynitrite, which attacks various biomolecules in the vascular endothelium [50]. Many of these changes trigger inflammatory mechanisms of cell stress and death. Inflammation, a potentially important pathogenic mechanism of DPN, induces activation of nuclear factor kappa B, activator protein 1, mitogen-activated protein kinases and the release of inflammatory cytokines [51]. The increase of inflammatory factors appears to be associated with both the incidence and progression of DPN [52].

Ultimately, the mechanisms outlined earlier result in mitochondrial dysfunction, endoplasmic reticulum (ER) stress and DNA damage, causing neuronal dysfunction and/or apoptosis, and subsequently manifesting as clinical DPN.

Vascular Factors

Microvascular disease of the vara nervorum has long been thought to be important in the pathogenesis of DPN [40, 53, 54]. The promotion of inflammatory cascades and other mechanisms, such as those described earlier, cause endoneurial capillary damage in DPN. This may in turn affect Schwann-cell function by disrupting access to oxygen and glucose (Figure 15.12.4.6). These microvascular changes are particularly severe in endoneurial capillaries, and include capillary basement membrane thickening, loss of pericyte coverage, and endothelial hyperplasia [55]; and the degree of microvascular changes has been correlated with both the extent of peripheral nerve fibre loss and the clinical severity of DPN [55]. Narrowing of individual capillaries might not prevent blood from passing through the endoneurial capillary bed, but the

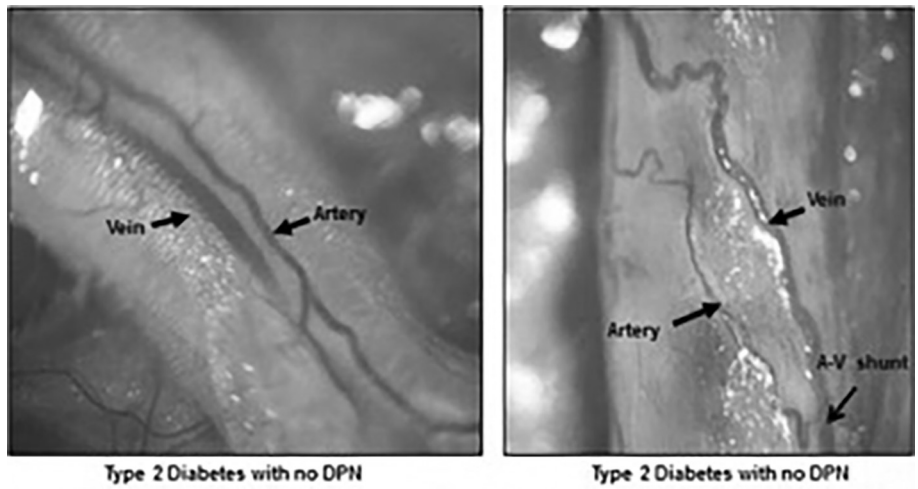


Figure 15.12.4.7 Sural nerve epineurial vessel anatomy in a patient without (left) and with DPN (right).

resulting increase in velocity of blood through endoneurial functional shunts prevents efficient oxygen extraction, causing nerve hypoxia [54, 56].

In vivo studies focused on the sural nerve in human subjects have given evidence of epineural arteriovenous shunting, that seems to cause a 'steal' phenomenon diverting blood from the nutritive endoneurial circulation (Figure 15.12.4.7) [57]. The fall in endoneural oxygen tension demonstrated in DPN [58] appears to be the result of the impairment of nerve blood flow [57].

In addition, there are several other studies providing indirect evidence implicating a vascular aetiology for diabetic neuropathy. In non-neuropathic diabetic subjects, strenuous exercise increases nerve blood flow, increasing nerve conduction velocity by an average of 4 metres/second [59]. But in neuropathic subjects whose nerve microvasculature is severely diseased, this increase in nerve conduction velocity with exercise is diminished [59]. Moreover, the correlation between nerve conduction velocity and lower limb transcutaneous oxygenation measurements in diabetes is strong; macrovascular disease appears to aggravate neuropathy and surgical restoration of perfusion improves nerve conduction velocity [60]. Several studies have also found a strong correlation between DPN and cardiovascular risk factors [5]. A population-based study of patients with type 2 diabetes found that microvascular disease conferred a risk for DPN, equivalent to conventional factors such as smoking, hypertension, and dyslipidaemia [61]. Additionally, the absence of at least one dorsalis pedis or one posterior tibial pulse compared with the presence of all peripheral pulses was associated with diabetic nephropathy and DPN [62].

Impairment of nerve blood flow has also been identified as an early manifestation in rats with streptozotocin diabetes [53]. Several vasodilators have also been demonstrated to improve nerve blood flow and nerve function in diabetic animals [53]. angiotensin-converting-enzyme (ACE) inhibitors have been found to improve nerve function in experimental [53] and human [63] DPN. Subjects with acute painful neuropathy of rapid glycaemic control (insulin neuritis, Treatment-Induced Neuropathy of Diabetes—TIND) have also been found to have severe epineurial microvascular changes and associated new vessel formation similar to that found in proliferative retinopathy, providing perhaps an even stronger evidence for the importance of microvascular factors in the pathogenesis of DPN [19].

Central Nervous System (CNS) Involvement in DPN

Recent work has suggested that DPN may not be as its name suggests and may also involve the CNS [64]. The use of advanced multimodal magnetic resonance imaging (MRI) is now making detailed CNS studies, in carefully phenotyped patients, possible. A detailed review of recent advances in this area is beyond the scope of this chapter but some of the discoveries in brief include:

- A reduction in spinal cord cross-sectional area (shrinkage) in DPN [65].
- Loss of grey matter volume (atrophy) in the primary somatosensory cortex in both painful and painless DPN [66].
- Evidence of neuronal dysfunction in the thalamus, an important gateway for most sensory information to the cerebral cortex, in patients with painless DPN [67].
- Thalamic hyperperfusion in painful DPN whereas there is a sluggish blood flow in those with painless DPN [68].
- Changes in higher brain areas, specifically the 'pain processing matrix' thought to be involved not only in detecting the location and intensity of pain but also the emotional (affective) responses [64, 69].

Autonomic Neuropathy

Abnormalities of autonomic function are very common in subjects with longstanding diabetes; however, clinically significant autonomic dysfunction is uncommon. Several systems are affected (Box 15.12.4.6). Autonomic neuropathy has a gradual onset and is slowly progressive. The prevalence of diabetic autonomic neuropathy depends on the type of population studied, and a number of tests of autonomic function employed. In the EURODIAB study the prevalence of autonomic neuropathy defined as the presence of two abnormal cardiovascular autonomic function tests, was 24%, and the

Box 15.12.4.6 Clinical consequences of autonomic neuropathy

- Cardiac autonomic neuropathy
  - sudden death
  - silent myocardial ischaemia
  - exercise intolerance
  - orthostatic hypotension
  - foot vein distension/AV shunting
- Gastrointestinal autonomic neuropathy
  - gastroparesis
  - diarrhoea or constipation
- Bladder hypomotility
  - urinary incontinence/retention
- Erectile dysfunction
- Gustatory sweating
- Reduced peripheral sweating

prevalence increased with age, duration of diabetes, poor glycaemic control and presence of cardiovascular risk factors [5].

Cardiovascular Autonomic Neuropathy

Cardiovascular autonomic neuropathy (CAN) is a serious complication of long standing diabetes and causes postural hypotension, and may be a cause of sudden death. Although CAN prevalence is very low in newly diagnosed patients with type 1 diabetes, CAN prevalence increases substantially with diabetes duration, and prevalence rates of at least 30% were observed in the DCCT/EDIC cohort after 20 years of diabetes duration [8]. In type 2 diabetes, the prevalence of CAN also increases with diabetes duration and may be present in up to 60% of patients with type 2 diabetes after 15 years. CAN may affect youth, especially young women and those with elevated A1C levels [8]. In addition, CAN is present in patients with pre-diabetes [8].

A timely diagnosis of CAN may have important clinical implications, as CAN is a major risk factor for cardiovascular mortality, arrhythmia, silent ischaemia, and major cardiovascular events.

Postural Hypotension

It is now generally accepted that a fall in systolic blood pressure of >20 mmHg is considered abnormal. Coincidental treatment with tricyclic antidepressants for neuropathic pain, and diuretics may exacerbate postural hypotension. The symptoms of postural hypotension can be disabling for some patients who may not be able to walk for more than a few minutes. Severely affected patients are prone to unsteadiness and falls. The degree of dizziness does not appear to correlate with the postural drop in blood pressure. There is increased mortality in subjects with postural hypotension.

The management of subjects with postural hypotension is challenging. Current treatments include: (1) removing any drugs that may result in orthostatic hypotension such as diuretics, β-blockers, etc.; (2) advising patients to get up from the sitting or lying position slowly, and crossing the legs; (3) increasing sodium intake up to 10 grams (185 mmol) per day and fluid intake of 2–2.5 litres/day (caution in elderly patients with heart failure); (4) the use of custom fitted elastic stockings extending to the waist; (5) treatment with fludrocortisone (starting at 100 μgm per day) while carefully monitoring electrolytes; and (6) in severe cases the alpha-1 adrenal receptor agonist, midodrine, or octreotide may be effective.

Cardiovascular Autonomic Function Tests

Five cardiovascular autonomic function tests are now widely used for the assessment of autonomic function. These tests are non-invasive, and all that is required is an electrocardiogram machine, an aneroid pressure gauge attached to a mouthpiece, a hand grip dynamometer, and sphygmomanometer. See Table 15.12.4.2 [70].

The American Diabetes Association Position Statement on diabetic neuropathy recommends screening for diabetic autonomic neuropathy at diagnosis in type 2 diabetes and 5 years after diagnosis in type 1 diabetes [8]. If negative, it recommends yearly screening subsequently and if, positive or symptomatic, to institute treatment [8].

A correct diagnosis of CAN is based on a combination of symptoms, signs and/or tests of autonomic function. These tests combine reflex responses to parasympathetic and sympathetic testing (Table 15.12.4.2). The recommended tests include the following during a period of electrocardiographic monitoring:

- A demonstration of a resting tachycardia greater than 100 beats per minute. It is useful as a risk stratification tool and a therapeutic target.
- An estimation of the beat–to-beat variation in heart rate under resting conditions while supine. This is best augmented by maximizing vagal activity by breathing at a rate of six breaths per minute. A difference between expiratory and inspiratory heart rates of greater than 15 beats per minute is expected. Less than ten beats is considered abnormal. Between these two figures is borderline. This variation with respiration can be represented as the ratio between the mean heart rates in expiration and inspiration. It is age dependent and varies between the sexes decreasing with advancing age and in females.

Table 15.12.4.2 Reference values for standard cardiovascular function tests

	Normal	Borderline	Abnormal
Heart rate tests Heart rate response to standing up (30:15 ratio)	≥ 1.04	1.01–1.03	≤ 1.00
Heart rate response to deep breathing (maximum minus minimum heart rate)	≥ 15 beats/min	11–14 beats/min	≤ 10 beats/min
Heart rate response to Valsalva manoeuvre (Valsalva ratio)	≥ 1.21	–	≤ 1.20
Blood pressure tests Blood pressure response to standing up (fall in systolic BP)	≤ 10mmHg	11–29 mmHg	≥ 30 mmHg
Blood pressure response to sustained handgrip (increase in diastolic BP)	≥ 16 mmHg	11– 15 mmHg	≤ 10 mmHg

- A measurement of the reflex response to standing which usually involves reflex tachycardia followed by bradycardia. It is usual to express this as the ratio between the R-R interval after 15 beats and at 30 beats. This is more conveniently known as the 30/15 ratio.
- Assessment of the Valsalva manoeuvre. The subject strains at 40 mmHg for 15 seconds during which there is a tachycardia and vasoconstriction. Bradycardia and a rise in blood pressure then occur following relaxation. This expressed as a ratio of the longest to shortest R-R interval. It is regarded as normal when greater than 1.03.
- The presence of postural hypotension measured after being supine for 2 minutes. A drop in systolic blood pressure of up to 10 mmHg is regarded as normal, between 10 and 29 mmHg is borderline and 30 or more is regarded as significant. There is, however, acceptance of postural hypotension as a drop in systolic greater than 20 mmHg and or a diastolic drop greater than 10 mmHg [8].
- ECG determination of corrected QT interval. This is normal below 440 ms and is considered a risk factor for ventricular arrhythmias when prolonged.
- Analysis of spectral heart rate variability (SHRV) expressed in frequency bands. This may manifest as reduced very low frequency, low frequency, and high frequency bands. A ratio of the low frequency to high frequency can be used as a measure of sympathovagal balance. SHRV is considered to provide more information than time domain analysis.

Gastrointestinal Autonomic Neuropathy

Gastroparesis

Autonomic neuropathy can reduce oesophageal motility (dysphagia and heartburn), and cause gastroparesis (reduced gastric emptying, vomiting, swings in blood sugar) [71]. The diagnosis of gastroparesis is often made on clinical grounds by the evaluation of symptoms and sometimes the presence of succussion splash, while barium swallow and follow through, and gastroscopy may reveal a large food residue in the stomach. Gastric motility and emptying studies may aid diagnosis.

Management of diabetic gastroparesis include: optimization of glycaemic control; the use of antiemetics (e.g. metoclopramide) and the use of the cholinergic agent which stimulates oesophageal motility (erythromycin which may enhance the activity of the gut peptide, motilin) [1]. Pyloric Botox injection may provide short-term benefit in some patients. Gastric electrical stimulation (GES) has recently been introduced as a treatment option in patients with drug refractory gastroparesis [1].

Severe gastroparesis causing recurrent vomiting, is associated with dehydration, swings in blood sugar and weight loss, and is an indication for hospital admission. The patient should be adequately hydrated with intravenous fluids and blood sugar should be stabilized, antiemetics could be given intravenously and if the course of the gastroparesis is prolonged, total parenteral nutrition, or feeding through a gastrostomy tube may be required.

Autonomic Diarrhoea

The usual presentation is that of diarrhoea which tends to be worse at night, or alternatively some may present with constipation. Both the diarrhoea and constipation respond to conventional treatment. Diarrhoea associated with bacterial overgrowth may respond to treatment with a broad spectrum antibiotic.

Abnormalities of Bladder Function

Autonomic bladder dysfunction is a rare complication of autonomic neuropathy and may result in hesitancy of micturition, increased frequency of micturition, and in serious cases with urinary retention associated with overflow incontinence. Such a patient is prone to urinary tract infections. Ultrasound scan of the urinary tract and urodynamic studies may be required. Treatments include mechanical methods of bladder emptying by applying suprapubic pressure, or the use of intermittent self-catheterization. Anticholinesterase drugs such as neostigmine or pyridostigmine may be useful.

Gustatory Sweating

Increased sweating usually affecting the face, and often brought about by eating (gustatory sweating) can be embarrassing to patients. Oral anticholinergic agents, including oxybutynin, propantheline, and glycopyrrolate, have improved symptoms; however adverse reactions limit their use. Clonidine has also been used with some success but is also limited by side effects including hypotension and dry mouth. Systemic side effects have led to the investigation of non-systemic approaches. Topical glycopyrrolate, a quaternary ammonium, antimuscarinic compound has been shown to significantly decrease the incidence, severity, and frequency of sweating with eating and is tolerated well [72]. Botulinum toxin has been used for gustatory sweating, though in most literature it is limited to use in unilateral, surgical-related cases.

Management of Painful DPN

Painful DPN can be extremely distressing to many patients but unfortunately, currently available treatment approaches may not completely abolish the pain [73]. An empathic approach is essential. The assessment and treatment of painful DPN should ideally involve a multidisciplinary team (MDT) that may include a diabetologist, a neurologist, the pain clinic team, specialist nurses, podiatrists, psychologists, physiotherapists, occupational therapists, and others. However, in most clinical settings this is not possible and the management falls mainly to the diabetes physician, the primary care physician, or neurologist. When treatment is started, a realistic objective would be to achieve around 50% reduction in pain intensity. However, being 'realistic' shouldn't be interpreted as less aggressive pursuit of maximum pain relief. Secondary objectives should include restoration or improvement in functional measures, quality of life, sleep, and mood.

Glycaemic Control

There is now little doubt that good blood sugar control prevents or delays the onset of DPN in type 1 diabetes [74]. Similar convincing data is lacking in type 2 diabetes and there is currently only Level B evidence [8]. The view that painful neuropathic symptoms may be improved by improving metabolic control, if necessary with the use of insulin in type 2 diabetes is not supported by evidence from controlled trials. Nevertheless, current consensus is that the first step in the management of painful neuropathy is an attempt at improving

glycaemic control where appropriate. Additionally, as cardiovascular disease is common in patients with DPN [5] and vascular risk factors (hypertriglyceridaemia, hypertension, visceral obesity, etc.) appear to be implicated in the pathogenesis of DPN [5], there is a good rationale for management of vascular risk factors beyond glycaemic control.

Pharmacotherapy

International consensus panels [73] including the National Institute for Health and Care Excellence (NICE) [75] broadly recommend that tricyclic compounds such as amitriptyline, serotonin noradrenaline reuptake inhibitors (SNRI) such as duloxetine, and the anticonvulsants, pregabalin and gabapentin as first-line agents for the management of painful DPN. Box 15.12.4.7 shows some of the commonly prescribed medications for painful DPN.

Tricyclic Compounds

Tricyclic compounds are regarded as one of the first-line treatment agents [73]. A number of double-blind clinical trials have confirmed their effectiveness beyond any doubt, although they are not licensed for use in painful DPN. However, there is a good case for their use off label and indeed this is the case. As these drugs do have unwanted side effects such as drowsiness, dry mouth, and postural hypotension, patients should be started on Imipramine or Amitriptyline at a low dose (10–25 mg taken before bed), the dose gradually titrated if necessary up to 100 mg per day. Caution should be taken in elderly patients and in those with cardiovascular disease [73]. The mechanism of action of tricyclic compounds in improving neuropathic pain is not fully understood.

Serotonin Noradrenaline Reuptake Inhibitors (SNRI)

SNRIs such as duloxetine relieve pain by increasing synaptic availability of 5-HT and noradrenaline in the descending pathways that are inhibitory to pain impulses. Duloxetine is licensed for the treatment of painful diabetic neuropathy. The efficacy of duloxetine in painful neuropathy has been investigated in several studies [73] with the 60 mg/day and 120 mg/day doses being effective in relieving painful symptoms. Duloxetine is contraindicated in those with liver disease.

Box 15.12.4.7 Pharmacological treatment of painful DPN

- **Tricyclic antidepressants (TCAs)**
  Amitriptyline 25–150 mg/day;
  Imipramine 25–150 mg/day
- **Serotonin noradrenaline reuptake inhibitors (SNRIs)**
  Duloxetine 60–120 mg/day
- **Anticonvulsants**
  Gabapentin 300–3600 mg/day
  Pregabalin 300–600 mg/day
- **Opiates**
  Tramadol 200–400 mg/day
  Tapentadol slow release 100–500 mg/day
  Oxycodone 20–80 mg/day
  Morphine sulphate SR 20–80 mg/day
- **Capsaicin cream**
  (0.075%) applied sparingly 3–4 times per day)
  8% patch applied under a specialist supervision
- **IV lignocaine**
  5 mg/kg given IV over 30 min with ECG monitoring

Anticonvulsants

Older anticonvulsants, including sodium valproate and carbamazepine, though effective, tend to have more side effects. Gabapentin and pregabalin bind to the α-2-δ subunit of the calcium channel reducing calcium flux, and thus resulting in reduced neurotransmitter release in the hyperexcited neurone. Gabapentin has been used to treat painful neuropathy for over two decades [73]. More recently pregabalin at 300–600 mg/day has been found effective in several clinical trials and is licensed for the treatment of painful DPN. Side effects of include dizziness, somnolence, and peripheral oedema.

Alpha-Lipoic Acid

Infusion of the antioxidant alpha-lipoic acid: at a dose of 600 mg per day orally or intravenously has also been found to be useful in reducing neuropathic pain [73].

Opiates

The opiate derivative tramadol (50–100 mg four times per day) has been found effective in relieving neuropathic pain [73]. Another opioid, oxycodone slow release and tapentadol slow release have also been shown to be effective in the management of neuropathic pain [73]. Because of their marked addiction potential, opioids should only be considered when all other drug combinations have failed.

Topical Capsaicin and Capsaicin Patch

Topical capsaicin works by depleting substance 'P' from nerve terminals, and there may be worsening of neuropathic symptoms for the first 2–4 weeks of application. Topical capsaicin (0.075%) applied sparingly 3–4 times per day to the affected area has also been found to relieve neuropathic pain [73]. In patients with painful DPN, capsaicin 8% patch treatment was found to provide modest pain relief and sleep quality improvements versus a placebo patch, similar in magnitude to other treatments with known efficacy, but without systemic side effects or sensory deterioration [76]. Recently, the European Commission has granted approval for a label extension for QUTENZA (capsaicin 8% patch) to include the treatment of adult diabetic patients with peripheral neuropathic pain, either alone or in combination with other medicinal products for pain. However, this treatment is provided in specialist centres as protective gear has to be worn during the application.

Intravenous Lignocaine

Intravenous lignocaine at a dose of 5 mg per kg body-weight with another 30 minutes with a cardiac monitor in situ, has been found to be effective in relieving neuropathic pain for up to 2–6 weeks [73]. This form of treatment is useful in subjects who are having severe pain which is not responding to these agents, although it does necessitate bringing the patient into hospital for a few hours.

Combination Treatment

Combinations of first-line therapies may be considered if there is pain, despite a change in first-line monotherapy [73]. If pain is still inadequately controlled, opioids such as tramadol and oxycodone may be added in a combination treatment [73]. A number of areas

relating to painful DPN warrant further investigation including population-based prevalence and natural history studies, trials using active comparators rather than placebo [77, 78], assessment of combination therapies in addition to placebo, and longer-term studies of the efficacy and durability of treatments of painful DPN [73].

Management of Disabling Painful Neuropathy Not Responding to Pharmacological Treatment

Neuropathic pain can sometimes be extremely severe. Unfortunately some patients are not helped by conventional pharmacological treatment. Such patients may respond to electrical spinal cord stimulation which relieves both background and peak neuropathic pains [79]. This treatment is provided in specialist centres and a recent two-centre prospective study has confirmed long-term efficacy [80].

REFERENCES

1. Tesfaye S, Boulton AJ, eds. *Diabetic Neuropathy*. Oxford: Oxford University Press, 2009.
2. Dyck PJ, Kratz KM, Karnes JL, *et al.* The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: the Rochester Diabetic Neuropathy Study. *Neurology*, 1993; **43**: 817–24.
3. Shaw JE, Zimmet PZ. The epidemiology of diabetic neuropathy. *Diabetes Rev*, 1999; **7**: 245–52.
4. Tesfaye S, Stephens L, Stephenson J, *et al.* The prevalence of diabetic neuropathy and its relation to glycaemic control and potential risk factors: the EURODIAB IDDM Complications Study. *Diabetologia*, 1996; **39**: 1377–84.
5. Tesfaye S, Chaturvedi N, Eaton SEM, Witte D, Ward JD, Fuller J. Vascular risk factors and diabetic neuropathy. *New Engl J Med*, 2005; **352**: 341–50.
6. Thomas PK. Metabolic neuropathy. *J Roy Coll Phys (Lond)*, 1973; **7**: 154–74.
7. Watkins PJ, Edmonds ME. Clinical features of diabetic neuropathy. In: Pickup J, Williams G, eds. *Textbook of Diabetes*, **Vol. 2**. Oxford: Wiley-Blackwell, 1997: 50.1–50.19.
8. Pop-Busui R, Boulton AJ, Feldman EL, *et al.* Diabetic neuropathy: a position statement by the American Diabetes Association. *Diabetes Care*, 2017; **40**: 136–54.
9. Tesfaye S, Boulton AJ, Dyck PJ, *et al.* Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care*, 2010; **33**: 2285–93.
10. Andreassen CS, Jakobsen J, Ringgaard S, Ejskjaer N, Andersen H. Accelerated atrophy of lower leg and foot muscles—a follow-up study of long-term diabetic polyneuropathy using magnetic resonance imaging (MRI). *Diabetologia*, 2009; **52**: 1182–91.
11. Tesfaye S, Boulton AJ, Dickenson AH. Mechanisms and management of diabetic painful distal symmetrical polyneuropathy. *Diabetes Care*, 2013; **36**: 2456–65.
12. Selvarajah D, Cash T, Sankar A, *et al.* The contributors of emotional distress in painful diabetic neuropathy. *Diab Vasc Dis Res*, 2014; **11**: 218–25.
13. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. *N Engl J Med*, 2017; **376**: 2367–75.
14. Rajbhandari SM, Jenkins R, Davies C, Tesfaye S. Charcot neuroarthropathy in diabetes mellitus. *Diabetologia*, 2002; **45**: 1085–96.
15. Papanas N, Ziegler D. New vistas in the diagnosis of diabetic polyneuropathy. *Endocrine*, 2014; **47**: 690–8.
16. Binns-Hall O, Selvarajah D, Sanger D, Walker J, Scott A, Tesfaye S. One-stop microvascular screening service: an effective model for the early detection of diabetic peripheral neuropathy and the high-risk foot. *Diabet Med*, 2018; **35**: 887–94.
17. Ellenberg M. Diabetic neuropathic cachexia. *Diabetes*, 1974; **23**: 418–23.
18. Gibbons CH. Treatment-induced neuropathy of diabetes. *Curr Diab Rep*, 2017; **17**: 127.
19. Tesfaye S, Malik R, Harris N, *et al.* Arteriovenous shunting and proliferating new vessels in acute painful neuropathy of rapid glycaemic control (insulin neuritis). *Diabetologia*, 1996; **39**: 329–35.
20. Vinik AI, Park TS, Stansberry KB, Pittenger GL. Diabetic neuropathies. *Diabetologia*, 2000; **43**: 957–73.
21. Singleton JR, Smith AG, Bromberg MB. Increased prevalence of impaired glucose tolerance in patients with painful sensory neuropathy. *Diabetes Care*, 2001; **24**: 1448–53.
22. Matikainen E, Juntunen J. Diabetic neuropathy: epidemiological, pathogenetic, and clinical aspects with special emphasis on type 2 diabetes mellitus. *Acta Endocrinol Suppl (Copenh)*, 1984; **262**: 89–94.
23. Garland H. Diabetic amyotrophy. *Br Med J*, 1955; **2**: 1287–90.
24. Asbury AK. Aldredge H, Hershberg R, Fisher CM. Oculomotor palsy in diabetes mellitus: a clinicopathological study. *Brain*, 1970; **93**: 555–7.
25. Ellenberg M. Diabetic truncal mononeuropathy—a new clinical syndrome. *Diabetes Care*, 1978; **1**: 10–13.
26. Weeks RA, Thomas PK, Gale AN. Abdominal pseudohernia caused by diabetic truncal radiculoneuropathy. *J Neurol Neurosurg Psychiatry*, 1999; **66**: 405.
27. Malik RA, Tesfaye S, Newrick PG, *et al.* Sural nerve pathology in diabetic patients with minimal but progressive neuropathy. *Diabetologia*, 2005; **48**: 578–85.
28. Malik RA, Newrick PG, Sharma AK, *et al.* Microangiopathy in human diabetic neuropathy: relationship between capillary abnormalities and the severity of neuropathy. *Diabetologia*, 1989; **32**: 92–102.
29. Said G, Slama G, Selva J. Progressive centripetal degeneration of axons in small fibre diabetic polyneuropathy. *Brain*, 1983; **106**(Pt 4): 791–807.
30. Bradley JL, Thomas PK, King RH, *et al.* Myelinated nerve fibre regeneration in diabetic sensory polyneuropathy: correlation with type of diabetes. *Acta Neuropathol*, 1995; **90**: 403–10.
31. Gibbons CH, Freeman R, Veves A. Diabetic neuropathy: a cross-sectional study of the relationships among tests of neurophysiology. *Diabetes Care*, 2010; **33**: 2629–34.
32. Callaghan BC, Cheng HT, Stables CL, Smith AL, Feldman EL. Diabetic neuropathy: clinical manifestations and current treatments. *Lancet Neurol*, 2012; **11**: 521–34.
33. Sugimoto K, Yasujima M, Yagihashi S. Role of advanced glycation end products in diabetic neuropathy. *Curr Pharm Des*, 2008; **14**: 953–61.
34. Duran-Jimenez B, Dobler D, Moffatt S, *et al.* Advanced glycation end products in extracellular matrix proteins contribute to the failure of sensory nerve regeneration in diabetes. *Diabetes*, 2009; **58**: 2893–903.
35. Vincent AM, Perrone L, Sullivan KA, *et al.* Receptor for advanced glycation end products activation injures primary sensory neurons via oxidative stress. *Endocrinology*, 2007; **148**: 548–58.
36. Obrosova IG. Increased sorbitol pathway activity generates oxidative stress in tissue sites for diabetic complications. *Antioxid Redox Signal*, 2005; **7**: 1543–52.

37. Feldman EL, Nave KA, Jensen TS, Bennett D. New horizons in diabetic neuropathy: mechanisms, bioenergetics, and pain. *Neuron*, 2017; **93**: 1296–313.
38. Vincent AM, Callaghan BC, Smith AL, Feldman EL. Diabetic neuropathy: cellular mechanisms as therapeutic targets. *Nat Rev Neurol*, 2011; **7**: 573–83.
39. Wiggin TD, Sullivan KA, Pop-Busui R, Amato A, Sima AA, Feldman EL. Elevated triglycerides correlate with progression of diabetic neuropathy. *Diabetes*, 2009; **58**: 1634–40.
40. Smith AG, Singleton JR. Obesity and hyperlipidemia are risk factors for early diabetic neuropathy. *J Diabetes Complications*, 2013; **27**: 436–42.
41. Vincent AM, Hayes JM, McLean LL, Vivekanandan-Giri A, Pennathur S, Feldman EL. Dyslipidemia-induced neuropathy in mice: the role of oxLDL/LOX-1. *Diabetes*, 2009; **58**: 2376–85.
42. Nowicki M, Muller K, Serke H, Kosacka J, Vilser C, Ricken A, Spanel-Borowski K. Oxidized low-density lipoprotein (oxLDL)-induced cell death in dorsal root ganglion cell cultures depends not on the lectin-like oxLDL receptor-1 but on the toll-like receptor-4. *J Neurosci Res*, 2010; **88**: 403–12.
43. Padilla A, Descorbeth M, Almeyda AL, Payne K, De Leon M. Hyperglycemia magnifies Schwann cell dysfunction and cell death triggered by PA-induced lipotoxicity. *Brain Res*, 2011; **1370**: 64–79.
44. Jang ER, Lee CS. 7-ketocholesterol induces apoptosis in differentiated PC12 cells via reactive oxygen species-dependent activation of NF-kappaB and Akt pathways. *Neurochem Int*, 2011; **58**: 52–9.
45. Mielke JG, Wang YT. Insulin, synaptic function, and opportunities for neuroprotection. *Prog Mol Biol Transl Sci*, 2011; **98**: 133–86.
46. Xu QG, Li XQ, Kotecha SA, Cheng C, Sun HS, Zochodne DW. Insulin as an *in vivo* growth factor. *Exp Neurol*, 2004; **188**: 43–51.
47. Kim B, Feldman EL. Insulin resistance in the nervous system. *Trends Endocrinol Metab*, 2012; **23**: 133–41.
48. Kim B, McLean LL, Philip SS, Feldman EL. Hyperinsulinemia induces insulin resistance in dorsal root ganglion neurons. *Endocrinology*, 2011; **152**: 3638–47.
49. Vincent AM, Russell JW, Low P, Feldman EL. Oxidative stress in the pathogenesis of diabetic neuropathy. *Endocr Rev*, 2004; **25**: 612–28.
50. Pacher P, Obrosova IG, Mabley JG, Szabo C. Role of nitrosative stress and peroxynitrite in the pathogenesis of diabetic complications. Emerging new therapeutical strategies. *Curr Med Chem*, 2005; **12**: 267–75.
51. Roman-Pintos LM, Villegas-Rivera G, Rodriguez-Carrizalez AD, Miranda-Diaz AG, Cardona-Munoz EG. Diabetic polyneuropathy in type 2 diabetes mellitus: inflammation, oxidative stress, and mitochondrial function. *J Diabetes Res*, 2016; **2016**: 3425617.
52. Herder C, Kannenberg JM, Huth C, *et al.* Proinflammatory cytokines predict the incidence and progression of distal sensorimotor polyneuropathy: KORA F4/FF4 Study. *Diabetes Care*, 2017; **40**: 569–76.
53. Cameron NE, Eaton SE, Cotter MA, Tesfaye S. Vascular factors and metabolic interactions in the pathogenesis of diabetic neuropathy. *Diabetologia*, 2001; **44**: 1973–88.
54. Ostergaard L, Finnerup NB, Terkelsen AJ, *et al.* The effects of capillary dysfunction on oxygen and glucose extraction in diabetic neuropathy. *Diabetologia*, 2015; **58**: 666–77.
55. Malik RA, Tesfaye S, Thompson SD, *et al.* Endoneurial localisation of microvascular damage in human diabetic neuropathy. *Diabetologia*, 1993; **36**: 454–9.
56. Goncalves NP, Vaegter CB, Andersen H, Ostergaard L, Calcutt NA, Jensen TS. Schwann cell interactions with axons and microvessels in diabetic neuropathy. *Nat Rev Neurol*, 2017; **13**: 135–47.
57. Tesfaye S, Harris N, Jakubowski JJ, *et al.* Impaired blood flow and arterio-venous shunting in human diabetic neuropathy: a novel technique of nerve photography and fluorescein angiography. *Diabetologia*, 1993; **36**: 1266–74.
58. Newrick PG, Wilson AJ, Jakubowski J, Boulton AJ, Ward JD. Sural nerve oxygen tension in diabetes. *Br Med J (Clin Res Ed)*, 1986; **293**: 1053–4.
59. Tesfaye S, Harris ND, Wilson RM, Ward JD. Exercise-induced conduction velocity increment: a marker of impaired peripheral nerve blood flow in diabetic neuropathy. *Diabetologia*, 1992; **35**: 155–9.
60. Young MJ, Veves A, Smith JV, Walker MG, Boulton AJ. Restoring lower limb blood flow improves conduction velocity in diabetic patients. *Diabetologia*, 1995; **38**: 1051–4.
61. Brownrigg JR, Hughes CO, Burleigh D, *et al.* Microvascular disease and risk of cardiovascular events among individuals with type 2 diabetes: a population-level cohort study. *Lancet Diabetes Endocrinol*, 2016; **4**: 588–97.
62. Mohammedi K, Woodward M, Zoungas S, *et al.* Absence of peripheral pulses and risk of major vascular outcomes in patients with type 2 diabetes. *Diabetes Care*, 2016; **39**: 2270–7.
63. Malik RA, Williamson S, Abbott C, *et al.* Effect of angiotensin-converting-enzyme (ACE) inhibitor trandolapril on human diabetic neuropathy: randomised double-blind controlled trial. *Lancet*, 1998; **352**: 1978–81.
64. Tesfaye S, Selvarajah D, Gandhi R, *et al.* Diabetic peripheral neuropathy may not be as its name suggests: evidence from magnetic resonance imaging. *Pain*, 2016; **157 Suppl 1**: S72–80.
65. Eaton SE, Harris ND, Rajbhandari SM,*et al.* Spinal-cord involvement in diabetic peripheral neuropathy. *Lancet*, 2001; **358**: 35–6.
66. Selvarajah D, Wilkinson ID, Maxwell M, *et al.* Magnetic resonance neuroimaging study of brain structural differences in diabetic peripheral neuropathy. *Diabetes Care*, 2014; **37**: 1681–8.
67. Selvarajah D, Wilkinson ID, Emery CJ, *et al.* Thalamic neuronal dysfunction and chronic sensorimotor distal symmetrical polyneuropathy in patients with type 1 diabetes mellitus. *Diabetologia*, 2008; **51**: 2088–92.
68. Selvarajah D, Wilkinson ID, Gandhi R, Griffiths PD, Tesfaye S. Microvascular perfusion abnormalities of the Thalamus in painful but not painless diabetic polyneuropathy: a clue to the pathogenesis of pain in type 1 diabetes. *Diabetes Care*, 2011; **34**: 718–20.
69. Watanabe K, Hirano S, Kojima K, *et al.* Altered cerebral blood flow in the anterior cingulate cortex is associated with neuropathic pain. *J Neurol Neurosurg Psychiatry*, 2018; **89**: 1082–7.
70. Ewing DJ, Martyn CN, Young RJ, Clarke BF. The value of cardiovascular autonomic function tests: ten years experience in diabetes. *Diabetes Care*, 1985; **8**: 491–98.
71. Horowitz M, Fraser R. Disordered gastric motor function in diabetes mellitus. *Diabetologia*, 1994; **37**: 543–51.
72. Shaw JE, Abbott CA, Tindle K, Hollis S, Boulton AJ. A randomised controlled trial of topical glycopyrrolate, the first specific treatment for diabetic gustatory sweating. *Diabetologia*, 1997; **40**: 299–301.

73. Tesfaye S, Vileikyte L, Rayman G, *et al.*; Toronto Expert Panel on Diabetic Neuropathy. Painful diabetic peripheral neuropathy: consensus recommendations on diagnosis, assessment and management. *Diabetes Metab Res Rev*, 2011; **27**: 629–38.
74. Diabetes Control and Complications Trial Research Group: the effect of intensive diabetes therapy on the development and progression of neuropathy. *Ann Int Med* 1995; **122**: 561–8.
75. Centre for Clinical Practice at NICE (UK). *Neuropathic Pain: The Pharmacological Management of Neuropathic Pain in Adults in Non-specialist Settings*. London: National Institute for Health and Care Excellence, (UK), 2013.
76. Simpson DM, Robinson-Papp J, Van J, *et al.* Capsaicin 8% patch in painful diabetic peripheral neuropathy: a randomized, double-blind, placebo-controlled study. *J Pain*, 2017; **18**: 42–53.
77. Tesfaye S, Wilhelm S, Lledo A, *et al.* Duloxetine and pregabalin: high-dose monotherapy or their combination? The 'COMBO-DN study'—a multinational, randomized, double-blind, parallel-group study in patients with diabetic peripheral neuropathic pain. *Pain*, 2013; **154**: 2616–25.
78. Selvarajah D, Petrie J, White D, *et al.* Multicentre, double-blind, crossover trial to identify the Optimal Pathway for TreatIng neurOpathic paiN in Diabetes Mellitus (OPTION-DM): study protocol for a randomised controlled trial. *Trials*, 2018; **19**: 578.
79. Tesfaye S, Watt J, Benbow SJ, Pang KA, Miles J, MacFarlane IA. Electrical spinal cord stimulation for painful diabetic peripheral neuropathy. *Lancet*, 1996; **348**: 1696–70.
80. van Beek M, Geurts JW, Slangen R, *et al.* Severity of neuropathy is associated with long-term spinal cord stimulation outcome in painful diabetic peripheral neuropathy: five-year follow-up of a prospective two-center clinical trial. *Diabetes Care*, 2018; **41**: 32–8.

15.13

Macrovascular Disease in Diabetes

15.13.1 Mechanisms of Macrovascular Disease in Diabetes

Mark T. Kearney, Peysh A. Patel, and Richard M. Cubbon

Introduction

Type 2 diabetes mellitus represents a growing global public health problem, driven by a demographic shift towards ageing in the population and increased prevalence of obesity. Over 25% of patients presenting to hospital with acute myocardial infarction have comorbid diabetes, and a substantial further proportion have lesser degrees of impaired glycaemic control (pre-diabetes) [1]. Cardiovascular complications remain the most common cause of death in people with diabetes, with major macrovascular events occurring 15 years earlier in those with diabetes than in the general population [2]. Atherogenesis refers to the pathological process wherein fibro-fatty deposits accrue within the arterial wall as a result of chronic injury and inflammation. Diabetes accelerates atherogenesis, resulting in more diffuse and complex lesions, and also impairs coronary vessel repair after percutaneous intervention [3].

Insulin resistance is defined, in principle, as a reduction in the metabolic and/or signalling responses to insulin at cell, tissue, or organismal level. Most commonly it is taken to denote impaired blood glucose lowering action of insulin [4]. Where pancreatic beta-cells are unable to compensate fully for the resulting increased insulin demand, hyperglycaemia ensues, ranging through glucose intolerance to the clinical phenomenon of type 2 diabetes mellitus. This chapter explores the key components of the systemic insulin resistance syndrome that potentially contribute to atherogenesis and atherosclerotic plaque rupture. We also discuss how local perturbation of insulin signalling in the arterial wall may promote atherogenesis, independent of changes in systemic glucose and lipid metabolism, noting that insulin receptors are abundantly expressed in vascular endothelium, vascular smooth muscle cells (VSMCs) and leucocytes. In particular, the notion of cell- and tissue-specific differences in insulin resistance extending beyond classical metabolic tissues, such as liver, adipose tissue, and muscle, is reviewed. Finally, a brief overview is provided of how diabetes impairs endogenous vascular reparative processes which serve to mitigate atherogenesis.

Atherogenesis: From Fatty Streak to Plaque Rupture

The macrovascular risk associated with diabetes mellitus primarily relates to accelerated atherosclerosis in large and medium-sized arteries. Atherogenesis is a multifaceted disease process, perhaps best characterized as an 'inflammatory' disorder of vessel walls associated with progressive accumulation of oxidized lipids in a plaque with a necrotic core. These lesions impinge upon the vascular lumen and are susceptible to rupture [5]. The pathology of atherosclerotic lesions in patients with diabetes is in many regards indistinguishable from that seen in patients without diabetes, although the lesions are generally more advanced and diffusely distributed [6]. A unifying concept thought to explain initiation of atherogenesis is 'endothelial dysfunction', which describes a complex disturbance of endothelial biology, associated with reduced nitric oxide (NO) bioavailability, oxidative stress, and leukocyte adhesion molecule expression. This homeostatic failure initiates the cascade of events that underpin the atherogenic cascade (**Figure 15.13.1.1**).

The earliest discernible vascular lesion is the so-called fatty streak, which is often present in young adults [7]. It occurs as a response to receptor-mediated transcytosis of low-density lipoproteins (LDL) and other apoB-containing lipoproteins across the endothelium, with subsequent trapping in the arterial intima. These are targets for modifications including oxidation, which have pro-inflammatory effects via recruitment of monocytes to the intima, where they terminally differentiate into macrophages. Oxidized LDL can then be internalized by macrophages and stored in cytoplasmic lipid droplets [8]. Progressive failure to clear these modified lipids results in

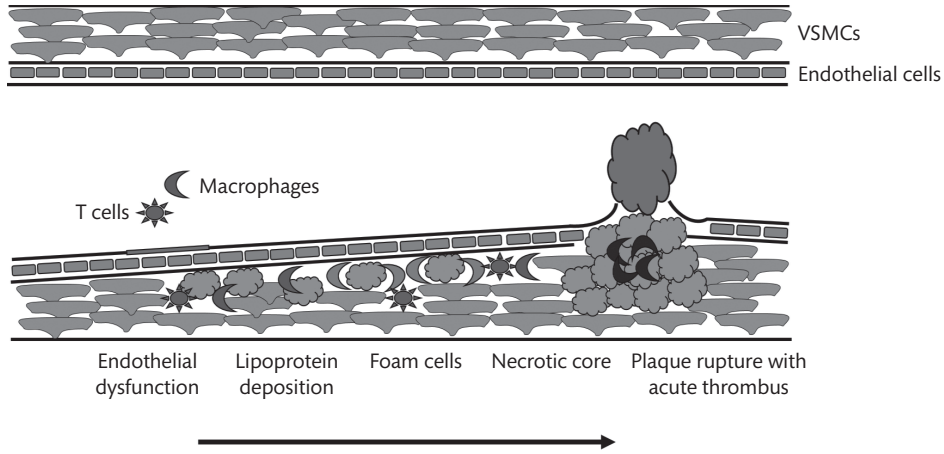


Figure 15.13.1.1 Processes that underpin atherothrombosis. The black arrow indicates time, spanning decades. VSMCs = vascular smooth muscle cells.

the formation of 'foam cells', which characterize lesion progression from early to intermediate stages and, in combination with T lymphocytes, are the histological hallmark of fatty streaks.

Foam cells release chemokines and cytokines, and express adhesion receptors which orchestrate further recruitment of macrophages and T cells, perpetuating the inflammatory process [9]. Supplementary release of growth factors such as platelet derived growth factor (PDGF) and fibroblast growth factor (FGF) results in migration of VSMC into the intima. This provides the major source of extracellular matrix (ECM) in atherosclerotic lesions [10].

Apoptosis of foam cells within the lesions results in formation of a necrotic core, which is essential for the development of the propensity for plaque rupture [11]. This core is moreover impregnated with oxidized lipids and tissue factors, rendering it highly thrombogenic when exposed to circulating blood. The lipid-rich core is covered by a fibrous cap formed by VSMCs and the ECM components that they secrete, which include collagen and proteoglycans. Such advanced lesions may remain stable for years, and be entirely asymptomatic if non-occlusive. However, activated leukocytes within the plaque produce matrix metalloproteinases (MMP), collagenases, and elastases, which collectively act to degrade the fibrous cap. Apoptosis of VSMCs also destabilizes the cap by reducing ECM synthesis. Rupture of the fibrous cap results in exposure of the underlying necrotic core and surrounding extracellular components, such as collagen and von Willebrand factor (vWF), to circulating blood. This results in activation of endogenous haemostatic cascades. Although these processes are crucial in physiological haemostasis, their activation in the context of acute plaque rupture generates thrombus within an already narrowed vascular lumen. If this causes sufficient impairment of blood flow, ischaemic injury occurs downstream, as typified by clinical presentations such as acute coronary syndromes.

The Metabolic Syndrome

Insulin resistance and central obesity are intimately associated with a cluster of cardiovascular risk factors, frequently collectively referred to as the 'metabolic syndrome' [12]. While various diagnostic criteria exist, these generally capture the coexistence of central obesity, dysglycaemia, dyslipidaemia, and hypertension, with broader descriptions also including factors such as inflammation, oxidative stress, and a pro-thrombotic tendency. The incremental value of such wider definitions of the syndrome beyond its constituent parts has been questioned, however there is broad consensus that it captures a unifying driver of multiple pro-atherogenic phenomena. Unpicking the causal relationships involved in these associations with vascular disease, and the mechanisms that underpin them are a major focus of current research.

The Insulin Signalling Cascade

Understanding of insulin resistance and its role both in the complex metabolic perturbation of the metabolic syndrome, and directly in the tissue pathology of the vessel wall, first requires understanding of the cellular pathways whereby insulin exerts its many effects on tissue. These are summarized in Figure 15.13.1.2. The insulin receptor is a transmembrane receptor tyrosine kinase (RTK). Upon ligand binding it undergoes autophosphorylation before recruiting and phosphorylating insulin receptor substrates (IRS) and other adaptor proteins [13]. These then bind to docking proteins containing Src-homology-2 (SH2) domains. In broad terms, proteins bearing SH2 domains can be subdivided into two classes: enzymes such as phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K), and adaptor proteins such as growth factor receptor-bound protein 2 (Grb2) [14]. Insulin signal transduction is discussed in more detail in Chapter 1.3, 'Molecular Aspects of Hormonal Regulation'.

Class IA PI3K are most strongly implicated in insulin's metabolic actions. Their primary substrate is phosphatidylinositol 4,5-bisphosphate (PtdIns(4,5)P_2), which is phosphorylated to yield PtdIns(3,4,5)P_3 (or PIP_3). This is a membrane-resident lipid second messenger that is able to initiate multiple downstream enzyme cascades by recruiting proteins with Pleckstrin homology (PH) domains specific for PIP_3 [15]. AKT is a serine/threonine kinase of particular relevance, functioning as a critical regulatory node in signalling [16]. Activated AKT has multiple downstream targets which mediate regulation of diverse cellular functions. In skeletal and cardiac muscle AKT action drives translocation of glucose

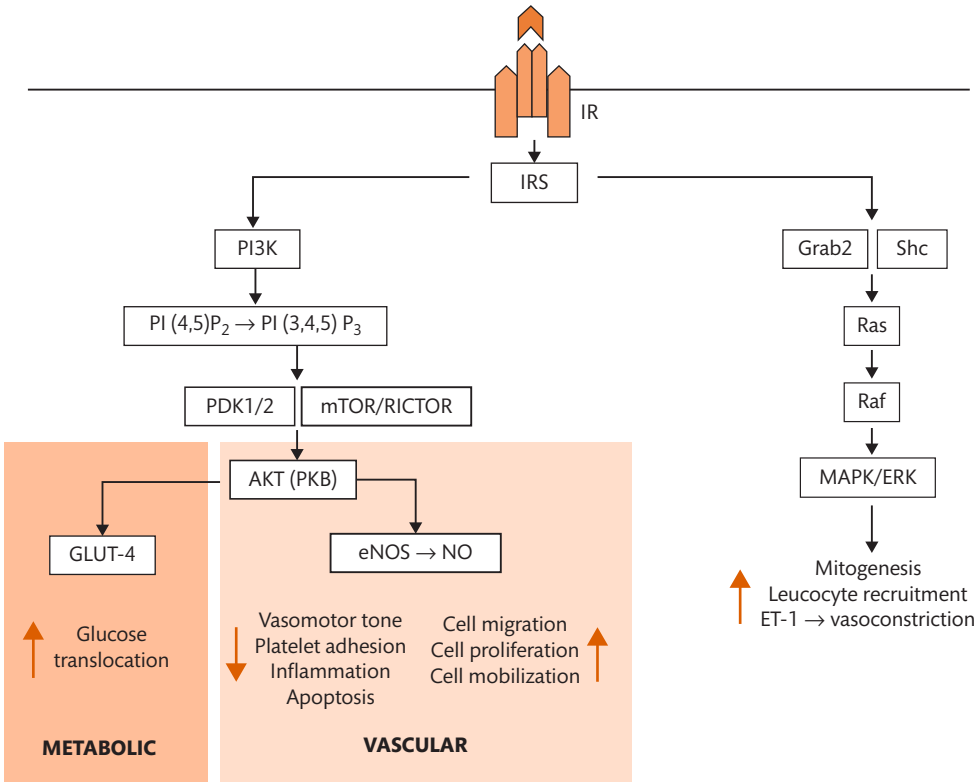


Figure 15.13.1.2 Simplified schematic of key elements of the insulin signalling network implicated in vascular disease. IR, insulin receptor; IRS, insulin receptor substrate; PI3K, phosphoinositide-3-kinase; $PI(4,5)P_2$, phosphatidylinositol 4,5-biphosphate, $PI(3,4,5)P_3$, phosphatidylinositol 3,4,5-triphosphate; PDK, phosphoinositide-dependent kinase; mTOR, mammalian target of rapamycin; RICTOR, rapamycin-insensitive companion on mTOR; GLUT, glucose transporter protein; NOS, nitric oxide synthase; NO, nitric oxide; Grb2, growth-factor-receptor-bound protein 2; ET-1, endothelin-1.

transporter protein 4 (GLUT4)-containing vesicles to the plasma membrane, stimulating cellular glucose uptake. In adipose tissue it plays a similar role, and is also critical in insulin's suppression of triglyceride breakdown to release fatty acids. In liver, insulin's roles include promotion of fatty acid and VLDL synthesis, and suppression of liver glucose output. While much work has focussed on the role of insulin in these 'canonical' insulin responsive tissues, insulin also plays critical roles in other tissues critical for vascular disease. In endothelial cells, for example, AKT activated by insulin phosphorylates endothelial nitric oxide synthase (eNOS) at serine 1177, augmenting catalytic activity and production of NO [17]. NO is considered one of the most important mediators of vascular homeostasis, regulating vascular tone, inflammation, thrombosis, cell growth, and migration.

As well as being important for metabolic regulation, insulin is also a potent growth factor. The Grb2/mitogen-activated protein kinase (MAPK) pathway is essential for insulin's mitogenic effects [18]. Grb2 binds to phosphorylated IRS or Shc, after insulin receptor activation. A phosphorylation cascade eventually results in activation of the kinases ERK-1 and 2. Activated ERK can phosphorylate a range of cellular targets that regulate gene expression, cellular differentiation, and mitogenesis. They also induce surface adhesion molecules on endothelial cells which facilitate leucocyte recruitment. ERK signalling also potentiates release of endothelin-1 (ET-1), which has potent vasoconstrictor and potentially pro-atherogenic effects.

Metabolic Syndrome Components and Atherogenesis

Dyslipidaemia

The dyslipidaemia associated with the metabolic syndrome comprises elevated triglycerides, reduced high-density lipoprotein (HDL) cholesterol and smaller, denser LDL cholesterol particles [19] (see Chapter 15.13.4 'Diabetic Dyslipidaemia'). Triglycerides are a source of free fatty acids (FFA), which in excess can result in cardiovascular lipotoxicity through various mechanisms. FFA-induced mitochondrial dysfunction, uncoupled eNOS (a shift from dimeric to monomeric structure, resulting in generation of superoxide instead of NO) arising from direct enzymatic modification or depleted cofactors, and activation of nicotinamide adenine dinucleotide phosphate (NADPH) oxidases (NOX) results in generation of superoxide (a form of reactive oxygen species, or ROS) and vascular oxidative stress [20]. These phenomena, with the associated pro-inflammatory signalling cascade activation, are directly implicated in insulin resistance, creating a vicious cycle [20]. Ceramide derived from saturated FFA metabolism also directly inhibits insulin-stimulated AKT activation. It impairs glucose uptake in human subjects with obesity and also contributes to oxidized LDL-induced calcification of human VSMCs, a feature of advanced atherosclerosis. Lastly, small, dense LDL cholesterol particles are particularly susceptible to oxidative modification which, as described earlier, is potently atherogenic.

Hyperglycaemia

Hyperglycaemia results in diverse alterations in vascular cell biology which can potentiate atherosclerosis. The formation of advanced glycation end products (AGE) by non-enzymatic glycation driven by high blood glucose levels indirectly promotes ROS production, a pro-inflammatory state, and impaired insulin signal transduction; moreover, by binding to AGE receptors (termed RAGE) on leukocytes, vascular endothelial cells, and VSMCs, AGEs can directly influence cell biology [21]. Higher AGE levels have been found in patients with type II diabetes than in healthy and type I diabetic patients. Moreover, their abundance appears to be increased in patients with diabetes and coronary disease compared to those without coronary disease. Glycated haemoglobin (HbA1c) levels appear to correlate with AGE, and pharmacotherapy with agents that lower blood glucose and AGE is associated with reductions in microvascular sequelae such as nephropathy and retinopathy, though effects on macrovascular complications are less established [22].

Shunting of glucose metabolism through the hexosamine biosynthesis pathway results in modification of proteins such as AKT via O-linked N-acetylglucosamine (O-GlcNAcylation), which is associated with increased vascular calcification in the context of diabetes [23]. Hyperglycaemia activates protein kinase C (PKC), which augments monocyte recruitment, foam cell formation, and promotes switching to a synthetic VSMC phenotype which enables cell migration and proliferation [24]. Consistent with this observation, atherosclerosis-prone mice with genetic deletion of PKC display reduced atherosclerosis. Endothelial expression of the transcriptional coactivator PGC-1α is induced in rodents and humans with hyperglycaemia, and potently inhibits Notch and AKT/eNOS signalling with deleterious effects on endothelial function [25]. Lastly, hyperglycaemia generates oxidative stress and a pro-inflammatory milieu, which is directly implicated in atherogenesis as detailed next.

Hyperinsulinaemia and Insulin Resistance

Insulin resistance is characterized by impaired intracellular insulin signal transduction and compensatory hyperinsulinaemia. A key feature of this complex perturbation is that while signalling via the PI3K/AKT cascade is impaired, the MAPK/ERK pathway remains unaffected [26]. This concept of 'pathway-specific resistance' is relevant because the compensatory hyperinsulinaemia required to sustain glucose homeostasis via PI3K/AKT leads to excessive MAPK/ERK signalling through a 'bystander' effect of high insulin levels acting on relatively unimpaired pathways leading to ERK activation. Since PI3K/AKT signalling promotes eNOS activation, vascular pathway-specific insulin resistance is characterized by reduced NO bioavailability, alongside increased vascular leukocyte adhesion molecule expression and ET-1 mediated vasoconstriction through MAPK signalling. Indeed, people with diabetes have elevated baseline levels of ET-1, and show improved endothelial function when treated with a pharmacological antagonist [27].

Insulin also plays a vital role in regulation of sympathetic nervous system (SNS) activity by acting on the arcuate nucleus of the hypothalamus. In the context of obesity, SNS hyperactivity arises from chronic hyperinsulinaemia. This promotes vasoconstriction and can secondarily activate the renin–angiotensin–aldosterone system (RAAS), contributing to systemic hypertension, which is an independent risk factor for development of atherosclerosis.

Oxidative Stress

Oxidative stress is intimately involved both in development of individual components of the metabolic syndrome, and directly with the pathophysiology of diabetes-associated macrovascular disease. The term refers to disrupted spatiotemporal balance between production of ROS and their scavenging by antioxidant defences, resulting in detrimental modification of lipids, proteins, and nucleic acids [28]. ROS are highly reactive derivatives of oxygen (O_2), loosely categorized into free radicals, such as superoxide (O_2^-) and peroxynitrite ($ONOO^-$), and non-radicals such as hydrogen peroxide (H_2O_2).

At physiological concentrations, ROS are fundamental to cellular homeostasis by acting primarily as a reversible signalling mechanism [29]. Excessive production or inadequate degradation of ROS results in maladaptive oxidative stress. One important mechanism of vascular oxidative stress in animal models of diabetes is increased NADPH oxidase activity [30]. eNOS uncoupling, for example due to oxidation of the enzyme's cofactor tetrahydrobiopterin (BH4), can also paradoxically result in generation of superoxide by eNOS. Dysfunctional mitochondrial lipid metabolism is an additional important source of ROS in the insulin-resistant vasculature [31]. Oxidative stress associated single-strand DNA breakage and poly(ADP-ribose) polymerase (PARP) activation may also subsequently promote diabetic endothelial dysfunction.

Mechanistic insights into the role of vascular ROS have been derived from *in vivo* manipulation of vascular insulin signalling. Mice with endothelium-targeted overexpression of an inhibitory mutant human insulin receptor (ESMIRO), or with loss of one of the two insulin receptor alleles (IRKO), have enhanced vascular superoxide production. This can be suppressed with a pharmacological Nox-2 inhibitor, and is associated with improved endothelium-mediated vasorelaxation [32]. Consistent with this notion, diabetic blood vessels in hamsters demonstrate improved endothelium-dependent relaxation when treated with superoxide dismutase (SOD), an antioxidant. However, conflicting data has arisen from studying mice with endothelium-specific overexpression of the human insulin receptor (hIRECO), a model of *increased* endothelial insulin signalling, which also exhibited reduced NO bioavailability and superoxide-mediated oxidative stress [33]. Similarly, mice with endothelial inactivation of Shc homology 2-containing inositol 5' phosphatase-2 (SHIP2), a negative regulator of insulin signalling, also exhibit oxidative stress and endothelial dysfunction [34]. Overall, these data may provide some explanation for the disappointing results of clinical trials that have sought to improve cardiovascular outcomes with broad-spectrum antioxidants, in addition to intensive insulin regimens.

Pro-Inflammatory Adipocytokines

Chronic inflammation is an important driver of atherosclerosis [35]. It is closely coupled with the glucotoxicity and lipotoxicity of the metabolic syndrome and is aggravated by oxidative stress. Dysfunctional visceral adipose depots are a particularly important driver of systemic inflammation, due to altered adipocytokine (or adipokine) secretion profiles. Adipokines are a diverse group of cytokines derived from adipose tissue, and include interleukin (IL)-1β, IL-6, tumour necrosis factor (TNF)-α, adiponectin, and leptin, among many others [36]. Their impact extends beyond

adipose tissue biology and wider metabolic homeostasis, with direct effects on vascular function also being apparent. However, our understanding of adipokine biology remains limited. For example, adiponectin is generally less abundant in insulin-resistant states, and has been demonstrated to have anti-inflammatory and antiatherosclerotic properties in mouse models of atherosclerosis [37]. While some data suggest its circulating concentration is inversely predictive of coronary plaque calcification and fibrous cap integrity, contradictory observations have also been made. This illustrates a recurring theme in the adipokine literature, probably reflecting our incomplete understanding of their context dependent effects.

Cell-Autonomous Insulin Resistance

Vascular Endothelial Cells

The vascular endothelium is a monolayer composed of approximately 10^{14} cells. It separates the vessel lumen and VSMCs and hence provides a direct mechanical interface between blood and the tissues it supplies. Beyond its role as a structural barrier, it is also metabolically active and has a critical role in regulating physiological homeostasis. Its functions include transport of nutrients and solutes across the endothelium, maintenance of vascular tone, and regulation of the thrombotic environment.

Endothelial dysfunction refers to a subtle alteration in endothelial phenotype that is implicated in early atherosclerosis, and precedes structural changes in the arterial wall. This unfavourable shift is accelerated by systemic contributors including components of the metabolic syndrome and arises through various mechanisms, perhaps most important of which is a reduction in NO bioavailability. NO has many antiatherogenic and antithrombotic properties that are required in normal physiology. In health there is continuous, basal vascular synthesis of NO which is essential for vascular function, for example promoting vasodilation, inhibiting platelet activation, suppressing inflammation and limiting leukocyte adhesion molecule expression [38]. Indeed, eNOS knockout mice are hypertensive, insulin-resistant, and develop accelerated atherosclerosis.

The macrovascular sequelae of endothelium-specific insulin resistance have been explored in various murine models. Importantly, mice with diet-induced obesity exhibit vascular insulin resistance associated with reduced expression of the insulin receptor [39]. In atherosclerosis-prone mice, genetic knockout of endothelial insulin receptors induces expression of the leukocyte adhesion molecule VCAM-1 and increases atherosclerotic lesion area [40]. ESMIRO mice have been shown to develop endothelial dysfunction, and when crossed with atherosclerosis-prone mice, develop accelerated atherosclerosis. This is independent of systemic metabolic perturbations, and associated with reduced NO bioavailability and superoxide-dependent oxidative stress [41]. As outlined earlier, however, murine models of increased endothelial insulin signalling also paradoxically exhibit enhanced superoxide abundance and oxidative stress-induced endothelial dysfunction. Nevertheless these findings, as well as highlighting the complexity in interplay between metabolic and vascular function, prove the principle that perturbated insulin resistance autonomous to vascular tissues may promote atherogenesis.

Macrophages

There is also evidence that myeloid-specific insulin resistance can accelerate atherogenesis. Macrophages from obese mice have reduced insulin receptor expression, alongside increased expression of surface receptors such as scavenger receptor A (SRA) and CD36 [42]; these are implicated in internalization and cytoplasmic storage of oxidized LDL. Similar changes have been observed in insulin receptor-deficient macrophages. Using murine bone marrow transplantation studies, it has been shown that macrophage insulin receptor deficiency increases endoplasmic reticulum (ER) stress-induced apoptosis and necrotic core formation in advanced atherosclerotic lesions [43]. However, conflicting data also exist indicating that myeloid lineage-specific insulin resistance can protect atherosclerosis-prone mice from the development and progression of atherosclerotic lesions [44]. This may relate to insulin receptor-deficient macrophages having impaired capacity to produce pro-inflammatory cytokines, such as IL-1β and IL-6.

Atherothrombosis

The propensity of atherosclerotic plaques to acute rupture is dependent primarily upon plaque constituents, rather than the extent of vessel stenosis [45]. Factors that enhance vulnerability to rupture are also associated with diabetes, and include thinning of the fibrous cap, increased macrophage infiltration, increased size of the necrotic core, and reduced abundance of VSMCs [46]. The extent of plaque vascularization has also been implicated in instability. In the early phases of atherogenesis, plaque neovascularization may have antiapoptotic effects, while facilitating lesion growth. In more advanced lesions, however, release of cytokines secondary to local inflammation can result in dysregulated neointimal proliferation of microvessels which are leaky, fragile, and of poor quality, rendering them susceptible to rupture. Studies in humans have reported higher proportions of neovessel formation in the most advanced atherosclerotic plaques, with increased abundance in those with diabetes. Expression of proteolytic enzymes, such as MMPs, is also increased by hyperglycaemia, and increases propensity for fibrous cap rupture [47].

Platelets initiate plaque thrombosis upon detection of ECM and core constituents, and their aggregation is augmented in the setting of insulin resistance by reduced NO bioavailability, altered calcium regulation, oxidative stress, and increased circulating von Willebrand factor. Notably, atheromatous plaques in people with diabetes have been shown to exhibit increased intracoronary thrombus compared to controls [48]. Normoglycaemic patients with insulin resistance, and those with type II diabetes, have altered circulating components of the coagulation and fibrinolytic pathways, with increased levels of plasminogen activator inhibitor (PAI)-1 particularly implicated in the pro-thrombotic tendency of the metabolic syndrome [49]. PAI-1 binds to tissue plasminogen activator (tPA) to inhibit its conversion of plasminogen to plasmin, an enzyme that is essential to degrade fibrin clots. Indeed, PAI-1 levels have been shown to be lowered in patients with DM by weight reduction and pharmacotherapies, such as metformin and thiazolidinediones. Hyperglycaemia can also directly cause post-translational modification of coagulation proteins, such as fibrinogen. This results in a

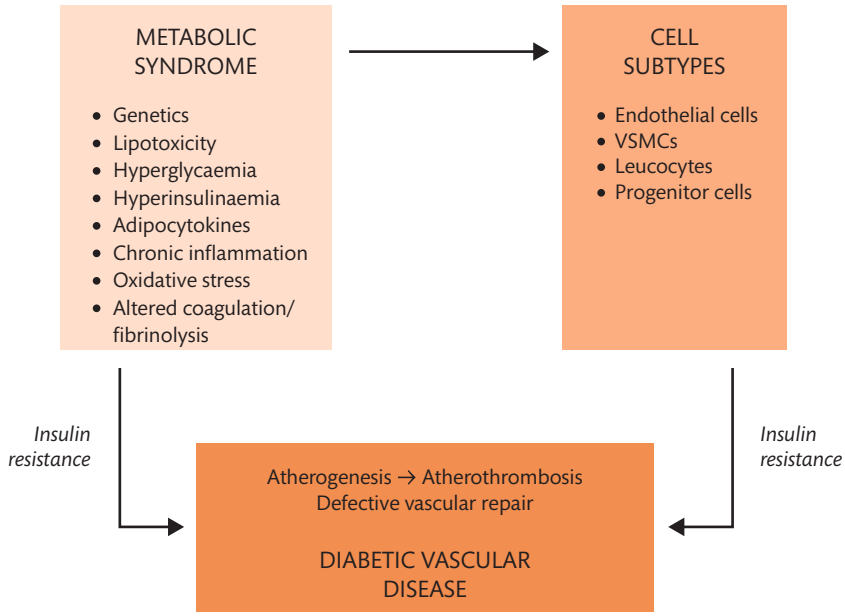


Figure 15.13.1.3 Interplay between systemic and cell-autonomous factors in pathogenesis of diabetic vascular disease.

fibrin clot which is denser, with thinner fibres, and is more resistant to plasmin-mediated lysis [50].

Defects in Vascular Repair

In response to vascular insults, endogenous reparative mechanisms are thought to mitigate injury, and so potentially retard atherogenesis. The term 'vascular repair' is typically used to describe re-endothelialization of existing conduit vessels that have been exposed to injurious insults, and requires complex interaction of local and circulating cells [51] (Figure 15.13.1.3). The endothelium is a critical player in this process, with proliferation and migration of local endothelial cells resulting in restoration of a seamless endothelial monolayer. While bone marrow-derived cells may promote this process via paracrine mechanisms, they do not directly contribute to the neo-endothelium.

There is emerging evidence to support the notion that an insulin-resistant phenotype has adverse effects on indices of endogenous vascular repair. IRKO mice demonstrate broadly preserved metabolic sensitivity to insulin, but markedly impaired re-endothelization of the femoral artery after denudation [52]. *In vivo* models of type 2 diabetes have an exaggerated neointimal response to vascular injury, which is retarded with rosiglitazone, an agent that enhances insulin sensitivity [53]. Myeloid angiogenic cells derived from patients with diabetes have impaired paracrine capacity to promote vascular repair, which correlates with increased superoxide abundance [54]. Outgrowth endothelial cells (OEC) derived from South Asian men, who are insulin resistant and have increased risk of cardiovascular sequelae, also exhibit impaired vascular repair capacity due to impaired AKT signalling [55].

eNOS expression also appears reduced in diabetic vascular progenitors, and this is associated with impaired migration and adhesion capacity [56]. Reduced endothelial NO bioavailability in the context of insulin-resistant states may be pertinent, promoting cell senescence and impairing migration in response to chemotactic stimuli, both factors likely to be relevant in promoting conduit repair [57]. Indeed, NO has more potent inhibitory effects on neointimal hyperplasia after arterial injury in rats with metabolic syndrome and frank diabetes, in comparison to normoglycaemic controls.

Conclusions

Despite contemporary pharmacotherapies and lifestyle modification, the presence of diabetes continues to confer a prognostic disadvantage with regards to atherosclerosis-mediated major cardiovascular events. Significant progress has been made in our understanding of insulin resistance and diabetes-associated vascular disease; however, the complexity of this association means that many mechanistic and translational questions remain. Addressing these uncertainties will be pivotal in the development of therapeutic strategies that safely and effectively promote macrovascular health, survival, and quality of life in patients across the spectrum of dysglycaemia.

REFERENCES

1. Norhammar A, Tenerz A, Nilsson G, *et al*. Glucose metabolism in patients with acute myocardial infarction and no previous diagnosis of diabetes mellitus: a prospective study. *Lancet*, 2002; **359**: 2140–4.
2. Booth GL, Kapral MK, Fung K, Tu JV. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *Lancet*, 2006; **368**: 29–36.
3. Elezi S, Kastrati A, Pache J, *et al*. Diabetes mellitus and the clinical and angiographic outcome after coronary stent placement. *J Am Coll Cardiol*, 1998; **32**: 1866–73.
4. Reaven GM. Role of insulin resistance in human disease (syndrome X): an expanded definition. *Annu Rev Med*, 1993; **44**: 121–31.

5. Ross R. Atherosclerosis—an inflammatory disease. *N Engl J Med*, 1999; **340**: 115–26.
6. Aronson D, Edelman ER. Revascularization for coronary artery disease in diabetes mellitus: angioplasty, stents and coronary artery bypass grafting. *Rev Endocr Metab Disord,* 2010; **11**: 75–86.
7. Napoli C, D'Armiento FP, Mancini FP, *et al*. Fatty streak formation occurs in human fetal aortas and is greatly enhanced by maternal hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atherosclerotic lesions. *J Clin Invest*, 1997; **100**: 2680–90.
8. Rader DJ, Pure E. Lipoproteins, macrophage function, and atherosclerosis: beyond the foam cell? *Cell Metab*, 2005; **1**: 223–30.
9. Boisvert WA, Santiago R, Curtiss LK, Terkeltaub RA. A leukocyte homologue of the IL-8 receptor CXCR-2 mediates the accumulation of macrophages in atherosclerotic lesions of LDL receptor-deficient mice. *J Clin Invest*, 1998; **101**: 353–63.
10. Doran AC, Meller N, McNamara CA. Role of smooth muscle cells in the initiation and early progression of atherosclerosis. *Arterioscler Thromb Vasc Biol*, 2008; **28**: 812–19.
11. Lee RT, Libby P. The unstable atheroma. *Arterioscler Thromb Vasc Biol*, 1997; **17**: 1859–67.
12. Grundy SM, Brewer HB Jr, Cleeman JI, *et al*. Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*, 2004; **109**: 433–8.
13. Becker AB, Roth RA. Insulin receptor structure and function in normal and pathological conditions. *Annu Rev Med*, 1990; **41**: 99–115.
14. Vinciguerra M, Foti M. PTEN and SHIP2 phosphoinositide phosphatases as negative regulators of insulin signalling. *Arch Physiol Biochem*, 2006; **112**: 89–104.
15. DiNitto JP, Cronin TC, Lambright DG. Membrane recognition and targeting by lipid-binding domains. *Sci STKE*, 2003; **2003**: re16.
16. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol*, 2006; **7**: 85–96.
17. Dimmeler S, Fleming I, Fisslthaler B, Hermann C, Busse R, Zeiher AM. Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation. *Nature*, 1999; **399**: 601–5.
18. Kido Y, Nakae J, Accili D. Clinical review 125: the insulin receptor and its cellular targets. *J Clin Endocrinol Metab*, 2001; **86**: 972–9.
19. Sniderman AD, Scantlebury T, Cianflone K. Hypertriglyceridemic hyperapob: the unappreciated atherogenic dyslipoproteinemia in type 2 diabetes mellitus. *Ann Intern Med*, 2001; **135**: 447–59.
20. Furukawa S, Fujita T, Shimabukuro M, *et al*. Increased oxidative stress in obesity and its impact on metabolic syndrome. *J Clin Invest*, 2004; **114**: 1752–61.
21. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. *Nature*, 2001; **414**: 813–20.
22. Haffner SM. Epidemiological studies on the effects of hyperglycemia and improvement of glycemic control on macrovascular events in type 2 diabetes. *Diabetes Care*, 1999; **22** Suppl 3: C54–6.
23. Heath JM, Sun Y, Yuan K, *et al*. Activation of AKT by O-linked N-acetylglucosamine induces vascular calcification in diabetes mellitus. *Circ Res*, 2014; **114**: 1094–102.
24. Fan HC, Fernandez-Hernando C, Lai JH. Protein kinase C isoforms in atherosclerosis: pro- or anti-inflammatory? *Biochem Pharmacol*, 2014; **88**: 139–49.
25. Sawada N, Jiang A, Takizawa F, *et al*. Endothelial PGC-1alpha mediates vascular dysfunction in diabetes. *Cell Metab*, 2014; **19**: 246–58.
26. Jiang ZY, Lin YW, Clemont A, *et al*. Characterization of selective resistance to insulin signaling in the vasculature of obese Zucker (fa/fa) rats. *J Clin Invest*, 1999; **104**: 447–57.
27. Mather KJ, Mirzamohammadi B, Lteif A, Steinberg HO, Baron AD. Endothelin contributes to basal vascular tone and endothelial dysfunction in human obesity and type 2 diabetes. *Diabetes*, 2002; **51**: 3517–23.
28. Betteridge DJ. What is oxidative stress? *Metabolism*, 2000; **49**(2 Suppl 1): 3–8.
29. Taniyama Y, Griendling KK. Reactive oxygen species in the vasculature: molecular and cellular mechanisms. *Hypertension*, 2003; **42**: 1075–81.
30. Konior A, Schramm A, Czesnikiewicz-Guzik M, Guzik TJ. NADPH oxidases in vascular pathology. *Antioxid Redox Signal*, 2014; **20**: 2794–814.
31. Du X, Edelstein D, Obici S, Higham N, Zou MH, Brownlee M. Insulin resistance reduces arterial prostacyclin synthase and eNOS activities by increasing endothelial fatty acid oxidation. *J Clin Invest*, 2006; **116**: 1071–80.
32. Sukumar P, Viswambharan H, Imrie H, *et al*. Nox2 NADPH oxidase has a critical role in insulin resistance-related endothelial cell dysfunction. *Diabetes*, 2013; **62**: 2130–4.
33. Viswambharan H, Yuldasheva NY, Sengupta A, *et al*. Selective enhancement of insulin sensitivity in the endothelium *in vivo* reveals a novel proatherosclerotic signaling loop. *Circ Res*, 2017; **120**: 784–98.
34. Watt NT, Gage MC, Patel PA, *et al*. Endothelial SHIP2 suppresses Nox2 NADPH oxidase-dependent vascular oxidative stress, endothelial dysfunction, and systemic insulin resistance. *Diabetes*, 2017; **66**: 2808–21.
35. Ridker PM, Everett BM, Thuren T, *et al*. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med*, 2017; **377**: 1119–31.
36. Kamata H, Honda S, Maeda S, Chang L, Hirata H, Karin M. Reactive oxygen species promote TNFalpha-induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell*, 2005; **120**: 649–61.
37. Okamoto Y, Kihara S, Ouchi N, *et al*. Adiponectin reduces atherosclerosis in apolipoprotein E-deficient mice. *Circulation*, 2002; **106**: 2767–70.
38. Napoli C, de Nigris F, Williams-Ignarro S, Pignalosa O, Sica V, Ignarro LJ. Nitric oxide and atherosclerosis: an update. *Nitric Oxide*, 2006; **15**: 265–79.
39. Imrie H, Abbas A, Viswambharan H, *et al*. Vascular insulin-like growth factor-I resistance and diet-induced obesity. *Endocrinology*, 2009; **150**: 4575–82.
40. Rask-Madsen C, Li Q, Freund B, *et al*. Loss of insulin signaling in vascular endothelial cells accelerates atherosclerosis in apolipoprotein E null mice. *Cell Metab*, 2010; **11**: 379–89.
41. Gage MC, Yuldasheva NY, Viswambharan H, *et al*. Endothelium-specific insulin resistance leads to accelerated atherosclerosis in areas with disturbed flow patterns: a role for reactive oxygen species. *Atherosclerosis*, 2013; **230**: 131–9.
42. Liang CP, Han S, Okamoto H, *et al*. Increased CD36 protein as a response to defective insulin signaling in macrophages. *J Clin Invest*, 2004; **113**: 764–73.
43. Han S, Liang CP, DeVries-Seimon T, *et al*. Macrophage insulin receptor deficiency increases ER stress-induced apoptosis and necrotic core formation in advanced atherosclerotic lesions. *Cell Metab*, 2006; **3**: 257–66.
44. Baumgartl J, Baudler S, Scherner M, *et al*. Myeloid lineage cell-restricted insulin resistance protects apolipoproteinE-deficient mice against atherosclerosis. *Cell Metab*, 2006; **3**: 247–56.
45. Lusis AJ. Atherosclerosis. *Nature*, 2000; **407**: 233–41.

46. Bornfeldt KE. 2013 Russell Ross memorial lecture in vascular biology: cellular and molecular mechanisms of diabetes mellitus-accelerated atherosclerosis. *Arterioscler Thromb Vasc Biol*, 2014; **34**: 705–14.
47. Uemura S, Matsushita H, Li W, Glassford AJ, *et al.* Diabetes mellitus enhances vascular matrix metalloproteinase activity: role of oxidative stress. *Circ Res*, 2001; **88**: 1291–8.
48. Kohler HP, Grant PJ. Plasminogen-activator inhibitor type 1 and coronary artery disease. *N Engl J Med*, 2000; **342**: 1792–801.
49. Juhan-Vague I, Roul C, Alessi MC, Ardissone JP, Heim M, Vague P. Increased plasminogen activator inhibitor activity in non insulin dependent diabetic patients—relationship with plasma insulin. *Thromb Haemost*, 1989; **61**: 370–3.
50. Ajjan RA, Gamlen T, Standeven KF, *et al.* Diabetes is associated with posttranslational modifications in plasminogen resulting in reduced plasmin generation and enzyme-specific activity. *Blood*, 2013; **122**: 134–42.
51. Cubbon RM, Mercer BN, Sengupta A, Kearney MT. Importance of insulin resistance to vascular repair and regeneration. *Free Radic Biol Med*, 2013; **60**: 246–63.
52. Wheatcroft SB, Shah AM, Li JM, *et al.* Preserved glucoregulation but attenuation of the vascular actions of insulin in mice heterozygous for knockout of the insulin receptor. *Diabetes*, 2004; **53**: 2645–52.
53. Phillips JW, Barringhaus KG, Sanders JM, *et al.* Rosiglitazone reduces the accelerated neointima formation after arterial injury in a mouse injury model of type 2 diabetes. *Circulation*, 2003; **108**: 1994–9.
54. Sorrentino SA, Bahlmann FH, Besler C, *et al.* Oxidant stress impairs *in vivo* reendothelialization capacity of endothelial progenitor cells from patients with type 2 diabetes mellitus: restoration by the peroxisome proliferator-activated receptor-gamma agonist rosiglitazone. *Circulation*, 2007; **116**: 163–73.
55. Cubbon RM, Yuldasheva NY, Viswambharan H, *et al.* Restoring Akt1 activity in outgrowth endothelial cells from South Asian men rescues vascular reparative potential. *Stem Cells*, 2014; **32**: 2714–23.
56. Ii M, Takenaka H, Asai J, *et al.* Endothelial progenitor thrombospondin-1 mediates diabetes-induced delay in reendothelialization following arterial injury. *Circ Res*, 2006; **98**: 697–704.
57. Hayashi T, Yano K, Matsui-Hirai H, Yokoo H, Hattori Y, Iguchi A. Nitric oxide and endothelial cellular senescence. *Pharmacol Ther*, 2008; **120**: 333–9.

15.13.2 Macrovascular Disease in Type 2 Diabetes

Naveed Sattar

Cardiovascular Risk in Type 2 Diabetes

While type 2 diabetes was once considered a coronary heart disease (CHD) risk equivalent, this notion has been superseded. The most comprehensive evidence comes from the emerging risk factor collaboration (ERFC), which produced a collaborative meta-analysis of 102 prospective studies using individual participant data [1]. This showed that prevalent diabetes confers on average a twofold excess risk for a wide range of vascular outcomes, independently of other risk factors (**Figure 15.13.2.1**). This work also showed that diabetes is around a third more strongly related to risk of fatal events than of non-fatal myocardial infarction (MI). Further analyses showed the excess relative risk of vascular events in people with versus without diabetes was greater in women and at younger ages. This fits well with findings in national datasets such as the QRISK2 study from the United Kingdom which reported type 2 diabetes to be associated with CVD risk levels of 2.20 and 2.54 in men and women, respectively, independent of other measured risk factors [2].

Lesser Degrees of Dysglycaemia and Cardiovascular Risk

In those without diabetes, ERFC analyses showed no clear association with CHD risk at fasting blood glucose (FBG) levels between 3.9 and 5.6 mmol/L, but a modest rise in risk thereafter [1]. However, when the hazard ratios for CHD were examined by clinically defined categories, they were only meaningfully elevated with known or undiagnosed diabetes, being modest at FBG levels below 7 mmol/L (**Figure 15.13.2.2**). By contrast, low-density lipoprotein (LDL) cholesterol and blood pressure had stronger and nearly log-linear associations with vascular risk. A subsequent meta-analysis of individual participant data for three different glycaemia measures (HbA1c, random or FBG and post load glucose) from multiple cohorts determined that in those without type 2 diabetes although the HbA1c association with CVD risk was equal to or greater than that seen for fasting, random, or post load glucose levels, no glycaemia measure meaningfully improved CVD risk prediction in this non-diabetic group [3]. Near identical findings were recently reported for a 11.5-year follow-up from the Whitehall II study [4], where CVD risk associations for all glycaemia measures in the pre-diabetes range turned out to be modest. Collectively, these results suggest that, while diabetes is meaningfully and independently related to CVD risk, lesser levels of dysglycaemia are not. In clinical terms, CVD risks in patients with pre-diabetes should be determined using usual CVD risk scores without any multiplication factor.

Factors Associated with Differential CVD Risk in Diabetes

Notably, a doubling of CVD risk represents only the average risk in *all* patients with type 2 diabetes, and wide heterogeneity of risk is seen, driven by several patient characteristics. For example, at the

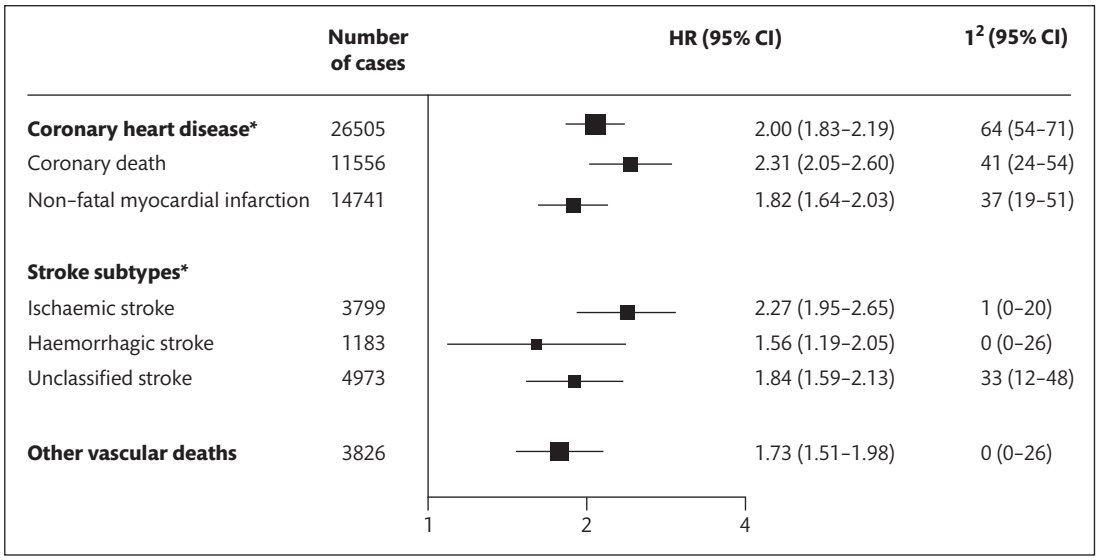


Figure 15.13.2.1 Hazard ratios (HRs) for vascular outcomes in people with versus those without diabetes at baseline. Analyses were based on 530 083 participants. HRs were adjusted for age, smoking status, body mass index, and systolic blood pressure, and, where appropriate, stratified by sex and trial arm. 208 coronary heart disease outcomes that contributed to the grand total could not contribute to the subtotals of coronary death or non-fatal myocardial infarction because there were fewer than 11 cases of these coronary disease subtypes in some studies. *Includes both fatal and non-fatal events.

point of diabetes diagnosis relative risk is well below 2, whereas once a patient has had several years of diabetes or exhibits microvascular complication, such as albuminuria, low glomerular filtration rate (eGFR), or moderate retinopathy, CVD risk levels are elevated many fold. The relevance of these and other risk factors to excess risk of CVD in type 2 diabetes is discussed in more detail next.

Age of Onset

Age of onset of type 2 diabetes is an important determinant of risk of adverse outcomes as well as years lost due to diabetes. Several lines of evidence support this notion. These include the finding from the Swedish diabetes registry of much higher hazard ratios (~2–3 fold) for CVD and total mortality in diabetes patients under 55, whereas hazard ratios (HRs) were around 1–1.2 fold in patients with diabetes above 75 years of age (Figure 15.13.2.3) [5]. Similarly, recent data from the Swedish diabetes registry showed that life-years lost due to type 2 diabetes were more than a decade when it was diagnosed in adolescence, to around 6 years lost when diagnosed in the 40s, and tapering to no loss of life-years when diagnosed above 80 years or so (Figure 15.13.2.4) [6]. More recently, using only two age groups, a Chinese study suggested a near doubling of risk of non-fatal cardiovascular in those with early (<40 years) versus later onset type 2 diabetes [7], while

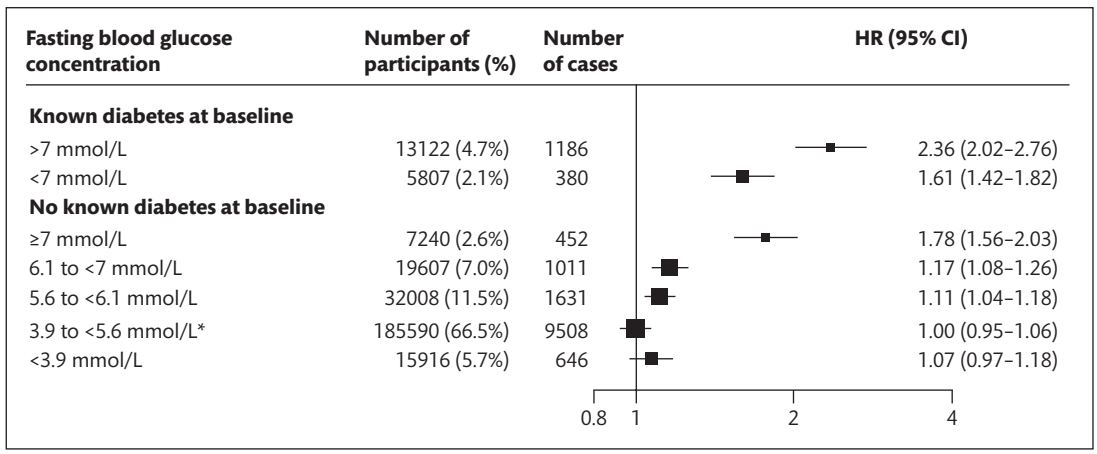


Figure 15.13.2.2 Hazard ratios (HRs) for coronary heart disease by clinically defined categories of baseline fasting blood glucose concentration. Analyses were based on 279 290 participants (14 814 cases). HRs were adjusted as described in Figure 15.13.2.1. HR (95% CI) in people with fasting glucose 5.60–6.99 mmol/L was 1.12 (1.06–1.18). *Reference group.

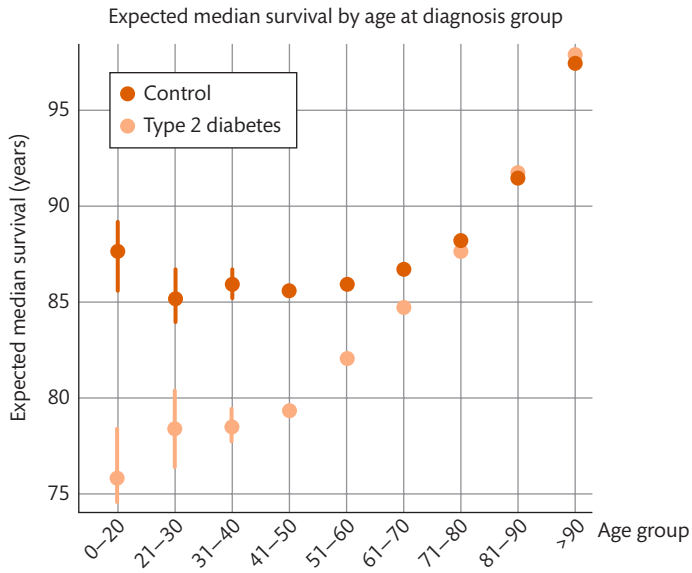


Figure 15.13.2.3 Age at diagnosis of type 2 diabetes mellitus and loss of life-years in persons without previous cardiovascular disease and without any restriction on the duration of type 2 diabetes mellitus.

work from Australia showed higher mortality risk, in particular CVD mortality, in younger onset diabetes patients in a study of just under three-quarters of a million Australians with diabetes [8]. Collectively, these studies support the notion that type 2 diabetes is a more potent driver of CVD risk when it develops at younger ages.

Other research has provided important mechanistic insights into reasons for higher CVD risk at younger age of diabetes onset. In an analysis of just over hundred thousand newly diagnosed patients, individuals with younger onset type 2 diabetes were much more obese and had worse glycaemia and lipid (LDL cholesterol, triglyceride, and high-density lipoprotein (HDL) cholesterol) levels than those developing type 2 diabetes much later in life (Figure 15.13.2.5a–e) [9]. Such differences persisted for several years' post diagnosis and could not be accounted for differential glycaemia or lipid-lowering treatments. There was also evidence from this latter study plus other work from Scotland

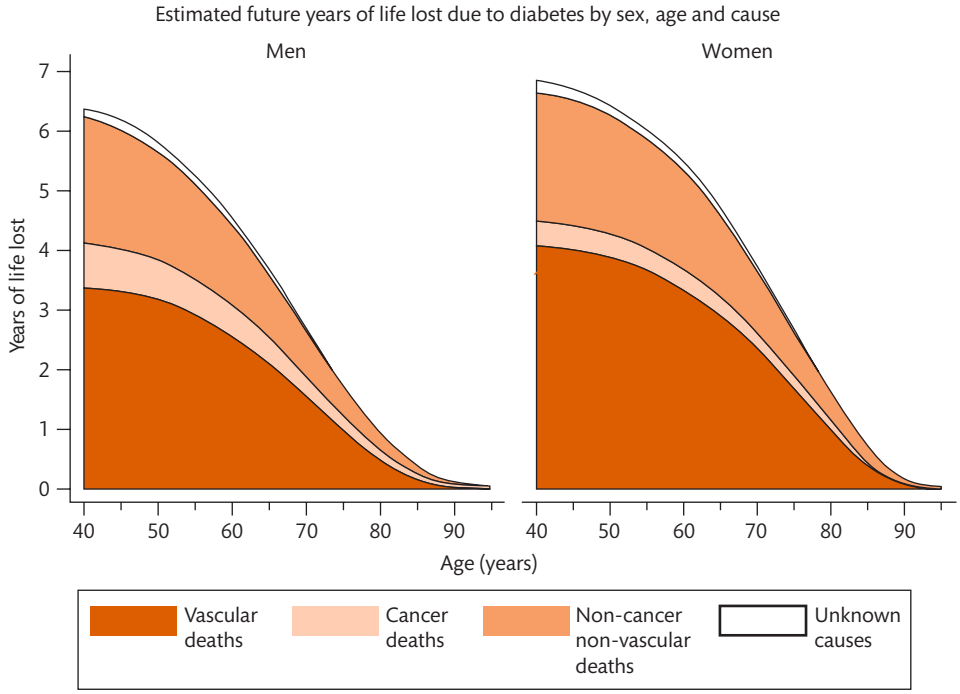


Figure 15.13.2.4 Estimated future years of life lost due to diabetes by sex, age, and cause.

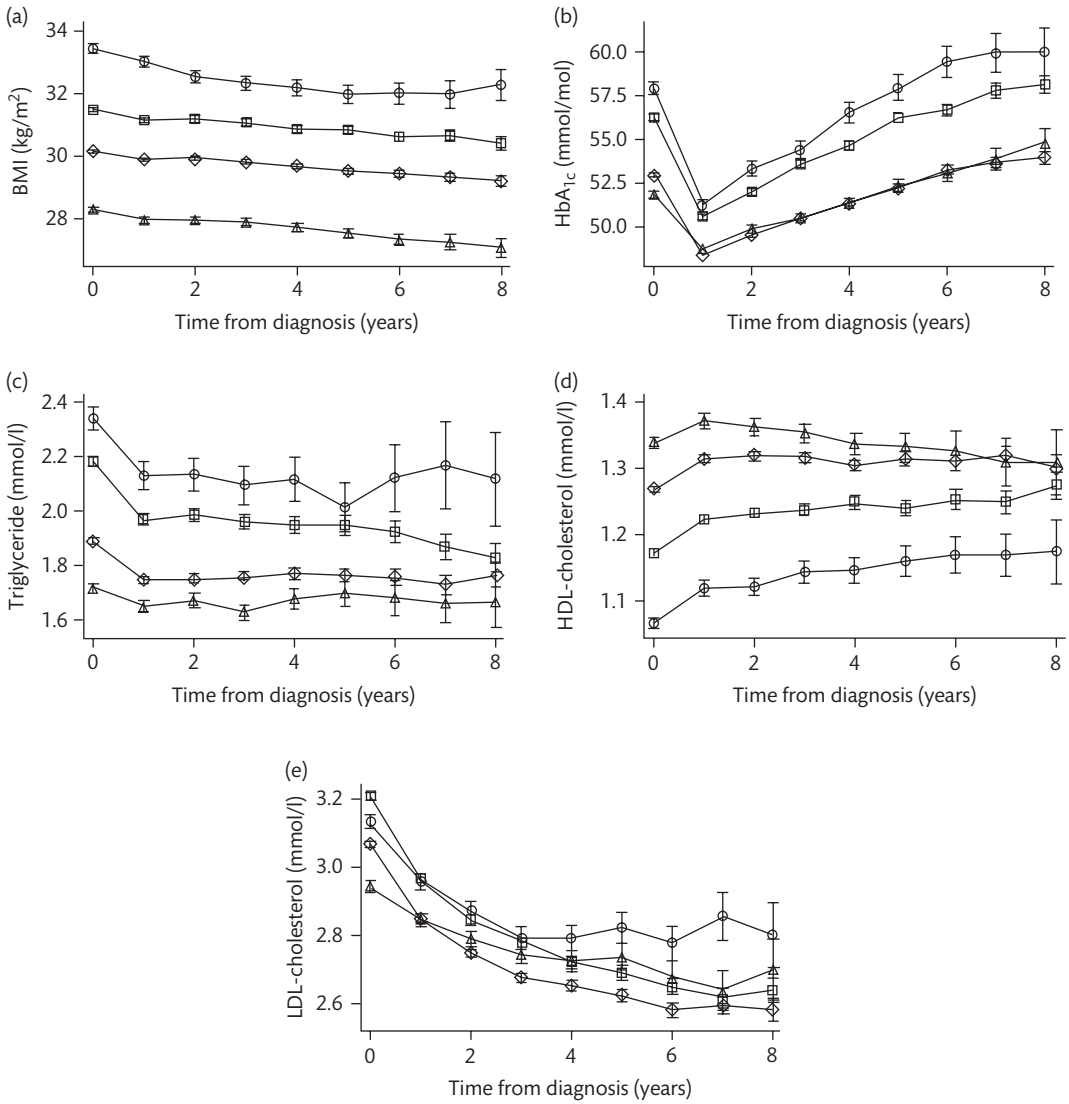


Figure 15.13.2.5 The trajectories of cardiometabolic risk factors over time by age at type 2 diabetes diagnosis. Yearly averages (95% CI) stratified by age group for (a) BMI, (b) HbA1c, (c) triglyceride, (d) HDL cholesterol and (e) LDL cholesterol. White circles, 18–44 years old; plus sign (+), 45–59 years old; crosses (×), 60–74 years old; white triangles, ≥75 years old. All analyses, $p < 0.05$ where 95% CI do not overlap.

for faster glycaemic deterioration in younger type 2 diabetes patients, with HbA1c changes being more modest in those diagnosed with diabetes when older [10]. Indeed, in the latter study, it was noted that 'The rate of deterioration in those diagnosed at over 70 years of age was very low, with 66% having a rate of deterioration of less than 1.1 mmol/mol HbA1c per year, and only 1.5% progressing more rapidly than 4.4 mmol/mol HbA1c per year'. Collectively, these epidemiological findings support age at type 2 diabetes diagnosis as an important stratifier for future CVD and mortality risks. They also suggest that gains from preventative therapies in terms of life-years free from vascular disease will likely be greatest in younger onset type 2 diabetes patients. This is due to the fact that gain in life-years will be greater when treating younger people at elevated risk, than when treating older people who are on average at lower excess risk relative to people without diabetes [11].

Duration

Duration of diabetes is linked to greater CVD risks. In analysis of data from the British Regional Heart Study [12], it was shown that men without prior MI and with diabetes of mean duration 16.7 years had subsequent CHD risks approaching those with prior MI but without diabetes, whereas those with short mean diabetes duration (mean 1.9 years) had much lower risks. One caveat in this type of analysis is that those with longer duration also have earlier onset disease so it is difficult to tease out to what extent duration explains risks with earlier onset disease. Regardless, these findings are understandable if frank hyperglycaemia is causally related to excess CVD risk in a manner consistent with cumulative exposure.

Presence of Microvascular Disease

In a recent report using the UK Clinical Practice Research Datalink [13], it was elegantly demonstrated that the presence of microvascular

complications, whether retinopathy, nephropathy, or neuropathy, signals excess vascular risk perhaps better than differences in usual risk factors (Figure 15.13.2.6). Furthermore, this work demonstrated that if two or three such complications were present then CVD risks were even higher (Figure 15.13.2.6) [13]. This latter observation accords with the notion that small vessel disease begets large vessel disease, perhaps directly by provoking vasculopathy [14]. Of note, several studies have shown that both low eGFR and evidence of proteinuria are very strong prognostic indicators of future CVD risks in diabetes. In a systematic review of event rates from diabetes trials, the presence of proteinuria seemed to be associated with future outcome risks to a similar extent to those with prevalent cardiovascular disease [15].

Of course, microvascular disease is also linked to longer diabetes duration and to cumulative hyperglycaemia exposure as well as other risk factors, in particular hypertension, so that several risk factors can accelerate both small and large vessel disease.

Ethnicity

It is well established that several ethnicities—notably South Asians, Chinese, and Afro-Caribbeans—are at particularly elevated risk for type 2 diabetes even when body mass index (BMI) is taken into account. The reasons for these excess risks are not fully established but could include, among other factors, altered body composition. In terms of cardiovascular risk, historical evidence suggests South Asians with diabetes have even greater risks of premature macrovascular disease, as previously reviewed [16]. More recent evidence suggests such risks may be coming down for immigrant South Asian populations [17], perhaps in part due to earlier diagnosis of diabetes and so earlier and more prolonged exposure to cardiovascular risk factor management. This suggestion remains speculative and more research is needed in this area (see Chapter 15.7.3, 'Type 2 Diabetes in Different Ethnic Groups').

Sex

Plentiful evidence suggests that relative risk of CVD increases more as women transit from non-diabetes to diabetes than as do men, with a recent meta-analysis suggesting that after controlling for major vascular risk factors, diabetes roughly doubles occlusive vascular mortality risk among men while tripling risk among women [18]. However, while relative risks are often greater in women with diabetes, their absolute risks appear similar. In other words, the presence of diabetes appears to narrow or abolish the gap in absolute vascular risks between men and women [18]. Recent work has suggested that the greater excess vascular risk imposed by diabetes in women may be due to women requiring greater weight gain to develop type 2 diabetes than men, since women start from a point of lower risk [19]. This latter observation may explain why risk factor changes (in particular blood pressure) are on average greater in women who develop diabetes than their male counterparts [20]. It also explains why there exists a higher prevalence of diabetes in middle aged men compared to women, a finding which seems true for different ethnicities [21].

Existing Vascular Disease

When diabetes occurs in the context of established vascular disease, then risks for premature mortality are much higher. In an individual participant meta-analysis the presence of diabetes together with premature MI or stroke at the age of 40 years led to a near doubling of life-years lost compared to diabetes alone, whereas when all three conditions coexisted, over 20 years of life were lost (Figure 15.13.2.7) [22]. At older ages, fewer years were lost but even at 65 years of age, having diabetes plus MI or stroke led to loss of nearly 10 years of life compared to 5 years with diabetes alone. These considerations, combined with emerging evidence for significant cardiovascular and total mortality benefits for some newer diabetes drugs [23], mean that it is important to check for occult diabetes in patients with existing vascular disease. Determining HbA1c, which can easily be done at any time of the day or night and is also informative when tested during acute illness, should substantially increase opportunities for diabetes diagnosis in those with existing vascular disease. Relevant guidelines have ratified use of HbA1c in such scenarios [24].

Changes in Cardiovascular and Mortality Risks Over Time in Type 2 Diabetes Relative to the General Population

In high income countries, such as the United States, a pattern of substantially reducing risks for MI and stroke in diabetes patients

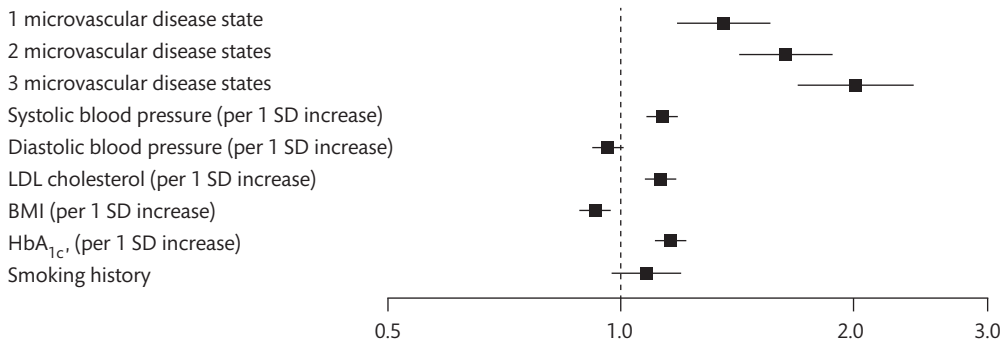


Figure 15.13.2.6 Adjusted hazard ratio for the primary outcome by cumulative burden of microvascular disease and per 1 SD difference in values for established risk factors.* The primary outcome measure was cardiovascular mortality, non-fatal myocardial infarction, or non-fatal ischaemic stroke. 1 SD of each established risk factor is: blood pressure 13.5/8.4 mmHg; LDL cholesterol 0.9 mmol/L; BMI 6.3 kg/m²; HbA1c 1.3% (14.2 mmol/mol).

* Adjusted for age, sex, systolic blood pressure, diastolic blood pressure, LDL cholesterol, HDL cholesterol, HbA1c, BMI, duration of diabetes, smoking status, antiplatelet therapy, lipid-lowering treatment, RAS blockade, other blood pressure treatment, ethnicity, and Index of Multiple Deprivation.

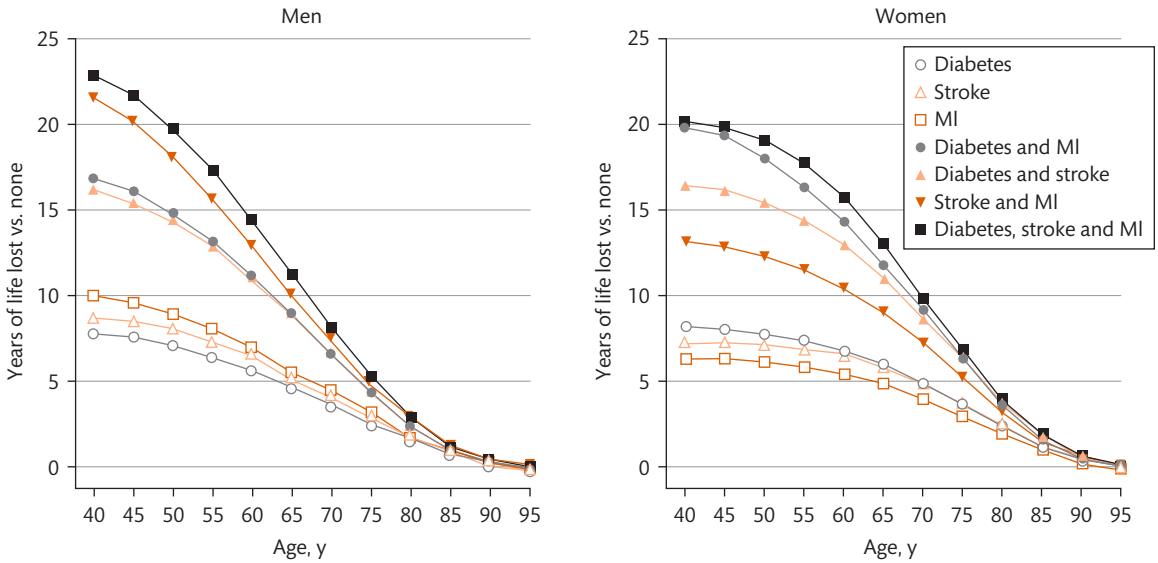


Figure 15.13.2.7 Life-years lost from diabetes, MI, or stroke alone, or in various combinations.

Reproduced with permission from Di Angelantonio E, Kaptoge S, Wormser D *et al.* Association of cardiometabolic multimorbidity with mortality. *JAMA—J Am Med Assoc* 2015;314:52–60.

has been noted over the last two or so decades [25], though such risks still remain far in excess of healthy controls (Figure 15.13.2.8a) [26]. Likewise, in recent longitudinal Swedish data covering the period between 1998 and 2013, type 2 diabetes patients experienced greater (44% vs. 29%) relative risk reductions for CVD hospitalizations than did their non-diabetes counterparts (Figure 15.13.2.8b) [27]. However, reduction over time in CVD deaths appeared to be similar so that there was no obvious narrowing in risks in diabetes. Consequently, recent attention has focused more on the excess risk for other 'vascular' outcomes in diabetes such as heart failure and peripheral arterial disease (Figure 15.13.2.9). In a recent analysis of data from the United Kingdom, these latter two outcomes were the most common first 'vascular' outcomes in diabetes patients, with MI and stroke [28] being less frequent. This greater success at preventing non-fatal CVD events in diabetes [27] suggests an incomplete understanding of the causes of excess CVD mortality in type 2 diabetes, an area which requires further research. That noted, as vascular deaths have declined in both individuals with diabetes and in those without known diabetes, the percentage of total deaths in the United States due to vascular causes has declined from around 48% to 34% in diabetes, and 45% to 31% in non-diabetes from 1988 to 2015 [29]. The percentage of deaths due to cancer was stable in both groups so that proportionately more deaths are now due to non-vascular and non-cancer causes [29]. With further dissection of these data by age, it was notable that only people with diabetes of younger onset (<44 years of age) did not experience reductions in mortality, suggesting a need for more particularly assertive risk factor management in this group.

Heart Failure Risks in Type 2 Diabetes

The excess risk for heart failure in diabetes has been known for many years. While this could be due to excess myocardial infarctions, the presence of a specific diabetic cardiomyopathy, linked in part to the altered metabolic milieu in diabetes, has been proposed for many years although its presence remains strongly debated. Recent research has shown that while heart failure risks are elevated in type 2 diabetes at all ages, relative risks are greatest in those with diabetes of earlier onset [30]. Overall, risks for heart failure are around double those in people without diabetes [31]. More recent research suggests a more prominent role for renal disease and obesity in heart failure pathogenesis in diabetes, whereas the usual risk factors—glycaemia, lipids, blood pressure—appear more important for incident myocardial infarction [32]. This finding, together with unexpected results from the EMPAREG outcome trial, where a substantial reduction in mortality and heart failure hospitalizations with treatment with an SGLT2 inhibitor occurred [33], suggests that pathways to heart failure in type 2 diabetes are more complex than previously thought. Perturbation of circulating volume may play a prominent role [34]. Recent thought-provoking reviews [35, 36] have pointed out a need for more attention to be directed at heart failure in future diabetes clinical trials.

Evidence for Better CVD Risk Factor Management in Diabetes

The aforementioned reductions in CVD risk in people with diabetes (greater relative reductions than in general population) and vascular mortality (reductions about the same as in general population) have occurred due to improvements in risk factors. There have been large reductions in cholesterol levels from around the year 2000 in all high income countries as a result of wide prescription of statins, driven by the view enshrined in many guidelines that diabetes confers high CVD risk. For example, Ford and colleagues showed average cholesterol levels in diabetes patients

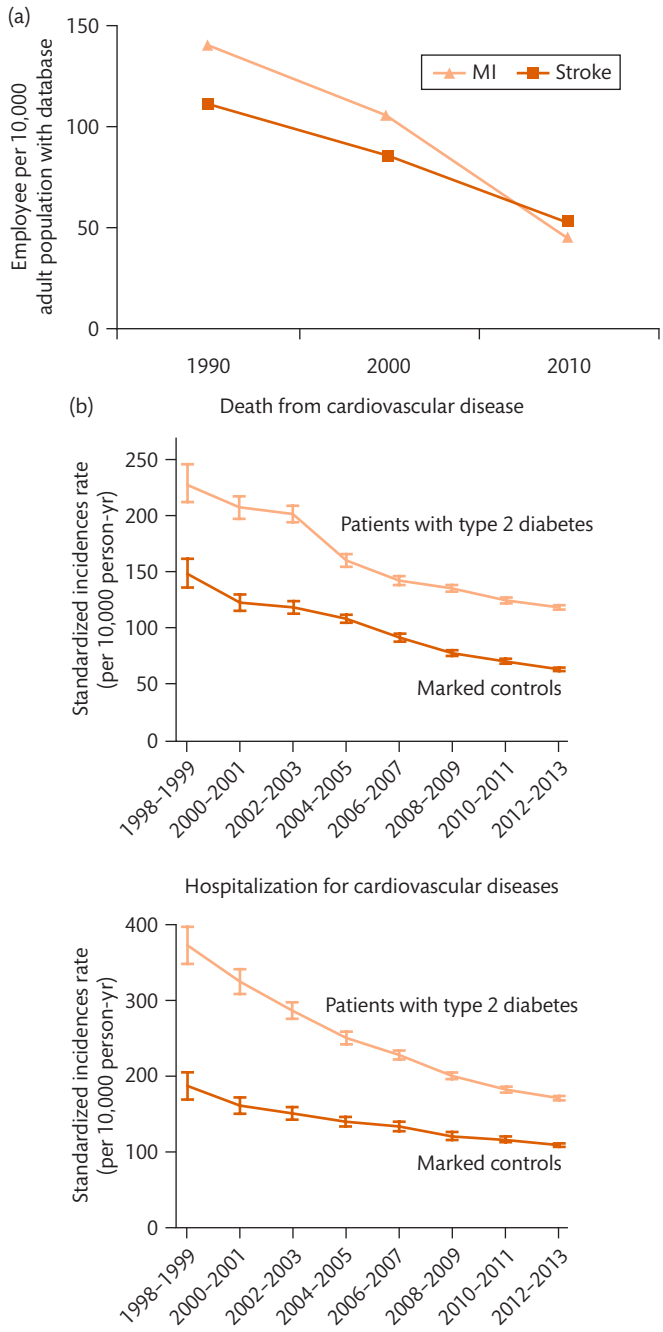


Figure 15.13.2.8 Changes in MI, stroke in United States in diabetes over 1990–2010 and in CVD deaths and CVD hospitalizations in Sweden in diabetes and age and sex-matched controls from 1998 to 2013.

(a) Reproduced with permission from Gregg EW, Cheng YJ, Srinivasan M *et al.* Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet (London, England)* 2018;391:2430–40. (b) Reproduced with permission from Rawshani A, Rawshani A, Franzén S *et al.* Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. *N Engl J Med* 2017;376:1407–18. Copyright © 2017, Massachusetts Medical Society.

declined from around 5.3 mmol/L in 1999/2000 to 4.6 mmol/L by 2007/8; over the same period smoking rates fell from 20% to 15% and average HbA1c from 7.7% to 7.3% [37]. Large improvements in blood pressure stemmed from 1990s onwards, and the call to even lower targets (to <130 mmHg in many), should continue to push average levels lower whereas glycaemia levels have improved to lesser extents and mostly in the last two decades [26]. Population level reductions in smoking have also contributed to such CVD trends as have earlier diagnosis of diabetes. The relative contribution of such risk factor changes to reductions in CVD outcomes in diabetes have not been formerly calculated but greater short-term outcome benefits have been seen in trials from intensive management of lipids and blood pressure than with intensive glycaemia management [38]. That noted, newer diabetes therapies with proven CVD benefits may start to influence this pattern, even if they may lessen risks partially or largely independent of glucose lowering. Of course, the patterns of risk factor treatments are far from optimal in low and middle income countries where limited resources mean that many patients go without oral antidiabetes drugs, statins, or antihypertensive drugs. This lack of resource explains the persisting very high mortality rates in diabetes in such countries. To reduce CVD risk in diabetes in low- and middle-income countries (LMIC) therefore requires a sustainable supply of cheap statins, antihypertensives and metformin/sulfonylureas. This is by no means an easy task and, as recently reviewed for South Asian countries, many difficult challenges (costs, infrastructure, care processes) need to be overcome/improved to make this happen [39].

Summary

CVD risk in type 2 diabetes has been declining for the last 2–3 decades in high income countries with greater reductions in non-fatal than in fatal outcomes. The main drivers of these improvements are clear: better management of multiple risk factors, including targeting and improvement of glycaemia and LDL cholesterol levels, and blood pressure, as well as a reduction in smoking rates. Better secondary prevention has also played a role. Consequently, patients with type 2 diabetes are living longer than ever, although they continue to suffer an excess (on average double) of cardiovascular risk relative to the general population. Thus, considerable work remains to lower rates further, especially of fatal events. There is also a need to improve prediction and management of other vascular outcomes in diabetes such as heart failure and peripheral arterial disease, which appear now at least as common as incident MI or stroke as vascular outcomes in high income countries. Subgroups of patients have even higher relative risks for CVD outcome dependent on their characteristics such as age of diagnosis (worse in younger), sex (higher relative risks in women, though absolute risks remain lower than in men), ethnicity (most non-whites at higher CVD risks), and presence of coexisting disease. Finally, the greatest future needs for better CVD risk factor management in high income countries is in younger patients, whereas all LMIC patients, require improved and sustainable access to cheap and cost–effective medications across all risk factor domains to lower CVD risks.

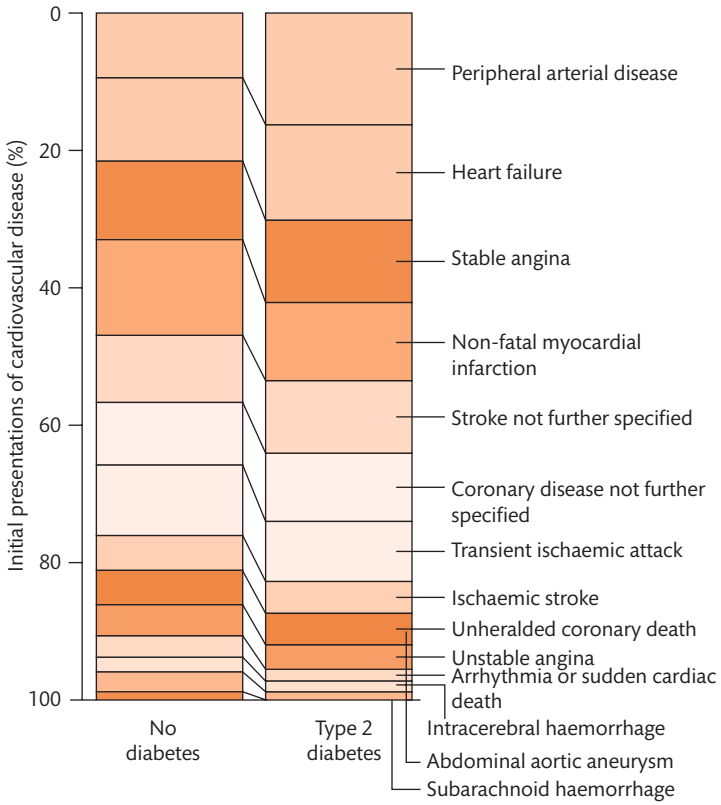


Figure 15.13.2.9 Distribution of initial presentations of cardiovascular diseases.

REFERENCES

1. Emerging Risk Factors Collaboration, Sarwar N, Gao P, *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet (London, England)*, 2010; **375**: 2215–22.
2. Hippisley-Cox J, Coupland C, Vinogradova Y, *et al.* Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*, 2008; **336**: 1475–82.
3. Di Angelantonio E, Gao P, Khan H, *et al.* Glycated hemoglobin measurement and prediction of cardiovascular disease. *JAMA*, 2014; **311**: 1225.
4. Vistisen D, Witte DR, Brunner EJ, *et al.* Risk of cardiovascular disease and death in individuals with prediabetes defined by different criteria: the Whitehall II study. *Diabetes Care*, 2018; **41**: 899–906.
5. Tancredi M, Rosengren A, Svensson A-M, *et al.* Excess mortality among persons with type 2 diabetes. *N Engl J Med*, 2015; **373**: 1720–32.
6. Rao Kondapally Seshasai S, Kaptoge S, Thompson A *et al.* Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med*, 2011; **364**: 829–41.
7. Huo X, Gao L, Guo L, *et al.* Risk of non-fatal cardiovascular diseases in early-onset versus late-onset type 2 diabetes in China: a cross-sectional study. *Lancet Diabetes Endocrinol*, 2016; **4**: 115–24.
8. Huo L, Magliano DJ, Rancière F, *et al.* Impact of age at diagnosis and duration of type 2 diabetes on mortality in Australia 1997–2011. *Diabetologia*, 2018; **61**: 1055–63.
9. Steinarsson AO, Rawshani A, Gudbjörnsdottir S, *et al.* Short-term progression of cardiometabolic risk factors in relation to age at type 2 diabetes diagnosis: a longitudinal observational study of 100,606 individuals from the Swedish National Diabetes Register. *Diabetologia*, 2018; **61**: 599–606.
10. Donnelly LA, Zhou K, Doney ASF, *et al.* Rates of glycaemic deterioration in a real-world population with type 2 diabetes. *Diabetologia*, 2018; **61**: 607–15.
11. JBS3 Board. Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). *Heart*, 2014; **100**: ii1–ii67.
12. Wannamethee SG, Shaper AG, Whincup PH, *et al.* Impact of diabetes on cardiovascular disease risk and all-cause mortality in older men: influence of age at onset, diabetes duration, and established and novel risk factors. *Arch Intern Med*, 2011; **171**: 404–10.
13. Brownrigg JRW, Hughes CO, Burleigh D, *et al.* Microvascular disease and risk of cardiovascular events among individuals with type 2 diabetes: a population-level cohort study. *Lancet Diabetes Endocrinol*, 2016; **4**: 588–97.
14. Gerstein HC, Werstuck GH. Dysglycaemia, vasculopenia, and the chronic consequences of diabetes. *Lancet Diabetes Endocrinol*, 2013; **1**: 71–8.
15. Preiss D, Sattar N, McMurray JJ. A systematic review of event rates in clinical trials in diabetes mellitus: the importance of quantifying baseline cardiovascular disease history and proteinuria and implications for clinical trial design. *Am Heart J*, 2011; **161**: 210–19.e1.

16. Sattar N, Gill JMR. Type 2 diabetes in migrant south Asians: mechanisms, mitigation, and management. *Lancet Diabetes Endocrinol*, 2015; **3**: 1004–16.
17. Johns E, Sattar N. Cardiovascular and mortality risks in migrant South Asians with type 2 diabetes: are we winning the battle? *Curr Diab Rep*, 2017; **17**: 100.
18. Prospective Studies Collaboration and Asia Pacific Cohort Studies Collaboration L, Herrington W, Halsey J, *et al.* Sex-specific relevance of diabetes to occlusive vascular and other mortality: a collaborative meta-analysis of individual data from 980 793 adults from 68 prospective studies. *Lancet Diabetes Endocrinol*, 2018; **6**: 538–46.
19. Logue J, Walker JJ, Colhoun HM, *et al.* Do men develop type 2 diabetes at lower body mass indices than women? *Diabetologia*, 2011; **54**: 3003–6.
20. Wannamethee SG, Papacosta O, Lawlor DA, *et al.* Do women exhibit greater differences in established and novel risk factors between diabetes and non-diabetes than men? the British Regional Heart Study and British Women's Heart Health Study. *Diabetologia*, 2012; **55**: 80–7.
21. Ferguson LD, Ntuk UE, Celis-Morales C, *et al.* Men across a range of ethnicities have a higher prevalence of diabetes: findings from a cross-sectional study of 500 000 UK Biobank participants. *Diabet Med*, 2018; **35**: 270–6.
22. Di Angelantonio E, Kaptoge S, Wormser D, *et al.* Association of cardiometabolic multimorbidity with mortality. *JAMA—J Am Med Assoc*, 2015; **314**: 52–60.
23. Sattar N, Petrie MC, Zinman B, *et al.* Novel diabetes drugs and the cardiovascular specialist. *J Am Coll Cardiol*, 2017; **69**: 2646–56.
24. Authors/Task Force Members, Rydén L, Grant PJ, *et al.* ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD: the Task Force on diabetes, pre-diabetes, and cardiovascular diseases of the European Society of Cardiology (ESC) and developed in collaboration with the European Association for the Study of Diabetes (EASD). *Eur Heart J*, 2013; **34**: 3035–87.
25. Gregg EW, Li Y, Wang J, *et al.* Changes in diabetes-related complications in the United States, 1990–2010. *N Engl J Med*, 2014; **370**: 1514–23.
26. Gregg EW, Sattar N, Ali MK. The changing face of diabetes complications. *Lancet Diabetes Endocrinol*, 2016; **4**: 537–47.
27. Rawshani A, Rawshani A, Franzén S, *et al.* Mortality and cardiovascular disease in type 1 and type 2 diabetes. *N Engl J Med*, 2017; **376**: 1407–18.
28. Dinesh Shah A, Langenberg C, Rapsomaniki E, *et al.* Type 2 diabetes and incidence of a wide range of cardiovascular diseases: a cohort study in 1.9 million people. *Lancet*, 2015; **385**: S86.
29. Gregg EW, Cheng YJ, Srinivasan M, *et al.* Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet (London, England)*, 2018; **391**: 2430–40.
30. Rosengren A, Edqvist J, Rawshani A, *et al.* Excess risk of hospitalization for heart failure among persons with type 2 diabetes. *Diabetologia*, 2018; **61**: 2300–9.
31. McAllister DA, Read S, Kerssens J, *et al.* Incidence of hospitalisation for heart failure and case-fatality among 3.25 million people with and without diabetes. *Circulation*, 2018; **138**: 2774–86.
32. Rawshani A, Rawshani A, Franzén S, *et al.* Risk factors, mortality and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*, 2018; **379**: 633–44.
33. Zinman B, Wanner C, Lachin JM, *et al.* Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*, 2015; **373**: 2117–28.
34. Packer M. Role of deranged energy deprivation signaling in the pathogenesis of cardiac and renal disease in states of perceived nutrient overabundance. *Circulation* 2020; **141**(25): 2095–105.
35. Greene SJ, Vaduganathan M, Khan MS, *et al.* Prevalent and incident heart failure in cardiovascular outcome trials of patients with type 2 diabetes. *J Am Coll Cardiol*, 2018; **71**: 1379–90.
36. McMurray JJ V, Gerstein HC, Holman RR, *et al.* Heart failure: a cardiovascular outcome in diabetes that can no longer be ignored. *Lancet Diabetes Endocrinol*, 2014; **2**: 843–51.
37. Ford ES. Trends in the risk for coronary heart disease among adults with diagnosed diabetes in the U.S.: findings from the National Health and Nutrition Examination Survey, 1999–2008. *Diabetes Care*, 2011; **34**: 1337–43.
38. Ray KK, Seshasai SRK, Wijesuriya S, *et al.* Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. *Lancet*, 2009; **373**: 1765–72.
39. Misra A, Sattar N, Tandon N, *et al.* Clinical management of type 2 diabetes in south Asia. *Lancet Diabetes Endocrinol*, 2018; **6**: 979–91.

15.13.3 Macrovascular Disease in Type 1 Diabetes

John R. Petrie

Introduction

As onset is most common early in life, type 1 diabetes is often thought of as a condition of children and adolescents. However, more than half of those affected are over 40 years of age and the 'typical' person with type 1 diabetes is a middle-aged adult who has had the condition for several decades [1].

'Macrovascular disease' is a collective term for the complications of diabetes as they affect medium or large diameter arteries (>100 μm), predominantly the coronary, cerebral and peripheral (lower limb) arteries. Also known as 'cardiovascular disease', it comprises angina, myocardial infarction, stroke, congestive heart failure, and peripheral artery disease. The glucose-dependent mechanisms and pathways mediating microvascular complications (i.e. retinopathy, nephropathy, and neuropathy) play a key role but additional

inflammatory, lipid, haemodynamic, and other pathways are also implicated.

Pathology and Pathophysiology

By activating enzymes that generate oxidative stress (e.g. by producing superoxide ions), hyperglycaemia initiates several parallel adverse biochemical processes within vascular cells, including activation of the polyol and hexosamine pathways, activation of Protein Kinase C (pro-inflammatory), and formation of advanced glycation end-products (AGEs) [2]. These glucose-mediated mechanisms account for much of the adverse cardiovascular risk in type 1 diabetes [3] but additional immune mechanisms specific to type 1 diabetes may also play a role.

As may also occur in the general population, atherosclerosis in type 1 diabetes is often associated with calcification of the layer of the arterial wall adjacent to the lumen, i.e. the tunica intima. In type 1 diabetes, this may be accompanied by an additional (and independent) process of calcification of the tunica media known as Monckeberg's arteriosclerosis. This is most often seen in the aorta, causing stiffening, elevated systolic BP, and decreased coronary artery perfusion—and in lower limb vessels in association with arterial insufficiency. Its significance is uncertain.

Epidemiology

Since the discovery of insulin a century ago, life expectancy of individuals with type 1 diabetes has dramatically improved, but remains substantially reduced: by 11 years and 13 years for a 20-year-old man and woman, respectively [4]. Metabolic complications (hypoglycaemia, ketoacidosis) and infections (osteomyelitis) are the commonest causes of death in younger individuals, but cardiovascular disease predominates after the age of 40 years [5].

A man in his mid-40s with type 1 diabetes living in Scotland, for example, has an approximately 12% risk of having a cardiovascular event (as defined earlier) over the next ten years, compared with 4% risk in the general population (i.e. a threefold elevation in *relative* risk). For a woman of the same age, the average ten-year risk is 8% compared with 1.6% in the general population (i.e. a fivefold elevation [5]). At younger ages, while the absolute rates are lower, the relative risks are even higher (due to the low risk of the background population). After the menopause, women with type 1 diabetes lose premenopausal protection from cardiovascular disease.

Heart failure is rare in young adults without type 1 diabetes, but recent population-based data from Scotland indicate a fivefold excess risk in men aged 20–29 years (3.6% vs. 0.7% ten-year risk) [6]. There are similar data from Sweden [7–9], where women (but not men) aged 40 years were also shown to have twice the risk of atrial fibrillation compared with the general population [10].

Following a cardiovascular event, outcomes are poorer in type 1 diabetes: for example, in Scotland 30 day mortality in men aged 30–49 years of age after admission to hospital with heart failure is 9.3% compared with 4.4% in the non-diabetic population (i.e. relative risk more than doubled) [6].

Risk Factors and Markers

Glucose: Much of our knowledge about type 1 diabetes comes from the Diabetes Control and Complications Trial (DCCT), a landmark randomized clinical trial initiated in North America in the early 1980s. It compared rates of complications in 1441 children and adolescents with type 1 diabetes on intensive (average HbA1c 7.2%) or conventional (average HbA1c 9.1%) insulin treatment [11]. Follow-up in the study for 6.5 years was insufficient to detect differences in rates of macrovascular complications between groups, but these became evident as participants were followed-up post-randomization in the Epidemiology of Diabetes and its Complications (EDIC study; average HbA1c 8.0% in both groups). In the most recent EDIC study follow-up, 30 years after initial DCCT randomization, there was a 30% reduction in the risk of cardiovascular events in those formerly allocated to intensive insulin therapy: 149 cardiovascular events (in 82 participants) in the intensive arm and 217 cardiovascular events (in 102 participants) in the conventional arm [12].

Other traditional risk factors: In younger individuals with type 1 diabetes the DCCT suggests that blood pressure and cholesterol on average contribute little to overall cardiovascular risk given the dominant effect of blood glucose. However, after the age of 40 years, or 15–20 years from diagnosis for those with earlier onset, blood pressure and cholesterol start to mediate an increasing proportion of the overall risk and require to be targeted directly [13]. In addition, smoking is a powerful risk factor for macrovascular disease in type 1 diabetes, as in the general population, but unfortunately remains prevalent in up to 25% of individuals [5], often receiving insufficient attention.

Microalbuminuria: Much of the excess cardiovascular risk in type 1 diabetes occurs in those who develop renal disease, whether subclinical (microalbuminuria) or clinical (diabetic nephropathy) [13]. Preventing microalbuminuria is an important therapeutic goal, particularly as remission from sustained microalbuminuria to normoalbuminuria in DCCT/EDIC did not lead to a significant improvement in cardiovascular (or renal) outcomes [14].

Novel risk markers: A number of other risk markers are associated with the development of macrovascular complications (e.g. advanced glycation end-products [15], high sensitivity C-reactive protein [16], and haptoglobin genotype [17]). However, these are not routinely measured in clinical practice [18] as risk can be quite accurately predicted by combining traditional risk factors [19].

Assessment

Cardiovascular disease: As in type 2 diabetes and the general population, routine screening for coronary and cerebrovascular disease is not currently undertaken in asymptomatic people with type 1 diabetes, even in those with microalbuminuria [20]. This may seem surprising given a marked elevation of cardiovascular risk. However, while some tests (e.g. measurement of coronary artery calcification by electron beam computed tomography) provide accurate risk stratification, no test has as yet been demonstrated to

have sufficient sensitivity and specificity to guide the planning of cost-effective interventions (i.e. to be suitable for use in widespread screening).

In the absence of screening, presentation with coronary and cerebrovascular disease is thus likely to occur as in the general population, either with the gradual onset of symptoms (angina, breathlessness) or with sudden development of angina, myocardial infarction, congestive heart failure, transient ischaemic attack, or stroke leading to emergency medical assessment. Assessment follows the same pathways as in the general population.

Peripheral vascular disease: In contrast, screening for peripheral arterial disease by palpating the posterior tibial and dorsalis pedis pulses is a routine component of annual foot ulcer risk assessment. Absence of both conveys a significantly higher risk for foot ulceration (and requirement for amputation) in that foot and is associated with a two-year mortality of 14% [21]. There is a further increase in risk if protective sensation (detected by light touch with a 10 g monofilament) is also absent. Individuals with absent foot pulses should be referred to a podiatrist for assessment and ongoing foot care.

Those who also have a history of intermittent claudication and reduced walking distance on direct questioning should be referred to a vascular surgeon for consideration of angiography and/or revascularization. At the time of writing, rates of lower limb amputation vary widely between different areas of the United Kingdom: this is thought to reflect, at least in part, the quality of local foot screening and referral systems [22].

Prevention and Risk Factor Management

Glucose control: Mean glucose concentration (as estimated by HbA1c) is the dominant risk factor for coronary and cerebrovascular disease in type 1 diabetes [3, 13]. Within- and between-day glucose variability predict hypoglycaemia, but do not appear to contribute independently to the risk of long-term complications [23]. For younger individuals, achieving and maintaining target HbA1c while minimizing the risk of hypoglycaemia is therefore the most rational strategy for preventing macrovascular disease.

Doing this in everyday life using multiple daily injections of insulin is challenging, particularly for younger individuals, and HbA1c targets are reached in only a minority [24]. Even in those who do sustain intensive glycaemic control, the cardiovascular benefits can be attenuated or even lost entirely over time by insulin-induced weight gain [25].

Blood pressure and cholesterol: There have been no randomized trials directly addressing management of these risk factors in relation to cardiovascular outcomes specifically in type 1 diabetes. Guidelines therefore draw on epidemiology and studies in type 2 diabetes for selecting therapies, as well as in setting appropriate thresholds and targets for their use (Table 15.13.3.1). In addition, there have been some outcome trials from which people with type 1 diabetes were not excluded (e.g. the Heart Protection Study with simvastatin): this reported a similar point estimate for reduction in risk as in other subgroups, with no evidence of statistical heterogeneity but wide and non-statistically significant confidence intervals for the treatment effect [26].

Smoking: There are few trials from which the benefits of smoking cessation in type 1 diabetes can be directly estimated, or indeed to guide how it can best be achieved and maintained. However, its likely benefits if achieved seem self-evident.

Antiplatelet therapy: Low dose aspirin was widely prescribed for primary prevention of cardiovascular disease for asymptomatic people with types 1 and 2 diabetes until publication of the Prevention Of Progression of Arterial Disease And Diabetes (POPADAD) trial in 2008 [27]. As there was no evidence of benefit over seven years, this practice was removed from guidelines. However, as the trial was not large enough to exclude a small benefit, the larger ASCEND study was undertaken with 12 times the number of participants, 6% of whom had type 1 diabetes. The results, published in 2018, showed a lower absolute rate of cardiovascular events with aspirin than with placebo (8.5% vs. 9.6%, $P = 0.01$) over 7.4 years but this was outweighed by an increased risk of major bleeding (4.1% vs. 3.2%, $P = 0.003$), supporting existing guidelines [28].

Given these uncertainties (Table 15.13.3.1), further clinical trials aimed specifically at optimizing cardiovascular prevention in type 1 diabetes are urgently required.

Treatment

At coronary angiography, people with type 1 diabetes are more likely to have diffuse disease affecting several rather than single arteries: in general, outcomes are better with coronary artery bypass grafting than with stenting individual vessels [29], but technology is developing rapidly and decision-making must be made on an individual basis.

Table 15.13.3.1 Risk factor management in type 1 diabetes: targets for primary prevention in adults

Risk factor	Test	Target	Evidence base	Notes
Glucose	HbA1c	≤7.0% (53 mmol/mol)	High level: DCCT and EDIC [3, 11]	
Lipids	LDL cholesterol	<2.6 mmol/L (<100 mg/dl)	Mainly extrapolation from type 2 diabetes [20]	Lifestyle measures plus statins*
Blood pressure	Semi-automated sphygmomanometer (clinic reading)	<130/ 80 mmHg [20]	Absence of high quality data [34]	Lifestyle measures plus ACE inhibitor or ARB as first line*

* in women at risk of pregnancy ensure adequate contraceptive measures in place.

Abbreviations: HbA1c, glycated haemoglobin; LDL, low density lipoprotein; ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker.

Future Perspectives

One approach to reducing rates of cardiovascular events in type 1 diabetes is 'adjunct therapy' (i.e. adding in non-insulin treatments that have been shown to be of benefit in type 2 diabetes). The REMOVAL trial recently showed over three years in 428 middle-aged people with type 1 diabetes that metformin decreases weight and LDL cholesterol and may reduce the rate of progression of carotid atherosclerosis [30]. Other candidate agents that require further testing in type 1 diabetes include glucagon-like peptide-1 agonists and sodium glucose transporter-2 inhibitors, ideally using cardiovascular outcomes rather than surrogate endpoints. Wider use of PCSK9 inhibitors in those who do not achieve lipid targets with statins may also be of benefit [31].

Increasing use of technology is another strategy that provides hope for improved cardiovascular outcomes in type 1 diabetes: lower rates of cardiovascular mortality have been observed in association with continuous subcutaneous insulin infusion (CSII, 'insulin pump') use in Sweden [32]. Increasing combination of this method of insulin delivery with continuous glucose monitoring systems, moving towards 'closed-loop' systems ('the artificial pancreas'), is likely to allow more individuals with type 1 diabetes to achieve target glycaemia, while avoiding hypoglycaemia and weight gain.

There have been attempts to design specific therapies to disrupt the downstream intracellular pathways activated by glucose (e.g. alagebrium, an 'AGE-breaker'), but these have so far shown either toxicity or insufficient efficacy, perhaps because glucose can cause vascular damage by multiple biochemical pathways [33].

Conclusion

Macrovascular (cardiovascular) disease decreases life expectancy by more than 10 years on average in type 1 diabetes. Current care focuses on its prevention via glucose control and avoidance of cigarette smoking, combined with controlling blood pressure and cholesterol with longer disease duration. However, this strategy is difficult to implement for many individuals and new approaches are urgently required to reduce macrovascular complications and bring life expectancy in line with the non-diabetic population.

REFERENCES

1. Scottish Diabetes Survey 2016. Available at: http://www.diabetesinscotland.org.uk/wp-content/uploads/2019/12/Diabetes-in-Scotland-website-Scottish-Diabetes-Survey-2016.pdf (accessed: 26 April 2020).
2. Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*, 2005; **54**: 1615–25.
3. Nathan DM, Cleary PA, Backlund JY, *et al.* Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*, 2005; **353**: 2643–53.
4. Ho CY, Shanahan CM. Medial arterial calcification: an overlooked player in peripheral arterial disease. *Arterioscler Thromb Vasc Biol*, 2016; **36**: 1475–82.
5. Livingstone SJ, Levin D, Looker HC, *et al.* Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008–2010. *JAMA*, 2015; **313**: 37–44.
6. McAllister DA, Read S, Kerssens J, *et al.* Incidence of hospitalization for heart failure and case-fatality among 3.25 million people with and without diabetes mellitus. *Circulation*, 2018; **138**: 2774–86.
7. Lind M, Bounias I, Olsson M, Gudbjörnsdottir S, Svensson AM, Rosengren A. Glycaemic control and incidence of heart failure in 20,985 patients with type 1 diabetes: an observational study. *Lancet*, 2011; **378**: 140–6.
8. Rawshani A, Rawshani A, Franzén S, *et al.* Range of risk factor levels: control, mortality, and cardiovascular outcomes in type 1 diabetes mellitus. *Circulation*, 2017; **135**: 1522–31.
9. Matuleviciene-Anängen V, Rosengren A, Svensson AM, *et al.* Glycaemic control and excess risk of major coronary events in persons with type 1 diabetes. *Heart*, 2017; **103**: 1687–95.
10. Dahlqvist S, Rosengren A, Gudbjörnsdottir S, *et al.* Risk of atrial fibrillation in people with type 1 diabetes compared with matched controls from the general population: a prospective case-control study. *Lancet Diabetes Endocrinol*, 2017; **5**: 799–807.
11. Diabetes Control and Complications Trial Research Group, Nathan DM, Genuth S, *et al.* The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*, 1993; **329**: 977–86.
12. Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) Study Research Group. Intensive diabetes treatment and cardiovascular outcomes in type 1 diabetes: the DCCT/EDIC study 30-year follow-up. *Diabetes Care*, 2016; **39**: 686–93.
13. Bebu I, Braffett BH, Pop-Busui R, Orchard TJ, Nathan DM, Lachin JM; DCCT/EDIC Research Group. The relationship of blood glucose with cardiovascular disease is mediated over time by traditional risk factors in type 1 diabetes: the DCCT/EDIC study. *Diabetologia*, 2017; **60**: 2084–91.
14. de Boer IH, Gao X, Cleary PA, *et al.*; Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Research Group. Albuminuria changes and cardiovascular and renal outcomes in type 1 diabetes: the DCCT/EDIC study. *Clin J Am Soc Nephrol*, 2016; **11**: 1969–77.
15. Libby P, Nathan DM, Abraham K, *et al.*; National Heart, Lung, and Blood Institute; National Institute of Diabetes and Digestive and Kidney Diseases Working Group on Cardiovascular Complications of Type 1 Diabetes Mellitus. Report of the National Heart, Lung, and Blood Institute—National Institute of Diabetes and Digestive and Kidney Diseases Working Group on cardiovascular complications of type 1 diabetes mellitus. *Circulation*, 2005; **111**: 3489–93.
16. Alman AC, Kinney GL, Tracy RP, *et al.* Prospective association between inflammatory markers and progression of coronary artery calcification in adults with and without type 1 diabetes. *Diabetes Care*, 2013; **36**: 1967–73.
17. Orchard TJ, Costacou T. When are type 1 diabetic patients at risk for cardiovascular disease? *Curr Diab Rep*, 2010; **10**: 48–54.
18. American Diabetes Association. 9. Cardiovascular disease and risk management: Standards of Medical Care in Diabetes—2018. *Diabetes Care*, 2018; **41**(Suppl 1): S86–S104.

19. Vistisen D, Andersen GS, Hansen CS, *et al.* Prediction of first cardiovascular disease event in type 1 diabetes mellitus: the steno type 1 risk engine. *Circulation*, 2016; **133**: 1058–66.
20. de Ferranti SD, de Boer IH, Fonseca V, *et al.* Type 1 diabetes mellitus and cardiovascular disease: a scientific statement from the American Heart Association and American Diabetes Association. *Diabetes Care*, 2014; **37**: 2843–63.
21. Vadiveloo T, Jeffcoate W, Donnan PT, *et al.* Amputation-free survival in 17,353 people at high risk for foot ulceration in diabetes: a national observational study. *Diabetologia*. 2018; **61**(12): 2590–7.
22. Holman N, Young RJ, Jeffcoate WJ. Variation in the recorded incidence of amputation of the lower limb in England. *Diabetologia*, 2012; **55**: 1919–25.
23. Lachin JM, Bebu I, Bergenstal RM, *et al.*; DCCT/EDIC Research Group. Association of glycemic variability in type 1 diabetes with progression of microvascular outcomes in the Diabetes Control and Complications Trial. *Diabetes Care*, 2017; **40**: 777–83.
24. McKnight JA, Wild SH, Lamb MJ, *et al.* Glycaemic control of type 1 diabetes in clinical practice early in the 21st century: an international comparison. *Diabet Med*, 2015; **32**: 1036–50.
25. Purnell JQ, Braffett BH, Zinman B, *et al.* Impact of excessive weight gain on cardiovascular outcomes in type 1 diabetes: results from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study. *Diabetes Care*, 2017; **40**: 1756–62.
26. Collins R, Armitage J, Parish S, Sleigh P, Peto R; Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. *Lancet*, 2003; **361**: 2005–16.
27. Belch J, MacCuish A, Campbell I, *et al.* The Prevention Of Progression of Arterial Disease And Diabetes (POPADAD) trial: factorial randomised placebo controlled trial of aspirin and antioxidants in patients with diabetes and asymptomatic peripheral arterial disease. *BMJ*, 2008; **337**: a1840.
28. ASCEND Study Collaborative Group, Bowman L, Mafham M, *et al.* Effects of aspirin for primary prevention in persons with diabetes mellitus. *N Engl J Med*, 2018; **379**: 1529–39.
29. Farkouh ME, Domanski M, Sleeper LA, *et al.* Strategies for multivessel revascularization in patients with diabetes. *N Engl J Med*, 2012; **367**: 2375–84.
30. Petrie JR, Chaturvedi N, Ford I, *et al.* Cardiovascular and metabolic effects of metformin in patients with type 1 diabetes (REMOVAL): a double-blind, randomised, placebo-controlled trial. *Lancet Diabetes Endocrinol*, 2017; **5**: 597–609.
31. Leiter LA, Cariou B, Müller-Wieland D, *et al.* Efficacy and safety of alirocumab in insulin-treated individuals with type 1 or type 2 diabetes and high cardiovascular risk: the ODYSSEY DM-INSULIN randomized trial. *Diabetes Obes Metab*, 2017; **19**: 1781–92.
32. Steineck I, Cederholm J, Eliasson B, *et al.*; Swedish National Diabetes Register. Insulin pump therapy, multiple daily injections, and cardiovascular mortality in 18,168 people with type 1 diabetes: observational study. *BMJ*, 2015; **350**: h3234.
33. MT Coughlan, JM Forbes ME Cooper. Role of the AGE crosslink breaker, alagebrium, as a renoprotective agent in diabetes. *Kidney Int*, 2007; **72**: S54–S60.
34. American Diabetes Association. 10. Cardiovascular disease and risk management: standards of medical care in diabetes—2019. *Diabetes Care*, 2019; **42**(Suppl 1): S103–23.

15.13.4 Diabetic Dyslipidaemia

Bruno Vergès

Introduction

Diabetic dyslipidaemia, which encompasses abnormalities in all lipoproteins, is a major contributor to the increased cardiovascular risk associated with type 2 diabetes (T2DM). In this chapter, the pathophysiology of diabetic dyslipidaemia in T2DM, its relationship with cardiovascular disease and its diagnosis and management will be discussed. At the end of this chapter, the main features of the diabetic dyslipidaemia observed in type 1 diabetes (T1DM) will be considered.

Diabetic Dyslipidaemia in T2DM

Pathophysiology of Diabetic Dyslipidaemia

Lipid abnormalities in T2DM are not only quantitative, but also qualitative and kinetic in nature [1]. Increased triglycerides and reduced HDL cholesterol are the main quantitative abnormalities whereas qualitative and kinetic abnormalities are observed for all lipoproteins (Table 15.13.4.1).

Many of the lipid abnormalities observed in patients with T2DM precede the onset of diabetes, and form an important part of the 'insulin resistance' or 'metabolic' syndrome [2]. This emphasizes the important role of insulin resistance in the pathophysiology of diabetic dyslipidaemia. Insulin is a key hormone in the regulation of lipid metabolism, exerting a major influence on several different tissues which collectively regulate lipid metabolism. For example, insulin inhibits hormone-sensitive lipase in adipose tissue, suppressing release and hepatic delivery of free fatty acids, and also inhibits hepatic VLDL production, both directly and indirectly via alteration of substrate flux. It also activates lipoprotein lipase (LPL), responsible for triglyceride catabolism and free fatty acid trapping in tissues, and increases the expression of LDL receptor and LRP (LDL receptor-related protein), responsible for hepatic uptake of chylomicron-remnants. The main lipid abnormalities in T2DM are shown in Figure 15.13.4.1.

Patients with T2DM show postprandial hyperlipidaemia, which is a pro-atherogenic phenomenon. This is due both to increased production and delayed catabolism of chylomicrons. Insulin resistance is likely to be involved in the increased chylomicron production, since the normal acute suppression of postprandial chylomicron secretion by insulin is absent in patients with T2DM [3]. The delayed catabolism of chylomicrons is attributable to decreased activity of LPL, the enzyme responsible for chylomicron hydrolysis, to increased plasma levels of ApoC-III, an inhibitor of

Table 15.13.4.1 Key changes of lipoproteins in type 2 diabetes

Lipoprotein	Quantitative changes in plasma concentration	Qualitative changes	Kinetic/metabolic changes
Chylomicron	Increased	• Very few data (decreased ApoE content in diabetic rabbits)	• Increased production • Decreased catabolism
VLDL	Increased	• Greater proportion of larger particles ($VLDL_1$) • Increased palmitic acid-containing species and diacylglycerol, reduced sphingomyelin • Glycation	• Increased production • Decreased catabolism
LDL	No change or slightly increased	• Greater proportion of small dense particles (triglyceride enrichment) • Increased LDL oxidation • Increased palmitic acid-containing species and diacylglycerol, reduced sphingomyelin • Glycation	• Decreased catabolism
HDL	Decreased	• Triglyceride enrichment • Increased palmitic acid-containing species and diacylglycerol, reduced sphingomyelin • Glycation • ↘ vasorelaxant effect • ↘ antioxidative effect	• Increased catabolism

Apo: apolipoprotein; VLDL: Very Low-Density Lipoprotein; LDL: Low-Density Lipoprotein; HDL: High-Density Lipoprotein.

LPL, and potentially to reduced expression of LRP, the receptor for chylomicron-remnants.

Fasting hypertriglyceridemia in T2DM is largely due to an increased number of VLDLs, particularly large $VLDL_1$ particles [4]. Both increased production and delayed catabolism of VLDL are responsible for the increased VLDL pool. Increased production of $VLDL_1$ has been shown to be correlated with indices of insulin resistance [1]. Several mechanisms are responsible for the increased production of $VLDL_1$. First, insulin resistance is associated with reduced inhibition of the hormone-sensitive lipase by insulin in adipose tissue, leading to increased lipolysis and thus augmented delivery of free fatty acids to the liver, which stimulates synthesis

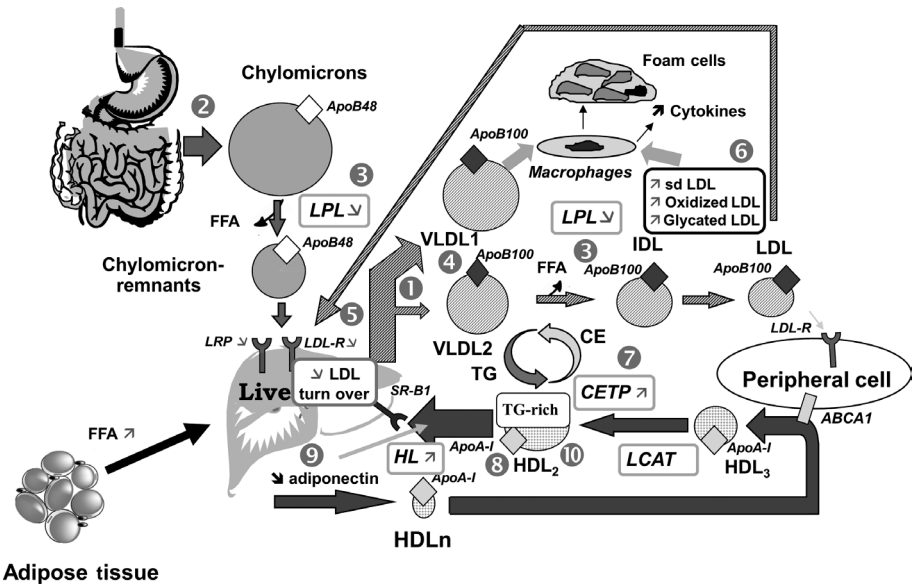


Figure 15.13.4.1 Main lipid abnormalities in type 2 diabetes. Triglycerides (hypertriglyceridemia, qualitative and kinetic abnormalities): ❶ increased VLDL production (mostly $VLDL_1$), ❷ increased chylomicron production, ❸ reduced catabolism of both chylomicrons and VLDLs (diminished LPL activity), ❹ increased production of large VLDL ($VLDL_1$), preferentially taken up by macrophages; **LDL** (qualitative and kinetic abnormalities): ❺ reduced LDL turnover (decreased LDL receptor), ❻ increased number of glycated LDLs, small dense LDLs (TG-rich) and oxidized LDLs, which are preferentially taken up by macrophages; **HDL** (low HDL cholesterol, qualitative and kinetic abnormalities): ❼ increased CETP activity (increased transfer of triglycerides from TG-rich lipoproteins to LDLs and HDLs), ❽ increased TG content of HDLs, promoting HL activity and HDL catabolism, ❾ low plasma adiponectin favouring the increase in HDL catabolism, ❿ Dysfunctional HDLs. Apo: apolipoprotein; FFA: free fatty acid; VLDL: very low-density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; sd LDL: small, dense LDL; HDL: high-density lipoprotein; HDL_n: nascent HDL; TG: triglyceride; CE: cholesterol esters; LDL-R: LDL receptor; LRP: LDL receptor-related protein; ABCA1: ATP-binding cassette A1 transporter; SR-B1: scavenger receptor B1; HSL: hormone-sensitive lipase; LPL: lipoprotein lipase; HL: hepatic lipase; CETP: cholesteryl ester transfer protein; LCAT: lecithin-cholesterol acyl transferase.

of triglycerides in hepatocytes. Second, the normal suppressant effect of insulin on VLDL (and more precisely $VLDL_1$) production is blunted due to hepatic insulin resistance [1]. Third, de novo lipogenesis is increased in individuals with insulin resistance [5]. Delayed catabolism of VLDL in diabetic dyslipidaemia is due to reduced activity of LPL secondary to insulin resistance and/or 'relative' insulin deficiency. Increased plasma levels of ApoC-III could also contribute to the decreased catabolism of VLDL in T2DM patients.

VLDLs from patients with T2DM also show qualitative abnormalities that are potentially atherogenic. Indeed, there is an increase in large $VLDL_1$ particles, enriched with cholesterol esters and phospholipids, which are preferentially taken up by macrophages, leading to the formation of foam cells in vessel walls and the production of pro-inflammatory cytokines. VLDLs from patients with T2DM have increased diacylglycerol content and reduced sphingomyelin, along with increased palmitic acid-containing species. In addition, glycation of apolipoproteins in VLDL (ApoB, ApoCs, ApoE) in poorly controlled diabetes may be detrimental [1].

Although plasma LDL cholesterol is usually normal in patients with T2DM, LDL metabolism is significantly modified, with a substantial reduction of its catabolism observed in parallel with a reduction of its production, secondary to diminution of LDL-receptors. This extends the lifespan of LDL in the plasma and may promote lipid deposition in artery walls. LDLs from patients with T2DM also show important qualitative abnormalities. Small, dense, triglyceride-rich LDL particles are more prevalent in T2DM as a direct consequence of hypertriglyceridemia and increased cholesteryl ester transfer protein (CETP) activity. Furthermore, an increased number of oxidized and glycated LDLs is also observed. Small dense LDLs, oxidized and glycated LDLs are pro-atherogenic since they induce the production of pro-inflammatory cytokines and adhesion molecules. They are preferentially taken up by macrophages, leading to the formation of foam cells. Moreover, LDLs from patients with T2DM have increased diacylglycerol content and reduced sphingomyelin, along with increased palmitic acid-containing species.

The decrease in HDL cholesterol in patients with T2DM is due to increased catabolism of HDLs. Hypertriglyceridemia, or more precisely the increased $VLDL_1$ pool, is an important factor responsible for accelerated HDL catabolism [1, 6]. Indeed, hypertriglyceridemia, observed in T2DM, promotes CETP-mediated triglyceride enrichment of HDL particles and, as a consequence, enhances HDL catabolism, because HDL-rich particles are very good substrates for hepatic lipase, the enzyme responsible for HDL catabolism. The reduction in plasma adiponectin levels often observed in individuals with insulin resistance and T2DM may also be directly involved in the accelerated HDL catabolism [7]. As for other lipoproteins, several qualitative abnormalities of HDL are observed in T2DM, including increased glycation of apoA-I, increased diacylglycerol content and reduced sphingomyelin, and an increased complement of palmitic acid-containing species. HDLs from patients with T2DM are known to be dysfunctional, with reduced capacity to stimulate *ex vivo* cholesterol efflux from cells, and reduced antioxidative and vasorelaxant properties [1].

Diabetic Dyslipidaemia and Cardiovascular Disease

Similarly to what is observed in the general population, elevated LDL cholesterol is a risk factor for cardiovascular disease (CVD) in T2DM, as shown in the United Kingdom Prospective Diabetes Study (UKPDS), the Strong Heart Study or a large Finnish epidemiological study [8–10]. Hypertriglyceridemia has been suggested to be an independent risk factor for CVD in T2DM in the WHO study, the Paris prospective study and the Schwabing study [11–13], while HDL cholesterol has been found to be inversely associated with CVD in patients with T2DM in the UKPDS, and the Strong Heart Study [8, 9]. In the Finnish epidemiological study and the long-term follow-up of the Strong Heart Study, coupling of hypertriglyceridemia and low HDL cholesterol was an independent risk factor for CVD [10, 14]. Collectively these findings suggest that hypertriglyceridemia and low HDL cholesterol, the main quantitative lipid abnormalities in T2DM, are likely to be *bona fide* risk factors for CVD (Table 15.13.4.2).

Many *in vitro* and *ex vivo* data argue that the qualitative lipoprotein abnormalities of diabetic dyslipidaemia are likely to be atherogenic, as just discussed. However, the direct association between these abnormalities and CVD has not been extensively studied. Although some epidemiological studies have shown that high concentrations of small dense LDL particles are an independent factor for CVD in large unselected populations, the question of a direct role of small dense LDL particles in development of CVD among patients with T2DM remains to be answered. Similarly, the direct role of other qualitative abnormalities such as increased levels of oxidized LDL, glycation of apolipoproteins or dysfunctional HDLs in the increased CVD risk in T2DM remains to be fully assessed.

Diagnosis of Diabetic Dyslipidaemia

Diagnosis of diabetic dyslipidaemia is based on quantitative lipid abnormalities, being made based on the presence of triglycerides ≥150 mg/dl (1.7 mmol/L) and/or HDL cholesterol <40 mg/dl (1.03 mmol/L) in men or <50 mg/dl (1.29 mmol/L) in women. However, vigilance is required, as not all lipid disorders observed in

Table 15.13.4.2 Independent association between plasma lipids and CVD risk in patients with type 2 diabetes. Data from epidemiological studies. CVD, cardiovascular disease.

LDL cholesterol	**Positive** independent association with CVD risk: • UKPDS [16] • Strong Heart Study [17] • Finnish epidemiological study [18]
HDL cholesterol	**Negative** independent association with CVD risk: • UKPDS [16] • Strong Heart Study [17] • Finnish epidemiological study [18], in association with triglycerides • Long-term follow-up of the Strong Heart Study [19], in association with triglycerides
Triglycerides	**Positive** independent association with CVD risk: • WHO study [20] • Paris Prospective study [21] • Schwabing study [22] • Finnish epidemiological study [18], in association with low HDL cholesterol • Long-term follow-up of the Strong Heart Study [19] in association with low HDL cholesterol

patients with T2DM are related to diabetic dyslipidaemia. Indeed, in situations with very high triglycerides (>500 mg/dl or 5.6 mmol/L), lipodystrophy or a primary hypertriglyceridemia associated with diabetes should be considered. If a significant elevation of plasma LDL-cholesterol is observed in a patient with T2DM, then a diagnosis of coincidental familial hypercholesterolemia must be considered.

Management of Diabetic Dyslipidaemia

Lifestyle modification focusing on weight loss (if indicated), dietary alterations including reduction of saturated fat and transfat, and increase of n-3 fatty acids and viscous fibre, and increased physical activity are recommended to improve the lipid profile in patients with diabetes. Improvement of glycaemic control is likely to reduce hypertriglyceridemia but will not, alone, normalize the lipid profile.

The clear cardiovascular benefit of LDL cholesterol reduction with statins in patients with diabetes has been shown in several prospective controlled trials [15]. In the Heart Protection Study (HPS), in the subgroup of 5963 patients with diabetes (mainly type 2), treatment with simvastatin reduced by 22% the risk for major CV events (P <0.0001) and by 26% the risk for major coronary events (P <0.0001). There was also a highly significant 33% reduction of major CV events (P <0.0001) among the 2912 diabetic patients without history of CVD. The primary prevention Collaborative Atorvastatin Diabetes Study (CARDS), which was undertaken in 2838 patients with T2DM, showed that treatment with atorvastatin 10 mg daily for 3.9 years induced a 37% reduction of major CV events (P = 0.001). In the Treating to New Targets (TNT) study, which compared the efficacy of atorvastatin 80 mg daily versus 10 mg daily, a significant 25% reduction of major CV events (P = 0.026) was observed with the highest atorvastatin dose in the subgroup of 1501 diabetic patients.

Data from all clinical trials indicate that the CV benefit obtained by LDL cholesterol reduction with a statin is similar in patients with diabetes and in non-diabetic individuals. The clear CV benefit of statin treatment in patients with diabetes has been confirmed in several meta-analyses. Indeed, data from a meta-analysis of 12 randomized prospective studies indicate that, in patients with diabetes, a significant 21% reduction of major CV event and a 19% reduction of CV mortality is observed for each 1 mmol/L (39 mg/dl) decrease in LDL cholesterol [8]. The Cholesterol Treatment Trialists' Collaborators meta-analysis, performed from 14 randomized prospective studies, which includes 17 220 patients with T2DM, shows a significant 21% reduction of major CV event for each 1 mmol/L (39 mg/dl) decrease in LDL cholesterol [9].

Because of the clear cardiovascular benefit of LDL cholesterol reduction in patients with T2DM, the use of statins is widely recommended in guidelines. In the 2016 European Guidelines on CVD prevention, lipid lowering agents (principally statins) are recommended in all patients with diabetes above the age of 40 years, since most diabetic patients are at high risk [10]. In diabetic patients with high-risk, LDL cholesterol target <2.6 mmol/L (<100 mg/dl) or a reduction of at least 50% if the baseline LDL cholesterol is between 2.6 and 5.1 mmol/L (100 and 200 mg/dl) is recommended. In diabetic patients at very high-risk (secondary prevention, target organ damage, severe kidney disease with GFR<30 ml/min or a calculated SCORE ≥10%), an LDL cholesterol target <1.8 mmol/L (<70 mg/dl), or a reduction of at least 50% if the baseline LDL cholesterol is between 1.8 and 3.5 mmol/L (70 and 135 mg/dl), is recommended [10]. American Diabetes Association (ADA) standards of care recommend moderate-intensity statins for all T2DM patients over the age of 40 years and in patients less than 40 years with additional CV risk factors [11]. For patients of all ages with diabetes and CVD, high-intensity statin therapy is recommended by the ADA [11]. If the LDL-target is not obtained with the optimal dose of statin, additional treatment with ezetimibe or PCSK9 inhibitor may be needed.

Post-hoc analyses of the FIELD and ACCORD-Lipid trials have shown a significant reduction of major cardiovascular events with fenofibrate in individuals with elevated triglycerides and low HDL cholesterol. This was confirmed in a meta-analysis demonstrating a 35% reduction for the risk of major cardiovascular events with fibrates in this specific population [12]. This is the reason why some recent guidelines suggest considering adding fenofibrate to a statin in high CV risk patients featuring fasting triglyceride levels ≥2.26 mmol/L (200 mg/dl) and low HDL cholesterol levels [13]. For patients with fasting triglycerides above 5.7 mmol/L (500 mg/dl), it is recommended to look for secondary or primary hypertriglyceridemia and to consider fibrate treatment to reduce the risk of pancreatitis [11, 13].

Diabetic Dyslipidaemia in T1DM

Type 1 diabetic patients with good glycaemic control and without nephropathy do not show quantitative lipid disorders. Plasma triglyceride and LDL cholesterol levels are normal or slightly reduced and HDL cholesterol level is normal or slightly increased. However, several qualitative and potentially atherogenic abnormalities of lipoproteins are observed in patients with T1DM, even in those with good metabolic control. These include increased cholesterol/triglyceride ratio within VLDLs, increased triglyceride content of LDLs and HDLs, augmented oxidation of LDLs, increased small dense LDL particles and reduced ceramides and sphingosine-1-phosphate in HDLs [14]. In the context of hyperglycaemia, glycation of apolipoproteins may occur. All these qualitative changes of lipoproteins are likely to impair their function. HDLs from type 1 diabetic patients show reduced antioxidative and vasorelaxant properties. The reasons for these qualitative abnormalities are not clear. However, some are due to increased CETP activity promoted by peripheral hyperinsulinemia secondary to the subcutaneous route of insulin administration [14].

The cardiovascular benefit of LDL cholesterol reduction with statins is similar in T1DM and T2DM (10). ADA standards of care recommend statin therapy for all T1DM patients over the age of 40 years and in those less than 40 years with additional CV risk factors [11].

For the Future: Questions and Perspectives

Many questions about diabetic dyslipidaemia remain unanswered. These relate to the pathophysiology and the cardiovascular consequences of the qualitative lipid abnormalities, the mechanisms and signalling pathways linking insulin resistance to lipid abnormalities, and the role played by adipose tissue dysfunction and adipocytokines. Additional studies focusing on these areas will be

required to improve our understanding of lipid disorders associated with type 2 diabetes, and these should lead to new specific therapeutic targets.

In the clinical trial arena, the important question of the ability of fibrate to reduce CVD in high-risk diabetic patients with triglycerides ≥2.26 mmol/L (200 mg/dl) and HDL cholesterol ≤1.03 mmol/L (40 mg/dl) will be assessed by the ongoing PROMINENT trial, which is assessing pemafibrate in addition to high-dose statins.

REFERENCES

1. Vergès B. Pathophysiology of diabetic dyslipidaemia: where are we? *Diabetologia*, 2015; **58**: 886–99.
2. Grundy SM. Hypertriglyceridemia, atherogenic dyslipidemia, and the metabolic syndrome. *Am J Cardiol*, 1998; **81**: 18B–25B.
3. Nogueira J-P, Maraninchi M, Béliard S, *et al.* Absence of acute inhibitory effect of insulin on chylomicron production in type 2 diabetes. *Arterioscler Thromb Vasc Biol*, 2012; **32**: 1039–44.
4. Taskinen M-R. Diabetic dyslipidaemia: from basic research to clinical practice. *Diabetologia*, 2003; **46**: 733–49.
5. Vergès B. Abnormal hepatic apolipoprotein B metabolism in type 2 diabetes. *Atherosclerosis*, 2010; **211**: 353–60.
6. Vergès B, Adiels M, Boren J, *et al.* Interrelationships between the kinetics of VLDL subspecies and HDL catabolism in abdominal obesity: a multicenter tracer kinetic study. *J Clin Endocrinol Metab*, 2014; **99**: 4281–90.
7. Vergès B, Petit JM, Duvillard L, *et al.* Adiponectin is an important determinant of apoA-I catabolism. *Arterioscler Thromb Vasc Biol*, 2006; **26**: 1364–9.
8. Costa J, Borges M, David C, Vaz Carneiro A. Efficacy of lipid lowering drug treatment for diabetic and non-diabetic patients: meta-analysis of randomised controlled trials. *BMJ*, 2006; **332**: 1115–24.
9. Cholesterol Treatment Trialists' (CTT) Collaborators, Kearney PM, Blackwell L, *et al.* Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. *Lancet Lond Engl*, 2008; **371**: 117–25.
10. Piepoli MF, Hoes AW, Agewall S, *et al.* 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts). Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J*, 2016; **37**: 2315–81.
11. American Diabetes Association. 9. Cardiovascular disease and risk management: standards of medical care in diabetes—2018. *Diabetes Care*, 2018; **41**(Suppl 1): S86–104.
12. Sacks FM, Carey VJ, Fruchart J-C. Combination lipid therapy in type 2 diabetes. *N Engl J Med*, 2010; **363**: 692–4; author reply 694–5.
13. French Society of Endocrinology (SFE), Francophone Society of Diabetes (SFD), New French Society of Atherosclerosis (NSFA), Béliard S, Bonnet F, Bouhanick B, *et al.* Consensus statement on the management of dyslipidaemias in adults. *Diabetes Metab*, 2016; **42**: 398–408.
14. Vergès B. Dyslipidemia in type 1 diabetes: AMaskedDanger. *Trends Endocrinol Metab.* 2020; **31**: 422–34.
15. Standl E. Statins and beyond: concurrent strategies for prevention of cardiovascular disease in patients with type 2 diabetes. *Diab Vasc Dis Res*, 2013; **10**: 99–114.
16. Turner RC, Millns H, Neil HA, *et al.* Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23). *BMJ*, 1998; **316**: 823–8.
17. Howard BV, Robbins DC, Sievers ML, *et al.* LDL cholesterol as a strong predictor of coronary heart disease in diabetic individuals with insulin resistance and low LDL: the Strong Heart Study. *Arterioscler Thromb Vasc Biol*, 2000; **20**: 830–5.
18. Lehto S, Rönnemaa T, Haffner SM, Pyörälä K, Kallio V, Laakso M. Dyslipidemia and hyperglycemia predict coronary heart disease events in middle-aged patients with NIDDM. *Diabetes*, 1997; **46**: 1354–9.
19. Lee JS, Chang P-Y, Zhang Y, Kizer JR, Best LG, Howard BV. Triglyceride and HDL-C dyslipidemia and risks of coronary heart disease and ischemic stroke by glycemic dysregulation status: the Strong Heart Study. *Diabetes Care*, 2017; **40**: 529–37.
20. West KM, Ahuja MM, Bennett PH, *et al.* The role of circulating glucose and triglyceride concentrations and their interactions with other 'risk factors' as determinants of arterial disease in nine diabetic population samples from the WHO multinational study. *Diabetes Care*, 1983; **6**: 361–9.
21. Fontbonne A, Eschwège E, Cambien F, *et al.* Hypertriglyceridaemia as a risk factor of coronary heart disease mortality in subjects with impaired glucose tolerance or diabetes. Results from the 11-year follow-up of the Paris Prospective Study. *Diabetologia*, 1989; **32**: 300–4.
22. Janka HU. Five-year incidence of major macrovascular complications in diabetes mellitus. *Horm Metab Res Suppl Ser*, 1985; **15**: 15–19.

15.13.5 Hypertension in Diabetes Mellitus

Bryan Williams

Epidemiology and Pathophysiology of Hypertension in Diabetes

Hypertension is very common in people with diabetes. There is moreover an association between hypertension and diabetes that tracks through life, while the blood glucose concentration of young non-diabetic individuals has been shown to predict risk of future

hypertension. Conversely, people with hypertension are twice as likely to develop type 2 diabetes over their lifetime.

In younger people with type 1 diabetes, the excess prevalence of hypertension compared to the non-diabetic population is closely associated with the development of albuminuria. Blood pressure rises by about 2–4 mmHg per year following the onset of microalbuminuria, tracking the increase in urinary albumin excretion. Thus, by the time overt proteinuria is established, hypertension is nearly inevitable.

The situation is different in people with type 2 diabetes, who tend to be older. Hypertension—defined by blood pressures above the arbitrary cut-off of 140/90 mmHg for clinic/office measurements—is at least twice as common as in the age-matched non-diabetic population, affecting approximately 80% of people with type 2 diabetes. It is also important to note that the pathophysiology and characteristics of hypertension are different in people with type 2 diabetes. Although the development of microalbuminuria also heralds the onset of hypertension, by far the commonest cause is an acceleration of the vascular and renal ageing process. Mechanisms include stiffening of large conduit arteries, such as the aorta, principally due to advanced glycation of vascular wall collagen and a reduction in endothelium-dependent vasorelaxation. This reduction in vascular compliance results in a widening of pulse pressure and acceleration of the age-related rise in systolic blood pressure, accompanied by a corresponding fall in diastolic blood pressure. This means that many people with type 2 diabetes develop systolic hypertension at least 10 years earlier than the general population.

Arterial stiffening is important because it renders the systolic blood pressure more resistant to treatment. The vascular changes are compounded by accelerated loss of glomerular filtration rate (GFR) and a shift in the pressure—natriuresis curve to the right. That is, higher pressures are required to generate any given level of natriuresis. Alongside these changes, there is often inappropriate activity of the renin–angiotensin system (RAS) and enhanced renal sympathetic tone, which promote sodium retention. Moreover, the insulin resistance that characterizes type 2 diabetes (and common in people with hypertension) does not extend to insulin's action on the renal tubule, so hyperinsulinaemia promotes further sodium retention. Together, these mechanisms conspire to produce an exquisitely salt-sensitive state. A vicious circle is created whereby volume expansion is poorly accommodated by the reduced vascular compliance, culminating in hypertension (Figure 15.13.5.1). Although separated for the purposes of this discussion, younger people with type 1 diabetes, even if they do not develop nephropathy, ultimately develop the same high risk of hypertension with ageing due to the mechanisms discussed for type 2 diabetes.

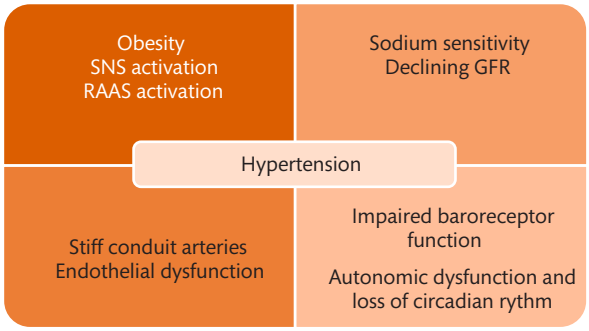


Figure 15.13.5.1 Mechanisms contributing to the pathogenesis of hypertension in people with diabetes. SNS, sympathetic nervous system; RAAS, renin–angiotensin–aldosterone system.

Vulnerability to Hypertensive Complications in Diabetes

Not only is hypertension remarkably common in people with diabetes, but there is also enhanced vulnerability to its cardiovascular and microcirculatory system sequelae [1]. There are two important reasons why people with diabetes are especially vulnerable to pressure-mediated macrovascular and microvascular damage:

1. Soon after onset of diabetes, subtle autonomic dysfunction develops in the majority of people with diabetes, disturbing the circadian rhythm of blood pressure regulation. In particular, there is a blunting of the usual nocturnal dip in blood pressure during sleep. This means that, for any given level of clinic blood pressure, people with type diabetes invariably have an increase in their 24-h blood pressure 'load'. Importantly, this means that 24-h blood pressure load is a more accurate predictor of structural damage to the heart and circulation than measurements in clinic. This explains why the magnitude of left ventricular mass and carotid intima-media thickness often appear disproportionate to clinic blood pressure levels.
2. To compound matters, blood flow autoregulation is impaired in people with diabetes. This means that any increase in circulatory pressure is more readily transferred to the delicate microcirculation, thereby explaining in large part the development of sometimes devastating microvascular disease (Figure 15.13.5.2).

This vulnerability is captured by the European Systematic COronary Risk Evaluation (SCORE) system which classifies the large majority of patients with diabetes and hypertension as being at very high (evidence of organ damage) or high risk of

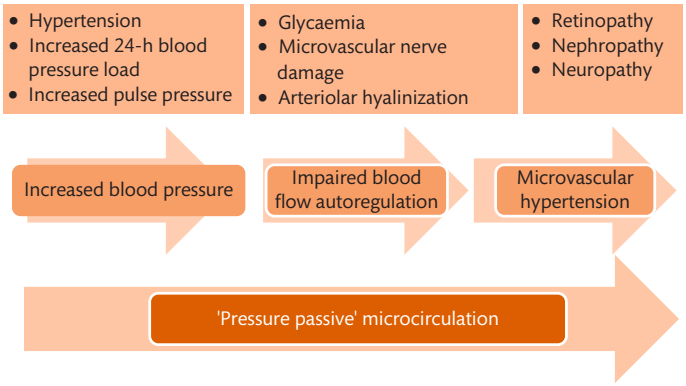


Figure 15.13.5.2 The special vulnerability of the microcirculation to hypertensive injury in people with diabetes. This arises from impairment of blood flow autoregulation, leading to increased pressure transmission to the microcirculation. This results in higher microvascular pressures than would otherwise be encountered for the level of systemic blood pressure, culminating in microvascular injury. Glycaemia is a major factor associated with blunting of autoregulation, compounded by microvascular denervation and poor responsiveness of hyalinized arterioles.

a fatal atherosclerotic event over 10 years [2]. Younger patients with diabetes, or those with no other risk factors may be at moderate risk only. Detection and treatment of hypertension in people with diabetes is thus critical in protection of both the macro- and microvasculature. The combination of increased blood pressure load and impaired microcirculatory defences have also provided the rational for advocating more aggressive blood pressure lowering.

Benefits of Blood Pressure Lowering in People with Diabetes

In patients with type 2 diabetes, there is a continuous relationship between systolic blood pressure and the risk of cardiac, microvascular, and all diabetes-related endpoints (Figure 15.13.5.3). The Framingham cohort study established that the presence of diabetes at least doubles the risk of all cardiovascular events relative to the non-diabetic population with the same levels of blood pressure [3], while the lower cardiovascular risk experienced by premenopausal women appeared to be eliminated by the presence of diabetes.

Since this link between cardiovascular risk and diabetes has been established, numerous randomized clinical trials and meta-analyses have demonstrated that blood pressure lowering reduces the risk of cardiovascular disease, microvascular disease, and premature mortality in people with diabetes. The United Kingdom Prospective Diabetes Study (UKPDS) first illustrated the magnitude of benefit associated with 'more' versus 'less' blood pressure control [4]. A difference in blood pressure achieved in that study of approximately 10/5 mmHg was associated with a reduction in diabetes-related endpoints by a quarter, reductions in death related to diabetes by almost a third, a reduction in stroke by almost half and in heart failure by more than half. Moreover, microvascular disease was reduced by about a third (mainly accounted for by delayed progression of retinopathy). The 'number needed to treat' to prevent one major complication over 10 years was only 6 patients, and to prevent a diabetes-related death was only 15 patients. Subsequent studies have confirmed the impact of blood pressure lowering on major clinical outcomes.

A feature of clinical outcome trials of blood pressure-lowering therapy in people with diabetes is that the treatment benefit is usually greater than anticipated from the epidemiological association between blood pressure and risk. This perhaps reflects the aforementioned enhanced vulnerability of patients with diabetes to pressure-mediated cardiovascular and microvascular disease.

The potency of blood pressure lowering as a strategy to reduce the cardiovascular risk of people with diabetes highlights the importance of early intervention to prevent the evolution of disease. The importance of primary prevention of cardiovascular disease in people with diabetes is underscored by the observation that when people with type 2 diabetes develop coronary heart disease, heart failure, or stroke, their prognosis was worse than that of the non-diabetic population.

Evaluation of Hypertension in People with Diabetes

The aims of evaluation of any patient with hypertension are:

- to establish whether blood pressure is elevated, and, if so, to determine the magnitude of the elevation;
- to determine the presence of any associated target organ damage. This may be manifest as established cardiovascular complications (e.g. ischaemic heart disease, cerebrovascular disease, cardiac

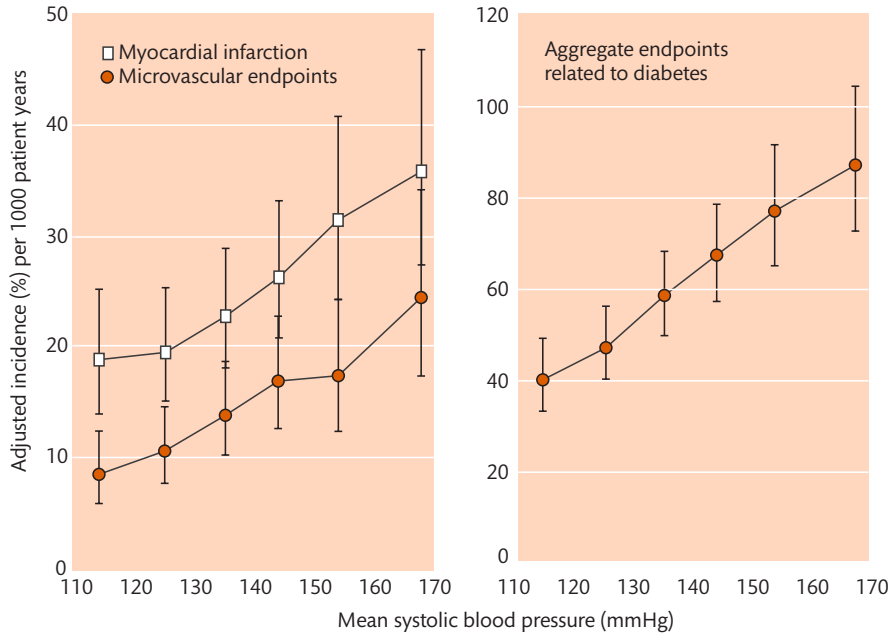


Figure 15.13.5.3 Incidence rates (95% CI) of various endpoints by category of updated mean systolic blood pressure adjusted for age, gender, ethnicity, smoking history, lipid levels, and albuminuria, expressed for white men aged 50–54 at diagnosis of diabetes and a mean duration of diabetes of 10 years in the UKPDS.

failure, peripheral vascular disease) or be subclinical and ascertained by investigation (e.g. retinopathy, renal dysfunction, left ventricular hypertrophy);

- to seek underlying causes of hypertension (e.g. Conn's adenoma, renovascular disease, chronic kidney disease unrelated to diabetes, phaeochromocytoma, Cushing's disease). Although diabetes is commonly associated with hypertension, secondary causes of hypertension must be considered, especially where the blood pressure is refractory to treatment and where there are specific suggestive signs or symptoms. Presentation and investigation of 'secondary' hypertension is beyond the scope of this chapter;
- to assess lifestyle (especially salt intake) or pharmacological (e.g. non-steroidal anti-inflammatory drugs) contributors to blood pressure elevation
- to document concomitant conditions (e.g. asthma, heart failure, ischaemic heart disease) relevant to selection of drug therapy.

Basic Investigations

These include measurement of renal function and electrolyte concentrations, a 12-lead ECG, urinalysis for albumin, urine albumin:creatine ratio (ACR) and blood lipid profile, as well as assessment of glycaemic control.

Blood Pressure Measurement

The method for estimating blood pressure assumes particular importance in people with diabetes. Traditional seated clinic blood pressure measurement has remained the standard approach, however recognition that people with diabetes often have an elevated 24-h blood pressure load and nocturnal hypertension has prompted increased use of 24-h ambulatory blood pressure recording; 24-h ambulatory recording is underused in characterizing the blood pressure elevation at diagnosis and in establishing the quality of blood pressure control on treatment for people with diabetes. Such 24-h ambulatory monitoring can also identify those patients who have either high clinic blood pressure measurements but normal blood pressure outside this environment ('white coat' hypertension) and those with normal clinic blood pressure measurements but high 24-h blood pressure levels ('masked' hypertension). Both variants of hypertension confer increased cardiovascular risk, with masked hypertension increasing risk of cardiovascular and renal events in diabetes especially where blood pressure is elevated nocturnally [5].

Home blood pressure monitoring (HBPM) is increasingly being validated as another useful clinical tool in assessment of blood pressure. It should be performed using a semiautomated, validated monitor, for 3–7 days prior to a clinic visit, and be undertaken morning and evening in a quiet room after 5 minutes of rest. It has shown more reproducible measurements than clinic blood pressure determinations, and these show a stronger relationship to hypertension-mediated end organ disease [6] and perhaps cardiovascular morbidity and mortality [7].

It should be noted that the normal ranges for 24-h ambulatory daytime and night-time averages, and for home blood pressure recordings, are substantially lower than those measured in clinic. That said, the normal ranges for 24-h ambulatory blood pressure have not been formally defined for people with diabetes and it is recommended that the ranges identified for the general population should be used for people with diabetes (Table 15.13.5.1).

Table 15.13.5.1 Definitions of hypertension according to the method of measurement (mmHg); similar values should apply to people with diabetes

	Systolic blood pressure	Diastolic blood pressure
Office or clinic	≥ 140 AND/OR	≥ 90
24-h average	≥ 130 AND/OR	≥ 80
Day	≥ 135 AND/OR	≥ 85
Night	≥ 120 AND/OR	≥ 70
Home	≥ 135 AND/OR	≥ 85

24 h, day, and night refer to the averages for these time periods as measured by ambulatory blood pressure monitoring. The home average refers to an average of readings, usually twice daily, over 7 days. If a lower blood pressure threshold for intervention and/or target for treatment is adopted, then the values for 24-h and home blood pressure averages will be lower, in proportion to the threshold/target adjustment.

Two additional caveats are important when measuring blood pressure in people with diabetes: (1) lying and standing blood pressure should be recorded at diagnosis and to assess the impact of treatment because of the potential of autonomic dysfunction to induce significant postural hypotension and supine hypertension; and (2) the correct cuff size must be used to measure blood pressure, either in the clinic, at home, or when using ambulatory devices. The use of blood pressure cuffs that are too small for the arm circumference (e.g. in people with obesity) leads to overestimation of blood pressure readings. Comprehensive information on blood pressure measurement, validated monitors and appropriate cuff sizes is available at British Hypertension Society website (https://www.bhsoc.org).

Treatment Thresholds for Hypertension in Diabetes

Blood pressure is a continuous variable and has a continuous relationship with cardiovascular disease risk, so 'cut offs' are arbitrary, although pragmatic and important in guiding practice. A 'normal' blood pressure is commonly taken to be less than 120/80 mmHg, but cardiovascular risk begins to rise when systolic blood pressure increases to greater than 115 mmHg. Treatment guidelines are based on data from clinical intervention trials, however, and, for the most part, these have not enrolled patients with blood pressure values less than 140/90 mmHg. Thus, the definition of hypertension and the threshold for intervention with blood pressure-lowering drugs for people with diabetes has been defined as a blood pressure of 140/90 mmHg or higher. This value is based on clinic blood pressure and lower thresholds apply for 24-h ambulatory blood pressure averages (Table 15.13.5.1).

Some guidelines over the past decade have recommended intervention with blood pressure-lowering medications at a lower threshold (i.e. 130–139/85–90 mmHg). For those with microalbuminuria or evidence of target organ damage and/or microvascular or macrovascular disease it could be argued that lower blood pressure thresholds for intervention with blood pressure-lowering drugs are justified because such patients already exhibit evidence of pressure-mediated damage. This illustrates the conflict between generalized evidence-based guidance and individualized patient care. Evidence supporting lower blood pressure intervention thresholds for those with manifest blood pressure-related disease is sparse, however

the approach is supported by a strong pathophysiological rational. Further clinical trials are required to conclusively demonstrate the benefits of initiating blood pressure-lowering therapy in those at high risk with blood pressure in the high normal range (i.e. 130–139/80–89 mmHg). It is thus suggested that this approach be considered in recent guidelines [8].

Blood Pressure Treatment Targets

Most available evidence on cardiovascular endpoints in people with diabetes relates to type 2 diabetes. Blood pressure targets for those with diabetes and treated hypertension have been contentious, with previous recommendations often ahead of the evidence from clinical trials. Important uncertainty remains, but the evidence guiding practice has solidified somewhat in recent years, in the form of randomized controlled trials, meta-analyses, and post-hoc subgroup analyses of controlled trials [8, 9].

There is a broad evidence base for a blood pressure target of less than 140/85 mmHg for all people with treated hypertension and diabetes. International guidance previously suggested that blood pressure targets for people with diabetes should be lower than those advocated for the treatment of hypertension in general, with an optimal treatment target of less than 130/80 mmHg recommended, especially for patients with manifest cardiovascular disease and/or nephropathy. However, no major trial of blood pressure lowering in diabetes had achieved such low levels of blood pressure (Figure 15.13.5.4). A preponderance of more recent evidence from trials, meta-analyses and post-hoc analyses of trials now suggests that targeting systolic blood pressures towards the lower end of the 130–130 mmHg is beneficial, and this is now enshrined in most guidelines (e.g. [8]).

Whether systolic blood pressure targets lower than 130 mmHg are appropriate is a matter of much greater debate [9]. There is no outcome evidence that lowering systolic blood pressure below 120 mmHg is beneficial, and indeed the ACCORD study suggested this may increase risk [10], and so this should be avoided. The evidence for cardiovascular benefits of targets between 120 and 130 is mixed, however. There is no conclusive evidence that lowering blood pressure to these levels improves general cardiovascular risk, especially where glycaemia is optimally controlled, however consistent evidence has emerged from several trials and meta-analyses that risk of stroke is reduced [9]. Targets in this range may thus be considered where risk of stroke is considered particularly high. Regarding diastolic blood pressure, current evidence suggests that a target of 80 mmHg, is safe, with little evidence on which to base a judgement about lower levels.

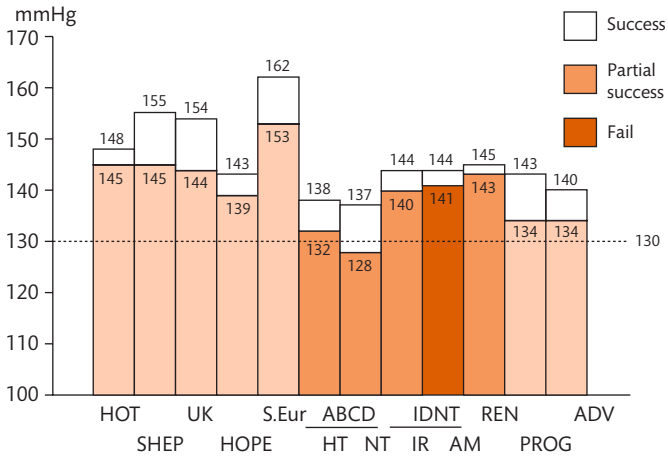


Figure 15.13.5.4 Clinical outcomes in blood pressure-lowering studies of people with diabetes; 130 refers to the recommended blood pressure target (mmHg) in most guidelines. Success refers to a positive clinical outcome in the study; partial success refers to positive outcomes in some endpoints, but not all; fail refers to no evidence of benefit. The various trials are annotated on the axis.

Blood Pressure Treatment Strategies

The ultimate objective of treating blood pressure is not only to lower the blood pressure but also to safely minimize cardiovascular and microvascular risk. This is best achieved through a combination of lifestyle intervention, blood pressure-lowering medication, and concomitant use of other medications to reduce cardiovascular risk, notably statins.

Lifestyle Interventions

These are important and should be implemented early. Lifestyle intervention can lower blood pressure and also reduce cardiovascular risk. Recommended lifestyle interventions for people with diabetes and hypertension are shown in Box 15.13.5.1. With regard to blood pressure reduction, one of the most effective lifestyle interventions for people with diabetes is restriction of dietary sodium intake. This is important because the pathophysiology of hypertension in people with diabetes points to a 'salt-sensitive state'. Moreover, most people with diabetes and treated hypertension will receive medication to inhibit the activity of the RAS, and sodium restriction increases the effectiveness of these medications with regard to blood pressure-lowering and the reduction of albuminuria. Other lifestyle interventions such as increasing intake of fresh fruit and vegetables, reduction in saturated fat intake, increased regular exercise, weight reduction, and moderation of alcohol intake, can all contribute to blood pressure lowering, as well as improving cardiovascular health.

Box 15.13.5.1 Lifestyle measures suggested for people with diabetes and high blood pressure

Lifestyle measures that lower blood pressure:
- Weight reduction
- Reduced salt intake
- Limitation of alcohol consumption
- Increased physical activity
- Increased fruit and vegetable consumption
- Reduced total fat and saturated fat intake

Measures to reduce cardiovascular disease risk
- Cessation of smoking
- Reduced total fat and saturated fat intake
- Replacement of saturated fats with monounsaturated fats
- Increased oily fish consumption

Blood Pressure-Lowering Medication

Medication is invariably required to lower blood pressure in most patients. Patients with diabetes often develop systolic hypertension and this can be particularly resistant to treatment. There has been much debate about the relative merits of specific medications with regard to cardiovascular and renal protection in people with hypertension and diabetes. What is beyond dispute is that the most important driver of benefit is the magnitude of blood pressure lowering. This must be the first objective.

Data from clinical trials have suggested that blockade of the RAS 'adds value' over and above the blood pressure lowering they produce in preventing cardiovascular and renal events in people with diabetes, although this has not been confirmed by all studies. There appears to be a clear advantage of RAS blockade in people with diabetes on key renal endpoints (albuminuria and progression of renal disease), however, which has prompted international guidelines to conclude that RAS blockade should be part of the cocktail of treatments used to lower blood pressure in people with diabetes.

It is emphasized that most people with diabetes will need two or more medications to control their blood pressure. Thus, modern treatment strategies would usually include RAS blockade along with a calcium channel blocker (CCB) and/or a thiazide-type diuretic, or all three, ideally in a single pill [8]. This is based on the higher cardiovascular risk conferred by hypertension in the context of diabetes, and the challenges in controlling hypertension in diabetes. Recent trials have suggested that sodium glucose cotransporter 2 inhibitors, used primarily for their hyperglycaemic effect, also lower clinic and ambulatory blood pressure by several mmHg, even in those concomitantly treated with antihypertensive drugs [11, 12]. This may help improve blood pressure control in people with diabetes, which is often challenging. Of note, ACE inhibitors and angiotensin receptor antagonists should not be combined, as was once suggested, due to trial evidence that this confers unacceptable excess risk of renal side effects without outcome benefit [13].

Concerns About Some Blood Pressure-Lowering Medications in Diabetes

In the 1990s, there was concern that CCBs, especially the dihydropyridines (e.g. nifedipine, amlodipine), might be less effective at protecting against cardiovascular and renal events when compared to alternative treatments. This is worthy of comment because such concern still lingers. This concern was surprising because the CCBs are arguably the most effective blood pressure-lowering agents in people with diabetes. It arose from post-hoc analyses of trials and case–control studies and was ill-founded. However, subsequent large-scale, prospective trials have conclusively demonstrated the effectiveness, safety, and importance of CCBs for the management of hypertension in people with diabetes.

There has also been longer standing concern about the use of thiazide-type diuretics for people with diabetes because of their potential to adversely impact on glucose homoeostasis. This in part reflects the glucocentric perspective of the clinical management of diabetes. Once again, data from large-scale trials have demonstrated that thiazide-type diuretics are very effective at reducing blood pressure and cardiovascular events in people with diabetes. Moreover, as sodium retention plays such a fundamental part in the pathogenesis of hypertension in diabetes, thiazide-type diuretics are often essential to control blood pressure in people with diabetes.

Finally, β-blockers have slipped down the hierarchy of prescribing for hypertension because they appear less effective at reducing cardiovascular events, especially stroke, when compared to aforementioned blood pressure-lowering medications. However, β-blockers would still be used where there is a specific indication for their use, i.e. symptomatic angina, post myocardial infarction, and in people with chronic heart failure.

An Algorithm for Use of Blood Pressure Lowering Medications in Diabetes

As indicated earlier, most guidelines, including prominent recent revisions [8] recommend the inclusion of RAS blockade (i.e. an ACE-inhibitor or angiotensin receptor blocker) as part of the treatment strategy for people with hypertension and diabetes. For many, RAS blockade alone will not be sufficient to control blood pressure and the addition of a CCB or thiazide-type diuretic will be required to improve blood pressure control [8]. The fact that two or more medications is almost inevitable to control blood pressure for most people with diabetes, has led to strengthening of the view that initial therapy for hypertension in people with diabetes should be with a combination of two medications, ideally combined in a single pill to reduce medication burden and increase compliance.

Which combination of blood pressure-lowering medications is preferred remains unclear, despite a growing number of randomized trials, summarized in recent guidelines [8]. The ACCOMPLISH study from the United States, for example, compared the combination of an ACE-inhibitor with either a CCB or thiazide-type diuretic in a hypertensive population in which a high proportion had diabetes [14]. This study suggested that despite similar blood pressure control, those treated with the ACE + CCB combination experienced fewer cardiovascular events than those treated with the ACE + thiazide combination. In reality, however, there is no single combination that is preferred or tolerated by all patients and thus choice is needed. The initial preferred choice is; RAS blockade + CCB, or RAS blockade + thiazide. For those not controlled with two drugs, RAS blockade + CCB + thiazide would be the most logical next step (Figure 15.13.5.5).

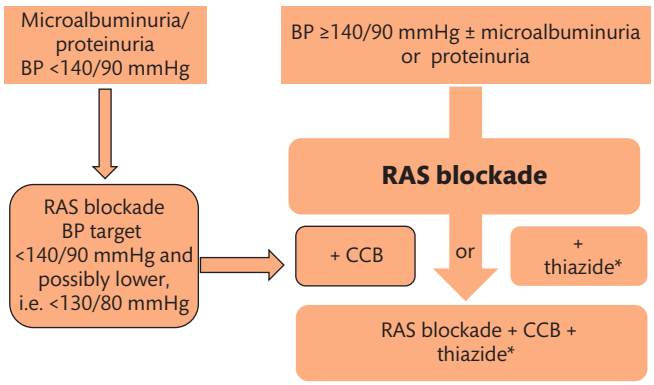


Figure 15.13.5.5 A suggested treatment algorithm for hypertension in people with diabetes. Guidelines concur that RAS blockade would usually be the foundation for treatment with a common requirement for additional therapy, i.e. a dihydropyridine CCB, or a thiazide-type diuretic, or all three in those with more resistant hypertension.

Resistant Hypertension

The blood pressure of some patients remains uncontrolled despite treatment with optimal doses of three medications. This is described as resistant hypertension [15]. This is more common in people with people with diabetes, in whom systolic blood pressure is often the most difficult to control. Specific causes of secondary resistant hypertension to be considered in diabetes include progressive nephropathy, obstructive sleep apnoea, atherosclerotic renovascular disease, certain medications or, more rarely, Cushing's syndrome or phaeochromocytoma, which also both often cause diabetes.

Treatment options for resistant hypertension have been poorly evaluated. Sodium restriction is a valuable adjunct to pharmacotherapy. Pharmacological options include further increasing the potency or dose of thiazide-type diuretic, or use of an additional diuretic such as low-dose spironolactone, or a loop diuretic at low eGFR. Careful monitoring of potassium and sodium are required. Other options are β-blockade or α-blockade and selective endothelin receptor antagonism. Referral to a specialist hypertension centre should be considered in patients in whom a secondary cause of hypertension is considered likely and for patients in whom blood pressure is resistant to treatment.

Treating Blood Pressure and Wider Cardiovascular Risk in Diabetes

People with diabetes and hypertension are at especially high risk of premature cardiovascular disease. It therefore seems illogical not to optimize their risk reduction by considering the use of statin therapy alongside blood pressure-lowering medications. The Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) used a factorial design to demonstrate that the addition of statin therapy to people with hypertension (many with diabetes) resulted in a reduction of ischaemic heart disease related events by over a third and stroke events by almost a quarter [16]. This was in addition to the benefits already accrued by blood pressure lowering. Moreover, the relative risk reduction was consistent irrespective of baseline cholesterol levels. This and other similar observations from the Heart Protection Study and CARDS supports the philosophy that if a decision has been taken to reduce blood pressure with medication, then the addition of statin therapy should be the rule rather than the exception for people with diabetes.

REFERENCES

1. Williams B. The unique vulnerability of diabetic subjects to hypertensive injury. In: Williams B, ed. *Hypertension in Diabetes*. London: Martin Dunitz, 2003: 99–108.
2. Piepoli MF, Hoes AW, Agewall S, *et al.* 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts). Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J*, 2016; **37**: 2315–81.
3. Garcia MJ, McNamara PM, Gordon T, Kannel WB. Morbidity and mortality in diabetics in the Framingham population. Sixteen year follow-up study. *Diabetes*, 1974; **23**: 105–11.
4. Adler AI, Stratton IM, Neil HA, Yudkin JS, Matthews DR, Cull CA, *et al.* Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. *BMJ*, 2000; **321**: 412–19.
5. Franklin SS, Thijs L, Li Y, *et al.* Response to masked hypertension in untreated and treated patients with diabetes mellitus: attractive but questionable interpretations and response to: is masked hypertension related to diabetes mellitus? *Hypertension*, 2013; **62**: e23–5.
6. Bliziotis IA, Destounis A, Stergiou GS. Home versus ambulatory and office blood pressure in predicting target organ damage in hypertension: a systematic review and meta-analysis. *J Hypertens*, 2012; **30**: 1289–99.
7. Ward AM, Takahashi O, Stevens R, Heneghan C. Home measurement of blood pressure and cardiovascular disease: systematic review and meta-analysis of prospective studies. *J Hypertens*, 2012; **30**: 449–56.
8. Williams B, Mancia G, Spiering W, *et al.* 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J*, 2018; **39**: 3021–104.
9. Mancia G, Grassi G. Blood pressure targets in type 2 diabetes. Evidence against or in favour of an aggressive approach. *Diabetologia*, 2018; **61**: 517–25.
10. Group AS, Cushman WC, Evans GW, *et al.* Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med*, 2010; 362: 1575–85.
11. Zinman B, Wanner C, Lachin JM, *et al.* Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med*, 2015; **373**: 2117–28.
12. Neal B, Perkovic V, Mahaffey KW, *et al.* Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*, 2017; **377**: 644–57.
13. Investigators O, Yusuf S, Teo KK, *et al.* Telmisartan, ramipril, or both in patients at high risk for vascular events. *N Engl J Med*, 2008; **358**: 1547–59.
14. Jamerson K, Weber MA, Bakris GL, *et al.* Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. *N Engl J Med*, 2008; **359**: 2417–28.
15. Williams B. Resistant hypertension: an unmet treatment need. *Lancet*, 2009; **374**: 1396–8.
16. Sever PS, Poulter NR, Dahlof B, *et al.* Reduction in cardiovascular events with atorvastatin in 2,532 patients with type 2 diabetes: Anglo-Scandinavian Cardiac Outcomes Trial—lipid-lowering arm (ASCOT-LLA). *Diabetes Care*, 2005; **28**: 1151–7.

15.14

The Diabetic Foot

15.14.1 Modern Management of Diabetes-Related Foot Disease

Frank Lee Bowling and Andrew J.M. Boulton

Introduction

Diabetic foot problems are preventable but still account for a high number of hospital admissions in developed countries. In 2005, The International Diabetes Federation (IDF) focused on the diabetic foot. The 'Put Feet First' campaign highlighted the high prevalence of amputations within the diabetic population worldwide. Parallel to this World Diabetes Day launch, *The Lancet* dedicated a large proportion of its issue to the diabetic foot: a first for any major non-specialist journal focusing on this global challenge [1].

The economic burden both nationally and internationally can be extremely high. UK data suggest an annual cost of over $749 million for diabetic foot complications [2]. The financial constraints are influenced by ulcer type as illustrated in the Eurodiale study [3] which reported an average spend of $13 000 for treating non-infected foot ulcers increasing to $18 000 for infected ulcers with concurrent peripheral arterial disease. A grossly infected lesion can be predicted to incur higher costs. Calculations that incorporate failed antibiotic regimens, inpatient admissions for intravenous antibiotics, limb salvage and/or major amputations with associated aftercare generated sums in excesses of $188 645 [4].

In many Western countries, neuroischaemic lesions account for the greatest majority and represent a major challenge both medically and surgically. Routine podiatric diabetic foot care in high-risk patients was emphasized by a recent observational study from Arizona where the state cut podiatric care as a cost-saving measure. This led to an annual saving of $351 000 but the cost of this action measured by increased hospitalization, length of stay, and amputations was $16.7 million per annum [5].

Armstrong *et al.* concluded that 34% of people with diabetes will develop a foot ulcer during the course of their life [6]; thus the overall treatment and healthcare management of diabetic foot complications can be sustained and time-consuming requiring the involvement of a multidisciplinary team (MDT) dedicated to foot health. It has recently been suggested that those with a previous foot ulcer history should be described as being in 'remission' rather than 'healed', emphasizing the seriousness of a foot ulcer history [7]. A non-healing lesion complicated by gross infection is estimated to precede 85% of all major lower limb amputations [8].

Epidemiology of Diabetic Foot Problems

In the United Kingdom, the annual incidence and prevalence of foot ulceration in patients with diabetes was calculated at 2.2% and 1.7%, respectively, in 2002 [9]. Later, a European multicentre study on Diabetes and the Lower Extremity (Eurodiale) [3] followed 1232 diabetes patients with a foot ulcer for 12 months, and found that 5% of these underwent a major amputation (above or below knee) during the follow-up period. Krishnan *et al.* [10] reported

an amputation rate of 16.5 per 10 000 people with diabetes in the United Kingdom. Data extracted from the General Physicians databases in Scotland identified that 2.5% of the diagnosed diabetes population had an active foot ulcer at the beginning of December 2010 [11]. Diabetic foot disease is associated with a risk of amputation 23 times that of a person without diabetes [12].

There are few databases that capture diabetic foot ulceration as a distinct entity but Diabetes UK used data from the Public Health Observatory and National Diabetes In-patient Audit [13], to estimate the cost of inpatient care for complicated diabetic foot ulcers (DFUs). For the period 2010 to 2011 expenditure was UK £219 million ($285 million) [14].

Prevalence of Major Amputations

The overall rate of major amputation in England in the general adult population from studies published over the last 15 years was approximately 5/100 000. The rate was approximately eight times higher in those with diabetes than those without while the rate for males was three times higher than for females [15]. In the absence of diabetes, amputation rates were higher in the black British versus white-British population; however, black British people with diabetes had a 50% lower rate of amputation compared with their white-British counterparts.

Vamos *et al.* [16] published two studies describing the prevalence of major amputation in England. The first described rates in 1996 and 2005 based on hospital data and found the overall rate in England to have decreased from 7.0 to 4.9/100 000. A subsequent study using similar methodology describing major amputation rates from 2004 to 2008 found the overall rate to have decreased from 7.7 in 2004 to 6.9/100 000 in 2008 [17]. Moxey *et al.* [18] described the major amputation rate in England between 2003–2008 using the same data and found it to be 5.1/100 000.

Prevalence of Major Amputation by Diabetic Status

Rayman *et al.* [19] compared the rates for major amputations in patients with and without diabetes between 1997 and 2000 in the south-east of England. At the end of the study period those with diabetes had an amputation rate of 108/100 000 compared to 3.0/100 000 for those without. McCaslin *et al.* [20], however, described the age-specific major amputation rate by diabetes status across the whole of England in 2004. Surprisingly, the rate was much lower for diabetes where the rates, per 100 000, for the age ranges 45–64, 65–74, and 75–84 were 1.9 (11.5), 5.0 (30.9), and 6.5 (51.5).

Canavan *et al.* [21] described a 5-year reduction of amputation by diabetes status between 1995 and 2000 in the north-east of England. At the end of the study period the rate was 75.8 for people with diabetes compared to 15.3/100 000 for those without diabetes.

Vamos *et al.* described the amputation rate by diabetes status in 1996 and 2005 using national data [16]. During the 9 year intervening period, the rate of amputation in type 1 diabetes reduced from 1.3 to 0.7/100 000 in contrast to the increased rate of amputation observed for type 2 diabetes rising from 2.0 to 2.7/100 000. Again the rate in the general population was similar to McCaslin *et al.* (i.e. higher in the non-diabetic population but decreased over the same period from 7.0 to 4.9/100 000). A further study by Vamos *et al.* reported amputation rates for the period 2004 to 2008 in adults over 17 years of age across England [17]. The rate for diabetes in 2008 was over ten times higher than the rate for people without diabetes at 102 versus 6.9/100 000, respectively.

Abbott *et al.* [22] reported rates for lower extremity amputation based on diabetes patients attending a specialist clinic in Manchester. Rates were 1.3%, 0.6%, and 1.4% for white-British, south Asian, and black British attendees, respectively.

Diabetic Peripheral Neuropathy (DPN)

The late sequelae of diabetic neuropathy (DPN) include foot ulceration, Charcot neuroarthropathy (CN), and amputation: likewise, peripheral vascular disease (PVD) is a major aetiological factor in diabetic foot disease. DPN increases the risk of foot ulceration through the loss of protective sensation by a sevenfold increase when comparing to the non-diabetic population [23]. The resultant soft tissue trauma results in the development of a diabetic foot lesion; in many cases the patient is unaware and continues to mobilize in some cases for days or even weeks. DPN is also inextricably linked to an increased risk of falls by a marked alteration in gait and foot architecture [24].

Distal Symmetrical Sensorimotor Peripheral Neuropathy

Sensorimotor peripheral neuropathy can be divided into painful and painless via the symptoms, although both may occur simultaneously. Symptoms include reduced or absent sensation, burning, tingling, stabbing sensation, pain, paraesthesia, and patients can also report the sensation of walking on marbles. In contrast, numbness, heaviness or in some cases asymptomatic, profound sensory loss renders the foot 'at risk'. Up to 50% may not experience any type of symptomatology at all [25]; thus careful examination of both feet is paramount.

Peripheral Sympathetic Autonomic Neuropathy

In some cases, autonomic dysfunction can manifest as anhidrosis predisposing the foot to callus and fissure formation beneath weight-bearing structures [26]. Sympathetic vasoconstrictor tone resulting in arteriovenous shunting and a warm well-perfused foot result in 'auto-sympathectomy'. Distended dorsal foot veins remain distended on elevation. Thus it is the warm, insensate, and painless foot that is the 'at-risk foot'.

Other Long-Term Risk Factors for Foot Ulceration

Diabetic foot lesions remain a major cause of morbidity in patients with renal failure and even in the early stages of renal disease microalbuminuria is a strong predictor of foot ulceration [27]. Patients who undergo simultaneous pancreas and kidney transplants (SPK) may have normoglycemia with near normal renal output but are still at high risk of ulceration and must be monitored long term. As their overall health status improves and activity levels

increase, reports of foot ulcers and even acute CN some years after the SPK have been reported [28].

Structural changes within the foot/ankle render the foot vulnerable to areas of high pressures, associated digital flexion deformities along with migration of the plantar fat pad distally increases the risk of tissue breakdown.

A number of studies have examined the role of psychosocial factors in the pathway to foot ulceration. They have demonstrated that patients' behaviour is not driven by the abstract designation of being 'at risk': it is more that the patients' perception of their risk [29]. A recent prospective study confirmed that depression predicts first, although not recurrent, DFUs [30].

Peripheral Vascular Disease

Critical Limb Ischaemia–Risk Factors and Epidemiology

Critical limb ischaemia (CLI) is a late complication of PVD with a European prevalence of 500–1000 per million [31]. It is defined as stage 3 in the Rutherford Classification [32]. The progression from intermittent claudication to CLI is related to diabetes (fourfold risk), smoking (threefold risk) and hypercholesterolaemia (twofold risk) [33]. Approximately half of all patients with CLI have some form of revascularization with 40% of those with no reconstructable disease undergoing amputation within 6 months and 20% dying within the same period [34].

The influence of different risk factors on the progression of PAD into CLI was described by Dormandy *et al.* [35] and is illustrated in Figure 15.14.1.1.

Identification of the High-Risk Foot

An annual foot examination by a clinician should occur in tandem with patient engagement (i.e. patients should be actively engaged through regular monitoring of their own feet). The foot should always be an integral part of the annual review. The clinician should never rely on symptoms alone to identify high-risk patients; 50% of patients with insensitive feet have no previous history of neuropathic symptoms, and claudication [36].

Assessment

Up to 50% of older, type 2 diabetic patients have signs of DPN [37]. The American Diabetes Association (ADA) document on the 'Comprehensive Diabetic Foot Examination (CDFE)' [37] provides a robust assessment tool.

Examination is the 'key' component of foot screening but needs to be placed in context via a thorough history that identifies specific risk factors.

History

- Past or present neuropathic symptoms
- Vascular (intermittent claudication/rest pain/past history of bypass surgery or angioplasty)
- History of ulcer or minor/major amputation
- Social factors (living alone, smoking)
- Visual impairment or end-stage renal failure (dialysis or post-transplant)

Clinical Examination

- Skin colour, callus, fissures, reduced sweating
- Bacterial/fungal infection
- Ulceration?
- Architecture/structural alterations, claw toes, prominent metatarsal heads
- Anhidrosis
- Skin temperature; a unilateral, warm, insensate foot should be considered to be acute CN until proven otherwise
- Footwear suitability

Neurological Assessment

The 10 g monofilament (Bailey Instruments Ltd, Manchester, UK) is widely used in clinical practice and designed to buckle when a 10 g force is applied. Testing at the first, third and fifth metatarsal heads along with plantar surface of hallux (a simple yes or no if a stimulus has been detected) is a reliable and highly accurate predictor of foot ulceration [38].

The Ipswich Touch Test requires a light touch via the index finger onto the 1st, 3rd, and 5th digits. A loss of sensation, neuropathy, is identified when detection fails at two or more sites (out of the total

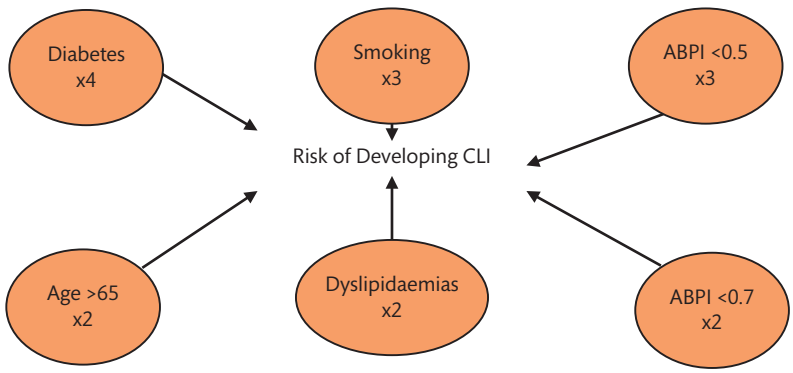


Figure 15.14.1.1 The influence of different risk factors on the risk of developing critical limb ischaemia.

six). This highly simplistic test demonstrates strong agreement with other validated tests such as the monofilament [39].

The Vibratip™, (McCallan Medical, Nottingham, UK) is a small, battery operated, disposable vibrating stylus that assesses vibration sensation. Levels of agreement with other, similar tests are highly positive [40].

The Vibration Perception Testing (VPT) is a hand-held device that generates a vibratory stimulus when applied to the hallux. VPT has demonstrated excellent sensitivity and specificity for neuropathy.

A 128 Hz tuning fork is a traditional method of assessing vibration perception when placed over the apices of the hallux bilaterally. The cold metal can also provide additional stimuli in the form of temperature.

A pin-prick test is a simple test sensation over the apex of the halluces. Ankle reflex testing is a standard component of neurological testing whereby the absence of ankle reflexes bilaterally is an abnormal response [41].

Vascular Assessment

Classification of Peripheral Arterial Disease

Peripheral arterial disease (PAD) is characterized by progressive arterial occlusive disease of the lower extremities. The worldwide prevalence has been estimated at 3–10% increasing to 15–20% in individuals over 70 [42]. The ratio of symptomatic to asymptomatic disease is up to one in three with as many as 50% never consulting a clinician. Those with PAD have a threefold increase in risk of mortality from major cardiovascular events (heart attack and stroke) compared to those without PAD [43].

It can be diagnosed based on symptoms—classically intermittent claudication (cramping pain in legs on walking) and signs such as absent peripheral pulses, ulcers, gangrene, and amputation. Non-invasive diagnosis can also be made using the ankle-brachial pressure index. This is a ratio of the blood pressure taken at the arm and the leg (at the level of the malleoli). A ratio of <0.9 is 95% sensitive and 100% specific for detecting angiogram positive disease [44].

PAD is classified by the Fontaine and Rutherford classification system [32] and ranges from asymptomatic, through to claudication and finally tissue loss and gangrene.

Screening for vascular disease can be difficult. Palpation for posterior tibial and dorsalis pedis pulses is important but detection can be influenced by the skill and room temperature. A femoral bruit can also be a strong indicator of PVD. Doppler ultrasound can be useful to assess flow signal waveforms although vessel wall calcification can lead to a falsely elevated reading of the ankle-brachial index (ABI).

The High-Risk Patient

If anomalies identified and/or relevant clinical history place the foot 'at risk', strategies aimed at-risk management must be implemented. Education promotes self-monitoring and foot hygiene; however, this needs to be supported by regular clinical reviews by the multidisciplinary team.

Wound Classification

A number of classification systems are currently in existence but all follow a similar theme: size, depth, appearance, and location. These provide the clinician with a good reference point for wound surveillance (Box 15.14.1.1). In this system, grades refer to the depth of the wound, and each grade has four stages, depending on the presence or absence of infection and/or ischaemia.

Box 15.14.1.1 Wound classification

The 'University of Texas Diabetic Wound Classification' is regularly used for staging diabetic foot ulcers. This classification grades and stages ulcers by their depth and the presence of any infection or ischaemia.

- Staging
  - **A:** No infection or ischaemia
  - **B:** Infection present
  - **C:** Ischaemia present
  - **D:** Infection and ischaemia present
- Grading
  - **0:** Epithelialized wound
  - **1:** Superficial wound
  - **2:** Wound penetrates to tendon or capsule
  - **3:** Wound penetrates to bone or joint

Grade 0

A UT grade 0 foot has no open lesions, but is at risk. The patient may have a history of foot ulcers or pre-ulcer lesions.

Grade 1

Superficial full-thickness loss: these lesions tend to be predominantly neuropathic with (UT 1B) or without (UT 1A) infection. Lesions commonly occur under points of high-pressure areas (e.g. the metatarsal heads, toes). The presence of ischaemia should be confirmed examination.

Grade 2

Deeper seated lesions that penetrate into the subcutaneous tissue, exposing tendon and/or capsule but without bony involvement.

Grade 3

Lesions penetrating to bone and/or joints. The ability to probe to bone has been shown to be sensitive for identifying osteomyelitis.

Evidence shows that UT grade 3B lesions may respond to antibiotic treatment without the need for surgical intervention. A recent randomized comparative trial (RCT) of antibiotics versus surgery for localized bone infection yielded similar outcomes in both groups [45]. A recent interim analysis of a large multicentre UK RCT of Oral Versus Intravenous Antibiotics (OVIVA) over a 6-week period for localized joint and bone infection showed oral to be inferior to IV antibiotics. It should be emphasized that only 16% of the study population had a diagnosis of diabetes mellitus [46].

Management of Diabetic Foot Ulceration

Neuropathic Ulcers

The mainstay of management of uncomplicated lesions consists of debridement of non-viable tissues along with an appropriate wound dressing; very little evidence exists for any one type of wound dressing over others and the decision in clinical practice has generally been dictated by the clinician's personal experience.

The EXPLORER trial is the first clinical study to indicate the efficacy of a dressing in the treatment of diabetic foot lesions [47]. The

double-blind RCT involved teams in 43 hospitals with specialized diabetic foot clinics in France, Spain, Italy, Germany, and the UK. The aim was to assess the effect of a sucrose octasulfate dressing versus a control dressing on wound closure in patients with neuroischaemic lesions. Ulcers were deemed neuroischaemic if peripheral neuropathy/PAD were both present. Over a period of three years, 240 individuals took part in the trial. Of these, 126 were treated with the sucrose octasulfate dressing under investigation and 114 with a control dressing of a type usually used to dress this type of wound.

After 20 weeks, 18% more wound closures were recorded in the sucrose octasulfate dressing group, compared to the control dressing group and there was an observed 30% improvement in healing. The expected time to wound closure for all participants—as determined by Kaplan–Meier analysis—was 180 days in the control group and only 120 days in the sucrose octasulfate dressing group. The results suggest a significant improvement in wound healing resulting from treatment with the sucrose octasulfate dressing. This is the first study to demonstrate proven efficacy of a dressing in wound healing.

Debridement

The formation of hyperkeratotic tissue (Callus) results from sheer pressure and removal reduces abnormally high plantar pressures. Lesions that show extensive bone and soft tissue involvement require aggressive debridement providing drainage of purulent discharge. Surgical removal/excision has been shown to significantly reduce time to healing when compared to conservative management [48].

Offloading

This is the key to healing plantar foot lesions, and outcomes are positive when offloading devices are used. Total contact casts (TCC) are the gold standard based on evidence of a 90% success rate for ulcer healing, supported by several RCTs [49–51]. Other devices, removable cast walkers, or adapted footwear, to date have not been successful in demonstrating similar results. Armstrong *et al.* [49] recorded the activity levels of patients prescribed removable cast walkers as treatment for neuropathic foot ulcers. Findings demonstrated that patients only wore the offloading device for 28% of their total daily activity. Further randomized controlled trials were carried out in Miami and Tucson. In Miami, patients with non-infected neuropathic plantar ulcers were randomized either to a total contact cast, or a removable cast walker, rendered irremovable by wrapping with a sheet of cast material. Not surprisingly, there were no differences in healing rates which were generally rapid [50]. By contrast, in the Tucson study, where patients were randomized either to a total contact cast or a removable cast walker, healing rates were much more rapid in the total contact cast group as the removable walkers were not used for much of the time [51, 52].

Neuroischaemic Ulcers

Conservative debridement in the case of neuroischaemic lesions may be necessary using minimal sharp techniques, using instead debriding agents. Conflicting advice exists with regards to casting neuroischaemic with or without osseous involvement and to date little evidence exists but offloading remains an essential part of the management plan for these patients [53]. However, some degree of vascular intervention may need to be considered in order to increase the potential for healing. Initial investigations should comprise a non-invasive assessment using Doppler ultrasound techniques. Prior to any vascular surgical intervention, arteriography is usually indicated. Care should be taken with the use of certain contrast media as many patients with foot ulceration have renal dysfunction. All patients with PVD should be seen by a vascular surgeon or interventional radiologist who would normally be a member of the MDT.

Infection

DFUs serve as a portal for pathogens and approximately 60% of DFUs are already infected on initial presentation. Gram-positive cocci, especially *Staphylococcus aureus,* streptococcus species, and coagulase-negative staphylococci have all been isolated from DFUs [54, 55]. The prevalence of Gram-negative bacteria, pseudomonas, and *Enterobacteriaceae* species is lower, but increases in chronic wounds previously treated with antibiotics. Anaerobic bacteria must also be considered, especially in neuroischaemic ulcers.

Antibiotic-resistant organisms have become an increasing problem in the management of DFUs over recent decades with the rise of methicillin-resistant *Staphylococcus aureus* (MRSA). Multidrug resistant (MDR) Gram-negative strains such as highly resistant pseudomonas, extended-spectrum β-lactamase (ESBL) and carbapenemase-producing Gram-negative bacilli are also being isolated from diabetic foot wounds [56].

The clinical signs, pain, warmth, erythema, raised temperature, and inflammatory markers can be reduced or absent in patients with neuropathy and or ischaemia with approximately 50% remaining asymptomatic. Characteristics such as onset of tenderness, prolonged healing, and wound malodour may be the only indicators of infection. Others include discharge, poor granulation tissue, and hyperglycaemia [54]. Tissue samples are the gold standard for culture and sensitivity to inform the choice of antibiotic regimen as superficial cultures are contaminated by colonizing bacteria.

Non-infected wounds do not require antibiotics; superficial wound swabs are inaccurate and misleading, only yielding contaminants; deep tissue specimens via a curettage or after aggressive debridement are the gold standard for microbiological testing. Initial antibiotic therapy should be empirical consisting of activity against *Staphylococcus aureus* and aerobic *streptococci*. Consider that agents against Gram-negative organisms may be required for patients with severe infections.

Tailor antibiotic regimen once culture and sensitivity results are available; a more specific regimen should be initiated being mindful of local resistance patterns. Intravenous antibiotic administration should only be indicated in severe infections; for mild to moderate infections oral antibiotics with a high bioavailability will suffice. 'It's not what you put on the ulcer, it's what you take off'. Antibiotics will be of little benefit if sharp debridement, drainage of purulent discharge, and appropriate offloading are not adhered to. Both the International Working Group on the Diabetic Foot [57] and the Infectious Disease

Society of North America [54] have provided useful guidelines to assist in the antibiotic treatment of infected DFUs.

Wound Dressings

Selection of the ideal wound dressing will depend upon specific characteristics of the lesion. Unfortunately, there is currently little evidence concerning wound dressings, such that selection relies more on clinical experience rather than evidence. Wound healing can be challenging and is further complicated by neuropathy and/or ischaemia. The wound environment is paramount in maintaining a moist protective occlusive layer at the wound interface. The basic make-up of wound dressings requires absorption of exudate, thermal insulation, gas permeability, and being impenetrable to microorganisms. Adherent products should not contact the wound bed, thus preventing removal of newly granulated tissue. As the wound progresses, a variety of different dressings can come into play (i.e. sloughy wounds require debridement agents, clean moist wounds absorbency properties). Products are divided into three broad categories; debriding, antiseptic-based, and moisture control, and are listed in Table 15.14.1.1.

Negative Pressure Wound Therapy

The application of Negative Pressure Wound Therapy (NPWT) is believed to accelerate healing through reducing oedema, removal of exudate, increased perfusion, self-proliferation, and the formation of granulation tissue [58]. Studies have demonstrated that wounds achieve faster closure over conventional dressing regimes [59]. Increased perfusion and promotion of granulation tissue have been reported via increasing cell mitosis. RCTs have suggested efficacy in rates of wound healing and reduced amputations in post-surgical wounds [60].

Hyperbaric Oxygen Therapy

Hyperbaric oxygen therapy (HBO) enhances the body's natural healing process by inhalation of 100% oxygen in a total body chamber, where atmospheric pressure is increased and controlled. HBO has been promoted as an effective treatment in diabetic foot wounds over many years. Unfortunately, early RCTs consisted of small patient numbers and methodological flaws. A well designed Swedish RCT suggested there was a benefit from HBO in neuroischaemic infected lesions without the possibility of revascularization [61]. Two large multicentre RCTs, one carried out in the Netherlands [62] and a further undertaken in Canada [63], showed no benefits of HBO of the patient groups. To date there is no definite clinical indication for the use of HBO in diabetic foot lesions and few data to support its efficacy.

Growth Factors and Skin Substitutes

Becaplermin is a recombinant platelet-derived growth factor (PDGF) ointment which has shown some slight benefit in wounds with delayed healing. Granulocyte colony stimulating factor (GCS-F) has been reported to improve resolution of infection and is claimed to reduce amputation rates [64]. Bioengineered skin (Apligraf) and human dermis (Dermagraft) are types of biologically active implants and contain human fibroblasts that deliver growth factors to the wound. The evidence base is weak for these expensive

Table 15.14.1.1 Wound management products

Dressing	Description	Contraindications	Example
Hydrocolloid	Facilitate re-hydration & autolytic debridement. Dry, sloughy, necrotic wounds. Promote granulation.	Infected wounds. Twice weekly change.	Aquacel: ConvaTec Deeside, Wales, UK. Comfeel: Coloplast Peterborough, UK.
Hydrogels	Donates liquid to dry wounds & absorbs exudates. Dry, sloughy wounds. Autyolitic debridement.	Hydrogel sheets avoided in infected wounds.	Intrasite gel: Smith & Nephew wound management, Hull, UK. Iodosorb: Smith & Nephew wound mgt, UK.
Silver	Antimicrobial. Colonization.	Sensitivity to silver.	Acticoat: Smith & Nephew wound mgt, USA.
Vapour-permeable	Provide a moist healing environment. Mild exude.	Heavily exudating wound.	Tegaderm: 3M, Reading, UK.
Foam dressing	Primary or secondary cover. Light & Heavy exudates.	Remove if strike through occurs.	Allevyn: Smith & Nephew wound mgt, Europe. Lyofoam: Molnlyck,. Oldham, UK.
Odour absorbent	Absorbs odour. Malodorous.	Silver (sensitivity).	Actisorb: Johnson & Johnson Medical Skipton, UK
Larval therapy	Debridement, promote granulation. Heavily sloughy necrotic wounds.	Increase in pain.	Maggots: BioMond Bridgend, Wales, UK.
Alginate	Haemostat. Heavy exudates.	Blockage. Loose fibres.	Kaltostate: ConvaTec, UK.
Skin substitutes	Living skin. Obstinate wounds.	Colonized. Infected wound.	Dermagraft: Smith & Nephew medical, Europe.
Iodine	Antibacterial. Exudating wounds.	Iodine (sensitivity) Renal/thyroid conditions.	Iodosorb: Smith & Nephew medical, Europe.
Honey	Antimicrobial. Sloughy necrotic wounds. Autolytic debridement.	Medical grade only.	L-Mesitran: Aspen medical Europe Ltd Ashby de la Zouch, Europe.

therapies and large-scale randomized controlled trials (RCTS) are needed. A recent systematic review concluded that there is little published evidence to justify the use of any of these therapies [65].

Charcot Neuroarthropathy

CN is non-infective arthropathy in a well-perfused, insensitive foot. CN is inextricably linked with distal symmetrical somatic and autonomic neuropathy although the exact pathogenetic mechanisms are unknown.

Advances into the pathogenesis of osteopenia and osteoporosis, and the role of the receptor activator of nuclear factor κB/osteoprotegerin (RANKL/OPG) signalling system have led to the theory that acute CN might be triggered in a susceptible individual by an event that leads to localized inflammation in the affected foot, resulting in an increasing cycle of osteoblastic/clastic activity.

Its main characteristics are osseous and joint destruction leading to a gross alteration in foot structure (Figure 15.14.1.2). Abnormalities may occur in the forefoot, mid-foot, peri-talar, or ankle regions with avulsion fractures affecting the posterior tuberosity of the calcaneus. Mobilization due to a lack of sensory awareness compounds further destruction. In the later stages complete mid-foot collapse can be recognized by a rocker bottom deformity [66] (Figure 15.14.1.3). The patient with acute CN continues to walk and more minor trauma occurs. Any patient with neuropathy who complains of swelling or discomfort, with or without a history of injury, should be assessed.

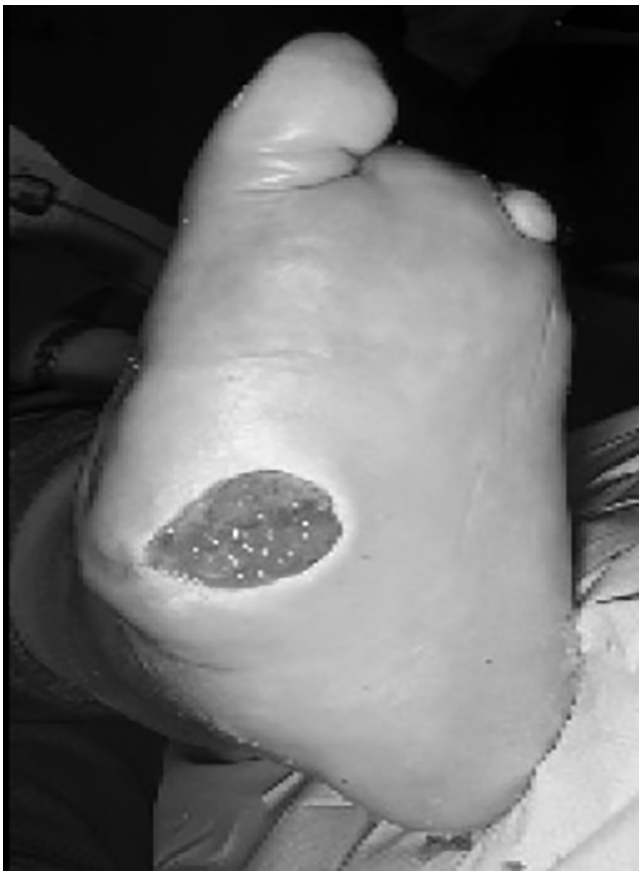


Figure 15.14.1.2 Chronic Charcot neuroarthropathy involving the mid-foot (cuneiform-metatarsal bone area). There was extensive deformity with a large plantar ulcerative lesion under the bony prominence.

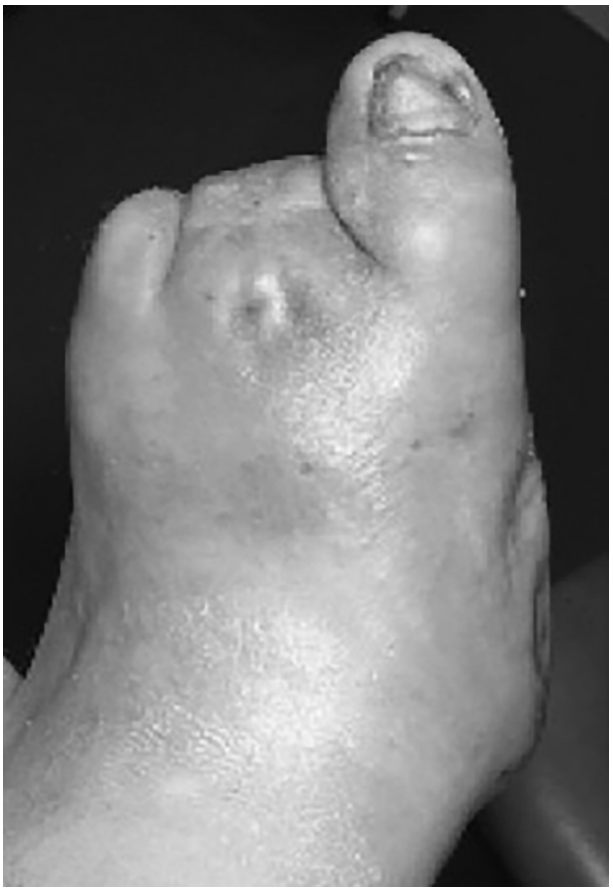


Figure 15.14.1.3 Chronic neuropathic foot problems with Charcot deformity and previous amputations of four toes.

Treatment in the initial stages of acute CN is offloading with a cast walker. Unfortunately, there are currently no proven medical or pharmacological approaches other than immobilization. Controversy exists over the timing of surgical intervention in the acute or later phase due to a lack of evidence. Simple exostectomies and tendon transfers can offer therapeutic benefit thereby reducing the risk of ulcerative lesions (Figure 15.14.1.4). Little evidence exists to show a surgically corrected Charcot deformity functions any better than a non-surgically corrected deformity [67].

Patient Education

Successful self-management requires motivation and compliance from the patient to accept a degree of responsibility for their own care. However, foot inspection can be problematic for obese individuals and those with visual impairment. Difficulty understanding the nature and relevance of neuropathy to the individual with diabetes has also been suggested as a barrier to engaging in the education process [68].

Some success with patient engagement in foot care has been demonstrated utilizing self-monitoring tools such as temperature monitoring [69], the Neuropad™ (an indicator of autonomic neuropathy

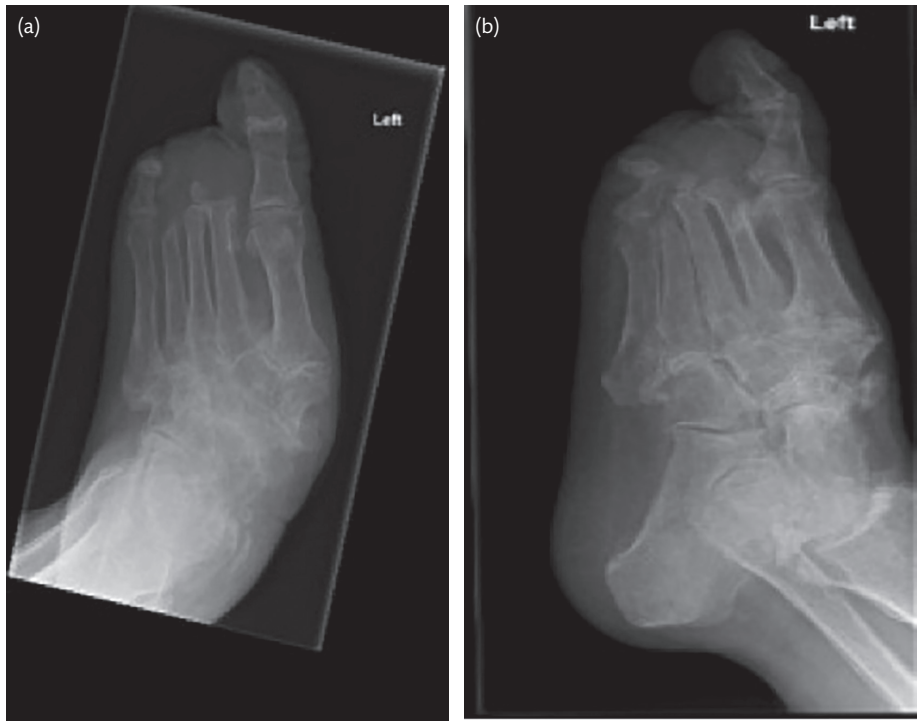


Figure 15.14.1.4 (a) Radiograph of a chronic Charcot foot demonstrating previous amputation of three digits, vascular calcification, and gross disruption in the cuneiform-metatarsal joints of the mid-foot. (b) Radiograph showing chronic Charcot neuroarthopathic changes in the mid-foot with peritalar destruction.

[70]), and a simple foot pressure mat for monitoring pressure changes under the foot [71]. Inappropriate footwear, either incorrect size or inadequate cushioning, has been observed to contribute to the development of ulcers. Tightly fitting shoes commonly lead to ulceration at dorsal deformities such as bunions or between the spaces of toes which have been crushed together. Loose shoes can also be problematic, leading to ulceration from frictional forces.

Additionally, patients should be advised about other associated risk factors such as controlling high blood pressure, cholesterol, smoking cessation, and obesity. Not only will these measures reduce patients' risk of ulcers, they will also lower their macrovascular complication risk.

Multidisciplinary Team Input

Successful management of diabetic foot complications depends upon achieving stability in all aspects of diabetes care. Patients requiring a total package of care from a specialist diabetes foot care team need a structured management plan in order to contend with the multiple comorbidities and complications associated with diabetes. Varying levels of care currently exist in different countries. At a minimum, a multidisciplinary foot-care team should consist of a diabetologist, surgeon (podiatric, vascular, or orthopaedist), diabetes specialist nurse and podiatrists (or in some countries the older title of chiropodist).

Improved outcomes, including reduced incidence of amputations, have been shown in a number of studies where there is a clear MDT involvement [10]. One study directly compared outcomes associated with care delivered by an established MDT comparing with another hospital lacking an MDT. Major amputations performed on patients treated by the MDT were lower (4.7%) vs. 21.7% without MDT input (P <0.0001). Mortality during hospitalization was also significantly different between the two groups at 2.5% for the MDT group and 9.4% for the controls (P <0.001) [72].

Conclusion

The Foot in Remission

Recurrence rates after ulcer healing may be as high as 40% after 1 year and 65% after 5. CN patients have up to 50% chance of developing the same scenario in the contralateral foot. Thus the concept of 'remission' may be preferable to referring to 'healed' [73, 74].

The Future

Diabetic foot complications involve multiple factors coming into play driven primarily by prolonged hyperglycaemia. The tools for reducing foot complications are already available to the MDT in the form of consensus and evidence based guidelines. Lack of patient engagement and self-empowerment represent the most significant barrier to future success in diabetic related foot complications.

The aetiopathogenesis and management of diabetic foot disease over the last three decades, in the mainstay comes from what we have learned from clinical practice still lacking evidence base. Most important of all however in the management of patients with diabetic foot disorders, is to remember that the patient has frequently

lost the 'gift of pain' that protects most of us from developing significant foot problems but, when absent, can lead to devastating consequences.

REFERENCES

1. Y Boulton AJM, Vileikyte L, Ragnaarson-Tennvall G, Apelqvist J. The global burden of diabetic foot disease. *Lancet*, 2005; **366**: 1719–24.
2. Jeffcoate WJ, Young RJ. National diabetic foot audit of England and Wales yields its first dividends. *Diabetic Med*, 2016; **33**: 1464–65
3. Prompers L, Huijberts M, Scaper N, *et al.* Resource utilisation and costs associated with the treatment of diabetic foot ulcers. Prospective data from the Eurodiale study. *Diabetologia*, 2008; **51**: 1826–34.
4. Cavanagh P, Attinger C, Abbas Z, Bal A, Rojas N, Xu ZR. The cost of treating diabetic foot ulcers in 5 different countries. *Diabetes Metab Res Rev*, 2012; **28**: 107–11.
5. Skrepnek GH, Mills JL, Armstrong DG. Foot-in-wallet disease: tripped up by 'cost saving' reductions? *Diabetes Care*, 2014; **37**: e196–7.
6. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. *N Engl J Med*, 2017; **376**: 2367–75.
7. Armstrong DG, Boulton AJM, Bus SA. The diabetic foot in remission: towards increasing ulcer-free days. *N Eng J Med*, 2017; **376**: 2367–75.
8. Pecoraro RE, Reiber GE, Burgess EM. Pathways to diabetic limb amputation: basis for prevention. *Diabetes Care*, 1990; **13**: 513–21.
9. Abbott CA, Carrington AL, Ashe H, *et al.* The North-West Diabetes Foot Care Study: incidence of, and risk factors for new diabetic foot ulceration in a community-based patient cohort. *Diabet Med*, 2002; **19**: 377–84.
10. Krishnan S, Nash F, Baker N, Fowler D, Rayman G. Reduction in diabetic amputations over 11 years in a defined UK population: benefits of multidisciplinary team work and continuous prospective audit. *Diabetes Care,* 2008; **31**: 99–101.
11. Leese GP, Stang D, McKnight JA, and The Scottish Diabetes Foot Action Group. A national strategic approach to diabetic foot disease in Scotland: changing a culture. *Br J Diabetes Vasc Dis*, 2011; **11**: 69–73.
12. Holman N, Young RJ & Jeffocate WJ. Variation in the recorded incidence of amputation of the lower limb in England. *Diabetologia*, 2012; **55**: 1919–25.
13. Health and Social Care Information Centre, National Diabetes In-patient Audit, 2012. (2013). Available at: https://catalogue.ic.nhs.uk/publications/clinical/diabetes/nati-diab-inp-audi-12/nati-diab-inp-audi-12-nat-rep.pdf (accessed: 26 April 2020).
14. Diabetes UK, Cost of Diabetes Report, 2014. Available at: https://www.diabetes.org.uk/resources-s3/2017-11/diabetes%20uk%20cost%20of%20diabetes%20report.pdf (accessed: 26 April 2020).
15. Ahmad N, Thomas G, Chan C, Gill P. Ethnic differences in lower limb revascularisation and amputation rates. Implications for the aetiopathology of atherosclerosis? *Atherscleroses*, 2014; **233**: 503–7.
16. Vamos EP, Bottle A, Majeed A, Millett C. Trends in lower extremity amputations in people with and without diabetes in England, 1996–2005. *Diabetes Res Clin Pract*, 2010; **87**: 275–82.
17. Vamos EP, Bottle A, Edmonds ME, Valabhji J, Majeed A, Millett C. Changes in the incidence of lower extremity amputations in individuals with and without diabetes in England between 2004 and 2008. *Diabetes Care*, 2010; **33**: 2592–7.
18. Moxey PW, Hofman D, Hinchliffe RJ, Jones K, Thompson MM, Holt PJE. Epidemiological study of lower limb amputation in England between 2003 and 2008. *BJS*, 2010; **97**: 1348–53.
19. Rayman G, Krishnan ST, Baker NR, Wareham AM, Rayman A. Are we underestimating diabetes-related lower-extremity amputation rates? Results and benefits of the first prospective study. *Diabetes Care*, 2004; **27**: 1892–6.
20. McCaslin JE, Hafez HM, Stansby G. Lower Limb revascularisation and major limb amputation rates in England. *BJS*, 2007; **94**: 835–9
21. Canavan RJ, Unwin NC, Kelly WF, Connolly VM. Diabetes- and nondiabetes-related lower extremity amputation incidence before and after the introduction of better organized diabetes foot care: continuous longitudinal monitoring using a standard method. *Diabetes Care*, 2008; **31**: 459–63.
22. Abbott CA, Carrington AL, Ashe H, *et al.* The North-West Diabetes Foot Care Study: incidence of, and risk factors for new diabetic foot ulceration in a community-based patient cohort. *Diabet Med*, 2002; **19**: 377–84.
23. Young MJ, Breddy JL, Veves A, Boulton AJM. The prediction of diabetic neuropathic foot ulceration using vibration perception thresholds. A prospective study. *Diabetes Care*, 1994; **17**: 557–600.
24. Brown SJ, Handsaker J, Bowling FL, Boulton AJM, Reeves ND. Diabetic peripheral neuropathy compromises balance during daily activities. *Diabetes Care*, 2015; **3**: 32–4.
25. Boulton AJM, Malik RA, Arezzo JL, Sosenko JM. Diabetic somatic neuropathies: technical review. *Diabetes Care*, 2004; **27**: 1458–86.
26. Tentolouris N, Marinou K, Kokotis P, Karanti A, Diakoumopoulou E, Katsilambros N. Sudomotor dysfunction is associated with foot ulceration in diabetes. *Diabet Med*, 2009; **26**: 302–5.
27. Ndip A, Rutter MK, Vileikyte L, *et al.* Dialysis is an independent risk factor for foot ulceration in patients with diabetes and chronic kidney disease. *Diabetes Care*, 2010; **33**: 1811–16.
28. Charcot neuroarthropathy after simultaneous pancreas-kidney transplantation: risk factors, prevalence and outcome. *Clin Transplant*, 2015; **29**: 712–19.
29. Vileikyte L. Psychosocial and behavioural aspects of diabetic foot lesions. *Curr Diab Rep*, 2008; **8**: 119–25.
30. Gonzalez JS, Vileikyte L, Ulbrecht JS, *et al.* Depression predicts first but not recurrent diabetic foot ulcers. *Diabetologia*, 2010; **53**: 2241–8.
31. Bitar FG, Garcia LA. Critical limb ischaemia: an overview of the epidemiologic and clinical implications. *Vasc Dis Manag*, 2010; **7**: 182–4.
32. Rutherford RB, Baker JD, Ernst C, *et al.* Recommended standards for reports dealing with lower extremity ischemia: revised version. *J Vasc Surg*, 1997; **26**: 517.
33. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. *Circ Res*, 2015; **116**: 1509–26.
34. Heikkila K, Loftus IM, Mitchell DC, *et al.* Population based study of mortality and major amputation following lower limb revascularisation. *Br J Surg*, 2018; **105**: 1145–54.
35. Dormandy JA, Betteridge DJ, Schernthaner G, *et al.* Impact of peripheral arterial disease in patients with diabetes—results from PROactive. *Atherscleroses*, 2009; **202**: 272–81.

36. Boulton AJM. The diabetic foot. In: DeGroot LJ, Chrousos G, Dungan K, *et al.*, eds. Endotext. South Dartmouth, MA: MD text.com, Inc.
37. Pop-Busui R, Boulton AJ, Feldman EL, *et al.* Diabetic neuropathy: a position statement by the American Diabetes Association. *Diabetes Care*, 2017; **40**: 136–54.
38. Valk GD, de Sonnaville JJ, van Houtum WH, *et al.* The assessment of diabetic polyneuropathy in daily clinical practice: reproducibility and validity of Semmes Weinstein monofilaments examination and clinical neurological examination. *Muscle Nerve*, 1997; **20**: 116–18.
39. Rayman G, Vas PR, Baker N, *et al.* The Ipswich Touch Test: a simple and novel method to identify in patients with diabetes at risk of foot ulceration. *Diabetes Care*, 2011; **34**: 1517–18.
40. Bowling FL, Abbott CA, Harris WE, *et al.* A pocket-sized disposable device for testing the integrity of sensation in the outpatient setting. *Diabet Med*, 2012; **29**: 1550–2.
41. Perkins B, Olaleye D, Zinman B, Bril V. Simple screening tests for peripheral neuropathy in the diabetes clinic. *Diabetes Care*, 2001; **24**: 250–6.
42. Creager MA, Belkin M, Bluth EI, *et al.* Key data elements and definitions for peripheral atherosclerotic vascular disease: a report of the American college of cardiology Foundation/American Heart association task force on clinical data standards. *J Am Coll Cardiol*, 2012; **59**: 294.
43. Fowkes FG, Rudan D, Rudan I, *et al.* Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *Lancet*, 2013; **382**: 1329.
44. Cassar A, Poldermans D, Rihal C, Gersh B. The management of combined coronary artery disease and peripheral vascular disease. *Eur Heart J*, 2010; **31**: 1565–1572.
45. Lazaro-Martinez JL, Aragon-Sanchez J, Garcia Morales E. Antibiotics versus conservative surgery for treating diabetic foot osteomyelitis: a randomised controlled comparative trial. *Diabetes Care*, 2014; **37**: 789–95.
46. Li HK, Scarborough M, Zambellas R, *et al.* Oral versus intravenous antibiotic treatment for bone and joint infections (OVIVA): study protocol for a randomised controlled trial. *Trials*, 2015; **16**: 583.
47. Edmonds M, Lazaro-Martinez J, Alfayate-Garcia J, *et al.* Sucrose octasulfate dressing versus control dressing in patients with neuroischaemic diabetic foot ulcers (EXPLORER): an international multicentre, double blind, randomised, controlled trial. *Lancet*, 2017; **20**: 85–6.
48. Vab Baal JG. Surgical treatment of the infected diabetic foot. *Clin Infect Dis*, 2004; **39**: S123–8.
49. Armstrong DG, Lavery LA, Kimbriel HR, Nixon BP, Boulton AJM. Activity patterns of patients with diabetic foot ulceration: patients with active ulceration may not adhere to a standard pressure off-loading regimen. *Diabetes Care*, 2003; **26**: 2595–7.
50. Armstrong DG, Lavery LA, Wu S, Boulton AJ. Evaluation of removable and irremovable cast walkers in the healing of diabetic foot wounds: a randomised controlled trial. *Diabetes Care*, 2005; **28**: 551–4.
51. Piaggesi A, Macchiarini S, Rizzo L, *et al.* An off the shelf instant contact casting device for the management of diabetic foot ulcers: a randomised prospective trial versus traditional fiberglass cast. *Diabetes Care*, 2007; **30**: 586–90.
52. Katz IA, Harlan A, Miranda-Palma B, *et al.* A randomized trial of two irremovable off-loading devices in the management of plantar neuropathic diabetic foot ulcers. *Diabetes Care*, 2005; **28**: 555–9.
53. Nabuurs-Franssen MH, Sleegers R, Huijberts MS, *et al.* Total contact casting of the diabetic foot in daily practice a prospective follow-up study. *Diabetes Care*, 2005; **28**: 243–7.
54. Lipsky BA, Berendt AR, Cornia PB, *et al.* Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. *Clin Infect Dis*, 2012; **54**: e132–73.
55. Citron DM, Goldstein EJ, Merriam CV, Lipsky BA, Abramson MA. Bacteriology of moderate-to-severe diabetic foot infections and *in vitro* activity of antimicrobial agents. *J Clin Microbiol*, 2007; **45**: 2819–28.
56. Uckay I, Gariani K, Pataky Z, Lipsky BA. Diabetic foot infections: state-of-the-art. *Diabetes Obes Metab*, 2014; **16**: 305–16.
57. Peters EJ, Lipsky BA, Aragon-Sanchez J, *et al.* Interventions in the management of infection in the foot in diabetes: a systematic review. *Diabetes Metab Res Rev*, 2016; **32**: 145–53.
58. Meloni, M, Izzo Valentina, Vainieri E, *et al.* Management of negative pressure wound therapy in the treatment of diabetic foot ulcers: a systematic review. *Diabetes Metab Res Rev*, 2016; **32**: 145–53
59. Lone AM, Zaroo M, Laway B, *et al.* Vacuum assisted closure versus conventional dressings in the management of diabetic foot ulcers: a prospective case control study. *Diabet Foot Ankle*, 2014; **8**: 5.
60. Armstrong DG, Lavery LA. Negative pressure wound therapy after partial diabetic foot amputation: a multicentre, randomised controlled trial. *Lancet*, 2005; **366**: 1704–10.
61. Londahl M. Hyperbaric oxygen therapy as treatment for diabetic foot ulcers. *Diab Metab Res Rev*, 2012; **28**: 78–84.
62. O'Reilly D, Linden R, Fedorko L, *et al.* A prospective double-blind, randomised, controlled clinical trial comparing standard wound care with adjunctive hyperbaric oxygen therapy (HBOT) to standard wound care only for the treatment of chronic, non-healing ulcers of the lower limb in patients with diabetes mellitus: a study protocol. *Trials*, 2011, **12**: 69.
63. Margolis DJ, Gupta J, Hoffstad O, *et al.* Lack of effectiveness of hyperbaric oxygen therapy for the treatment of diabetic foot ulcer and the prevention of amputation: a cohort study. *Diabetes Care*, 2013; **36**: 1961–6.
64. Cruciani, M, Lipsky BA, Mengoli C, de Lalla F. Granulocyte-colony stimulating factors as adjunctive therapy for diabetic foot infections. *Cochrane Database Syst Rev*, 2013; **10**: CD006810.
65. Game FL, Apelqvist J, Attinger C, *et al.* Effectiveness of interventions to enhance healing of chronic ulcers of the foot in diabetes: a systematic review. *Diabetes Metab Res Rev*, 2016; **32**: 154–68.
66. Rogers LC, Frykberg RG, Boulton AJM. The Charcot foot in diabetes. *Diabetes Care*, 2011; **34**: 2123–9.
67. Shen W, Wukich D. Orthopaedic surgery and the diabetic Charcot foot. *Med Clin N Amer*, 2013; **97**: 873–82.
68. Vileikyte L. Psychosocial and behavioural aspects of diabetic foot lesions. *Curr Diab Rep*, 2008; **8**: 119–25.
69. Lavery LA, Higgins KR, Lanctot DR, *et al.* Preventing diabetic foot ulcer recurrence in high risk patients. Use of temperature monitoring as a self assessment tool. *Diabetes Care*, 2007; **30**: 114–20.
70. Tentolouris N, Achtsidis V, Marinou K, Katsilambros N. Evaluation of the self-administered indicator plaster neuropad for the diagnosis of neuropathy in diabetes. *Diabetes Care*, 2008; **31**: 236–7.
71. van Schie CH, Abbott CA, Vileikyte L, Shaw JE, Hollis S, Boulton AJM. A comparative study of the Podotrack, a simple

semi-quantitative plantar pressure measuring device, and the optical paedobarograph in the assessment of pressures under the diabetic foot. *Diabet Med*, 1999; **16**: 144–59.
72. Weck M, Slesaczeck T, Paetzold H, *et al.* Structured healthcare for subjects with diabetic foot ulcers results in major reduction of amputation rates. *Cardiovasc Diabetol*, 2013; **12**: 45.
73. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. *N Engl J Med*, 2017; **376**: 2367–75.
74. Weck M, Slesaczeck T, Paetzold H, *et al.* Structured healthcare for subjects with diabetic foot ulcers results in major reduction of amputation rates. *Cardiovasc Diabetol*, 2013; **12**: 45

15.15

Delivery of Diabetes Care

15.15.1 Diabetes Service Organization

Jonathan Valabhji

This section was written in June 2018. Several programmes which are discussed to offer context are described as they were at this time, namely:

- The Healthier You: NHS Diabetes Prevention Programme
- The NHS England Digital Diabetes Programme
- The NHS England Diabetes Treatment and Care Programme.

Introduction

The World Health Organization has estimated that globally, 422 million adults were living with diabetes in 2014, compared to 108 million in 1980. This represents an almost doubling of the age-standardized global prevalence of diabetes since 1980, rising from 4.7% to 8.5% in the adult population over this period [1]. Separate global estimates of diabetes prevalence for type 1 and type 2 do not exist. However, the majority of people with diabetes have type 2 diabetes.

Overweight and obesity is the major modifiable risk factor for type 2 diabetes, and the global increase in prevalence of type 2 diabetes has tracked the global increase in prevalence of overweight and obesity. The World Health Organization has estimated that worldwide, obesity has nearly tripled since 1975; that in 2016, more than 1.9 billion adults were overweight, and of these over 650 million were obese; and that in 2016, 39% of adults were overweight and 13% were obese [2].

In England, with a population of approximately 55 million, 3.8 million people aged over 16 had diabetes in 2015, around 9% of the adult population. Approximately 90% had type 2 diabetes, and the model estimated that 940 000 of those with type 2 diabetes were as yet undiagnosed [3].

The direct health costs for type 1 and type 2 diabetes in the United Kingdom were estimated to be £9.8billion in financial year 2010/11, approximately 10% of the total health resource expenditure. The majority of the direct costs, 80%, related to the management of the complications of diabetes [4]. The National Audit Office estimated in 2012 that National Health Service (NHS) spending on diabetes services in England in 2009/10 was at least £3.9 billion [5], and in 2015 estimated that the cost of diabetes to the NHS in England in 2010/11 was £5.6 billion [6]. The differences in estimates reflect different assumptions made, and reflect the difficulties in accurately assessing health costs related to diabetes.

Increasing Prevalence Has Required Evolution of Healthcare Delivery Systems

Although different healthcare systems have different means of funding, increasing prevalence of diabetes has required evolution of healthcare delivery systems irrespective of funding methods. While the lens of the current article is the NHS in England, a taxpayer funded system that provides universal care free at the point of delivery, the challenges posed by increasing diabetes prevalence, and the potential solutions, will to some extent be common to most healthcare systems. Whereas in the 1980s, diabetes care delivery was often the preserve of hospital-based specialist diabetes services, increasing patient numbers now require the majority of diabetes care to be delivered by generalist healthcare professionals outside of hospital settings, with more complex care, and care related to the complications of diabetes usually delivered by hospital-based specialists, with an integrated interface between generalist and specialist services. The nature of the interface, and the differing relative care distributions between generalist and specialist teams, vary across local healthcare systems and local health economies in England. Type 1 diabetes and rarer forms of diabetes are usually cared for by specialist teams even in the absence of complications.

In some low and middle income countries, there may be additional and distinct problems, such as access to medicines, including insulin,

that will not be covered specifically in this chapter. Where both resources and diabetes expertise are in short supply, care delivery is often centralized. However, although approaches may be quite different, the principles of providing the best possible services within a specific and sometimes limited resource envelope remain the same.

Sustainability of Healthcare Systems

With health inflation at around 7% per annum and ageing populations, sustainability is a concern for most healthcare systems internationally. In England, a strategy document called the Five Year Forward View was published in 2014 that suggested a direction of travel for healthcare, with an important emphasis on long-term sustainability of the NHS in England [7]. Areas of focus relevant to diabetes care prioritized either within that document or subsequently are:

- To establish a national evidence-based type 2 diabetes prevention programme for those already identified to be at high risk, in order to reduce the current national trajectory of type 2 diabetes incidence.
- To invest in diabetes treatment and care in order to reduce future complication burden, and so potentially realize long-term return on investment as well as improvements in care and outcomes.
- To empower individuals with diabetes to better self-manage their condition, through better access to information, support, and structured education.
- To harness digital technologies for healthcare provision and to support individuals to better self-manage their condition through digital health platforms.
- To better integrate services, across primary care and hospital interfaces, across physical and mental healthcare interfaces, and across health and social care interfaces.

The Healthier You: NHS Diabetes Prevention Programme

Based on best evidence [8], an intervention was developed to prevent or delay onset of type 2 diabetes in those already identified to be at high risk. The intervention encourages behavioural change focussing on weight loss, increased physical activity, and better quality nutrition. The intervention is for those with non-diabetic hyperglycaemia (also referred to as pre-diabetes) and involves at least 13 group-based face-to-face sessions, across at least a 9 month period, constituting at least 16 hours of contact time. During the first year of national roll-out (financial year 2016/17), the programme was implemented across approximately 50% of England. Referral numbers and percentage uptake were found to be in excess of prior modelled values, suggesting that general practices and people at high risk of type 2 diabetes value the prevention opportunities provided. Attendance rates for men, Asian, Afro-Caribbean, mixed, and other ethnic groups and participants from areas in the most deprived quintile suggest that, in its early stages, the programme is reaching both those who are at greater risk of developing type 2 diabetes and those who typically access healthcare less effectively [9]. During the second year of national roll-out (financial year 2017/18), implementation reached 75% of England, with early data demonstrating encouraging weight loss seen in participants that have so far completed the intervention, as well as demonstrating encouraging completion rates, with over 50% of those that initiated an intervention attending at least eight of the tailored support sessions. By summer 2018, implementation reached 100% of England. At full roll-out, the programme is planned to deliver to 100 000 people each year, although more recent policy decisions (announced 30 November 2018), based on the initial encouraging results, suggest that this will be extended to 200 000 people each year. The National Institute of Health Research has supported an independent evaluation of the programme, scheduled to report in 2021, that will assess programme impact on the current national trajectory of type 2 diabetes incidence.

Interventions for individuals at high risk of type 2 diabetes will be complemented by population level interventions, outlined in the UK Government's Childhood Obesity Plan [10], such as a levy on sugar-sweetened beverages, and plans to encourage product reformulation to reduce sugar and total calorie content.

An overview of systematic reviews and meta-analysis has recently evaluated the efficacy of population-wide obesity and diabetes prevention programmes. Increased price of sugar-sweetened beverages and decreased price of fruit and vegetables, menu labelling, grocery-store interventions and multicomponent interventions were associated with small reductions in body mass index (BMI) or weight. While the reach of such interventions was many times greater than interventions for those at high risk, the effect size was smaller by a similar factor [11]. So a combination of both approaches is desirable.

The NHS England Digital Diabetes Programme

The two promises of digital health interventions are greater reach and lower unit cost, and these may be another route to deliver behaviour change services with large reach. However, while evidence of effectiveness of digital interventions in the prevention of type 2 diabetes is currently emerging, this has not yet reached a consensus position. For this reason, NHS England is currently piloting a Digital Diabetes Prevention programme, trialling five digital products in those at high risk of type 2 diabetes in the real world setting, with a formal evaluation to assess clinical effectiveness and reach [12].

Digital health interventions also offer opportunities for delivery of behaviour change services around diabetes self-management [13], and this is a further area of investment and development in England.

The NHS England Diabetes Treatment and Care Programme

The focus here is to reduce the future diabetes complication burden, with up-front investment. Cases for potential return on investment, by way of improvements in care and outcomes that evidence indicated should result in reinvestable savings, were made in the following areas:

- Achievement of treatment targets for HbA1c, blood pressure, and cholesterol and tackling variation around that achievement in order to reduce the risk of onset of complications.
- Attendance at structured education to empower individuals with diabetes to better self-manage their condition and so support improved glycaemic control and psychosocial well-being.

- Multidisciplinary diabetic foot care, and more specifically the existence of and sufficient capacity within a multidisciplinary foot service. Multidisciplinary diabetic foot care has been shown to reduce amputations by up to 50% [14] and a strong economic case has been made [15].
- Inpatient care, and more specifically the existence of and sufficient capacity within a diabetes inpatient specialist nurse service which has been shown to support reduction in lengths of inpatient stay. One in six of all hospital beds in England are now occupied by someone with diabetes.

A Model of Service Provision That Promotes Integrated Care Delivery

In the latter part of 2014, NHS England, in collaboration with stakeholders, produced a sample service specification for diabetes services [16]. The principals involved are broadly applicable—the model outlines the provision of high quality care for all those with diabetes, and differentiates the care needs of those with type 1 from those with type 2 diabetes where those care needs differ. It details the entire care pathway for people with diabetes, including those with the long-term complications of diabetes. The model divides the care pathway into two broad elements:

- Generalist care and
- Specialist care.

In the example (see Figure 15.15.1.1), a community based multidisciplinary team (MDT) that interfaces between general practice-based and specialist services is also included.

In the model described, the generalist practice-based service will have primary responsibility for the person with diabetes. Specialist services and the community based MDT will have responsibility for the episodes of care provided in those settings. However, accountability for the incidence of onset of complications and incidence of hard clinical endpoints such as amputation and blindness across the

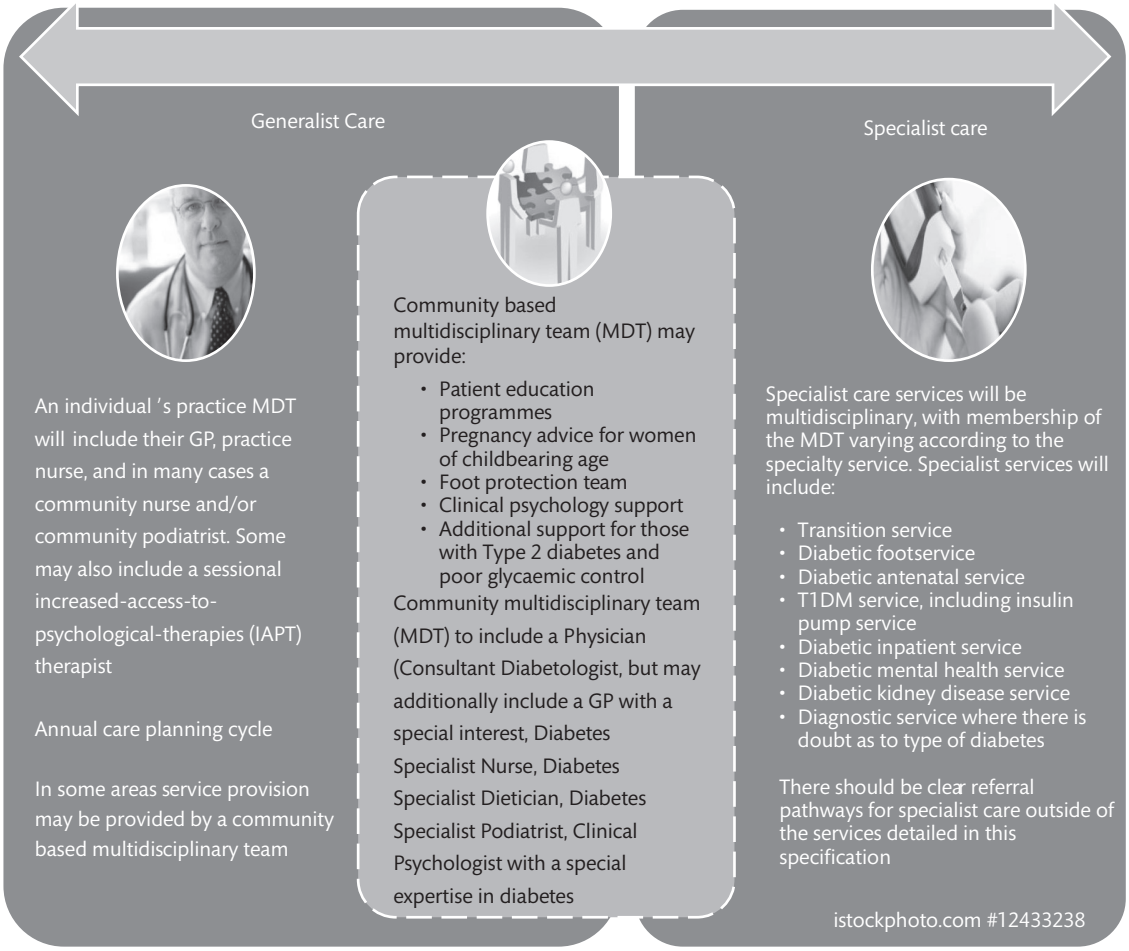


Figure 15.15.1.1 An exemplar care pathway. Care delivery is based on a multidisciplinary approach whether the care setting is the GP practice, a community centre, or a hospital. The example illustrates a community based multidisciplinary team (MDT) at the interface between generalist practice-based care and specialist care. The community based multidisciplinary team can support delivery of parts of the pathway that could not be delivered in every GP practice—an example could be delivery of structured education. Such teams should be innovatively exploring ways of generalists and specialists working together in the community using information technology and new technologies to ensure patient care is delivered in an appropriate setting local to the patient when possible. They need to have strong networking links and channels of communication with generalist and specialist colleagues working in the hospital and community.

health economy should be shared by all providers of diabetes care. Therefore both generalist and specialist services will be jointly accountable for clinical outcomes.

A community based MDT may act as the link between generalist clinicians and hospital-based specialists by representing both. Hospital-based specialists should spend a proportion of their time in the community advising and facilitating the work of the community based MDT. The presence of specialists in the MDT will facilitate fast-tracking of complications once diagnosed up to appropriate specialist settings and allow the team to provide more routine aspects of specialist care closer to the patient's home. Examples of services that may be delivered by the community based MDT include:

1. Structured education for those with type 2 diabetes, for people whose GP practice does not provide this in-house.
2. Structured education for those with type 1 diabetes.
3. Type 2 diabetes with poor glycaemic control despite best efforts in primary care.
4. Pregnancy advice for women of childbearing age.
5. Foot protection—assessing and advising those identified to be at high risk of diabetic foot disease (e.g. those already identified to have peripheral neuropathy or/and peripheral vascular disease).
6. Type 1 diabetes care when the MDT includes a consultant diabetologist.
7. Clinical psychology support within the MDT environment for those with depression and anxiety that is related to their diabetes.

All people with T1DM should have access to specialist services if they so choose, given the relative rarity of type 1 diabetes and the associated specific care needs. People with other forms of diabetes, such as monogenic diabetes (e.g. maturity-onset diabetes of the young (MODY), mitochondrial diabetes), diabetes due to chronic pancreatitis or total pancreatectomy, should also have access to specialist services given their specific care needs.

For the purposes of this model the providers of specialist care have the following designated responsibilities:

1. Provision of a transition diabetes service (ages 13–25 years).
2. Provision of a diabetic foot service.
3. Provision of a diabetic antenatal service.
4. Provision of a diabetic kidney service, prior to renal replacement therapy.
5. Provision of a T1DM service, including an insulin pump service.
6. Provision of a diabetic inpatient service.
7. Provision of a diabetic mental health service.
8. Provision of a diagnostic service where there is doubt as to the type of diabetes—if there is difficulty differentiating type 1 from type 2 diabetes, or if a rarer form of diabetes, such as MODY or mitochondrial diabetes, is suspected.

There may be additional services provided by the specialist provider, depending on local requirement that are not covered here. There will also be additional services that contribute to comprehensive diabetes care, that are dealt with through broader population based services, such as ophthalmology/medical retinal services and retinal screening services. Where this is the case, it is however important that such services are still integrated within the diabetes care pathways—for example, that the recognition of significant diabetic retinopathy is associated with greater input and efforts to improve glycaemic and blood pressure control by diabetes care pathway generalists and/or specialists as appropriate. It is also important to ensure that the number screening positive with retinal screening is matched by appropriate capacity in ophthalmology/medical retinal services, even if such services are not included within the diabetes service specification.

This model of care is reliant on the seamless integration of generalist and specialist services. To achieve this it is essential that patient records are integrated—and wherever possible shared or owned by the person with diabetes—and the two elements have good communication mechanisms to allow for continuity of care. Integration can be further supported by formal arrangements for specialists to support generalists through:

- Email advice (e.g. a specified one working day turn-around for email advice);
- Telephone contact support (e.g. a dedicated daily time window for taking calls for advice).

For older people, many of whom will have complications of diabetes and hence will have multiple comorbidities and may also suffer frailty, there will need to be coordination of health and social care.

Competency of Healthcare Professionals and Continuous Professional Development

It is important to ensure that all staff involved in designing and delivering the service are trained in line with any national/professional recommendations and curricula to achieve key competencies that have been identified in their job role to deliver appropriate diabetes care. Time should be made available in job plans to support relevant initial, and then continuous professional development for all staff contributing to the diabetes clinical pathway. This provides an opportunity to foster further interaction between generalists and specialists. Diabetes specialist physicians, nurses, dieticians, podiatrist, and psychologist members of the MDT should provide continuing diabetes-specific education to members of the generalist teams. These specialist members of the MDT can also provide support, advice, and mentorship in diabetes management to members of the generalist teams.

Other Models of Service Provision

While the broad principles are similar, other models of diabetes care provision have been described and implemented in the NHS in England. All share the principles of patient-centred care, and integration of diabetes services [17].

The NHS RightCare Diabetes Pathway

NHS RightCare is a national NHS England supported programme committed to delivering the best care to patients, making the funding available go as far as possible and improving patient outcomes [18]. NHS RightCare advises local health economies to:

- Make the best use of their resources—by tackling overuse and underuse of resources.
- Understand their performance—by identifying variation between demographically similar populations so they can adopt and implement optimal care pathways more efficiently and effectively.

- Talk together about the same things—about population healthcare rather than organizations and encouraging joint decision-making.
- Focus on areas of greatest opportunity by identifying priority programmes which offer the best opportunities to improve healthcare for people and ensuring taxpayer money goes as far as possible.
- Use tried and tested evidence-based processes to make sustainable improvement to reduce unwarranted variation.

The RightCare diabetes pathway defines the core components of an optimal diabetes service, leaning heavily on the principles already outlined in the model of service delivery here, for people with type 1 and type 2 diabetes or those at risk of developing type 2 diabetes that delivers better value in terms of outcomes and cost. The pathway shows the core components of an optimal diabetes service, evidence of the opportunity to reduce variation and improve outcomes and the key evidence-based interventions which the system should focus on for greatest improvement, supported by practice examples from across the NHS [19]. The pathway allows local health economies to think about their existing diabetes service and compare it with an optimal diabetes service. It provides guidance for commissioners about the scale of improvements that could be delivered through optimization of local pathways.

Conclusions

Increasing diabetes prevalence has required evolution of healthcare delivery systems, with a view to long-term sustainability. With a focus on patient-centredness, much of the care now provided is by generalist healthcare professionals, necessitating effective integration of generalist and specialist services.

REFERENCES

1. World Health Organization. Global Report on Diabetes, 2016. Available at: http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=0B417A3232978EF7DE3F5E397B58701B?sequence=1 (accessed: 27 March 2018).
2. World Health Organization. Media Centre Factsheet, 2018. Available at: http://www.who.int/mediacentre/factsheets/fs311/en/ (accessed: 27 March 2018).
3. Public Health England. GOV.UK, 2016. Available at: https://www.gov.uk/government/news/38-million-people-in-england-now-have-diabetes (accessed: 27 March 2018).
4. Hex N, Bartlett C, Wright D, Taylor M, Varle D. Estimating the current and future costs of type 1 and type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med*, 2012; **29**: 855–62.
5. National Audit Office. The management of adult diabetes services in the NHS, 2012. Available at: https://www.nao.org.uk/wp-content/uploads/2012/05/121321.pdf (accessed: 27 March 2018).
6. National Audit Office. The management of adult diabetes services in the NHS: progress review 2015. Available at: https://www.nao.org.uk/wp-content/uploads/2015/10/The-management-of-adult-diabetes-services-in-the-NHS-progress-review.pdf (accessed: 27 March 2018).
7. NHS England. Five Year Forward View, 2014. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf (accessed: 27 March 2018).
8. Public Health England. A systematic review and meta-analysis assessing the effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes mellitus in routine practice. 2015. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/456147/PHE_Evidence_Review_of_diabetes_prevention_programmes-_FINAL.pdf
9. Barron E, Clark R, Hewings R, Smith J, Valabhji J. Progress of the Healthier You: NHS Diabetes Prevention Programme: referrals, uptake and participant characteristics. *Diabet Med,* 2017; **35**: 513–18.
10. HM Government. Childhood Obesity; a plan for action. 2016. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/546588/Childhood_obesity_2016__2__acc.pdf (accessed: 28 March 2018).
11. Roberts S, Pilard L, Chen J, Hirst J, Rutter H, Greenhalgh T. Efficacy of population-wide diabetes and obesity prevention programmes: an overview of systematic reviews on proximal, intermediate, and distal outcomes and a meta-analysis of impact on BMI. *Obes Rev*, 2019; **20:** 947–63.
12. NHS England. NHS Diabetes Prevention Programme—digital stream. 2018. Available at: https://www.england.nhs.uk/diabetes/digital-innovations-to-support-diabetes-outcomes/nhs-diabetes-prevention-programme-digital-stream/ (accessed: 27 March 2018).
13. Murray E, Sweeting M, Dack C, *et al. BMJ Open,* 2017; **7**: e016009.
14. Edmonds ME, Blundell MP, Morris ME, Thomas EM, Cotton LT, Watkins PJ. Improved survival of the diabetic foot: the role of a specialized foot clinic. *Q J Med*, 1986; **60**: 763–71.
15. Diabetes UK. Improving footcare for people with diabetes and saving money: an economic study in England. Available at: https://diabetes-resources-production.s3-eu-west-1.amazonaws.com/diabetes-storage/migration/pdf/Improving%2520footcare%2520economic%2520study%2520%28January%25202017%29.pdf (accessed: 29 March 2018).
16. NHS England. Sample Diabetes Service Specification, 2014. Available at: https://diabetes-resources-production.s3-eu-west-1.amazonaws.com/diabetes-storage/migration/pdf/FINAL%2520Diabetes%2520Sample%2520Specification%2520V19%252029%2520July.pdf (accessed: 27 April 2020).
17. Diabetes UK. Improving the delivery of adult diabetes care through integration. 2017. Available at: https://www.diabetes.org.uk/resources-s3/2017-11/integrated%20diabetes%20care%20%28pdf%2C%20648kb%29.pdf (accessed: 28 March 2018).
18. NHS England. NHS RightCare, 2018. Available at: https://www.england.nhs.uk/rightcare/what-is-nhs-rightcare/ (accessed: 27 March 2018).
19. NHS England. NHS RightCare Pathway: Diabetes, 2017. Available at: https://www.england.nhs.uk/rightcare/wp-content/uploads/sites/40/2018/07/nhs-rightcare-pathway-diabetes.pdf (accessed: 27 March 2018).

15.15.2 Health Economics of Diabetes Care and Prevention

Philip Clarke and Thomas Lung

Introduction

It has been estimated that the cost of diagnosed diabetes in 2017 was $327 billion accounting for 1 in 4 healthcare dollars [1]. A recent review of studies globally indicates an even greater relative cost in low- and middle-income countries [2]. For the majority of countries, overall healthcare expenditure has been increasing over time. Estimates from the Organisation for Economic Co-operation and Development (OECD) indicate that in 2019 the United States had the highest health expenditure amongst the developed nations of the world, at 17.0% of GDP. This has increased from 12.5% in the year 2000. Over the same period, health expenditure in the United Kingdom was 9.8% of GDP in 2019, up from 7.3% in 2000.

Like other types of healthcare, costs associated with diabetes has been rising over time. In the United States, the population diagnosed with diabetes has grown by approximately 700 000 people annually between 2012 and 2017 (24.7 million people in 2017) and medical costs associated with diabetes has increased by 26% (from $188 billion to $237.3 billion) during that same period [3]. 36.3% of these medical costs are attributed to insulin and prescription medications.

Sustainability

Studies examining key drivers of changes in health expenditure over time attribute the rise in many countries to population ageing, increasing demand for more and better healthcare, rising incomes as well as technological change. If real expenditure (i.e. spending adjusted for inflation) is rising faster than allocated expenditure (i.e. receipts from taxation), the increase in health expenditure must be offset by either increasing the allocated expenditure through raising the level of taxation, or reducing expenditure on other government programmes (e.g. education, defence). It is also important to efficiently allocate resources to ensure spending on healthcare represents value for money. Health economics is an important feature of healthcare decision-making to ensure sustainability of the health system.

Variations in Drug Prices and Its Impact on Sustainability

The changing demand for insulins in the United States is an example of how increasing health expenditure is impacted by technological change [4]. The introduction of newer medications has resulted in many individuals substituting analogue for human insulin, which is costlier. Figure 15.15.2.1 shows the cost of antihyperglycemic medications in the United States based on data from the Medicare Expenditure Panel Survey (MEPS) over time. The estimated cost of insulin per patient has increased from $231 in 2002 to $722 in 2012 (after adjusting for inflation), while the price of older oral therapies has decreased to $111 in 2012 (Figure 15.15.2.1). Given the substantial variation and increasing discrepancies in costs over time between these therapies, it is important to assess whether the more expensive therapies provide additional benefits to justify the additional costs (i.e. provide value for money).

Technological Change—An Ongoing Challenge for Health Systems

New technologies for the treatment of diabetes such as analogue insulin can provide additional benefits to patients but are often more expensive than the medical care they replace. The substantial costs in research and development incurred when developing new pharmaceuticals and medical devices means that higher prices are charged, particularly when patents protect new products.

Government budgetary constraints mean that expenditure on new drugs need to be assessed to determine if it represents value for money relative to existing expenditure and the communities' willingness to pay for more healthcare relative to other types of goods.

Economic evaluation is a tool that involves quantifying the healthcare costs and health outcomes of the new drug relative to a comparator drug. Normally, the comparator refers to the drug that is currently in use and will be replaced by the new drug. What matters most in an evaluation is the **incremental** (or **additional**) outcomes and costs between the new drug and its comparator, which is relevant for use in an economic evaluation [5].

The drug being evaluated and its comparator may incur positive and negative impacts on both outcomes and costs, highlighted in Table 15.15.2.1.

Positive outcomes associated with a new drug are reductions in the risk of morbidity (such as diabetes-related complications) and mortality and therefore an increase in patient's life expectancy, as well as their quality of life. Negative side effects can potentially increase the risk of adverse events, worsen quality of life, and elevate mortality risk.

When quantifying costs it is important to take account of the full cost of the intervention, so if a pharmaceutical drug was being evaluated, not only is it important to cost the drug itself, but also the cost of administration (e.g. the time taken by a practice nurse or diabetes educator when patients start insulin therapy). The use of

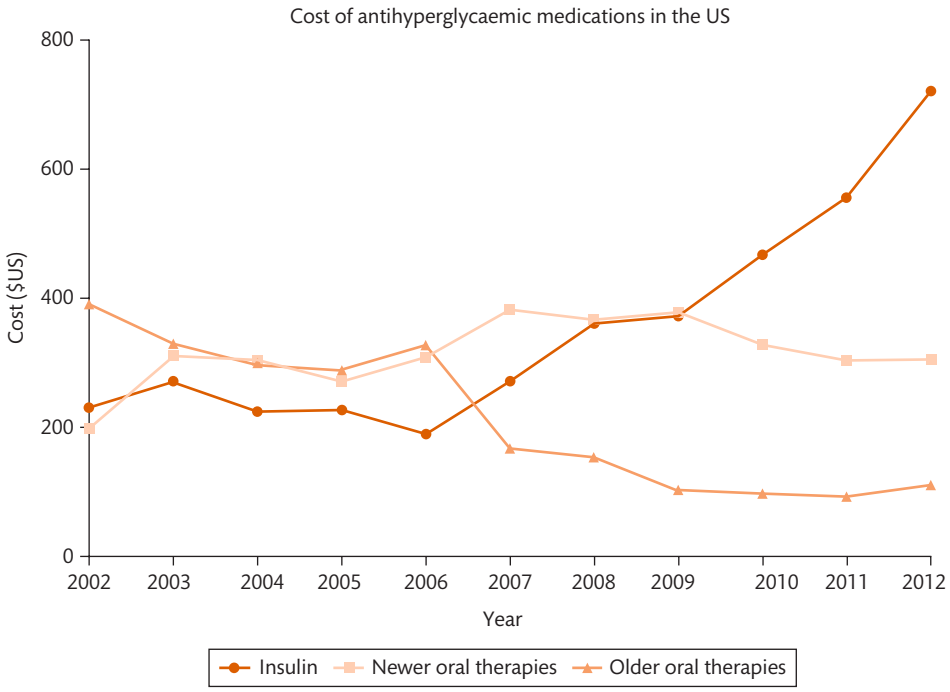


Figure 15.15.2.1 Cost of antihyperglycemic medications in the United States, 2002–2012.

Insulin = human and analogue. Newer oral therapies = thiazolidinediones, dipeptidyl peptidase-4 inhibitors, and combinations. Older oral therapies = metformin, sulfonylureas, α-galactosidase inhibitors, and non-sulfonylurea secretagogues.

some types of healthcare may have an impact on downstream costs. For example, drugs that prevent diabetes complications may reduce rates of hospitalizations and other medical services. Reductions in healthcare costs are known as **cost savings** and represents a monetary gain to society.

Economic evaluation essentially compares the costs and benefits of the intervention in question (new or existing) to an alternative intervention; it is dependent on the availability of information on the costs and effectiveness associated with an intervention.

Economic evaluation involves the calculation of a ratio of incremental costs to incremental outcomes, known as the **incremental cost-effectiveness ratio** (ICER) using the formula:

$$\text{Incremental cost} - \text{effectiveness ratio} = \frac{\text{Cost of intervention A} - \text{Cost of intervention B}}{\text{Outcomes of intervention A} - \text{Outcomes of intervention B}}$$

Table 15.15.2.1 Outcomes and costs to be considered when assessing a new diabetes intervention

Costs • Drug price and duration of use; • Cost of administration/prescription (e.g. visits to doctor) • Cost of healthcare utilization (hospitalizations, medical services, use of other drugs)	**Positive outcomes (Benefits)** • Increases in life expectancy • Improvements in quality of life • Reduced hospitalization events
Savings • Reduced healthcare utilization due to the drug	**Negative outcomes (Side effects)** • Adverse events including elevated risk of mortality • Adverse impacts on patient reported outcomes including quality of life

The ICER is a measure of the additional cost per additional unit of health outcome produced by an intervention compared with another.

There are three main types of economic evaluation: **cost-benefit analysis**, **cost-effectiveness analysis, and cost-utility analysis**. These differ only in the metric used to qualify outcomes. Cost-benefit analysis values outcomes in monetary terms, which is often challenging due to the subjectivity of putting a dollar value on the health of individuals. Cost-effectiveness analysis values outcomes in the same unit of health gain. Examples of health outcomes are life expectancy, cardiovascular disease event prevented, and hospital admission averted or life-years gained. A limitation of cost-effectiveness analysis is its reliance on a single, specific measure of outcome. Cost-utility analysis enables comparison of interventions of different health outcomes of interest because it accounts for both length and quality of life using an outcome known as **quality-adjusted life-years** (QALYs).

QALYs are a measure of health outcomes in which life expectancy, in terms of life-years, is weighted by an index of quality of life and measured on a scale in which 1 represents full health and zero represents health states equivalent to death. For example, if an intervention that results in a 10-year gain in life expectancy but the quality of life for each of those years was valued at 0.5, then the gain over a 10-year period is calculated to be 5 QALYs.

Health Technology Assessments

Economic evaluation is a tool used in health technology assessment (HTA) to help inform decision-making and is one of the criteria for allocating public funds. In many countries, HTA research is undertaken to evaluate any intervention used in the treatment,

prevention, or diagnosis of disease that has the potential to directly benefit patients. Normally, some evidence of efficacy already exists for an intervention. However evidence of cost-effectiveness has yet to be determined. There are a wide range of interventions for diabetes management, from treatment-specific interventions (pharmaceutical and medical devices) that lower blood glucose, improve metabolic control, to prevention, health promotion, and broader community interventions.

Economic evaluations of early intensive blood glucose-lowering trials, notably the DCCT and UKPDS, suggested that such policies are cost-effective. For example, an analysis of UKPDS intensive blood glucose control based on policies of using insulin and sulphonylureas reported an ICER of around £6000 per QALY, while blood glucose control with metformin was cost-saving [6]. Similarly, large clinical trials of better control of blood pressure [7] and lipid lowering [8] show that these strategies are also cost-effective and (in some cases) cost-saving. Lifestyle management interventions such as the Diabetes Prevention Program (DPP) have also been found to be cost-effective (discussed later) [9].

It is important to note that the cost-effectiveness of interventions may vary over time or across different healthcare settings, particularly if the cost of therapies change (see Figure 15.15.2.1). In such an environment, decision makers need to consider whether different strategies for prevention and treatment of diabetes are considered value for money and hence should be funded. Economists use the calculated ICER to implicitly place a value on extending and improving the quality of life. This is done by identifying a benchmark or threshold to determine whether an intervention is worth funding relative to outcomes.

What threshold should be used in the US? While there is no universally accepted amount, many studies often quote around $US50 000 per QALY. The origins of this threshold come from an American study in the 1970s that estimated the funding of dialysis patients with end-stage renal disease, the argument being that if there was public funding provided for dialysis then public funding should also be provided for interventions that had lower cost-effectiveness ratios [10]. In the United Kingdom, the National Institute for Health and Clinical Excellence (NICE) has an explicit threshold of £30 000 per QALY [11] although some have recently questioned this level as analyses of actual funding decisions indicates lower levels are currently being spent on care in the United Kingdom.

Use of Simulation Modelling

While there are now many randomized controlled trials of therapies for people with diabetes, economic evaluations just based on the evidence from such studies do not capture all the costs and benefits associated with these interventions due to the limited timeframe of the trial. Figure 15.15.2.2 illustrates the lifetime outcomes (measured in QALYs) of intensive blood glucose therapy based on evidence from the UKPDS [6, 12]. Although the UKPDS was a study of comparatively long duration having a median follow-up of over 10 years, the majority of the incremental outcomes (and costs) of intensive therapy over conventional therapy occurred after completion of the trial.

Quality-Adjusted Life Year in the US

Simulation modelling can be used to extrapolate beyond the results of a clinical trial to capture the costs and health outcomes accumulated over a lifetime associated with an intervention. Models such as

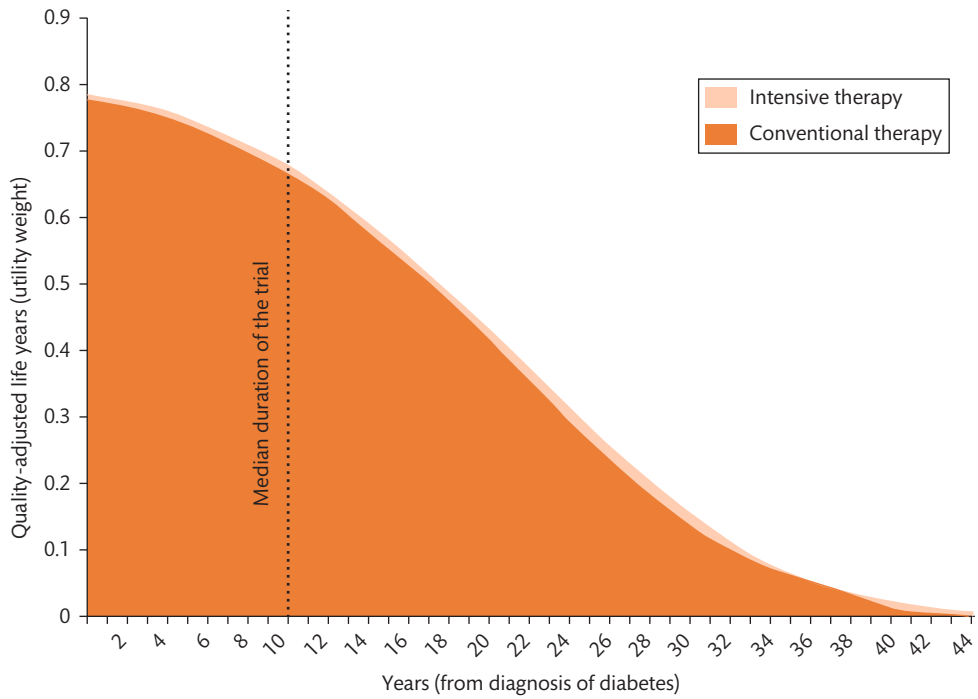


Figure 15.15.2.2 Average quality-adjusted life year estimates per year following the diagnosis of diabetes for individuals randomized to the intensive blood glucose and conventional therapy group of the United Kingdom Prospective Diabetes Study (UKPDS).
Reproduced with permission from AM Gray, PM Clarke, JL Wolstenholme, S Wordsworth. *Applied methods of cost-effectiveness analysis in healthcare*, Oxford University Press (2010).

the UKPDS Outcomes model have been used in type 2 diabetes to explore head-to-head comparisons of new interventions to an existing treatment, developed using survival analysis techniques on data from the trial to estimate lifetime health outcomes and costs for type 2 diabetes patients [13, 14]. This has been used worldwide to investigate the cost-effectiveness of interventions that could potentially lower the risk and healthcare costs associated with diabetes-related complications, as well as prevent premature mortality.

Case Study—Economic Evaluation Comparing a Lifestyle Intervention Versus Metformin for the Prevention of Diabetes

People with impaired glucose tolerance (IGT) are at high risk of progressing to type 2 diabetes. The DPP was a large randomized controlled clinical trial that tested whether either lifestyle (diet, exercise, and behaviour modification) changes or metformin could prevent or delay the onset of type 2 diabetes in over 3000 people with IGT [15]. Over a 3-year period, individuals reduced their risk of developing diabetes by 58% in the lifestyle intervention group and 31% in the metformin group when compared to the placebo group. DPP participants were followed up for an additional 7 years to assess the long-term effects of the interventions on health and healthcare utilization.

An economic evaluation was conducted comparing the lifestyle and metformin interventions with the placebo group. Both intervention groups experienced savings due to lower medical care costs, with less doctor and hospital visits, and other types of medications compared to the placebo group. Over a 10-year period the average cost per person was calculated to be an additional $928 in the lifestyle group compared to placebo, while metformin was cheaper than the placebo group ($321 per person).

Outcomes were assessed in terms of the cumulative number of QALYs over a 10-year period. The total increase in QALYs was approximately 0.14 and 0.02 for the lifestyle and metformin groups, respectively. It is important to remember that both outcomes and costs in economic evaluation are represented as the average or mean change, and may vary considerably between individual patients. For example, blood-pressure medication may have a large impact on an individuals' survival if it prevents them having a stroke, but may be of no benefit on survival for other patients.

Using the information provided on the costs and outcomes of both interventions (see Table 15.15.2.2):

Then, using the formula to calculate an ICER for the lifestyle intervention group compared to the placebo group:

$$\textit{Incremental cost} - \textit{effectiveness ratio} = \frac{\$928}{0.14} = \$6629.$$

Table 15.15.2.2 Incremental costs and incremental QALYs of the lifestyle and metformin interventions, when compared to the placebo group over a 10-year period

	Incremental Costs	Incremental QALYs
Lifestyle intervention	$928	0.14
Metformin intervention	-$321	0.02

Note that no incremental cost-effectiveness ratio was calculated for the metformin group. That is because the metformin intervention is both **cheaper** and **more effective** compared to placebo, which is a cost-saving scenario and an easy decision to fund for policymakers. This is known as a **dominant** intervention.

This means it costs on average around $6600 for each additional QALY gained through the use of a lifestyle intervention. Is it worth society spending the extra money to ensure access to the lifestyle intervention for diabetes prevention in adults with IGT?

REFERENCES

1. Association AD. Economic costs of diabetes in the US in 2017. *Diabetes Care*, 2018; **41**: 917.
2. Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. *Pharmacoeconomics*, 2015; **33**: 811–31.
3. Association AD. Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting. *Diabetes Care*, 2007; **30**: 1638–46.
4. Hua X, Carvalho N, Tew M, Huang ES, Herman WH, Clarke P. Expenditures and prices of antihyperglycemic medications in the United States: 2002–2013. *JAMA*, 2016; **315**: 1400–2.
5. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press, 2015.
6. Clarke P, Gray A, Briggs A, Stevens R, Matthews D, Holman R. Cost-utility analyses of intensive blood glucose and tight blood pressure control in type 2 diabetes (UKPDS 72). *Diabetologia*, 2005; **48**: 868–77.
7. Glasziou PP, Clarke P, Alexander J, *et al.* Cost-effectiveness of lowering blood pressure with a fixed combination of perindopril and indapamide in type 2 diabetes mellitus: an ADVANCE trial-based analysis. *Med J Aust*, 2010; **193**: 320–4.
8. Collaborative HPS. Lifetime cost effectiveness of simvastatin in a range of risk groups and age groups derived from a randomised trial of 20 536 people. *BMJ*, 2006; **333**: 1145.
9. Raikou M, McGuire A, Colhoun H, *et al.* Cost-effectiveness of primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes: results from the Collaborative Atorvastatin Diabetes Study (CARDS). *Diabetologia*, 2007; **50**: 733.
10. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness—the curious resilience of the $50,000-per-QALY threshold. *N Engl J Med*, 2014; **371**: 796–7.
11. Claxton K, Martin S, Soares M, *et al.* Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*, 2015; **19**: 1–503.
12. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. *Applied Methods of Cost-Effectiveness Analysis in Healthcare*. Oxford: Oxford University Press; 2011.
13. Clarke P, Gray A, Briggs A, *et al.* A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia*, 2004; **47**: 1747–59.
14. Hayes A, Leal J, Gray A, Holman R, Clarke P. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia*, 2013; **56**: 1925–33.
15. Group DPPR. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*, 2002; **346**: 393–403.

Index

Tables, figures, and boxes are indicated by an italic *t*, *f*, and *b* following the page number.